МОДЕЛИ КОГНИТИВНЫХ ПРОЦЕССОВ
(Вычислительные системы)
2001 год **Выпуск 168**

УДК 519

ЛОГИКО-ПРАГМАТИЧЕСКИЙ АНАЛИЗ ПРОБЛЕМЫ КОРРЕКТНОГО ПРИМЕНЕНИЯ СТАТИСТИКИ

В.М. Резников

В работе исследованы некоторые факторы, препятствующие корректному использованию статистики в практике научных исследований. Первая проблема корректного использования статистики и теории вероятностей связана с тем, что структура знания в этих дисциплинах имеет определенную особенность. Эта особенность заключается в том, что методы доказательства всех основных результатов предполагают, что исследуемые объекты всегда обладают некоторыми свойствами. Проверка наличия таких свойств технически сложна. Вторая проблема вызвана тем, что некоторые методы приписывают определенные и весьма разумные свойства изучаемому миру явлений. В то же время верификация адекватности исходных предпосылок невозможна по техническим причинам. Третья проблема возникает из-за несоизмеримости результатов, получаемых с помощью методов, считающихся весьма респектабельными. К такого рода методам, например, относится метод максимального правдоподобия. Четвертая проблема связана с верификацией гипотез. Верификация опирается на ряд положений не являющихся реалистичными. Предполагаются известными распределение данных и распределение критериальной статистики. Распределение представляет собой максимальную информацию о статистическом объекте и поэтому в предложениях редко бывает доступной. Пятая проблема связана с неправильной трактовкой классических предельных теорем.

Начнем изложение с первой проблемы, вызванной спецификой вероятностного знания. С целью выявления этой специфики дадим определение базового свойства объектов в математической статистике и теории вероятностей.

ОПРЕДЕЛЕНИЕ. Свойство называется базовым для определенной математической дисциплины, если оно удовлетворяет двум положениям:

1) все основные результаты в этой области получены в предположении, что это свойство имеет место;

2) это свойство логически независимо от любой комбинации других свойств.

Примерами таких свойств являются статистическая однородность в статистике и независимость в теории вероятностей.

А.Н.Колмогоров считал, что свойство независимости является значимым для исследований в области теории вероятностей. Он писал: "исторически независимость испытаний и случайных величин явилась тем математическим понятием, которое придало теории вероятностей своеобразный отпечаток ... Соответственно этому одной из важнейших задач философии естественных наук, после разъяснения пресловутого вопроса о сущности самого понятия вероятности, являются выяснение и уточнение тех предпосылок, при которых можно какие–либо данные или действительные явления рассматривать как независимые" [1, с. 18]. Серьезной экспликации свойства независимости не было получено, так как независимость является базовым свойством, и для его исследования необходимо использовать содержательный анализ.

В арсенале математической статистики имеются разные способы проверки свойства независимости в эмпирических объектах. Эти методы отличаются друг от друга тем, что для их применения требуется разная дополнительная информация. Общей особенностью этих методов является то, что для их применения для больших массивов данных требуются интенсивные вычисления. На практике достаточно редко используют формальные методы для проверки независимости. Обычно ограничиваются содержательным анализом. Острый критический анализ содержательных рассуждений, обосновывающих принятие модели независимых испытаний, дан в работах Ю.И.Алимова [2–3]. Часто принятие модели независимых экспериментов ограничивается следующими рассуждениями. Пусть из содержательных соображений известно, что эксперименты проводились при независимых, контролируемых условиях. Тогда результаты экспериментов тоже будут независимыми, и, следовательно, имеет смысл принять модель независимых экспериментов. В чем слабость данных рассуждений?

Во-первых, за пределами физики трудно гарантировать полный контроль условий проводимых экспериментов. Даже в фундаментальной физике, где эксперименты чище, чем в других областях, не всегда это удавалось.

Во-вторых, в этом рассуждении имеется определенный логический изъян. Из интуитивной, содержательной независимости автоматически не следует формальная независимость. Правильнее сказать, что если формальным образом выявлена независимость результатов экспериментов, то имеются основания полагать, что данные эксперименты проводились в независимых условиях.

К этому, по сути, совершенно справедливому анализу Ю.И.Алимова необходимо сделать следующее замечание. Строго формальные методы недостаточны для верификации сложных моделей. Верификация и принятие любой достаточно сложной модели осуществляется путем синтеза знания, полученного с помощью как формальных, так и содержательных соображений. Специально выделим следующую возможную ситуацию. Предположим, что эксперименты действительно проводились в независимых условиях, и мы знаем об этом факте. В то же время для нашей конкретной ситуации современные статистические средства являются слишком грубыми для выявления независимости. Поэтому принятие модели с независимыми экспериментами в данной ситуации является вполне обоснованным решением.

Вторая проблема корректного использования статистики связана с тем, что некоторые статистические методы не позволяют ни верифицировать используемые в этих методах предпосылки, ни обеспечить соизмеримость полученных с их помощью результатов.

Рассмотрим метод максимального правдоподобия. Метод максимального правдоподобия опирается на следующее предположение. Согласно этому методу считается, что совместная вероятность осуществившихся событий должна быть максимальной. Сама по себе предпосылка этого метода вполне разумна. Тем не менее, метод не позволяет формальным путем ни исследовать правильность исходного предположения, ни оценить погрешность от его принятия. Другой особенностью этого метода является то, что он фактически предназначен для объяснения типа ad hoc, так как получаемые с его помощью результаты не являются соизмеримыми в рамках формальных методов. Предположим, что посредством метода максимального правдоподобия мы получили две различные оценки. Возникает естественный вопрос: "Какая из них ближе к искомому параметру?" Метод максимального правдоподобия не дает ответа на этот вопрос. В статистической литературе к достоинствам этого метода относят то, что он позволяет получить оптимальные асимптотические оценки. Оценка называется Θ асимптотически эффективной если она удовлетворяет двум условиям:

1) $\lim\limits_{n\to\infty} \Theta_n \to \Theta$;

2) оценка Θ имеет наименьшую дисперсию оценок среди оценок, удовлетворяющих условию 1.

В работе П.Е.Эльясберга значимость асимптотических оценок в области ответственных приложений справедливо названа одним из мифов 20 столетия [4, с. 101].

Третья проблема корректного применения статистики связана с верификацией гипотез. В любом учебнике, посвященном статистике, имеется раздел, имеющий название проверки гипотез. По-видимому, название не отражает сути используемых в этом разделе методов. Более правильным названием является фальсификация гипотез. Дело в том, что результатом применения этих методов является или фальсификация гипотезы на данном уровне значимости, или не фальсификация этой гипотезы для данного уровня значимости. Фальсификация гипотез в статистике опирается на ряд нереалистичных предположений и принципов, неадекватных для использования в статистической практике. Предполагается что нам точно известно распределение данных и распределение критериальной статистики, что не является реалистичным. Во-вторых, фальсификация гипотез опирается на принцип практической достоверности [5]. Согласно этому принципу гипотеза отвергается независимо от ее правильности, если при первом же проведенном эксперименте происходит событие, имеющее низкую вероятность появления для данной гипотезы. В чем слабость данного принципа? Во-первых, особенностью маловероятных событий является их низкая частота появления, а не невозможность появления в первом проведенном эксперименте. Во-вторых, если реалистично предполагать, что мы не знаем априори распределение данных и пытаемся каким-либо справедливым экспериментальным способом проверить вероятность события порядка 10^{-N}, то по оценке Ю.И.Алимова для дискретных распределений необходимо осуществить 10^{N+1} экспериментов [2]. Осуществление окончательной фальсификации гипотез в статистике принципиально невозможно, корректная процедура фальсификации с помощью событий с малыми вероятностями сложна по техническим причинам. Отсюда из дидактических соображений более адекватной для приложений представляется не фальсификация гипотез путем апелляции к событиям с малыми вероятностями, а верификация гипотез для событий с большими и средними вероятностями.

Четвертая проблема характеризуется тем, что корректные в математическом отношении теории некорректно применяются. Необходимо подчеркнуть, что критикуется ни теория статистики, а способы приложений теории. Так, например, теория доверительных интервалов является вполне корректной. Она предполагает задание вероятности попадания неизвестного параметра, который является константой, в интервал со случайными концами. Исследователя часто интересует: попадет ли параметр в конкретную числовую область. Поэтому задается доверительная вероятность попадания параметра в фиксированный интервал. Такая процедура для вероятностей, отличных от 0 и 1, является бессмысленной. Константа или попадает в фиксированный интервал с вероятностью равной единице, или не попадает в него с единичной вероятностью.

Пятая проблема связана с неправильной интерпретацией в теории вероятностей так называемых предельных теорем. В литературе, начиная с работы Бернулли, этим теоремам часто приписывается не только внутриматематическая, но даже эпистемологическая значимость [6,7]. По нашему мнению предельные теоремы не имеют ни особой прикладной, ни эпистемологической значимости. Рассмотрим для примера теорему закона больших чисел. Считается, что эта теорема обосновывает устойчивость частот. Существует множество формулировок этой теоремы. Приведем достаточно стандартную формулировку теоремы закона больших чисел.

ТЕОРЕМА. *Пусть проводится n испытаний, в каждом из которых может наступить или не наступить некоторое событие a, и пусть Y_n — количество наступлений события a в этой серии испытаний. Далее полагаем, что имеют место следующие условия:*

1) вероятность $P(a)$ события (a) в каждом испытании равна p;

2) испытания независимы.

Тогда для вероятности

$$P(n, \varepsilon) = P(|\omega_n(a) - p|) \leq \varepsilon \qquad (1)$$

того, что частота $\omega_n = \dfrac{1}{n} Y_n$, отклонится от p не более, чем на ε, справедливо равенство

$$\lim_{n\to\infty} P(n, \varepsilon) = 1, \text{ для любого } \varepsilon > 0. \qquad (2)$$

Стандартная интерпретация этой теоремы заключается в том, что она обосновывает устойчивость частот. Такая трактовка расходится с общепринятыми представлениями о роли математики в прикладных исследованиях. Математический аппарат обеспечивает переход от истинных утверждений к другим истинным утверждениям. Посредством исключительно математических соображений невозможно установить истину исходных посылок. Стандартная интерпретация означает, что заключение теоремы позволяет обнаружить истинность исходной посылки о сходимости частоты к вероятности.

Теперь предположим, что эта теорема не была доказана, и проверка устойчивости будет проводиться эмпирически. Для решения не полностью формализуемых задач математика не способна дать окончательный ответ, но ее использование часто позволяет сделать обоснованные прогнозные оценки. Для эмпирической проверки устойчивости частот в рамках мизесовской частотной концепции может быть применена одна из двух схем [2]. Первая схема названа Ю.И.Алимовым схемой многих выборок. Данные разбиваются на q равных групп. Для каждой группы вычисляются эмпирические частоты: Z_i, $i = \overline{1, q}$. Далее вычисляются $Z_{\min}$ и $Z_{\max}$. Если

$$(|Z_{\max} - Z_{\min}|) < \varepsilon, \tag{3}$$

где ε — заданная точность, то гипотеза об устойчивой эмпирической частоте представляется вполне обоснованной, причем в качестве представителя среднего может быть выбрано любое Z_i из интервала $Z_{\min}, Z_{\max}$. Пусть N и n соответственно объем новой и старой выборок данных, а Q и q — число групп, на которые были разбиты новая и старая выборки данных. Если при $N > n$ и $Q > q$, новые эмпирические частоты будут удовлетворять неравенству (3), то гипотеза об устойчивости частот будет оправдана. Если данных мало, то используем схему удлиняющей выборки. Часть данных разбивается минимум на две группы, и для этих групп вычисляются эмпирические частоты. Затем в выделенные группы добавляем новые данные, и для полученных групп вычисляем частоты. В остальном алгоритмы для этих схем идентичны.

Прикладника интересует, имеет ли место устойчивость частоты $\omega_n(a)$ события a. Для эмпирической проверки сходимости он использует одну из ранее описанных схем. К достоинствам математики в приложениях относится то, что ее применение уменьшает объем рутинной работы для получения искомых результатов. Какой выигрыш мы получаем, используя заключение теоремы закона о больших числах? Для того, чтобы с помощью частотной концепции на эмпирическом материале проверить справедливость неравенства (1) необходимо иметь не n данных, как в схеме многих выборок для проверки устойчивости частоты, а гораздо больше порядка nq, где q — число групп, использованных в схеме многих выборок. Таким образом, для верификации неравенства (1) на эмпирическом материале (для того, чтобы только опосредованным образом с помощью верификации неравенства (2) убедиться в устойчивости частот) необходимо в существенной степени иметь больше данных, чем при прямой проверке, и соответственно возрастает объем вычислений. А.Н.Колмогоров назвал такое положение дел логическим зацикливанием [8]. В работе Ю.И.Алимова эта ситуация получила название первого парадокса теории вероятностей [2]. Для того, чтобы убедиться в устойчивости частот с помощью теоремы закона больших чисел, необходимо знать a priori вероятность события, являющейся теоретической величиной. Назовем это вторым парадоксом теории вероятностей. В приведенной формулировке теоремы говорится об n наблюдениях одного события a. По существу, данные наблюдений представляют одно событие, или одну случайную величину. Для того, чтобы воспользоваться аппаратом теории вероятностей и проверить независимость наблюдений, необходимо одно реально наблюдаемое событие, или одну случайную величину, трансформировать в множество событий, или соответственно множество случайных величин. Классическая статистика не предлагает конструктивных правил для такого рода операций. Дадим конструктивную формулировку теоремы закона больших чисел. Для этого необходимо алгоритмически задать выбор последовательностей, на которых осуществляется проверка независимости. Для того, чтобы показать нетривиальность проверки независимости, а также того, что в отличие от классической статистики независимость последовательностей не влечет устойчивости результатов, дадим конструктивное определение независимости по Виллю. Пусть задан протокол-индикатор события a : $\{X_a(s)\}$, $s = 1\ \infty$. Протокол-индикатор: $\{X_a(s)\}$, $s = 1, \infty$, — это в простейшем случае последовательность нулей и единиц. Единица соответствует появлению события a, нуль — не появлению события a. Формируем на каждом шаге $r = 1, 2, \dots, L, \dots$ выборку объема n следующим образом:

$$S(r, n) = \{X_a(r - 1 + j) | j = \overline{1, n}\},\ r = 1, 2, \dots \qquad (4)$$

Рассмотрим комбинацию δ_{n-1}, из $n-1$ двоичных цифр. Рассчитаем условную вероятность: $P(1|\delta_{n-1})$ следования 1 за фиксированной комбинацией δ_{n-1} в выборке $S(r,n)$.

Пусть $M_r(\delta_{n-1})$ означает встречаемость комбинации δ_{n-1} на шаге r.

Соответственно с помощью $M_r(1|\delta_{n-1})$ обозначим встречаемость 1 после комбинации δ_{n-1}.

Теперь вычисляем частоту следования 1 на шаге r после комбинации δ_{n-1} :

$$W_r(1|\delta_{n-1}) = \frac{M_r(1|\delta_{n-1})}{M_r(\delta_{n-1})}.$$

Если $W_r(1|\delta_{n-1})$, $r=\overline{1,q}$, близки с заданной точностью, то с помощью алгоритма многих выборок или удлиняющихся выборок, используя $W_r(1|\delta_{n-1})$, $r=\overline{1,q}$, находим $W_q(1|\delta_{n-1})$, котоpя является точечной эмпирической частотой.

Если устойчивость сохраняется при больших значениях q, то найдем $P(1|\delta_{n-1}) = \lim\limits_{q\to\infty} W_q(1|\delta_{n-1})$.

Последовательность называется иррегулярной по Виллю–Постникову, если имеет место:

$$P(1|\delta_{n-1}) = P(1), \ \forall n, \forall \delta_{n-1}. \tag{5}$$

Таким образом, в отличие от классической статистики, где независимость автоматически влечет устойчивость, конструктивная формулировка независимости возможна, если предварительно была установлена устойчивость. Устойчивость оказывается более фундаментальным понятием, чем независимость. Теперь приведем конструктивную формулировку теоремы.

ТЕОРЕМА. *Пусть в результате проведенных испытаний получен протокол-индикатор $\{X_a(s)\}$, $s=\overline{1,\infty}$.*

Пусть далее выполнены следующие условия:

1) для последовательности $\{X_a(s)\}$, $s=\overline{1,\infty}$, существует предел

$$p = P(a) = \lim_{m\to\infty} \omega_m(a), \tag{6}$$

2) испытания являются независимыми; это означает, что они удовлетворяют условию (5).

Пусть далее $W_q(n, \varepsilon)$ — частота выполнения неравенства $W_{rn}(a) - p(a)| \le \varepsilon$ для серии выборок вида $S(r, n) = \{X_a(r - 1 + j) | j = 1, n\}$, $r = 1, 2, \ldots, q$. Тогда

$$\lim_{n \to \infty} P(n, \varepsilon) = 1, \ \forall \varepsilon > 0, \tag{7}$$

где $P(n, \varepsilon) = \lim\limits_{q \to \infty} W_q(n, \varepsilon)$.

При метрологической, конструктивной интерпретации теоремы устойчивость частоты не доказывается, а декларируется условием 1. Как видно из выше изложенного, в конструктивном варианте теорема закона больших чисел достаточно тривиальна по смыслу и весьма нетривиальна для проверки. В конструктивном изложении теорема закона больших чисел теряет статус теоремы. Во-первых, заключение теоремы, записанное с помощью выражения (7), не содержит новой информации по сравнению с условием теоремы, записанным с помощью выражения (6). Во-вторых, второе условие в конструктивном варианте теоремы не является обязательным для проверки устойчивости. В руководствах по классической статистике наоборот считается, что она не тривиальна по смыслу, и не уделяется внимание проблемы верификации условий и заключения этой теоремы.

Таким образом, многие результаты классической статистики имеют исключительно внутри математическую ценность, и по принципиальным соображениями не могут быть использованы в приложениях. Несмотря на это математики не уделяют внимание исследованиям альтернативных концепций статистики, в частности мизесовскому анализу. Как справедливо заметил В.Н.Тутубалин, с одной стороны, мизесовский язык мертв, на нем математики не говорят, с другой стороны, практики применяют статистический анализ по Мизесу [9]. Для изменения ситуации необходима заинтересованность работающих математиков в развитии мизесовского анализа. В частности, необходимо сформулировать на современном математическом языке нерешенные Мизесом проблемы, решение которых сделает мизесовскую концепцию обоснованной современной математической теорией, обладающей, в отличие от многих других теорий, несомненной практической значимостью.

Литература

1. КОЛМОГОРОВ А.Н. Основные понятия теории вероятностей. — М., 1974.

2. АЛИМОВ Ю.И., КРАВЦОВ Ю.А. Является ли вероятность "нормальной" физической величиной? //Успехи физических наук. — 1992. — Т. 162, № 7. — С. 149–182.

3. АЛИМОВ Ю.И. Альтернатива методу математической статистики. — М., 1980.

4. ЭЛЬЯСБЕРГ П.Е. Измерительная информация: сколько ее нужно? Как ее обрабатывать? — М., 1986.

5. КОЛМОГОРОВ А.Н. Основные понятия теории вероятностей. — М., 1975.

6. ГНЕДЕНКО Б.В. Курс теории вероятностей. — М., 1988.

7. БЕРНУЛЛИ Я. О законе больших чисел. — М., 1986.

8. КОЛМОГОРОВ А.Н. Теория вероятностей //Математика, ее содержание, методы и значение. — М., 1956.

9. ТУТУБАЛИН В.Н. Теория вероятностей. — М., 1972.

Поступила в редакцию
27 декабря 2000 года