СТАНДАРТНАЯ И МЕТРОЛОГИЧЕСКАЯ КОНЦЕПЦИИ СТАТИСТИКИ:
эпистемологический и логико-прагматический анализ адекватности приложений

В.М. Резников

Среди математических дисциплин теория вероятностей и математическая статистика занимают особое место. С одной стороны, благодаря достижениям ряда математиков, в том числе Бореля, Лебега, Бернштейна и Колмогорова, теория вероятностей относится к так называемой чистой математике. С другой стороны, благодаря интенсивному использованию теории вероятностей и математической статистики в приложениях эти дисциплины принадлежат к области прикладной математики. Широкое применение теории вероятностей и особенно математической статистики приводит к вопросу о том, оказывается ли эффективным использование вероятностных подходов при решении проблем, возникающих в практике научных исследований, или оно является лишь свидетельством моды на математизацию.

По нашему мнению, применение этих подходов обусловлено в большей степени внешними причинами, в том числе требованиями, согласно которым в публикациях должны использоваться методы статистического анализа. Так, например анализ примерно двухсот кандидатских и докторских диссертаций в области медицины и биологии, проведенный В.П.Леоновым и П.В.Ижевским, показал, что в подавляющем числе работ статистические методы использовались неоптимальным образом или просто неправильно, без учета границ применимости соответствующих методов [1]. Авторы полагают, что это связано с недостатками в обучении студентов методам теории вероятностей, а также с тем, что статистический метод как средство принятия решений в условиях неопределенности долгое время не был вполне легитимным, о чем свидетельствует запрет на преподавание биостатистики, когда генетика была объявлена лженаукой.

Кроме медицины и биологии наименее обоснованной областью применения математической статистики является область стандартизации. Многие ГОСТы в технике, разработанные с использованием статистических методов, содержат грубые ошибки. Несмотря на острую и справедливую критику, эти ГОСТы остаются без изменений [2].

В настоящее время существует несколько концепций теории вероятностей и математической статистики. Наиболее известными вероятностными концепциями являются классическая, частотная, логическая, субъективистская, байесовская, конструктивистская и некоторые другие. В вузовских учебниках излагается преимущественно стандартная, или фишеровская, концепция статистики, и соответственно она чаще всего применяется в прикладных исследованиях. В литературе, особенно в методологической, практически неизвестна метрологическая концепция, которая представляет собой альтернативу стандартной концепции. В настоящей статье исследуется обоснованность некоторых принципов фишеровского статистического анализа, а в качестве альтернативы стандартному подходу рассматривается метрологическая концепция статистики.

Проблема обоснованности фишеровской статистики изучалась нами на примере параметрического раздела статистики. Рассматривались следующие вопросы: 1) адекватность проблем, решаемых в параметрическом разделе статистики, методологическим принципам естествознания; 2) обоснованность базовых принципов, на которые опираются методы оценки параметров; 3) адекватность критериев оценки параметров эмпирическим исследованиям; 4) эпистемологический статус предельных теорем теории вероятностей.

Проблема оценки параметров может трактоваться как самостоятельная проблема или как часть общей проблемы проверки гипотез. В общем случае кроме задачи оценивания параметров решается задача проверки соответствия распределения с найденными параметрами имеющимся данным. Проблема проверки гипотез, несомненно, заслуживает самостоятельного анализа, и в настоящей работе она не рассматривается. Проблема оценки параметров является классическим направлением статистики, и, по нашему мнению, необходим критический анализ этого раздела статистики с точки зрения современных требований к эффективности математики.

В любом случае в задачах параметрического раздела предполагается, что задано вероятностное распределение исследуемых явлений.

Некоторые параметры распределения неизвестны. Декларируется "оптимальное" определение искомых параметров.

Как известно, в вероятностном распределении содержится вся возможная информация об объектах вероятностной природы. Считается, что распределение является вероятностным законом природы. Для анализа адекватности проблемы оценивания параметров практике естественных наук полезно вначале рассмотреть аналогичную детерминистскую постановку задачи.

Рассуждая по аналогии, предположим, что мы каким-то образом в существенной степени открыли закон природы, например второй закон Ньютона. Мы знаем, что ускорение пропорционально приложенной силе, но коэффициент пропорциональности точно не известен. В разделе параметрической статистики предлагаются универсальные методы оценивания параметров "оптимальным образом". Не вызывает сомнения, что предположение о существовании универсальных "оптимальных" теоретических способов оценки неизвестного коэффициента в законе Ньютона и других законах будет единодушно отклонено специалистами в области физики.

Перейдем от детерминистских аналогий к собственно вероятностной проблематике. Параметрами часто являются моменты распределений. Так, например, параметрами нормального распределения являются первый и второй моменты, – это соответственно математическое ожидание и дисперсия. Известна теорема о том, что знание всех моментов распределения не позволяет восстановить это распределение. Другими словами, в распределении содержится в огромной степени больше информации, чем в неизвестных параметрах. Весьма сложная теория оценивания параметров в итоге добавляет немного информации к уже имеющейся. Огромная разница в информативности распределения и параметров порождает вопросы эпистемологического характера: каким образом априори стало известно распределение? рационально ли для оценивания неизвестного параметра вначале решать более сложную задачу установления типа распределения?

Сформулируем некоторые аргументы против адекватности постановки задач в параметрическом разделе статистики для технических приложений. Во-первых, в распределении содержится максимально возможная информация об изучаемом процессе и знание априори распределения исследуемого процесса является нетипичной ситуацией. Во-вторых, часто исследователь не имеет достаточно данных для проверки гипотезы о типе распределения. В-третьих, параметры распределения, особенно на первой стадии исследования, например средние, дисперсии и др., представляют интерес сами по себе, а не как параметры распределения. В-четвертых, по-видимому нерационально для решения весьма простой задачи определения среднего предварительно решать гораздо более трудную проблему поиска распределения.

Рассмотрев постановку задач в параметрическом анализе, перейдем к исследованию обоснованности базовых положений теории оценивания. Наиболее общим методом оценивания параметров является метод максимального правдоподобия (ММП), предполагающий, что данные могут быть описаны множеством независимых случайных величин. Методологической основой ММП является принцип максимального правдоподобия. Этот принцип приписывает максимальную величину совместной плотности вероятности независимых случайных величин. Он приписывает максимальную совместную вероятность появления множеству данных, рассматриваемых как единое целое. В сущности, этот принцип означает: то, что произошло, непременно должно было произойти. Принцип максимального правдоподобия был введен в практику статистических исследований английским статистиком Фишером. В методологической литературе Фишер считается представителем частотной концепции [3]. Эта оценка не вполне правильная. В чисто частотной концепции Мизеса вероятность может быть приписана только коллективу событий, обладающему свойством статистической устойчивости. В концепции Рейхенбаха вероятность может быть получена и для отдельного события, при этом предполагается, что рассматриваемое событие является членом коллектива, которому была задана вероятность.

Принцип максимального правдоподобия очевидно не вписывается ни в концепцию Мизеса, ни в концепцию Рейхенбаха. Согласно частотному принципу события с большой вероятностью будут появляться чаще, чем события с малой вероятностью. Частотный принцип не предполагает, в отличие от фишеровского подхода, что в первом и единственном проведенном эксперименте произойдет событие с наибольшей вероятностью. Поэтому принцип максимального правдоподобия не является частотным.

В рамках ММП не исследованы многие вопросы, без решения которых адекватность применения этого метода на практике не представляется обоснованной. Наиболее важными являются следующие вопросы: а) почему функция правдоподобия, построенная на единственной выборке, примет максимальное значение? б) как изменится величина найденного параметра, если отказаться от условия, что функция правдоподобия примет максимальное значение?

Как известно, область применения статистики как научного метода – это массовые повторяющиеся явления. Верификация регулярности в массовых явлениях необходимо предполагает, что исследуемые статистические характеристики суть характеристики устойчивые. Применение принципа правдоподобия не позволяет верифицировать устойчивость статистических оценок. Во-первых, в известных руководствах по математической статистике предписывается использовать все имеющиеся данные для определения оценки параметра, проверка же устойчивости полученного результата на новых данных не предполагается. Во-вторых, проверка независимости осуществляется преимущественно на интуитивных основаниях.

Теперь от анализа принципа правдоподобия перейдем к краткому рассмотрению критериев оценивания. В фишеровской концепции статистики метод максимального правдоподобия считается обоснованным, так как его использование приводит к получению статистических оценок, обладающих свойствами состоятельности, несмещенности и эффективности. Оценка t_n параметра θ называется состоятельной, если имеет место сходимость по вероятности этой оценки к искомому параметру [4]. Другими словами, для $\forall\, \varepsilon > 0$

$$\lim_{n\to\infty} P\{|\, t_n - \theta| < \varepsilon\} = 1 \qquad (1)$$

Свойство состоятельности не позволяет определить величину отклонения оценки, найденной с помощью ММП по конечной выборке, от оценки, удовлетворяющей выражению (1).

Оценка t_n параметра θ называется несмещенной, если математическое ожидание оценки совпадает со значением искомого параметр [5]:

$$M\,[t_n] = \theta \qquad (2)$$

Математическое ожидание является теоретической величиной. Равенство (2) не означает, что математическое ожидание оценки t_n существует. Верно обратное, а именно: если математическое ожидание оценки существует, то выполняется выражение (2).

Чисто теоретические конструкции, используемые в критериях оценивания, стали объектом внимания в теории вероятностей после Бернулли [6]. Бернулли полагал, что доказанная им в простейшем случае теорема закона больших чисел устанавливает связь субъективных и объективных вероятностей. В современной математической литературе считается, что теорема закона больших чисел устанавливает эмпирическую сходимость средних величин и частот соответственно к математическому ожиданию и вероятности. В наиболее простом случае теорема закона больших чисел формулируется следующим образом.

Пусть μ – число наступлений события A в n независимых испытаниях и p есть вероятность наступления A в каждом из испытаний. Тогда каково бы ни было $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|\,\mu/n - p(A)|<\varepsilon\}=1\,. \qquad (3)$$

Почему этой теореме придается большое значение? Например, в работе Б.В.Гнеденко приведена следующая аргументация: “В практической деятельности, да и в общетеоретических задачах большое значение имеют события с вероятностями, близкими к единице или нулю. Отсюда становится ясным, что одной из основных задач теории вероятностей должно быть установление закономерностей, происходящих с вероятностями, близкими к единице, при этом особую роль должны играть закономерности, возникающие в результате наложения большого числа независимых или слабозависимых случайных факторов. Закон больших чисел является одним из таких предложений теории вероятностей, и при том важнейшим” [7]. И далее автор пишет: “На самом же деле непреходящая научная ценность исследований Чебышева, Маркова и других исследователей в области закона больших чисел состоит не в том, что они подметили эмпирическую устойчивость средних, а в том, что они нашли те общие условия, выполнение которых обязательно влечет за собой статистическую устойчивость средних” [8]. Кроме того, в литературе отмечается значимость предельных теорем теории вероятностей для математической статистики.

Фундаментальное различие между теорией вероятностей и математической статистикой состоит в том, что аппарат теории вероятностей обеспечивает корректное вычисление вероятностей событий, если задано вероятностное пространство, т.е. множество элементарных событий, множество событий и вероятностная мера. Задачи математической статистики являются в некотором смысле обратными и заключаются в определении первичных вероятностных характеристик исследуемых объектов.

Возникает вопрос о статусе теоремы закона больших чисел: является ли она теоретико-вероятностной, т.е. позволяет ли вычислять вероятности событий на основании вероятностей элементарных, или эта теорема позволяет выявить и верифицировать первичные вероятностные характеристики, например определить устойчивость частот, и тем самым относится к математической статистике?

В теореме закона больших чисел теоретическая величина “вероятность успеха” задана априори, и тем самым теорема не может быть использована для определения этой уже известной вероятности. Заключение теоремы говорит о вероятности события, представляющего собой разность вероятности события A и частоты события A. Под событием в прикладной теории вероятностей понимается результат любого опыта. Вероятность события является теоретической величиной, поэтому она не может быть результатом реального испытания.

При эмпирической интерпретации вероятность события определена тогда и только тогда, когда соответствующая частота является устойчивой. При эмпирическом подходе внешняя вероятность в выражении (3) должна быть заменена на частоту, тогда заключение примет следующий вид:

$$\omega n\{|\,\mu/n - p(A)| < \varepsilon\} = 1\,, \tag{4}$$

где ω – частота, а n – большое число.

Заключение теоремы, записанное с помощью выражения (4), говорит о частоте отклонения абсолютной разности вероятности события A и частоты этого же события при условии, что эта разность меньше ε. Эвристическая ценность заключения теоремы для определения устойчивости частоты при ее эмпирической интерпретации вызывает большие сомнения. Действительно, если вероятность события A близка к нулю, то в заключении теоремы говорится о частоте частоты события A. Предположим, что теорема имеет эвристическое значение, тогда оно заключается в том, что по частоте частоты события A предполагается возможным определить устойчивость частоты события A.

Очевидно, что методологическая значимость такой интерпретации теоремы явно переоценивается. Во-первых, невозможно определить частоту отклонения частоты события A от вероятности, предварительно не определив частоту события A. Во-вторых, для того, чтобы использовать частоту частоты события с целью определения устойчивости события, предварительно необходимо убедиться в наличии устойчивости частоты от частоты события. Если продолжить аналогию, то для того, чтобы убедиться в устойчивости частоты от частоты события A, необходимо использовать более сложную конструкцию, а именно, требуется частота от частоты события B, при этом последнее событие, в свою очередь, есть частота события A. Налицо логический круг [9].

Особенностью современной математической статистики является доминирование теоретических величин над эмпирическими [10]. Существует вполне обоснованное предположение о том, что супертеоретическая направленность стандартной статистики препятствует корректным применениям. Поэтому вполне естественным представляется интерес к эмпирическому подходу, обосновывающему корректное применение математики, и в частности математической статистики.

Рассмотрим основные положения метрологической концепции как альтернативы фишеровскому подходу.

Эмпирический критерий адекватности утверждений математики области приложений заключается в следующем. Если исходные утверждения, сформулированные на языке какой-либо математической дисциплины, правильно описывают положение дел, то и утверждение, формальным образом полученное из исходных утверждений с помощью правил вывода и аксиом использованного раздела математики, тоже корректно описывает действительность. Эмпирический критерий корректного применения научных методов для исследования естественно-научных явлений был предложен Контом [11]. Конт относил математику к естественно-научным дисциплинам. Типичное применение научного метода, опирающегося на математику, по Конту, состоит в использовании декартовского принципа “разделяй и властвуй”. Решение исследуемой проблемы достигается путем последовательного решения ряда вспомогательных задач.

Предполагается, что решение сводится к определению совокупности переменных величин

$$X_1, X_2, \ldots, X_n. \tag{5}$$

Для того, чтобы это осуществить, предварительно определяется совокупность переменных

$$P_1, P_2, \ldots, P_s. \tag{6}$$

Переменные последовательности (6) сущностно связаны с искомыми переменными, и они могут быть экспериментально исследованы с заданной точностью. Использование математики предполагает определение функций F_1, F_2, …, F_m, связывающих измеряемые в эксперименте переменные с искомыми переменными.

Метрологическая концепция статистики является развитием и модификацией контовской программы с учетом вероятностной специфики [12]. В метрологической концепции переменные $P_1, P_2, \ldots, P_s$ непосредственно неизмеримы. Они имеют статус полуэмпирических величин. Величины $P_1, P_2, \ldots, P_s$ являются статистическими характеристиками непосредственно измеряемых первичных величин $Z_1, Z_2, \ldots, Z_k$. Величины Z_i, $i = 1, k$ делятся на s групп. P_i, $i = 1, s$ является статистической характеристикой i группы.

Обозначим посредством $P_{\min}$ и $P_{\max}$ соответственно минимальное и максимальное значения P_i, $i = 1, s$. Величины $P_1, P_2, \ldots, P_s$ являются статистически устойчивыми, если $P_{\max} - P_{\min} < \varepsilon$, где ε – заданная точность вычислений.

Критерий адекватности в метрологической статистической концепции имеет прогнозный характер и предполагает, что формализация является корректной, если величины $P_1, P_2, \ldots, P_s$ остаются устойчивыми при $K \geq k$, а также $S >> s$. Если полуэмпирические величины $P_1, P_2, \ldots, P_s$ являются устойчивыми, то при изложении результатов исследований полуэмпирической величины P, а также при вычислении величины X переходят к теоретическим величинам P, X.

В отличие от стандартной фишеровской концепции, в которой теоретические величины выступают объектами исследования, в метрологической концепции они имеют второстепенное значение. Одни теоретические величины используются для вычисления других, связанных с ними функциональным образом, при условии, что упомянутые теоретические величины имеют эмпирическую интерпретацию. Кроме того, они используются для того, чтобы осуществить компактную запись полученных результатов исследования.

Трудно говорить о прямом влиянии методологии Конта на развитие прикладной математики, и в частности на метрологическую теорию. В работах Алимова, Тутубалина, Крылова и других исследователей, так или иначе способствовавших созданию метрологической концепции, нет ссылок на работы Конта. Это вполне объяснимо. Дело в том, что работы позитивистов в России никогда не были широко известными. Основные принципы контовской методологии являются естественными для корректных приложений математики в технических науках и были, по всей вероятности, сформулированы независимо от французского философа методологами в области прикладной математики.

Влияние мизесовской статистической концепции на метрологическую концепцию, напротив, непосредственное. Метрологическая теория – это адаптированный для приложений вариант концепции Мизеса. Она свободна от мизесовского фундаментализма. Согласно Мизесу, вероятность может быть определена для бесконечных коллективов [13]. Последовательность данных, представляющих собой коллектив, сходится к предельной величине, являющейся частотой события. Для любой произвольным образом выбранной последовательности требуется сходимость членов последовательности к одному и тому же для всех последовательностей предельному значению. В отличие от мизесовской концепции метрологическая концепция применяется для данных конечных объемов и опирается на ряд положений, соответствующих естественно-научной традиции:

а) первичные характеристики объекта должны быть измерены с заданной точностью;

б) статистические характеристики являются полуэмпирическими и должны быть устойчивыми;

в) теоретические величины представляют устойчивые полуэмпирические величины;

г) устойчивость есть основание для исследования поведения изучаемых процессов в будущем, т.е. для прогнозирования.

Очевидно, что жесткая ориентация на получение правильного прогноза – общий момент для всех частотных концепций. Другие статистические теории, например фишеровская, в меньшей степени ориентированы на получение точного прогноза, поэтому в современных технических приложениях адекватной является метрологическая концепция.

Дополнительным аргументом в пользу адекватности метрологического подхода служит алгоритмическая сложность определения устойчивости с помощью эмпирической интерпретации закона больших чисел и с помощью непосредственного применения метрологической концепции.

Рассмотрим алгоритмическую сложность определения устойчивости частоты события A в двух случаях. В первом случае применяется методология, разработанная Алимовым. Пусть используются исходные данные X_1, X_2, …, X_n. Разбивая их на k групп, при этом $n >> k$, получаем k характеристик w_1, w_2, …w_k для определения устойчивости по вышеприведенному алгоритму.

Во втором случае используется эмпирический вариант теоремы закона больших чисел. Здесь для определения устойчивости частоты события A, опираясь на частоту от частоты события A, надо определить частоту частоты события с такой же надежностью, как и саму частоту. Для этого необходимо иметь n значений частот события A. Для надежного определения одного значения частоты необходимо иметь $[n/k]$ данных. Символ $[z]$ означает целую часть числа z. Отсюда для получения n значений частот необходимо иметь $[n/k] \times n$ данных. Поэтому трудоемкость при определении устойчивости частоты на основе определения устойчивости частоты от частоты события принципиально превышает трудоемкость определения устойчивости частоты события.

Очевидно, что ни по идейным, ни по техническим соображениям теорема закона больших чисел не имеет эффективной эмпирической интерпретации для определения устойчивости частоты.

Итак, нами получено, что постановки задач параметрического раздела статистики, принципы, на которых основаны методы оценивания параметров, не являются адекватными для ответственных приложений. Принцип правдоподобия имеет не вероятностный, а детерминистский характер. Все, что произошло в конкретном эксперименте, должно было произойти – так в силу принципа правдоподобия совместному появлению актуализированных данных приписывается максимальная вероятность. Общая методологическая слабость стандартных статистических принципов проявляется в формализованном описании результатов сингулярного эксперимента. Не убедительны ни принцип практической достоверности, в силу которого событие с малой вероятностью не может произойти в первом проведенном эксперименте, ни фишеровский принцип правдоподобия, ни другие принципы, обосновывающие переход от общестатистической закономерности к закономерности, проявляющейся в единичном опыте.

В технических науках роль классической математической статистики невелика. Первичные статистические характеристики определяют экспериментально. Используя теорию вероятностей, определяют характеристики вторичных величин. Для того чтобы прикладная математическая статистика играла бóльшую роль в приложениях, должен быть решен ряд вопросов, в том числе должны быть выявлены принципы, проливающие свет на связь общестатистической закономерности и сингулярной статистической закономерности, и найдены принципы связи полуэмпирических и теоретических статистических величин.

Примечания

1. См.: *Леонов В.П., Ижевский П.В.* Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям // Бюл. Гос. высш. аттестацион. комитета Рос. Федерации. – 1997. – № 3. – С. 56–61.

2. См.: *Алимов Ю.И., Кравцов Ю.А.* Является ли вероятность “нормальной” физической величиной? // Успехи физ. наук. – 1992. – Т. 162. – № 7.- С. 149–182; *Орлов А.И.* Сертификация и статистические методы // Заводская лаборатория. – 1997. – Т. 63. № 3. – С. 55–62.

3. См.: *Hacking I.* Logic of statistical inference. – Cambridge, 1965; *Кендалл М., Стьюарт А.* Статистические выводы и связи. – М., 1973.

4. См.: *Кендалл М, Стьюарт А.* Статистические выводы и связи.

5. Там же.

6. См.: *Бернулли Я.* О законе больших чисел. – М., 1986.

7. *Гнеденко Б.В.* Курс теории вероятностей. – М., 1988. – С. 185.

8. Там же. – С. 186.

9. См.: *Колмогоров А.Н.* Теория вероятностей // Математика, ее содержание, методы и значение. – М., 1956.

10. См.: *Алимов Ю.А.* Альтернатива методу математической статистики. – М., 1980; *Алимов Ю.И., Кравцов Ю.А.* Является ли вероятность “нормальной” физической величиной?

11. См.: *Конт О.* Курс положительной философии. – СПб., 1899.

12. См.: *Колмогоров А.Н.* Теория вероятностей; *Алимов Ю.И., Кравцов Ю.А.* Является ли вероятность “нормальной” физической величиной?

13. См.: *Мизес Р.* Вероятность и статистика. – М.;Л., 1930; *Mises R.* Theory of probability. – California, 1960.

Институт философии и права
СО РАН, г. Новосибирск

Reznikov V.M. **Standard and metrological conceptions of statistics: epistemological and logical-pragmatic analysis of application adequacy**

The paper studies methodology of Fisher’s statistics. Is grounded that the basic theses of Fisher’s statistics – such as the *credibility* principle (which is used for estimating distribution parameters) and the principle of *practical confidence* of *hypotheses falsification* – notwithstanding traditional notions do not conform to the frequency paradigm. Is stated that the large number theorem has no epistemological significance, nor pragmatic one. One the one hand, the metrological conception of statistics is the alternative for Fisher’s theory. On the other hand, the robustness criterion which is used in metrological theory may be adequately used within Fisher’s methodology. Is shown that there exists ideological connection between metrological conception and *Kont*’s methodology of application of mathematics, as well as *Mizes*’s and *Reuchenbach*’s frequency theories.