

С. Д. ХАЙТУН

НАУКОМЕТРИЯ
СОСТОЯНИЕ
И ПЕРСПЕКТИВЫ

АКАДЕМИЯ НАУК СССР
ИНСТИТУТ ИСТОРИИ ЕСТЕСТВОЗНАНИЯ
И ТЕХНИКИ

С. Д. ХАЙТУН

**НАУКОМЕТРИЯ
СОСТОЯНИЕ
И ПЕРСПЕКТИВЫ**



ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА 1983

В монографии анализируются место и роль наукометрии в ряду науковедческих дисциплин, рассматривается проблема измерения науки, производится обзор отдельных наукометрических методов и полученных с их помощью результатов и наукометрических исследований в целом. Значительное внимание автор уделяет обоснованию основных наукометрических закономерностей, в том числе неадекватности научной деятельности. Наукометрические методы и результаты анализируются в общем контексте науковедческих исследований. Монография снабжена большим графическим и библиографическим материалом и может в определенной степени служить наукометрическим справочником. Книга предназначена исследователям науки, а также всем интересующимся науковедческой проблематикой.

Ответственный редактор
доктор физико-математических наук
Г. М. Идлис

Сергей Давыдович Хайтун
НАУКОМЕТРИЯ
состоянии и перспективы

Утверждено и печати
Институтом истории естествознания и техники АН СССР

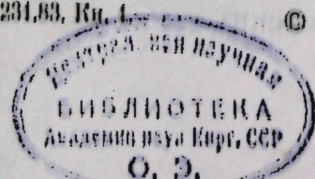
Редактор Л. И. Языкина. Художник С. В. Воробьев. Художественный редактор И. Ю. Истерова. Технические редакторы Р. М. Денисова, М. И. Комарова. Корректор Л. Н. Лукичева

ИБ № 22380

Сдано в набор 05.08.82. Подписано к печати 01.12.82. Т-20729. Формат 60×84¹/₁₆. Бумага типографская М 1. Гарнитура обыкновенная. Печать высокая. Усл. печ. л. 10,00. Усл. кр.-стр. 10,00. Уч.-изд. л. 23,4. Тираж 2000 экз. Тип. ван. 2003. Цена 3 р. 50 к.

Издательство «Наука» 117804, ГСП-7, Москва, Н-465, Профсоюзная ул., 90
2-я типография издательства «Наука» 121000, Москва, Г-90, Шубинский пер., 10

1403000000 015
X 042(02)-82



© Издательство «Наука», 1983 г.

544398

ВВЕДЕНИЕ

Наука превращается сегодня в непосредственную производственную силу общества. Реализация достижений науки обеспечивает в развитых странах до 3/4 прироста национального продукта. Необратимость процесса увеличения роли науки и создания общественного продукта закрепляется завершающимся в наши дни новым крупным общественным разделением труда, начало которого было зафиксировано К. Марксом: наука как наука, примененная к производству, отделяется от непосредственного труда [0.48, с. 118].

Наука, с которой все обстоит благополучно в стратегической перспективе, в наши дни, однако, переживает серьезные изменения: от стадии преимущественно экстенсивного роста она переходит к стадии преимущественно интенсивного роста. Преимущественно экстенсивный рост науки означает, что увеличение роли науки в развитии общественного производства реализуется в основном за счет ее объема. Число лиц, занятых в науке, нарастает по экспоненте, удваивается примерно каждые 10-15 лет, тогда как население в целом удваивается примерно за 35 лет. Темпы роста ассигнований на науку в большинстве промышленных стран также опережают темпы роста национального продукта.

Экспоненциальный рост науки, опережающий рост других областей общественного производства, не может продолжаться сколь угодно долго. Если бы нынешние темпы роста науки сохранились и в будущем, то рано или поздно наступил бы момент, когда все население Земли было бы занято научными исследованиями (см. рис. 1).

О невозможности продолжения процесса экстенсификации науки свидетельствует и удорожание научных исследований. Например, за последние пятьдесят лет удвоение новых научных результатов потребовало 15-кратного увеличения персонала науки и более чем 3-кратного увеличения ассигнований [0.48, с. 128].

Одним из отрицательных последствий продолжающегося преимущественно экстенсивного роста науки явились антисафентистские (от английского *science* — наука) течения в философской и общественной мысли в развитых капиталистических странах, получившие широкое распространение в середине 70-х годов нашего века.

Итак, в ближайшие десятилетия *экстенсивный рост науки должен замедлиться*. Снижение темпов роста ассигнований на науку и численности научного персонала может быть, однако, компенсировано. И компенсировано единственным образом — повышением *эффективности* науки, что означает переход к преимущественно интенсивному ее росту. Сегодня этот переход — проблема первостепенной важности, стоящая в одном ряду с экологической, энергетической, демографической проблемами. Человечество рефлектировало на создавшуюся в результате экстенсивного развития науки критическую ситуацию научными *исследованиями науки*, быстро развивающимися последние 20–25 лет.

Характеризуя эти исследования, С. Р. Микулинский отмечает «две довольно четко выраженные тенденции. В Советском Союзе и большинстве социалистических стран науковедение (под разными названиями) конституировалось в виде единой самостоятельной области *комплексного* изучения науки, которая существовала и развивалась в тесном взаимодействии с изучением различных аспектов науки... но не сводилось ни к одному из них...

Иначе обстояло дело в США и в большинстве западноевропейских стран, в том числе и в Англии, где, несмотря на энергичные усилия Дж. Бернала по созданию «науки о науке» как самостоятельной научной дисциплины, преобладало разрозненное изучение различных аспектов функционирования науки в рамках отдельных дисциплин. В США это были прежде всего информатика, эмпирическая социология и научный менеджмент. В Великобритании эти исследования велись под флагом социологии науки и социологии познания, отражающих определенную оппозицию позитивистской интерпретации науки. В ФРГ изучение науки опиралось на традиции в разработке «теории науки» с попытками ее историко-культурной интерпретации.

Если же обратиться к современному положению вещей, то можно констатировать, что эволюция этих двух путей исследований науки довольно явственно свидетельствует в пользу правильности первого, т. е. комплексного, подхода к науке и ее развитию. И хотя название исследовательской области, занимающейся изучением науки, в большинстве стран сохраняет на себе печать традиции, предмет исследований все больше приобретает комплексный характер» [0. 41, с. 41].

Таким образом, можно говорить об идущем сейчас процессе формирования особой комплексной науки — *науковедения*. Именно науковедение, использующее достижения социологии, экономики, психологии, философии и других дисциплин, объединяющее усилия ученых разных специальностей, более, чем какая-либо другая наука, призвано подвести научную базу под реорганизацию науч-

ной деятельности в направлении ее интенсификации. «Главной задачей науковедения, — говорит академик А. М. Румянцев, — является сейчас повышение эффективности исследовательской работы в условиях научно-технической революции. Эта задача приобретает особую актуальность в связи со стремительным ростом затрат на науку во всех развитых странах и столь же быстрым увеличением численности научных кадров» [0. 46, с. 15].

Процесс формирования науковедения развивается довольно успешно и здесь налицо серьезные сдвиги. «В настоящее время, — пишет С. Р. Микулинский, — когда определились основные направления науковедческого исследования, соответствующие его структурному расчленению как сложной комплексной области исследований, стало реальным оформление предмета различных науковедческих дисциплин (социологии, экономики, психологии науки и т. д.). Если еще недавно изучение науки социологами, экономистами, философами, психологами напоминало отдельные партизанские вылазки с разрозненных баз, то теперь началось формирование регулярных соединений, осознавших важность разработки единого плана операций и тесного взаимодействия» [0. 42, с. 47].

Можно, по-видимому, утверждать, что комплексный подход при исследовании науки реализован пока в плане синтеза *предмета* науковедения, *методические* же средства остаются в значительной степени разобщенными. Пути к созданию науковедения как самостоятельной дисциплины лежат через комплексное применение различных науковедческих методов.

Основная задача, которую ставил автор данной монографии, — *исследование того, что могут дать наукометрические методы для науковедения в целом в результате комплексного применения различных науковедческих методов*. Другими словами, здесь производится попытка «наведения мостов» между различными науковедческими дисциплинами с наукометрического «берега».

«Наведение мостов» между науковедческими дисциплинами — необходимый этап на пути к созданию науковедения как единой науки, выбор же именно наукометрии в качестве точки отсчета определяется профессиональными интересами автора.

Чтобы лучше уяснить задачу книги, обоснуем ее иначе. *Интенсификация* науки означает ее оптимизацию, *оптимизация* на определенном ее этапе предполагает число, а *число* предполагает, в свою очередь, *количественные методы* анализа науки. *Науковедение как научная база интенсификации науки должно опираться на количественные методы*.

Как считал К. Маркс, «наука только тогда достигает совершенства, когда ей удастся пользоваться математикой» [0. 3, с. 66]. И, по-видимому, любой объект поддается в той или иной степени мате-

математическому описанию. «Принципиально область применения математического метода, — говорит А. Н. Колмогоров, — не ограничена» [0.35, с. 464]. Отсюда, однако, не следует, что любой объект и любая сторона объекта поддаются какой угодно математизации. Как писал Ф. Энгельс, «если захочешь добиться математической достоверности в вещах, не допускающих этого, нельзя не впасть в пеленость или варварство» [0.1, т. 1, с. 638]. С. Р. Микулинский и Н. Н. Родный, процитировав известные слова Н. Винера о том, что приписывать «неопределенным по самой своей сути величинам какую-то особую точность бесполезно и нечестно и, каков бы ни был предмет, применение точных формул к этим слишком волюно определенным величинам есть не что иное, как обман и пустая трата времени» [0.13, с. 98], добавляют: «Короче говоря, в развитии науки имеются такие аспекты, которые или не могут быть предметом количественной обработки, или для последней требуются не конирование имеющихся методов, а разработка адекватных для этой цели способов» [0.43, с. 57].

Таким образом, науковедение в целях выполнения своей основной задачи интенсификации науки, с одной стороны, должно базироваться на количественных методах анализа; с другой — далеко не все стороны развития науки поддаются математическому описанию и потому науковедение должно использовать и методы качественного анализа. Иными словами, арсенал науковедческих средств должен содержать весь спектр методов — от точных до качественных и описательных.

Фактически так оно и есть на самом деле. Однако до сих пор отсутствует ясность относительно того, на что могут претендовать *количественные* методы, а на что они претендовать не могут. Неясно пока, чем различаются эти методы, в чем специфика каждого. Без такого анализа невозможно сколько-нибудь эффективное применение количественных методов.

Таким образом, мы пришли к другой формулировке задачи книги: *исследование места и роли наукометрии в ряду науковедческих дисциплин*. Другими словами, здесь производится попытка выявления круга науковедческих задач, подлежащих наукометрическому анализу, вне которого наукометрические методы не работают.

Чтобы исследовать место наукометрии в науковедении, необходимо, прежде всего, уточнить содержание понятий *науковедения* и *наукометрии*.

Начнем с понятия *науки*. «Долгое время наука рассматривалась только как система знаний, — пишет Б. М. Кедров и С. Р. Микулинский. — В настоящее время стало очевидно, что наука — это не только совокупность систематизированных знаний, но и специ-

фическая целостная система и особая форма деятельности, подчиняющаяся в своем развитии особым закономерностям» [0.46, с. 48].

Если предметом исследования таких наук, как, например, история науки, является наука как система знаний, то «науковедение... рассматривает науку как особую форму деятельности» [0.46, с. 48]. Конечно, историки науки привлекают для своего анализа деятельности аспекты науки, а науковеды — знаниевые (предметные), но здесь важны акценты. История науки учитывает деятельностные аспекты лишь постольку, поскольку это необходимо для изучения развития научного знания. Напротив, науковедение учитывает знаниевые аспекты, но лишь в той мере, в какой это необходимо для изучения научной деятельности.

Существуют различные определения наукометрии. Так, В. В. Налимов и З. М. Мульченко вводят такое определение: «Будем называть *наукометрией* (курсив наш. — С. Х.) количественные методы изучения науки как информационного процесса» [1.46, с. 12]. Б. Я. Брусилковский предлагает называть наукометрией «математически корректные методы изучения науки» [0.41, с. 6]. Таким образом, согласно этим определениям, всякое количественное исследование науки — наукометрическое. Сужению Налимовым, Мульченко определения наукометрии рамками информационного процесса не играет здесь существенной роли, поскольку информационные процессы в науке можно изучать и количественными методами, основанными на использовании данных анкетирования, экспертных оценок и т. д.

На наш взгляд, отнесение к наукометрическим всех количественных методов анализа науки, как это делают и некоторые другие авторы [см., например, 1.12; 1.13; 1.22], не совсем оправдано, поскольку в этом случае наукометрические методы анализа науки приравниваются к социологическим методам измерения, психометрическим и т. д. Кроме того, такое расплывчатое определение наукометрии не позволяет отделить ее от математического моделирования науки или от теории принятия решений, затрудняет постановку задач для наукометрических исследований.

По нашему мнению, определение наукометрии следует сузить и называть *наукометрией науковедческую дисциплину, осуществляющую воспроизводимое измерение научной деятельности и выявляющую ее объективные количественные закономерности*.

Заметим, что воспроизводимость результатов часто относят к числу критериев научного исследования. Особенно жесткие требования к воспроизводимости результатов предъявляются в естественных науках. В науках о человеке ситуация другая. По-видимому, и здесь идеалом является воспроизводимость результатов, однако в отличие от естественных наук науки о человеке далеко не

во всех своих методах ощутимо приблизились к этому идеалу. Существуют разные мнения относительно того, является ли слабая воспроизводимость результатов анализа наук о человеке следствием их относительной неразвитости, или же она порождена природой объектов анализа этих наук. Как бы то ни было, следует констатировать факт, что измерение объектов наук о человеке, в частности объектов науковедческого анализа, может быть сегодня воспроизводимым и может быть невозможным. Ответственность за воспроизводимое измерение научной деятельности мы и возлагаем на наукометрию.

К определению наукометрии можно подойти и через определение ее предмета. Будем называть наукометрией науковедческую дисциплину, имеющую своим предметом объективные количественные закономерности научной деятельности. В науках о человеке, как уже говорилось, не все методы обеспечивают хорошо воспроизводимое измерение. Потому и имеет смысл, например, в исследованиях науки выделить особую дисциплину — наукометрию, занимающуюся выявлением объективных количественных закономерностей. Поскольку такие закономерности могут быть выявлены лишь на базе воспроизводимого измерения, постольку определения наукометрии по ее предмету и как науковедческой дисциплины, осуществляющей воспроизводимое измерение, эквивалентны.

В литературе измерители, дающие хорошо воспроизводимые результаты, называют иногда *твердыми*, а дающие слабо воспроизводимые — *мягкими*. Точнее, следует говорить о спектре твердости измерителей параметров науки, который при определении наукометрических методов должен быть произвольным образом рассеян так, чтобы отсеченная «твердая» часть спектра и соответствовала наукометрическим методам. Мы производим это рассечение таким образом, что в числе наукометрических оказываются: статистический метод (измерители — число открытий, число журналов, число организаций, число премий, число ученых, частота соавторства и некоторые другие); метод подсчета числа публикаций (измеритель — число научных продуктов); цитат-индекс (измеритель — число ссылок); контент-анализ (измеритель — число так называемых символов); тезаурусный метод (измеритель — число терминов); сленговый метод (измеритель — число так называемых сленговых слов). Это — определение наукометрических методов. Продуктивно ли оно или? Может быть, следовало захватить более широкую или, напротив, более узкую часть спектра твердости измерителей параметров науки? Об этом можно судить лишь по результатам его последовательного применения. Пока у нас нет оснований сомневаться в данном определении, тем более что мы опираемся на существующие в наукометрической литературе тра-

диции. Так, Д. Прайс в своей классической книге «Малая наука, большая наука», В. В. Налимов и З. М. Мульченко в своей монографии (признающие, кстати говоря, более широкое определение наукометрии) и журнал «Scientometrics» фактически не выходят за пределы нашего определения.

Основы наукометрии, очерченной указанными выше рамками, были заложены Ф. Гальтоном [1.102], Д. Берналом [0.6—0.7], Г. Холтоном [1.108; 1.115], Д. Прайсом (см., например, [1.58—1.59]), В. В. Налимовым и З. М. Мульченко [1.46]. Отечественные наукометрические исследования систематически ведутся со второй половины 60-х годов XX в., а отдельные наукометрические работы появились у нас в начале века. Например, русский академик П. И. Вальден применял цитат-индекс в 1911 г. при анализе развития химии в России [4.1]. Советский физик Т. Райнов в 1929 г. провел сравнительный анализ временной динамики числа открытий по физике для ряда стран [2.56].

С 1979 г. в Венгрии выходит международный журнал «Scientometrics». В мировой литературе существует ряд обзорных и монографических исследований, в которых затрагиваются различные аспекты наукометрии. Приведем работы Д. Прайса [1.140; 1.58], Г. М. Доброва [0.22], Д. Крейн [1.90], Дж. Коула и С. Коула [1.88; 1.48], Ю. Гарфилда [4.78], Г. Цукерман [2.76], А. И. Яблонского [0.65], [0.112], Ю. В. Грановского [1.15]. Однако, несмотря на все значение этих работ для развития наукометрии, ни одна из них не является обобщающей наукометрической монографией. Такой монографией является пока лишь книга В. В. Налимова и З. М. Мульченко [1.46]. В ней впервые в мировой литературе выделен предмет исследований наукометрии как относительно самостоятельной науковедческой дисциплины и сжато резюмировано состояние наукометрических исследований на 1969 г.

Основная особенность монографии Налимова и Мульченко заключается в изолированности в ней наукометрических методов анализа от всех других науковедческих количественных и качественных методов. Эта ее особенность проявляется, прежде всего, в определении науки как самоорганизующейся системы. «Наука — это самоорганизующаяся система, — пишут Налимов и Мульченко, — развитие которой управляется ее информационными потоками. Внешние условия — ассигнования, отпускаемые на развитие науки, организационные формы, сложившиеся в той или иной стране, идеологическое давление, секретность — все это лишь элементы той среды, в которой развивается наука. Среда может быть благоприятной или неблагоприятной для развития науки, но она не в силах заставить науку развиваться в каком-либо органически чуждом ей направлении» [1.46, с. 11].

Модель науки как самоорганизующейся системы действительно предполагает изолированность наукометрических методов от всех остальных. В самом деле, экономические, социальные и многие другие факторы невозможно рассматривать, оставаясь в пределах самой науки, их можно учитывать, только беря науку как подсистему общей социальной системы. В противном случае к анализу, помимо наукометрических, Налимову и Мульченко пришлось бы подключать еще и эконометрические, психометрические и другие методы, что противоречило бы нацеленности авторов на выделение наукометрических методов.

Прежде чем объединить наукометрические методы в конкретных науковедческих исследованиях с остальными науковедческими методами, их следует выделить из остальных. Именно эта задача выделения наукометрических методов, придания им самостоятельного статуса и стояла в 1969 г. перед Налимовым и Мульченко.

Сейчас, спустя десятилетие, стоит задача объединения наукометрических методов с другими науковедческими методами. В этом случае модель науки как самоорганизующейся системы является уже тормозом, так как изолированные наукометрические исследования начинают себя уже исчерпывать, нуждаются в их объединении с другими.

Можно констатировать, таким образом, что сегодня, когда в значительной степени благодаря В. В. Налимову и З. М. Мульченко наукометрия выделена как самостоятельная науковедческая дисциплина, созрели необходимые предпосылки для решения стоящих перед автором данной монографии задач по выявлению возможностей комплексного применения наукометрии и других науковедческих дисциплин.

Содержание и структура книги определяются ее задачами. В значительной своей части она представляет собой обзор наукометрической литературы. Материал структурирован не по решаемым наукометрией науковедческим проблемам, а по методам и измерителям. Этот обзор не является исчерпывающим, так как мы не ставили перед собой задачу дать детальный отчет о всей быстро увеличивающейся наукометрической литературе, но постарались лишь уловить в этом потоке все сколько-нибудь существенные для дальнейшего развития наукометрии течения, все сколько-нибудь важные тенденции и идеи.

Глава I посвящена проблеме измерения при изучении научной деятельности. Здесь на базе представлений об индикаторах (измерителях) и латентных (скрытых) переменных, которые эти индикаторы косвенно характеризуют, формируется представление об опорных исследованиях по установлению соответствия индикаторов латентным переменным. Здесь же обсуждаются особенности

измерения при исследовании научной деятельности. На результатах анализа этих особенностей в определенном смысле строится вся книга. Сам же этот анализ базируется на результате, полученном в итоге обобщения наукометрических эмпирических данных. Именно на тезисе о негауссовости научной деятельности и являющемся его обобщением тезисе о негауссовости социальных явлений. Эти тезисы формулируются в гл. 7 и являются одним из основных выводов всей книги в целом.

Главы 2–6 посвящены отдельным наукометрическим методам. Рубрикация этих глав выдержана в одном ключе. Сначала излагается содержание метода и освещаются связанные с ним опорные исследования, затем приводятся полученные с помощью данного метода результаты в последовательности: нестационарные распределения (распределения, содержащие время), стационарные распределения (не содержащие времени) и структурные исследования. Завершает каждую из этих глав анализ перспектив развития и применения данного метода.

В гл. 7 рассматриваются основные наукометрические закономерности. Последовательно обсуждаются закономерности, которые позволяют выявить анализ нестационарных и стационарных распределений. Первые дают прайсовскую концепцию экспоненциально-логистического роста науки, вторые — концепцию негауссовости научной деятельности.

В заключении суммируются выводы относительно места и роли наукометрии в науковедении в настоящем и в перспективе.

Книга снабжена большим графическим материалом, который подразделен на части, соответствующие основным направлениям наукометрических исследований: нестационарные распределения, стационарные распределения и структурные исследования. Особый, четвертый раздел графиков посвящен комплексному применению наукометрических методов. Разделы, посвященные нестационарным и стационарным распределениям, построены так, чтобы показать действующие здесь закономерности — закон экспоненциально-логистического роста и негауссовость научной деятельности. Вместе с тем мы старались по возможности, чтобы эти графики имели и чисто познавательный интерес.

Библиографию в соответствии с общей направленностью книги мы предпочли структурировать. В основу положены методологические, а не предметные признаки.

В заключение следует отметить, что § 2 гл. 6 написан в соавторстве с Ю. М. Низовцевым.

Работа выполнена в секторе проблем научного творчества Института истории естествознания и техники АН СССР. При обсуждении концепции негауссовости научной деятельности на ран-

ней стадии работы ценными были критика и информация А. И. Яблонского и А. А. Игнатова. Большое значение для автора на разных стадиях работы имели замечания К. Г. Борескова, Вл. П. Визгина, Б. В. Гнеденко, Л. Б. Окуня, В. И. Рогова, Б. А. Розенфельда, Ю. И. Тюриня. Интересным было обсуждение этой концепции в секторах истории физики и математики Института истории естествознания и техники АН СССР.

Обсуждение с И. С. Алексеевым, А. В. Ахутиным, А. П. Огурцовым вопросов измерения, затронутых в §§ 1 и 2 гл. 1, прояснило некоторые существенные моменты. Э. П. Андреев, В. С. Библер, Р. С. Гиляревский, Ю. В. Грановский, Г. М. Идлис, А. И. Яблонский высказали ряд важных соображений по всей рукописи.

В. С. Кухарчуку, В. Л. Рабиновичу и всем, кто в той или иной форме оказывал содействие, автор выражает свою глубокую признательность.

Глава первая

НАУКОМЕТРИЯ И ПРОБЛЕМА ИЗМЕРЕНИЯ: НЕГАУССОВОСТЬ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ

Проблема измерения при изучении научной деятельности — центральная проблема наукометрии. Эта проблема актуальна и для других науковедческих дисциплин, но в области наукометрии метрологические вопросы стоят особенно остро. При изучении науки вообще трудно измерять, однако особенно трудно измерять *воспроизводимо*. А именно на такое измерение претендует наукометрия.

Без решения проблемы измерения, без ответа на вопросы «что измерять?» и «как измерять?» невозможно очертить круг подведомственных наукометрии исследовательских задач, т. е. нельзя очертить *предмет* наукометрических исследований, а следовательно, невозможно решить задачу по выявлению места наукометрии в науковедении.

Измерение в исследованиях науки имеет свою специфику: во-первых, как измерение, отличное от измерения в естественных науках; во-вторых, как измерение, отличное от измерения в исследованиях других социальных объектов.

Приведем несколько примеров. Б. Ф. Скиннер заложил в 30-х годах основы теории поведения животных. Однако его обоснованно критиковали за использование в экспериментах слишком малого количества животных. Вняв критике, Скиннер увеличил число особей в каждой серии с 4 до 24. Ничего хорошего из этого не вышло, данные «поплыли». В конце концов он решил «уйти от механической статистики и сосредоточиться на изучении единичных случаев» [0.108].

Американский психолог Г. А. Дэвис с коллегами искали корреляцию между творческими способностями и . . . верой в сверхчувственное восприятие. Странная постановка вопроса, не правда ли? Тем не менее искомая корреляция, хоть и небольшая, была найдена [0.77].

Что это? Следствие некомпетентности исследователей? Но этого нельзя сказать ни о Скиннере, ни о Дэвисе с коллегами, ни о боль-

министе других ученых, работающих в науках о человеке. Однако результаты, получаемые ими, часто противоречат друг другу, а иногда, как в случае с Дэвисом, и здравому смыслу.

Неустойчивость количественных результатов, получаемых в науках о человеке, породила пессимистические настроения относительно возможностей количественного анализа социальных явлений. На наш взгляд, продуктивнее не отрицать количественные методы анализа в применении к социальным явлениям, а постараться разобраться, в чем корни трудностей измерения при изучении социальных явлений.

1. ИЗМЕРЕНИЕ: ВОСПРОИЗВОДИМОСТЬ, АДЕКВАТНОСТЬ, ТРУДОЕМКОСТЬ

Основы теории измерения в применении к физике и другим естественным наукам заложили в конце XIX — начале XX в. Г. Гельмгольц, О. Гёльдер, П. Р. Кемпбелл. Теорию социального измерения разрабатывают с начала 30-х годов С. Стивенс, К. Кумбс, П. Суппес и Дж. Зиннес, П. Ф. Лазарсфельд, И. Пфандагль и др. [см. 0.53; 1.86; 1.71; 0.41; 1.10; 1.35; 0.112; 0.54].

Измерение предполагает *сравнение* объектов в определенном отношении. Сравнение предполагает выделение в объектах некоторых *свойств*, по которым и производится сравнение. «Различные вещи становятся количественно сравнимыми, — пишет К. Маркс, — лишь после того, как они сведены к одному и тому же единству. Только как выражение одного и того же единства они являются одноименными и, следовательно, сравнимыми величинами» [0.1, т. 23, с. 58—59].

Мы расходимся с теми авторами (в основном психологами и социологами), которые говорят о возможности так называемого *уникального измерения*, не предполагающего *сравнения* объектов. На наш взгляд, такая ситуация оторвана от реальности. Когда, скажем, я измеряю размеры стола, чтобы пронести его через дверь, я сравниваю размеры стола с размерами двери. Когда я измеряю силу землетрясения, то имею в виду сравнение данного землетрясения и его разрушительных последствий с другими (в том числе и будущими) землетрясениями и их последствиями, чтобы научиться строить сейсмостойкие сооружения, и т. д. Другими словами, *результаты измерения всегда предназначены для их сравнения в процессе человеческой деятельности*.

В теории измерения выделяют четыре основные шкалы — *номинативную*, *порядковую* (ординальную), *интервальную* и *пропорциональную* (шкалу отношений). Номинативная шкала предполагает лишь различение объектов: объект А отличается от об-

ъектов В, С и т. д. Порядковая шкала не только различает объекты, но и выстраивает их в ряд в отношении измеряемого свойства: объект А превосходит по этому свойству объект В, однако, насколько именно превосходит, не устанавливается. Интервальная шкала не только упорядочивает объекты, но и соотносит каждый из них с точки зрения измеряемого свойства с некоторым *числом*, определяемым с точностью до некоторого интервала. И, наконец, пропорциональная шкала не только упорядочивает объекты, но и соотносит каждый из них с некоторым *числом* «абсолютно точно».

Принято разделять *качественное* и *количественное* измерение. Качественное измерение соответствует номинативной и порядковой шкалам, количественное — интервальной и пропорциональной.

В *наукометрии*, как и в естественных науках, используется в основном количественное измерение. И этим она отличается, скажем, от психологии и социологии, где используется в основном качественное измерение.

Количественное измерение предполагает оценку выделяемых свойств объектов в *сопоставимых* числах, т. е. мы должны оценивать одно и то же свойство разных объектов в числах *единым образом*, пользуясь одной методикой. При этом определяется отношение измеряемой величины к другой величине, принимаемой за *единицу измерения*.

Измерения свойств разных объектов могут быть разнесены во времени и в пространстве, могут производиться разными исследователями. При этом могут использоваться разные процедуры измерения. Поскольку, однако, научное исследование всегда нацелено на выявление *объективных связей* между объектами (сравнение), т. е. связей, свободных от всех привходящих обстоятельств, постольку мы должны обеспечить *воспроизводимость* измерения. *Степень воспроизводимости результатов — основная характеристика измерения*.

Заметим, что установка на уникальное измерение предполагает установку на невоспроизводимость измерения. Именно необходимость сравнения различных результатов измерения в процессе человеческой деятельности (в частности, исследовательской) выдвигает требование воспроизводимости измерения.

Измерению предшествует *качественный анализ* объектов, в ходе которого, с учетом конечной *цели исследования* и наличных методологических и методических средств, выбираются для измерения свойства объектов и процедуры определения соответствующих им числовых значений.

Качественный анализ выступает на сцену и после того, как измерение произведено. Теперь он призван оценить степень аде-

кватности (соответствия) результатов измерения объектам измерения в соотносительности с целями исследования. Одно и то же измерение может быть адекватно относительно одной исследовательской задачи и неадекватно относительно другой.

Ясно, что адекватность измерения в большой степени зависит от выбора измеряемого свойства объекта. Данное измерение тем более адекватно целям исследования, чем в большей степени измеряемое свойство характеризует весь объект, «представляет» его в рамках данной исследовательской задачи. *Адекватность измерения — вторая основная его характеристика.*

При фиксированном уровне научных исследований повышение воспроизводимости результатов измерения часто бывает связано с понижением их адекватности, и наоборот. Поэтому, чтобы повысить одновременно и воспроизводимость, и адекватность измерения зачастую приходится переходить к новому, более высокому уровню исследования, разрабатывать принципиально новые методы измерения.

Третий основной критерий измерения — его трудоемкость. При фиксированном уровне научных исследований повышение адекватности измерения при сохранении данного уровня воспроизводимости связано, как правило, с увеличением трудоемкости измерения. Уменьшение трудоемкости измерения при фиксированных его адекватности и воспроизводимости связано с переходом к более высокому уровню исследований (с использованием новых приборов, ЭВМ и т. д.).

При сравнении различных методов измерения в целях выбора одного или нескольких из них для данного конкретного исследования приходится учитывать все три критерия измерения — воспроизводимость, адекватность, трудоемкость.

2. ЛАТЕНТНЫЕ ПЕРЕМЕННЫЕ И ИХ ИНДИКАТОРЫ

В теории измерения вводятся переменные двух типов: то, что мы *хотим измерить* — массу тела, научный вклад ученого и т. д., часто называют *латентными (скрытыми)* переменными (иначе — латентными параметрами, гипотетическими или промежуточными переменными, факторами, следами); то, что мы *непосредственно измеряем* — показания весов, число ссылок и т. д., часто называют *индикаторами* или *измерителями*. Индикаторы позволяют нам косвенно судить о латентных переменных.

В общем случае латентная переменная связана со своим индикатором (индикаторами) *вероятностным* образом. Каждому значению индикатора соответствует вероятностное распределение возможных значений латентной переменной, и наоборот. Для

индикаторов, имеющих лишь два возможных значения (например, в социологии, когда респонденту задают дихотомические вопросы), такое распределение иногда, следуя П. Лазарсфельду [0.41, с. 362—412], называют графиком (trace line) индикатора, а иногда — его *операциональной характеристикой*, как это делает, например, Б. Ф. Грин (там же, с. 227—287). Мы будем использовать первый из этих терминов, сохраняя его и в том случае, если индикатор имеет непрерывный спектр значений, т. е. характеризуется бесконечным числом своих графиков.

Если для данного индикатора известны все его графики, то принципиально вопрос об измерении латентной переменной решен. Именно такие графики позволяют сказать, какова вероятность того, что объект с измеренным значением индикатора имеет значение латентной переменной в любом заданном интервале. Увеличение числа индикаторов позволяет в этом случае повысить точность измерения.

К сожалению, способов прямого, непосредственного установления графиков индикаторов не существует и существовать не может. Чтобы понять, в чем тут дело, необходимо учесть некоторые особенности взаимодействия объекта и субъекта познания. Объект в самом факте своего существования в любой из своих конкретных определенностей, в любой из своих «срезов» независим от познающего субъекта. Но выбор «среза» объекта, выбор конкретной определенности из всего «веера» конкретных определенностей, которую принимает во внимание познающий субъект, детерминирован конкретной определенностью познающего субъекта. Поэтому конкретная определенность объекта познания *как она предстает перед познающим субъектом* зависит от субъекта.

Латентная переменная — это та конкретная определенность объекта исследования, которую познающий субъект, руководствуясь теоретической концепцией и целями исследования, выбирает, фиксирует для своего анализа. И фиксация эта происходит в сознании познающего субъекта. Таким образом, латентная переменная, отражающая реальное свойство реального объекта, непосредственно тем не менее не наблюдаема.

Познающий субъект использует для описания данного свойства объекта исследования переменные разной природы. Латентная переменная — это *представление о свойстве*, и потому непосредственно измерена быть не может. Индикатор же непосредственно измерим, это — *наблюдаемая, эмпирическая переменная*. Латентные переменные и индикаторы находятся, образно говоря, в разных плоскостях используемого исследователем аппарата, непосредственно не пересекаются. Тогда каким же образом соотносить индикаторы и латентные переменные?

Теория социального измерения знает различные *метрические модели*, соотносящие значения индикаторов со значением латентных переменных: шкала А. Бине; сигма-шкала; Т-шкала; модели Л. Терстоуна, М. Сэффира, Р. Ликерта, Л. Гуттмана; модель факторного анализа (линейная); модель П. Ф. Лазарсфельда (алгебраический многочлен) и др. Однако, поскольку до сих пор отсутствует критерий объективной оценки достоверности той или другой модели, они остаются в известной степени схоластическими. Проблема соотнесения индикаторов и латентных переменных остается центральной в социальных науках, в частности в наукометрии. Далее излагаются некоторые соображения относительно возможных путей решений этой проблемы.

Соответствие между латентными переменными и индикаторами устанавливается в ходе рефлексивного осознания человеком своей деятельности, ее целей, средств, результатов. На каждой стадии процесса познания субъект отождествляет с латентными переменными некоторые теоретические конструкции, образованные им из наблюдаемых переменных, т. е. из индикаторов, постепенно корректируя эти теоретические конструкции под влиянием практики.

Теоретические конструкции, отождествляемые на очередном этапе познания с латентными переменными, могут быть определены через указание процедуры их измерения.

Операциональным определением латентной переменной может служить тесно связанная группа взвешенных индикаторов. Отождествление латентной переменной с ее операциональным референтом выражается в том, что мы устанавливаем графики индикаторов относительно референта и отождествляем их с графиками индикаторов относительно исследуемой латентной переменной.

Переменную, конструируемую указанным образом из индикаторов, будем называть операциональной латентной переменной.

В тривиальном случае (нулевое приближение) за операциональную латентную переменную принимается один из индикаторов. Скажем, нас интересует латентная переменная «научный вклад ученого». В нулевом приближении в качестве операционального определения этой латентной переменной может быть принят индикатор «число ссылок на ученого». Точнее говоря, ранжировка ученых по числу ссылок на них отождествляется нами в рамках нашего нулевого приближения с их ранжировкой по указанной латентной переменной. Ранжировка ученых на базе другого, скажем социологического, индикатора может служить другим операциональным определением данной латентной переменной.

Конечно, операциональное определение латентной переменной по одному индикатору малопродуктивно. Если мы хотим иметь более точное описание действительности, мы должны определять операциональные латентные переменные с помощью группы тесно связанных между собой индикаторов. Эти группы индикаторов выявляются, например, факторным анализом. Данную систему индикаторов можно согласовать определенным образом — так, чтобы ранжировка данной группы объектов обеспечивала максимальную корреляцию с ранжировками по каждому из индикаторов в среднем по всей системе индикаторных распределений (ранжировок) при фиксированном способе определения этого среднего. Например, сначала мы ранжировали 1000 ученых по числу публикаций, затем — по ученым степеням и наградам, затем — по числу ссылок и т. д. Заметим, что порядок расположения ученых во всех этих ранжировках будет разный. После этого мы находим оптимизирующую ранжировку, имеющую минимальное в среднем расхождение с отдельными индикаторными ранжировками. При определении средней корреляции оптимизирующей ранжировки с отдельными индикаторными ранжировками по всей группе этих ранжировок в целом отдельные индикаторы (индикаторные распределения) фиксируются с определенными весами. Выбранная нами система индикаторов вкупе со способом определения средней корреляции разных индикаторных ранжировок и будет служить определением данной операциональной латентной переменной. Полученное же на выходе оптимизирующее распределение ученых будет служить эталонным распределением ученых по этой операциональной латентной переменной, относительно которого мы и будем определять графики отдельных индикаторов.

Таким образом, проблема измерения выступает как проблема разработки системы операциональных латентных переменных, привязки отдельных индикаторов к этой системе, ее перманентной проверке на «жизнеспособность» и доводке в процессе включения результатов исследования в человеческую практику. Другими словами, в процессе измерения мы должны: во-первых, исходя из некоторой концепции и целей исследования разработать систему операциональных латентных переменных, т. е. определить для каждой латентной переменной ее операционального референта путем фиксации соответствующим образом взвешенной системы индикаторов; во-вторых, соотнести систему индикаторов с этой системой операциональных латентных переменных, т. е. определить для каждого индикатора графики для каждой из операциональных латентных переменных; в-третьих, постоянно соотносить результаты измерений с получаемым на

их основе практическим эффектом и производить на базе этого соотношения постоянную «доводку» системы операциональных переменных в сторону увеличения этого эффекта.

Итак, количественные исследования социальных явлений будут иметь смысл только в том случае, если будут базироваться на специальных *опорных исследованиях* по разработке систем индикаторов и операциональных латентных переменных в их соотносительности. Проблема измерения в силу ее сложности и комплексности должна решаться в социальных исследованиях централизованно. Только на базе опорных исследований результаты количественных исследований социальных явлений можно будет на научной основе соотносить с объектами измерений. В противном случае измерение оказывается построенным на песке и невозможно знать, в какой степени его результаты относятся к объектам измерения.

Следует заметить, что опорные исследования фактически в социальных исследованиях давно уже ведутся, хотя и не систематически. Определяются не графики индикаторов относительно операциональных латентных переменных, а в основном лишь корреляции отдельных индикаторных распределений друг с другом. Такие корреляции — лишь плацдарм для структурных исследований индикаторов, которые только и позволяют средствами, скажем, факторного анализа выявлять тесно связанные группы индикаторов и соотносить затем с ними те или другие латентные переменные, т. е. вводить собственно операциональные латентные переменные.

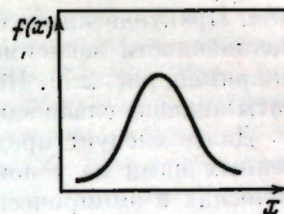
3. НЕГАУССОВСТЬ СОЦИАЛЬНЫХ ЯВЛЕНИЙ

(Предварительные соображения)

Начиная со второй половины XX в. в науке зародились и получили широкое распространение вероятностные представления, являющиеся обобщением детерминистических. Мы не будем излагать курс теории вероятностей и математической статистики, предполагая знакомство читателя, хотя бы в общих чертах, с этим курсом. Сделаем только несколько замечаний, существенных для дальнейшего.

В общем случае в силу специфической природы исследуемых объектов или в силу недостаточности нашего знания о них количественные величины, которыми мы описываем эти объекты и которые называются в теории вероятностей *случайными*, характеризуются вероятностными распределениями, говорящими о том, какова вероятность события, соответствующего данному значению случайной величины x .

Вероятностные распределения характеризуются набором параметров (характеристик), построенных на базе функции распределения $F(x)$ или плотности вероятности $f(x)$.



О статистических закономерностях, действующих в данной генеральной совокупности исследуемых объектов, мы судим по выборкам объектов, извлекаемым из этой генеральной совокупности. Эти выборки могут иметь разные объемы, так как производятся разными исследователями и часто бывают разнесены во времени и в пространстве. Мы же хотим, естественно, выявлять объективные связи между объектами, свободными по возможности от этого привходящего обстоятельства. Другими словами, мы хотим, чтобы получаемые результаты были максимально воспроизводимы и, в частности, по возможности, независимы от объема выборки.

Это требование учтено в математической статистике, применяемой в естественных науках. Она базируется на *моментах* статистического распределения

$$M^{(n)} = \int (x - X)^n f(x) dx.$$

Здесь $M^{(n)}$ — момент n -го порядка относительно значения случайной величины $x = X$. Момент первого порядка относительно $X = 0$ — это среднее $\bar{x}$. Момент второго порядка относительно $X = \bar{x}$ — это дисперсия σ^2 и т. д.

Аппарат моментов связан с *распределением Гаусса* (нормальным распределением). Эта связь вытекает из центральной предельной теоремы математической статистики, которая говорит о том, какую форму имеет распределение выборочного среднего $\bar{x}$. Иными словами, ставится следующий вопрос: если мы сделаем некоторое число разных выборок из генеральной совокупности и для каждой такой выборки рассчитаем среднее $\bar{x}$, то как будут распределены эти средние на множестве наших выборок? Оказывается, что если для распределения $f(x)$ существуют конечные среднее и дисперсия, то с ростом объема выборки распределение выборочного среднего приобретает форму распределения Гаусса.

Распределения, для которых выполняется центральная предельная теорема, будем называть *гауссовыми*. Для гауссовых распределений, таким образом, моменты (точнее — первые два момента) при устремлении объема выборки к бесконечности устремляются к конечным пределам.

Выборка бесконечного объема — это математическая абстрак-

т. е. быть полезным с точки зрения включения этого феномена в человеческую практику. В данном случае объяснение особенностей измерения при изучении социальных явлений должно указывать, каким должен быть количественный анализ социальных явлений в отличие от количественного анализа естественных. В настоящем параграфе мы попытаемся изложить особенности измерения при изучении социальных явлений, дать им объяснение в свете негауссовости этих явлений и сформулировать следствия из этого объяснения для количественного анализа в социальных науках.

Первая особенность: большая, чем в естественных науках многофакторность явлений

Эта особенность проявляется прежде всего в слабой связанности различных переменных, разных факторов, влияющих на объекты исследования. Здесь «все связано со всем», но связано слабо. Поведение социального объекта детерминировано большим числом факторов по сравнению с природным, но каждый из них влияет слабее.

Обратимся опять к примеру с Г. А. Дэвисом [0.77]. На своем массиве из 140 студентов он нашел корреляцию между творческими способностями, определяемыми по тесту Торренса, и верой в сверхчувственное восприятие. Нашел он и корреляцию теста Торренса с тестом «поиск сильных ощущений». Правда, и та и другая корреляции не слишком велики (0,5 и 0,65 соответственно).

Еще пример. Творческие способности ученого, как это явствует из общих соображений, должны влиять на его профессиональные характеристики. Однако Р. Е. Коннор [0.74] и Ф. Эндриус [0.69] не нашли корреляции между творческими способностями, определяемыми по тесту Медника на отдаленность ассоциирования, профессиональной компетентностью 64 ученых, занятых в групповых исследованиях (Коннор), и степенью новаторства научного вклада (Эндриус). Отсюда авторы делают выводы о том, что творческие способности ученого не влияют на его научный вклад. Заведомо это не так. Но корреляция здесь слабая и выявить ее можно лишь на достаточно большой выборке.

Другой пример. Казалось бы, творческие способности индивиду и его интеллект должны быть тесно связаны. Однако Д. Гилфорд [0.82; 0.83] нашел такую корреляцию весьма скромной.

Феномен многофакторности социальных явлений может быть объяснен их негауссовостью. Предварительно следует определенным образом формализовать понятие «фактор».

Любой объект испытывает различные воздействия, образующие в общем случае непрерывный спектр и определяющие в совокупности его поведение. Без ограничения общности рассмотре-

ния можно ввести переменную x , характеризующую эти воздействия в отличие друг от друга. Разность отдельных значений этой переменной $x_2 - x_1$ характеризует степень различия воздействий, соответствующих x_1 и x_2 .

Обозначим через $n(x)dx$ число воздействий на объект, описываемых значениями переменной x , лежащими в интервале $x, x + dx$, $n(x)$ — это распределение воздействий, определяющих поведение данного объекта по величине x , т. е. по степени различия воздействий.

Понятие «фактор», несомненно, *дискретное*: для описания поведения объекта исследования всегда используется конечное число «разных» факторов. Поэтому, чтобы от *воздействий*, непрерывно переходящих от одного к другому, перейти к *факторам*, следует от непрерывного распределения $n(x)$ перейти к дискретному. Поступим так, как часто поступают в подобных случаях. Разделим область изменения x на равные так называемые классовые интервалы Δx . Воздействия, находящиеся в одном таком классовом интервале, различать не будем, относим их к одному *фактору*, воздействующему на объект. Если максимальное значение x обозначить через J , то число факторов, действующих на объект, равно $J/\Delta x$. Распределение $n(x)$ имеет теперь смысл распределения факторов, определяющих поведение данного объекта, по их разнообразию, по их отличию друг от друга.

Представляется оправданным распространение концепции негауссовости социальных явлений и на эти (стационарные) распределения факторов по их разнообразию. Иными словами, будем считать, что эти распределения для человеческой деятельности негауссовы, а для природы — гауссовы. Но негауссовые распределения имеют более длинные «хвосты», для них J больше, чем для гауссовых. Поэтому и $J/\Delta x$, т. е. в нашем случае *число факторов для негауссовых (социальных) распределений больше, чем для гауссовых (природных)*.

Вследствие многофакторности социальных явлений и анализ их должен быть многофакторным. Конечно, анализ любого объекта — и природного и социального — должен быть многофакторным. Речь идет о том, что анализ социальных явлений должен быть существенно более многофакторным, чем анализ естественных.

Требование многофакторности анализа социальных явлений отнюдь не декларативно. Иллюстрацией жесткой необходимости многофакторного анализа социальных явлений могут служить, например, известные рассуждения П. Лазарсфельда. Излагая результаты социологического обследования американских солдат во время второй мировой войны, проведенного коллективом во главе

с С. А. Стоуффером, Лазарсфельд приводит несколько выводов Стоуффера и объясняет, почему они очевидны.

«1. Солдаты с более высоким уровнем образования проявляли больше психоневротических симптомов, чем их менее образованные товарищи. (Психическая нестабильность интеллектуала в сравнении с более инертной психологией «человека с улицы» часто является предметом обсуждения.)

2. Солдаты-выходцы из сельских местностей обычно находились в хорошем настроении чаще, чем солдаты-выходцы из городов (в конце концов первые более привычны к трудностям).

3. Солдаты-южане лучше переносили жаркий климат островов Южного моря, чем солдаты-северяне (естественно, ведь южане более привычны к жаркой погоде).

4. Рядовые-белые больше стремились стать унтер-офицерами, чем рядовые-негры (отсутствие у негров честолюбия вошло в поговорку)...

...[Возникает вопрос], почему для установления подобных данных тратится так много средств и энергии, ведь они столь очевидны? Не лучше ли принимать их без доказательства и сразу переходить к более углубленному уровню анализа? Возможно, это и было бы лучше, если бы не одно «но», касающееся приведенных примеров. Каждое из этих утверждений прямо противоположно тому, что было обнаружено в действительности... Если бы мы с самого начала привели подлинные результаты исследования, читатель и их бы назвал «очевидными». Очевидно, что-то не в порядке с самим доводом очевидности. Его следует поставить с головы на ноги. Поскольку всегда можно представить себе любой тип человеческого поведения, крайне необходимо знать, какие из них и при каких условиях проявляются чаще всего» [1.35, с. 147—148].

Понять этот пример Лазарсфельда можно следующим образом. Поведение социального объекта детерминировано множеством факторов (многофакторность социальных явлений!). Если нет никаких количественных данных о тех или иных факторах, определяющих поведение объекта анализа, субъект анализа выбирает из этого множества факторов тот, который к нему «ближе лежит». Но, если такие количественные данные есть, субъект анализа, не испытывая никаких сомнений в их достоверности (к этому нас приучили естественные науки), неосознанно приспособливает свою точку зрения на явление к этим данным, выбирает из множества по-разному влияющих на объект факторов тот, который обеспечивает согласие с числами. Так рождается миф об «очевидности» результатов количественного анализа социальных объектов. В естественных науках этого не происходит потому, что здесь число факторов, детерминирующих поведение объекта, относительно не-

велико (гауссовость природы) и субъект исследования не располагает большой свободой при своем их выборе.

Трудность описания социальных явлений как раз и состоит в том, что здесь ни один латентный параметр явно не доминирует. Здесь все связи «размыты». В условиях такой связанности «всего со всем» ученые, количественно анализирующие социальные явления, приходят к выводу о необходимости многофакторного анализа. Д. Гилфорд, например, пишет: «Поскольку творческий талант состоит из многих компонент, нет и не может быть одного исчерпывающего надежного теста. Требуется «множественный» подход» [0.82]. Ш. Эдвардс и М. У. Маккэри в своем обзоре методов оценки труда ученого [1.96] приходят к выводу о том, что многоаспектность труда ученого отвергает применение любого одномерного критерия, о необходимости многокатегориального оценивания. Г. М. Добров и соавт. в связи с обсуждением кривых роста науки говорят об ограниченности отдельных измерителей доз науки [0.27, с. 55].

Исследователи социальных явлений не только говорят о необходимости многофакторного, многомерного анализа, но и направляют в эту сторону свои усилия, не дожидаясь помощи профессиональных математиков. Не случайно первый математический аппарат такого рода — факторный анализ — родился именно в науках о человеке: его автором был психолог С. Спирмен [0.109]. Известны также латентно-структурный анализ П. Лазарсфельда, многомерное шкалирование и другие методы многомерного анализа.

Практически исследователь имеет дело с системой индикаторов социальных явлений. При изучении данного объекта необходимо выявить группу существенных для определения поведения этого объекта операциональных латентных переменных-факторов. Математические средства многомерного анализа и нацелены на эту задачу.

В § 2 мы говорили, что анализ социальных явлений должен основываться на *опорных исследованиях* по соотносению систем индикаторов и латентных переменных. Такие исследования необходимы, конечно, и в естественных науках, и они проводятся, хотя и не выделены в самостоятельную область. Необходимость проведения опорных исследований в социальных науках возрастает, и, более того, здесь появляется необходимость в выделении их в относительно самостоятельную область. Обусловлено это именно большей многофакторностью социальных явлений по сравнению с природными. Соответственно системы латентных переменных и индикаторов социальных явлений более обширны и более «размыты», чем в естественных науках. Поэтому соотношение этих

систем требует больших и более координированных усилий исследователей *разных специальностей*, чем в естественных науках. Без придания опорным исследованиям самостоятельного статуса невозможен сколько-нибудь продуктивный количественный анализ социальных явлений.

Вторая особенность: большая, чем в естественных науках, субъективность результатов измерения

Эта особенность проявляется в трудности разделения объективно наблюдаемого факта и влияния наблюдателя. «В общественных науках, — говорит, например, по этому поводу Н. Винер, — связь между наблюдаемым явлением и наблюдателем трудно свести к минимуму. С одной стороны, наблюдатель может оказывать значительное влияние на явление, привлекая его внимание... С другой стороны, ученый-социолог не может взглянуть на свои объекты с исходных позиций нечности и недействительности... Другими словами, в общественных науках мы имеем дело с короткими статистическими рядами и не можем быть уверены, что значительная часть наблюдаемого не создана нами самими» [0.14, с. 202].

Проявляется субъективность измерения в социальных науках двойко. Во-первых, в *«перекосе»* спектра количественных методов в сторону относительно слабовоспроизводимых — социологических методов измерения, психометрических и др. Во-вторых, в наличии *объектом* мышления исследователей, в недостаточно развитом у них умении рефлексировать на наблюдателя даже при использовании воспроизводимых измерителей.

Приведем пример наличия объективного мышления, когда результат измерения относится исследователем целиком к объекту измерения без анализа того, что в нем, быть может, привнесено субъектом измерения. Р. Стерджис и Ф. Клементе [2.64] изучали продуктивность американских университетов в зависимости от престижа. Продуктивность оценивалась по числу публикаций. Оказалось, что наиболее престижные университеты отнюдь не отличаются большим числом публикаций. «В ходе этого исследования, — отмечают авторы, — возник интересный вопрос. Если допустить первым предположение, что в университеты с более высоким престижем поступают более способные студенты, то в результате чего же тогда эта способная молодежь впоследствии проявляет столь низкую научную продуктивность. И, наоборот, что происходит с менее общающимися студентами, поступающими в не столь престижные университеты и затем оказывающимися более плодовитыми авторами?» [2.64].

Совершенно очевидно, что авторы данной работы не задумываются над своими измерителями продуктивности. Но без анализа

измерителей, без анализа того, что в даваемых результатах относится к объекту, а что, быть может, привнесено исследователем, нельзя серьезно обсуждать эти результаты.

Субъективность социальных измерений означает фактически следующее: при исследовании человеческой деятельности труднее, чем при исследовании естественного мира, получать *воспроизводимые* характеристики объектов, *определяющие в их совокупности* поведение последних.

Получение *воспроизводимых* характеристик социальных объектов затруднено по той причине, что моменты стационарных распределений здесь уже, вообще говоря, не могут служить характеристиками распределений в силу негауссовости социальных явлений. Поэтому, если даже распределение построено на базе воспроизводимого индикатора, мы имеем в принципе меньше, чем в естественных науках, характеристик распределений и, следовательно, объектов. Если же, игнорируя негауссовость социальных явлений, все-таки использовать моменты, как сейчас и делается, результаты измерения окажутся существенно зависящими от объема выборки, меняющегося от исследователя к исследователю.

Что же касается характеристик социальных объектов, *определяющих в своей совокупности* с требуемой точностью поведение этих объектов, то получение их сложнее, чем в естественных науках. Обусловлено это многофакторностью человеческой деятельности: ведь число характеристик социальных объектов, определяющих в своей совокупности их поведение, в силу многофакторности существенно больше, чем число тех же характеристик для объектов природы. Здесь же следует указать, что именно в силу этой многофакторности воспроизводимые характеристики социальных объектов не могут уже один, как то часто имеет место в естественных науках, с требующейся для решения практических задач точностью и полнотой описывать объекты анализа. Необходимо привлечение значительного числа невоспроизводимых характеристик, что, конечно, снижает объективность анализа.

Чтобы сделать социальное измерение максимально объективным, во-первых, необходимо развивать по возможности количественные методы, дающие *воспроизводимые* результаты. Иными словами, развивать целостную систему индикаторов социальных явлений, которую требует их многофакторность, необходимо отделить приоритет более воспроизводимым индикаторам. Базирование количественного анализа социальных явлений на *негауссовой статистике*, снимающей зависимость результатов измерения от объема выборки, также будет способствовать увеличению объективности анализа. Во-вторых, необходимо сделать само мышление исследователя социальных явлений *менее объективным*, для чего

следует (парадоксы науки) явным образом вводить субъект измерения в процедуру измерения. Проще говоря, нужно научиться вычленять субъект измерения из результатов измерения.

Третья особенность: меньшая, чем в естественных науках, применимость математических формул

Этот факт нашел отражение в литературе по социальным наукам. «...В социологии нет законов, аналогичных законам Ньютона и Эйнштейна, — пишут, например, Э. П. Андреев и Г. В. Осипов, — и для области социальных явлений нет математической теории, подобной классической или квантовой механике» [0.6, с. 105].

«И, по-видимому, вряд ли когда-либо конкретные социальные явления приведут к открытию социального закона, выраженного на языке формул...» [0.49, с. 6—7].

Заметим, что в математике величины, подчиняющиеся аксиомам арифметики, называют *аддитивными* и *основные разделы математики приложимы лишь к аддитивным величинам*. Это касается алгебры, дифференциального и интегрального исчисления и т. д. Между тем в социальных исследованиях часто возникают *неаддитивные* данные. В этом и состоит одна из существеннейших причин малой эффективности применения математики в социальных науках.

Феномен негауссовости социальных явлений помогает и здесь. И помогает не только *понять* корни возникающих трудностей, но и *преодолеть* их.

В социальных исследованиях используются шкалы *открытые* и *закрытые*. Закрытые шкалы используют баллы, ограниченные сверху. Примером закрытой шкалы могут служить школьные отметки. Открытые шкалы не ограничивают баллы сверху. Таковы наукометрические шкалы.

Анализ встречающихся в социальных исследованиях шкал и переменных позволяет заметить следующее кардинально важное обстоятельство: *открытые шкалы дают аддитивные переменные, закрытые — неаддитивные*. Более строго: социальные переменные, распределения по значениям которых негауссовы, аддитивны; социальные переменные, распределения по значениям которых гауссовы, — неаддитивны. В этом, собственно, и состоит феномен негауссовости социальных явлений.

Выбор латентных переменных в ходе исследования данного явления, как было показано в § 2, — дело во многом интуитивное. Человеческая практика за многие сотни и тысячи лет показала, что целесообразнее всего выбирать *аддитивные* латентные переменные. Именно такими переменными и оперируют чаще всего естественные науки.

Проблему неаддитивности социальных данных мы предлагаем «перевернуть». Вместо того чтобы принимать эту неаддитивность как факт и «бороться» с возникающими отсюда трудностями, попробуем отказаться от неаддитивных переменных и применять только аддитивные. Практически это означает, что *нужно отказаться от закрытых шкал и использовать только открытые*.

Применение аддитивных переменных существенно расширит возможности применения математики в социальных науках.

Сегодня в социальных исследованиях можно выделить три уровня применения математики: измерение, математическое описание (моделирование), теория принятия решения (исследование операций, теория игр и др.). Здесь на основе полученных на двух нижних уровнях эмпирических значений переменных и связывающих их математических соотношений выбираются оптимальные решения из наборов альтернатив. Пока что нижний и два верхних уровня математических приложений в социальных исследованиях существуют во многом независимо друг от друга. Формулы пишутся в расчете (осознанном или неосознанном) на аддитивные переменные, а результаты измерения часто бывают неаддитивными. Отсюда и низкая эффективность математики. Ограничение открытыми шкалами решит эту проблему.

Вероятностность и негауссовость социальных явлений определяют математический аппарат их количественного анализа. Статистический язык — единственно возможный здесь язык, а негауссовая статистика — грамматика этого языка.

В основе всякого количественного анализа лежат количественные закономерности, позволяющие не только определенным образом унифицировать описание объектов исследования, но и учесть их *специфику*. Если исследуемые объекты описываются, скажем, все одной и той же количественной закономерностью, то отдельные объекты отличаются друг от друга значениями параметров этой объединяющей их закономерности. Измеряя значения параметров, мы и фиксируем специфику объектов.

Унификация описания тем более необходима, чем более многофакторно это описание. Многофакторная картина данного явления до тех пор останется мозаикой, трудно поддающейся расшифровке, пока языки описания факторов разной природы не будут унифицированы и пока, таким образом, не появится возможность учета специфики отдельных факторов. В силу большей многофакторности социальных явлений требование унифицированности описания здесь существенно более актуально, чем в естественных науках. Таким образом, *в социальных науках роль количественных закономерностей возрастает*. И поскольку выявление действующих в области социального количественных закономерностей

существенно затруднено по сравнению с естественными науками, поскольку социальные исследования требуют особых усилий в этом направлении.

Негауссовость социальных явлений служит свою службу и в этом отношении. Как говорилось в § 3, все стационарные распределения здесь — и это есть проявление их негауссовости — имеют при больших значениях переменных вид распределения Ципфа. Это — количественная закономерность, действие которой распространяется, по крайней мере, на социальные явления любой природы.

Вывод из параграфа, отсюда, что в предлагаемых нами рекомендациях, вытекает из применяемой в настоящем тексте концепции, что — и это является ранее положенным. Многофакторность социальных исследований следует из простой констатации многофакторности социальных явлений, необходимость же выделения *опорных исследований* в относительно самостоятельную область социальных исследований вытекает уже из развитой здесь концепции измерения. Необходимость уделения большого внимания количественным методам, дающим воспроизводимые результаты, вытекает из известных соображений о большей субъективности социальных исследований в сравнении с естественнонаучными, необходимости же базирования количественного анализа социальных явлений на *негауссовой статистике* вытекает из представления о негауссовости социальных явлений. Из этих же представлений вытекает вывод о том, что количественный анализ социальных явлений должен базироваться на *открытых шкалах*. То, что *стабильность стационарных распределений* является одной из основных количественных закономерностей, распространяющейся, по крайней мере, на социальные явления любой природы, снова-таки, вытекает из представления о негауссовости социальных явлений.

Вывод из параграфа, отсюда, что в предлагаемых нами рекомендациях, вытекает из применяемой в настоящем тексте концепции, что — и это является ранее положенным. Многофакторность социальных исследований следует из простой констатации многофакторности социальных явлений, необходимость же выделения *опорных исследований* в относительно самостоятельную область социальных исследований вытекает уже из развитой здесь концепции измерения. Необходимость уделения большого внимания количественным методам, дающим воспроизводимые результаты, вытекает из известных соображений о большей субъективности социальных исследований в сравнении с естественнонаучными, необходимости же базирования количественного анализа социальных явлений на *негауссовой статистике* вытекает из представления о негауссовости социальных явлений. Из этих же представлений вытекает вывод о том, что количественный анализ социальных явлений должен базироваться на *открытых шкалах*. То, что *стабильность стационарных распределений* является одной из основных количественных закономерностей, распространяющейся, по крайней мере, на социальные явления любой природы, снова-таки, вытекает из представления о негауссовости социальных явлений.

6. Сущность измерения научной деятельности

В параграфе выдана сформулированная и предложенная концепция измерения научной деятельности (§§ 1-2) и обоснована необходимость измерения научной деятельности в отрыве от ее содержания (§§ 3-4). Будет ли измерение научной деятельности и ее результатов, научной дея-

тельности. Кратко обсудим и особенности науковедческого измерения в отличие от измерения в других социальных науках.

1. (§ 1). При количественном анализе научной деятельности, при выборе конкретной методики измерения необходимо принимать во внимание три основных критерия процедуры измерения — *воспроизводимость, адекватность, трудоемкость*.

2. (§ 2). Решение проблемы измерения при изучении научной деятельности состоит в следующем. Во-первых, на очередном этапе познания латентные переменные (понятия об измеряемых свойствах объектов, существующие в нашем сознании) отождествляются с их операциональными референтами — *операциональными латентными переменными*, конструируемыми из групп фиксированным образом взвешенных индикаторов (непосредственно измеряемых, наблюдаемых величин). Во-вторых, система индикаторов соотносится с системой операциональных латентных переменных, т. е. для каждого индикатора определяются вероятностные зависимости его числовых значений от числовых значений операциональных латентных переменных, так называемые графики. Производимое в рабочем порядке отождествление латентных переменных с их операциональными референтами и выражается в отождествлении графиков индикаторов относительно операциональных латентных переменных с графиками индикаторов относительно латентных переменных. В-третьих, соответствие латентных переменных и их операциональных референтов устанавливается в ходе рефлексивного осознания человеком своей деятельности, ее целей, средств, результатов. Под давлением практики и происходит постепенное, от одного этапа познания к другому, совершенствование системы операциональных латентных переменных.

3. (§ 2). Продуктивный количественный анализ научной деятельности невозможен без придания *опорным исследованиям* по разработке систем индикаторов и латентных переменных науки в их соотносительности самостоятельного статуса.

4. (§ 3). Феномен негауссовости стационарных индикаторных распределений научной деятельности позволяет сформулировать гипотезу о негауссовости стационарных латентных распределений научной деятельности, т. е. о *негауссовости научной деятельности*. Этот феномен может быть должным образом осмыслен лишь при его сопоставлении с феноменом гауссовости природных явлений.

5. (§ 3). Научная деятельность вероятностна по своей природе. Это — эмпирический факт, который может быть объяснен *негауссовостью научной деятельности*.

6. (§ 4). Научную деятельность характеризует большая, чем в естественных науках, *многофакторность явлений*, поэтому и ана-

лиз научной деятельности должен быть многофакторным. Для его осуществления необходимо, в частности, объединять усилия специалистов разного профиля. В этом заключается одна из основных причин необходимости существования науковедения как самостоятельной комплексной дисциплины.

7. (§ 4). Измерение при изучении научной деятельности отличается большей, чем в естественных науках, субъективностью результатов. В целях преодоления этой субъективности нужно развивать по возможности воспроизводимое количественное измерение. Наукометрические методы, базирующиеся на наиболее воспроизводимых индикаторах научной деятельности, должны лежать в основе количественного анализа научной деятельности. Использование негауссовой статистики, снимающей зависимость результатов измерения от объемов выборок (вообще говоря, разных у различных исследователей), также будет способствовать увеличению объективности науковедческого измерения. В этих же целях необходимо развивать у исследователей науки менее объектное мышление, т. е. способность рефлексировать на субъект познания, вычленять его из результатов измерения.

8. (§ 4). Количественный анализ научной деятельности характеризуется меньшей, чем в естественных науках, применимостью математических формул. Данная особенность заставляет при количественном анализе научной деятельности использовать лишь открытые шкалы, отказавшись от закрытых, которые дают индикаторы, неадекватные аддитивным по своей природе латентным переменным.

9. (§ 4). В силу вероятностности научной деятельности, статистический язык — единственно возможный язык количественного ее анализа. В силу негауссовости научной деятельности, негауссова математическая статистика — грамматика этого языка.

10. (§ 4). В основу количественного анализа научной деятельности должны быть положены количественные закономерности. Только такие закономерности, позволяющие определенным образом унифицировать количественное описание объектов науковедческого анализа, позволяют учесть специфику отдельных объектов. Поскольку выявление количественных закономерностей научной деятельности существенно затруднено по сравнению с естественными науками в силу указанных выше причин, постольку исследования научной деятельности требуют особых усилий. Одну такую закономерность дает феномен негауссовости научной деятельности: все стационарные распределения здесь при больших значениях переменных имеют с точностью до медленно меняющихся функций вид распределения Ципфа. Это явление мы называем ципфовостью научной деятельности, а распределения, имеющие

при больших значениях переменных вид распределения Ципфа, — ципфовыми.

11. Количественные закономерности в процессе их выявления должны быть максимально «очищены» от привходящих обстоятельств, отрицательно влияющих на устойчивость результатов измерения: оттого что замеры свойств объектов могут быть разнесены во времени и в пространстве, могут производиться разными исследователями и при этом могут использоваться разные процедуры измерения. Количественные закономерности можно выявлять лишь на базе воспроизводимого измерения. Поэтому именно наукометрия, в задачи которой входит воспроизводимое измерение научной деятельности, «отвечает» и за выявление здесь количественных закономерностей.

12. Научной деятельности в силу ее направленности на производство нового знания в большей мере, чем многим другим видам человеческой деятельности, присущи элементы творчества. Между тем имеются эмпирические указания в пользу того, что, чем в большей степени данная человеческая деятельность носит творческий характер, чем в большей степени она является «человеческой», тем в большей степени она негауссова. Поэтому особенности измерения при изучении научной деятельности в отличие от измерения при изучении явлений природы выражены ярче, чем эти же особенности для менее творческих видов человеческой деятельности. Отсюда вытекает, что для науковедения практические следствия, вытекающие из анализа особенностей измерения ее объектов, еще более актуальны, чем для социальных наук, изучающих менее творческие виды человеческой деятельности, например для экономики.

13. В отличие от других творческих видов человеческой деятельности (например, искусства) научная деятельность характеризуется фиксированностью продукта труда ученого в научных текстах, имеющих дискретную структуру. Это обстоятельство создает предпосылки для воспроизводимого измерения параметров научной деятельности на основе счета отдельных элементов научных текстов. Воспроизводимое измерение при изучении научной деятельности облегчено по сравнению с другими видами творческой деятельности. Поэтому наукометрия имеет благоприятные перспективы, а наукометрические индикаторы базируются, как правило, на научном тексте. В настоящее время ежегодно публикуется свыше 3 млн. журнальных статей только по естественным и техническим дисциплинам, более 75 тыс. книг, более 450 тыс. описаний патентов и авторских свидетельств [0.59], а общее число опубликованных к настоящему моменту работ составляет уже несколько десятков миллионов. Каких-либо существенных изменений в этом

отношении, кроме, быть может, некоторых возможных перемен в характере научной публикации, пока не предвидится [1.136; 4.98; 1.100; 1.02; 1.159; 1.103 и др].

14. В силу вероятностной природы научной деятельности наукометрические методы, как и другие количественные методы анализа научной деятельности, *вероятностны* в своей основе. «Суть моего метода, — пишет, например, один из основателей наукометрии Д. Прайс, — это статистический... анализ общих проблем формы и объема науки... Иными словами, я не буду говорить о каких-либо определенных открытиях, их использовании, их взаимосвязи, но, рассматривая науку как измеримую целостность, я попытаюсь разбить исчисление научного персонала, литературы, таланта, а также расходов на науку в национальном и международном масштабах...

Метод, которым я буду пользоваться, сродни тем методам термодинамики, которыми исследуют поведение газа в различных условиях температуры и давления. При таком подходе не обращают внимания на поведение одной какой-нибудь молекулы (названной, например, Джордж), которая движется со своей особой скоростью и находится в заданный момент в своем особом месте. Рассматриваются только средние значения всего ансамбля, в котором некоторые молекулы движутся быстрее других, а распределены они в пространстве случайным образом и движутся в разных направлениях. Основываясь на таком безличном среднем, можно узнать довольно много полезного о поведении газа в целом. Именно этим методом безличного среднего я попытаюсь исследовать науку как целостность» [1.58, с. 282—283].

В этом описании Прайса нетрудно узнать вероятностный подход к количественному анализу научной деятельности. Только его аналогия систем науки с газом молекул не совсем верна: системы науки негауссовы, а молекулярный газ — гауссов. Различие это настолько существенно, что сами эти «безличные средние», о которых говорит Прайс, для научной деятельности имеют ограниченное значение в силу их зависимости от объемов выборок.

15. Все наукометрические методы имеют одну — статистическую — природу, в их основе лежит одна математическая теория — негауссова статистика, и базируются они на одной эмпирической базе — на научном тексте. Образно говоря, совокупность наукометрических методов представляет собой один «наукометрический кулак» с «пальцами-индикаторами». *Индикаторы — это единственное, что отличает один наукометрический метод от другого.*

Наукометрические индикаторы различаются «дробностью», т. е. величиной той части массива публикаций, которой соответствует единица данного индикатора. В порядке возрастания «дробности»

наукометрические индикаторы располагаются примерно так: число научных направлений, журналов, организаций, ученых, публикаций, ссылок, сленговых слов. В самом деле, научное направление представлено обычно несколькими журналами, один журнал, как правило, приходится на несколько научных организаций, научная организация состоит из ученых, ученых имеет несколько публикаций, на одну публикацию приходится в среднем несколько (порядка 14) ссылок, на одну публикацию приходится в среднем порядка 1 тыс. сленговых слов. Думаю, читатель теперь и сам сможет упорядочить по «дробности» любые наукометрические индикаторы.

В следующих далее пяти главах освещаются отдельные наукометрические методы и даваемые ими результаты. Материал всюду структурирован так, что более «дробные» индикаторы следуют за менее «дробными».

Глава вторая

СТАТИСТИЧЕСКИЙ МЕТОД

1. СОДЕРЖАНИЕ МЕТОДА И ОПОРНЫЕ ИССЛЕДОВАНИЯ

Отдельные наукометрические методы определяются используемыми в них измерителями (индикаторами). *Статистический метод мы определяем как метод, использующий в качестве наукометрических индикаторов все измерители, кроме числа публикаций, ссылок и отдельных слов*, о которых речь пойдет в гл. 3—6. Такое определение, конечно, условно и оправдывается только задачами данного исследования. В будущем, возможно, на основе некоторых относимых нами сегодня к статистическому методу измерителей выделятся самостоятельные наукометрические методы. Название *статистического* метода также условно — как уже говорилось в предыдущей главе, все наукометрические методы имеют статистическую природу.

В число измерителей науки, охватываемых статистическим методом, попадают такие измерители, как число ученых, журналов, заказов на годовые комплекты журналов в библиотеках и информационных центрах, открытий и другие, имеющие относительно крупные «единицы измерения», либо несоизмеримые с другими наукометрическими измерителями (например, энергия ускорителя элементарных частиц).

Статистический метод, как и вся наукометрия, вырос из обыкновенной статистики. Статистический учет различных областей человеческой деятельности, в том числе и науки, ведется примерно с XVII в. Дж. Бернал [0.6; 0.7] был в числе первых, кто вел *систематический* сбор количественной информации о науке, «достаточно репрезентативной и обработанной для анализа, что дало возможность дать точную формулировку нескольким положениям, которые раньше рассматривались как результат качественного умозрительного анализа» [1.59, с. 249]. Среди предшественников Дж. Бернала, ведших спорадические исследования науки на основе статистических данных, можно назвать, например, Ф. Гальтона [0.16; 1.102], определившего, в частности, что в Англии на 1 млн. ученых приходится 5—10 выдающихся.

В СССР справочники по научным работникам выпускаются с 1920 г. [см. 2.32]. Численность научных работников в СССР достигает сегодня примерно 1,3 млн. человек, а в мире — порядка 4 млн. человек, так что массив ученых представляет собой статистическую совокупность, достаточную для рассмотрения целого ряда задач. «Можно с уверенностью сказать, что величины смертности, выхода на пенсию, миграций между наукой и производством, притока молодых специалистов и прочих параметров, определяющих динамику развития этой социальной группы (научных работников.— С. Х.), подчиняются статистическим закономерностям» [1.50, с. 177—179]. Различных научных журналов выпускается сейчас в мире свыше 70 тыс. [1.42, с. 217—218], что также является достаточной статистической совокупностью для рассмотрения некоторых науковедческих задач. То же самое верно и в отношении других рассматриваемых в настоящей главе измерителей. Вопрос заключается в определении круга задач, при решении которых возможно применение этих измерителей, степени полезности получаемой с их помощью информации, в выборе статистики, необходимой для получения информации. Иными словами, речь идет о соотношении рассматриваемых здесь индикаторов с другими индикаторами и латентными переменными. Для решения этой задачи необходимо проведение соответствующих *опорных исследований*. Как говорилось, опорные исследования сегодня практически выливаются в измерение корреляций между различными индикаторными распределениями. Возникает вопрос об отнесении данных о корреляции к той или иной рубрике нашего исследования. Этот вопрос необходимо решать вместе с вопросом более широкого плана: при рассмотрении какого индикатора из двух задействованных, следует рассматривать данные о корреляции двух соответствующих им распределений. Строго говоря, в каждом конкретном случае из двух задействованных распределением индикаторов лишь один является в полном смысле индикатором. Именно его значения откладываются по оси абсцисс. Распределение не может быть «перевернуто» без изменения его смысла: скажем, распределения ученых по числу публикаций и публикаций по числу ученых — разные по смыслу. В целях упрощения рубрикации монографии мы пошли по пути ее формализации и, не вникая в содержание каждого распределения, относили его к рубрике, содержащей более «дробный» индикатор. Аналогично данные о корреляции двух распределений мы излагаем в главах, посвященных более «дробному» измерителю. Поэтому данные по опорным исследованиям относительно связи индикаторов, относящихся к статистическому методу, с другими наукометрическими индикаторами следует искать в следующих главах.

Мы не встретили работ, в которых специально рассматривалась бы связь анализируемых в настоящей главе наукометрических индикаторов с латентными переменными. На качественном же уровне высказываются соображения о возможности измерения с помощью этих индикаторов *роста науки* в целом и отдельных ее направлений (индикаторы — число журналов, ученых и т. д.), *тенденций коллективности* научного труда (с помощью анализа сотей соавторства [см. 2.30; 1.6 и др.], *перспективности* отдельных научных направлений (на основе анализа числа ученых разных категорий [2.22], спроса на литературу [3.3 и др.]).

Стационарные распределения также могут быть использованы для выявления перспективных научных направлений. В. И. Горькова [1.11], проанализировав более 30 различных наукометрических стационарных распределений, пришла к выводу, что перспективные, или, как она говорит, *проблемные*, направления характеризуются распределениями, имеющими более короткие «хвосты», нежели традиционные.

2. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: НЕСТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

Эмпирические данные, анализируемые в этой и последующих главах, мы будем разбивать на три группы: эмпирические данные, содержащие временные зависимости; статистические распределения, не содержащие времени; эмпирические данные, отражающие структуру исследуемых объектов науки. Соответственно будут выделяться отдельные параграфы. В настоящем параграфе рассматриваются данные по нестационарным распределениям, полученные на основе измерителей науки, «подведомственных» статистическому методу. Материал структурирован по измерителям.

Временная динамика числа открытий

Рост во времени числа открытий рассматривается в работах Т. Райнола [2.56], Г. Лемана [2.53], П. Оже [1.73], Д. Прайса [1.58], Е. Д. Гражданникова [2.9] и др. Т. Райнол, рассматривая в 1929 г. временную динамику числа открытий в физике за XVIII—XIX вв., на своей относительно небольшой статистике получил медленный, почти линейный рост числа открытий (см. рис. 1.2). Г. Леман (1947), П. Оже (1961) и Д. Прайс (1963) получили для роста числа открытий экспоненту. Период удвоения числа открытий равен примерно 10 годам [1.73], числа *важных* открытий — 20 годам. Здесь мы сталкиваемся с зависимостью результатов измерения от выбора процедуры измерения. В данном

случае от выбора критериев, по которым научное достижение считается (но считается) научным открытием или важным научным открытием. Ужесточение этих критериев приводит, как и в случае роста числа ученых (см. далее), к увеличению периода удвоения.

Е. Д. Гражданников аппроксимирует кривую роста числа открытий, точнее — пиков этой кривой, геометрической прогрессией [2.9].

Временная динамика числа журналов

Рост числа журналов рассматривают в своих работах Е. Клайн [1.121], Ф. Пич [1.137], М. Конрад [2.41], П. Оже [1.73], Б. Гласс [2.45], Д. Прайс [1.58], И. Сенгупта [1.155], А. И. Михайлов с соавт. [1.42] и др. Для больших статистических совокупностей и при охвате достаточно большого промежутка времени кривые роста числа журналов могут быть аппроксимированы экспонентой (см. рис. 1.3). При рассмотрении этого графика необходимо, однако, «сделать оговорку, что в нем учитывалось суммарное число когда-либо выходивших журналов, включая и те, которые прекратили свое существование. А что такая оговорка весьма существенна, можно судить по тому факту, что ежедневно возникают три новых журнала по естественным, точным и прикладным наукам, а один журнал перестает выходить» [1.42, с. 172].

Период удвоения числа журналов составляет для мира в целом около 15 лет [1.58, с. 289]. Что касается отдельных наук, то такой период известен, пожалуй, только для биологии, где он составляет 15 [2.45]—18 [2.41] лет.

Экспоненциальный рост числа журналов не может по естественным причинам продолжаться до бесконечности и потому в настоящее время он замедляется. А. И. Михайлов с соавт. нашли, например, что на протяжении двух последних десятилетий ежегодное приращение числа журналов оставалось постоянным и составляло около 2 тыс. журналов [1.42, с. 171].

Временная динамика числа ученых

Анализ динамики роста числа ученых содержится в работах П. Оже [1.73], Д. Прайса [1.58, 1.59], Г. М. Доброва [0.23], Л. Нового и Я. Фолты [1.52], В. И. Клименюка [1.26], Г. М. Доброва и А. А. Коренного [1.21], В. В. Налимова и З. М. Мульченко [1.46], Я. Влахи [2.69], Г. М. Бастракова с соавт [2.1], С. Солдатова [1.61], У. Менарда [1.125], В. Д. Комарова [2.19], Г. Даниельса [0.76], Д. Крайн [1.90], В. С. Сомпского и Э. М. Торфа [2.33], В. Д. Осетрова [2.29] и др. Согласно их результатам, при больших статисти-

ческих совокупностях кривые роста числа ученых могут быть аппроксимированы экспонентами (см., например, рис. 1.4б). При этом период удвоения числа ученых составляет для стран капиталистического мира в целом 15 лет [0.24, с. 17], для США — 10 лет [1.73; 1.58, с. 289] и для СССР — 6–7 лет [0.46, с. 17]. Период удвоения численности выдающихся ученых больше периода удвоения численности всех ученых. «10-летний период удвоения получают в случаях поточных замеров, где не учитывается качество работ, и в основу определения критерия научности берут минимум требований, — пишет Д. Прайс. — 15-летний период удвоения возникает в тех случаях, когда отбор производится более строго и учитывают лишь те научные работы или только тех авторов, которые удовлетворяют более узким определениям научности. Если строгость отбора возрастает и учитывают только научные труды очень высокого качества, период удвоения начинает увеличиваться и приближается к 20 годам» [1.58, с. 189].

Как видим, выбор процедуры измерения самым прямым образом сказывается на результате измерения. Процедура измерения в данном случае устанавливает способ «обрезания» выборки: устанавливая разные критерии, по которым конкретные индивиды попадают (не попадают) в выборку «ученые», мы тем самым меняем период удвоения роста объема нашей выборки. Чтобы результаты измерения были воспроизводимыми, необходимо, чтобы воспроизводимыми были и критерии научности. К сожалению, эти вопросы в литературе пока не обсуждались.

Очевидно, что возрастание числа ученых по экспоненте с периодом удвоения 10–15 лет — очень быстрое. Население удваивается, как известно, за 35 лет. Согласно оценкам Дж. Бернала [0.6], в 1896 г. в мире насчитывалось не более 15 тыс. ученых, сейчас их около 4 млн. В России в конце XIX в. было около 2 тыс. ученых, сейчас их в СССР — порядка 1 млн. 300 тыс. [1.22, с. 99]. Д. Прайс говорит об этом как об эффекте «непосредственной данности». Численность ученых, живущих в настоящее время, составляет 80–90% от численности всех ученых, когда-либо живших на Земле. «Эта непосредственная данность есть настолько сильная и доминирующая черта науки, что она выявляет себя как основу отношения к науке и ученых и дилетантов. Именно благодаря этой черте наука всегда представляется новой и современной» [1.58, с. 285].

Экспоненциальный рост числа ученых наблюдается в течение двух-трех столетий [1.58, с. 288]. Точнее, наблюдался. «В реальном мире, — пишет Прайс, — не бывает так, чтобы вещи росли и росли до бесконечности. Экспоненциальный рост постепенно приближается к какому-то пределу, процесс замедляется и останавливается, не достигая абсурдных значений» [1.58, с. 301]. Замедление тем-

пов роста численности ученых и выход экспоненты на более пологие кривые заметен, например, на рис. 1.4а для США и в какой-то мере Франции, а на рис. 1.4б — для США.

При меньших статистических совокупностях кривые роста числа ученых уже труднее аппроксимировать экспонентой (см., например, рис. 1.4г, 1.18). Однако, и здесь экспоненциальная аппроксимация может быть в ряде случаев полезна: кривые роста, как говорит Прайс [1.145], пульсируют около экспоненты (см. рис. 1.4д).

Иногда кривые роста числа научных работников аппроксимируются не одной, но несколькими экспонентами [см., например, 1.52; 2.1].

Временная динамика соавторства

Анализ динамики роста соавторства содержится в работах Д. Прайса [1.58], Б. Кларка [2.40], Г. Цукерман [2.72], Я. Влахи [2.70], Г. М. Доброва и А. А. Коренного [1.22], Г. Цукерман и Р. Мертон [0.117], Н. М. Мончева [2.26], Ю. М. Червинского [1.65], В. Н. Крутько [2.21], Г. М. Доброва и Л. А. Игнатов [2.10], Д. Бивера, Р. Розена [1.74] и др. Первые столкнулся с проблемой соавторства Б. Берельсон в 1960 г. [2.38]. Однако внимание специалистов к этому феномену привлек не Берельсон, а Д. Прайс [1.58], детально обсуждающий этот феномен.

Ученые отметили возрастание удельного веса работ с двумя и более авторами. Так, почти 80% работ по химии, физике, биологии написаны в соавторстве (см. рис. 1.5а, б, в) (табл. 2.2.1).

Разные области науки имеют свои темпы роста соавторства, а в некоторых наблюдается стабилизация этого роста, связанная с эффектом насыщения [2.10].

Исследования структуры соавторских коллективов (см. § 4 настоящей главы) показали, что на соавторство определенным образом влияют отношения научного руководства — лидеры научных коллективов зачастую являются и лидерами соавторских коллективов. Причем в отношениях соавторства отношения научного руководства фиксируются явным образом. Вот эта явность и делает соавторство одним из интересных объектов научного анализа.

Д. Прайс объясняет рост соавторства особенностями динамики продуктивности. «Одним из способов понимания движения к массовому сотрудничеству, — пишет он, — может служить представление о нем, как об естественном расширении роста, который вызван постоянным сдвигом распределения Парето для научной продуктивности (т. е. распределения Лотки ученых по числу публикаций — см. гл. 3, § 3. — С. Х.). Существует постоянное

Таблица 2.2.1

Временная динамика доли статей, написанных несколькими авторами, % [0.117]

Год публикации	Физические и биологические науки	Общественные науки	Гуманитарные науки
1900-1909	25	—	—
1910-1919	31	—	—
1920-1929	49	16	1
1930-1939	56	11	1
1940-1949	66	16	2
1950-1959	83	32	1
Итого:	66	20	1

стремление к повышению продуктивности наиболее выдающихся авторов и к увеличению числа авторов с минимальной продуктивностью. Когда и в том и в другом направлении мы приходим к пределу, чем-то, естественно, приходится жертвовать. Наиболее выдающиеся авторы увеличивают продуктивность, становясь лидерами научных коллективов, в совместной работе с которыми они могут сделать больше, чем в одиночку. Малопродуктивная группа становится своего рода источником: ее членам не позволяют расти до тех пор, пока они не достигнут зрелости, то есть будут в состоянии дать ценную собственную статью. Возникновение класса частичных авторов (ученых, производящих только какую-то часть статьи) удерживает в хвосте распределения значительно большее число минимально продуктивных авторов, то есть обеспечивает большой объем науки» [1.58, с. 361].

Другими словами, Прайс говорит здесь вот о чем: число выдающихся ученых растет медленнее общего числа ученых (см. с. 42). И вся эта масса ученых, не могущих давать самостоятельную продукцию, хочет ее давать. Среди больших же ученых всегда есть группа, не могущая реализовать все свои идеи из-за ограниченности производительности труда отдельного человека. В соавторстве удовлетворяются и те и другие. Из-за увеличения доли самостоятельно непродуктивных ученых и имеет место, по Прайсу, рост соавторства.

Заметим, что такое объяснение роста соавторства неполно. В самом деле, легко можно найти другие факторы, действующие как в направлении увеличения соавторства, так и в противоположном направлении. К факторам, содействующим росту соавторства, можно отнести, например, усиление экспериментальной базы науки. При этом получает естественное объяснение и различный

рост соавторства в разных дисциплинах, тогда как интерпретация Прайса оставляет этот вопрос открытым.

К факторам, препятствующим росту соавторства, можно отнести, например, усиление конкуренции в науке и связанное с этим стремление отдельных ученых выделиться, опубликовать самостоятельные статьи. Или другой фактор — усилившаяся дифференциация научного знания: в условиях, когда даже работающие в одной области ученые зачастую не понимают друг друга, легче публиковать самостоятельные статьи, чем в условиях, когда все работающие в данной области тебя понимают.

Здесь мы сталкиваемся с необходимостью многофакторного анализа научной деятельности. Такой анализ предприняли в своем недавнем исследовании Д. Бивер и Р. Розен [1.74], выдвинувшие и проверившие несколько различных гипотез о природе соавторства.

Временная динамика возрастной структуры научных кадров

Анализ проводился в работах Л. Нового и Я. Фолты [1.52], Т. З. Козловой и А. Г. Кузнецова [2.18] и др. Л. Новый и Я. Фолта в результате анализа временной динамики среднего возраста математиков к моменту опубликования первой работы выявили тенденцию к омоложению авторов первой работы.

Т. З. Козлова и А. Г. Кузнецов рассматривают изменение возрастной структуры научных кадров как вероятностный процесс, что соответствует общей вероятностной природе науки. Для прогнозирования такого вероятностного процесса авторы аппроксимируют его с помощью марковских цепей: используя марковские модели через прогнозирование матриц переходных вероятностей авторы осуществляют прогнозирование процесса изменения возрастной структуры ученых СССР. При этом они делают несколько шагов назад во времени и линейно экстраполируют в будущее. Такая линейность оправдана, конечно, только при прогнозировании на близкое будущее.

Динамика роста «экзотических» индикаторов науки

Здесь отметим работы М. Ливингстона и Дж. Блуэтта [2.55], Г. Холтона [1.115], Д. Прайса [1.58], С. С. Бацанова и Ф. А. Брусенцова [2.2], В. В. Налимова и З. М. Мульченко [1.46], С. С. Бацанова и В. Н. Ручкиной [2.3], В. М. Потапова и И. Н. Лукиной [5.13], В. Н. Бернадского и В. В. Журавкова [2.4]. В этих

работах для анализа динамики роста параметров науки используются, как правило, измерители, «привязанные» к определенной предметной области. Одним из первых результатов здесь был знаменитый график, показывающий временную динамику энергии элементарных частиц, достигаемую в ускорителе (см. рис. 1.11). Этот график впервые дали М. Ливингстон и Дж. Блуэнт [2.55], затем его перепечатал Г. Холтон [1.115], но широкую известность он получил благодаря Д. Прайсу, поместившему его в своей книге [1.58]. Указанная зависимость аппроксимируется, как показывает график, экспонентой.

Другой знаменитый график построен Г. Холтоном. Это — временная динамика числа известных химических элементов, аппроксимируемая суммой экспонент (см. рис. 1.10).

С. С. Бацанов и Ф. А. Брусенцов [2.2] проводят статистический анализ кристаллоструктурных работ. Рис. 1.12—1.15 демонстрируют динамику роста некоторых индикаторов, характеризующих исследования в этой области. Эти графики показывают, что индикаторы отдельных предметных областей науки не всегда растут по экспоненте.

Д. Прайс утверждает, что точность инструментов растет со временем по экспоненте с периодом удвоения порядка 20 лет [1.58, с. 289]. В. В. Налимов и З. М. Мульченко приводят результаты выполненного ими совместно с М. П. Бурмистровым исследования роста чувствительности эмиссионного спектрального анализа при определении содержания редких элементов и цветных металлов (см. рис. 1.16). Малая статистика не позволяет утверждать, что мы имеем здесь дело с экспонентой.

Существуют также данные о динамике роста таких индикаторов науки, как число университетов, число известных химических соединений, число известных астероидов [1.58, с. 289], листах диссертаций [0.22] и листах журналов [2.48; и 3.103], рабочая площадь [2.69] и т. д.

Характерно, что статистический анализ динамики роста «экзотических» индикаторов науки, имеющих хождение лишь в пределах узких предметных областей, адресуется зачастую работникам этих областей [см., например, 2.4]. Это и понятно: узкоспециализированные индикаторы не могут быть использованы для количественного сравнения разных предметных областей науки.

Возрастная динамика открытий

Зависимость числа открытий от возраста ученого, в котором они делаются, рассматривали П. И. Вальден (см. [0.46]), Г. Леман [3.68—3.71], И. Кацаров [2.15] и др. Основной результат: распре-

деление числа открытий по возрасту ученого имеет пик, смещенный к началу научной деятельности. Русский химик академик П. И. Вальден установил, что большинство крупнейших открытий в физике и химии конца XVIII — начала XX в. было сделано учеными в возрасте 25—30 лет [0.46, с. 16].

Наибольшую известность получили исследования Г. Лемана, первое из которых появилось в 1953 г. [3.68]. Леман обнаружил не только пик творческой активности ученого, но и три немаловажных обстоятельства: во-первых, пик тем более смещен к началу научной деятельности, чем более крупными открытиями ограничена выборка (см. рис. 1.21а—с). Кроме того, для более крупных открытий этот пик выражен резче (см. рис. 1.21а). С точки зрения результатов Лемана остается открытым вопрос о существовании пика творческой активности для *среднего* ученого.

Во-вторых, пик тем более смещен к началу научной деятельности, чем больше важных вкладов сделано этим ученым за жизнь в целом, т. е. чем выше ранг ученого (см. рис. 1.21а).

И, наконец, в-третьих, положение пика творческой активности ученого зависит от предметной области, в которой он работает: в абстрактных науках (математике, теоретической физике) этот пик более смещен к началу научной деятельности по сравнению с науками, ориентированными на эмпирию (геология, биология).

Здесь мы снова имеем дело с зависимостью результата измерения от выбора процедуры этого измерения: положение пика творческой активности ученого, его выраженность и самый факт существования зависят от степени жесткости критериев, по которым мы возводим научные достижения в ранг учитываемых нами «открытий» или относим к числу «существенных вкладов». Если эти критерии невоспроизводимы, то результат измерения имеет низкую научную ценность. Это понимал сам Леман: если в первых его работах, например в работе [3.68], отбор «существенных вкладов» производился целиком на основе качественных соображений, то в последующих своих публикациях Леман до известной степени формализует процедуру их отбора. Например, в работе [3.70] «существенными» он считал только те вклады химиков, которые обсуждались не менее чем в 15 из 44 книг по истории химии.

Возрастная динамика соавторства

Возрастную динамику соавторства Нобелевских лауреатов и взятой для сравнения с ними группы «средних» ученых изучали Г. Цукерман и Р. Мортон [0.117] (табл. 2.2.2).

Авторы отмечают два факта. Во-первых, согласно данным табл. 2.2.2, Нобелевские лауреаты чаще соавторствуют. И, во-вто-

Таблица 2.2.2

Возрастная динамика соавторства Нобелевских лауреатов и сравниваемых с ними ученых, % [0.117]

Возраст в момент публикации	Статьи, написанные в соавторстве	
	лауреатами	сравниваемыми учеными
20-29	58(523) *	40(288)
30-39	65(1382)	55(756)
40-49	66(1641)	53(590)
50-59	60(1198)	51(622)
60 и более	55(768)	48(264)
Всего	62(5612)	51(2520)

* В скобках указано число статей, на которых установлена данная цифра.

рых, слабо выраженный максимум для ученых обеих групп приходится на 40—49 лет. Что касается первого факта, то нельзя, по-видимому, говорить о том, что он установлен окончательно: судя по указанным в скобках цифрам, число сравниваемых с лауреатами ученых равно (или порядка) числу лауреатов, т. е. составляет примерно 80 человек. Трудно сказать, в какой степени такая малая выборка ученых репрезентативна.

Что же касается второго факта, то здесь Г. Цукерман и Р. Мертон дают объяснение, на наш взгляд, столь же произвольное, как и приведенные выше рассуждения Прайса о причинах роста соавторства. «Молодые ученые, — пишут они, — которые вообще, как правило, сотрудничают с менторами, обыкновенно работают вместе только с одним из них в какое-то данное время. Кроме того, молодые ученые, иногда под нажимом со стороны своих наставников, мотивированы на самостоятельные работы и публикации собственных статей, позволяющих заявить о своей личности. Совсем не парадоксально, что мотивация начинающих ученых на публикацию подписанных своим именем статей может только усиливаться в результате большого увеличения числа статей, подписываемых несколькими, а иногда большим числом «авторов». Индивидуальный вклад отдельного ученого теряется в толпе ученых, подписывающих статьи своими именами, а это, как они знают, особенно вредит молодым ученым, не имеющим самостоятельных опубликованных работ, свидетельствующих об их способностях. Незначительная доля статей, написанных в результате сотрудничества нескольких авторов и опубликованных на ранних этапах карьеры, объясняется частично действием развившейся в науке системы вознаграждений» [0.117]. «Уменьшение к концу карьеры числа

опубликованных совместно исследований может означать, что сотрудничающие стороны не являются более полезными друг другу. Именно на этом этапе ученые часто обращаются к более широкому «философскому» или «социологическому» темат, не оставляющим места для сотрудничества» [0.117].

Несостоятельность этих доводов — в их бездоказательности. С такой же легкостью могли бы быть сформулированы доводы в пользу *минимума* возрастных кривых соавторства, если бы таковой наблюдался. Например, наблюдавшееся бы в этом случае большое число соавторств в молодые годы объяснялось бы потребностью молодого ученого публиковаться — молодые ученые испытывают при публикации своих работ определенные трудности и фамилия известного ученого на статье зачастую служит пропуском в солидное научное издание. Наблюдавшееся бы в этом случае возрастание соавторств к концу жизни можно было объяснить уменьшением творческих возможностей ученого и его желанием сохранить число публикуемых им статей на прежнем уровне.

Здесь снова возникает потребность в многофакторном анализе явлений науки, в синтезе наукометрических и других науковедческих методов анализа. Социологические и другие данные позволили бы выявить факторы, детерминирующие тот или другой наукометрический феномен, обнажить «игру» этих факторов, вскрыть действующие здесь внутренние закономерности.

Следует заметить, что Цукерман и Мертон дают объяснение причин роста соавторства иное, чем Прайс. Последний, как мы помним, объяснял это явление опережающим ростом числа всех ученых по сравнению с ростом числа выдающихся ученых. Цукерман и Мертон же пишут следующее: «Поскольку первые годы получения образования в науке предусматривают взаимодействие между возрастными группами, то и в последующем это взаимодействие приобретает развитую форму в ходе сотрудничества при осуществлении исследования. Эта форма социализации приобретает растущее значение, так как социальная организация научного исследования значительно изменяется, когда вопросы сотрудничества и создания исследовательских групп все больше ставятся на повестку дня» [0.117]. Нам представляется, что фактор, о котором говорят здесь Цукерман и Мертон, существенно фактора, о котором говорит Прайс. Однако объективное соотношение этих факторов можно установить с помощью социологических и других науковедческих методов анализа.

Распределение запросов на информационные документы в зависимости от года их публикации

Падение числа запросов на литературу с увеличением возраста последней называется старением литературы. Это явление изучают, например, Е. С. Кашафутдинова и соавт. [2.17] (см. рис. 1.27).

3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: СТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

В этом параграфе рассматриваются эмпирические данные, касающиеся статистических распределений, не содержащих времени. Материал структурирован по измерителям, причем при наличии двух измерителей предпочтение отдается более «дробному».

Распределение числа открытий

Дж. Бен-Давид [2.37] изучает распределение числа открытий по странам.

Распределение числа журналов

Распределение числа журналов по странам, в которых они издаются, дают Ч. Готтшпалк и У. Десмонд [2.46], Д. Прайс [1.58], М. Карпентер, Ф. Нарин [2.39] и др. Это распределение, как видно из рис. 2.8, может быть аппроксимировано цифровым распределением.

Распределение числа премий

Статистику премий изучают Г. Цукерман [2.73], К. М. Варшавский и Н. П. Панаева [2.6], Г. Цукерман и Р. Мертон [0.117], Ю. В. Грановский и соавт. [1.16] и др. Известны распределения Нобелевских премий по городам и странам [1.16], Ленинских премий — по уровню квалификации и должностям лауреатов [2.6].

Изучая распределение американских нобелевских лауреатов по университетам, в которых они обучались, Г. Цукерман установила, что из 82 американцев-лауреатов 58 получили образование в 16 университетах, а половина — всего в 4 [2.73]. Оказалось, также, что половина Нобелевских лауреатов работала, будучи молодыми учеными, под началом Нобелевских лауреатов же. В работе Г. Цукерман и Р. Мертона содержатся такие скорректированные цифры: 42 из 83 американских лауреатов работали под началом в общей сложности 54 старших лауреатов [0.117]. В США насчитывается около 800 тыс. ученых и то, что половина будущих но-

белевских лауреатов прошла школу именно у Нобелевских лауреатов, говорит о необычной *концентрации интеллектов* в современной науке. Естественно предположить, что аналогичная ситуация имеет место на любом уровне, не только на уровне Нобелевских лауреатов.

Распределение числа ученых

Интересные сведения о распределении ученых, достигших определенных успехов в науке, содержатся в работе Дж. Коула [1.33]: оказывается, что свыше 40% всех физиков — докторов наук в США выпускается всего девятью ведущими американскими университетами. Это явление также говорит о большой *концентрации интеллектов* в науке (см. обсуждение статистики премий).

Распределения ученых по возрасту анализируются в работах Г. М. Доброва [0.22], Я. Влахи [3.100], А. А. Савельева [2.31], Г. Г. Шляхина и Г. В. Пилиди [1.69], К. Варшавского [2.5], М. В. Василенко и В. И. Клименюка [2.7], Т. З. Козловой и А. Г. Кузнецова [2.18], В. Н. Клименюка [1.27] и др. (см., например, рис. 1.26). Отмечаемое в ряде работ (см. например, [1.27]) постарение научных кадров может быть связано с замедлением темпов экстенсивного роста науки. Некоторые авторы предлагают различные меры по их омоложению. Например, при продвижении на посты руководителей отделов (секторов) отдавать предпочтение кандидатам наук и ввести при этом возрастной ценз [2.31]. Понятно, что такая голая «наукометризация» кадровой научной политики только повредит делу, здесь необходим всесторонний многофакторный анализ кандидатов на должности, в котором, разумеется, должны учитываться и возрастные данные.

Распределения числа ученых по научным отраслям рассматриваются в работах Я. Влахи [2.68], С. А. Кугеля [2.22], В. С. Соминского и Э. М. Торфа [2.33; 1.49; 1.154], И. И. Набокова [2.28], М. В. Василенко и В. И. Клименюка [2.7], В. Д. Осетрова [2.29] и др. В качестве единиц счета выступают как число всех ученых, так и число ученых разных категорий — магистры, кандидаты, доктора наук и пр. Эти распределения предполагается использовать в качестве исходной информации для перераспределения средств и усилий в целях увеличения темпов роста ведущих направлений науки [см., например, 2.22].

Территориальное распределение ученых встречаем в работах Дж. Саржента [2.59], Г. Г. Шляхина и Г. В. Пилиди [1.69], М. В. Василенко и В. И. Клименюка [2.7], В. И. Дуженкова [2.12], Х. Инхэйбера [2.49] и др. Распределение ученых по стра-

нам (см. рис. 2.9) или по регионам (см. рис. 2.10) имеет вид распределения Ципфа. В распределениях ученых по городам для отдельных стран самая высокая концентрация отмечается для СССР и Японии, самая низкая — для США [2.49].

Распределение ученых по уровню их квалификации обсуждают Ф. Гальтон [1.102], Д. Прайс [1.58], Г. М. Добров [0.22] и др. Ф. Гальтон определял число выдающихся ученых на 1 млн. населения. При этом к числу выдающихся он относил ученых, которые были не только членами Королевского общества (речь шла о Великобритании), но либо имели кафедру в университете, либо были награждены медалью научного общества, либо же занимали в таком обществе официальное положение, либо были членами избранного научного клуба академического ранга. Гальтон установил, что на 1 млн населения приходится 5—10 великих ученых. Д. Прайс [1.58, с. 314], опираясь на данные Вишера 1947 г., получил на 1 млн населения США примерно 11—12,4 выдающихся ученых. «Обе цифры, — пишет Прайс, — величинны того же порядка, что и у Гальтона». Ясно, однако, что число выдающихся ученых определяется жесткостью используемых критериев. Эти критерии для Великобритании XIX в. и США XX в. не могли быть одинаковыми и, более того, сравнимыми. Поэтому совпадение цифр Гальтона и Вишера (Прайса) следует считать случайным.

Заметим еще по этому поводу, что непосредственных воспроизводимых измерителей квалификации ученых пока нет. Существующая сегодня квалификационная градация ученых на бакалавров, магистров, докторов, кандидатов наук и т. д. обеспечивается работой соответствующих экспертных комиссий, передвигающих ученых с одной квалификационной ступеньки на другую. Измерение комиссией уровня квалификации ученого невоспроизводимо, т. е. разные комиссии будут, вообще говоря, разного мнения об одном и том же ученом. Но, будучи зафиксировано соответствующими протоколами и дипломами, решение экспертной комиссии становится воспроизводимым в науковедческом смысле, т. е. разные исследователи науки прочтут решение комиссии одинаково. Только поэтому квалификационные распределения и могут рассматриваться как наукометрические. Такая ситуация часто встречается в наукометрии.

В литературе встречаются распределения ученых: по языкам, которыми они владеют [0.22, с. 154—155; 2.34]; по числу предметных областей, в которых они работали в течение жизни [2.36]; по принадлежности к основному и вспомогательному научному персоналу [0.22, с. 100—102]; по должности [1.63] и др. Несколько особняком стоят исследования возможной связи

между сиротством и научной гениальностью. Здесь идет речь о распределении выдающихся ученых по двум классам. К первому относятся ученые, рано потерявшие родителей, ко второму — не испытывавшие такой утраты. У. Уодвард [2.71] корректнее других анализирует этот вопрос. В отличие от своих предшественников, ограничивавшихся рассмотрением только самых выдающихся ученых, Уодвард сравнивает их со всей человеческой популяцией. По его данным, в генетической популяции в целом в 1933 г. было 18,8% сирот, т. е. людей, потерявших по крайней мере одного из родителей до 15 лет. Среди крупных ученых XX в. число сирот составляет 22,9%. Поэтому, говорит Уодвард, гипотеза, согласно которой среди больших ученых больше сирот, чем среди остального населения, «незащитима для двадцатого столетия». Однако остается фактом, что «большой процент ученых в прежние времена терял одного из родителей в детстве и дальнейшие исследования этого важного личностного фактора вполне оправданы» [2.71]. По-видимому, в будущих исследованиях влияния сиротства на научные способности следует изучить статистику сиротства в зависимости от ранга ученого.

Распределение числа соавторов

Распределение авторов по числу соавторов рассматривают в своих работах Д. Прайс и Д. Бивер [2.30], Ю. П. Адлер и соавт. [1.1.], Г. М. Добров и Л. А. Игнатов [2.10] и др. Как показано на рис. 2.11, эти распределения также могут быть аппроксимированы цифровыми распределениями.

Распределение числа авторов на статью по отраслям науки анализируют В. Берельсон [2.38], С. Кинан и Р. Ачертон [2.51], Г. Цукерман [2.72], Л. А. Игнатов [2.14], Г. Цукерман и Р. Мерстон [0.117] и др. В этих работах дается распределение по научным дисциплинам публикаций с одним автором (см., например, рис. 2.12), статей, написанных несколькими авторами, по годам (см. табл. 2.2.1), среднего для данной дисциплины числа соавторов на работу [см., например, 2.14] и т. д.

М. К. Лучник и В. Д. Осетров [1.39] установили средние числа соавторов на одну работу в США и СССР и нашли их примерно одинаковыми. Г. М. Добров и Л. А. Игнатов [2.10] выявили, что число соавторов на одно изобретение практически не влияет на его внедрение (среднее число авторов для внедренных изобретений составляет в 1970 г. 3,2, для невнедренных — 3,1).

Здесь же мы упомянем работу Г. Цукерман [2.74], в которой сравнивается распространенность алфавитного порядка расположения имен авторов на титуле публикации среди лауреатов Нобе-

левской премии и работ нелауреатов. Оказалось, что алфавитный порядок принят в 39% работ лауреатов и в 46% нелауреатов, что лауреаты реже, чем нелауреаты, ставят свою фамилию первой (27 против 39% статей). Этот вопрос изучает и Э. Радд [2.58].

Распределение по числу информационных запросов

Распределение журналов по числу запросов на них в библиотеках и информационных центрах анализируют Д. Уркварт [2.67], Д. Прайс [1.58], Г. М. Добров и соавт. [0.25], С. Наранан [1.131], Е. С. Кашафутдинова и соавт. [2.16; 2.17] и др. Такое распределение показано на рис. 2.13, 2.14. Это эмпирическое распределение носит название распределения Уркварта.

Распределение инженеров по числу запросов ими научной литературы изучает П. Коул [1.89].

Распределение спроса на литературу по научным направлениям М. А. Берг [3.3] использует для оценки сравнительной актуальности этих направлений. Анализа степени соответствия данного индикатора данному латентному параметру Берг, однако, не дает.

Распределение заказов на журналы в зависимости от языка публикации находим в работах Е. С. Кашафутдиновой [2.16, 2.17]. Это распределение, показанное на рис. 2.15, может быть аппроксимировано распределением Ципфа.

Распределение научных публикаций по объему

Распределение ученых по суммарному объему публикаций дается в работе Е. Д. Гражданникова и А. М. Щербакова [1.13] (см. рис. 2.16). Распределение среднего объема статьи по странам и научным дисциплинам дают З. Б. Барнинова и соавт. [4.3] (табл. 2.3.1).

Г. М. Добров [0.22] фиксирует больший объем диссертаций по точным наукам, чем по историческим. Этот наукометрический феномен он объясняет следующим образом: «В науках, имеющих относительно лучше функционирующий информационный сервис (физико-математические, технические), в какой-то мере удается обеспечивать оптимально достаточный объем необходимой исследователю научной информации, сохраняя или даже уменьшая общий объем научных работ. В научных же дисциплинах с плохим информационным обслуживанием трудно обеспечить оптимальную полноту действительно необходимой информации; исследователи невольно вынуждены компенсировать ее менее концентрированными материалами и сведениями, а это неизбежно приводит к ясно

Таблица 2.3.1

Распределение объемов статей по странам и научным дисциплинам, усл. ед. [4.3]

Дисциплины	СССР	США	Великобритания	ФРГ
Молекулярная спектроскопия	1	1,6	1,3	—
Физическая химия	1	1,8	—	1,4
Металловедение	1	2,6	1,7	2,1
Аналитическая химия	1	1,7	—	—

наметившейся тенденции дальнейшего увеличения и без того больших объемов диссертационных работ» [0.22, с. 149].

Такое объяснение нам представляется в известной степени произвольным. Во-первых, неточно говорить о «статистически складывающихся изменениях объемов диссертаций» [0.22, с. 147], так как в нашей стране, к которой относятся данные Г. М. Доброва, эти объемы складываются не вполне *статистически*, т. е. случайно, но имеет место детерминирование объемов диссертаций централизованным органом — ВАК. Во-вторых, мыслимы и другие механизмы, обеспечивающие больший объем диссертаций в исторических науках — меньшая формализованность языка, необходимость цитирования источников и т. д. А есть факторы, влияющие в другом направлении. Скажем, известно (см. гл. 4, § 3), что в гуманитарных науках число ссылок на статью меньше, чем в точных, т. е. объем литературы, перерабатываемой автором диссертации, здесь, по-видимому, меньше, что, конечно, должно влиять на объем диссертаций в сторону ее уменьшения, и т. д.

Опять мы приходим к выводу о необходимости многофакторного анализа наукометрических феноменов, который только и может вскрыть «игру» всей совокупности факторов, детерминирующих эти феномены.

Распределение временных лагов (задержек) публикаций

Эти распределения изучаются в работах З. Б. Барниновой и соавт. [4.3], В. В. Налимова и З. М. Мульченко [1.46], Г. М. Доброва и соавт. [0.25], У. Гарвея и соавт. [2.44], К. Субрамания и Дж. Шаффера [2.65] и др. Здесь изучаются распределения по времени задержки отдельных журналов [см., например, 0.8; 2.65], по отраслям науки журнальных публикаций (так, например, в работе [2.44] сравниваются в этом отношении социальные и физические науки) и отдельным странам (см., например, [1.46], с. 153—160).

С. С. Бацанов и Ф. А. Брусенцов [2.2] при статистическом анализе кристаллоструктурных работ дают распределения по некоторым, характеризующим исследования в этой области, индикаторам.

4. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: СТРУКТУРНЫЕ ИССЛЕДОВАНИЯ

Количественный структурный анализ объектов предполагает, прежде всего, измерение связанности двух любых элементов данного объекта. Если мы измерим взаимосвязь всех рассматриваемых элементов объекта, то, используя, скажем, методы факторного анализа, мы сможем восстановить и всю структуру этого объекта. Таким образом, измерение взаимосвязи двух частей объекта науковедческого анализа — это «элементарный акт» количественного анализа структуры этого объекта.

Разные индикаторы науки обеспечивают разные способы измерения взаимосвязи частей объектов науки. Далее, как и везде в настоящем исследовании, материал структурирован по индикаторам науки.

Структурный анализ предметного поля науки (индикатор — число открытий).

Опираясь на книгу П. Сорокина [2.63], который распределил 12761 научное открытие и изобретение по 9 научным дисциплинам, Д. Симонто [2.61; 2.62] соотнес все эти открытия и изобретения с соответствующими каждому из них временными интервалами. 27 веков были разбиты на 44 временных отрезка разной длины: от 800 до н. э. до 1000 г. н. э. каждый отрезок был равен 100 годам, от 1000 до 1500 г. — 50 годам и с 1500 по 1900 г. — 25 годам. Определяя корреляцию временных распределений, соответствующих разным научным дисциплинам, Симонто ищет связь между успехами в разных областях науки. Он дает, в частности, таблицу корреляций, показывающую, как влияют успехи в одних научных дисциплинах в g -й временной интервал на успехи в других в $(g+1)$ -временной интервал (табл. 2.4.1). В табл. 2.4.1 отражен период с 1500 по 1900 гг., разбитый на 16 интервалов по 25 лет каждый. Десятичные запятые в этой таблице опущены, так что, например, —74 означает —0,74.

На основании рассмотрения своих таблиц Симонто делает общий вывод, что научные открытия и изобретения «имеют тенденцию группироваться по крайней мере в три независимые сферы и не представляют собой единого массива» [2.61]: 1) физика, химия и биология; 2) астрономия и математика; 3) техника,

Таблица 2.4.1

Корреляции научных дисциплин и военных мер [2.62]

Поколение g	Поколение $g+1$										
	1	2	3	4	5	6	7	8	9	10	11
1. Математика	31	01	-53	-44	-17	-64	-43	-69	-46	-18	44
2. Астрономия	46	46	-49	-81	-15	02	-66	-49	-19	05	34
3. Биология	-62	42	33	-07	-11	25	-23	37	42	-16	-34
4. Медицина	-62	20	54	05	-22	02	-01	38	38	-12	-23
5. Химия	-39	02	52	30	32	34	59	55	52	46	08
6. Физика	-55	-20	38	64	47	34	59	60	52	-07	-08
7. Геология	-60	18	86	51	-01	26	34	68	50	16	-45
8. Техника	-74	04	62	46	42	62	51	85	82	14	-34
9. География	-63	02	43	32	48	63	56	72	71	37	-22
10. Длительность войн	-04	39	52	42	11	28	20	33	24	04	-43
11. Военные людские потери	32	11	-07	16	01	-10	14	-15	-35	12	-30

география и геология. Медицина примыкает к первой и второй сферам.

Положительное влияние результатов научных исследований одного 25-летнего периода на успехи следующего Симонто объясняет «переносом информации» [2.62, с. 60]. Отрицательное влияние, например, развития астрономии на развитие биологии, медицины, геологии и техники и т. д. Симонто объясняет «конкурентной борьбой за кадры» и ассигнования. Ясно, что здесь мы также сталкиваемся с необходимостью многофакторного анализа наукометрических феноменов и недостаточностью простого их качественного истолкования.

Одновременно Симонто исследует влияние «военной активности» на развитие отдельных научных дисциплин в течение текущего и последующего 25-летнего периода. Корреляции между показателями военной активности (продолжительностью войн и боевыми людскими потерями воюющих сторон) и числом научных открытий и изобретений в последующее 25-летие он не нашел (см. табл. 2.4.1). «Единственное влияние, — пишет Симонто, — которое оказывает война на научные усилия, выражается в том, что она тормозит открытие и изобретение лекарственных препаратов, раз-

работку хирургических методик, диагностических приемов и т. д.» [2.62]. Причиной отрицательного влияния войн на развитие медицины, возможно, являются, по мнению Симонтона, отток кадров и переориентация медиков на практическую помощь в ущерб исследовательской деятельности.

Работы Симонтона являются, на наш взгляд, сильными анти-милитаристскими доводами, поскольку они разрушают имеющиеся в определенных кругах представления о якобы позитивном влиянии войн на развитие науки.

Структурный анализ соавторских коллективов

Для измерения степени связанности двух ученых используется число подписанных ими совместно работ. В этом направлении работают Д. Прайс и Д. Бивер [2.30], Ю. П. Адлер и соавт. [1.1], В. В. Максимов [2.24], В. В. Налимов и З. М. Мульченко [1.46], М. Е. Голостенов [2.8], Ю. В. Грановский и соавт. [1.17], З. М. Мульченко и соавт. [1.43] и др. Типичный граф соавторства показан на рис. 3.1.

Первым и наиболее интересным было исследование Д. Прайса и Д. Бивера [2.30]. Авторы вскрывают внутреннюю структуру исследуемого ими конкретного *невидимого колледжа* — группы обмена информацией № 1 (ГОИ) по окислительной фосфорилиции и направленному переносу электронов, которая была организована Отделом грантов при Национальном институте здравоохранения (США). Исследование велось на материале 533 работ, разосланных членам группы за 1961—1965 гг.

Интересен путь, каким Прайс и Бивер приходят к выводу о существовании в данном незримом колледже стабильного ядра, окруженного мобильной оболочкой. Они приводят матрицу «продуктивность (измеряемая как число работ на одного автора) × × число соавторов на автора», элементами которой являются числа авторов, имеющие данную продуктивность и данное число соавторов в совокупности, т. е. по всем работам исследуемого массива. Оказалось, что продуктивность и сотрудничество (понимаемое в терминах соавторства) коррелируют. «Наиболее продуктивные авторы, — пишут Прайс и Бивер, — и сотрудничают энергично. Три из четырех следующих по продуктивности авторов представлены также и в списке наиболее часто сотрудничающих. Ясно, конечно, что если множественное авторство является правилом, то человек увеличивает число сотрудников-соавторов с написанием каждой новой работы, но с учетом этого обстоятельства представляется малым разброс от основной диагонали матрицы. Это означает, что для всего спектра продуктивности почти не наблюдается отходов

от правила вовлекать примерно одного нового соавтора в написание каждой новой работы. Иными словами, здесь не видно людей, которые обнаруживали бы тенденцию создавать жестко оформленные авторские группы или тенденции избегать прежних соавторов-коллег. Особо следует отметить, что никто из тех, кто писал без соавторов или только с одним соавтором, не преуспел в написании более четырех работ за 5-летний период, тогда как каждый, в активе которого числится более 12 сотрудников-соавторов, представил за тот же период 14 и более работ» [2.30].

«Естественным объяснением этого явления, — продолжают Прайс и Бивер, — может быть представление о чрезвычайно активном ядре исследователей, которые окружены большой и изменчивой массой людей, сотрудничающих с исследователями в одной-двух работах, а затем исчезающих и уже не подающих о себе знать снова» [2.30].

При изучении этого явления «набор карточек с индексами авторов был вручную разложен таким способом, чтобы рядом с каждым автором располагались его соавторы, а с последними те, кто в свою очередь соавторствует с ними и т. д.». Таким образом, оказалось возможным выделить пять наиболее крупных групп, в которые «входит примерно треть общей популяции авторов, причем каждая группа содержит одного или нескольких индивидов, удерживающих в ней рекорд как по продуктивности, так и по числу сотрудничающих с ними авторов. Каждая группа, как правило, включает небольшое число очень активных авторов и большое число авторов, которые сотрудничают в малом числе статей» [2.30].

Далее Прайс и Бивер соотносят авторов с их адресами. Оказалось, что каждая группа привязана к институту или месту работы своих ведущих членов, хотя и не ограничена локальными связями. «Наиболее крупная группа, — пишут авторы, — концентрируется вокруг самого большого исследовательского института в данной отрасли, руководителем которого является высший и по продуктивности и по объему сотрудничества член данной группы... основатель своего механизма группы...» [2.30]. Это обстоятельство, как о том говорилось в § 2 настоящей главы, позволяет сделать вывод, что структура соавторских коллективов в какой-то степени отражает структуру научных коллективов, причем отражает явным образом. В этом и состоит основное значение феномена соавторства для исследователей науки.

Структурный анализ системы научных организаций с числом библиографических заказов в качестве индикатора науки.

А. А. Коренной и О. Г. Левченко [2.20] изучали распределение библиографических запросов по одной из проблем охраны сре-

ды — 733 запроса из 330 организаций на 284 отчета. Сила связи между двумя организациями — авторами отчетов определялась по числу общих отчетов в списке, сила связи между организациями, присылающими запросы, — по наличию общих запросов. Авторами была построена сеть связей.

5. ПЕРСПЕКТИВЫ

Оценивая состояние статистического метода, следует сказать, что он находится пока в самом начале своего развития. Это подтверждают и данные табл. 2.5.1, показывающей, сколько наукометрических работ, попавших в наше поле зрения в настоящей главе, используют тот или другой наукометрический индикатор. Индикаторы 1—10 — наукометрические, 11, 12 включены сюда для общности рассмотрения, остальные — ненаукометрические. Диагональ таблицы, за исключением пересечения 10-й строки и 10-го столбца, соответствует структурным исследованиям, 11-й и 12-й столбцы — исследованиям нестационарных распределений науки, остальная часть таблицы отвечает исследованиям стационарных распределений. Табл. 2.5.1 позволяет получить качественное

Таблица 2.5.1

Число упомянутых в настоящей главе работ, использующих отдельные наукометрические индикаторы

Индикатор	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Число открытий	2																
2 Число журналов							5		2		5	5		1			
3 Число организаций				2	1						8			2			
4 Число премий																	
5 Число ученых						3	1				17	8	9	4	3	1	2
6 Частота соавторства						7					10	1	5	1			
7 Число библиографических запросов							1						1	1			2
8 Объем публикаций																	
9 Временной лаг													1	1			
10 «Экзотические»										1	7		1	1			
11 Время												2					
12 Возраст																	
13 Предметные																	
14 Географические																	
15 Квалификационные																	
16 Должностные																	
17 Языковые																	

представление о состоянии разработок отдельных индикаторов. Табл. 2.5.1 заполнена хаотически и неполно, что свидетельствует о необходимости исследований по разработке системы наукометрических индикаторов, охватываемых статистическим методом.

Следует также обратить внимание на то, что в освещенных в настоящей главе исследованиях практически не использовался аппарат ошибок. До сих пор, пока велась «разведка» наукометрических объектов анализа, это было допустимо. Сейчас, когда пришло время систематического их исследования, игнорирование аппарата ошибок уже недопустимо. Как говорилось в гл. 1, этот аппарат должен быть основан на негауссовой математической статистике.

Перспективы применения статистического метода определяются результатами опорных исследований отвечающих ему наукометрических индикаторов. Как отмечалось в § 1, измерители, на которых базируется статистический метод, могут служить индикаторами *тенденций коллективности* научного труда (число соавторств), *перспективности* научных направлений (число ученых равных категорий, спрос на литературу и др.). «Часто новые быстро растущие направления», пишут В. В. Налимов и З. М. Мульченко, — долго остаются незамеченными, так как они тонут в массе работ старого, но теперь уже медленно растущего направления. Кривые роста легко позволяют выявить потенциально интересные направления» [1.46, с. 186]. При этом, чем выше скорость роста характеризующих данное направление наукометрических индикаторов, тем актуальнее и перспективнее считается это направление. Для оценки *перспективности* или *проблемности* научных направлений могут использоваться также стационарные распределения — более перспективные направления, согласно В. И. Горьковой [1.11], характеризуются распределениями, имеющими более короткие «хвосты». Поскольку стационарные распределения научной деятельности диффовы (об этом речь пойдет в гл. 7), то длину их «хвостов» можно характеризовать показателем негауссовости диффового распределения α .

Таким образом, статистический метод имеет широкое поле применения в исследованиях научной деятельности, следует только помнить о статистическом характере всех даваемых им оценок. Однако практическое использование результатов применения статистического метода должно предварять опорные исследования лежащих в его основе наукометрических индикаторов. Увязка этих индикаторов с латентными переменными до сих пор производится на основе умозрительных наблюдений, в связи с чем отсутствует уверенность в том, что трактовать получаемые с помощью статистического метода результаты следует именно так, как их трактуют авторы наукометрических исследований.

Глава третья

МЕТОД ПОДСЧЕТА ЧИСЛА ПУБЛИКАЦИЙ

1. СОДЕРЖАНИЕ МЕТОДА И ОПОРНЫЕ ИССЛЕДОВАНИЯ

Измеритель (индикатор), определяющий рассматриваемый в настоящей главе наукометрический метод, — это число научных продуктов, под которыми здесь понимаются книги, статьи, отчеты и т. д. Иногда при расчете общего объема научной продукции ее типы не различаются, иногда же книги, оригинальные статьи, обзорные статьи и прочее берутся с разными весами. Различные способы «взвешивания» публикаций (так для упрощения изложения мы будем иногда называть научные продукты, которые, конечно, не всегда бывают опубликованными) дают Г. Мэрис [3.76], Л. Мельцер [3.77], Д. Крэйн [3.51], А. Байер и Дж. Фольгер [4.38], А. Картер [3.47], А. Рое [3.91], П. С. Ильиницкий и В. И. Михайленко [3.12] и др. Так, Д. Крэйн приравнивает одну книгу четырем статьям, А. Картер — шести, Г. Морис и Л. Мельцер — восемнадцати и т. д. «Взвешиванию» могут подвергаться и отдельные публикации одного типа. Формальной основой такого «взвешивания» может служить, например, ранг издания, в котором опубликована данная работа, определяемый, например, средствами цитатного или социологического анализа.

По-разному может учитываться и соавторство. Во-первых, можно за публикацию с n авторами каждому из них начислять $1/n$ публикации. В этом случае принято говорить о *частичной продуктивности автора*. А можно и начислять одну публикацию независимо от числа авторов. В этом случае принято говорить о *полной продуктивности автора*. Эти индикаторы довольно тесно коррелируют между собой. Например, согласно данным Б. Шоу [3.92], корреляция между распределениями ученых по полной и частичной продуктивности составляет 0,71.

Научная публикация — продукт труда ученого. Используя его для количественного анализа деятельности человека науки, наукометрия только следует общим традициям научного анализа деятельности человека.

Ежегодно публикуется около 3 млн. журнальных статей только по естественным наукам и технике, 70–75 тыс. книг, 450 000 описаний к патентам и авторских свидетельств [0.59], так что статистические совокупности, возникающие при использовании в качестве индикатора науки числа публикаций, достаточно велики для исследования целого ряда проблем. И здесь вопрос заключается только в определении круга задач, при решении которых адекватен обсуждаемый метод, какую информацию из этого можно получить и какая статистика потребна для извлечения этой информации в каждом конкретном случае, т. е. сводится к соотношению числа публикаций с другими индикаторами и латентными переменными науки. Другими словами, и здесь речь идет об *опорных исследованиях* данного индикатора.

Конечно, всякое многофакторное исследование фиксированного объекта науки — это еще и опорное исследование с установлением корреляций между распределениями, соответствующими разным индикаторам науки. Поэтому и говорилось в гл. 2 § 3, что переход в процессе исследования науки ко все более цельной и согласованной ее картине означает и переход ко все более адекватной системе латентных параметров, т. е. прогресс в опорных исследованиях. В связи с этим, помимо рассмотренных в настоящем параграфе работ, отнесенных нами к опорным, в какой-то степени опорными являются и исследования, приводимые в заключении.

Опорные исследования числа публикаций как индикатора науки выполняли У. Деннис [3.52], Л. Мельцер [3.77], А. Грасберг [3.59], Д. Пельц и Ф. Эндрюс [0.51], С. Коул и Дж. Коул [4.62], В. Шоу [3.92], У. Менард [1.125], П. Пиганоль [3.82], У. Хагстром [4.92], Ш. Эдвардс и М. Маккэри [1.96] и др.

Корреляцию между распределениями ученых по числу публикаций, авторами которых они являлись, и мерами признания этих ученых нашли У. Деннис, С. и Дж. Коулы и У. Менард. У. Деннис [3.52] в качестве меры признания ученого брал членство в Национальной академии наук, упоминание в Британской энциклопедии и т. д. Оказалось, что половина из 10% наиболее продуктивных авторов являются членами Национальной академии и упоминаются в Британской энциклопедии. При ограничении выборки пятью процентами подобное признание имели уже 70%. И, наоборот, ни один человек в выборке У. Денниса, бывший членом Национальной академии или упоминавшийся в Британской энциклопедии, не имел менее 27 публикаций; а среднее количество работ — порядка 200.

С. и Дж. Коулы приводят корреляции распределений ученых по числу работ и мерам их признания (табл. 3.1.1). В этой табли-

Таблица 3.1.1

Корреляция между числом работ ученого-физика и мерами его признания [4.62]

Наукометрические индикаторы	Мера признания			Доля ученых-физиков, которым известны работы данного ученого
	престиж самой высокой награды	число наград	ранг учреждения	
Число работ	0,35	0,46	0,24	0,49
Всего в год	0,28	0,32	0,19	0,43
Качество работ * (число ссылок)	0,41	0,67	0,33	0,64

* Обсуждается в гл. 4 § 1.

де налицо, хотя и не очень большая, корреляция между числом публикаций и мерами признания ученого.

У. Менард [1.125] на своей выборке геологов также находит, что число публикаций хорошо отражает *известность* ученого. В его списке из 50 наиболее продуктивных геологов 14 — члены Национальной академии наук (США), 44 были президентами либо медалистами Американского геологического общества, а 35 были удостоены одним из этих видов признания.

Корреляцию числа публикаций с *вкладом в науку*, определяемым методом экспертной оценки, находят Д. Пельц и Ф. Эндрюс [0.51]. Эта корреляция варьировалась от 0,15 до 0,39.

А. Грасберг [3.59] находит корреляцию между ранжировками ученых по числу публикаций и *общей продуктивности*, определяемой с помощью экспертов, равной 0,74. Его выборку составили 60 известных ученых, работающих в большой индустриальной исследовательской лаборатории. Следует сказать, что эта корреляция подозрительно высока и объясняется, возможно, подбором крупных ученых. С. и Дж. Коулы, Д. Пельц и Ф. Эндрюс, чьи выборки не были связаны с какими-либо априорными соображениями относительно известности ученых, получили более скромные значения аналогичной корреляции.

У. Хагстрем нашел значимую корреляцию определяемого экспертным путем *престижа* научного учреждения с числом выпускаемых в его стенах работ [4.92].

В литературе зафиксирован также ряд, так сказать, опорных соображений качественного порядка. Д. Прайс связывает число публикаций ученого с его *элитностью* [1.58], А. И. Яблонский — с *активностью*, *интенсивностью* работы. [0.63]. По мнению

Ю. В. Грановского и соавт. [1.18], большое число монографий в данном научном направлении свидетельствует о его *насыщенности*, а журнальных статей — о его *конфликтности*. Дж. Донохью [1.95], Л. Фигуайрадо [2.43], О. С. Радбильт и И. Л. Турькина [1.60] и другие определяют *ценности* журналов по частотам отражения их публикаций во вторичных изданиях, т. е. по числам соответствующих каждому журналу рефератов публикаций в реферативных изданиях.

Скорость роста числа публикаций — индикатор науки, производный от числа публикаций. Пока его связывают с латентными переменными науками, по-видимому, только умозрительно. Г. М. Добров [0.22, с. 64], Г. Н. Аптекман [3.2], А. Д. Кордо и В. С. Гарышкин [3.17], Н. В. Шершова [3.36], Н. С. Гаврилов и соавт. [1.9] и другие связывают скорость роста числа публикаций с *актуальностью* данного научного направления, В. В. Налимов и З. М. Мульченко [1.46], А. А. Васильковский и И. В. Дзиековская [3.5], А. Д. Кордо и В. С. Гарышкин [3.17], В. И. Горькова [3.6], Н. С. Гаврилов и соавт. [1.9], И. В. Дзиековская [3.13], Г. В. Самсонов [3.31] и другие — с его *перспективностью*. Г. М. Добров монотонный рост числа публикаций в данной предметной области связывает с отсутствием коренных *изменений* «в методах и принципах решений данной проблемы, а также в исходных стимулах постановки подобного рода исследований» ([0.22], с. 64). Ю. В. Грановский и его соавторы [1.19] предлагают по скорости роста числа публикаций данного научного коллектива в данный момент, соотносенной с общим видом кривой их роста для этого коллектива, определять фазы развития направления, в котором работает коллектив.

Подчеркнем, что связь числа публикаций как индикатора науки с тем или другим латентным параметром по самой своей природе носит статистический характер. С самого начала и с величайшей готовностью соглашусь, пишет Д. Прайс, что число публикаций — «это очень плохая шкала. В самом деле, кто мог бы осмелиться попытаться уравновесить одну статью Эйнштейна по теории относительности хотя бы сотней статей бакалавра Джона Доу о константах эластичности для различных древесных пород в лесах Нижнего Базутоленда, по одной константе на статью» [1.58, с. 317]. «...Если бы мы располагали данными для истинного исчисления качества, а не данными для приближенных грубых количественных подсчетов... ранги индивидуумов ни в коем случае не оказались бы идентичными по положению в этих двух различных табелях о рангах, но мы можем с уверенностью сказать, что существует значимая корреляция между качественной и количественной элитностью. Поскольку, однако, и к счастью для

нас, мы не берем на себя обязательство предложить качественные мерки для индивидуумов, для нас достаточно знания того, что статистическая картина для научного персонала и литературы будет и в том и другом случае следовать общему закону» [1.58].

По абсолютной величине корреляция числа публикаций с латентными переменными науки, насколько можно судить по имеющимся данным, хотя и значима, но ниже, чем, скажем, корреляция числа ссылок с теми же переменными (см., например, табл. 3.1.1 и § 1 гл. 4). Поэтому применение числа публикаций в качестве индикатора науки эффективно лишь, во-первых, на статистических совокупностях, больших, чем при использовании в качестве индикатора науки, скажем, числа ссылок, и, во-вторых, в сочетании с другими индикаторами науки.

В целом опорные исследования числа публикаций в настоящее время находятся в зачаточном состоянии, поскольку вместо цельной картины корреляций числа публикаций с каждой латентной переменной и каждым индикатором науки мы имеем только отдельные корреляции. Кроме того, в проведенных опорных исследованиях статистическая достоверность получаемых значений этих корреляций или не оценивается вовсе, или оценивается средствами гауссовой математической статистики. В сущности, следует говорить о необходимости заново проводить количественные опорные исследования числа публикаций, поскольку полученные до сих пор результаты имеют лишь качественное значение.

2. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: НЕСТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

Динамика роста числа публикаций

В наше поле зрения попало около 70 работ, в которых рассматривается динамика роста числа публикаций по науке в целом и по отдельным научным направлениям. Перечислять здесь все эти работы не имеет смысла, общее же число нам необходимо для дальнейшего использования в некоторых таблицах.

Наиболее известным результатом исследований динамики роста публикаций является возможность аппроксимации многих кривых роста публикаций экспонентой (см., например, работы Д. Прайса [1.58], Л. Антони [3.39], В. В. Налимова и З. М. Мульченко [1.46], А. А. Васильковского и И. В. Дзиековской [3.5], А. Н. Алексеева [3.1] и др. и рис. 1.6б, 1.6е). Иногда в этих целях используется сумма экспонент (см., например, работы Д. Бэкера [3.40], Л. Нового и Я. Фолты [1.52], У. Менарда [1.125],

В. И. Громина [3.8], Н. Л. Крашенинниковой [3.18], Ан. Н. Несмеянова [1.51] и др., а также рис. 1.6в, г, и, к).

При увеличении числа «колен» эмпирической кривой роста числа публикаций авторы, не желающие отказываться от языка экспонент, говорят о так называемой *скользящей экспоненте* (см., например, [4.22]). Понятно, что такая аппроксимация лишена смысла, если нам не удастся установить вид зависимости параметров этой скользящей экспоненты от положения точки на кривой роста. Если же такая зависимость будет найдена для данной кривой роста, то это будет означать аппроксимацию этой кривой не экспонентой, а более сложной зависимостью.

К линейной аппроксимации эмпирических данных по росту числа публикаций прибегают Б. И. Спасский и Л. В. Заржицкая [3.10], Г. Магяр [1.123], А. И. Михайлов и соавт. [1.42] и др. (см. рис. 1, 6а, д). Как правило, при этом речь идет о коротком временном интервале. При расширении временного интервала для кривых роста часто получают зависимости так называемого *логистического* (об этом речь пойдет в гл. 7, § 1) типа (см., например, рис. 1, 6ж—з, м). В общем же случае кривые роста числа публикаций могут иметь самый разнообразный вид (см., например, рис. 1.18 и 1.6л).

При обсуждении кривой роста числа публикаций по физике лазеров Г. М. Добров пишет: «...мы наблюдаем... бурный „всплеск“ количества работ в 1947—1955 гг. (с периодом удвоения порядка двух лет), а затем — непосредственно после создания первых квантовых генераторов (лазеров) — заметное снижение численности публикаций. Последнее обстоятельство И. Малецкий и Е. Ольшевский объясняют... тем, что в период, предшествовавший созданию лазеров, в международном масштабе параллельно проводилось множество исследований. Можно высказать предположение, что спад последней ветви рассматриваемой кривой объясняется не только снижением количества параллельно ведущихся поисковых работ, но также и заскреживанием результатов по определившейся после 1955 г. особо перспективной проблематике. Возможно также, что на характере этого участка кривой отразилось то обстоятельство, что анализ относящихся к ней данных проводился в непосредственной близости от времени самого явления. В этом случае заново и целенаправленно ориентированный поток информации в научных исследованиях не успел отразиться на статистике» [0.22, с. 64—65]. Чтобы эти и аналогичные предположения относительно факторов, влияющих на наукометрические кривые роста, имели научное значение, их необходимо подтвердить, как и при анализе других наукометрических феноменов, ненаукометрическими исследованиями.

Возрастная динамика числа публикаций

Распределение числа работ с возрастом ученого изучают У. Деннис [3.53], В. Эйдусон [3.55], Д. Пельц и Ф. Эндрюс [0.51], Л. Новый и Я. Фолта [1.52], Г. Цукерман [2.74], Я. Влахи [3.100], Е. Д. Гражданников и А. И. Щербаков [3.7], авторы сборника [1.49], Л. Зусне [3.106] и др.

Мы видели (гл. 2, § 2), что Г. Леман, исследуя возрастную динамику открытий, обнаружил не только пик творческой активности, приходящийся на первую половину творческой деятельности ученого, но и установил, что этот пик тем менее выражен, чем скромнее принимаемые во внимание результаты (см. рис. 1.21а-в). Следует поэтому, ожидать, что *возрастное распределение числа публикаций не будет иметь явно выраженного максимума*. И, действительно, об этом свидетельствуют эмпирические данные, например представленные на рис. 1.22.

В некотором диссонансе с этим выводом как будто бы находятся данные Д. Пельца и Ф. Эндрюса [0.51]. Эти авторы, основываясь на результатах социологического и наукометрического анализа ученых, выделяют в жизни последних два максимума, первый из которых соотносят с максимумом дивергентной деятельности ученого по производству нового знания, второй — с максимумом конвергентной деятельности по упорядочению наличного знания. Не оспаривая социологических результатов Д. Пельца и Ф. Эндрюса, осмелимся утверждать, что их наукометрические данные не позволяют сами по себе говорить о наличии двух максимумов возрастного распределения числа публикаций. В самом деле, рис. 1.23, на котором показаны эти данные для опубликованных статей, мало чем отличается от рис. 1.22 Я. Влахи. Правда, данные по неопубликованным отчетам (см. рис. 1.24) показывают два несильно выраженных максимума, но число неопубликованных отчетов в качестве наукометрического индикатора вызывает определенные сомнения в отношении воспроизводимости результатов. Следует заметить также, что Д. Пельц и Ф. Эндрюс получили свои *наукометрические* данные из *социологических* анкет, что также отнюдь не увеличивает воспроизводимости полученных этими авторами наукометрических данных.

Мы не можем утверждать, что возрастная динамика творчества ученого не имеет двух максимумов, на которых настаивают Д. Пельц и Ф. Эндрюс. Может статься, что такие два максимума и существуют (см. гл. 6, § 2), только из данных по возрастной динамике числа публикаций этого не следует, что, возможно, объясняется недостаточной адекватностью числа публикаций как наукометрического индикатора проблеме возрастной динамики

Таблица 3.2.1

Распределение доли статей, в которых фамилия автора стоит первой, в зависимости от возраста ученых, в % [2.74]

Возраст, лет	Лауреаты Нобелевской премии	Нелауреаты
20-29	49 (103) *	38 (34)
30-39	40 (304)	38 (130)
40-49	26 (578)	56 (80)
50-59	20 (374)	37 (131)
60+	18 (222)	25 (64)
Всего	27 (1581)	39 (439)

* В скобках дается число публикаций, на которых получены табличные значения.

научного творчества. Нас же в данном случае волнует только чистота наукометрического анализа, т. е. чтобы наукометрическим данным не приписывали то, чего из них не следует.

Положительным в исследовании Д. Пельца и Ф. Эндрюса является использование и наукометрических и социологических методов анализа. Правда, здесь эти методы идут еще *рядом*, увеличивая надежность выводов, вместо того чтобы социологический анализ вскрывал природу наукометрических феноменов. Другими словами, мы имеем здесь *сумму* наукометрических и социологических методов вместо их *синтеза*.

Искажение картины возрастной динамики научного творчества при использовании такого индикатора, как число публикаций, объясняется, возможно, тем, что «монотонно возрастающей функцией возраста являются должность, ученая степень и звание. Продвижение же по квалификационно-должностной лестнице изменяет статус, престиж научного работника, вследствие чего повышаются публикационные возможности» [1.49, с. 102]. Впрочем, это тоже из области предположений, решение же вопроса требует многофакторного анализа проблемы.

Идею Г. Лемана о существовании в первой половине творческой жизни ученого пика творческой активности количественно описал Л. Зусне [3.106]. В выборке из 213 известных психологов он усреднил даты первой и последней публикаций ученого и даты самого важного его открытия. Оказалось, что эта средняя дата близка *средней гармонической* первой и последней дат; что подтверждает вывод Г. Лемана. Среднее арифметическое и среднее геометрическое первой и последней дат оказались близки друг другу и далеки от средней даты.

Г. Цукерман [2.74] установила уменьшение с возрастом доли

публикаций, в которых фамилия данного ученого стоит первой в списке авторов (табл. 3.2.1).

Л. Новый и Я. Фолта [1.52] обнаружили тенденцию к понижению среднего возраста математиков, в котором они публикуют свои первые работы (см. рис. 1.25).

Динамика поступления статей в журналы в течение года

Этим вопросом занимался В. Д. Осетров [3.27]. Как и следовало ожидать, минимум в этом распределении приходится на период летних отпусков.

3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: СТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

Распределение числа публикаций по журналам

Распределение числа публикаций на заданную тематику по журналам и журналов по числу запросов на них в библиотеках (см. гл. 2, § 3) — это соответственно входная и выходная характеристики журналов. Первое описывает, как ученые выбирают журналы для своих работ, а второе — как они выбирают журналы для чтения.

Эмпирическое распределение публикаций фиксированного профиля по журналам называют *распределением Брэдфорда* по имени английского ученого. Это распределение часто изучают по ссылкам, которые содержатся в публикациях журналов данного профиля, и потому его правомерно было бы освещать и в следующей главе.

Распределению Брэдфорда посвящено очень большое число работ. Наше внимание привлекли около 60 из них. По-видимому, первым о рассеянии публикации по журналам говорит Ф. Кемпбелл [3.65]. Различные аппроксимации предложили С. Брэдфорд [3.42–3.43], П. Коул [3.50], М. Кендэлл [1.119], Ф. Лаймюлер [3.72–3.73], Б. Брукс [3.44–3.46], С. Нараян [1.130], Е. О'Нил [3.81], А. Т. Мицевич [3.23–3.24], Д. Прайс [1.143], Дж. Хьюберт [3.62]. Обзор аппроксимаций дает Дж. Хьюберт [3.62].

Аппроксимации распределения публикаций по журналам распадутся на две ветви. В первой данные, следуя Брэдфорду, представлены в ранговой интегральной форме (см. рис. 2.17г) и получаемая эмпирическая кривая описывается различными формульными выражениями.

Брэдфорд дал две формулировки своего закона рассеяния публикаций. Первая (так называемая зональная) [3.42; 3.43]: если научные журналы расположить в порядке уменьшения числа помещенных в них статей по какому-либо предмету, то в полученном списке можно выделить ядро журналов, посвященных непосредственно этому предмету, и несколько групп или зон, каждая из которых содержит столько же статей, что и ядро. Тогда числа журналов в ядре и последующих зонах будут относиться как $1:m:m^2\dots$, где m , вообще говоря, меняется от распределения к распределению.

Другая формулировка [3.42; 3.43]:

$$X(r) = a + b \log r, \quad (3.3.1)$$

где r — ранг журнала; $X(r)$ — кумулятивное число статей; a и b — параметры аппроксимации Брэдфорда. Эта аппроксимация соответствует распределению Ципфа с $\alpha=1$, когда эффект рангового искажения мал. Действительно, распределение Ципфа в его ранговой интегральной форме для $\alpha=1$ имеет вид (см. выражение (7.2.13)):

$$X(r) = A \ln \frac{r+B}{1+B} = -A \ln(1+B) + \frac{A}{\log e} \log(r+B).$$

Для $r \ll B$ это дает:

$$X(r) \approx -A \ln(1+B) + \frac{A}{\log e} \log r,$$

что и соответствует аппроксимации Брэдфорда (3.3.1) с

$$a = -A \ln(1+B); \quad b = \frac{A}{\log e}.$$

Для $\alpha=1$, $x_0=1$, $N \gg 1$ и $I \gg 1$ имеем (см. выражение (7.2.15))

$$A = \frac{N-1}{1-\frac{1}{I}} \approx N; \quad B = \frac{N-1}{I-1} - 1 \approx \frac{N}{I} - 1,$$

так что

$$a \approx -N \ln \frac{N}{I}; \quad b \approx \frac{N}{\log e}.$$

Таким образом, параметры аппроксимации Брэдфорда определяются параметрами выборки и не характеризуют, собственно, рассеяние публикаций по журналам.

П. Коул [3.50], как и Брэдфорд, аппроксимирует график, показанный на рис. 2.17г, прямой:

$$\frac{X(r)}{X(N)} = 1 + \operatorname{tg} \psi \log \frac{r}{N}, \quad (3.3.2)$$

где N — число журналов; ψ — угол наклона прямой к оси абсцисс. Эта аппроксимация эквивалентна аппроксимации Брэдфорда (3.3.1) и к ней может быть отнесено все, что говорилось выше о последней.

Ф. Лаймкюлер [3.72–3.73] использует аппроксимацию

$$X(r) = \frac{a \log(1+br)}{\log(1+b)}. \quad (3.3.3)$$

Эта аппроксимация предназначена для описания ситуации, с которой не справляется аппроксимация Брэдфорда и которая наиболее часто встречается на практике. Именно, когда график, показанный на рис. 2.17г, существенно отклоняется в области небольших значений r от прямой, т. е. когда эффект рангового искажения велик. Аппроксимация Лаймкюлера соответствует распределению Ципфа с $\alpha=1$ и большим ранговым искажением, $B \gg 1$. Действительно, распределение Ципфа для $\alpha=1$ имеет вид (см. выражение (7.2.13)):

$$X(r) = -A \ln(1+B) + A \ln(r+B).$$

Это соответствует аппроксимации Лаймкюлера

$$X(r) = \frac{a \ln b}{\ln(1+b)} + \frac{a}{\ln(1+b)} \ln\left(r + \frac{1}{b}\right)$$

с $b=1/B$ при условии согласованности двух уравнений для параметра аппроксимации Лаймкюлера a :

$$\frac{a}{\ln(1+b)} = A$$

и

$$\frac{a \ln b}{\ln(1+b)} = -A \ln(1+B).$$

Первое из них дает:

$$a = A \ln \frac{1+B}{B},$$

второе:

$$a = A \left(\ln \frac{1+B}{B} \right) \frac{\ln(1+B)}{\ln B}.$$

Эти выражения согласуются, если $B \gg 1$, что требовалось доказать.

Б. Брукс [3.44–3.46] аппроксимирует начало графика, показанного на рис. 2.17г, экспонентой, а остальную часть — прямой:

$$X(r) = \begin{cases} ar^\beta, & 1 \leq r \leq c \\ N \ln \frac{r}{s}, & c \leq r \leq N, \end{cases} \quad (3.3.4)$$

где c — число журналов в «ядре»; α, β, s — параметры аппроксимации Брукса. Эта аппроксимация представляет собой сочетание двух распределений Ципфа с разными значениями α . Первая часть аппроксимации Брукса

$$x(r) = \frac{dX(r)}{dr} = \frac{\alpha\beta}{r^{1-\beta}}$$

соответствует распределению Ципфа с $\alpha=1/(1-\beta)$ и $B=0$ (см. выражение (7.2.12)). Вторая — распределению Ципфа с $\alpha=1$ и $B=s-1 \leq r$ (см. выражение (7.2.13)).

А. Т. Мицевич [3.23–3.24] аппроксимирует показанную на рис. 2.17г кривую алгебраическим многочленом:

$$\frac{X(r) - X(1)}{X(N) - X(1)} = a \frac{\ln^3 r}{\ln^3 N} + b \frac{\ln^2 r}{\ln^2 N} + c \frac{\ln r}{\ln N} + d. \quad (3.3.5)$$

Для этой аппроксимации

$$\left. \frac{d \ln x(r)}{d \ln r} \right|_{r=1} = \frac{2}{\ln N} - 1, \quad (3.3.6)$$

т. е. при $x \rightarrow \infty (N \rightarrow \infty)$ эта аппроксимация является ранговым распределением Ципфа с $\alpha=1$, когда эффект рангового искажения мал (см. выражение (7.2.27) с $B=0$).

График 2.17г не очень хорошо описывается формульными выражениями. Именно поэтому и существуют разные его аппроксимации. В частности, трудно поддается объяснению отклонение эмпирической кривой в верхней части графика от прямой (так называемый «доор-эффект», или эффект Грооса, — по имени впервые обратившего на него внимание ученого О. В. Грооса [3.60]. По этим причинам параллельно с представлением кривых рассеяния публикаций в ранговом интегральном виде их представляют в серии работ в более традиционной дифференциальной (частотной или ранговой) форме.

Итак, ко второй ветви относятся работы, в которых распределение Брэдфорда связывается с распределением Ципфа. Эту связь обсуждают Р. Фэйтхорн [1.97; 1.98], Б. Брукс [3.44], Л. Мэрфи

[3.79], Р. Койл (см. [3.49]), Дж. Хьюберт (см. [3.62]), А. Букштайн [1.77–1.79], Д. Прайс [1.143], А. И. Яблонский [0.63–0.65], Ю. Гарфилд [4.81] и др. Здесь также существуют разные аппроксимации.

Дж. Хьюберт [3.62] предложил аппроксимацию

$$x(r) \sim 1/r^\gamma, \quad (3.3.7)$$

где $x(r)$ — число публикаций заданного профиля в журнале ранга r ; γ — параметр. Это — ранговое распределение Ципфа с произвольным $\alpha=1/\gamma$, когда эффект рангового искажения пренебрежимо мал (см. выражение (7.2.12) с $B=0$).

С. Нарабан [1.131] и А. Букштайн [1.77–1.79] используют аппроксимацию

$$n(x) \sim 1/x^\alpha, \quad (3.3.8)$$

где $n(x)$ — число журналов с x статьями данного профиля. Эта аппроксимация является распределением Ципфа с произвольным $\alpha=s-1$ (см. выражение (7.2.10)).

Р. Фэйтхорн [1.98] и А. И. Яблонский [0.63] аппроксимируют распределение Брэдфорда распределением

$$n(x) \sim 1/x^{1+\alpha}, \quad (3.3.9)$$

которое является распределением Ципфа с произвольным α в современном его написании (см. выражение (7.2.10)).

М. Кендалл [1.119] аппроксимирует распределение Брэдфорда выражением

$$n(x) = N/(a+x^2). \quad (3.3.10)$$

При больших значениях x эта аппроксимация соответствует распределению Ципфа с $\alpha=1$.

Д. Прайс [1.143] аппроксимирует распределение Брэдфорда бета-функцией

$$n(x) = N(m+1)B(x, m+2) = N(m+1) \frac{\Gamma(x)\Gamma(m+2)}{\Gamma(x+m+2)} \quad (3.3.11)$$

($B(p, q)$ — бета-функция, $\Gamma(z)$ — гамма-функция). При больших значениях x эта аппроксимация является распределением Ципфа

$$n(x) \sim 1/x^{m+2} \quad \text{с } \alpha=m+1. \quad (3.3.12)$$

Суммируя обсуждение распределения Брэдфорда, можно сказать, что оно является *цифровым* с α порядка 0,8–1,0 (см. рис. 2.17, 2.18). Отклонение распределения Брэдфорда от распределения

Ципфа в области малых значений x (эффект Грооса) существенно при числе журналов, большем или порядка 300. При меньшем числе журналов распределение Брэдфорда вполне может быть аппроксимировано распределением Ципфа. Когда же эффект Грооса существен, следует использовать другие цифровые аппроксимации, возможно бета-функцию.

Аппроксимации (3.3.1–3.3.3), соответствующие распределению Ципфа с фиксированным значением $\alpha=1$, представляются в свете современных данных не вполне корректными. Не вполне корректно и рассмотрение распределения Брэдфорда в координатах последнего (рис. 2.17г), как чрезмерно усложняющее вопрос об аппроксимации этого распределения и маскирующее его негауссовую природу.

Качественно распределение Брэдфорда означает концентрацию большей части массива публикаций по фиксированной тематике в небольшом ядре ведущих журналов и рассеяние остальной части массива публикаций в большом числе периферийных журналов. Поэтому иногда закон Брэдфорда называют *законом концентрации и рассеяния публикаций*. В качестве количественной меры концентрированности и рассеянности публикаций данного профиля по источникам берут в зависимости от выбранной аппроксимации либо показатель распределения Ципфа α , либо параметры распределения Брукса, либо энтропию эмпирического распределения, как это делают, например, Т. Н. Муравьева и Т. П. Колесникова [3.26].

Распределение журналов по числу рефератов на помещенные в них статьи

П. Г. Лойферман и соавт. [2.23, 3.21] изучает распределение журналов по частоте обращения к ним вторичных (например, реферативных) изданий.

Распределение публикаций по организациям

Это распределение исследуют Дж. Вестбрук [4.142] и Р. Фишер [3.57]. Данные Вестбрука, построенные на небольшой статистике, допускают аппроксимацию распределением Ципфа с $\alpha \approx 1,7$ (см. рис. 2.19).

Распределение ученых по числу публикаций

Число публикаций принято рассматривать в качестве индикатора *продуктивности* ученого. Поэтому изучение распределения

ученых по числу публикаций одновременно есть изучение распределения ученых по продуктивности. Этим и объясняется чрезвычайно большое число работ, посвященных этому распределению (см. например, библиографию [1.158] и обзор [3.87]). Число публикаций расщепляется в этих работах на два индикатора, соответствующие *полной продуктивности* (когда ученому приписывается целая публикация независимо от числа соавторов) и *частичной продуктивности* (когда за публикацию с n авторами каждому начисляется $1/n$ публикации).

Распределение ученых по числу публикаций впервые встречается, по-видимому, в работе А. Дрездена [3.54], приведшего свои данные без какой-либо аппроксимации. Вероятно, это и явилось причиной того, что работа Дрездена большой известности не получила в отличие от работы А. Лотки [3.74]. Поэтому распределение ученых по числу публикаций часто называют распределением Лотки.

Аппроксимации распределения Лотки распадаются на две ветви. В первой это распределение связывается с распределением Ципфа. Эту связь исследуют А. Лотка [3.74], П. Нагнал и соавт. [3.80], М. Кендэлл [1.120], Д. Прайс [1.58; 1.143], Г. М. Добров и А. А. Коренной [1.21], Р. Фэйтхорн [1.98], С. Наранан [1.131], Я. Влахи [3.101], Л. Мэрфи [3.79], Х. Воос [3.102], А. Букштайн [1.77–1.79], А. И. Яблонский [0.63; 0.114] и др. Здесь существуют разные аппроксимации.

А. Лотка [3.74] аппроксимировал распределение ученых по полной продуктивности распределением

$$n(x) \sim C/x^2, \quad (3.3.13)$$

где $n(x)$ — число ученых с x публикациями; C — нормировочная постоянная. Это — распределение Ципфа с $\alpha=1$. Распределение (3.3.13) часто называют *законом Лотки* (см. рис. 2.20).

Д. Прайс [1.58, с. 331] формулирует для распределения Лотки так называемый закон «квадратного корня», согласно которому число высокопродуктивных авторов, написавших половину массива публикаций, равно квадратному корню из общего числа ученых:

$$\sum_{m+1}^I n(x) = \sqrt{N}, \quad (3.3.14)$$

где $I=x_{\max}$ — максимальная продуктивность ученого на данной выборке; N — число ученых; m определяется условием:

$$\sum_1^m xn(x) = \sum_{m+1}^I xn(x). \quad (3.3.15)$$

П. Алисон и др. [3.37] изучали связь закона «квадратного корня» с законом Лотки. Заменяя суммирование в (3.3.14) и (3.3.15) интегрированием, получаем $m \approx \sqrt{I}$, а условие соответствия закона «квадратного корня» закону Лотки приближенно записывается в виде

$$I \approx N. \quad (3.3.16)$$

Авторы, не заменяя интегрированием суммирование, получили условие $I \approx C$. Поскольку для $I \gg 1$ и $\alpha=1$ имеет место $C \approx N$ (см. выражение (7.2.15)), постольку эти условия совпадают. Прайс, являющийся одним из авторов этой статьи, делает персональную сноску, в которой выражает свои сомнения относительно соблюдения (3.3.16). И он прав: в действительности связь I и N носит вероятностный характер (см. выражение (7.2.22)), так что выборка ученых заданного объема N может иметь, вообще говоря, любое значение максимальной продуктивности I . Поэтому закон «квадратного корня», фиксирующий в законе Лотки значение I (равное N) и соответствующий в силу этого частному случаю этого закона, в общем случае не вытекает из него. Так как распределение Лотки в общем случае не подчиняется закону Лотки, то закон «квадратного корня», являющийся частным случаем закона Лотки, в общем случае тем более не соблюдается.

Во многих работах показывается, что не всегда $\alpha=1$. С. Наранан [1.131] и А. Букштайн [1.77–1.79] используют аппроксимацию

$$n(x) \sim 1/x^s. \quad (3.3.17)$$

Эта аппроксимация является распределением Ципфа с произвольным $\alpha=s-1$.

М. Кендэлл [1.120], Г. Фэйтхорн [1.98] и А. И. Яблонский [0.63] используют для описания распределения Лотки аппроксимацию

$$n(x) \sim 1/x^{1+\alpha}, \quad (3.3.18)$$

которая является распределением Ципфа с произвольным α в современном написании этого распределения.

Дж. Хьюберт [1.117] аппроксимирует распределение ученых по числу публикаций распределением (3.3.7). Это — ранговое распределение Ципфа с произвольным $\alpha=1/\gamma$, когда эффект рангового искажения пренебрежимо мал (см. выражение (7.2.12) с $B=0$).

Д. Прайс [1.58, с. 323] аппроксимирует распределение Лотки выражением

$$F(x) = (1/N)(ak/(ax+x^2)), \quad (3.3.19)$$

где $F(x)$ — функция распределения, a и k — параметры. При больших x эта аппроксимация имеет вид распределения Ципфа

$$n(x) \sim 1/x^3 \quad \text{с } \alpha=2. \quad (3.3.20)$$

Д. Прайс [1.143] предлагает аппроксимировать распределение Лотки бета-функцией:

$$n(x) = N(m+1)B(x, m+2) = N(m+1) \frac{\Gamma(x)\Gamma(m+2)}{\Gamma(x+m+2)}. \quad (3.3.21)$$

При больших x эта аппроксимация является распределением Ципфа с $\alpha=m+1$.

Во второй ветви аппроксимаций распределения Лотки используются более традиционные, гауссовы нецифровые распределения. Е. Вильямс [3.105] использует логарифмическое распределение:

$$n(x) \sim 0^x/x, \quad 0 < 0 < 1. \quad (3.3.22)$$

Е. Вильямс [3.105] предложил для аппроксимации распределения Лотки геометрическое распределение:

$$n(x) \sim 0^x, \quad 0 < 0 < 1. \quad (3.3.23)$$

Эта аппроксимация является частным случаем отрицательного биномиального распределения (см. далее).

Лауреат Нобелевской премии по физике 1956 г. У. Шокли предложил в 1957 г. аппроксимировать распределение Лотки логнормальным распределением [1.156] (см. рис. 2.80). Ему последовали Д. Пельц и Ф. Эндрюс [0.51], Д. Прайс и Д. Бивер [2.30] (см. рис. 2.81), Р. Ф. Васильев [1.6], А. А. Игнатьев и А. И. Яблонский [0.31]. Шокли аппроксимирует логнормальным распределением

$$n(x) = \frac{N}{x\sqrt{2\pi}} e^{-1/2 \left(\frac{\ln x - \mu}{\sigma} \right)^2}, \quad x \geq 0 \quad (3.3.24)$$

распределение ученых по частичной продуктивности, Прайс и Бивер, Игнатьев и Яблонский — распределение ученых и по полной и частичной продуктивности. Игнатьев и Яблонский полагают при этом, что информационные массивы (распределение публикаций по именам) подчиняются распределению Ципфа, а массивы ученых (распределение индивидуальных вкладов) — логнормальному распределению.

Л. Мантелл [3.75] прибегает к аппроксимации распределения Лотки распределением Пуассона:

$$n(x) = Ne^{-\lambda} \lambda^x / x!, \quad \lambda > 0, \quad x = 1, 2, 3, \dots \quad (3.3.25)$$

(см. рис. 2.79).

К. Равичандра Рао [3.87] предлагает аппроксимировать распределение Лотки отрицательным биномиальным распределением:

$$n(x) = \frac{\Gamma(x+k)}{\Gamma(k)\Gamma(x+1)} 0^x, \quad 0 < 0 < 1, \quad k > 0, \quad x = 1, 2, 3, \dots \quad (3.3.26)$$

где $\Gamma(z)$ — гамма-функция. Геометрическое распределение (см. аппроксимацию Вильямса (3.3.23)) является отрицательным биномиальным распределением с $k=1$.

Несколько слов о позиции Д. Прайса. В 1963 г. Прайс [1.58] в какой-то степени принял закон Лотки, т. е. распределение Ципфа с $\alpha=1$. Однако, будучи не вполне им удовлетворен, Прайс тогда же, во-первых, формулирует закон «квадратного корня» (являющийся, как было показано, частным случаем закона Лотки) и, во-вторых, несколько модифицирует закон Лотки, позволяя показателю распределения Ципфа меняться от 1 до 2 (см. аппроксимацию (3.3.19)). Но и такой, модифицированный закон Лотки не удовлетворяет Прайса. В 1966 г. в статье, написанной совместно с Д. Бивером [2.30], он явно отдает предпочтение уже логнормальному распределению. Однако и это распределение ненадолго удовлетворяет Прайса. В 1974 г. он допускает одновременное использование гиперболического, логнормального и отрицательного биномиального распределений в качестве равноправных аппроксимаций распределения Лотки [1.142, с. 183]. В 1976 г. Прайс приходит как будто бы к окончательному решению — бета-функция и только она одна может использоваться в этой роли [1.143]. Однако в 1980 г. он снова допускает возможность использования наряду с бета-функцией отрицательного биномиального распределения [1.146, с. 47].

Суммируя обсуждение распределения Лотки, можно сказать, что оно является *цифровым* со значениями α , лежащими в диапазоне примерно 1–2,5 (см., например, рис. 2.20–2.23). Отклонение распределения Лотки от распределения Ципфа в области малых значений x существенно для распределений по частичной продуктивности и менее существенно — по полной. Различия эмпирических распределений ученых по числу публикаций в области малых значений переменных являются в большой степени *артефактами*, порожденными невоспроизводимым «обрезанием» выборки. Форма этих распределений характеризует не только распределение ученых по латентной переменной «продуктивность», но и критерии, которыми руководствуются при формировании анализируемой вы-

борки *ученых*, «обрезая» малопродуктивную в научном отношении ее часть. В распределении всего населения по латентной переменной научной продуктивности нулевому значению этой продуктивности будет, очевидно, соответствовать бесконечно большое число индивидов, не имеющих отношения к науке. Ученый же по определению производит ненулевой научный вклад, и потому в распределении *ученых* по научной продуктивности нулевому значению последней не будет соответствовать ни один индивид, т. е. число ученых с нулевой продуктивностью будет равно нулю. Поэтому мы и должны на выборках *ученых* получать в области малых значений научной продуктивности отклоняющиеся от распределения Ципфа выборочные распределения. Именно такие распределения мы и получаем, когда работаем с индикаторами научной продуктивности, хотя бы и грубо, но фиксирующими научные вклады, меньшие, чем одна публикация. Так обстоит дело, например, с распределениями ученых по *частичной* продуктивности (см. рис. 2.23).

Если же мы используем индикаторы, не фиксирующие малые научные вклады, то получаем распределения, очень мало отклоняющиеся от распределения Ципфа. Так обстоит дело с распределениями ученых по *полной* продуктивности (см. рис. 2.20–2.22).

Критерии, позволяющие отделить *ученых* от *неученых*, слабо воспроизводимы. Поэтому слабо воспроизводим и любой конкретный способ «обрезания» выборки *ученых*, а форма распределения Лотки в области малых значений переменных характеризует не только и не столько объекты исследования, т. е. ученых и их публикации, сколько критерии исследователя-научомера, которыми он руководствуется при формировании выборки *ученых*. Это снижает ценность отдельных ципфовых аппроксимаций распределения Лотки, в частности бета-функции, которая, таким образом, почти наверное не является единственно корректной ципфовой аппроксимацией распределения Лотки.

Если говорить конкретно о распределении ученых по *частичной* продуктивности, то на его базе может быть построено распределение, совсем мало или вовсе не отличающееся от распределения Ципфа. Для этого следует только максимально уменьшить заложенный в эмпирических данных произвол, связанный с «обрезанием» выборки в области малых значений переменных. Скажем, взять распределение *отдельных значений частичной продуктивности* по частоте, с которой они встречаются в той же выборке ученых. Здесь, очевидно (сравните рис. 2.23б и 2.24), переменной уже является число ученых, а роль элементов, из которых формируется выборка, играют не ученые, а отдельные значения *частичной* продуктивности. Поскольку эта выборка формировалась

нами не прямо, а опосредованно, через выборку ученых, постольку здесь влияние указанного выше произвола существенно ослаблено. Так что следует ожидать, что данное распределение будет мало отличаться от распределения Ципфа. Эмпирические данные подтверждают этот вывод (см. рис. 2.24). Аналогичная ситуация имеет место и с распределением отдельных значений *частичной* продуктивности, деленной на полную продуктивность (см. рис. 2.25). Гауссовы аппроксимации (3.3.22)–(3.3.26) в применении к распределению Лотки некорректны. «Короткохвостовые» гауссовы распределения просто не могут быть с удовлетворительной точностью совмещены с данными наблюдениями, как это имеет место, например, для распределения Пуассона (см. рис. 2.22). «Длиннохвостовые» же гауссовы распределения, особенно на небольших выборках, часто могут быть с хорошей точностью совмещены с эмпирическими данными. Аппроксимации гауссовыми распределениями некорректны и в этих случаях. И некорректны *в силу существенной зависимости для них моментов от объемов выборок* для эмпирических значений параметров этих выборок. Это иллюстрируют рис. 2.80–2.82 для логнормального распределения. Подробнее об этом говорится в гл. 7, § 2. Здесь же заметим, что гауссовы распределения для тех конечных выборок, на которых они могут быть хорошо совмещены с эмпирическими точками, еще не являются, строго говоря, гауссовыми, так как не успевают сойтись к распределению Гаусса в том смысле, в каком они сходятся к нему с ростом объема выборки до бесконечности.

Распределение публикаций среди ученых по полам

Это распределение изучают Л. Гайер и Л. Фидели [3.61], Н. Коган [3.66], А. Фэрримен [3.56] и др. Отмечается, что, если статистически устранить влияние факторов, не связанных с полом, разница в количестве публикаций между мужчинами и женщинами становится пренебрежимо малой [3.61]. Таким образом, если пол и влияет на научную продуктивность, то это влияние невелико. Чтобы установить степень этого влияния в сравнении с другими факторами, влияющими на научную продуктивность, необходимо, как и везде, вести многофакторные исследования.

Распределение числа работ по числу соавторов

Эти распределения анализируют в своих работах Д. Прайс [1.58], Д. Прайс и Д. Бивер [2.30], Ю. П. Адлер и соавт. [1.1], Я. Влахи [2.70], В. А. Вознесенский [1.8], Я. Мончев [2.27], В. Б. Лукьянов и соавт. [3.22], Г. Магяр [1.123] и др. Д. Прайс

и Д. Бивер нашли, что число работ с n авторствами в первом приближении пропорционально $1/(n-1)$, кроме работ с одним автором [2.30]. Данные Прайса и Бивера, приведенные на рис. 2.26, доказывают возможность аппроксимации распределения числа работ по числу соавторов цифровым распределением.

Распределение публикаций по времени их задержки в журналах

Распределение изучают З. Б. Барнинова и соавт. [4.3], Г. М. Добров и соавт. [0.25], В. А. Маркусова [4.15], К. Субраманиян и Дж. Шаффер [2.65] и др. Приводящиеся здесь гистограммы используются для сравнения отдельных журналов по этому индикатору.

«Экзотические» распределения публикаций

А. П. Борисова и соавт. [1.4] изучают распределение биохимических работ по числу химических элементов, которые в них рассматриваются. Э. П. Фридман [3.32–3.34] дает распределение публикаций о приматах по видам использованных в ходе исследования обезьян. И. В. Дзиековская [3.9], проводя наукометрическое исследование литературы в области сверхпроводящих тонких пленок, дает распределение числа публикаций по сверхпроводящим однослойным структурам на основе одного химического элемента, сплавов, соединений и многослойных структур.

Распределение публикаций по предмету

Распределение публикаций по научным направлениям изучают Р. Фишер [3.57], М. М. Карпов и соавт. [3.13], В. Викери [3.98], Э. М. Мирский [1.41], В. И. Горькова [3.6], Г. И. Коншин и В. Б. Ляцкий [3.15], М. К. Петров [3.28], В. А. Конышев и М. Г. Смирнова [3.16], К. С. Лагерева и соавт. [1.34], Э. П. Фридман [3.33–3.34], В. Л. Мигдал и др. [2.25] и др. Сопоставление таких предметных спектров для отдельных стран позволяет получить представление о степени относительной развитости отдельных отраслей науки в этих странах, что может быть полезным при выработке политики научного сотрудничества (см. например, [3.13]).

Распределение публикаций по странам

Это распределение изучается по крайней мере с 1889 г. (см. [2.50]). Им занимаются Д. Прайс [1.58], Р. Тримбл [3.95], Р. Фи-

шер [3.57], З. Б. Барнинова и соавт. [4.3], М. М. Карпов и др. [3.13], Дж. Бен-Давид [3.41], В. И. Горькова [3.6], М. К. Петров [3.28], Э. П. Фридман [3.32–3.34], А. П. Борисова и соавт. [1.4], С. А. Заремба и И. М. Алабян [2.13], К. С. Лагерева и соавт. [1.34], В. Лабаускайте и А. Канцлерис [3.20], Г. Магяр [1.123], М. К. Князева [3.14], Г. Инхэйбер [3.64], М. Роше [3.90] и др. На рис. 2.27–2.28 показаны типичные распределения числа научных работ по странам. Здесь, как и в других случаях, оказывается возможной аппроксимация цифровым распределением.

Распределение публикаций по языкам

Эти данные приводят в своих работах Р. Тримбл [3.95], В. А. Полушкин и Г. С. Жданова [1.56], А. П. Борисова и соавт. [1.4], М. Г. Куренкова [3.19], Н. А. Чайка [3.35] и др. Рис. 2.29 показывает, что здесь также уместна аппроксимация распределением Ципфа.

Распределение публикаций по типу издания и характеру

Данные такого рода приводят Ю. Т. Бурбуля и Б. П. Коварская [3.4], И. В. Дзиековская [3.9], М. Г. Куренкова [3.19], М. М. Черная и О. Ф. Руденкова [1.67] и др. Публикации здесь делятся: по характеру — обзоры, теоретические, экспериментальные и т.п.; по типу издания — журналы, сборники, труды конференций, отечественные, зарубежные и т.п. (см., например, рис. 1.7).

4. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: СТРУКТУРНЫЕ ИССЛЕДОВАНИЯ

К данному параграфу мы, следуя принятой в настоящем исследовании схеме подачи материала, не смогли отнести ни одной работы из попавших в наше поле зрения.

5. ПЕРСПЕКТИВЫ

Наукометрический индикатор число научных публикаций разрабатан больше, чем какой-либо из индикаторов предыдущей главы. Тем не менее общее состояние разработанности метода подсчета числа публикаций нельзя считать удовлетворительным. Действительно, табл. 3.5.1 показывает количество наукометрических работ, рассмотренных в настоящей главе, в которых используется тот или другой индикатор.

Таблица 3.5.1

Распределение числа упомянутых в настоящей главе работ в зависимости от использования отдельных индикаторов

№ п/п	Индикаторы	Число работ	№ п/п	Индикаторы	Число работ
1	Число открытий	—	11	«Экзотические»	5
2	Число журналов	50	12	Время	70
3	Число организаций	2	13	Возраст	9
4	Число премий	—	14	Предметные	13
5	Число ученых	60	15	Географические	18
6	Частота соавторства	8	16	Квалификационные	—
7	Число публикаций	—	17	Должностные	—
8	Число запросов	—	18	Языковые	5
9	Объем публикаций	9	19	Прочие	4
10	Временной лаг	4			

Видим, что табл. 3.5.1 местами не заполнена или заполнена в известной степени случайно. Следовательно, необходимы дальнейшие исследования по разработке числа публикаций как наукометрического индикатора. В частности, должна быть разработана формализованная процедура «взвешивания» публикаций разных типов и отдельных публикаций. Основой такого «взвешивания» может служить ранг публикаций, издания или типа издания, в котором помещена эта публикация. Ранг же может определяться средствами цитатного, социологического или какого-либо другого анализа.

Перспективы применения метода подсчета числа публикаций определяются результатами опорных исследований числа публикаций как наукометрического индикатора латентных переменных научной деятельности. Число публикаций может служить, как показывают изложенные в § 1 результаты таких исследований, индикатором признания ученого, его известности, вклада в науку, продуктивности, престижа научного учреждения.

Согласно зафиксированным в литературе, так сказать, опорным соображениям качественного порядка, число публикаций может служить индикатором элитности ученого, его активности, интенсивности. Большое число монографий в данном научном направлении связывают с его насыщенностью, а число журнальных статей — с его конфликтностью. Ценность журнала определяют по частотам отражения их публикаций во вторичных изданиях.

Скорость роста числа публикаций связывают с актуальностью данного научного направления или с его перспективностью.

По скорости роста числа публикаций данного научного коллектива в данный момент, соотношенной с общей логистической кривой их роста для этого коллектива, предлагается определять фазы развития данного коллектива. Коллективы, находящиеся в разных фазах своего развития, перспективны в разной степени.

Стационарные распределения числа публикаций также могут быть использованы для выявления проблемных научных направлений [1.11], которые характеризуются распределениями, имеющими более короткие «хвосты», нежели традиционные направления. В проблемном направлении публикации рассеяны, например, в меньшем числе журналов, чем в непроблемном.

Таким образом, возможности применения числа публикаций в качестве индикатора научной деятельности, как показывают результаты опорных исследований, в принципе обширны. Однако до широкого практического использования результатов этих исследований еще далеко, поскольку не установлена полная картина корреляций числа научных публикаций ни с системой индикаторов, ни с системой латентных параметров науки. Кроме того, уже вычисленные корреляции устанавливались на основе аппарата гауссовой математической статистики, тогда как распределения, между которыми устанавливались эти корреляции, имеют сугубо негауссову природу. Все это означает, что опорные исследования числа публикаций должны вестись фактически заново средствами негауссовой математической статистики с широким систематическим охватом всей системы индикаторов и латентных переменных науки.

Следует напомнить здесь, что все наукометрические оценки носят статистический характер. Фиксируемые опорными исследованиями связи числа публикаций как наукометрического индикатора с латентными переменными научной деятельности также носят поэтому статистический характер. Это означает, что, не гарантируя верности оценки в каждом конкретном случае, среднестатистически метод подсчета числа публикаций при его использовании совместно с другими методами даст положительный эффект. Между тем аппарат ошибок в освещенных в настоящей главе исследованиях использовался слабо.

Глава четвертая

МЕТОД «ЦИТАТ-ИНДЕКС»

1. СОДЕРЖАНИЕ МЕТОДА И ОПОРНЫЕ ИССЛЕДОВАНИЯ

А. Эйнштейн, опубликовав в 1905 г. свою первую и основную работу по специальной теории относительности, не сопроводил ее библиографией. Эту оплошность молодого физика-непрофессионала (Эйнштейн тогда работал еще в Бернском бюро патентов) счел необходимым исправить знаменитый уже в то время академик М. Планк, который собственноручно составил по теме статьи библиографию и даже сопроводил ее обзором, который и опубликовал.

Планк оказал эту услугу малоизвестному автору не только потому, что признавал важность идей Эйнштейна, но и потому, что понимал важность *аппарата библиографических ссылок*.

В наши дни, как и три четверти века назад, упоминание в ссылках работ предшественников является нормой, обязательной для научной публикации (см. например, [0.6, 4.85, 4.2]). Наукометрический метод «цитат-индекс» базируется именно на этой обязательности ссылок в научных публикациях, в основе его лежит наукометрический индикатор «число цитат или ссылок». Иногда, следуя Д. Прайсу [4.126], цитаты и ссылки различают, понимая под числом ссылок число пунктов приводимой в статье библиографии, а под числом цитат — число упоминаний статьи другими авторами. Мы же в дальнейшем изложении эти термины различать не будем.

В принципе ссылки, как и публикации, могут подвергаться «взвешиванию», например, в соответствии с определяемым средствами социологического или цитатного анализа рангом публикации, издания или автора, делающего эту ссылку.

При цитировании может учитываться по-разному и соавторство. Можно учитывать только первую фамилию на титуле работы, а можно всех соавторов. Эти индикаторы тесно коррелируют между собой. Так, например, согласно данным Дж. Коула и С. Коула [4.58], корреляция между числом ссылок только по первому автору и общим числом ссылок равна 0,95.

Цитатный анализ возник в таких науках, как теория информации-поисковых систем, библиометрия и др., направленных на удовлетворение информационных потребностей ученых. Практическая база для использования этого метода появилась с созданием указателей научных ссылок. Современные указатели ссылок восходят к «Парижской библиографии», изданной в 1771—1772 гг. [4.41]. В этой библиографии рецензий, опубликованных в двух журналах, на книги, изданные в Париже в 1770 г., материал упорядочивается не по рецензиям, а по рецензируемым книгам.

Другим предшественником указателей научных ссылок является изданный в 1873 г. «Указатель ссылок Федерального законодательства» Шепарда (США) [4.132], включающий в себя списки публикаций, содержащих ссылки на данные судебные решения, интересные как прецеденты.

В 1949 г. вышел указатель к американскому журналу «Анналы математической статистики», одна из частей которого содержит список статей из других журналов, на которые не менее двух раз ссылались в публикациях данного журнала. Цитируемые статьи расположены в указателе в алфавитном порядке фамилий авторов [см. 4.10].

В 1963 г. в экспериментальном порядке, а с 1964 г. — регулярно стал выходить «Индекс научных ссылок» (SCI, Science Citation Index), издаваемый Институтом научной информации в Филадельфии, возглавляемом Ю. Гарфилдом. Индекс Гарфилда (так часто называют этот указатель) и лежит в основе современных применений метода «цитат-индекс». Он состоит из нескольких частей. *Указатель ссылок*, представляющий собой основную часть Индекса Гарфилда, позволяет установить, кто цитирует фиксированную работу данного автора (предусмотрен также учет анонимных работ). Фиксируется лишь фамилия первого автора работы, название работы не приводится, название журнала дается в сокращенном виде (аббревиатурой). Название работы, фамилии и адреса всех авторов можно найти в *указателе источников*, представляющем собой вторую часть Индекса Гарфилда. Третья часть Индекса — *пермутационный предметный указатель* — позволяет найти авторов, в заголовках работ которых встречается данное слово. И, наконец, последняя часть Индекса Гарфилда — *Journal Citation Reports* — содержит результаты библиометрического анализа охватываемых Индексом Гарфилда цитирующих и цитируемых журналов. Каждый журнал характеризуется рядом показателей, среди которых — число ссылок в нем и на него в отдельных журналах и в сумме, так называемый импакт-фактор

Таблица 4.1.1

Распределение научных периодических и продолжающихся изданий, учитываемых Индексом научных ссылок, по странам (1977 г.)

Страна	Число изданий	Страна	Число изданий	Страна	Число изданий
США	1372	Франция	145	Дания	40
Великобритания	382	Япония	79	Италия	40
ФРГ	253	СССР	61	Швеция	36
Нидерланды	178	Канада	49	ГДР	35
Швейцария	148				

(impact-factor) (см. далее) и др. Приводятся ранжировки журналов по этим показателям.

Собственно Индекс научных ссылок охватывает естественные науки. Институтом научной информации Ю. Гарфилда с 1970 г. выпускается и Индекс ссылок по социальным наукам (SSCI, Social Science Citation Index), но много меньшего объема. При составлении Индекса Гарфилда просматривают наиболее известные (по мнению издателей Индекса, см. об этом далее) научные журналы и другие периодические и продолжающиеся (serial) издания, а также книги и патенты. Данные вводятся в ЭВМ, которая и дает в переработанном виде всю указанную выше информацию. При наличии соответствующего заказа программой предусматривается получение информации, не содержащейся в публикуемом Индексе. Например, можно получить сведения о цитировании данного автора или данной группы авторов, о распределении публикаций на заданную тематику (заголовки которых содержат данный набор терминов) по журналам или странам и т. д.

К сожалению, составители Индекса Гарфилда руководствуются, прежде всего, национальными интересами. Это приводит к тому, что различные страны представлены в нем отнюдь не пропорционально их вкладам в мировую науку (табл. 4.1.1).

В 1977 г. Индексом Гарфилда было учтено 3 115 изданий, входящих в 41 стране. Советские ученые представлены в нем, как видим, всего 61 изданием, в то время как в 1977 г. в СССР вышло более 4 тыс. периодических и продолжающихся изданий. Понятно, что если опираться только на данные Индекса, мы получим заведомо искаженные представления о вкладах в мировую науку ученых разных стран.

Первое упоминание об отечественном указателе ссылок встречается в работе З. А. Футько [4.31], в которой обсуждается указатель ссылок, охватывающий вопросы строительства. В 1972 г. в Азербайджанском НИИ научно-технической информации были опубликованы указатели цитированной литературы «УЦЛ-нефть» и «УЦЛ-экономика», а в 1974 г. — «УЦЛ-информатика».

Из первых исследований науки средствами цитатного анализа назовем работы, которые выполнили русский химик академик П. И. Вальден [см. 4.1] и в 1927 г. — П. Гросс и Е. Гросс [4.89]. В 1938 г. Дж. Бернал в своей книге «Наука и общество», заложившей основы современного науковедения, высказал идеи, касающиеся применения цитатного анализа при исследованиях науки с использованием теории графов и матричных методов. В наши дни цитатный метод широко используется не только науковедами. К нему все чаще обращаются и историки науки при оценках темпов роста и старения научной литературы, а также идентификации наиболее важных и влиятельных публикаций, ученых и формальных и неформальных научных коллективов.

Описания и правила пользования указателями научных ссылок содержатся, например, в работах [4.33, 1.46, 4.10, 4.27, 4.32]. Наиболее удобным при науковедческом анализе остается до сих пор Индекс Гарфилда, сканирующий примерно 2—3% всех издающихся в мире периодических изданий. Он позволяет учесть около 50% литературы по любой данной тематике.

Как и метод подсчета числа публикаций, метод «цитат-индекс» используется для измерения параметров науки: продукт труда ученого — научную публикацию, только делается это здесь иначе. Число ссылок — более «дробная» мера публикации: на одну публикацию приходится в среднем около 15 ссылок (см. § 3). Ссылки фиксируют связь данной работы с работами предшественников и коллег, указывают место этой работы во всем массиве публикаций, неявно, таким образом, оценивая не только цитируемые публикации, но и самую работу автора. С этой точки зрения, «цитат-индекс» — это завуалированный метод экспертной оценки, причем не в лучшем его варианте, ибо экспертизу работы проводит непосредственно исполнитель. Естественно, что это порождает субъективность метода ссылок и картина связей научных работ искажается вследствие либо личных отношений, либо пристрастия и «местного» патриотизма, либо языковых барьеров и т. д.

Субъективность метода ссылок в определенной степени гасится на больших массивах публикаций в результате усреднения экспертиз разных авторов. Однако увеличение массива публикаций не снимает некоторых систематических ошибок. Прежде всего, имеются в виду *преувеличение роли больших ученых и недооценка малых*. При равном качестве статей работы ученых вы-

сокого ранга, как установлено [4.61], цитируются чаще, чем низкого. Можно понять, почему это происходит. Ссылаясь на больших ученых, автор тем самым ставит свою работу в один ряд с их работами. Такое преувеличение значимости своих работ свойственно любому ученому. Факты истории науки неопровержимо свидетельствуют, что открытия в их большинстве совершаются людьми с «установкой на открытие». Эта особенность научного мышления и проявляется в склонности ученого к преувеличению своих достижений. Если бы «цитат-индекс» не давал преимущества маститым ученым, то это, возможно, облегчило бы жизнь наукометрам, но затруднило бы функционирование всей науки.

Именно на непонимании этого вносимого «цитат-индексом» систематического «перекоса» в пользу больших ученых основаны обсуждающиеся далее (см. § 3) заключения некоторых авторов о необходимости удаления из науки малоцитируемых авторов как не способствующих ее развитию.

Как и метод подсчета числа публикаций, метод ссылок применяется на массивах публикаций. Но в силу большей «дробности» числа ссылок как наукометрического индикатора по сравнению с числом публикаций (на одну публикацию приходится в среднем, как говорилось, порядка 11–15 ссылок) при фиксированном уровне статистической достоверности результатов счета метод ссылок предполагает использование массивов публикаций примерно в 11–15 раз меньших, чем при использовании метода подсчета публикаций. Поэтому в статистическом смысле метод ссылок — более валидный, чем счет публикаций [см. например, 4.38].

Среди недостатков метода ссылок, кроме отмеченных выше, можно назвать трудность учета отрицательного цитирования, запаздывание с цитированием новых идей, детерминирование числа ссылок числом исследователей в данной области и др. [см., например, 1.46, 1.14]. Наличие этих недостатков побуждает некоторых авторов говорить о невозможности применения этого метода [см., например, 4.65]. Иногда же, напротив, несмотря на недостатки, оценки метода ссылок чрезмерно оптимистичны [см., например, 4.26]. Вопрос же заключается, на наш взгляд, в определении круга задач, при решении которых адекватен метод «цитат-индекс». Другими словами, число ссылок должно быть соотносено в *опорных исследованиях* с другими индикаторами и латентными параметрами науки.

Опорные исследования метода ссылок фактически ведутся уже довольно давно, хотя и не называются так. Помимо рассмотренных далее работ, к числу опорных могут быть отнесены некоторые работы, рассматриваемые в заключении.

Таблица 4.1.2

Корреляция между разными мерами продуктивности социологов [4.52]

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1		0,972	0,923	0,746	0,562	0,617	0,496
X_2			0,916	0,699	0,516	0,617	0,442
X_3				0,474	0,461	0,561	0,427
X_4					0,552	0,589	0,468
X_5						0,863	0,779
X_6							0,779

Корреляцию распределений по числу ссылок и по числу публикаций устанавливали Л. Мельтцер [3.77], Дж. Вестбрук [4.142], Дж. Коул и С. Коул [4.58], Е. Лайтфилд [4.102], Д. Чубин [4.52], З. М. Мульченко и соавт. [1.43], Р. Ф. Васильев [4.5] и др. Л. Мельтцер [3.77] находит корреляцию распределений ученых по числу публикаций и числу ссылок равной 0,51, Е. Лайтфилд [4.102] — 0,75, а Д. Чубин, изучавший социологов, дает таблицу коэффициентов корреляций между разными мерами продуктивности ученого [4.52] (табл. 4.1.2).

В табл. 4.1.2 X_1 — X_4 — меры продуктивности, полученные на основе числа публикаций, а X_5 — X_7 — на основе числа ссылок. Например, X_1 — общее число работ, X_2 — число статей, X_3 — число моносоавторств и т. д.

Дж. Коул и С. Коул [4.58], найдя положительную корреляцию между числом цитат на работы физиков и числом их опубликованных работ, не нашли, однако, какой-либо связи между размерами предметной области и числом ссылок на отдельных ученых: в 1966 г. физики, работающие в области физики твердого тела, имели в среднем в год по 17 ссылок, а работающие в области физики элементарных частиц — по 19, хотя первых было примерно в 2,5 раза больше, чем вторых.

В опорных исследованиях существен объем выборки, на которых они проводятся. Так, Р. Ф. Васильев [4.5] на группе, состоящей из 26 физико-химиков, нашел, что число публикаций и число ссылок дают «резко различные ранжировки». Ясно, что анализируемая им выборка слишком мала, чтобы можно было делать какие-либо выводы. Приведенные же выше данные, полученные на больших выборках, говорят о наличии такой корреляции, хотя и не очень большой по абсолютной величине.

Дж. Вестбрук искал корреляцию между цитируемостью лабораторий по исследованию керамических материалов и числом вы-

исследования в этих работ. На своей небольшой статистике он нашел некую корреляцию слабой [4.142].

Ю. В. Грановский и соавт. [1.16] находят слабо выраженную корреляцию между распределениями по числу Нобелевских премий и по числу ссылок на ее лауреатов.

Корреляцию между распределениями ученых по числу ссылок на них и мерами признания этих ученых обсуждают Р. Ф. Васселов [4.6], С. Коул и Дж. Коул [4.62; 4.63], Дж. Коул [1.33], К. Майерс [4.118], Дж. Коул и С. Коул [4.58], У. Менард [1.125], Дж. Коул и С. Коул [4.39], Р. Эдберт [4.34] и др. Как показывают, например, приведенная в предыдущей главе табл. 3.1.1, эта корреляция выше, чем для числа публикаций.

У. Менард [1.125], данные которого по 50 персонам частично уже обсуждались в § 1 предыдущей главы, также нашел, что число ссылок хорошо отражает признание ученого, определенное им по членству в Национальной академии наук (США) и по тому, были ли они президентами или медалистами Американского психологического общества.

А. Байер и Дж. Фольгер [4.38] нашли положительную корреляцию между числом ссылок на ученого и его индивидуальным вкладом в науку.

Корреляцию престижа ученого с числом ссылок на него нашел Дж. Коул [1.33], корреляцию престижа научного учреждения с числом ссылок на вышесказанные в его стенах работы установили О. Харгрен [4.92] и М. Ороманор [4.123].

К. Майерс [4.118] нашел корреляцию, равную 0,89, между ссылами на 45 психологов и цитированием их по качеству исследований 23 экспертами-коллегами.

Цитирование работы с ее ценностью связывают Е. Бродмэн [4.45], А. Коренной [4.13], Т. В. Сорокина и Е. Д. Гражданников [4.27] и др. Е. Бродмэн [4.45] в своей ранней работе не получила значимой корреляции оценок сотрудниками факультета психологии Колумбийского университета ценности работ психологов с распределением этих работ по числу ссылок на них в четырех ведущих журналах этой области. Используемая ею статистика невелика.

К. Кларк [4.54] находит корреляцию распределений журнальных статей по числу ссылок на них и оценкам экспертов об уровне (eminence) работы равной 0,67 и на этом основании относит «цитат-индекс» к числу лучших из наличных индикаторов веса (worth) работ психологов.

М. Мартин [4.111] на небольшой статистике сравнивает цитируемость отдельных работ с экспертной оценкой их значимости. На вербальном уровне число ссылок связывают: с полезностью

работы — В. В. Налимов и З. М. Мульченко [1.46], с ее влиянием на ученых — А. Байер и Дж. Фольгер [4.38] и В. В. Налимов и З. М. Мульченко [1.46], с важностью работы — А. Байер и Дж. Фольгер [4.38], с желанием ученого публиковаться — Н. Лин и К. Нельсон [4.103] и т. д.

О работах, в которых изучается совпадение цитатных связей со связями, выявленными в ходе содержательного историко-научного анализа, см. в § 4.

В целом опорные исследования числа ссылок находятся в зачаточном состоянии, поскольку не только нет общей картины корреляций числа ссылок с другими индикаторами и латентными переменными научной деятельности, но и в проведенных опорных исследованиях, как правило, не определялась статистическая достоверность расчетных значений этих коэффициентов корреляций.

2. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: НЕСТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

Динамика роста числа ссылок

Рост во времени числа цитат рассматривают Д. Прайс [1.58], З. Б. Баринава и др. [4.3], Б. Тагликоззо [4.137], В. М. Егоров [1.23], У. Менард [1.125], Х. Ротман и М. Вудхид [4.129], Д. Крайн [1.90], М. Моравчик [1.127], Ю. В. Грановский и соавт. [1.18], С. Кот [4.100] и др. Типичные графики показаны на рис. 1.8, 1.19а, б. Кривые роста числа ссылок, как и кривые роста числа публикаций, могут быть грубо аппроксимированы экспонентами с различными периодами удвоения для разных предметных областей. Предполагается, что сравнительный анализ темпов роста числа ссылок в разных научных направлениях поможет выбрать наиболее перспективные из них [см., например, 1.18].

Следует заметить, что, пока скорость роста числа ссылок как индикатор перспективности научного направления не связана с латентными переменными перспективности направления в опорных исследованиях, трудно говорить о таком использовании метода ссылок. Но даже, если бы опорные исследования были здесь уже проведены, обсуждаемый анализ должен производиться осмотнительно.

Ю. В. Грановский и соавт. [1.18], например, сравнивают несколько научных направлений по удельной цитируемости (число ссылок, деленное на число работ), взятой по годам. Чтобы учесть то обстоятельство, что Индекс Гарфилда, которым пользовались эти авторы, сканирует только часть мировой литературы, авторы умножили соответствующие показатели для разных направлений

на коэффициенты, обратно пропорциональные числам сканируемых Индексом журналов по данным направлениям. Коэффициент для биохимии равен 1,0, для математики — 0,50, для физики — 0,25, биологии — 0,125 и т. д. Это означает, что сканируется в 2 раза больше журналов по математике, в 4 — по физике, в 8 раз — по биологии, чем по биохимии.

Ю. В. Грановский и соавт. установили, что биохимия опережает по рассматриваемому параметру остальные направления. Но и число сканируемых по биохимии журналов меньше всего. Такое совпадение неслучайно: когда Индекс Гарфилда «отсекает» какое-то число журналов данного направления, он захватывает наиболее цитируемые издания. Но журналы распределены по цитируемости согласно распределению Ципфа (см. наст. гл., § 3). Поэтому пропорциональности между числом ссылок и числом журналов нет. В предположении линейности этой зависимости, принятом Ю. В. Грановским и соавт., из направлений, производящих равные массивы литературы, предпочтение автоматически отдается наименее полно сканируемым. Чтобы утверждать, что биохимия опережает другие научные направления по удельной цитируемости, следовало бы убедиться в том, что литературы в этой области меньше, чем в других, причем настолько, что степень охвата Индексом Гарфилда биохимической литературы не уступает степени охвата Индексом других направлений, а если и уступает, то не настолько, чтобы это могло исказить соотношение направлений по удельной цитируемости. Ю. В. Грановский и его соавторы не делают этого. В целях увеличения валидности проводимой этими авторами корректировки данных следовало выбрать поправочные коэффициенты с учетом формы распределения журналов по цитируемости, т. е. распределения Ципфа.

Данные по росту числа цитат можно получить, в частности, и путем фиксирования года издания литературы, ссылки в которой затем распределяются по годам издания цитируемой литературы. Подобные данные (см., например, рис. 1.8, 1.19а) могут рассматриваться и как данные по возрастному распределению ссылок (см. далее).

Возрастное распределение ссылок (старение публикаций)

В наше поле зрения попала 61 работа, в которых изучается старение публикаций. Типичные эмпирические данные, приведенные на рис. 1.28 и 1.29, показывают, что, хотя вероятность цитирования данной статьи растет с ее возрастом, новые статьи цитируются значительно чаще, чем старые, если не считать, ко-

нечно, тех статей, с которыми научная общественность еще не успела ознакомиться.

Некоторые эмпирические данные по старению публикаций оказываются возможным, как это показали П. Коул [4.60], Д. Прайс [1.58], Б. Брукс [4.4] и др., аппроксимировать экспонентой. На рис. 1.30 и 1.31 показаны соответствующие данные.

Из предположения об экспоненциальном уменьшении числа ссылок на публикации с возрастом последних, дополненного предположением об экспоненциальном же росте числа публикаций со временем с тем же периодом удвоения (см. гл. 3, § 2), следует закон постоянства числа ссылок в одной статье на другие работы, сформулированный Д. Прайсом [1.58] и М. Гомпертсом [4.86].

Если распределение числа ссылок y на работы, опубликованные за t лет до текущего момента, описывается экспоненциальной зависимостью

$$y = ae^{-\lambda t}, \quad (4.2.1)$$

то скорость старения литературы можно характеризовать параметром λ . Пользуются, однако, параметром $(\ln 2)/2$, равным числу лет, за которое количество ссылок падает вдвое. Этот параметр называют *временем полужизни* публикаций или периодом удвоения числа ссылок. Временем полужизни описывают старение публикаций часто и в том случае, когда это старение описывается не экспонентой, а как-то иначе (о других аппроксимациях кривых старения см. далее). Так поступают, например, Р. Бартон и Р. Кеблер.

Время полужизни публикаций зависит от предметной области, как это показывает табл. 4.2.1, в которой суммированы результаты некоторых работ.

Иногда оказывается невозможным аппроксимировать эмпирические данные по старению литературы одной экспонентой, но бывает достаточно суммы двух или большего числа экспонент (см., например, рис. 1.32). Суммой двух экспонент пользуются Р. Бартон и Р. Кеблер [4.50], Д. Прайс [4.24], А. Мидоус [4.113], А. Аврамеску [4.35], Б. Брукс [4.4], В. А. Маркусова [4.16] и др. Аналитический вид такой аппроксимации, предложенной впервые Р. Бартоном и Р. Кеблером, следующий [4.50]:

$$y = 1 - (a/e^x + b/e^{2x}), \quad (4.2.2)$$

где x — время в десятилетиях; y — доля литературы по данному предмету, цитируемая через определенное время. Соответственно в научной литературе выделяют фронты «классической» медлен-

Таблица 4.2.1

Время полужизни публикаций в различных науках

Наука	Время полужизни (в годах)	Источник	Наука	Время полужизни (в годах)	Источник
Социальная практика [customs]	37,0	[4.71]	Биология	5,0	[4.141a]
Экономика	33,0	[4.71]	Социальное обеспечение	5,0	[4.71]
География	16,0	[1.141]	Социология	5,0	[4.108]
Геология	11,8	[4.50]	Вычислительная математика	4,95	[4.11]
Ботаника	10,0–11,8	[4.50]–[4.137]	Техническая химия	4,8	[4.50]
Химия	8,1	[4.50]	Философия	4,5	[4.4]
Математика	5,42–10,5	[4.11]–[4.50]	Физика	3,9–4,6	[4.137]–[4.50]
Физиология	7,2	[4.50]	Металлургия	3,9	[4.50]
Инженерная механика	5,2	[4.50]	Биомедицина	3,0	[4.108]

но меняющейся и «эфемерной» или оперативной быстро стареющей литературы.

Модернизация выражения Р. Бартон и Р. Кеблера (4.2.2) содержится в работе А. Аврамеску [4.36], который старение «классической» литературы по-прежнему описывает экспонентой:

$$y_c = y_0^c \exp(-\tau/T_c), \quad (4.2.3)$$

где y_c — доля старых ссылок старше τ ; $y_0^c = 0,02$ и $T_c = 20$ годам; старение же оперативной литературы Аврамеску аппроксимирует кривой, дающей более резкий спад:

$$y_n = \exp[\ln y_0^{(n)} \exp(-\tau/T_n)], \quad (4.2.4)$$

где y_n — доля новых ссылок старше τ ; $y_0^{(n)} = 6$ и $T_n = 10$ годам. Для старения всех статей отсюда получаем выражение:

$$y = y_0^{(c)} \exp(-\tau/T_c) + \exp[\ln y_0^{(n)} \exp(-\tau/T_n)], \quad (4.2.5)$$

где y — доля всех ссылок старше τ .

* Аврамеску, вместо того, чтобы складывать кривые (4.2.3) и (4.2.4), как это делают, например, Бартон и Кеблер в линейных координатах, скла-

дывает их в координатах логарифмических, приводя графики в логарифмических координатах и получая вместо (4.2.5) выражение

$$y = y_0^{(c)} \exp[-\tau/T_c + \ln y_0^{(n)} \exp(-\tau/T_n)].$$

Он так и пишет: «Чтобы сложить кривые с n и n (т.е. (4.2.3) и (4.2.4) соответственно), «мы должны умножить» (4.2.3) и (4.2.4) [4.36, с. 354]. Это неверно, так как доля всех статей старше τ равна сумме долей старых и новых статей старше τ , а не их произведению, т.е. $y = y_c + y_n$ и $y \neq y_c y_n$.

Существует ряд более сложных, чем экспонента или сумма экспонент, аппроксимаций для описания старения литературы. Обзор аппроксимаций можно найти в работах [4.105; 0,65]. Мы же здесь назовем только основные из них.

Р. Койл [4.56], Д. Макрей [4.108], М. Лайн [4.104], Т. Краузе и К. Халлингер [4.101] и другие учитывают влияние на старение литературы роста массива публикаций. Д. Макрей [4.108], предполагая рост литературы экспоненциальным, а вероятность цитирования статьи — не зависящей от объема растущего массива работ, описывает старение литературы кривой

$$f(t) = K \exp[(k+m)(t-t_p)], \quad (4.2.6)$$

где t — текущее время; t_p — время опубликования цитируемой статьи; k — мера роста объема литературы; m — мера предпочтения в цитировании более недавней информации. При этом Д. Макрей полагает, что вероятность цитирования падает со временем по экспоненте.

М. Лайн и Т. Краузе и К. Халлингер отказываются от предположения Д. Макрея о независимости вероятности цитирования статьи от объема массива литературы, но делают это по-разному. М. Лайн [4.104] считает, что новая литература доступнее, и потому ученый ссылается на нее при прочих равных условиях чаще. Т. Краузе и К. Халлингер [4.101] полагают, напротив, что число прочитанных ученым работ, определяющее вероятность цитирования, растет медленнее, чем массив публикаций, и потому вероятность прочтения, а следовательно, и цитирования статьи падает с ростом объема массива публикаций.

Учет роста популяции ученых наряду с ростом литературы содержится в работах Б. Брукса [4.46–4.48]. При этом и рост массива публикаций, и рост числа ученых Б. Брукс берет экспоненциальным. Характеризуя старение литературы зависимостью $s=a^t$, где коэффициент старения $a < 1$ и в любой год число ссылок на работу составляет долю a от числа ссылок на нее в предыдущем году, так что общее число ссылок на работу за все время ее

содержится в работах Б. Брукса [4.46–4.48]. При этом и рост массива публикаций, и рост числа ученых Б. Брукс берет экспоненциальным. Характеризуя старение литературы зависимостью $s=a^t$, где коэффициент старения $a < 1$ и в любой год число ссылок на работу составляет долю a от числа ссылок на нее в предыдущем году, так что общее число ссылок на работу за все время ее

жизни определяется выражением

$$n(1+a+a^2+\dots)=n \cdot 1/(1-a), \quad (4.2.7)$$

(n — число ссылок в момент публикации работы; лаг между временем публикации и временем, за которое она поступает к потребителю, не принимается во внимание), Б. Крукс получает выражение

$$1-a_1=(1-a_0)(g/s)', \quad (4.2.8)$$

связывающее коэффициенты старения в годы t и 0. Здесь g и s — параметры роста соответственно объема литературы и популяции ученых. М. Оливер нашла g и s для физики полупроводников равными соответственно 13,4 и 15,0% в год [4.121].

Как марковский процесс описывают старение литературы Е. Морз и К. Элстон [0.101], Д. Крафт и Р. Полачек [1.122] на основе системы дифференциальных уравнений строят математическую модель, охватывающую рост и старение литературы и рассеяние публикаций по журналам. Упомянем еще качественную работу Б. Нолтинка [4.119], в которой даются пять мысленных кривых старения литературы, отвечающих работам разной значимости — малозначимым, фундаментальным теоретическим, гениальным и пр. [1.46, с. 102].

Иногда эмпирические данные по старению литературы приводятся вне связи с какой-либо их определенной аппроксимацией. Например, Д. Прайс [4.25] вводит индекс, представляющий собой отношение количества ссылок на литературу не более чем 5-летней давности к общему количеству ссылок. По науке в целом индекс Прайса оказывается равным 0,5, для физики и биохимии — 0,6–0,7, для радиологии — 0,58, для социологии — 0,465 и т. д. Д. Прайс заметил, что в точных науках его индекс имеет большие значения, чем в гуманитарных. Поэтому он предлагает использовать свой индекс как количественный критерий, позволяющий различать точные и неточные науки. По Прайсу, точные науки имеют индекс Прайса больше 0,42.

Г. М. Добров [4.8], Н. Лин и К. Нельсон [4.103], Р. Бродус [4.44] и другие, обсуждая эмпирические данные по старению литературы, приводят средний возраст ссылок. Иногда же, как это делают У. Деннис [4.67], Л. Кофнофек [4.99], П. Атанасу [4.2], Р. Бродус [4.44], Р. Мартино [4.112], Дж. Баугман [4.37], Э. К. Ваттер [4.6] и другие, данные по старению литературы никак не параметризуются и приводятся в их «первозданном» виде, т. е. эти авторы говорят, что такой-то процент ссылок старше m лет, а такой-то — n .

К. Борн [4.43], О. Гроос [4.88] и другие отмечают слабую согласованность разных мер старения литературы между собой. Это, а также существование большого числа разных аппроксимаций кривых старения говорит об *отсутствии на сегодняшний день удовлетворительной аппроксимации кривых старения*, позволяющей описывать единым образом все эмпирические данные. И это вполне объяснимо. В основу различных математических моделей старения литературы кладутся экспоненты роста объема массива публикаций и числа ученых, и падение интереса ученых к литературе также полагается падающим со временем по экспоненте. Трудно сказать что-либо определенное относительно характера кривой падения интереса, кривые же роста литературы и популяции ученых только в немногих случаях являются экспонентами (см. гл. 2, § 2; гл. 3, § 2; гл. 7, § 1). Наиболее «чистый» вид имеют кривые роста для очень больших массивов, скажем для науки в целом. И в этом случае они имеют не экспоненциальный, а логистический вид, лишь на определенном интервале дающий экспоненту. Для отдельных же предметных областей кривые роста экспонентой, вообще говоря, не описываются. Поэтому и математические модели, построенные на экспонентах роста (а почти все существующие сейчас модели старения литературы таковы), имеют ограниченное значение.

На наш взгляд, отсутствие удовлетворительной аппроксимации эмпирических данных по старению научной литературы не означает бесперспективности вообще исследований старения литературы. Если даже такая удовлетворительная аппроксимация и не будет найдена, эти данные могут быть использованы в интегральном виде. Информация, поставляемая таким образом, представляет собой интерес для информационных служб науки. Уже сейчас, например, из того факта, что научная литература, как показали исследования ее старения, очень быстро выбывает из сферы потребления ее учеными, сделан вывод о необходимости максимально возможного ускорения процесса публикации работ.

3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: СТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

Распределение журналов по их цитируемости

Из попавших в наше поле зрения наукометрических работ цитируемость журналов обсуждается в 21, авторами которых являются П. Гросс и Е. Гросс [4.89], П. Гросс и А. Вудфорд [4.90], Х. Фаслер [4.73], Р. Койл [4.57], Р. Бартон [4.51], Л. Райзинг [4.128], Ю. Гарфилд и Дж. Шер [4.83], Ю. П. Адлер и соавт.

[1.1], П. Ерл и Б. Викери [4.71], Д. М. Мехтиев и соавт. [4.20], Ю. Гарфилд [4.77], А. П. Борисова и соавт. [1.4] и др.

До появления статьи Л. Райзига [4.128] строились лишь распределения журналов по числу ссылок на них. Типичные данные такого рода показаны на рис. 2.32. Начиная с Л. Райзига, строятся еще и распределения журналов по числу ссылок на них, деленному на число содержащихся в этих журналах публикаций. Данные такого рода приведены на рис. 2.33. Цитатные распределения журналов могут быть аппроксимированы цифровыми распределениями.

Распределение научных учреждений по их цитируемости

Цитируемость научных учреждений рассматривают Дж. Вестбрук [4.142], Дж. Коул [1.33] и др. Дж. Вестбрук дал распределение ссылок в работах, посвященных исследованию керамических материалов, по лабораториям, в стенах которых они выполнены. Это распределение, как показывает рис. 2.34, может быть аппроксимировано распределением Ципфа с $\alpha \approx 1$, говорящим о непропорционально большой роли в науке ведущих научных подразделений по сравнению с рядовыми.

О непропорционально большой концентрации научных интеллектов в ведущих исследовательских центрах еще более ярко говорят данные Дж. Коула по цитированию американскими физиками 1200 американских университетов и колледжей. «60% ссылок, — пишет Коул, — делаемых физиками, которым принадлежат работы низшего качества (здесь: малоцитруемые работы. — С. Х.), относятся к физикам, работающим в одном из 37 высших подразделений. 36% этих ссылок относятся к физикам, сосредоточенным в девяти высших подразделениях, по сравнению со всего лишь 18% ссылок на ученых из низших подразделений» [1.33, с. 406].

Еще более выборочно цитируют коллег крупные ученые. «Возьмем для примера, — говорит Дж. Коул, — Нобелевских лауреатов и членов национальной академии... Представители этой группы лишь в 9% случаев ссылаются на ученых из подразделений низшего ранга, а в 58% случаев цитируют ученых, так или иначе связанных с «высшей девяткой». Подобным же образом, цитирующие из видных подразделений и ученые высокого ранга (здесь: часто цитруемые ученые. — С. Х.) преимущественно используют работы сотрудников наиболее выдающихся подразделений. Наконец, только 10% ссылок у физиков, проводящих исследования высшего качества (здесь: часто цитруемые. — С. Х.),

относятся к ученым, работающим в более мелких подразделениях, в то время как около 60% ссылок делается на ученых из основных подразделений» [1.33, с. 405].

Данные о том, что цитируются чаще всего работы ученых из немногочисленных ведущих научных подразделений, носят оптимистический характер: в интеллектуальном отношении ученые разных подразделений не могут существенно отличаться, поэтому, очевидно, популяция ученых в целом таит в себе громадные интеллектуальные резервы.

Распределение ученых по их цитируемости

Исследование цитируемости ученых проводят П. И. Вальден [см. 4.1], Д. Прайс [1.58], З. Б. Баринаева и соавт. [4.3], Р. Ф. Васильев [1.6], С. Коул и Дж. Коул [4.62], И. М. Ориент [4.23], Ю. П. Адлер и соавт. [1.1], В. В. Налимов и З. М. Мульченко [1.46], Дж. Коул [1.33], С. Коул [4.61], Ю. Гарфилд [4.76], Дж. Коул и С. Коул [4.59], М. Ороманер [4.123], Ю. В. Грановский и соавт. [1.20], Н. Вэйд [4.138], М. П. Великова и З. Г. Залесова [1.3], Д. Прайс [1.143], Ю. Гарфилд [4.79; 4.80] и др. Типичные данные приведены на рис. 2.35, 2.36. Д. Прайс предлагает распределение ученых по числу ссылок на них в числе других цитатных распределений аппроксимировать бета-функцией [4.143]. Утверждая, что все эти распределения следует описывать цифровыми аппроксимациями, к числу которых относятся и бета-функция, мы не разделяем, однако, уверенности Прайса в том, что бета-функция имеет какой-либо приоритет. Дело в том, что, хотим мы этого или не хотим, форма эмпирических цитатных распределений (как и вообще наукометрических стационарных распределений) в области малых значений переменных существенно зависит от способа формирования выборки. Детальнее об этом мы будем говорить при обсуждении других цитатных распределений и в гл. 7, § 2.

Качественно цифровое распределение говорит о непропорционально высоком цитировании крупных ученых. Дж. Коул [1.33] и Дж. Коул и С. Коул [4.59] приводят данные, особенно ярко говорящие о непропорционально большой роли в науке ведущих ученых (табл. 4.3.1, 4.3.2).

Комментируя табл. 4.3.1, Дж. Коул говорит: «Если исследовать процент ссылок на ученых, работы которых цитируются с той же частотой, что и работы Нобелевских лауреатов (60 и более ссылок), то обнаружатся устойчивые и важные результаты. Почти половина ссылок, приводимых в работах высшего качества (т. е. создаваемых наиболее продуктивными, знаменитыми физи-

Таблица 4.3.1

Цитируемость авторов и работ [1.33]

Цитируемость цитирующих авторов	Цитируемые авторы с 60 и более ссылками, %	Цитируемые авторы с 15 и более ссылками, %
60 и более ссылок	51	81
15-59 ссылок	39	68
0-14 ссылок	35	60

Таблица 4.3.2

Цитируемость цитируемых авторов и цитирующих работ [1.33]

Цитируемость цитируемых авторов	Число ссылок на «лучшую» статью цитируемого автора		
	20 и более	10-19	0-9
60 и более ссылок	54	48	33
15-59 ссылок	28	30	36
0-14 ссылок	18	22	31
	100%	100%	100%

ками), приходится на тех ученых, которые по крайней мере по уровню цитирования считаются равными лауреатам и членам Национальной академии наук. Разумеется, многие из цитируемых авторов являются лауреатами. Во 2-м столбце табл. 4.3.2, если принять в расчет долю ссылок разных типов ученых на работы среднего уровня (15 и более ссылок), мы получим еще более поразительные результаты. 82% ссылок в высшей степени «заметных» ученых, 78 — лауреатов и академиков, 81 — авторов работ высшего качества по нашей шкале и 74% ссылок сотрудников наиболее крупных исследовательских подразделений приходится исключительно на ученых, чьи работы упоминаются в «Science Citation Index» за 1965 г. по крайней мере 15 раз» [1.33, с. 408].

Без проведения опорных исследований числа ссылок как индикатора вклада ученого в науку невозможно определенно сказать, в какой степени данные по цитируемости ученых характеризуют их вклады. Могут быть высказаны, например, соображения, связанные с так называемым эффектом Матфея [0.100], в пользу того, что цитатные данные завышают роль ведущих ученых. В самом деле, в применении к цитатам этот эффект проявляется,

как это констатировал С. Коул [4.61], в том, что при равном качестве статьи высокоранжированных ученых цитируются чаще, чем статьи низкоранжированных (см. § 1 рассуждения о вносимом методом ссылок субъективном «перекосе»).

Кроме того, часть наукометрических данных Дж. Коула, о которых мы здесь говорим, получена методом анкетирования: 2036 физикам, рассредоточенным по 86 ведущим подразделениям, были разосланы анкеты, на которые 1308 физиков прислали ответы [1.33, с. 397]. При такой постановке дела в используемой Дж. Коулом выборке ученых возможен «перекос» в сторону завышения доли крупных физиков: во-первых, корреспонденты невольно могут исказить истинное положение дела в свою пользу; во-вторых, как правило, не отвечают на анкеты те, кто «слабее». Из-за этой методологической особенности данных Дж. Коула (наукометрические данные вообще нецелесообразно получать социологическими средствами, это снижает объективность результатов) в их трактовку вносится дополнительная неопределенность, снятие которой, возможно, привело бы к их еще более резкой стратификации.

На основании своих данных Дж. Коул и С. Коул [4.59] приходят к слишком, на наш взгляд, сильным выводам. Они полемизируют с известным испанским философом Хосе Ортега-и-Гассетом [0.102]. «Приходится наставлять на том удивительном, но неоспоримом факте, говорит Ортега, — что экспериментальные науки развиваются преимущественно благодаря работе людей поразительно посредственных... Другими, словами, современная наука, корень и символ существующей цивилизации, находит место для интеллектуально ординарных людей и позволяет им работать с успехом. Таким образом, большинство ученых способствует общему прогрессу науки, замыкаясь в тесных кельях своих лабораторий, подобно пчеле в ячейке ее улья...» (цит. по: 4.59)].

Оспаривая этот взгляд, Дж. Коул и С. Коул пишут, что подавляющее большинство ссылок в работах часто цитируемых авторов составляют ссылки на ученых с индексом цитирования, также значительно превышающим средний. Даже в статьях относительно малого значения преимущественно используются работы выдающихся (по частоте цитирования) ученых. Эти данные позволяют Коулам подвергнуть сомнению точку зрения Ортеги. Более обоснованной, по их мнению, представляется противоположная позиция: относительно небольшое число физиков своими работами закладывает базу для дальнейшего развития науки. Поэтому Дж. Коул и С. Коул предполагают, что темпы научного прогресса в физике могут остаться на прежнем уровне, даже если число

активных физиков-исследователей будет резко сокращено [4.59].

Дж. Коул и С. Коул, на наш взгляд, убедительно демонстрируют опасность слепой, не подкреплённой опорными исследованиями веры в результаты количественного, в данном случае наукометрического, анализа. Число ссылок — это только один из наукометрических индикаторов индивидуальных вкладов ученых в науку. И делать столь категоричные выводы на основании эмпирических данных по цитатному анализу нельзя. Несомненно, учет других индикаторов вклада ученых и результатов опорных исследований по связи этих индикаторов с латентной переменной научного вклада приведет к выводу о существенно большей роли средних ученых, чем это представляется при рассмотрении одних только данных цитатного анализа, вносящих, как говорилось выше, систематический «перекос» в пользу крупных ученых.

Но этим не ограничивается ошибка Дж. и С. Коулов. Дело в том, что *ведущие ученые*, действительно вносящие непропорционально большой вклад в науку, не могут работать иначе, как *будучи погруженными в массу средних ученых*.

Следует сказать, что концепции Ортеги и Коулов характерны для научной мысли Запада. Согласно концепции Ортеги, индивид — винтик механизма науки; в концепции Коулов, напротив, ученый полностью противопоставлен науке. Такая амплитуда и такие крайности есть следствие общего недостатка философской мысли Запада: непонимания того, что субъект не противопоставлен обществу, что следует говорить об *общественном субъекте*. Поэтому задача изучения *противопоставления* индивидуального и коллективного в науке и научном творчестве должна быть заменена задачей изучения *соотношения* индивидуального и коллективного в науке.

В частности, из факта непропорционально большой роли в науке ведущих ученых следует вполне оптимистический вывод о необходимости уделения максимального внимания научным талантам. Если мы действительно хотим интенсифицировать науку, то различные средства, направляемые на реализацию этой задачи, должны распределяться пропорционально индивидуальным вкладам ученых в науку, т. е. в соответствии с цифровым распределением, которым, судя по всему, описывается распределение ученых по научному вкладу. Другими словами, *в науке чрезвычайно выгодно вкладывать средства в научные таланты*, в их диагностику, прогнозирование и интенсификацию.

Распределение числа ссылок на работу по числу соавторов

Нам известны две работы такого рода. Б. Липетц [2.54] нашел, что работы коллективов авторов цитируются в 1,5 раза чаще работ солоавторов. С. Коул [4.61] установил, что соавторы часто цитируемых ученых при равном качестве работ цитируются чаще других ученых.

Распределение публикаций по числу содержащихся в них ссылок

Распределения такого рода анализируют Д. Прайс [1.58], Н. Каплан [4.94], Д. Прайс [4.24], А. А. Коренной [1.32], У. Менард [1.125], Ю. В. Грановский и соавт. [1.17], Ю. В. Тимошин [4.28], С. А. Заремба и И. М. Алабян [2.13], Д. М. Мехтиев и соавт. [4.21], В. А. Маркусова [4.16], А. Ф. Клементьев [4.11] и др. Типичные эмпирические данные показаны на рис. 2.37. Эти данные, как утверждает Д. Прайс [1.58, с. 350–351], для большого числа ссылок могут быть аппроксимированы обратной степенной зависимостью с показателем $n=2$, т. е. это распределение может быть аппроксимировано цифровым распределением с $\alpha=1$.

В своих последних работах [1.143; 1.144; 4.127] Д. Прайс утверждает, что распределение публикаций по числу ссылок в них, как и другие цитатные распределения, следует аппроксимировать бета-функцией. Такое предпочтение какому-нибудь одному цифровому распределению нам представляется не вполне оправданным. Дело в том, что в области малых значений переменных (а различные цифровые распределения отличаются друг от друга именно поведением в этой области) форма распределения публикаций по числу ссылок в них в большой степени является *артефактом*, порожденным невоспроизводимым «обрезанием» выборки. В самом деле, индикатор «число ссылок в публикациях» измеряет, очевидно, латентную переменную «в какой степени индивид опирается при высказывании своих соображений на соображения предшественников» или, короче, «преемственность соображений». Но ученый тем и отличается от неученого, что сознательно использует достижения предшественников, это — один из принятых критериев научности высказанных данным индивидом соображений. Обычный человек не склонен в своих суждениях систематически опираться на предшественников и тем более ссылаться на них. Поэтому соответствующее латентное распределение даст бесконечное число индивидов, соображения которых имеют нулевую осознанную преемственность с соображениями

Таблица 4.3.3

Среднее число ссылок на публикацию в некоторых научных дисциплинах

Дисциплина	Среднее число ссылок	Источник
Химия	18,7	[4.94]
Математика	6,1–8,2	[4.11]
Обогащение полезных ископаемых	6,4	[2.13]
Медицина	4,3–4,5	[4.94]
Геофизическая разведочная техника	4,03	[4.11]

предшественников. Таким образом, легко понять, что распределение индивидов (не обязательно ученых) по числу ссылок на соображения своих предшественников будет иметь практически бесконечно большое число индивидов с нулевым числом таких ссылок. Следовательно, распределение окажется, судя по всему, распределением Ципфа.

Если же взять выборку ученых, которые по своему статусу должны указывать предшественников развиваемых ими положений, то, очевидно, их распределение по тому же индикатору будет иметь заведомо небольшое число индивидов с нулевым числом ссылок. И потому такое распределение в области малых значений числа ссылок отклонится от распределения Ципфа. Насколько отклонится, зависит от того, каким образом мы формируем выборку ученых, насколько жесткими окажутся наши критерии ученого в отличие от неученого. Таким образом, форма распределения публикаций по числу ссылок в них в области малых значений числа ссылок определяется не только и не столько природой исследуемых объектов (ученых, публикаций, ссылок), сколько используемой данным исследователем конкретной методикой формирования выборки ученых.

В ряде работ определяется среднее для данной научной дисциплины число ссылок на статью (табл. 4.3.3).

Некоторые авторы (см., например, работы Д. Прайса [4.24] и А. А. Коренного [1.32]) говорят о снижении нормы цитирования при переходе от естественных наук к гуманитарным.

В целом по науке Д. Прайс [1.58] нашел среднее число ссылок на статью равным 15, В. А. Маркусова [4.16] и А. Ф. Клейментьев [4.11], учитывая данные по естественным, точным и техническим наукам, — 10,3.

Следует сказать, что все эти данные в принципе некорректны.

Среднее число — это момент первого порядка статистического распределения, распределение же статей по числу ссылок в них — ципфовое с $\alpha \approx 1$, и потому этот момент стремится к бесконечности с увеличением объема выборки. Другими словами, величина среднего для данного распределения зависит от объема выборки, что делает его применение некорректным.

О зависимости среднего числа ссылок в статье от объема выборки говорит Д. Прайс в своих последних работах [1.143], [1.144] и [4.127]. Вот что он пишет о средних числах ссылок, получаемых для Индекса Гарфилда по естественным наукам (SCI) и Индекса Гарфилда по общественным наукам (SSCI): «Во всяком случае, среднее число ссылок в статье определяется объемом доступного архива литературы в данной области. Именно число ссылок на статью должно быть равно маленькой константе (меньшей единицы) плюс натуральный логарифм объема архива. Натуральный логарифм миллиона составляет примерно 14 и вот почему число ссылок таково.

Для Индекса ссылок по социальным наукам (SSCI) соответствующее число составляет около 11 ссылок на статью, что соответствует архиву около 60 000 статей в каждой области. Как для SCI, так и для SSCI число ссылок на статью увеличивается с ростом архива» [4.127, с. 12–13].

Заметим, что в приведенном пассаже Прайс не различает среднее число ссылок на статью и среднее число ссылок в статье. Он это делает сознательно, исходя из предположения, что эти средние равны, как он говорит, для «общей цитатной сети для всех времен» [1.143, с. 301]. То, что бесспорно верно для «общей цитатной сети», вообще говоря, неверно для произвольной выборки публикаций. Приравнивая указанные средние друг другу для архивов SCI и SSCI, Прайс тем самым предполагает, что соответствующие выборки публикаций достаточно полны. Полны в том смысле, что ссылки, содержащиеся в выборке, не выходят за ее пределы, замкнуты целиком на эту выборку. А это, надо сказать, довольно-таки сильное предположение. Но не в нем суть.

Качественно Прайс описал картину зависимости среднего числа ссылок от объема выборки верно. Однако даваемый им способ расчета этой зависимости не совсем корректен.

В самом деле, для общего числа ссылок G на данной выборке публикаций объема N , $G = \bar{x}N$ ($\bar{x}$ — среднее число ссылок) Прайс дает выражение

$$G \approx N(C - 1 + \ln N) \quad (4.3.1)$$

где $C = 0,577215665...$ — постоянная Эйлера. Отсюда для $\bar{x}$ имеем

$$\bar{x} \approx C - 1 + \ln N. \quad (4.3.2)$$

Именно это выражение и описывает Прайс в приведенном выше отрывке. Он не дает его вывода, однако легко понять, как Прайс его получает. Он аппроксимирует распределение публикаций по числу ссылок в них бета-функцией

$$n(x) = N(m+1) \frac{\Gamma(x) \Gamma(m+2)}{\Gamma(x+m+2)} \quad (4.3.3)$$

с $m=0$, т. е. распределением

$$n(x) = N \cdot 1/[x(x+1)]. \quad (4.3.4)$$

Находим

$$x = \sum_1^I \frac{x n(x)}{N} = \sum_1^I \frac{1}{x+1} \approx C - 1 + \ln(I+1), \quad (4.3.5)$$

где $I=x_{\max}$ — максимальное на данной выборке значение x .

Видим, что выражение Прайса (4.3.2) получается, если полагать

$$I \approx N. \quad (4.3.6)$$

Прайс так и говорит: «Для общей цитатной сети для всех времен ясно, что максимальное число успехов на статью равно общей совокупности статей» [1.143, с. 301]. Под «успехами» Прайс понимает при этом число цитирований статьи плюс единица, относящаяся к самому факту ее публикации.

Некорректность выкладок Прайса заключается в замене I на N в выражении (4.3.5) (ср. с выражением (3.3.16)). В общем случае I зависит от N вероятностным образом, так что данному объему выборки N может отвечать, вообще говоря, любое значение I . Поэтому в (4.3.5) следует подставлять не какое-нибудь, но выборочное, характерное для данной выборки значение I , т. е. в данном случае максимальное на выборке значение Индекса Гарфилда (как одного, так и другого) число ссылок в одной публикации. Это число может составлять примерно 5 тыс., но никак не 1 млн. Подставляя в (4.3.5) это значение I , находим $\bar{x} \approx 8$, что уже, конечно, не очень хорошо согласуется с фактическими 14 ссылками в среднем в одной статье.

Расхождение теории и практики в данном случае объясняется, видимо, некорректностью аппроксимации распределения публикаций по числу ссылок в них бета-функцией. В самом деле, согласно этой бета-функции число статей, в которых нет ни одной ссылки, бесконечно велико, а число статей с одной ссылкой составляет 33,3% от общего числа статей (см. выражение (4.3.4)). Фак-

тически же доля статей с малым содержанием ссылок существенно скромнее (см. рис. 2.37). Именно за счет занижения доли статей с малым числом ссылок бета-функция и запирает среднее число ссылок, содержащихся в статье.

Распределение публикаций по их цитированию

Цитируемость публикаций изучают Ю. Гарфилд [4.75–4.84], Д. Прайс [1.58; 4.125; 4.24; 1.143; 1.144; 4.127], Дж. Марголис [4.109], Дж. Коул [1.33], С. Наранан [1.131], Ю. В. Грановский и соавт. [1.48], Дж. Коул и С. Коул [4.59], Д. Крэйи [1.90], М. Ороманер [4.123], Г. Магир [1.123], М. К. Петров [1.54], М. П. Великова и З. Г. Залесова [1.3], Д. Дикс и Х. Чанг [4.69], С. Д. Хайтун [4.32], Ю. В. Грановский [1.15] и др. Типичные данные показаны на рис. 2.38–2.40.

Только немногие авторы аппроксимируют данное распределение гауссовыми распределениями. М. П. Великова и З. Г. Залесова [1.3] аппроксимируют свои данные, построенные на небольших выборках, распределением Пуассона точно так же, как они это делают и с распределением ученых по числу публикаций (см. гл. 3, § 3). Эта аппроксимация некорректна по тем же соображениям, которые были высказаны в отношении аппроксимации распределения Лотки распределением Пуассона.

К. Кроули [4.66] аппроксимирует распределение публикаций по их цитированию отрицательным биномиальным распределением.

Гораздо большее число авторов связывают обсуждаемое распределение с распределением Ципфа. А. Платц [1.138] и Г. Магир [1.123] говорят о возможности применения здесь закона Лотки, т. е. распределения Ципфа с $\alpha=1$. Р. Фэйтхорн [1.98], С. Наранан [1.131], Дж. Букштайн [1.78–1.79], Д. Дикс и Х. Чанг [4.69], А. И. Яблонский [0.63] описывают распределение публикаций по их цитированию распределением Ципфа в общем его виде, т. е. с произвольным значением показателя α . Д. Прайс [1.143; 1.144] предлагает аппроксимировать данное распределение бета-функцией.

По-видимому, все эмпирические распределения публикаций по числу ссылок на них могут быть успешно аппроксимированы при больших значениях переменной *распределением Ципфа* с α в диапазоне 0,8–1,8. Значение, принимаемое α , и форма распределения в области малого числа ссылок существенно зависят от способа формирования выборки. В частности, есть основания полагать, что, чем более жесткими критериями мы руководствуемся при формировании выборки публикаций, тем меньше α и тем

Таблица 4.3.4

Цитируемость цитирующих и цитируемых работ [1.33]

Цитируемость цитирующих работ	Цитируемые работы с более чем 15 ссылками, %	Цитируемые работы с 0—14 ссылками, %	Итого, %
60 и более ссылок	86	14	100
15—59 ссылок	75	25	100
0—14 ссылок	67	33	100

больше эмпирическое распределение отличается в области малого числа ссылок от распределения Ципфа. Рис. 2.38—2.40 иллюстрируют это положение: с переходом от выборки просто публикаций к выборке публикаций наиболее цитируемых ученых мы наблюдаем указанные здесь изменения в распределениях. Поэтому отдавать предпочтение какой-либо одной цифровой аппроксимации (например, бета-функции) вряд ли целесообразно. Возможность описания эмпирических распределений публикаций по их цитируемости распределением Ципфа говорит о непропорционально большой роли ведущих публикаций. Эта непропорциональность подчеркивается данными Дж. Коула [1.33].

«Временной лаг» цитирования

З. Б. Барина с соавт. [4.3] и другие изучают так называемую зону библиографического молчания для разных стран. В нашей стране «зона молчания» зарубежных публикаций достигает нескольких лет. Это явление связывают обычно с языковыми барьерами и значительной в некоторых изданиях задержкой публикаций (см. гл. 3, § 3).

«Экзотические» распределения ссылок

Ю. П. Адлер и соавт. [1.1] дают распределение пар публикаций по числу совпадающих ссылок в них. Как показано на рис. 2.41, это распределение может быть аппроксимировано цифровым распределением.

Распределение ссылок по предмету

Распределения ссылок по предмету изучают Х. Фасслер [4.73], Р. Бартоп [4.51], З. Б. Барина и соавт. [4.3], И. М. Ориент [4.23], П. Ерл и Б. Викери [4.70; 4.71], Е. Крэйг [4.64], А. Митра [4.115] и др. Типичные эмпирические данные такого рода по-

казаны на рис. 2.42 и 2.43, где они аппроксимированы цифровыми распределениями.

Географические распределения ссылок

Распределения ссылок по странам изучают Х. Фасслер [4.73], Р. Бартоп [4.51], З. Б. Барина и соавт. [4.3], П. Атанаску [4.2], В. В. Налимов и З. М. Мульченко [4.46], С. Дхаван [4.68], М. Г. Фролова [4.30], М. К. Лучник и В. Д. Осетров [1.39], М. П. Беликова и З. Г. Залесова [1.3], Г. М. Добров и А. А. Коренной [2.11] и др. Типичные эмпирические данные показаны на рис. 2.30. Эти распределения могут быть аппроксимированы цифровыми распределениями.

Распределение ссылок по языку

Данные такого рода приводят К. Луттит [4.107], Е. Крэйг [4.64], В. А. Полушкин и Г. С. Жданова [1.56], Дж. Баутман [4.37], Д. М. Мехтнев и соавт. [4.21] и др. Эти данные, показанные на рис. 2.31, могут быть аппроксимированы цифровым распределением.

Распределение ссылок по типу издания

Данные по распределению цитируемых работ по типу издания дают У. Гуттсман [4.91], Д. Прайс [4.24], П. Атанаску [4.2], П. Ерл и Б. Викери [4.71], Х. Ест и А. Вейман [4.72], В. А. Маркусова [4.16], Дж. Баутман [4.37], Д. М. Мехтнев и соавт. [4.21] и др.

4. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ: СТРУКТУРНЫЕ ИССЛЕДОВАНИЯ

В наше поле зрения попало 35 работ, в которых структура объектов науки изучается средствами цитатного анализа.

Как говорилось в гл. 2, § 4, количественный анализ структуры данной системы предполагает измерение связанности любых двух подсистем. В данном случае объектом анализа выступает поток публикаций, связь между частями которого и фиксируется посредством числа ссылок. Взаимосвязь элементов потока публикаций фиксируется при этом пока тремя способами.

I. Сила связи одного элемента потока публикаций (например, отдельной публикации) с другим определяется числом ссылок, содержащихся в первом из них, на работы, содержащиеся во втором. Связь при этом имеет векторный характер.

Таблица 4.4.1

Доля самоцитирования в различных предметных областях [4.70]

Предметная область	Самоцитирование, %	Предметная область	Самоцитирование, %
Математика	79	Ботаника	86
Астрономия	62	Зоология	38
Физика	75	Медицина	61
Химия	73	Инженерные науки	70
Геология и метеорология	52		

Такого рода цитатную связь используют Дж. Бернал [0.70], Ю. Гарфилд и соавт. [4.84], П. Ерл и Б. Викири [4.70], В. В. Налимов и З. М. Мульченко [1.46], Ю. Гарфилд и А. Коукелл [4.82], А. А. Коренной [4.12–4.13], Е. Лайтфилд [4.102], А. Мидоус и Дж. О'Коннор [4.114], Т. Волдхарт [4.139], А. Мидоус [1.24] и др.

Сопоставление сетей цитирования с сетями связей, выявляемыми средствами качественного историко-научного анализа, производят Дж. Бернал [0.70], Ю. Гарфилд и соавт. [4.84], А. А. Коренной [4.12; 4.13], и др. На рис. 3.2 представлены результаты одной из таких работ, показавших, что число ссылок является хорошим индикатором взаимовлияния научных работ, позволяющим не только выявлять большую часть связей, различаемых историками науки, но и фиксировать связи, упускаемые ими из виду.

Самоцитирование разных предметных областей изучают П. Ерл и Б. Викири [4.70], А. Мидоус и Дж. О'Коннор [4.114] и др. (табл. 4.4.1). При этом относительная доля самоцитирования рассматривается как индикатор быстрорастущих научных направлений, в которых эта доля выше, чем в традиционных областях. Например, такую бурно развивающуюся область науки, как литература о пульсарах, характеризует большее самоцитирование, чем такую традиционную науку как астрономия [4.114].

Заметим, что самоцитирование может быть связано с ростом научного направления отнюдь не так однозначно, как это представляется авторам указанных работ. А. П. Огурцов, например, связывает рост самоцитирования научной школы с ее угасанием. «Естественно, — говорит он, — что приверженцы одной теории или идеи, объединенные в некоторую научную школу, ссылаются

друг на друга намного чаще, чем на ученых из других научных школ. Эта тенденция достигает своей карикатурной формы при разложении научной школы. Здесь взаимные ссылки на самих себя, на лидера научной школы представляют единственный цитируемый материал...» [0.47, с. 254]. Ясно, что без соответствующих опорных исследований применение самоцитирования как индикатора роста (угасания?) научного направления некорректно.

Самоцитирование определяют иногда и для журналов [см., например, 4.102].

Анализ перекрестного цитирования науки и техники выявил тесное взаимодействие социумов ученых и инженеров. Более того, как показал, например, Т. Волдхарт [4.139], инженеры чаще цитируют научные работы, чем технические, — отношение ссылок в технических журналах на технику к числу ссылок на науку составило 0,51.

II. Сила связи двух элементов потока публикаций определяется числом общих для этих элементов ссылок, в них содержащихся. Эта связь носит скалярный характер.

Цитатная связь такого рода была предложена впервые М. Кесслером [4.95–4.97], который апробировал свой метод на материале 334 статей 112-го тома журнала «Physical Review». При этом он сравнил две классификационные системы этих статей, одна опиралась на метод ссылок, а вторая — на предметный индекс, содержащий 73 раздела. Результаты показали хорошее совпадение обеих классификационных систем [4.97].

Метод измерения цитатной связи Кесслера используют З. Б. Барина и соавт. [4.3], Ю. П. Адлер и соавт. [1.1], В. В. Налимов и З. М. Мульченко [1.46], В. В. Максимов [4.14] и др. Так, З. Б. Барина и соавт. [4.3] сравнивают степень взаимосвязанности публикаций в математической статистике и философии порознь. Сопоставляя эти данные с данными М. Кесслера по физике, В. В. Налимов и З. М. Мульченко заключают, что плотнее связи в физике, далее — в математической статистике и затем — в философии.

III. Сила связи двух элементов потока публикаций определяется числом работ, одновременно цитирующих эти элементы. Эта связь также носит скалярный характер.

Цитатная связь такого рода была предложена независимо И. В. Маршаковой [4.17–4.19] и Г. Смоллом [4.133–4.136; 4.87]. И. В. Маршакова предлагает называть таким образом определяемую цитатную связь *перспективной*, а цитатную связь Кесслера — *ретроспективной*. На рис. 3.3 показана сеть ссылок, выявленная посредством перспективной цитатной связи.

Сети цитатных связей математически описываются либо с помощью *графов*, либо средствами *матричного анализа*. К графам прибегают Дж. Марголис [4.109], Л. А. Хурсин и соавт. [0.57], Н. Г. Вишеченко и А. Е. Юрцунь [0.15], Ю. Гарфилд и А. Коукелл [4.82], Г. М. Добров и соавт. [0.26], А. А. Коренной [4.12–4.13] и др. Примеры графов цитат дают рис. 3.2, 3.3. Матричные методы используют Д. Прайс [4.124], Яги Ери [4.144], А. А. Коренной [4.12–4.13] и др.

Формализация цитатного анализа структуры науки средствами теории графов и матричного анализа позволяет учитывать не только прямые, но и косвенные связи между работами, учеными и т. д. Учет косвенных связей на языке теории матриц производят Яги Ери [4.144], А. А. Коренной [4.12] и др. При этом переход от прямых связей к косвенным осуществляется возведением в квадрат, куб и т. д. матрицы связей. На языке теории графов косвенные связи учитывают Дж. Марголис [4.109], Л. А. Хурсин и соавт. [0.57] и др.

Сами по себе сети цитирования, пусть даже математически описанные указанными средствами, дают материал в основном для качественного анализа. Математически строгое выделение по сетям цитирования структурных образований разной величины и «автономности» — научных школ, направлений и пр. — может быть произведено, например, с помощью факторного анализа. Именно так и поступает Б. Рассет [4.130], выделяющий среди 68 специалистов в области международных отношений 12 групп («школа»).

П. Зунде и В. Сломек, изучая связи цитирования между 13 научными дисциплинами, вместо факторного анализа используют аппарат теории марковских переходов [4.145]. Произвольность методологических оснований такого подхода отмечают Г. Борениус и С. Шварц [4.42].

Цитатная структура науки является индикатором предметной ее структуры. Анализ же предметной структуры науки представляет интерес, конечно, только в ее сопоставлении с организационной, социально-психологической и другими структурами. Поэтому анализ структуры науки методом ссылок и другими наукометрическими методами должен вестись совместно с социологическими, социально-психологическими и другими науковедческими методами.

5. ПЕРСПЕКТИВЫ

Наукометрический индикатор «число ссылок» — наиболее разработанный среди других наукометрических индикаторов, хотя и этот уровень разработанности нельзя считать удовлетворитель-

Таблица 4.5.1

Использование индикаторов, упомянутых в настоящей главе

Индикатор	Число работ	Индикатор	Число работ	Индикатор	Число работ
Число открытий	—	Число запросов	—	Предметные	8
Число журналов	21	Объем публикаций	—	Географические	10
Число организаций	2	Временной лаг	1	Классификационные	—
Число премий	—	Число ссылок*	35	Должностные	—
Число ученых	15	«Экзотические»	1	Языковые	5
Частота соавторства	2	Время	10	Прочие	8
Число публикаций	11	Возраст	61		

* Строка соответствует структурным исследованиям.

ным. Данные табл. 4.5.1, отражающие реальную, далеко не полную и в значительной мере случайную, свидетельствуют о необходимости дальнейшего исследования этого индикатора.

Число ссылок — самый «дробный» измеритель из рассмотренных выше, что определяет и его преимущества и его недостатки. Преимуществом является то, что с его помощью можно проводить исследования не только на тех массивах, на которых используются менее «дробные» измерители, но и на существенно меньших, на которых эти менее «дробные» измерители дают статистически недостоверные результаты. Возможность перехода к меньшим массивам свидетельствует о наличии невыявленных резервов анализа качественно новых задач (см., например, § 4).

Недостатком рассматриваемого измерителя науки является его большая по сравнению с менее «дробными» измерителями трудоемкость, что определяет нецелесообразность его использования на больших массивах, на которых применимы менее «дробные» измерители.

Большая «дробность» числа ссылок как измерителя науки определяет и необходимость разработки для него специального аппарата: мы видели, что метод ссылок основан на использовании указателей научных ссылок, прежде всего Индекса Гарфилда. Последний, как уже отмечалось, сканирует преимущественно естественнонаучные журналы. Кроме того, отдельные страны представлены в нем непропорционально их вкладом в мировую науку. В частности, непропорционально слабо отражаются Индексом издания социалистических стран. Следует поэтому поставить вопрос о создании странами-членами СЭВ Указателя научных

ссылок, дополняющего Индекс Гарфилда. Возможно, имеет смысл издание Всемирного указателя научных ссылок, причем должен быть учтен опыт издания Индекса Гарфилда. При этом следует договориться о единых правилах транскрипции фамилий ученых разных национальностей, об унификации встречающихся в указателях обозначений и сокращений, о перечислении всех соавторов работ непосредственно в Указателе ссылок и т. д.

В целях уменьшения зоны «библиографического молчания» следует внести в практику использование научных документов, содержащих оперативную информацию [4.10]. Возможно, следует узаконить ссылки на реферативные издания.

В современных применениях «цитат-индекса» публикации, ученые, научные коллективы и т. д. сравниваются, как правило, по абсолютному числу ссылок. Между тем заведомо ясно, например, что ссылки, сделанные на данного ученого высоко- и низкоранжированными журналами, неравнозначны. Поэтому числа ссылок следует умножать на определенные коэффициенты, различные для разных предметных областей, журналов, ученых и т. д. Это — так называемое *взвешивание ссылок*. Поскольку распределения ссылок для большого их числа описываются распределением Ципфа, постольку и при «взвешивании» ссылок следует опираться на это распределение.

Перспективы развития метода «цитат-индекс» в большой мере связаны с развитием структурных исследований науки на основе изучения сетей цитирования. Последние, в свою очередь, связаны с изучением временной динамики науки [см., например, 1.85]. Чтобы сделать картину связей более «контрастной», имеет смысл отсекаать ссылки, встречающиеся во всех рассматриваемых работах. Эти ссылки характеризуют весь массив в целом и поэтому затушевывают связи между отдельными его частями. «Ссылки на статьи по физическим методам исследований, математике, вычислительной технике и т. п., — пишет В. И. Горькова, — будут появляться достаточно часто у всех авторов всех отраслей науки, но опять-таки не будут фиксировать связи между объектами исследований. Тематические связи, имеющие место через общенаучные и общетехнические журналы, только сглаживают картину связей — являются «вуалирующей пеленой» и уменьшают контрастность картины связей» [1.11, с. 21]. Целесообразно, на наш взгляд, разработать формализованную процедуру такого отсека общенаучных, общедисциплинарных и других ссылок, аналогичную, например, излагаемой в гл. 6 процедуру отсека слов общенаучного характера при слоговом анализе публикаций.

В целях увеличения эффективности цитатного анализа иногда предлагается [см., например, 4.55], организовать обучение пра-

вильному цитированию. На наш взгляд, этот вопрос следует рассматривать в рамках более общего вопроса о введении учебного курса по науковедению. Но здесь возникают определенные трудности. Так, следует установить, какая доля ссылок на крупных специалистов оптимальна. Необходимость ответа на этот и аналогичные вопросы породила исследования, в которых наукометрический анализ ссылок сопровождается смысловым, качественным их анализом. Здесь могут быть названы Л. Бен-Ами [4.39], М. Моравчик и П. Муругеян [4.117], И. М. Ориент [1.53], В. А. Зябров и Т. В. Скляр [4.9], Д. Е. Чубин и С. Мойтра [4.53] и др. Но такие исследования проводятся пока в недостаточном объеме.

Перспективы развития «цитат-индекса» связаны в определенной мере и с применением *негауссовой статистики*, на которой, в частности, должны основываться аппараты ошибок и корреляционный.

Опорные исследования числа ссылок как наукометрического индикатора ведутся интенсивнее, чем исследования других индикаторов. Эти исследования показали, что число ссылок может служить индикатором многих латентных переменных науки. Для удобства изложения объединим эти латентные переменные в группы.

Латентные переменные научного направления и отдельных участков его проблематики

Индикатором *актуальности и перспективности* направления или предметной области могут служить темпы роста числа ссылок. Возможен и обратный заход — поиск не перспективных, а *тупиковых* направлений. «Нам кажется, — пишут, например, В. В. Налимов и З. М. Мульченко, — что правомернее будет иная постановка задачи: изучение тупиковых, или лучше — бесконфликтных направлений в науке. В науке легко заметить такие направления, которые, развиваясь, не создают конфликтных ситуаций — не приводят к выявлению внутренних противоречий и созданию новых концепций. За развитием бесконфликтных направлений легко следить по спектру цитирования. Если для какой-либо научной школы мы будем строить графики, откладывая по осям абсцисс удельную цитируемость (число цитирований на одну публикацию), то на таких графиках в бесконфликтных направлениях не будет резких всплесков» [1.47, с. 60]. Иными словами, тупиковые направления предлагается выявлять по кривым старения публикаций.

Стационарные распределения ссылок также могут быть использованы для выявления перспективных направлений. В. И. Горькова [1.11], проанализировав большой массив эмпирических данных,

приходит к выводу, что перспективные или *проблемные* направления характеризуются распределениями, имеющими более короткие «хвосты», чем традиционные направления (с. 19–20). Например, в работах проблемного направления ссылки даются на работы меньшего числа ученых, чем в работах непроблемного.

Однако остается неясным, какие критерии положены в основу различения проблемных и непроблемных направлений. Используемый В. И. Горьковой для описания длины «хвоста» распределения количественный параметр «проблемность» [1.11, с. 19], на наш взгляд, не совсем удобен в употреблении (возможно, более удобным окажется здесь показатель сходимости цитфовых распределений α (см. гл. 7, § 2).

Актуальность и перспективность научных направлений, проблемных областей и отдельных проблем могут оцениваться и на основе анализа *сетей цитирования*. В. Гриффит и Н. Маллинз [0.36, с. 131–151] проводили структурные исследования науки на материале так называемых сплоченных групп ученых, возникающих для и во время работы над кардинальными проблемами науки. «Наиболее существенным признаком этих групп, — пишут они, — было то, что их члены работали над своими научными проблемами в тесном и непрерывном контакте друг с другом» [0.36, с. 138]. Характеризуясь существенно более высоким уровнем организации и коммуникации, чем представители «нормальной» науки (авторы пользуются терминологией Т. Куна), эти сплоченные группы отличались, однако, друг от друга. «Как и следовало ожидать, степень организации во всех группах не была однородной. Более высокоорганизованные группы являлись поистине «революционными»: они сознательно противопоставляли себя нечленам группы, занимавшим обычно более привилегированное положение в университетах и соответствующих научных дисциплинах» [0.36, с. 138–139].

Продолжительность существования сплоченных групп 10–15 лет. «Даже наиболее продуктивные группы разрастаются в конце концов до такой степени, что это приводит к уменьшению, а затем и к ликвидации межличностных связей и личного влияния. Институционализация (например, создание факультета молекулярной биологии или отделения по проблемам оперантного обусловливания) уменьшает степень антагонизма группы по отношению к другим ученым и ослабляет групповую идентичность, а это обычно означает возвращение достигшей успеха группы к состоянию, нормальному для науки, то есть к структуре со слабой сетью связей. Таким образом, платой за успех группы... является ее смерть как особой социальной и интеллектуальной единицы» [0.36, с. 149–150].

Выводы Б. Гриффита и Н. Маллинза о сплоченных группах, согласующиеся с выводами В. И. Горьковой о проблемных научных направлениях, не могут, однако, считаться исчерпывающими. Угасающая научная школа также может, по-видимому, характеризоваться замыканием цитатной сети на самое себя (см. § 4). Чрезмерная координация усилий ученых в процессе работы над общей проблемой не может не сказаться отрицательно на их деятельности. Речь должна идти о некотором *оптимальном* уровне координированности научного труда, превышающем, если справедливы наблюдения Б. Гриффита и Н. Маллинза, в перспективных научных направлениях средний уровень координированности труда ученых по науке в целом. Поэтому, в целях выявления перспективных научных направлений имеет смысл проведение исследования наукометрическими средствами структуры науки с выявлением относительно замкнутых групп ученых, которые как раз и могут работать в перспективных направлениях. Построив, например, с помощью метода ссылок *сеть связей* работ, ученых, научных коллективов и т. д., мы получаем своеобразный «сетевой график», в котором относительная *актуальность* вопросов, связанных с данным узлом этой сети, определяется степенью связанности последнего со всей сетью. Сами по себе эти сети, даже описанные средствами теории графов или матричного анализа, дают материал в основном для качественных наблюдений. Корректно выделение по сетям цитирования структурных образований разного типа и разной степени автономности — проблемных групп, научных школ и пр. — может производиться с помощью специальных математических методов, например с помощью факторного анализа (точнее, его не-гауссового эквивалента).

Ряд авторов, как мы видели в § 4, анализируя сети цитирования, определяют относительную долю самоцитирования работ в данной предметной области. Такое *самоцитирование* рассматривается ими как индикатор *быстрорастущих направлений*, в которых относительное самоцитирование выше, чем для традиционных областей. Как говорилось в § 4, самоцитирование может быть связано с ростом научного направления отнюдь не так однозначно.

При решении вопросов о распределении средств между *фундаментальными* и *прикладными* исследованиями могут быть полезными *кривые старения публикаций*. При прочих равных условиях глубина цитирования, т. е. возраст цитируемых работ, в прикладных исследованиях, живущих сегодняшним днем, ниже, чем для фундаментальных, живущих днем завтрашним. Если в двух научных направлениях глубины цитирования разные, то при прочих равных условиях большая глубина цитирования одного из этих направлений может служить индикатором большей роли в ней

фундаментальных исследований. Это подтверждается данными, согласно которым гуманитарные науки имеют большие глубины цитирования, чем точные (см., например, индекс Прайса в § 2).

Латентные переменные научных журналов

Большое самоцитирование научных журналов считают иногда признаком их высокого статуса (см. § 4). Число цитирований данного журнала берется в качестве меры его *престижности* или *влиятельности*. Более валидной мерой этих латентных параметров считается число цитирований данного журнала, деленное на число опубликованных в нем статей (см., например, так называемый «Impact factor» [4.77]).

Латентные переменные ученого

Как показали опорные исследования (см. § 1), метод «цитат-индекс» дает индикаторы признания ученого, его *индивидуального вклада* в науку, *престижа*, *желания ученого публиковаться*.

«Цитат-индекс» может помочь и в измерении степени *координированности научного труда*. Мерой такой координированности может служить энтропия характеризующих деятельность ученых стационарных цитатных распределений, или же показатель сходимости этих стационарных распределений α , который Ю. А. Шрейдер [см., например, 1.70] трактует для научных коллективов как показатель их *творческойности* *.

Возможны и другие индикаторы координированности научного труда. Так, например, А. И. Яблонский [0.64] предлагает в качестве меры *степени организации* коллектива величину

$$R = 1 - S/S_{\max}, \quad (4.6.1)$$

где S — энтропия характеризующего данный коллектив наукометрического стационарного распределения; $S_{\max}$ — максимально возможная энтропия этого распределения. Следуя К. Шеннону, А. И. Яблонский называет величину (4.6.1) *избыточностью*.

Латентные переменные научной работы

Как показали опорные исследования (см. § 1), «цитат-индекс» дает индикаторы *качества* научной работы, ее *ценности*, *значимости*, *полезности*, *важности*.

Подчеркнем, что, как и для других наукометрических измерителей, связь числа ссылок с латентными переменными носит

* Строго говоря, Ю. А. Шрейдер пользуется не α , а $\gamma = 1/\alpha$.

статистический характер. Поэтому использование «цитат-индекса» может дать положительный эффект лишь на счетных массивах достаточной величины. В каждом же отдельном случае «цитат-индекс» может дать и неверный результат, тем более, что на сегодняшний день опорные исследования числа ссылок проведены в недостаточных объемах, не установлена еще полная картина корреляций числа ссылок ни с системой индикаторов, ни с системой латентных переменных науки. Эти исследования числа ссылок должны вестись фактически заново с привлечением аппарата негауссовой статистики, с широким охватом всей системы индикаторов и латентных переменных науки.

МЕТОДЫ «КОНТЕНТ-АНАЛИЗ» И ТЕЗАУРУСНЫЙ

В этой и следующей главах анализируются наукометрические индикаторы, при использовании которых научный текст рассматривается на уровне слов.

Такой анализ, как и всякий количественный анализ науки, опирается на вероятностные представления о деятельности (см. гл. 1, § 3). При этом вероятностная концепция текстовой деятельности ученого выступает как часть вероятностной концепции речевой и текстовой деятельности вообще, являющейся, в свою очередь, частью еще более общей вероятностной концепции человеческой деятельности.

Вероятностная концепция языка развивается Р. М. Фрумкиной [6.9], В. В. Налимовым [6.5] и другими авторами. Согласно этой концепции, «в речевых механизмах человека существует определенная иерархическая организация элементов текста в соответствии с частотой их встречаемости в речевой деятельности индивида» [6.10]. Каждое слово, «хранящееся в памяти индивида, имеет некий «индекс частоты», отражающий частоту встречаемости данного слова в прошлом речевом опыте индивида, а словарь в целом «организован» в соответствии с «индексом частоты» отдельных слов. Вероятностная организация словаря претерпевает определенные изменения, отражая процесс текущей речевой деятельности. Можно думать, что в множестве лиц-носителей некоторого языка можно выделить множество индивидов, у которых отдельные классы слов имеют сходные «индексы частоты», то есть выделить подмножества лиц с примерно одинаковой (в указанном смысле) «организацией» словаря и, соответственно, подмножества лиц с различной его организацией» [6.10, с. 9].

Мы отмечали, что вероятностность человеческой деятельности — эмпирический факт. Это утверждение полностью приложимо и к языковой деятельности человека как разновидности деятельности вообще. Иначе, как на языке вероятностных представ-

лений, этот вид деятельности удовлетворительно описать не удастся. Иначе, как вероятностным, не может быть количественное описание и текстовой деятельности ученого.

Мы выделим три наукометрических метода анализа текстов на уровне слов — «контент-анализ», тезаурусный и сленговый.

1. СОДЕРЖАНИЕ МЕТОДА «КОНТЕНТ-АНАЛИЗ»

Наукометрический метод «контент-анализ» происходит от социологического контент-анализа средств массовой коммуникации, развитого в основном американскими социологами Г. Лассуэлом [5.21–5.22], Ч. Осгудом [5.23], О. Холсти [5.19] и др. Из работ советских авторов назовем работы В. С. Коробейникова [5.6; 5.7], В. М. Соковнина [5.16], В. Е. Семенова [5.15].

Анализируя далее интересующую нас количественную ветвь контент-анализа, обозначенную работами Г. Лассуэла и его последователей, мы будем опираться на работу В. С. Коробейникова [5.7].

Процедура контент-анализа заключается в сведении рассматриваемого текста к ограниченному набору определенных элементов, которые затем подвергаются счету и анализу.

Цель лассуэловской процедуры контент-анализа «заключалась в том, чтобы зафиксировать в материалах средств массовой коммуникации частоту появления различных символов (фактически отдельных слов или, реже, предложений), определить соотношение во времени (сравнивая разные периоды) и пространстве (сравнивая различные источники информации). По сравнению с прежними процедурами метод Лассуэла представлял собой новый этап в развитии анализа содержания. Характерными особенностями этого этапа было выделение в содержании элементов (ими оказались символы) и фиксация частоты появления этих элементов» [5.7, с. 9].

Лассуэловская процедура распадается на три этапа: 1) получение из материалов массовой коммуникации совокупности определенных данных; 2) интерпретация; 3) получение описания поведения коммуникатора или респондента.

На первом этапе составляется список интересующих исследователя символов. «В этот список включаются как отдельные слова (единичные символы — unit symbols), так и высказывания (взаимосвязанные символы — sequence symbols), частоту которых нужно зафиксировать. Затем этот список уточняется с тем, чтобы одно и то же лицо или предмет, называемое в тексте по-разному, обозначалось бы одним символом» [5.7, с. 40]. «Затем определяется так называемая «фиксированная часть» текста. Ее смысл состоит

в том, что в пределах этой части текста частота символа фиксируется только один раз, хотя в действительности он может появиться в этой части и несколько раз... Например, одной из широко применяемых «фиксированных частей» является кусок текста, состоящий из трех предложений (иногда одного или двух предложений)» [5.7, с. 40]. «После определения символов и выбора «фиксированной части» начинается основная часть процедуры, заключающаяся в фиксации частоты символов. Эта часть работы проводится кодировщиками. Перед каждым исследованием проводится специальная тренировка кодировщиков на идентичность работы» [5.7, с. 40]. «Считается вполне нормальным, когда совпадение результатов кодировки (двумя кодировщиками — С. Х.) достигает 80 процентов» [5.7, с. 41].

На втором этапе производится интерпретация получаемых результатов на основе введения так называемых *категорий*, представляющих собой классы символов. При этом для Лассуэла на первый план выступает величина внимания, уделяемого символам той или другой категории, т. е. частота употребления символов разных категорий в сравнении друг с другом.

Третьего этапа процедуры мы здесь касаться не будем.

Подчеркнем, что метод Лассуэла основан, по существу, на вероятностной концепции: «Процедура Лассуэла имела своей целью в основном выявление эффективности воздействия пропагандистских материалов на широкие массы. В этой ситуации частота заняла господствующее положение, так как априорно предполагалось, что, чем чаще данный символ воздействует на аудиторию, тем более он эффективен» [5.7, с. 68—69].

После второй мировой войны, начиная с работы Ч. Осгуда и Т. Собекса 1951 г. американские социологи развивают частотный контент-анализ Лассуэла в так называемый «анализ связанности символов», изучающий, как часто данный символ встречается в связи с другими символами. На основании данных о связанности символов затем рассчитывают связанность *категорий*.

Сам Лассуэл называл свой метод количественным, критикуя качественные методы, а между тем формализация контент-анализа в лассуэловской процедуре является чисто внешней. Уже на первом этапе список подвергаемых затем частотному анализу символов составляется совершенно произвольно, в зависимости от стоящих перед исследователем задач. Произвольно выбирается и «фиксированная часть» текста. Но наибольший произвол, перечеркивающий по сути дела всю производимую Лассуэлом формализацию контент-анализа, содержится в той части его процедуры, в которой анализируемые символы распределяются по категориям: это распределение осуществляется на основе умозритель-

ных соображений, что делает возможным получение любых конечных результатов.

Таким образом, лассуэловский контент-анализ имеет как позитивные, так и негативные стороны. «Факты показывают, что за долгие годы работы над развитием процедуры анализа содержания американские социологи изучали различные формы использования этого метода, хорошо отработали технику его применения. Особенно важны результаты, касающиеся необходимости фиксации частоты появления выделенных элементов, частоты связанности одних элементов с другими, отношения к этим элементам со стороны коммуникаторов (знак высказывания)» [5.7, с. 13]. Негативным же моментом лассуэловской процедуры является отмеченная выше ее субъективность.

Если количественный контент-анализ развивается в США вот уже полвека, то только благодаря сочетанию заключенной в нем формализованности и неформализованности. Формализация метода, основанная на фиксации частот символов, позволяла политическим деятелям США, осуществлявшим финансирование исследований в этой области, получать объективные данные об эффективности органов пропаганды. Неформализованность же метода позволяла этим же деятелям получать в свое распоряжение «научно обоснованные» и желательные для них выводы, запускаемые в ту же самую пропагандистскую машину, но уже на «научной» основе.

В последние годы область применения контент-анализа начинает расширяться, распространяясь, например, на исторические исследования [см. 0.40]. Существенную роль в этом сыграла переориентация контент-анализа с изучения текстов самих по себе на изучение по ним производящих эти тексты систем (см., например, принцип восхождения от текста к источнику К. Криппендорфа [5.20]). В применении к анализу науки, например, это означает изучение научной деятельности по научным текстам.

В связи с этим проникновение контент-анализа в количественный анализ науки представляется естественным процессом. В. В. Налимов и З. М. Мульченко так говорят о возможностях количественного анализа текста на уровне слов: «Особый интерес представляет статистическое изучение языка научных публикаций. По мере своего обособления и углубления каждая область знаний вырабатывает свой особый язык — своеобразный научный «сленг», делающий ее малодоступной для специалистов в других областях. Этот процесс представляется вполне естественным, поскольку в каждой, даже узкой области знаний возникают понятия, имеющие свой особый и очень глубокий смысл — с их помощью кодируются целые научные концепции. Такая си-

стема кодирования научных концепций облегчает обмен информацией в пределах отдельных научных направлений, но затрудняет общение со смежными дисциплинами. Толковые словари здесь не приносят пользы — они не могут заменить собой руководства, разъясняющие научные концепции. Нам представляется интересным проследить при помощи статистического анализа за историей развития специфических языков в отдельных областях знаний и оценить их роль в развитии системы научной информации. Таким способом можно, по-видимому, решать и отдельные частные задачи. Одна из подобных задач — изучение проникновения каких-либо методов исследования и мышления в смежные области. Например, можно количественно определить степень проникновения методов математической статистики в различные области знаний, оценивая частоту появления специфических терминов данной дисциплины в публикациях, посвященных другим разделам науки. Такой анализ позволяет оценить не только широту, но и глубину проникновения новых идей и методов. Последняя будет задаваться частотой появления слов, которыми кодируются сложные концепции. Этим способом, видимо, удастся следить и за развитием отдельных идей или научных школ в пределах одной области знаний... Работа в этом направлении только начинается» [1.46, с. 15–16].

2. ПРИМЕНЕНИЕ МЕТОДА «КОНТЕНТ-АНАЛИЗ»

Контент-анализ в наукометрических исследованиях применяют И. М. Ориент [4.23], Ю. П. Адлер и соавт. [1.1], В. В. Налимов и З. М. Мульченко [1.46], В. В. Максимов [5.8–5.9], Дж. Роекелайн [2.57], В. А. Владыкин и Ю. П. Воронов [5.3], В. М. Потапов и И. Н. Лукина [5.13], Г. Б. Преображенская и соавт. [5.14] и др.

Сравнительным анализом временной динамики разных предметных областей занимаются В. М. Потапов и И. Н. Лукина [5.13]. Они изучают динамику объема материала, уделяемого данному понятию в предметных указателях РЖ «Химия» и «Chemical Abstract».

Статистические распределения элементов текста дают Дж. Роекелайн [2.57], Г. В. Преображенская и соавт. [5.14] и др. На рис. 2.44 показано распределение имен отдельных ученых по частоте встречаемости в 8-ми учебниках психологии. Как показывает рис. 2.44 данные могут быть аппроксимированы цифровым распределением.

Структурные исследования науки с помощью контент-анализа

ведут И. М. Ориент [4.23], Г. В. Преображенская и соавт. [5.14] и др. Эти авторы изучали, например, проникновение математических статистических методов в аналитическую химию.

3. СОДЕРЖАНИЕ ТЕЗАУРУСНОГО МЕТОДА

В основе тезаурусного наукометрического метода лежит тезаурусный метод, широко применяемый в теории и практике информационного поиска. Информационные службы нуждаются не столько в выявлении объективных воспроизводимых характеристик изучаемых объектов, сколько в отборе среди публикаций текстов, релевантных заказу. Поэтому и упор в тезаурусном методе делается на содержательный анализ.

Под терминном «тезаурус» понимается словарь наименований понятий и их классификационных связей. «...Основным назначением тезауруса является составление с его помощью поисковых образов документов и запросов при их вводе в информационно-поисковые системы» [5.17, с. 4]. «Тезаурусы... создаются на основе лексико-статистического анализа текстовых сообщений и представляют собой сборники упорядоченной и нормализованной терминологии» [5.17, с. 3]. При отборе терминов основными критериями являются: относительная частота появления термина в различных документах-источниках; относительная частота использования термина в действующей системе; предполагаемая значимость термина для поиска информации; точность и однозначность термина [5.17, с. X].

«Если в обычном словаре по слову находят его значение, обычно определяемое рядом примеров его применения, то в тезаурусе по значению, которое записывается некоторым способом, находят слово или несколько слов, выражающих искомое значение» [5.12, с. 215]. Тезаурус разделяется на ряд «тем, в каждую из которых входят слова, имеющие в некотором смысле одинаковое значение. Практически название темы представляет как бы контекст для входящих в нее слов» [5.12, с. 215]. В тезаурусе каждое слово определяется последовательностью названий тем, включающих данное слово. «Различать слова, определенные совпадающими цепочками», невозможно [5.12, с. 216].

Мы видели выше, что частота появления слова в поисковых документах используется как один из критериев при составлении тезаурусов. Но вероятностный подход может использоваться при построении тезаурусов и более сильным образом, как это показывает В. А. Москович [6.4]. «Составление смысловых словарей-тезаурусов, — пишет он, — предприятие трудоемкое и дорогостоящее. Построение смыслового словаря требует глубокого анализа со-

держания текстов данной области знания, четкой классификации понятий, характерных для этой области, отбора наиболее существенных из этих понятий для включения в тезаурус» [6.4, с. 1]. Обычно тезаурусы составляются «группой специалистов данной области науки или техники. Поэтому этот метод построения тезаурусов называют методом экспертной оценки. Недостатками метода являются возможная субъективность решений при отборе слов и их группировке и большой объем трудозатрат» [6.4, с. 2].

«При построении тезаурусов,— продолжает В. А. Москович,— полезными могут оказаться следующие сведения, извлекаемые путем автоматического анализа текстов: информация о частоте слов в тексте и их дистрибуции (совместной встречаемости с другими словами в тексте).

Наличие слова в тексте и знание его частоты указывает на целесообразность его включения в тезаурус. Выявление списка слов, с которыми часто встречается вместе данное слово, служит ценным указанием о возможных смысловых связях данного слова с другими словами» [6.4, с. 2—3].

Предлагая автоматизировать процесс составления тезаурусов, В. А. Москович исходит фактически из вероятностных представлений, которые формулируются им в применении к речевой деятельности следующим образом: «По соотношению частотностей слов, принадлежащих к одному и тому же семантическому полю, можно судить о строении языка» [6.3, с. 15]. «Между частотностью слова и его активностью в языке существует тесная взаимозависимость... Стратификация семантического поля по признаку частотности совпадает с его стратификацией по признаку активности» [6.3, с. 28]. Под активностью при этом В. А. Москович понимает многочисленность и силу связей данного слова со всей системой слов.

Используя одни только количественные частотные методы, построить тезаурус невозможно. «Разумеется,— пишет В. А. Москович,— машина не может заменить человека при интерпретации характера выявленных смысловых связей, однако материал для интерпретации она дает. Сочетая статистический метод с методом экспертной оценки, можно рационализировать процедуру построения тезауруса» [6.4, с. 3].

4. ПРИМЕНЕНИЕ ТЕЗАУРУСНОГО МЕТОДА

В наукометрических целях тезаурусный метод используют Н. Равер [5.24], Г. М. Добров и В. В. Борисов [5.4], А. Н. Алексеев [3.1, 5.1], Л. А. Богаев и соавт. [5.2], И. В. Маршакова [5.10], В. М. Потапов и И. Н. Лукина [5.13], И. Л. Турыкина

[5.18], И. С. Дубовиков [5.5], В. М. Потапов и Э. К. Кочетова [1.57], М. П. Ширшова [1.68] и др.

Динамику роста отдельных предметных областей изучают А. Н. Алексеев [3.1, 5.1], И. В. Маршакова [5.10], В. М. Потапов и И. Н. Лукина [5.13], И. Л. Турыкина [5.18], В. М. Потапов и Э. К. Кочетова [1.57], М. П. Ширшова [1.68] и др. При этом количественной мерой скорости роста данной предметной области служит скорость роста частоты употребления характеризующих эту область терминов. Пример данных такого рода показан на рис. 1.9 а, б.

Распределения частот применения отдельных терминов по предметным областям изучают Н. Равер [5.24], В. М. Потапов и Э. К. Кочетова [1.57] и др. Н. Равер, например, находит, что частота использования дескрипторов в предметном индексировании подчиняется распределению Ципфа.

Структурные исследования посредством тезаурусного метода проводят Г. М. Добров и В. В. Борисов [5.4], изучая вопрос, из каких смежных областей знания черпают информацию ученые-электроники.

5. ПЕРСПЕКТИВЫ

Методы «контент-анализ» и тезаурусный как наукометрические нашли себе применение меньшее, чем какой-либо другой наукометрический метод. И дело здесь не в непригодности индикатора «число слов», а в особенностях его использования: и в «контент-анализе», и в тезаурусном методе процедуры измерения не обеспечивают воспроизводимости результатов счета, достаточной для наукометрических исследований, направленных на выявление количественных объективных закономерностей науки.

Будущее наукометрического индикатора «число слов» связано с дальнейшей формализацией процедуры измерения с целью увеличения воспроизводимости результатов счета. Один из возможных путей формализации реализован в сленговом методе, которому посвящена следующая глава.

Глава шестая

СЛЕНГОВЫЙ МЕТОД

1. СОДЕРЖАНИЕ МЕТОДА

Здесь, как и в предыдущей главе, обсуждается измерение науки путем анализа научных текстов на уровне слов. Но в качестве индикатора берется не число «символов», как это делается в контент-анализе, и не число «терминов» (дескрипторов, ключевых слов), как это делается в тезаурусном методе, но просто число слов, точнее, так называемых *словарных слов*, которые следует отличать от *словоформ* или *словоупотреблений*. Под последними понимаются слова как они встречаются в тексте, т. е. последовательности букв между двумя пробелами, тогда как словарное слово объединяет по определенным формализуемым правилам соответствующую ему группу словоформ.

Сленговый метод, как и другие наукометрические методы, по-явится на вероятностных представлениях и является статистическим. Являясь результатом развития «контент-анализа» и тезаурусного методов, сленговый метод опирается, кроме того, на традиции статистической лингвистики, изучающей частотные распределения слов и создавшей частотные словари. Первый частотный словарь появился в 1898 г. — это был словарь немецкого языка Кеддинга. С тех пор создано более 300 различных частотных словарей [6.8]. Среди русских частотных словарей отметим «Частотный словарь современного русского литературного языка» Э. А. Штейнфельда [6.16], «Частотный словарь общенаучной лексики» [6.14], «Частотный словарь русского языка» [6.15]. Создаются частотные словари по отдельным научным специальностям — электронике, техническим наукам, биологии и т. д.

Практические приложения статистической лингвистики чрезвычайно интересны [см., например, 6.6], но, к сожалению, далеки от науковедения. Р. М. Фрумкина, например, следующим образом формулирует целевую направленность частотных словарей «...1) Рационализация изучения иностранного или родного языка... 2) Усовершенствование различных кодовых систем... 3) Изучение словаря отдельных литературных памятников или авторов...»

[6.8, с. 5]. Однако и здесь мы находим идеи, интересные с точки зрения их возможного применения при решении задач количественного анализа науки.

А. Н. Колгушин [6.2] предполагает использовать частотные словари для анализа взаимосвязи между различными структурными единицами войскового управления. Он пишет в связи с этим: «Методы ранговой корреляции (зависимости) применяются в статистической лингвистике для сравнения близости (отдаленности) частотных словарей, разработанных как на родственной, так и на разнородной лексической основе, для оценки достоверности словарей, зависимости достоверности словарей от объема выборки и т. д.»

«Коэффициент близости частотных словарей, — продолжает А. Н. Колгушин, — можно использовать и как меру их достоверности, а также для количественного измерения связи между различными уровнями иерархической структуры, в том числе и военной. Прочность связей между органами управления войсками и внутри них неодинакова и тем сильнее, чем больше они выполняют совместных функций, чем интенсивнее взаимный обмен информацией.

Однородные функции описываются однородными лексическими единицами, дающими близкие по составу словари. Близость (отдаленность) словарей указывает на близость (отдаленность) функций, выполняемых отдельными ячейками управления...

Величину близости словарей словарной лексики можно использовать также для определения характера взаимосвязей между ячейками управления и глубины взаимодействия между ними...» [6.2, с. 49–50].

По сути дела А. Н. Колгушин предлагает проводить структурный анализ военной организации через структурный анализ массива военных документов посредством частотного анализа последних на уровне слов. Заменяя массив военных документов массивом научных публикаций, мы, следуя А. Н. Колгушину, получим возможность структурного анализа научной деятельности посредством частотного анализа публикаций на уровне слов.

Идеи, высказанные Колгушиным, предельно просты, но при количественном анализе науки до сих пор не получили широкого распространения. Обусловлено это особенностями истории статистической лингвистики, до сих пор имеющей дело почти исключительно с текстами, написанными так называемым естественным языком. В этой науке распределение слов по частоте их встречаемости в текстах описывается распределением Ципфа. Об этом распределении уже говорилось и еще будет говориться в гл. 7, § 2. Здесь же существенно то, что, следуя Г. Ципфу [6.23; 1.160],

$$f(x) = (\alpha/x_0) (x_0/x)^{1+\alpha}, \quad x \geq x_0, \alpha > 0$$

в статистической лингвистике связывают исключительно со значением α , равным 1. И, действительно, для естественных языков $\alpha \approx 1$ (см., например, рис. 2.48, 2.49).

Таким образом, с точки зрения статистической лингвистики, изучающей тексты, написанные естественным языком, все такие тексты *тождественны*, имеют одинаковые частотные организации. Эта их тождественность усиливается тем обстоятельством, что универсальность распределения Ципфа с $\alpha = 1$ для всех текстов означает и тождественность *словарных составов*, так как, например, лишь такая тождественность *словарей* двух текстов может обеспечить тождественность *формы* частотного распределения суммы этих текстов с формами частотных распределений этих текстов порознь.

Не следует понимать слишком буквально утверждение о тождественности разных текстов с точки зрения статистической лингвистики. Выше отмечалось, что одной из целей последней является изучение словаря отдельных литературных памятников или авторов, которое предполагает, конечно, наличие индивидуальных особенностей разных текстов. Говоря о тождественности разных текстов в статистической лингвистике, мы имеем в виду сложившийся в этой науке под действием представлений об универсальном характере распределения Ципфа с $\alpha = 1$ *стереотип лингвостатистического мышления*, предполагающий акцент не на отличии одного текста от другого, а на общем в различных текстах. Этот стереотип и тормозит реализацию идей о возможности исследования *средствами частотного анализа текстов на уровне слов структуры, порождающих эти тексты систем*, в частности, военных организаций, о которых пишет А. Н. Колгушин, или объектов науки, интересующих нас в контексте настоящего исследования в первую очередь. Такие структурные исследования предполагают упор на различии отдельных объектов анализа, но не на их сходстве.

Говорить об универсальности стандартного распределения Ципфа с $\alpha = 1$ в применении к науке было бы просто неверно: словари, а следовательно, и формы частотных распределений слов научных текстов отличаются от текста к тексту. И это различие, вообще говоря, тем больше, чем дальше отстоят друг от друга предметные области сравниваемых текстов и чем больше различаются эти тексты по авторским концепциям: в науке каждой специально вырабатывается определенный «сленг», зачастую кодирующий в отдельных терминах целые научные концепции [см., например, 1.46, с. 15–16]. Поэтому частотный анализ научных тек-

тов на уровне слов имеет смысл существенно больший, чем такой анализ текстов, написанных естественным языком. И, следовательно, *большой смысл имеют исследования структуры объектов науки через частотный анализ научных публикаций на уровне слов.*

Эффективность этих исследований увеличивается при использовании процедуры, которая может быть названа *сленгированием* и направлена на индивидуализацию научных текстов. Исходя из представлений, аналогичных тем, на которые опирается В. И. Горькова, предлагая «отсекать» библиографические ссылки, сравнительно часто встречающиеся в работах всего исследуемого массива публикаций (см. гл. 4, § 5), т. е. представлений о том, что одинаково часто встречающиеся в публикациях слова общенаучного характера маскируют картину связей между частями потока научных публикаций, можно предложить *отсекать* такие слова.

Для формализации процедуры сленгирования достаточно из словарного состава данного научного текста выбросить все слова, занимающие в частотном словаре общенаучного или, скажем, литературного языка первые 500, 1000 или другое число мест, в зависимости от необходимой *степени сленгирования*. Оставшиеся слова будем называть *сленговыми*.

В научных текстах широко варьируют, как говорилось выше, как словарные составы, так и величины α , характеризующие форму частотных распределений слов в этих текстах. Поэтому здесь нет того жесткого соответствия между формой распределения слов и словарем, какое свойственно текстам, написанным естественным языком. По этой причине то обстоятельство, что сленгирование лишь несущественно изменяет *форму* распределений слов по их встречаемости в научных текстах (а для распределения Ципфа и вовсе ее не меняющую), не означает, что оно не изменяет *словари* этих текстов. Напротив, индивидуализация научных текстов осуществляется при сленгировании исключительно за счет индивидуализации словарей сленгируемых текстов.

Статистическая лингвистика достигла ныне уровня практической полной формализованности процедуры составления частотных словарей. Операция сленгирования, как мы видели, также полностью формализуема, поэтому сленговый метод анализа научных публикаций представляет собой количественный формализованный метод и должен давать *воспроизводимые* результаты.

Для количественной характеристики распределений сленговых слов по частоте встречаемости, как и для характеристики других наукометрических распределений, не применим аппарат моментной (см. гл. 1 § 3, гл. 7). Поэтому здесь следует использовать энтропию распределения или показатель распределения Ципфа α . Для распределений сленговых слов научных текстов часто $\alpha < 1$ (см., на-

пример, рис. 2.45, 2.46), и поэтому применимы как частотная, так и ранговая энтропия (см. гл. 7 § 2). Мы использовали ранговую:

$$S = - \sum_{r=1}^N \frac{x(r)}{G} \log_2 \frac{x(r)}{G}. \quad (6.1.1)$$

Здесь S — ранговая энтропия; r — ранг сленгового слова; $x(r)$ — частота появления в тексте r -го слова; N — число разных сленговых слов; G — общее число сленговых слов.

Абсолютная ошибка ΔS ранговой энтропии для $\alpha < 1$ может вычисляться по формуле

$$\Delta S = \sigma / \sqrt{N}, \quad (6.1.2)$$

где σ — среднеквадратическое отклонение величины $-\log_2(x(r)/G)$ от своего среднего значения S .

При измерении корреляции распределений сленговых слов аппарат гауссовой статистики также неприменим. Поэтому целесообразно пользоваться энтропийным коэффициентом взаимосвязи, который для $\alpha < 1$ может браться, например, в следующем виде:

$$K = 1 - \frac{S_{\Sigma} - \frac{1}{2}(S_1 + S_2)}{\log_2 2}, \quad (6.1.3)$$

где S_1 — ранговая энтропия распределения слов первого текста; S_2 — второго; S_{Σ} — ранговая энтропия распределения суммарного текста. Выражение (6.1.3) является приближением более строгого выражения

$$K = \frac{S_{\Sigma}^{\max} - S_{\Sigma}}{S_{\Sigma}^{\max} - S_{\Sigma}^{\min}}, \quad (6.1.4)$$

где $S_{\Sigma}^{\max}$ и $S_{\Sigma}^{\min}$ — соответственно максимально и минимально возможные ранговые энтропии суммарного текста при данных S_1 и S_2 .

Энтропию как измерение взаимосвязи привлекают У. Мак-Гилл [0.99], Н. В. Хвостова [0.56] и др.

2. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ *

Сленговый метод предложен автором [6.11–6.13] и развивается им совместно с Ю. М. Низовцевым.

В работах [6.11; 6.12] проводится сленговый анализ заголовков публикаций М. Планка и А. Эйнштейна: сленгирование здесь

* Параграф написан совместно с Ю. М. Низовцевым.

Таблица 6.2.1

Значения энтропии частотных распределений сленговых слов в заголовках публикаций Планка и Эйнштейна по десятилетиям их научной деятельности [6.12]

Ученые	Десятилетия					
	I	II	III	IV	V	VI
Планк	1,63± 0,03	1,71± 0,035	1,85± 0,02	1,87± 0,03	1,83± 0,03	1,71± 0,03
Эйнштейн	1,86± 0,01	2,03± 0,03	1,96± 0,005	1,92± 0,03	1,66± 0,04	

проводится еще неформализованно, к разряду сленговых («смысловых») относятся все встречающиеся в заголовках слова, за исключением артиклей, предлогов, союзов и слов типа «введение в», «к вопросу о» и пр. Массив сленговых слов составил: для Планка — 777 слов, для Эйнштейна — 1207 слов.

Были построены частотные распределения сленговых слов по отдельным десятилетиям научной деятельности Планка и Эйнштейна и рассчитаны соответствующие им значения энтропии (табл. 6.2.1).

В табл. 6.2.2 и 6.2.3 также представлены значения энтропии, но для распределений, в которых отброшены все сленговые слова, за исключением трех и пяти слов-лидеров, наиболее часто встречающихся в данном десятилетии у данного ученого.

Как явствует из табл. 6.2.1, практически на протяжении всей научной деятельности Эйнштейна характеризуют большие значения энтропии, чем Планка. Мы связываем это с большей *разносторонностью* научных интересов Эйнштейна.

Однако из табл. 6.2.2, 6.2.3 видно, что значения энтропии распределения трех и пяти слов-лидеров в заголовках публикаций Эйнштейна меньше, чем в заголовках публикаций Планка, что связывается нами с большей *концентрированностью* научных интересов Эйнштейна.

Характерное для Эйнштейна сочетание большой *разносторонности* научных интересов с большой *концентрированностью* его на нескольких основных для него понятиях в работах [6.11–6.12] связывается с тем обстоятельством, что Эйнштейн, в отличие от Планка был «генератором теорий».

Наукометрический анализ структуры научной деятельности предполагает измерение взаимосвязи отдельных пар элементов

Таблица 6.2.2

Значения энтропии частотных распределений трех сленговых слов-лидеров в заголовках публикаций Планка и Эйнштейна по десятилетиям их научной деятельности [6.12]

Ученые	Десятилетия					
	I	II	III	IV	V	VI
Планк	0,48± 0,00	0,475± 0,01	0,46± 0,02	0,47± 0,02	0,47± 0,015	0,47± 0,02
Эйнштейн	0,46± 0,03	0,43± 0,015	0,41± 0,02	0,46± 0,025	0,48± 0,025	

Таблица 6.2.3

Значения энтропии частотных распределений пяти сленговых слов-лидеров в заголовках публикаций Планка и Эйнштейна по десятилетиям их научной деятельности [6.12]

Ученые	Десятилетия					
	I	II	III	IV	V	VI
Планк	0,70± 0,00	0,70± 0,01	0,68 0,025	0,68± 0,02	0,68± 0,03	0,675± 0,03
Эйнштейн	0,67± 0,015	0,66± 0,02	0,60± 0,03	0,69± 0,02	0,68± 0,03	

этой структуры и последующий, скажем, факторный, анализ матриц взаимосвязи. Поэтому, если будет доказано, что сленговый метод позволяет адекватно измерять взаимосвязь двух структурных единиц потока публикаций, то будет доказана и его пригодность для структурного анализа научной деятельности вообще. В работе [6.13] полигоном для апробации сленгового метода выбрана проблема оценки вкладов соавторов в совместную публикацию. Для сравнительной оценки вкладов двух соавторов в совокупную публикацию используются коэффициенты взаимосвязи работ каждого из соавторов, написанных ими до совместной работы, с общей работой.

По рекомендации Вл. П. Визгина мы остановились на случае соавторства А. Эйнштейна и И. Лауба при написании статьи «Об основных электродинамических уравнениях движущегося тела» ([0.61], с. 115–122), опубликованной в 1908 г.

Как историк физики Вл. П. Визгин является специалистом в области теории относительности и, таким образом, может рассматриваться как эксперт в интересующем нас случае. Мнение Вл. П. Визгина: Эйнштейн произвел основной вклад в интересующую нас статью. Но не будем ограничиваться ссылкой на мнение эксперта и кратко изложим предметную сторону вопроса.

Статья, о которой мы здесь говорим, посвящена выводу уравнений преобразования Максвелла в веществе.

До написания общей с Эйнштейном работы Лауб занимался вторичными катодными лучами и оптикой движущихся сред. По каждому вопросу он имел две индивидуальные статьи. В статьях по оптике движущихся сред [0.92], не обращаясь к общим уравнениям электродинамики и оперируя лишь уравнениями преобразования координат и времени Лоренца-Эйнштейна, Лауб дает теорию некоторых частных задач, например эффекта Доплера. При этом он явно исходит из знаменитой работы Эйнштейна по теории относительности (1905 г.).

Эйнштейн до совместной работы с Лаубом занимался многими вопросами — статистической механикой, теорией броуновского движения, теорией фотоэффекта и, конечно, теорией относительности. Целый ряд статей по теории относительности написан им в этот период — в собрании научных трудов Эйнштейна насчитывается восемь таких статей. В этих работах Эйнштейн излагает общие вопросы своей теории и развивает ее в приложении к различным областям физики — электродинамике, кинематике, механике, оптике, термодинамике и др. В его статье «К электродинамике движущихся сред» (1905) есть § 6 «Преобразование уравнений Максвелла-Герца для пустого пространства. О природе электродвижущих сил, возникающих при движении в магнитном поле» и § 9 «Преобразование уравнений Максвелла-Герца с учетом конвекционных токов» [0.61, с. 22–25 и 30–31]. Здесь дается теория преобразований уравнений Максвелла в общем виде. С этих общих позиций Эйнштейн рассматривает и некоторые частные задачи оптики, дает, например, теорию абберации и эффекта Доплера. Правда, в статье «О принципе относительности и его следствиях» (1907) Эйнштейн рассматривает некоторые оптические задачи [0.61, с. 77–79], но используя уравнения Максвелла непосредственно на основе формул преобразования координат и времени (так делает и Лауб). Но это для Эйнштейна отнюдь не характерно.

Для нас здесь существенно, что Эйнштейн — универсал в физике. Он последовательно рассматривает следствия теории относительности в различных областях физики, идя, как правило, от общих положений к частным, не пренебрегая, однако, и частными способами рассмотрения. Лауб же — ученый меньшего масштаба, занимается частными вопросами и идет при этом, как правило, от частного же. Другими словами, Эйнштейн своими иссле-

дованиями «перекрывает» и тематику, и способ мышления Лауба. То обстоятельство, что Эйнштейн до Лауба не распространил релятивистскую оптику на вещество, объясняется, конечно же, не тем, что он не смог этого сделать без соавтора, — он, безусловно, сделал бы это, если бы располагал временем рассматривать все частные следствия своей теории.

При непосредственном анализе совместной работы Эйнштейна и Лауба [0.61, с. 115—122] видно, что она написана в эйнштейновской манере: она построена по принципу от общего к частному. В первом параграфе, который называется «Вывод основных уравнений электродинамики для движущегося тела», дается в общем виде вывод уравнений преобразования электродинамических уравнений в веществе с учетом релятивистских поправок, во втором («Об электромагнитных свойствах движущегося диэлектрика. Опыт Вильсона») — пример использования этих общих уравнений к специальному случаю движущегося диэлектрика.

Посмотрим теперь, что может дать для сравнительной оценки вкладов Эйнштейна и Лауба в рассматриваемую работу выход из предметной сферы в плоскость личных отношений этих ученых. При этом будем опираться на любезно предоставленную Вл. П. Визгиным диссертацию Л. Р. Пайнсона [0.103].

Закончив в 1904 г. обучение в Геттингенском университете, Лауб приступает к работе над докторской диссертацией у В. Вина в Вирцбурге. В то время, когда Вин, редактировавший «Анналы физики», отсутствовал, в редакцию поступила знаменитая впоследствии статья Эйнштейна по теории относительности (1905). Диссертант Вина Лауб организовал коллоквиум по статье Эйнштейна и принял ее в печать. Заметим, что утверждение статьи Эйнштейна к печати не было просто формальным актом, поскольку эта статья была отнюдь не тривиальна по содержанию и не совсем обычна по форме. Во всяком случае, когда Эйнштейн представил свою статью как основание для занятия им вакантной должности доцента кафедры физики в Берне, ему отказали.

В 1905 г. Лауб под руководством Вина экспериментально исследовал эмиссию вторичных катодных лучей, результатом чего и явились его две статьи на эту тему, о которых упоминалось выше. Статья Эйнштейна натолкнула Лауба на другую тему, которой он и занимался параллельно с работой над катодными лучами. (Имеется в виду оптика движущихся сред, по которой Лауб опубликовал большую статью в двух частях, упоминавшуюся выше.) Докторскую диссертацию Лауб защитил в 1906 г., продолжал работать после этого ассистентом у Вина.

В феврале 1908 г. Лауб послал из Вирцбурга Эйнштейну почтовую открытку с просьбой разрешить поработать с ним в Берне три месяца по специальной теории относительности. Если учесть, что Лауб к тому вре-

мени уже два года был доктором философии, а Эйнштейн по-прежнему работал в Бернском бюро патентов, эта просьба Лауба не покажется стандартной. Об ученическом характере этой просьбы Лауба говорит уже то обстоятельство, что в этой же открытке он спрашивает Эйнштейна, вправе ли он, Лауб, думать, что специальная теория относительности применима не только к электродинамике (напомним, что первая статья Эйнштейна по теории относительности называлась «К электродинамике движущихся сред»), но и ко всей физике. Для Лауба это был отнюдь не праздный вопрос, поскольку он занимался в то время, как говорилось, применением теории относительности к оптике движущихся сред. Получив согласие Эйнштейна, Лауб приехал в Берн, где они в сотрудничестве написали две статьи — соответственно в апреле и мае 1908 г., первая из которых и рассматривается нами.

Отметим в связи с изложенным следующие обстоятельства. Во-первых, интерес Лауба к оптике движущихся сред зародился, несомненно, под влиянием статьи Эйнштейна по теории относительности (1905), одним из первых читателей которой был Лауб. Во-вторых, лучшие работы Лауба, т. е. работы, оставившие заметный след в физике, — это работы, написанные под руководством Вина или в соавторстве с Эйнштейном (это соображение подсказано нам Вл. П. Визгиным). Последующие его работы не играли сколько-нибудь заметной роли в физике. Это говорит, как нам представляется, о недостаточной самостоятельности Лауба как физика, — судя по всему, для успешной работы ему нужно было руководство более сильного ученого, который направлял бы его исследования. В первом случае таким руководителем был Вин, во втором — Эйнштейн. Во втором случае ученичество Лауба не было закреплено формально, но это не меняет сути дела и Лауб может рассматриваться (и иногда рассматривается) как первый (по времени) ученик Эйнштейна.

Таким образом, историко-научно вопрос об относительных вкладах Эйнштейна и Лауба в рассматриваемую совместную работу решается вполне однозначно: *вклад Эйнштейна несомненно превышает вклад Лауба*. Если бы проведенный с помощью сленгового метода количественный анализ не показал бы того же, это говорило бы только о неадекватности этого метода.

Чтобы вычисляемые коэффициенты взаимосвязи действительно служили оценками индивидуальных вкладов соавторов, необходимо принимать во внимание не все предшествующие работы соавторов (в этом случае вклад каждого соавтора будет неопределенным образом занижен), но достаточно большие в статистическом смысле их части, обеспечивающие для каждого данного соавтора максимальные коэффициенты взаимосвязи.

Из четырех статей Лауба, написанных им до сотрудничества с Эйнштейном, были выбраны две [0.92], посвященные той же проблематике, что и статья, опубликованная в соавторстве с Эйнштей-

ном. Из работ Эйнштейна, предваряющих общую с Лаубом статью, была выбрана первая и основная его статья по специальной теории относительности «К электродинамике движущихся сред», точнее § 6 и § 9 [0.61, с. 22–35, 30–31], рассматриваемые в совокупности как один текст. Текст общей статьи [0.61, с. 115–122] фигурировал в анализе дважды — § 1 статьи и вся статья целиком.

При слепгировании был использован частотный словарь русского литературного языка Э. Штейнфельд [6.16], содержащий 2500 слов. В результате слепгирования оставалось несколько более половины всех встречающихся в тексте словарных слов. В текстах Лауба словарные слова переводились с немецкого на русский. Формулы принимались во внимание: каждое из встречающихся в тексте обозначений кодировалось набором слов на основе определения этого обозначения, данного в тексте. Например, в тексте статьи Эйнштейна и Лауба встречается обозначение «К» и дается его определение — «инерциальная система координат». Оно кодировалось набором словарных слов: «инерция», «система» и «координата». В результате слепгирования слово «система» исчезает и остается набор из двух сленговых слов: «инерция» и «координата». Математические обозначения (π , может быть, напрасно) во внимание не принимались.

Распределения сленговых слов в текстах Лауба и Эйнштейна (см. рис. 2.45, 2.46) могут быть аппроксимированы распределением Ципфа. Для Эйнштейна величина α меньше, чем для Лауба. Это может быть связано с большей творчеством Эйнштейна по сравнению с Лаубом (см. рис. 2.76–2.78).

Вспомним результаты, полученные в раннем исследовании [6.11; 6.12] (см. табл. 6.2.1–6.2.3 и связанный с ними текст). Большие значения ранговой энтропии распределения «смысловых» слов в заголовках публикаций в сочетании с меньшей ранговой же энтропией распределения трех и пяти слов-лидеров для Эйнштейна по сравнению с Планком связывались с тем, что Эйнштейн в отличие от Планка был «генератором теорий». Ранговая энтропия распределения Ципфа для $\alpha < 1$ возрастает с α (см. выражение (7.2.25)). Латентные переменные ученого «творчество» и «способность генерировать теории» достаточно близки, так что налицо, по-видимому, некоторое несоответствие результатов, полученных на этих разных эмпирических массивах. Мы пока не знаем, чем объяснить это несоответствие. Может быть, оно является следствием того, что в одном случае обобщались заголовки публикаций за целые десятилетия, а в другом — отдельные статьи. Так или иначе, сам факт существования этого несоответствия говорит о том, что необходимы серьезные опорные исследования сленговых наукометрических индикаторов.

Таблица 6.2.4

Результаты обчета соавторства И. Лауба и А. Эйнштейна
(уточненные данные работы [6.13])

	А	Б	В	Г	АГ	БГ	АВ	БВ
G	675	1131	1023	1191				
N	41	70	67	81				
S	1,144	1,279	0,976	1,172	1,310	1,391	1,111	1,195
ΔS	0,019	0,017	0,017	0,021	0,014	0,013	0,013	0,013
K					0,495	0,450	0,831	0,776
ΔK					0,066	0,062	0,060	0,059
$K_{AB}-K_{AG}$						0,335	$\pm 0,077$	
$K_{BV}-K_{BG}$						0,325	$\pm 0,076$	

В приведенной табл. 6.2.4 приняты следующие обозначения: А — § 1 совместной статьи Эйнштейна и Лауба; Б — совместная статья Эйнштейна и Лауба целиком; В — § 6 + § 9 статьи Эйнштейна; Г — две статьи Лауба в сумме; S — ранговая энтропия распределения сленговых слов; ΔS — статистическая ошибка расчетного значения энтропии; K — коэффициент взаимосвязи, рассчитываемый по формуле (6.1.3); ΔK — статистическая ошибка расчетного значения коэффициента взаимосвязи; G — число сленговых слов; N — число разных сленговых слов. При использовании формул (6.1.1) — (6.1.3) логарифмы брались десятичными.

Значения коэффициентов взаимосвязи совместной статьи Эйнштейна и Лауба со статьями Эйнштейна (АВ и БВ) оказываются больше аналогичных значений коэффициентов взаимосвязи для Лауба (АГ и БГ). Разность достигает 4,3 статистической ошибки этой разности. Это означает, что большее значение коэффициентов взаимосвязи для Эйнштейна по сравнению с коэффициентами взаимосвязи для Лауба установлено со статистической достоверностью порядка 99,998%.

Исследование заголовков публикаций 42 известных физиков, проведенное Ю. М. Низовцевым, носит поисковый характер, и его результаты являются предварительными. Однако мы излагаем далее эти результаты и делаем это потому, что они имеют определенную эвристическую ценность, показывая возможные направления применений сленгового метода. При обчете заголовков за сленговые принимались все слова, за исключением предлогов, артиклей, союзов и слов типа «введение в», «к вопросу о» и пр. Строились частотные распределения сленговых слов по десятилетиям для отдельных ученых. Энтропия этих распределений трактовалась как мера предметно-логической разносторонности данного ученого

в соответствующем десятилетии его публикационной деятельности. Совокупность значений энтропии за все десятилетия образует возрастную кривую разносторонности ученого.

Кроме энтропии распределения всей совокупности сленговых слов, рассчитывалась энтропия распределения трех и пяти наиболее часто встречающихся в данном десятилетии у данного ученого слов-лидеров, которая трактовалась как мера предметно-логической концентрированности ученого в данное десятилетие.

Сленговые слова разделялись на «общенаучные» и «специальные». Процесс отбора «общенаучных» слов из всей совокупности сленговых был формализован с помощью специальных англо-русских словарей — политехнического (за отсутствием физического), химического, биохимического, медицинского и словаря по ядерной физике. К «общенаучным» слово относилось, во-первых, если оно не встречалось в специальном словаре, соответствующем профилю ученого (в данном случае в политехническом) и, во-вторых, если оно встречалось во всех специальных словарях. Доля отобранных таким образом «общенаучных» слов варьировала от 5 до 30 % для разных ученых. Оставшиеся слова отнесены к «специальным». Их распределения мало отличаются от распределений сленговых слов.

Энтропия частотного распределения «общенаучных» слов трактовалась как мера общенаучной разносторонности ученого. Как и в предыдущем случае, строились возрастные кривые теперь уже общенаучной разносторонности 42 физиков. Рассчитывалась корреляция между возрастными кривыми «специальной» разносторонности и концентрированности для каждого из ученых с помощью рангового коэффициента корреляции Кендэла. Ученые разбились на две группы (для первой эта корреляция положительна, а для второй — отрицательна), с «буферной» группой посредине.

Мы не претендуем здесь на содержательный анализ работ рассматриваемых ученых. Наша задача — проверить предположение, что большая величина нашего индикатора говорит (среднестатистически) о большей способности к созданию новых теорий. Для этого с помощью Большой советской энциклопедии (2-е изд.) было подсчитано, авторами скольких теорий считается каждый из 42-х физиков. Эти данные сгруппированы в табл. 6.2.5.

В табл. 6.2.5 K_1 — корреляция возрастных кривых «специальной» разносторонности, определяемой по энтропии распределения «специальных» сленговых слов, и концентрированности, определяемой по энтропии распределения пяти слов-лидеров. K_2 — аналогичен K_1 , только здесь берутся три слова-лидера.

Табл. 6.2.5 показывает, что на одного ученого I группы при-

Таблица 6.2.5
Распределение 42 физиков в соответствии с коэффициентом корреляции возрастных кривых «специальной» разносторонности и концентрированности и числом теорий, авторами которых они считаются

Кодовые имена ученых	K_1	K_2	Число теорий, n
I группа			
A-1	1.00	1.00	3
A-2	1.00	1.00	1
A-3	1.00	1.00	0
A-4	1.00	1.00	2
A-5	1.00	1.00	0
A-6	0.80	0.20	2
A-7	0.74	0.74	1
A-8	0.66	0.00	0
A-9	0.60	0.60	4
A-10	0.33	0.33	3
A-11	0.33	0.33	3
A-12	0.33	0.33	0
A-13	0.33	0.33	0
A-14	0.33	0.33	1
A-15	0.20	0.60	0
			n=1.33
«Буферная» группа			
B-1	0.66	0.20	0
B-2	0.66	0.00	2
B-3	1.00	-1.00	0
B-4	0.50	-0.33	0
B-5	0.33	0.00	1
B-6	0.00	0.00	0
B-7	0.00	0.00	0
B-8	0.00	0.00	1
B-9	-0.33	1.00	0
B-10	-0.33	0.00	5
B-11	-0.33	0.00	0
II группа			
C-1	-0.20	-0.33	1
C-2	-0.20	-0.40	0
C-3	-0.33	-0.33	0
C-4	-0.33	-0.33	0
C-5	-0.33	-0.33	0
C-6	-0.33	-0.33	0
C-7	-0.33	-0.33	0
C-8	-0.33	-0.33	1
C-9	-0.33	-0.33	0
C-10	-0.33	-0.33	1
C-11	-0.33	-0.33	0
C-12	-0.33	-0.33	1
C-13	-0.40	-0.40	0
C-14	-0.66	-0.66	0
C-15	-1.00	-0.33	0
C-16	-1.00	-1.00	0
			n=0.25

ходится в среднем, согласно БСЭ, 1,33 теории, тогда как на ученого II группы — 0,25. Коэффициент ранговой корреляции Спирмена двух последовательностей ученых, ранжированных, во-первых, по величине $(K_1 + K_2)/2$ и, во-вторых, по величине n , оказался равным 0,45. Учитывая крайнее несовершенство такого индикатора рассматриваемой нами в данном случае латентной переменной, как «„показания“ БСЭ», совершенно на это не рассчитанной, обнаруженная здесь по заголовкам публикаций положительная корреляция (о ее величине гауссовый по своей природе коэффициент Спирмена судить не позволяет) вселяет определенные надежды. Есть основания полагать, что переход к обсчету полных текстов публикаций и к формализованной процедуре счета, какая, например, имела место при обсчете соавторства Эйнштейна и Лауба, приведет к более тесной связи данного индикатора с данной латентной переменной.

Значения «общенаучной» разносторонности ученых за отдельные десятилетия их публикационной деятельности усреднялись и вычислялось среднее для данного ученого за жизнь значение «общенаучной» разносторонности. И здесь ученые разбились на две группы (для первой эти значения больше, а для второй — меньше), с «буферной» группой посредине.

Мы высказываем предположение, что большая величина данного индикатора свидетельствует (среднестатистически) о большей способности «генерировать» идеи. Для проверки этого снова была взята БСЭ и подсчитывалось, авторами скольких идей считается каждый из 42 ученых. Поскольку в БСЭ идеи как таковые не фиксируются, данный ученый считался автором идеи, если, согласно БСЭ, он автор либо нового метода, либо новой модели, либо новой формулы и т. п. Эти данные сгруппированы в табл. 6.2.6.

Табл. 6.2.6 показывает, что на одного ученого I группы приходится в среднем, согласно БСЭ, 2,12 идей, тогда как на ученого второй — 1,00. Коэффициент ранговой корреляции Спирмена в данном случае составляет 0,38. Таким образом, эти данные показывают, что даже при ограничении заголовками публикаций индикатор «энтропия распределения „общенаучных“ слов» положительно коррелирует с таким несовершенным индикатором латентной переменной «способность ученого генерировать идеи», как «„показания“ БСЭ», явно на это ее применение не рассчитанной. Есть основания полагать, что и здесь переход к обсчету полных текстов и к формализованной процедуре счета приведет к более тесной связи данного индикатора с данной латентной переменной.

Информационную продуктивность ученого Ю. М. Низовцев определяет умножением энтропии распределения сленговых слов на число этих слов. Построенные для 42 физиков возрастные кри-

Таблица 6.2.6

Распределение 42 физиков в соответствии со средним за жизнь ученых значением «общенаучной» разносторонности и числом идей, авторами которых они считаются

Кодовые имена ученых	«Общенаучная» разносторонность	Число идей, n	Ученые	«Общенаучная» разносторонность	Число идей, n
I группа			«Буферная» группа		
A-1	1,357	5	B-5	1,006	0
A-2	1,337	2	B-6	0,970	1
A-3	1,251	3	B-7	0,967	5
A-4	1,250	3	B-8	0,931	2
A-5	1,240	2	B-9	0,886	4
A-6	1,208	1	B-10	0,877	1
A-7	1,172	3	B-11	0,869	1
A-8	1,170	2	II группа		
A-9	1,165	2	C-1	0,865	0
A-10	1,150	1	C-2	0,839	2
A-11	1,144	3	C-3	0,832	0
A-12	1,142	1	C-4	0,813	1
A-13	1,129	1	C-5	0,799	0
A-14	1,117	4	C-6	0,778	0
A-15	1,115	0	C-7	0,723	1
A-16	1,111	1	C-8	0,695	2
$\bar{n}=2,12$			C-9	0,695	3
«Буферная группа»			C-10	0,695	1
B-1	1,070	3	C-11	0,628	1
B-2	1,065	0	C-12	0,628	2
B-3	1,063	2	C-13	0,560	0
B-4	1,028	4	C-14	0,522	2
			C-15	0,506	0
			$\bar{n}=1,00$		

вые информационной «специальной» и «общенаучной» продуктивности имеют по одному максимуму, причем эти максимумы смещены друг относительно друга. Сравнение этих данных с кривыми Д. Пельца и Ф. Эндрюса (см. гл. 3, § 2) показывает, что максимум кривой «специальной» продуктивности приходится примерно на тот же возраст, что и первый максимум кривых этих авторов, а максимум кривой «общенаучной» продуктивности — на тот же

возраст, что и второй максимум Д. Пельца и Ф. Эндриуса. Поэтому выдвинутая ими гипотеза о «дивергентной» (творческой) природе первого их максимума и «конвергентной» (синтезирующей) природе второго получает в приведенных здесь предварительных данных сленгового анализа подтверждение.

3. ПЕРСПЕКТИВЫ

Сленговый метод, являясь результатом развития методов «контент-анализ» и тезаурусного, имеет, на наш взгляд, право на статус самостоятельного наукометрического метода, имея свой индикатор — «число слов сленга» — и являясь, в отличие от своих «родителей» *формализованным*, что обеспечивает воспроизводимость даваемых им результатов.

Число слов сленга — самый «дробный» наукометрический измеритель, что определяет основное его преимущество: он позволяет проводить исследования на тех массивах, где применимы другие наукометрические измерители, и на существенно меньших, где менее «дробные» измерители дают статистически недостоверные результаты или вовсе не применимы. Более того, сленговый метод обладает определенной *гибкостью*: «свертывая» в ходе сленгирования исходные массивы публикаций, мы получаем счетные массивы меньшего объема. Варьируя степень свертывания, мы получаем возможность варьирования объема исследуемого массива при заданном объеме трудозатрат.

Недостатком сленгового метода является то, что он в сравнении с другими наукометрическими методами требует больших затрат труда, поэтому нецелесообразно его использовать там, где применимы менее «дробные» измерители, т. е. на больших массивах публикаций.

Впрочем, вопрос этот не однозначен. Во-первых, сленговый метод, как говорилось, *гибок*, т. е. дает возможность *варьировать* объемы массивов публикаций при заданном объеме счетных массивов посредством варьирования степени сленгирования. Во-вторых, сленговый метод по самому своему содержанию требует использования ЭВМ, что существенно расширит границы его применения.

Подчеркнем, что необходимость использования для сленгового метода аппарата частотных словарей (к счастью, уже разработанного для нужд статистической лингвистики) и ЭВМ, что является его недостатком, определяется большой «дробностью» лежащего в его основании наукометрического индикатора «число слов сленга», что является его достоинством. Ближайший к сленговому по шкале «дробности» индикатор наукометрический метод — «цитат-

индекс» — также потребовал специального аппарата — указателей научных ссылок, — который пришлось специально разрабатывать. Опыт статистической лингвистики говорит, что автоматизация сленгового анализа на базе ЭВМ является вполне реальной задачей.

Опорные исследования числа слов сленга проводились пока в совершенно недостаточном объеме и носили предварительный характер. Поэтому излагаемые далее соображения о связи числа слов сленга как наукометрического индикатора с латентными переменными науки также носят предварительный характер.

Сленговые индикаторы науки могут строиться, прежде всего, на использовании *отдельно взятого распределения слов сленга*, характеристики которого и выступают в качестве таких индикаторов. Мы имеем в виду *энтропию* распределения и *показатель сходимости α* этих распределений (как и другие наукометрические стационарные распределения, сленговые распределения являются *цифровыми* — см. гл. 7). Энтропия и α между собой однозначно связаны (см., например, для распределения Ципфа выражения (7.2.24), (7.2.25)). Поэтому энтропия и α в известной мере равноправны как сленговые индикаторы.

Как было показано в § 2, *энтропия* сленгового распределения (следовательно, и α) может служить мерой *концентрированности* или, напротив, *разносторонности* научных интересов ученого. Если не ограничиваться текстами одного ученого, а принимать во внимание тексты, скажем, научного коллектива, то *энтропия* и α могут служить мерами *координированности* труда ученых. Они могут выступать и мерами *творчесткости* коллектива (ср. наблюдение Ю. А. Шрейдера, приведенное в гл. 4, § 6 о том, что величина $\gamma=1/\alpha$ характеризующего данный коллектив распределения может служить мерой его «творчесткости»). Это согласуется с приведенными в § 2 предварительными данными о том, что *энтропия* (следовательно, и α) может выступать мерой *способности ученого генерировать новые идеи*. Эта же энтропия, как показано там же, может служить и индикатором *способности ученого создавать новые теории*.

В. И. Горькова [1.11], как говорилось, например, в гл. 4, § 5, пришла к выводу о возможности по длине «хвоста», характеризующего научное направление стационарного распределения, судить об его *перспективности* или, если воспользоваться ее терминологией, *проблемности* научного направления. Очевидно, это должно распространяться и на сленговые распределения. Длину «хвоста» распределения можно характеризовать его показателем *сходимости α* или энтропией. Поэтому *энтропия* распределения сленговых слов или α могут служить, согласно выводам В. И. Горьковой,

индикаторами *проблемности* данного научного направления. При этом, конечно, следует прибегать к сленгированию минимальной степени свернутости, чтобы охватить характеризующий данное научное направление массив публикаций.

Сленговые индикаторы науки могут строиться и на использовании *двух сленговых распределений*, характеристика связи которых и выступает в качестве таких индикаторов. Мы имеем в виду, прежде всего, энтропийные коэффициенты корреляции (см. выражения (6.1.3), (6.1.4)), которыми здесь следует пользоваться взамен неприменимых к ципфовым распределениям обычных квадратичных коэффициентов корреляции (см. гл. 7).

Как показано в § 2, коэффициенты корреляции сленговых распределений могут служить мерой *индивидуального вклада ученого в соавторскую публикацию*. Если не ограничиваться текстами двух ученых, а принимать во внимание тексты, скажем, научного коллектива, то коэффициент корреляции между сленговыми распределениями текстов данного ученого, с одной стороны, и текстами данного коллектива — с другой, может служить мерой степени *связанности тематики исследований и научных интересов ученого и научного коллектива*. Аналогично с помощью сленгового метода можно измерить силу связи предметных интересов ученого и научных коллективов с планируемыми исследованиями.

Сленговые индикаторы науки могут строиться, наконец, на использовании *данного множества сленговых распределений*. При этом путем измерения средствами сленгового анализа взаимосвязи отдельных выполненных в данной предметной области исследований, отдельных ученых, коллективов и т. д. воссоздается *сеть связей* этих единиц анализа, а далее с помощью, скажем, факторного анализа — сленговая структура этой данной области исследований.

Сленговая структура научного коллектива явится *«слепоком» предметной его структуры*, которую полезно, например, сравнить со структурой организационной. Как уже говорилось в гл. 4, § 5 в применении к сетям цитирования, *актуальность и перспективность* исследований, связанных с данным узлом сленговой структуры, измеряется степенью связанности этого узла со всей структурой. Если мы построим сленговую структуру в крупном масштабе, используя при сленгировании малую степень свернутости, то получим возможность судить об относительной *перспективности научных направлений*. Уменьшив масштаб, т. е. перейдя к сленгированию большой степени свернутости, мы сравним относительную *перспективность исследований отдельных научных коллективов*. Еще более уменьшив масштаб сленговой структуры, мы сравним относительную *перспективность исследований отдельных ученых*. Эти оценки явятся и оценками эффективности деятель-

ности научных коллективов и отдельных ученых. В отличие, например, от «цитат-индекса» или методов экспертной оценки сленговый метод дает возможность оценивать *действительную актуальность* отдельных исследований, а не мнения ученых об этом, пусть даже зафиксированные в ссылках. Поэтому сленговый метод предоставляет в принципе возможность выявления важных для исследований в данной предметной области работ, еще не известных научной общественности, и таким образом наметить пути решения задачи *выявления потенциальных лидеров науки*.

По-видимому, указанные здесь связи между сленговыми индикаторами и латентными переменными научной деятельности действительно существуют. Вопрос заключается в том, *насколько* они существенны, т. е. насколько больше нуля коэффициенты корреляции между сленговыми и латентными распределениями. Если связи есть, но слабо выражены, то их существование трудно использовать в исследованиях науки. Здесь, как и для всех наукометрических индикаторов, необходимы специальные опорные исследования, которые должны вестись на базе негауссовой статистики с широким охватом всей системы индикаторов и латентных переменных науки.

ОСНОВНЫЕ НАУКОМЕТРИЧЕСКИЕ ЗАКОНОМЕРНОСТИ

Наукометрические закономерности, позволяя унифицировать результаты измерений отдельных объектов науки и создавая тем самым предпосылки для сравнения последних, представляют собой основу дальнейшего развития наукометрии и науковедения в целом. Современные наукометрические данные дают возможность говорить о двух таких закономерностях: 1) так называемый закон экспоненциально-логистического роста науки, характеризующий временную динамику науки; 2) негауссовость (цифровость) стационарных наукометрических распределений, характеризующая ее статистику.

1. ЭКСПОНЕНЦИАЛЬНО-ЛОГИСТИЧЕСКИЙ РОСТ НАУКОМЕТРИЧЕСКИХ ИНДИКАТОРОВ ВО ВРЕМЕНИ

Обсуждая эмпирические данные по временной динамике науки, полученные с помощью различных методов (см. гл. 2–5), мы видели, что часть этих данных может быть аппроксимирована экспонентой, тем лучше соблюдающейся, чем большие счетные массивы рассматриваются. Мы видели также, что в ряде случаев заметно замедление экспоненциального роста. Эти данные и легли в основу концепции экспоненциально-логистического роста науки, автором которой является Д. Прайс [1.58].

«Исходным пунктом наших рассуждений, — пишет Прайс, — будут эмпирические свидетельства статистики, почерпнутые из множества справочников по разным отраслям и аспектам науки. Все они с удивительной настойчивостью и постоянством показывают, что если найден более или менее удовлетворительный способ измерить какой-либо достаточно большой сегмент науки, то этот сегмент в нормальных условиях растет экспоненциально» [1.58, с. 287]. Наиболее любопытное и важное свойство экспоненциального закона, по мнению Прайса, «состоит в том, что в отличие от множества других процессов, допускающих интерпретацию кри-

вых, экспоненциальный рост науки подтверждается в течение двух-трех столетий... это порождает сильнейшее подозрение, что при должном выборе индексов и параметров можно было бы показать (как раз это мы и пытаемся сделать), что для подобного анализа существует солидное теоретическое обоснование» [1.58, с. 288].

«Второй важной чертой роста науки, — продолжает Д. Прайс, — является необычно высокий темп этого роста, каким бы способом он ни измерялся... В зависимости от того, что именно и как измеряется, объем науки по числу ученых или по числу публикаций удваивается за период от 10 до 15 лет. 10-летний период удвоения получают в случаях поточных замеров, где не учитывается качество работ, и в основу определения критерия научности берут минимум требований. 15-летний период удвоения возникает в тех случаях, когда отбор производится более строго и учитывают лишь те научные работы или только тех авторов, которые удовлетворяют более узким определениям научности. Если строгость отбора возрастает и учитывают только научные труды очень высокого качества, период удвоения начинает увеличиваться и приближаться к 20 годам» [1.58, с. 288–289].

Концепция экспоненциального роста науки является для Д. Прайса частью логистической концепции ее роста. Экспоненциальный рост науки, пишет Д. Прайс, «представляет собой в действительности весьма ненормальное положение вещей. В реальном мире не бывает так, чтобы вещи росли и росли до бесконечности. Экспоненциальный рост постепенно приближается к какому-то пределу, процесс замедляется и останавливается, не достигая абсурдных значений. Эта функция, которая более полно отражает реальное положение дел, хорошо известна как логистическая кривая, которая дана в нескольких различающихся математических формах. Здесь следует снова заметить, что на данной стадии, когда мы еще очень мало знаем о феноменах науки, нас не должны отвлекать и заботить математические детали и точность формулировки. В первом приближении... вполне достаточно будет учитывать лишь общую тенденцию роста» [1.58, с. 301].

«Таким образом, даже без ссылок на современное положение дел или на оценки того, когда и где объявится «потолок», ограничивающий рост, для нас становится очевидным, что в пределах жизни одного человеческого поколения наука должна будет отказаться от традиционного экспоненциального роста» [1.58, с. 303].

Свои рассуждения Д. Прайс иллюстрирует рисунком (см. рис. 1.16).

Экспоненциальный рост науки по ряду показателей был известен и до Д. Прайса («то, о чем мы пока говорили, — пишет в этой

связи Д. Прайс, — в общем-то хорошо известно и не вызывает серьезных разногласий среди тех, кто исследует науку» [1.58, с. 294]). Самое важное в его концепции касается прежде всего необходимости перехода науки в процессе своего роста от стадии экспоненты к стадии насыщения.

В. В. Налимов и соавт. [1.7], М. М. Карпов [1.25], В. В. Налимов и З. М. Мульченко [1.45—1.46], Г. М. Добров и соавт. [0.27], Д. Прайс [1.58—1.59], Р. Л. Хамблин [0.85] и другие дают вывод экспоненты и логисты роста науки, основанный на использовании дифференциальных уравнений. Именно при получении экспоненты скорость роста данного показателя науки dy/dt полагается пропорциональной этому показателю y

$$dy/dt = ky, \quad k > 0. \quad (7.1.1)$$

Решение уравнения (7.1.1) есть экспонента:

$$y = y_0 e^{k(t-t_0)}. \quad (7.1.2)$$

При выводе логистической кривой задаются дифференциальным уравнением

$$dy/dy = ky(b-y), \quad (7.1.3)$$

решением которого является логистическая зависимость

$$y = \frac{b}{1 + y_0 \exp(-kbt)}. \quad (7.1.4)$$

Экспоненциально-логистическая концепция роста науки обсуждается в целом ряде работ. Анализируя математическую форму, которую имеет эта концепция в уравнениях (7.1.1) — (7.1.4), Н. М. Сагалович [3.30], Э. Янч [0.66], Е. Д. Гражданников [1.12], Б. Я. Брусиловский [1.5; 0.11] и другие показывают, что лежащие в основании этих уравнений предположения отнюдь не бесспорны и поэтому дифференциальные уравнения (7.1.1) и (7.1.3) сами нуждаются в выводе. Э. Янч, например, обращает внимание на то, что предположение о пропорциональности скорости роста параметров науки этим параметрам, на котором основано уравнение (7.1.1), справедливо лишь: а) при условии существования идеальной связи между всеми исследованиями и элементами возможностей, обусловленных этой связью. Конечно, говорит Э. Янч, оба эти условия не соблюдаются. Более правомерным представляется ему предположение о пропорциональности скорости роста параметров науки не этим параметрам, а их логарифмам, т. е. $dy/dt = k \lg y$ вместо (7.1.1). В этом случае кривая роста науки име-

ла бы вид прямой линии, если знания растут по экспоненте, и вид экспоненты, если знания растут по двойной экспоненте.

Б. Я. Брусиловский [0.11, с. 182] делает другое предположение, записывая вместо (7.1.1) выражение

$$dy/dt = \alpha(y)y, \quad (7.1.5)$$

откуда, раскладывая $\alpha(y)$ в ряд Тейлора, получает различные кривые роста.

Существующие математические модели роста науки, как рассмотренные выше, так и оставшиеся вне нашего рассмотрения (см. о таких моделях, например, в работе А. И. Яблонского [0.64]), отличаются произвольностью лежащих в их основе предположений*. Теория не в состоянии дать сегодня удовлетворительный вывод экспоненты, логисты или какой-либо другой кривой роста науки.

Теоретические трудности, встающие в связи с экспоненциально-логистической концепцией роста науки, становятся понятными, если обратиться к эмпирическим данным роста науки (см. рис. 1.1—1.20). Как показано, например, на рис. 1.4б, 1.8, 1.19а, 1.20, в ряде случаев, если мы рассматриваем большие промежутки времени (порядка столетия или больше) и имеем достаточно большие счетные массивы (скажем, целые страны или регионы), экспонента соблюдается лишь в первом приближении. На нее накладываются, как говорит Прайс [1.145], «пульсации», которые пока не поддаются математическому описанию. Более того, часто (см., например, рис. 1.18) кривые роста науки вообще не поддаются какой-либо аппроксимации. Поэтому количественные формулировки экспоненциально-логистической концепции роста науки имеют ограниченное значение. И это прекрасно понимает Прайс, призывающий, как мы видели, не обращать внимания на «математические детали и точность формулировки» концепции.

Тем не менее экспоненциально-логистическая концепция дает основания для чрезвычайно важных выводов: о том, что наука, опережающая по темпам роста на протяжении последних двух-трех столетий другие сферы человеческой деятельности, должна

* Ряд моделей s -образных кривых роста можно найти, например, в книге Э. Янча [0.66]. Логиста (7.1.4) как средство описания научно-технического прогресса появляется уже в работе Л. Райденура 1951 г. Р. Ленц выводит логисту на основе аналогии развития науки и техники с развитием эмбриона. В качестве s -образной кривой роста берут нормальную функцию распределения, а также функцию распределения логнормального распределения. Свои модели кривой роста дают Р. Айзенсон и Л. Гартман.

в течение ближайших десятилетий резко снизить темпы роста; о надвигающемся переходе науки от преимущественно экстенсивного развития к преимущественно интенсивному; о *предстоящих качественных преобразованиях науки*. Кроме этих общих выводов, из обсуждаемой концепции следуют и некоторые частные, например вывод об ожидающемся постарении научного социума [см., например, 1.46, с. 64—65].

Л. Новый и Я. Фолта [1.52], В. В. Налимов и З. М. Мульченко [1.46], Г. М. Добров и др. [0.27] и другие обсуждают вывод о замедлении экстенсивного роста науки. Для этих авторов ясно, что снижение темпов роста науки и даже выход на плато наукометрических индикаторов не означает угасания науки. Индикаторы характеризуют лишь отдельные стороны жизни науки, и замедление темпов их роста не есть замедление темпов развития науки. Другими словами, число ученых, число публикаций и другие наукометрические индикаторы лишь приблизительно отражают латентные переменные «развитие науки». Замедление темпов роста этих индикаторов науки сопровождается по-прежнему высоким ростом этих латентных переменных. Задачей наукометрии является поиск новых наукометрических индикаторов развития науки, более адекватно отражающих это развитие. Эта задача порождает другую — подключение к наукометрическим другим науковедческим индикаторов развития науки, так как ясно, что только такое комплексное использование измерителей науки различной природы, обеспечивающее всестороннее рассмотрение науки во временной ее динамике, может гарантировать нас от неверных выводов относительно будущего науки.

К выводу о необходимости комплексного подхода к анализу развития науки приводит обсуждение вопроса о механизмах, обеспечивающих замедление экстенсивного роста науки [1.46, с. 41—47]. В. В. Налимов и З. М. Мульченко приводят два таких фактора: рост науки, по их мнению, замедляет, во-первых, потоки научной информации и, во-вторых, прохождение научной идеи своего пути из-за затрудненности поиска ею адресатов.

Можно указать и другие факторы, замедляющие рост науки, скажем, уменьшение доли талантливых ученых в общей массе научных работников.

Более того, можно указать и факторы, ускоряющие рост науки. Скажем, улучшение образования, использование новой техники коммуникаций, уменьшение временных познаний между открытием и его внедрением и т. д.

При анализе развития науки необходимо далее учитывать и внешние по отношению к науке факторы, модель науки как самоорганизующейся системы здесь недостаточна. Д. Прайс при

обсуждении своей концепции экспоненциально-логистического роста науки использует, как мы видели, в качестве аналогии биологический организм. Влияние внешних факторов очевидно и в этом случае, несмотря на ограниченность аналогии. Действительно, замедление темпов роста биологического организма связано не только с его внутренними особенностями, но и с внешними условиями: гравитацией, наличием корма и конкурирующих видов и т. д. Темпы экстенсивного развития науки также определяются не только внутренней логикой ее развития, но и уровнем развития производства, общей численностью населения и т. д.

Таким образом, мы имеем еще одно свидетельство отсутствия приемлемого теоретического обоснования экспоненциально-логистической концепции роста науки: ясно, что в таком обосновании должны учитываться если не все факторы, то по крайней мере важнейшие. А этого до сих пор не делалось в силу трудности такого многофакторного анализа, выходящего за рамки обычных математических построений.

2. НЕГАУССОВСТЬ СТАЦИОНАРНЫХ НАУКОМЕТРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

В настоящем параграфе мы исследуем *стационарные* наукометрические распределения. Эти распределения имеют необычайно длинные «хвосты» и описываются с помощью «длиннохвостых» аппроксимаций. В первых двух разделах параграфа излагаются элементы теории таких аппроксимаций, а в третьем дается их обзор, называются распределения: Ципфа, Парето, Мандельброта, логнормальное, Пуассона, бета-функция, отрицательное биномиальное и другие — общим числом свыше двадцати.

Эмпирические данные могут быть аппроксимированы разными теоретическими распределениями. Статистические гипотезы проверяют с помощью специальных критериев. Однако эта проверка чаще всего не дает однозначных результатов, поэтому, во-первых, разные авторы часто используют разные аппроксимации, чем и объясняется их обилие; во-вторых, авторы, понимающие ограниченность критериев статистических гипотез, также допускают использование разных аппроксимаций. Так, Д. Прайс [1.142] говорит: «Распределения по числу публикаций и их качеству, измеренному с помощью ссылок, следуют математическим законам, которые с прекрасной точностью представлены гиперболическим, логнормальным или отрицательным биномиальным распределением — не видно аргументов в пользу предпо-

тительного использования какой-либо одной аппроксимации»*.

Мы попытаемся показать, что в этой схеме есть существенные недостатки**, что для описания стационарных наукометрических распределений пригодны не все теоретические распределения, а лишь имеющие при больших значениях переменных вид распределения Ципфа. (Такие распределения мы называем *ципфовыми*). Из приведенных выше в принципе не пригодны для аппроксимации распределения: Пуассона, логнормальное, отрицательное биномиальное, логарифмическое.

Мы опираемся на точку зрения А. И. Яблонского [0.63; 0.114], который развивает, в свою очередь, подход Б. Мандельброта [0.97; 0.98]. Этот подход излагается в четвертом разделе параграфа.

Вообще говоря, моменты распределения Ципфа бесконечны, поэтому Яблонский указывает на непригодность аппарата моментов для его описания. Кроме того, распределение Ципфа сходится к устойчивым распределениям, отличным от нормального. Распределения, сходящиеся к устойчивым, отличающимся от распределения Гаусса, далее называются *негауссовыми*. Таким образом, распределение Ципфа негауссово. Яблонскому принадлежит заслуга привлечения к анализу роли распределения Ципфа в наукометрии теоремы (сформулированной независимо Б. В. Гнеденко в 1939 г. и В. Дебленим в 1940 г.) негауссовой статистики. На основании этой теоремы он установил, что для того, чтобы асимптотика данного распределения совпадала с распределением Ципфа, необходимо, чтобы оно было негауссовым. В связи с этим Яблонский предполагает, что распределение Ципфа играет ту же универсальную роль предельного распределения в соответствующих областях, что и нормальное распределение в неорганических процессах***.

В параграфе предпринимается попытка уточнить точку зрения Яблонского и развить ее. Для Яблонского распределение Ципфа при всей его универсальности остается все же лишь одной из аппроксимаций. Например, наряду с распределением Ципфа он использует логнормальное распределение, которое является гауссо-

вым неципфовым (гауссовыми здесь и далее называются распределения, сходящиеся к распределению Гаусса).

В пятом разделе будет показано, что недооценка Яблонским роли распределения Ципфа объясняется тем, что он использует частную формулировку теоремы Гнеденко-Дебленина. Из общей ее формулировки следует: для того чтобы данное распределение было негауссовым, необходимо и достаточно, чтобы его асимптотика совпадала (с точностью до медленно меняющихся функций) с распределением Ципфа. Таким образом, негауссовые распределения могут быть (с точностью до медленно меняющихся функций) только ципфовыми.

В связи с этим оказывается, что универсальность распределения Ципфа отличается от универсальности распределения Гаусса. Если гауссовые распределения сходятся к распределению Гаусса, то негауссовые имеют распределение Ципфа своей асимптотикой при больших значениях переменных.

Развивая идеи Яблонского, мы используем подход, противоположный подходу Г. Юла. Последний [6.22] отказывается от распределения Ципфа по той причине, что его среднее бесконечно. Мы отказываемся от гауссовых распределений по той причине, что их моменты конечны.

Для гауссовой генеральной совокупности, т. е. для совокупности с конечными моментами, выборочные моменты несущественно зависят от объема выборки. Для негауссовой генеральной совокупности, т. е. для совокупности с бесконечными моментами, выборочные моменты существенно зависят от объема выборки. В последнем случае аппарат моментов гауссовой статистики дает невоспроизводимые результаты, т. е. результаты, меняющиеся от исследователя к исследователю, поскольку разные исследователи имеют, вообще говоря, выборки разных объемов. Поэтому применять гауссовую статистику моментов на негауссовой генеральной совокупности некорректно.

Рассмотрение в шестом разделе параграфа большого числа стационарных наукометрических распределений показывает, что их моменты существенно зависят от объема выборок. На этом основании формулируется тезис о *негауссовости научной деятельности*, который обобщается в седьмом разделе параграфа на дополнительном эмпирическом материале в тезис о *негауссовости социальных явлений*. Эти тезисы обсуждаются затем в восьмом — двенадцатом разделах.

Негауссовость генеральных совокупностей для научной деятельности приводит к некорректности использования при их анализе гауссовой статистики моментов. Но атрибутами гауссовой статистики являются не только моменты и построенные на их

* Это высказывание относится к 1974 г. Через два года Прайс отдал предпочтение одной аппроксимации — бета-функции, а еще через четыре он опять вернулся к использованию наряду с бета-функцией отрицательного биномиального распределения (см. гл. 3, § 3).

** Основные положения, развиваемые в параграфе, изложены в работе [1.111].

*** К. Скарротт говорит, что распределение Ципфа может оказаться столь же фундаментальным (и, возможно, столь же важным), «как и другие статистические распределения, такие, как распределение Гаусса» [0.104]. Но этот его тезис остается просто догадкой, вербальным.

основе конструкты, например, коэффициенты корреляции, но и сами теоретические гауссовы распределения, например: логнормальное, Пуассона, отрицательное биномиальное, логарифмическое и др. Применение этих распределений для аппроксимации стационарных наукометрических распределений поэтому также некорректно.

Необходимость использования при количественном анализе науки негауссовой статистики — только одно из следствий феномена негауссовости научной деятельности, носящее, так сказать, прикладной характер. Другие следствия, в том числе и более общие, были рассмотрены в гл. 1.

1. Частотная и ранговая формы статистических распределений

Соотношение частотной и ранговой форм статистического распределения обсуждается в ряде работ (см., например [1.2; 1.78; 1.7; 1.116; 0.114]). Нам приходилось сталкиваться с заблуждением, что ранговая форма существует только для некоторых распределений (например, для распределения Ципфа), не встречающихся в естественных науках. Это не так, и любое выборочное статистическое распределение имеет частотную и ранговую формы.

Частотная дифференциальная форма выборочного статистического распределения определяется $n(x)$ — частотой появления данного значения случайной величины x в выборке объема N :

$$\sum_{x_0}^I n(x) = N, \quad (7.2.1)$$

где x_0 и I — соответственно минимальное и максимальное выборочные значения случайной величины x .

Частотная интегральная форма определяется

$$F(x) = \frac{1}{N} \sum_{x_0}^x n(\xi) \quad (7.2.2)$$

— функцией распределения выборочного статистического распределения.

Ранговая форма выборочного статистического распределения вводится соотношением:

$$r = \sum_x^I n(\xi), \quad 1 \leq r \leq N. \quad (7.2.3)$$

Ранговая дифференциальная форма $x(r)$ определяется непосредственно соотношением (7.2.3). Ранговая интегральная форма:

$$X(r) = \sum_1^r x(\xi), \quad \sum_1^N x(\xi) = G. \quad (7.2.4)$$

Ранг r имеет простой смысл. Это — порядковый номер данного значения случайной величины x , когда все эти значения расположены в порядке убывания частоты их встречаемости $n(x)$.

Мы различаем *ранговое распределение* и *ранговую последовательность*. В ранговом распределении объекты ранжируются по величине x , причем значения x фиксируются и непосредственно фигурируют в ранговом распределении $x(r)$ или $X(r)$. В ранговой же последовательности объекты также ранжируются по величине x , однако значения x не фиксируются и нигде не фигурируют. Ранговая последовательность не является ранговым аналогом статистического распределения.

В ранговом распределении $x(r)$ все объекты имеют *разные* ранги. Некоторые авторы присваивают объектам, имеющим равные значения x , *равные* ранги и обозначают через $n(r)$ число объектов ранга r . С этими (частотными) распределениями $n(r)$ они и работают, вычисляют их характеристики, относя эти характеристики к исследуемым объектам. Это не совсем корректно, так как форма и, следовательно, характеристики распределения $n(r)$ зависят не только и не столько от природы объектов, сколько от *точности* нашего измерения. Если, скажем, точность достаточно велика, т. е. если мы замечаем мельчайшие различия между объектами по величине x , то все объекты имеют *разные* ранги, так что $n(r) \equiv 1$. С уменьшением точности значения $n(r)$, вообще говоря, растут, и распределение $n(r)$ приобретает, как и все социальные распределения (см. далее), форму ципфоваго распределения. Показатель α ципфоваго распределения в этом случае растет, характеризуя выбранную методику измерения.

Следует заметить в этой связи, что форма распределений $n(x)$ или $x(r)$ также зависит от точности измерения, однако эта зависимость в известном смысле обратна зависимости от точности измерения для распределений $n(r)$. Первые приобретают все более четкую форму с *ростом точности* и теряют ее с *уменьшением* последней. Распределения $n(r)$, напротив, приобретают более четкую форму с *уменьшением точности* и теряют ее с *увеличением* последней. Ясно, что нас должны интересовать описания объектов, рельефность которых растет с *увеличением точности* измерения. Если рельефность описания растет с *уменьшением точности* измерения, то она приносится наблюдателем. Именно поэтому имеет

смысл применение распределений $n(x)$ и $x(r)$, а не $n(r)$. Последнее лучше применять в тех случаях, когда нас интересует именно методика измерения.

По аналогичным причинам не имеет смысла относить классификационные распределения (см., например, рис. 2.75) к объектам классификации. Здесь также с увеличением точности фиксации классификационных признаков распределение теряет форму. В пределе, когда мы с нулевой ошибкой фиксируем и отражаем в классификационной схеме мельчайшие различия между объектами, каждый объект попадает в свою рубрику и пребывает в ней в единственном числе. С увеличением ошибки измерения в каждую классификационную рубрику попадает все большее число объектов. Форма классификационных распределений описывает, таким образом, методику классификации и имеет, как и все социальные распределения (см. далее), форму ципфоваго распределения.

Возникает вопрос: почему в естественных науках доминируют частотные формы распределений, а в социальных — ранговые? Частотная форма удобна (применима), когда выборочная относительная частота $n(x)/N$ служит статистически достоверной оценкой плотности вероятности $f(x)$, т. е. когда относительная средняя квадратическая ошибка $\sigma_{(n/N)}/(n/N)$ распределения случайной величины $n(x)/N$, имеющей своим средним $n(x)/N = f(x)$ мала:

$$\frac{\sigma_{(n/N)}}{n/N} \ll 1. \quad (7.2.5)$$

Это имеет место, когда

$$\frac{\sigma_n}{n} \approx \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}} \ll 1, \quad (7.2.6)$$

т. е. когда

$$\frac{I - x_0}{\Delta x} \ll N, \quad (7.2.7)$$

где Δx — величина классовых интервалов, на которые разделен размах $I - x_0$ случайной величины x на данной выборке.

Ранговая форма удобна (применима), когда с достаточной статистической достоверностью определяются значения случайной величины x , т. е. когда

$$\frac{\sigma_{(x/\Delta x)}}{x/\Delta x} \approx \frac{\sqrt{x(r)/\Delta x}}{x(r)/\Delta x} = \frac{1}{\sqrt{x(r)/\Delta x}} \ll 1, \quad (7.2.8)$$

где Δx — снова классовый интервал. Условие (7.2.8) дает

$$\frac{I + x_0}{\Delta x} \gg 2. \quad (7.2.9)$$

Условия (7.2.7) и (7.2.9) существенно различаются. Частотная форма оказывается тем применимее, чем больше объем выборки N и меньше I . Ранговая форма применима при любом объеме выборки и тем применимее, чем больше I .

В естественных науках, как правило, объемы выборок велики. Поэтому частотная форма здесь удобна. Максимальные значения переменной I здесь, как правило, невелики (стационарные распределения в естественных науках имеют, как правило, короткие «хвосты» — см. десятый раздел параграфа). Поэтому ранговая форма здесь часто неудобна.

В социальных науках, напротив, объемы выборок, как правило, невелики, так что частотная форма здесь бывает неудобна; I , напротив, как правило, велико (стационарные социальные распределения имеют длинные «хвосты» — см. разд. 6 и 7). Поэтому ранговая форма здесь удобна.

Заметим, что соотношение (7.2.9) дает простой рецепт повышения статистической достоверности результатов при работе с ранговыми распределениями: следует уменьшать Δx , что связано с увеличением точности измерения случайной величины x . Эта процедура заменяет увеличение объема выборки, полезное в случае частотных распределений.

2. Распределение Ципфа: эффект рангового искажения.

Зависимость моментов от объема выборки

Теорию распределения Ципфа рассматривают Р. М. Фрумкина [6.8], Н. Хомский и Дж. Миллер [2.35], В. И. Горькова [1.11], Е. Тома и Д. Пуркару [6.7], Л. С. Козачков [1.28], А. И. Яблонский [0.63], [0.114] и др. Однако эту теорию нельзя пока считать завершенной.

Частотная дифференциальная форма распределения Ципфа:

$$n(x) = \frac{C}{x^{1+\alpha}}, \quad 0 < \alpha < \infty. \quad (7.2.10)$$

Частотная интегральная форма:

$$F(x) = \frac{C}{\alpha N} \left(\frac{1}{x_0^\alpha} - \frac{1}{x^\alpha} \right) \approx 1 - \frac{x_0^\alpha}{x^\alpha}. \quad (7.2.11)$$

Ранговая дифференциальная форма распределения Ципфа*:

$$x(r) = \frac{A}{(r+B)^{\gamma}}. \quad (7.2.12)$$

Ранговая интегральная форма:

$$X(r) = \begin{cases} A \ln \frac{r+B}{1+B}, & \gamma = 1; \\ \frac{A}{\gamma-1} \left[\frac{1}{(1+B)^{\gamma-1}} - \frac{1}{(r+B)^{\gamma-1}} \right], & \gamma \neq 1. \end{cases} \quad (7.2.13)$$

Граничные условия (7.2.3) дают:

$$x_0 = \frac{A}{(N+B)^{\gamma}}; \quad I = \frac{A}{(1+B)^{\gamma}}. \quad (7.2.14)$$

Для $x \leq x_0$, по определению распределения Ципфа, $n(x) = 0$. N , x_0 , I — параметры выборки. На множестве выборок из данной генеральной совокупности эти параметры связаны статистически (см. выражение (7.2.22)). Для отдельной выборки их следует рассматривать как независимые. Скажем, I может быть при данных N и x_0 , вообще говоря, любым. Для определения пяти параметров распределения Ципфа (α , γ , A , B и C) имеем четыре уравнения (7.2.10), (7.2.12) и (7.2.14). Еще одно уравнение дают данные наблюдения. Они позволяют, например, графическим путем определить α (см. выражения (7.2.26)). Тогда получаем:

$$\gamma = 1/\alpha; \quad C = \frac{\alpha(N-1)}{1/x_0^{\alpha} - 1/I^{\alpha}}; \\ A = \left(\frac{N-1}{1/x_0^{\alpha} - 1/I^{\alpha}} \right)^{1/\alpha}; \quad B = \frac{N-1}{(I/x_0)^{\alpha} - 1} - 1. \quad (7.2.15)$$

Казалось бы, при получении выражений для A , B и C можно пользоваться вместо выражений (7.2.14) выражениями (7.2.1) и (7.2.3), заменив суммирование на интегрирование. Однако это было бы не вполне корректно. Дело в том, что в выражениях

* Прежде писали

$$x(r) \sim 1/r^{\gamma}.$$

(*)

А. И. Яблонский [0.114] дал выражения (7.2.12) и (7.2.13) (последнее — для $\gamma=1$). Ранговым аналогом частотного распределения Ципфа является именно распределение Мандельброта (7.2.12), а не «ранговое распределение Ципфа» (*). К этим выражениям и выводам автор пришел независимо (см. далее об эффекте рангового искажения).

(7.2.1) и (7.2.3) при переходе от суммирования к интегрированию теряется небольшая часть суммы. Например, для выражения (7.2.3) эта часть суммы для $x=I$ достигает единицы:

$$r = \sum_x n(x) \Big|_{x=I} \equiv 1$$

(если $n(I) > 1$, то это последнее слагаемое суммы раскладывается на несколько других, каждое из которых равно единице), а

$$\int_x^I n(x) dx \Big|_{x=I} \equiv 0.$$

Ясно, что минимальный ранг равен 1, а не 0. Используя выражения (7.2.1) и (7.2.3), получаем, таким образом, для A , B и C выражения:

$$A = \left(\frac{N}{1/x_0^{\alpha} - 1/I^{\alpha}} \right)^{1/\alpha}; \quad B = \frac{N}{(I/x_0)^{\alpha} - 1}; \quad C = \frac{\alpha N}{1/x_0^{\alpha} - 1/I^{\alpha}}$$

вместо правильных (7.2.15). Этого обстоятельства не учитывает, например, А. И. Яблонский [0.114], производящий указанную выше замену суммирования на интегрирование и получающий вследствие этого $r \geq 0$ вместо верного $r \geq 1$. Эта неточность особенно сказывается в социальных науках (в том числе и в наукометрии), где данные часто приходится откладывать в логарифмических координатах, в которых $\log 0 = -\infty$, а $\log 1 = 0$.

Подчеркнем: I следует рассматривать как параметр выборки, независимый от N , что иногда упускается из виду. Далее будет говориться о том, что I связано с N вероятностным образом, поэтому данному объему выборки N при фиксированной форме распределения может соответствовать, вообще говоря, любое значение I .

В общем случае в логарифмических координатах ранговое распределение Ципфа в отличие от частотного при $r \leq B$ отклоняется от прямой линии (см. выражение (7.2.12)). Этот эффект может быть назван *эффектом рангового искажения*. Он существует при $B \geq 1$, т. е. при

$$I/x_0 \leq [(N+1)/2]^{1/\alpha}, \quad (7.2.16)$$

и не существует при $B \ll 1$; т. е. при

$$I/x_0 \gg [(N+1)/2]^{1/\alpha} \quad (7.2.17)$$

(см. выражение (7.2.15) и рис. 2.5).

Эффект рангового искажения имеет место не только для рас-

предела Ципфа, но и в более общем случае ципфоваго распределения (см. рис. 2.1, 2.3), т. е. распределения, имеющего при больших значениях x (малых значениях r) вид распределения Ципфа. Этот эффект не зависит от поведения распределения в области малых значений x . Условия (7.2.16), (7.2.17) сохраняют силу и в этом общем случае, только здесь следует вместо N подставлять максимальный ранг, при котором данное ципфовое распределение еще имеет вид прямой линии, а вместо x_0 — значение x , соответствующее этому значению ранга.

В социальных распределениях часто $I \gg x_0 > 0$, а величина x_0 определяется минимально измеримым значением x . Здесь x_0 может быть отождествлено с ошибкой измерения случайной величины x (классовым интервалом) Δx в условиях (7.2.7) и (7.2.9), а $I - x_0 \approx I$. Для распределения Ципфа в этом случае условие (7.2.17) несущественности рангового искажения подобно условию (7.2.9) применимости ранговой формы статистического распределения. Поэтому в социальных явлениях, как правило, эффект рангового искажения проявляется слабо.

Выборочные моменты порядка n для любого статистического распределения в общем случае зависят от I и x_0 :

$$M^{(n)} = \frac{\sum_{x_0}^I (x - \bar{x})^n n(x)}{\sum_{x_0}^I n(x)} \quad (7.2.18)$$

Однако для гауссовых распределений эта зависимость несущественна, поскольку с ростом I (при фиксированном x_0) моменты быстро устремляются к своим конечным пределам. Например, для «правого» распределения Гаусса:

$$f(x) = \frac{2}{\sigma \sqrt{2\pi}} e^{-x^2/2\sigma^2};$$

$$x = \frac{\sigma \sqrt{\frac{2}{\pi}} [1 - e^{-(I/\sigma \sqrt{2})^2}]}{2\Phi(I/\sigma)} \quad (7.2.19)$$

Для логнормального распределения:

$$f(x) = \frac{1}{x\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma}\right)^2}, \quad x \geq 0;$$

$$\bar{x}^n = \frac{e^{\frac{1}{2} n^2 \sigma^2 + n\mu} \left[0,5 + \operatorname{sign} \left(\frac{\ln I - \mu}{\sigma} - n\sigma \right) \Phi \left(\frac{\ln I - \mu}{\sigma} - n\sigma \right) \right]}{0,5 + \operatorname{sign} \left(\frac{\ln I - \mu}{\sigma} \right) \Phi \left(\frac{\ln I - \mu}{\sigma} \right)} \quad (7.2.20)$$

(В этих выражениях $\Phi(u)$ — интеграл вероятности, $0 \leq \Phi(u) \leq 0,5$). Эти величины, действительно, с ростом I сходятся к конечным пределам (см., например, рис. 2.7).

С распределением Ципфа ситуация другая. Его моменты:

$$M^{(n)} = \begin{cases} \frac{\alpha \ln(I/x_0)}{1/x_0^\alpha - 1/I^\alpha}, & \alpha = n \\ \frac{\alpha}{\alpha - n} \frac{1/x_0^{\alpha-n} - 1/I^{\alpha-n}}{1/x_0^\alpha - 1/I^\alpha}, & \alpha \neq n. \end{cases} \quad (7.2.21)$$

При расчете выражений (7.2.19) — (7.2.21) и (7.2.24) (см. далее) суммирование заменено интегрированием, причем предполагается, что выборочные распределения совпадают с распределением генеральной совокупности.

Для бесконечной совокупности моменты порядка n распределения Ципфа, как видим, бесконечны при $\alpha \leq n$. Для конечной выборки эти моменты растут с увеличением I , и эта зависимость тем существеннее, чем меньше I и чем меньше α (см., например, рис. 2.6).

С ростом объема выборки I и моменты распределения, вообще говоря, растут. Однако зависимость моментов от N носит вероятностный характер. Размах $w = I - x_0$ выборки объема N имеет плотность распределения

$$f_w(w) = N(N-1) \int_{-\infty}^{\infty} [F(x+w) - F(x)]^{N-2} f(x) f(x+w) dx \quad (7.2.22)$$

[см. 0.37, с. 619]. Соотношения (7.2.21) и (7.2.22) позволяют в принципе найти вероятностную зависимость моментов распределения Ципфа от N . Однако здесь возникают определенные математические трудности. Например, для $\alpha = 1$ имеем:

$$f_w(w) = \frac{N(N-1)x_0 w^{N-2}}{2N-1} F(N, 2N-1; 2N; -w/x_0), \quad (7.2.23)$$

где $F(\alpha, \beta; \gamma; z)$ — гипергеометрическая функция. Таким образом, искомая зависимость моментов распределения Ципфа от объема выборки может быть найдена лишь путем численного интегрирования.

Энтропия распределения Ципфа в отличие от его моментов

с ростом I быстро сходится к конечному пределу:

$$S = - \sum_{x_0} \frac{n(x)}{N} \log_2 \frac{n(x)}{N} = \frac{1}{\ln 2} \left[\ln \frac{1}{\alpha} + \frac{1}{\alpha} + 1 + \ln \left(\frac{1}{x_0^\alpha} - \frac{1}{I^\alpha} \right) + \right. \\ \left. + \frac{1+\alpha}{\alpha} \frac{\ln(x_0^\alpha/x_0^\alpha) - \ln(I^\alpha) \cdot I^\alpha}{1/x_0^\alpha - 1/I^\alpha} \right]; \quad (7.2.24)$$

$$\lim_{N \rightarrow \infty} S = \frac{1}{\ln 2} \left(\ln \frac{1}{\alpha} + \frac{1}{\alpha} + 1 + \ln x_0 \right).$$

Иногда в качестве характеристики распределения Ципфа используют энтропию его ранговой формы:

$$S_r = - \sum_1^N \frac{x(r)}{G} \log_2 \frac{x(r)}{G} = \\ = \begin{cases} (1/\ln 2) \{ \ln [1/(\gamma - 1)] + 1/(\gamma - 1) + 1 + \ln [1/(1+B)^{\gamma-1} - \\ - 1/(N+B)^{\gamma-1}] + \frac{\gamma}{\gamma-1} \frac{\frac{\ln(1+B)^{\gamma-1}}{(1+B)^{\gamma-1}} - \frac{\ln(N+B)^{\gamma-1}}{(N+B)^{\gamma-1}}}{1/(1+B)^{\gamma-1} - 1/(N+B)^{\gamma-1}} \}, \gamma \neq 1 \\ \frac{1}{\ln 2} \left[\ln \frac{N+B}{1+B} + \frac{1}{2} \ln ((1+B)(N+B)) \right], \gamma = 1 \end{cases} \quad (7.2.25)$$

$$\lim_{N \rightarrow \infty} S_r = \\ = \begin{cases} \frac{1}{\ln 2} \{ \ln [1/(\gamma - 1)] + 1/(\gamma - 1) + 1 + \ln(1+B) \}, & \gamma > 1 \ (\alpha < 1) \\ \infty, & \gamma \leq 1 \ (\alpha \geq 1) \end{cases}$$

Эта величина, как видим, при $N \rightarrow \infty$ стремится к бесконечности для $\alpha \geq 1$. Поэтому ее использование корректно лишь для $\alpha < 1$.

Если для данного распределения первые два момента конечны, то оно сходится к распределению Гаусса и является гауссовым. Иногда моменты заданного аналитически распределения вычислить трудно, а знать, с каким распределением мы имеем дело, гауссовым или негауссовым, необходимо. В этом случае полезно определить асимптотику распределения при $x \rightarrow \infty$. Для распределения Ципфа:

$$\frac{d \ln n(x)}{d \ln x} = -(1 + \alpha); \quad \frac{d \ln x(r)}{d \ln r} = -\gamma \frac{1}{1+B/r}. \quad (7.2.26)$$

Для ципфического распределения:

$$\lim_{x \rightarrow \infty} \frac{d \ln n(x)}{d \ln x} = -(1 + \alpha);$$

$$\lim_{\substack{x \rightarrow \infty \\ (N \rightarrow \infty)}} \frac{d \ln x(r)}{d \ln r} \Big|_{r=1} = -\gamma \frac{1}{1+B} \quad (7.2.27)$$

(см. рис. 2.1, 2.3). При $\alpha > 2$ ципфическое распределение гауссово. Для гауссового неципфического распределения:

$$\lim_{x \rightarrow \infty} \frac{d \ln n(x)}{d \ln x} = -\infty \quad (7.2.28)$$

(см. рис. 2.2).

Выражения (7.2.26) — (7.2.28) позволяют различать гауссовые и негауссовые распределения не только в том случае, если они даны аналитически, но и в том, когда они даны графически. Особенно легко это делать, когда распределения даны в частотной форме. Для ципфического распределения угол наклона асимптоты при $x \rightarrow \infty$ к оси x острый и его величина согласно выражению (7.2.27) прямо определяет α . Для гауссового распределения этот угол для бесконечной выборки равен прямому. Для гауссового же распределения, построенного на конечной выборке, этот угол не прямой, однако расчетное значение α оказывается равным, хоть и не бесконечной, но все же большой величине.

Для ранговой формы гауссового распределения асимптотика при $x \rightarrow \infty$ имеет менее ясную картину, чем для частотной (см. рис. 2.2 и 2.4). Однако и здесь, если пренебречь областью малых рангов, гауссовые распределения можно отличать от негауссовых по большему углу наклона асимптот к оси x .

Переход от гауссовых распределений к негауссовым на конечных выборках, разумеется, непрерывен, тем более что он непрерывен и для теоретических распределений.

3. РАЗЛИЧНЫЕ АППРОКСИМАЦИИ

Здесь дается краткий обзор аппроксимаций, используемых для описания стационарных наукометрических распределений. Многие из приведенных аппроксимаций названы в работах [6.21], [3.62], [3.87]. Полезна также библиография [1.158]. Разные авторы часто используют разные обозначения и разные формы статистических распределений. Мы используем унифицированную систему обозначений, которая была дана в предыдущих разделах.

1. Распределение Ципфа (7.2.10) используют многие авторы. Первоначально α берется фиксированным, чаще всего $\alpha = 1$.
1.1. Ж.-Б. Эсту [6.18], Г. Ципф [6.23; 6.24] и другие описывают распределением

$$x(r) \sim 1/r \quad (7.2.25)$$

частотные распределения слов. Это — ранговое распределение Ципфа при $\alpha = 1$, когда эффект рангового искажения пренебрежимо мал (см. выражение (7.2.12) с $B = 0$). Г. Ципф [1.160], Г. Скарротт [0.104] и другие расширяют применения этой аппроксимации.

1.2. А. Лотка [3.74] берет $\alpha = 1$ для описания распределения ученых по числу публикаций (см. выражение (3.3.13)).

1.3. С. Брэдфорд [3.42; 3.43] использует аппроксимацию (3.3.1) для описания распределения журналов по числу статей данного профиля. Эта аппроксимация соответствует распределению Ципфа с $\alpha = 1$, когда эффект рангового искажения пренебрежимо мал (см. выражение (7.2.13) с $B = 0$).

1.4. П. Коул [3.50] использует аппроксимацию (3.3.2) для описания распределения журналов по числу статей данного профиля. Эта аппроксимация также соответствует распределению Ципфа с $\alpha = 1$, когда эффект рангового искажения мал.

1.5. Ф. Лаймкюлер [3.72; 3.73] прибегает к аппроксимации (3.3.3) для описания распределения журналов по числу статей данного профиля. Эта аппроксимация также соответствует распределению Ципфа с $\alpha = 1$, но принимает во внимание эффект рангового искажения (см. выражение (7.2.13) с произвольным B).

1.6. С. Наранан [1.131] и А. Букштайн [1.77–1.79] используют аппроксимацию (3.3.8) для описания частотных распределений слов, распределений ученых по числу публикаций, журналов по числу статей данного профиля, цитатных распределений. Эта аппроксимация соответствует распределению Ципфа с произвольным $\alpha = s - 1$.

1.7. Дж. Хьюберт [3.62; 1.117] предложил аппроксимацию (3.3.7) для описания распределения журналов по числу работ данного профиля и ученых по числу публикаций. Это — ранговое распределение Ципфа с произвольным α , когда эффект рангового искажения пренебрежимо мал (см. выражение (7.2.12) с $B = 0$).

1.8. Б. Мандельброт [0.94] предложил аппроксимацию

$$x(r) = A/(r+B)^1 \quad (7.2.30)$$

для описания частотных распределений слов. Закон Мандельброта — точный ранговый аналог частотного распределения Ципфа (см. выражение (7.2.12) с произвольным B).

1.9. М. Кендалл [1.120], Р. Фэйтхорн [1.98], А. И. Яблонский [0.63] описывают распределением Ципфа в его общей форме (7.2.10), распределения слов по частоте встречаемости, ученых по числу публикаций, журналов по числу статей данного профиля и др.

История распределения Ципфа сложна и вокруг нее ведутся приоритетные споры. В этой истории определенную роль сыграли исследования Ф. Хайта [1.109], Л. С. Козачкова и Л. А. Хурсина [1.30–1.31], М. Бакленда и А. Хиндла [1.83] и др. Отдельную главу представляет проникновение распределения Ципфа в цитатные распределения. Здесь могут быть названы А. Платц [1.138], Д. Прайс [4.125], С. Наранан [1.131], Р. Фэйтхорн [1.98], Г. Магяр [1.123], Дж. Букштайн [1.78; 1.79], Д. Дикс и Х. Чапг [4.69], А. И. Яблонский [0.63], С. Д. Хайтуп [4.32] и др.

2. Б. Брукс [3.44–3.46] использует аппроксимацию (3.3.4) для описания распределения журналов по числу статей данного профиля. Эта аппроксимация представляет собой сочетание двух распределений Ципфа с разными значениями α , когда эффект рангового искажения пренебрежимо мал (см. выражение (7.2.13) с $B = 0$).

3. А. Т. Мицевич [3.23; 3.24] предложил аппроксимацию (3.3.5) для описания распределения журналов по числу статей данного профиля. Эта аппроксимация является ранговым ципфовым распределением с $\alpha = 1$, когда эффект рангового искажения мал.

4. К. Босанквет [6.17] использует аппроксимацию

$$f(x) = ab/(ax + bx^2) \quad (7.2.31)$$

для описания частотного распределения слов. При больших значениях x эта аппроксимация имеет вид распределения Ципфа:

$$f(x) \sim 1/x^2, \quad \alpha = 1. \quad (7.2.32)$$

5. Д. Прайс [1.58] предложил аппроксимацию (3.3.19) для описания распределения ученых по числу публикаций. При больших значениях x эта аппроксимация имеет вид распределения Ципфа с $\alpha = 2$.

6. Бета-функция (см. выражение (3.3.21)) была первоначально предложена Г. Юлом [0.115] для описания распределений в математической генетике. Г. Саймон [0.106] использовал ее для распределений слов по частоте встречаемости, Д. Прайс [1.143] — для описания распределений слов, ученых по числу публикаций, журналов по числу статей данного профиля и цитатных распределений. При больших x эта аппроксимация является

распределением Ципфа:

$$n(x) \sim 1/x^{m+2}, \quad \alpha = m+1. \quad (7.2.33)$$

6.1. М. Кэндэлл [1.119] аппроксимирует распределение Брэдфорда выражением (3.3.10), которое представляет собой бета-функцию при $m=0$.

7. Отрицательное биномиальное распределение (3.3.26) используют К. Кроули [4.66] для описания цитатных распределений, а К. Равичандра Рао [3.87] — для описания распределения ученых по числу публикаций. Эта аппроксимация имеет конечные моменты:

$$\bar{x} = \frac{k\theta}{(1-\theta)[1-(1-\theta)^k]};$$

$$\bar{x}^2 = \frac{k\theta^2(1+k)}{(1-\theta)^2[1-(1-\theta)^k]} + \frac{k\theta}{(1-\theta)[1-(1-\theta)^k]} \quad (7.2.34)$$

и потому является гауссовой.

7.1. Е. Вильямс [3.105] использует геометрическое распределение

$$f(x) = [(1-\theta)/\theta]^x; \quad x=1, 2, 3, \dots; \quad 0 < \theta < 1; \quad (7.2.35)$$

$$\bar{x} = 1/(1-\theta); \quad \bar{x}^2 = (1+\theta)/(1-\theta)^2$$

для описания распределения Лотки. Эта аппроксимация является отрицательным биномиальным распределением с $k=1$.

8. Е. Вильямс [3.105] использует логарифмическое распределение

$$f(x) = \frac{1}{\ln(1-\theta)} \frac{\theta^x}{x}; \quad x=1, 2, 3, \dots; \quad 0 < \theta < 1;$$

$$\bar{x} = \frac{\theta}{(1-\theta) \ln(1-\theta)}; \quad \bar{x}^2 = \frac{\theta}{(1-\theta)^2 \ln(1-\theta)} \quad (7.2.36)$$

для описания распределения Лотки. Это — гауссовое распределение.

9. У. Шокли [1.156], Д. Пельц и Ф. Эндрюс [0.51], Д. Прайс и Д. Бивер [2.30], Р. Ф. Васильев [1.6], А. А. Игнатьев и А. И. Яблонский [0.31] и другие используют логнормальное распределение (7.2.20) для описания распределения Лотки. Это — гауссовое распределение.

10. Распределение Пуассона (3.3.35) используют Л. Мантелл [3.75] для описания распределения Лотки, М. П. Великова и З. Г. Залесова [1.3] — для аппроксимации цитатных распределе-

ний. Моменты распределения Пуассона конечны:

$$\bar{x} = \bar{x}^2 = \lambda, \quad (7.2.37)$$

поэтому оно гауссово.

11. С. Зихель [6.21] реализовал идею Г. Юла [6.22] и Г. Херда-на [6.19] об аппроксимации частотных распределений слов обобщенным распределением Пуассона:

$$f(x) = \frac{[(1-\theta)^{1/2}]^x}{K_v[\alpha(1-\theta)^{1/2}]} \frac{(\alpha\theta/2)^x}{x!} K_{x+v}(\alpha);$$

$$-\infty < v < \infty; \quad 0 < \theta < 1; \quad \alpha > 0, \quad (7.2.38)$$

где $K_v(z)$ — модифицированная функция Бесселя второго рода (называемая иногда модифицированной функцией Бесселя третьего рода, или функцией Бесселя от мнимого аргумента третьего рода). Поведение этого распределения при $x \rightarrow \infty$ зависит от значений параметра θ :

$$f(x) \sim \frac{\Gamma(x+v)}{\Gamma(x+1)} \theta^x, \quad (7.2.39)$$

где $\Gamma(z)$ — гамма-функция и

$$\lim_{x \rightarrow \infty} \frac{d \ln f(x)}{d \ln x} = \begin{cases} v-1, & \theta=1 \\ -\infty, & \theta < 1. \end{cases} \quad (7.2.40)$$

Сравнивая (7.2.40) с выражениями (7.2.27), (7.2.28), видим, что при $\theta < 1$ эта аппроксимация является гауссовой нецифовой, а при $\theta=1$ — цифровой с $\alpha=-v$. С. Зихель [6.21] утверждает, что ряд распределений, таких, как Пуассона, отрицательное биномиальное, геометрическое, логарифмическое, бета-функция, Гуда (см. далее выражения (7.2.48) и (7.2.50)), Ципфа, являются различными частными или предельными формами обобщенного распределения Пуассона. Действительно, что касается распределения Пуассона (3.3.35), то оно является частным случаем выражения (7.2.38). Что же касается других названных распределений, то они возникают как частные случаи асимптотики (7.2.39). Например, при $\theta < 1$ и $v > 0$ выражение (7.2.39) дает отрицательное биномиальное распределение (3.3.26) с $k=v$, при $\theta=1$ и $v < 0$ — бета-функцию (3.3.21) с $m=-(v+1)$, при $\theta=1$ и $v < 0$ — распределение Ципфа (7.2.10) с $\alpha=-v$ и т. д.

Итак, ситуация с аппроксимациями стационарных наукометрических распределений сложилась довольно сложная, тем более что используемые здесь критерии статистических гипотез (на-

пример, известный критерий χ^2), к сожалению, не дают однозначных результатов. Дискуссии продолжаются [см. например, 6.24; 3.49; 3.84; 3.86; 3.87].

Мы убеждены, однако, что в эту проблему может быть внесена определенная ясность. Ключ к ее решению дают Б. Мандельброт и А. И. Яблонский. Развиваемый ими подход, в котором распределение Ципфа связывается с теорией негауссовых устойчивых распределений, позволяет придать наблюдению о том, что наукометрические стационарные распределения имеют необычайно длинные «хвосты», количественную определенность. Этот подход рассматривается в следующем разделе.

4. РАСПРЕДЕЛЕНИЕ ЦИПФА И НЕГАУССОВЫЕ УСТОЙЧИВЫЕ РАСПРЕДЕЛЕНИЯ

Распределение Ципфа с позиции теории негауссовых устойчивых распределений рассматривают Б. Мандельброт [0.97; 0.98], Н. Хомский и Дж. Миллер [2.35] и А. И. Яблонский [0.63, 0.114]. Они исходят из того факта, что для распределения Ципфа существуют моменты лишь порядка $n < \alpha$.

В основании современных приложений математической статистики и теории вероятностей лежат предельные теоремы о сходимости распределений сумм одинаково распределенных случайных независимых величин к определенным конечным так называемым устойчивым распределениям.

Устойчивыми называются распределения, свертка которых с такими же распределениями приводит к распределению того же вида, т. е. если для любых $a_1 > 0$, b_1 , $a_2 > 0$ и b_2 найдутся такие $a > 0$ и b , что при всех x

$$F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b). \quad (7.2.41)$$

Свертка двух распределений $f_1(\xi_1)$ и $f_2(\xi_2)$

$$\int f_1(y - x) f_2(x) dx \quad (7.2.42)$$

является распределением суммы $\eta = \xi_1 + \xi_2$ случайных величин ξ_1 и ξ_2 , описываемых свертываемыми распределениями.

Распределения, сходящиеся к устойчивым, разделим на два класса — гауссовы и негауссовы. Гауссовыми, следуя А. И. Яблонскому [0.63], назовем распределения, сходящиеся в указанном смысле к распределению Гаусса, негауссовыми — распределения, сходящиеся к другим устойчивым распределениям. Распределение Гаусса — гауссовое, но гауссовое распределение не обязательно

является распределением Гаусса. Гауссовыми являются, например, распределения 7—10 предыдущего раздела.

Чтобы математическая статистика могла быть использована в приложениях, необходимо, чтобы были конечными используемые в ней характеристики распределений. Гауссовы распределения согласно центральной предельной теореме имеют конечные моменты, точнее — конечные первые два момента. Гауссовая статистика — это статистика моментов.

Негауссовы распределения имеют, вообще говоря, бесконечные моменты. Б. Мандельброт [0.97; 0.98], по-видимому, первым указал на сходимость распределения Ципфа при $\alpha < 2$ к устойчивым распределениям, отличным от нормального, говоря конкретно о его принадлежности к семейству устойчивых распределений Леви. Н. Хомский и Дж. Миллер [2.35] также говорят о принадлежности распределения Ципфа к числу распределений, имеющих бесконечные моменты, но тем не менее сходящихся, как они говорят, к стационарным предельным распределениям.

Наиболее полную на сегодня разработку вопросов, связанных с негауссовостью распределения Ципфа, находим у А. И. Яблонского [0.63; 0.114]. Во-первых, Яблонский указывает на непригодность традиционного аппарата гауссовой математической статистики при описании статистических систем, подчиняющихся распределению Ципфа. «Поэтому, — говорит он, — особое значение приобретает оценка не средних значений, а таких параметров, как квантили и особенно характеристический показатель α ... При этом, оценка этих параметров должна производиться не методом наименьших квадратов, а более общими методами, например, методом максимума правдоподобия» [0.63, с. 37].

Теоретический вывод о необходимости привлечения аппарата негауссовой статистики подтверждается эмпирическими исследованиями, в которых авторы приходят к выводу о недостаточности гауссовой статистики при обработке эмпирических данных, описываемых распределением Ципфа [см., например, 6.1; 0.9; 0.93].

Во-вторых, Яблонский обращает внимание на связь распределения Ципфа с теоремой негауссовой статистики, которую он приводит со ссылкой на курс теории вероятностей [0.52] в следующей редакции: «Сходимость нормированной суммы случайных величин к устойчивым распределениям, отличным от нормального, имеет место, в частности, если распределение этих величин $F(x)$ при $x \rightarrow \infty$ удовлетворяет следующим условиям:

$$\begin{aligned} F(-x) &\sim C_1/|x|^\alpha; & 1-F(x) &\sim C_2/x^\alpha; \\ C_1 &\geq 0; & C_2 &\geq 0; & C_1+C_2 &> 0. \end{aligned} \quad (7.2.43)$$

Сам закон Ципфа совпадает по форме с асимптотикой *устойчивых негауссовых* (курсив наш.—С. Х.) распределений. Поэтому закон Ципфа является не одним из многих эмпирических распределений, случайно подогнанных с тем или иным успехом под результаты измерений, а теоретическим законом, несколько специфическим, но имеющим надежную математическую базу в виде теории устойчивых негауссовых распределений» [0.63, с. 31].

Можно даже предположить, пишет Яблонский далее, что распределение Ципфа «действительно играет ту же универсальную роль предельного распределения в соответствующих областях (ассоциируемых, например, со сложными информационными системами) как закон Гаусса в стохастических проблемах с конечной дисперсией (ассоциируемых, как правило, с неорганическими процессами). Это также подтверждается удивительной распространенностью закона Лотки (так же, как и его рангового аналога — закона Ципфа) в самых разнообразных областях, особенно относящихся к человеческому поведению (здесь Яблонский ссылается на известную работу М. Кендэлла [1.120].—С. Х.), что оправдывает интенсивное изучение взаимодействия этих закономерностей и устойчивых негауссовых распределений» [0.63; 0.114].

Установление Мандельбротом и особенно Яблонским связи распределения Ципфа с негауссовыми устойчивыми распределениями представляется весьма ценным. Однако для Яблонского распределение Ципфа при всей его «универсальности» — лишь одна из возможных аппроксимаций стационарных наукометрических распределений. Например, в работе [0.31], опубликованной совместно с А. А. Игнатьевым, он аппроксимирует распределение ученых по частичной продуктивности логнормальным распределением. «Таким образом,— говорят авторы,— данные о продуктивности ученых позволяют построить две различные эмпирические кривые, которые могут быть аппроксимированы разными вероятностными распределениями. Именно распределение публикаций по именам (данные Лотки и др.) подчиняются закону Лотки, а распределение индивидуальных вкладов (данные Шокли, а также Пельца и Эндриуса) — логнормальному закону» [0.31, с. 67].

Эту точку зрения Яблонский развивает и в работе [0.65]: «В зависимости от выбора эмпирической картины,— пишет он,— мы приходим к различным распределениям. Оценка научной продуктивности по информационному массиву приводит к закону Лотки (информационная модель науки), а анализ продуктивности на массиве ученых (социальная модель науки) приводит к логнормальному распределению. Анализ специфики этих распреде-

лений позволяет предположить количественно обоснованное наличие различных структур (проекций) науки: «информационной» и «социальной», определяемых разными закономерностями (менее жесткой для «социальной» проекции науки и более жесткой для «информационной»)» [0.65; с. 86]. В последующих публикациях, например в [0.114], Яблонский не отказывается от этих взглядов.

Выводы Яблонского о роли распределения Ципфа опираются на соответствующую теорему негауссовой статистики. Однако он приводит редакцию этой теоремы, являющуюся лишь частной редакцией более общей теоремы Гнеденко-Деблина. В следующем разделе приводится более полная редакция этой теоремы, которая позволяет сделать в отношении распределения Ципфа выводы, более сильные, чем выводы Яблонского.

5. РАСПРЕДЕЛЕНИЕ ЦИПФА И НЕГАУССОВЫЕ РАСПРЕДЕЛЕНИЯ

Теорема негауссовой статистики, фигурирующая в теории распределения Ципфа, была сформулирована независимо Б. В. Гнеденко [0.18] и В. Деблиным [0.78]. Согласно редакции теоремы, данной И. А. Ибрагимовым и Ю. В. Линником [0.28] и Ю. В. Прохоровым и Ю. А. Розановым [0.52], для сходимости распределений нормированных сум одинаково распределенных независимых случайных величин (т. е. распределений выборочных средних) к устойчивым распределениям, отличным от нормального, необходимо и достаточно, чтобы при $x \rightarrow \infty$ имело место:

$$F(-x) \sim C_1 \frac{h_1(x)}{|x|^\alpha}; \quad 1 - F(x) \sim C_2 \frac{h_2(x)}{x^\alpha} \quad (7.2.44)$$

$$C_1 \geq 0; \quad C_2 \geq 0; \quad C_1 + C_2 > 0; \quad 0 < \alpha < 2.$$

Здесь $h_i(x)$ — функции, медленно меняющиеся в смысле Караматы, т. е. такие, что для всех $t > 0$ имеет место:

$$\lim_{x \rightarrow \infty} \frac{h_i(tx)}{h_i(x)} = 1. \quad (7.2.45)$$

Сравнивая распределение Ципфа (см. выражение (7.2.11)) с соотношением (7.2.44), находим, что с точностью до медленно меняющихся функций $h_i(x)$ распределение Ципфа совпадает с асимптотикой негауссовых распределений.

Яблонский, как говорилось в предыдущем разделе, пришел к аналогичному выводу. Здесь несущественно, что он говорил

в 1976 г. [0.63] о совпадении распределения Ципфа с асимптотической *негауссовых устойчивых* распределений, тем более что в 1980 г. [0.114] Яблонский говорит уже о совпадении распределения Ципфа с асимптотикой *негауссовых* распределений, убирая слово «устойчивых». Существенно другое: в приводимой Яблонским частной формулировке выпали функции $h_i(x)$. Отсюда с необходимостью следует занижение им роли распределения Ципфа. Ведь если брать частную формулировку теоремы Гнеденко-Деблина, которую приводит Яблонский (см. выражение (7.2.43) и связанный с ним текст), то из нее следует лишь, что негауссовость данного распределения является *необходимым, но не достаточным* условием того, чтобы асимптотика этого распределения совпадала с распределением Ципфа. Следовательно, возможны негауссовы распределения, не совпадающие в своей асимптотике с распределением Ципфа. Именно в этом корень допущения Яблонским других распределений, например логнормального, в качестве аппроксимаций стационарных наукометрических распределений.

Полная формулировка теоремы Гнеденко-Деблина говорит, что эти «другие» распределения, сходящиеся к отличному от нормального, могут отличаться в их асимптотике от распределения Ципфа лишь медленно меняющимися функциями $h_i(x)$. Таким образом, с точностью до этих функций $h_i(x)$ множество *негауссовых распределений исчерпывается распределениями, имеющими в асимптотике больших значений переменной вид распределения Ципфа*.

Определение: распределения, имеющие при $x \rightarrow \infty$ вид распределения Ципфа, для $\alpha < \infty$ назовем *ципфовыми*. Как и распределение Ципфа, ципфовое распределение гауссово при $\alpha > 2$ и негауссово при $\alpha < 2$.

Таким образом, общая формулировка теоремы Гнеденко-Деблина позволяет сделать заключение, что негауссовы распределения с точностью до медленно меняющихся функций могут быть только ципфовыми.

Яблонский говорит (см. предыдущий раздел), что распределение Ципфа играет «ту же универсальную роль предельного распределения» в соответствующих областях, что и нормальное распределение в задачах с конечной дисперсией. Это не совсем так. Универсальность распределения Ципфа отличается от универсальности распределения Гаусса. Если гауссовы распределения сходятся (в указанном в предыдущем разделе смысле) к распределению Гаусса, то негауссовы имеют распределение Ципфа своей асимптотикой при больших значениях переменной.

Гауссовы распределения имеют конечные моменты, негаус-

совы — бесконечные. Но это — на бесконечной совокупности. Для выборки конечного объема и гауссовы и негауссовы распределения имеют моменты, зависящие, вообще говоря, от объема выборки (см., например, рис. 2.6, 2.7). Чем же руководствоваться при выборе аппроксимаций?

Определение. *Гауссовой совокупностью* назовем совокупность, на которой моменты (точнее — первые два момента) *существенно* зависят от объема выборки. *Негауссовой совокупностью* назовем совокупность, на которой моменты зависят от объема выборки *несущественно*.

Существенно или несущественно — определяется в конечном счете задачами, стоящими перед исследователем данной совокупности, используемыми им методами, точностью проводимых им измерений. «Абсолютных» критериев здесь не существует.

Гауссовы совокупности следует аппроксимировать гауссовыми распределениями, негауссовы — негауссовыми. *Аппроксимация негауссовых совокупностей гауссовыми распределениями некорректна* по той причине, что на негауссовых совокупностях гауссовы распределения (согласно центральной предельной теореме) с ростом объема выборки не сходятся к распределению Гаусса. Не сходятся в пределах точности, определяемых данной конкретной исследовательской задачей.

На негауссовых совокупностях некорректна не только аппроксимация гауссовыми распределениями, но и вся гауссовая статистика, частью которой являются гауссовы распределения. Моменты здесь дают результаты, зависящие от объема выборки и потому трудно воспроизводимые. Чтобы получить воспроизводимые результаты, следует отказаться в этом случае от моментов и использовать другие характеристики статистических распределений, которые слабо зависят для негауссовой совокупности от объема выборки. Вместе с моментами в этом случае следует отказаться и от их производных, таких, как, например, коэффициенты корреляции.

В следующем разделе на большом эмпирическом материале показывается, что стационарные наукометрические распределения имеют моменты, существенно зависящие от объема выборки. Другими словами, в наукометрии мы имеем дело с негауссовыми стационарными совокупностями.

6. НЕГАУССОВЫЕ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ

Чтобы сделать вывод о гауссовости или негауссовости генеральной совокупности стационарных наукометрических распределений, мы возьмем максимально большую выборку таких рас-

пределений, при образовании которых отсутствовала бы какая-либо направленность. Для оценки степени зависимости выборочных моментов отдельных распределений от объема выборки мы будем аппроксимировать данные наблюдения ципфовыми распределениями (гауссовы и ципфовые распределения имеют $\alpha > 2$, гауссовы неципфовые — $\alpha = \infty$ и негауссовы — $\alpha < 2$) и графическим путем определять α . По величине α , используя рис. 2.6 для распределения Ципфа, мы и будем приближенно судить о зависимости моментов от объема выборки.

Наша выборка стационарных наукометрических распределений насчитывает 53 распределения. 42 из них показаны на рис. 2.8–2.46, еще 11 распределений были построены на основании данных, содержащихся в работах [1.119; 4.3; 3.88; 3.101; 0.45; 2.39]. В этой выборке распределений представлены все разделы современной наукометрии. Все распределения нанесены на графики в логарифмических координатах, к эмпирическим кривым проведены асимптоты при больших значениях x (малых значениях r). Значения α определялись графически по углу наклона асимптот к оси x (см. выражение (7.2.26), (7.2.27)).

На рис. 2.76 дается распределение значений α для нашей выборки, которое, в свою очередь, может быть аппроксимировано ципфовым распределением с небольшим значением α : $\alpha \approx 1,09$. Это приводит к тому, что основная часть нашей выборки распределений имеет небольшие значения α : в 49,1% случаев $\alpha \leq 1$, в 73,6% — $\alpha \leq 2$ и лишь в 1,9% — $\alpha > 10$.

Было бы неверным утверждать, что 73,6% распределений нашей выборки негауссовы, а 26,4% — гауссовы. Критерием негауссовости является, как говорилось, зависимость моментов от объема выборки. Для ципфовых распределений эта зависимость тем существеннее, чем меньше объем выборки и чем меньше α (см. рис. 2.6). Переход от гауссовых распределений к негауссовым совершается непрерывно. Поэтому для большей части распределений нашей выборки, значительно превышающей 73,6%, моменты существенно зависят от объема выборки. Следовательно, подавляющая часть нашей выборки распределений негауссова.

Таким образом, в той мере, в какой наша выборка в 53 распределения репрезентативна относительно генеральной совокупности стационарных наукометрических распределений, эта совокупность негауссова.

Наукометрия как раздел науковедения, отмечали мы в гл. 1, «отвечает» за воспроизводимое измерение научной деятельности. Поэтому, говорилось в этой главе, наукометрические закономерности в большой степени исчерпывают количественные закономерности научной деятельности. Так что негауссовость генераль-

ной совокупности стационарных наукометрических распределений означала бы негауссовость научной деятельности.

Научная деятельность является лишь одним из видов человеческой деятельности. Трудно представить, чтобы научная деятельность была негауссовой, а другие виды человеческой деятельности — гауссовыми. И наоборот. Поэтому, эмпирический феномен негауссовости стационарных наукометрических распределений может быть должным образом осмыслен лишь при его включении в более широкий материал по стационарным распределениям человеческой деятельности. Такое расширение анализа тем более необходимо и отвечает целям настоящего исследования, что без него мы не в состоянии уяснить особенности измерения при изучении научной деятельности в отличие от измерения при изучении других видов человеческой деятельности. В следующем разделе осуществляется этот краткий экскурс за пределы научной деятельности.

7. НЕГАУССОВОСТЬ СОЦИАЛЬНЫХ ЯВЛЕНИЙ

Ситуация со стационарными социальными распределениями аналогична таковой в наукометрии. Здесь также распределения имеют необычайно длинные «хвосты». И здесь также для их аппроксимации используют самые разные вероятностные распределения. Назовем в дополнение к изложенному в третьем разделе некоторые аппроксимации.

В. Парето [1.135] предложил для описания распределения индивидов по доходам распределение (7.2.10), которое в настоящем тексте называется распределением Ципфа и которое в экономике называется распределением Парето.

Дж. Уиллис [0.113] описывает частотным распределением Ципфа (7.2.10) с $\alpha = 0,5$ таксономическое распределение биологических родов по числу видов. Рис. 2.75, на котором представлены его данные, показывает, что их с учетом эффекта рангового искажения лучше аппроксимировать ципфовым распределением с $\alpha \approx 1,5$.

Т. Дэвис [1.93] использует для описания распределения индивидов по доходам обобщенное распределение Парето:

$$f(x) = a/x^\beta \cdot 1/(e^{b/x} - 1). \quad (7.2.46)$$

При больших x это распределение является распределением Ципфа

$$f(x) \approx (a/b) (1/x^{\beta-1}) \text{ с } \alpha = \beta - 2. \quad (7.2.47)$$

Отрицательное биномиальное распределение (3.3.26) широко используется для описания статистики несчастных случаев (см., например, [0.90]).

И. Гуд [0.80] использует аппроксимацию

$$f(x) = \{0/[0 + (1 - \theta) \ln(1 - \theta)]\} \theta^x / [x(x - 1)];$$

$$x = 1, 2, 3, \dots; \quad 0 < \theta < 1 \quad (7.2.48)$$

для описания распределения видов. Первые два ее момента

$$\bar{x} = \frac{0 + \ln(1 - \theta)}{0 + (1 - \theta) \ln(1 - \theta)}; \quad x^2 = 1/(1 - \theta) \quad (7.2.49)$$

конечны, и потому данная аппроксимация — гауссовая.

И. Гуд [0.80] предлагает аппроксимацию

$$f(x) = (-\ln \theta)^{-1} \Gamma(1 - s, -x_0 \ln \theta) \cdot x^{-s} \theta^x;$$

$$x \geq x_0; \quad 0 < \theta < 1 \quad (7.2.50)$$

($\Gamma(v, m)$ — неполная гамма-функция) для описания распределения видов. Среднее и дисперсия этой аппроксимации

$$\bar{x} = \frac{1}{(-\ln \theta)} \cdot \frac{\Gamma(2 - s, -x_0 \ln \theta)}{\Gamma(1 - s, -x_0 \ln \theta)};$$

$$\bar{x}^2 = \frac{1}{(-\ln \theta)^2} \cdot \frac{\Gamma(3 - s, -x_0 \ln \theta)}{\Gamma(1 - s, -x_0 \ln \theta)} \quad (7.2.51)$$

конечны, и потому она гауссовая.

М. Бальмер [0.71] предлагает логнормальное распределение Пуассона

$$f(x) = \frac{(2\pi V)^{-1/2}}{\pi!} \int \lambda^{x-1} e^{-\lambda} e^{-(\ln x - M)/2V} d\lambda; \quad x = 0, 1, 2, \dots \quad (7.2.52)$$

для описания распределения видов по распространенности. Для этого распределения, как это можно показать,

$$\lim_{x \rightarrow \infty} \frac{d \ln f(x)}{d \ln x} = -\infty, \quad (7.2.53)$$

так что данная аппроксимация — гауссовая нецифровая (см. выражение (7.2.28)).

Логнормальное распределение (7.2.20) особенно часто используется в экономике для описания распределения индивидов по доходам как конкурент распределения Парето [см., например, 0.68].

Наша выборка стационарных ненаукометрических социальных распределений насчитывает 57 распределений. 30 из них показаны на рис. 2.47–2.75, еще 27 были построены на основании данных, которые содержатся в работах [0.45] и др. В этой выборке представлены многие виды человеческой деятельности — экономика, медицина, лингвистика, искусство, игры, демография и др. Отбор распределений не носил направленного характера, за одним исключением: мы не рассматривали распределения, при построении которых использовались *ограниченные сверху* баллы, поскольку при таком способе построения распределения «обрезаются» большие значения переменной x и нет поэтому возможности изучать асимптотику (о таких распределениях см. разд. 9).

На рис. 2.77 дается распределение значений α для нашей выборки в 57 ненаукометрических распределений. Это распределение может быть аппроксимировано цифровым распределением с $\alpha \approx 1,37$. В результате основная часть нашей выборки распределений имеет небольшие значения α : в 33,3% случаев $\alpha \leq 1$, в 54,4 — $\alpha \leq 2$ и лишь в 8,8% — $\alpha > 10$.

На рис. 2.78 показано распределение значений α для всей нашей выборки стационарных социальных распределений, включая 53 наукометрических. Параметры этого распределения имеют значения, промежуточные между значениями параметров распределений, показанных на рис. 2.76 и 2.77.

Таким образом, в той мере, в какой наша выборка в 110 распределений репрезентативна относительно генеральной совокупности стационарных социальных распределений, эта *совокупность* — *негауссовая*.

Эти данные увеличивают достоверность результатов, полученных в предыдущем разделе, говорящих о негауссовости научной деятельности.

Сравнивая рис. 2.76–2.78 видим, что, если наши выборки репрезентативны, то научная деятельность более негауссова, чем другие виды человеческой деятельности и чем деятельность в целом. Это мы связываем с *творческим* характером научной деятельности (см. разд. 11).

8. ОБСУЖДЕНИЕ: ЭФФЕКТ РАНГОВОГО ИСКАЖЕНИЯ

В предыдущих двух разделах значения выборочных α определялись графическим путем по углу наклона асимптоты при больших значениях x в логарифмических координатах. Для частотной формы здесь не встречается особых затруднений. Для ранговой

же формы дело осложняется эффектом рангового искажения (см. второй раздел параграфа).

Посмотрим, как влияет этот эффект при измерении α . Допустим, мы им полностью пренебрегли и опираемся на наклон асимптоты при $r=1$, т. е. пользуемся выражением (7.2.27) с $B=0$. Тогда измеренное и действительное значения α соотносятся следующим образом:

$$\alpha_{\text{изм}}/\alpha \approx 1+B, \quad (7.2.54)$$

где B определяется выражением (7.2.15).

Эффект рангового искажения существует при $B>0$. Поэтому

$$\alpha_{\text{изм}} > \alpha. \quad (7.2.55)$$

Другими словами, неучет эффекта рангового искажения может лишь в известной степени демаскировать феномен негауссовости научной деятельности и социальных явлений вообще. Это влияние обсуждаемого эффекта, однако, не может быть значительным, так как в социальных науках, как это было показано во втором разделе параграфа, эффект рангового искажения проявляется не очень сильно.

Мы по возможности учли эффект рангового искажения, т. е., прежде всего, мы старались не принимать распределение Ципфа с большим ранговым искажением за ципфовое распределение с большим значением α (см. рис. 2.5).

Особенные трудности возникают, когда имеешь дело с ципфовым распределением с большим ранговым искажением. Прояснить картину помогает в этом случае частотный график, который мы строили, если позволял объем выборки. Если же и в этом случае не удавалось отдать предпочтение одному из альтернативных значений α , фиксировалось большее из них как ослабляющее тезис о негауссовости исследуемых распределений. Таким образом, уточнение картины в последующих исследованиях может только усилить этот тезис.

9. ОБСУЖДЕНИЕ:

ИНДИКАТОРНЫЕ И ЛАТЕНТНЫЕ РАСПРЕДЕЛЕНИЯ

В седьмом разделе были исключены из рассмотрения распределения, построенные на ограниченных сверху баллах, т. е. с помощью *закрытых* шкал. В качестве примера рассмотрим распределения людей по коэффициенту интеллекта IQ (см. рис. 2.89). Эти распределения имеют «хвосты» нормальной длины и аппроксимируются обычными гауссовыми неципфовыми распределениями, например распределением Пуассона. Как это увязать с нашим тезисом о негауссовости социальных явлений?

Величина коэффициента IQ устанавливается в результате тестирования. Респонденту задают определенное число вопросов, за ответы на которые он получает баллы по закрытой, скажем, пятибалльной, шкале. Суммарный получаемый респондентом балл превышает поэтому некоторого максимального, характерного для данного конкретного теста. Это необходимо ведет к *деформации* латентного распределения эмпирическим индикаторным распределением людей по IQ . Эта деформация становится неприемлемой при оценках абсолютной величины интеллекта. Искажения, вносимые указанной деформацией, весьма значительны, особенно в области больших значений переменной, т. е. для лиц с большим интеллектом. Согласно рис. 2.89, например, самый интеллектуальный человек всего лишь в 1,5 раза интеллектуальнее среднего. Как ни неопределенны оценки в этой области, представляется бесспорным, что, скажем, гений интеллектуальнее среднего индивида не в 1,5 раза, а значительно более. Если бы баллы не ограничивались сверху, т. е. использовалась бы *открытая* шкала, то такое индикаторное распределение вернее отражало бы истинное положение вещей и имело бы длинный «хвост». Иными словами оно было бы негауссовым.

К аналогичной деформации латентных распределений приводят и другие индикаторы человеческой деятельности, построенные с помощью закрытых шкал. Латентные переменные *аддитивны* по своей природе. Индикаторы же, базирующиеся на закрытых шкалах, *неаддитивны*. И уже в этом проявляется их несоответствие латентным переменным (об аддитивности социальных переменных см. разд. 12 настоящего параграфа). Чтобы установить «истинное» положение вещей, мы должны от закрытых шкал перейти к открытым. В идеале — от индикаторных распределений к латентным.

Это положение, как уже говорилось в гл. 1, носит общий характер. Любой эмпирический феномен устанавливается на индикаторах и затем переносится на латентные переменные. Оправдано ли такое перенесение, в каждом конкретном случае в конечном счете показывает практика.

Перенесенный с социальных стационарных индикаторных распределений на латентные наш тезис о негауссовости социальных явлений из относительно достоверного эмпирического факта превращается в *гипотезу*. Однако таков статус любого эмпирического наблюдения. И проверить нашу гипотезу, как и любую другую эмпирического характера, может только человеческая практика, опыт. Пока же будем полагать, что наша гипотеза о негауссовости научной деятельности и социальных явлений вообще справедлива.

Остается выяснить, почему в наукометрии мы не встречаем распределений, построенных на ограниченных сверху баллах. Дело

в том, что вносимая такими баллами деформация распределений трудно воспроизводима, так как разные процедуры измерения вызывают разную деформацию «хвостов» распределений. Поэтому указанная деформация возникает лишь при использовании не очень хорошо воспроизводимых индикаторов. Воспроизводимость же индикаторов следует уже из определения наукометрических, поэтому наукометрические стационарные распределения и не деформированы при *больших* значениях переменных. Однако они зачастую невоспроизводимо деформированы при *малых* значениях переменных. И происходит это уже вследствие невоспроизводимого обрезания сверху не значений индикаторов, а объемов выборок. Это явление обсуждается в следующем разделе.

10. ОБСУЖДЕНИЕ: РАСПРЕДЕЛЕНИЕ ЦИПФА И ЕГО КОНКУРЕНТЫ

Что следует из феномена негауссовости стационарных наукометрических распределений для аппроксимации выборочных распределений? Можно ли навести порядок в аппроксимациях, которые были приведены выше? Да, по некоторой.

Во-первых, становится понятным, почему все перечисленные аппроксимации исчерпываются гауссовыми и ципфовыми распределениями. Это объясняется теоремой Гнеденко-Деблина (см. разд. 5).

Гауссовы и ципфовы распределения образуют в совокупности множество распределений, сходящихся к устойчивым распределениям. Напомним, теорема Гнеденко-Деблина говорит, что если данное распределение принадлежит этому множеству и не сходится к распределению Гаусса, то оно (с точностью до медленно меняющихся функций) является ципфовым. То обстоятельство, что все эмпирические стационарные распределения могут быть описаны гауссовыми или ципфовыми распределениями, означает, что эмпирические стационарные распределения могут быть описаны только распределениями, *сходящимися к устойчивым*. Это — эмпирическое наблюдение. Верно оно или нет — покажет будущее.

Во-вторых, гауссовы неципфовы аппроксимации здесь не годятся. Из названных выше не годятся отрицательное биномиальное, логарифмическое распределения, обе аппроксимации Гуда, логнормальное, Пуассона и логнормальное Пуассона распределение, а также обобщенное распределение Пуассона при $\theta < 1$.

При небольших выборках данные наблюдения часто могут быть с хорошей точностью совмещены с кривыми гауссовых распределений (см., например, рис. 2.80—2.83). Аппроксимации гауссовы-

ми распределениями некорректны в этих случаях в силу существенной зависимости моментов от объемов выборок (см. разд. 5). При больших объемах выборок эмпирические данные уже просто не удастся совместить с кривыми гауссовых неципфовых распределений (см., например, рис. 2.84—2.87).

В-третьих, здесь применимы ципфовы распределения. Из названных выше — распределение Ципфа, аппроксимации Брукса, Мицевича, Босанквета, Прайса, обобщенное распределение Парето, бета-функция, а также обобщенное распределение Пуассона при $\theta = 1$.

В-четвертых, на ципфовы аппроксимации могут накладываться медленно меняющиеся функции (см. разд. 5).

Колебания данных наблюдения около теоретических кривых заметны на графиках, приводимых в некоторых известных работах [см., например, 3.60]. Такие колебания данных наблюдения около асимптот заметны и на рис. 2.33, 2.35, 2.48, 2.53, 2.57, 2.63, 2.66, 2.76. Без теоремы Гнеденко-Деблина в полной ее формулировке природа этих колебаний оставалась бы неразъясненной.

Ципфовы аппроксимации различаются поведением в области малых значений переменных и значениями α . При этом α зависит от выбора латентной переменной, стационарное распределение значений которой изучается. Например, число статей, число монографий, число нововведений, число открытий и им подобные индикаторы описывают разные латентные переменные научной продуктивности. Поэтому α зависит от выбора индикатора. Именно это обстоятельство, в сущности, фиксируют некоторые авторы [см., например, 3.67; 1.58; 3.78], которые говорят, что доля ученых, производящих большинство научных продуктов, очень сильно зависит от выбора индикатора.

Обсуждая различия отдельных выборочных стационарных наукометрических распределений в области малых значений переменной, следует учитывать, что эти различия зачастую являются артефактами, порожденными невоспроизводимым «обрезанием» выборок. (Это было показано в главах, посвященных отдельным наукометрическим методам, на примерах распределений ученых по числу публикаций и публикаций по числу ссылок в них и на них.)

Процедура «обрезания» выборок необходимо несет на себе отпечаток субъекта исследования, и потому отклонения ципфовых распределений от распределения Ципфа в области малых значений переменных, в свою очередь, должны как-то описываться ципфовым распределением. Иллюстрацией может служить распределение ученых по частичной продуктивности (см. рис. 2.23). Это ципфовое распределение не является распределением Ципфа. Однако распределение отдельных значений частичной продуктивности,

встречающихся в той же выборке ученых, уже описывается распределением Ципфа (см. рис. 2.24, 2.25). Это преобразование ципфоваго распределения, не являющегося распределением Ципфа, в распределение Ципфа, как говорилось при обсуждении распределения Лотки, объясняется тем, что произвольность, связанная с формированием выборки, при таком переходе существенно ослаблена.

В следующем разделе проводится краткое обсуждение теоретических оснований феномена негауссовости научной деятельности и социальных явлений в целом. При этом в силу особого статуса распределения Ципфа его обоснованию уделяется места несколько больше, чем обоснованию других аппроксимаций.

11. ОБСУЖДЕНИЕ: ПОЧЕМУ СОЦИАЛЬНЫЕ ЯВЛЕНИЯ НЕГАУССОВЫ?

Тот факт, что стационарные распределения научной деятельности имеют необычайно длинные «хвосты», требует объяснения. В ряде работ на основе разных предположений выводят различные аппроксимации.

Распределение Ципфа с помощью вариационного энтропийного принципа, следуя традициям статистической физики, выводят Б. Мандельброт [0.94–0.96], Ю. А. Шрейдер [0.60], Л. С. Лозинский [0.39], Л. С. Козачков [1.28], А. И. Яблонский [0.63] и другие. Это распределение выводят также И. Гуд [0.81] и А. Букштайн [1.77].

Бета-функцию (и распределение Ципфа как ее асимптомику) выводят с помощью теории марковских процессов Г. Юл. [0.115], Д. Чемпернауп [0.73], Г. Саймон [0.106; 0.107], И. Стейндл [0.110], С. Крамер [0.75], А. И. Яблонский [0.63] и др. С помощью комбинаторики и статистики неразличимых объектов Бозе-Энштейна бета-функцию выводит В. Хилл [0.86–0.88].

Отрицательное биномиальное распределение выводит Дж. Равичандра Рао [3.87].

Выводы, в основе которых лежат допущения, сами нуждающиеся в обосновании, не позволяют отдать предпочтение той или другой аппроксимации. Это тем более верно, что разные аппроксимации, которые выводятся на основе разных предположений, различаются лишь поведением функции распределения (плотности вероятности) в области малых значений переменной. Здесь же, как об этом говорилось в предыдущем разделе параграфа, различия разных аппроксимаций в большой степени являются артефактами, порожденными невозпроизводимым «обрезанием» выборки.

Тем не менее ценность исследований, в которых выводятся различные аппроксимации стационарных распределений научной деятельности, бесспорна. Только состоит она не в выводах той или другой аппроксимации, а в исследовании в процессе поисков этих выводов связей между «короткохвостовыми» (гауссовыми) и «длиннохвостовыми» (негауссовыми) распределениями, с одной стороны, и явлениями объективного мира — с другой. Особенно ценным в этом отношении представляется вывод распределения Ципфа с помощью вариационного энтропийного принципа, идущего от Л. Больцмана. Рассмотрим его.

Здесь вид распределения, которым описывается данная объективная система, определяется условием максимизации энтропии системы при фиксированных условиях, отвечающих природе объекта:

$$S = - \int f(x) \ln f(x) dx = \max;$$

$$\int E(x) f(x) dx = E; \quad \int f(x) dx = 1 \quad (7.2.56)$$

(последнее условие означает нормировку распределения).

При выводе конкретного распределения для $E(x)$ постулируется конкретная зависимость. $E(x)$ — мера усилий, необходимых для перевода данного элемента системы в состояние x ; E — мера усилий, необходимых для перевода всей системы в состояние, описываемое распределением $f(x)$.

В социальных исследованиях для $E(x)$ постулируется логарифмическая зависимость

$$E(x) = \rho \ln x, \quad (7.2.57)$$

где ρ — константа. Это и дает распределение Ципфа с $\alpha = \rho/E$ [см., например, 0.63, с. 22]. $E(x)$, например, для распределения ученых по числу публикаций имеет смысл усилий, необходимых, чтобы один ученый написал x работ.

Таким образом, вариационный энтропийный принцип дает распределение Ципфа, когда для $E(x)$ имеет место логарифмическая зависимость вида (7.2.57). Постулирование такого рода зависимости для исследований человеческой деятельности имеет под собой определенную эмпирическую базу. А. И. Яблонский [0.65] указывает на подобие выражения (7.2.57) психологическому закону Хика-Хаймена для среднего времени реакции индивида в задачах выбора (это среднее время, очевидно, примерно пропорционально прилагаемому индивидом усилиям).

В естественных науках, как правило, $E(x)$ не имеет логарифмической зависимости от x . В физике, например, $E(x)$ имеет

смысл энергии и часто принимает вид степенной зависимости. Например, $E(x) = mx^2/2$ (кинетическая энергия молекул газа) приводит к максвелловскому распределению молекул по скорости, а $E(x) = mgx$ (энергия частицы в поле тяжести) — к больцмановскому распределению частицы по высоте. Именно по этой причине мы и встречаемся в естественных науках, как правило, с гауссовыми распределениями.

На наш взгляд, следует пытаться вывести, точнее говоря — объяснить, не ту или другую аппроксимацию, а один общий факт «длиннохвостости» (негауссовости) стационарных распределений научной деятельности. Здесь могут быть названы следующие работы [см. 3.78].

Дж. Бриттон развивает гипотезу «священной вспышки». «Длиннохвостость» социальных распределений он выводит из «длиннохвостости» распределений индивидов по их способностям. Последний факт берется за исходный и не объясняется.

Гипотезу «успех порождает успех» развивают В. Мертон [0.100], С. Коул и Дж. Коул [1.88], Р. Аллисон и Дж. Стюарт [3.38], Д. Прайс [1.143] и др. Как и другие концепции, данная гипотеза позволяет получить разные количественные аппроксимации. Д. Прайс [1.143] на ее основе получает бета-функцию, Дж. Равичандра Рао [3.87] — отрицательное биномиальное распределение.

«Альтернативную гипотезу» развивают П. Бурде [0.72] и К. Кнорр [см. 3.78]. Они исходят из того факта, что ученые вовлечены в конкурентную борьбу за ассигнования.

На наш взгляд, слабость названных концепций состоит в том, что стационарные распределения научной деятельности рассматриваются в них изолированно от других. Между тем, чтобы понять явление, следует выйти за его пределы. Сравним стационарные природные и социальные распределения.

Доминирование в природе среди стационарных распределений, распределений, сходящихся к нормальному, т. е. гауссовых, хорошо известно и не нуждается в специальном обосновании. Было бы неверно, однако, говорить о гауссовости *всех* стационарных природных распределений. Примерами негауссовых природных стационарных распределений могут служить, например, распределения химических элементов в земной коре по содержанию [0.10], малых космических тел по массе [0.29], космических частиц по энергии [1.99, 1.107].

Итак, природа — гауссова, социальные явления — негауссовы. Могут быть высказаны разные предположения относительно того, почему гауссовые и ципфовые распределения именно так разделили сферы влияния. Автор склоняется к предположению, свя-

занному с принципом Циглера [0.58; 0.55], согласно которому из возможных путей эволюции любой объективной системы реализуется путь, отвечающий максимальному производству (скорости порождения) в ней энтропии. Существуют аргументы в пользу того, что негауссовые системы обеспечивают большие скорости порождения энтропии, чем гауссовые. Изложим здесь один из них.

Негауссовые распределения являются ципфовыми с $\alpha < 2$. Ципфовые распределения с $2 < \alpha < \infty$ являются уже гауссовыми, а гауссовые неципфовые распределения (Гаусса, Пуассона и т. д.) имеют $\alpha = \infty$, т. е. негауссовые распределения имеют меньшие значения α , чем гауссовые. Но энтропия распределения Ципфа тем больше, чем меньше α (см. выражение (7.2.24) и рис. 2.88). Таким образом, распределения негауссовой совокупности стационарных распределений человеческой деятельности имеют большую энтропию, чем распределения гауссовой совокупности стационарных природных распределений. Более того, мы видели (см. рис. 2.76, 2.78), что стационарные распределения научной деятельности как творческой имеют в целом меньшие значения α (большие значения энтропии), чем стационарные распределения менее творческих видов человеческой деятельности.

Если верно данное предположение, то материя в ходе своей эволюции в сторону возрастания энтропии перешла от гауссовых систем природы к негауссовым системам социального мира. Человеческое общество, развивая все более творческие виды деятельности, также, оказывается, развивается в сторону дальнейшего возрастания энтропии. В рамках этого предположения разумно допустить, что немногие встречающиеся в природе ципфовые явления могли послужить «катализатором» будущих ципфовых систем жизни и человека. Спровоцировать появление жизни на Земле могло, в частности, космическое излучение. И сегодня вероятности появления отдельных мутаций в генофонде человеческой популяции в какой-то степени определяются параметрами такого известного мутагенного фактора, как космическое излучение с его ципфовым энергетическим спектром. В литературе прочно закрепилась точка зрения о «мутагенной» природе творческого процесса.

12. ОБСУЖДЕНИЕ: ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В дополнение к изложенным в гл. 1 следствиям из негауссовости научной деятельности для исследований науки скажем несколько слов об аппарате *негауссовой статистики*, который должен лежать в основании количественного анализа науки. Ясно, что в его основе должны лежать характеристики статистических распределений, слабо зависящие от объемов выборок и сходящиеся (с устремлением этих объемов к бесконечности) к конечным пределам. Такими характеристиками могут служить *показатель ципфоваго распределения α , энтропия распределения и его квантили*.

О сходимости к конечному пределу энтропии распределения Ципфа с ростом объема выборки говорилось во втором разделе. Что же касается квантилей распределения, то они также сходятся к конечным пределам. В самом деле, квантилем порядка P называется такое значение x_P случайной величины x , для которого $F(x_P) = P$. Для распределения Ципфа (7.2.11) находим:

$$x_P = [1/x_0^\alpha - P(1/x_0^\alpha - 1/I^\alpha)]^{-1/\alpha} \xrightarrow[N \rightarrow \infty]{(x \rightarrow \infty)} x_0 (1 - P)^{-1/\alpha}. \quad (7.2.58)$$

В гауссовой статистике коэффициенты корреляции базируются, как и аппарат ошибок, на моментах второго порядка. В негауссовой статистике коэффициенты корреляции должны быть заменены *коэффициентами взаимосвязи*, строящимися, например, на основе энтропии распределения (см. гл. 6). Возможно, здесь окажется полезной развиваемая С. Кульбаком [0.38] математическая статистика, в которой коэффициент корреляции двух распределений $f_1(x)$ и $f_2(y)$ заменен коэффициентом расхождения:

$$I_{1,2} = \iint f(x, y) \log \frac{f(x, y)}{f_1(x) f_2(y)} dx dy. \quad (7.2.59)$$

Соответственно должен быть видоизменен и факторный анализ. Для оценки параметров распределения Ципфа пригоден, как о том говорит А. И. Яблонский [0.63], метод максимального правдоподобия. В нем ищутся значения параметров распределения

$$\eta_k = (x_1, x_2, \dots, x_N) \quad (7.2.60)$$

($k=1, 2, \dots, m$; m — число параметров; $x_1, x_2, \dots, x_N$ — данная выборка значений случайной величины x объема N), максимизи-

рующие функцию правдоподобия

$$L(x_1, x_2, \dots, x_N) = \varphi(x_1) \varphi(x_2) \dots \varphi(x_N), \quad (7.2.61)$$

где $\varphi(x)$ — предполагаемая плотность распределения величины x на генеральной совокупности. Для этого решают систему m уравнений максимального правдоподобия:

$$\frac{\partial}{\partial \eta_k} \ln(x_1, x_2, \dots, x_N; \eta_1, \eta_2, \dots, \eta_m) = 0. \quad (7.2.62)$$

Этот метод сохраняет силу в негауссовой статистике.

В целом в связи с незавершенностью разработки негауссовой статистики вопросов возникает значительно больше, чем мы решаем здесь. Все эти вопросы мы адресуем к математикам, которые должны удовлетворить потребности социальных исследований, в частности науковедения, в такой статистике.

Так, нуждается в более строгой проработке вопрос о зависимости различных характеристик негауссовых распределений, включая их моменты, от объема выборки. Эти характеристики имеют свои распределения, форма которых зависит в общем случае от объема выборки. Во втором разделе параграфа для моментов мы заменили такой анализ приближенным расчетом их зависимостей от максимального выборочного значения переменной I . При этом предполагалось, что выборка значений случайной величины x описывается тем же распределением, что и генеральная совокупность. Между тем распределение выборки может отклоняться от распределения генеральной совокупности и тем значительнее, чем меньше объем выборки. Поэтому, строго говоря, моменты зависят от I также вероятностным образом, а даваемые нами зависимости (7.2.19) — (7.2.21) — всего лишь оценки зависимостей наиболее вероятных значений моментов от I . Какова статистическая достоверность таких оценок, это еще предстоит установить.

Феномен негауссовости социальных явлений также порождает много вопросов. Так, при соответствующей замене переменных негауссовые распределения становятся гауссовыми, и наоборот. В связи с этим, может быть, вместо разработки негауссовой статистики, следует просто использовать переменные, распределения по которым для социальных явлений гауссовы? На наш взгляд, это было бы неверно. Как показала практика, целесообразнее выбирать *аддитивные* (в математическом смысле) переменные*. Между

* Содержание понятия аддитивности в математике отличается от содержания этого понятия, скажем, в физике. Например, энергия в физическом смысле в общем случае неаддитивна. В математическом же — аддитивна. В математике аддитивны переменные, для которых определены опера-

тем, можно заметить, что всегда, когда используется закрытая шкала, социальные переменные неаддитивны. Более того, справедливо, по-видимому, и более сильное наблюдение: социальные переменные аддитивны тогда и только тогда, когда их распределения негауссовы. В этом и состоит суть тезиса о негауссовости социальных явлений.

В теории социального измерения (С. Стивенс, К. Кумбс, П. Суплес и Дж. Зиннес, П. Ф. Лазарсфельд, И. Пфанцгаль и др.) неаддитивность социальных данных принимается как факт и отношение между индикаторами и этими латентными переменными описывается с помощью теории множеств. В результате теория зашла в тупик: для неаддитивных данных дело ограничивается, фактически, качественным измерением. Практический выход из этой ситуации, по нашему мнению, может быть довольно простым: отказаться от неаддитивных переменных. Работать следует, по возможности, с аддитивными, т. е. негауссовыми, переменными. В частности, *следует отказаться от закрытых шкал.*

То обстоятельство, что в социальных науках аддитивны лишь переменные, распределение которых имеет форму цифрового распределения, может быть положено в основу критерия аддитивности социальных переменных. Скажем, показатель распределения Циффа α аддитивен (см. рис. 2.78), а энтропия распределения — неаддитивна и т. д.

Другой важный вопрос. Абстрактно рассуждая, можно прийти к выводу, что гауссовы распределения — не единственные конкуренты негауссовым. Во-первых, не обязательно ограничиваться гауссовыми и негауссовыми распределениями, т. е. не обязательно ограничиваться распределениями, для которых распределение среднего $\bar{x}$ с неограниченным ростом объема выборки стремится к некоторой предельной форме. Не обязательно вообще накладывать какие-либо ограничения на распределения выборочных средних.

Во-вторых, не обязательно ограничиваться распределениями независимых случайных величин.

Все это так, но верно лишь на уровне абстракций. Если данное явление может быть описано более простой моделью, зачем прибегать к сложной. Негауссовая статистика — простейшая из математических статистик, не являющихся гауссовыми. Ее возможности не только не исчерпаны, но пока еще и не выявлены должным образом. Сначала следует внедрить негауссовую ста-

ция сложения и, следовательно, другие арифметические операции. На этих операциях базируются алгебра, дифференциальное и интегральное исчисления и, вообще, практически вся применяемая сегодня в социальных науках математика.

истику в социальные науки взамен гауссовой, поднять с ее помощью воспроизводимость результатов и т. д. Только тогда, быть может, социальным наукам окажется тесно в рамках негауссовой статистики и можно будет поставить вопрос о переходе от негауссовой статистики к более сложным математическим статистикам. Пока же следует сосредоточиться на негауссовой статистике.

Таким образом, *предлагается постулировать, что гауссовыми и негауссовыми распределениями могут быть описаны все стационарные распределения, встречающиеся в естественных и социальных науках.*

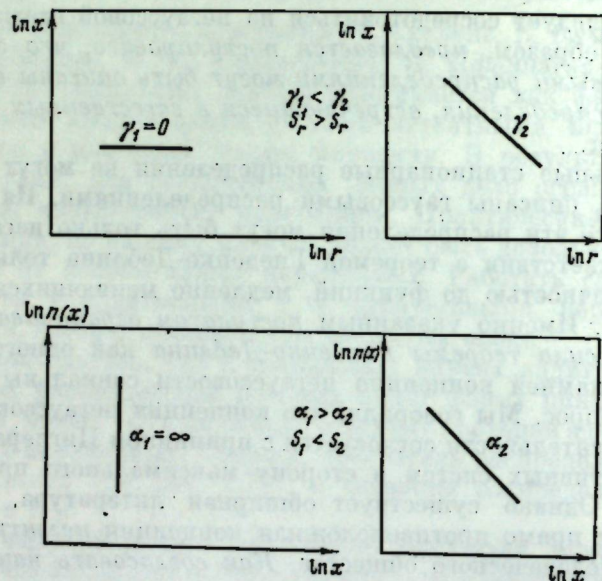
Социальные стационарные распределения не могут быть, как говорилось, описаны гауссовыми распределениями. Из постулата следует, что эти распределения могут быть только негауссовыми, т. е. в соответствии с теоремой Гнеденко-Деблина только цифровыми (с точностью до функций, медленно меняющихся в смысле Карамата). Именно указанным постулатом ограничивается доказательная сила теоремы Гнеденко-Деблина как одного из краеугольных камней концепции негауссовости социальных явлений.

Еще вопрос. Мы говорили, что концепция негауссовости человеческой деятельности согласуется с принципом Циглера об эволюции объективных систем в сторону максимального производства энтропии. Однако существует обширная литература, в которой проводится прямо противоположная концепция негэнтропийности жизни и человеческого общества. *Как согласовать нашу концепцию с наблюдающимся в социальных системах ростом организованности, структурированности?*

Заметим, прежде всего, что концепция негэнтропийности базируется на качественных оценках связи переменных «структурированность», «организованность» и т. д. с энтропией. Принято считать, что структурирование, упорядочение сопровождаются уменьшением энтропии. Поскольку в социальном мире наблюдается рост организованности, структурированности и т. д., постольку и делается вывод об уменьшении здесь энтропии. Между тем связь структурированности, организованности и т. д. с энтропией не так однозначна.

Допустим, мы рассматриваем некоторый коллектив (например, ученых). И пусть этот коллектив представлял собой сначала однородную массу сотрудников, а затем из них выделились лидеры разных рангов, коллектив приобрел определенную структуру, стал более «организованным», более «структурированным». Что произойдет при этом с энтропией распределения членов коллектива по их статусу? Согласно концепции негэнтропийности социальных систем она должна уменьшиться. И, действительно, энтропия ранговой формы распределения членов коллектива по их статусу

уменьшится (см. выражение (7.2.25)), ранговое распределение станет более неоднородным. Однако энтропия частотной формы распределения членов коллектива по их статусу возрастает (см. выражение (7.2.24)), частотное распределение станет более одно-

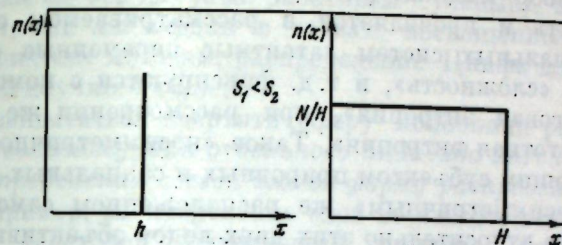


родным ($n(x)$ — число сотрудников со статусом x).

Частотная и ранговая формы математически эквивалентны, однако в законе возрастания энтропии и принципе Циглера фигурирует именно *частотная* энтропия, а не ранговая. Поэтому в нашем примере с ростом структурированности социальной системы энтропия также растет. Этот пример иллюстрирует положение о том, что рост организованности, структурированности и т. д. социальных систем сопровождается вопреки традиционным представлениям *ростом энтропии*. Это подрывает основы концепции негэнтропийности социального мира.

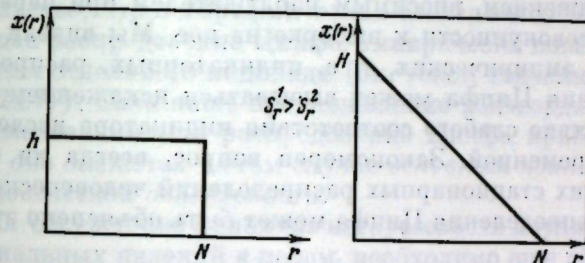
В чем же корни мифа об уменьшении энтропии с ростом структурированности социальных систем? Этот миф коренится в естественных науках. Возьмем для примера природную систему. Допустим, мы имеем в сосуде идеальный газ, находившийся поначалу в таком внешнем поле, что все его N молекул были сосредоточены на некоторой высоте h . Затем мы «выключили» поле и получили однородный газ. В подобных случаях принято говорить об уменьшении упорядоченности, структурированности

системы. А что произошло при этом с энтропией распределения молекул по высоте? *Частотная* энтропия возросла.



($n(x)$ — число молекул в слое единичной толщины на высоте x), частотное распределение стало более однородным.

Энтропия же *ранговой* формы уменьшилась. ($r(x)$ — ранг молекулы, находящейся на высоте x , т. е. ее порядковый номер в ря-



ду молекул, расположенных в порядке убывания их высоты. Ранговое распределение стало менее однородным. Этот пример иллюстрирует то положение, что рост энтропии естественнонаучных систем сопровождается в полном соответствии с традиционными представлениями *уменьшением структурированности*, упорядоченности природных систем. На этом базируется концепция «энтропийности» природы.

Таким образом, для социальных систем рост энтропии сопровождается ростом структурированности, упорядоченности и т. д., а для естественнонаучных систем — их уменьшением. Почему? Мы имеем здесь дело с *артефактом*, порожденным субъективной природой наблюдателя. Дело в том, что структурированность, организованность, упорядоченность и т. д. — это *латентные* переменные. Отражая реальные свойства реальных объектов, латентные переменные тем не менее *выбираются* познающим субъектом, который осознанно или неосознанно руководствуется определенной теоретической концепцией, традициями и т. д. (см. гл. 1, § 2). Выбор латентных переменных, таким образом, в известной степени

произволен, субъективен. Выбрав индикатор (группу индикаторов) для измерения латентной переменной, мы тем самым фиксируем ее. Выбор индикатора в том же смысле произволен. Эта произвольность и проявляется в рассматриваемом случае. При анализе социальных систем латентные переменные «структурированность», «сложность», и т. д. фиксируются с помощью индикатора «ранговая энтропия», при рассмотрении же природных систем — «частотная энтропия». Такое «несимметричное» рассмотрение познающим субъектом природных и социальных систем объясняется «несимметричным» же расположением самого познающего субъекта относительно этих двух видов объективных систем: человек принадлежит социальному миру.

Последний вопрос, который мы здесь рассмотрим, это вопрос о статусе распределения Ципфа относительно всех цифровых распределений. Мы видели, что в ряде случаев отклонение эмпирических распределений от распределения Ципфа может быть объяснено искажением, вносимым наблюдателем при переходе от генеральной совокупности к выборке из нее. Мы видели также, что отклонение эмпирических, т. е. индикаторных, распределений от распределения Ципфа может вызываться искажением, возникающим вследствие слабого соответствия индикатора исследуемой латентной переменной. Закономерен вопрос, всегда ли отклонение эмпирических стационарных распределений человеческой деятельности от распределения Ципфа может быть объяснено этими двумя причинами?

Отвечая на этот вопрос, следует сказать, что могут быть указаны социальные стационарные латентные распределения на генеральной совокупности, которые не описываются распределением Ципфа. Речь идет о распределениях значений случайных величин, описывающих свойства, неотъемлемые от каждого объекта данной генеральной совокупности, в известном смысле *определяющие* эту совокупность. Таким распределением является, например, распределение индивидов по числу слов в словаре каждого. Если бы это распределение было распределением Ципфа, мы имели бы бесконечное число индивидов, не имеющих в запасе ни одного слова. Заведомо это не так (мы исключаем младенцев). Заведомо это распределение в области малых значений переменной будет существенно отклоняться от распределения Ципфа. Человека характеризует, прежде всего, умение говорить. Это свойство каждого объекта данной генеральной совокупности. Отклонения здесь аномальны.

Если же переменная характеризует только часть объектов генеральной совокупности, то возникает распределение Ципфа. Возьмем снова в качестве генеральной совокупности множество инди-

видов. Но в качестве латентной переменной возьмем уже частную характеристику — *научный вклад индивида*. Не все индивиды — ученые (даже не все научные работники — ученые), поэтому здесь и возникает, как мы видели в главах, посвященных отдельным наукометрическим методам, распределение Ципфа для различных индикаторов научного вклада.

Можно попытаться очертить сферу монополии распределения Ципфа. *По-видимому, для отдельного индивида латентные стационарные распределения всегда имеют форму распределения Ципфа.* Таково, например, распределение слов по частоте встречаемости в речевой и текстовой деятельности индивида (см., например, рис. 2.45, 2.46). Таково распределение цифр по их предпочтению индивидом (см. рис. 2.74). Таково, видимо, вообще распределение объектов данного вида, с которыми имеет дело индивид (слов, цифр, публикаций, знакомых ему людей и т. д.), по степени предпочтения, которое он им оказывает. Неслучайным выглядит в этой связи то обстоятельство, что логарифмическая зависимость $E(x) = \rho \ln x$, генерирующая распределение Ципфа, эмпирически пока обнаружена именно для отдельного индивида (см. текст, связанный с выражением (7.2.57)). Если наше предположение подтвердится в дальнейших исследованиях, то распределение Ципфа приобретет особый статус: оно окажется в этом случае *основным количественным законом человеческой деятельности*.

Завершая обсуждение феномена негауссовости научной деятельности и социальных явлений в целом, необходимо еще раз подчеркнуть, что для дальнейшего обоснования и проверки развиваемых здесь представлений нужны систематические опорные исследования по разработке все более адекватных систем латентных переменных и соотносению с ними системы индикаторов. Переход ко все более адекватной системе латентных переменных, как говорилось, означает переход ко все более эффективной деятельности человека. Поэтому проверять обсуждаемое эмпирическое наблюдение (в числе других) — это значит вести во всем объеме исследования природы и социальных явлений, включая науку.

НАУКОМЕТРИЯ В ОБЩЕМ КОНТЕКСТЕ ИССЛЕДОВАНИЙ НАУКИ (вместо заключения)

Многофакторность науки диктует необходимость комплексного применения различных методов ее исследования. Для наукометрии это означает комплексное применение, во-первых, отдельных наукометрических методов и, во-вторых, наукометрических и пенаукометрических методов. Исследование места наукометрических методов в общем ряду науковедческих методов предполагает предварительное изучение возможностей каждого наукометрического метода. Такая попытка и была предпринята нами. Теперь предстоит выполнить основную задачу книги — уяснить место наукометрии в науковедении.

1. КОМПЛЕКСНОЕ ПРИМЕНЕНИЕ НАУКОМЕТРИЧЕСКИХ МЕТОДОВ (взгляд на наукометрию «изнутри»)

Синтез наукометрических методов может осуществляться на различных уровнях. *Использование различных наукометрических методов для получения одного и того же результата* — это синтез низшего порядка. Такое дублирование необходимо не только для увеличения надежности получаемых результатов, но и для анализа самих наукометрических индикаторов в ходе *опорных исследований*.

Среди работ, в которых проводится подобное дублирование наукометрических методов, помимо названных ранее, можно отметить следующие. Ф. Чаней [3.48] сравнивает распределения ученых по числу патентов и по числу технических статей. Корреляция между этими распределениями оказывается равной 0,05, т. е. очень малой.

Дж. Вестбрук [4.142] определяет эффективность лабораторий по исследованию керамических материалов с помощью «цитат-индекса» и по числу публикаций. Корреляцию между этими индикаторами он находит слабой.

У. Менард [1.125] для измерения роста научных исследований использует наряду с числом ссылок число публикаций.

Ю. В. Грановский и соавт. [1.16] сопоставляют распределения Нобелевских премий по городам и странам с аналогичными распределениями, получаемыми на основе «цитат-индекса». Обнаружена слабо выраженная корреляция.

Синтез более высокого уровня — *комплексное использование различных наукометрических методов для получения результатов, которые не могут быть получены при раздельном их использовании*.

Х. Тимани [2.66] исследует временную динамику численности научных коллективов с определенным средним возрастом их членов (33 и 36 лет) для разных возрастов выхода из коллектива. Полученные данные в противовес данным Я. Влахи [3.100] дают основание утверждать, что в сильных исследовательских коллективах средний возраст сотрудников сохраняется постоянным (33—36 лет).

Я. Влахи [3.100] изучает зависимость продуктивности группы ученых, определяемой по числу публикаций, от среднего возраста членов группы. Найденная им зависимость (см. рис. 4.2) носит линейный характер.

Я. Влахи [2.70], Д. Прайс и Д. Бивер [2.30] изучали корреляцию между числом работ данного ученого и числом его соавторов. В соответствии с их данными (см. рис. 4.3, 4.4), эта корреляция положительна, т. е. соавторствующие ученые чаще имеют большее число публикаций.

Г. М. Добров [0.22] дает график (см. рис. 4.5), показывающий зависимость числа повторных изобретений от числа авторов. Согласно графику, чем больше авторов, тем выше вероятность новизны изобретения.

М. Кочен и Р. Таглиакоццо [2.52] ищут связь между производительностью автора, определяемой им по числу публикаций, и рассеянием его работ по различным журналам. Оказалось, что производительные авторы (5 или более статей в год) сосредотачивают свои работы в меньшем числе журналов.

Дж. Валлмарк и соавт. [4.141; 4.140] изучают зависимость продуктивности группы ученых, определяемой с помощью метода ссылок, от ее численности. Для эффективности группы, определяемой как число ссылок на работы группы, деленное на число ее членов, была найдена экспоненциальная зависимость от размера группы (см. рис. 4.1).

Р. Ф. Васильев [1.6] и Ю. М. Червинский [1.66] изучали зависимость цитирования работ от числа их авторов. Согласно, напри-

мер, данным Червинского, коллективные работы цитируются в 2,8 раза чаще, чем индивидуальные.

М. Бакленд [4.49] исследует связь старения литературы, которое характеризуется долей цитируемых публикаций старше 10 лет, с рассеянием публикаций по периодическим изданиям (см. рис. 4.6).

Оценивая ситуацию, следует отметить, что наукометрические исследования, в которых бы комплексно применялись различные индикаторы, только начинают развиваться. А между тем такое комплексное применение наукометрических индикаторов может дать весьма многое при решении науковедческих задач.

Предметный уровень науковедческой проблематики и наукометрия

Здесь стоит задача разработки критериев относительной перспективности различных участков научной проблематики.

Структурные исследования. Наукометрия позволяет построить «слепок» предметной структуры данного участка исследований. Для этого путем измерения взаимосвязи отдельных выполняемых исследований или их фрагментов воссоздается наукометрическими средствами *сеть связей* этих структурных единиц анализа и далее с помощью, скажем, факторного анализа (точнее — его негауссово-го варианта) воссоздается искомая предметная структура. При этом могут использоваться самые разные индикаторы — частота соавторства, число ссылок, число сленговых слов. В полученной сети связей относительная *перспективность* проблематики, отождествляемой с данным узлом этой сети, определяется *степенью связанности* этого узла со всей сетью.

Другой мерой *перспективности* данного участка проблематики является определенная по сети связей этого участка *степень структурированности* или координированности исследований. Согласно некоторым данным, в перспективных направлениях уровень координированности исследований превышает таковой по науке в целом. В общем случае следует говорить об оптимальном уровне таковой координированности. См. также гл. 4, § 5.

Исследования стационарных распределений. Эти распределения также могут быть использованы для выявления *перспективных* участков проблематики. Некоторые данные говорят о том, что перспективные направления характеризуются распределениями, имеющими более короткие «хвосты», чем традиционные, т. е. они более *негауссовы*.

См. также гл. 2—4 (§§5) и гл. 6, § 3.

Исследования нестационарных распределений. Показателем

перспективности данного научного направления может служить *скорость* роста различных описывающих его индикаторов — число ученых, журналов, публикаций, ссылок.

По скорости роста наукометрических индикаторов в данный момент, отнесенной к общей логистической кривой роста, может быть определена *фаза развития* данного направления. Направления, находящиеся в начале своего развития или на закате, перспективны, конечно, в разной степени. См. также гл. 4, § 5.

Межличностный уровень науковедческой проблематики и наукометрия

Здесь могут быть выделены три круга проблем. Во-первых, проблемы, связанные с разработкой критериев, в соответствии с которыми можно было бы на определенные участки проблематики подбирать конкретные научные коллективы. Во-вторых, проблемы, связанные с разработкой оптимальной организационной структуры научных коллективов. В-третьих, проблемы, связанные с оценкой их эффективности.

Структурные исследования. Наукометрия позволяет построить «слепок» предметной структуры данного научного коллектива или данной системы коллективов. Для этого следует построить наукометрическими средствами *сеть связей* между текстами отдельных структурных единиц — ученых, научных подразделений и т. д. В качестве текстов могут выступать либо публикации, либо рефераты, написанные специально для этого анализа. Попросив научных работников изложить в таких рефератах, какой теме они отдали бы предпочтение, если бы им была предоставлена свобода выбора, можно сравнить полученную таким образом *«потенциальную»* предметную структуру научного коллектива с *реализованной* сегодня, как она выглядит по публикациям. Аналогичным путем можно получить представление о *«перспективной»* предметной структуре данного коллектива, как она мыслится составляющими его учеными. В полученных сетях связей относительная *перспективность* работ ученых или коллективов определяется *степенью связанности* данного участка этих сетей со всеми сетями.

Другой мерой *перспективности* исследований данного коллектива является количественная характеристика внутренней связанности сети связей, соответствующей этому коллективу (см. также гл. 4, § 5).

Исследование стационарных распределений. Коэффициент корреляции между наукометрическими стационарными распределениями, описывающими данное подразделение научного коллектива и весь коллектив в целом, может служить индикатором *соответ-*

вия тематики исследований этого подразделения тематике исследований коллектива. Аналогично можно оценить соответствие исследований данного коллектива планируемым исследованиям или данному научному направлению.

Отдельные стационарные распределения могут использоваться для оценки перспективности исследований данного коллектива или отдельных его подразделений (см. также гл. 2—4 (§§ 5) и гл. 6, § 3).

Исследования нестационарных распределений. Скорость роста во времени различных индикаторов объема научного выхода данного научного коллектива — числа публикаций, ссылок и др. — дает возможность судить о перспективности последнего. По полученной таким образом кривой роста можно судить о фазе развития, в которой он находится (См. также гл. 6, § 5).

Личностный уровень науковедческой проблематики и наукометрия

Здесь также можно выделить три круга проблем. Во-первых, проблемы, связанные с оптимальной расстановкой научных кадров. Во-вторых, связанные с созданием оптимального режима работы для научных исследователей. И, в-третьих, проблемы, связанные с оценкой эффективности деятельности ученых.

Структурные исследования. Используя наукометрические индикаторы, и прежде всего наиболее «дробные», например число слов сленга, можно построить «слепок» предметной структуры исследований отдельного ученого или группы ученых. По связанности соответствующего узла *сети связей* со всей сетью можно оценить относительную перспективность (эффективность) в целом исследований ученого или отдельных фрагментов таких исследований. При этом не обязательно отталкиваться от публикаций, в основу анализа могут быть положены и *рефераты*, специально для этого написанные. Использование сленгового метода позволяет подойти к решению проблемы *диагностики потенциальных лидеров* науки: работы никому не известного ученого могут оказаться центральными в сленговом «слепке» предметной структуры, скажем данного научного направления.

Проблема создания оптимального режима работы для ученого связана с поисками оптимального уровня *концентрированности* тематики его исследований и *координированности* его работы с исследованиями коллег. Характеристики наукометрических сетей связи полезны для решения и этой проблемы. Измерения концентрированности предметной структуры исследований ученого окажутся полезными и при нахождении оптимального уровня его *профессиональной мобильности* — характеристики, отражающей

способность ученого переключаться на новую проблематику и тесно связанной с его *разносторонностью*.

Структурный анализ системы индикаторов, характеризующих отдельных ученых, позволяет подойти к решению проблемы *типологии ученых*. Используя, например, факторный анализ (точнее — негауссовый эквивалент факторного анализа, который пока, к сожалению, не разработан), можно разбить индикаторы на группы-факторы, объединяющие тесно коррелирующие индикаторы. Различным вариантам разбиения индикаторов на группы, допускаемым эмпирическими данными, будут соответствовать разные типологии (см. также гл. 6, § 3).

Исследования стационарных распределений. Коэффициент корреляции двух стационарных наукометрических распределений, из которых одно относится к данному ученому, а другое — к данному научному коллективу, является индикатором *связанности* тематики исследований ученого и научного коллектива. Таким путем можно установить степень соответствия в предметной плоскости ученых их рабочим местам — реальным или гипотетическим. При этом можно брать не только публикации, но и *рефераты*, написанные специально для такого анализа.

Отдельные стационарные наукометрические распределения могут использоваться для оценки разносторонности данного ученого, его творческой и т. д. (см. также гл. 2—4 (§ 5) и гл. 6, § 3).

Исследования нестационарных распределений. По скорости роста во времени наукометрических индикаторов данного ученого, соотношенной с кривой возрастной динамики научного творчества, можно судить о *фазе развития*, в которой находится данный ученый.

Высказанные здесь соображения о возможных применениях наукометрических методов порознь и комплексно носят во многом *гипотетический характер*. Дело не в том, что предполагаемые связи между латентными переменными и их наукометрическими индикаторами могут не существовать. Связи эти, несомненно, существуют в социальных явлениях. Вопрос заключается в том, *насколько эти связи существенны*. Если они слабо выражены, то их трудно использовать. По-видимому, часть обсуждавшихся выше наукометрических приложений не реализуется, однако, какая именно часть, сказать сейчас, без проведения в достаточном объеме *опорных исследований*, невозможно.

О достоверности наукометрического измерения, определяемой уровнем опорных исследований, говорится в следующем параграфе. Здесь же следует сказать несколько слов о том, какими соображениями нужно руководствоваться при выборе того или другого наукометрического индикатора.

Проведенный в предыдущих главах анализ отдельных наукометрических методов показывает, что все они имеют единую основу: это — статистические методы с одним и тем же аппаратом негассусовой статистики. Отдельные наукометрические методы различаются степенью «дробности» лежащих в их основе индикаторов. Один и тот же исследуемый массив — публикации, тексты данной группы ученых — для разных наукометрических индикаторов дает счетные массивы разного объема. Если ученых N , то публикаций в счетном массиве публикаций уже aN , где a — среднее число публикаций на одного ученого. Число ссылок в счетном массиве ссылок составляет уже abN , где b — среднее число ссылок на публикацию. Число сленговых слов в счетном массиве сленговых слов равно acN , где c — среднее число таких слов на публикацию. При заданном уровне репрезентативности, фиксирующем объем счетных массивов, объемы исследуемых массивов разные для различных индикаторов. Если взять (что соответствует действительности) $a \approx 1,5$, $b \approx 15$ и $c \approx 1000$, то получим, что объем исследуемых массивов для метода подсчета числа публикаций примерно в 1,5 раза меньше, чем для метода подсчета числа ученых, для метода ссылок — меньше примерно в 22,5 раза и для сленгового — примерно в 1500 раз.

Оптимальные границы применимости отдельных наукометрических методов определяются, с одной стороны, приведенными выше статистическими соображениями о репрезентативности выборки, ограничивающими объемы счетных массивов снизу, а с другой — соображениями о трудоемкости методов, ограничивающими эти массивы сверху. Методы, построенные на измерителях малой «дробности», оптимальны при анализе научных направлений или больших научных коллективов. При анализе небольших предметных областей и средних по размеру научных коллективов выгоднее «цитат-индекс» и сленговый метод, примененный к анализу текстов большой степени свернутости. Для анализа же малых научных коллективов и отдельных ученых удобен сленговый метод, примененный к анализу текстов малой степени свернутости.

Верхние границы применимости наукометрических методов, определяемые их трудоемкостью, не имеют столь жесткого характера, как нижние, определяемые требованиями статистической достоверности. Это дает возможность маневрирования наукометрическими методами, сочетая в случае необходимости грубые наукометрические индикаторы малой «дробности», использование которых может обеспечить получение приближенной картины исследуемых явлений, с тонкими индикаторами большой «дробности», использование которых может обеспечить получение картины более тонкой структуры этих явлений.

Функции научного руководства процессами интенсификации науки, возлагаемые на науковедение, требуют, чтобы оно имело «глубоко эшелонированную оборону». Первый ее рубеж — прикладные исследования, дающие мгновенную, пусть даже небольшую отдачу при небольших затратах средств. Второй рубеж — исследования тактической перспективы, отдача от которых ожидается в ближайшем будущем. И третий рубеж — исследования стратегической перспективы, отдача от которых ожидается не скоро, но которые обещают наиболее полное и адекватное рассмотрение. Прикладные наукометрические исследования удобно вести методами минимальной трудоемкости, основанными на индикаторах малой «дробности» — числе журналов, ученых, публикаций. Исследования тактической перспективы удобно вести методами, основанными на индикаторах средней «дробности», к числу которых может быть отнесен «цитат-индекс». И, наконец, исследования стратегической перспективы следует вести методами, использующими индикаторы большой «дробности», к числу которых относится, прежде всего, сленговый метод.

2. КОМПЛЕКСНОЕ ПРИМЕНЕНИЕ НАУКОМЕТРИЧЕСКИХ И ПЕНАУКОМЕТРИЧЕСКИХ МЕТОДОВ (взгляд на наукометрию «изнутри»)

Синтез наукометрических и пенаукометрических методов, осуществляемый таким образом, что эти методы используются параллельно, — это синтез низшего порядка. Примеры такого рядоположного применения наукометрических и пенаукометрических методов содержат работы У. Гарвея и Б. Гриффита [1.104] и Д. Крэйна [1.90].

Другой род синтеза наукометрических и пенаукометрических методов — использование их для получения одного и того же результата. Такое дублирование проводится, прежде всего, в целях увеличения достоверности получаемых результатов. Необходимо оно и для опорных исследований наукометрических индикаторов.

В числе работ, в которых производится дублирование наукометрических и пенаукометрических методов, кроме перечисленных ранее, назовем следующие. Г. Залтман [0.116] контролирует результаты социологического анализа неформальных контактов ученых статистическим методом — подсчитывается число получаемых респондентами препринтов, отчетов других организаций и учреждений, число телефонных разговоров и пр.

Д. Пельц и Ф. Эндрюс [0.51] используют метод подсчета числа публикаций для контроля социологических методов определения продуктивности ученого (см. рис. 1.23, 1.24).

С. Коул [4.61] использует метод ссылок для подтверждения так называемого эффекта Матфея, согласно которому публикации высокоцитируемых ученых при прочих равных условиях привлекают большее внимание, чем публикации малоцитируемых.

Д. Алик [1.72] использует ряд наукометрических индикаторов (число журналов, редактируемых данным ученым; число написанных им обзоров и др.) для проверки корректности выделения им социологическими средствами научной элиты в 13 человек из 199 химиков.

М. Эдвардс и М. Маккаррей [1.96] обсуждают многопараметрическое исследование продуктивности ученого, в ходе которого используются и наукометрические индикаторы.

Наибольший интерес представляет такой синтез наукометрических и ненаукометрических методов, когда они используются для получения результатов, которые не могут быть получены при раздельном их использовании.

В работе [1.133] фиксируется линейная связь между ассигнованиями на науку в данной стране и числом в ней научных работников (см. рис. 4.7а).

М. М. Кураши [1.151; 1.152], М. Моравчик [1.128], П. Бласко [1.75] и другие изучали зависимость между национальным доходом страны и численностью в ней ученых и другими наукометрическими индикаторами объема науки. Оказалось, что эта зависимость носит степенной характер (см. рис. 4.7б).

Д. Прайс и С. Гюрсей [1.149] изучали связь между производством в данной стране электроэнергии и наукометрическими индикаторами объема науки — числом ученых и пр. Эта зависимость также носит степенной характер.

С. Коул и Дж. Коул [4.63] и Г. Цукерман [2.74] исследовали связь порядкового номера фамилии данного автора в заголовке научной работы с его известностью и индивидуальным вкладом в эту работу. Заметной корреляции обнаружено не было.

У. Хирш и Дж. Синглтон [2.47] обнаружили зависимость числа авторов работы от наличия финансовой поддержки (1,38—1,60 автора) и ее отсутствия (1,16 автора).

В работе [2.48] сопоставляются бюджеты и объемы 12 ведущих химических журналов.

Д. Прайс [3.85], Л. Тондл [3.93; 3.94] и Я. Влахи [2.70] исследовали корреляцию между распределениями стран по национальному доходу и по числу публикуемых в них работ. Эта зависимость оказалась степенной (см. рис. 4.7в).

П. Винцент и А. Миракор [3.99] изучали связь между продуктивностью ученого, измеряемой ими по числу публикаций, и влияющими на продуктивность факторами — возрастом, стажем, удо-

влетворенностью работой, уровнем образования и др. (см. рис. 4.8).

Р. Ван Зельст и У. Керр [3.96] на массиве в 574 научных сотрудника двух научных учреждений США изучали корреляцию между числом публикаций ученого и отдельными чертами его характера, устанавливаемыми в ходе психологических исследований.

Р. Б. Стерджис и Ф. Клементе [2.64] изучали связь продуктивности 2120 американских социологов — выпускников различных университетов, определяя ее по числу публикаций, с престижем этих университетов. Искомая корреляция оказалась равной 0,65 в 1966 г. и 0,58 — в 1970 г.

С. Кроуфорд [1.91] изучает вопрос о связи между числом публикаций ученого и числом завязываемых им контактов с коллегами. Обнаружена значимая корреляция — 0,62. Эти данные согласуются с аналогичными данными Д. Пельца и Ф. Эндрюса [0.51].

Г. Боулт и соавт. [1.76], Н. Коган [3.66] и другие, изучая сравнительную продуктивность ученых — мужчин и женщин, определяемую ими с помощью числа публикаций, ищут влияние на эту продуктивность таких внешних по отношению к научной деятельности факторов, как материальное благосостояние семьи, наличие супруга и пр. Установлено, что научная карьера женщин в большей степени зависит от внешних обстоятельств, чем мужчин.

О. Хэгстрем [0.84] находит связь между числом публикаций ученого и его готовностью давать ответы социологам: из химиков, не имевших публикации в течение двух последних лет, на вопросы социологов ответили 90,3%; из тех, кто имел 4 публикации, — 96,2; 16 и более публикаций — только 75%. Отсюда вытекает обычно не учитываемое социологами обстоятельство, что при подсчете публикаций на основании анкет завышается доля малопродуктивных ученых и занижается доля высокопродуктивных.

С. Кроуфорд [1.91] находит, что цитирование коррелирует с завязываемыми (0,71) и уже налаженными (0,60) контактами. Кроме того, ученые, с которыми чаще консультируются коллеги, имеют наиболее читаемые статьи (0,80), т. е., ученые, стоящие в центре коммуникаций, чаще цитируются.

Р. Гарольд [1.112] исследует, как связаны между собой 14 измеримых характеристик (индикаторов) военной исследовательской лаборатории. Одну из полученных автором корреляционных матриц мы приводим в нижеследующей таблице.

Заметим, что в этой работе предпринята попытка комплексного анализа объекта — исследовательской военной лаборатории — с помощью системы индикаторов, в том числе и наукометрических.

Оценивая состояние современных наукометрических исследований, в которых наукометрические методы применялись бы сов-

Таблица
Корреляции 11 измеримых характеристик военной
лаборатории [1.112]

1	2	3	4	5	6	7	8	9	10	11
1 1,00	0,56	0,15	0,17	0,08	0,43	0,17	0,69	0,48	0,01	0,69
2	1,00	-0,12	0,04	0,08	-0,23	0,67	0,73	0,56	0,37	0,85
3		1,00	0,74	0,77	0,57	0,11	-0,06	-0,12	0,43	-0,07
4			1,00	0,63	0,20	0,27	-0,19	-0,26	0,26	0,10
5				1,00	0,34	0,07	0,19	0,26	0,35	-0,01
6					1,00	-0,43	0,21	0,00	0,00	0,05
7						1,00	0,21	0,13	0,60	0,41
8							1,00	0,80	0,14	0,67
9								1,00	0,16	0,32
10									1,00	0,11
11										1,00

Примечания: 1 — общее число статей и изобретений, 2 — число сотрудников — профессионалов-исследователей, 3 — средний их стаж (в годах), 4 — средний стаж их работы в данной лаборатории, 5 — средний возраст, 6 — средний оклад, 7 — число уволенных или перемещенных сотрудников, 8 — число аспирантов, 9 — число семинаров, проводимых в лаборатории, 10 — число сотрудников, участвующих в заседаниях технического общества, 11 — объем ассигнований на лабораторию.

местно с ненаукометрическими, следует сказать, что и такие исследования находятся пока в самом зачаточном состоянии. Между тем именно данное направление наукометрических исследований представляется особенно перспективным.

С этой точки зрения определенный интерес представляет работа Д. Прайса [1.146], в которой он ставит задачу разработки единой системы индикаторов науки. В качестве необходимого этапа он выдвигает задачу максимального упрощения, упорядочения существующих сегодня необозримых данных, полученных в результате использования множества самых разных индикаторов. Прайс, прежде всего, предлагает «стратегию последовательной аппроксимации». Встречающиеся распределения следует аппроксимировать по возможности единым образом, используя наиболее простые зависимости. В качестве примера он приводит экспоненты роста науки, применение которых, по его мнению, облегчает описание отклонений от них (так называемые «пульсации» науки, о которых говорят сам Прайс, С. Г. Раутман [3.29] и др.).

Для упрощения описания всевозможных табличных данных по стационарным наукометрическим распределениям Прайс [1.146—1.148] предлагает некоторые формальные процедуры. Эти процедуры не связаны с какими-либо конкретными аппроксимациями,

посят в известной степени схоластический характер и потому, на наш взгляд, вряд ли будут полезны. Однако сам курс на использование количественных закономерностей (аппроксимаций в терминологии Прайса) для унифицированного описания объектов наукометрического анализа, конечно же, взят совершенно верно (см. гл. 1, § 4).

Далее Прайс предлагает эмпирическим путем выявлять связи между различными индикаторами. Если найдены соотношения между разными индикаторами, то эти соотношения вкупе с самими индикаторами и образуют систему индикаторов. Выше уже говорилось о корреляции между различными индикаторами науки. Прайс же выделяет три группы индикаторов — индикаторы численности научного персонала (число ученых разных категорий и т. п.), научного выхода (число публикаций, патентов, ссылок и пр.) и экономические индикаторы (ассигнования на науку и др.). Первые две группы индикаторов связаны известными соотношениями — Лотки, цитатными и др. Эти распределения позволяют, утверждает Прайс [1.150; 1.146], по данным о числах публикаций, ссылок и т. д. вывести эквивалентные числа ученых, работающих в данной области, т. е. связать эти индикаторы определенными соотношениями.

Что же касается связи индикаторов численности научного персонала и экономических, то здесь Прайс ссылается на данные типа представленных на рис. 4.7.

Прайс полагает, что он «очертил контуры» [1.146, с. 49] единой системы индикаторов науки. Констатируя, что Прайс работает в нужном направлении, следует все же сделать несколько замечаний.

Во-первых, связи между индикаторами по самой своей природе стохастические, что существенно усложняет анализ.

Во-вторых, пока что выявлены связи между весьма немногими индикаторами науки. Далеко не во всех случаях установлена степень связанности. Между тем, например, связь наукометрических и экономических индикаторов выявлена пока лишь для весьма больших исследуемых массивов, скажем целых стран. Для меньших массивов картина, возможно, менее однозначна.

В-третьих, до сих пор связи между индикаторами выявляются средствами гауссовой статистики. В применении к негауссовым по своей природе объектам наукометрического анализа это некорректно и существенно искажает действительные корреляции.

И последнее. Прайс ставит задачу разработки единой системы индикаторов науки. Это, конечно, правильно. Однако решение этой задачи изолированно от задачи создания единой системы латентных переменных науки лишено смысла. Не будучи привязаны

к латентным переменным, индикаторы сами по себе ничего не говорят об объектах исследования.

Таким образом, *перспективы комплексного применения наукометрических и ненаукометрических методов связаны с опорными исследованиями индикаторов науки*. В ходе этих исследований формируются *метрические модели*. Для этого конструируются операциональные референты латентных переменных — операциональные латентные переменные, представляющие собой группы определенным образом «взвешенных» индикаторов. Вот эта система операциональных латентных переменных, постоянно корректируемая под давлением практики, и соотносится в ходе опорных исследований с системой индикаторов. В принципе полное определение связи индикатора с операциональной латентной переменной предполагает определение вероятностного спектра возможных значений этой переменной для каждого значения индикатора. В первом приближении достаточно знать, как *коррелируют* распределения исследуемых объектов по значениям индикаторов и операциональной латентной переменной, как коррелируют индикатор и его операциональный референт.

Вокруг проблемы применимости наукометрических методов на протяжении вот уже ряда лет ведется полемика. Одна сторона (Дж. Коул и др.) полагает, что наукометрические методы применимы в принципе лишь для исследования достаточно больших популяций ученых, но не могут считаться надежными при рассмотрении творчества отдельных ученых. Приверженцы наиболее крайних позиций этой точки зрения утверждают, что наукометрические методы неприменимы вообще. Другая сторона (Д. Прайс и др.) полагает, что, напротив, наукометрические методы применимы и для оценки вкладов отдельных ученых. Наиболее горячие сторонники этой точки зрения утверждают, что наукометрические методы применимы всегда и везде.

Постараемся разобраться в этом вопросе. В конечном счете важна *достоверность* получаемых наукометрией результатов измерения. Эта достоверность имеет две компоненты. Первая — *содержательная достоверность*, характеризующая то, в какой мере можно полагаться на значение латентной переменной, определяемое по косвенно его характеризующему значению индикатора. Вторая — *статистическая достоверность*, определяемая статистической природой измеряемых случайных величин.

Содержательная достоверность измеренного значения латентной переменной зависит, прежде всего, от «качества» ее операционального референта. Последнее прямым оценкам не поддается. Мы можем только, опираясь на практику, переходить от «худшей» операциональной латентной переменной к «лучшей».

Если же, как бывает на каждом конкретном этапе опорных исследований, латентная переменная отождествлена с определенным ее операциональным референтом, *содержательная достоверность в принципе поддается оценке*. На нее влияют два фактора: во-первых, объемы используемых в опорных исследованиях выборок, определяющие статистическую достоверность расчетных значений коэффициентов корреляций индикаторных и (операциональных) латентных распределений; во-вторых, величина этих коэффициентов корреляций (чем больше коэффициенты корреляции, тем выше искомая содержательная достоверность).

Статистическая достоверность измеренного значения латентной переменной определяется объемами выборок, используемых в ходе текущих наукометрических измерений (чем больше эти выборки, тем выше статистическая достоверность).

Итак, в приближении, в котором латентные переменные отождествляются с их операциональными референтами, *содержательная и статистическая достоверности измеряемых значений латентных переменных в принципе поддаются количественной оценке средствами математической статистики*.

Математическая статистика для ненулевых объемов выборок дает, вообще говоря, ненулевые достоверности расчетных значений величин. Таким образом, *в общем случае измеряемые наукометрией значения латентных переменных имеют ненулевые достоверности*, разные для различных измерений.

Вернемся к полемике относительно проблемы применимости наукометрических методов. Проведенный анализ показывает, что на качественном уровне, на каком эта полемика ведется, она носит в известной мере схоластический характер. В каждом конкретном случае результат наукометрического измерения имеет конкретную достоверность, уровень которой и определяет применимость наукометрического метода. В этом плане *исследование отдельного ученого ничем не отличается от исследования их популяции: в обоих случаях необходимы конкретные расчеты достоверностей*. И только эти расчеты могут показать, какие наукометрические индикаторы и для решения каких исследовательских задач применимы. Существенную роль в решении этих вопросов играют объемы выборок и «дробности» индикаторов (см. § 1).

Выше были обрисованы *перспективы комплексного применения наукометрических и ненаукометрических методов анализа науки*, тесно связанные, как оказалось, с развитием *опорных исследований*. Чтобы эти перспективы реализовались, необходимы: во-первых, проведение *опорных исследований*, т. е. систематических, постоянно корректируемых практикой исследований связи индикаторов и латентных переменных; во-вторых, *негауссова статисти-*

стика, на которую только и могут опираться наукометрические исследования, включая опорные. Пока мы не научимся вычислять средствами негауссовой статистики средние коэффициенты корреляций и прочее, все очерченные здесь перспективы реализованы быть не могут.

3. МЕСТО И РОЛЬ НАУКОМЕТРИИ СРЕДИ НАУКОВЕДЧЕСКИХ ДИСЦИПЛИН (взгляд на наукометрию «извне»)

Суммируем выводы относительно места и роли наукометрии в науковедении. Понятно, что этого нельзя сделать без хотя бы краткого рассмотрения задач и основных черт науковедения в целом.

Основной задачей науковедения является создание научной базы для реорганизации науки в направлении ее интенсификации. Решить эту задачу необходимо в ближайшие десятилетия, так как именно в этот временной интервал наука перейдет, как показывают данные по временной динамике ее параметров, от стадии преимущественно экстенсивного роста к стадии преимущественно интенсивного роста. Решение этой задачи требует, чтобы науковедение опиралось на количественные методы.

Можно выделить три основные науковедческие количественные дисциплины. *Наукометрия* поставляет количественные закономерности и (в числе других эмпирических науковедческих дисциплин) результаты измерения различных параметров объектов науковедческого анализа. *Математическое моделирование науки*, используя выявленные наукометрией закономерности, описывает результаты измерения математическими зависимостями (моделями). *Теория принятия решений* (исследование операций, теория игр и др.), используя эти результаты измерения и математические зависимости, ищет значения параметров объектов исследования, оптимизирующие эти объекты в направлении, задаваемом целями данного конкретного исследования.

Объект исследования науковедения — научная деятельность — имеет особенности, отличающие его от объектов исследования других наук. Проведенный в гл. 1 анализ этих особенностей, опирающийся на представления о негауссовости научной деятельности и социальных явлений в целом, привел к следующим выводам относительно того, каким должен быть анализ науки.

1. Анализ объектов науковедческого исследования должен быть существенно более многофакторным, чем анализ естественных явлений. Реально мы имеем дело с системой индикаторов науки, соотносимой в *опорных исследованиях* с системой

ее латентных переменных. Для данного объекта науковедческого анализа системой существенных факторов выступает часть системы латентных переменных науки. Связь между отдельными факторами здесь значительно слабее, чем в естественных науках, и потому острее чувствуется необходимость выявления группы латентных переменных-факторов, существенных для определения поведения данного объекта. Математические средства многомерного анализа позволяют в принципе выделить группу (группы) факторов, определяющих поведение данного объекта. Поскольку разные факторы, разные латентные переменные науки «приписаны», вообще говоря, к разным науковедческим дисциплинам, постольку требование многофакторности описания науки означает требование *комплексности* науковедческого исследования, требование синтеза науковедческих дисциплин.

Размытость связей между отдельными индикаторами и латентными переменными науки («все связано со всем», но связано слабо) требует больших и более координированных усилий исследователей разных специальностей, чем в естественных науках.

Комплексный характер носят по самой своей природе и *опорные исследования* по соотношению систем индикаторов и латентных переменных. Без придания таким опорным исследованиям относительно самостоятельного статуса невозможен, на наш взгляд, сколько-нибудь продуктивный количественный анализ науки.

Комплексное исследование науковедческих проблем должно сочетаться с развитием отдельных науковедческих дисциплин. Науковедческие организации, видимо, должны сочетать две наложенные друг на друга сетки — дисциплинарную и проблемную. Один и тот же исследователь должен иметь возможность сочетать свою работу в относительно стационарном дисциплинарном подразделении с работой в относительно мобильном проблемном подразделении, создающемся для решения конкретной проблемы и распускаемом, когда проблема исчерпает себя.

2. Большая, чем в естественных науках, субъективность результатов анализа, т. е. большие трудности с разделением объективно наблюдаемого факта и влияния наблюдателя, обязывает при исследовании науки прилагать большие усилия к развитию объективных методов анализа, дающих *воспроизводимые* результаты. Другими словами, развивая целостную систему индикаторов науки, которую требует ее многофакторность, необходимо особое внимание уделять развитию *воспроизводимых индикаторов*. Науковедение как научная база интенсификации науки должно опираться на «твердые» индикаторы, дающие воспроизводимые результаты.

3. Количественный анализ науки должен базироваться, во-

первых, на *открытых шкалах* и, во-вторых, на *негауссовой математической статистике*. Научная деятельность вероятностна по своей природе, и потому ее количественный анализ должен опираться на математическую статистику. Научная деятельность негауссова и потому ее следует описывать с помощью негауссовой статистики.

Научную деятельность, направленную на производство нового знания, в большей мере, чем многие другие виды человеческой деятельности, характеризует *творческий характер*. Это приводит к тому, что *негауссовость научной деятельности выражена сильнее по сравнению с другими видами человеческой деятельности*. Поэтому особенности количественного анализа научной деятельности в отличие от количественного анализа в естественных науках выражены сильнее, чем это характерно для исследований других видов человеческой деятельности, например для экономики. Поэтому актуальность практических следствий для количественного анализа, вытекающих из негауссовости человеческой деятельности, для науковедения выражена особенно сильно.

Итак, стоящие перед науковедением задачи и особенности объектов анализа предъявляют к нему вполне определенные требования. Посмотрим, что в этих требованиях относится к собственно наукометрии, каковы в общих чертах ее задачи и черты как науковедческой дисциплины.

Многофакторность науковедческого анализа и наукометрия

Применение наукометрических индикаторов изолированно от индикаторов другой природы имеет ограниченное значение.

В основу наукометрических индикаторов (число журналов, публикаций, ссылок и др.) положен научный текст. *Наукометрический анализ — это статистический анализ научных текстов.*

Научный текст — это продукт труда ученого, овестьвленный предмет его деятельности. Поэтому наукометрические индикаторы — предметны по своему содержанию, а наукометрия измеряет параметры предметной структуры науки.

Разные науковедческие дисциплины исследуют, вообще говоря, разные *специальные структуры науки* — предметную, организационную, психологическую, социально-психологическую и т. д., образующие в своем единстве *общую структуру науки*. Основная количественная характеристика каждой специальной структуры — степень ее *структурированности*. Так, речь может идти о степени *централизованности* организационной структуры, о степени *координированности* проблематики исследований ученых и

научных коллективов и т. д. Каждая специальная структура науки имеет свой оптимальный для текущего момента уровень структурированности, определяемый как внутренними, так и внешними по отношению к науке в целом факторами. Существование этих оптимумов проявляется в виде характерных для каждой из специальных структур противоречий — между централизованностью и децентрализованностью организационной структуры, между координированностью и автономностью проблематики исследований и т. д.

Существует и оптимум структурированности общей структуры науки, т. е. оптимальный для текущего момента относительный *сдвиг* специальных структур: организационная структура не может совпадать с предметной и психологической и т. д. Оптимум структурированности общей структуры науки и, следовательно, и оптимальные относительные сдвиги специальных структур определяются как внутренними, так и внешними по отношению к науке факторами.

Необходимость интенсификации науки, т. е. ее оптимизации, диктует необходимость поиска оптимальных степеней структурированности специальных структур науки и оптимальных сдвигов между ними. Таким образом, перед науковедением стоит задача систематических замеров параметров специальных структур науки, определения степеней их структурированности и величин сдвигов между ними.

Перед наукометрией в связи с этим стоит задача систематических замеров параметров предметной структуры научной деятельности и определения ее структурированности.

Таким образом, на наукометрию приходится, казалось бы, довольно частный сектор науковедческих задач — воспроизводимое измерение параметров одной из специальных структур науки. Следует учесть, однако, что именно предметная структура науки является основной, характеризующей научную деятельность именно как *научную* и никакую иную.

Необходимость воспроизводимого измерения параметров науки и наукометрия

В условиях становления наукометрии как самостоятельной науковедческой дисциплины первоочередной ее задачей было выявление *количественных закономерностей науки*. Эта задача остается и сегодня, однако перестает быть единственной. Развитие наукометрии на проведение совместно с другими науковедческими дисциплинами комплексных проблемных исследований диктует и разворот наукометрии на *воспроизводимое измерение*

параметров отдельных объектов науковедческого анализа, к выявлению того, что характеризует один такой объект в отличие от другого. При этом сама возможность такой индивидуализации объектов обеспечивается существованием количественных закономерностей, на основе которых и оказывается возможным введение единой системы параметров, в качестве которых выступают параметры количественных закономерностей. Значения этих параметров и определяются в процессе воспроизводимого измерения, индивидуализирующего отдельные такие объекты.

Измерение параметров научной деятельности, осуществляемое наукометрией в числе других науковедческих дисциплин, представляет интерес не только и не столько само по себе, сколько как материал для последующего анализа средствами математического моделирования науки и теории принятия решений. Таким образом, наукометрия вдвойне является фундаментом количественного анализа научной деятельности. Во-первых, она составляет количественные закономерности, на которых базируется весь такой анализ. Во-вторых, она в числе других науковедческих дисциплин дает сырье для последующего количественного анализа, отвечая конкретно за воспроизводимое измерение параметров ее предметной, т. е. собственно научной, структуры.

Необходимость базирования количественного анализа науки на негауссовой математической статистике и наукометрии

К выводу о необходимости перехода от гауссовой статистики к негауссовой позволили прийти именно наукометрические исследования, так как наукометрия «отвечает» в науковедении за воспроизводимое измерение.

Этот вывод имеет значение для всего науковедения. Неустойчивость, слабая воспроизводимость получаемых сегодня при исследовании науки количественных данных не в последнюю очередь объясняются применением некорректной здесь гауссовой статистики. Использование негауссовой статистики приведет к более устойчивой, более воспроизводимой картине явлений. Существующие на сегодняшний день количественные данные нуждаются в соответствующем их пересмотре и коррекции.

Исследования научной деятельности, как и другие социальные исследования, должны быть многовариантными, с тем чтобы выбор оптимального решения производился на основе анализа альтернатив. Многовариантным, в частности, должен быть и анализ перспектив развития наукометрии. В какой мере автору удалось осветить возможные направления развития наукометрии — судить читателю.

ПОСЛЕСЛОВИЕ

Анализ наукометрии как науковедческой дисциплины носит в настоящем исследовании в основном *теоретический* характер. Между тем по своей направленности наукометрия — *прикладная* наука. Как же обстоит дело с приложениями наукометрии?

Ситуация здесь сложилась сложная. У наукометрии имеются свои сторонники и противники. Наиболее горячие ее приверженцы убеждены, что результаты наукометрических измерений должны уже сейчас в возможно более широких масштабах внедряться в практику управления научными исследованиями. По мнению же противников наукометрии, ее методы пока не следует применять вовсе, поскольку возможный положительный эффект от их применения гораздо меньше, как они утверждают, отрицательных побочных последствий. Наиболее ортодоксальные из них полагают, что уже само применение наукометрических индикаторов (числа публикаций, ссылок и др.) *предполагает* неверные выводы.

Наиболее полярные мнения относительно возможностей применения наукометрических индикаторов в равной мере базируются, как нам представляется, на игнорировании основных положений *теории измерения*, разделом которой в теоретическом плане является наукометрия. В этой теории выделяются переменные двух видов — *индикаторы* (непосредственно измеряемые величины — число публикаций, ссылок и т. д.) и *латентные переменные* (которые мы хотим измерить — продуктивность ученого, его научный вклад и т. д.). Индикаторы лишь *косвенно* характеризуют латентные переменные, связь их носит вероятностный характер, т. е. данному значению индикатора отвечает не одно значение латентной переменной, а целый спектр их, описываемый вероятностным распределением. Таково положение дел не только в наукометрии, но и в психологии, социологии и других науках. *Метрические модели* связывают значения индикаторов со значениями латентных переменных (модели Сэффира, Гуттмана, Ликерта,

Хорста, Терстоуна, Лазарсфельда и др.). Однако проблема связи индикаторов с латентными переменными остается центральной в теории измерения. Остается центральной она и в наукометрии. В настоящей монографии развиваются представления об *опорных исследованиях*, в ходе которых вырабатываются эти метрические модели, постоянно корректируемые под давлением практики.

В наукометрических работах имеют место случаи отождествления индикаторов с латентными переменными (число публикаций с продуктивностью ученого, число ссылок с его научным вкладом и т. д.). При этом игнорируется тот факт, что число публикаций, ссылок и прочее — лишь *косвенные* характеристики научной деятельности. Именно такая *абсолютизация* индикаторов и приводит зачастую к неверным выводам. Это в большей степени характерно, например, для известного американского наукометра Д. Прайса. Эти ошибки Прайса и других хорошо известны и получили должное освещение в литературе. Обсуждаются они и в настоящей монографии.

Противников наукометрии характеризует другая крайность. Понимая, что наукометрические индикаторы нельзя отождествлять с латентными переменными, они считают, что эти индикаторы *никак не характеризуют* латентные переменные.

Истина как всегда посередине. Наукометрические индикаторы, конечно, характеризуют латентные переменные, но характеризуют их лишь в *определенной степени*. Эта степень может быть выявлена лишь в специальных опорных исследованиях.

Наукометрические индикаторы в этом отношении не отличаются от индикаторов, используемых в других социальных науках — психологии, социологии и т. д. Ответы респондента на вопросы социолога весьма косвенно характеризуют действительное его (респондента) отношение к затрагиваемым проблемам. Дело не только в том, что респондент может сознательно дезинформировать исследователя, но и в том, что у респондента зачастую не выработалось определенного отношения к этим проблемам. Это хорошо известно, и именно по этой причине исследователи обращаются сегодня к теории измерения и создают различные метрические модели.

Почему же мы не видим таких горячих противников психологических, социологических и других индикаторов, каких имеют индикаторы наукометрические? Ответ заключается, на наш взгляд, в *кажущейся доступности*, внешней легкости наукометрического измерения. Неспециалисту трудно провести социологическое или психологическое измерение. Некоторые же наукометрические индикаторы (число ученых, журналов, публикаций и др.) лежат «на поверхности». Создается впечатление, что с ними мо-

жет работать даже неспециалист, получая необходимые выводы. Однако выводы, не основанные на знании теории измерения и природы наукометрических индикаторов, чаще всего ошибочны. Таким образом, критики заслуживают не *наукометрические индикаторы* сами по себе, а их легковесная трактовка.

Противники наукометрии, оставляя за ней право на жизнь в форме *исследований*, отказывают ей сегодня в праве на жизнь в форме *практических приложений*. Такая позиция, однако, оторвана от реальности, так как практики все же используют наукометрические индикаторы, ориентируясь на их внешнюю доступность, но, лишённые теории, используют неграмотно. Авторский гонорар рассчитывается по объему публикации, план научным работникам обозначается в авторских листах и т. д. В результате увеличение количества и объема публикаций зачастую становится едва ли не главным стимулом деятельности не только отдельных ученых, но и целых научных коллективов [0.48, с. 114].

Кроме того, растет поток *стихийных* наукометрических работ, выполняемых учеными — специалистами в области конкретных дисциплин (физики, химии и т. д.), не обладающими специальными наукометрическими знаниями. Получаемые ими зачастую неадекватные выводы используются руководителями соответствующих научных организаций, также отнюдь не являющимися специалистами по наукометрии.

Таким образом, *наукометрические методы уже нашли практическое применение и этот процесс необратим*. В этих условиях основным должен быть не вопрос о *возможности* применения наукометрических методов, а проблема сведения к минимуму потерь от стихийного их применения. Другими словами, *стихийные наукометрические исследования должны быть поставлены на профессиональную основу*. Неотложной задачей является в этих условиях создание наукометрии как институционализированной науковедческой дисциплины со своими эмпирической базой, теорией, квалифицированными кадрами.

Развивая наукометрию как относительно самостоятельную дисциплину, следует, однако, постоянно помнить, что она занимает вполне определенное место в науковедении, определяемое *прикладным* ее характером. Как и в других социальных науках, применение математических методов в науковедении ведется на трех уровнях. Первый уровень — *уровень измерения* — ставляет значения переменных и количественные закономерности, характеризующие объекты анализа. Второй уровень — *математическое моделирование* (описание). Здесь, используя данные измерения и закономерности, выявленные на первом уровне, переменные связывают математическими зависимостями. Третий уровень — *уро-*

вень принятия решений (методы исследования операций, сетевые и др.), на котором, используя результаты измерения и математические зависимости между переменными, определяют значения последних, оптимизирующие объекты анализа в направлении, задаваемом конкретными целями прикладных науковедческих исследований.

Наукометрия относится к уровню измерения, отвечая за наиболее воспроизводимые индикаторы науки. Наукометрические методы, как и другие методы измерения науки, следует оценивать, прежде всего, с точки зрения последующего их включения в методы математического моделирования и принятия решения. Между тем, сегодня три указанных уровня количественного анализа науки взаимодействуют друг с другом совершенно недостаточно. Исследователи, работающие на первом уровне, зачастую не задумываются, куда и зачем идут получаемые ими данные. Исследователи же, работающие на втором и третьем уровнях, часто не вникают в природу используемых ими данных измерения. Вот эта разобщенность исследователей разных специальностей и привела к тому, что результаты измерения науки в той форме, в какой они сегодня поставляются на первом уровне, и современные математические методы, используемые на втором и третьем уровнях, несовместимы. Математика, используемая при математическом моделировании науки и в методах принятия решений — алгебра, дифференциальное и интегральное исчисление и т. д., — требует аддитивности переменных. Кроме того, используемая здесь математическая статистика является статистикой моментов, т. е. гауссовой. Между тем, одна часть результатов измерения науки сегодня негауссова, а другая — неаддитивна. Именно, при использовании открытых шкал результаты измерения негауссовы, при использовании закрытых шкал — неаддитивны. Таким образом, развитие методов измерения науки и, в частности, наукометрической статистики и, во-вторых, с отказом от закрытых шкал, что означает в совокупности коренную перестройку всего комплекса современных методов количественного анализа научной деятельности. Решение этой сложной проблемы требует координации усилий исследователей науки разных специальностей.

ГРАФИКИ

1. НЕСТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

РИС. 1а. Рост населения Земли и числа ученых на ней во времени

1 — численность ученых, 2 — численность населения. Координаты — полулогарифмические, прямая соответствует экспоненте. Население Земли растет в первом приближении по экспоненте с периодом удвоения, равным примерно 35 лет. Число ученых удваивается примерно каждые 12,5 лет. Если бы такое положение вещей сохранилось, то примерно через 170 лет все население Земли, включая грудных младенцев, состояло из ученых. Абсурдность этого доказывает неизбежность предстоящего замедления темпов роста численности ученых. Должен замедлиться и рост других экстенсивных индикаторов объема науки (см. рис. 1.16). На этом обстоятельстве базируется тезис о переходе науки от преимущественно экстенсивного развития к преимущественно интенсивному развитию

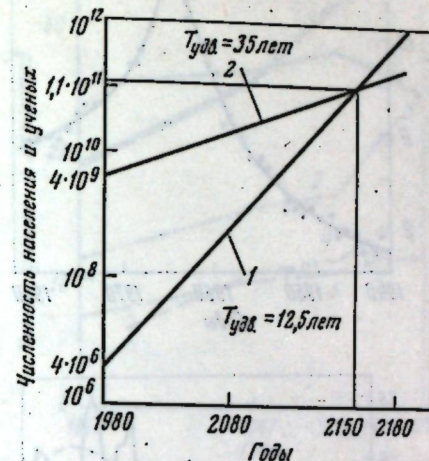


РИС. 1.16. Логистическая кривая [1.58]

1 — рост с насыщением, 2 — чистый экспоненциальный рост, 3 — положение сегодня. Таков, согласно концепции экспоненциально-логистического роста науки Д. Прайса, в первом приближении рост во времени экстенсивных индикаторов объема науки

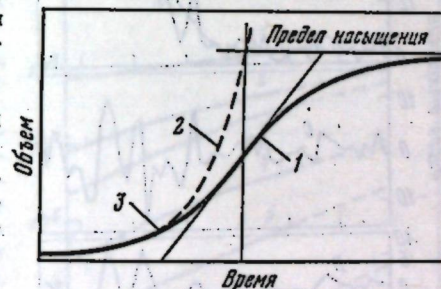
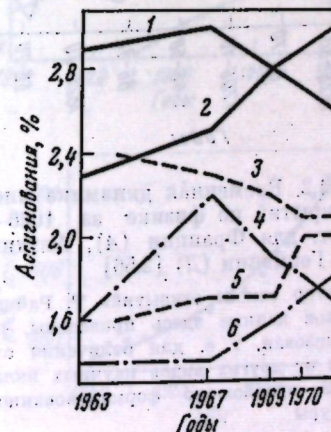


РИС. 1.1в. Временная динамика ассигнований на исследования в отдельных странах (1963—1971 гг.), % от валового национального продукта [1.154]

1 — США, 2 — СССР, 3 — Великобритания, 4 — Франция, 5 — ФРГ, 6 — Япония. В некоторых промышленно развитых странах заметна тенденция к снижению темпов роста ассигнований на науку. Этот и подобные факты служат симптомами ожидающегося замедления темпов экстенсивного роста науки



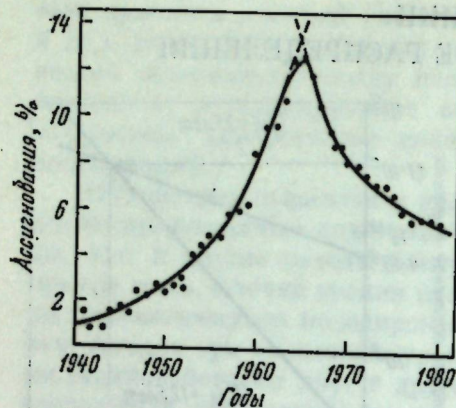


РИС. 1.1г. Временная динамика ассигнований (1940—1980 гг.) на науку, % от Федерального бюджета США [1.146]. Заметна тенденция к снижению ассигнований на науку (см. рис. 1.1а—в).

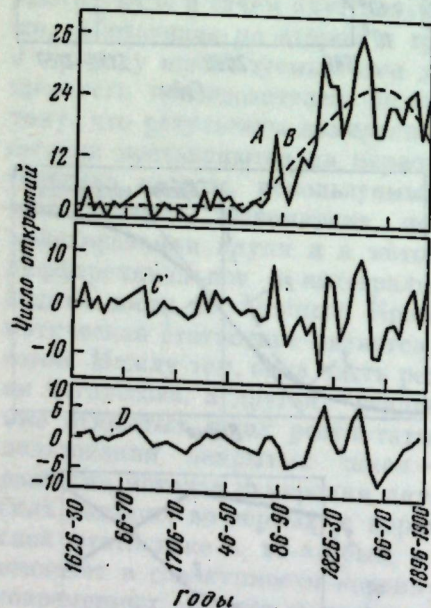


РИС. 1.2. Временная динамика числа открытий по физике за 1626—1900 гг. для Франции (А), Англии (С) и Германии (Д) [2.56].

Индикатор «число открытий» Т. Райновым, чьи данные здесь приведены, не формализован, т. е. для отличия открытия от других видов научных вкладов не предложено формализованной процедуры.

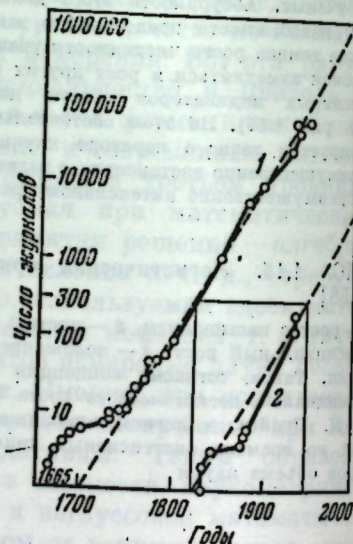


РИС. 1.3. Рост общей массы журналов [1.58]

1 — журналы, 2 — РЖ. Координаты — полулогарифмические, прямая соответствует экспоненте. Данные относятся ко всем журналам независимо от того, выходили они в настоящее время, а также и к журналам, которые имеют только научные разделы.

РИС. 1.4а. Временная динамика числа ученых и инженеров (на 10 000 человек населения), занятых в исследованиях, в отдельных странах [1.154]

1 — СССР, 2 — США, 3 — Япония, 4 — ФРГ, 5 — Франция. Для некоторых развитых капиталистических стран заметно снижение темпов роста этого индикатора объема науки (ср. рис. 1.1)

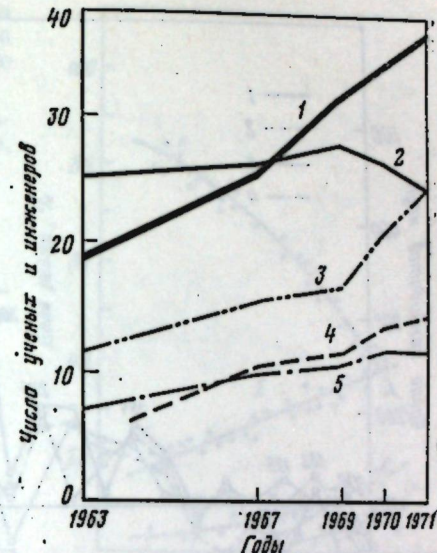


РИС. 1.4б. Временная динамика числа научных работников и населения США [1.59]

1 — население, 2 — выпускники школ, 3 — выпускники высшей школы, 4 — выпускники колледжей, 5 — ученые и инженеры, 6 — бакалавры. Координаты — полулогарифмические, прямая соответствует экспоненте. Чем выше квалификация, тем выше темпы роста. Заметна также тенденция приближения объема групп к «потолку», расположенному параллельно кривой роста населения. Эти данные говорят об опережающих по темпам роста науки по сравнению с темпом роста населения, а также о замедлении экстенсивного роста науки (см. рис. 1.1)

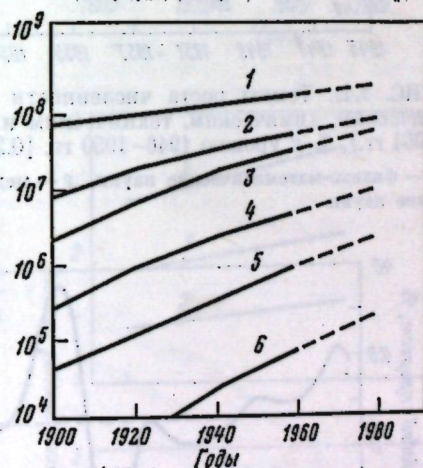
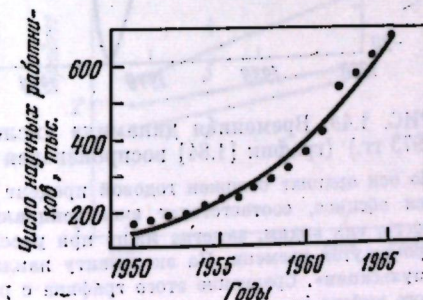


РИС. 1.4в. Временная динамика роста числа научных работников в СССР [1.46]



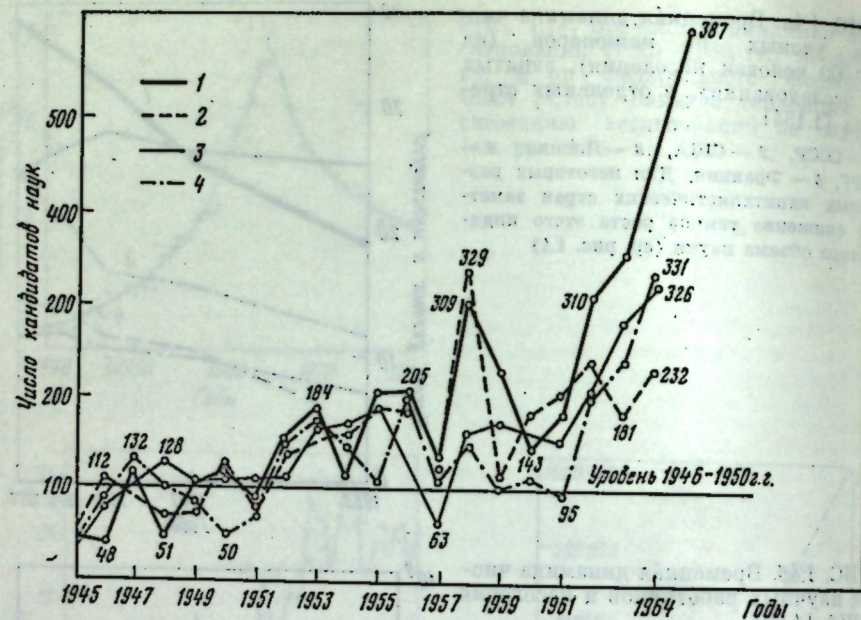


РИС. 1.4а. Темпы роста численности кандидатов наук по физико-математическим, химическим, техническим и медицинским наукам в СССР (1945-1964 гг.), % к уровню 1946-1950 гг. [0.23]

1 — физико-математические науки, 2 — медицинские, 3 — технические, 4 — химические науки

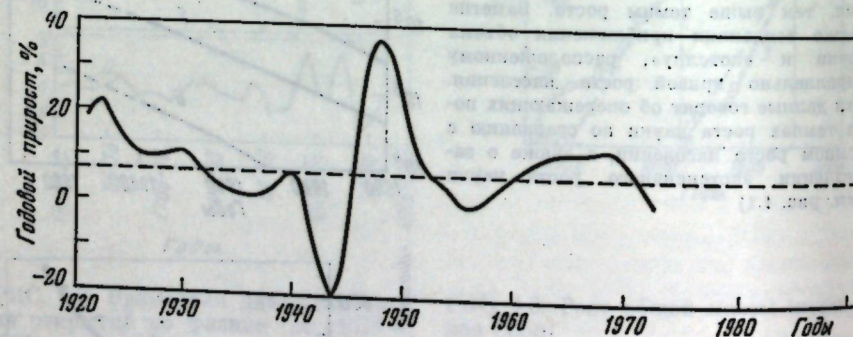


РИС. 1.4б. Временная динамика численности докторов наук США (1920-1973 гг.) (график [1.84] воспроизведен по [1.146])

По оси ординат отложен годовой процент прироста, так что прямая, параллельная оси абсцисс, соответствует экспоненциальному росту. Экспоненциальная зависимость, как видим, заметна лишь при рассмотрении данных за достаточно большой промежуток времени. На экспоненту накладываются, как говорит Д. Прайс [1.146], «пульсации». Сравнение этого графика с рис. 1.4а демонстрирует значение правильного выбора координат

РИС. 1.5а. Развитие коллективности в авторство публикаций по химии (1910-1960 гг.) (график [1.22] по данным [1.58])

По оси ординат откладывается доля работ, имеющих данное число соавторов, % от общего числа работ.

1 — один автор,
2 — два,
3 — три,
4 — четыре автора.

Эти и подобные данные (см. рис. 1.5б — в) говорят о тенденции роста во времени числа соавторов научной публикации

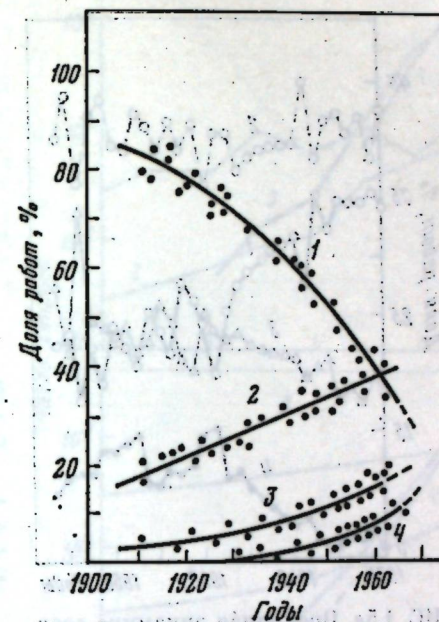
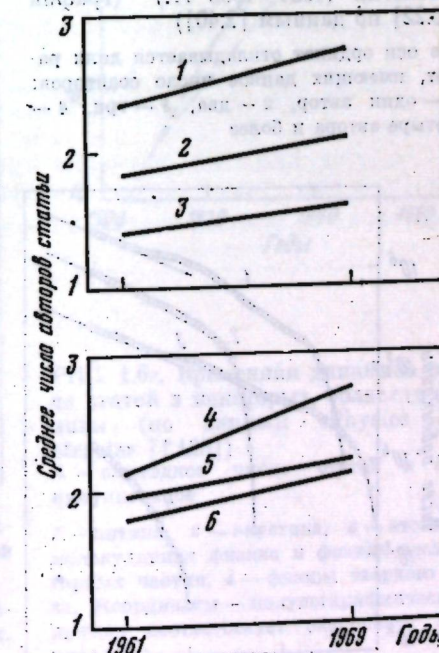


РИС. 1.5б. Временная динамика среднего числа авторов на статью для отдельных областей физики [2.70]

1 — экспериментальная физика,
2 — прикладная,
3 — теоретическая,
4 — ядерная,
5 — техника физического эксперимента,
6 — физика твердого тела.

Вообще говоря, при анализе наукометрических данных следует брать не среднее, а медиану. Однако в данном случае соответствующее стационарное распределение (см. рис. 2.26) имеет достаточно большое значение показателя цитированности, так что зависимость среднего от объема выборки несущественна (см. рис. 2.6)



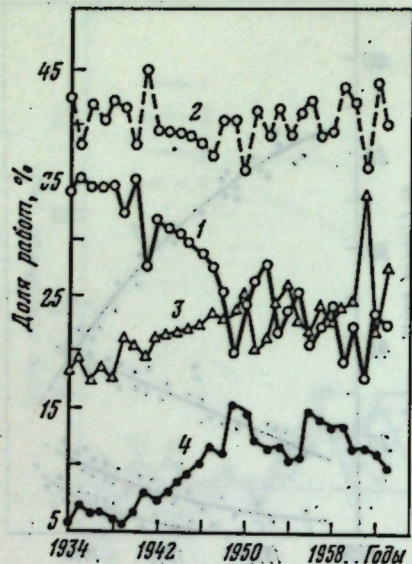


РИС. 1.5а. Временная динамика соавторства в научных публикациях по биологии (1934–1963 гг.) (график [0.22] по данным [2.40])

По оси ординат откладывается доля работ, имеющих данное число соавторов: 1 — один автор, 2 — два, 3 — три, 4 — четыре автора и более

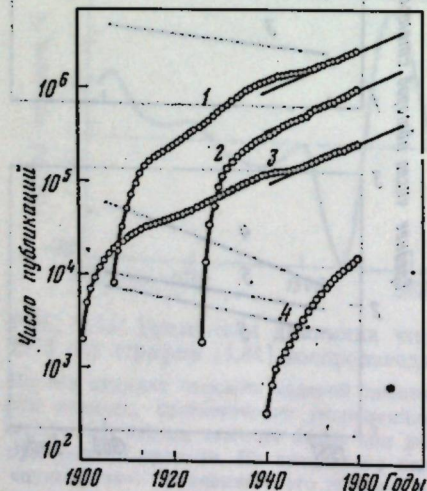


РИС. 1.6а. Временная динамика общего числа публикаций в реферативных журналах по разным дисциплинам [1.58]

1 — химия, 2 — биология, 3 — физика, 4 — математика. Координаты — полулогарифмические, прямая соответствует экспоненте. После начального периода ускоренного роста устанавливается постоянный экспоненциальный рост с периодом удвоения примерно в 15 лет (см. рис. 1.6а)

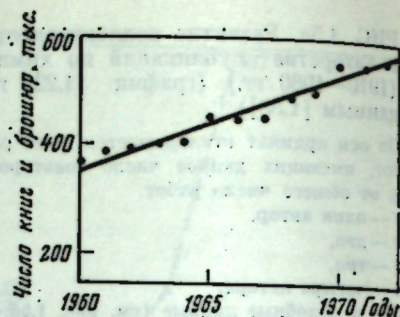


РИС. 1.6а. Временная динамика числа научных книг и брошюр в мире (1960–1972 гг.) (по данным [1.42])

На этом относительно небольшом временном интервале допустима, как видим, линейная аппроксимация. Эти и подобные данные (см. рис. 1.6б — м) свидетельствуют о логистическом в первом приближении характере роста во времени числа научных публикаций (см. рис. 1.16). На графиках 1.6а — м представлены различные участки логисты: соответствующие экспоненциальному росту (см. рис. 1.6б — г, е — ж, з, м); для небольших временных интервалов они могут быть описаны приблизительно линейной аппроксимацией (см. рис. 1.6а, б, м); соответствующие первоначальному ускоренному росту (см. рис. 1.6б, ж, з, м); соответствующие замедлению экспоненциального роста (см. рис. 1.6ж — к, м). Вся логиста фигурирует на рис. 1.6ж, з, м

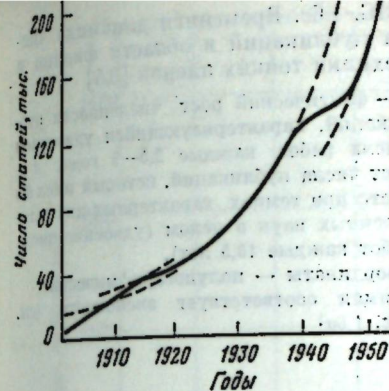


РИС. 1.6а. Общее число публикаций в «Физическом реферативном журнале» с 1 января 1900 г. (график [3.40] воспроизведен по [1.59])

Сплошная линия дает общее число публикаций, пунктиром показаны экспоненты аппроксимации. Параллельные экспоненты демонстрируют влияние войны (см. рис. 1.6а)

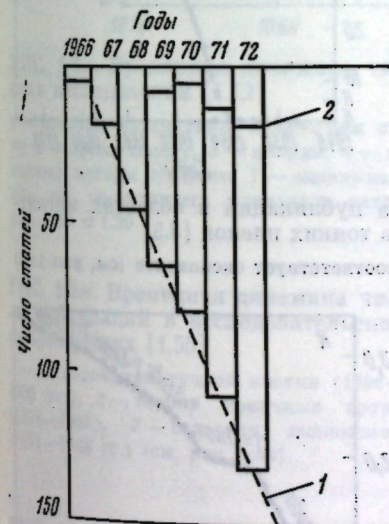


РИС. 1.6а. Временная динамика числа публикаций по физике лазеров [1.123]

1 — общее число статей, 2 — число статей по прикладной физике лазеров (см. рис. 1.6а)

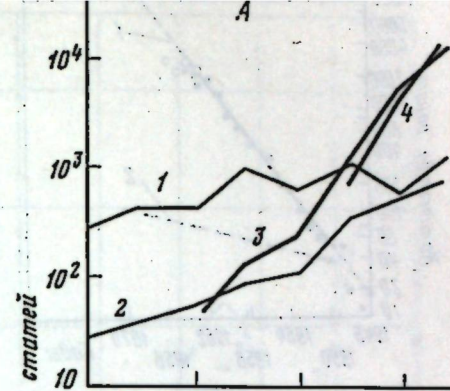


РИС. 1.6а. Временная динамика числа статей в некоторых областях физики (по данным «Physics Abstracts» [1.125])

А — ежегодное число статей и В — кумулятивное

1 — оптика, 2 — акустика, 3 — атомная, молекулярная физика и физика элементарных частиц, 4 — физика твердого тела. Координаты — полулогарифмические, прямая соответствует экспоненте (см. рис. 1.6а)

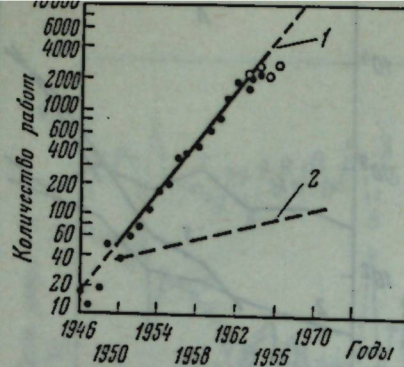


Рис. 1.6б. Временная динамика числа публикаций в области физики и техники тонких пленок [3.5]
1 — фактический рост численности публикаций, характеризующийся удвоением числа работ каждые 2,5—3 года; 2 — рост числа публикаций, который имел бы место при темпах, характерных для естественных наук в целом (удвоение числа работ каждые 13,5 лет).
Координаты — полулогарифмические, прямая соответствует экспоненте (см. рис. 1.6а)

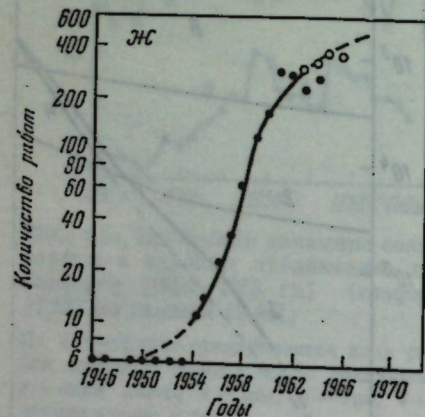


Рис. 1.6ж, з. Временная динамика числа публикаций в областях магнитных (ж) и гальваномагнитных (з) свойств тонких пленок [3.5]
Координаты — полулогарифмические, прямая соответствует экспоненте (см. рис. 1.6а)

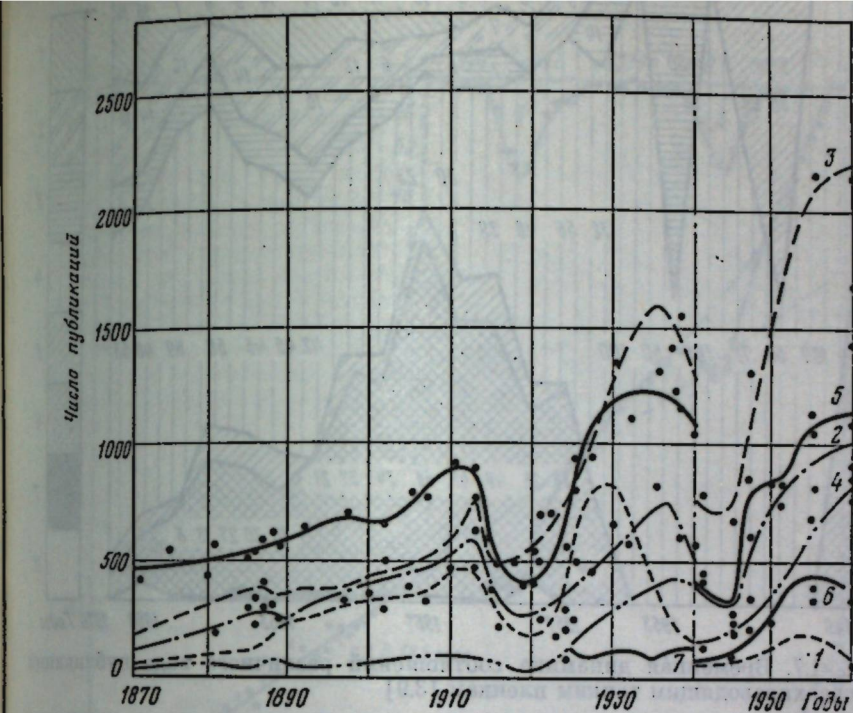
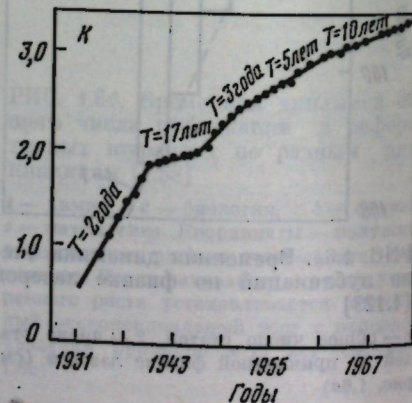
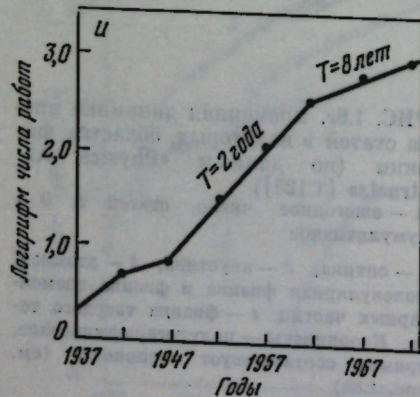
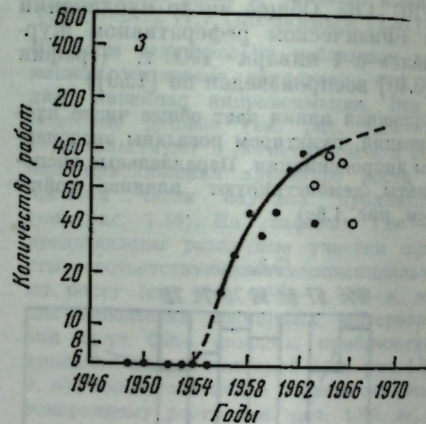


Рис. 1.6в. Временная динамика числа публикаций в областях математических исследований [1.52]

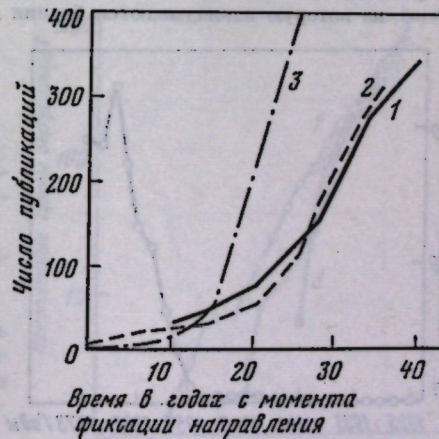
1 — история математики, педагогики и философии, 2 — алгебра (включая арифметику и теорию чисел), 3 — анализ, 4 — теория вероятностей, 5 — геометрия, 6 — численные методы графики, 7 — машинная математика. Чтобы эти данные «заговорили», их, по-видимому, следовало дать в координатах, типа представленных на рис. 1.4б и 1.20

Рис. 1.6м. Временная динамика числа публикаций в исследовательских направлениях [1.55]

1 — исследования тучной клетки (1890—1920 гг.), 2 — теория конечных групп (1934—1968), 3 — сельская социология (1941—1966 гг.) (см. рис. 1.6а)

Рис. 1.6и, к. Накопление работ по химии получения радиоактивных изотопов (и) и горячих атомов (к) [1.51]

Координаты — полулогарифмические, прямая соответствует экспоненте (см. рис. 1.6а)



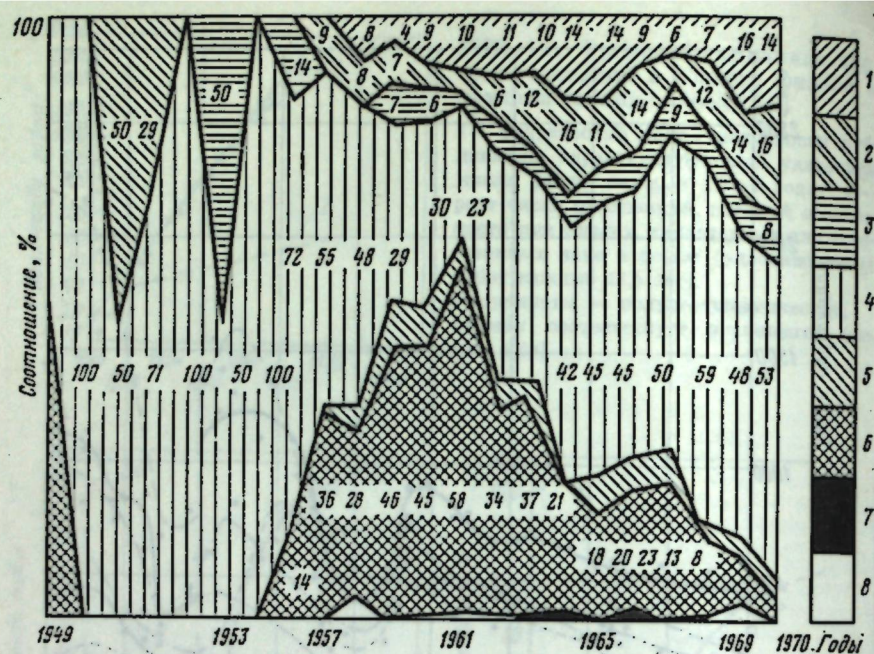


РИС. 1.7. Временная динамика соотношений различного вида публикаций по сверхпроводящим тонким пленкам [3.9]

1 — обзоры, 2 — теоретические, 3 — экспериментально-теоретические, 4 — экспериментальные, 5 — экспериментально-практические, 6 — практические, 7 — рекламные, 8 — разное

РИС. 1.8. Доля работ, выполненных в каждый из годов в интервале с 1862 по 1961 г., которые были процитированы в 1961 г. (отнесены к общему числу работ, опубликованных в 1961 г.) ([4.24] по данным Ю. Гарфилда)

Координаты — полулогарифмические, прямая соответствует экспоненте (ср. рис. 1.20). В первом приближении данные описываются, как видим, экспоненциальной зависимостью, на которую накладываются биения (ср. также рис. 1.40, 1.19а)

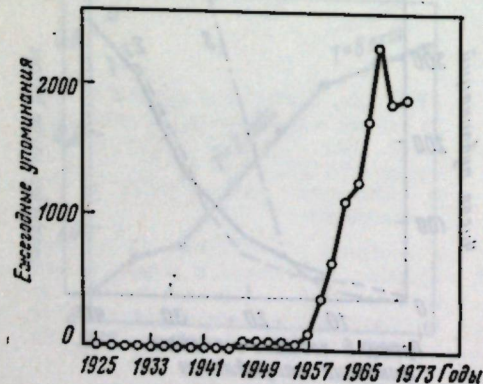


РИС. 1.9а. Временная динамика частоты упоминания понятия «кинетика реакций полимеризации» в предметном указателе «СА» [1.68]

Эти данные трудно аппроксимировать какой-либо определенной зависимостью, что иллюстрирует тезис, согласно которому экспоненциально-логистическая концепция роста науки Д. Прайса допускает иногда лишь весьма приближенные математические формулировки

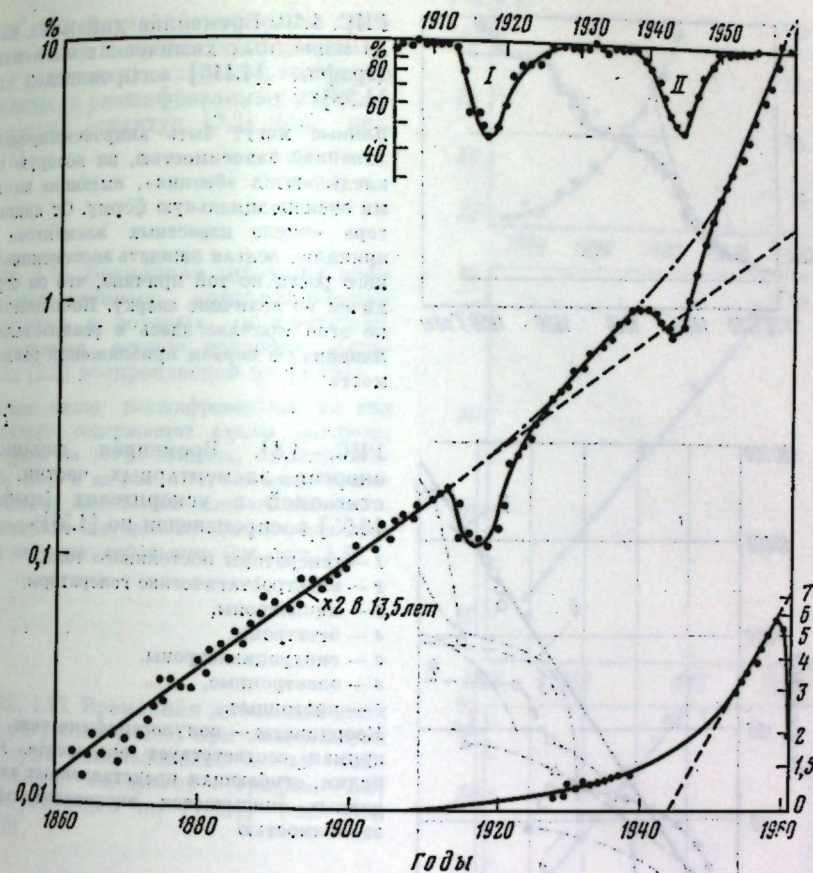
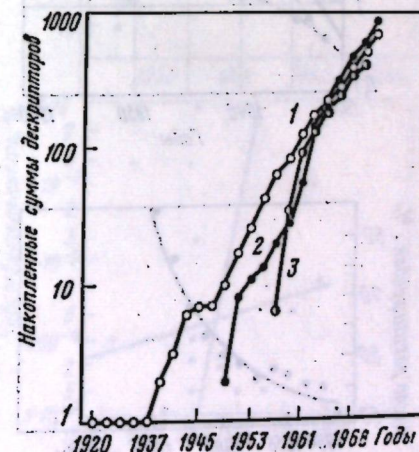


РИС. 1.9б. Временная динамика частоты упоминания дескрипторов «радикальная», «ионная» и «стереоспецифическая» полимеризации [1.68]

Координаты — полулогарифмические, прямая соответствует экспоненте. Здесь трудно говорить об определенной аппроксимации, но тем не менее при желании можно говорить о выходе данных на экспоненту



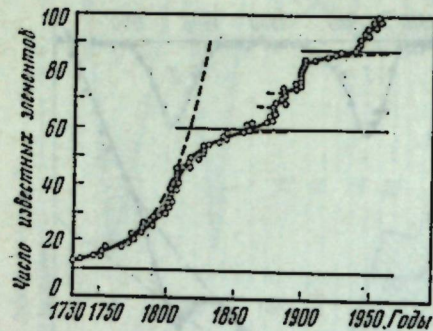


Рис. 1.10. Временная динамика числа известных химических элементов (график [1.115] воспроизведен по [1.58])

Данные могут быть аппроксимированы линейной зависимостью, на которую накладываются «биения», имеющие места экспоненциальную форму. От индикатора «число известных элементов» в принципе нельзя ожидать экспоненциального роста по той причине, что он ограничен по величине сверху. По-видимому, по этой причине здесь и реализовалась линейная в первом приближении зависимость

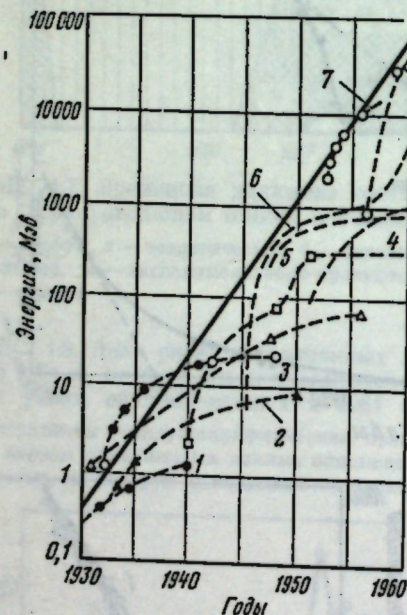


Рис. 1.11. Временная динамика энергии элементарных частиц, достигаемой в ускорителях (график [2.55] воспроизведен по [1.58])

- 1 — генераторы постоянного тока,
- 2 — электростатические генераторы,
- 3 — циклотроны,
- 4 — бетатроны,
- 5 — синхротроны,
- 6 — электронные,
- 7 — протонные.

Координаты — полулогарифмические, прямая соответствует экспоненте. Как видим, огибающая представленных здесь данных описывается экспоненциальной зависимостью

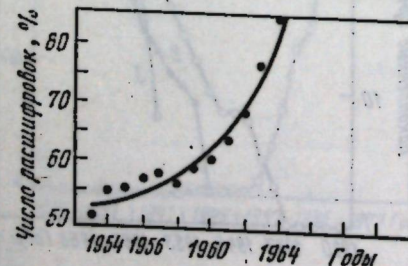


Рис. 1.12. Временная динамика числа структурных расшифровок, выполненных по трехмерным данным [2.2]

Зависимости, которым подчиняются данные, представленные здесь и на рис. 1.13, похожи на экспоненциальные. Картина стала бы яснее, если бы координаты были выбраны полулогарифмическими

Рис. 1.13. Временная динамика среднего числа координатных параметров $n(x, y, z)$, характеризующего сложность расшифрованных кристаллических структур [2.2] (см. рис. 1.12)

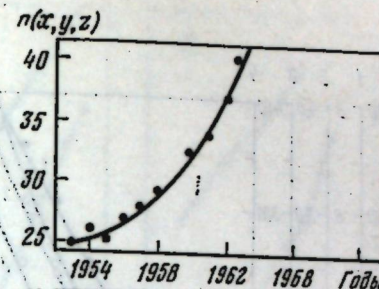


Рис. 1.14. Временная динамика доли расшифрованных кристаллических структур, в которых определялось положение атомов водорода (график [2.2] воспроизведен по [1.46])

Общее число расшифрованных за год структур, содержащих атомы водорода, принято за 100%. Данные, представленные здесь и на рис. 1.15, как видим, описываются линейными аппроксимациями. Временные интервалы, охваченные этими данными, небольшие (см. рис. 1.6)

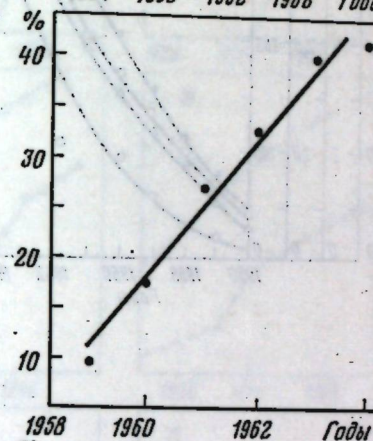


Рис. 1.15. Временная динамика доли структурных работ, при выполнении которых использовался метод наименьших квадратов (график [2.2] воспроизведен по [1.46]) (см. рис. 1.14)

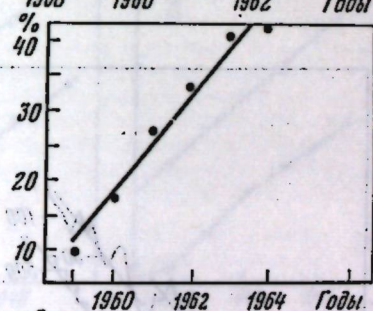
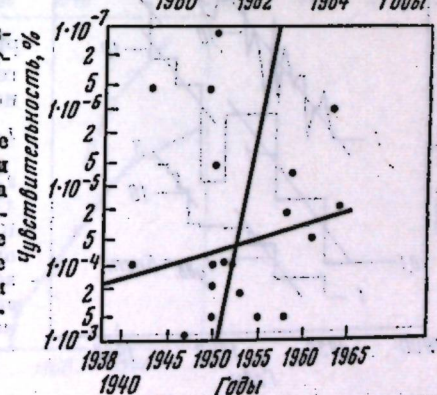


Рис. 1.16. Временная динамика чувствительности эмиссионного спектрального анализа таллия [1.46]

Координаты — полулогарифмические, прямая соответствует экспоненте. Две прямые соответствуют двум эмпирически найденным линиям регрессии, величина угла между которыми служит мерой жесткости линейной связи: чем меньше угол, тем жестче связь. В данном случае угол достаточно велик, так что едва ли можно говорить об экспоненциальном росте



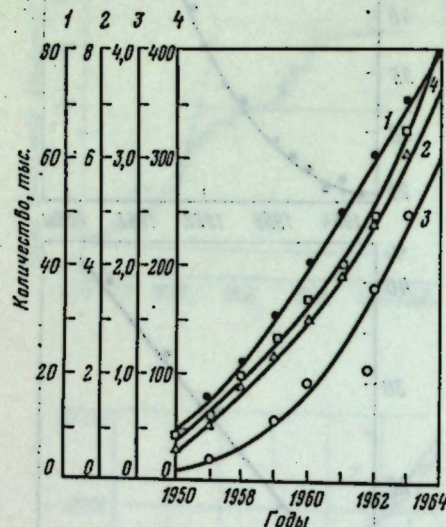


РИС. 1.17. Кумулятивные кривые роста основной «научной продукции» [1.21]

1 — научные публикации,
2 — доктора наук,
3 — заявки на научные открытия,
4 — кандидаты наук.
Координаты — линейные, поэтому трудно судить, описываются ли данные экспонентами

РИС. 1.18. Временная динамика числа докторских степеней и публикаций в отдельных областях наук о Земле [1.125]

Координаты — полулогарифмические, прямая соответствует экспоненте. Вообще говоря, эти данные трудно описать какой-либо определенной зависимостью (см. рис. 1.9a). Точки светлые — публикации, полужирные — докторские степени.

- I — геохимия,
- II — геофизика,
- III — океанография,
- IV — гляциология,
- V — стратиграфия,
- VI — микропалеонтология,
- VII — палеонтология беспозвоночных,
- VIII — вулканология,
- IX — экспериментальные науки о Земле
- X — полевая геология,
- XI — экономика

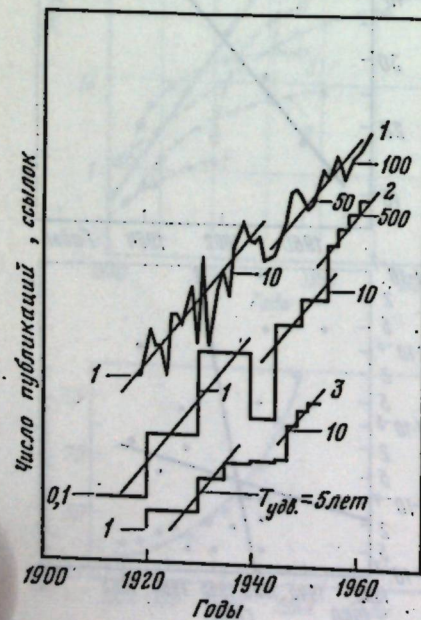
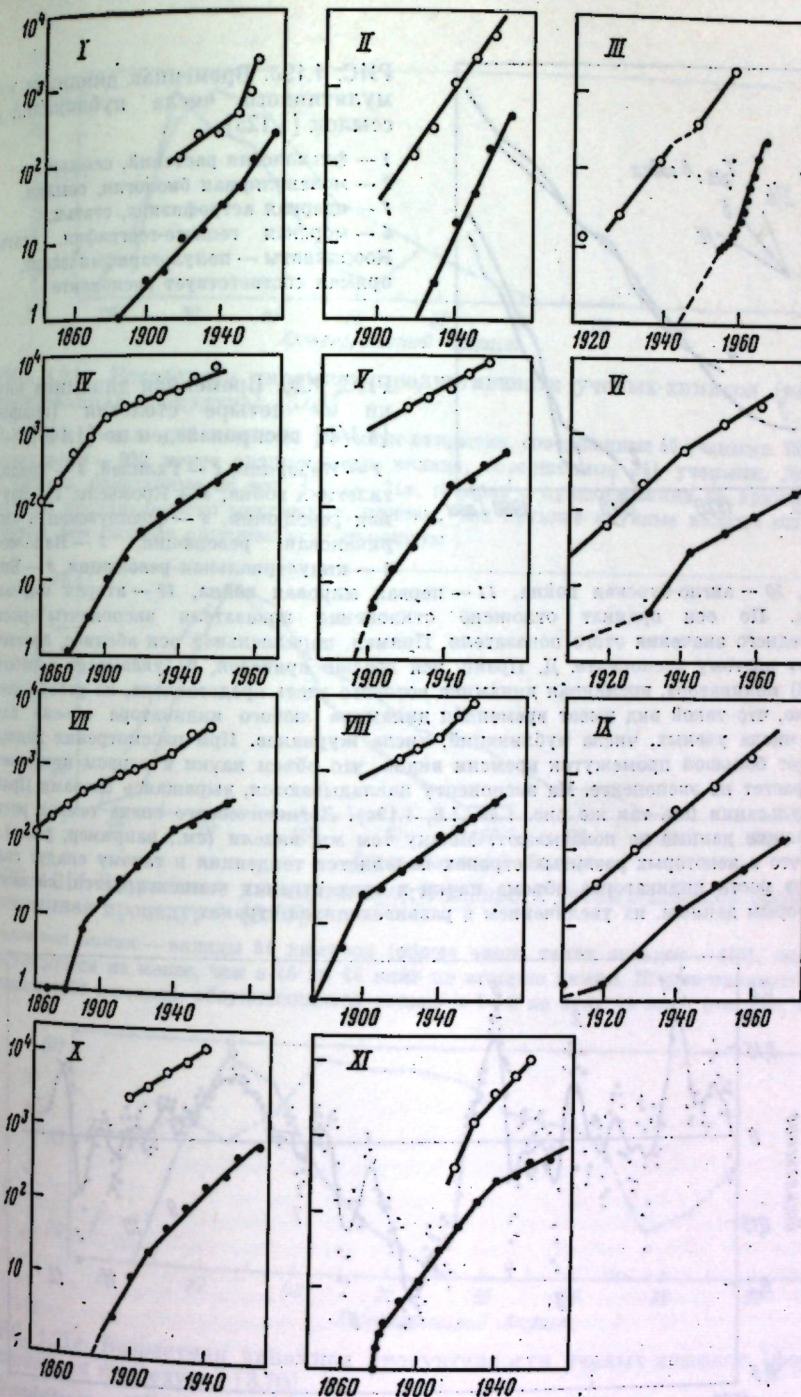


РИС. 1.19a. Временная динамика ежегодного производства статей в ядерной астрофизике [1.125]

1 — ссылки в статьях, опубликованных в разное время,
2 — ссылки в 1963 г.,
3 — ссылки в 1953 г.
Координаты — полулогарифмические, прямая соответствует экспоненте. В первом приближении рост числа статей, как видим, экспоненциальный, на экспоненты накладываются «бления»



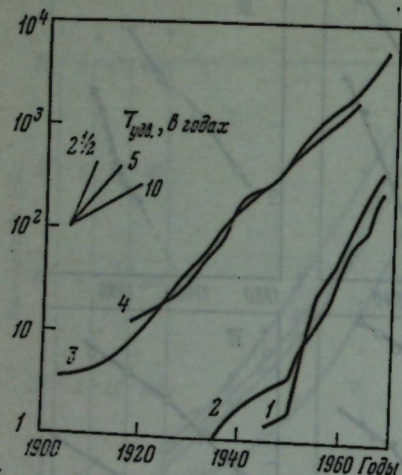


Рис. 1.196. Временная динамика кумулятивного числа публикаций и ссылок [1.125]

1 — физиологии растений, ссылки,
2 — молекулярная биология, ссылки,
3 — ядерная астрофизика, статьи,
4 — морская геолого-география, статьи
Координаты — полулогарифмические,
прямая соответствует экспоненте

Рис. 1.20. Временная динамика науки за четыре столетия (график [1.145] воспроизведен по [1.146])

1 — Коперник, 2 — Галилей, 3 — Тридцатилетняя война, 4 — Кромвель, 5 — научная революция, 6 — французская и американская революции, 7 — Наполеон, 8 — индустриальная революция, 9 — Вис-

марк, 10 — англо-бурская война, 11 — первая мировая война, 12 — вторая мировая война. По оси ординат отложено отклонение показателя экспоненты роста от среднего значения этого показателя. Прямая, параллельная оси абсцисс, соответствует поэтому экспоненте. Д. Прайс, чей график приведен, не указывает в работе [1.146] индикатора, временная динамика которого здесь представлена, но утверждает, однако, что такой вид имеет временная динамика любого индикатора объема науки за этот большой промежуток времени видно, что объем науки в первом приближении растет по экспоненте. На экспоненту накладываются, выражаясь словами Прайса, пульсации (см. так же рис. 1.40, 1.8, 1.19a). Логистического спада темпов роста настоящие данные не показывают. Между тем мы видели (см., например, рис. 1.1, 1.4), что в некоторых развитых странах появляется тенденция к такому спаду: спад темпов роста индикаторов объема науки в этих странах компенсируется, согласно некоторым данным, их увеличением в развивающихся странах

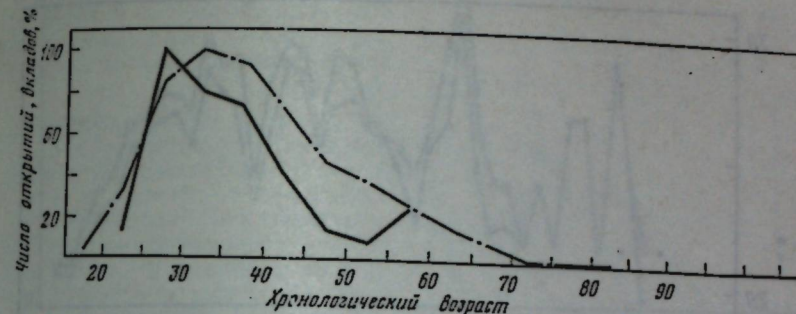
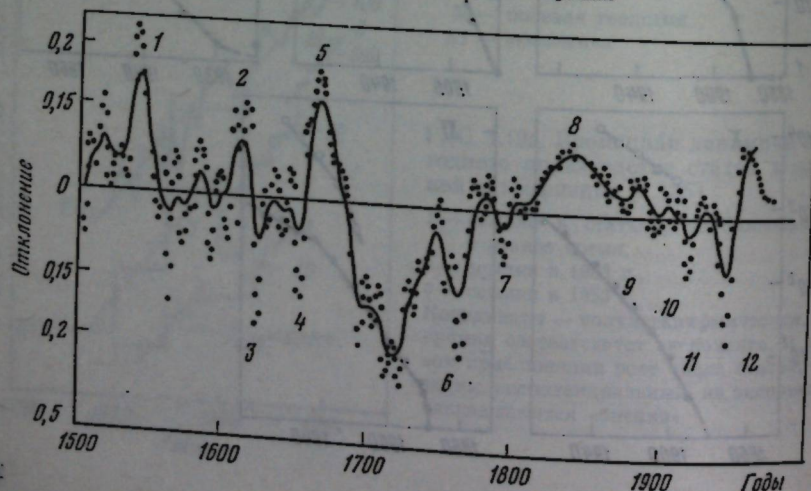


Рис. 1.21a. Возрастная динамика продуктивности ученых-химиков (неформализованная процедура) [3.68]

Сплошная линия — 52 наиболее крупные открытия, совершенные 45 учеными. Штрих-пунктирная — 993 менее значительные вклада, совершенные 344 учеными. Данные Г. Лемана, показанные на рис. 1.21a — 21e, говорят о существовании на кривой возрастной продуктивности максимума, причем, чем крупнее научные вклады мы фиксируем, тем сильнее выражен этот максимум

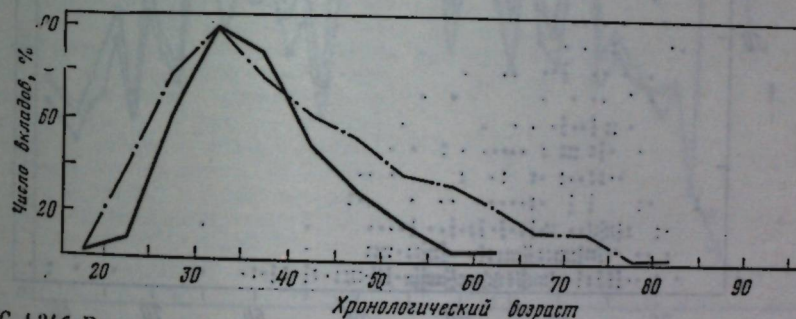


Рис. 1.21b. Возрастная динамика продуктивности ученых-химиков (формализованная процедура) [3.70]

Сплошная линия — вклады 54 химиков (общее число таких вкладов — 155), которые обсуждаются не менее, чем в 15 из 44 книг по истории химии. Штрих-пунктирная — вклады 1166 химиков, обсуждающиеся только в 1—2 из этих 44 книг (см. рис. 1.21a)

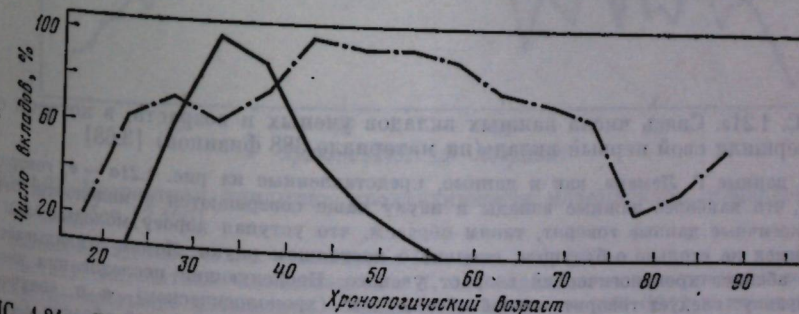


Рис. 1.21c. Возрастная динамика продуктивности ученых-химиков (формализованная процедура) [3.70]

Штрих-пунктирная линия — 550 относительно малых вкладов этой же группы химиков, обсуждавшихся не более чем в 1—2 из 44 книг по истории химии (см. рис. 1.21a)

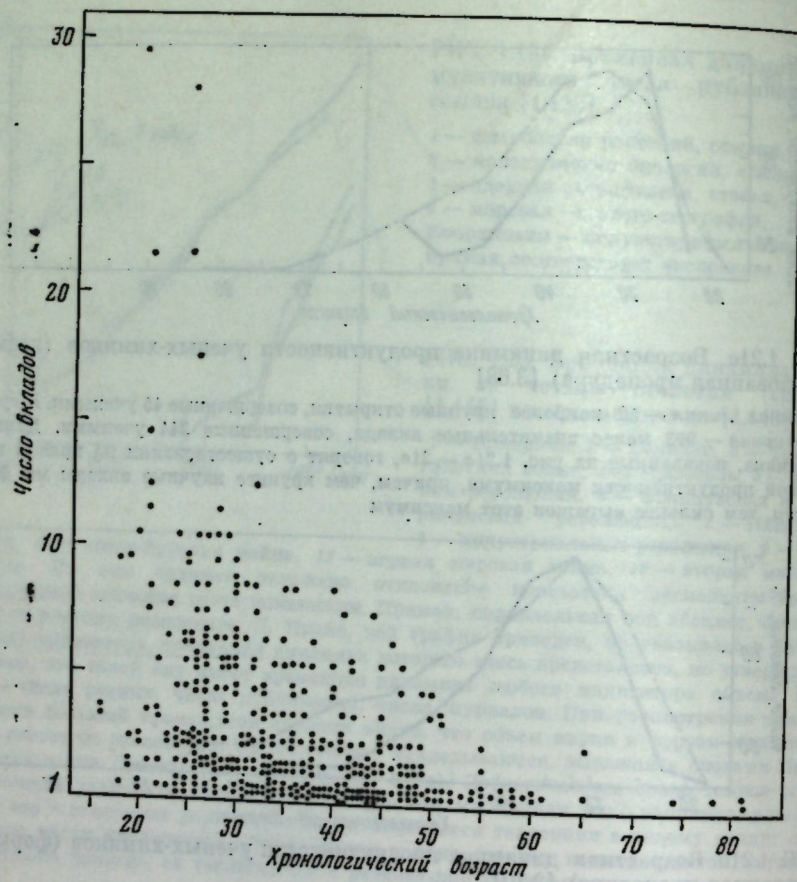


РИС. 1.21а. Связь числа важных вкладов ученых и возраста, в котором они совершили свой первый вклад (на материале 388 физиков) [3.68]
Эти данные Г. Лемана, как и данные, представленные на рис. 1.21а — е, говорят о том, что наиболее важные вклады в науку чаще совершаются в молодости. Эти и аналогичные данные говорят, таким образом, что уступая дорогу молодым, мы заботимся не столько о будущем, сколько о настоящем науки. Леман откладывает по оси абсцисс хронологический возраст ученого. Последующие исследования внесли поправку: следует говорить не об абсолютном (хронологическом), а о «научном» возрасте ученого — о его стаже работы в данной исследовательской области. Согласно некоторым данным, более половины научных открытий совершается новичками в данной области исследований

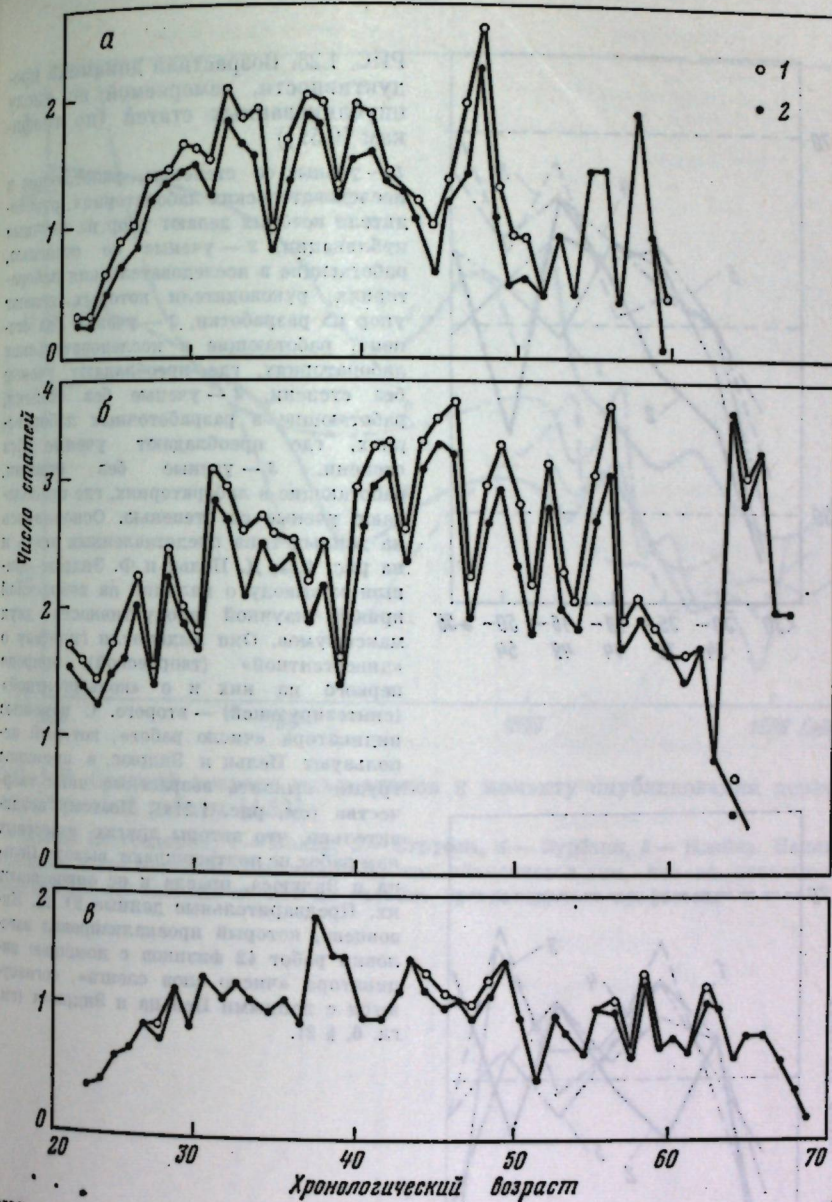


РИС. 1.22. Возрастная динамика продуктивности математиков, измеряемой по числу работ [3.100]

1 — полная продуктивность, 2 — частичная (см. рис. 2.23, 2.25)
а — Институт математики Чехословацкой АН, б — выдающиеся чехословацкие математики, в — советские математики — члены-корреспонденты АН СССР и академики. Данные не показывают наличия возрастного пика продуктивности. Это и естественно, так как здесь используется индикатор «число публикаций», нивелирующий научные вклады разной степени важности (см. рис. 1.21а)

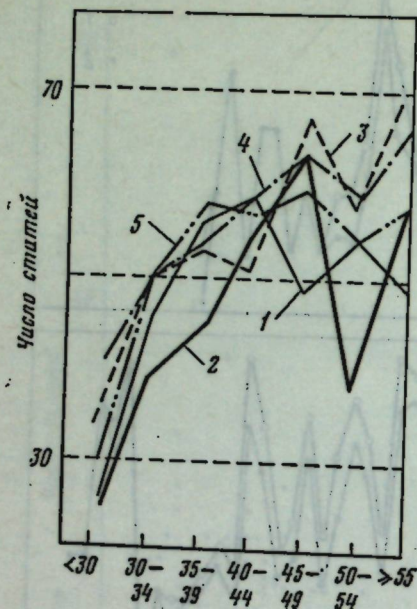


РИС. 1.23. Возрастная динамика продуктивности, измеряемой по числу опубликованных статей (по графикам [0.51])

1 — ученые со степенью, работающие в исследовательских лабораториях, руководители которых делают упор на научные публикации, 2 — ученые со степенью, работающие в исследовательских лабораториях, руководители которых делают упор на разработки, 3 — ученые без степени, работающие в исследовательских лабораториях, где преобладают ученые без степени, 4 — ученые без степени, работающие в разработочных лабораториях, где преобладают ученые без степени, 5 — ученые без степени, работающие в лабораториях, где преобладают ученые со степенью. Основываясь на данных типа представленных здесь и на рис. 1.24, Д. Пельц и Ф. Эндрюс пришли к выводу о наличии на возрастной кривой научной продуктивности двух максимумов. Они выдвинули гипотезу о «дивергентной» (творческой) природе первого из них и о «конвергентной» (синтезирующей) — второго. С помощью индикатора «число работ», который используют Пельц и Эндрюс, в принципе трудно выявить возрастные пики творчества (см. рис. 1.21а). Поэтому неудивительно, что авторы других известных нам работ не подтверждают выводы Пельца и Эндрюса, правда и не опровергают их. Предварительные данные Ю. М. Низовцева, который проанализировал заголовки работ 42 физиков с помощью индикатора «число слов сленга», согласуются с данными Пельца и Эндрюса (см. гл. 6, § 2).

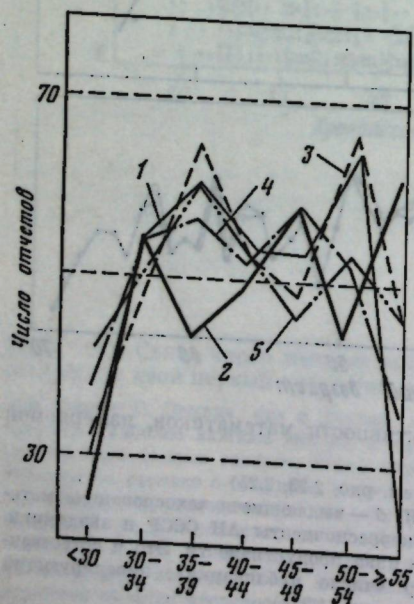


РИС. 1.24. Возрастная динамика продуктивности ученых, измеряемой по числу неопубликованных отчетов (по графикам [0.51])

Обозначения те же, что и на рис. 1.23

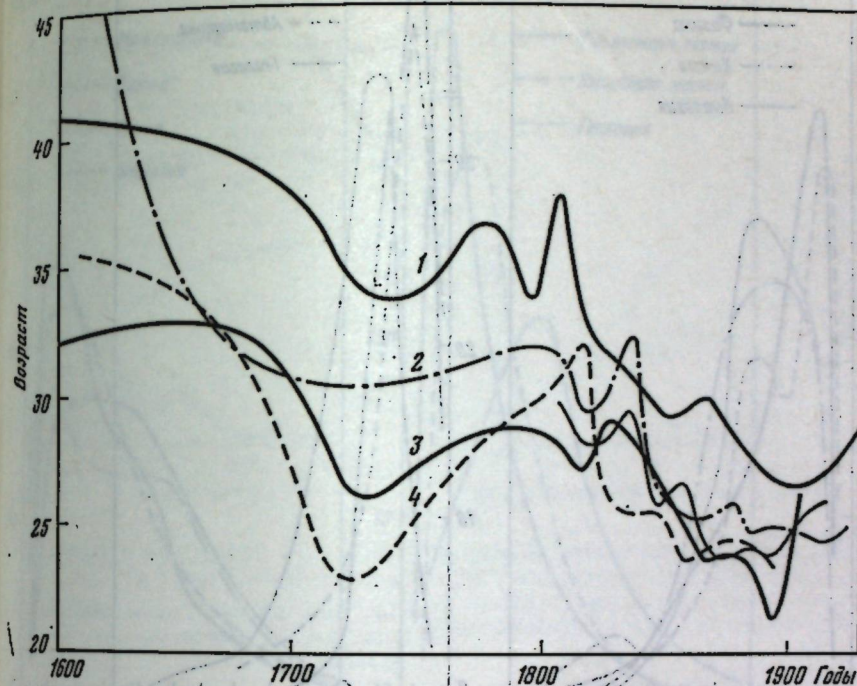


РИС. 1.25. Средний возраст математиков к моменту опубликования первой математической работы [1.52]

1 — данные Поггендорфа, 2 — Белла, 3 — Струйка, 4 — Бурбаки, 5 — Клейна. Налицо тенденция к омоложению. По-видимому, это объясняется тем, что за охваченные графиком 300 лет занятия наукой из «хобби» превратились в профессию

РИС. 1.26. Распределение ученых по возрасту [3.100]

а — Чехословакия, б — США. Видим, что в среднем ученые США «старее», чем их чехословацкие коллеги. По-видимому, это следует связать с наблюдающимся торможением темпов роста численности американских ученых (см. рис. 1.4а (с. 242), б (с. 243))

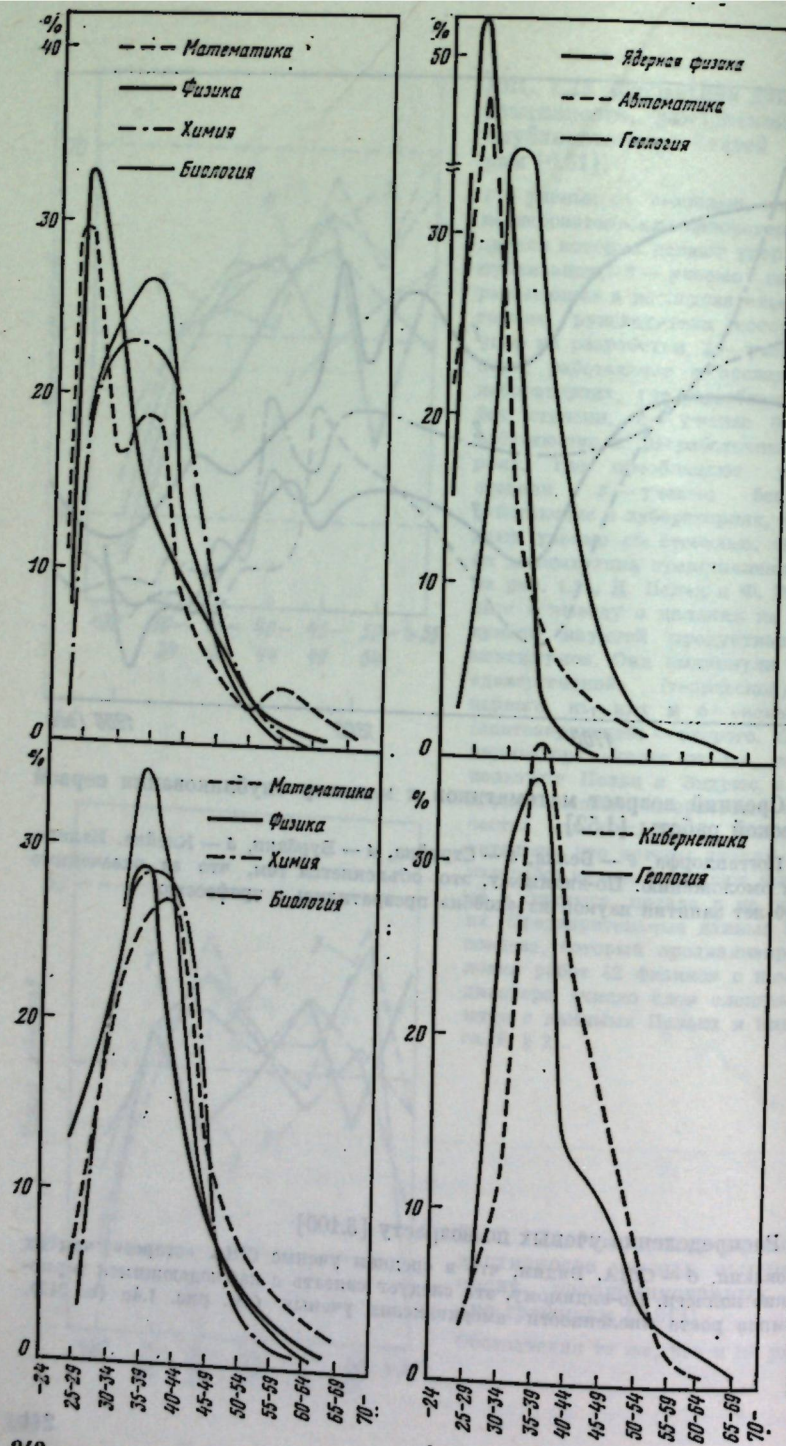


РИС. 1.265

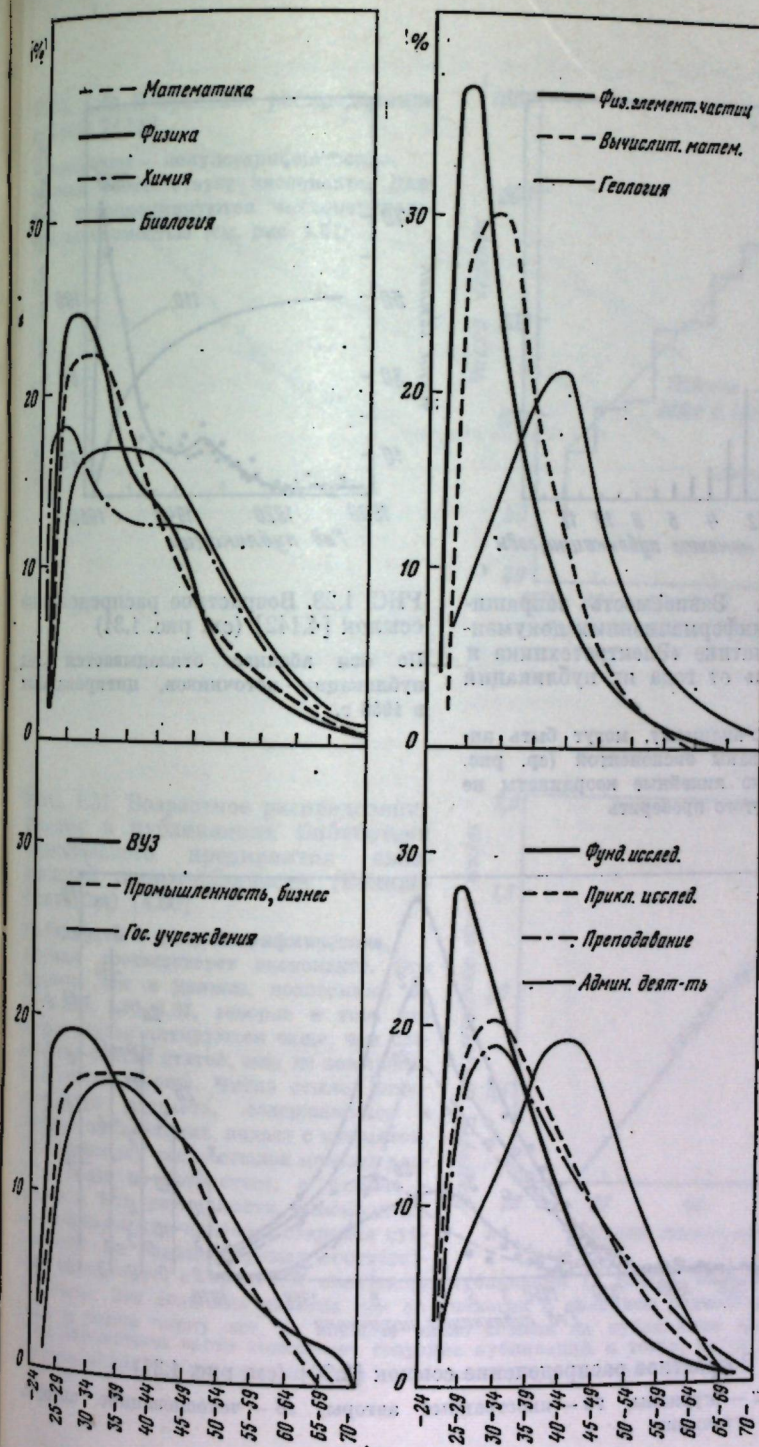


РИС. 1.266

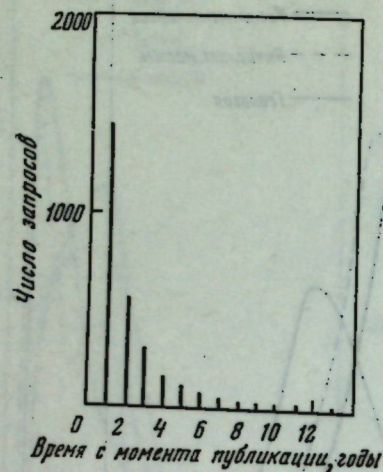


Рис. 1.27. Зависимость запрашиваемости информационных документов по тематике «Электротехника и энергетика» от года их публикации [2.17]

Данные, по-видимому, могут быть аппроксимированы экспонентой (см. рис. 1.30), однако линейные координаты не позволяют этого проверить

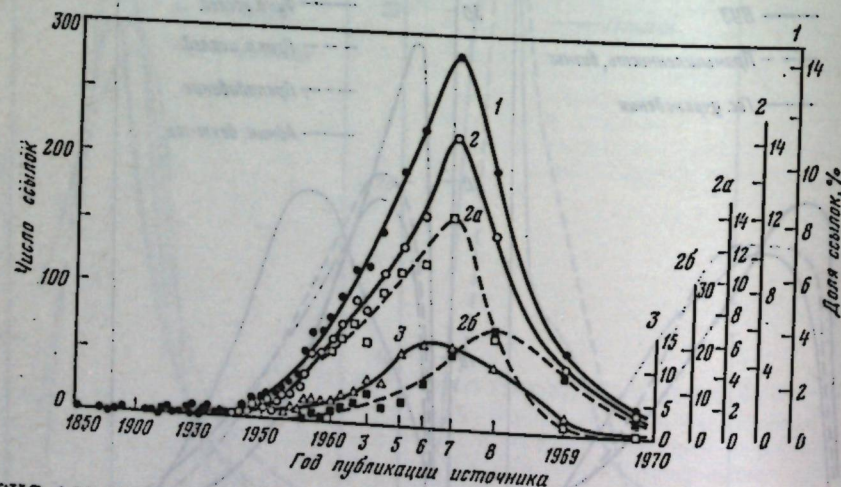


Рис. 1.29. Возрастное распределение ссылок [2.70] (см. рис. 1.31)
1 — всего, 2 — журналы, 2а — иностранные авторы, 2б — чехословацкие авторы, 3 — прочие источники

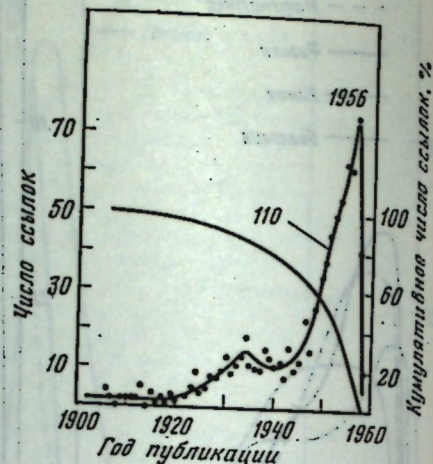


Рис. 1.28. Возрастное распределение ссылок [4.142] (см. рис. 1.31)

По оси абсцисс откладывается год публикации источников, цитированных в 1958 г.

Рис. 1.30. Возрастное распределение ссылок [1.58]

Координаты — полулогарифмические, прямая соответствует экспоненте. Данные аппроксимируются экспоненциальной зависимостью (см. рис. 1.31)

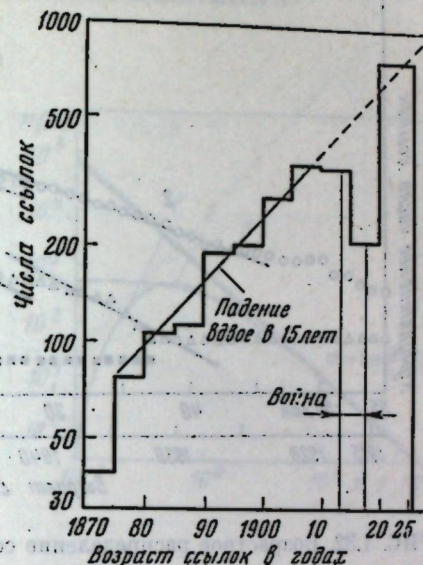
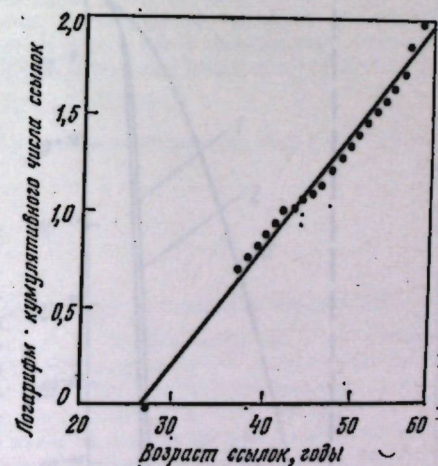


Рис. 1.31. Возрастное распределение ссылок в публикациях Библиотеки Королевского предприятия авиационной промышленности (Великобритания) [4.60]

Координаты — полулогарифмические, прямая соответствует экспоненте. Эти данные, как и данные, показанные на рис. 1.28, 1.30, 1.32, говорят о том, что новые статьи цитируются чаще, чем старые (не считая статей, еще не вошедших в научный оборот). Число ссылок определенного возраста, содержащихся в данных публикациях, падает с возрастом. Суммарное же число ссылок моложе данного возраста возрастает, а старше — убывает. Эти зависимости описываются так называемыми кривыми старения публикаций. На графике прямая соответствует экспоненте с временем полужизни публикаций (в годах), определяем ее наклоном. Эта величина названа так по аналогии с явлением радиоактивного распада и равна числу лет, за которые число ссылок на публикацию падает вдвое. Этим параметром часто описывают старение публикаций и тогда, когда возрастные распределения ссылок описываются не экспонентой, а как-либо иначе



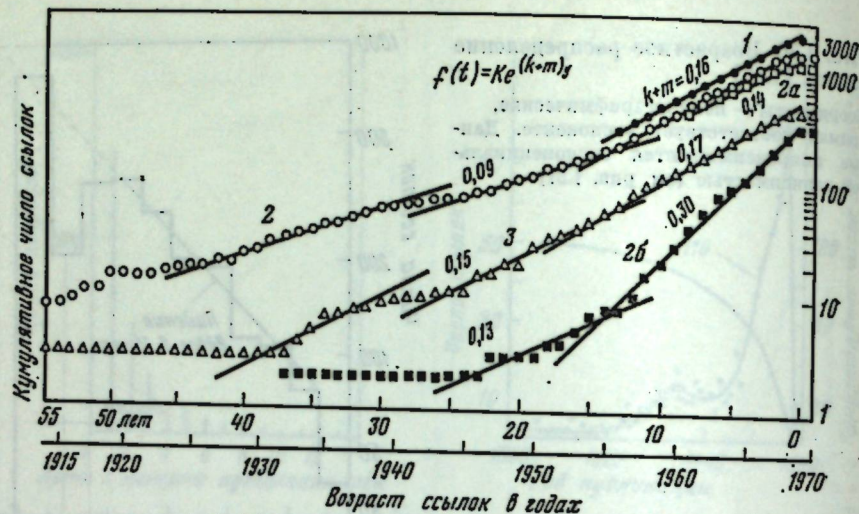
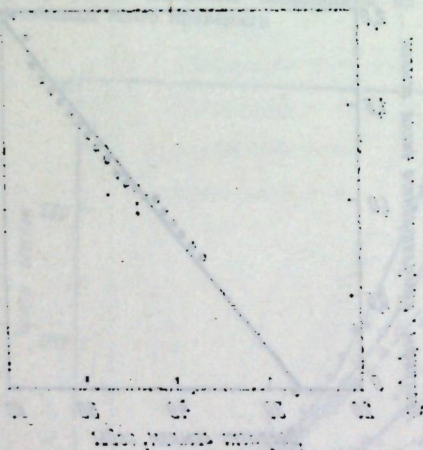


РИС. 1.32. Возрастное распределение ссылок [2.70]

Координаты — полулогарифмические, прямая соответствует экспоненте. Данные аппроксимируются суммой экспонент
1 — всего, 2 — журналы, 2a — иностранные авторы, 2b — чехословацкие авторы, 3 — прочие источники



2. СТАЦИОНАРНЫЕ РАСПРЕДЕЛЕНИЯ

РИС. 2.1. Частотная форма ципфовых распределений

Координаты — логарифмические. 1 — обобщенное распределение Парето: $n(x) = \frac{aN}{x^{1+\alpha}}; x \geq 0, a=1, \alpha=1.5; N=1000, a/b=0.793$. 2 — распределение Коши: $n(x) = \frac{2N}{\pi d} \frac{1}{1 + (\frac{x}{d})^2}; x \geq 0,$

$N=1000; 2/\pi d=0.793$, 3 — распределение Ципфа: $n(x) = \frac{cN}{x^{1+\alpha}}; \alpha=1, c=0.793, N=1000$. В этих координатах ципфовые распределения при $x \rightarrow \infty$ имеют асимптоты, наклоненные к оси x под острым углом φ , величина которого определяет показатель ципфового распределения α

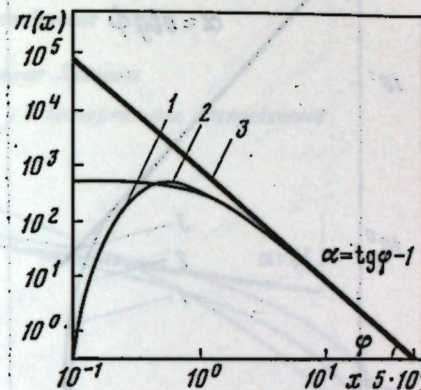


РИС. 2.2. Частотная форма гауссовых распределений

Координаты — логарифмические. 1 — «правое» распределение Гаусса: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}; x \geq 0, \bar{x}=1.5,$

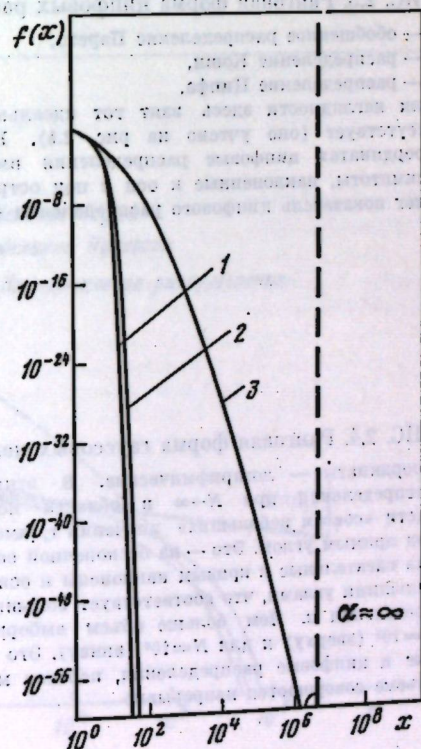
$\sigma=\bar{x}\sqrt{\frac{\pi}{2}}$, 2 — распределение Пуассона: $f(x) = e^{-\lambda} \frac{\lambda^x}{x!}; x \geq 0; \bar{x}=1.5; \lambda=\bar{x}$. 3 — логнормальное распределение: $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\ln x - \mu}{\sigma})^2};$

$x \geq 0, \bar{x}=1.5, \mu=0, \sigma = \sqrt{2 \ln \bar{x}}$

В этих координатах гауссовые неципфовые распределения при $x \rightarrow \infty$ имеют касательные, наклоненные к оси x под прямым углом. Это — на бесконечных совокупностях. На выборках конечного объема касательные к кривым при больших значениях переменных наклонены к оси x не под прямыми, но тем не менее под значительными углами, что соответствует большим значениям показателя ципфового распределения α . Это позволяет различать гауссовые неципфовые и ципфовые распределения, переход между которыми на выборках конечного объема совершается непрерывно

В этих координатах гауссовые неципфовые распределения при $x \rightarrow \infty$ имеют касательные, наклоненные к оси x под прямым углом. Это — на бесконечных совокупностях. На выборках конечного объема касательные к кривым при больших значениях переменных наклонены к оси x не под прямыми, но тем не менее под значительными углами, что соответствует большим значениям показателя ципфового распределения α . Это позволяет различать гауссовые неципфовые и ципфовые распределения, переход между которыми на выборках конечного объема совершается непрерывно

В этих координатах гауссовые неципфовые распределения при $x \rightarrow \infty$ имеют касательные, наклоненные к оси x под прямым углом. Это — на бесконечных совокупностях. На выборках конечного объема касательные к кривым при больших значениях переменных наклонены к оси x не под прямыми, но тем не менее под значительными углами, что соответствует большим значениям показателя ципфового распределения α . Это позволяет различать гауссовые неципфовые и ципфовые распределения, переход между которыми на выборках конечного объема совершается непрерывно



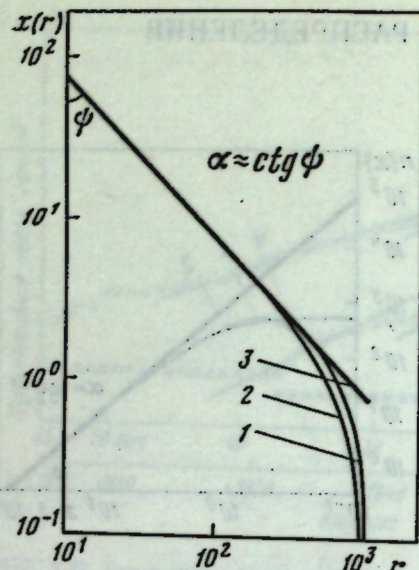


РИС. 2.3. Ранговая форма ципфовых распределений

- 1 — обобщенное распределение Парето,
- 2 — распределение Коши,
- 3 — распределение Ципфа.

Для наглядности здесь взят тот идеальный случай, когда ранговое искажение отсутствует (оно учтено на рис. 2.5). Координаты — логарифмические. В этих координатах ципфовые распределения имеют при $x \rightarrow \infty$ (т. е. при $N \rightarrow \infty$ и $r \rightarrow 1$) асимптоты, наклоненные к оси x под острым углом ϕ , величина которого определяет показатель ципфоваго распределения α

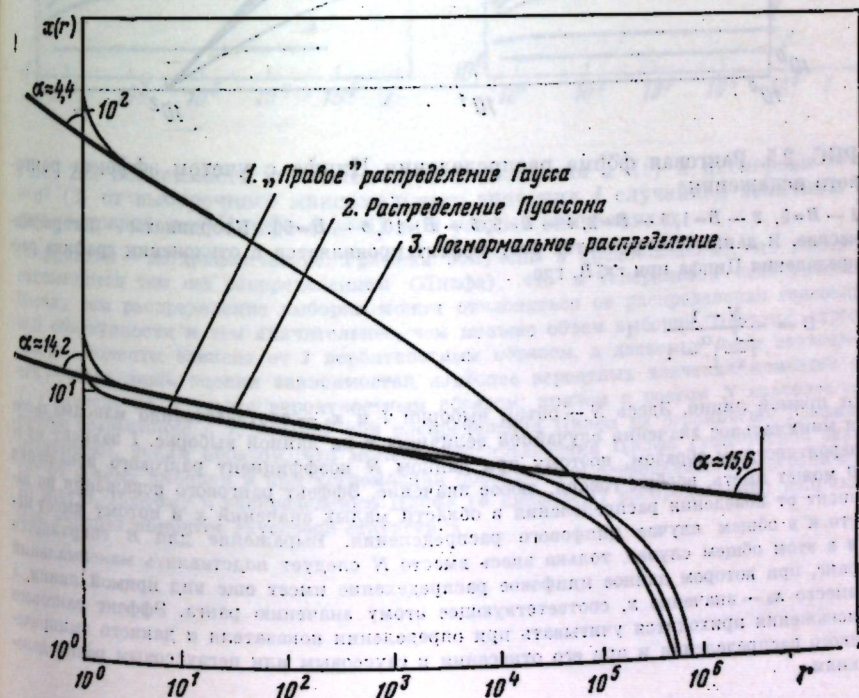
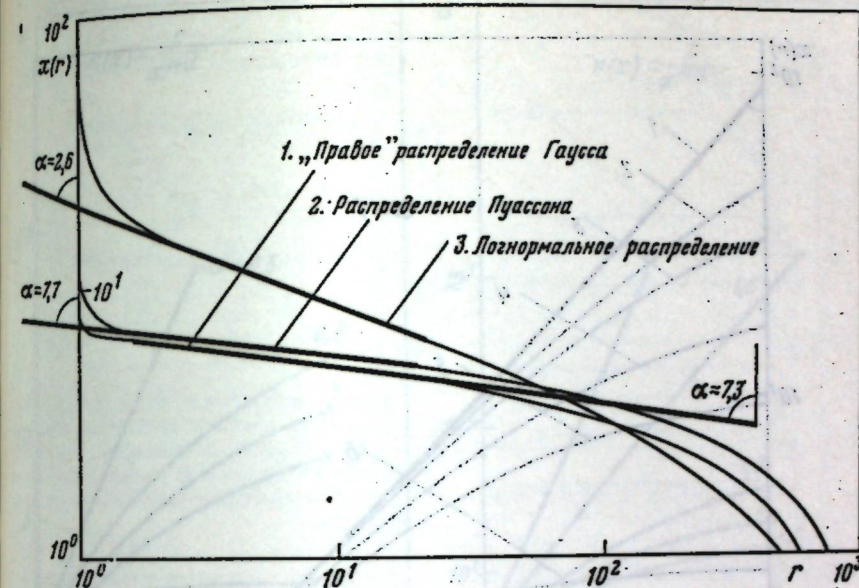


РИС. 2.4. Ранговая форма гауссовых неципфовых распределений

Координаты — логарифмические. В этих координатах гауссовые неципфовые распределения при $N \rightarrow \infty$ в области небольших значений ранга, кроме области «совсем небольших» значений r , имеют касательные, наклоненные к оси x под прямым углом. Это — на бесконечной совокупности. На выборках конечного объема касательные к кривым наклонены к оси x не под прямыми, но тем не менее под большими углами, что соответствует большим значениям показателя ципфоваго распределения α . Чем больше объем выборки, тем больше эти значения — ср. для $N=10^3$ (вверху) и для $N=10^6$ (внизу). Это позволяет различать гауссовые неципфовые и ципфовые распределения, переход между которыми на выборках конечного объема совершается непрерывно

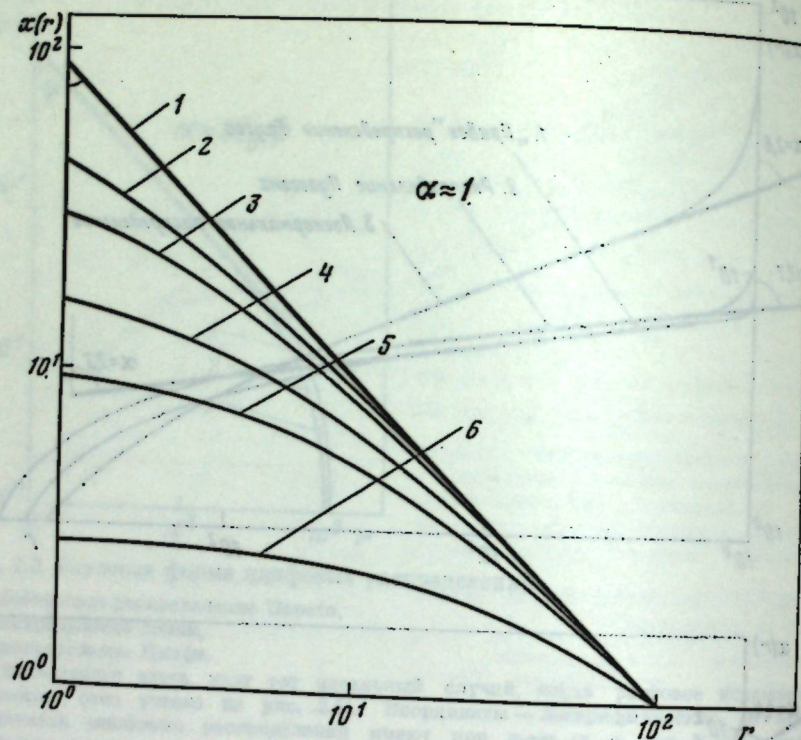


РИС. 2.5. Ранговая форма распределения Ципфа с учетом эффекта рангового искажения
1 — $B=0$, 2 — $B=1$, 3 — $B=2$, 4 — $B=5$, 5 — $B=10$, 6 — $B=50$. Координаты — логарифмические. В данных координатах этот эффект проявляется в отклонении графика распределения Ципфа при $r \lesssim B$, где

$$B = \frac{N-1}{\left(\frac{I}{x_0}\right)^\alpha - 1}$$

от прямой линии. Здесь N — объем выборки I и x_0 — соответственно максимальное и минимальное значения случайной величины x на данной выборке. I зависит от N вероятностным образом, поэтому при данном N коэффициент рангового искажения B может иметь, вообще говоря, любое значение. Эффект рангового искажения не зависит от поведения распределения в области малых значений x и потому имеет место и в общем случае ципфоваго распределения. Выражение для B сохраняется и в этом общем случае, только здесь вместо N следует подставлять максимальный ранг, при котором данное ципфовое распределение имеет еще вид прямой линии, а вместо x_0 — значение x , соответствующее этому значению ранга. Эффект рангового искажения приходится учитывать при определении показателя α данного эмпирического распределения и при его отношении к гауссовым или негауссовым распределениям.

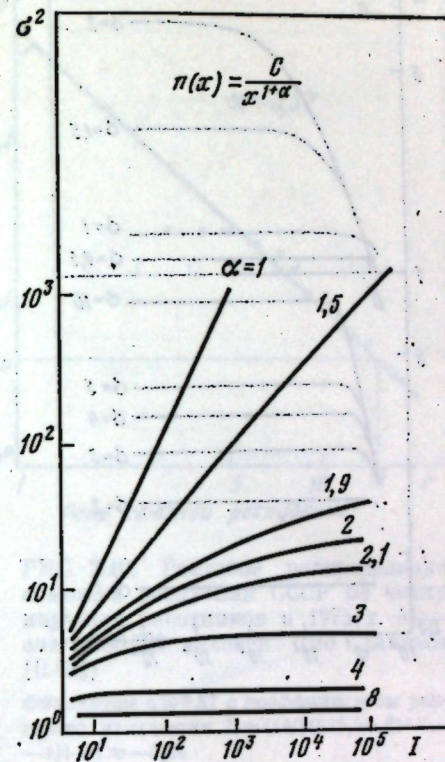
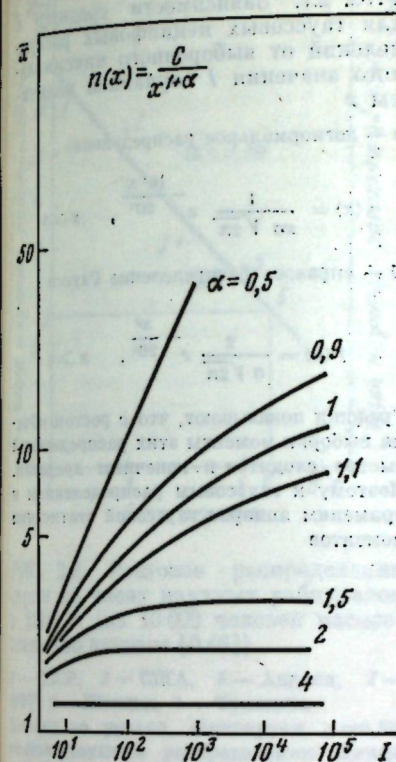


РИС. 2.6. Зависимости математического ожидания $\bar{x}$ (1) и дисперсии $\bar{x}^2 = \sigma^2$ (2) от выборочного максимального значения I случайной величины x для распределения Ципфа ($x_0=1$) при разных значениях α

Координаты — логарифмические. Графики получены в предположении, что выборки описываются тем же распределением (Ципфа), что и генеральная совокупность. Между тем распределение выборки может отклоняться от распределения генеральной совокупности и тем значительнее, чем меньше объем выборки. Поэтому, строго говоря, моменты зависят от I вероятностным образом, а даваемые нами зависимости — всего лишь оценки зависимостей наиболее вероятных значений моментов от I . I зависит от N также вероятностным образом, причем с ростом N наиболее вероятно возрастание и I , которое для распределения Ципфа не ограничено. Графики показывают, таким образом, что моменты распределения Ципфа существенно зависят от объема выборки и при устремлении последнего к бесконечности неограниченно растут. Поэтому к распределению Ципфа (и ципфовому распределению) неприменим аппарат моментов гауссовой статистики

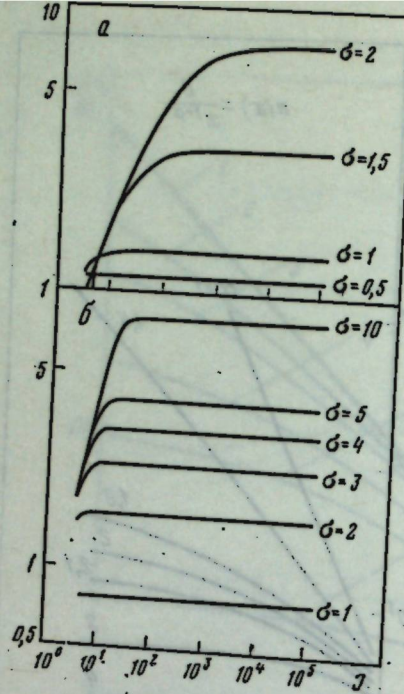


РИС. 2.7. Зависимости среднего $\bar{x}$ для гауссовых нецифровых распределений от выборочного максимального значения l случайной величины x

а — логнормальное распределение

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{\ln^2 x}{2\sigma^2}}, \quad x \geq 0$$

б — «правое» распределение Гаусса

$$f(x) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0$$

Графики показывают, что с ростом объема выборки моменты этих распределений быстро сходятся к конечным пределам. Поэтому к гауссовым распределениям применим аппарат гауссовой статистики моментов

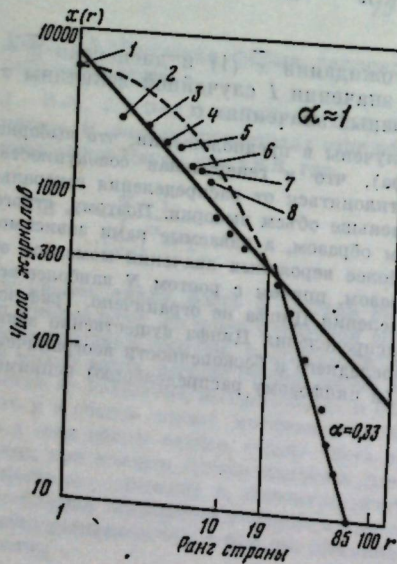


РИС. 2.8. Ранговое распределение стран по числу издаваемых в них научных журналов [1.58]

1 — США, 2 — ФРГ, 3 — Франция, 4 — Англия, 5 — СССР, 6 — Италия, 7 — Япония, 8 — Бельгия.

Массив не указан. Координаты на рис. 2.8—2.78, за исключением особо оговариваемых случаев, логарифмические. Гипотеза, что данное распределение является распределением Ципфа с $\alpha \approx 0,33$ и коэффициентом рангового искажения $B \approx [84/((6000/10)^{0,33}-1)]-1 \approx 10,6$, по-видимому, недостоверна, так как соответствующая кривая (пунктирная на графике) плохо описывает данные в области первых рангов. Поэтому фиксируем $\alpha \approx 1$, соответствующее ципфовому распределению с $B \approx [18/((6000/380)^1-1)]-1 \approx 0,22$

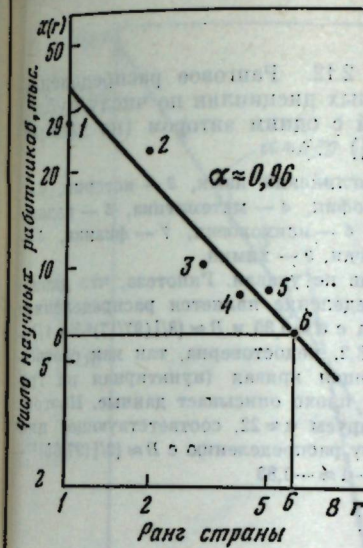


РИС. 2.9. Ранговое распределение стран по числу научных работников в 1965 г. (на 10 000 человек населения) (по данным [0.46])

1 — СССР, 2 — США, 3 — Англия, 4 — ФРГ, 5 — Швеция, 6 — Франция. Массив не указан. Фиксируем $\alpha \approx 0,96$, соответствующее распределению Ципфа с $B \approx [5/((29/6)^{0,96}-1)]-1 \approx 0,41$

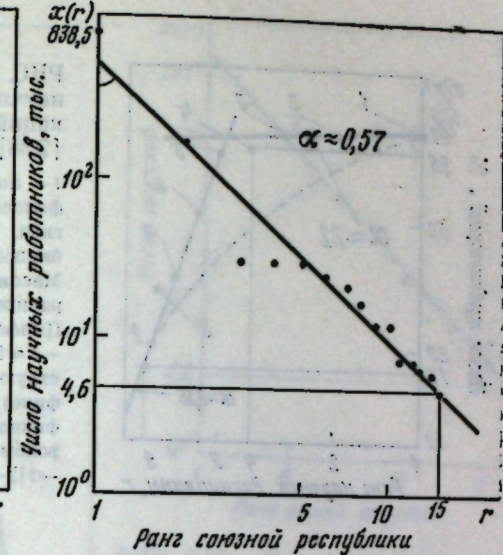


РИС. 2.10. Ранговое распределение союзных республик СССР по числу научных работников в 1975 г. Массив 838 500 человек (по данным [0.45])

Фиксируем $\alpha \approx 0,57$ с коэффициентом рангового искажения $B \approx [14/((838,5/4,6)^{0,57}-1)]-1 \approx -0,24$

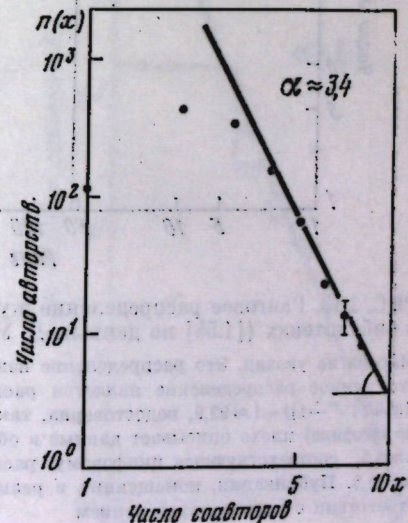


РИС. 2.11. Распределение числа авторств по числу соавторов на статью (по данным [2.30])

Массив 1 239 авторств

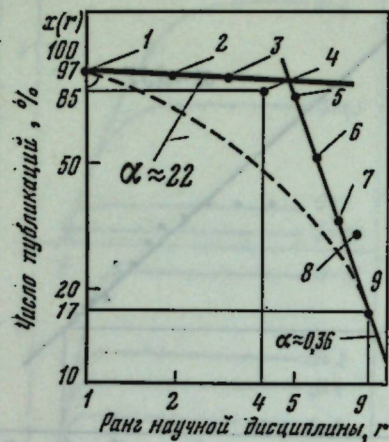


РИС. 2.12. Ранговое распределение научных дисциплин по числу публикаций с одним автором (по данным [2.38])

1 — английский язык, 2 — история, 3 — философия, 4 — математика, 5 — педагогика, 6 — психология, 7 — физика, 8 — биология, 9 — химия.

Массив не указан. Гипотеза, что данное распределение является распределением Ципфа с $\alpha \approx 0,36$ и $B \approx (8/[(97/17)^{0,36} - 1]) - 1 \approx 8,2$, недостоверна, так как соответствующая кривая (пунктирная на графике) плохо описывает данные. Поэтому фиксируем $\alpha \approx 22$, соответствующее ципфовому распределению с $B \approx (3/[(97/85)^{22} - 1]) - 1 \approx -0,83$

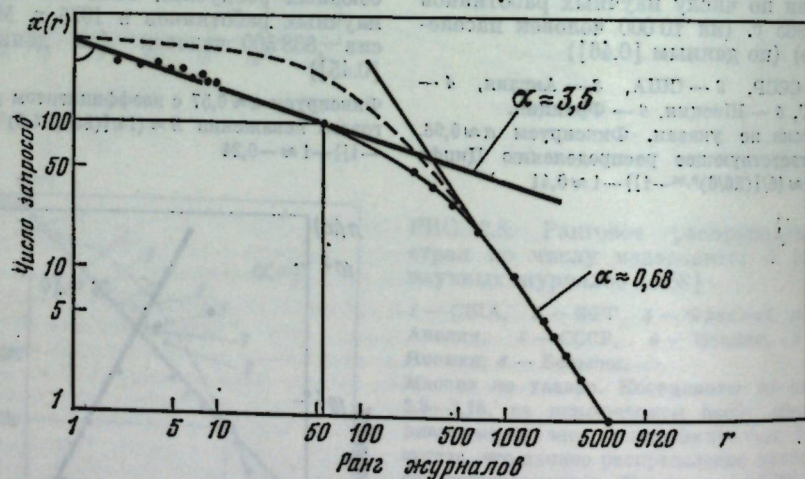


РИС. 2.13. Ранговое распределение журналов по числу запросов на них в год в библиотеках ([1.58] по данным Д. Уркварта [2.67])

Массив не указан. Это распределение называют распределением Уркварта. Гипотеза, что данное распределение является распределением Ципфа с $\alpha \approx 0,68$ и $B \approx (4999/[(350/1)^{0,68} - 1]) - 1 \approx 93,9$, недостоверна, так как соответствующая кривая (пунктирная на графике) плохо описывает данные в области первых рангов. Поэтому фиксируем $\alpha \approx 3,5$, соответствующее ципфовому распределению с $B \approx (59/[(400/100)^{3,5} - 1]) - 1 \approx -0,5$. Публикации, помещенные в разных журналах, могут «взвешиваться» в соответствии с этим распределением

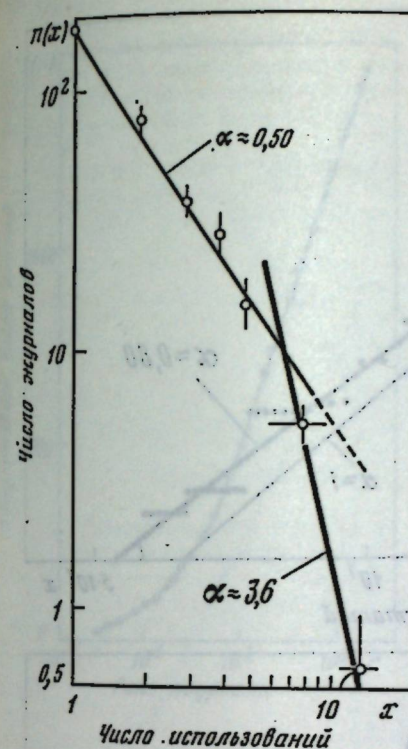


РИС. 2.14. Распределение журналов по числу библиотечных запросов [1.131]

Массив 376 журналов, 858 использований. Фиксируем значение $\alpha \approx 3,6$, согласующееся с данными, представленными на рис. 2.13

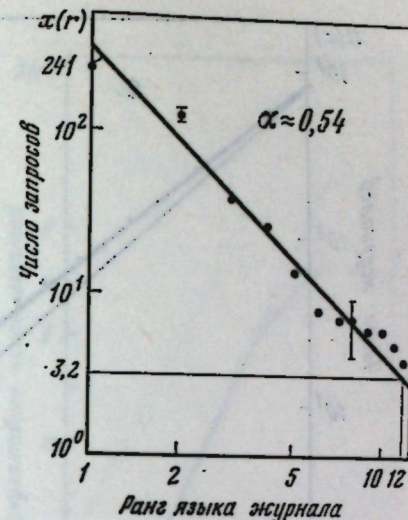


РИС. 2.15. Ранговое распределение языков журналов по числу библиотечных запросов (по данным [2.17])

Массив 490 запросов. Фиксируем $\alpha \approx 0,54$, соответствующее распределению Ципфа с $B \approx (11/[(241/3,2)^{0,54} - 1]) - 1 \approx 0,18$

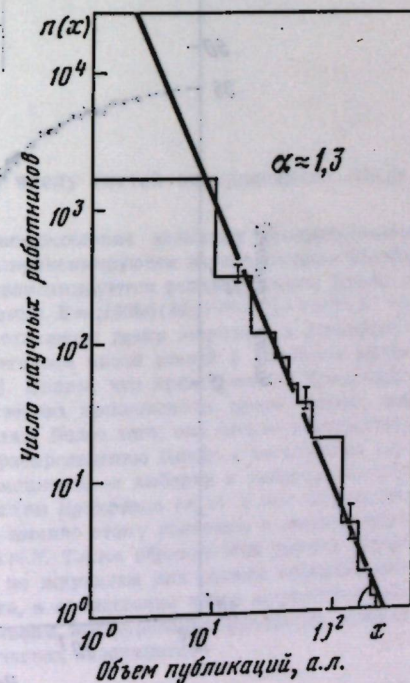


РИС. 2.16. Распределение научных работников Сибирского отделения АН СССР по суммарному объему публикаций (по данным [1.13])

Массив 2650 научных работников

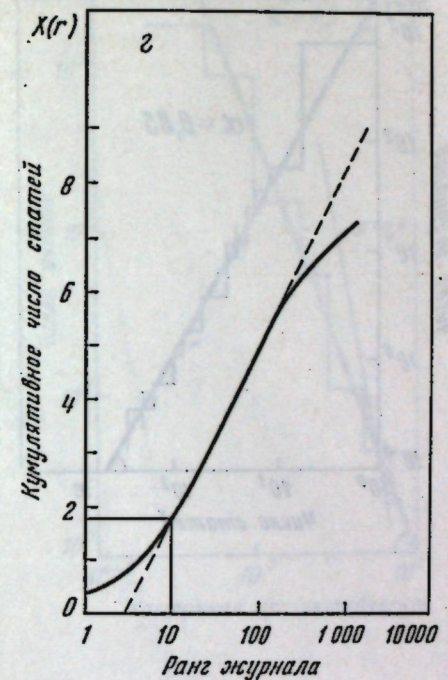
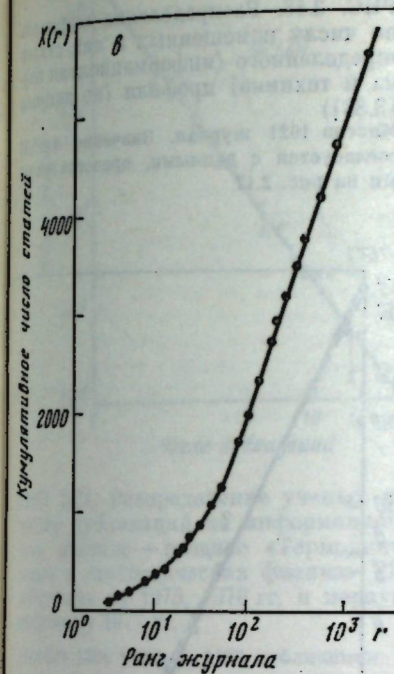
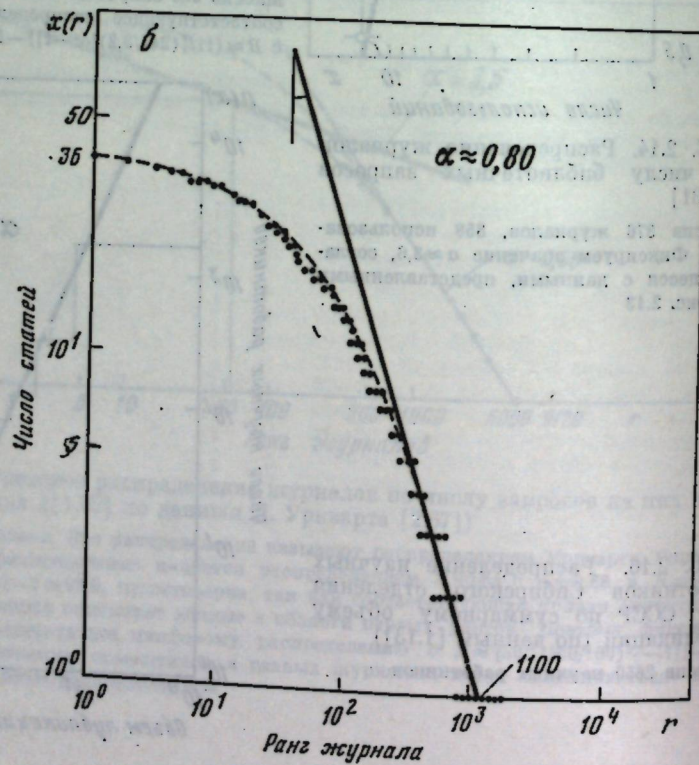
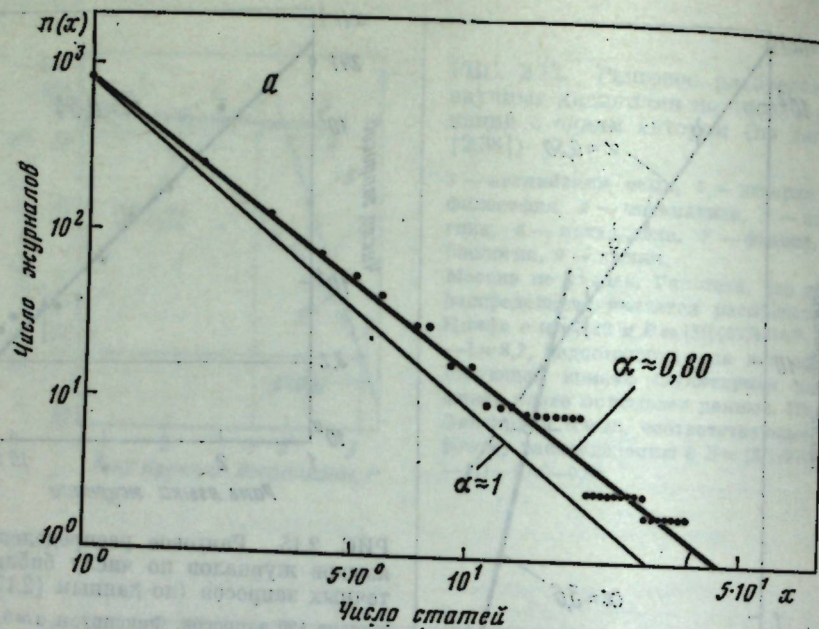


РИС. 2.17. Распределение журналов по числу статей определенного (медицинского) профиля (по данным [3.58])

Массив 1653 журнала, 5591 статья. Это распределение называют распределением Бродфорда. (а) Частотная форма. Данные аппроксимируются распределением Ципфа с $\alpha \approx 0,80$. (б) Ранговая форма. Данные аппроксимируются распределением Ципфа с $\alpha \approx 0,80$ и коэффициентом рангового искажения $B \approx \{1099 / [(36/1)^{0,80} - 1]\} - 1 \approx 65,3$. (в) Представление Бродфорда: по оси абсцисс отложены ранги журналов в логарифмическом масштабе, по оси ординат — кумулятивное число статей в линейном масштабе. (г) Закон рассеяния Бродфорда [3.83]. Видим, что представление Бродфорда, часто используемое сегодня в библиометрических приложениях, менее удобно, чем представление в логарифмических координатах. Более того, оно просто некорректно, так как прямая линия в нем соответствует распределению Ципфа с постоянным значением $\alpha = 1$. Между тем значения α могут меняться от выборки к выборке. Тангенс угла наклона линейного участка аппроксимации Бродфорда (в, г) к оси абсцисс равен A (см. выражение (7.2.13)). Для $\alpha = 1$ (а именно этому значению α соответствует аппроксимация Бродфорда) и $I \gg 1$ имеем $A \approx N$. Таким образом, этот наклон характеризует не степень рассеяния публикаций по журналам для разных информационных массивов (выборок), как принято думать, а единственно число журналов в рассматриваемой выборке. По всем этим причинам представление Бродфорда должно уступить место представлению в логарифмических координатах

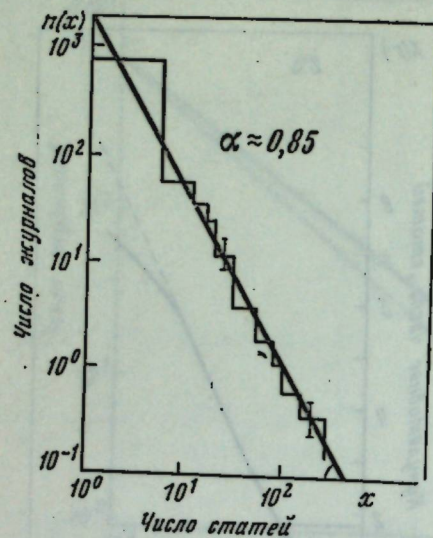


РИС. 2.18. Распределение журналов по числу помещенных в них статей определенного (информационная наука и техника) профиля (по данным [3.83])
Массив 1021 журнал. Значение $\alpha \approx 0,85$ согласуется с данными, представленными на рис. 2.17

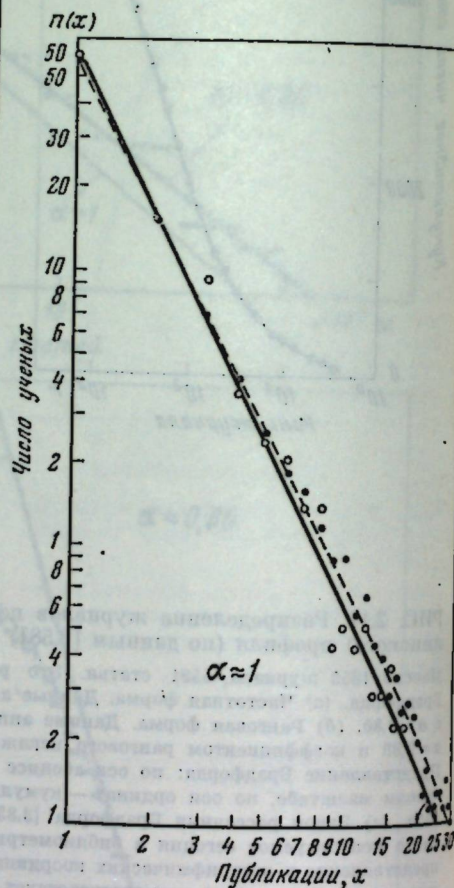


РИС. 2.20. Распределение химиков по числу публикаций (на материале «Auerbach's Tables» и «Chemical Abstracts» за 1907—1919 гг.) [3.74]
Массив не указан. По имени А. Лотки, чьи данные здесь представлены, распределение ученых по числу публикаций часто называют распределением Лотки

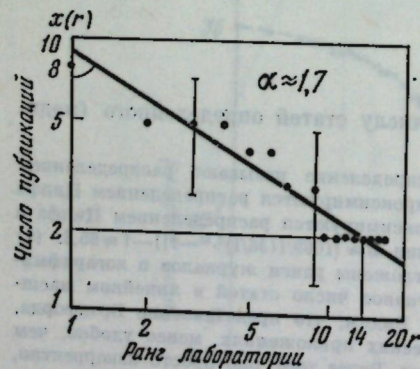


РИС. 2.19. Ранговое распределение лабораторий в области исследования керамических материалов по числу публикаций (по данным [4.142])

Массив 100 публикаций. Фиксируем $\alpha \approx 1,7$, соответствующее распределению Ципфа с $B \approx (13/[(8/2)^{1,7}-1]) - 1 \approx 0,32$

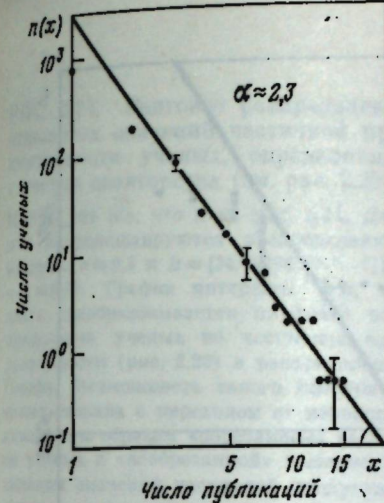


РИС. 2.21. Распределение ученых по числу публикаций на информационном массиве — разделе «Термодинамика и статистическая физика» РЖ «Физика» за 1975, 1976 гг. и первую половину 1977 гг.

Массив 1204 ученых, 1192 публикации

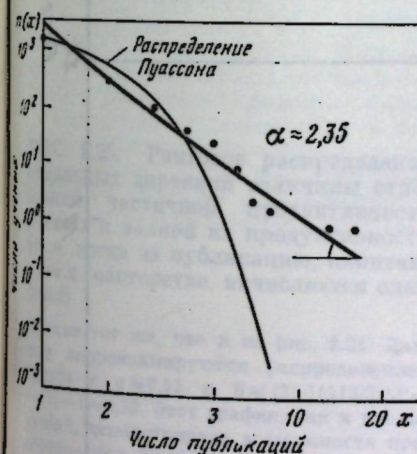


РИС. 2.22. Распределение ученых по числу публикаций в области военной техники (по данным [3.75])

Данные, как видим, могут быть аппроксимированы распределением Ципфа с $\alpha \approx 2,35$. Логарифмический масштаб позволяет увидеть непригодность распределения Пуассона, используемого Л. Мэнтеллом, чьи данные здесь представлены, и показанного на графике

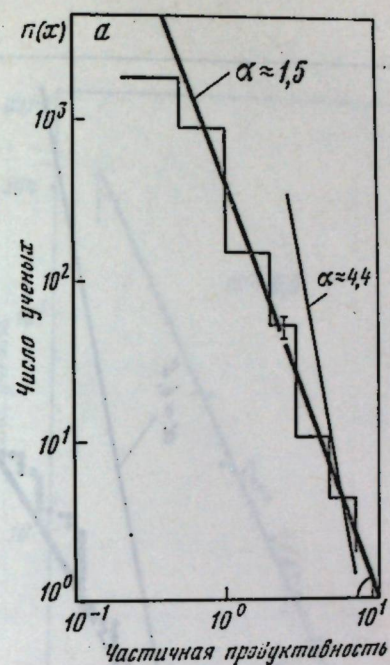


РИС. 2.23. Распределение ученых по частичной продуктивности (за публикацию, написанную n авторами, каждому начисляется балл $1/n$)

Массив тот же, что на рис. 2.21(а). Частотная форма. Данные аппроксимируются цифровым распределением с $\alpha \approx 1,5$. (б) Ранговая форма. Гипотеза, что данные описываются распределением Ципфа с $\alpha \approx 4,4$ и коэффициентом рангового искажения $B \approx (26/[(7,583/3,833)^{4,4}-1]) - 1 \approx 0,36$, несостоятельна с частотным графиком. Гипотеза, что данные описываются цифровым распределением с $\alpha \approx 1,5$ и $B \approx (264/[(7,583/1,00)^{1,5}-1]) - 1 \approx 12,3$, удовлетворительно описывает данные в ранговом представлении и согласуется с частотным графиком. Поэтому фиксируем $\alpha \approx 1,5$ (РИС. 2.23б см. на стр. 260)

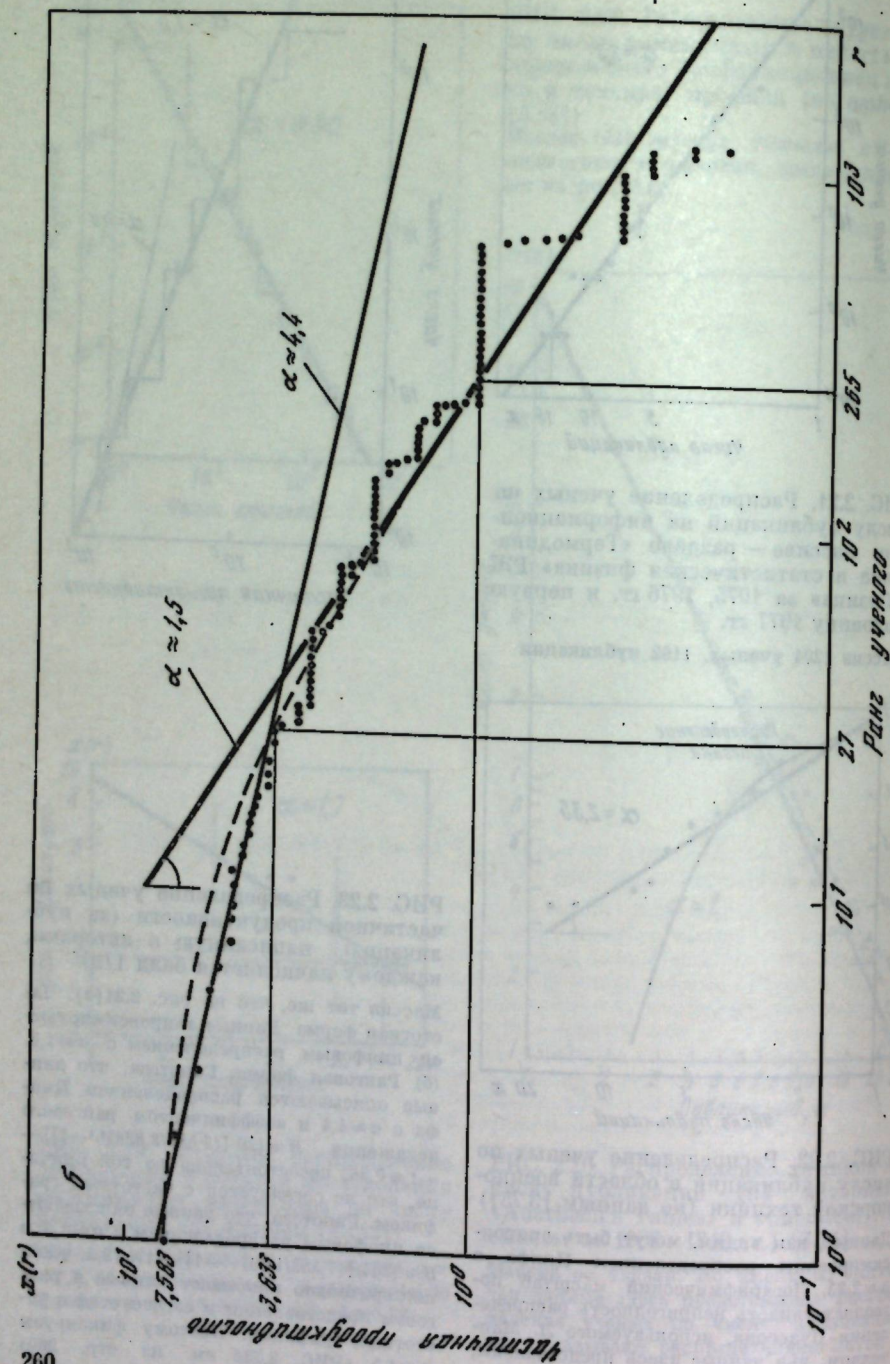


РИС. 2.24. Ранговое распределение отдельных значений частичной продуктивности ученых, определяемой с учетом соавторства (см. рис. 2.23)

Массив тот же, что и на рис. 2.21. Данные аппроксимируются распределением Ципфа с $\alpha \approx 0,5$ и $B \approx \{34/[(394/1)^{0,5} - 1] \} - 1 \approx 0,80$. График интересен тем, что здесь преобразовывается, цифровое распределение ученых по частичной продуктивности (рис. 2.23) в распределение Ципфа. Возможность такого преобразования связана с переходом от невоспроизводимым образом «обрезанной» выборки ученых к «необрезанной» выборке отдельных значений частичной продуктивности

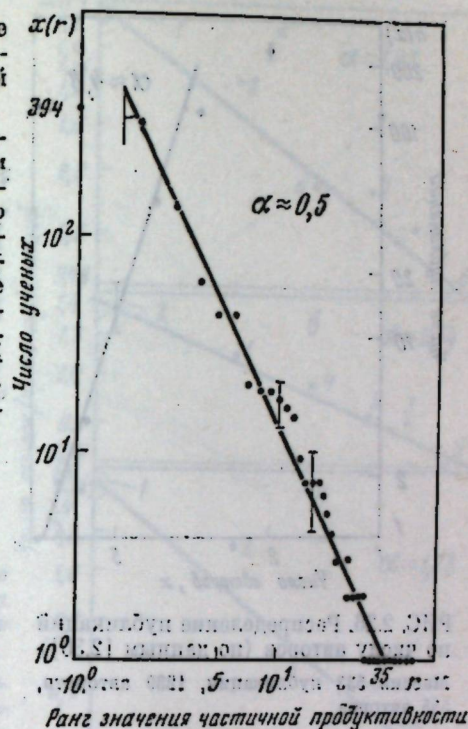
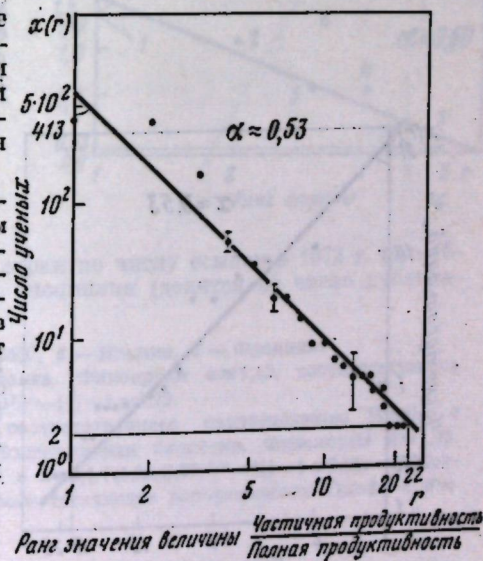


РИС. 2.25. Ранговое распределение отдельных значений величины отношения частичной продуктивности ученых к полной их продуктивности (т. е. когда за публикацию, написанную в соавторстве, начисляется один балл)

Массив тот же, что и на рис. 2.21. Данные аппроксимируются распределением Ципфа с $\alpha \approx 0,53$ и $B \approx \{21/[(413/2)^{0,53} - 1] \} - 1 \approx 0,32$. Этот график, как и предыдущий, демонстрирует возможность преобразования цифрового распределения в распределение Ципфа путем перехода от «обрезанной» выборки к «необрезанной»



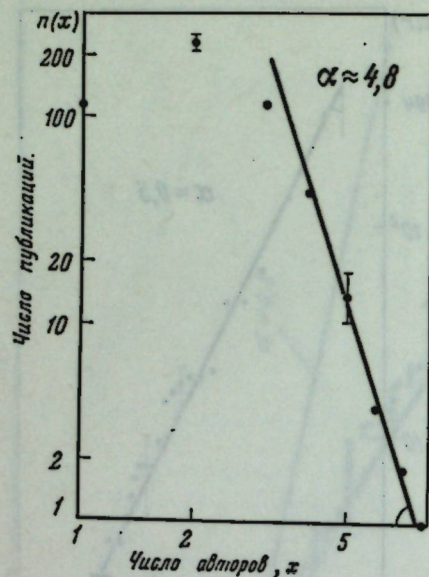


РИС. 2.26. Распределение публикаций по числу авторов (по данным [2.30])
Массив 533 публикации, 1239 авторов, 555 авторов

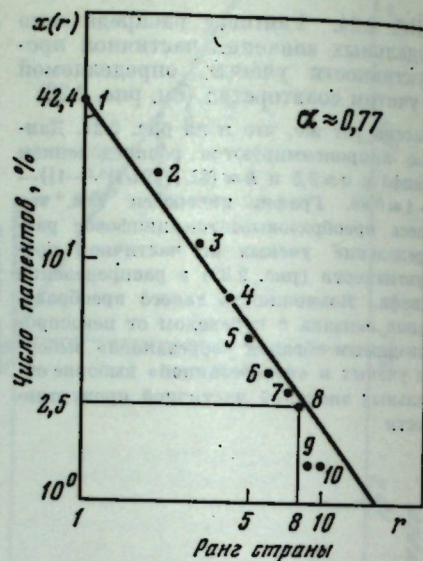


РИС. 2.27. Ранговое распределение 10 стран по числу патентов (по данным [1.118])

1 — США, 2 — ФРГ, 3 — Великобритания, 4 — Франция, 5 — Нидерланды, 6 — Италия, 7 — Швеция, 8 — Япония, 9 — Бельгия, 10 — Канада.
Массив не указан. Фиксируем $\alpha \approx 0,77$, соответствующее распределению Ципфа с коэффициентом рангового искажения $B \approx (7/((424/2,5)^{0,77} - 1)) - 1 \approx -0,11$

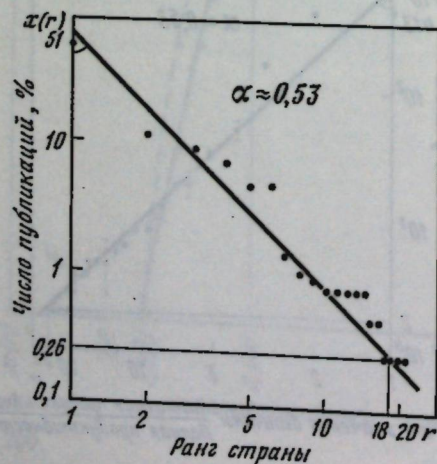


РИС. 2.28. Ранговое распределение стран по числу публикаций в области технологии изготовления рессор (по данным [3.14])

Массив не указан. Фиксируем $\alpha \approx 0,53$, соответствующее распределению Ципфа с $B \approx (17/((51/0,26)^{0,53} - 1)) - 1 \approx 0,10$

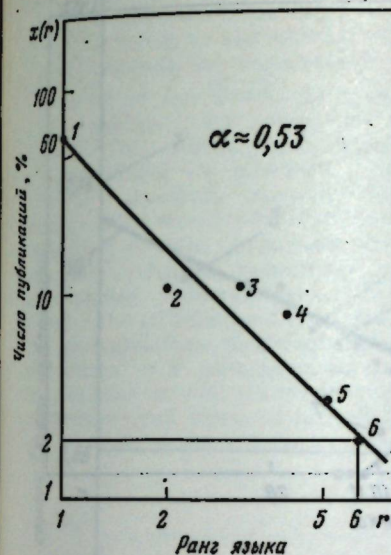


РИС. 2.29. Ранговое распределение языков публикаций по числу этих публикаций в мировой печати в 1957 г. (по данным [2.60])

1 — английский язык, 2 — русский, 3 — немецкий, 4 — французский, 5 — японский, 6 — испанский язык.
Массив не указан. Фиксируем $\alpha \approx 0,53$, соответствующее распределению Ципфа с $B \approx (5/((60/2)^{0,53} - 1)) \approx -0,01$

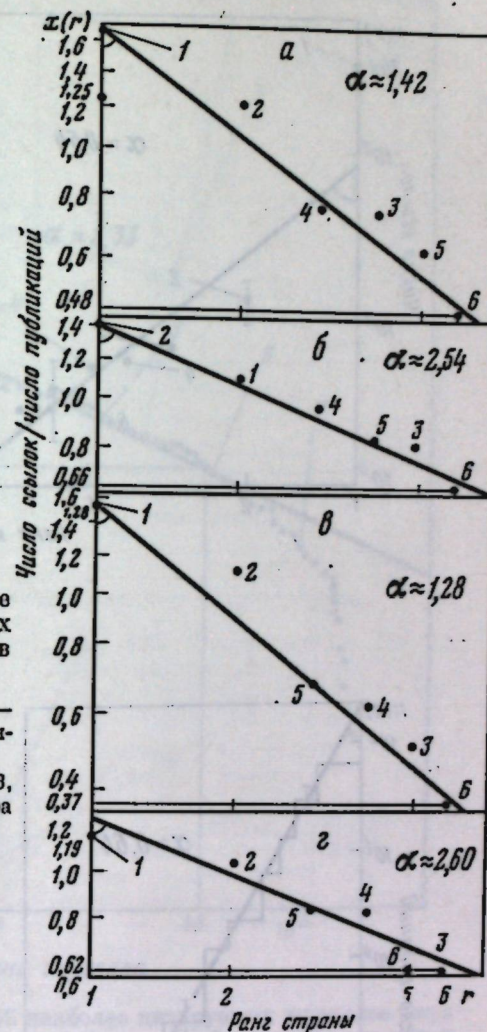


РИС. 2.30. Ранговое распределение стран по числу ссылок в 1972 г. на публикации этих стран для отдельных дисциплин (делится на число публикаций) (по данным [1.154])

1 — США, 2 — Англия, 3 — СССР, 4 — ФРГ, 5 — Япония, 6 — Франция.
Массив не указан. (а) Физика и геофизика. Фиксируем $\alpha \approx 1,42$, соответствующее распределению Ципфа с $B \approx (5/((1,25/0,48)^{1,42} - 1)) - 1 \approx 0,72$.
(б) Математика. Фиксируем $\alpha \approx 2,54$, соответствующее распределению Ципфа с $B \approx (5/((1,42/0,66)^{2,54} - 1)) - 1 \approx 0,20$. (в) Молекулярная биология. Фиксируем $\alpha \approx 1,28$, соответствующее распределению Ципфа с $B \approx (5/((1,28/0,37)^{1,28} - 1)) - 1 \approx 0,28$. (г) Технические науки. Фиксируем $\alpha \approx 2,60$, соответствующее распределению Ципфа с $B \approx (5/((1,19/0,62)^{2,60} - 1)) - 1 \approx 0,12$

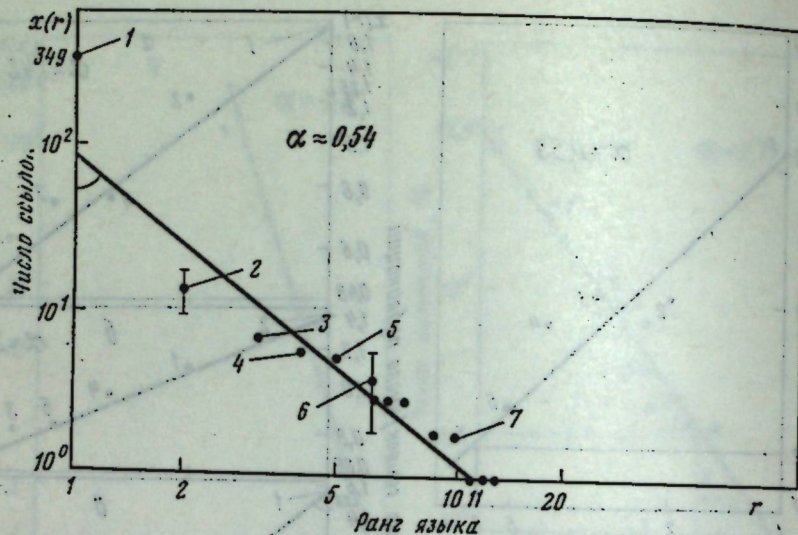


РИС. 2.31. Ранговое распределение языков цитируемых источников по числу ссылок в геологической литературе библиотек США (по данным [4.64])

1 — английский язык, 2 — французский, 3 — японский, 4 — немецкий, 5 — русский, 6 — итальянский, 7 — датский язык. Массив 397 ссылок. Фиксируем $\alpha \approx 0,54$, соответствующее распределение Ципфа с коэффициентом рангового искажения $B \approx (10/[(349/1)^{0,54} - 1]) - 1 \approx -0,56$

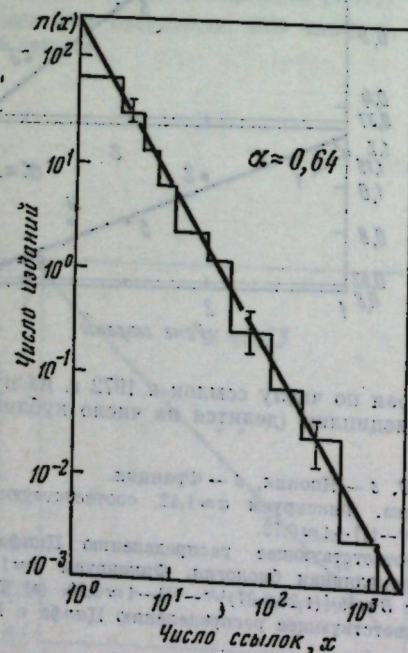


РИС. 2.32. Распределение изданий в области нефтяной науки и техники по их цитированию (число ссылок не делится на число публикаций в издании) (по данным [4.20])

Массив 4290 ссылок; 179 изданий

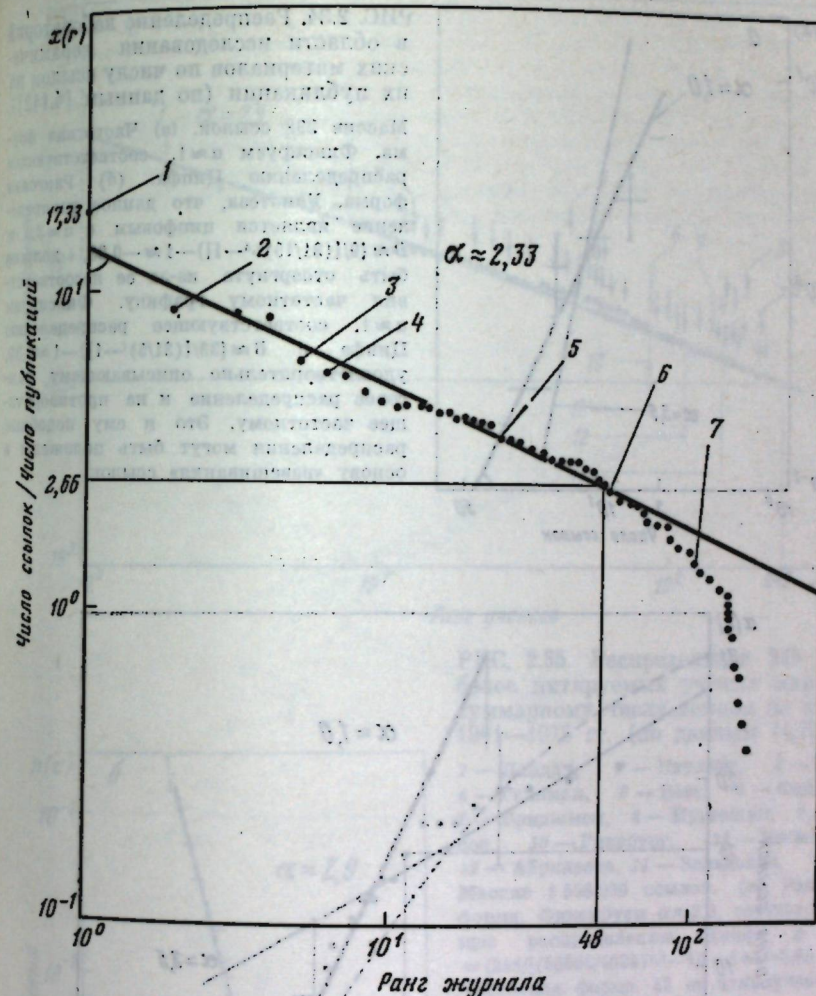


РИС. 2.33. Ранговое распределение 152 наиболее цитируемых журналов мира по числу ссылок, сделанных в 1969 г. на статьи, опубликованные в них в 1967—1968 гг., деленному на число этих публикаций (по так называемому импакт-фактору (impact-factor)) (по данным [4.77])

1 — Physiological Rev., 2 — Mol. Biol., 3 — Biol. Chem., 4 — Amer. Chem. Soc., 5 — Phys. Rev., 6 — IEEE, 7 — Radiology.

Массив 130 000 публикаций, 450 000 ссылок. Фиксируем $\alpha \approx 2,33$, что соответствует ципфовому распределению с $B \approx (48/[(17,33/2,66)^{2,33} - 1]) - 1 \approx -0,38$. Число ссылок делится на число публикаций, если нужно, чтобы при оценках цитирования толстые журналы не имели преимущества перед тонкими. Данное и ему подобные распределения могут быть положены в основу «взвешивания» ссылок

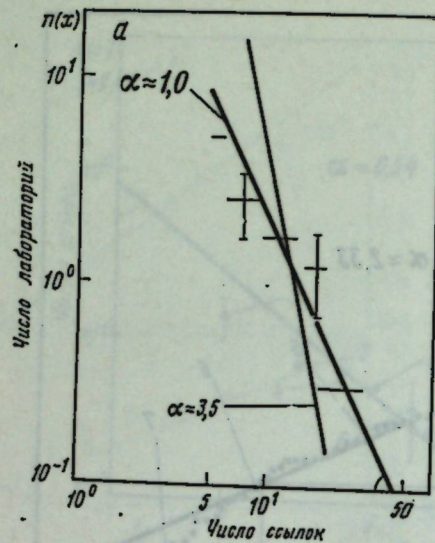


РИС. 2.34. Распределение лабораторий в области исследования керамических материалов по числу ссылок на их публикации (по данным [4.142])

Массив 998 ссылок. (а) Частотная форма. Фиксируем $\alpha \approx 1$, соответствующее распределению Ципфа. (б) Ранговая форма. Гипотеза, что данное распределение является ципфовым с $\alpha \approx 3,5$ и $B \approx [9/(31/15)^{3,5} - 1)] - 1 \approx -0,23$, должна быть отвергнута из-за ее несоответствия частотному графику. Фиксируем $\alpha \approx 1$, соответствующее распределению Ципфа с $B \approx [33/((31/5)^1 - 1)] - 1 \approx 5,35$, удовлетворительно описывающему ранговое распределение и не противоречащее частотному. Это и ему подобные распределения могут быть положены в основу «взвешивания» ссылок

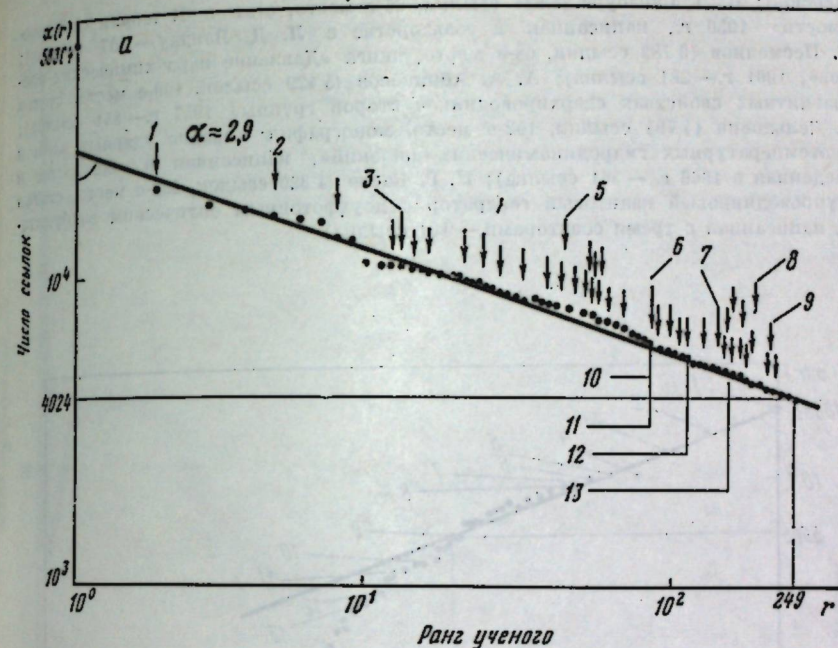
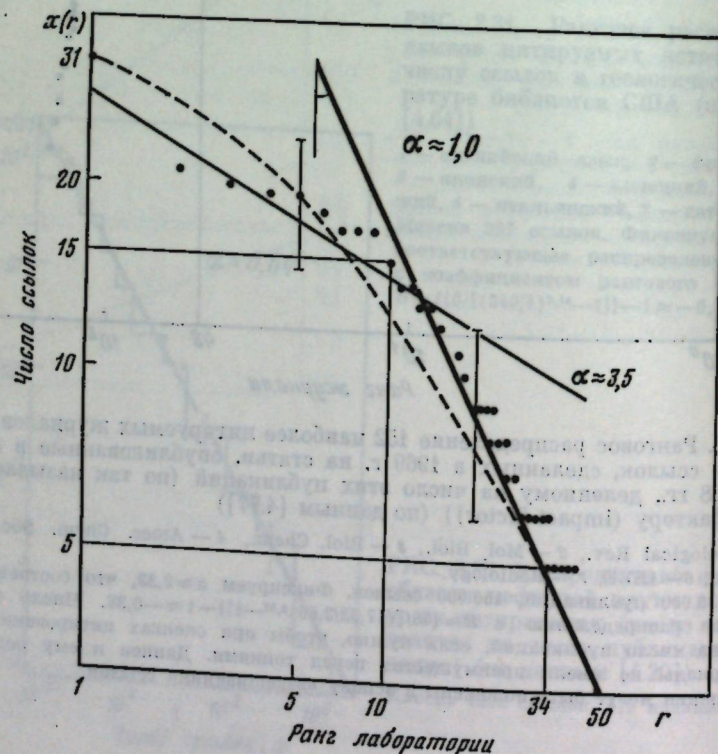
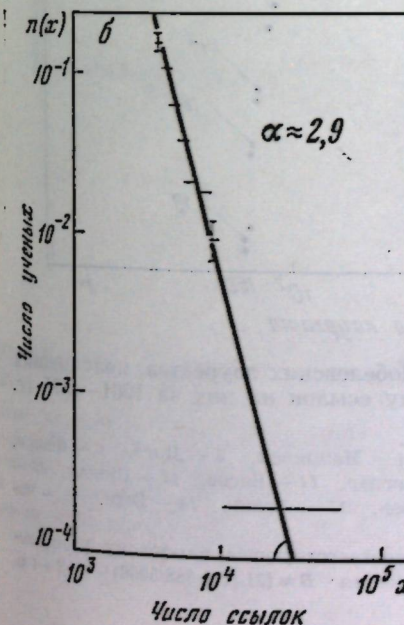


РИС. 2.35. Распределение 249 наиболее цитируемых ученых мира по суммарному числу ссылок на них в 1961—1975 гг. (по данным [4.79])

1 — Ландау, 2 — Паулинг, 3 — Мотт, 4 — Геллман, 5 — Бето, 6 — Фейнман, 7 — Бриджмен, 8 — Пригожин, 9 — Басов, 10 — Гинзбург, 11 — Носмеилов, 12 — Абрикосов, 13 — Зельдович.

Массив 1 556 000 ссылок. (а) Ранговая форма. Фиксируем $\alpha \approx 2,9$, соответствующее распределению Ципфа с $B \approx [248/(58804/4024)^{2,9} - 1)] - 1 \approx -0,89$. (б) Частотная форма. 42 из этих ученых — нобелевские лауреаты (они помечены на графике стрелками). Возглавляют список О. Лоури (58 304 ссылки, 1-о место. Наиболее цитируемая его публикация — статья «Протениновое изоморфие с фенилфеноловым реагентом», написанная в 1951 г. с тремя соавторами (50 010 ссылок)). Список наиболее цитируемых советских ученых возглавляют: Л. Д. Ландау (18 888 ссылок, 2-о место, наиболее цитируемая работа — «Теория сверхтекучести гелия II», 1941 г.)



214 ссылок); В. Л. Гинзбург (6 834 ссылки, 66-е место, статья «К теории сверхпроводимости» 1950 г., написанная в соавторстве с Л. Д. Ландау.— 407 ссылок); А. Н. Несмеянов (6 783 ссылки, 68-е место, книга «Давление пара химических элементов», 1961 г.— 281 ссылка); А. А. Абрикосов (5 429 ссылок, 109-е место, статья «О магнитных свойствах сверхпроводников второй группы» 1957 г.— 851 ссылка); Я. Б. Зельдович (4 794 ссылки, 152-е место, монография «Физика ударных волн и высокотемпературных гидродинамических явлений», написанная в соавторстве и переведенная в 1966 г.— 361 ссылка); Н. Г. Басов (4 320 ссылок, 202-е место, статья «Полупроводниковый квантовый генератор с двухфотонным оптическим возбудителем», написанная с тремя соавторами,— 101 ссылка)

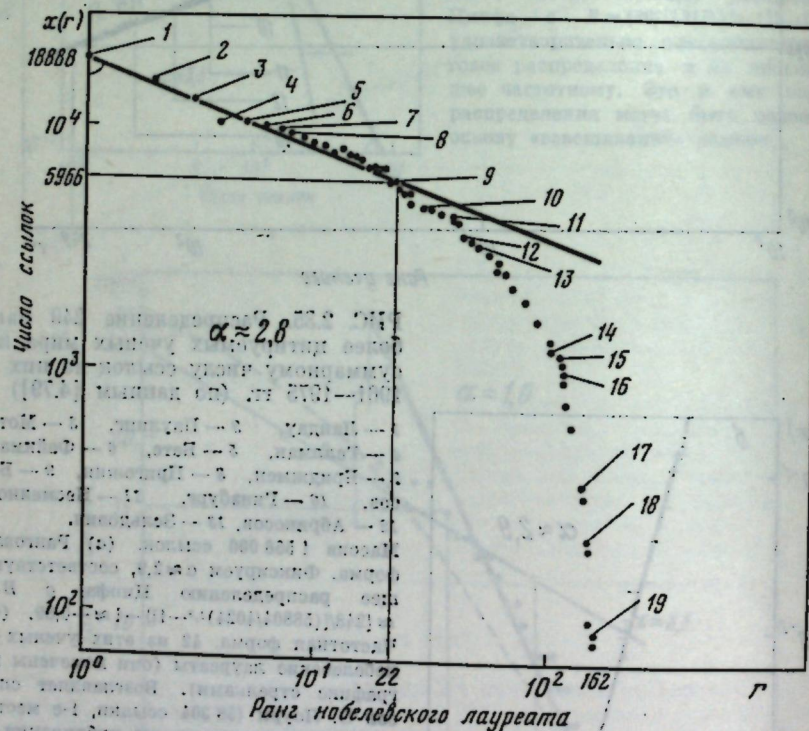


РИС. 2.36. Ранговое распределение 162 Нобелевских лауреатов, получивших свои премии, начиная с 1950 г., по числу ссылок на них за 1961—1975 гг. (по данным [4.79])

1 — Ландау, 2 — Паулинг, 3 — Герцберг, 4 — Малликен, 5 — Мотт, 6 — Флори, 7 — Энклес, 8 — Бори, 9 — Фейнман, 10 — Вигнер, 11 — Басов, 12 — Шокли, 13 — Бор, 14 — Семенов, 15 — Тамм, 16 — Прохоров, 17 — Франк, 18 — Бете, 19 — Черенков

Массив 475 000 ссылок. Фиксируем $\alpha \approx 2,8$, соответствующее цифровому распределению с коэффициентом рангового искажения $B \approx \{21/[(18\,888/5966)^{2,8} - 1]\} - 1 \approx -0,13$

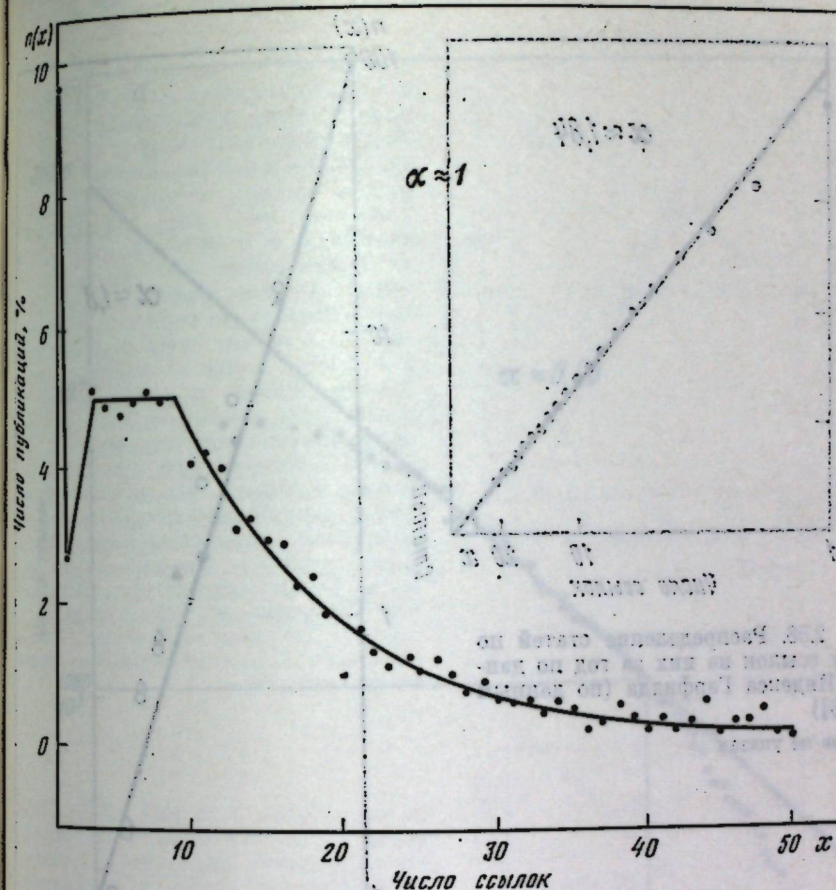


РИС. 2.37. Распределение публикаций по числу содержащихся в них ссылок [4.24]

Массив не указан. Координаты — линейные. Согласно Д. Прайсу, дающему это распределение, для большого числа ссылок оно является распределением Цифа с указанным на графике значением α . Форма распределения в области малых значений x в определенной степени объясняется невоспроизводимым «обрезанием» выборок научных публикаций

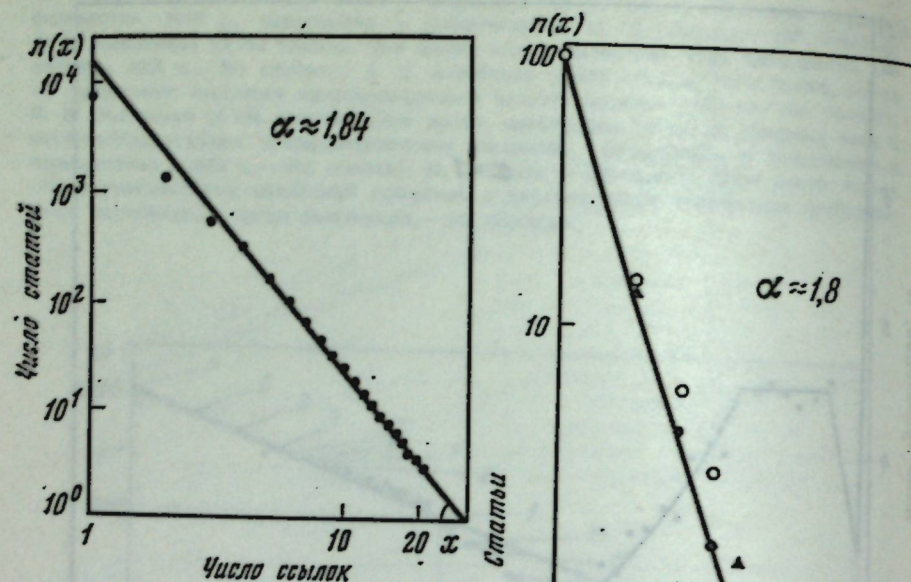


РИС. 2.38. Распределение статей по числу ссылок на них за год по данным Индекса Гарфилда (по данным [1.143])

Массив не указан

РИС. 2.39. Распределение статей по генетике по их цитированию в 1961 г. [1.131]

Массив не указан. С. Наранан, чьи данные здесь приведены, использовал для их получения специализированные указатели научных ссылок по генетике

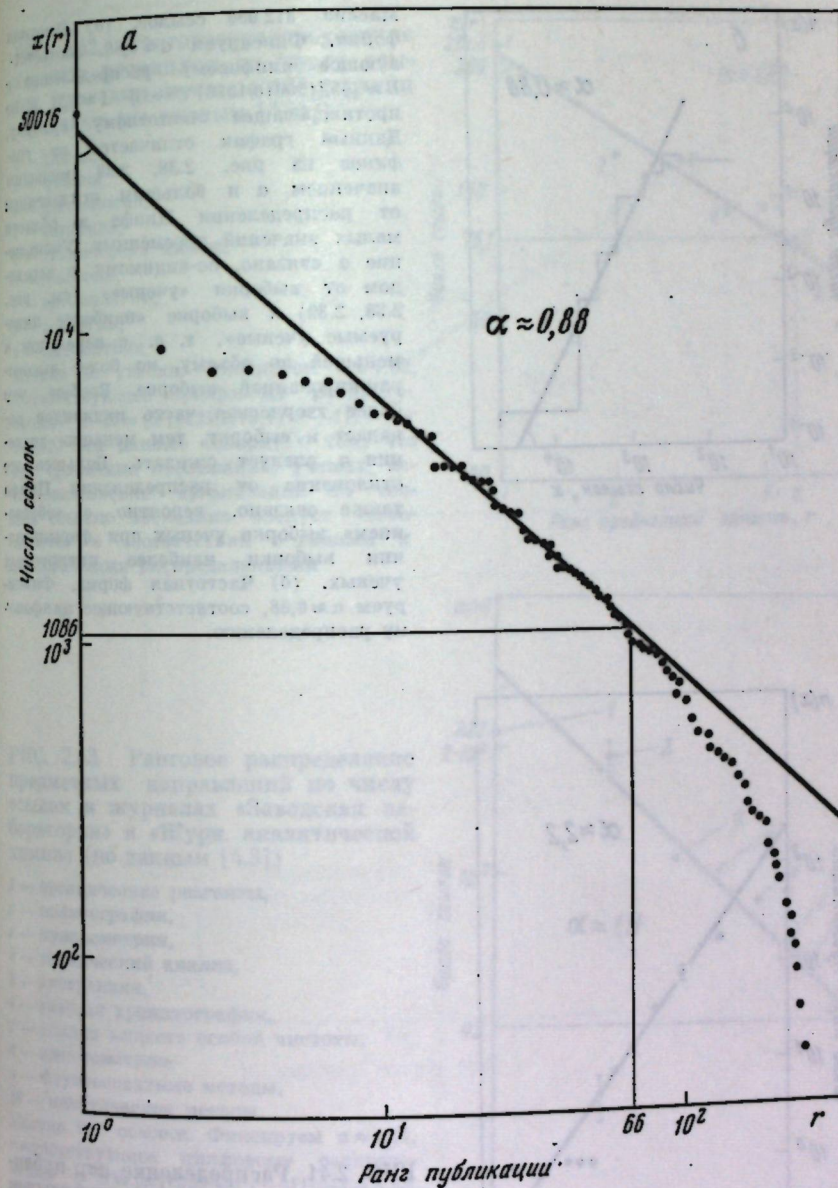
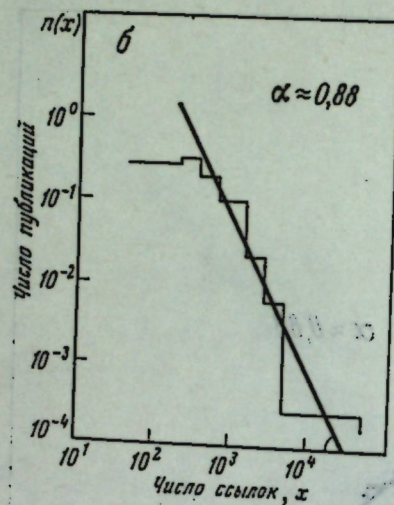


РИС. 2.40. Распределение 249 публикаций наиболее цитируемых в 1961—1975 гг. ученых (по одной наиболее цитируемой публикации каждого ученого) по числу ссылок на эти публикации (по данным [4.79])



Массив 312 000 ссылок. (а) Ранговая форма. Фиксируем $\alpha \approx 0,88$, соответствующее ципфовому распределению с $B \approx (65/[(50016/1086)^{0,88} - 1]) - 1 \approx 1,31$ и не противоречащее частотному графику. Данный график отличается от графиков на рис. 2.38, 2.39 меньшим значением α и большим отклонением от распределения Ципфа в области малых значений переменных. Уменьшение α связано, по-видимому, с переходом от выборки «ученые» (см. рис. 2.38, 2.39) к выборке «наиболее цитируемые ученые», т. е. с переходом к меньшей по объему, но более высоко-ранжированной выборке. Вообще, чем более творческая часть индивидов попадает в выборку, тем меньше значения α следует ожидать. Большее же отклонение от распределения Ципфа также связано, вероятно, с «обрезанием» выборки ученых при формировании выборки наиболее цитируемых ученых. (б) Частотная форма. Фиксируем $\alpha \approx 0,88$, соответствующее ципфовому распределению.

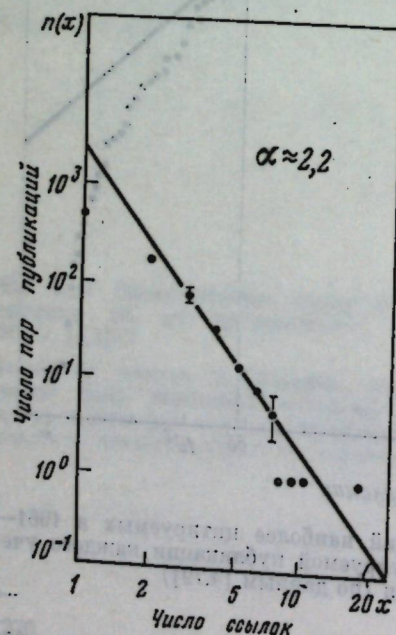


РИС. 2.41. Распределение пар публикаций по числу совпадающих ссылок в них (по данным [1.1]).

Массив 841 пара публикаций

РИС. 2.42. Ранговое распределение научных дисциплин по числу ссылок, приходящемуся на 100 страниц периодики (по данным [4.51])

- 1 — химия,
- 2 — физика,
- 3 — физиология,
- 4 — ботаника,
- 5 — техническая химия,
- 6 — геология,
- 7 — энтомология,
- 8 — зоология,
- 9 — математика

Массив не указан. Фиксируем $\alpha \approx 1,62$, соответствующее ципфовому распределению с $B \approx (5/[(224,8/75,1)^{1,62} - 1]) - 1 \approx 0,02$. Эти данные говорят о том, что при сравнении публикаций, ученых, исследовательских организаций по числам ссылок последние следует «взвешивать» в соответствии с данным и аналогичными распределениями

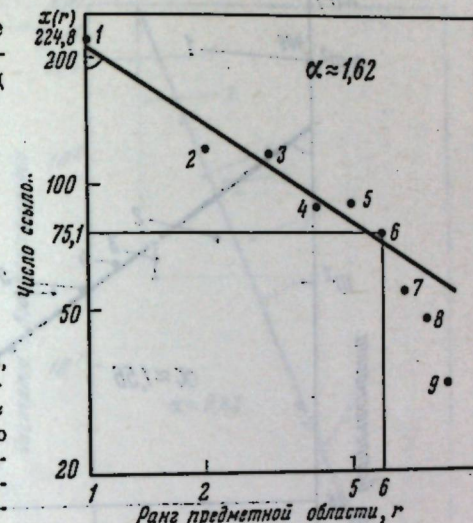
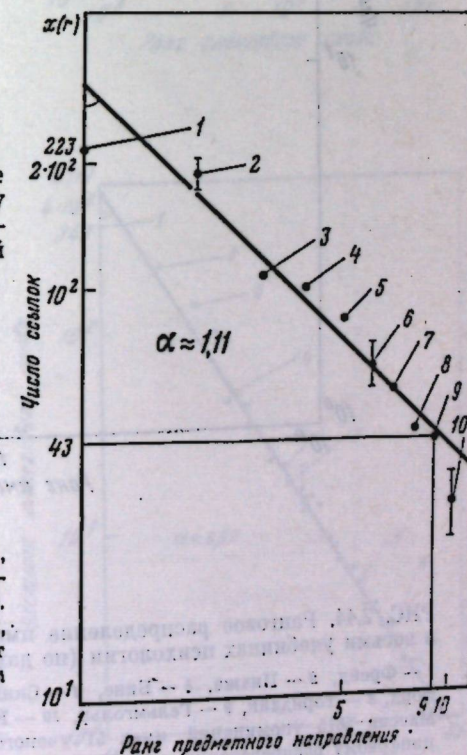


РИС. 2.43. Ранговое распределение предметных направлений по числу ссылок в журналах «Заводская лаборатория» и «Журн. аналитической химии» (по данным [4.3])

- 1 — органические реагенты,
- 2 — полярография,
- 3 — кулонометрия,
- 4 — технический анализ,
- 5 — экстракция,
- 6 — газовая хроматография,
- 7 — анализ веществ особой чистоты,
- 8 — амперометрия,
- 9 — флуоресцентные методы,
- 10 — кинетические методы

Массив 962 ссылки. Фиксируем $\alpha \approx 1,11$, соответствующее ципфовому распределению с $B \approx (8/[(223/43)^{1,11} - 1]) - 1 \approx 0,53$. Это распределение, как и на рис. 2.42, показывает, что числа ссылок следует «взвешивать» в соответствии с данным и подобными распределениями



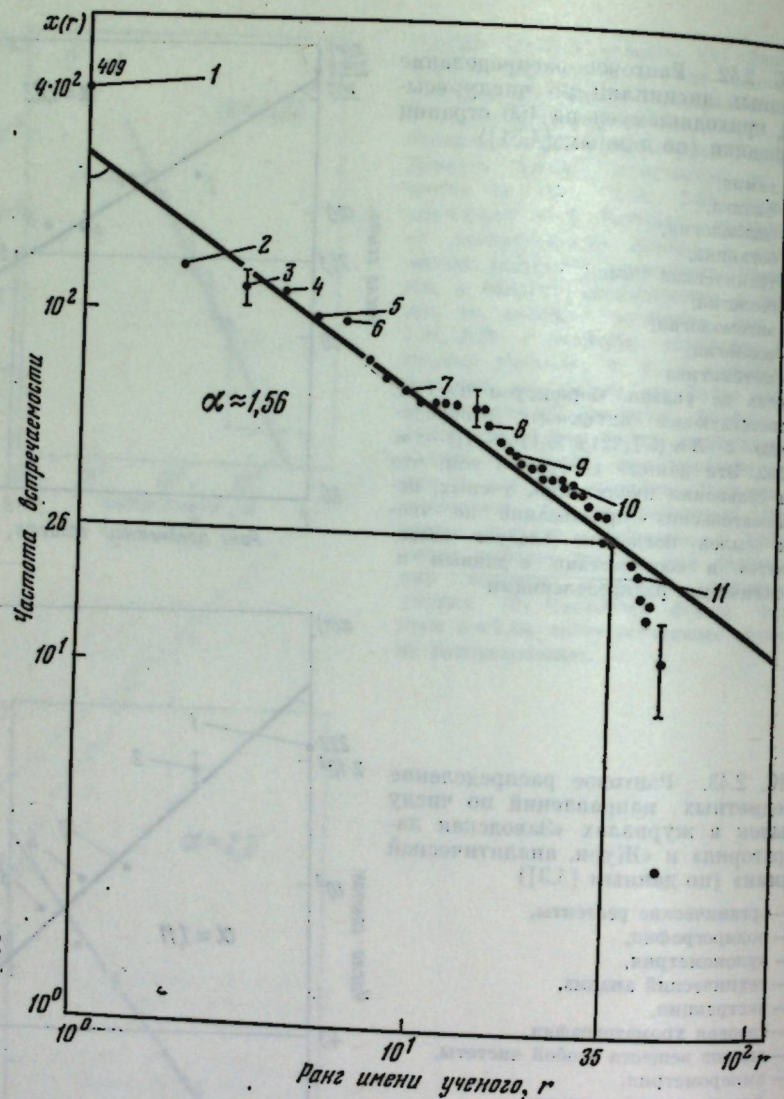


РИС. 2.44. Ранговое распределение имен ученых по частоте встречаемости в восьми учебниках психологии (по данным [2.57])
 1 — Фрейд, 2 — Пиаже, 3 — Бине, 4 — Скиннер, 5 — Павлов, 6 — Миллер, 7 — Гилфорд, 8 — Торндайк, 9 — Гельмгольц, 10 — Кеннон, 11 — Фехнер
 Массив 2615 упоминаний имен 51 ученого. Фиксируем $\alpha \approx 1,56$, соответствующее цифровому распределению с $B \approx (34/(409/26)^{1,56} - 1)) - 1 \approx -0,53$

РИС. 2.45. Ранговое распределение сленговых (определение см. гл. 6) слов по частоте встречаемости в §§ 6 и 9 статьи А. Эйнштейна ([0.61], с. 7—35)

- 1 — координата,
- 2 — вектор,
- 3 — магнитный,
- 4 — с (скорость света),
- 5 — заряд,
- 6 — уравнение,
- 7 — измерение

Массив 1023 словоупотребления 45 разных сленговых слов. Фиксируем $\alpha \approx 0,43$, соответствующее распределению Ципфа с $B \approx (28/((304/1)^{0,43} - 1)) - 1 \approx 1,62$

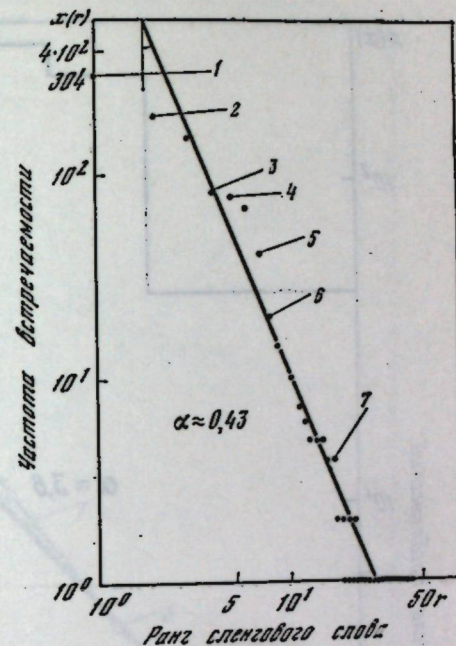
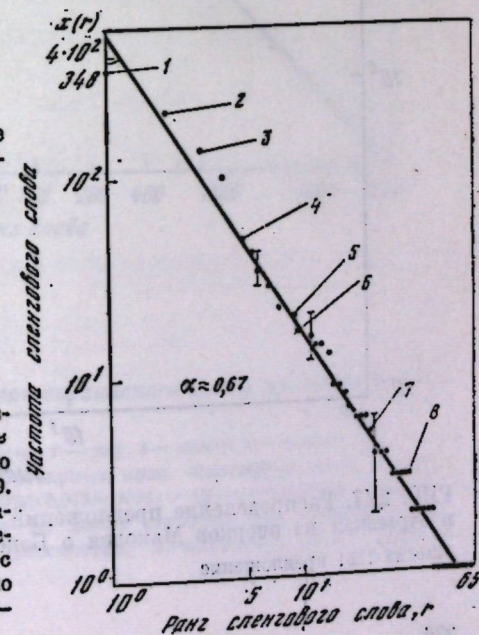


РИС. 2.46. Ранговое распределение сленговых слов по частоте встречаемости в статьях И. Лауба [0.92]

- 1 — координата,
- 2 — с (скорость света),
- 3 — преломление,
- 4 — абсолютный,
- 5 — диэлектрик,
- 6 — вакуум,
- 7 — динамика,
- 8 — Допплер.

Массив 1190 словоупотреблений 81 разного сленгового слова. Фиксируем $\alpha \approx 0,67$, соответствующее распределению Ципфа с $B \approx (64/((348/1)^{0,67} - 1)) - 1 \approx 0,29$. Это распределение отличается от показанного на рис. 2.45 большим значением α , что связано, по-видимому, с большей творческостью Эйнштейна по сравнению с Лаубом (ср. рис. 2.76—2.78)



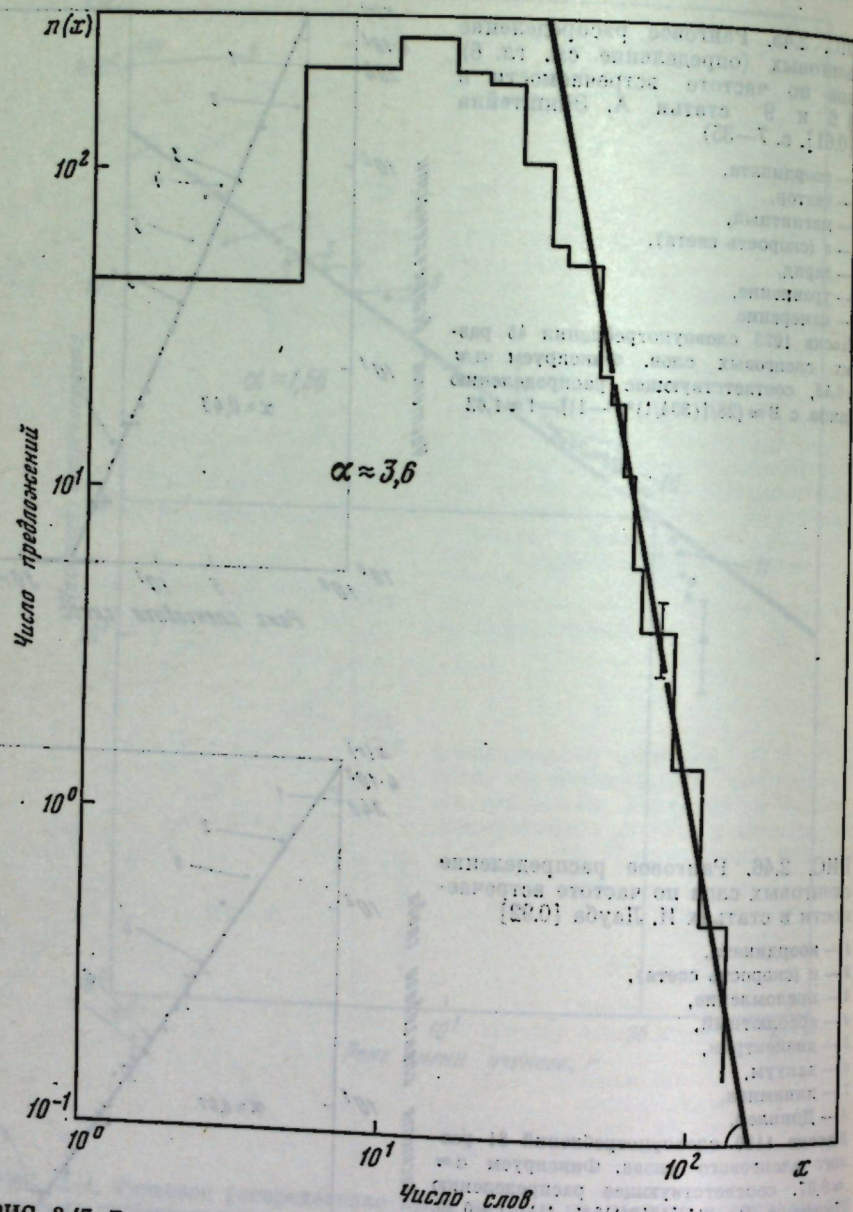


РИС. 2.47. Распределение предложений по числу содержащихся в них слов в отрывках из очерков Маколей о Бэконе и Чатаме (по данным [0.34])
Массив 1251 предложение

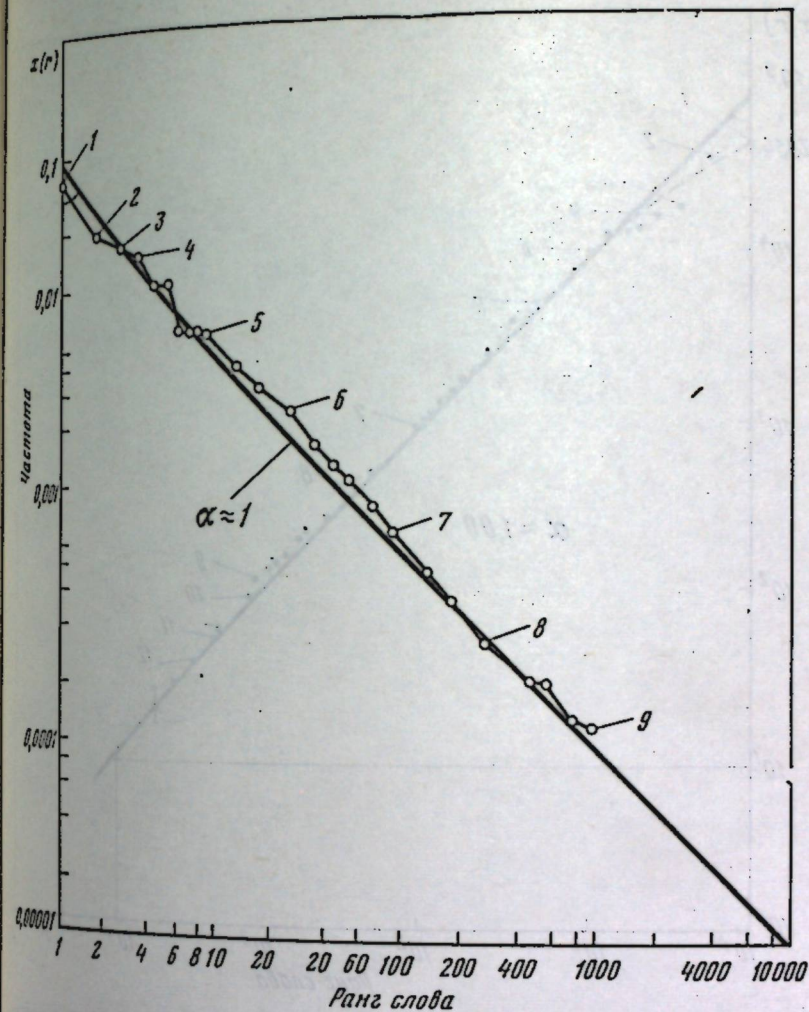


РИС. 2.48. Ранговое распределение слов английского языка по частоте встречаемости [6.20]

1 — the, 2 — of, 3 — and, 4 — to, 5 — I, 6 — or, 7 — say, 8 — really, 9 — quality
Массив 80 000 словоупотреблений 8727 словарных слов. Фиксируем $\alpha \approx 1$, соответствующее распределению Ципфа с $B \approx (999 / ((0,075 / 0,0001)^1 - 1)) - 1 \approx 0,33$. Это распределение называют распределением Ципфа по имени детально исследовавшего его Гатто. Впервые это распределение рассмотрел, по-видимому, Ж.-Б. Эсту в 1916 г.

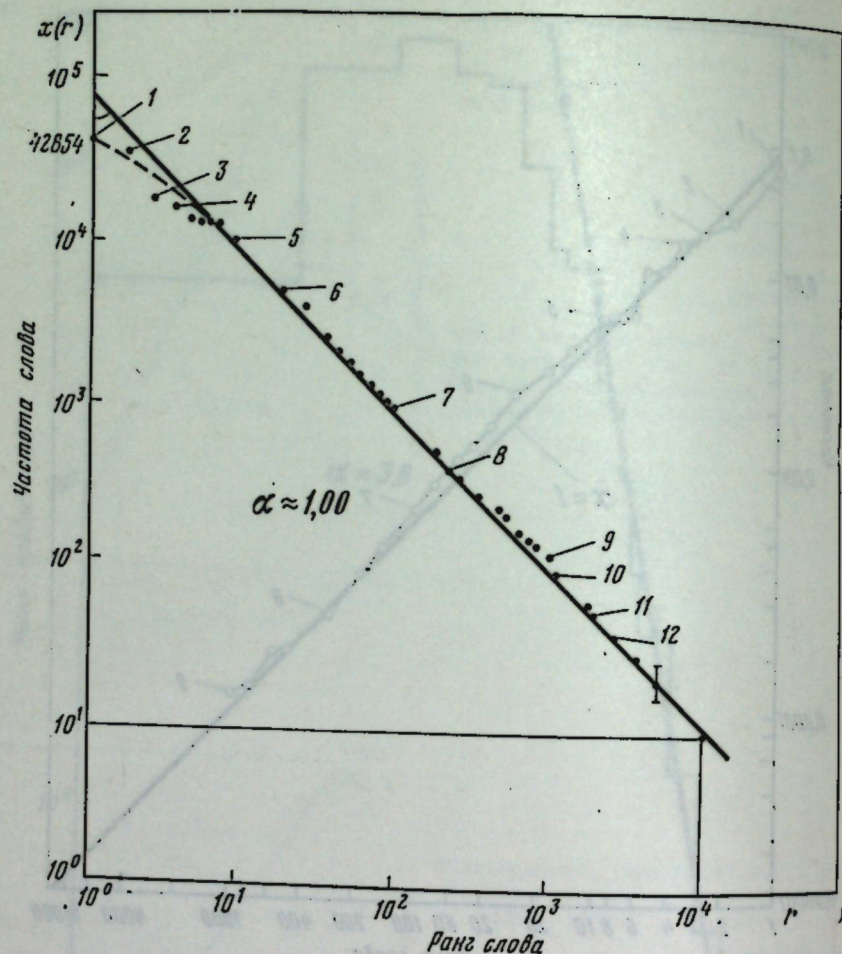


РИС. 2.49. Ранговое распределение слов русского языка по частоте встречаемости (по данным [6.15])

1 — в (во), 2 — и, 3 — не, 4 — на, 5 — а, 6 — кан, 7 — каждый, 8 — огромный, 9 — красота, 10 — могила, 11 — любитель, 12 — поучать. Массив 1 056 382 словоупотребления 39 268 разных лексем. Фиксируем $\alpha \approx 1$, соответствующее распределению Ципфа с коэффициентом рангового искажения $B \approx (9999/[(42\,854/10)^{1.00} - 1]) - 1 \approx 1.33$. Данное распределение отклоняется при малых значениях x от «рангового распределения Ципфа» $x(r) \sim 1/r^\alpha$ (которым оперировал Г. Ципф), представляющего собой в логарифмических координатах прямую линию. Б. Мандельброт предложил для описания этого отклонения свое распределение $x(r) \sim 1/(r+B)^\alpha$, получившее название закона Мандельброта и являющееся ранговой формой распределения Ципфа с отличным от нуля эффектом рангового искажения (см. гл. 7). Пунктирная линия на графике описывает этот эффект

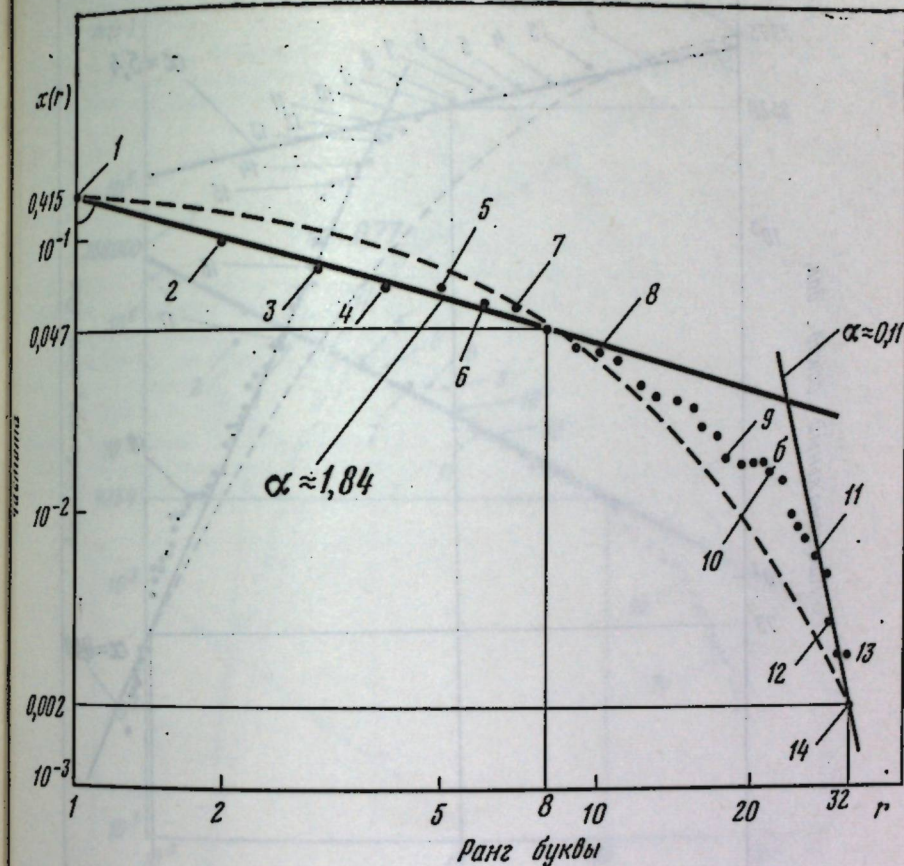


РИС. 2.50. Ранговое распределение букв русского языка по частоте встречаемости (по данным [0.12])

1 — (тире), 2 — о, 3 — е, 4 — а, 5 — и, 6 — т, 7 — н, 8 — с, 9 — л, 10 — б, 11 — ю, 12 — ц, 13 — у, 14 — ф. Массив не указан. Гипотеза, что данное распределение является распределением Ципфа с $\alpha \approx 0.11$ и $B \approx (31/[(0.145/0.002)^{0.11} - 1]) - 1 \approx 50.5$, недостоверна, так как плохо описывает данные (пунктирная линия на графике). Фиксируем $\alpha \approx 1.84$, соответствующее ципфовому распределению с $B \approx (7/[(0.145/0.047)^{1.84} - 1]) - 1 \approx 0.01$

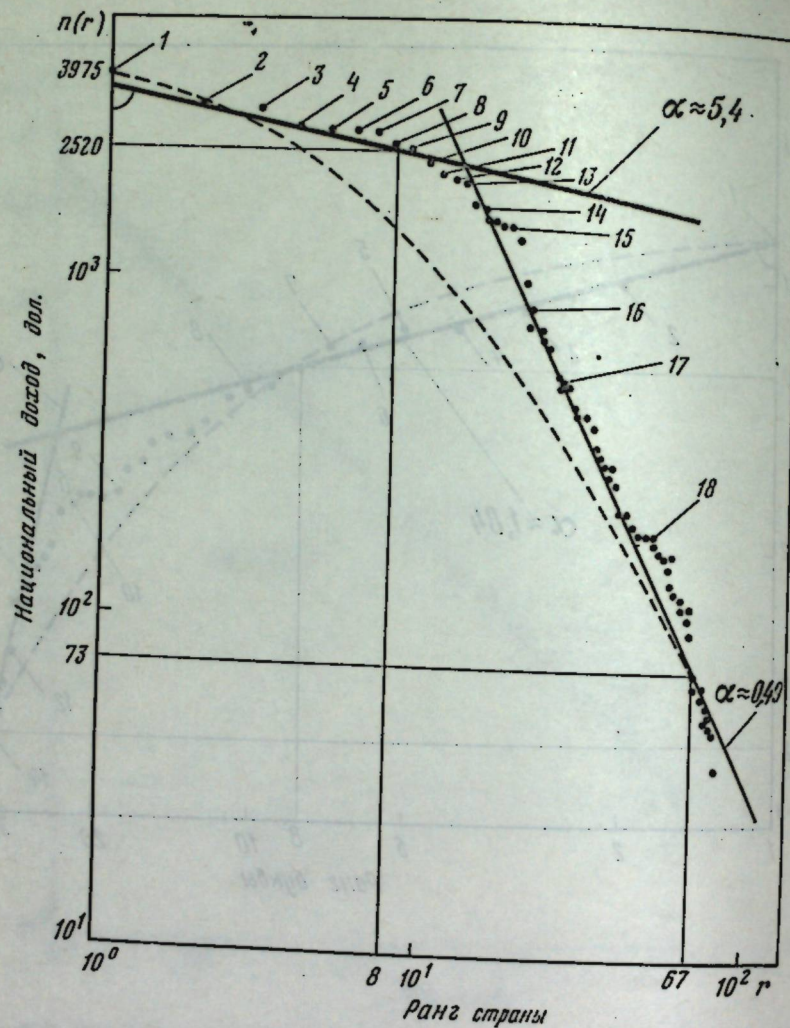


РИС. 251. Ранговое распределение стран по национальному доходу на душу населения (по данным [0.45], с. 123—125)

1 — США, 2 — ФРГ, 3 — Дания, 4 — Канада, 5 — Швеция, 6 — Нидерланды, 7 — Франция, 8 — Швейцария, 9 — Бельгия, 10 — Япония, 11 — Норвегия, 12 — Австралия, 13 — СССР, 14 — Великобритания, 15 — Италия, 16 — Испания, 17 — Турция, 18 — Бразилия

Национальный доход рассчитывается здесь по официальному курсу валют согласно методике, принятой в статистике СССР, т. е. без учета доходов, полученных в производственной сфере. Массив не указан. Гипотеза, что данное распределение является распределением Ципфа с $\alpha \approx 0,49$ и $B \approx \{66 / [(3975/73)^{0,49} - 1]\} - 1 \approx 9,84$, несостоятельна, так как плохо описывает данные (пунктирная линия на графике). Фиксируем $\alpha \approx 5,4$, соответствующее ципфовому распределению с $B \approx \{7 / [(3975/2520)^{5,4} - 1]\} - 1 \approx -0,34$

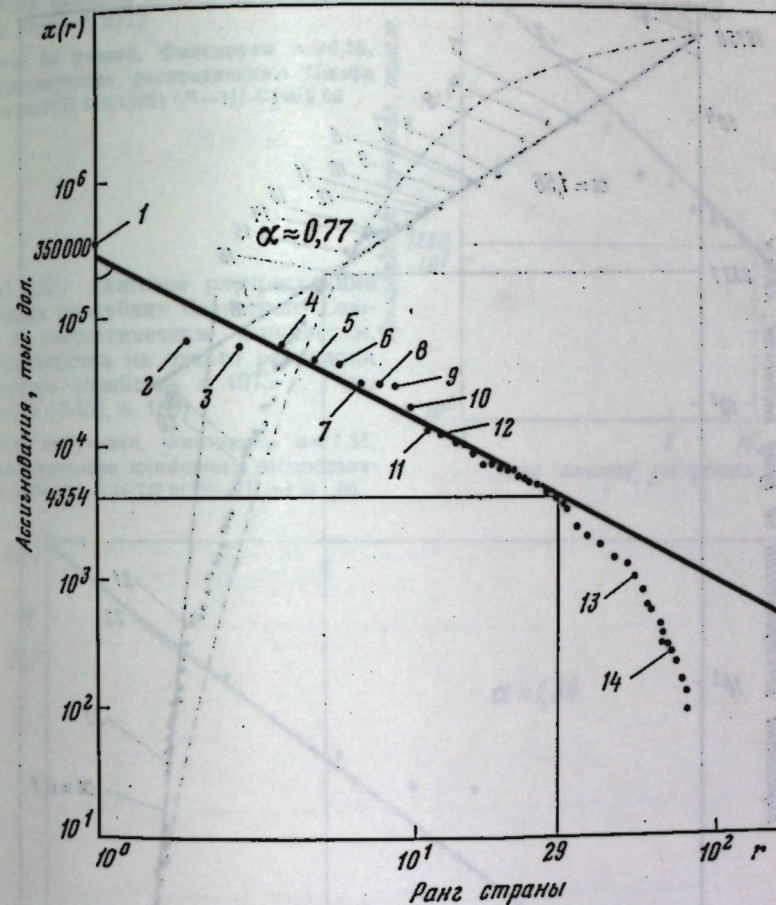


РИС. 252. Ранговое распределение стран по ассигнованиям на исследования в сельском хозяйстве (по данным [2.42])

1 — США, 2 — Япония, 3 — Австралия, 4 — ФРГ, 5 — СССР, 6 — Канада, 7 — Франция, 8 — Великобритания, 9 — Нидерланды, 10 — ЮАР, 11 — Италия, 12 — Индия, 13 — Мозамбик, 14 — Бирма

Массив не указан. Фиксируем $\alpha \approx 0,77$, соответствующее ципфовому распределению с $B \approx \{28 / [(350\,000/4354)^{0,77} - 1]\} - 1 \approx -0,01$.

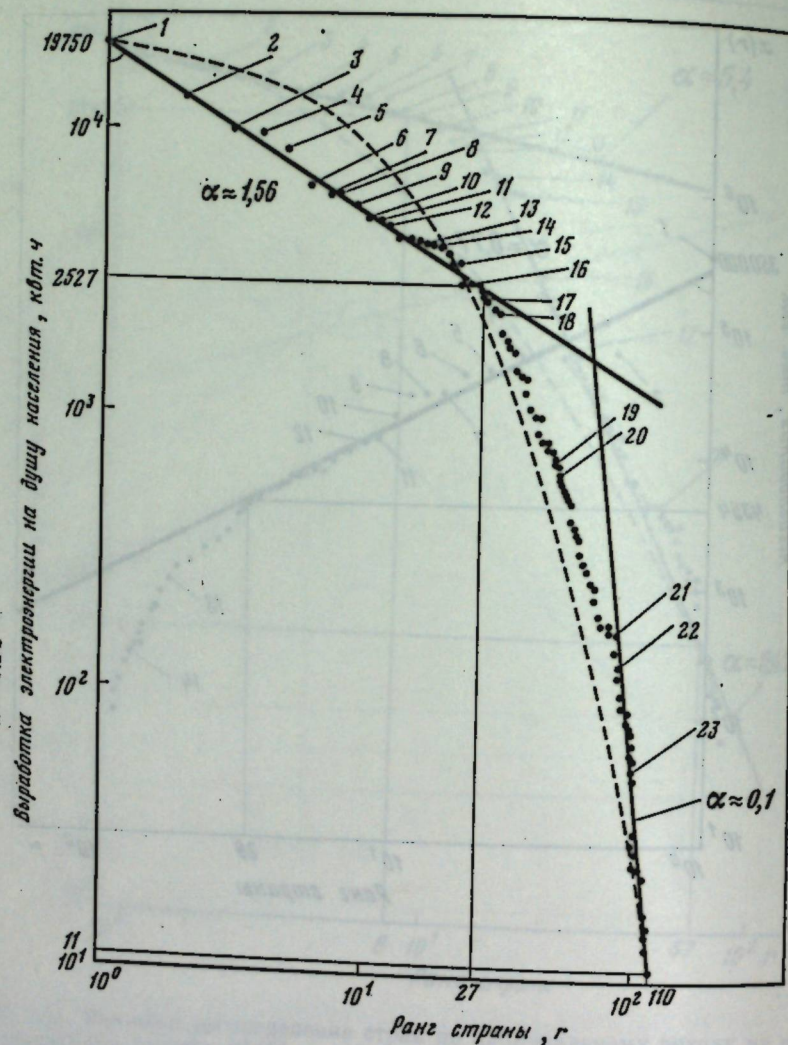


РИС. 2.53. Ранговое распределение стран по выработке электроэнергии на душу населения (по данным [0.45], с. 142—145)
 1 — Норвегия, 2 — Канада, 3 — Исландия, 4 — США, 5 — Швеция, 6 — Новая Зеландия, 7 — Швейцария, 8 — Финляндия, 9 — Австралия, 10 — Великобритания, 11 — ГДР, 12 — ФРГ, 13 — Япония, 14 — СССР, 15 — Франция, 16 — ЮАР, 17 — Италия, 18 — Испания, 19 — Куба, 20 — Бразилия, 21 — Папуа Новая Гвинея, 22 — Индия, 23 — Вьетнам
 Массив не указан. Гипотеза, что это распределение является распределением Ципфа с $\alpha \approx 0,1$ и $B \approx (109 / [(19\ 750 / 11)^{0,1} - 1]) - 1 \approx 96,7$ несостоятельна, так как оно плохо описывает данные (пунктирная линия на графике). Фиксируем $\alpha \approx 1,56$, соответствующее ципфовому распределению с $B \approx (26 / [(19\ 750 / 2527)^{1,56} - 1]) - 1 \approx 0,10$

РИС. 2.54. Ранговое распределение союзных республик СССР по производству цемента в 1975 г. (по данным [0.45], с. 277)
 Массив не указан. Фиксируем $\alpha \approx 0,58$, соответствующее распределению Ципфа с $B \approx (10 / [(73\ 119 / 1260)^{0,58} - 1]) - 1 \approx 0,05$

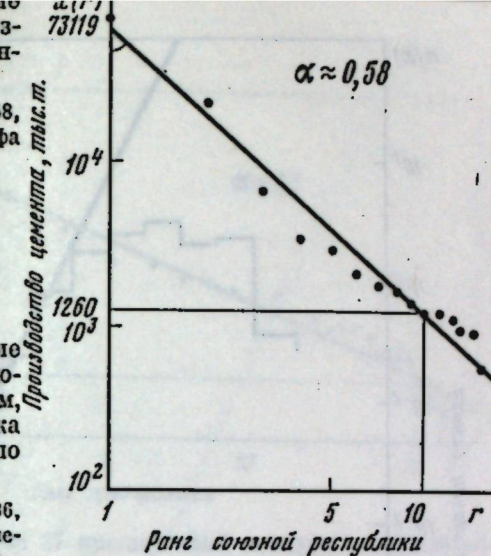
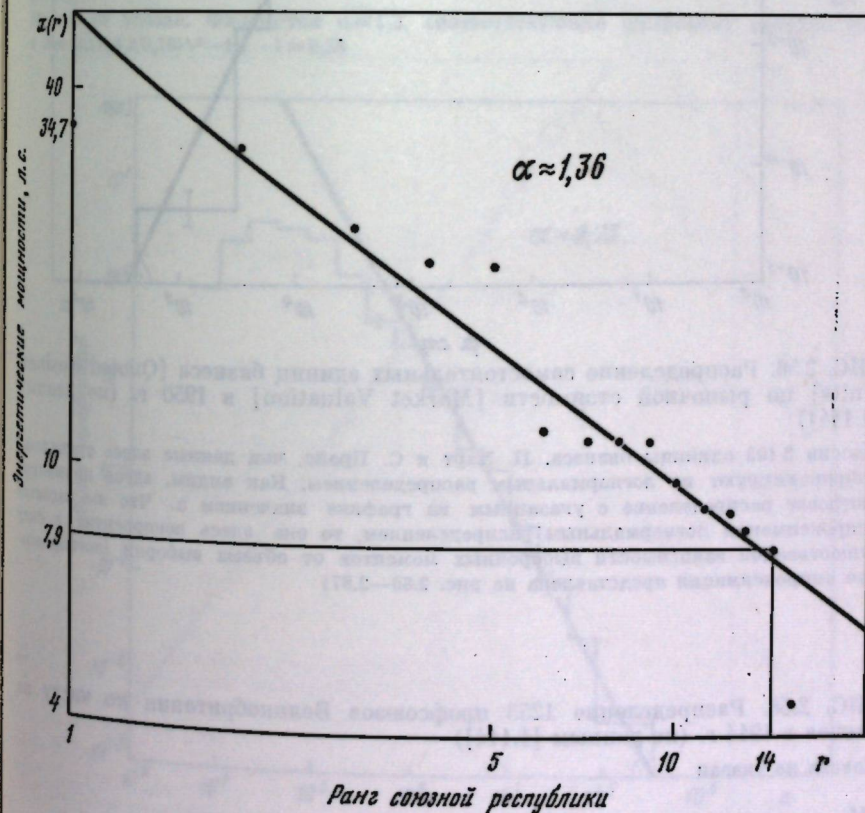


РИС. 2.55. Ранговое распределение союзных республик Советского Союза по энергетическим мощностям, приходящимся на одного работника сельского хозяйства в 1975 г. (по данным [0.45], с. 180)
 Массив не указан. Фиксируем $\alpha \approx 1,36$, соответствующее ципфовому распределению с $B \approx (13 / [(34,7 / 7,9)^{1,36} - 1]) - 1 \approx 1,00$



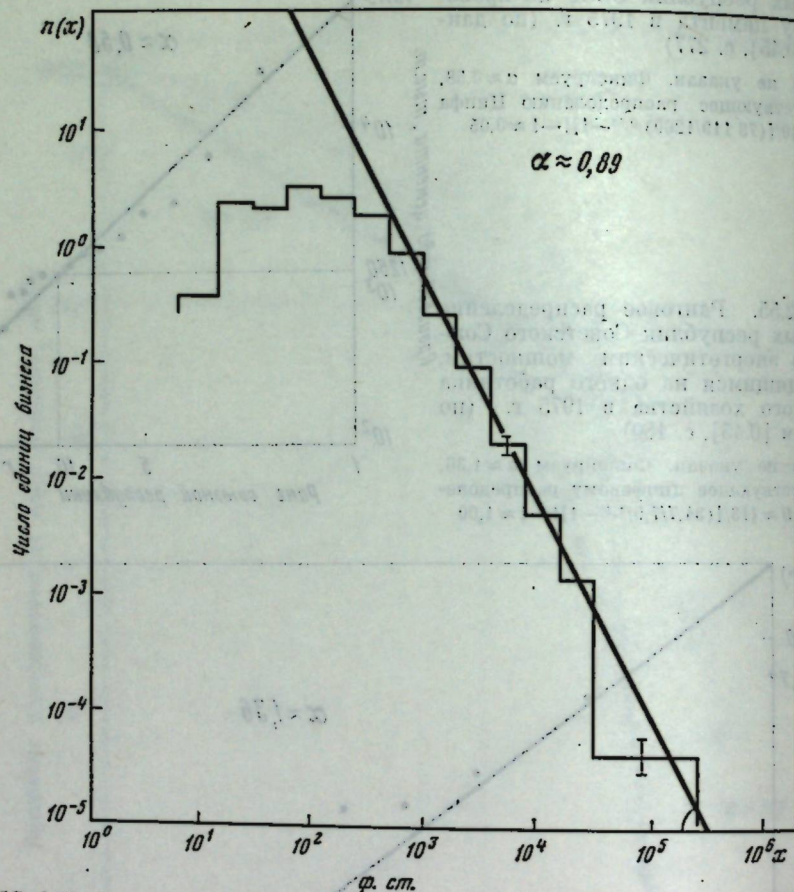


РИС. 2.56. Распределение самостоятельных единиц бизнеса [Quoted Business Units] по рыночной стоимости [Market Valuation] в 1950 г. (по данным [1.114])

Массив 2 103 единицы бизнеса. П. Харт и С. Прайс, чьи данные здесь приведены, аппроксимируют их логнормальным распределением. Как видим, здесь применимо ципфовое распределение с указанным на графике значением α . Что же касается аппроксимации логнормальным распределением, то она здесь некорректна в силу существенной зависимости выборочных моментов от объема выборки (логнормальная аппроксимация представлена на рис. 2.80—2.87)

РИС. 2.58. Распределение 1253 профсоюзных Великобритании по числу их членов в 1914 г. (по данным [1.114])

Массив не указан

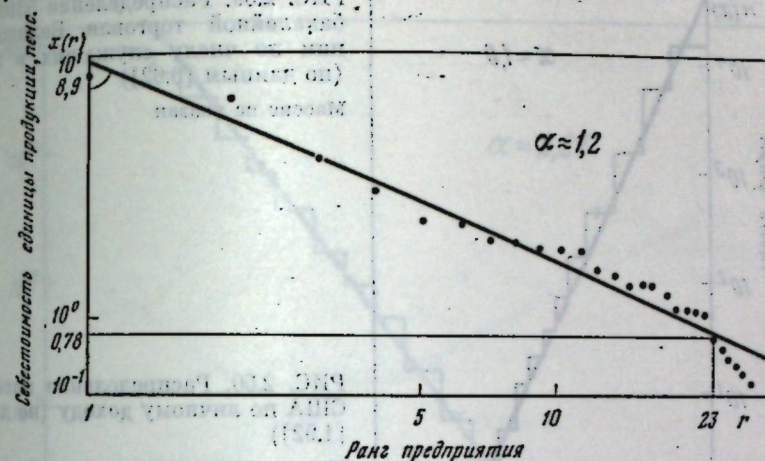
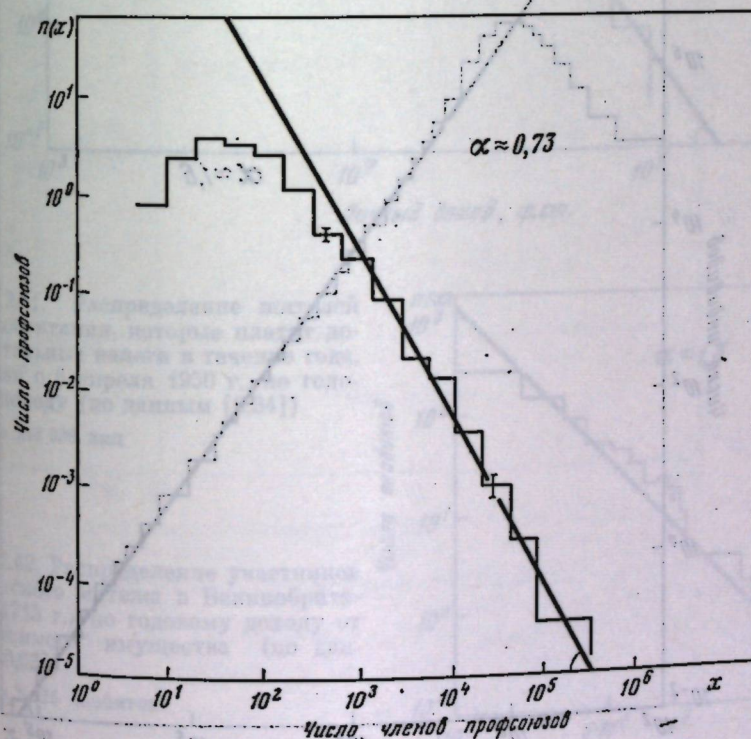


РИС. 2.57. Ранговое распределение 27 предприятий электропромышленности Великобритании по себестоимости продукции в 1933—1934 гг. (по данным [0.62])

Массив не указан. Фиксируем $\alpha \approx 1,2$, соответствующее ципфовому распределению с $B \approx [22/(8,9/0,78)^{1,2} - 1)] - 1 \approx 0,25$



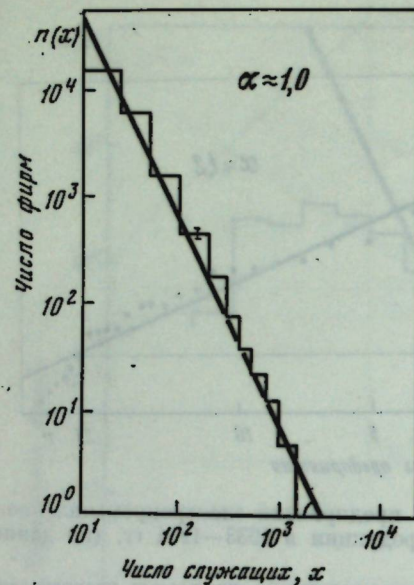


РИС. 2.59. Распределение 5310 фирм бакалейной торговли Великобритании по числу служащих в 1930 г. (по данным [0.62])

Массив не указан

РИС. 2.60. Распределение населения США по личному доходу (по данным [1.92])

Массив 37 569 000 человек. Это распределение часто называют распределением Парето по имени впервые исследовавшего его в 1897 г. итальянского экономиста

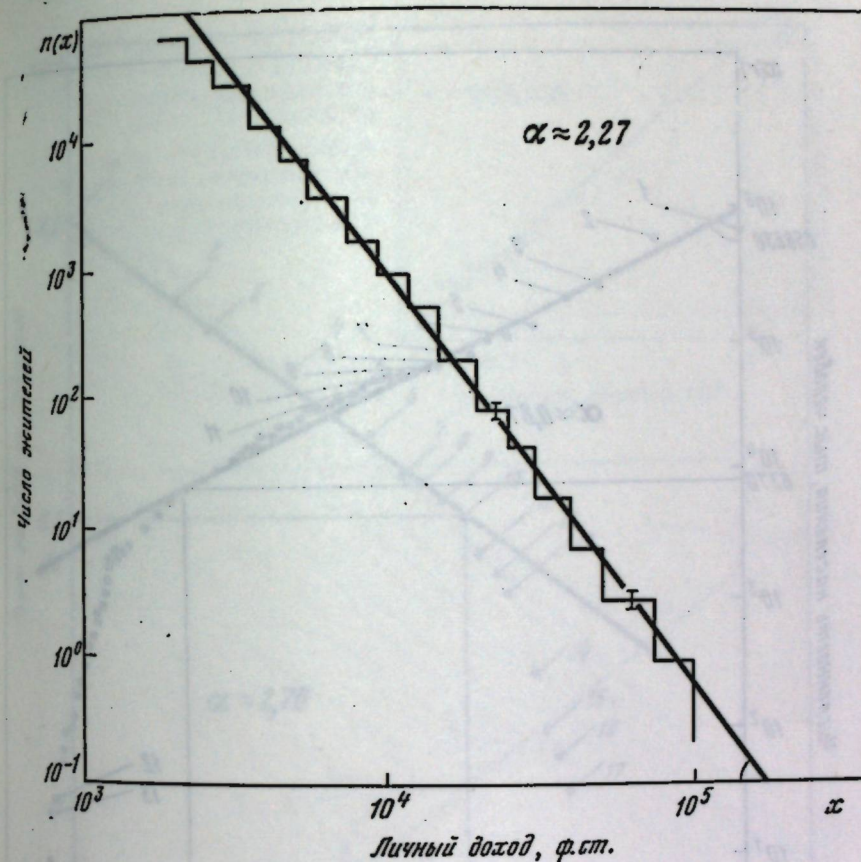
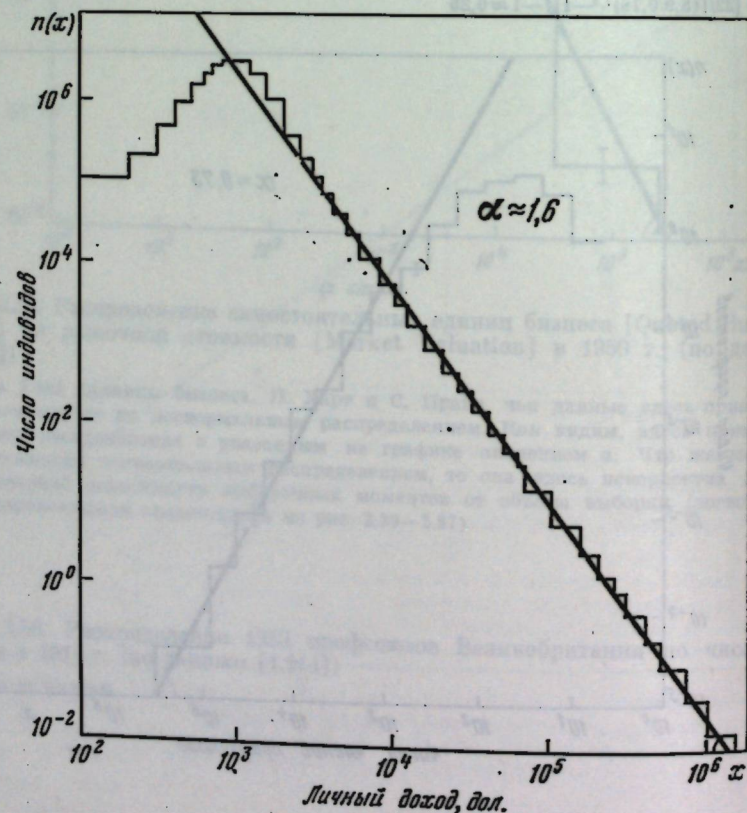
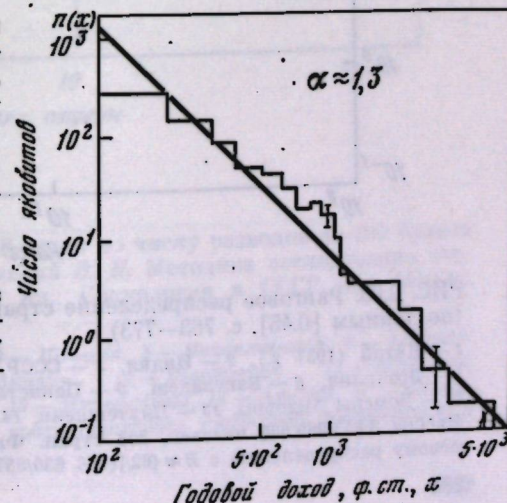


РИС. 2.61. Распределение жителей Великобритании, которые платят дополнительные налоги в течение года, начиная с 6 апреля 1950 г., по годовому доходу (по данным [0.34])

Массив 213 938 лиц

РИС. 2.62. Распределение участников якобитского мятежа в Великобритании (1715 г.) по годовому доходу от недвижимого имущества (по данным [0.62])

Массив 2 476 якобитов



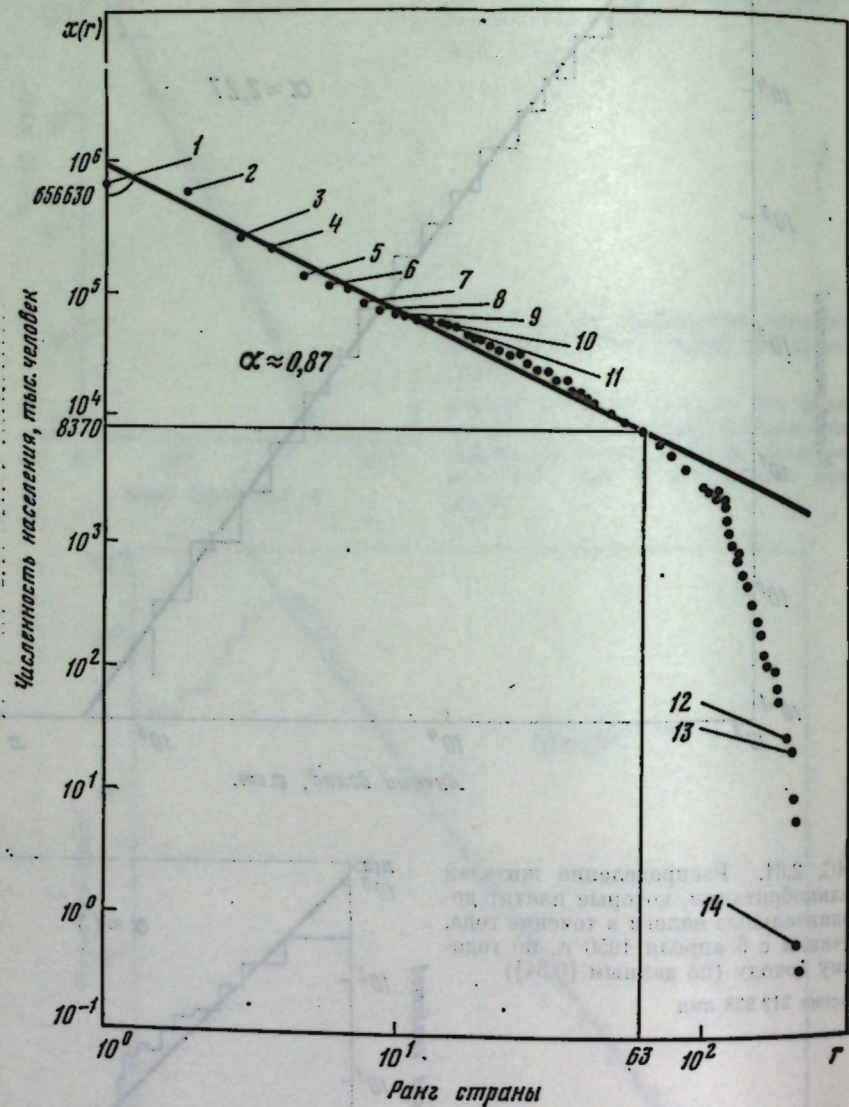


РИС. 2.63. Ранговое распределение стран по численности населения в 1975 г. (по данным [0.45], с. 763–773).

1 — Китай (1957 г.), 2 — Индия, 3 — СССР, 4 — США, 5 — Индонезия, 6 — Япония, 7 — Бразилия, 8 — Бангладеш, 9 — Пакистан, 10 — Нигерия, 11 — Великобритания, 12 — Домены Андорры, 13 — Лихтенштейн, 14 — Ватикан
 Массив 3 843 989 000 человек, 208 стран. Фиксируем $\alpha \approx 0,87$, соответствующее цифровому распределению с $B \approx \{62 / [(656\ 630 / 8370)^{0,87} - 1]\} - 1 \approx 0,42$

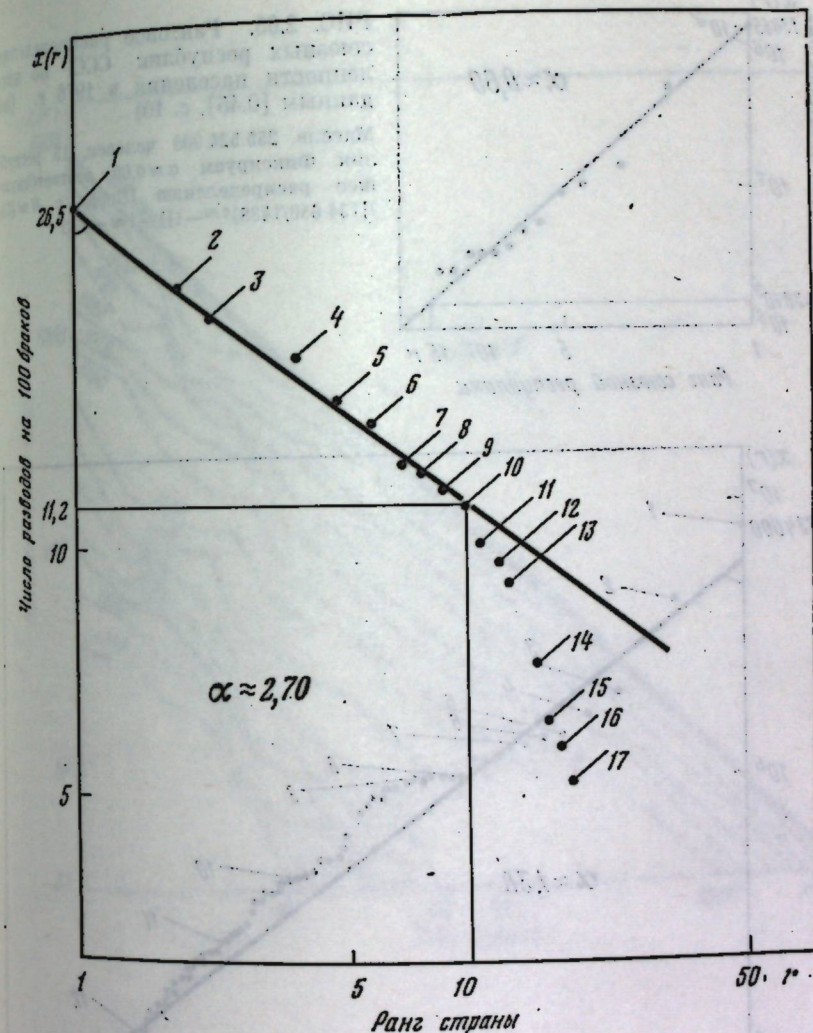


РИС. 2.64. Ранговое распределение 18 стран по числу разводов на 100 браков в 1959–1964 гг. (по данным: Старовский В. П. Методика исследования элементов роста народонаселения. — В кн.: Социология в СССР. М.: Мысль. 1966, т. 1, с. 242–262)

1 — США, 2 — Венгрия, 3 — Румыния, 4 — Швеция, 5 — Чехословакия, 6 — Австрия, 7 — Югославия, 8 — СССР, 9 — Финляндия, 10 — Швейцария, 11 — Норвегия, 12 — Франция, 13 — ФРГ, 14 — Бельгия, 15 — Нидерланды, 16 — Англия, 17 — Канада
 Массив не указан. Фиксируем $\alpha \approx 2,70$, соответствующее цифровому распределению с $B \approx \{9 / [1 / (26,5 / 11,2)^{2,7} - 1]\} - 1 \approx -0,02$

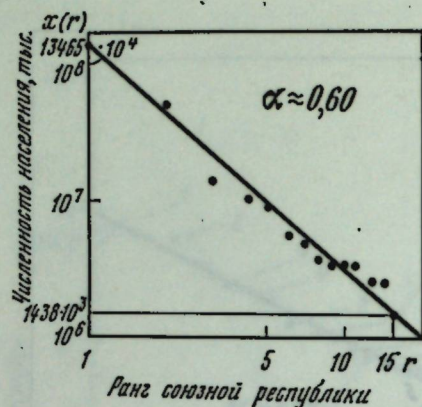


РИС. 2.65. Ранговое распределение союзных республик СССР по численности населения в 1976 г. (по данным [0.45], с. 10)

Массив 255 524 000 человек, 15 республик. Фиксируем $\alpha \approx 0,60$, соответствующее распределению Ципфа с $B \approx (14 / (134\ 650 / 1438)^{0,60} - 1)) - 1 \approx -0,42$

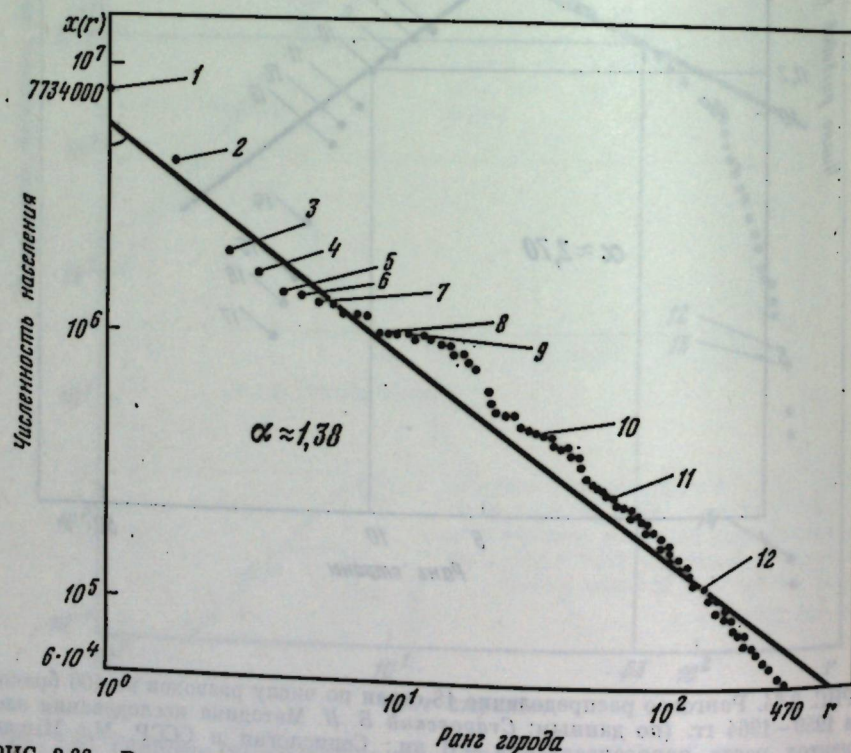


РИС. 2.66. Ранговое распределение городов СССР с населением 50 000 человек или более по численности населения в 1976 г. (по данным [0.45], с. 22—31)

1 — Москва, 2 — Ленинград, 3 — Киев, 4 — Ташкент, 5 — Баку, 6 — Харьков, 7 — Горький, 8 — Тбилиси, 9 — Омск, 10 — Душанбе, 11 — Сочи, 12 — Абакан
Массив 107 113 000 человек, 470 городов. Фиксируем $\alpha \approx 1,38$, соответствующее распределению Ципфа с $B \approx (469 / (7\ 734\ 000 / 50\ 000)^{1,38} - 1)) - 1 \approx -0,55$

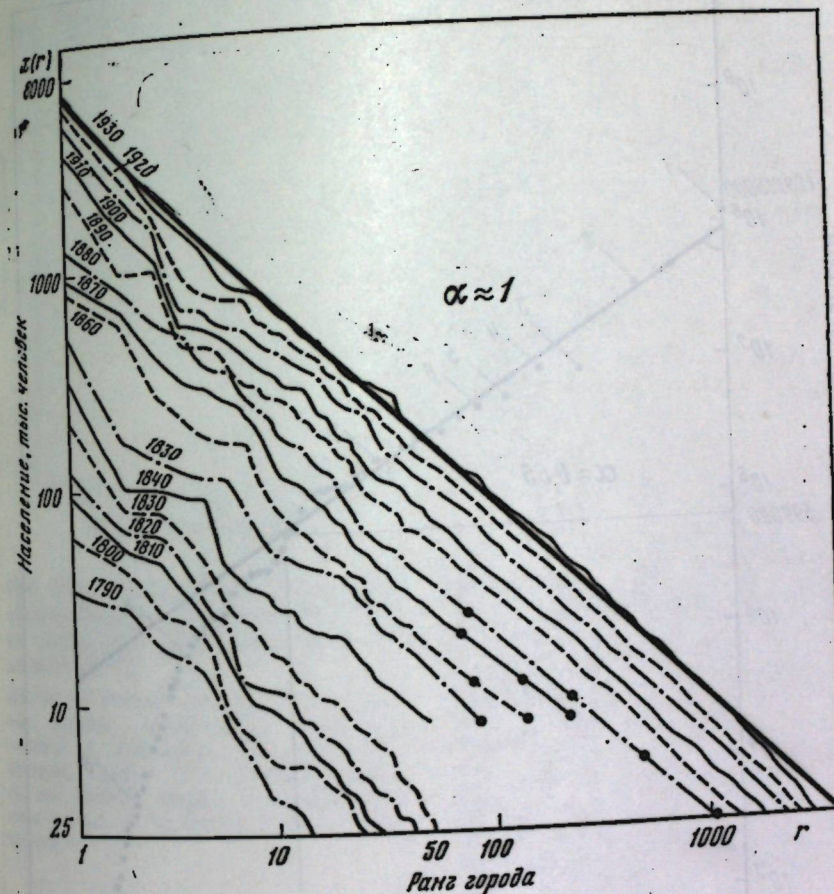


РИС. 2.67. Ранговое распределение городов США по численности населения в 1790—1930 гг. ([1.46], Г. Ципфа)

Массив не указан. Фиксируем $\alpha \approx 1$, соответствующее распределению Ципфа

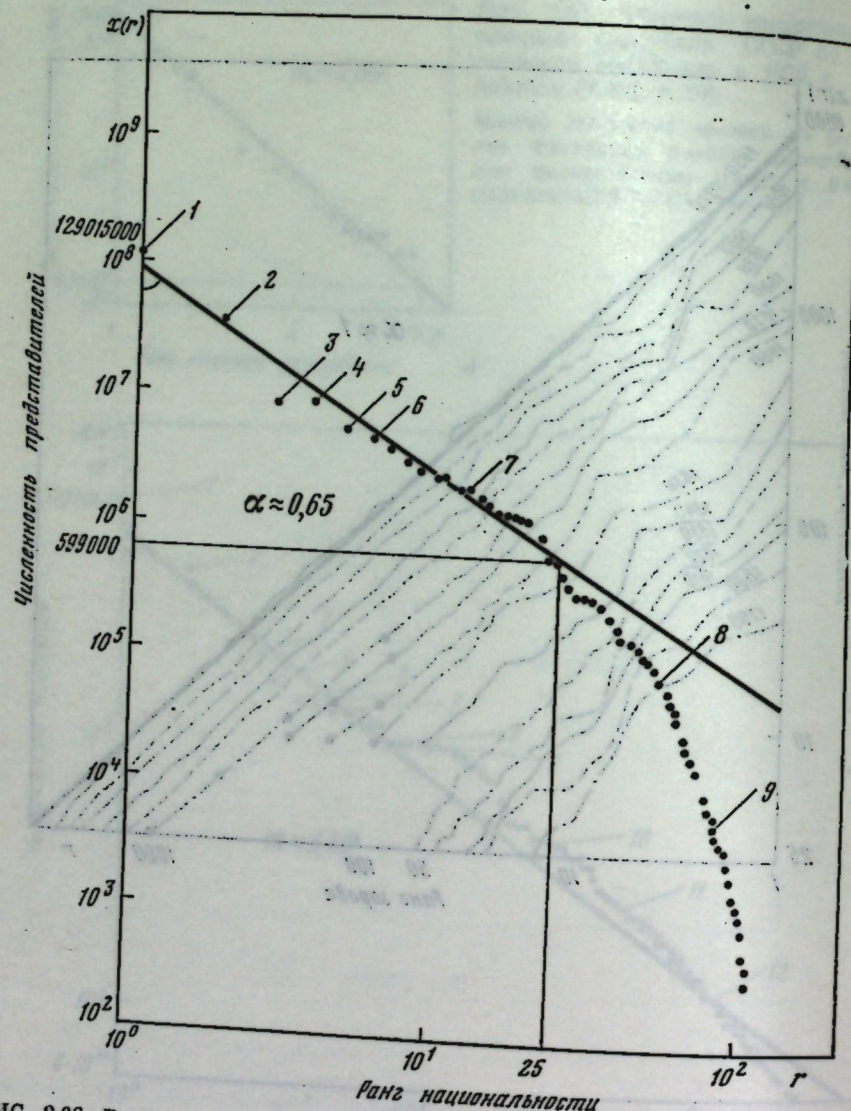


Рис. 2.68. Ранговое распределение национальностей СССР по численности в 1970 г. (по данным [0.45], с. 32—35)
1 — русские, 2 — украинцы, 3 — узбеки, 4 — белорусы, 5 — татары, 6 — казахи, 7 — таджики, 8 — абхазы, 9 — удинцы
Массив 241 568 000 человек, 104 национальности. Фиксируем $\alpha \approx 0,65$, соответствующее ципфовому распределению с коэффициентом рангового искажения $B \approx \approx (24 / (129\ 015\ 000 / 599\ 000))^{\alpha} - 1) - 1 \approx -0,25$

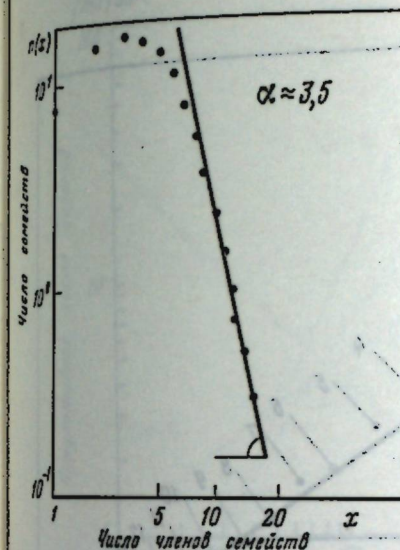


Рис. 2.69. Распределение семейств области Эмилия (Италия) по числу их членов на 1 декабря 1926 г. (по данным [0.21])

Массив не указан. На это распределение, помимо социальных, оказывают влияние и природные (биологические) факторы, однако оно аппроксимируется, как видим, ципфовым распределением и даже с не очень большим значением α

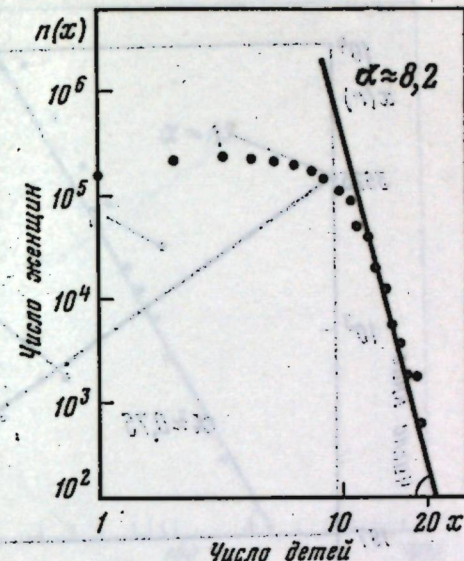
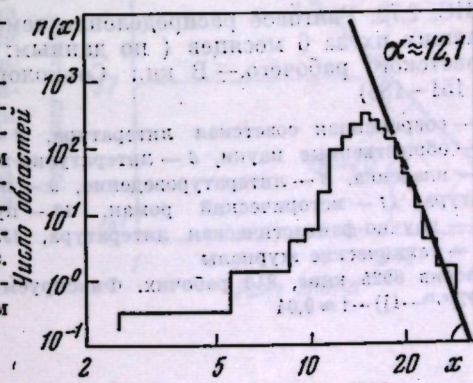


Рис. 2.70. Распределение итальянских женщин, разошедшихся или разведшихся с мужьями (в католической Италии до последнего времени развод был чрезвычайно затруднен), по числу детей в 1931 г. (по данным [0.21])

Массив 1 718 000 женщин. На это распределение, помимо социальных, оказывают действие и природные (биологические) факторы. Поэтому, здесь величина α довольно значительна. Тем не менее относительно большая часть женщин, оказывается, имеет много детей: 500 женщин имеют по 19 детей, 1500 — по 18 и т. д.

Рис. 2.71. Распределение 1475 областей Англии и Уэльса по числу новорожденных (по данным [0.34])

Массив не указан. На это распределение, помимо социальных, влияют природные факторы, поэтому оно аппроксимируется ципфовым распределением с большим α , т. е. является гауссовым. Переход от негауссовых социальных распределений к гауссовым природным совершается непрерывно (см. рис. 2.76—2.78). И эта непрерывность отражается соответствующим непрерывным изменением α



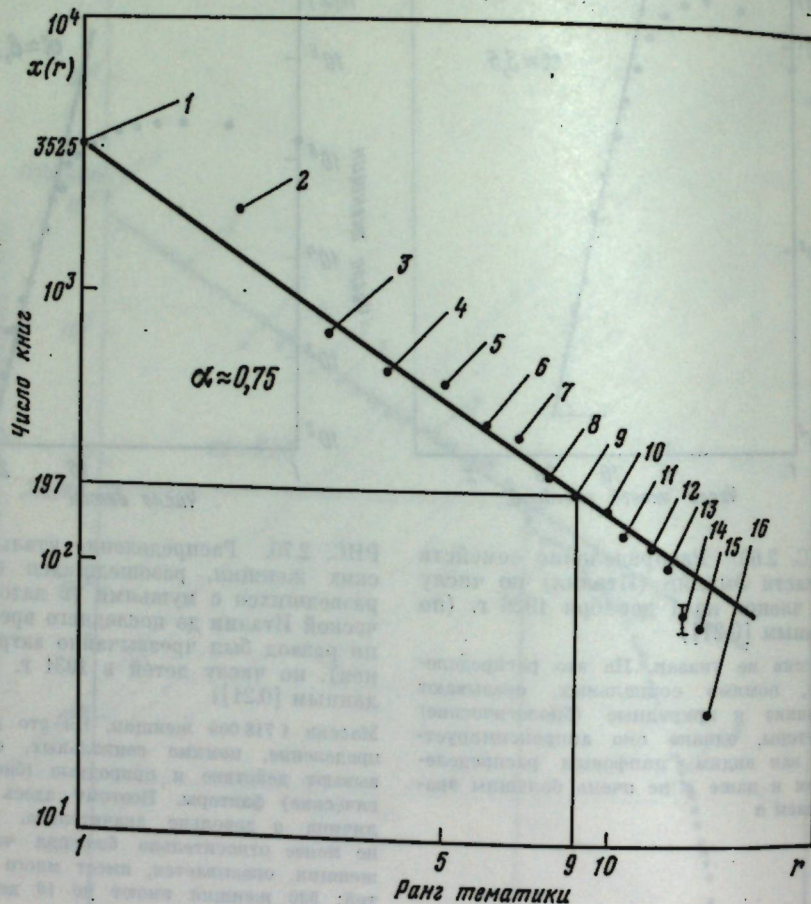


РИС. 2.72. Ранговое распределение тематик книг по числу рабочих, прочитавших их за 6 месяцев (по данным: Гурьянов С. Т. Духовные интересы советского рабочего. — В кн.: Социология в СССР. М.: Мысль, 1966, т. 2, с. 151—186)

1 — современная советская литература, 2 — зарубежная литература, 3 — техника, 4 — общественные науки, 5 — литературно-художественные журналы, 6 — история, 7 — классика, 8 — литературоведение, 9 — искусство, 10 — приключенческая литература, 11 — исторический роман, 12 — путешествия, 13 — военная литература, 14 — научно-фантастическая литература, 15 — общественно-политические журналы, 16 — технические журналы
Массив 8925 книг, 840 рабочих. Фиксируем $\alpha \approx 0,75$, соответствующее $B \approx (8/(3525/197)^{1/(1-\alpha)} - 1) - 1 \approx 0,04$

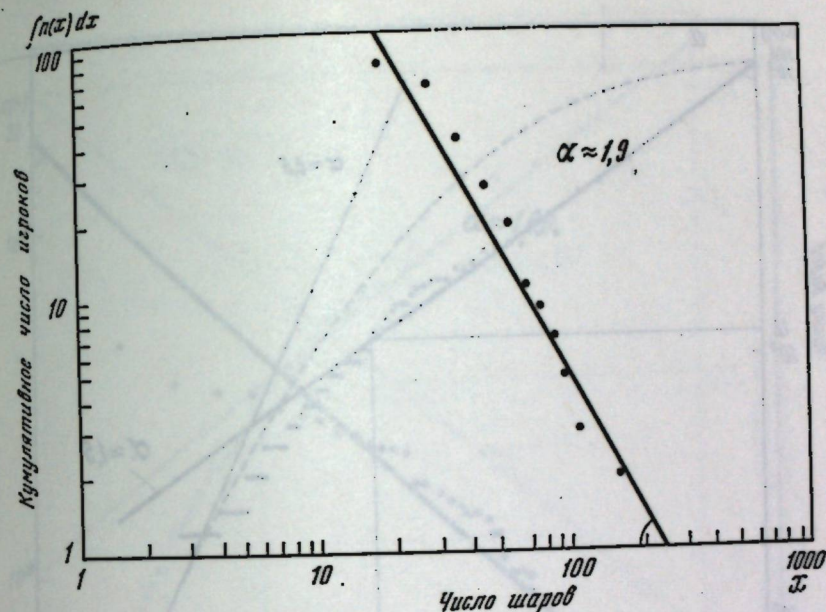
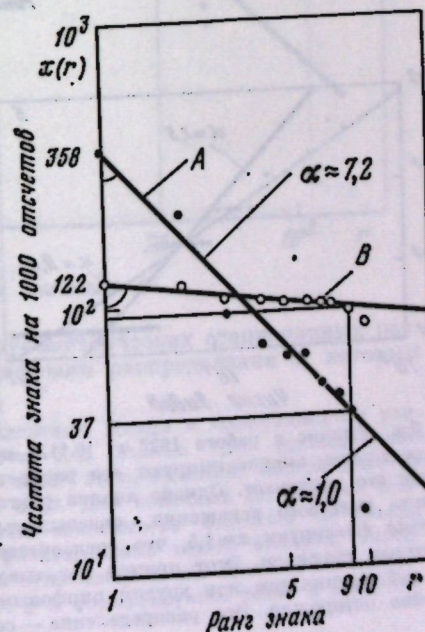


РИС. 2.73. Кумулятивное распределение игроков по числу забитых ими в лузу шаров при игре в бильярд за 50 подач [1.92]

Массив 3540 шаров

РИС. 2.74. Ранговое распределение последних знаков в отсчетах испытуемых А и В по частоте появления этих знаков (индивиду, снимая показания приборов особенно последние знаки этих показаний, находящиеся на грани точности, вносят определенный произвол, отражаемый данным распределением) (по данным [0.62])

Массив А — 1000 и В — 3000 знаков. Данные показывают большую зависимость α от личности испытуемого



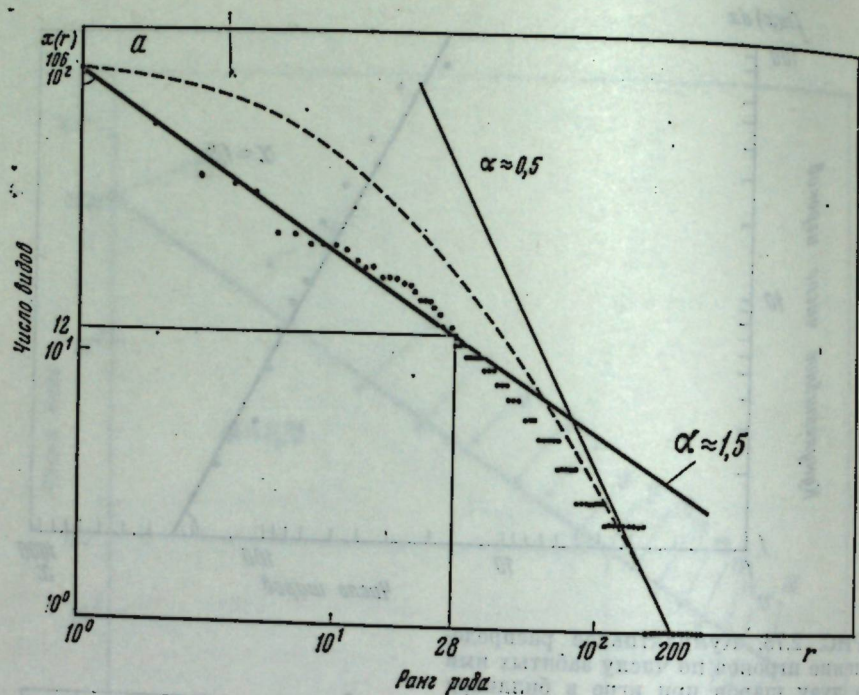


РИС. 2.75. Таксономическое распределение биологических родов по числу видов для ящериц (по данным Дж. Уиллиса, приведенным в [0.86])

(а) Ранговая форма. Гипотеза, что данное распределение является распределением Ципфа с $\alpha \approx 0,5$ и коэффициентом рангового искажения $B \approx (199/[106/1]^{0,5} - 1) - 1 \approx 20,4$, несостоятельна, так как плохо описывает данные (см. пунктирную кривую на графике). Фиксируем $\alpha \approx 1,5$ соответствующее ципфовому распределению с $B \approx (27/[106/12]^{1,5} - 1) - 1 \approx 0,07$.

(б) Частотная форма. Дж. Уиллис в работе 1922 г. [0.113] и другие авторы, исследовавшие данное распределение, аппроксимируют его распределением Ципфа с $\alpha \approx 0,5$. И частотный эффект рангового искажения, показывая некорректность этой аппроксимации. Поэтому фиксируем $\alpha \approx 1,5$, что согласуется с ранговым графиком и допускается частотным графиком. Этот пример показывает, что вопрос об аппроксимации данных наблюдения тем или другим ципфовым распределением следует решать чрезвычайно осторожно. Это распределение — социальное (см. раз. 1 § 2, гл. 7)

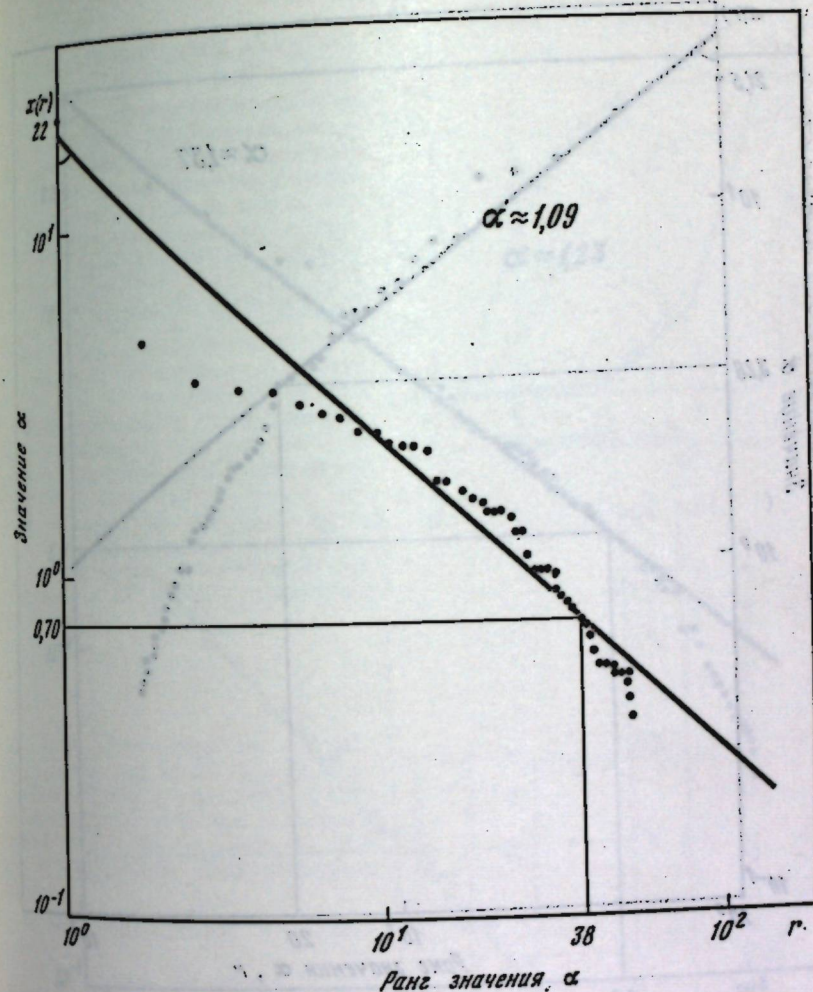


РИС. 2.76. Ранговое распределение 53 наукометрических стационарных распределений по величине показателя ципфового распределения α , которым они аппроксимируются

Фиксируем $\alpha \approx 1,09$, соответствующее распределению Ципфа с коэффициентом рангового искажения $B \approx (37/[(22/0,7)^{1,09} - 1]) - 1 \approx -0,12$. Только 1,9% распределений имеют $\alpha > 10$, для 49,1% $\alpha \leq 1$ и для 73,6% $\alpha \leq 2$. Распределения, построенные на бесконечных совокупностях, негауссовы при $\alpha \leq 2$. Для конечных же выборок критерием негауссовости распределений является зависимость моментов от объема выборки. Для подавляющего большинства распределений нашей выборки, значительно большего, чем 73,6%, эти зависимости существенны, что означает их негауссовость. На этой эмпирической базе в гл. 7, § 2 формулируется тезис о негауссовости научной деятельности

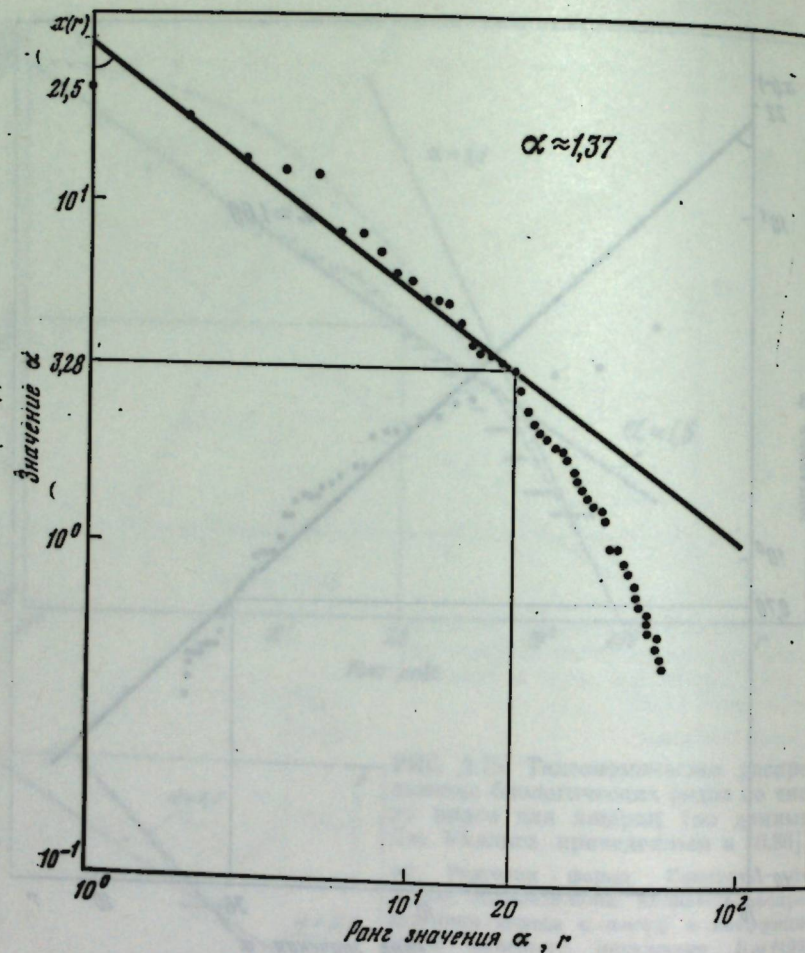


РИС. 2.77. Ранговое распределение 57 ненаукометрических социальных стационарных распределений по величине α

Фиксируем $\alpha \approx 1,37$, что соответствует дигфовому распределению с $B \approx (19/(21,5/3,28)^{1,37} - 1)) - 1 \approx 0,56$. 8,8% распределений имеют $\alpha > 10$, для 33,3% $\alpha \leq 1$ и для 54,4% $\alpha \leq 2$. Эти данные говорят о негауссовости социальных явлений. Сравнение данного графика с графиком, показанным на рис. 2.76, говорит, что для научной деятельности эта негауссовость выражена сильнее, чем для ряда других видов человеческой деятельности. Это объясняется, по-видимому, творческой природой научной деятельности, направленной на производство нового знания

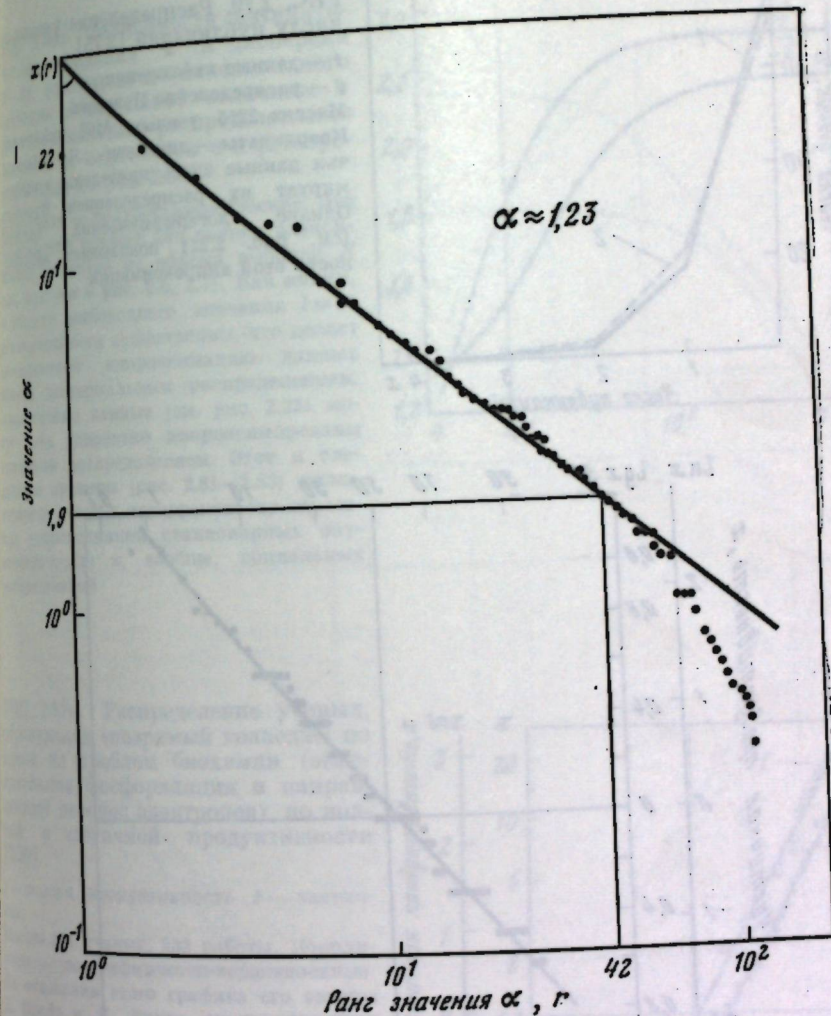


РИС. 2.78. Ранговое распределение 110 социальных стационарных распределений (включая 53 наукометрических, отраженных на рис. 2.76) по величине α

Фиксируем $\alpha \approx 1,23$, что соответствует дигфовому распределению с $B \approx (41/(22/1,9)^{1,23} - 1)) - 1 \approx 1,12$. 5,4% распределений имеют $\alpha > 10$, для 40,9% $\alpha \leq 1$ и для 63,6% $\alpha \leq 2$. Для подавляющего большинства распределений нашей выборки, превышающее 63,6%, имеет место существенная зависимость моментов от объема выборки. На этой эмпирической базе в гл. 7, § 2 формулируется тезис о негауссовости социальных явлений. Параметры распределения, показанного на данном графике, по понятным причинам имеют значения, промежуточные между значениями параметров распределений, показанных на рис. 2.76 и 2.77

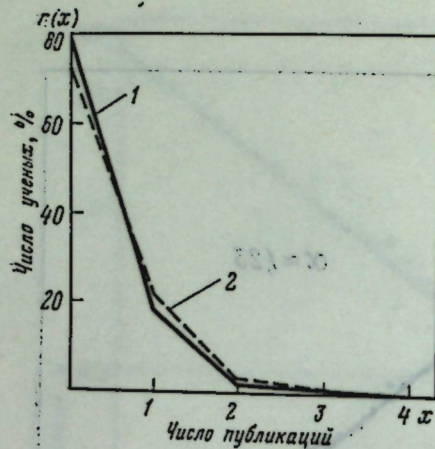


РИС. 2.79. Распределение ученых по числу публикаций [3.75]

1 — данные наблюдения, 2 — распределение Пуассона. Массив 2255 ученых, 3162 публикаций. Координаты — линейные. Л. Мантелл, чьи данные здесь приведены, аппроксимирует их распределением Пуассона. Однако логарифмический масштаб (см. рис. 2.22) показывает непригодность этой аппроксимации

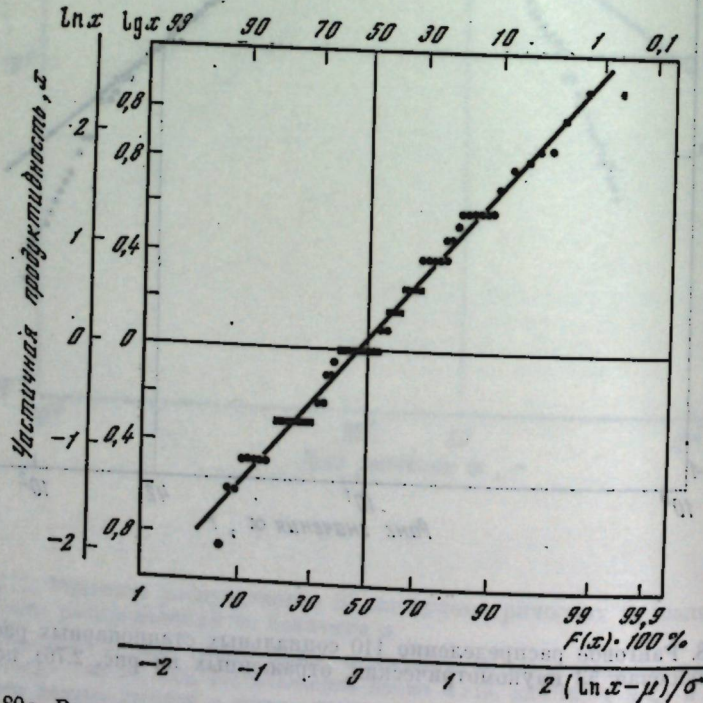


РИС. 2.80а. Распределение ученых Брукгейвенской лаборатории (США) по частичной продуктивности (за публикацию с n авторами каждому начисляется балл $1/n$) [1.156]

Массив 180 ученых. Координаты — логарифмически-вероятностные. В этих координатах прямая линия соответствует логнормальному распределению. На основании этого графика его автор, лауреат Нобелевской премии по физике У. Шокли, аппроксимирует свои данные логнормальным распределением. График позволяет определить параметры этого логнормального распределения: $\sigma=1,14$ и $\mu=0,18$

РИС. 2.80б. Зависимости математического ожидания $\bar{x}$ и дисперсии $\bar{x}^2$ от выборочного максимального значения I случайной величины x для логнормального распределения с параметрами σ и μ , определенными по рис. 2.80а

Координаты — полулогарифмические. Эти зависимости служат для оценки зависимости моментов от объема выборки N (см. подписи к рис. 2.6, 2.7). Как видим, в области выборочного значения $I \approx 10$ эти зависимости существенны, что делает неадекватной аппроксимацию данных Пуассона логнормальным распределением. Аналогичные данные (см. рис. 2.23) могут быть корректно аппроксимированы ступенчатым распределением. Этот и следующие примеры (рис. 2.81—2.83) иллюстрируют тезис о некорректности гауссовых аппроксимаций стационарных наукометрических и, вообще, социальных распределений

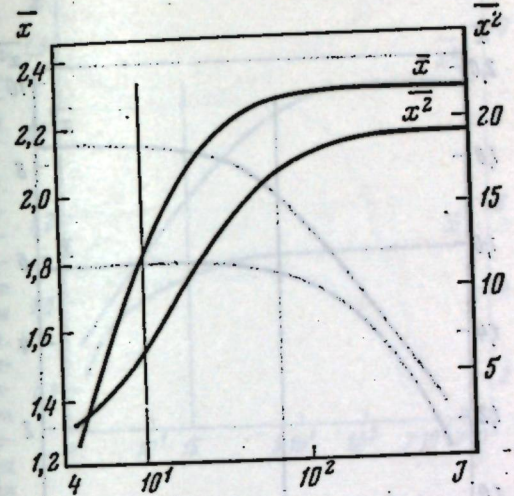
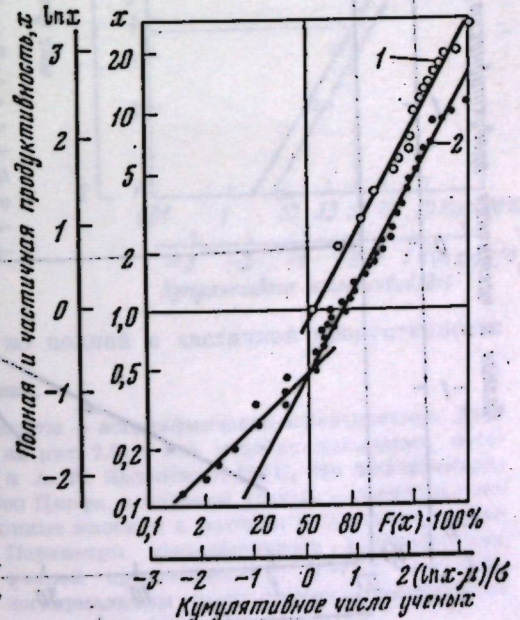


РИС. 2.81а. Распределение ученых, образующих «незримый колледж» по одной из проблем биохимии (окислительная фосфорилизация и направленный перенос электронов), по полной и частичной продуктивности [2.30]

1 — полная продуктивность 2 — частичная

Массив 555 ученых, 533 работы. Координаты — логарифмически-вероятностные. На основании этого графика его авторы Д. Прайс и Д. Бивер аппроксимируют свои данные логнормальными распределениями. Данные по полной продуктивности, действительно, удовлетворительно описываются логнормальной прямой. Параметры этого логнормального распределения, определенные по графику: $\sigma=1,11$, $\mu=-0,16$. Логнормальная аппроксимация полной продуктивности, однако, некорректна (см. рис. 2.81б). Данные же по частичной продуктивности логнормальной прямой описываются, как видим, плохо



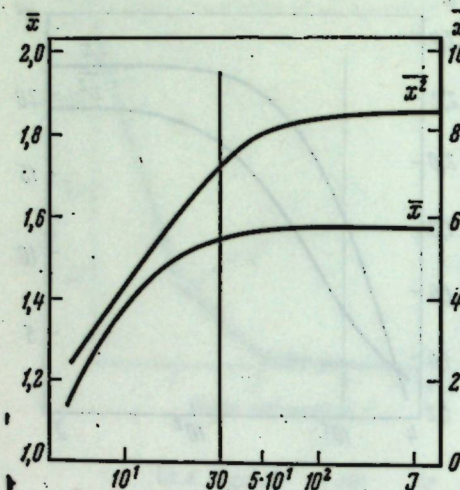


РИС. 2.816. Зависимости математического ожидания $\bar{x}$ и дисперсии $\bar{x}^2$ от I для логнормального распределения со значениями параметров σ и μ , определенными по графику Д. Прайса и Д. Бивера полной продуктивности ученых (см. рис. 2.81а)

Координаты — полулогарифмические. Как видим, в области выборочного значения ($I \approx 30$) эти зависимости существенны, что делает некорректной логнормальную аппроксимацию в данном случае. Аналогичные им данные (см. рис. 2.23) вполне корректно аппроксимируются цифровым распределением

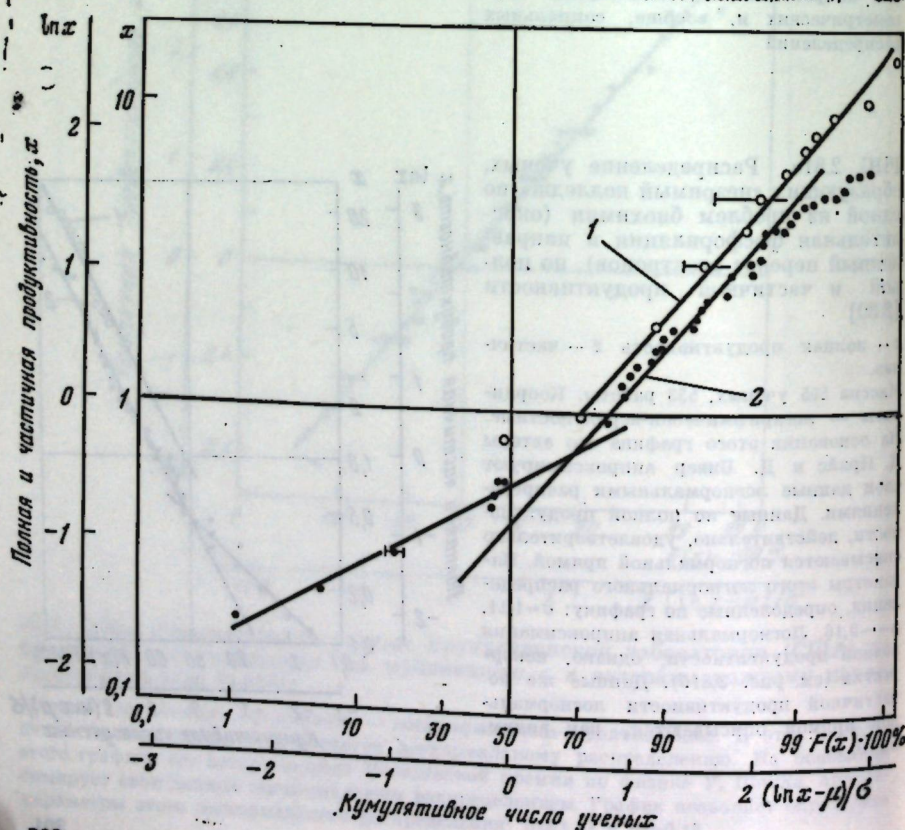


РИС. 2.826. Зависимости математического ожидания $\bar{x}$ и дисперсии $\bar{x}^2$ от I для логнормального распределения со значениями параметров, определенными по графику, показанному на рис. 2.82а, для полной продуктивности

Координаты — полулогарифмические. Как видим, в области выборочного значения $I \approx 15$ эти зависимости существенны, что делает некорректной логнормальную аппроксимацию. Эти данные вполне корректно аппроксимируются цифровым распределением с $\alpha \approx 2,3$ (см. рис. 2.21). Сопоставление этой цифровой аппроксимации с графиком, показывающим зависимость моментов распределения Цифа от объема выборки при разных значениях α (рис. 2.6), подтверждает вывод о существенности этой зависимости для данного выборочного распределения

РИС. 2.83а. Распределение населения США в 1944 и 1950 гг. по личному доходу [0.68]

1 — 1950, 2 — 1944.

Массив не указан. Координаты — логарифмически-вероятностные. Данные, как видим, удовлетворительно описываются логнормальным распределением. Параметры одного из них (1950 г.): $\sigma = 0.66$; $\mu = -1.94$. Логнормальные аппроксимации здесь, однако, некорректны. Во-первых, на данном графике не показан «хвост» распределения, соответствующий большим личным доходам и как раз отклоняющийся от логнормальной прямой (см. рис. 2.86). Во-вторых, в той области значений x , где имеет место логнормальная прямая, эта аппроксимация все равно некорректна (см. рис. 2.83б)

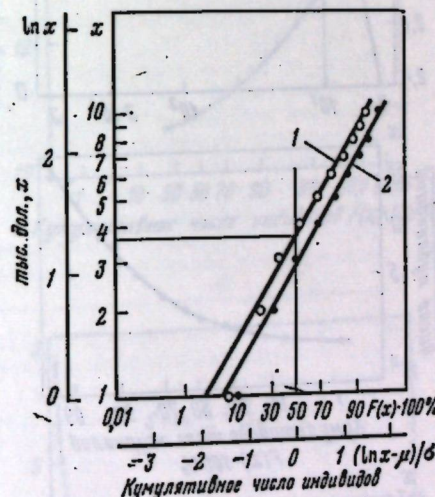


РИС. 2.82а. Распределение ученых по полной и частичной продуктивности на информационном массиве
1 — полная продуктивность, 2 — частичная.

Массив тот же, что на рис. 2.21. Координаты — логарифмически-вероятностные. Данные, подобные приведенным, показаны на рис. 2.81а. Это подобие доказывает несостоятельность тезиса А. А. Игнатова и А. П. Яблонского [0.31], что информационные массивы подчиняются распределению Цифа, а массивы ученых — логнормальному распределению, т. е. что информационные массивы и массивы публикаций описываются разными аппроксимациями. Параметры логнормального распределения, описывающего здесь распределение полной продуктивности: $\sigma = 1,16$; $\mu = -0,58$. Аппроксимация этого распределения логнормальным распределением некорректна (см. рис. 2.82б)

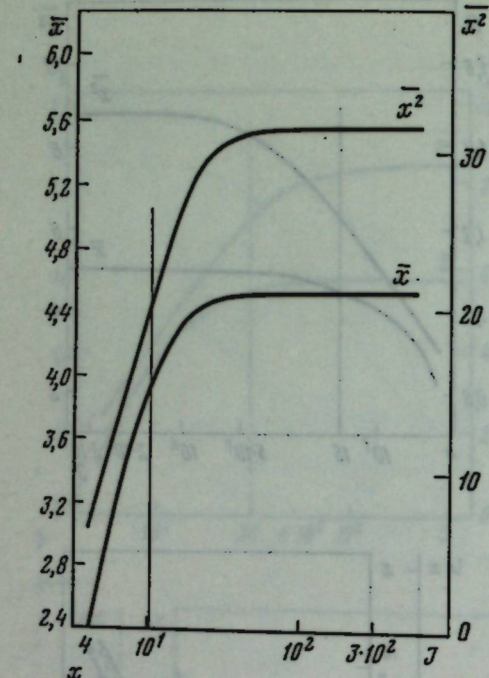


РИС. 2.836. Зависимости математического ожидания $\bar{x}$ и дисперсии x^2 от I для логнормального распределения со значениями параметров σ и μ , определенными графически по рис. 2.83а

Координаты — полулогарифмические. Как видим, в области выборочного значения ($I \approx 10$) эти зависимости существуют, что делает аппроксимацию логнормальным распределением некорректной

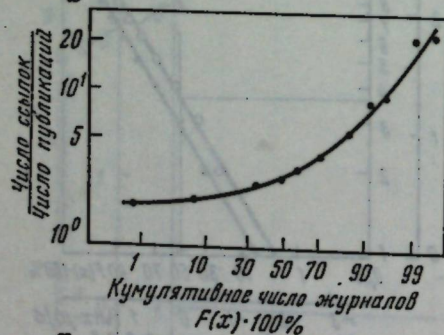


РИС. 2.84. Распределение журналов по числу ссылок на них, деленному на число содержащихся в этих журналах публикаций

Массив тот же, что и на рис. 2.33. Координаты — логарифмически-вероятностные. Отклонение точек от прямой доказывает невозможность их аппроксимации логнормальным распределением. Эта ситуация характерна для всех наукометрических и, вообще, социальных стационарных распределений, построенных на больших выборках (см. рис. 2.85—2.87). На небольших же выборках аппроксимации логнормальным распределением, как и другими гауссовыми распределениями, некорректны из-за зависимости моментов от объема выборки (см. рис. 2.80—2.83)

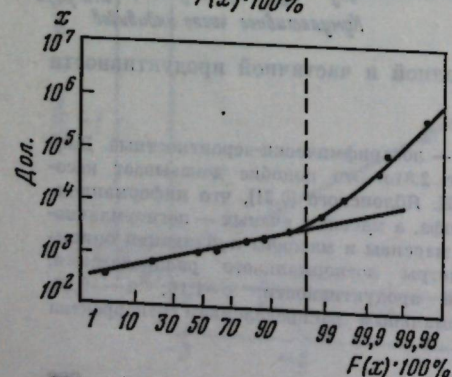


РИС. 2.85. Распределение ученых по их цитированию

Массив тот же, что на рис. 2.35. Координаты — логарифмически-вероятностные. Отклонение точек от прямой доказывает невозможность их аппроксимации логнормальным распределением

РИС. 2.86. Распределение населения ЛА по личному доходу

Массив тот же, что на рис. 2.60. Отклонение точек от прямой доказывает невозможность их аппроксимации логнормальным распределением

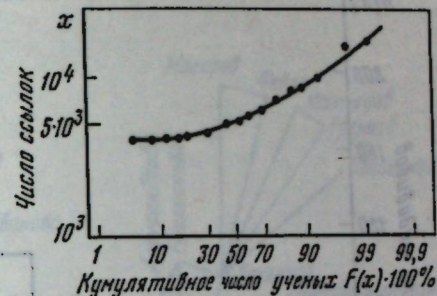


РИС. 2.87. Распределение населения Великобритании по годовому доходу

Массив тот же, что на рис. 2.61. Отклонение точек от прямой доказывает невозможность их аппроксимации логнормальным распределением

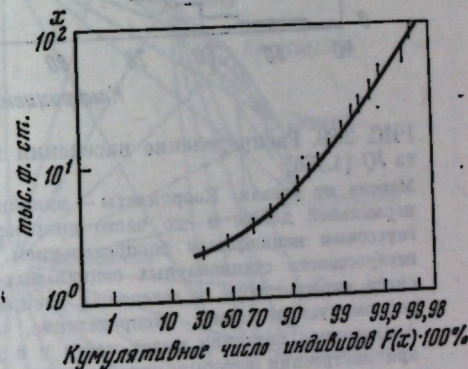
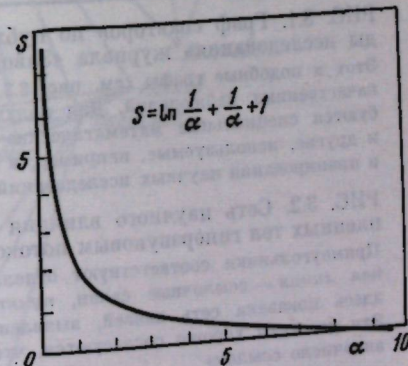


РИС. 2.88. Зависимость энтропии S распределения Ципфа ($x_0=1$) от α

Как видим, энтропия и α связаны монотонной зависимостью, причем с ростом энтропии уменьшается. Энтропия и α — характеристики статистического распределения, быстро устремляющиеся для негуссовых распределений (не говоря уже о гауссовых) к конечным пределам с $\alpha \rightarrow \infty$



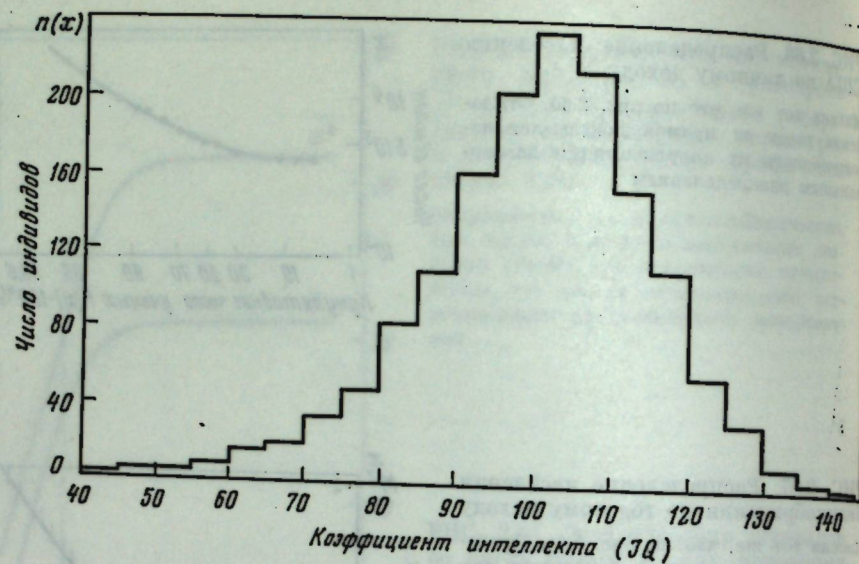


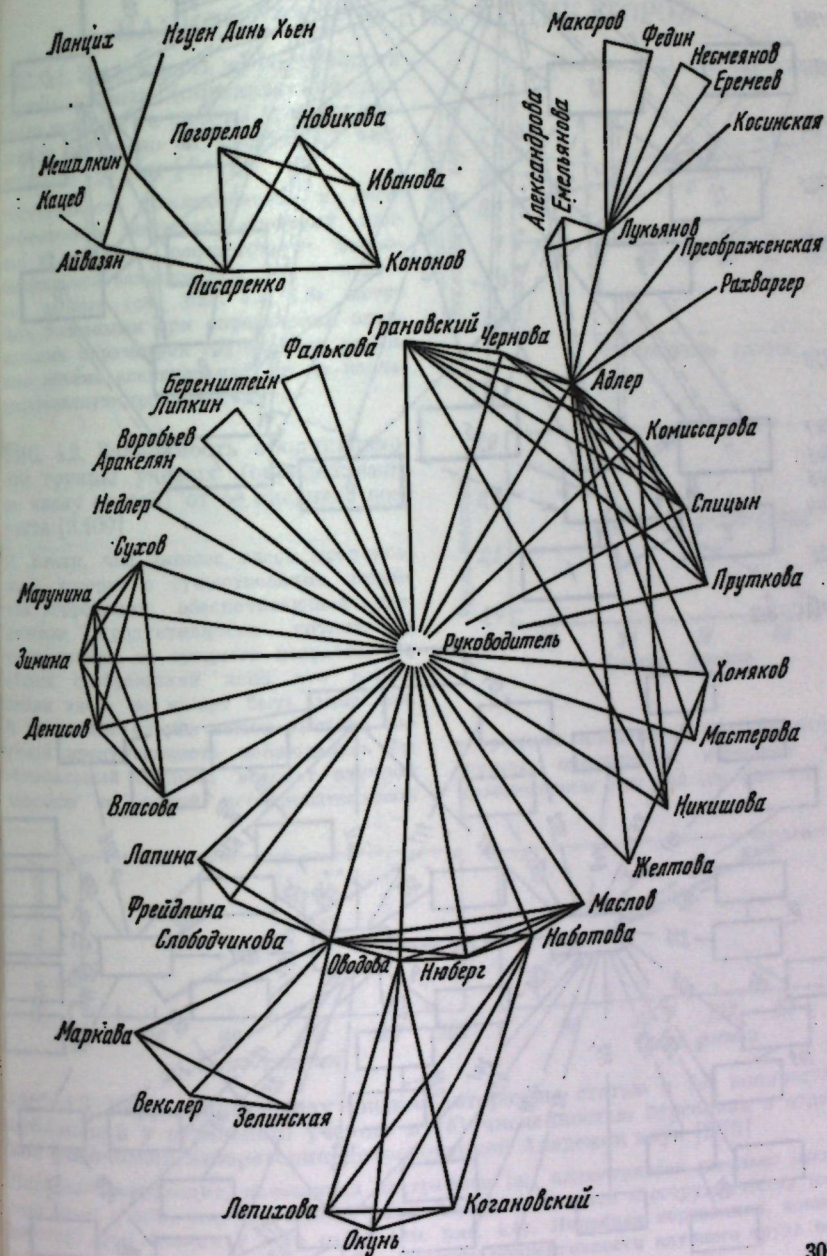
РИС. 2.89. Распределение населения 16—75 лет по коэффициенту интеллекта JQ [1.126]

Массив не указан. Координаты — линейные. Данное распределение имеет «хвост» нормальной длины и его часто аппроксимируют распределением Пуассона, т. е. гауссовым нецифровым распределением. Это, однако, не противоречит тезису о негауссовости стационарных социальных распределений, так как последнее относится, строго говоря, к латентным распределениям. В данном же случае индикатор деформирует латентное распределение, «обрезая» большие значения переменной. Аналогичная ситуация имеет место и с другими стационарными распределениями, при построении которых используются баллы, ограниченные сверху, т. е., применяются так называемые закрытые шкалы

РИС. 3.1. Граф соавторов по публикациям раздела «Математические методы исследования» журнала «Заводская лаборатория» (1962—1966 гг.) [1.1] Этот и подобные графы (см. рис. 3.2, 3.3) сами по себе дают материал лишь для качественных наблюдений. Для получения количественных результатов здесь требуются специальные математические методы — теория графов, матричный анализ и другие, используемые, например, в так называемых сетевых методах управления и планирования научных исследований

РИС. 3.2. Сеть научного влияния при решении проблемы обтекания затупленных тел гиперзвуковым потоком [4.13] Прямоугольники соответствуют отдельным исследователям (см. рис. 3.1). Сплошная линия — смысловые связи, пунктирная — исторические. Помимо сети смыслов здесь показана сеть связей, выявленных средствами историко-научного анализа. Эти две сети хорошо согласуются, что говорит в пользу индикатора научной связи «число смыслов»

РИС. 3.3. Сеть биомедицинских работ, построенная Г. Смоллом и Б. Гриффитом по 100 000 публикациям, цитированным 15 раз и более в 1973 г., на основе Индекса Гарфилда [4.138] Прямоугольники соответствуют отдельным публикациям. Числа на линиях, соединяющих попарно работы, соответствуют числу смыслов (см. рис. 3.1). (РИС. 3.2 и 3.3 см. на стр. 308)



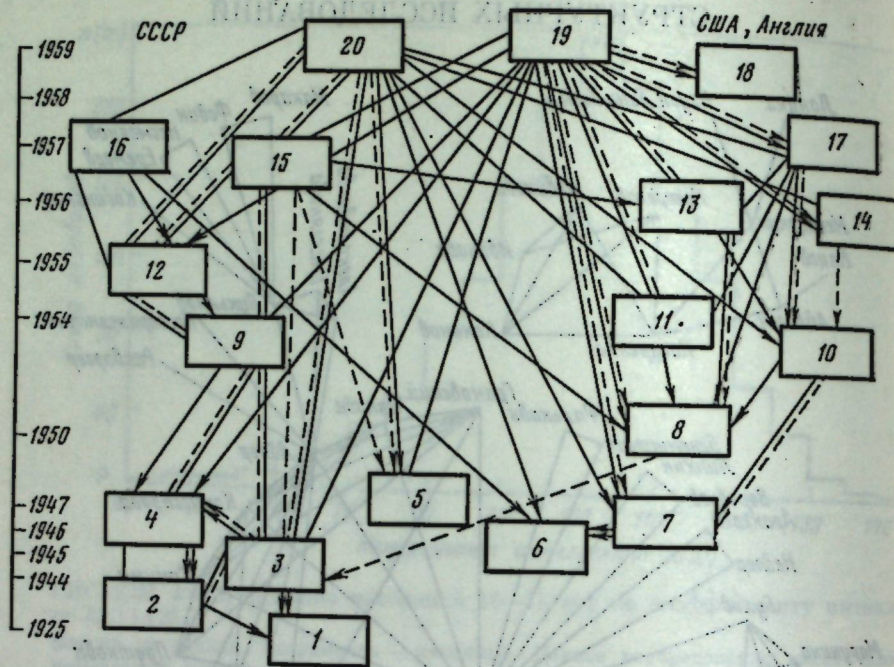


Рис. 3.2

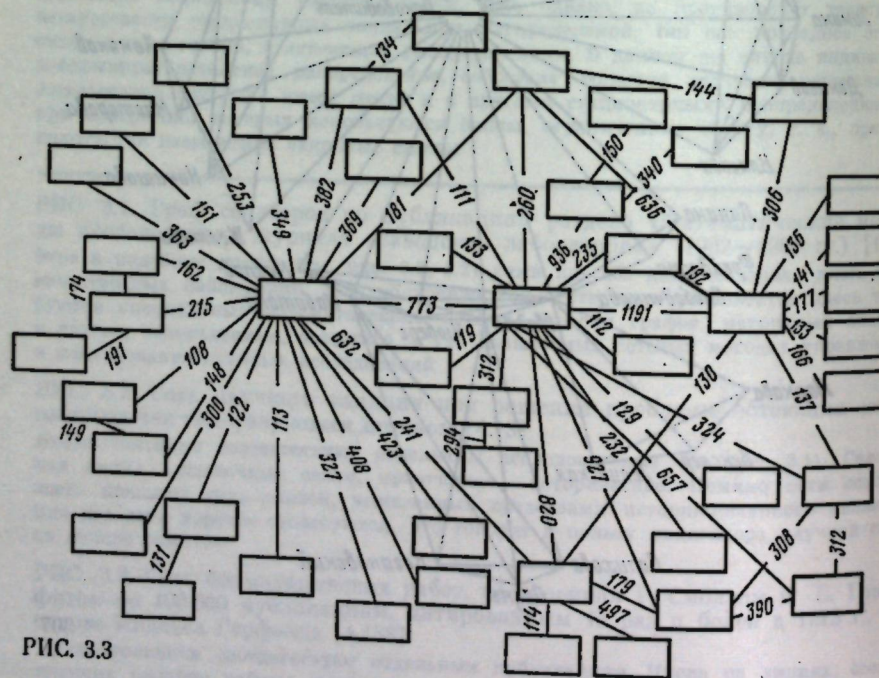


Рис. 3.3

4. РЕЗУЛЬТАТЫ КОМПЛЕКСНОГО ПРИМЕНЕНИЯ НАУКОМЕТРИЧЕСКИХ ИНДИКАТОРОВ

Рис. 4.1. Зависимость эффективности группы ученых (определяемой как число ссылок на работы группы, деленное на число ее членов) от численности группы [4.141]

Эффективность откладывается в логарифмическом масштабе, поэтому прямая на графике соответствует экспоненциальной зависимости. Эти и подобные данные (см. рис. 4.2, 4.8) могут быть полезными при определении оптимальных параметров группы для заданного объема ассигнований на ее исследовательскую проблематику

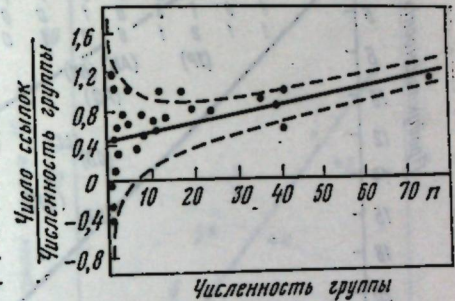
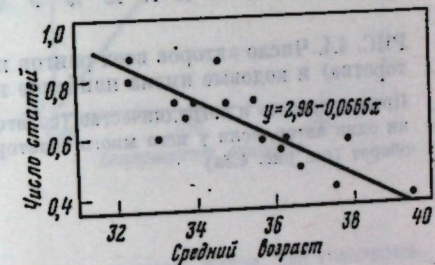


Рис. 4.2. Зависимость продуктивности группы ученых (определяемой по числу статей) от ее среднего возраста [3.100]

Я. Влахи, чьи данные здесь представлены, говорит о существовании линейной корреляции, обеспечивающей увеличение продуктивности группы с уменьшением ее среднего возраста. Из общих соображений ясно, что корреляция здесь не может быть линейной. В противном случае самая высокая научная продуктивность наблюдалась бы оптимальный средний возраст научной переход от одной исследовательской



у грудных младенцев. Следует искать группы, меняющейся, возможно, при проблематики к другой (см. рис. 4.1)

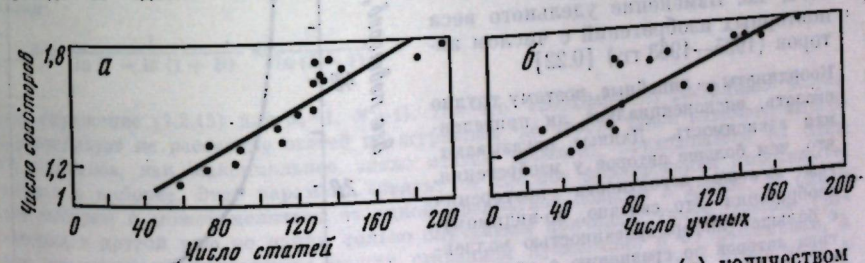


Рис. 4.3. Корреляция между числом авторов на статью и (а) количеством публикаций у отдельного ученого и (б) численностью персонала в отдельной физической лаборатории Чехословацкой Академии наук [2.70]

Линейная корреляция, показанная на графике (а), иллюстрирует довольно очевидный тезис о том, что, чем больше данный ученый склонен к сотрудничеству (соавторству), тем больше у него работ (см. рис. 4.4). Линейная корреляция, показанная на графике (б), говорит, что увеличение коллективности научного труда ведет к увеличению коллективности продукта труда — научной публикации

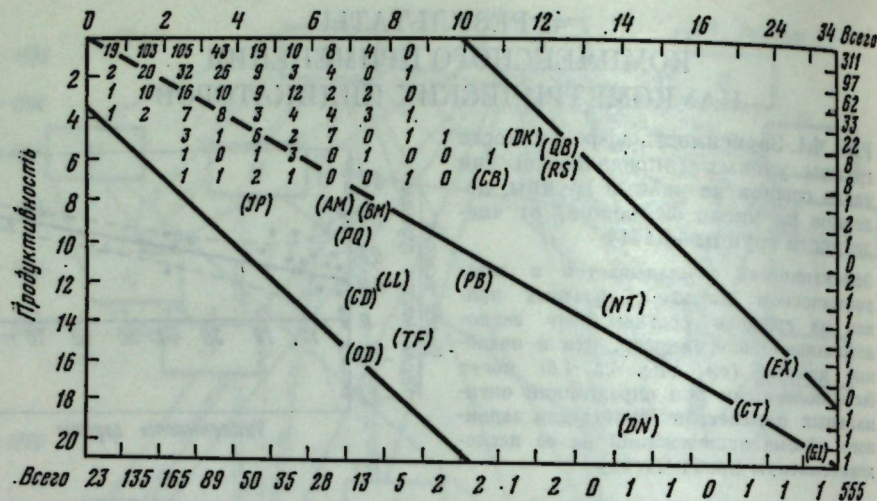


Рис. 4.4. Число авторов всех рангов продуктивности и сотрудничества (соавторства) и кодовые имена наиболее продуктивных авторов [2.30]

Продуктивность и сотрудничество (соавторство) хорошо коррелируют друг с другом: ни один автор, если у него много соавторов, не имеет низкой продуктивности, и наоборот (см. рис. 4.3а)

Рис. 4.5. Изменение удельного веса повторных изобретений с числом авторов (1935—1963 гг.) [0.22]

Координаты — линейные, поэтому трудно сказать, экспоненциальна ли приведенная зависимость. Данные показывают, что, чем больше авторов у изобретения, тем меньше вероятность повторения изобретения. Это связано, по-видимому, с большей информированностью коллектива авторов по сравнению с одним автором

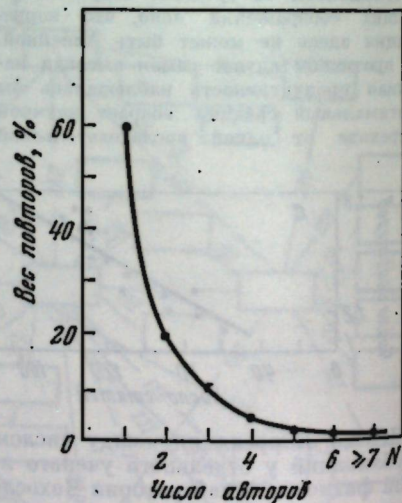


Рис. 4.6. Корреляция между старением публикаций и их рассеянием по журналам для отдельных научных дисциплин [4.49]

- 1 — зоология,
- 2 — энтомология,
- 3 — геология,
- 4 — ботаника,
- 5 — математика,
- 6 — химия,
- 7 — физиология,
- 8 — физика.

Коэффициент рассеяния k , откладываемый по оси абсцисс, определяется при этом аппроксимацией П. Коула

$$\frac{X(r)}{X(N)} = 1 + k \ln(r/N)$$

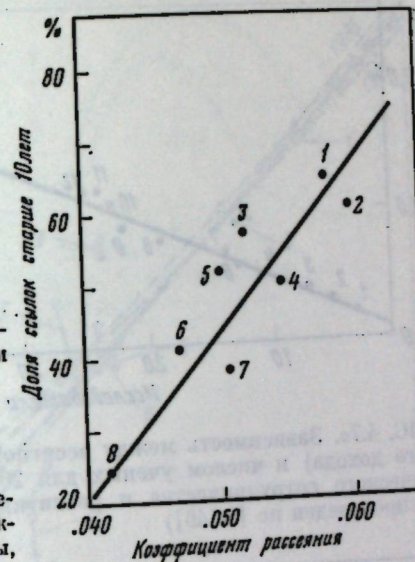
распределения Брэдфорда (см. выражение (3.3.2)). При расчете k М. К. Бакленд, чьи данные здесь представлены, берет $X(r)/X(N) = 0.90$. Другими словами наклон аппроксимации Коула к оси абсцисс определяется на участке, для которого выполняется условие $B \ll r$, где B — коэффициент рангового искажения. Аппроксимация Коула, как говорилось в тексте, соответствует распределению Ципфа с $\alpha = 1$ и $B \ll r$:

$$X(r) = -A \ln(1+B) + A \ln r$$

(см. выражение (7.2.13) для $\alpha = 1$ и $B \ll r$). Сравнивая с аппроксимацией Коула, находим:

$$k = \frac{1}{\ln N - \ln(1+B)} \approx \frac{1}{\ln(I-1)}$$

(см. выражение (7.2.15) для $\alpha = 1$, $N \gg 1$). Так что «коэффициент рассеяния» Коула характеризует не рассеяние статей по журналам, но просто такой параметр выборки (ссылки), приходящееся на один журнал в выборке. Этот параметр зависит от способа формирования исследователем выборки и может меняться от одной предметной области к другой, от одной выборки к другой уже по одним только случайным причинам. Таким образом, график описывает не корреляцию между старением публикаций и их рассеянием, а корреляцию между старением публикаций и максимальным числом публикаций в данной выборке на один журнал. Почему эта корреляция линейна — это особый вопрос. Ответ на него осложняется прежде всего тем, что мы не знаем, какова в действительности зависимость старения публикаций от их рассеяния по журналам для разных дисциплин, рассеяния, измеряемого показателем распределения Ципфа α . Для ответа на этот вопрос необходимо также знать, как конкретно формировал свои выборки Бакленд



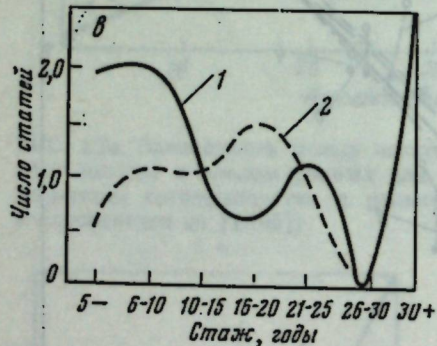
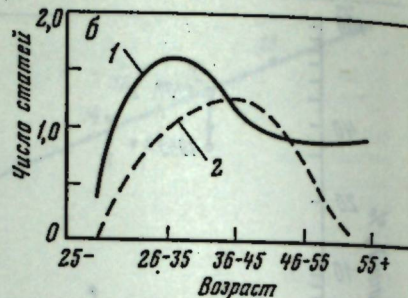
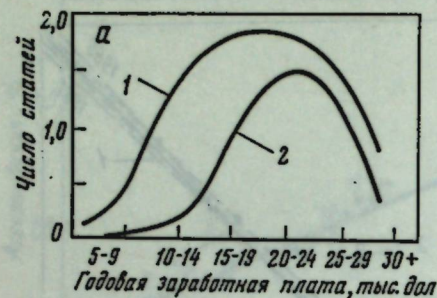


РИС. 4.8. Число опубликованных статей, приходящееся на одного человека, для групп с высоким и низким уровнями удовлетворенности в зависимости от (а) годовой заработной платы, (б) возраста и (в) рабочего стажа [3.99]

Данные отличаются от представленных на рис. 4.1 и 4.2 тем, что они показывают наличие некоторых оптимальных значений параметров, в зависимости от которых меняется продуктивность исследовательской группы

ЛИТЕРАТУРА

О. Наукометрия и науковедение

- 0.1. Маркс К., Энгельс Ф. Сочинения. 2-е изд., т. 1, 23.
- 0.2. Ленин В. И. Полное собрание сочинений.
- 0.3. Воспоминания о Марксе и Энгельсе. М.: Госполитиздат, 1956.
- 0.4. Алексеев И. С. Концепция дополнительности: Ист.-методол. анализ. М.: Наука, 1978.
- 0.5. Арапов М. В., Шрейдер Ю. А. Закон Ципфа и принцип симметрии системы. — В кн.: Семантика и информатика. М.: ВИНТИ, 1978, вып. 9, с. 74—95.
- 0.6. Бернал Дж. Д. Наука и общество. М.: Изд-во иностр. лит., 1953.
- 0.7. Бернал Дж. Д. Наука в истории общества. М.: Изд-во иностр. лит., 1969.
- 0.8. Бешелев С. Д., Гурвич Ф. Г. Экспертные оценки. М.: Наука, 1973.
- 0.9. Большаков И. А. Статистические оценки отдельных параметров частотных словарей. — НТИ. Сер. 2, 1972, № 11.
- 0.10. БСЭ. 3-е изд., 1971, т. 6, с. 330—337.
- 0.11. Врусиловский Б. Я. Математические модели в прогнозировании и организации науки. Киев: Наук. думка, 1975.
- 0.12. Вентцель Е. С. Теория вероятностей. М.: Физматгиз, 1962.
- 0.13. Винер Н. Творец и робот. М.: Прогресс, 1965.
- 0.14. Винер Н. Кибернетика. М.: Сов. радио, 1968.
- 0.15. Винниченко И. Г., Юрцунь А. Е. К вопросу анализа сетей информационных связей. — В кн.: Науковедение и информатика. Киев: Наук. думка, 1970, вып. 2, с. 106—108.
- 0.16. Гальтон Ф. Наследственность таланта, ее законы и последствия. СПб.: Ред. журн. Знания, 1875.
- 0.17. Глушков В. М. Науковедение и фундаментальные исследования. — В кн.: Науковедение и информатика. Киев: Наук. думка, 1976, вып. 15, с. 3—11.
- 0.18. Гнеденко Б. В. К теории предельных теорем для сумм независимых случайных величин. — Изв. АН СССР. Сер. мат., 1939, с. 181—232, 643—657.
- 0.19. Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. М.; Л.: Гостехиздат, 1949.
- 0.20. Голдинг У. Беседы Уильяма Голдинга. — Иностран. лит., 1973, № 10, с. 204—219.
- 0.21. Джини К. Средние величины. М.: Статистика, 1970.
- 0.22. Добров Г. М. Наука о науке. Киев: Наук. думка, 1966.
- 0.23. Добров Г. М. Наука о науке: Введение в общее науковедение. Киев: Наук. думка, 1970.
- 0.24. Добров Г. М., Клименюк В. Н., Смирнов Л. П. и др. Потенциал науки. Киев: Наук. думка, 1969.
- 0.25. Добров Г. М., Клименюк В. Н., Одрин В. М., Савельев А. Л. Организация науки. Киев: Наук. думка, 1970.

- 0.26. Добров Г. М., Коренной А. А., Юрцунь А. Е. Математическая модель классификации элементов по силе связи между ними.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1970, вып. 2, с. 95—100.
- 0.27. Добров Г. М., Коренной А. А., Юрцунь А. Е. Математическое обеспечение информационного анализа в науковедении.— В кн.: Материалы по науковедению. Киев: СОПС УССР, 1971, вып. 15, с. 50—98.
- 0.28. Ибрагимов И. М., Линник Ю. В. Независимые и стационарно связанные величины. М.: Наука, 1965.
- 0.29. Идлис Г. М. Космические силовые поля и некоторые вопросы структуры и эволюции галактической материи.— Изв. АФИ АН Каз. ССР, 1957, т. 4, с. 3—159.
- 0.30. Идлис Г. М. Математическая теория НОТ и оптимальной структуры НИИ. Алма-Ата: Наука, 1970.
- 0.31. Игнатъев А. А., Яблонский А. И. Аналитические структуры научной коммуникации.— В кн.: Системные исследования — 1975. М.: Наука, 1975, с. 64—81.
- 0.32. Каныгин Ю. М. О статистическом вероятностном подходе к оценкам научно-технического прогресса.— В кн.: Современный детерминизм и наука. Новосибирск: Ин-т истории, филологии и философии СО АН СССР, 1971, с. 175—177.
- 0.33. Кара-Мурза С. Г. Цитирование в науке и подходы к оценке научного вклада.— Вестн. АН СССР, 1981, № 5, с. 68—75.
- 0.34. Кендэлл М., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973.
- 0.35. Колмогоров А. Н. Математика.— В кн.: БСЭ. 2-е изд., 1954, т. 26, с. 464—483.
- 0.36. Коммуникация в современной науке. М.: Прогресс, 1976.
- 0.37. Корн Г., Корн Т. Справочник по математике. М.: Наука, 1974.
- 0.38. Кульбак С. Теория информации и статистика. М.: Наука, 1967.
- 0.39. Лозинский Л. С. Об одной модели оптимизации при формировании речи.— Кибернетика, 1970, № 2, с. 105—107.
- 0.40. Математические методы в исторических исследованиях. М.: Наука, 1972.
- 0.41. Математические методы в современной буржуазной социологии. М.: Прогресс, 1966.
- 0.42. Микулинский С. Р. Науковедение: проблемы и исследования 70-х годов.— Вопр. философии, 1975, № 7, с. 40—52.
- 0.43. Микулинский С. Р., Родный Н. И. История науки и науковедение.— В кн.: Очерки истории и теории науки. М.: Наука, 1969, с. 44.
- 0.44. Моделирование социальных процессов. М.: Наука, 1970.
- 0.45. Народное хозяйство СССР в 1975 г. М.: Статистика, 1976.
- 0.46. Наука сегодня. М.: Мол. гвардия, 1969.
- 0.47. Огурцов А. П. Научная школа как форма кооперации ученых.— В кн.: Школы в науке. М.: Наука, 1977, с. 248—262.
- 0.48. Организация научной деятельности. М.: Наука, 1968.
- 0.49. Осипов Г. В., Андреев Э. П. Проблемы формирования научного знания в процессе конкретных социальных исследований.— В кн.: Социология и математика. Новосибирск: Ин-т социологии СО АН СССР, 1970, с. 3—17.
- 0.50. Остаальд В. Великие люди. СПб.: Вестн. знания, 1910.
- 0.51. Пельц Д., Эндрюс Ф. Ученые в организации. М.: Прогресс, 1972.
- 0.52. Прохоров Ю. В., Розаков Ю. А. Теория вероятностей. М.: Наука, 1973.
- 0.53. Психологическое измерение. М.: Мир, 1967.
- 0.54. Суннес П., Зиннес Дж. Основы теории измерений.— В кн.: Психологическое измерение. М.: Мир, 1967, с. 9—110.

- 0.55. Хайтун С. Д. К вопросу о понятии энтропии.— В кн.: Проблемы истории и методологии научного познания. М.: Наука, 1974, с. 281—294.
- 0.56. Хвостова К. В. Некоторые вопросы применения количественных методов изучения социально-экономических явлений средневековья: (По материалам византийских источников XII—XIV вв.).— В кн.: Математические методы в исторических исследованиях. М.: Наука, 1972, с. 15—88.
- 0.57. Хурсин Л. А., Юрцунь А. Е., Коренной А. А. Ориентированный граф как математическая модель информационных связей.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1969, вып. 1, с. 90—97.
- 0.58. Циглер Г. Экстремальные принципы термодинамики необратимых процессов и механика сплошной среды. М.: Мир, 1966.
- 0.59. Черный А. И. Введение в теорию информационного поиска. М.: Наука, 1975.
- 0.60. Шрейдер Ю. А. О возможности теоретического вывода статистических закономерностей текста.— Пробл. передачи информ., 1967, № 1, с. 57—63.
- 0.61. Эйништейн А. Собрание научных трудов: В 4-х т. М.: Наука, 1965, Т. 1.
- 0.62. Юл Дж. Э., Кендэлл М. Дж. Теория статистики. М.: Гостехиздат, 1960.
- 0.63. Яблонский А. И. Стохастические модели научной деятельности.— В кн.: Системные исследования: Ежегодник, 1975. М.: Наука, 1976, с. 5—42.
- 0.64. Яблонский А. И. Структура и динамика современной науки: (Некоторые методол. пробл.) — В кн.: Системные исследования: Ежегодник, 1976. М.: Наука, 1977, с. 66—90.
- 0.65. Яблонский А. И. Модели и методы математического исследования науки: Науч.-аналит. обзор. М.: ИНИОН АН СССР, 1977.
- 0.66. Янч Э. Прогнозирование научно-технического прогресса. М.: Прогресс, 1974.
- 0.67. Aigner D. J., Goldberger A. S. Estimation of Pareto's law from grouped observations.— J. Amer. Statist. Assoc., 1970, vol. 65, p. 712—723.
- 0.68. Aitchison J., Brown J. A. The lognormal distribution. Cambridge: Univ. press, 1957.
- 0.69. Andrews F. Social and psychological factors which influence the creative process.— In: Perspectives in creativity. Chicago, 1975, p. 117—144.
- 0.70. Bernal J. D. Science and industry XIXth century. L., 1953.
- 0.71. Bulmer M. C. On fitting the Poisson lognormal distribution to species-abundance data.— Biometrics, 1974, vol. 30, p. 101—110.
- 0.72. Burdieu P. The specificity of the scientific field and the social condition of the progress of reason.— Soc. Sci. Inform., 1975, vol. 14, p. 19—47.
- 0.73. Champernownes D. G. A model of income distribution.— Econ. J., 1953, vol. 63, p. 318.
- 0.74. Connor P. E. Scientific research competence as a function of creative ability.— IEEE Trans. Eng. Manag., 1974, vol. EM-21, p. 2—8.
- 0.75. Cramer I. S. Empirical econometrics. Amsterdam; London, 1969.
- 0.76. Daniels G. H. Science in American society: a social history. N. Y., 1971.
- 0.77. Davis G. A., Peterson J. M., Farley F. H. Attitude, motivation, sensation seeking and belief in ESP as predictors of real creative behaviour.— J. Creat. Behav., 1974, vol. 8, p. 31—39.
- 0.78. Doeblin W. Sur l'ensemble de puissances d'une loi de probabilité.— Studia math., 1940, vol. 9, s. 71—96.
- 0.79. Goffman W., Morris T. G. Bradford's law and library acquisition.— Nature, 1970, vol. 226, p. 922—923.
- 0.80. Good I. J. The population frequencies of species and the estimation of population parameters.— Biometrics, 1953, vol. 40, p. 237—264.

- 1.12. *Гражданников Е. Д.* Динамические закономерности наукометрии.— В кн.: Материалы по науковедению. Киев: СОПС УССР, 1969, вып. 3, с. 93—104.
- 1.13. *Гражданников Е. Д., Щербаков А. И.* Элементарное введение в наукометрию. Новосибирск, 1970. 102 с. (Рукопись деп. в ИНИОН АН СССР 27.XI.70, № 67).
- 1.14. *Грановский Ю. В.* Наукометрические исследования в химии.— В кн.: Наукометрические исследования в химии. М.: Изд-во МГУ, 1974, с. 8—15.
- 1.15. *Грановский Ю. В.* Наукометрический анализ информационных потоков в химии. М.: Наука, 1980.
- 1.16. *Грановский Ю. В.* и др. Использование статистики премий и патентов для оценки развития исследований.— В кн.: Наукометрические исследования в химии. М.: Изд-во МГУ, 1974, с. 122—130.
- 1.17. *Грановский Ю. В., Страхов А. Б., Мурашова Т. И.* «Незримый» коллектив по планированию эксперимента.— НТИ. Сер. 2, 1972, № 12, с. 10—13.
- 1.18. *Грановский Ю. В.* и др. Выявление преимущественных направлений в развитии науки.— В кн.: Наукометрические исследования в химии. М.: Изд-во МГУ, 1974, с. 16—66.
- 1.19. *Грановский Ю. В., Мурашова Т. М., Страхов А. Б.* Некоторые вопросы оценки эффективности исследований.— В кн.: Наукометрические исследования в химии. М.: Изд-во МГУ, 1974, с. 131—136.
- 1.20. *Грановский Ю. В.* и др. Анализ информационной деятельности отечественных исследователей—докторов химических наук.— В кн.: Наукометрические исследования в химии. М.: Изд-во МГУ, 1974, с. 80—101.
- 1.21. *Добров Г. М., Коренной А. А.* Информационная база наукометрии.— В кн.: Теоретические проблемы информатики. М.: ВИНТИ, 1968, с. 174—202.
- 1.22. *Добров Г. М., Коренной А. А.* Опыт наукометрического анализа.— В кн.: Материалы по науковедению. Киев: СОПС УССР, 1971, вып. 16, с. 71—135.
- 1.23. *Егоров В. М.* Наукометрия потоку інформації лабораторії біологічного профілю.— Вісн. АН УССР, 1970, № 9, с. 84—88.
- 1.24. *Зеврев В. М.* Из опыта наукометрических наблюдений в области истории социологии в России (1860—1917 гг.).— В кн.: Методологические и методические проблемы контент-анализа. М.; Л.: ИСИ АН СССР, 1973, вып. 2, с. 122—124.
- 1.25. *Карпов М. М.* Закон ускоренного развития естественных наук.— Вopr. философии, 1963, № 4, с. 106—111.
- 1.26. *Клименюк В. И.* Анализ основных составляющих научного потенциала и методика выбора средств переработки информации по истории науки и техники: Автореф. дис. ...канд. филос. наук. М.: ИИЕТ АН СССР, 1967.
- 1.27. *Клименюк В. И.* Кадры экономической науки.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1975, вып. 14, с. 47—56.
- 1.28. *Козачков Л. С.* Системы потоков научной информации. Киев: Наук. думка, 1973.
- 1.29. *Козачков Л. С.* Язык информационного анализа. Киев: ИК АН УССР, 1975.
- 1.30. *Козачков Л. С., Хурсин Л. А.* Об основном вероятностном распределении науковедения.— В кн.: Анализ закономерностей и прогнозирования развития науки и техники. Киев: Укр НИИТИ, 1967, вып. 1, с. 42—45.
- 1.31. *Козачков Л. С., Хурсин Л. А.* Основное вероятностное распределение

- в системах информационных потоков.— НТИ. Сер. 2, 1968, № 2, с. 3—12.
- 1.32. *Коренной А. А.* Анализ библиографических связей и оценка значимости научных работ.— В кн.: Проблемы повышения эффективности научно-исследовательской работы: (Материалы науч.-практ. конф.). Новосибирск: Ин-т горного дела СО АН СССР, 1968, ч. 2, с. 90—96.
- 1.33. *Коул Дж. Р.* Схемы интеллектуального влияния в научных исследованиях.— В кн.: Коммуникация в современной науке. М.: Прогресс, 1976, с. 390—425.
- 1.34. *Лагерева К. С., Лифшиц С. Е., Матусов Ю. А.* Некоторые наукометрические измерения тенденций в области научно-технического прогнозирования.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1973, вып. 5, с. 343—345.
- 1.35. *Лазарсфельд П. Ф.* Измерение в социологии.— В кн.: Американская социология. М.: Прогресс, 1972.
- 1.36. *Лазарсфельд П. Ф.* Латентно-структурный анализ и теория тестов.— В кн.: Математические методы в социальных науках. М.: Прогресс, 1973, с. 42—58.
- 1.37. *Лебин Б. Д.* Возраст ученого и научный потенциал.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1970, вып. 3, с. 110—115.
- 1.38. *Логика развития и наукометрический анализ отдельных направлений в химии.* М.: Изд-во МГУ, 1976.
- 1.39. *Лучник М. К., Осетров В. Д.* Сравнительный наукометрический анализ советских и американских журналов по океанологии.— В кн.: Материалы V симпозиума по науковед. и науч.-техн. прогнозированию. Киев: УкрНИИТИ, 1974, ч. 3, с. 61—62.
- 1.40. *Методологические проблемы теории измерений.* Киев: Наук. думка, 1966.
- 1.41. *Мирский Э. М.* Некоторые количественные параметры журнала «Вопросы философии».— Вopr. философии, 1968, № 6, с. 145—154.
- 1.42. *Михайлов А. И., Черный А. И., Гиляровский Р. С.* Научные коммуникации и информатика. М.: Наука, 1976.
- 1.43. *Мульченко З. М., Грановский Ю. В., Страхов А. Б.* О наукометрических характеристиках информационной деятельности ведущих ученых.— В кн.: Проблемы удовлетворения информационных потребностей. М.: ВИНТИ, 1975, с. 127—147.
- 1.44. *Налимов В. В., Адлер Ю. П., Грановский Ю. В.* Информационная система по математической теории эксперимента.— В кн.: Кибернетика и документалистика. М.: Наука, 1966, с. 138—149.
- 1.45. *Налимов В. В., Мульченко З. М.* Еще раз к вопросу о концепции экспоненциального роста.— НТИ. Сер. 2, 1968, № 8, с. 13—14.
- 1.46. *Налимов В. В., Мульченко З. М.* Наукометрия: Изучение развития науки как информационного процесса. М.: Наука, 1969.
- 1.47. *Налимов В. В., Мульченко З. М.* Наука и биосфера: опыт сравнения двух систем.— Природа, 1970, № 11, с. 55—68.
- 1.48. *Наукометрические исследования в химии.* М.: Изд-во МГУ, 1974.
- 1.49. *Научно-техническая революция и изменение структуры научных кадров СССР.* М.: Наука, 1973.
- 1.50. *Несостайлов Г. А.* О методике оценки деятельности академического института.— В кн.: Новая техника и оценка эффективности научных исследований. Минск: ИТМО, 1974.
- 1.51. *Несмеянов А. И.* Развитие радиохимии и работы кафедры радиохимии.— В кн.: Логика развития и наукометрический анализ отдельных направлений в химии. М.: Изд-во МГУ, 1976, с. 6—18.
- 1.52. *Новый Л., Фолга Я.* О применении количественных методов в исследованиях по истории математики.— В кн.: Анализ тенденций и про-

- гнозирование научно-технического прогресса. Киев: Наук. думка, 1967, с. 235—262.
- 1.53. Ориент И. М. Наукометрические исследования в аналитической химии.—Завод. лаб., 1975, т. 41, № 9, с. 1071—1077.
 - 1.54. Петров М. К. Институт публикаций и научная деятельность. В кн.: Тр. 13 Междунар. конгр. по истории науки. Секция 1А, Москва, 1971. М.: ИИЕТ АН СССР, 1974, с. 189—191.
 - 1.55. Петрова Т. М. Методологические особенности количественного выделения структурных единиц науки.—В кн.: Системные исследования: Ежегодник, 1975. М.: Наука, 1976, с. 43—53.
 - 1.56. Полушкин В. А., Жданова Г. С. Информационные барьеры и возможности их количественного изучения. М.: ВИНТИ, 1970.
 - 1.57. Потанов В. М., Кочетова Э. К. Указатели реферативных журналов как источник наукометрической информации.—В кн.: Логика развития и наукометрический анализ отдельных направлений в химии. М.: Изд-во МГУ, 1976, с. 95—100.
 - 1.58. Прайс Д. Малая наука, большая наука.—В кн.: Наука о науке. М.: Прогресс, 1966, с. 281—384.
 - 1.59. Прайс Д. Наука о науке.—В кн.: Наука о науке. М.: Прогресс, 1976, с. 236—254.
 - 1.60. Радбиль О. С., Турыкина И. Л. Об оценке информативности и профильности зарубежных журналов (на примере изданий по гастроэнтерологии).—ИТИ. Сер. 1, 1976, № 3, с. 33—34.
 - 1.61. Солдатов С. Количественный анализ истории науки.—Изв. АН ЭССР, 1970, т. 19. Сер. Обществ. науки, № 1, с. 83—94.
 - 1.62. Сорокин Ю. Н. О возможной перспективе первичных и вторичных научных публикаций.—ИТИ. Сер. 1, 1969, № 1, с. 3—5.
 - 1.63. Фейзин С. Е., Голованова М. А. Оценка творческих способностей и квалификация научных коллективов.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1971, вып. 4, с. 287—293.
 - 1.64. Хайтун С. Д. Анализ системы связей между учеными с помощью величины энтропии.—В кн.: Социально-психологические проблемы науки. М.: Наука, 1973, с. 232—236.
 - 1.65. Червинский Ю. М. Коллективность научного труда с позиций наукометрии.—В кн.: Актуальные проблемы физического воспитания и спорта. М.: ГЦОЛИФК, 1974, с. 25—26.
 - 1.66. Червинский Ю. М. Наукометрический анализ развития научных исследований по физической культуре и спорту (на примере журнала «Теория и практика физической культуры» за 1962—1973 гг.).—Теория и практика физ. культуры, 1975, № 1, с. 65—67.
 - 1.67. Черная М. М., Руденкова О. Ф. Некоторые параметры информационных потоков по лесному хозяйству.—ИТИ. Сер. 1, 1976, № 12, с. 15—17.
 - 1.68. Шишова М. П. Наукометрический анализ некоторых аспектов развития химии ВМС.—В кн.: Логика развития и наукометрический анализ отдельных направлений в химии. М.: Изд-во МГУ, 1976, с. 35—48.
 - 1.69. Шляхун Г. Г., Пилиди Г. В. Некоторые результаты количественного исследования математических публикаций.—В кн.: Вопросы истории естествознания и техники. М.: Наука, 1979, вып. 4, с. 48—58.
 - 1.70. Шрейдер Ю. А. К вопросу о структуре научных коллективов.—В кн.: Вопросы теории и практики информационных систем. Киев: УкрНИНТИ, 1967, с. 21—42.
 - 1.71. Adams E. W. Elements of a theory of inexact measurement.—Phil. Sci., 1965, vol. 32, p. 205—228.
 - 1.72. Amick D. J. Scientific elitism and the information system of science.—J. Amer. Soc. Inform. Sci., 1973, vol. 24, N 5, p. 317—327.

- 1.73. Auger P. Tendences: actuelles de la recherche scientifique. P., 1961.
- 1.74. Beaver D. de B., Rosen L. Studies in scientific collaboration. Pt. I. The professional origin of scientific co-authorship.—Scientometrics, 1978, vol. 1, p. 65—85; Pt II. Scientific co-authorship research productivity and visibility in the french scientific elite, 1799—1830. Ibid., 1979, vol. 1, p. 133—149; Pt. III. Professionalization and the national history of modern scientific co-authorship.—Ibid., 1979, vol. 1, p. 231—246.
- 1.75. Blasco P. G. Historia y Sociologia de la Ciencia. Madrid: Alianza Editorial S. a., 1979.
- 1.76. Boalt G., Lantis H., Herlin H. The academic patterns. A comparison between researchers and non-researchers, men and women. Stockholm: Almqvis and Wiksell, 1973.
- 1.77. Bookstein A. Bibliometric symmetry and the Bradford-Zipf Law. Chicago: Univ. of Chicago, Graduate Library School, 1975.
- 1.78. Bookstein A. The bibliometric distributions.—Libr. Quart., 1976, vol. 45, p. 416—423.
- 1.79. Bookstein A. Patterns of scientific productivity and social change: A discussion of Lotka's law and bibliometric symmetry.—J. Amer. Soc. Inform. Sci., 1977, vol. 28, p. 206—210.
- 1.80. Bountry G.—A. Quantity versus quality in scientific research. II: The paper explosion.—Impact Sci. Soc., 1970, vol. 20, p. 195—206.
- 1.81. Brookes B. C., Griffiths J. M. Frequency-rank distributions.—J. Amer. Soc. Inform. Sci., 1978, vol. 29, p. 5—13.
- 1.82. Brožek V., Karen P. Dynamics of information flow in the field of rare earth carbides research.—Scientometrics, 1979, vol. 1, p. 339—358.
- 1.83. Bucland M. K., Hindle A. Library Zipf.—J. Doc., 1969, vol. 25, p. 52—57, p. 154.
- 1.84. A Century of doctorates: Data analysis of growth and change. NAS, Wash.: D. C., 1978.
- 1.85. Chubin D. E., Studer K. E. Knowledge and structures of scientific growth of a cancer problem domain.—Scientometrics, 1979, vol. 1, p. 171—193.
- 1.86. Chuchman C. W., Ratoosh P. (eds.). Measurement: Definition and theories / Eds. C. W. Chuchman, P. Ratoosh. N. Y., 1959.
- 1.87. Cohen J. E. Publication rate as a function of laboratory size in a biomedical research institution.—Scientometrics, 1980, vol. 2, p. 35—52.
- 1.88. Cole J. R., Cole S. Social stratification in science. Chicago; London: Univ. Chicago press, 1973, XIV.
- 1.89. Cole P. F. Analysis of reference question records as a guide to the information requirements of scientists.—J. Doc., 1958, vol. 14, N 4, p. 197—207.
- 1.90. Crane D. Invisible colleges. Diffusion of knowledge in scientific communities. Chicago, 1972.
- 1.91. Crawford S. Communication centrality and performance.—In: Proceeding of the American society for information science. Philadelphia, 1970, p. 45—48.
- 1.92. Davis H. T. The analysis of economic time series. Bloomington (Ind.): Principia press, 1941, XIV.
- 1.93. Davis H. T. The summation of series. San Antonio (Tex.), Principia press Trinity Univ., 1962, IX.
- 1.94. Dingle H. A theory of measurement.—British J. Philos. Sci., 1950, vol. 1, p. 20.
- 1.95. Donohue J. C. A bibliometric analysis of certain information science literature.—J. Amer. Soc. Inform. Sci., 1972, vol. 23, N 5, p. 313—317.
- 1.96. Edwards Sh. A., McCarrey M. W. Measuring the performance of researchers.—Res. Manag., 1973, vol. 16, N 1, p. 34—41.
- 1.97. Fairthorne R. A. Algebraic representation of storage and retrieval lan-

- guage.—In: Proc. of the Intern. Conf. on Sci. Inform., National Academy of Science, Wash., 1958.
- 1.98. Fairthorne R. A. Empirical hyperbolic distribution (Bradford — Zipf — Mandelbrot) for bibliometric description and prediction.— J. Doc., 1969, vol. 25, N 4, p. 319—343.
 - 1.99. Fermi E. On the origin of the cosmic radiation.— Phys. Rev., 1949, vol. 75, N 8, p. 1169—1174.
 - 1.100. Finneston H. M. The future of the journals.— Nature, 1967, vol. 214, N 5083, p. 47—48.
 - 1.101. Frame J. D. Measuring scientific activity in lesser developed countries.— Scientometrics, 1980, vol. 2, p. 133—145.
 - 1.102. Galton F. Hereditary genius. L.: Macmillan, 1869; Galton F. English men of science. L., 1874.
 - 1.103. Garfield E. Is there a future for the scientific journals? — Sci. and Public Policy, 1975, vol. 2, p. 498—501.
 - 1.104. Garvey W., Griffith B. Scientific communication: its role in the conduct of research and creation of knowledge.— Amer. Psychol., 1971, vol. 26, N 4, p. 349—362.
 - 1.105. Gilbert G. M. Measuring the growth of science: A review of scientific growth.— Scientometrics, 1979, vol. 1, p. 9—34.
 - 1.106. Gilbert G. N., Woolgar S. The quantitative study of science: An examination of the literature.— Sci. Stud., 1974, vol. 4, N 3, p. 279—294.
 - 1.107. Ginzburg V. L. The astrophysics of cosmic rays.— Sci. Amer., 1969, vol. 220, N 2, p. 51—63.
 - 1.108. Graubard S. R., Holton G. Excellence and leadership on a democracy. N. Y., 1962, p. 94—131.
 - 1.109. Haight F. A. Some statistical problems in connection with word association data.— J. Math. Psychol., 1966, vol. 3, N 1, p. 217—233.
 - 1.110. Hattun S. D. Scientific investigations in the USSR.— Scientometrics, 1980, vol. 2, p. 65—84.
 - 1.111. Hattun S. D. The stationary scientometric distributions. Part I. The different approximations.— Scientometrics, 1982, vol. 4, p. 5—25; Part II. The non-Gaussian nature of scientific activities.— Ibid., p. 89—104; Part III. The role of the Zipf distribution.— Ibid., p. 183—196.
 - 1.112. Harrold R. W. An evaluation of measurable characteristics within army laboratories.— IEEE Trans. Eng. Manag., 1969, vol. EM-16, N 1, p. 16—23.
 - 1.113. Hart P. E., Phelps Brown E. H. The sizes of trade unions: A study in the laws of aggregation.— Econ. J., 1957, vol. 67, p. 1—15.
 - 1.114. Hart P. E., Prajs S. J. The analysis of business concentration: a statistical approach.— J. Roy. Statist. Soc., 1956, vol. A-119, N 2, p. 150—180.
 - 1.115. Holton G. Scientific research and scholarship: Notes toward the design of proper scales.— Proc. Amer. Soc. Arts and Sci. (Daedalus), 1962, vol. 91, p. 362—399.
 - 1.116. Hubert J. J. A relationship between two forms of Bradford's law.— J. Amer. Soc. for Inform. Sci., 1978, vol. 29, p. 159—161.
 - 1.117. Hubert J. J. A rank-frequency model for scientific productivity.— Scientometrics, 1981, vol. 3, p. 191—202.
 - 1.118. Inhaber H., Lipsett M. S. Gaps in «Gaps in technology» and other innovation inventories.— Scientometrics, 1979, vol. 1, p. 405—417.
 - 1.119. Kendall M. G. The bibliography of operational research.— Oper. Res. Quart., 1960, vol. 11, p. 31—36.
 - 1.120. Kendall M. G. Natural law in the social sciences.— J. Roy. Statist. Soc., 1961, vol. 124, p. 4—16.
 - 1.121. Klein E. Vorlesungen über die Entwicklung der Mathematik in 19. Jahrhundert. B., 1926.

- 1.122. Kraft D. H., Polacsek R. A. Biomedical literature dynamics.— Meth. Inform. Med., 1974, vol. 13, S. 242—248.
- 1.123. Magyar G. Bibliometric analysis of a new research subfield.— J. Doc., 1974, vol. 10, p. 32—40.
- 1.124. Meadows A. J. Communication in science. L.: Burretworths, 1974.
- 1.125. Menard H. W. Science: growth and change. Cambridge (Mass.), 1971.
- 1.126. Montroll E. M., Badger W. W. Introduction to quantitative aspects of social phenomena N. Y.: Gordon and Breach Science, 1974.
- 1.127. Moravcsik M. J. Measures of scientific growth.— Res. Policy, 1973, vol. 2, p. 266—275.
- 1.128. Moravcsik M. J. Science development: The building of science in less developed countries: Program for advanced studies in institution building and technical assistance methodology PASITAM. International Development Research Center. Bloomington (Ind.): Indiana Univ., 1975.
- 1.129. Mulchenko Z. M., Granovsky Yu. V., Strakhov A. B. On scientometrical characteristics of information activities of leading scientists.— Scientometrics, 1979, vol. 1, p. 307—326.
- 1.130. Naranan S. Bradford's law of bibliography of science: on interpretation.— Nature, 1970, vol. 227, p. 631—632.
- 1.131. Naranan S. Power law relations in science bibliography: a selfconsistent interpretation.— J. Doc., 1971, vol. 27, p. 83—97.
- 1.132. Narin F. Objectivity versus relevance in studies of scientific advance.— Scientometrics, 1979, vol. 1, p. 35—42.
- 1.133. OECD. Patterns of resources devoted to research and experimental development on the OECD Area 1963—71 SPT(74)12, OECD. P., 1974.
- 1.134. Oromaner M. Professional age and the reception of sociological publications: a Test of the Zuckerman-Merton hypothesis.— Soc. Stud., 1977, vol. 7, p. 381—388.
- 1.135. Pareto V. Cours d'économie politique. P., 1897.
- 1.136. Phelps R. H., Herlin J. P. Alternatives to the scientific periodical.— UNESCO Bull. Libr., 1960, vol. 14, p. 61—67.
- 1.137. Pietsch F. Arbeitsgemeinschaft für Forschung des Nordrhein.— Westfalen Heft, 1954, vol. 38.
- 1.138. Platz A. Psychology of the scientist. XI. Lotka's law and research visibility.— Psychol. Repts, 1965, vol. 16, p. 566—568.
- 1.139. Price D. J. de S. Quantitative measures of development of science.— Arch. Inform. Hist. Sci., 1951, vol. 14, p. 85—93.
- 1.140. Price D. J. de S. Science since Babilon. New Haven: Yale Univ. press, 1961.
- 1.141. Price D. J. de S. Measuring the size of science.— Proc. Israel Acad. Sci. and Humanities, vol. 1 (1969—1970), p. 98—111.
- 1.142. Price D. J. de S. The productivity of research scientists.— In: Yearbook of Science and the Future 1975. Chicago etc.: Encyclopedia Britannica, Inc., 1974.
- 1.143. Price D. J. de S. A general theory of bibliometric and other cumulative advantage.— J. Amer. Soc. Inform. Sci., 1976, vol. 27, p. 292—306.
- 1.144. Price D. J. de S. Cumulative advantage Urn games explained: a reply to Kantor.— J. Amer. Soc. Inform. Sci., 1978, vol. 29, p. 204—206.
- 1.145. Price D. J. de S. Ups and downs in the pulse of science and technology.— In: The sociology of science / Ed. J. Gaston. San Francisco, 1978, ch. 8.
- 1.146. Price D. J. de S. Towards a comprehensive system of science indicators.— Sci. Yugosl., 1980, vol. 6(1/4), s. 45—65.
- 1.147. Price D. J. de S. Research notes: The analysis of scientometric matrices for policy implications.— Scientometrics, 1981, vol. 3, p. 47—53.
- 1.148. Price D. J. de S. The analysis of square matrices of scientometric transaction.— Ibid., p. 55—63.

- 1.149. Price D. J. de S., Gürsey S. Some statistical results for the numbers of authors in the States of the United States and the nations of the world: Preface to ISI's Who is publishing in science 1975 Annual. Philadelphia: P. A., 26—34, 1975.
- 1.150. Price D. J. de S., Gürsey S. Studies in scientometrics. Part I. Transians and continuance in scientific authorship.—Intern. Forum Inform. and Doc., 1976, vol. 1, p. 17—24; Pt III. The relation between source author and cited author population.—Ibid., p. 19—22.
- 1.151. Qurashi M. M. An analysis of defence spending in relation to the gross national product and degree of sophistication. Part I.—J. Pakistan Defence Sci. Organization, 1969, vol. 3, p. 22—32.
- 1.152. Qurashi M. M. The mechanics of R & D and technology transfer. Islamabad: Pakistan Acad. Sci., 1978.
- 1.153. Ruff I., Brown T. A tudománytörténeti elemzés a tudományágazati elemzésre.—Művelődés, 1977, vol. 22, N 3, s. 216—230.
- 1.154. Science Indicator 1972: Report of the National Science Board. Wash.: Gov. print. off., 1973, VIII.
- 1.155. Sengupta I. N. Choosing physiology periodicals: a recent study of the growth of its literature.—Ann. Libr. Sci. and Doc., 1973, vol. 20, N 1—4, p. 39—57.
- 1.156. Shokley W. On the statistics of individual variation of productivity in research laboratories.—Proc. IRE, 1957, vol. 45, N 3, p. 279—290.
- 1.157. Vlachý J. Survey materials for scientific boards and members of the presidium. Praha: Czechosl. Acad. Sci., October 1969.
- 1.158. Vlachý J. Frequency distribution of scientific performance. A bibliography of Lotka's law and related phenomena.—Scientometrics, 1978, vol. 1, p. 109—131.
- 1.159. Wooster H. The future of scientific publishing.—J. Washington Acad. Sci., 1970, vol. 60, N 2, p. 44—45.
- 1.160. Zipf K. Human behaviour and the principle of least effort. Cambridge (Mass.), 1949.

2. Наукометрия: статистический метод

- 2.1. Бастраков Г. М., Гражданников Е. Д., Щербаков А. И. Динамика численности научных работников в СССР.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1970, вып. 3, с. 136—138.
- 2.2. Бацапов С. С., Брусенцов Ф. А. Опыт статистического анализа кристаллоструктурных работ.—Кристаллография, 1966, т. 11, № 3, с. 393—397.
- 2.3. Бацапов С. С., Ручкина В. Н. О развитии неорганической химии: (Количественное сопоставление по данным РЖ «Химия» за 1962 и 1966 гг.).—В кн.: Вопросы истории естествознания и техники. М.: Наука, 1971, вып. 2, с. 44—46.
- 2.4. Бернадский В. Н., Журавков В. В. Выбор перспективных направлений и формирование системы целей научных работ в области сварочной науки и техники.—В кн.: Наукоедение и информатика. Киев: Наук. думка, 1976, вып. 15, с. 61—76.
- 2.5. Варшавский К. Будущее возрастной структуры научных кадров.—В кн.: 13-й Междунар. конгр. по истории науки. Секция 1А, II, Москва, 1971. М.: ИИЕНТ, 1974, с. 292—294.
- 2.6. Варшавский К. М., Панасова Н. П. Статистика премий как науковедческий источник.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1971, вып. с. 214—221.

- 2.7. Василенко И. В., Клименко В. Н. Анализ уровня использования трудовых ресурсов в педагогической науке.—В кн.: Наукоедение и информатика. Киев: Наук. думка, 1974, вып. 10, с. 62—69.
- 2.8. Голостенев М. Е. Анализ авторских коллективов и проблемы проектирования научных центров.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1971, вып. 4, с. 380—386.
- 2.9. Гражданников Е. Д. Хронологические спектры.—В кн.: Анализ тенденций и прогнозирование научно-технического прогресса. Киев: Наук. думка, 1967, с. 318—328.
- 2.10. Добров Г. М., Игнатюк Л. А. Организационные структуры авторских коллективов изобретателей.—В кн.: Наукоедение и информатика. Киев: Наук. думка, 1976, вып. 16, с. 63—67.
- 2.11. Добров Г. М., Коренной А. А. Научно-техническая информация и эффективность управления наукой.—В кн.: Об эффективности научно-информационной деятельности. М.: МФД, 1976, с. 47—72.
- 2.12. Дуженков В. И. Проблемы территориальной организации научной деятельности.—В кн.: Проблемы деятельности ученого и научных коллективов. М.: Л.: Наука, 1977, вып. 6, с. 47—53.
- 2.13. Заремба С. А., Алабян И. М. Структура информационного потока по обогащению полезных ископаемых.—НТИ. Сер. 1, 1973, № 12, с. 17—19.
- 2.14. Игнатюк Л. А. К вопросу о коллективности в техническом творчестве.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1971, вып. 4, с. 147—152.
- 2.15. Кацаров И. Возраст и возможности за развитие и творчество на научных работниках.—В кн.: Формы за повишаване квалификация на научните работници: Балкански симпозиум. Варна, 1972. Варна: Съюз на научните работници в България, 1973, с. 131—141.
- 2.16. Кашафуддинова Е. С., Гришина Ф. Е., Мельдизина М. Н. Анализ запрашиваемости журналов по электротехнике из фонда информационно-справочного центра ВИНТИ.—НТИ. Сер. 2, 1975, № 8, с. 8—11.
- 2.17. Кашафуддинова Е. С., Гришина Ф. Е., Мельдизина М. Н. Закономерности рассеяния и старения информации в изучении запрашиваемости журналов.—Электротехн. пром-сть. Общепромышленные вопросы, 1976, № 11, с. 26—29.
- 2.18. Козлова Т. З., Кузнецов А. Г. Возрастная структура научных кадров СССР: тенденции и перспективы.—В кн.: Наукоедение и информатика. Киев: Наук. думка, 1974, вып. 11, с. 22—28.
- 2.19. Комаров В. Д. Об одном методе изучения динамики численности выдающихся ученых.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1974, вып. 4, с. 234—237.
- 2.20. Коренной А. А., Левченко О. Г. О формализованных приемах анализа массивов опубликованных научно-технических документов.—В кн.: Материалы V симпозиума по науковедению и науч.-техн. прогнозированию. Киев: УкрНИИТИ, 1974, ч. 3, с. 55—57.
- 2.21. Крутько В. Н. Сравнение тенденций в количественности авторства публикаций систем «наука», «техника» и «производство».—В кн.: Информационные и управляющие системы в науковедении. Киев: ИК АН СССР, 1975, с. 20—29.
- 2.22. Кугель С. А. Изменение профессиональной революции.—В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1971, вып. 4, с. 249—255.
- 2.23. Лойферман П. Г., Волков В. Н., Дуженко В. С. Исследование рангового распределения документальных источников по их роли в формировании информационных потоков.—НТИ. Сер. 2, 1975, № 12, с. 8—13.

- 2.24. Максимов В. В. Некоторые количественные оценки соавторской деятельности ученых.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1969, вып. 2, с. 158—162.
- 2.25. Миздал В. Л., Храброва О. П., Федоскина В. В. Распределение статей по некоторым радиобиологическим аспектам состояния системы кровообращения и биохимических процессов в периодических изданиях.— НТИ, Сер. 1, 1977, № 5, с. 22—24.
- 2.26. Мончев Н. М. Тенденция к коллективности в изобретательской деятельности Всесоюзного электротехнического института имени В. И. Ленина.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1973, вып. 5, с. 206—210.
- 2.27. Мончев Н. М. Анализ изменения числа авторов изобретений Всесоюзного электротехнического института им. В. И. Ленина.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1973, вып. 9, с. 64—69.
- 2.28. Набоков И. И. Профессиональная структура кадров и особенности перемены труда в науке (на материалах Карельской АССР).— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1973, вып. 5, с. 176—180.
- 2.29. Осетров В. Д. Авторский коллектив «Украинского ботанического журнала»: Наукометрическая характеристика новых авторов.— Укр. ботан. журн., 1976, т. 33, № 4, с. 407—408.
- 2.30. Прайс Д. Дж. де., Бишер Д. де. Сотрудничество в «невидимом колледже».— В кн.: Коммуникация в современной науке. М.: Прогресс, 1976, с. 335—350.
- 2.31. Савельев А. А. Структура руководящих кадров в научных учреждениях.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1971, вып. 4, с. 171—176.
- 2.32. Сидорова Э. М. Из истории изучения структуры научных кадров в 20—30-е годы.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука, Ленингр. отд-ние, 1970, вып. 3, с. 199—204.
- 2.33. Солинский В. С., Торф Э. М. Состав и структура кадров в сфере науки и научного обслуживания.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1972, вып. 6, с. 40—46.
- 2.34. Фокенкан Дж. Машинный перевод и изучение научного языка.— Мир науки, 1960, № 2, с. 25—27.
- 2.35. Хомский Н., Миллер Дж. А. Конечные модели использования языка.— Кибernet. сб. Новая сер., 1967, № 4, с. 141—218.
- 2.36. Щелищ П. Б. О «широкой» и «узкой» специализации ученых.— В кн.: Проблемы деятельности ученого и научных коллективов. М.; Л.: Наука, 1977, вып. 6, с. 127—131.
- 2.37. Ben-David J. Scientific productivity and academic organization in 19th century medicine.— Amer. sociol. rev., 1960, vol. 25, N. 6.
- 2.38. Berelson B. Graduate education in the United States. N. Y.: McGraw-Hill Book Co, 1960, p. 55.
- 2.39. Carpenter M. P., Narin F. The subject comparison of the worlds scientific journals.— Scientometrics, 1980, vol. 2, p. 53—63.
- 2.40. Clarke B. L. Multiple authorship trends in scientific papers.— Science, 1964, vol. 143, N 3608, p. 822—824.
- 2.41. Conrad M. G. Growth of biological literature and the future of «Biological Abstracts».— Fed. Proc., 1957, vol. 16, N 3, p. 711—712.
- 2.42. Evenson R. E., Kislsey Y. Agricultural research and productivity. New Haven: Yale Univ. press, 1975.
- 2.43. Figueiredo L. M. Distribuanda literatura geologica brasileira: estudo bibliometrico.— Sci. Inform., 1973, vol. 2, N 1, p. 27—40.
- 2.44. Garvey W. D., Lin N., Nelson C. E. Communication in the physical and the social science.— Science, 1970, vol. 170, N 3963, p. 1166—1173.

- 2.45. Glass B. Information critic in biology.— Bull. Atom. Sci., 1962, vol. 18, N 6, p. 6—12.
- 2.46. Gottschalk Ch. M., Desmond W. S. Worldwide census of scientific and technical serials.— Amer. Doc., 1963, vol. 14, N 3, p. 188—194.
- 2.47. Hirsch W., Singleton J. F. Research support, multiple authorship and publication in sociological journals, 1936—1964: Prepr. N. Y., 1965, June.
- 2.48. Increased subscription rates due in 1967.— Chem. and Eng. News, 1966, vol. 44, N 6, p. 88—90.
- 2.49. Inhaber H. Changes in centralization of science.— Res. Policy, 1977, vol. 6, blz. 178—193.
- 2.50. Jahrbuch über Fortschritte der Mathematik. B., 1889, vol. 21, S. 2—4.
- 2.51. Keenan S., Atherton P. The journal literature of physics. N. Y., 1964.
- 2.52. Kochen M., Tagliacozzo R. Matching authors and readers of scientific papers.— Inform. Storage and Retrieval, 1974, vol. 10, N 5/6, p. 197—210.
- 2.53. Lehman H. C. The exponential increase of man's cultural output.— Social Forces, 1947, vol. 25, N 3, p. 281—290.
- 2.54. Lipetz B. A. Information sciences.— Amer. Doc., 1962, vol. 13, p. 251—266.
- 2.55. Livingston M. J., Blewett J. P. Particle accelerators. N. Y.: McGraw-Hill Co Inc., 1962.
- 2.56. Rainoff T. J. Wave — life fluctuations of creative productivity in the development of West — European physics in the XVIIth and XIX centuries.— Intern. Rev. Devoted to the Hist. Sci. and Civilization (ISIS), 1929, vol. 12, N 38, p. 287—397.
- 2.57. Roeckelein J. E. Eponymy in psychology.— Amer. Psychol., 1972, vol. 27, N 7, p. 657—659.
- 2.58. Rudd E. The effect of alphabetical order of author listing on the careers of scientists.— Social Stud. Sci., 1977, vol. 7, p. 268—269.
- 2.59. Sargent J. R. The distribution of scientific manpower.— In: Science and technology in economic growth. London; Basingstoke, 1973, p. 360—374.
- 2.60. Scientific and technical translating and other aspects of the language problem. Geneva: UNESCO, 1958.
- 2.61. Simonton D. K. Invention and discovery among the sciences: a P-technique factor analysis.— J. Vocat. Behav., 1975, vol. 7, N 3, p. 275—281.
- 2.62. Simonton D. K. Interdisciplinary and military determinants of scientific productivity: a crosslagged correlation analysis.— J. Vocat. Behav., 1976, vol. 9, N 1, p. 53—62.
- 2.63. Sorokin P. Social and cultural dynamics. N. Y., 1937.
- 2.64. Sturgis R. B., Clemente F. The productivity of graduates of 50 sociology departments.— Amer. Sociol., 1973, vol. 8, N 4, p. 169—180.
- 2.65. Subramanyan K., Schaffer C. J. Effectiveness of «Letters J».— New Libr. World, 1974, vol. 75, N 894, p. 258—259.
- 2.66. Thiemann H. Changing dynamics in research and development.— Science, 1970, vol. 168, N 3938, p. 1427—1431.
- 2.67. Urquhart D. J. Use of scientific periodicals.— In: Intern. Conf. on Sci. Inform., Wash., 1958, p. 277—290.
- 2.68. Vlachy J. Planning in physics.— Czechosl. J. Phys., 1968, vol. B18, N 11, s. 1479—1501.
- 2.69. Vlachy J. Physical institutes of the Czechoslovak Academy of Science in 1953—1967.— Czechosl. J. Phys., 1969, vol. B19, N 1, s. 138—148.
- 2.70. Vlachy J. Physics journal in retrospect and comparison.— Czechosl. J. Phys., 1970, vol. B20, N 4, p. 501—526.
- 2.71. Woodward W. R. Scientific genius and loss of a parent.— Sci. Stud., 1974, vol. 4, N 3, p. 265—277.
- 2.72. Zuckerman H. Nobel laureates in science: patterns of productivity. Collaboration and authorship.— Amer. Sociol. Rev., 1967, vol. 32, p. 391—403.

- 2.73. Zuckerman H. The sociology of the Nobel prizes.— Sci Amer., 1967 vol. 217, p. 25—33.
- 2.74. Zuckerman H. Patterns name ordering among authors of scientific papers: A study of social symbolism and its ambiguity.— Amer. J. Sociol., 1968, vol. 74, N 3, p. 276—291.
- 2.75. Zuckerman H. Stratification in american science.— Sociol. Inquiry, 1970, vol. 40, p. 235—257.
- 2.76. Zuckerman H. Scientific elite: Nobel laureates in the United States. N. Y.: Free press, 1977, XX.

3. Наукометрия: метод подсчета числа публикаций

- 3.1 Алексеев А. И. О некоторых тенденциях в развитии общественных наук: (Библиографические и наукометрические наблюдения).— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1971, вып. 4, с. 77—83.
- 3.2 Антеклиан Г. И. Исследование и разработка отраслевой системы информационного обеспечения и управления НИР в угольной промышленности: Автореф. дис. ... канд. техн. наук. М.: ЦИТИ угольной пром-сти, 1969.
- 3.3 Берг М. А. О возможности оценки тем научно-исследовательских работ по информационным критериям.— В кн.: Теория и практика научно-технической информации. М.: МДНТИ, 1975, с. 144—150.
- 3.4 Бурбуля Ю. Т., Коварская Б. П. Структура информационного потока по механике: (колебания упругих тел).— НТИ. Сер. 1, 1975, № 11, с. 21—24.
- 3.5 Васильковский А. А., Дзиековская И. В. Динамика роста числа публикаций по физике и технике тонких пленок и прогнозирование дальнейшего развития этой области знаний.— В кн.: Науковедение. Прогнозирование. Информатика. Киев: Наук. думка, 1970, с. 259—268.
- 3.6 Горькова В. И. Слежение за тенденциями развития науки по данным научно-технической информации.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1971, вып. 4, с. 93—102.
- 3.7 Гражданников Е. Д., Щербатов А. И. Проблема прогнозирования индивидуального научного творчества. М.: ССА, 1970, 13 с.
- 3.8 Громин В. И. Оценка развития экспериментальной тектоники по публикациям.— Геотектоника, 1974, № 1, с. 124—127.
- 3.9 Дзиековская И. В. Структура и динамика научных исследований в области сверхпроводящих тонких пленок.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1974, вып. 11, с. 72—80.
- 3.10 Заржицкая Л. В., Спасский В. И. Количественные методы исследования развития отдельных разделов физики.— В кн.: История и методология естественных наук. М.: Изд-во МГУ, 1971, вып. 10, с. 10—14.
- 3.11 Зверев В. М. Из опыта наукометрических наблюдений в области истории социологии в России (1860-е гг.—1917 г.).— В кн.: Методологические и методические проблемы контент анализа. М.; Л.: ИСИ АН СССР, 1973, вып. 2, с. 122—124.
- 3.12 Ильиницкий П. С., Михайличенко В. Е. О роли вузов в подготовке научных кадров.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1974, вып. 11, с. 8—13.
- 3.13 Карпов М. М., Потемкин А. В., Петров М. К. Национальный и мировой стандарты распределения научной деятельности.— В кн.: Всесоюз. совещ. по количеств. методам в социологии. М., 1967, с. 25—26.
- 3.14 Князева М. К. Выявление источников, составляющих «ядро» изданий по тематике выборки.— НТИ. Сер. 1, 1977, № 7, с. 25—29.
- 3.15 Коншин Г. И., Ляцкий В. В. Метод статистического многофакторного

- анализа документальной геологической информации как средство выявления состояния и тенденций развития геологических знаний.— НТИ. Сер. 1, 1974, № 6, с. 14—17.
- 3.16 Коньшова В. А., Смирнова М. Г. Анализ вторичных источников информации по нутрициологии (науке о питании).— НТИ. Сер. 1, 1973, № 12, с. 20—23.
- 3.17 Кордо А. Д., Тырышкин В. С. К вопросу о статистической оценке тенденций развития радиотехники и ее отдельных направлений.— В кн.: Материалы по науковедению. Киев: СОПС УССР, 1970, вып. 5, с. 43—48.
- 3.18 Крашенинникова И. Л. Динамика развития реферативного журнала «Механика» (общая механика, гидромеханика).— НТИ. Сер. 1, 1976, № 1, с. 24—30.
- 3.19 Куренкова М. Г. Анализ информационных потоков по РЖ ВИНТИ «Трактора и сельскохозяйственные машины и орудия».— НТИ. Сер. 1, 1976, № 9, с. 27—29.
- 3.20 Лабинаускайте В., Канцлерис А. Теоретическая моноинформация: (Анализ лит. потоков).— Библиотеку дарбас, 1974, № 10, с. 15—17.
- 3.21 Лойферман П. Г., Волков В. И. Связь между ранговым распределением документальных источников в нескольких годичных выборках.— НТИ. Сер. 2, 1977, № 2, с. 19—21.
- 3.22 Лукьянов В. В., Симонов Е. Ф., Коробков В. И. Статистический анализ информационных потоков в радиохимии. Методы приготовления препаратов для измерения радиоактивности.— В кн.: Наукометрические исследования в химии. М.: Изд-во МГУ, 1974, с. 113—121.
- 3.23 Мицевич А. Т. Исследования структуры научно-технической информации по машиностроению.— НТИ. Сер. 2, 1975, № 5, с. 3—16.
- 3.24 Мицевич А. Т. Автоматизация процессов проектирования судов и информационное обеспечение.— В кн.: Судостроение. М.: Ин-т науч. информ. АН СССР, 1975, т. 6.
- 3.25 Муравьева Т. П., Колесникова Т. П. Распределение публикаций по металлургии и ее отраслям в периодических и продолжающихся изданиях (по данным РЖ ВИНТИ «Металлургия»).— НТИ. Сер. 1, 1975, № 5, с. 8—16.
- 3.26 Муравьева Т. П., Колесникова Т. П., Бабаева А. П. Некоторые результаты исследования информационных потоков по металлургии.— НТИ. Сер. 1, 1974, № 12, с. 20—26.
- 3.27 Осетров В. Д. Распределение творческой активности в течение года.— В кн.: Материалы V симпози. по науковед. и науч.-техн. прогнозированию. Киев, 1974, ч. 3, с. 64—66.
- 3.28 Петров М. К. Журнал «Вопросы философии» в цифрах: (Колич. характеристики за 1967—1970 гг.).— Вопр. философии, 1972, № 3, с. 140—148.
- 3.29 Раутиан С. Г. Тенденции развития современной оптики.— Природа, 1978, № 11.
- 3.30 Сагалович Н. М. К вопросу о тенденциях роста числа научных публикаций.— НТИ, 1963, № 10, с. 3—4.
- 3.31 Самсонов Г. В., Златокова Т. П., Черногоренко В. Б. Наукометрический анализ и прогноз информации по оксидам.— В кн.: Науковедение и информатика. Киев: Наук. думка, 1975, вып. 13, с. 132—136.
- 3.32 Фридман Э. П. Некоторые науковедческие характеристики мировых исследований (публикаций) на приматах за 1968 г.— Вопр. антропологии, 1972, № 40, с. 179—188.
- 3.33 Фридман Э. П. Дальнейший анализ мировых исследований на приматах: Публ. 1969 г.— Вопр. антропологии, 1973, № 45, с. 191—194.
- 3.34 Фридман Э. П. Обзор исследований на приматах по публикациям 1970 г.— Вопр. антропологии, 1974, № 46, с. 146—153.

335. Чайка Н. А. Языковой барьер в научной медицинской литературе.— ИТН. Сер. 1, 1977, № 3, с. 17—20.
336. Шершова Н. В. Формирование и анализ информационного массива по гидроагрегатостроению для определения перспектив развития технических устройств (на примере электрогидравлических усилителей мощности): Автореф. дис. ... канд. техн. наук. М.: ВИНТИ, 1971.
337. Allison P. D., Price D. J. de S., Griffith B. C., Moravcsik M. J., Stewart J. A. Lotka's law: a problem in the its interpretation and application.— Sci. Stud., 1976, vol. 6, p. 269—276.
338. Allison P. D., Stewart J. A. Productivity differences among scientists: evidence for accumulative advantage.— Amer. Sociol. Rev., 1974, vol. 39, p. 596—606.
339. Anthony L. J., East H., Slater M. J. The growth of the literature of physics.— Repts Progr. Phys., 1969, vol. 32, p. 713.
340. Baker D. B. Growth of chemical literature. Past, present and future.— Chem. and Eng. News, 1961, vol. 39, N 29, p. 78—81.
341. Ben-David J. Fundamental research and the universities. Organization for economic co-operation and development. P., 1968.
342. Bradford S. C. Sources of information on specific subjects.— Engineering, 1934, vol. 26, Jan.
343. Bradford S. C. Documentation. L.: Crosby Lockwood and Son LTD, 1948.
344. Brookes B. C. The derivation and application of the Bradford — Zipf distribution.— J. Doc., 1968, vol. 24, p. 247—265.
345. Brookes B. C. Bradford's law and the bibliography of science.— Nature, 1969, vol. 224, N 5223, p. 953—956.
346. Brookes B. C. Theory of the Bradford law.— J. Doc., 1977, vol. 33, p. 180—209.
347. Carter A. M. An assisment of quality in graduate education. Wash.: D. C. Amer. Council of Education, 1966.
348. Chaney F. B. A cross-cultural study of industrial research performance.— J. Appl. Psychol., 1966, vol. 50 N 3, p. 206—210.
349. Coile R. C. Lotka's frequency distribution of scientific productivity.— J. Amer. Soc. Inform. Sci., 1977, vol. 28, p. 366—370.
350. Cole P. F. A new look at reference scattering.— J. Doc., 1962, vol. 18, N 2, p. 58—64.
351. Crane D. Scientists at major and minor universities: a study of productivity and recognition.— Amer. Sociol. Rev., 1965, vol. 30, p. 699—714.
352. Dennis W. Bibliographies of eminent scientists.— Scientific Monthly, 1954, vol. 79, Sept., p. 180—183.
353. Dennis W. Age and productivity among scientists.— Science, 1956, vol. 123, p. 724—725.
354. Dresden A. A report on the scientific work of the Chicago section 1897—1922.— Bull. Amer. Math. Soc., 1922, vol. 28, p. 303—307.
355. Eiduson B. T. Productivity rate in research scientists.— Amer. Sci., 1966, vol. 54, N 1, p. 57—63.
356. Ferriman A. Women academics «publish less than men».— Impact Sci. Soc., 1975, vol. 25, N 2, p. 153—154.
357. Fischer R. B. Trends in analytical chemistry.— Anal. Chem., 1965, vol. 37, N 13, p. 27a—34a.
358. Goffman W., Warren K. S. Dispersion of paper among journals based on a mathematical analysis of the diverse medical literatures.— Nature, 1969, vol. 221, N 5187, p. 1205—1207.
359. Grasberg A. G. Merit rating and productivity in an industrial research laboratory: a case study.— IRE Trans. Eng. Manag., 1959, vol. EM-6, N 1, p. 31—37.

360. Groos O. V. Bradford's law and the Keenan-Atherton data.— Amer. Doc., 1967, vol. 19, N 1, p. 46.
361. Guyer L., Fidell L. Publications of men and women psychologists. Do women publish less? — Amer. Psychol., 1973, vol. 28, N 2, p. 157—160.
362. Hubert J. J. Bibliometric models for journal productivity.— Social Indicators Res., 1977, vol. 4, p. 441—473.
363. Hurt C. D. A correlation study of the journal article productivity of environmental scientist.— Inform. Res. and Manag., 1977, vol. 13, p. 305—309.
364. Inhaber H. Where scientists publish? — Social Stud. Sci., 1977, vol. 7, p. 383—394.
365. Kempbell F. The theory of the national and international bibliography. L., 1896, p. 86—97.
366. Kogan N. Creativity and sex differences.— J. Creat. Behav., 1974, vol. 8, N 1, p. 1—14.
367. Leavens D. H. Communication from Dikson H. Leavens.— Econometrics, 1959, vol. 21, p. 630.
368. Lehman H. C. Age and achievement. Princeton (N. J.): Princeton Univ. press, 1953.
369. Lehman H. C. Mens creative production rate at different ages and in different countries.— Scientific Monthly, 1954, vol. 78, N 5, p. 321—326.
370. Lehman H. C. The chemists most creative years.— Science, 1958, vol. 127, N 3308, p. 1213—1222.
371. Lehman H. C. The age decrement in outstanding scientific creativity.— Amer. Psychol., 1960, vol. 15, p. 128—134.
372. Leimkuhler F. F. The Bradford distribution.— J. Doc., 1967, vol. 23, p. 197—207.
373. Leimkuhler F. F. Operational analysis of library systems.— Inform. Processing and Manag., 1977, vol. 13, p. 79—93.
374. Lotka A. J. The frequency distribution of scientific productivity.— J. Washington Acad. Sci., 1926, vol. 16, N 12, p. 317—323.
375. Mantell L. On law of special abilities and the production of scientific literature.— Amer. Doc., 1966, vol. 17, N 1, p. 8—16.
376. Maris G. J. Some academic influence upon publication productivity.— Social Forces, 1951, vol. 29, p. 267—272.
377. Meltzer L. Scientific productivity in organizational settings.— J. Social Iss., 1956, vol. 12, N 2, p. 32—40.
378. Mittermerr R., Knorr K. D. Scientific productivity and accumulative advantage: a thesis reassessed in the light of informational data.— R and D Manag., 1979, vol. 9, p. 235—239.
379. Murphy L. J. Lotka's law in humanities.— J. Amer. Soc. Inform. Sci., 1973, vol. 24, N 6, p. 461—462.
380. Nagpal P. S., Bhatva J. R., Rahman A., Bhargava R. K. Trends of research, series 1 and 2, CSJR, New Delhi, 1956.
381. O'Neill E. T. Limitations of the Bradford distribution.— Proc. Amer. Soc. Inform. Sci., 1973, vol. 10, p. 177—178.
382. Piganol P. The competitive pressure: the effect of «publish or perish».— Impact Sci. Soc., 1971, vol. 21, N 2, p. 163—171.
383. Pope A. Bradford's law and the periodical literature of information science.— J. Amer. Soc. Inform. Sci., 1975, vol. 26, N 4, p. 207—212.
384. Praunlich P., Kroll M. Bradford's distribution: a new formulation.— J. Amer. Soc. Inform. Sci., 1978, vol. 29, p. 51—55.
385. Price D. J. de S. Research of Research.— In: Journeys in Science: Small Steps-Great Strided /Ed. D. L. Arn. Rep. 12th. Air Force Office of Scientific Research Science Seminar. Mexico: Univ. Mexico press, 1967, p. 1—21.

- 3.86. Radhakrishnan T., Kernizan R. Lotka's law and computer science literature.— J. Amer. Soc. Inform. Sci., 1979, vol. 30, p. 51—64.
- 3.87. Ravichandra Rao I. K. The distribution of scientific productivity and social change.— J. Amer. Soc. Inform. Sci., 1980, vol. 31, p. 111—122.
- 3.88. Rice E. W. Research clinical chemists to United States — a statistical analysis— Clin. Chem., 1970, vol. 16, N 11, p. 877—881.
- 3.89. Robertson S. E. The cost of Bradford.— В кн.: Об эффективности научно-информационной деятельности. М.: Междунар. федерация по документации, 1976, с. 135—239.
- 3.90. Roche M. Productivity of Latin American science.— Interciencia, 1978, vol. 3, p. 134—135.
- 3.91. Roe A. Patterns in productivity of scientists.— Science, 1972, vol. 176, N 4037, p. 940—941.
- 3.92. Shaw B. T. The use of quality and quantity of publication as criteria for evaluating scientists. Wash.: ARS, US Department of Agriculture, 1967.
- 3.93. Tondl L.— Teor. a metoda, 1969, vol. 1, N 2, s. 9—42.
- 3.94. Tondl L.— Věda jako faktor plánování.— Vesmir, 1970, vol. 49, N 2, s. 35—37.
- 3.95. Trimble R. F. The journals of inorganic chemistry.— J. Chem. Doc., 1963, vol. 3, p. 79—81.
- 3.96. Van Zelst R. H., Kerr W. A. Personality self-assessment of scientific and technical personal.— J. Appl. Psychol., 1954, vol. 38, N 3, p. 145—147.
- 3.97. Vickery B. C. Bradford's law of scattering.— J. Doc., 1948, vol. 4, N 3, p. 198—203.
- 3.98. Vickery B. C. Statistics of scientific and technical articles.— J. Doc., 1968, vol. 24, N 3, p. 192—195.
- 3.99. Vincent H. P., Mirakhor A. Relationship between productivity, satisfaction, ability, age and salary in a military R and D organization.— IEEE Trans. Eng. Manag., 1972, vol. EM-19, N 2, p. 45—53.
- 3.100. Vlachý J. Remarks on the productive age.— Teor. a metoda, 1970, vol. 2, N 3, s. 121—150.
- 3.101. Vlachý J. Variable factors in scientific communities (observations of Lotka's law).— Teor. a metoda, 1972, vol. 4, N 1, s. 91—120.
- 3.102. Voos H. Lotka and information science.— J. Amer. Soc. Inform. Sci., 1974, vol. 25, N 4, p. 270—272.
- 3.103. Webb E. C. Communications in biochemistry.— Nature, 1970, vol. 225, N 5228, p. 132—135.
- 3.104. Wilkinson E. A. The ambiguity of Bradford's law.— J. Doc., 1972, vol. 28, p. 122—128.
- 3.105. Williams E. B. The numbers of publications written by biologists.— Ann. Eugenics, 1944, vol. 12, p. 143—144.
- 3.106. Zusne L. Age and achievement in psychology. The harmonic mean as a model.— Amer. Psychol., 1976, vol. 31, N 11, p. 805—807.

4. Наукометрия: метод «цитат-индекс»

- 4.1. Альшуллер С. В. П. И. Вальден и язык научных ссылок.— Природа, 1969, № 8, с. 119.
- 4.2. Атанасиу П. Цитаты как инструмент оценки первичных документов.— В кн.: Междунар. конгр. по науч. информ., Москва, 16—18 сент. 1968 г. М.: ВИНТИ, 1969.
- 4.3. Баранова З. Б., Налимов В. В. и др. Изучение научных журналов как каналов связи. Оценка вклада, вносимого отдельными странами в мировой научный информационный поток.— НТИ. Сер. 2, 1967, № 12, с. 3—11.

- 4.4. Брукс Б. Старение научной литературы.— В кн.: Проблемы информатики. М.: ВИНТИ, 1973, с. 74—102.
- 4.5. Васильев Р. Ф. О критерии цитируемости в науковедении.— В кн.: История науки и науковедение. Рига: Зинатне, 1975, с. 152—153.
- 4.6. Ваттер Э. К.-Ф. О структуре научных ссылок.— НТИ. Сер. 2, 1975, № 5, с. 12—20.
- 4.7. Гиляревский Р. С., Мульченко З. М., Теремин А. Т., Черный А. И. Опыт изучения «Science Citation Index».— В кн.: Прикладная документалистика. М.: Наука, 1968, с. 32—53.
- 4.8. Добров Г. М. Информационная основа истории научно-технического развития.— В кн.: Прикладная документалистика. М.: Наука, 1968, с. 5—31.
- 4.9. Зябрев В. А., Скляр Т. В. Исследование степени использования научной информации в области электроники с помощью анализа библиографических ссылок.— НТИ. Сер. 1, 1974, № 1, с. 19—22.
- 4.10. Киселев О. М. Использование указателей научных ссылок. Л.: ЛДНТИ, 1970.
- 4.11. Клементьев А. Ф. О некоторых характеристиках цитируемости публикаций по математике (по материалам отечественных изданий).— НТИ. Сер. 1, 1975, № 7, с. 33—34.
- 4.12. Коренной А. А. Экспериментальное построение и математический анализ сетей научного влияния.— В кн.: Науковедение. Прогнозирование. Информатика. Киев: Наук. думка, 1970, с. 109—115.
- 4.13. Коренной А. А. Методы и опыт исследования информационных связей в научных системах.— В кн.: Материалы по науковедению. Киев: СОПС УССР, 1971, вып. 16, с. 3—70.
- 4.14. Максимов В. В. Социометрия «цитат-поведения» ученого.— В кн.: Проблемы деятельности ученого и научных коллективов. Л.: Наука. Ленингр. отд-ние, 1973, вып. 5, с. 327—334.
- 4.15. Маркусова В. А. О некоторых тенденциях распределения и старения публикаций в области теоретической электрохимии.— НТИ. Сер. 1, 1971, № 5, с. 31—37.
- 4.16. Маркусова В. А. Сравнительная характеристика цитируемости научно-технических публикаций (по материалам отечественных изданий).— НТИ. Сер. 1, 1973, № 1, с. 27—31.
- 4.17. Маршакова И. В. Система связей между документами, построенная на основе ссылок (по указателю «Science Citation Index»).— НТИ. Сер. 2, 1973, № 6.
- 4.18. Маршакова И. В. Проспективная связь в системе научных публикаций.— В кн.: Системные исследования: Ежегодник, 1976. М.: Наука, 1977, с. 38—54.
- 4.19. Маршакова И. В. Классификация документов и авторов методом проспективной библиографической связи.— В кн.: Проблемы деятельности ученого и научных коллективов. М.; Л.: Наука, 1977, вып. 6, с. 203—209.
- 4.20. Мехтиев Д. М., Бабеев А. М., Рустамов А. М. Указатель цитированной литературы по нефтяной науке и технике (УЦЛ—нефть). Баку: АЗНИИНТИ, 1971.
- 4.21. Мехтиев Д. М., Аракелов Р. К., Меллион С. П. и др. Использование указателя цитированной литературы для анализа состояния и тенденций развития информатики.— НТИ. Сер. 2, 1977, № 2, с. 9—18.
- 4.22. Мульченко З. М. Исследования информационных потоков в науке на основе библиографических ссылок: Автореф. дис. ... канд. пед. наук. М.: МГИК, 1970.
- 4.23. Ориент И. М. Статистическое изучение цитируемости работ по аналитической химии.— Завод. лаб., 1967, т. 33, № 11, с. 1383—1385.

- 4.24. Прайс Д. Система научных публикаций.— УФН, 1966, т. 90, № 2, с. 347—359.
- 4.25. Прайс Д. Квоты цитирования в точных и естественных науках, технике и не-науке.— Вopr. философии, 1971, № 3, с. 153.
- 4.26. Преображенская Г. Б. Смысловый анализ ссылок в статьях по металлургии.— НТИ. Сер. 2, 1969, № 10, с. 10—11.
- 4.27. Сорокина Т. В., Гражданников Е. Д. Тематический библиографический поиск с помощью указателя цитированной литературы.— В кн.: Наукoведение и информатика. Киев: Наук. думка, 1975, вып. 13, с. 64—68.
- 4.28. Тимошин Ю. В., Семиховский Г. С. Особенности системы цитирования в литературе по геофизической разведочной технике.— НТИ. Сер. 1, 1972, № 4, с. 24—26.
- 4.29. Указатель цитированной литературы по информатике (УЦЛ-информатика): В 3-х ч. Баку: АЗНИНТИ, 1976.
- 4.30. Фролова М. Г. Сравнительный анализ материала в советских и зарубежных научных публикациях по контактной электросварке.— НТИ. Сер. 1, 1972, № 8, с. 20—22.
- 4.31. Футько З. А. Изучение информационного потока по вопросам строительства путем составления указателя научных ссылок. Минск: БелНИНТИ и техн.-эконом. исслед. 1968.
- 4.32. Хайтун С. Д. Что такое «цитат-индекс»? — Природа, 1980, № 3, с. 40—51.
- 4.33. Adair W. C. Citation index for scientific literature? — Amer. Doc., 1955, vol. 6, N 1, p. 31—32.
- 4.34. Albert R. S. Towards a behavioral definition of genius.— Amer. Psychol., 1975, vol. 30, N 2, p. 140—151.
- 4.35. Avramescu A. Möglichkeiten einer Wertordnung an Wissenschaftliche Veröffentlichungen.— In: XI Intern. wissenschaftliches Kolloquium 17—18 Okt., 1966. Berlin: Technische hochschule Limenan, 1966.
- 4.36. Avramescu A. Science citation distribution and obsolescence.— Stud. și cerc. doc., 1973, vol. 15, N 4, p. 345—356.
- 4.37. Baughman J. C. A structural analysis of the literature of sociology.— Libr. Quart., 1974, vol. 44, N 4, p. 293—308.
- 4.38. Bayer A. E., Folger J. Some correlates of a citation measure of productivity in science.— Sociol. Educ., 1966, vol. 39, N 4, p. 381—390.
- 4.39. Ben-Ami L. Improvement of the selectivity of citation index to science literature through inclusion of citation relationship indicators.— Amer. Doc., 1965, vol. 16, N 2, p. 81—90.
- 4.40. Bernal J. D. Science citation index.— Sci. Progr., 1965, vol. 53, N 211, p. 455—459.
- 4.41. Bibliographic parisienne. P.; Desnos, 1771—1772.
- 4.42. Borenus G., Schwarz S. Remarks on the use of citation data in predictive models of scientific activity.— Inform. Storage and Retrieval, 1972, vol. 8, N 4, p. 171—75.
- 4.43. Bourne C. P. Some user requirements stated quantitatively in terms of the 90 percent library.— In: Electric information handling / Ed. A. Kent. Wash.: D. C., 1965, p. 93—100.
- 4.44. Broadus R. N. The literature of the social sciences: a survey of citation studies.— Intern. Social. Sci. J., 1971, vol. 23, N 2, p. 236—243.
- 4.45. Brodman E. Choosing physiology journals.— Med. libr. assoc. Bull., 1944, vol. 32, p. 479—483.
- 4.46. Brookes B. C. The growth, utility, and obsolescence of scientific periodical literature.— J. Doc., 1970, vol. 26, N 4, p. 283—294.
- 4.47. Brookes B. C. Obsolescence of special library periodicals: sampling errors and «utility contours».— J. Amer. Soc. Inform. Sci., 1970, vol. 21, N 5, p. 320—329.
- 4.48. Brookes B. C. Numerical methods bibliographic analysis.— Libr. Trends, 1973, vol. 22, N 1, p. 18—43.
- 4.49. Buckland M. K. Are obsolescence and scattering related?— J. Doc., 1972, vol. 28, N 3, p. 242—245.
- 4.50. Burton R. E., Kebler R. W. The «half-life» of some scientific and technical literature.— Amer. Doc., 1960, vol. 11, N 1, p. 18—22.
- 4.51. Burton R. E. Citations in American engineering journals I. Chemical engineering.— Amer. Doc., 1959, vol. 10, p. 70—73; II. Mechanical engineering.— Ibid., p. 135—137; III. Metallurgical engineering.— Ibid., p. 209—213.
- 4.52. Chubin D. On the use of the «Science citation index» in sociology.— Amer. Sociol., 1973, vol. 8, N 4, p. 187—191.
- 4.53. Chubin D. E., Moitra S. D. Content analysis of references: adjunct or alternative to citation counting? — Sci. Stud., 1975, vol. 5, N 4, p. 423—441.
- 4.54. Clark K. E. American psychologist: a survey of a growing profession: Wash.: D. C., Amer. Psychol. Assoc., 1967.
- 4.55. Cleverdon C. Citation idiosyncrasies.— Nature, 1970, vol. 228, N 5278, p. 1356.
- 4.56. Coile R. C. Information sources for electrical and electronic engineers.— IEEE Trans. Eng., Writting and Speech, 1949, vol. EWS-12, Oct., p. 71—78.
- 4.57. Coile R. C. Periodical literature for electrical engineers.— J. Doc., 1952, vol. 8, p. 209—226.
- 4.58. Cole J. R., Cole S. Measurings the quality of sociological research problems in the use of the «Science citation index».— Amer. Sociol., 1971, vol. 6, Febr., p. 23—29.
- 4.59. Cole J. R., Cole S. The Ortega hypothesis. Citation analysis suggests that only a few scientists contribute to scientific progress.— Science, 1972, vol. 178, N 4059, p. 368—375.
- 4.60. Cole P. F. Journal usage versus age of journals.— J. Doc., 1963, vol. 19, N 1, p. 1—11.
- 4.61. Cole S. Professional standing and the recognition of scientific discoveries.— Amer. J. Sociol., 1970, vol. 76, p. 286—306.
- 4.62. Cole S., Cole J. R. Scientific output recognition: A study in the operation of the reward system in science.— Amer. Sociol. Rev., 1967, vol. 32, p. 377—390.
- 4.63. Cole S., Cole J. R. Visibility and the structural bases of awareness of scientific research.— Amer. Sociol. Rev., 1968, vol. 33, N 3, p. 397—413.
- 4.64. Craig Jr. J. E. C. Characteristics of use of geology literature.— Coll. and Res. Libr. 1969, vol. 30, N 3, p. 230—236.
- 4.65. Croom D. L. Danger in the use of the science citation index.— Nature, 1970, vol. 227, p. 1173.
- 4.66. Crowley C. J. The distribution of citations to scientific papers: A model. Chicago, 1975, Apr. (Unpublished).— Инт. no: Price D. J. de S. A general theory of bibliometric and other cumulative advantage.— J. Amer. Soc. Inform. Sci., 1976, vol. 27, p. 292—306.
- 4.67. Dennis W. The age decrement in outstanding scientific contributions: fact or artifact? — Amer. Psychol., 1950, vol. 13, N 8, p. 457—460.
- 4.68. Dhawan S. M., Yedav R. S. A study on most frequently cited journals by Indian scientists in physics.— Ann. Libr. Sci. and Doc., 1972, vol. 19, N 4, p. 185—193.
- 4.69. Dieks D., Chang H. Difference in impact of scientific publication: Some indices derived from a citation analysis.— Social Stud. Sci., 1976, vol. 6, p. 427—467.
- 4.70. Early P., Vickery B. Subject relations in science / technology literature.— Aslib Proceedings, 1969, vol. 21, № 6, p. 237—243.

- 4.71. *Early P., Vickery B.* Social science literature in the UK as indicated by citations.—J. Doc., 1969, vol. 25, N 2, p. 123—141.
- 4.72. *East H., Weyman A.* A study in the source literature of plasma physics.—Aslib Proc., 1969, vol. 21, N 1, p. 160—171.
- 4.73. *Fussler H. H.* Characteristics of the research literature by chemists and physicists in the United States.—Libr. Quart., 1949, vol. 19, N 1, p. 19—35.
- 4.74. *Fussler H. H.* Characteristics of the research literature by chemists and physicists in the United States.—Libr. Quart., 1949, vol. 19, N 2, p. 119—145.
- 4.75. *Garfield E.* Citation indexes for science: a new dimension in documentation through association of ideas.—Science, 1955, vol. 122, N 3159, p. 108—111.
- 4.76. *Garfield E.* Citation indexing for studying science.—Nature, 1970, vol. 227, N 5259, p. 669—671.
- 4.77. *Garfield E.* Citation analysis as a tool in journal evaluating.—Science, 1972, vol. 178, N 4060, p. 471—479.
- 4.78. *Garfield E.* Essays of an information scientists. Philadelphia: ISI press, 1977.
- 4.79. *Garfield E.* The 250 most-cited primary authors, 1961—1975.—Curr. Contents, 1977, N 49, p. 5—15; N 50, p. 5—15; N 51, p. 5—20.
- 4.80. *Garfield E.* The 300 most-cited primary authors, 1961—1976, including co-authors at last.—Curr. Contents, 1978, N 28, p. 5—17; N 35, p. 5—30; N 46, p. 5—12.
- 4.81. *Garfield E.* Bradford's law and related statistical patterns.—Curr. Contents, 1980, N 19, p. 5—12.
- 4.82. *Garfield E., Cawkell A. E.* Location of milestone papers through citation networks.—J. Libr. Hist., 1970, vol. 5, N 2, p. 184—188.
- 4.83. *Garfield E., Sher I. H.* New factors in the evaluation of scientific literature through citation indexing.—Amer. Doc., 1963, vol. 14, N 3, p. 195.
- 4.84. *Garfield E., Sher I. H., Torpic R. J.* The use of citation data in writing the history of science. Philadelphia: ISI press, 1964.
- 4.85. *Glass B.* The ethical basis of science.—Science, 1965, vol. 150, N 3701, p. 1254—1261.
- 4.86. *Gomperts M. C.* The law of constant citation for scientific literature.—J. Doc., 1968, vol. 24, N 2, p. 113—117.
- 4.87. *Griffith B. C., Small H. G., Stonehill D. C.* The structure of scientific literatures. II: Toward a macro- and microstructure of science.—Sci. Stud., 1974, vol. 4, N 4, p. 339—365.
- 4.88. *Groos O. V.* Citation characteristics of astronomical literature.—J. Doc., 1969, vol. 25, N 4, p. 344—347.
- 4.89. *Gross P. L. K., Gross E. M.* College libraries and chemical education.—Science, 1927, vol. 66, Oct. 28, p. 385—389.
- 4.90. *Gross P. L. K., Woodford A. D.* Serial literature uses by american geologist.—Science, 1931, vol. 73, p. 660—664.
- 4.91. *Guttsman W. L.* The literature of the social sciences and provision for research in them.—J. Doc., 1966, vol. 22, p. 186—194.
- 4.92. *Hagstrom W. O.* Input, output, and the prestige of university science departments.—Sociol. Educ., 1971, vol. 44, N 4, p. 375—397.
- 4.93. *Half-life of physics articles.*—Amer. Doc., 1960, vol. 11, N 2, p. 199.
- 4.94. *Kaplan N.* The norms of citation behaviour: prolegomena to the footnote.—Amer. Doc., 1965, vol. 16, N 3, p. 179—183.
- 4.95. *Kessler M. M.* An experimental study of bibliographic coupling between technical papers.—IEEE Trans. Inform. Theory, 1963, vol. 9, Jan., p. 50—51.

- 4.96. *Kessler M. M.* Bibliometric coupling between scientific papers.—Amer. Doc., 1963, vol. 14, N 1, p. 10—21.
- 4.97. *Kessler M. M.* Comparison of the results of bibliographic coupling and analytic indexing.—Amer. Doc., 1965, vol. 16, N 3, p. 223—233.
- 4.98. *Kofnovec L.* Das Veralten Wissenschaftlicher-technischer Information.—Dokumentation, 1962, vol. 9, N 2, S. 42—49.
- 4.99. *Kofnovec L.* Citations in Soviet metallurgical literature. Prague, 1965. Paper given at the F.I.D. Congr. Held in Washington, Oct., 1965.
- 4.100. *Kot S. M.* Prognozy rozwoju dyscypliny naukowej (na przykladne fizyki).—Przegl. statyst., 1974, vol. 21, N 3, s. 489—510.
- 4.101. *Krauze T. K., Hillinger C.* Citation, references and the growth of scientific literature: a model of dynamic interaction.—J. Amer. Soc. Inform. Sci., 1971, vol. 22, N 5, p. 333—336.
- 4.102. *Lightfield E. T.* Output and recognition of sociologists.—Amer. Sociol., 1971, vol. 6, p. 128—133.
- 4.103. *Lin N., Nelson C. E.* Bibliographic reference patterns in core sociological journals, 1965—1966.—Amer. Sociol., 1969, vol. 4, p. 47—50.
- 4.104. *Line M. B.* The half-life of periodical literature: Apparent real obsolescence.—J. Doc., 1970, vol. 26, N 1, p. 46—54.
- 4.105. *Line M. B., Sandison A.* «Obsolescence» and changes in the use literature with time.—J. Doc., 1974, vol. 30, N 3, p. 283.
- 4.106. *Loosjes Th. P.* On documentation of scientific literature. L.: Batherworths, 1967, Cop. 4, p. 26—43; cop. 9, p. 67—82.
- 4.107. *Louttit C. M.* The use of foreign language by psychologists, chemists, and physicists.—Amer. J. Psychol., 1957, vol. 70, June, p. 314—316.
- 4.108. *MacRoe Jr. D.* Growth and decay curves in scientific citations.—Amer. Sociol. Rev., 1969, vol. 34, N 5, p. 631—635.
- 4.109. *Margolis J.* Citation indexing and evaluation of scientific papers.—Science, 1967, vol. 155, N 3767, p. 1213—1219.
- 4.110. *Martin J.* Examination of citation indexes.—Aslib Proc., 1965, vol. 17, N 6, p. 184—196.
- 4.111. *Martin M. W.* The measurement of value of scientific information.—In: Operational research in R and D. N. Y.: L., 1963, p. 97—123.
- 4.112. *Martino J. P.* Citation indexing for research and development management.—IEEE Trans. Eng. Manag., 1971, vol. EM-18, N 4, p. 146—151.
- 4.113. *Meadows A. J.* The citation characteristics of astronomical research literature.—J. Doc., 1967, vol. 23, N 1, p. 28—33.
- 4.114. *Meadows A. J., O'Connor J. G.* Bibliographical statistics as a guide to growth points in science.—Sci. Stud., 1971, vol. 1, N 1, p. 95—99.
- 4.115. *Mitra A. C.* Literature cited by Indian scientists: A study of the use pattern of literature and its english-language, foreign-language and domestic components published in various periods.—Ann. Libr. Sci. and Doc., 1972, vol. 19, N 3, p. 119—145.
- 4.116. *Moravcsik M. J., Muregesan P.* Citation patterns in scientific revolutions.—Scientometrics, 1979, vol. 1, p. 161—169.
- 4.117. *Moravcsik M. J., Muregesan P.* Some results on the function and quality of citations.—Sci. Stud., 1975, vol. 5, N 1, p. 86—92.
- 4.118. *Myers C. R.* Journal citations and scientific eminence in contemporary psychology.—Amer. Psychol., 1970, vol. 25, p. 1041—1048.
- 4.119. *Noltingk B. E.* The human element in research management. Amsterdam, 1959.
- 4.120. *Noltingk B. E.* The art of research. Amsterdam etc., 1965.
- 4.121. *Oliver M. R.* The effect of growth on the obsolescence of semiconductor physics literature.—J. Doc., 1971, vol. 27, N 1, p. 11—18.
- 4.122. *Oliver P. T. P.* Citation indexing for studying science.—Nature, 1970, vol. 227, N 5260, p. 870.

- 4.123. *Oromaner M.* Career contingencies and the fate of sociological research.— *Social. Sci. Inform.*, 1973, vol. 12, N 2, p. 97—111.
- 4.124. *Price D. J. de S.* Why does science cumulate.— In: *Frontiers of science and philosophy*: Vol. 2. Pittsburg: Univ. of Pittsburg press, 1965.
- 4.125. *Price D. J. de S.* Networks on scientific papers.— *Science*, 1965, vol. 149, N 3683, p. 510—515.
- 4.126. *Price D. J. de S.* Citation measures of hard science and soft science, technology and nonscience.— In: *Communication among scientists and engineers*. Lexington (Mass.), 1970, p. 3—22.
- 4.127. *Price D. J. de S.* The citation cycle.— In: *Key papers in information science* / Ed. B. C. Griffith. N. Y.: White Plains, Knowledge Industry Publ., 1980, p. 195—210.
- 4.128. *Raisig L. M.* Mathematical evaluation of the scientific serial.— *Science*, 1960, vol. 131, p. 1417—1419.
- 4.129. *Rothman H., Woodhead M.* The use of citation counting to identify research trends.— *J. Doc.*, 1971, vol. 27, p. 287—294.
- 4.130. *Russett B. M.* Methodological and theoretical schools in international relations.— In: *A design for international research: Scope, theory, methods, and relevance* / Ed. N. D. Palmer. Philadelphia: Amer. Acad. of Polit. and Social Sci., 1970, p. 87—105.
- 4.131. *Science citation index. Guide and journal lists.* Philadelphia: Inst. for Sci. Inform., 1964.
- 4.132. *Shepard's Citation.* Chicago, 1873.
- 4.133. *Small H. G.* Co-citation in the scientific literature: a new measure of the relationship between two documents.— *J. Amer. Soc. Inform. Sci.*, 1973, vol. 24, p. 265—269.
- 4.134. *Small H. G.* Multiple citation patterns in scientific literature: The circle and bill models.— *Inform. Storage and Retrieval*, 1974, vol. 10, p. 393—402.
- 4.135. *Small H. G.* A co-citation models of scientific speciality: A longitudinal study of collagen research.— *Social Stud. Sci.*, 1977, vol. 7, p. 139—166.
- 4.136. *Small H. G., Crane D.* Specialties and disciplines in science and social science: an examination of their structure using citation indexing.— *Scientometrics*, 1979, vol. 1, p. 445—461.
- 4.137. *Tagliacozzo B.* Citations and citation indexes: A review.— *Meth. Inform. Med.*, 1967, vol. 6, S. 136—142.
- 4.138. *Wade N.* Citation analysis: A new tool for science administrators.— *Science*, 1975, vol. 186, N 4187, p. 429—432.
- 4.139. *Waldhart T. J.* Utility of scientific research. The engineer's use of the products of science.— *IEEE Trans. Prof. Commun.*, 1974, vol. P-17, p. 33—35.
- 4.140. *Wallmark J. T., Eckarstein S., Holmqvist H. E. S.* The increase in efficiency with size of research teams.— *IEEE Trans. Eng. Manag.*, 1973, vol. EM-20, p. 80—86.
- 4.141. *Wallmark J. T., Sallerberg B.* Efficiency vs. size of research teams.— *IEEE Trans. Eng. Manag.*, 1966, vol. EM-13, p. 113—142.
- 4.141a. *Weiss P.* Knowledge: a growth process.— *Science*, 1960, vol. 131, June, 1716—1719.
- 4.142. *Westbrook J. H.* Identifying significant research.— *Science*, 1960, vol. 132, Oct., p. 1229—1234.
- 4.143. *Xhignesse L., Osgood Ch.* Bibliographical citation characteristics of the psychological journal networks in 1950 and 1960.— *Amer. Psychol.*, 1967, vol. 22.
- 4.144. *Yagu Eri.* An application of a type of matrix to analyze citations of scientific papers.— *Amer. Doc.*, 1965, vol. 16, p. 10—19.
- 4.145. *Zunde P., Slamecka V.* Predictive models scientific progress.— *Inform. Storage and Retrieval*, 1971, vol. 7, p. 103—109.

5. **Наукометрия: методы «контент-анализ» и тезаурусный**
- 5.1. *Алексеев А. Н.* Об одном способе исследования тенденций развития общественных наук (частотный словарь на базе указателя пермутационного типа).— В кн.: *Наука и техника*. Л.: ЛО ИИЕТ АН СССР, 1972, вып. 7, ч. 1, с. 68—72.
- 5.2. *Богачев А. А., Дубинин В. Г., Клименко В. П.* Применение перфокарт для анализа структуры научных кадров.— В кн.: *Науковедение и информатика*. Киев: Наук. думка, 1972, вып. 5, с. 82—85.
- 5.3. *Владыкин В. А., Воронов Ю. П.* Контент-анализ и рецензирование научной литературы.— В кн.: *Методологические и методические проблемы контент-анализа*. М.; Л.: ИСИ АН СССР, 1973, вып. 2, с. 71—76.
- 5.4. *Добров Г. М., Борисов В. В.* Формализованный метод анализа содержания используемой информации.— *НТИ*. Сер. 2, 1970, № 8, с. 3—4.
- 5.5. *Дубовиков И. С., Маркова Л. П., Солец Н. А.* Применение статистического метода к анализу документальных информационных потоков (ДНП).— *НТИ*. Сер. 2, 1976, № 1, с. 22—25.
- 5.6. *Коробейникова В. С.* Анализ содержания массовой коммуникации.— *Вопр. философии*, 1969, № 4, с. 100—110.
- 5.7. *Коробейникова В. С.* Анализ содержания массовой коммуникации: Автореф. дис. ... канд. филос. наук. М.: ИКСИ АН СССР. 1969.
- 5.8. *Максимова В. В.* О социально-психологическом анализе научных текстов.— В кн.: *Проблемы деятельности ученого и научных коллективов*. Л.: Наука. Ленингр. отд-ние, 1971, вып. 4, с. 350—357.
- 5.9. *Максимов В. В.* Применение контент-анализа для исследования научных отношений.— В кн.: *Методологические и методические проблемы контент-анализа*. М.; Л.: ИСИ АН СССР, 1973, вып. 2, с. 84—87.
- 5.10. *Маршак И. В.* Исследование частотного словаря ключевых слов.— *НТИ*. Сер. 2, 1974, № 11, с. 7—13.
- 5.11. *Методологические и методические проблемы контент-анализа* М.; Л.: ИСИ АН СССР, 1973, вып. 1 и 2.
- 5.12. *Паркер-Рудс А. Ф., Уорд С.* Применение тезаурусного метода при машинном переводе с помощью существующей вычислительной техники.— В кн.: *Математическая лингвистика*. М.: Мир, 1964, с. 214—228.
- 5.13. *Потанов В. М., Лукина И. И.* Наукометрические анализ некоторых аспектов развития органической химии.— В кн.: *Наукометрические исследования в химии*. М.: Изд-во МГУ, 1974, с. 102—112.
- 5.14. *Преображенская Г. Б., Пругкова И. М., Грановский Ю. В.* Анализ статистических терминов в публикациях по спектральному анализу и аналитической химии.— *Защит. лаб.*, 1974, т. 40, № 10, с. 1240—1244.
- 5.15. *Семенов В. Е.* Применение метода контент-анализа в социально-психологических исследованиях: Дис. ... канд. психол. наук. Л.: ЛГУ, 1975.
- 5.16. *Соколов В. М.* Об объективности исследования в контент-анализе.— В кн.: *Методологические и методические проблемы контент-анализа*. М.; Л.: ИСИ АН СССР, 1973, вып. 1, с. 58—60.
- 5.17. *Тезаурус научно-технических терминов*. М.: Воен. книга, 1972.
- 5.18. *Турыкина И. Л.* Комплексное изучение состояния научных исследований в области гастроэнтерологии и внедрения их результатов в практику здравоохранения (с использованием ЭВМ): Автореф. дис. ... канд. мед. наук. М.: Моск. мед. стоматол. ин-т, 1974.
- 5.19. *Holsti O. R.* Content analysis for the social sciences and humanities. Addison—Wesley, Reading, (Mass.), 1969.
- 5.20. *Krippendorff K.* Models of messages: three prototypes.— In: *The analysis of communication content. Developments in scientific theories and computer techniques*. N. Y., 1969, p. 69—106.
- 5.21. *Lasswell H. D., Pool de S.* The comparative study of symbols. Stanford. Calif., 1952.

- 5.22. *Lasswell H. D.* The future of political science. L.: Tovistock publ., 1964.
 5.23. *Osgood Ch. E.* The representational model and relevant research methods.— In: Trends in content-analysis, Urbana, 1959, p. 33—88.
 5.24. *Raver N.* Performance of IR systems.— In: Informational retrieval: A critical view / Ed. by G. Schechter, based on the third Annual colloquium on information retrieval, 12—13 may 1966, Philadelphia, Pa., Wash., Thompson, 1967, p. 131—142.

6. Наукометрия: сленговый метод

- 6.1. *Калинин В. М.* О статистике литературного текста.— *Вопр. языкознания*, 1965, № 1, с. 123—127.
 6.2. *Колгушин А. Н.* Лингвистика в военном деле: (Разработка и использование частотных словарей военной лексики). М.: Воениздат, 1970.
 6.3. *Москович В. А.* Статистика и семантика. М.: Наука, 1969.
 6.4. *Москович В. А.* Дистрибутивно-статистический метод построения тезаурусов: Современное состояние и перспективы. М.: Исслед. совет по комплекс. пробл. «Кибернетика», 1971.
 6.5. *Налимов В. В.* Вероятностная модель языка. М.: Наука, 1974.
 6.6. *Орлов Ю. К.* О статистической структуре сообщений, оптимальных для человеческого восприятия (постановка вопроса).— *НТИ*. Сер. 2, 1970, № 8, с. 11—16.
 6.7. *Тома Е, Пуркару Д.* Об информационном значении формализованных словарей.— В кн.: Междунар. симпоз. стран-членов СЭВ «Теоретические основы информатики». М.: ВИНТИ, 1970.
 6.8. *Фрумкина Р. М.* Статистические методы изучения лексики. М.: Наука, 1964.
 6.9. *Фрумкина Р. М.* Прогноз в речевой деятельности. М.: Наука, 1974.
 6.10. *Фрумкина Р. М.* Вероятность элементов текста и прогноз в речевой деятельности: Автореф. дис. ... д-ра психол. наук. Л.: Изд-во ЛГУ, 1974.
 6.11. *Хайтун С. Д.* Количественная методика определения концентрированности научной тематики.— В кн.: Социально-психологические проблемы науки. М.: Наука, 1973, с. 226—232.
 6.12. *Хайтун С. Д.* Об одной из методик изучения научной биографии.— В кн.: Человек науки. М.: Наука, 1974, с. 214—228.
 6.13. *Хайтун С. Д.* О возможностях наукометрического исследования соавторских публикаций.— В кн.: Проблемы деятельности ученого и научных коллективов. М.; Л.: Наука, 1979, вып. 7, с. 303—308.
 6.14. *Частотный словарь общенаучной лексики.* М.: Изд-во МГУ, 1970.
 6.15. *Частотный словарь русского языка.* М.: Рус. яз., 1977.
 6.16. *Штейнфельд Э. А.* Частотный словарь современного русского литературного языка. М.: Прогресс, 1973.
 6.17. *Bosanquet C. H.* Distribution of word frequencies.— *Nature*, 1957, vol. 179, N 4559, p. 596.
 6.18. *Estoup J. B.* Gamma stenographique. 4^e ed. P., 1916.
 6.19. *Herdan G.* Language as choice and chance. Groningen: Noordhoff, 1956.
 6.20. *Shannon C. E.* Prediction and entropy of printed english.— *Bell System Techn. J.*, 1951, vol. 30, p. 50—64.
 6.21. *Stichel H. S.* On a distribution law for word frequencies.— *J. Amer. Statist. Assoc.*, 1975, vol. 70, (352, pt 1), p. 542—547.
 6.22. *Yule G.* The statistical study of literature vocabulary. Cambridge (England): Cambridge Univ. press, 1944.
 6.23. *Zipf G. K.* Selected Studies of the principle of relation frequency on language. Cambridge (Mass.): Harvard Univ. press; 1932.
 6.24. *Zipf G. K.* Psycho-biology of language. Houghton: Mufflin, 1935.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
ГЛАВА 1. НАУКОМЕТРИЯ И ПРОБЛЕМА ИЗМЕРЕНИЯ: НЕГАУССОВСТЬ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ	13
1. Измерение: воспроизводимость, адекватность, трудоемкость	14
2. Латентные переменные и их индикаторы	16
3. Негауссовость социальных явлений. (Предварительные соображения)	20
4. Особенности измерения при изучении социальных явлений	23
5. Особенности измерения при изучении научной деятельности	32
ГЛАВА 2. СТАТИСТИЧЕСКИЙ МЕТОД	38
1. Содержание метода и опорные исследования	38
2. Результаты применения: нестационарные распределения	40
3. Результаты применения: стационарные распределения	50
4. Результаты применения: структурные исследования	56
5. Перспективы	60
ГЛАВА 3. МЕТОД ПОДСЧЕТА ЧИСЛА ПУБЛИКАЦИЙ	62
1. Содержание метода и опорные исследования	66
2. Результаты применения: нестационарные распределения	70
3. Результаты применения: стационарные распределения	83
4. Результаты применения: структурные исследования	83
5. Перспективы	86
ГЛАВА 4. МЕТОД «ЦИТАТ-ИНДЕКС»	86
1. Содержание метода и опорные исследования	93
2. Результаты применения: нестационарные распределения	99
3. Результаты применения: стационарные распределения	111
4. Результаты применения: структурные исследования	114
5. Перспективы	343

ГЛАВА 5. МЕТОДЫ «КОНТЕНТ-АНАЛИЗ» И ТЕЗАУРУСНЫЙ	122
1. Содержание метода «контент-анализ»	123
2. Применение метода «контент-анализ»	126
3. Содержание тезаурусного метода	127
4. Применение тезаурусного метода	128
5. Перспективы	129
ГЛАВА 6. СЛЕНГОВЫЙ МЕТОД	130
1. Содержание метода	130
2. Результаты применения	134
3. Перспективы	146
ГЛАВА 7. ОСНОВНЫЕ НАУКОМЕТРИЧЕСКИЕ ЗАКОНОМЕРНОСТИ	150
1. Экспоненциально-логистический рост наукометрических индикаторов во времени	150
2. Негауссовость стационарных наукометрических распределений	155
НАУКОМЕТРИЯ В ОБЩЕМ КОНТЕКСТЕ ИССЛЕДОВАНИЙ НАУКИ (ВМЕСТО ЗАКЛЮЧЕНИЯ)	198
1. Комплексное применение наукометрических методов (взгляд на наукометрию «изнутри»)	198
2. Комплексное применение наукометрических и пенаукометрических методов (взгляд на наукометрию «изнутри»)	205
3. Место и роль наукометрии среди науковедческих дисциплин (взгляд на наукометрию «извне»)	212
ПОСЛЕСЛОВИЕ	217
ГРАФИКИ	221
1. Нестационарные распределения	221
2. Стационарные распределения	247
3. Результаты структурных исследований	307
4. Результаты комплексного применения наукометрических индикаторов	309
ЛИТЕРАТУРА	315
0. Наукометрия и науковедение	315
1. Наукометрия: общие вопросы	319
2. Наукометрия: статистический метод	326
3. Наукометрия: метод подсчета числа публикаций	330
4. Наукометрия: метод «цитат-индекс»	335
5. Наукометрия: методы «контент-анализ» и тезаурусный	341
6. Наукометрия: сленговый метод	342

2p. 50k.



INVENTEDO «HAYKA»