Июль 1993 г. Том 163, № 7

УСПЕХИ ФИЗИЧЕСКИХ НАУК

МЕТОДИЧЕСКИЕ ЗАМЕТКИ

ВЕРОЯТНОСТЬ, КОМПЬЮТЕРЫ И ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

В.Н. Тутубалин

(Московский государственный университет им. М.В. Ломоносова)

(*Статья поступила 15.02.93 г.*)

СОДЕРЖАНИЕ

1. Введение (93).
2. Понятие научной пользы (94).
3. Благодать исторического развития и теория вероятностей (95).
4. Верна ли классическая модель? (98).
5. Понята ли модель теории ошибок в современной практике преподавания и научных исследований? (99).
6. Польза вероятностно-статистических методов обработки (100).
6.1. Философское введение: парадигма прикладной статистики.
6.2. Пример взаимодействия эксперимента и статистической обработки на уровне физического практикума.
7. Заключение: основные принципы преподавания теории ошибок (107).
Список литературы (109).

1. Введение. Поводом для написания данной статьи послужило знакомство автора с очень интересной статьей Ю.И. Алимова и Ю.А. Кравцова "Является ли вероятность "нормальной" физической величиной?" [1]. Вообще-то, я в основном согласен с критическими взглядами на применения вероятностных методов, высказанными в этой статье. Но имеются определенные расхождения в общетеоретических посылках, которые ведут к существенным различиям в практических выводах. Вкратце они заключаются в следующем.

Грешно, конечно, сомневаться в том, что в фундаментальной физике вероятность является "нормальной" физической величиной, т.е. настолько истинной и/или благородной, насколько вообще бывают физические величины. Существует также целый мир других физических явлений, отличный от мира фундаментальной физики, в котором естественно возникают статистические ансамбли тех или иных процессов и операция усреднения по ансамблю. С помощью корреляций, вычисляемых усреднением по ансамблю, в этом мире возможен новый подход к случайности и детерминированности (см. конец работы [1], а также [2, 3]). Конечно, и в этом мире вероятность является нормальной физической величиной.

Но существует еще третий мир — мир ошибок измерений и наблюдений, которому, в основном, и посвящена работа [1]. К этому миру понятия и методы теории вероятностей прилагаются уже 200 лет. Как мне кажется, исторический опыт совершенно определенно показывает, что в этом мире вероятность *не является* нормальной физической величиной, так как не обладает в должной мере ни истинностью, ни благородством. Если оценивать вероятностно-статистические методы обработки результатов эксперимента с точки зрения физической парадигмы, то их все придется отбросить как недостаточно надежные. Между тем, я совершенно убежден и постараюсь показать в данной статье на примере, что эти методы чрезвычайно полезны при правильном к ним отношении. Иначе говоря, отказ от применения многих понятий и методов математической статистики (который предлагается в [1]), предлагается заменить подходом в рамках иной парадигмы, которую я и пытаюсь как-то описать в данной статье.

На тему о применениях теории вероятностей я уже писал довольно много, а недавно издательство МГУ даже доставило мне редкую по нынешним временам роскошь публикации книги [4]. Но в самые последние годы мне выпало — даже не знаю, как сказать: счастье или роскошь? — поработать статистическими методами в новой для меня области — клинической физиологии. Эта область оказалась, как это ни странно в наше время, совершенно девственной в смысле понимания и применения статистических методов: конечно,

© В.Н. Тутубалин 1993

каждое экспериментальное среднее сопровождается его стандартной ошибкой, но трактуются эти ошибки либо неправильно, либо вообще никак; а на самом деле они несут ценнейшую информацию, позволяющую наметить дальнейшие исследования. В результате, хотя мое мнение об истинности и благородстве вероятностных методов осталось прежним, об их полезности — сильно возвысилось. В этом и состоит смысл новой публикации.

Стержнем дискуссии предлагается сделать вопрос о том, что из теории вероятностей и математической статистики должно включаться в вузовское преподавание. Я не раз убеждался в том, что под именем "теории ошибок" или "теории обработки результатов эксперимента" студентам может преподноситься довольно причудливая смесь утверждений, некоторые из которых неверны математически, другие — бессмысленны, а третьи не подтверждаются опытом исторического развития науки. Первопричина этого состоит, как мне кажется, в том, что вероятностная теория ошибок действительно задумывалась Гауссом и Лапласом как физическая теория, но в качестве физической испытания временем не выдержала, однако не была и отброшена, так как много раз доказывала свою полезность. Поэтому получился некий разброд и недоумение в мыслях, не говоря уж о неизбежной человеческой ограниченности, которую предлагается рассмотреть в некоторых деталях.

2. Понятие научной пользы. Согласно словарю Даля мудрец — это тот, кто соединяет истину с пользой. Как всякий человек, я, конечно, хочу быть мудрецом, хотя и знаю, что мне это не удастся, конкретно — хочу соединить горькую истину о том, что вероятностные модели в вопросах обработки наблюдений не бывают адекватными, с тем убеждением, что они весьма полезны. Что же понимается под пользой для науки?

Я затрудняюсь в выборе термина — о философии или о религии идет речь. Если термин "философия" брать в его первоначальном смысле — "мудролюбие", то это будет философия. Но термин "философия науки" занят под обозначение ряда научных дисциплин, каждая из которых отвечает на свои собственные вопросы, но только не на вопрос о том, что есть благо и польза. Если, с другой стороны, брать слово "религия" в смысле Л.Н. Толстого, как учение о том, что есть благо, то это будет религия; однако Л.Н. Толстой считал науку таким не стоящим внимания пустяком, что просто оскорбился бы применением термина "религия науки". Во всяком случае, я с интересом вижу, что не могу избежать калькирования христианской догматики, и это есть еще одно доказательство того, что в глубине своей человеческое мышление едино для всех времен.

Итак, что есть аналог первородного греха для человека, занимающегося наукой? Ответ: конечно, то обстоятельство, что знания, умения, опыт, навыки каждого отдельного человека составляют лишь ничтожную долю в общем объеме научных знаний, короче — грех глупости. Глупость наша обнаруживается неумолимо и постоянно, стоит только данному человеку чуть выйти за пределы им освоенного. Личный интеллект бывает, конечно, различен. Например, я много раз убеждался в том, что А.Н. Колмогоров соображал на порядок быстрее, чем окружающие его математики, тоже ведь не лишенные интеллекта. Но ведь были и у А.Н. Колмогорова поступки, о которых лучше всего возможно скорее и прочнее забыть.

А что такое благодать для ученого? Ну, конечно, исторический ход развития науки, в процессе которого истинное утверждается, а ложное отвергается и забывается. Ведь не будем же мы спорить о том, что современные научные теории более глубоки, чем существовавшие прежде, а факты более обширны, и с позиций современности мы можем довольно правильно судить о прошлом. (Мы вскоре применим этот принцип к теории вероятностей.) Но эта благодать представляет собой настоящее чудо, поскольку прямо противоречит нашему повседневному опыту. Действительно, каким же образом грех глупости преодолевается и искупляется в науке путем коллективных усилий, если из повседневного опыта мы отлично знаем, что уж если в любом коллективе — от детского сада до Ученого совета и от Ученого совета до Верховного — возникает дискуссия, то она ведется на уровне самого глупого (т.е. наименее компетентного в данном вопросе) ее участника?!

Молчание, т.е. неучастие в дискуссии, есть часто не только золото, но и добродетель. Знаменитый П.Л. Чебышёв обычно не выступал на заседаниях Академии наук (см. [5], с. 275), кроме, разумеется, тех случаев, когда его просили дать научный отзыв. Но незадолго до смерти интеллект, видимо, ему несколько изменил, и он участвовал в обсуждении следующего вопроса (см. [5], с. 288). В России тех времен закон требовал, чтобы диссертация была напечатана в определенном числе экземпляров за счет диссертанта. Диссертанты попросили, чтобы в тех случаях, когда основой диссертации являлась статья, опубликованная в изданиях Академии наук, автору выдавались не обычные 50 оттисков, а 100,

чтобы их можно было считать публикацией диссертации. Поскольку типографский набор и вся подготовка к печати все равно производились, то стоить дополнительные оттиски могли сущие пустяки, и по тону протокола, опубликованного в [5], можно предположить, что готовилось положительное решение. Но П.Л. Чебышев спросил: "... Сообразно ли будет с достоинством Академии...", если она будет вообще думать о такой меркантильной ерунде, как защита диссертаций? Академики физико-математического отделения этому удивились и перенесли рассмотрение вопроса на общее собрание. В результате решили "не входить в рассмотрение, какое назначение желают давать авторы" их публикациям. Получился срам, состоявший в том, что простая и разумная просьба была отклонена как несообразная с достоинством Академии.

Короче говоря, любить в науке самих себя мы не можем, ибо грешны до мерзости (как и в христианской догматике), и остается нам любить только таинственную и чудесную благодать исторического развития, благодаря которой каждое следующее поколение ученых отбирает и удерживает истину из всех сложных дискуссий предшествующего поколения, как бы особенно и не вдаваясь во все сложности этих дискуссий. Роль вероятностных методов обработки информации в этом процессе заключается в том, что с их помощью перерабатываются и вводятся в научный оборот сравнительно большие массивы наблюдательных данных (мы увидим, что два-три десятка чисел — это уже большой массив), которые иначе осмыслить невозможно. Верно, что при этом сравнительно велик риск неправильных выводов, но если об этом помнить и уповать на благодать исторического развития, то этот риск, пожалуй, не столь *уж* важен.

3. Благодать исторического развития и теория вероятностей. Возможно, имеет смысл напомнить структуру основных математических понятий теории вероятностей вообще и вероятностной теории ошибок в частности. Лаплас представлял себе случайное событие в виде некоего извлечения шара из урны. Случайная величина — это функция от шара (т.е. каждому находящемуся в урне шару поставлено в соответствие некоторое число). Насколько я могу судить, у Лапласа не было понятия плотности распределения случайной величины, которая может принимать любые вещественные значения.

Через сто лет после Лапласа его урна была значительно усовершенствована математиками: теперь в урну можно положить все вещественные числа, а также векторы, матрицы, кватернионы (и что угодно) и даже функции вещественной или иной переменной. Равновероятный выбор из подобных множеств объектов теряет смысл (впрочем, Лаплас нисколько не настаивал на равновероятности), но для того, чтобы математически корректно думать о каком-либо случайном выборе, следует обратиться к математической теории меры. Математическая кухня теории меры и измеримых функций кажется многим слишком экзотической и действительно является таковой, по крайней мере, в отношении меры на пространстве функций, т.е. теории случайных процессов (подробнее см. [4]). Но я лично думаю, что эта кухня неизбежна не только в образовании профессиональных математиков, но также и в образовании физиков и вообще в сколько-нибудь расширенном математическом образовании. Действительно, забудем о теории вероятностей и подумаем о том, что такое самосопряженный оператор. Мы должны определить его (в простейшем случае), допустим, на пространстве функций, суммируемых с квадратом. Но в это пространство придется включить все измеримые функции, иначе оно не будет полным, а это чрезвычайно досадно. Итак, если мы хотим обучать студента квантовой механике, то ему следует знать, что такое измеримая функция. Но тогда и при изложении теории вероятностей естественно пользоваться этим понятием и мы приходим к тому или иному варианту аксиоматики Колмогорова. Посмотрим, как выглядит с этой точки зрения вероятностная модель ошибок измерений.

(Здесь следует сделать одну оговорку. В настоящее время многими авторами теоретически и практически изучаются модели ошибок в виде зависимых случайных величин, т.е. в виде случайных процессов (см., например, [6, 7]). На мой взгляд, однако, достигнутые успехи не настолько значительны, чтобы могла идти речь о внедрении подобных результатов в студенческое образование. Очень часто при обработке фактических данных у нас бывает слишком мало информации, чтобы модель случайного процесса для ошибок могла быть полезной. Поэтому в данной статье рассматривается лишь модель независимых случайных ошибок.)

Мысль о том, что измерения той или иной величины следует несколько раз повторить, чтобы извлечь какую-то информацию о точности каждого отдельного измерения, является древней. Во времена Лапласа, в частности, на примере астрономических и геодезических измерений было

отлично известно, что при высокой точности измерений результаты повторных измерений никогда в точности не повторяются, а обычно колеблются хаотически, напоминая результат последовательного извлечения из некоей таинственной урны шаров с написанными на них числами. Итак, каждое наблюдение x_i имеет вид

$$x_i = a + \delta_i, \quad i = 1, ..., n, \tag{1}$$

где a — неизвестное нам истинное значение измеряемой величины (в простейшем случае, который мы скоро оставим, величина a одна и та же для всех и наблюдений), δ_i — ошибка i-го наблюдения. Именно ошибки колеблются хаотически. Скажем условно, что Лаплас и Гаусс сделали решительный шаг, объявив ошибки наблюдений случайными величинами (более корректно было бы сказать, что в эпоху Лапласа и Гаусса в науке сложился этот взгляд). Современный вариант мыслится следующим образом. Имеется некоторое пространство элементарных событий Ω с вероятностной мерой на нем. Под элементом $\omega \in \Omega$, т.е. отдельным элементарным событием, многие статистики понимают прямо результат наблюдения (или значение ошибки), но, по моему мнению, это неортодоксально, а правильно под ω понимать набор всех мыслимых параметров, влияющих на ошибку наблюдения, которая, таким образом, оказывается функцией от ω, т.е. случайной величиной. Спрашивается, конечно, зачем нужно думать о всех таких параметрах, о которых обычно мы имеем лишь самые смутные представления? Но дело в том, что вторым важнейшим элементом концепции Лапласа и Гаусса является предположение о статистической независимости ошибок отдельных наблюдений. Если ошибка δ_1 первого наблюдения есть функция от набора мешающих параметров ω_1, а ошибка δ_2 второго наблюдения есть функция от набора ω_2, то мы фактически предполагаем, что между первым и вторым наблюдением прошло столь много времени, что наборы ω_1 и ω_2 статистически независимы. Итак, если думать об Ω так, как предлагается здесь, то мы совершенно не удивимся тому, что к радиолокационным наблюдениям, которые могут делаться очень часто, классическая теория ошибок неприменима (см. [6]). О наборе же всех n наблюдений мы должны думать как о функции на n-кратном прямом произведении вероятностного пространства на самого себя (подробнее см., например, [4]). Наконец, третьим практически не подлежащим изменению элементом классической концепции является предположение об отсутствии систематических ошибок: считается, что математическое ожидание $\mathbf{M}\delta_i = 0$. Часто еще высказывается предположение о нормальном распределении вероятностей для ошибок, но в общем-то мало что изменится, если предположить, что все ошибки имеют одинаковое распределение вероятностей, хотя бы и не нормальное.

Что вытекает из всех перечисленных предпосылок, если они верны? Вытекает, притом если не прямо на математическом уровне строгости, то на близком к нему (и превышающем требования физической парадигмы), ряд весьма замечательных вещей. Например, в качестве оценки для неизвестной нам величины a мы берем, естественно, $\bar{x} = \sum x_i/n$. Желаете знать, как велика при этом может быть ошибка? Пожалуйста: вычислим

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}.$$

Тогда с вероятностью 0,95

$$|\bar{x} - a| \le \frac{1,96s}{n^{1/2}}, \tag{2}$$

а если желательна надежность не 0,95, а более высокая, скажем, 0,99, то нужно просто заменить 1,96 на 2,57. Если (2) записать в виде

$$\bar{x} - \frac{1,96s}{n^{1/2}} \le a \le \bar{x} + \frac{1,96s}{n^{1/2}},$$

то мы получаем так называемый *доверительный интервал,* т.е. интервал со случайными концами (зависящими от результатов наблюдений), который ловит неслучайное, но нам неизвестное значение a с вероятностью 0,95. Доверительными интервалами именно в таком (современном для нас) понимании широко пользовался уже Лаплас. В 1845 г. в магистерской диссертации Чебышева вычислен по данным Кавендиша доверительный интервал для средней плотности Земли с надежностью 0,9924794 (см. [5], с. 85). Поэтому авторы [1] не вполне правы, называя доверительные интервалы фишеровскими. Стьюдент и Фишер в начале нашего века внесли лишь небольшие уточнения в теорию и практику вычисления доверительных интервалов, существенные лишь при малом числе наблюдений (менее десяти) и при нормальном распределении ошибок.

Прежде чем обратиться к более сложной ситуации, следует подчеркнуть следующую основную особенность классической теории ошибок, равно как и любых других методов статистической обработки информации: для их применения совершенно не нужно знать, каким методом проводились измерения, да и, вообще, что именно измерялось: достаточно знать лишь сами числа x_1 ..., x_n, В этом основная сила (в одних ситуациях)

и основная слабость (в других ситуациях) статистических методов.

Неправильно было бы сводить классическую теорию ошибок (как это иногда легкомысленно делается) к обработке измерений одной и той же величины. Главное творение Лапласа и Гаусса в области обработки результатов эксперимента — это метод наименьших квадратов. Это, как мне кажется, и есть необходимый минимум статистического образования студентов физических и других специальностей с расширенной математической подготовкой. Дело в том, что уже в начале прошлого века (и тем более теперь) основной интерес представляли не так называемые "прямые" наблюдения, когда в каждом эксперименте непосредственно измеряется значение какой-то интересующей нас величины, а измерения косвенные, когда интересующие нас величины входят в качестве параметров в непосредственно измеряемые зависимости. Например, астронома интересуют параметры орбиты того или иного тела Солнечной системы, а измеряет он углы, характеризующие его положение на небесной сфере в ряд моментов времени. Геодезиста интересует длина дуги земного меридиана, а измеряет он стороны и углы некоторой триангуляции. Далее в этой статье приводится пример, когда автора интересовали электродвижущая сила и внутреннее сопротивление некоторого источника тока, а измерял он вольтамперную характеристику. Основная черта косвенных измерений состоит в том, что имеются некоторые законы природы, т.е. уравнения, которым должны были бы удовлетворять измерения, если бы они были совершенно точными, а основная идея Гаусса и Лапласа заключается в том, что по "невязкам" или "остаткам", т.е. по тому, насколько нарушаются уравнения для реальных измерений, можно:

1) оценить величину ошибки отдельных наблюдений;

2) подобрать наиболее правильные значения параметров, входящих в исследуемые зависимости, оценив точность их определения;

3) понять, действительно ли достаточно точно выполняются рассматриваемые законы природы или они нуждаются в корректировке.

Требуется, конечно, чтобы число наблюдений было больше числа оцениваемых параметров. Основная вычислительная идея состоит в том, что сначала небольшое число (по предположению, достаточно высоко точных наблюдений) используется для приближенной оценки значений параметров, а затем уравнения, выражающие законы природы линеаризуются в окрестности этих оценок и тем самым произвольные уравнения заменяются линейными. В результате возникает следующая математическая модель метода наименьших квадратов:

$$x_i = a_i + \delta_i, \quad i = 1, \ldots, n, \tag{3}$$

где a_i — точное значение величины, измерявшейся в i-ом опыте, δ_i — ошибка этого измерения. Использованные законы природы в каждом случае поразному, но всегда довольно естественным и несомненным образом переходят в предположение, что вектор $a = (a_1, \ldots, a_n)$ принадлежит некоторому известному подмногообразию n-мерного пространства. Например, для модели (1) — это, так сказать, "биссектриса", выражаемая уравнениями $a_1 = a_2 = \ldots = a_n$. Ошибки δ_i предполагаются независимыми случайными величинами с $\mathbf{M}\delta_i = 0$. Что же касается их дисперсий $\mathbf{D}\delta_i$, то классики заложили следующую весьма интеллигентную модель:

$$\mathbf{D}\delta_i = \frac{\sigma^2}{w_i},$$

где σ^2 — некоторый параметр, подлежащий определению по невязкам (только один неизвестный параметр для дисперсий), w_i — *известные* каким-то образом числа, называемые весами наблюдений. В простейшем случае (например, для модели (1)) веса принимаются равными (тогда их можно принять равными единице), но вообще-то веса определяются, исходя из каких-то представлений или статистических данных о величине некоторых первичных для измерений ошибок и условий линеаризации законов природы.

Математика, связанная с методом наименьших квадратов, может быть в вычислительной части довольно сложной и громоздкой. К счастью, для выполнения вычислений теперь можно воспользоваться математическим обеспечением компьютеров, а принципиальная часть может быть изложена просто. Первое такое изложение на русском языке дано А.Н. Колмогоровым [8] (оригинальная публикация в 1946 г.). Можно пойти еще чуть дальше и почти полностью освободить изложение от математических формул, используя лишь простые факты из линейной алгебры (см., например, [4]). Однако в 1946 г. в работе [8], адресованной широкой читательской аудитории, можно было излагать теорию метода наименьших квадратов, ничего не сказав о том, верна или нет положенная в его основу модель. Теперь же, почти через пятьдесят лет, накоплены новые факты (см. для примера [6,7]), и предполагать по умолчанию, что здесь все в порядке, уже совершенно невозможно. Я собираюсь рассмотреть

следующие вопросы:

1) Верна ли для реальных ошибок классическая вероятностная модель теории ошибок?

2) Понята ли хотя бы эта модель в современной практике преподавания и научных исследований?

3) В чем, тем не менее, может состоять ее польза?

4. Верна ли классическая модель? Не надо, конечно, представлять себе дело так, что классики думали, что теория ошибок применима к любым измерениям. В широком движении за совершенствование качества измерений нечего делать с одной лишь математической обработкой: прежде всего нужно совершенствовать средства измерений, в чем участвуют и физика, и химия, и материаловедение; на определенном этапе включается метрология; и не мешает, конечно, хорошо обучать наблюдателей. О вероятностной модели можно говорить в применении к тем ошибкам, которые все же остаются после того, как доступные меры по улучшению приборов и подготовке наблюдателей уже приняты. Классический подход предполагает, что при хороших наблюдениях основную роль начинают играть случайные ошибки. Однако если встать на ту точку зрения, с которой физика судит о правильности тех или иных теоретических моделей, то неизбежно придется признать, что классическая модель, бесспорно задуманная как физическая, не выдержала испытания временем. Вкратце аргументация состоит в следующем.

С общетеоретической точки зрения физик вряд ли может признавать какие-то субъективные или иные вероятности и предпочитает говорить о вероятности в тех случаях, когда имеется некий ансамбль экспериментов, обладающий определенной статистической устойчивостью (см., например, [1]). За 200 лет, прошедших со времен Лапласа и Гаусса, наука не добилась продвижения в фундаментальном вопросе — когда возникает статистическая устойчивость. Узнать, так обстоит дело или нет, мы можем только из эксперимента, и совершенно правильно авторы [1] настаивают на тщательном рассмотрении протоколов экспериментов. Казалось бы странно, что статистик начинает свою работу с рассмотрения протокола, не вдаваясь в экспериментальную методику, но если учесть грех глупости, то становится ясным, что этот классический ход задуман весьма тонко. Действительно, происходит передача экспериментального материала от одного специалиста (физика-экспериментатора), который научился измерять на данном приборе, но не может подвергнуть свои данные какой-то совокупной обработке, к другому специалисту (по математической статистике), который владеет методами обработки, но непременно сожжет электрический тестер, как только он попадет ему в руки, так как по рассеянности скоро включит его неправильно. Анализ протоколов измерений — это что-то вроде приемо-сдаточной процедуры.

Но есть все основания полагать, что в большинстве случаев эта процедура закончится печально: статистическая устойчивость будет отвергнута. Возьмем пример геодезического характера: пусть измерения состояли в том, что с площадки одного геодезического знака измерялся угол между направлениями на два другие знака. Злые языки говорят, что существует явление горизонтальной, или боковой, рефракции, которое состоит в том, что из-за горизонтального градиента показателя преломления воздуха луч света может отклоняться в горизонтальном направлении на ту или иную величину, в зависимости от погодных условий. В таком случае два протокола измерений, сделанных при разной погоде, дадут достаточно резко различающиеся результаты, и статистическая устойчивость будет отвергнута. Дело тут не в ошибках прибора и/или наблюдателя, а в том, что исследуемый объект (угол между двумя лучами зрения) вообще не существует с той точностью, с какой мы хотели бы его изучать. Осознав проблему, мы определяем истинный угол как среднее значение угла по всем погодным условиям. Такое среднее можно определить из наблюдений, если наблюдать в течение нескольких сезонов по многу раз в сезон, но как знать, может быть, за это время земная кора несколько подвинется? Теперь можно попытаться усреднить по движениям земной коры, но на это, наверно, потребуется несколько столетий. Можно, пожалуй, сформулировать основное практическое разногласие между мнением авторов работы [1] и моим: допустим, для какой-то геодезической сети оказалось, что ситуация именно такова, как здесь описано — нет статистической однородности; имеет ли смысл уравнивать наблюдения методом наименьших квадратов? Насколько я понимаю, авторы [1] ответили бы на этот вопрос определенно отрицательно; я же считаю, что уравнивать стоит — не ради доверительных интервалов для углов сети (в которых, действительно, мало смысла), а в надежде определить, для каких именно треугольников наиболее существенна боковая рефракция.

Модель теории ошибок, как и любую другую, можно проверять и по выводам из нее. Доверительные интервалы вида (2) иногда блестяще подтверждаются последующими более точными

наблюдениями. Например, доверительный интервал Чебышева для средней плотности Земли (5,48 г·см$^{-3}$ ± 0,1) действительно содержит (и с большим запасом) современное нам значение 5,52. Но чаще доверительные интервалы, получаемые в ходе все более точных измерений той или иной величины, представляют восхитительную картину: они в самом деле все укорачиваются, но никакой следующий не содержится ни в каком предыдущем, а лежит в несколько другом месте числовой оси.

Кроме предпосылки случайности, т.е. статистической однородности, классическая модель содержит еще предпосылки независимости ошибок и определенных соотношений между их дисперсиями (в простейшем случае дисперсии должны быть равны), а также предпосылки отсутствия систематической ошибки. Приведем некоторые фактические сведения из книги [7], из которых вытекает, что для электроизмерительных приборов подобные вещи невозможны.

Для электроизмерительных приборов существует метрологическая служба, которая их периодически поверяет (путем сравнения с более точными приборами) и регулирует, чтобы удерживать ошибку в заданных пределах (которые нормируются в процентах наибольшего деления шкалы — не от измеряемого значения!). Согласно [7], выявляемые таким методом ошибки в первом приближении являются линейными функциями от измеряемой величины, параметры которых меняются со временем (коррекция нуля, возможность которой предусматривается конструкцией прибора, скорее вредна, чем полезна). Когда в процессе периодических поверок обнаруживается, что наибольшая ошибка прибора (которая обычно достигается где-то вблизи верхнего деления шкалы) близка к выходу за заданный процент, прибор регулируют с большим запасом, чтобы он мог прослужить еще несколько лет до следующей регулировки. Польза компьютера (если данные все равно вносятся в компьютер) могла бы состоять в том, что регулировка прибора заменялась бы введением в компьютер таблицы поправок. В любом случае истинные значения ошибок зависят как от значения измеряемой величины, так и от времени, прошедшего со времени предыдущей поверки (а также от дополнительных мешающих факторов, которые могут иметь место при использовании прибора в сравнении с условиями поверочной лаборатории).

Короче говоря, с точки зрения физической парадигмы за вероятностной моделью ошибок наблюдений вряд ли можно признать иной статус, чем статус некоего мифа. Лет сто назад понятия "наука" и "миф" радикально противопоставлялись: наука считалась благородным делом, а миф — мерзким. Но сейчас мысль о некотором глубоком единстве всех проявлений человеческого мышления все больше входит в силу, так что вряд ли кто-нибудь обидится, если, например, назвать современную космологию вариантом мифа о сотворении мира, отвечающим уровню науки (и вообще культуры) XX в. По моим наблюдениям, для студенческой молодежи сейчас подобная философия вполне приемлема, и я говорю в своих лекциях о теории ошибок примерно следующее.

Человек не может думать о любых вещах и/или действовать с ними, не создавая для себя каких-то идеальных моделей их сущности. У антропологов это называется анимизмом, т.е. наделением душой. Например, "душа охотника бежит на душе лыж по душе снега за душой лося". В этом смысле такие математические понятия, как случайные величины, независимость, математическое ожидание и т.д. являются душой, которой мы наделяем конкретные результаты наблюдений, чтобы получить возможность работать с ними. Но в XX в. мы хотим получать из обработки наблюдений научные, т.е. сравнительно достоверные и по меньшей мере заслуживающие внимания, результаты. Поэтому настоящая задача изучения теории ошибок заключается не в том, чтобы выучить те или иные формулы и рекомендации по обработке наблюдений (можно считать, что это уже сделал для вас ваш компьютер), а в том, чтобы научиться понимать, что из результатов обработки, выдаваемых компьютером, заслуживает доверия и в какой степени.

5. Понята ли модель теории ошибок в современной практике преподавания и научных исследований? Нет, не понята. Например, мне приходилось видеть, как в физическом практикуме преподаватель стремился создать некую всеобъемлющую теорию ошибок, в которой любые ошибки, например

1) неточность значения той или иной величины, взятой с недостаточным числом знаков из справочника,

2) систематическая ошибка прибора,

3) ошибка, допущенная студентом при отсчете длины или времени, — все характеризовались одинаково, каждая своей "сигмой", которые складывались бы по правилу "корень квадратный из суммы квадратов". В каком-то большом ансамбле наблюдений, например, всех студентов и всех задач практикума, нечто похожее, быть может,

и возможно, но все-таки, наверно, мало интересно, какой может быть ошибка выполнения задачи практикума, усредненная по всем наблюдениям, всем студентам и всем задачам практикума. В рамках же отдельного студента и отдельной задачи я смысла в таком подходе не вижу.

В качестве другого примера обратимся к преподаванию классической науки геодезии, для нужд которой и создавалась прежде всего теория ошибок (информацию я извлекал из учебника [9]). В геодезии очень много математических вопросов, в частности вычислительных вопросов применения методов наименьших квадратов, и ни на что другое сил уже не остается. Вообще нет упоминания о том, что есть математическое понятие случайной величины и независимости, в рамках которых только и имеет смысл само уравнивание: лишь бы довести до счастливого конца саму процедуру уравнивания с неизвестно откуда взявшимися весами наблюдений. Между прочим, в наше время вполне возможно с помощью радиодальномера измерить довольно большое количество расстояний между узлами той или иной триангуляционной сети. Каждое новое измерение расстояния добавляет не одно лишнее соотношение для уравнивания, а много, потому что известные отрезки можно связывать с помощью весьма различных цепей, состоящих из треугольников сети. У меня создалось впечатление, что рекомендуется измерять поменьше длин, чтобы не усложнять уравнивание. Также непонятно, как в модели весов наблюдений сопоставляется вес измерения угла и вес измерения расстояния: в методе наименьших квадратов положен лишь один неизвестный параметр для дисперсий. В общем, как мне кажется, в геодезии надо было бы стремиться не столько к спокойствию души, вытекающему из того обстоятельства, что уравненные значения длин и углов, наконец, не противоречат геометрическим теоремам, сколько к более полезным приложениям модели ошибок, например к выделению тех наблюдений, ошибки которых, по-видимому, велики, чтобы их при возможности повторить.

Если теперь говорить об общенаучной судьбе теории ошибок, то, конечно, за 200 лет стало ясно, что вероятностная модель может быть и неверна. Но общепринятая альтернативная модель достаточно убога: она сводится к тому, что ошибки наблюдений δ_i могут иметь отличные от нуля математические ожидания: $\mathbf{M}\delta_i = c_i$, где c_i и называются систематическими ошибками. По вычитании систематических ошибок должны опять остаться независимые одинаково распределенные случайные величины (случайные ошибки). При этом делать c_i произвольно зависящими от i нельзя, так как такая модель будет бесполезной (все что угодно объясняется систематическими ошибками). Проще всего их сделать равными: $c_1 = \ldots = c_n = c$, и тогда c по-английски называется inaccuracy, а по-русски — систематической ошибкой (в единственном числе), Величина $(\mathrm{D}\delta_1)^{1/2} = \ldots = (\mathrm{D}\delta_n)^{1/2} = \sigma$ по-английски называется imprecision, а по-русски — сигма, или стандартное отклонение. Сложилась парадигма, согласно которой любые данные наблюдений должны быть прилично снабжены их inaccuracies и inprecisions. Но ведь после всего того, что было сказано (на основании [7]) о том, какими в реальности могут быть ошибки наблюдений, совершенно ясно, что систематические и случайные ошибки в действительности не существуют. И однако такая форма вероятностного мифа тоже может быть весьма полезной при обработке больших массивов информации.

Если я предлагаю признать за вероятностной моделью статус мифа о душе результатов наблюдений, то надо быть последовательным и за статистическими процедурами обработки информации признать статус гадания. Случайность и гадание, как мы отлично понимаем, в глубинной психологии тесно связаны: можно гадать по результатам любого случайного (в смысле непредсказуемости) "эксперимента": по картам и костям, по полету птиц и внутренностям жертвенных животных и т.д. Не мешает отметить, что и здесь всегда существовала своя специализация: одни гадали по полету птиц, другие по внутренностям — ни один ведь человек не может освоить всего сразу. В наше время при обработке экспериментальной информации гадают по числам, представляющим собой результаты наблюдений, и это есть искусство, отдельное от самого искусства измерений. Но в XX в. мы хотим гадать на научном уровне, и пора сказать, каким образом это может быть возможным и полезным.

6. Польза вероятностно-статистических методов обработки.

6.1. Философское введение: парадигма прикладной статистики. На протяжении примерно двенадцати лет в Московском университете существовало некое любопытное образование, в конце концов назвавшееся Межфакультетской лабораторией статистических методов. Замыслил ее (вероятно, по образцу Индийского статистического института) и заведовал ею А.Н. Колмогоров, а заместителем заведующего был В.В. Налимов. Лаборатория в конце концов была

расформирована: статистический институт не получился по некоторым глубоким причинам, анализировать которые я вряд ли способен, но интересно, что в ней происходило. Дело в том, что В.В. Налимов в момент приглашения в лабораторию занимался прикладной статистикой, в частности планированием эксперимента, основанным как раз на классической модели ошибок наблюдений, но затем его интересы стали меняться. Пока это была наукометрия, т.е. анализ количества научных работ и числа их цитирований, это было вполне терпимо для статистической лаборатории. Но когда пошли индийская философия, трансперсональная психология, вероятностная модель языка... короче, ряд тогда молодых сотрудников возмутился такими неблагородными интересами и начал атаку на занятия подобного рода. Атака, впрочем, окончилась лишь тем, что А.Н. Колмогоров постановил, что В.В. Налимов, как достаточно крупный ученый, имеет право заниматься, чем хочет. Но теперь, через двадцать с лишним лет, я вижу с полной ясностью, что прикладная деятельность в области математической статистики может быть успешной только в том случае, если математика Колмогорова сочетается в ней с философией Налимова.

Философия здесь понимается в смысле того, чего мы ожидаем от использования тех или иных методов и что считаем хорошим и плохим. Философия прикладной статистики самого А.Н. Колмогорова сводилась, в общем, к физической парадигме. Им даны и замечательные образцы работ, удовлетворяющих этим требованиям, из которых наиболее известными являются работы по статистической теории турбулентности. Но очень большое число работ А.Н. Колмогоров просто браковал на том основании, что они "не вышли", так сказать, вероятность не оказалась "нормальной" физической величиной. Из этих работ, никогда не увидевших публикации, мне особенно жалко работу по солнечным пятнам, потому что предварительный анализ проблемы и постановка задачи, сделанные на семинаре А.Н. Колмогоровым, произвели на меня неизгладимое впечатление. В общем, философия А.Н. Колмогорова, которой в молодости я, разумеется, слепо следовал, теперь кажется мне излишне жесткой.

Философию В.В. Налимова в полном объеме я излагать не берусь, но в сравнительно тривиальной части, касающейся философии науки, она сводится к тому, что по отношению к исследуемому объекту научная работа обычно может быть только метафорой, которая в одних отношениях похожа, а в других и совсем не похожа на объект. Так, вероятностная модель теории ошибок, так сказать, на микроуровне ошибок отдельных наблюдений на объект явно непохожа. Но на макроуровне, т.е. в рамках довольно большого массива экспериментальной информации она может давать некоторое среднее описание объекта в виде стандарта случайных ошибок и/или систематической ошибки. Узнать, является ли это описание достаточно надежным, т.е. научным, можно, разумно разделив имеющуюся информацию на несколько частей и посмотрев, устойчивы ли значения вычисляемых характеристик при переходе от одной части к другой, либо (если возможно) обработав теми же приемами новую информацию: должны получиться примерно прежние характеристики. Польза же статистического исследования, в общем, зависит от обратной связи с экспериментатором: если результаты статистической обработки подсказывают экспериментатору какие-то новые идеи для более совершенного ведения эксперимента, то это очень полезно. Возможны любые сочетания надежности, т.е. устойчивости результатов, статистического описания с его полезностью; например, описание может быть ненадежным, но полезным. Скажем, если по методике [1] обнаружилась статистическая неустойчивость протоколов эксперимента, то это может быть весьма полезным для улучшения его постановки. Статистическая обработка может также помочь заметить (или сделать более убедительным) какой-то новый эффект, т.е. дать выигрыш, равносильный приобретению более точного прибора.

Благодаря внедрению компьютеров теперь не нужно создавать вычислительное бюро для проведения вычислений. Однако правильно истолковать результаты статистического анализа (хотя бы применения метода наименьших квадратов) может лишь человек, понимающий вероятностные модели, из которых вытекают алгоритмы обработки: правильное истолкование нельзя просто купить как часть пакета статистических программ. Таким образом, издержки, связанные со статистической обработкой, состоят не в приобретении компьютера (он, наверное, уже и так есть), а в привлечении нового человека.

6.2. Пример взаимодействия эксперимента и статистической обработки на уровне физического практикума. Как-то мне понадобился для некоторых бытовых целей источник постоянного тока. Я собрал его из того, что было под рукой: трансформатор попался 220/24 В огромной для моих целей

мощности 250 В•А, нашлись и диоды для диодного моста, дроссель (сопротивление постоянному току я измерил примерно 70 Ом, индуктивность 4,2 Гн), а конденсатор попался вообще прекрасный — на 22000 мкФ. В целом получился прибор весом почти 8 кг, и я стал думать, каковы же свойства такого чуда и начал снимать вольтамперную характеристику. Тут я сообразил, что хоть это и не нужно для бытовых целей, но неплохо было бы посмотреть, что дадут статистические методы при обработке такого материала, и результат, как мне кажется, получился интересный для учебных целей.

Результаты всех измерений сведены в табл. I. Первые три столбца R, I и $V1$ были получены следующим образом. По радиолюбительскому тестеру ТЛ-4М сопротивление реостата, имитировавшего нагрузку (и подключаемого параллельно конденсатору фильтра), устанавливалось на то или иное значение (столбец R), затем тестер в режиме амперметра включался последовательно с нагрузкой, и измерялась сила тока (столбец I); наконец, тестер переключался в режим вольтметра, подключался параллельно реостату, и измерялось падение напряжения на нагрузке (столбец $V1$), Для начала считалось, что измерительный прибор не вносит искажений. Данные столбцов I и $V1$ представлены точками на рис. 1.

Получив такую вольтамперную характеристику, я удивился (ожидалась прямая линия) и стал думать, каковы мои возможности теоретического анализа явления. Когда-то меня учили, что диодный мост из синусоидальных колебаний напряжения переменного тока (с выхода трансформатора) делает модуль синусоиды, а затем нужно составить с помощью законов Кирхгофа систему линейных дифференциальных уравнений с постоянными коэффициентами, описывающую RCL-цепь, причем в качестве правой части в одно из уравнений подставить модуль синусоиды. Но я не измеряю никаких переходных режимов (оказалось из-за большой бойкости конденсатора, что в моей цепи эти режимы при каждом новом включении цепи длительны — до минуты): при измерениях я дожидаюсь установившегося режима. Изучение установившегося режима, как мне известно, упрощается: нужно правую часть разложить в ряд Фурье, и прохождение каждой гармоники через фильтр изучать отдельно. Однако мой измерительный прибор в режиме постоянного тока вообще не чувствует никаких гармоник, в чем можно убедиться, включив его прямо в сеть в режиме вольтметра постоянного тока. Поэтому я могу принять, что на вход моего фильтра действует только постоянное напряжение E, равное интегральному среднему от модуля синусоиды, и ничего больше. Но тогда закон природы, который я исследую, обязан иметь вид

$$V = E - rI, \tag{4}$$

где V и I — напряжение и сила тока во внешней цепи, r — сопротивление дросселя постоянному току. Кроме того, должен выполняться закон Ома для внешней цепи

$$V = RI, \tag{5}$$

где R — внешнее сопротивление.

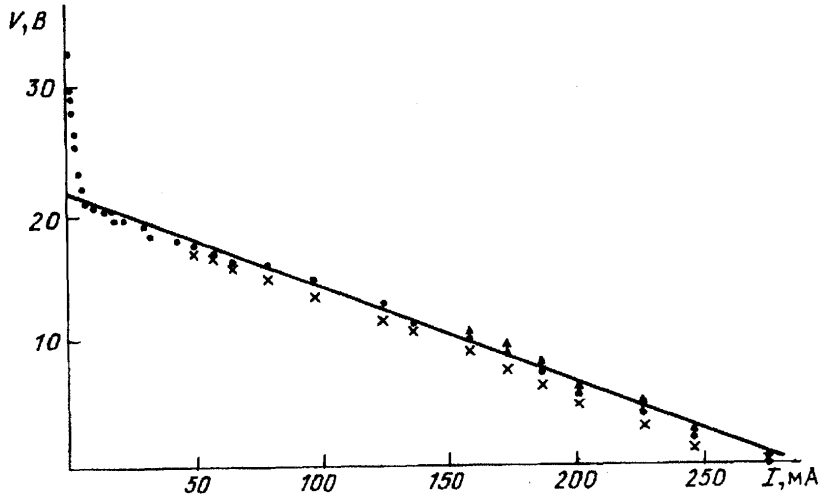


Рис. 1. Результаты измерений вольтамперной характеристики. Точками показаны результаты измерений грубым прибором при возрастании тока I. Прямая линия — результат глазомерного сглаживания точек в линейной области ($I \geq 7$ мА). Она практически совпадает с результатом линейного регрессионного анализа. Характер отклонений точек от прямой (недостаточно хаотический позволяет предположить наличие нелинейности. Однако при изменении порядка проведения эксперимента (измерения при постепенном снижении силы тока) получаются существенно другие значения напряжения (показаны крестиками). Таким образом, вольтамперная характеристика не определена условиями эксперимента с такой точностью, с какой можно было бы попытаться ее изучать, имея только наблюдения, показанные точками

Таблица I. Результаты измерений вольтамперной характеристики

Сопротивление	Сила тока	Напряжение по вариантам опыта, В				
R, Ом	I, мА	$V1$	$V2$	$V3$	$V4$	$V7$
	0	32	32	32,6	—	—
25000	1,2	30,1	30,1	29,4	—	—
20000	1,45	29,4	29,4	29,1	—	—
15000	1,85	28,5	28,5	28,5	—	—
10300	2,58	26,7	26,7	27	—	—
7500	3,3	25,4	25,4	25,6	—	—
5000	4,5	23,4	23,4	23,8	—	—
4000	5,4	22,2	22,2	22,6	—	—
3000	7	21,3	21,3	21,1	21,1	21,1
2000	10,5	21	21	20,8	20,8	20,8
1500	13,8	20,7	20,7	20,7	20,7	20,7
1200	17,4	20,5	20,5	20,3	20,3	20,3
1000	18,6	19,9	19,9	29,2	20,2	20,2
800	22,8	19,8	19,8	20	20,1	19,975
600	33	19,2	19,2	19,3	19,1	19,125
500	34	18,6	18,6	19,2	19,1	19,05
400	45	18	18	18,4	18,1	18,175
350	51	17,8	17,9	18,1	17,6	17,775
300	60	17	17,1	17,4	17	17,1
250	67	16,5	16,6	16,9	16,3	16,45
200	81	15,9	16	15,9	15,4	15,475
150	99	14,6	14,7	14,7	14	14,05
100	126	12,5	12,7	12,8	11,8	12,05
80	138	11,4	11,6	11,8	10,7	10,925
60	160	10	10,3	10,1	8,9	9,2
50	175	8,9	9,2	8,7	7,5	7,95
40	189	7,7	8,1	7,7	6,4	6,8
30	204	6,1	6,5	6,6	5	5,55
20	227	4,4	5	4,8	3,2	3,7
10	247	2,5	3,2	2,9	1,6	1,975
0	276	0	0,8	0,05	0,24*)	0,23*)

$V1$ —измерения грубым прибором, $V2$ — данные $V1$, поправленные на внутреннее сопротивление прибора, $V3$ — измерения точным комплектом приборов, $V4$ — измерения тем же комплектом, но в обратном порядке опыта (т.е. от больших значений I к малым), $V7$ — среднее из четырех опытов.

*) $I = 266$ мА

По данным рис. 1 закон природы (4) явно нарушается, что ясно без какой-либо статистической обработки. Выходит, что я открыл новое явление, но, разумеется, оно ново только для моей глупости, а вообще в радиотехнике наверняка хорошо известно. (Но если делать на подобном материале физпрактикум, то преподаватель, разумеется, должен объяснить студентам, в чем дело.) Однако закономерность (4) — физическая и, в отличие от математической теоремы, должна быть не абсолютно верной, и не всегда, а служить лишь хорошим приближением. Судя по рис. 1, линейность выполняется в очень широкой области токов $I \geq 7$ мА, которая только лишь и интересна для практических целей. Итак, непосредственно из эксперимента мы получаем вполне надежные сведения об области применимости и точности действия физической закономерности (4), и по отношению к этим сведениям термин "гадательные" совершенно неуместен.

Спросим себя, а можем ли мы из рис. 1 или откуда-нибудь еще получить столь же ясные и надежные сведения относительно применимости вероятностной модели теории ошибок к тем возможным ошибкам измерений, которые, может быть, обусловливают некоторые отклонения данных рис. 1 от прямой линии (в области $I \geq 7$ мА)? Конечно же нет, и на этом примере различие между физическим измерением и вероятностной "душой" наблюдений (а также между физическим измерением и статистическим "гаданием") должно, как мне кажется, стать совершенно ясным. Но теперь мы посмотрим, не может ли чем-нибудь гадание помочь измерению.

Концепция наименьших квадратов требует, вообще говоря, полного уравнивания измерений в

Таблица II. Результаты регрессионного анализа данных табл. I

Варианты опыта	Степень многочлена	Константа	Коэффициенты при			Коэффициент корреляции	Оценка ошибки измерений
			I	I^2	I^3		
$V1$	1	21,73	−0,0762			0,9987	0,34
	2	21,64	−0,0662	$-39\cdot10^{-6}$		0,9993	0,25
	3	21,65	−0,0792	$84\cdot10^{-6}$	$-3\cdot10^{-7}$	0,9999	0,21
$V2$	1	21,64	−0,0735			0,9989	0,30
	2	21,4	−0,0659	$-30\cdot10^{-6}$		0,9993	0,25
	3	21,63	−0,0781	$86\cdot10^{-6}$	$-2,8\cdot10^{-7}$	0,9995	0,22
$V3$	1	21,85	−0,0757			0,9989	0,30
	2	21,45	−0,0634	$-49\cdot10^{-6}$		0,9999	0,11
	3	21,56	−0,069	$4,9\cdot10^{-6}$	$-1,3\cdot10^{-7}$	0,9999	0,091
$V7$	1	21,8	−0,0797			0,9979	0,14
	2	21,61	−0,0738	$-23\cdot10^{-6}$		0,9999	0,057

Производилось стандартное сглаживание данных многочленом с помощью метода наименьших квадратов с одинаковыми весами наблюдений; для многочлена степени 1 константа имеет физический смысл ЭДС, коэффициент при I, взятый с противоположным знаком — внутреннего сопротивления, причем поскольку V измеряется в В, а I в мА, внутреннее сопротивление выражается в кОм; последний столбец представляет собой оценку ошибки отдельного наблюдения (стандартное отклонение), измеряемую в В

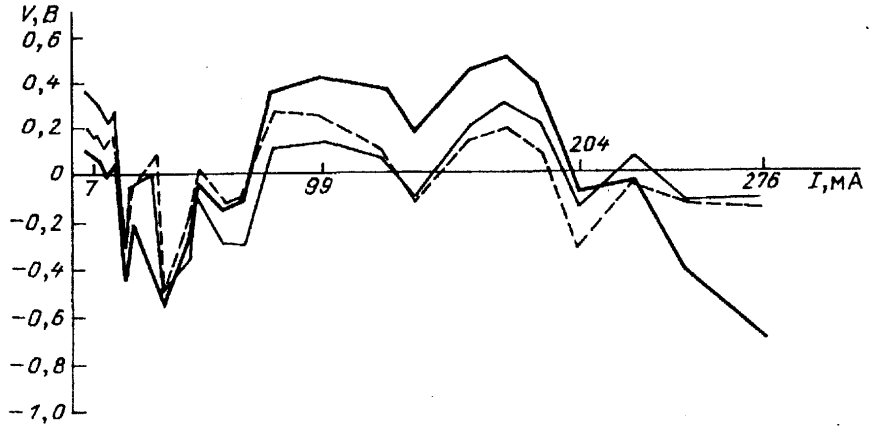
сответствии с законами природы (4) и (5) (всюду в дальнейшем речь идет только об области линейности $I \geq 7$ мА). Но при подстановке в (5) конкретных значений из первых трех столбцов табл. I обнаруживаются очень большие невязки, что, наверное, следует приписать грубости измерения сопротивления тестером в режиме омметра. Я не нашел разумного способа произвести полное уравнивание (где взять веса наблюдений?) и предпочел трактовать эксперимент таким образом, что (якобы) сначала с помощью реостата устанавливалось значение тока во внешней цепи, показанное в столбце I, а затем измерялось падение напряжения, показанное в столбце $V1$. (Данные столбца R ниже используются иначе. Ток и напряжение можно было бы переставить, разрешив уравнение (4) относительно I, но в этом случае коэффициенты уравнения имели бы чуть-чуть менее непосредственный физический смысл, чем коэффициенты уравнения (4).)

Итак, мы имеем дело с линейной регрессией значений столбца $V1$ на значения столбца I. Можно провести линию регрессии по точкам рис. 1 на глаз (именно такая прямая нанесена на рис. 1), а можно воспользоваться компьютером (первая строчка табл. II): результаты неотличимы. Коэффициент корреляции составляет 0,9987 (я радуюсь, что в моем любительском эксперименте получился такой коэффициент корреляции, но дело здесь в относительной простоте техники).

Полагаю, что большинство физиков, не искушенных в статистическом гадании, получив такой коэффициент корреляции, закончили бы измерения и их обработку: чего же искать еще, кроме линейной зависимости? Правда, значения E и r, оцененные по уравнению регрессии: E = 21,73 В, r = 76,2 Ом несколько отличаются от ожидаемых. Действительно, на выходе трансформатора было измерено напряжение переменного тока 25,4 В, что соответствует амплитудному значению 36 В и среднему интегральному значению 22,9 В. Сопротивление дросселя постоянному току было измерено как 70 Ом. Но причину этих различий надо, вероятно, искать там же, где и причину того, что при разомкнутой внешней цепи конденсатор заряжается не до среднего значения напряжения, а почти до амплитудного.

Но человек, знающий, что такое статистическое гадание, так просто не успокоится. Так и гадатели, сопровождавшие императора Юлиана в несчастный поход, все время смотрели, все ли в порядке, весело ли клюют зерно священные куры. Статистик будет рассматривать рис. 1 не как все люди, держа листок бумаги перед глазами, а наискось: поместив глаз почти в плоскость рисунка на продолжение линии регрессии. Тогда он увидит, что отклонения экспериментальных точек от прямой не ведут себя, как положено, независимым случайным величинам (хаотически): точки извиваются вокруг прямой как бы по синусоиде. Ужас: вероятностная модель не выполняется, священные куры не веселы.

Нужно искать причину. Для начала я поинтересовался внутренним сопротивлением тестера и режиме вольтметра и амперметра и нашел, что в режиме вольтметра оно достаточно велико, а в режиме амперметра недостаточно мало. Используя значения R в первом столбце табл. I, вычисляем поправки (вот когда пригодилась избыточная информация) и показываем их стрелками на рис. 1 Развеселились ли священные куры? Точки, находящиеся в правой части чертежа и находившиеся дальше других от прямой регрессии, действительно



Рис, 2. Графики остатков, получаемых при сглаживании наблюдений регрессионными многочленами (грубые измерения: столбец *И*1 табл. I). Жирная сплошная линия — остатки при сглаживании многочленом первой степени, тонкая сплошная линия — остатки для многочлена второй степени, штриховая линия — остатки для многочлена третьей степени. При повышении степени многочлена остатки убывают нередко и их графики явно коррелируют между собой. Это означает, что нелинейность (на которую указывает недостаточно хаотический характер жирной линии) не улавливается многочленами второй и третьей степени

сдвинулись почти на прямую, но сдвинулись и другие точки. Видно, что прямую регрессии придется теперь провести несколько по-иному, а характер отклонений от нее после этого останется прежним. Нет, священные куры не развеселились. Более того, искушенному глазу видно, что их не развеселить, если даже вместо линейной регрессии взять полином второй-третьей степени. Но проделаем соответствующие компьютерные вычисления, чтобы последнее обстоятельство сделать ясным и для любого ученика в искусстве гадания.

Компьютер (конкретно использовался пакет STATGRAPHICS) выдает довольно много всякой (полезной в принципе) информации. Но поскольку вероятностная модель все время под вопросом, большая часть этой информации не нужна (как, например, уровни значимости коэффициентов регрессии либо их стандартные ошибки; кстати, вряд ли кто из неспециалистов знает разницу между обычным уровнем значимости и предельным уровнем, или *p*-значением, которое и выдает компьютер). Но есть одно очень важное число — средний квадрат остатков; извлекая из него квадратный корень, получаем оценку стандартной ошибки наблюдений (в предположении, что регрессионная модель верна). В используемом пакете есть также очень важный график остатков, т.е. разностей между линией регрессии и наблюдениями: он заменяет для не очень искусного гадателя рассматривание наблюдений под косым углом. Рассмотрим рис. 2, на котором графики остатков показаны для столбца *И*1 и для многочленов регрессии степеней 1, 2 и 3. Видно, что остатки уменьшаются с увеличением степени многочлена, но как-то нерешительно, и графики остатков остаются коррелированными. Также и в табл. II оценки ошибки наблюдения убывают, но нерезко. Это означает, что с увеличением степени многочлена никакого раскрытия истины не происходит, а происходит лишь увеличение точности приближения в связи с увеличением числа подгоночных параметров. Ужасно: истинный вид отклонений от линейности с помощью многочлена не угадывается. Священные куры отказываются принимать пищу в виде увеличения порядка регрессии.

Картина для столбца *И*2 почти такая же и здесь не приводится (кроме регрессии в табл.II). Поправки неэффективны.

Возникает ряд гипотез. Может быть, систематические ошибки тестера при измерении тока и напряжения как-нибудь скомбинировались так, что получилось отклонение от линейности сложной формы? А может быть, причиной являются колебания напряжения в сети? Этот гордиев узел сомнений можно разрубить только при помощи новых измерений. Теперь питающее напряжение контролируется своим вольтметром и стабилизируется, насколько возможно, при помощи ручного автотрансформатора. В измерительную цепь включаются одновременно электронный цифровой вольтметр и стрелочный амперметр класса 0,5. Реостатом устанавливаются прежние значения тока, а с вольтметра списываются значения напряжения (столбец *И*3). График остатков приводится на рис. 3, а результаты регрессионного анализа — в табл. II.

Ну, теперь священные куры, кажется, довольны. При увеличении степени многочлена до 2 остатки падают довольно резко, а при увеличении до 3 почти не меняются, Ошибка наблюдений при этом оказывается равной 0,11, что весьма разумно, так как было замечено, что десятые доли вольта на

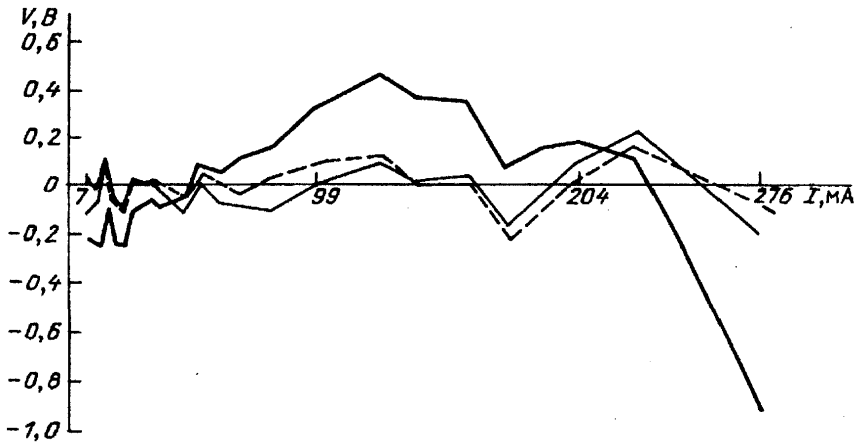


Рис. 3. То же самое, что рис. 2, но для наблюдений более точными приборами (столбец *V*3 табл. I). Нелинейность улавливается многочленом второй степени

электронном вольтметре неустойчивы (что, по-видимому, объясняется кратковременными колебаниями напряжения в сети с частотой порядка 1 Гц, которые нельзя убрать автотрансформатором). Третья степень регрессионного многочлена, очевидно, не нужна (при введении члена с I^3 коэффициент при I^2 компьютером даже трактуется как незначимый). Казалось бы, все ясно: надежно установлен нелинейный эффект, описываемый многочленом второй степени.

Одно лишь облачко есть на этом ясном горизонте. Все значения в столбцах *V*2 и *V*3 довольно близки, кроме самого последнего. Была сделана попытка повторить последнее наблюдение столбца *V*2, и при этом обнаружилось ... что оно не повторяется. Оказывается, что если опыт вести, увеличивая силу тока *I*, то получаются одни значения напряжения (столбцы *V*1, *V*2 и *V*3), а если, наоборот, реостатом постепенно уменьшать силу тока, то получаются другие значения напряжения (столбец *V*4, крестики на рис. 1), которые могут отличаться от первых на величину порядка 1 В. Исследуемое явление (вольтамперная характеристика) не определено условиями опыта с такой точностью, с какой я пытался его исследовать. Причина этого не вполне ясна, но по порядку величины вполне возможно объяснение небольшим нагревом током обмотки дросселя. Итак, описание устройства с помощью электродвижущей силы и внутреннего сопротивления является довольно грубым даже для практических целей, потому что у него могут, оказывается, быть некоторые динамические свойства, нежелательные для источника питания.

Какими же все-таки значениями *E* и *r* пользоваться? Для ответа на этот вопрос были сделаны еще два опыта с возрастанием и уменьшением тока и результаты их усреднены со столбцами *V*3 и *V*4. Полученные средние даны в столбце *V*7, а результаты регрессионного анализа — в двух последних строчках табл. II. Если данные столбца *V*7 нанести на большой лист миллиметровой бумаги (формата А3), то нетрудно увидеть, что они не совсем хорошо приближаются прямой линией, но прекрасно — параболой второго порядка, выпуклость которой направлена вверх. Данные табл. II с этим согласуются: коэффициент при I^2 отрицателен, а оценка ошибки заметно падает при переходе от прямой к параболе. Заметим, что ошибка 0,057 почти вдвое меньше ошибки 0,11, получаемой при сглаживании параболой столбца *V*3: дело можно представить так, что при усреднении данных четырех опытов выявляется некоторая средняя парабола, а случайные ошибки, усредняясь, становятся, как и полагается, меньше в корень квадратный из числа опытов.

Резюмируем все, что происходило. После первой серии измерений (столбец *V*1) была выделена область линейности вольтамперной характеристики. По данным сглаживания наблюдений в этой области оценивается э.д.с. и внутреннее сопротивление источника тока, но возникает вопрос о точности оценок, т.е. о доверительных интервалах; но доверительные интервалы, выдаваемые компьютером, не стоят ничего, если нарушена вероятностная модель, а она и в самом деле нарушена, потому что отклонения наблюдений от прямой ведут себя недостаточно хаотично. Возникает физически интересный вопрос, действительно ли вольтамперная характеристика в (якобы) линейной области на самом деле нелинейна или этот эффект следует приписать ошибкам измерений. Измерения, насколько возможно, поправляются (столбец *V*2), но к радикальным изменениям это не приводит. Нелинейность какая-то странная: не улавливается многочленами второй и третьей степени. Это

соображение вкупе с мыслью о возможных колебаниях напряжения в сети обесценивает имеющиеся наблюдения (на предмет выявления нелинейности). К счастью, более точные измерительные приборы мне достаточно просто снять с полки (не зря я в свое время выменял на водочные талоны на толкучке неплохой лабораторный амперметр). Более точные измерения дают нелинейность, которая улавливается параболой второго порядка, но объясняется это, по-видимому, не тем, что приборы более точны, а тем, что при использовании двух приборов вместо одного электрическая цепь во время серии измерений не разрывается, и потому температурная динамика становится более гладкой (во времени). Статистический прием усреднения четырех серий измерений дает вообще прекрасный результат, не оставляющий сомнений в реальности нелинейности. Практически надо пользоваться любыми значениями э.д.с. и внутреннего сопротивления, найденными хоть из первой серии измерений ($V1$), хоть из последней ($V7$), но теперь ясно, что такое описание достаточно грубо и с практической точки зрения мое чудо весом в 8 кг, построенное на принципах вековой давности, не вполне совершенно: ему не помешал бы стабилитрон.

Все это взаимодействие измерений и их обработки происходило без применения компьютера, которого в нужный момент под рукой не было (в отличие от амперметра): просто путем разглядывания листа миллиметровой бумаги. Но применение компьютера делает весь процесс простым, ясным и потому общедоступным: графики остатков вместо разглядывания чертежа под косым углом. Если бы речь шла о каких-то многомерных зависимостях, то без компьютера статистик оказался бы совершенно бессильным.

7. Заключение: основные принципы преподавания теории ошибок. Вероятностная теория ошибок эксперимента представляет для студента, действительно, крепкий орешек, тем более что и для преподавателя (как мы видели) здесь не всегда существует достаточная ясность. Она замышлялась классиками и может сколько-нибудь разумно излагаться лишь как приложение теории вероятностей, которая в математическом отношении довольно сложна (так как даже в простейшем случае независимости фактически имеет дело с прямым произведением большого числа вероятностных пространств). Но теоретические науки тем и хороши, что они сравнительно легко и быстро могут быть поняты (конечно, при соответствующих склонностях и подготовке). Хорошо известно (см., например, [4]), каким образом теория вероятностей вместе с основами метода наименьших квадратов может быть изложена (для достаточно подготовленных студентов) за семестровый курс, т.е. примерно за 15 лекций. Это составляет 30 академических, или 22,5 астрономических часа, или 3 рабочих дня. Между тем, бухгалтера самым ускоренным методом подготавливают примерно за месяц.

Однако, начав с чисто математической науки, студент, к своему удивлению, внезапно попадает в приложениях не в сферу науки физической (как ему было бы естественно ожидать и как старались думать примерно в течение 150 лет), а в сферу статистической науки, которая по стилю мышления сближается с такими явлениями нашей культуры, как экономическая или (в лучшем случае) демографическая статистика. Для студента физической, физико-технической, естественно-научной, да и вообще почти любой специальности статистическая парадигма чужда и должна разъясняться.

Первое, чего сразу не поймет студент, — это почему не существует физической теории ошибок. Например, если наука уже 200 лет говорит о стандартной ошибке прибора σ, то почему нельзя было научиться ее определять (с той или иной точностью, как физическую величину) и записывать в инструкцию к прибору, как записываются величины основных и дополнительных ошибок. Тут студенту следует разъяснить, что в инструкции записывают весьма несовершенные величины — те, которые более или менее гарантируются производителями приборов для всех приборов данной конструкции. Обычно реальные ошибки значительно меньше. Но ведь ученому нежелательно как недооценивать, так и переоценивать величину ошибок: может случиться, что он спишет некоторый новый эффект на ошибки наблюдений и напрасно не прославится. Когда же мы начинаем исследовать реальные ошибки, то попадаем в некие дебри: они зависят и от конкретного экземпляра прибора данной конструкции, и от времени, прошедшего с момента последней поверки, возможно, и от искусства пользователя, но главное — исследуемое явление может не вполне однозначно воспроизводиться в условиях нашего эксперимента, и тут-то совершенно ясно, что заранее учесть подобные ошибки и записать их в паспорт прибора совершенно невозможно. Это просто счастье, что классики придумали подгоночный параметр σ, который хоть в каком-то среднем смысле характеризует точность воспроизведения и наблюдения каких-то зависимостей; он, кроме того, полезен для спокойствия нашей души, так как обычно позволяет

любые экспериментальные данные согласовать с вероятностной моделью ошибок без особенно кричащих противоречий.

Мысль, состоящую в том, что если противоречия со статистической моделью все же окажутся, то анализ их причин может принести несомненную пользу физическому исследованию, я считаю совершенно ясной и, собственно, тривиальной для нормального студента. В этом и состоит постоянно действующая причина того, что вероятностный анализ ошибок эксперимента не забыт и не отброшен наукой, а наоборот, для него написаны пакеты компьютерных программ.

Но, конечно, усредненная характеристика соответствия наших наблюдений известным законам природы, выражаемая параметром σ, — это сущность, возможная не в рамках физической, а в рамках статистической парадигмы. Противопоставлять эту парадигму физической, употребляя слова типа "миф", "гадание", а также "священные куры" из п. 6.2, по моему мнению, весьма полезно: это поможет студенту понять ситуацию не за 150 лет, а за 15 минут. Неплохо приводить примеры модельных представлений из других наук. Я воспользуюсь некоторым опытом работы в клинической физиологии, чтобы привести следующий пример.

Существует такая наука — физиология дыхания, которая, в частности, служит необходимой и весьма полезной теоретической базой для осуществления разнообразнейших видов аппаратного дыхания: от акваланга до дыхания в космосе чистым кислородом при пониженном давлении, и от дыхания в космосе до применения искусственного кровообращения и искусственной оксигенации крови при хирургических операциях. Эта наука рекомендует оценивать степень недостаточно эффективной работы легких больного при помощи показателя венозного примешивания, или шунта. В учебниках физиологии этот показатель вводится следующим образом. Временно забывается, что легкое имеет 300 млн. альвеол, и рассматривается модель легкого, состоящая из одной альвеолы и двух капилляров. Первый капилляр расположен в стенках альвеолы, так что проходящая по нему кровь (якобы) оксигенируется полностью. Второй находится в стороне, и проходящая по нему кровь остается венозной. Следует простенькая алгебраическая выкладка, и доля крови, проходящей через второй капилляр (это и есть шунт), оказывается выраженной через концентрации кислорода в артериальной и венозной крови, а последние величины при современной технике могут быть сравнительно легко измерены у постели больного. На самом деле эффективность оксигенации крови в легких зависит от тех соотношений между вентиляцией и кровотоком, которые складываются в каждой из 300 млн. альвеол (рассуждение об этом есть на другой странице того же учебника физиологии), так что на самом деле все происходит не так, как в модели: нет ни полного сброса венозной крови, ни полной оксигенации. Но смысл показателя шунта состоит в том, что он согласован с существующими возможностями измерений. Известно, что в некоторых ситуациях догматическое использование шунта ведет к ошибочным выводам. Этот показатель напоминает мне чем-то параметр σ в методе наименьших квадратов, который ведь тоже хорош тем, что определяется по наблюдениям.

Физический эксперимент можно иногда строить с той или иной ориентацией на вероятностную модель ошибок. Например, можно создавать, как это предлагается в [1], коллективы из повторных наблюдений одной и той же величины и проверять их статистическую устойчивость. Это имеет смысл делать тогда, когда статистическая устойчивость или неустойчивость имеет существенное физическое значение. Другим примером является так называемое "планирование эксперимента", когда нас не особенно интересует точная зависимость некоторой функции (например, выхода полезного продукта) от условий эксперимента, но мы хотели бы возможно скорее максимизировать значение этой функции. Возможно планировать эксперимент для скорейшего определения параметров той или иной теоретической зависимости, ориентируясь на ту или иную вероятностную модель ошибок, но это мне лично представляется делом ненадежным из-за заведомой нефизичности моделей ошибок. Кроме того, так можно прозевать самое интересное: например, в эксперименте п. 6.2, если свято веровать в линейную модель (4), эксперименты нужно поставить при $R = \infty$ и $R = 0$, а на самом деле это весьма неразумно. Мне кажется, что в большинстве случаев эксперимент должен планироваться, исходя из соображений физического исследования, но очень неплохо, если результаты наблюдений потом анализируются статистически, например при помощи наименьших квадратов. Редко бывает так, чтобы статистическое исследование не приносило ничего интересного.

Какую же цену придется заплатить за внедрение статистического анализа там, где он ранее не применялся? Практически речь идет о том, что должен возникнуть новый человек, который будет заниматься тем, что с помощью компьютера и, в основном, стандартных пакетов статистических

программ будет заниматься статистическим анализом данных, которые и без него вносились бы в компьютер. Это может быть математик или физик по образованию, но он должен понимать математическую теорию вероятностей, так как алгоритмы статистической обработки всегда основываются на той или иной вероятностной модели для фактических данных, и без понимания того, насколько эта модель может быть неверна, нельзя интерпретировать выдаваемые компьютером результаты. На какую же роль в науке может претендовать такой человек? Все, разумеется, ясно, если речь идет о математике, который в наше трудное время просто кормится таким образом, поскольку за научную работу в области математики его никто кормить не хочет.

Но что такая работа может дать физику? Для ответа на этот вопрос посмотрим сначала, сколько времени могут жить в науке физические понятия и теории. В физике было, например, понятие флогистона, время жизни которого должно было бы быть коротким, потому что школьная традиция требовала его всячески поносить и оплевывать. Но обратимся к уравнению теплопроводности, которое в учебниках выводится из закона Фурье (количество энергии, переносимой через площадку, пропорционально градиенту температуры). Я не понимаю, чем количество тепловой энергии принципиально лучше, чем количество флогистона. Следовательно, физическое понятие вполне может прожить лет триста (конечно, видоизменяясь). Время же активной научной работы отдельного ученого можно, пожалуй, оценить по порядку величины лет в тридцать. Следовательно, занимаясь физикой, можно практически себя увековечить, оставив след в науке на время на порядок больше, чем продолжительность собственной работы. Сколько же времени может быть интересен для науки результат статистической обработки?

Милликен определил заряд электрона в 1909 г., дав, как полагается, и доверительный интервал: точность определения заряда электрона оценивалась в 0,1 % истинной величины. Этот интервал (как и многие другие) потом не подтвердился: ошибка составляет около 0,6 % от истинной величины (считается, что Милликен использовал не совсем верное значение вязкости воздуха). Но в книге [10], первое издание которой вышло в 1935 г., М.П. Бронштейн приводит именно милликеновскую оценку точности. Итак, на этом примере продолжительность жизни статистического результата оценивается лет в тридцать, т.е. на протяжении жизни одного поколения ученых. Это весьма разумно, потому что статистические оценки точности могут сохранять значение до появления существенно более совершенных методов и средств измерения. Таким образом, если физика сравнить с каменщиком, труд которого на века остается в здании науки, то статистика следует сравнить с водопроводчиком или электриком (водопроводные трубы или электропроводку ведь не мешает заменить лет через тридцать, отнюдь не перестраивая здания). Роль исторически более скромная, чем роль каменщика, но весьма полезная и необходимая с точки зрения того понятия о пользе науки, которым начинается данная статья.

СПИСОК ЛИТЕРАТУРЫ

1. *Алимов Ю.И., Кравцов Ю.А,* Является ли вероятность "нормальной" физической величиной? // УФН. 1992. Т. 162. С. 149.
2. *Кравцов Ю.А.* Случайность, детерминированность, предсказуемость. // УФН. 1989. Т. 158. С. 92.
3. *Кравцов Ю.А.* Случайность и предсказуемость динамического хаоса.//Нелинейные волны.Т,2: Динамика и эволюция. М.:Наука, 1981. С. 81.
4. *Тутубалин В.Н.* Теория вероятностей и случайных процессов. Основы математического аппарата и прикладные аспекты. М.:Изд-во МГУ, 1992.
5. *Чебышев П.Л.* Полное собрание сочинений. Т. 5. М.;Л.: Изд-во АН СССР, 1951.
6. *Эльясберг П.Е.* Измерительная информация: сколько ее нужно? как ее обрабатывать? М.: Наука, 1983.
7. *Новицкий П.В., Зограф И.А., Лабунец В.С* Динамика погрешностей средств измерений. Л.: Энергоатомиздат, 1990.
8. *Колмогоров А.Н.* К обоснованию метода наименьших квадратов.// Сборник работ А,Н. Колмогорова "Теория вероятностей и математическая статистика". М.: Наука, 1986. С, 267.
9. *Хаимов З.С.* Основы высшей геодезии. Учебник для вузов, М,: Недра, 1984.
10. *Бронштейн М.П.* Атомы и электроны. М.: Наука, 1980.

Вероятность, компьютеры и обработка результатов эксперимента. Тутубалин В.Н. "Успехи физических наук". Июль 1993 г. Т. 163, № 7. — 93— 109. (Методические заметки). В статье обсуждается вопрос о том, как следует относиться к результатам статистической обработки данных физических экспериментов и как, в частности, преподавать статистические методы студентам. Дело в том, что надежды, по традиции возлагаемые на те или иные статистические приемы и методы, часто бывают обмануты. В конечном счете это происходит из-за того, что общепринятые вероятностные модели ошибок наблюдений, как правило, неверны в статусе физической модели. Однако, если результаты обработки оценивать в рамках статистической парадигмы, то вполне возможно их полезное для физики использование. Ил. 3. Библиогр. ссылок 10.

Probability, computers, and a processing of experimental results. Tutubalin V.N. Uspekhi Fizicheskikh Nauk, July 1993. Vol. 163, No. 7. — 93— 109. (Methodological notes), A right evaluation of the statistical processing results for physical experimental data is in question; in particular, how statistical methods should be exposed for students. The point is that the hopes traditionaly rested on statistical procedures are often deceived. Ultimately, the reason for this is that the commonly accepted probability models for measurement errors are usually false if considered as physical models. However, if the processing results are evaluated in the frame of the statistical paradigm, they can be succesfully used for physical purposes. 3 figs. 10 refs.