Июль 1992 г. Том 162, № 7

УСПЕХИ ФИЗИЧЕСКИХ НАУК

МЕТОДИЧЕСКИЕ ЗАМЕТКИ

53.088

# ЯВЛЯЕТСЯ ЛИ ВЕРОЯТНОСТЬ «НОРМАЛЬНОЙ» ФИЗИЧЕСКОЙ ВЕЛИЧИНОЙ?

*Ю.И. Алимов, Ю.А. Кравцов*

(Акционерное общество ЦИТРОН, Екатеринбург, Малое предприятие ГРОТ, Институт общей физики РАН, Москва)

СОДЕРЖАНИЕ



## 1. Введение. Неформальные аспекты вероятности

Любой физик видит в вероятности как физическую, так и математическую величины. Однако если математическая концепция вероятности представляется ему солидной и непререкаемой, то с вероятностью как с физической величиной часто ощущаются какие-то неудобства и недоговоренности. В связи с этим и возникает вопрос, вынесенный в заголовок: является ли вероятность настоящей, "нормальной" физической величиной, т.е. величиной, допускающей физические измерения, или же это неизмеримая, вымышленная величина?

Если это в самом деле нормальная физическая величина, то почему постоянно возникают сомнения в результатах статистической обработки результатов измерений? Если же вероятность чем-то отличается от обычных физических величин, то в чем именно выражается ее «ненормальность»?

Вопросы этого круга стали составной частью науки уже более 200 лет назад. Каждая эпоха стремилась ответить на него в меру своих представлений о предмете. В той или иной форме о физическом смысле вероятности высказывались почти все крупные физики и математики.

И все же вопрос не закрыт по сей день. Почему же так сложно договориться о физическом смысле вероятности и о точности ее измерения? Может быть, этому препятствует нечто существенное, но трудноуловимое, постоянно остающееся за кадром?

Мы склонны думать, что существенным элементом, незримо присутствующим при физической интерпретации вероятности, является система трудно формализуемых гипотез, соглашений, домысливаний, как бы естественно, традиционно привязанных к формальному аппарату теории вероятностей, а в действительности являющихся самостоятельными гипотезами, требующими верификации.

Авторы с известным чувством удивления обнаружили, что эвристических элементов в физико-вероятностных построениях гораздо больше, чем можно было бы предположить, даже имея за плечами многолетний опыт работы в математической статистике.

В данной работе мы хотели бы обсудить именно неформальные аспекты понятия вероятности, возникающие в физическом контексте. На примере определения математического ожидания мы проанализируем роль эвристических, внелогических утверждений, казались бы, в хорошо освоенной области знаний. В сущности, речь идет о несостоятельности фишеровской процедуры обработки результатов измерений, которая укоренилась в естественнонаучной практике несмотря на сомнительные домысливания, и об описании альтернативной процедуры нахождения многовыборочных доверительных интервалов, основанной на простых и интуитивно приемлемых положениях (см. раздел 5).

Кроме того, в разделе 6 мы обсудим широкий перечень причин, по которым приходится прибегать к домысливанию в статистике, а в разделе 7 изложим взгляд на вероятность с позиций концепции частичной детерминированности (предсказуемости), которая, как нам представляется, лучше согласуется с логикой физического эксперимента, чем концепция алгоритмической сложности.

## 2. Верифицирующий эксперимент

**2.1.** Математические «жернова» и эвристическая «засыпка» естественнонаучных теорий. При обсуждении вероятности как физической величины, т.е. величины, допускающей измерения, полезно напомнить о роли, которую играет в естествознании математическая теория в целом и теория вероятностей — в частности. Последняя функционирует в естествознании так же, как другие математические дисциплины.

Математика представляет собой богатый специальный язык, в грамматику которого вмонтирована мощная формальная логика. «Непостижимая эффективность математики» и ее незаменимость как уникального языкового средства неоднократно отмечалась естествоиспытателями (пожалуй, наиболее известная статья на эту тему принадлежит Е. Вигнеру [1]).

В то же время понятно, что практическая деятельность не сводится к деятельности языковой. Соответственно естественнонаучные теории, будучи нацелены в конечном счете на практические приложения, не сводятся к математическим построениям. Ограниченность роли математики в естествознании хорошо отражена в известном высказывании Т. Гексли (T. Huxley): «Математика, подобно жернову, перемалывает то, что под него засыпают, и как засыпав лебеду, вы не получите пшеничной муки, так исписав целые страницы формулами, вы не получите истины из ложных предпосылок» [2, с. 106].

Тот факт, что математические жернова требуют эвристической «засыпки» извне, обусловлен природой формально-логического заключения. Подобные заключения — в том числе математические выкладки — всегда имеют импликативную форму «если..., то...». Официальные обязанности математики в приложениях сводятся в общих чертах к решению следующей задачи: по заданной исходной величине $X$ и заданному (возможно, в неявной форме) отображению $F$ вычислить выходную величину $Y = F(X)$. Отображение $F$ как раз является «жерновом» — математической моделью реальности. Детализируя высказывание Гексли, подчеркнем, что неизбежная эвристическая «засыпка», или, как более сухо сказано в [3], *внелогический компонент* включает выбор не только расчетного значения исходной величины, но и математических «жерновов» При плохих жерновах не получить хорошей муки даже из пшеницы.

Математика играет в приложениях и неофициальную роль, предоставляя естествоиспытателям сам язык для описания величин $X$, $Y$ и отображения $F$. Хотя математика не несет тут прямой ответственности, эта ее чисто языковая роль не менее важна, чем вышеупомянутая официальная вычислительная обязанность.

Выбор расчетного значения исходной величины $X$ оказывается, по существу, эвристическим даже тогда, когда это значение получают путем измерений. Ведь измерения и вообще эксперименты не являются полностью формализованными операциями по той простой причине, что они касаются реальности, а не какой-то формальной системы. Впрочем, исходное расчетное значение $X$ исследователь вправе постулировать и не прибегая к измерениям именно этой величины. Относительно же величин $Y$, считающихся выходными, в естествознании сложилась более жесткая традиция: вычисленное значение $Y$ всегда необходимо сопоставлять с действительностью с помощью измерений, В таком сопоставлении состоит экспериментальная проверка адекватности — *верификация* — всей теории, в том числе и выбора расчетных значении $X$. Иначе теория будет не естественнонаучной, а всего лишь математической конструкцией.

**2.2.** Метрологические установки верифицирующего эксперимента. В конечных результатах верифицирующего эксперимента — образно говоря, в постановлении суда окончательной инстанции — все должно быть интуитивно ясно, непосредственно убедительно, по крайней мере для большинства специалистов в данной области исследований. Для этого прежде всего необходимо придерживаться следующей фундаментальной метрологической установки:

1°. *Тщательно разрабатывать модель объекта измерения.* В противном случае останется неясным, достаточно ли близки по своему смыслу математическая величина $Y$ и физическая величина, измеряемая в верифицирующем эксперименте. Разработка модели объекта измерения является органической составной частью верификации. Остановимся на этом подробно.

Выполняя любые измерения, экспериментатор мысленно заменяет реальный объект измерения моделью. В качестве измеряемых величин выступают, сторого говоря, параметры этой модели, а не реального объекта. Под истинным значением измеряемой величины понимается такое значение параметра модели, которое получилось бы в результате безупречного эксперимента над идеальным объектом, являющимся точной материализацией модели. Однако между свойствами реального объекта и тем, что подразумевают под измеряемой величиной, всегда возможно некое качественное несоответствие. Оно порождает «погрешность несоответствия модели объекту» [4, с. 14], включаемую в методические, а иногда и в систематические погрешности. Если погрешность несоответствия слишком велика, переходят к другой модели.

Адекватная модель объекта измерения призвана гарантировать саму осмысленность понятия истинного значения измеряемой величины. Представим, например, что измеряется размер шарообразного тела — скажем, дробинки или Земли. Пусть в качестве модели взят шар. Соответственно ставится задача оценить единственное значение диаметра тела. Пусть, однако, измерения нескольких диаметров шарообразного тела дали результаты, которые различаются на величины, превышающие погрешность выбранных средств измерения. Тогда придется констатировать, что в данной конкретной обстановке модель в виде шара неадекватна объекту. Стало быть, поставленная задача оценивания единственного истинного значения диаметра тела лишена смысла: в рамках имеющейся точности измерений такого значения нет. Подобный отрицательный вывод, полученный при разработке модели объекта измерения, сам по себе может указывать на неадекватность всей верифицируемой теории.

Приведем еще два примера. Рассмотрим генератор электрических сигналов с неизвестной нам структурой и постараемся восстановить объект (структуру генератора) по форме генерируемых импульсов. Если импульсы по форме напоминают, скажем, процесс заряда и разряда конденсатора, то в качестве первичной модели естественно взять релаксационный генератор. Такая модель может оказаться удовлетворительной при грубой подгонке параметров цепи (емкости $C$ и сопротивления $R$) к форме генерируемых импульсов. Если впоследствие выяснится, что различие между реальными и модельными импульсами превышает погрешности измерительного прибора, то модель простого $RC$-генератора следует признать неадекватной наблюдаемому процессу, что потребует пересмотра модели в сторону ее усложнения.

В этом примере проявляется общая характерная особенность обратных задач: по мере накопления данных сначала происходит уточнение параметров простой модели, а затем наступает момент радикальной перестройки модели в целом.

Второй пример относится к измерению показателя ослабления волнового поля в поглощающей или рассеивающей среде. Пусть на опыте измерена зависимость интенсивности $I$ от дистанции $x$, представленная в логарифмическом масштабе на рис. 1. Допуская, что интенсивность меняется по экспоненциальному закону $I(x) = I_0\exp(-\mu x)$ (закон Бугера), показатель ослабления $\mu$ можно найти, подгоняя линейную зависимость $-\mu x$ к (логарифмированным) экспериментальным точкам.

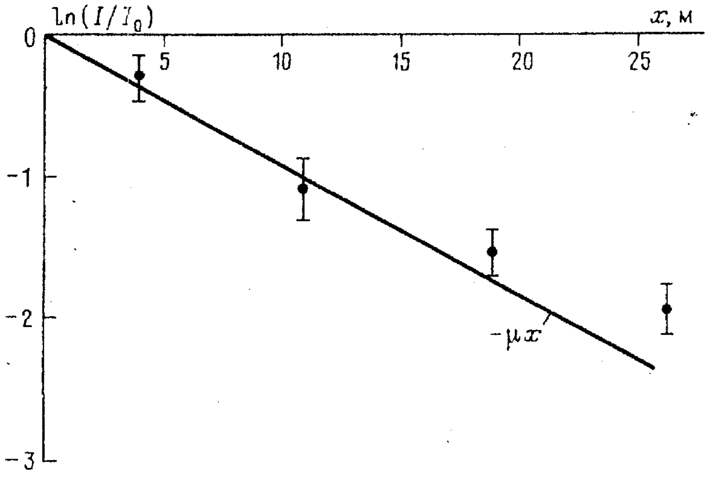


Рис. 1. К вопросу о выборе линейной модели — $\mu x$ для экспериментальной зависимости $\ln(I(x)/I_0)$. Слишком большой разброс экспериментальных точек свидетельствует о непригодности линейной модели и о неприменимости представления о постоянном коэффициенте поглощения $\mu$

Если разница между экспериментальными значениями $\ln(I/I_0)$ и линейной функцией $-\mu x$ превысит ошибку измерения, то интерпретация $\mu$ как коэффициента ослабления однородной среды теряет смысл. В этом случае требуется пересмотреть саму модель среды и/или повторить эксперимент с большей детальностью или с повышенной точностью (кстати, ограничения точности могут быть не только с физической, но и с экономической стороны).

Разработка модели объекта измерения является в целом неформальной операцией, требующей глубоких предметных знаний. Расчеты иногда столь тесно переплетаются с экспериментальными исследованиями, что появляется возможность говорить о едином *теоретико-экспериментальном процессе выработки модели объекта.*

Предположим, далее, что адекватная модель объекта измерения разработана и используется в верифицирующем эксперименте. Теперь для достижения интуитивной ясности и убедительности результатов этого эксперимента необходимо следить за тем, чтобы конечные количественные показатели адекватности теории были простыми, не связанными со сколько-нибудь сложными формальными построениями. Речь идёт о показателях, характеризующих точность окончательной принимаемой оценки истинного значения величины $Y$. Ведь дело дошло именно до эмпирического обоснования справедливости теории!

В противоположность схоластике в естествознании сразу был принят основной прагматический тезис: *справедливость теории в конечном счете обосновывается экспериментом, а не какой-либо другой теорией* (скажем, трудами Аристотеля или К. Маркса).

Представим, что для «обсчета» конечных результатов верифицирующего эксперимента мы все же строим достаточно сложную математизированную теорию. Такая вторичная теория породит соответственно непростые количественные показатели адекватности исходной, первичной теории. Эти показатели имеют отношение к действительности, очевидно, в той мере, в какой в целом адекватна вторичная теория. Возникает логическое зацикливание: в процессе верификации первичной теории мы обрели более трудную, а главное, бесперспективную проблему верификации вторичной теории. Данная благоприобретенная проблема труднее исходной задачи как раз тогда, когда вторичная теория математизирована и достаточно сложна. Бесперспективность же обусловлена логическим зацикливанием: если мы станем верифицировать в прежнем стиле и вторичную теорию, то придем к необходимости верифицировать возникающую непростую формализованную теорию «третьего поколения», т.е. теорию показателей адекватности вторичной теории. Ясно, что обрывая цепочку, мы рискуем поставить под сомнение даже первичную теорию.

Словом, во избежание логического зацикливания модель сопоставления конечных результатов верифицирующего эксперимента с вычисленными выходными результатами естественнонаучной теории обязана быть достаточно простой и неформальной. По своему характеру модель сопоставления относится к эвристической «засыпке», внелогическому компоненту, а не к математическим «жерновам».

Можно сказать, что на пути от теоретической деятельности к верификации неизбежна редукция применяемого языка до максимально деформализованного «языка наблюдения». Иначе невозможно достигнуть интуитивной, неформальной убедительности результатов верификации. Согласно естественнонаучной традиции желательно, чтобы наряду с требованием 1° выполнялись еще два требования:

2°. *Убедительно демонстрировать воспроизводимость соответствующего экспериментального результата.*

3°. *Убедительно оценивать систематические погрешности в воспроизводимом экспериментальном результате.*

Иногда возникает необходимость оценить с максимально возможной точностью истинные значения входных или каких-то промежуточных величин теории (а не только выходной величины $Y$). Так получается, например, если хотят верифицировать, или, как часто говорят, идентифицировать саму модель объекта, т.е. «жернова». В подобных ситуациях тоже необходимо следовать фундаментальным метрологическим установкам 1° — 3° и сопутствующему им требованию простоты и неформальности окончательных показателей точности результата измерений.

Таким образом, математика и метрология служат одинаково незаменимыми опорами всего количественного естествознания. Требования 1° — 3° оказываются обязательными при выполнении любых измерений, в том числе при измерении вероятностей или, более общо, математических ожиданий.

Мы тут не ломимся в открытые двери. На действующие сейчас метрологические нормативы [5] сильно повлияла теория статистической обработки данных, называемая математической статистикой и связываемая с именем Р. Фишера. Эта теория не вполне согласуется с традиционными метрологическими установками (критику фишеровской статистики см., например, в работах [12 — 18]). Так, при верификации фишеровскаяргатистика предлагает вычислять доверительные интервалы в рамках отнюдь не простой математической модели верифицирующего эксперимента.

Идеология фишеровской математической статистики подвергается анализу ниже в разделе 5. Отдельные замечания в ее адрес делаются по ходу изложения и раньше, они адресованы читателям, знакомым с фишеровской математической статистикой.

**2.3.** Чем занимается теория вероятностей. С точки зрения естествоиспытателя, теория вероятностей занимается тем же, что и любой другой раздел математики, а именно: изобретает и предоставляет пользователям математические «жернова» для логического перемалывания эвристической «засыпки».

Характер «засыпки», «жерновов» и «муки» в сфере действия теории вероятностей несколько специфичен. Основные теоретико-вероятностные понятия, работающие в приложениях, — это математическое ожидание, вероятность, распределение вероятностей. Назовем их собирательно *вероятностными характеристиками* (случайных величин или случайных событий).

В рамках статистической интерпретации вероятностные характеристики относятся к некоторому реально несуществующему объекту, который в физике называется ансамблем, а в математической статистике и теории вероятностей — генеральной совокупностью объектов или испытаний (наблюдений).

Специфика теории вероятностей сводится к тому, что «засыпкой» в ее математические «жернова» и выходным продуктом обычно считаются вероятностные характеристики. Иными словами, теория вероятности пересчитывает одни вероятностные характеристики в другие.

Проиллюстрируем сказанное простейшим примером теоретико-вероятностного соотношения для суммы случайных величин $Y = \Sigma\, a_i X_i$:

$$M[Y] = M[\sum_{i=1}^{m} a_i X_i] = \sum_{i=1}^{m} a_i M[X_i]; \tag{2.1}$$

здесь $M$ — оператор математического ожидания, $X_i$ — случайные величины, $a_i = \mathrm{const}$. «Засыпкой» в (2.1) служат задаваемые значения $M[X_1], \ldots, M[X_m]$ и значения $a_1, \ldots, a_m$ параметров модели, а выходом — вычисленное значение математического ожидания линейной функции $Y = \sum a_i X_i$.

Для того чтобы выявить механизм возникновения и сам смысл теоретико-вероятностных отношений, необходимо обратиться к физическому анализу основных понятий. Этим центральным для данной публикации вопросам посвящены разделы 3 и 4. Теория вероятностей, как и любая математическая дисциплина, заведомо не может претендовать на расчет точности и достоверности «засыпки» и «жерновов» (т.е. исходных величин) используемой модели и конечных результатов верифицирующего эксперимента. В том, как подходит к проблеме верификации фишеровская математическая статистика, можно усмотреть именно такую претензию.

## 3. Гипотезы, скрывающиеся в тени экспериментов

**3.1.** Принцип многократного воспроизведения: «как много раз бывало, так, по-видимому, и будет». Вернемся к метрологическим установкам 1° — 3° и рассмотрим их важнейшую черту: эти установки обеспечивают ни много ни мало выполнение естественными науками их главной задачи — прогнозировать результаты предстоящих экспериментов.

Количественные естественнонаучные прогнозы имеют форму следующего утверждения: реализация в эксперименте определенных контролируемых условий $U$ всегда приведет к одному и тому же (в пределах оговоренной точности) измеряемому результату $Y$. Условимся обозначать подобный прогноз символом $U \to Y$ (от $U$ к $Y$). Верифицировать теорию — значит верифицировать доставляемый ею прогноз.

Описание и реализация условий $U$, а также измерение прогнозируемого результата $Y$ осуществляются в соответствии с требованиями 1°, 3°, ориентирующими исследователя на тщательную разработку модели объекта измерения, всемерное устранение систематических погрешностей и оценку той их части, которая не поддается устранению. Без выполнения этих требований вообще невозможно говорить о верификации прогноза $\boldsymbol{U} \to \boldsymbol{Y}$.

Согласно требованию 2° верификация прогноза должна включать многократное воспроизведение условий $U$ и соответственно многократное измерение получаемых результатов $Y$. По традиции в естествознании вполне всерьез принимается только такой прогнозируемый результат, который воспроизводится без непонятных сбоев в достаточно длинной серии однородных испытаний. Весь опыт естествознания учит видеть в повторении испытаний необходимое, хотя и не абсолютное противоядие против случайных погрешностей, грубых промахов и подтасовок.

Верификацию прогноза, даже выполненную в соответствии, с традиционным требованием многократной воспроизводимости, конечно, нельзя считать безупречной. Такая верификация прогноза сама представляет собой прогноз «как много раз подряд было, так, видимо, и будет», или короче — «завтра как сегодня», связываемый с понятием эмпирической индукции. Хорошо известно — хотя бы из повседневной жизни — что этот эмпирико-индуктивный прогноз далек от безупречности. В частности, этот принцип может представлять известную опасность при анализе уникальных, дорогостоящих, незапланированных и некоторых других экспериментов, скажем, экспериментов в космосе, в глубинах океана и в других трудно доступных условиях. Этот принцип «завтра как сегодня» несет в себе значительный консервативный заряд и органически не приспособлен (не «настроен») к выявлению новых тенденций и изменений.

Несмотря на это, способов верификации, более убедительных и надежных, чем многократное повторение испытаний с соблюдением требований 1° — 3°, наука не знает. Впрочем, естественные науки и не претендуют на непогрешимость.

**3.2.** Принцип разумной достаточности. При проведении серии однородных испытаний возникает вопрос о том, какую минимальную длину $Q$ должна иметь эта серия, чтобы доказательство воспроизводимости экспериментального результата считалось убедительным. Важно осознавать, что данный вопрос касается эвристической «засыпки» и, таким образом, неразрешим формально-логическими средствами. По-видимому, время от времени чувствуется потребность напоминать о неизбежности эвристических-компонентов в естественнонаучном исследовании. Так, Е.Л. Фейнберг [3, с. 35] по поводу внелогического суждения о достаточности данной опытной проверки говорил: «... Оно сопутствует любому, даже вполне рядовому эксперименту, который остается ограниченным, как бы он ни модифицировался, ни расширялся и сколько бы раз ни повторялся. Все равно наступает момент, когда исследователь должен сказать: «Довольно, теперь я *убежден,* что такая-то закономерность верна». Это «я убежден» выражает внелогическое суждение в науке, которое является неустранимым и принадлежит к основам процесса познания».

Можно только гадать о моменте, когда догадка, гипотеза переходит в уверенность. Скорее всего, срабатывает какой-то внутренний критерий типа *разумной достаточности.* Значимыми признаются суждения, признаваемые сообществом специалистов, так что разумная достаточность прямо или косвенно ориентируется на господствующую в данный момент научную парадигму.

Выбор допуска $\Delta Y$ на рассеяние результатов измерения величины $Y$ в проводимой серии испытаний и оценивание систематических погрешностей тоже приходится выполнять, по существу, внелогически. Коротко говоря, все это делается по ближайшим прецедентам. Если прецедентов много и они однородны, налицо добротная эмпирическая индукция.

**3.3.** Статистический прогноз: «лучше прогнозировать что-нибудь, чем ничего». Измерение вероятностных характеристик особенно тесно связано с прогнозированием. Мы проследим эту связь в разделе 4 на примере математического ожидания. В общих чертах ситуация здесь такова: измерение вероятностных характеристик величины $Y$ приходится начинать с измерения соответствующих *статистических характеристик* $S_Y$, которые часто называют также *выборочными статистиками,* а в физике — *эмпирическими характеристиками*. Таким образом, вероятностные характеристики выступают как своего рода надстройки над эмпирическим материалом — статистическими (выборочными) характеристиками.

С позиций прогнозирования логика здесь вполне понятна. Исследователь, убедившись в невозможности достаточно хорошо предсказывать величину $Y$, приходит от неудавшегося «динамического» прогноза $U \to Y$ к более грубому статистическому прогнозу $U \to S_Y$, надеясь на то, что последний окажется *удачным* и притом в какой-то степени *полезным*. Словом, здесь следуют прагматической идее: *«лучше прогнозировать что-нибудь, чем ничего»*.

Статистические характеристики величин, непредсказуемых в динамическом смысле, в самом деле нередко оказываются воспроизводимыми, т.е. почти постоянными при переходе от одной достаточно большой выборки к другой, и именно поэтому оказываются поддающимися количественному прогнозированию. Эту разновидность воспроизводимости экспериментального результата называют его *статистической устойчивостью*.

Верификация прогноза статистической характеристики тоже должна в идеале выполняться в соответствии с требованием многократного воспроизведения экспериментального результата в однородных испытаниях. Не видно никаких оснований ставить статистические прогнозы в исключительное положение по отношению к данному традиционному естественнонаучному требованию. Особенность здесь заключается только в том, что при верификации статистического прогноза $U \to S_Y$ каждое испытание оказывается составным, вторичным по сравнению с исходными испытаниями, в которых измерялись значения $Y$. Оно состоит в вычислении статистической характеристики $S_Y$ для *серии* исходных испытаний. Словом, многократное повторение вторичных испытаний означает получение и обработку многих выборок. Назовем такую верификацию *многовыборочной*. Отметим, что фишеровская математическая статистика как раз ориентирована на многовыборочную верификацию статистических прогнозов.

Не видно оснований и для того, чтобы делать другие исключения для статистических прогнозов $U \to S_Y$ по сравнению с прогнозами $U \to Y$. Все, что сказано выше о верификации естественнонаучных теорий и доставляемых ими прогнозов, следует полностью относить и к случаям, когда эти прогнозы — статистические. Тут тоже необходима интуитивная ясность конечных результатов верифицирующего эксперимента, требующая описания этих результатов на простом языке и обращения к прецедентам как конечной аргументации. В. частности, здесь тоже внелогичны по своей сути выбор нижней оценки

$Q_{\min}$ для необходимой длины серии проверочных испытаний и выбор допуска $\Delta S_Y$ на рассеяние результатов измерений $S_Y$. Фишеровская математическая статистика пренебрегает и этими важными обстоятельствами.

## 4. Математическое ожидание и вероятность как физические величины

**4.1.** Выборочное среднее. Измерение математического ожидания $M[Y]$ случайной величины $Y$ начинают с измерения статистического (выборочного, эмпирического) среднего

$$M_n[Y] \overset{\text{not}}{=} \frac{1}{n}\sum_{s=1}^{n} Y(s); \tag{4.1}$$

здесь для четкости изложения введен знак not — «равно по обозначению»; через $s$ обозначен номер испытания, а через $Y(s)$ — измеренное значение величины $Y$ в $s$-м испытании. Последовательность

$$\{Y(1), \ldots, Y(n)\} \overset{\text{not}}{=} \{Y(s)\}_1^n \tag{4.2}$$

именуют выборкой объема $n$.

Пусть в качестве величины $Y(s)$ выступает индикатор $I_A(s)$ события $A$, определяемый как

$$I_A(s) = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}, \text{ если в } s\text{-м испытании событие } A \begin{Bmatrix} \text{не наступило} \\ \text{наступило} \end{Bmatrix}. \tag{4.3}$$

Тогда величина

$$\omega_n(A) \overset{\text{not}}{=} M_n[I_A] = \frac{1}{n}\sum_{s=1}^{n} I_A(s) = \frac{n(A)}{n}, \tag{4.4}$$

представляющая собой выборочное среднее индикатора $I_A$, имеет смысл относительной частоты события $A$ в выборке $\{I_A(s)\}_1^n$. Эта величина равна отношению числа испытаний $n(A)$, в которых произошло событие $A$, к общему числу испытаний $n$. Все, что будет сказано по поводу $M_n[Y]$, применимо и к рассматриваемому специальному случаю относительной частоты $\omega_n$.

Величина $M_n[Y]$ представляет собой одну из статистических характеристик величины $Y$, так что согласно сказанному в разделе 3.3 о статистических характеристиках нам предстоит экспериментально оценить воспроизводимость выборочных средних. В идеале для этого нужно при условиях эксперимента $U$ получить достаточно большое число $Q$ равноценных выборок $\{Y_{(k)}(s)\}_1^n$, $k = \overline{1, Q}$, объема $n$. Здесь $Y_{(k)}(s)$ — измеренное значение величины $Y$ в $s$-м первичном испытании из $k$-й выборки, так что вторичные испытания мы нумеруем индексом $k$.

Равенство объемов всех выборок является необходимым условием однородности вторичных испытаний: всю имеющуюся первичную выборку $\{Y(s)\}_1^N$ мы разбиваем на $Q$ равноценных подвыборок, так что общее число всех измерений (испытаний) составляет $N = Qn$. Число подвыборок $Q$ должно быть не меньше некоторого порогового значения $Q_{\min}$, представляющего собой нижнюю оценку количества вторичных испытаний: $Q > Q_{\min}$. Значение $Q_{\min}$ определяется соображениями внелогического характера и принимается по прецедентам.

Завершается $k$-е вторичное испытание вычислением подвыборочного среднего по алгоритму (4.1):

$$M_{k,n}[Y] \overset{\text{not}}{=} \frac{1}{n}\sum_{s=1}^{n} Y_{(k)}(s), \quad k = \overline{1, Q}. \tag{4.5}$$

В случае, когда измеряемой величиной $Y$ служит индикатор $I_A$ события $A$, $Q$ серий испытаний дадут нам $Q$ значений относительно частот $\omega_n^1, ..., \omega_n^Q$ (рис. 2). Эти значения испытывают вариации от серии к серии, как это присуще результатам физических измерений.

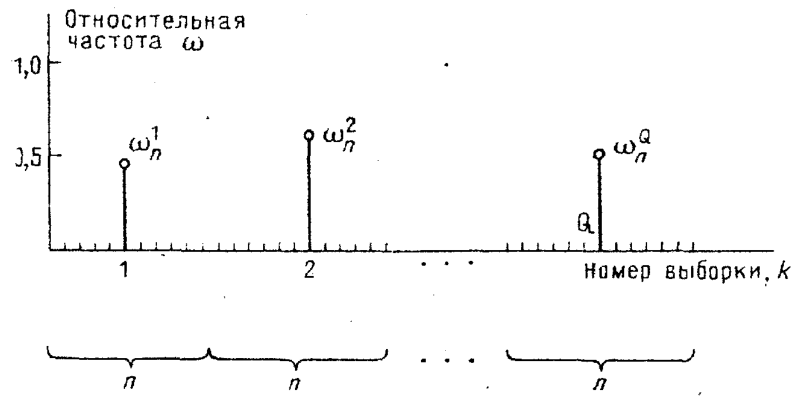


Рис. 2. Относительная частота меняется от одной выборки к другой. С ростом объема выборки $n$ разброс величин $\omega_n^k$, $k = 1, ..., Q$, **уменьшается**

Таким образом, относительная частота $\omega_n^k$ ведет себя как обычная физическая величина, т.е. величина, доступная измерениям. «Нормальные» физические свойства относительной частоты становятся менее очевидными, как только мы попытаемся придать $\omega_n$ вероятностную интерпретацию. К этому вопросу мы обратимся в разделе 4.5.

**4.2.** Уменьшение рассеяния с ростом объема выборки. В рассматриваемой схеме совершается переход от первичной непредсказуемой величины $Y(s)$ ко вторичной величине $M_{k,n}[Y]$, которая, вообще говоря, тоже непредсказуема. Смысл перехода заключается в том, что вторичная величина оказывается все же более предсказуемой, поскольку ее значения при $k = \overline{1, Q}$ и $n \gg 1$ обычно рассеяны существенно меньше, нежели значения $Y(s)$ при $s = \overline{1, N}$.

Степень рассеяния во вторичной выборке

$$\{M_{k,n}[Y]\}_1^Q \overset{\text{not}}{=} \{M_{1,n}[Y], ..., M_{Q,n}[Y]\} \tag{4.6}$$

целесообразно описать при помощи какой-нибудь простой количественной характеристики. Простейшим способом характеризовать разброс совокупности чисел (4.6) является интервал $[\overline{Y}_{\min}, \overline{Y}_{\max}]$, в котором заключены все значения величины $M_{k,n}[Y]$ в выборке (4.6):

$$\overline{Y}_{\min} = \min_{k=\overline{1,Q}} M_{k,n}[Y], \quad \overline{Y}_{\max} = \max_{k=\overline{1,Q}} M_{k,n}[Y]. \tag{4.7}$$

**Таким образом,**

$$M_{k,n}[Y] \in [\overline{Y}_{\min}, \overline{Y}_{\max}], \quad k = \overline{1, Q}. \tag{4.8}$$

Граничные значения $\overline{Y}_{\min}$ и $\overline{Y}_{\max}$ зависят от числа испытаний $n$ в одной выборке и от общего числа выборки $Q$.

Придерживаясь позиции, изложенной в разделе 2.3, мы подчеркиваем, что характеристика рассеяния подвыборочных средних *обязана быть простой,* поскольку речь идет уже об эмпирической оценке справедливости теории. Сама теория может быть формализованной и сколь угодно сложной. Однако при ее верификации уже неуместно пользоваться сложными формализмами.

Фактические границы изменения величины $M_{k,n}[Y]$ во вторичной выборке для задачи о бросании игрального кубика

| $Q$ | $n$ | $N = Qn$ | $\overline{Y}_{\min}$ | $\overline{Y}_{\max}$ | $\Delta\overline{Y}$ |
|---|---|---|---|---|---|
| 8 | 40 | 320 | 2,98 | 3,72 | 0,74 |
| 8 | 160 | 1280 | 3,31 | 3,60 | 0,29 |

При $n \gg 1$ размах $\Delta\overline{Y} = \overline{Y}_{\max} - \overline{Y}_{\min}$ колебаний подвыборочных средних (4.8) нередко в самом деле оказывается гораздо меньше размаха колебаний первичной величины $Y$. С ростом объема испытаний $n$ в подвыборке величина $\Delta\overline{Y}$ заметно уменьшается даже при фиксированном числе $Q$ подвыборок, поскольку общий объем $N = Qn$ обрабатываемой выборки растет.

Пример такого проявления устойчивости подвыборочных средних дается в таблице, заимствованной из отчета [14, п. 1.2] о лабораторной работе по «нефишеровскому» курсу теории вероятностей [14 — 16]. Первичным испытанием являлось бросание обычного игрального кубика и протоколирование выдавшего числа очков $Y(s)$. Из таблицы видно, что значения $M_n[Y]$ сосредоточены в окрестности классического среднего $\overline{Y}_{\text{class}} = (1/6)\,(1 + 2 + 3 + 4 + 5 + 6) = 3{,}5$, однако рассеяние относительно этого значения уменьшается с ростом объема выборки с $\Delta\overline{Y} = 0{,}74$ при $n = 40$ и $Q = 8$ до $\Delta\overline{Y} = 0{,}29$ при $n = 160$ и $Q = 8$.

**4.3.** Интервальный статистический прогноз. Прежде всего констатируем, что соотношение (4.8) выполнялось в проведенной серии однородных вторичных испытаний многократно и без единого сбоя (читателям, которых может смутить наш выбор интервала в правой части (4.8), рекомендуем следующее: представить, что этот выбор сделан еще до начала эксперимента). Опираясь на принцип эмпирической индукции «как много раз подряд было, так, по-видимому, и будет» предполагаем, что соотношение (4.8) будет выполняться и при продолжении вторичных испытаний:

$$M_{k,n'}[Y] \in [\overline{Y}_{\min}, \overline{Y}_{\max}], \quad \forall\, k > Q \text{ и } \forall\, n' > n\,. \tag{4.9}$$

Будущим подвыборкам мы присвоили номера $k = Q + 1,\ Q + 2,\ \ldots$

Интервальный статистический прогноз (4.9), в сущности, и является главным выходным продуктов многовыборочного эмпирического оценивания (устойчивости) средних. Этот эвристический прогноз вырабатывался на $Q$ подвыборках объема $n$. Сформулирован же он для объемов $n' > n$ потому, что при увеличении объема подвыборок воспроизводимость средних обычно улучшается (в частности, это улучшение демонстрирует таблица).

**4.4.** Точечный статистический прогноз. Математическое ожидание. Пусть $\Delta\bar{Y}$ достаточно мало, т.е. $\Delta\bar{Y}$ меньше допуска на рассеяние подвыборочных средних, выбираемого по прецедентам. Именно такую благоприятную ситуацию условимся называть дальше устойчивостью средних или, более общо, статистической устойчивостью. В данной ситуации заменим интервальную оценку-прогноз (4.9) несколько более грубой, но зато более простой точечной оценкой

$$M_{k,n}[Y] \cong M[Y], \quad \forall\, k > Q \text{ и } \forall\, n' > n; \tag{4.10}$$

здесь $M[Y]$ — число, выбранное из интервала $[\bar{Y}_{\min}, \bar{Y}_{\max}]$ в качестве его типичного представителя. С позиций экспериментатора оно и является тем, что называют измеренным значением математического ожидания случайной величины $Y$.

Правило выбора числа $M[Y]$ должно быть простым, апеллирующим лишь к здравому смыслу и прецедентам, но не к каким-либо формальным моделям. В качестве типичного представителя интервала можно брать его центр. В нашем случае это означает выбор по правилу $M[Y] = (\bar{Y}_{\min} + \bar{Y}_{\max})/2$. Чаще пользуются правилом $M[Y] = M_N[Y]$, где $N = Qn$ — объем всей выборки, полученной в эксперименте. «Мудрствовать» при выборе значений $M[Y]$ лишено смысла: все равно погрешность измерения $M[Y]$ приходится считать примерно равной $\Delta\bar{Y}$.

**4.5.** Статистическая (эмпирическая) вероятность как математическое ожидание. Статистическая (на физическом жаргоне — эмпирическая) вероятность $p(A)$ случайного события $A$ представляет собой математическое ожидание индикатора $I_A$, введенного соотношением (4.3):

$$p(A) \overset{\text{not}}{=} M[I_A]. \tag{4.11}$$

Словом, с точки зрения экспериментатора, статистическая вероятность $p(A)$ выступает как частный случай математического ожидания. Это согласуется с относительно новой тенденцией строить теорию вероятности не на основе вероятностной меры, как в теоретико-множественной аксиоматике Колмогорова, а на базе операции усреднения. Эта тенденция наиболее полно выражена в книге Х. Уиттла [19]. Для полноты картины укажем измерение распределения вероятностей случайной величины: оно фактически сводится к измерению конечного набора вероятностей, т.е. опять-таки к измерению математических ожиданий.

В соответствии со свойствами математического ожидания при увеличении длины выборки (и при однородности условий испытаний $U$) разброс величины $p(A)$ часто уменьшается. Это согласуется с частотной концепцией Р. Мизеса [20]. Можно только согласиться с мнением В.Н. Тутубалина, что «условия практической применимости теории вероятности сейчас трактуются по Р. Мизесу» [21, с. 143].

Вообще, мизесовское построение теории вероятностей не вписывается в аксиоматику А.Н. Колмогорова [22]. Критика этой аксиоматики содержится, в частности, в работах [12, 13, 23, 24], принадлежащих одному из авторов данной статьи.

Принимая мизесовское соответствие между относительной частотой и вероятностью

$$\lim_{N\to\infty} \omega_N(A) = p(A), \tag{4.12}$$

мы одновременно переносим на вероятность гласные и негласные допущения, касающиеся относительной частоты. Не все из этих допущений естественно вписываются в теоретико-множественного трактовку Колмогорова [22], которая тоже содержит систему соглашений допущений. Достаточно упомянуть, например, соглашение о существовании статистического ансамбля (генеральной совокупности), удовлетворяющего определенной вероятностной мере, или соглашение о случайной функции (сложности и условности этого понятия описаны, например, А.М. Ягломом [25]).

Сосуществование абстрактного теоретико-множественного подхода А.Н. Колмогорова и частотной интерпретации Р. Мизеса, ориентированной на эксперимент, приводит и к обратному процессу — переносу допущений, принятых в теоретико-множественных построениях, в сферу действия практической статистики. На деле это означает, например, домысливание ансамбля значений (генеральной совокупности) к ограниченной выборке, полученной в эксперименте. Ясно, что такое домысливание ансамбля к конкретному эксперименту открывает широкие возможности для произвола, если не для злоупотреблений, и далеко не всегда ведет к положительным результатам.

**4.6.** Случайные и неопределенные величины. Итак, в описанную многовыборочную процедуру измерения математического ожидания $M[Y]$ органически входит эмпирическое оценивание устойчивости средних $M_{k,n}[Y]$. В прикладной теории вероятностей случайными теперь принято называть только те непредсказуемые величины, подвыборочные средние которых обнаружили в эксперименте устойчивость. Стало быть, математическое ожидание существует (точнее, считается существующим) не у всякой непредсказуемой величины, а только у случайной величины.

Результат эмпирического оценивания устойчивости средних бывает и отрицательным: интервал в прогнозе (4.9) иногда оказывается слишком большим для того, чтобы можно было переходить к точечной оценке (4.10). Если же $Q$ мало, то дело вообще не доходит до принятия прогноза (4.9). За непредсказуемыми величинами, подвыборочные средние которых не признаны по результатам эксперимента устойчивыми, уже закрепилось название *неопределенные величины* [21, с. 6 — 7; 144,145; 26, с. 24]. У таких величин попросту нет определенного математического ожидания, точнее, его существование представляется сомнительным. Здесь складывается примерно такая же ситуация, как при упоминавшейся выше попытке измерить единственное истинное значение диаметра тела, форма которого, как выясняется в ходе эксперимента, заметно отлична от шаровой.

Схема охарактеризованной разветвляющейся процедуры измерения математического ожидания представлена на рис. 3. Через $U$ обозначены контролируемые условия эксперимента, дополненные описанием способа формирования равноценных подвыборок.

Словом, к неопределенной величине неприменимы многие традиционные

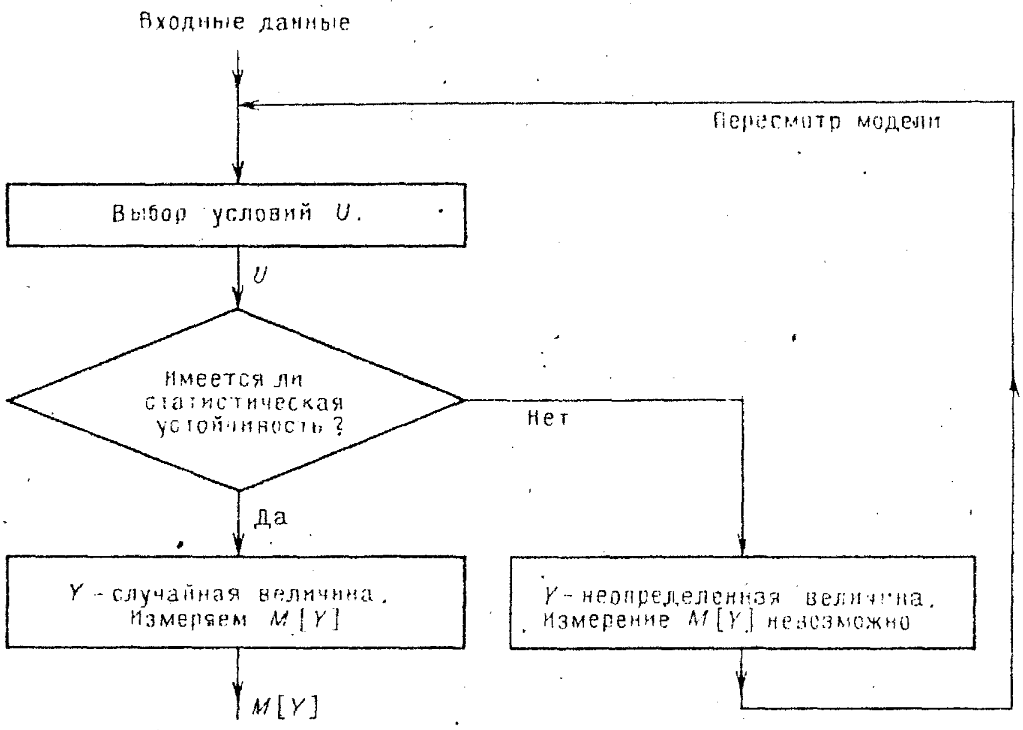


Рис. 3. Процедура измерения математического ожидания $M[Y]$. Процедура предусматривает отказ от измерений, если $Y$ оказывается неопределенной величиной. В этом случае рекомендуется пересмотреть основы модели явления

понятия теории вероятностей и математической статистики, начиная с понятий математического ожидания и генеральной совокупности. Вот как высказывался по этому поводу В.Н. Тутубалин [27, с. 7]: «Все мыслимые эксперименты можно разделить на три группы. К первой группе относятся хорошие эксперименты, в которых обеспечивается полная устойчивость исхода опыта. Ко второй группе относятся эксперименты похуже, где полной устойчивости нет, но есть статистическая неустойчивость. К третьей группе относятся совсем плохие эксперименты, когда нет и статистической устойчивости. В первой группе все ясно без теории вероятностей. В третьей группе она бесполезна. Вторая группа составляет настоящую сферу применения теории вероятностей, но вряд ли мы когда-нибудь можем быть вполне уверены, что интересующий нас эксперимент относится ко второй, а не к третьей группе».

Соглашаясь с приведенным высказыванием по существу, отметим, что и в третьей группе экспериментов (имеющих дело с неопределенными величинами) статистические (выборочные, эмпирические) характеристики все же могут быть достаточно полезными, хотя вероятностные характеристики здесь неприменимы. Статистические характеристики тоже можно относить к разряду теоретико-вероятностных понятий. Все зависит от точки зрения на то, чем и как занимается теория вероятностей (см. п. 2.3).

В литературе встречаются высказывания типа «если испытания однородны, то статистическая устойчивость налицо». Подобное высказывание является осмысленным, очевидно, лишь тоща, когда имеется в виду однородность (т.е. постоянство) *контролируемых условий эксперимента.* Говорить о постоянстве всех условий эксперимента абсурдно.

Однако факты указывают на то, что иногда рад существенных обетоятельств выпадает из поля зрения экспериментатора. И тогда устойчивости — даже статистической — нет, хотя все контролируемые условия эксперимента постоянны.

Такое случается даже в фундаментальной физике, где эксперименты в целом значительно чище, чем в других областях. В лаборатории Резерфорда результаты эксперимента с новым радиоактивным веществом однажды стали форменно хаотичными, хотя условия эксперимента выглядели, как и раньше. Исследователи не сразу догадались, что новое вещество — газ (теперь его называют радоном). Все ранее открытые радиоактивные вещества газообразными не были. Условия эксперимента были дополнительно рафинированы: устранили сквозняки, перестали поблизости курить. После этого статистическая устойчивость восстановилась. Этот эпизод описан в книге Д. Данина «Резерфорд».

Экспериментатор никогда полностью не гарантирован от подобных неожиданностей. В технике (например, в проблемах надежности и управления качеством [26, с. 24 — 27; 16]) неопределенные величины, к сожалению, не редкость. По-видимому, еще чаще с ними приходится иметь дело в экономике и социологии. По данной причине Н. Винер даже исключал эти области из сферы компетенции кибернетики как высоко математизированной «межотраслевой» дисциплины. В своем последнем научном мемуаре [28] он писал: «Успехи математической физики вызывали у социологов чувство ревности к силе ее методов, чувство, которое едва ли сопровождалось отчетливым пониманием интеллектуальных"истоков этой силы... Подобно тому, как некоторые отсталые народы заимствовали у Запада его обезличенные, лишенные национальных примет одежды и парламентские формы, смутно веря, будто эти магические облачения и обряды смогут их сразу приблизить к современной культуре и технике, так и экономисты принялись облачать свои весьма неточные идеи в строгие формулы интегрального и дифференциального исчислений... Как ни труден отбор надежных данных в физике, гораздо сложнее собрать обширную информацию экономического или социологического характера, состоящую *из многочисленных серий однородных данных* [курсив наш. — *Авт.*]. В этих обстоятельствах безнадежно добиваться слишком точных определений величин, вступающих в игру. Приписывать таким неопределенным по своей сути величинам какую-то особую точность бесполезно, и, каков бы ни был предлог, применение точных формул к этим слишком вольно определяемым величинам есть не что иное, как обман и пустая трата времени».

Пример с открытием радона показывает, что *именно выходную характеристику эксперимента — наличие статистической устойчивости,* а не его входную характеристику — стабильность контролируемых условий — приходится считать окончательным критерием однородности испытаний. Другими словами, вполне приемлемы разве что инверсные высказывания типа «если статистическая устойчивость налицо, то испытания однородны». Такие инверсные высказывания ни на что существенное не претендуют. Они лишь вводят еще один термин — «однородность испытаний» — для понятия статистической устойчивости.

При всем том опыт естественных наук и их приложений учит, что главным способом повышения воспроизводимости экспериментального результата является обеспечение стабильности максимально возможного числа условий эксперимента. Иначе к неизбежному рассеянию добавится еще и вклад, обусловленный неряшливостью. Так, по поводу математизации управления качеством [27, с. 5] ярко сказано: «...для применения теории вероятностей к анализу качества продукции необходимо создать сначала хорошо налаженное производство, в котором бы не было ни пьянства, ни прогулов, ни штурмовщины, ни негодного сырья, ни изношенного технологического оборудования и т.п. Теория вероятностей есть нечто вроде масла в каше: сначала надо иметь кашу».

Заметим попутно, что концепция рандомизации эксперимента, характерная для фишеровской математической статистики, не согласуется с этим традиционным взглядом на вещи.

Другое дело, что в прикладных исследованиях контролируемые условия эксперимента целесообразно выбирать на завершающей стадии лабораторного исследования такими, какими они ожидаются при будущей эксплуатации разрабатываемого изделия.

4.7. Промежуточные итоги. Подведем итог сказанному о выборочных средних и математическом ожидании, воспользовавшись метрологическими категориями. Вычисление величины $M_n[Y]$ согласно простой формуле (4.1) и по измеренным значениям $Y(s)$, $s = \overline{1, n}$, первичной величины $Y$ представляет собой разновидность косвенных измерении. Алгоритм (4.1) измерения $M_n[Y]$ отличается, таким образом, четкостью и бесспорностью (если, разумеется, значения $Y(s)$ уже измерены): он в точности совпадает с формальной дефиницией выборочного среднего. Разрабатывать модель объекта измерения здесь не требуется. Алгоритм (4.1) применим к любым величинам $Y$ независимо от того, имеется или отсутствует статистическая устойчивость.

Измерение математического ожидания $M[Y]$, являясь более косвенным, проводится по существенно менее четкому алгоритму, имеющему к тому же ветвящуюся структуру. Этот алгоритм включает эмпирико-индуктивное оценивание устойчивости подвыборочных средних $M_{k,n}[Y]$. Такое многовыборочное оценивание является, в сущности, разработкой модели объекта измерения. Результат этой разработки, как и в любом ином случае, может оказаться отрицательным, т.е. указывающим на неадекватность понятия истинного значения измеряемой величины в данной конкретной экспериментальной обстановке. Здесь подобный результат означает обнаружение неустойчивости подвыборочных средних и соответственно невозможность в данный момент осмысленно приписать математическому ожиданию определенное значение.

Своеобразие математического ожидания как измеряемой величины заключается лишь в том, что разработка модели объекта измерения сводится тут к реализации фундаментальной метрологической установки 2° — в форме оценивания воспроизводимости подвыборочных средних. При измерении других физических величин разработка модели объекта измерения обычно оказывается более специфичной экспериментально-теоретической операцией. Правда, если при измерении математического ожидания включать в разработку модели анализ и выбор экспериментальных условий $U$, способных, по мнению исследователей, обеспечить статистическую устойчивость, то разница вообще исчезнет (выше отмечалось, что окончательную оценку степени однородности испытаний дает многовыборочный эксперимент),

Нам приходится подчеркивать метрологическую ординарность измерений математического ожидания потому, что фишеровская математическая статистика абстрагируется от необходимости разрабатывать модель объекта измерения. Она полагает, что математическое ожидание существует всегда и его всегда можно оценить по результатам одновыборочного эксперимента.

Иногда задаются целью измерить с максимально возможной точностью и надежностью математические ожидания входных или каких-либо промежуточных величин теории (а не только выходной величины $Y$). Все сказанное об измерении математического ожидания полностью относится и к таким случаям, ибо эксперименты здесь по существу носят такой же характер, как при верификации.

## 5. Многовыборочные и фишеровские доверительные интервалы

**5.1.** Многовыборочный доверительный интервал. Доверительный интервал является ключевым понятием фишеровской математической статистики. Используя высказанные выше соображения, сопоставим его с интервалом $[\overline{Y}_{\min}, \overline{Y}_{\max}]$ в (4.9), возникающим при многовыборочном оценивании воспроизводимости средних. Последний можно назвать многовыборочным 100%-ным доверительным интервалом для статистических средних. Если он достаточно мал, из него выбирают численное значение $M[Y]$ математического ожидания. В таком благоприятном случае данный интервал можно назвать доверительным и для математического ожидания. В этом смысле измерение многовыборочного доверительного интервала $[\overline{Y}_{\min}, \overline{Y}_{\max}]$ **является** этапом верификации теоретико-вероятностного прогноза $U \to M[Y]$.

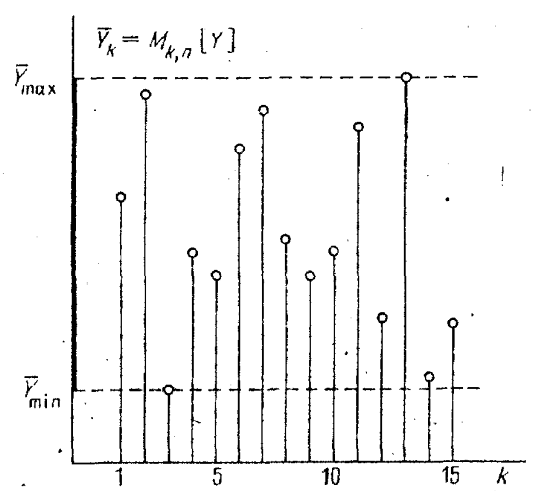


Рис. 4. Многовыборочный доверительный **интервал** $[Y_{\max}, Y_{\min}]$ относится к совокупности $Q$ выборок (в данном случае $Q$ = 15) [29]

На рис. 4 многовыборочный доверительный интервал показан жирной линией на оси $Y$. Вторичная выборка (4.6) представлена решеточным графиком зависимости $\overline{Y}_k = M_{k,n}[Y]$, $k = \overline{1, Q}$. При желании можно ввести, скажем, **90%-ный** многовыборочный доверительный интервал, отсекая по 5% от общего числа точек в верхней и нижней частях решеточного графика,

В соответствии с приводившимся антисхоластическим тезисом естествознания (справедливость теории в конечном счете обосновывается экспериментом, а не другой теорией) идеальный отчет об измерении многовыборочного доверительного интервала мог бы состоять из следующих трех пунктов [12, с. 49, 13 — 18]:

1) при контролируемых условиях $U$ эксперимента (следует подробное описание этих условий) получены $Q$ подвыборок объема $n$;

2) выяснилось, что подвыборочные средние $M_{k,n}[Y]$, $k$= 1, $Q$, лежат в определенном интервале $[\overline{Y}_{\min}, \overline{Y}_{\max}]$; принят эмпирико-индуктивный прогноз, что они будут находиться там и при всех $k > Q$;

3) ближайшие прецеденты: ...здесь желательно провести сопоставление с прежними измерениями многовыборочного доверительного интервала в условиях, аналогичных $U$. Если эти измерения выполнялись при верификации теорий, то следует дать обзор практических результатов, достигнутых с по-

мощью принимавшихся статистических прогнозов.

Например, в случае проверки физической теории целесообразно привести подтверждения полученных данных другими группами исследователей или же указать положительные (отрицательные) результаты родственных экспериментов. В случае прикладных теорий, относящихся к техническим устройствам, речь может идти о сведениях, касающихся безотказной работы или, наоборот, аварий.

Обратим внимание на то, что никакие теоретико-вероятностные гипотезы в подобном отчете не фигурируют. Если верификацию теоретико-вероятностного прогноза $U \to M[Y]$ проводить в рамках опять-таки теоретико-вероятностной модели, то получится логическое зацикливание: возникает надобность в верификации этой модели. Ведь не видно оснований для того, чтобы при условиях $U$ верить вторичной теоретико-вероятностной модели «на слово», коль скоро признается необходимой верификация первичной теоретико-вероятностной модели, доставившей прогноз $U \to M[Y]$.

**5.2.** Фишеровские доверительные интервалы. Обратимся к фишеровскому доверительному интервалу для математического ожидания $M[Y]$. На практике такой интервал

$$[\bar{Y}_N - t(N,\mathscr{P})S_N,\ \bar{Y}_N + t(N,\mathscr{P})S_N] \tag{5.1}$$

вычисляют по всей имеющейся выборке $\{Y(s)\}_1^N$; здесь $\bar{Y}_N \overset{\text{not}}{=} M_N[Y]$ — выборочное среднее, $\mathscr{P}$ — выбранное значение доверительной вероятности,

$$S_N \overset{\text{not}}{=} \Big[\sum_{s=1}^{N}(Y(s) - \bar{Y}_N)^2(N-1)^{-1}\Big]^{1/2} \tag{5.2}$$

— выборочное среднеквадратичное отклонение, $t(N,\mathscr{P})$ — коэффициент Стьюдента.

Пользователи фишеровской математической статистики обычно считают, что вычисленный конкретный доверительный интервал (5.1) накрывает неизвестное математическое ожидание $M[Y]$ с вероятностью $\mathscr{P}$. Метрологические стандарты [5] требуют именно такой трактовки. Скажем, ГОСТ 8.207—76 предписывает рассматривать (5.1) в качестве интервала, «в котором с установленной вероятностью находится суммарная погрешность измерений».

Между тем подобная трактовка интервала (5.1) лишена смысла даже чисто синтаксически. Ведь конкретный, вычисленный по единственной выборке интервал (5.1) либо накрывает, либо не накрывает неизвестное значение математического ожидания $M[Y]$, коль скоро оно считается существующим. Так что вероятность накрытия здесь либо равна единице, либо нулю, и не может равняться $\mathscr{P}$, если $0 < \mathscr{P} < 1$.

На самом деле фишеровские доверительные интервалы типа (5.1) все же имеют определенный смысл. Пусть с помощью независимых испытаний из нормальной генеральной совокупности извлечено $Q \gg 1$ подвыборок $\{Y_{(k)}(s)\}_1^n$, $k = \overline{1, Q}$, объема $n$. Пусть для каждой подвыборки вычислен фишеровский доверительный интервал

$$[\bar{Y}_{k,n} - t(n,\mathscr{P})S_{k,n},\ \bar{Y}_{k,n} + t(n,\mathscr{P})S_{k,n}],\quad k = \overline{1, Q}; \tag{5.3}$$

здесь $\bar{Y}_{k,n} \overset{\text{not}}{=} M_{k,n}[Y]$ — подвыборочное среднее, а $S_{k,n}$ — подвыборочное среднеквадратичное отклонение, вычисленное по формуле, аналогичной (5.2). При сделанных допущениях неравенства

$$\bar{Y}_{k,n} - t(n, \mathcal{P})S_{k,n} \le M_{k,n}[Y] \le \bar{Y}_{k,n} + t(n, \mathcal{P})S_{k,n} \tag{5.4}$$

окажутся справедливыми примерно для $\mathcal{P} \cdot Q$ «хороших» случайных интервалов (5.3), тогда как для остальных, «плохих» интервалов данные неравенства нарушаются. Точнее говоря,

$$\mathrm{Prob}[\bar{Y}_{k,n} - t(n, \mathcal{P})S_{k,n} \le M_{k,n}[Y] \le \bar{Y}_{k,n} + t(n, \mathcal{P})S_{k,n}] = \mathcal{P}. \tag{5.5}$$

Теория не содержит указаний на то, какие именно случайные интервалы из совокупности (5.3) являются хорошими.

Сказанное иллюстрирует рис. 5, заимствованный из работы [29, с. 235]. Как и на рис. 4, кружками показаны значения подвыборочиых средних. Фишеровские доверительные интервалы представлены вертикальными отрезками жирных линий. Эти доверительные интервалы найдены для $\mathcal{P} = 0{,}5$, $Q = 15$. Значение $M[Y]$ показано горизонтальной штриховой линией.

Рис. 4 и 5 иллюстрируют существенное отличие многовыборочного доверительного интервала от фишеровских доверительных интервалов: первый измеряется в единственном числе для всего имеющегося набора подвыборок, тогда как фишеровских доверительных интервалов получается столько, сколько имеется подвыборок.

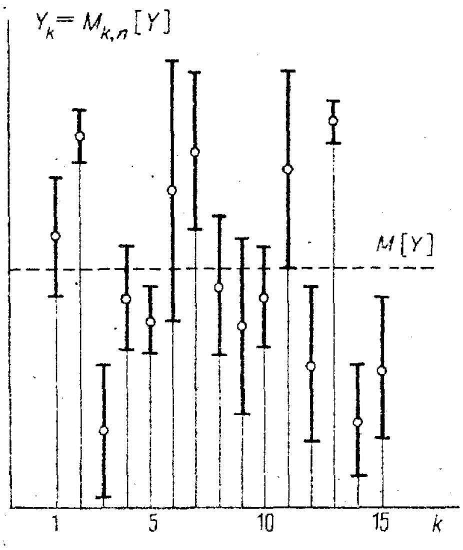


Рис. 5. Фишеровские доверительные интервалы для данных, представленных, на рис. 4. Из данного рисунка видно, что в данном примере многовыборочное среднее $M[Y]$ попадает внутрь фишеровских доверительных интервалов лишь для половины всех выборок. Именно это обстоятельство свидетельствует о неблагополучии в фишеровской процедуре определения доверительных интервалов

На рис. 5 видно, что неравенства (5.4) справедливы примерно для половины фишеровских доверительных интервалов. Значение $M[Y]$ оказалось здесь известным, вероятно, потому, что оно заранее закладывалось в датчик нормально распределенных псевдослучайных чисел, с помощью которого имитировались подвыборки. При натурном эксперименте значение $M[Y]$, естественно, неизвестно. Напомним к тому же, что заранее неизвестно, применимо ли в данной ситуации понятие математического ожидания.

На случайность положения и размера фишеровских доверительных интервалов, вычисленных для добора подвыборок из одной и той же генеральной совокупности, указывают многие источники. Однако наглядная графическая иллюстрация, помогающая до конца осмыслить эти указания, приводится редко.

Показательно, что в справочнике [9] по прикладной статистике, содержащим большое число иллюстративных примеров расчета, отсутствуют примеры в разделе, посвященном фишеровским доверительным интервалам. Расчет индивидуального фишеровского доверительного интервала авторы справочника не стали проводить, видимо, из-за его бессмысленности. Расчет же совокупности доверительных интервалов (5.3) для нескольких подвыборок авторы не привели скорее всего потому, чтобы не вызвать недоумения у пользователей: не ясно, что, собственно, надлежит делать с комплектом доверительных интервалов, подобным представленному на рис. 4 (к тому же вычисление комплекта фишеровских доверительных интервалов ГОСТами не предписывается).

Ход мысли, приводящий к ошибочной интерпретации фишеровского доверительного интервала как некоего фиксированного интервала, в который что-то попадает с вероятностью $\mathscr{P}$, можно проследить, например, по работе [30]. Сначала рассматривают вероятность

$$\mathscr{P} = \mathrm{Prob}(Y_1 \leq Y \leq Y_2) = \int_{Y_1}^{Y_2} w(Y)\,\mathrm{d}Y \tag{5.6}$$

попадания случайной величины $Y$, имеющей плотность вероятностей $w(Y)$, в фиксированный интервал $[Y_1,\ Y_2]$. В формулу (5.6) действительно можно подставить любые конкретные числа $Y_{1,2}$, и при этом она не теряет смысла (вероятность $\mathscr{P}$, вычисляемую по бесспорной формуле (5.6), в [30] тоже именуют доверительной). Затем соотношение (5.5) (в [30] — равенство (36)) считают полностью аналогичным формуле (5.6) и в него подставляют конкретные числовые значения $\overline{Y}_N$, $S_N$, найденные для имеющейся выборки $\{Y(s)\}_1^N$. В итоге в [30] переходят от соотношения типа (5.5), например, к равенству

$$\mathrm{Prob}(31{,}0 \leq M[Y] \leq 31{,}4) = 0{,}86 \tag{5.7}$$

[обозначения наши. — *Авт.*], лишенному смысла, поскольку $M[Y] = \mathrm{const}$.

Вот более простой пример подобной неувязки: если в равенство $\mathrm{Prob}(Y < \mu) = \mathscr{P}$, $0 < \mathscr{P} < 1$, где $\mu$ и $\mathscr{P}$ — константы (может быть, неизвестные), $Y$ — случайная величина, подставить конкретную реализацию данной величины, скажем, $Y$=31,0, то получится лишенная смысла запись $\mathrm{Prob}(31{,}0 < \mu) = \mathscr{P}$. В самом деле, число 31,0 либо меньше, либо не меньше константы $\mu$, так что вероятность выполнения неравенства $31{,}0 < \mu$ равняется либо единице, либо нулю, но никак не числу 0,86.

Рассматривая фишеровские доверительные интервалы, мы столкнулись с характерной чертой фишеровской математической статистики. С одной стороны, основные понятия этой теории являются вполне осмысленными, подобно доверительным интервалам, только в приложении к многим подвыборкам. К одной выборке они неприменимы буквально синтаксически, хотя, увы, и применяются. С другой стороны, интерпретация, которую фишеровские понятия получают в случае многих подвыборок, вряд ли подходит пользователю. Так, едва ли кому-нибудь подойдет целый комплект (5.3) фишеровских доверительных интервалов. В этом и выражается упоминавшаяся недостаточная приспособленность фишеровской математической статистики к многовыборочной схеме верифицирующего эксперимента.

**5.3.** Внелогические компоненты фишеровской статистики. Подведем итоги. Вычисление фишеровского доверительного интервала (ФДИ) для $M[Y]$ основано на следующем комплексе теоретико-вероятностных гипотез (обозначим его через $H$):

А) величина $Y$ случайна, так что ее математическое ожидание существует;

В) более того, существует распределение вероятностей случайной величины $Y$, и притом оно имеет определенный вид: нормальный для непрерывной случайной величины $Y$, биномиальный для двоичной величины-индикатора $Y = I_A$;

С) испытания независимы в теоретико-вероятностном смысле.

В результате верификация прогноза $U \to M[Y]$ расслаивается на три компонента, которые обозначим для четкости изложения символами $U \Rightarrow H$, $H \Rightarrow$ ФДИ, ФДИ $\Rightarrow M[Y]$. Стрелка $\Rightarrow$ обозначает формально-логическую операцию импликации. Принятие комплекса гипотез $H$ на основании анализа условий $U$ — внелогическое. Вычисление фишеровского доверительного интервала в компоненте $H \Rightarrow$ ФДИ выполняется в рамках формальной модели. Принятие окончательной оценки для $M[Y]$ на основании рассчитанного ФДИ (операция ФДИ $\Rightarrow M[Y]$) снова внелогическое хотя бы потому, что таковым является принятие значения $\mathcal{P}$ доверительной вероятности.

Словом, от внелогических компонентов мы, естественно, не ушли, а лишь вклинили между ними формально-логический вывод из отнюдь нетривиальной посылки $H$. Она сама нуждается в верификации, причем верифицируется значительно труднее, нежели прогноз $U \to M[Y]$ (который, кстати, в рассматриваемом контексте только один и интересует нас). К тому же в процедуру аккуратной верификации посылки $H$ обязательно войдет измерение $M[Y]$. Его необходимо выполнить по изложенной ветвящейся процедуре (см. рис. 3) уже при верификации гипотезы А. После такого измерения следовало бы остановиться, так как прогноз $U \Rightarrow M[Y]$ оказался бы, собственно, верифицированным. Между тем оставались бы неверифицированными гипотезы В и С...

Относительно верификации гипотезы В заметим следующее. Пусть нужно надежно установить меру сходства между статистическим распределением выборки $\{Y(s)\}_1^N$ и гипотетическим, допустим, нормальным распределением вероятностей с плотностью

$$w(Y) = \frac{1}{\sqrt{2\pi}\sigma[Y]} \exp[-(Y - M[Y])^2/2\sigma^2[Y]]. \tag{5.5}$$

По здравому смыслу данную задачу невозможно решить, предварительно не оценив по выборке $\{Y(s)\}_1^N$ со всей возможной точностью параметры $M[Y]$, $\sigma[Y]$ сдвига и масштаба плотности (5.5). Это означает, что сначала все-таки необходимо верифицировать гипотезу А (но тогда опять-таки дело будет уже сделано, и гипотезы В и С окажутся излишними).

Фишеровская математическая статистика не принимает во внимание эту естественную иерархию задач оценивания. Она предлагает, можно сказать, кое-как проверять сходство между упомянутыми выборочными и гипотетическим распределениями, а затем считает допустимым основывать на результатах такой грубой проверки дальнейшие достаточно сложные умозаключения теории фишеровских доверительных интервалов. Приведем в данной связи высказывание И. Грековой (литературный псевдоним Е.С. Вентцель — известного специалиста по авиационным приложениям теории вероятностей): для вычисления ФДИ «разработан довольно тонкий аппарат, основанный на допущении, что нам известен закон распределения наблюдаемой случайной величины (нормальный). И опять возникает вопрос: а откуда, собственно, это известно? И с какой точностью? И какова, наконец, практическая ценность самого «продукта» — доверительного интервала? Мало опытов — значит, мало информации, и дело наше плохо. А будет при этом доверительный интервал немного больше или меньше, не так уж и важно, тем более что и доверительная вероятность назначена произвольно» [31, с. 111] (последние фразы в приведенном высказывании намекают на то, что теория ФДИ ориентирована прежде всего на «обсчет» малых выборок).

Сказанное И. Грековой можно дополнить разве лишь напоминанием, что индивидуальный фишеровский доверительный интервал (5.1) вообще не имеет того смысла, который в нем ищут пользователи математической статистики.

Наиболее сложной на поверку оказывается гипотеза С хотя бы потому, что в определении теоретико-вероятностной независимости присутствуют многомерные, а не только одномерные распределения вероятностей. К тому же проверка независимости испытаний оказывается много сложнее, чем теоретико-вероятностное доказательство независимости случайных величин (см., например, [12]).

Более подробный критический анализ фишеровских доверительных интервалов и принципов математической статистики излагался в работах [12 — 18, 23, 24]. Конструктивную альтернативу фишеровской математической статистике составляют приведенные в разделе 5.1 пункты 1 — 3 программы измерения $M[Y]$ [12 — 18, 23, 24].

## 6, Достаточно ли имеющихся данных для надежного прогноза?

**6.1.** Неполнота системы гипотез. Перечислим и по возможности прокомментируем другие внелогические отношения, вовлекаемые в исчисление вероятностей. Начнем с общего вопроса о роли гипотез в вероятностных расчетах.

Можно утверждать, что любая система гипотез не полна. Этот принцип неполноты находит множество подтверждений в жизни: сколь ни были бы мы предусмотрительны и скрупулезны в разработке сценариев поведения сложной системы, все равно можно указать много (даже бесконечно много!) факторов, которые тоже могут воздействовать на результирующее поведение.

Примером неполноты гипотез могут служить практически все неполадки на атомных электростанциях, космических системах и прочих сложных технических устройствах. При оценках надежности АЭС, казалось бы, принимаются во внимание все мыслимые причины отказов и аварий. Оценки дают столь малые вероятности отказов, что приходится только удивляться, как вообще возникают неполадки.

А все дело в том, что исчисление вероятностей производится на базе всегда *неполной* системы гипотез. Показателен пример возникновения пожара на одной из американских АЭС, который приводит П.Л. Капица; причиной послужили: электрическая лампочка, которая перегорела в комнате, где случилась протечка водопроводного крана, и то, что слесарь не нашел ничего лучшего, как зажечь свечку в темном помещении, и тем самым создал очаг пожара.

При строительстве и эксплуатации Чернобыльской АЭС вряд ли пришла бы кому-нибудь в голову мысль о включении в число существенных факторов некомпетентность персонала и его неспособность понять аморальность несанкционированных экспериментов. Кстати, это не единственный пример, когда моральная ответственность служит столь же (если не более) важным фактором безопасности, как и технические характеристики сложных устройств. Каким коротким оказывается путь от моральных принципов к надежности функционирования сложных систем! И как непроизвольно возникают параллели с деятельностью прежних союзных преступных структур, которые не могли поступиться принципами...

**6.2.** Субъективные оценки вероятностей. К субъективным (экспертным) оценкам вероятностей чаще всего прибегают тогда, когда имеется слишком много неопределенных (в тутубалинском смысле) факторов, которые нельзя вложить ни в детерминированную, ни в статистическую модель прогноза.

Субъективные оценки опираются на предыдущий, зачастую практически не поддающийся формализации опыт эксперта. Такие оценки, как правило, бессмысленно подвергать верификации на уровне, принятом в естествознании. Верификации подвергаются скорее практические результаты деятельности — экономические, технические и т.д. Критерии такой верификации обычно размыты, как, впрочем, и прогнозные оценки экспертов, и их трудно включить в естественнонаучную парадигму. Именно поэтому никак не удается (да и нужно ли это?) построить мост между естественными науками, с одной стороны, и астрологами и экстрасенсами — с другой.

**6.3.** Домысливание статистического ансамбля. Восстановление, воссоздание статистического ансамбля (генеральной выборки) по ограниченной экспериментальной выборке представляет собой один из сложнейших вопросов практической теории вероятностей.

Пусть измерены статистические характеристики процесса на интервале $[0, T]$. Вся сложность проблемы восстановления сводится к домысливанию статистических характеристик вне интервала $[0, T]$. В большинстве случаев поступают простейшим способом — применяют принцип «завтра как сегодня», предполагая неизменность статистических характеристик при $0 < t < T$ и при $t > T$.

Разумеется, никто не может гарантировать такую неизменность в течение неопределенно долгого времени, и поэтому всюду, где это возможно и целесообразно, следует постоянно обновлять статистическую информацию и контролировать появление значимых изменений.

Перенесение экспериментально измеренных **временны́х** средних на характеристики статистического ансамбля получило название гипотезы эргодичности. В сущности, свойство эргодичности отражает не более чем нашу веру в правомерность использования **временны́х** средних в качестве параметров гипотетического статистического ансамбля. Любое изменение временных средних должно служить сигналом к пересмотру характеристик статистического ансамбля.

Разумеется, это хлопотно, но если вместо кропотливого анализа свойств реальной системы произнести слова об эргодичности системы, то исследователь как бы получает индульгенцию на случай возможных отклонений от принятой гипотезы и тем самым избавляет себя от необходимости подстраивать воображаемый статистический ансамбль под изменяющиеся условия.

Рассмотрим некоторые опасности, которые ожидают исследователя, «поставившего» на определенный статистический анализ.

**6.4.** Нестационарность. Говоря о выявлении нестационарности статистических характеристик процесса, приходится с самого начала ограничить себя случаем медленной нестационарности. Дело в том, что статистические характеристики нелокальны, они формируются за более или менее длительные промежутки времени.

Пусть, скажем, многовыборочное среднее $M[Y]$ определяется за время $T$ с неопределенностью $\Delta\bar{Y}$. Тогда экспериментально за время $T$ выявляются только нестационарные изменения, не меньшие, чем $\Delta\bar{Y}$. Таким образом, минимальный уровень нестационарности, «разложенный» на интервал $[0, T]$, составляет

$$\min\left|\frac{dM[Y]}{dt}\right| \sim \frac{\Delta\bar{Y}}{T}. \qquad (6.1)$$

В то же время интервал наблюдения $T$ должен быть мал по сравнению с временем нестационарного эволюционного изменения

$$t_{evol} \sim |M[Y]|/|dM[Y]/dt|. \qquad (6.2)$$

При помощи условия $T < t_{evol}$ из (6.1) получаем неравенство

$$\frac{\min\left|\frac{dM[Y]}{dt}\right|}{\left|\frac{dM[Y]}{dt}\right|} \gtrsim \frac{\Delta\bar{Y}}{|M[Y]|}, \qquad (6.3)$$

которое означает, что погрешность определения производной $dM[Y]/dt$ всегда больше, чем погрешность измерения $M[Y]$.

Таким образом, интервал многовыборочного измерения не может быть ни слишком коротким (в этом случае увеличивается погрешность $\Delta\bar{Y}$), ни слишком длинным, иначе можно пропустить искомый эффект нестационарности. Минимальное время измерения нестационарности (оно оценивается из (6.1) при заданном извне значении $\min|dM/dt|$) как раз и налагает ограничения на скорость нестационарности.

В целом этот вопрос еще не получил адекватного освещения в литературе. Отметим только, что принимая гипотезу эргодичности, мы можем приписать (примыслить) статистический ансамбль не только стационарному, но и нестационарному процессу.

**6.5.** Неустойчивости. Неустойчивости особенно опасны для прогнозов, поскольку они сначала скрытно развиваются под уровнем шумов, сопровождающих любое измерение. В первые моменты после зарождения неустойчивости быстро растущая экспонента еще скрывается под шумами. В результате процесс развития неустойчивости становится обнаруженным только через некоторое время после того, как он превысит уровень шума. После этого неустойчивый процесс продолжает стремительно развиваться и достигает макроскопических значений за конечное время, иногда буквально за несколько отсчетов. Соответственно меняются и статистические характеристики исследуемого процесса.

Заблаговременное выявление неустойчивостей той или иной природы представляет собой задачу огромней важности во многих областях науки и техники, скажем, в проблеме управляемого термоядерного синтеза. Иногда неустойчивость носит не экспоненциальный, а взрывной характер, и тогда время развития неустойчивости может быть еще короче. Наконец, упомянем еще процессы типа землетрясений, которые возникают в результате накопления статических напряжений и проявляются в форме кратковременных выделений запасенной энергии.

**6.6.** Редкие явления. Во многих задачах не приходится надеяться не только на воспроизводимость изучаемых процессов или явлений с целью верификации их статистических характеристик, но даже просто на повторное наблюдение. Речь идет о некоторых природных явлениях в космосе (Сверхновые), в океане (необычные течения) и атмосфере (редкие оптические и погодные явления), а также о редких явлениях в лабораторных экспериментах (регистрация редких превращений в физике высоких энергий» регистрация космических частиц большой энергии).

В подобных случаях статистические прогнозы принято строить с соблюдением моральных правил физического экспериментирования, т.е. с максимальной самокритичностью и с анализом всех альтернативных гипотез. Именно благодаря таким негласным правилам анализ даже данных и гипотез, вызывающих законные сомнения, протекает обычно в спокойной атмосфере, с желанием добиться максимальной объективности. Так, или примерно так, проходило обсуждение не подтвердившейся впоследствии (но и не отвергнутой!) гипотезы о наличии массы покоя у нейтрино.

Разумеется, бывают и печальные исключения, как, скажем, в случае с публикациями о холодном ядерном синтезе и о биологическом действии чистой воды.

**6.7.** «Тарелочки». Если в рамках современной научной парадигмы имеются достаточные силы для поддержания «морального здоровья» при верификации даже редких и уникальных явлений, то в околонаучной среде, зацикленной на НЛО, телекинезе, Бермудском треугольнике и экстрасенсах, сплошь и рядом встречаются утверждения, по отношению к которым термин «лженаука» не кажется преувеличением.

Речь идет не столько о смелости обсуждаемых гипотез (смелые гипотезы отнюдь не отвергаются современным естествознанием), а *об уровне их обсуждения*. Именно вольное обращение с гипотезами и полное игнорирование моральных правил экспериментирования отодвигает псевдопроблемы на далекую периферию естествознания, на уровень средневекового мышления[1*]. Разумеется, проблемы метрологическою обеспечения, верификации гипотез и вероятностного истолкования здесь вообще не возникают — добросовестному естествоиспытателю просто нечего здесь верифицировать.

**6.8.** Классические вероятности. Вероятности, отвечающие классическому определению (отношение числа благоприятных исходов к полному числу возможных исходов), с современной точки зрения выступают просто как простейшие гипотезы относительно частоты появления в идеализированных системах — монета, игральный кубик и т.д.

В зависимости от условий проведении эксперимента реальные частоты появления могут быть отличными от классических. Так, фактические результаты зависят от положения центра тяжести игрального кубика, от площади его граней, от сглаженности углов, от качества поверхности стола и от других физических характеристик объекта и эксперимента. Кроме того, как недавно проанализировал Дж. Келлер на примере бросания монеты, результат зависит от начальных линейной и угловой скоростей бросания [33]. Как выяснилось, равновероятность выпадения герба и решетки выполняется только асимптотически, при больших начальных скоростях.

Таким образом, даже классические вероятности иногда требуют экспериментальной верификации, особенно если речь идет о подозрениях в нечестной игре.

**6.9.** Фикции закона больших чисел. В физической литературе до сих пор господствует убеждение, что с ростом числа испытаний (длины выборки) $N$ в силу центральной предельной теоремы относительная частота стремится к своему пределу, который и представляет собой эмпирическую вероятность. Между тем эксперименты сплошь и рядом свидетельствуют о том, что при больших $N$ рассеяние данных не уменьшается, а начиная с некоторого значения $N_0$, наоборот, увеличивается. Наиболее полное и деловое изложение этой проблемы дано П.Е. Эльясбергом [26].

Если не гнаться за точностью формулировок, причина увеличения рассеяния заключается в том, что при $N > N_0$ становится заметной систематическая ошибка, обусловленная тем, что в базовой гипотезе не был учтен какой-либо существенный фактор, который на первых порах, при малом объеме выборки, давал незначительный вклад, а затем, по мере накопления данных, становился все более заметным.

Таким образом, если модель явления не включает какой-то существенный систематический фактор, то увеличение объема выборки вовсе не обязательно влечет за собой уменьшение рассеяния.

Другой возможной причиной неубывания (или недостаточно быстрого убывания) рассеяния может оказаться характер флуктуаций. Условия, предусматриваемые центральной предельной теоремой, хотя и не чересчур обременительны, но все же удовлетворяются отнюдь не автоматически. Поэтому верификация условий применимости центральной предельной теоремы во многих случаях является не только желательной, но просто обязательной процедурой.

Можно констатировать, таким образом, что уменьшение флуктуаций с ростом $N$, рассматриваемое как физический факт, не является тривиальным следствием центральной предельной теоремы, а имеет место только при выполнении определенных условий, требующих специального контроля.

## 7. Алгоритмическая сложность и частичная детерминированность

**7.1.** Физический эксперимент и концепция алгоритмической сложности. Система соглашений возникает не только при измерении вероятностей, но даже на еще более ранней стадии — на этапе определения понятия случайности. Теоретико-множественный подход относит к случайным величины, снабженные вероятностной мерой. Прикладная теория вероятностей выделяет класс случайных величин по признаку устойчивости статистических характеристик.

Алгоритмическая теория вероятности [34, 35] отождествляет случайность с алгоритмической сложностью. Наконец, в теории частично-детерминированных процессов случайность трактуется как непредсказуемость [36]. Как видим, даже в вопросе о том, что называть случайным, имеется по крайней мере несколько соглашений (более полный список см. в [36]).

Выше мы уже обсудили взаимоотношение между эмпирическими и теоретико-множественными средними. Затронем кратко соотношение *с* алгоритмической теорией вероятности и теорией частичной детерминированности.

Вопреки ожиданию, алгоритмическая теория вероятности собственно исчислением вероятности не занимается. Ее задача установить критерии случайности, трактуемой как алгоритмическая сложность последовательности чисел (результаты измерений как раз и представляют собой последовательности чисел). В контексте данной статьи алгоритмическая теория вероятности интересна тем, что она явно связывает случайность с алгоритмами, т.е. в конце концов с гипотезами и домысливанием.

Сложность определяется как наименьшая длина алгоритма $l_{\min}$ (количество операций), переводящего одну последовательность чисел $\{x\}$ в другую $\{y\}$. Отметим, что здесь возникает новая условность, связанная с существованием *множества* гипотетических алгоритмов, переводящих $\{x\}$ в $\{y\}$, и с необходимостью поиска кратчайшего из них.

Если длина алгоритма $l(N)$ при $N \to \infty$ остается конечной, точнее, если

$$\frac{l(N)}{N} \underset{N\to\infty}{\to} 0, \tag{7.1}$$

то последовательность объявляется алгоритмически простой и, стало быть, не случайной. Если же $l(N)/N \to \mathrm{const}$, то мы имеем дело с *алгоритмически сложной* и тем самым — со *случайной* последовательностью.

Несмотря на всю привлекательность концепции случайности как алгоритмической сложности, предложенной А.Н. Колмогоровым [34] и развитой его последователями, в целом она вряд ли представляет интерес для проблемы физических измерений. Во-первых, эта концепция имеет дело с пределом $N \to \infty$ и требует осуществления бесконечно удлиняющихся тестов на выявление сначала простых, а потом все более и более сложных алгоритмов. Такая последовательность удлиняющихся тестов должна была бы в конце концов приблизиться к «универсальному тесту» Мартин-Лёфа [35]. Ясно, что такая операция тестирования на все мыслимые и немыслимые алгоритмы практически неосуществима, тогда как сама концепция алгоритмической сложности принципиально требует перехода к пределу $N \to \infty$.

Во-вторых, концепция случайности как алгоритмической сложности противоречит тому взгляду на природу явлений, который развивает физика. Естествоиспытателю трудно признать последовательность (процесс) случайной, если ее алгоритм известен, хотя и сложен.

В-третьих, имеется едва ли преодолимая трудность, связанная с наличием шумов при измерениях. При физических измерениях, как правило, производится фильтрация (дискриминация) шумов, иначе объектом измерения будут шумы, а не исследуемый процесс. При алгоритмическом подходе, если его приложить к измерениям, различие между шумом и измеряемым процессом не производится, во всяком случае нам такие попытки не известны. В итоге даже при относительно простом алгоритме самого исследуемого процесса смесь «сигнал + шум» приобретает сложность шума.

Таким образом, логика физического эксперимента в существенном едва ли совместима с концепцией алгоритмической сложности.

**7.2.** Физический эксперимент и концепция частичной детерминированности. В той части, которая касается выявления динамических (детерминированных) закономерностей в исследуемом процессе, логика эксперимента ближе к концепции частичной детерминированности [36]. Последняя выносит суждение о детерминированности или случайности наблюдаемого процесса $y(t)$ по степени его сходства с модельным, прогностическим процессом $z(t)$.

Количественной характеристикой сходства может служить степень детерминированности [36]

$$D(\tau) = \frac{\{y(t)\cdot z(t)\}}{(\{y(t)\cdot y(t)\}\{z(t)\cdot z(t)\})^{1/2}}, \tag{7.2}$$

где $\tau = t - t^0$ — время, протекшее после начала наблюдения, а фигурные скобки означают операцию сравнения (проектирования). При сравнении наблюдения $y(t)$ с модельным процессом $z(t)$ предполагается, что начальные условия для $z(t)$ такие же, как для зарегистрированного процесса $y(t)$ а именно

$$z(t^0) = y(t^0). \tag{7.3}$$

Благодаря этому при $\tau = 0$ степень детерминированности (7.2) равна единице, как бы ни была определена операция сравнения $\{\cdot\}$.

Равенство $D = 1$ отвечает *полной детерминированности* (полной предсказуемости) наблюдаемого процесса $y(t)$ относительно модельного процесса $z(t)$. В противоположность этому обращение $D$ в нуль интерпретируется как *полная случайность* (непредсказуемость) $y(t)$ по отношению к $z(t)$. Значения $0 < |D| < 1$ описывают *частичную детерминированность*. Интервал, в течение которого степень детерминированности $D$ превышает определенный уровень, скажем, $D \geq 1/2$, характеризует время детерминированного поведения или, что в данном контексте то же самое, время предсказуемости $\tau_{pred}$.

В оригинальных работах [36, 37] в качестве операции, сравнения была выбрана операция статистического (эмпирического) усреднения произведения $y(t)z(t)$:

$$\{y(t)\cdot z(t)\} = \langle y(t)z(t)\rangle, \tag{7.4}$$

так что степень детерминированности $D$ представляет собой коэффициент корреляции между $y(t)$ и $z(t)$. Статистическое усреднение можно скомбинировать с временным интегрированием,

$$\{y(t)\cdot z(t)\} = \int_{t^0}^{t^0+\tau} \langle y(t)z(t)\rangle dt, \tag{7.5}$$

и тогда мера (7.2) характеризует не локальное, как (7.4), а *интегральное* на всем отрезке $[t^0, t^0 + \tau]$ сходство между наблюдением и прогнозом.

Концепция частичной детерминированности формализует фактически существующие отношения между экпериментатором и экспериментальным материалом. Экспериментатор выдвигает те или иные гипотезы $z(t)$ и проверяет на них экспериментальные данные $y(t)$. В реальном мире ни одна модель $z(t)$ не может претендовать на бесконечное время предсказания, и поэтому для *любых физических процессов*

$$\tau_{\text{pred}} < \infty.$$

Благодаря этому мы имеем возможность несколько иными глазами взглянуть на алгоритмический подход. Прежде всего с физической точки зрения бессмысленно подвергать процесс испытанию на бесконечно длинные тесты — достаточно ограничиться конечными временами $\tau < \tau_{\text{pred}}$.

Во-вторых, не нужно гнаться за универсальностью (по Мартин-Лёфу, это все мыслимые тесты и все те, что могут предложить последующие поколения людей). Гораздо практичнее ориентироваться на имеющиеся тесты ( $\equiv$ гипотезы, модели), т.е. оценивать сходство наблюдения $y(t)$ не со всем мыслимым множеством гипотез, а только с гипотезами, фактически имеющимися у экспериментатора.

Наконец, в-третьих, концепция частичной детерминированности радикально решает проблему шумов, всегда присутствующих в наблюдении $y(t)$, именно, операция сравнения (7.5) включает в себя *фильтрацию*. Удельный вес тех компонентов наблюдения $y(t)$, которые обязаны своим происхождением шумам, при длительном накоплении с весом $z(t)$ уменьшается, как, впрочем, и тех компонентов $y(t)$, которые не согласуются с принятой моделью $z(t)$. Поэтому многократное накопление данных позволяет выделить сигнал (т.е. «осмысленную» часть $y(t)$) из-под шумов. В алгоритмическом же подходе, как мы видели, шумы не подвергаются осмысленной фильтрации. С этими существенными дополнениями концепция частичной детерминированности может рассматриваться как развитие алгоритмического подхода на реальные физические объекты исследования.

**7.3.** Эмпирическая вероятность как степень детерминированности. Концепция частичной детерминированности оказывается весьма гибкой и универсальной. Как свидетельство этой гибкости укажем, что при определении операции $\{y(t)\cdot z(t)\}$ как *числа совпадений* между значениями $y(t)$ и $z(t)$ степень детерминированности (7.2) превращается в *эмпирическую вероятность.*

Чтобы убедиться в этом, произведем дискретизацию отсчетов и по времени ($s$-й отсчет берется в момент времени $s\Delta t$ после $t^0$), и по величине: значения $y$ и $z$ берутся с дискретностью $\varepsilon$. Определим операцию сравнения $\{y\cdot z\}$ как число $\varepsilon$-совпадений между $y$ и $z$ в пределах коридора шириной $\varepsilon$. Его можно выразить как число событий $A$, состоящих в том, что разность $y - z$ по модулю не превышает $\varepsilon/2$:

$$\{y\cdot z\} = n_A(\tau), \text{ где } A\colon |y - z| < \varepsilon/2. \tag{7.6}$$

Если ввести индикаторную функцию $I_A(s)$, определенную соотношением (4.3), число $\varepsilon$-совпадений $n_A(\tau)$ на интервале $[t^0, t^0 + n\Delta t]$ выразится суммой

$$\{y(t)\cdot z(t)\} = n_A(\tau) = \sum_{s=1}^{n} I_A(s), \quad \tau = n\Delta t. \tag{7.7}$$

Поскольку при $y = z$ всегда $I_A(s) = 1$ и $\{y\cdot y\} = \{z\cdot z\} = n,$ степень детерминированности (7.2) превращается в отношение числа совпадений $n_A(\tau)$ к общему числу отсчетов за время $\tau = n\Delta t$:

$$D(\tau) = \frac{n_A(\tau)}{n} = \frac{\sum_{s=1}^{n} I_A(s)}{n} = \omega_n(\varepsilon). \qquad (7.8)$$

Отношение (7.8), как легко видеть, представляет собой относительную частоту $\omega_n(\varepsilon)$ совпадений между $y$ и $z$ в пределах коридора $\varepsilon$, т.е. эмпирическую вероятность $p$.

Несмотря на существование непосредственной связи $D(\tau)$ с выборочной вероятностью $p$, величина $D$, как мера сходства между $y$ и $z$, обладает все же некоторой дополнительной гибкостью, которая состоит в том, что значения $z_s = z(t_s) = z(t^0 + s\Delta t)$ в процессе сравнения могут меняться вместе с изменением наблюдаемой величины $y_s = y(t_s) = y(t^0 + s\Delta t)$. Если модельное (предсказуемое) значение постоянно, $z = z_*$ (в этом случае нужно отказаться от требования (7.3) на начальное условие $z(t^0)$), то величина (7.8) представляет собой просто относительное время (относительную частоту) пребывания наблюдаемой величины $y(t_s)$ в **$\varepsilon$-окрестности** фиксированной величины $z_*$:

$$D(\tau) = \omega_* = \omega(|y - z_*| < \varepsilon/2). \qquad (7.9)$$

Предположим теперь, что у нас есть возможность предсказывать значения $y_s$, т.е. имеется удовлетворительный закон (правило, алгоритм, догадка, конспиративные сведения и т.д.) построения прогноза $z_s$. Тогда **$\varepsilon$-совпадение между** всеми членами последовательностей

$$y_1, y_2, \ldots, y_n \quad \text{и} \quad z_1, z_2, \ldots, z_n \qquad (7.10)$$

будет обозначать, что

$$D(\tau) = 1. \qquad (7.11)$$

Сравнивая значения (7.9) и (7.11), можно углядеть существование скрытых условностей, казалось бы, в простейших операциях, типа подсчета числа событий. Речь идет о сравнении частоты определенного фиксированного события, скажем, выпадения шестерки в случае игрального кубика, с частотой совпадения между наблюдаемой величиной $y_s$ и (переменным) прогнозом $z_s$. Если продолжить пример с кубиком, то в случае «честного» кубика при достаточно большом числе бросаний $n$ частота выпадения шестерки ($z_* = 6$) будет приближаться к 1/6, Это означает, что числа из наблюдаемой последовательности $y_1, \ldots, y_n$ в среднем один раз из шести совпадают с $z_* = 6$.

Иначе обстоит дело в случае кубика, управляемого, скажем, при помощи магнита, когда манипулятор организует известные ему значения $y_s$. Если эти значения берутся из таблицы случайных чисел, то совпадение с фиксированным значением $z_* = 6$ по-прежнему будет наблюдаться примерно в одной шестой части бросаний. Объективный (или, правильнее сказать, честно наивный) наблюдатель, судящий с позиций «честного» кубика, скажет, что результаты бросаний отвечают его интуитивным ожиданиям. В то же время манипулятор, т.е. информированный наблюдатель, для которого результат бросания $y_s$ заранее известен, будет предсказывать результат, основываясь на «нечестной» модели $y_s = z_s$, где $z_s$ — это число, устанавливаемое им же. В этом случае

$$D(\tau) = 1.$$

В промежуточном случае, когда манипулятор не вполне управляет результатом, величина $D$ может лежать между значениями $\omega_* = 1/6$ и $D = 1$.

Манипуляция может состоять также в использовании "нечестного" кубика, для которого $\omega_* > 1/6$. Тогда стратегия «нечестного» игрока сведется к более частному предсказанию шестерки. Интересно, что в случае ложной информации, получаемой манипулятором от управляющего устройства, предсказания могут дать значения $D < 1/6$. Например, если устройство «вредное», т.е. *всегда* дает не то, что от него требуется, то вообще $D = 0$.

Мы намеренно привлекли в качестве иллюстрации основной мысли пример из азартных игр. Дело в том, что труд естествоиспытателя во многом подобен работе дешифровщика, который из множества получаемых сообщений пытается извлечь закономерности, действующие в природе. Получаемую информацию экспериментатор осмысливает на основе тех или иных гипотез. Только что рассмотренный пример показывает, что объект, демонстрирующий «случайное» поведение (вероятность выпадения грани равна 1/6), все же может подчиняться какому-то алгоритму, детерминированному закону поведения.

Здесь мы еще раз сталкиваемся с многозначностью терминов. В данном случае *детерминированность,* трактуемая как подчинение процесса более ниш менее сложной закономерности, не противоречит случайности, трактуемой как равновероятность выпадения любой грани игральной кости.

В общем плане это означает глубокое различие между *внутренними* вероятностными характеристиками процесса $y(t)$ (математическое ожидание $M[y]$, высшие моменты $M[y_i y_j y_k \ldots]$) и *внешними* характеристиками, выявляемыми при сравнении $y(t)$ с модельными процессами $z(t)$.

Как это ни покажется странным, физики интересуются не только внутренними (в указанном выше смысле), но очень часто еще и внешними характеристиками, т.е. интересуются *степенью соответствия* между наблюдением и модельным процессом (теорией). Обсуждаемая здесь мера сравнения (7.2) (с корреляционной (7.4), интегральной (7.5) или вероятностной, точнее, "совпадательной" (7.7) операциями сравнения) как раз и отвечает требованиям эксперимента.

Операция сравнения при интерпретации эксперимента всегда подразумевает наличие гипотезы, модели и даже догадки (внелогическое, по Фейнбергу, начало в естествознании). Это вполне отвечает центральной теме нашей статьи — важнейшей роли гипотез и домысливания при обработке данных эксперимента. Напомним еще, что именно «Искусством угадывания» (Ars conjectaridi) Якоб Бернулли назвал свою первую в мире книгу по теории вероятностей [38].

## 8. Заключение

Сформулируем основные выводы из проведенного рассмотрения в прямолинейной школьной форме:

1) Относительная частота появления $\omega_N$, трактуемая как статистическая

вероятность, а также статистическое математическое ожидание $M_N[Y]$ или $\boldsymbol{M_{k,n}[Y]}$ являются *нормальными физическими величинами* (т.е. их можно измерить и оценить их погрешность), когда мы имеем дело со случайными (в смысле Тутубалина) величинами, которые, в отличие от неопределенных величин, обладают устойчивыми статистическими характеристиками.

2) «Ненормальность» эмпирической вероятности и математического ожидания заключается в том, что они больше, чем другие физические величины, нагружены условностями и гипотезами, которые требуют специальной проверки (верификации).

3) В фишеровской математическом статистике имеются гипотезы, не подлежащие экспериментальной проверке (или не выдерживающие ее) и даже синтаксическому анализу, что свидетельствует о ее несостоятельности(2*).

4) Традиционной и интуитивной приемлемой альтернативой фишеровской статистике служит многовыборочная обработка данных ([12 — 14]; разделы 4 и 5), которая опирается на разумный минимум допущений.

5) Домысливания вовлекаются в практическое исчисление вероятностей во многих случаях, в том числе:

— при перечислении (и умолчании!) факторов, влияющих на надежность сложных систем;

— при субъективной оценке вероятностей;

— при восстановлении статистического ансамбля по ограниченной экспериментальной выборке;

— при допущении свойства эргодичности;

— при прогнозировании в условиях нестационарности и неустойчивости;

— при интерпретации редких явлений;

— при употреблении классических вероятностей как моделей физических явлений;

— при привлечении закона больших чисел к анализу физических явлений.

## ПРИМЕЧАНИЯ

(1) Публикацию В.Л. Гинзбурга на эту тему в газете «Известия» за 1991 г. следует признать весьма уместной.

(2) Интересно, сколько времени понадобится российским властям на исправление ГОСТа, рекомендующего фишеровскую процедуру? Управятся ли они до конца этого тысячелетия?

## СПИСОК ЛИТЕРАТУРЫ


[1] *Вигнер Е.* Непостижимая эффективность математики в естественных науках//УФН. 1968. Т. 94. С. 535.

2. *Крылов А.Н.* Мои воспоминания. — Л.: Судостроение. 1979.

3. *Фейнберг Е.Л.* Интеллектуальная революция на путях единения «двух культур»//Вопр. философии. 1986. № 8. С. 33.

4. Основы метрологии и электрические измерения/Под ред. Е.М. Душина. — Л.: Энергоатомиздат, 1987.

5. Основополагающие стандарты в области метрологического обеспечения. — М.: Госкомстандарт, 1983.

6. *Кендалл М., Стьюарт А.* Теория распределений. — М.: Наука, 1966.

7. *Кендалл М., Стьюарт А.* Статистические выводы и связи. — М.: Наука, 1973.

8. *Кендалл М., Стьюарт А.* Многомерный статистический анализ и временные ряды. — М.: Наука, 1976.

9. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.

10. *Айвазян С.А, Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Исследование зависимостей. М.: Финансы и статистика, 1985.

[11] *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989.
12. *Алимов Ю.И.* Альтернатива методу математической статистики. М.: Знание, 1980.
13. *Алимов Ю.И.* Утилитарная логика построения теории вероятностей/Семиотика и информатика. — М.: ВИНИТИ. 1985. — № 24. С. 58.
14. *Алимов Ю.И.* Оценка фишеровской теории оценивания с позиций метрологии//Статистическая обработка экспериментальных данных. — Новосибирск: Новосиб. электротехн. ин-т связи, 1986. — С. 15.
15. *Алимов Ю.И.* Измерение моментов системы случайных величин. — Свердловск: Изд-во УПИ, 1984.
16. *Алимов Ю.И.* Измерение спектров и статистических вероятностей. — Свердловск: Изд-во УПИ, 1986.
17. *Алимов Ю.И.* Прогнозирование распределений вероятностей. — Свердловск: Изд-во УПИ, 1986.
18. *Алимов Ю.И,, Шаевич А.Б.* Методологические особенности оценивания результатов количественного химического анализа//Ж. аналит. химии. 1988. Т. 44. С. 1983.
19. *Уиттл Н.* Вероятность. — М.: Наука, 1982.
20. *Миэес Р., фон.* Вероятность и статистика. — М.; Л.: Госиздат, 1930.
[21] *Тутубалин В.Н.* Теория вероятностей. — М.: Изд-во Моск. ун-та, 1972.
22. *Колмогоров А.Н.* Основы теории вероятностей. — М.: Наука, 1974.
23. *Алимов Ю.И.* О проблемах приложения теории вероятностей, рассмотренных в работах В.Н. Тутубалина// Автоматика. 1978. № 1. С. 71.
24. *Алимов Ю.И.* Еще раз о реализме и фантастике в приложениях теории вероятностей//Автоматика. 1979. № 4. С. 83.
25. *Яглом А.М.* Корреляционная теория стационарных случайных функций. — Л.: Гидрометеоиздат, 1981.
26. *Эльясберг П.Е.* Измерительная информация: Сколько ее нужно? Как обрабатывать? — М.; Наука, 1983.
27. *Тутубалин. В.Н.* Теория вероятностей в естествознании. — М.: Знание, 1972.
28. *Винер Н.* Творец и робот. — М.: Прогресс, 1966.
29. *Смирнов Н.В., Дунин-Барковский И.В.* Курс теории вероятностей и математической статистики. — М.: Наука, 1965.
30. *Зайдель А.Н.* Погрешности измерений физических величин. — Л.: Наука, 1985.
131] *Грекова И.* Методологические особенности прикладной математики на современном этапе ее развития//Вопр. философии. 1976. №6. С. 104.
32. *Капица П.Л.* Теория, практика, эксперимент. — М.: Наука, 1984.
33. *Keller J.B.* The probability of heads/Mm. Math. Monthly. 1986. V. 93. P. 191.
34. *Колмогоров А.Н.* On the tables of random numbers//Sankhya Indian J. Statist. Ser. A. 1963. V. 25. P. 369; Три подхода к определению понятия «количество информации»//Пробл. передачи информации. 1965. Т. 1, № 1, С. 3; то же: *Колмогоров А.Н.* Теория информации и теория алгоритмов. — М.: Наука, '987. — С. 204 и 215 (обе статьи).
**35**. ***Martin-Löf P.*** The definition of random sequences//Inform. and Control. 1966. V. 9. P. 602.
36. *Кравцов Ю.А.* Случайность, детерминированность, нредсказуемость//УФН. 1989. Т. 158. С. 92.
37. *Кравцов Ю.А.* Случайность и предсказуемость динамического хаоса//Нелинейные волны. *Т. 2:* Динамика и эволюция. — М.: Наука, 1981. — С. 81.
38. *Бернулли Я.* О законе больших чисел. — М.: Науки, 1986. — Ч. 4.

Статья поступила 6.03.92 г.