ISSN 0869-8325

ТОМ 1
Выпуск 2

Серия «ВЕРОЯТНОСТЬ И СТАТИСТИКА»

ОБОЗРЕНИЕ ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ МАТЕМАТИКИ

Научное издательство «ТВП» • МОСКВА
1994

ОБОЗРЕНИЕ ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ МАТЕМАТИКИ
Том 1 **Выпуск 2**
1994

ВОЛОДИН И. Н., НОВИКОВ А. А., СИМУШКИН С. В.

ГАРАНТИЙНЫЙ СТАТИСТИЧЕСКИЙ КОНТРОЛЬ КАЧЕСТВА: АПОСТЕРИОРНЫЙ ПОДХОД

Введение

Статистический контроль качества включает в себя два аспекта: контроль за стабильностью технологического процесса производства и принятие решения о кондиционности выпускаемой продукции или, более широко, аттестацию ее качества. Мы будем рассматривать только второй аспект этой проблемы, традиционно решаемой в рамках теории проверки статистических гипотез и теории оценивания.

С. Н. Бернштейн [2], по-видимому, был первым, по крайней мере в отечественной литературе, кто обратил внимание на неудовлетворительное решение проблемы гарантийности статистического контроля качества в рамках классического (небайесовского) подхода, основанного на управлении доверительным коэффициентом при аттестации качества выпускаемой продукции. Аналогичная ситуация, естественно, имеет место и при решениях о кондиционности продукции с гарантированными ограничениями на риски изготовителя и потребителя (вероятности ошибок первого и второго рода). Суть претензий можно пояснить на следующем простом примере, который обычно приводил Л. Н. Большев в своих лекциях по вероятностным моделям (1972–1974 гг., Московский университет, не опубликовано).

Предприятие производит выборочные обследования выпускаемых за каждый технологический цикл партий своей продукции, используя некоторую процедуру контроля с риском потребителя 1%. Если предприятие всегда производит некондиционную продукцию, то ее определенная доля, но не более 1%, будет, тем не менее, поступать потребителю. Но в таком случае с точки зрения потребителя процедура контроля не является гарантийной, поскольку ему всегда поставляется продукция, не удовлетворяющая стандарту на качество, — истинный, а не мифический риск потребителя равен 100%. Естественно, случись такое на реальном производстве, оно будет немедленно остановлено. Однако, это произойдет не в соответствии с решающим правилом процедуры контроля, и тем самым будет признано, что существующая процедура неприемлема с точки зрения ее гарантийности — необходимо управлять не вероятностью пропуска продукции неудовлетворительного качества, а долей кондиционной продукции среди принятой, т.е. управлять выходным уровнем качества $Q_{\text{вых}}$. Решить такую задачу можно лишь в рамках байесовской модели производства продукции и проверки ее качества.

Одно из возможных решений проблемы реальной гарантийности процедуры контроля дает метод «последующих оценок» $Q_{\text{вых}}$, которые вычисляются периодически по мере накопления архива испытаний (см. [1]). Эти оценки предназначаются для корректировки действующей процедуры контроля таким образом, чтобы в последующем $Q_{\text{вых}}$ процедуры было приемлемо с точки зрения потребителя. Естественно, все это делается в предположении, что входной уровень $Q_{\text{вх}}$ качества производимой продукции остается неизменным, а процесс производства статистически подконтрольным.

В подходе Л. Н. Большева к этой проблеме предлагается изначально строить процедуру контроля с гарантированным уровнем выходного качества ($Q_{\text{вых}} \geqslant 1 - \alpha_0$, где α_0 — априори заданное малое число), оптимизируя при этом некоторые другие характеристики процедуры (ряд подобных задач решается в [12], [14]). Последующие оценки в такой постановке позволяют строить эмпирические (по Роббинсу) аналоги оптимальных процедур гарантийного контроля при известном априорном распределении контролируемого параметра.

В данной статье делается обзор публикаций авторов за последнее десятилетие, посвященных развитию идей Л.Н.Большева в области контроля качества и аттестации выпускаемой продукции. Кроме теоретических аспектов этого подхода приводятся результаты по решению следующих представляющих практический интерес задач: 1) построение процедур контроля с фиксированным объемом испытаний, которые гарантируют заданный уровень $1 - \alpha_0$ выходного качества $Q_{\text{вых}}$ и в то же время максимизируют уровень контроля $Q_{\text{конт}}$ — долю некондиционной продукции среди отклоненной; 2) характеризация и построение процедур с минимальным объемом испытаний, который требуется для достижения заданных уровней $1 - \alpha_0$ и $1 - \alpha_1$ для $Q_{\text{вых}}$ и $Q_{\text{конт}}$; 3) построение процедур аттестации выпускаемой продукции, которые гарантируют вероятность ее кондиционности в случае приобретения потребителем и обладают минимальной вероятностью аттестационной ошибки, состоящей в том, что продукт в действительности кондиционен, но его паспортные данные не удовлетворяют потребителя.

Следующие примеры, один из которых имеет непосредственное отношение к реальной практике машиностроительного производства, иллюстрируют круг задач, обсуждаемых (и частично решаемых) в данном обзоре.

1. Контроль плавки металла по предельной прочности. Плавка металла или металлическое изделие с припуском, из которого отбираются образцы, контролируется по предельной прочности — минимальной нагрузке на мм^2, при которой происходит разрыв образца. Металл плавки считается кондиционным, если значение θ его предельной прочности превосходит предписанное значение θ_0, называемое обычно «нормой». В условиях стабильного производства естественно предположить, что значение θ для каждой конкретной плавки представляет реализацию некоторой случайной величины ϑ и изменчивость значений θ от плавки к плавке обуславливается вариацией условий, при которых производится металл. Точно так же можно считать, что результат x испытания конкретного образца металла есть реализация случайной величины X и изменчивость x от образца к образцу в теле плавки обуславливается, в основном, так называемыми «ошибками испытаний».

Любая процедура контроля состоит из некоторого правила, которое определяет количество отбираемых образцов, и правила принятия решения о кондиционности металла контролируемой плавки по результатам их испытаний. Согласно объявленной выше концепции гарантийности мы считаем основной характеристикой процедуры контроля вероятность $Q_{\text{вых}}$ события $\vartheta \geqslant \theta_0$ при условии, что плавка металла прошла контроль (принято решение о кондиционности металла). При любом количестве испытуемых образцов правило принятия решения следует выбирать таким образом, чтобы $Q_{\text{вых}}$ было не меньше заданного уровня $1-\alpha_0$. Вероятность $Q_{\text{вых}}$ обычно называют *выходным уровнем качества* и сопоставляют с *входным уровнем качества* $Q_{\text{вх}}$ — априорной (безусловной) вероятностью события $\vartheta \geqslant \theta_0$; отношение $Q_{\text{вых}}/Q_{\text{вх}}$ определяет эффективность процедуры контроля. Если обратиться к частотной интерпретации этих характеристик, то $Q_{\text{вх}}$ определяет долю кондиционных плавок, изготовляемых в процессе производства, а $Q_{\text{вых}}$ — долю кондиционных плавок среди принятых, т.е. отосланных потребителю. Следовательно, $1 - Q_{\text{вых}}$ определяет средний размер выплат по рекламациям. Другой, не менее важной характеристикой контроля является вероятность события $\vartheta < \theta_0$ при условии, что плавка не принята. Эта условная вероятность $Q_{\text{конт}}$ называется *уровнем контроля* и трактуется как доля некондиционных плавок среди отклоненных при контроле.

Процедура контроля называется *гарантийной по выходному уровню качества* (*уровню контроля*), если $Q_{\text{вых}} \geqslant 1 - \alpha_0$ (соответственно $Q_{\text{конт}} \geqslant 1 - \alpha_1$), где $1 - \alpha_0$ (соответственно $1 - \alpha_1$) — заданное ограничение на указанную характеристику. Процедура контроля называется *гарантийной*, или *полностью гарантийной*, если выполняются оба неравенства $Q_{\text{вых}} \geqslant 1 - \alpha_0$ и $Q_{\text{конт}} \geqslant 1 - \alpha_1$.

Естественно, для того чтобы вычислить $Q_{\text{вых}}$ и $Q_{\text{конт}}$ для действующей процедуры или построить новую процедуру, обладающую каким-либо преимуществом по сравнению с действующей, необходимо знать совместное распределение случайных величин ϑ и X. Информацию об этом распределении поставляет постепенно накапливаемый архив данных по испытаниям плавок, так что в принципе о гарантийности процедуры можно говорить только по истечении определенного промежутка времени с начала производства. К данным архива следует также подключить информацию о физических явлениях, определяющих изменчивость θ от плавки к плавке и изменчивость x от образца к образцу, т. е. построить вероятностную модель изменчивости предельной прочности и ее измерений (с последующей проверкой адекватности модели данным испытаний).

Прочность металла в каждой плавке при устоявшейся технологии его производства целиком зависит от структурных особенностей его кристаллической решетки, более того, установлено, что предельная прочность прямо пропорциональна числу «правильных» кристаллов в плавке, что позволяет выдвинуть гипотезу о нормальности маргинального распределения случайной величины ϑ, которое в рамках подобных моделей принято называть априорным. Изменчивость же результатов испытаний различных образцов одной и той же плавки обуславливается не только (точнее, не столько) различием числа кристаллических дефектов в испытуемых образцах, а, в основном, их неидентичностью по отношению друг к другу, неравномерностью подачи нагрузки при их растяжении и прочими «ошибками» испытаний и измерений. Поэтому естественно предположить, что условное распределение случайной величины X при каждом фиксированном значении параметра θ также является нормальным.

Итак, мы пришли к гипотезе о справедливости так называемой *модели $\mathcal{N}$–$\mathcal{N}$*: распределение ϑ есть $\mathcal{N}(\mu, \tau^2)$ с плотностью

$$g(\theta) = \frac{1}{\sqrt{2\pi}\tau} \exp\left\{ -\frac{(\theta-\mu)^2}{2\tau^2} \right\},$$

а условное распределение X при $\vartheta = \theta$ есть $\mathcal{N}(\theta, \sigma^2)$ с плотностью

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\theta)^2}{2\sigma^2} \right\}.$$

В этой модели параметры μ и τ характеризуют общее состояние прочностных свойств металла данной марки, а параметр σ — точностные свойства метода испытаний. Не исключено, что в σ заложена изменчивость прочности по телу плавки, и в таком случае норму θ_0 следует «подправить». Например, рабочую норму $\widetilde{\theta}_0$ по заданной θ_0 следует находить из уравнения

$$\mathbf{P}\{X > \theta_0 \mid \vartheta = \widetilde{\theta}_0\} = 1 - \Phi\left(\frac{\theta_0 - \widetilde{\theta}_0}{\sigma}\right) = 1 - \gamma,$$

где γ — заданная малая вероятность. Тогда $\widetilde{\theta}_0 = \theta_0 + \sigma\Phi^{-1}(1-\gamma)$. Здесь $\Phi(x)$ — функция стандартного нормального распределения, $\Phi^{-1}(p)$ — квантиль уровня p этого распределения.

Проверку модели $\mathcal{N}$–$\mathcal{N}$ можно провести по данным архива

$$\begin{array}{cccc} x_{11}, & x_{12}, & \ldots, & x_{1n_1} \\ \ldots & \ldots & \ldots & \ldots \\ x_{m1}, & x_{m2}, & \ldots, & x_{mn_m} \end{array}$$

контроля m плавок с испытанием n_k образцов в k-й плавке, $k = 1, \ldots, m$. Для этого достаточно сформировать выборку объема m, скажем, из данных испытаний первых образцов $x_{11}, \ldots, x_{m1}$ в каждой плавке и проверить ее на нормальность. Согласно модели $\mathcal{N}$–$\mathcal{N}$ эти данные представляют выборку из нормального распределения с параметрами μ и $\sigma^2 + \tau^2$, и если такая гипотеза подтверждается, то в силу известной теоремы Крамера (см. [13, с. 66, теорема 19]) подтверждается как гипотеза о нормальности априорного распределения ϑ, так и гипотеза о нормальности условного распределения X, т. е. подтверждается справедливость модели $\mathcal{N}$–$\mathcal{N}$.

Мы располагали архивом данных по контролю $m = 560$ плавок. Из каждой плавки первоначально отбирались два образца. Если оба результата испытаний превосходили заданное число C, то наблюдения прекращались. В противном случае дополнительно испытывались еще четыре образца. Таким образом, в нашем архиве каждое n_k, $k = 1, \ldots, m$, являлось реализацией случайной величины и равнялось либо 2, либо 6.

Мы проверяли гипотезу о нормальности выборок $x_{11}, \ldots, x_{m1}$ и $x_{12}, \ldots, x_{m2}$ по критерию хи-квадрат. Согласие с моделью $\mathcal{N}$–$\mathcal{N}$ (критический уровень значимости) для первой выборки было равно 20%, для второй — 60%.

В настоящее время для контроля металла и металлических изделий по механическим характеристикам на большинстве машиностроительных предприятий используется следующая двухступенчатая процедура контроля: сначала испытываются n_1 образцов (обычно $n_1 = 2$) и полученные данные сравниваются с приемочным числом C. Если $\min_{1 \leqslant i \leqslant n_1} x_i \geqslant C$, то плавка принимается, если же выполняется противоположное неравенство, то дополнительно испытываются $n_2 - n_1$ образцов (обычно 4, так что $n_2 = 6$) и данные $x_{n_1+1}, \ldots, x_{n_2}$ испытаний этих образцов сравниваются с тем же числом C. Если $\min_{n_1+1 \leqslant i \leqslant n_2} x_i \geqslant C$, то плавка принимается, в противном же случае принимается решение о некондиционности плавки по значению наблюдаемой механической характеристики.

Вероятность приемки плавки с прочностью θ при нормальном условном распределении наблюдаемой случайной величины равна

$$L(\theta) = \mathbf{P}_\theta\{X_1 \geqslant C, \ldots, X_{n_1} \geqslant C\} + \left[1 - \mathbf{P}_\theta\{X_1 \geqslant C, \ldots, X_{n_1} \geqslant C\}\right]$$

$$\times \mathbf{P}_\theta\{X_{n_1+1} \geqslant C, \ldots, X_{n_2} \geqslant C\} = \left[1 - \Phi\left(\frac{C - \theta}{\sigma}\right)\right]^{n_1}$$

$$+\left[1-\left(1-\Phi\left(\frac{C-\theta}{\sigma}\right)\right)^{n_1}\right]\left[1-\Phi\left(\frac{C-\theta}{\sigma}\right)\right]^{n_2-n_1}.$$

Функцию $L(\theta)$ обычно называют ***оперативной характеристикой*** процедуры контроля. Другие, более важные с точки зрения нашего подхода к проблеме гарантийности характеристики двухступенчатой процедуры вычисляются с использованием формулы Байеса:

$$\begin{aligned}
Q_{\text{вх}} &= \mathbf{P}\{\vartheta \geqslant \theta_0\} = 1-\Phi\left(\frac{\theta_0-\mu}{\tau}\right),\\
Q_{\text{вых}} &\doteq \mathbf{P}\{\vartheta \geqslant \theta_0 \mid \text{плавка принята}\}\\
&= \frac{\frac{1}{\sqrt{2\pi}\tau}\int_{\theta_0}^{\infty} L(\theta)\exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}d\theta}{\frac{1}{\sqrt{2\pi}\tau}\int_{-\infty}^{\infty} L(\theta)\exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}d\theta},\\
Q_{\text{конт}} &= \mathbf{P}\{\vartheta < \theta_0 \mid \text{плавка отклонена}\}\\
&= \frac{\frac{1}{\sqrt{2\pi}\tau}\int_{-\infty}^{\theta_0}(1-L(\theta))\exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}d\theta}{\frac{1}{\sqrt{2\pi}\tau}\int_{-\infty}^{\infty}(1-L(\theta))\exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}d\theta}.
\end{aligned}$$

Для оценки параметров μ, τ^2 и σ^2 использовался тот же, что и при проверке модели $\mathcal{N}$–$\mathcal{N}$, архив данных. По результатам испытаний только первых двух образцов каждой плавки в рамках сбалансированной однофакторной модели II дисперсионного анализа со случайными факторами (см., например, [3, § 10.5, с. 293–295]) получены следующие значения оценок для параметров априорного распределения: $\widehat{\mu} = 122,07$, $\widehat{\tau} = 2,67$, $\widehat{\sigma} = 1,18$.

При норме $\theta_0 = 119,00$ подстановка этих значений в уравнения для $Q_{\text{вх}}$, $Q_{\text{вых}}$ и $Q_{\text{конт}}$ (см. выше) дала следующие значения состоятельных (с объемом архива $m \to \infty$) оценок характеристик двухступенчатой ($n_1 = 2$, $n_2 = 6$, $C = 119,00$) процедуры контроля: $\widehat{Q}_{\text{вх}} = 0,875$, $\widehat{Q}_{\text{вых}} = 0,989$, $\widehat{Q}_{\text{конт}} = 0,595$.

Если отказаться от испытаний дополнительных 4-х образцов на второй ступени и использовать для контроля одноступенчатую процедуру по минимальному значению в выборке из двух наблюдений, то $\widehat{Q}_{\text{вых}} = 0,989$, $\widehat{Q}_{\text{конт}} = 0,587$. Следовательно, введение второй ступени практически не оказывает влияния на характеристики процедуры.

Как показано в [17] (см. также пример после теоремы 2.3 в § 2), в случае нормального распределения величины X при фиксированном θ и любого априорного распределения ϑ процедура с фиксированным объемом испытаний n, гарантирующая заданный уровень $1-\alpha_0$ выходного качества и максимизирующая уровень контроля $Q_{\text{конт}}$, должна быть основана на выборочном среднем $\overline{x}$: плавка принимается, если $\overline{x} > C$, где приемочное число C определяется из уравнения $Q_{\text{вых}}(C) = 1-\alpha_0$.

В случае модели $\mathcal{N}$–$\mathcal{N}$ с использованием оценок параметров априорного распределения вместо их точных (неизвестных) значений такая процедура асимптотически гарантирует заданный уровень $1-\alpha_0$ для $Q_{\text{вых}}$ и асимптотически максимизирует величину $Q_{\text{конт}}$ при неограниченном возрастании объема архива m. Для данных нашего архива с $n=2$ и $1-\alpha_0=0,99$ (т.е. для фактического уровня выходного качества у рассмотренной выше двухступенчатой процедуры) приемочное число оптимальной процедуры C есть 119,5, а $Q_{\text{конт}}=0,639$.

Процедура контроля, основанная на $\overline{x}$, является также оптимальной (асимптотически оптимальной) среди всех полностью гарантийных процедур с фиксированным объемом испытаний. Если потребовать $1-\alpha_0=0,99$, $1-\alpha_1=0,90$, то по данным нашего архива необходимое число испытаний n равно 12, а приемочное число C равно 119,02; при ограничениях $1-\alpha_0=0,95$ и $1-\alpha_1=0,90$ необходимый объем выборки n равен 2, а $C=118,32$ (!). Мы обсудим этот замечательный факт (приемочное число ниже нормы) в § 5; там же см. продолжение данного примера в части гарантийной аттестации плавки металла по предельной прочности.

О последовательных гарантийных процедурах контроля, обладающих высокой вероятностью остановки после испытаний первого образца, см. § 4.

Отметим, что приведенные выше расчеты носят в основном иллюстративный характер. В действительности норма θ_0 равна 110, так что оценка $\widehat{Q}_{\text{вх}}$ равна 0,999, из чего можно сделать вывод о нецелесообразности контроля плавок по предельной прочности по крайней мере до тех пор, пока технологический процесс статистически подконтролен. Однако, на тех же образцах в процессе испытаний замеряются еще три механические характеристики металла, две из которых (относительные сужение и удлинение) имеют резко асимметричные априорные распределения, так что гарантийный контроль качества в рамках модели $\mathcal{N}$–$\mathcal{N}$ становится невозможным.

2. Контроль по альтернативному признаку. Классическая вероятностная модель контроля по альтернативному (или, как иногда говорят, по качественному) признаку связана со случайным отбором изделий из партии конечного объема N с последующим разделением отобранных изделий на два класса — годные и дефектные. Общее число D дефектных изделий в партии неизвестно; имеется ограничение θ_0 на допустимую долю $\theta=D/N$ дефектных изделий, и статистическая задача контроля качества состоит в различении двух гипотез: $\theta>\theta_0$ (партия отклоняется от приемки) и $\theta\leqslant\theta_0$ (партия принимается — отсылается потребителю).

Если формирование партий изделий не соотносится с технологическими циклами в их изготовлении, а происходит путем простого накопления определенного количества «штампуемых» изделий, то вполне разумно выдвинуть гипотезу о биномиальном априорном распределении величины D. При справедливости такой гипотезы каждое изделие с одинаковой вероятностью p может быть дефектным, причем, пока действует выбранная процедура контроля, величина p должна оставаться постоянной. Различные планы контроля в рамках такой гипергеометрическо-биномиальной модели подробно разбираются в монографии [1] (советуем особо обратить внимание на утверждения § 2.3). Замечательно, что в рамках этой модели $Q_{\text{вх}} = Q_{\text{вых}} = p$ при любой процедуре контроля, т.е. любой контроль (кроме, естественно, контроля за стабильностью технологического процесса производства изделий) обессмысливается — он не способен изменить долю дефектных изделий в принятой партии. К сожалению, до сих пор на практике бытует большое число стандартизированных методик контроля по альтернативному признаку, в которых умалчивается об априорном распределении D, а гарантийность контроля соотносится с ограничениями на риски потребителя и изготовителя.

Существенно другая вероятностная картина возникает в ситуации, когда формирование партии изделий соотносится с определенным технологическим циклом, например, когда партия содержит металлические изделия одной и той же плавки и контролируемый признак связан с прочностными свойствами металла. Рассмотрим для простоты случай, когда объем N партии значительно превосходит объем инспектируемых изделий. В таком случае процедура контроля осуществляется в схеме испытаний Бернулли с ностоянной вероятностью θ «успешного» испытания (вероятностью дефектности наугад взятого образца из партии изделий). Решение о принятии партии признается безошибочным, если $\theta \leqslant \theta_0$, где θ_0 — предельная допустимая доля дефектных изделий в партии.

В отличие от предыдущих примеров в данной ситуации зачастую не представляется возможным использовать какие-либо физические законы при спецификации априорного распределения θ. Однако на отрезке [0,1] возможных значений θ существует удивительно богатое семейство распределений, члены которого могут обладать правой или левой асимметрией, отрицательным или положительным эксцессом, могут иметь U-образную форму и пр. Это семейство бета-распределений с плотностью

$$g(\theta) = \frac{1}{B(\lambda_1, \lambda_2)} \theta^{\lambda_1 - 1}(1-\theta)^{\lambda_2 - 1}, \qquad 0 \leqslant \theta \leqslant 1; \quad \lambda_1 > 0, \quad \lambda_2 > 0$$

(об этой модели и других, используемых в практике контроля, см. монографию [1]). Параметры λ_1 и λ_2 бета-распределения довольно просто оцениваются по архиву данных с помощью метода моментов.

Как следует из результатов статьи [17] (см. также § 2 данного обзора), при любом (последовательном, многоступенчатом и пр.) плане отбора изделий из партии решение о ее приемке должно быть основано на числе X дефектных изделий в выборке: партия принимается, если $X \leqslant C$. При фиксированном плане отбора приемочное число C определяется по заданному ограничению $1-\alpha_0$ на $Q_{\text{вых}}$; полностью гарантийная процедура требует спецификации плана отбора и определения C по совокупности ограничений $1-\alpha_0$ и $1-\alpha_1$ на $Q_{\text{вых}}$ и $Q_{\text{конт}}$ соответственно.

Отметим, что в этом примере при оптимизации характеристик контроля, как и в теории Неймана–Пирсона (см. [10, гл. 9]), существенную роль играет рандомизация принятия решения о приемке партии в случае $X = C$. Более подробное описание рандомизационной процедуры контроля в рамках модели Bi–Be (биномиальное распределение X и бета-распределение ϑ) с выводом приближенных формул для расчета n и C в случае, когда априорное бета-распределение имеет ярко выраженный максимум в граничной точке θ_0, дано в § 4 данного обзора.

Приведенные примеры дают представление лишь о той части нашего обзора, которая касается приложений апостериорного подхода к проблемам гарантийности контроля качества, в то время как его основное содержание посвящено математическим обоснованиям понятия асимптотической (по мере накопления архива) гарантийности и построению оптимальных процедур при «известном априорном распределении». Мы выделили в кавычки последнюю фразу, понимая, что это исключительно редкий случай в практике контроля качества, но это тот фундамент, как критерий Неймана–Пирсона различения простых гипотез, опираясь на который, можно строить эмпирические процедуры контроля, основанные на оценках априорного распределения по архиву данных и обладающие определенной асимптотической оптимальностью.

Наши публикации в этой области носят, в основном, весьма абстрактный характер и направлены на построение общей теории статистического вывода на последовательностях статистических экспериментов (см. для сравнения близкую нам в идейном плане статью Хилла [18]). Статистический контроль качества — это тот объект приложения наших построений, в котором статистик действительно имеет дело с реальной (а не воображаемой, как в большинстве «научных» приложений) последовательностью статистических экспериментов. Поэтому основная цель данного обзора — изложить общую теорию апостериорного подхода применительно к простому и наглядному объекту статистических исследований.

В § 1 вводится понятие адаптивной стратегии — последовательности статистических процедур, применяемых для обследования последовательно поступающих объектов контроля. Каждая процедура определяется правилом остановки статистического эксперимента при контроле очередного объекта и правилом принятия решения о его кондиционности. Эти правила зависят как от результатов текущего эксперимента, так и от данных обследования предшествующих объектов контроля. Такие данные образуют «архив», постепенное накопление которого дает возможность с возрастающей точностью и надежностью оценить априорное распределение контролируемого параметра и тем самым обеспечить гарантийность адаптивной стратегии контроля качества.

Понятие гарантийности основано на определении d-рисков стратегии: имея дело с реальной последовательностью статистических экспериментов и принимаемых решений, мы вводим d-риски как доли ошибочных решений в подпоследовательностях экспериментов, завершаемых соответственно принятием (решение d^0) или отклонением (решение d^1) контролируемой продукции. Стратегия является гарантийной, если эти доли не превосходят заданных уровней α_0 и α_1, или, что то же, $Q_{\text{вых}} \geqslant 1 - \alpha_0$, $Q_{\text{конт}} \geqslant 1 - \alpha_1$.

В этом же параграфе дается постановка задачи на оптимальный выбор стратегии и указывается схема решения такой задачи. По своей сути эта схема восходит к эмпирическому байесовскому подходу Роббинса. Сначала мы решаем задачу на оптимум при известном априорном распределении, а затем строим эмпирический, основанный на оценках d-рисков по архиву данных, аналог «простой» оптимальной стратегии, т.е. стратегии, состоящей из последовательности одинаковых процедур с правилами, зависящими от данных только текущего эксперимента. Таким образом, простая стратегия является по существу обычной процедурой статистического вывода, оптимизация которой открывает путь к построению оптимальной стратегии контроля.

В § 2 рассматривается задача оптимизации простых стратегий — процедур контроля — при известном априорном распределении. Устанавливается своеобразный аналог «леммы Неймана–Пирсона», в котором указывается вид правила принятия решения, гарантирующего при любом заданном правиле остановки требуемый уровень $1 - \alpha_0$ выходного качества и максимизирующий $Q_{\text{конт}}$. Понятно, что это решает проблему построения гарантийной процедуры контроля с минимальным объемом выборки в классе всех процедур, правила остановки которых не зависят от результатов наблюдений (процедур с фиксированным объемом испытаний). Для некоторых частных случаев в рамках модели $\mathcal{N}$–$\mathcal{N}$ (см. п. 1 введения) выводятся явные формулы для необходимого объема выборочных обследований.

Что же касается минимизации среднего объема наблюдений в классе всевозможных последовательных гарантийных процедур контроля, то этот аналог классической задачи Вальда до сих пор не имеет решения. В § 2 устанавливаются параллели с соответствующими последовательными процедурами отношения правдоподобия в классической теории проверки гипотез — такой параллелизм проливает свет на основную трудность в решении задачи Вальда в рамках апостериорного подхода к проблеме гарантийности процедур контроля качества.

В § 3 устанавливается правомочность выбранной схемы построения оптимальных стратегий: показывается, что введение эмпирических аналогов простых стратегий позволяет строить оптимальные адаптивные стратегии в соответствии с определениями § 1. Основное содержание параграфа посвящено анализу асимптотической (при неограниченном возрастании объема архива) гарантийности эмпирических стратегий и наследованию оптимальных свойств адаптивными стратегиями, «порожденными» эмпирическими аналогами оптимальных простых стратегий.

В § 4 мы возвращаемся к простым стратегиям и исследуем асимптотику необходимого объема наблюдений в случае, когда априорное распределение имеет резко выраженный максимум в точке θ_0, разграничивающей проверяемые гипотезы $\theta \leqslant \theta_0$ и $\theta > \theta_0$, т.е. когда априорное распределение сосредотачивается на границе «нормы» и неопределенность в выборе решения о приемке продукции достигает своего апогея. Этот случай аналогичен схеме сближающихся гипотез (контигуальных альтернатив) в классической теории статистического вывода, и, может быть, именно поэтому здесь удается продвинуться в решении оптимальных задач значительно дальше, чем в § 2. Находится асимптотика необходимого объема наблюдений в классе гарантийных процедур с фиксированным числом испытаний. В классе всех последовательных гарантийных процедур приводится суррогат асимптотического решения задачи Вальда — минимизируется не средний объем наблюдений, а некоторые функционалы, характеризующие количество информации, поступающей из данных эксперимента в пользу гипотезы кондиционности (соответственно, некондиционности) контролируемой продукции.

Наконец, § 5 посвящен вопросам аттестации (паспортизации) выпускаемой продукции. В этом разделе мы пунктуально следуем идеям С. Н. Бернштейна [2] в постановке задачи доверительного оценивания; решение проблемы оптимальной доверительной аттестации есть компромисс между требованиями покупателя и желанием производителя сбыть максимум продукции. Для простоты рассматривается только задача построения верхней доверительной границы. Как и в классическом случае, в апостериорном подходе задача доверительного оценивания двойственна задаче проверки гипотез: периодически приобретая выпускаемую продукцию, покупатель должен быть уверен, что доля «хорошей» продукции среди приобретенной не опускается (в среднем) ниже номинального уровня $1-\alpha_0$, в то время как производитель стремится минимизировать долю «хорошей» продукции среди отклоненной покупателем по причине ошибочной аттестации. В такой постановке задача оптимизации доверительного оценивания по существу решена в § 2. Асимптотические результаты § 4 дают решения соответствующих задач по построению асимптотически оптимальных доверительных границ. Исследуются также соотношения найденных границ с байесовскими доверительными границами, которые определяются как квантили апостериорного распределения. Байесовские границы далеки от оптимальных, что особенно заметно при больших объемах испытаний.

Данный обзор посвящен только результатам, имеющим отношение к проблемам контроля качества и аттестации выпускаемой продукции. О других исследованиях в области апостериорного подхода к проблеме статистического вывода см. [6], [7], [19], [20]. В своем обзоре мы не касаемся других многочисленных исследований по проблемам контроля качества. Традиционная тематика этой области приложений математической статистики достаточно полно отражена в монографии [1]: новые исследования, за исключением разве лишь применений теории полезности и субъективного байесовского вывода, лежат в русле идей, затронутых в этой книге.

§ 1. Гарантийность статистической процедуры контроля. Постановка задачи оптимального вывода

Характерной особенностью контроля качества серийного производства является то, что одна и та же статистическая задача возникает многократно, и, как показывают примеры из введения, решение о качестве контролируемого в данный момент объекта должно зависеть не только от результатов обследования этого объекта, но и от данных ранее проведенных испытаний аналогичных объектов. Чтобы пояснить смысл нашей концепции гарантийности процедур такого типа, обратимся к классической (в духе Неймана–Пирсона) теории статистического вывода, в которой функция риска определяется как математическое ожидание потерь. Обычная интерпретация этого понятия и возникающего в связи с ним определения гарантийности состоит в искусственном домысливании некой бесконечной последовательности статистических экспериментов и в последующем сопоставлении величины риска процедуры с пределом средних арифметических потерь. При статистическом контроле серийной продукции такая бесконечная (или почти бесконечная) последовательность экспериментов существует реально, и поэтому представляется не только естественным, но и необходимым определить риск статистической процедуры контроля непосредственно через частоту ошибочно принятых решений о качестве продукции. Строгое определение соответствующего понятия гарантийности, а также формулировка возникающих в связи с этим оптимизационных задач контроля качества составляют содержание данного параграфа.

Рассмотрим последовательность контролируемых объектов и соответствующую ей последовательность статистических экспериментов, в каждом из которых производится обследование только одного из объектов. Результат k-го эксперимента (обследования k-го объекта) представляет собой выборку $x^{(n_k)} = (x_1, \dots, x_{n_k})$ объема n_k из распределения $F(\cdot \mid \theta_k)$, принадлежащего некоторому параметрическому семейству распределений $\{F(\cdot \mid \theta), \theta \in \Theta\}$. Параметр θ, индексирующий распределения семейства, характеризует качество контролируемого объекта: объект считается кондиционным, если значение θ принадлежит некоторой подобласти Θ_0 параметрического пространства Θ, некондиционному объекту соответствует значение $\theta \in \Theta_1 = \Theta \backslash \Theta_0$. Решение о приемке (кондиционности) объекта обозначим d^0, о некондиционности — d^1.

Предполагается, что значение θ_k параметра θ в k-м эксперименте есть реализация случайной величины ϑ с априорным распределением G, о котором известно только то, что оно принадлежит некоторому семейству распределений $\mathcal{G}$ на Θ.

Процедура контроля k-го объекта состоит из статистических правил остановки статистического эксперимента (прекращения наблюдений) и принятия решения о кондиционности объекта; этим правилам соответствует пара статистик $\langle \nu_k, \delta_k \rangle$, где ν_k — момент остановки (объем выборки), а δ_k — решающая функция, принимающая одно из двух «значений»: d^0 или d^1.

Последовательность $\phi = \{\langle \nu_k, \delta_k \rangle,\, k \geqslant 1\}$ статистических процедур, в которых при любом $k \geqslant 1$ статистики ν_k и δ_k зависят от результатов наблюдений $x^{(n_1)}, \ldots, x^{(n_{k-1})}$ предшествующих экспериментов с номерами $1, 2, \ldots, k-1$, а также от результатов текущего k-го эксперимента, называется *адаптивной стратегией* контроля.

Таким образом, вся последовательность статистических экспериментов характеризуется последовательностью $(\theta_k, x^{(n_k)}, d_k)$, $k = 1, 2, \ldots$, состоящей из ненаблюдаемых значений θ_k параметра θ, результатов испытаний $x^{(n_k)} = (x_{k1}, \ldots, x_{kn_k})$ и принятых решений d_k ($= d^0$ или d^1).

Определим для каждой фиксированной последовательности $\{(\theta_k, d_k),\, k \geqslant 1\}$ возрастающую последовательность $\{k_j,\, j \geqslant 1\}$ номеров тех экспериментов, в которых было принято решение d^0. Кроме того, положим $l_j = 0$, если для эксперимента с номером k_j значение параметра θ_{k_j} принадлежит Θ_0 (принятое решение d^0 правильно), и $l_j = 1$, если $\theta_{k_j} \in \Theta_1$ (решение d^0 ошибочно). В случае, когда последовательность $\{k_j,\, j \geqslant 1\}$ конечна и решение d^0 принималось, скажем, J раз, полагаем $l_j = 0$ для всех $j > J$.

Риском адаптивной стратегии ϕ от принятия решения d^0 назовем величину

$$R_G(d^0; \phi) = \limsup_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} l_j.$$

Отметим, что в общем случае риск представляет собой случайную величину, зависящую от всей бесконечной последовательности реализаций ненаблюдаемого параметра θ и результатов испытаний.

Использованное во введении понятие выходного уровня качества теперь конкретизируется и понимается как $Q_{\text{вых}} = 1 - R_G(d^0; \phi)$.

По определению адаптивная стратегия ϕ *гарантирует уровень α_0 для доли ошибочных решений d^0*, если $R_G(d^0; \phi) \underset{\text{п.н.}}{\leqslant} \alpha_0$ для любого априорного распределения $G \in \mathcal{G}$. Аналогично определяются *гарантийность и риск* $R_G(d^1;\phi)$ $(= 1 - Q_{\text{конт}})$ *стратегии* ϕ *с уровнем* α_1 *для доли ошибочных решений* d^1. Стратегия ϕ, гарантирующая оба уровня α_0 и α_1, называется *гарантийной стратегией контроля.*

Наряду с введенными терминами «риск» и «гарантийность» в дальнейшем будут использоваться также термины «*d-риск*» и «*d-гарантийность*», которые постоянно использовались в наших публикациях с целью противопоставления понятиям классической (с ограничениями на функцию риска) и апостериорной (с ограничениями на апостериорный риск) гарантийности процедур статистического вывода (см. определения в статье [4]).

В рамках данных определений d-риска и d-гарантийности будут решаться следующие традиционные задачи теории оптимального статистического вывода.

(A) Построение стратегий, гарантирующих либо один из уровней $Q_{\text{вых}}$ или $Q_{\text{конт}}$, либо оба уровня одновременно.

(B) Для заданной последовательности моментов остановки построение такой последовательности решающих функций $\{\delta_k,\ k \geqslant 1\}$, что адаптивная стратегия $\phi = \{(\nu_k,\delta_k),\ k \geqslant 1\}$ гарантирует уровень α_0 и доставляет минимум величине риска $R_G(d^1;\phi)$ (гарантирует уровень выходного качества и максимизирует уровень контроля).

(C) Построение стратегии $\phi = \{(\nu_k,\delta_k),\ k \geqslant 1\}$, гарантирующей оба уровня α_0 и α_1 и минимизирующей некоторую характеристику стратегии, связанную с объемом испытаний, например, предел по вероятности

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \nu_k,$$

если он существует.

Если $\mathcal{G}$ — достаточно широкий класс априорных распределений, то один из способов решения сформулированных задач базируется на эмпирическом байесовском подходе Роббинса (см., например, [10, § 6.9]). Реализуя этот подход, мы решаем сначала задачи (A)–(C) в предположении, что априорное распределение известно, а затем строим эмпирический аналог оптимальной стратегии, основанный на оценках априорного распределения G или непосредственно рисков $R_G(d^i,\phi)$ по накопленному архиву данных.

Отметим, что задачу (A) можно решить вне рамок эмпирического байесовского подхода. Пусть Θ_0 — интервал с границей, имеющей априорную меру нуль $\forall G \in \mathcal{G}$, и $\widehat{\theta}_n$ — состоятельная оценка θ по выборке объема n. Тогда любая стратегия, объем выборки которой возрастает до бесконечности с ростом номера эксперимента (например, $\nu_k = k$), а решающая функция $\delta_k = d^i$, если $\widehat{\theta}_k \in \Theta_i$, будет гарантировать любые наперед заданные уровни. Более того, для риска этой стратегии имеем

$R_G(d^i;\phi) \underset{\text{п.н.}}{=} 0,\ i = 0,1.$

§ 2. Характеризация оптимальных стратегий при известном априорном распределении

Следуя предложенной в предыдущем параграфе роббинсовской схеме построения оптимальных адаптивных стратегий, предположим, что априорное распределение G полностью определено. В связи с этим введем класс *простых стратегий*, в которых момент остановки ν и решающая функция δ не изменяются от эксперимента к эксперименту и зависят только от результатов текущего эксперимента; простую стратегию будем обозначать $\phi = \{\nu, \delta\}$. В этом параграфе приводится ряд утверждений, характеризующих решающие функции простых оптимальных стратегий. С помощью этих утверждений удается полностью решить задачу (В), а также задачу (С) в классе процедур с фиксированным объемом испытаний.

Заметим сначала, что в соответствии с законом больших чисел (см. подробности в [17]) для риска простой стратегии ϕ

$$R_G(d^0;\phi) \underset{\text{п.н.}}{=} \mathbf{P}\{\vartheta \notin \Theta_0 \mid \delta = d^0\},$$

$$R_G(d^1;\phi) \underset{\text{п.н.}}{=} \mathbf{P}\{\vartheta \in \Theta_1 \mid \delta = d^1\},$$

где условные вероятности вычисляются по совместному распределению ϑ и δ. Под величиной риска простой стратегии в дальнейшем будем понимать соответствующие правые части последних соотношений.

Аналогично классической теории Неймана–Пирсона различения простых гипотез, где ключевую роль играет критерий отношения правдоподобия, в нашем случае ту же роль выполняет решающая функция

$$\delta^* = \begin{cases} d^0, & \text{если} \quad \mathfrak{R}(d^0 \mid x^{(n)}) < C, \\ d^1, & \text{если} \quad \mathfrak{R}(d^0 \mid x^{(n)}) > C, \end{cases} \tag{2.1}$$

где $\mathfrak{R}(d^i \mid x^{(n)}) = \mathbf{P}\{\vartheta \notin \Theta_i \mid x^{(n)}\}$ — апостериорный риск от принятия решения d^i, $i = 0,1$, постоянная C неотрицательна. Если же апостериорный риск $\mathfrak{R}(d^0 \mid x^{(n)})$ равен C, то решающая функция δ^* принимает решение d^0 с вероятностью γ.

Для реализации такой «рандомизированной» процедуры в последнем случае проводится дополнительно одно испытание в схеме Бернулли с вероятностью «успеха» γ. В случае успешного испытания принимается решение d^0, при неудаче — решение d^1. Очевидно, что рандомизация принятия решения будет играть существенную роль только для экспериментов с дискретным распределением выборки, например, когда имеет место упомянутая во введении модель Bi–Be.

Замечание. Легко видеть, что решающая функция δ^* может быть определена в терминах отношения «взвешенных» правдоподобий

$$f_n(x^{(n)} \mid \Theta_i) = \mathbf{E}\{f_n(x^{(n)} \mid \vartheta) \mid \vartheta \in \Theta_i\}, \qquad i = 0, 1. \tag{2.2}$$

Однако принятый нами для δ^* вид более отвечает существу d-гарантийного различения гипотез, поскольку риск любой простой стратегии есть условное среднее апостериорного риска:

$$R_G(d^i; \{\nu, \delta\}) = \mathbf{E}\{\mathfrak{R}(d^i \mid X^{(\nu)}) \mid \delta = d^i\}, \qquad i = 0, 1. \tag{2.3}$$

Следующая теорема, аналогичная фундаментальной лемме Неймана–Пирсона, показывает, что не изменяя момента остановки ν, нельзя улучшить одновременно оба риска простой стратегии $\{\nu, \delta^*\}$ с решающей функцией δ^* вида (2.1).

Теорема 2.1 [17]. *Пусть $\phi^* = \{\nu, \delta^*\}$ — простая стратегия с некоторым моментом остановки ν и решающей функцией δ^*, определенной в* (2.1).

(i) *Если $\phi = \{\nu, \delta\}$ — другая простая стратегия с тем же самым моментом остановки ν, причем $R_G(d^i; \phi) \leqslant R_G(d^i; \phi^*)$, $i = 0, 1$, то решающие функции δ и δ^* совпадают почти наверное на множествах* $\{x^{(n)}\colon \mathfrak{R}(d^0 \mid x^{(n)}) \neq C, \nu = n\}$, $n = 1, 2, \ldots$.

(ii) *Если ограничение α_0 на долю ошибочных решений d^0 таково, что априорная вероятность* $\mathbf{P}\{vartheta \notin \Theta_0\}$ *больше α_0, то найдутся такие постоянные C и γ, что риск $R_G(d^0; \phi^*)$ равен α_0.*

(iii) *Если $\mathbf{P}\{\vartheta \in \Theta_0\} \leqslant \alpha_0$, то простая стратегия $\phi_0 = \{\nu, \delta_0\}$ с решающей функцией $\delta_0 \equiv d^0$ имеет риск $R(d^0; \phi_0) \leqslant \alpha_0$, $R(d^1; \phi) = 0$.*

Таким образом, при решении задач оптимальной остановки с ограничениями на риски простых стратегий достаточно ограничиться решающими функциями вида (2.1). Кроме того, из теоремы 2.1 непосредственно следует решение задачи (B).

Теорема 2.2 [17]. *Пусть $\mathbf{P}\{\vartheta \notin \Theta_0\} > \alpha_0$ и решающая функция простой стратегии $\phi^* = \{\nu, \delta^*\}$ удовлетворяет* (2.1) *с такими C и γ, что $R_G(d^0; \phi^*) = \alpha_0$. Тогда среди всех простых стратегий $\phi = \{\nu, \delta\}$, гарантирующих уровень α_0 и имеющих один и тот же момент остановки ν, стратегия ϕ^* минимизирует долю ошибочных решений d^1, т. е. минимизирует $R_G(d^1; \phi)$. Обратно, решающая функция любой простой стратегии ϕ, минимизирующая риск $R_G(d^1; \phi)$ при заданном ограничении α_0 на риск $R_G(d^0; \phi)$ и имеющая тот же момент остановки ν, почти наверное удовлетворяет* (2.1) *при некотором C; риск этой стратегии $R_G(d^0; \phi)$ равен α_0.*

Ряд вероятностных моделей (например, нормальная, биномиальная, показательная), применяемых в практике статистического контроля качества, обладает замечательным свойством монотонности отношения правдоподобия: $f_n(x^{(n)} \mid \theta')/f_n(x^{(n)} \mid \theta'')$ при любых $\theta' > \theta''$ является возрастающей функцией некоторой статистики $T_n = T_n(x^{(n)})$. Если, кроме того, проверяемые гипотезы имеют интервальный вид: $\Theta_0 = (-\infty, \theta_0]$ или $\Theta_0 = [\theta_0, \infty]$, то теоремы 2.1 и 2.2 могут быть сформулированы в терминах статистики T_n.

Пусть $\Theta_0 = (-\infty, \theta_0]$, $\Theta_1 = (\theta_0, \infty)$ и плотность $f_n(x^{(n)} \mid \theta)$ распределения выборки $X^{(n)}$ при каждом $n = 1, 2, \ldots$ имеет монотонное относительно статистики T_n отношение правдоподобия. Рассмотрим решающую функцию

$$\delta^* = \begin{cases} d^0, & \text{если} \quad T_n < C, \\ d^1, & \text{если} \quad T_n > C, \end{cases} \tag{2.4}$$

принимающую значение d^0 с вероятностью γ, если $T_n = C$.

Следствие [17]. *Для процедуры $\phi^* = \{\nu, \delta^*\}$ справедливы утверждения теорем* 2.1 *и* 2.2 *с заменой множества* $\{x^{(n)}\colon \Re(d^0 \mid x^{(n)}) \neq C, \nu = n\}$ *на множество* $\{x^{(n)}\colon T_n(x^{(n)}) \neq C, \nu = n\}$, $n = 0, 1, \ldots$.

Оптимальные свойства процедур контроля, основанных на выборочном среднем, в примерах пп. 1 и 2 введения устанавливаются с помощью этого следствия.

Рассмотрим теперь задачу (С) оптимизации момента остановки при ограничениях (α_0, α_1) на обе доли ошибочных решений. Предварительно дадим несколько определений.

Момент остановки ν назовем *достаточным* (д.м.о.) для гарантийного принятия решения, если для него существует такая решающая функция δ, что стратегия $\phi = \{\nu, \delta\}$ гарантийна. Момент остановки ν будем называть *минимально достаточным*, или *необходимым*, в классе моментов остановки $\mathcal{M}$, если он является д.м.о. и не существует другого д.м.о. из класса $\mathcal{M}$, стохастически строго меньшего ν. *Оптимальным* в классе $\mathcal{M}$ называется д.м.о. ν^*, для которого в классе $\mathcal{M}$ не существует д.м.о. ν с безусловным средним $\mathbf{E}\,\nu < \mathbf{E}\,\nu^*$.

Утверждения теорем 2.1 и 2.2, а также следствия ограничивают множество возможных решающих функций в оптимальных стратегиях до решающих функций вида (2.1) (соответственно, (2.4)). Кроме того, как следует из второго утверждения теоремы 2.1, решающую функцию гарантийной стратегии ϕ^* с д.м.о. можно подобрать таким образом, чтобы выполнялось хотя бы одно из равенств: $R_G(d^0; \phi^*) = \alpha_0$ или $R_G(d^1; \phi^*) = \alpha_1$.

Теорема 2.3 [17]. *Если ν — д.м.о. и априорное распределение таково, что $\alpha_1 < \mathbf{P}\{\vartheta \in \Theta_0\} < 1 - \alpha_0$, то существует такая решающая функция δ^* вида* (2.1), *что стратегия $\phi^* = \{\nu, \delta^*\}$ гарантийна, а постоянные C и γ могут быть выбраны таким образом, что* $R_G(d^0; \phi^*) = \alpha_0$.

Итак, если ограничиться простыми стратегиями с фиксированным объемом испытаний ($\nu \equiv n$ с вероятностью 1 для некоторого целого $n \geqslant 0$), то теорема 2.3 указывает способ построения оптимальной простой стратегии: для этого достаточно найти C, γ и n из соответствующих ограничений на риски стратегии с решающей функцией (2.1).

В качестве примера к утверждениям теорем 2.1–2.3 рассмотрим задачу различения гипотез $\Theta_0 = \{\theta \leqslant \theta_0\}$ и $\Theta_1 = \{\theta > \theta_0\}$ о среднем значении θ нормального распределения с дисперсией σ^2 и нормальным априорным распределением ϑ (модель $\mathcal{N}$–$\mathcal{N}$, см. пример п. 1 во введении). Для этой модели построение гарантийной стратегии представляет интерес лишь в том случае, когда среднее значение μ и дисперсия τ^2 априорного распределения, а также ограничения α_0 и α_1 на риски от принятия решений удовлетворяют неравенствам $0 < [\Phi((\theta_0-\mu)/\tau)-\alpha_1] < 1-\alpha_0-\alpha_1$. В противном случае в силу утверждения (iii) теоремы 2.1 соответствующее решение может приниматься всегда, причем без проведения наблюдений.

Для нормальной вероятностной модели отношение правдоподобия монотонно относительно выборочного среднего $\overline{x}_n$, и решающая функция, минимизирующая долю ошибочных решений d^1 при заданном ограничении α_0 на долю ошибочных решений d^0, принимает значение d^0, если $\overline{x}_n \leqslant C$ (без рандомизации), где постоянная C находится из уравнения

$$\frac{1}{\sqrt{2\pi}\tau}\int_{\theta_0}^{\infty}\Phi\left(\frac{C-\theta}{\sigma}\sqrt{n}\right)\exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}d\theta = \alpha_0\Phi\left(\frac{C-\mu}{\sqrt{\tau^2 n+\sigma^2}}\sqrt{n}\right).$$

Минимально достаточный объем выборки (м.д.о. в классе неслучайных и нерандомизированных моментов остановки) n^* определяется как наименьшее значение n, при котором

$$\Phi(t)\alpha_0 \geqslant \int\limits_{q(t)}^{\infty}\Phi\left(\left(t\sqrt{1+\sigma^2/n\tau^2}-\theta\right)\tau\sqrt{n}/\sigma\right)d\Phi(\theta),$$

где

$$t = \Phi^{-1}\left((\Phi((\theta_0-\mu)/\tau)-\alpha_1)/(1-\alpha_0-\alpha_1)\right),$$
$$q(t) = \Phi^{-1}\left((1-\alpha_0-\alpha_1)\Phi(t)+\alpha_1\right),$$

и постоянная C, определяющая оптимальное правило контроля, равна $\mu+t\sqrt{\tau^2+\sigma^2/n}$. В частном случае $\alpha_0=\alpha_1=\alpha$ $(<1/2)$, $\mu=\theta_0$ оптимальный объем выборки равен (см. [17]) наименьшему целому n, удовлетворяющему неравенству

$$n \geqslant \frac{\sigma^2}{\tau^2}\operatorname{ctg}^2\alpha\pi, \tag{2.5}$$

а решающая функция имеет вид

$$\delta^* = \begin{cases} d^0, & \text{если} \quad \overline{x}_{n^*} \leqslant \mu, \\ d^1, & \text{если} \quad \overline{x}_{n^*} > \mu. \end{cases}$$

При $\alpha \to 0$ для требуемого (минимального) числа наблюдений имеем: $n^* \sim \sigma^2/(\tau\alpha\pi)^2$. Напомним, что в классическом подходе при заданных ограничениях $a_0 = a_1 = a$ на риски поставщика и потребителя необходимый объем выборки n^* есть $O(\ln a)$, и величина n^* в основном определяется шириной области безразличия $\theta_1 - \theta_0$: $n^* = O\big((\theta_1 - \theta_0)^2\big)$ при $\theta_1 \to \theta_0$.

Отметим, что вопрос об оптимальном моменте остановки в классе всех «последовательных» простых стратегий остается открытым. Некоторый свет на эту проблему проливает еще одно характеризационное свойство оптимальной стратегии, которое позволяет связать задачу оптимальной остановки с ограничениями на доли ошибочных решений с классической задачей Вальда проверки простых гипотез с ограничениями на вероятности ошибок первого и второго рода.

Теорема 2.4 [16]. *Если $\phi = \{\nu, \delta\}$ — гарантийная простая стратегия с оптимальным или минимально достаточным моментом остановки в классе всевозможных моментов остановки, и априорное распределение G таково, что $\alpha_1 < \mathbf{P}\{\vartheta \in \Theta_0\} < 1 - \alpha_0$, то $R_G(d^i; \phi) = \alpha_i$, $i = 0, 1$.*

Как уже отмечалось, эта теорема позволяет свести рассматриваемую нами задачу оптимизации момента остановки к классической задаче Вальда. Для этого рассмотрим две простые гипотезы H_i: плотность распределения $X^{(n)}$ равна $f_n(x^{(n)} \mid \Theta_i)$ (см. (2.2)), $i = 0, 1$, $n = 1, 2, \dots$. Любая простая стратегия $\phi = \{\nu, \delta\}$ может быть использована для различения этих гипотез, если определить d^i как решение в пользу H_i. Пусть $r_i(\phi) = \mathbf{P}\{\delta = d^{1-i} \mid H_i\} = \mathbf{P}\{\delta = d^{1-i} \mid \vartheta \in \Theta_i\}$, $i = 0, 1$, — вероятности ошибок первого и второго рода для стратегии ϕ.

Теорема 2.5 [16]. *Стратегия ϕ имеет оптимальный (или минимально достаточный) момент остановки среди всех гарантийных стратегий с уровнями α_0 и α_1 тогда и только тогда, когда она имеет оптимальный (соответственно, минимально достаточный) момент остановки среди всех последовательных критериев различения гипотез H_0 и H_1, удовлетворяющих ограничениям $r_i(\phi) \leqslant a_i$, $i = 0, 1$, на вероятности ошибок первого и второго рода, причем*

$$a_0 = \frac{\alpha_1 p_1}{\alpha_1 p_1 + (1 - \alpha_0) p_0}, \quad a_1 = \frac{\alpha_0 p_0}{\alpha_0 p_0 + (1 - \alpha_1) p_1},$$

$$p_i = \mathbf{P}\{\vartheta \in \Theta_i\}, \qquad i = 0, 1.$$

Заметим, что известное доказательство оптимальности критерия Вальда (см. [10], § 9.5) здесь не может быть повторено, поскольку для плотностей $f_n(x^{(n)} \mid \Theta_i)$, $i = 0, 1$, компоненты выборки $X^{(n)}$ зависимы. Интересно также отметить, что приближенный критерий Вальда, который не обязательно удовлетворяет ограничениям $r_i(\phi) \leqslant a_i$, $i = 0, 1$, в нашем случае будет гарантийным, так как он совпадает с процедурой первого перескока (см. § 4). В § 4 мы сформулируем также некоторое оптимальное свойство этого критерия в случае априорных распределений, сильно сосредоточенных вблизи границы, разделяющей проверяемые гипотезы.

Итак, результаты этого параграфа дают решения ряда задач на построение оптимальных стратегий в случае известного априорного распределения. Следующий шаг — построение их эмпирических аналогов.

§ 3. Асимптотическая гарантийность; эмпирические и адаптивные стратегии контроля качества

Результаты предыдущего параграфа дают возможность находить для каждого априорного распределения $G \in \mathcal{G}$ простую стратегию ϕ_G, обладающую тем или иным оптимальным свойством. С целью формализации задачи построения аналогов оптимальных простых стратегий при неизвестном априорном распределении введем в рассмотрение так называемые *эмпирические стратегии*. Момент остановки $\nu^{[s]}$ и решающая функция $\delta^{[s]}$ такой стратегии не изменяются, начиная с эксперимента с номером $s+1$, и зависят только от результатов текущего эксперимента и данных $x^{[s]} = (x^{(n_1)}, \dots, x^{(n_s)})$ первых s экспериментов (s фиксировано). Эмпирическую стратегию будем обозначать $\phi^{[s]}$, а совокупность данных $x^{[s]}$ называть *архивом объема* s.

Необходимость рассмотрения эмпирических стратегий объясняется двумя факторами. Во-первых, на практике нередки ситуации, когда бесконечное накопление архива данных затруднено какими-либо техническими причинами. Во-вторых, эмпирические стратегии представляют собой наиболее удобный математический инструмент для доказательства утверждений относительно асимптотической оптимальности стратегий.

В силу закона больших чисел риск эмпирической стратегии $\phi^{[s]}$ равен

$$R_G(d^i; \phi^{[s]}) = \mathbf{P}\{\vartheta \notin \Theta_i \mid \delta^{[s]} = d^i; x^{[s]}\}, \qquad i = 0, 1, \tag{3.1}$$

и зависит как от неизвестного априорного распределения G, так и от архива данных $x^{[s]}$.

Последовательность эмпирических стратегий $\{\phi^{[s]}, s \geqslant 0\}$ называется *асимптотически гарантийной*, если для любого $G \in \mathcal{G}$

$$\limsup_{s\to\infty} R_G(d^i; \phi^{[s]}) \underset{\text{п.н.}}{\leqslant} \alpha_i, \qquad i = 0, 1.$$

Последовательность эмпирических стратегий $\{\phi^{[s]}, s \geqslant 0\}$ называется *эмпирическим аналогом* семейства $\{\phi_G = \{\nu_G, \delta_G\}, G \in \mathcal{G}\}$ простых стратегий, если для любого $G \in \mathcal{G}$

$$\lim_{s\to\infty} R_G(d^i; \phi^{[s]}) \underset{\text{п.н.}}{=} R_G(d^i; \phi_G), \qquad i = 0, 1,$$

$$\lim_{s\to\infty} \mathbf{E}\{\nu^{[s]} \mid x^{[s]}\} \underset{\text{п.н.}}{=} \mathbf{E}\nu_G.$$

Заметим, что процедура $\langle\nu_{s+1},\delta_{s+1}\rangle$, используемая в $(s+1)$-м эксперименте как компонента адаптивной стратегии $\phi=\{\langle\nu_k,\delta_k\rangle,\ k\geqslant 1\}$, формально может рассматриваться как «эмпирическая стратегия» в том смысле, что процедура $\langle\nu_{s+1},\delta_{s+1}\rangle$ используется во всех последующих экспериментах с номерами $k\geqslant s+1$. В частности, для нее может быть вычислена функция риска по формулам, аналогичным (3.1). В связи с этим возникает вопрос о свойствах адаптивной стратегии, процедуры которой или 1) асимптотически гарантийны, или 2) представляют собой эмпирический аналог простых оптимальных стратегий. В теореме 3.1 (см. ниже) утверждается, что в первом случае адаптивная стратегия будет гарантийна, а во втором случае риск адаптивной стратегии и средний объем испытаний на один эксперимент совпадает с соответствующими характеристиками оптимальной стратегии.

Рассмотрим последовательность $\varphi=\{\phi^{[s]},s\geqslant 0\}$ эмпирических стратегий, *согласованных* таким образом, что статистические процедуры $\phi^{[k]}$ совпадают с процедурами $\phi^{[m]}$ при любых $m\leqslant k$ для экспериментов с номерами $1,2,\ldots,m+1$; $k=0,1,\ldots$. Такая последовательность φ порождает адаптивную стратегию ϕ, которая совпадает с любой стратегией $\phi^{[s]}$ во всех экспериментах с номерами $1,2,\ldots,s+1$; $s=0,1,\ldots$.

Теорема 3.1 [17]. *Пусть $\varphi=\{\phi^{[s]},s\geqslant 0\}$ — согласованная последовательность эмпирических стратегий и ϕ — адаптивная стратегия, порожденная φ.*

(i) *Если φ асимптотически гарантийна, то ϕ гарантийна.*

(ii) *Если φ есть эмпирический аналог семейства простых стратегий $\{\phi_G,\ G\in\mathcal{G}\}$, то при адаптивной стратегии ϕ риск удовлетворяет соотношению $R_G(d^i;\phi)\underset{\text{п.н.}}{=}R_G(d^i;\phi_G)$ при любом $G\in\mathcal{G}$. Если, кроме того,*

$$\sum_{s=1}^{\infty}\frac{1}{s^2}\mathbf{E}\big\{(\nu^{[s]})^2\mid x^{[s]}\big\}\underset{\text{п.н.}}{<}\infty,$$

то средний объем наблюдений стратегии ϕ равен

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m}n_k\underset{\text{п.н.}}{=}\mathbf{E}\,\nu_G.$$

В силу этой теоремы построение оптимальной гарантийной стратегии контроля качества, действительно, может быть начато с отыскания оптимальной процедуры при известном G с дальнейшим построением ее эмпирических аналогов. В какой-то мере эта задача решена в статье [17], где приводятся некоторые методы построения адаптивных стратегий с использованием оценок апостериорных рисков $\mathfrak{R}(d^i\mid x^{(n)})$, $i=0,1$. Ввиду незаконченности результатов в этом направлении и сложности их математических формулировок, мы не будем останавливаться на этом вопросе и отошлем читателя к публикации [17].

§ 4. Оптимальные стратегии контроля при островершинных априорных распределениях

Изучим поведение минимального объема выборки, гарантирующего заданные уровни $1-\alpha_0$ и $1-\alpha_1$ выходного качества и контроля, когда априорное распределение имеет резко выраженный минимум на границе различаемых гипотез. Как видно из рассмотренного в § 2 примера (см. формулу (2.5)), увеличение объема испытаний происходит в двух случаях: накладываются более жесткие ограничения на риски или дисперсия априорного распределения τ^2 стремится к нулю, т.е. априорное распределение стягивается в точку θ_0, разграничивающую гипотезы (формула (2.5) получена в предположении $\mu = \theta_0$). Задача исследования асимптотического поведения необходимого объема выборки при $\alpha_0 \to 0$ и $\alpha_1 \to 0$ не получила к настоящему времени сколь-нибудь общего решения, в то время как для случая «вырождающихся» априорных распределений получен ряд результатов общего характера (см. [5], [15], [21]). Этот случай во многом аналогичен ситуации различения сближающихся гипотез в рамках классической теории Неймана–Пирсона.

Семейство вырождающихся в точку θ_0 априорных распределений формально можно задать с помощью априорной плотности вида $\tau^{-1}g((\theta-\theta_0)/\tau,\tau)$, где τ — «малый» параметр, значение которого характеризует степень сосредоточенности априорного распределения в точке θ_0; ясно, что предельное при $\tau \to 0$ распределение в этом случае приписывает точке θ_0 вероятность 1.

Естественно, $g(u,\tau)$ должно вести себя при $\tau \to 0$ достаточно регулярно, поэтому будем предполагать, что

$$\lim_{\tau\to 0} g(u,\tau) = g(u,0) \overset{\mathrm{df}}{=} g(u), \tag{4.1}$$

причем $g(u)$ является плотностью некоторого непрерывного распределения $G(\cdot)$.

Напомним (см. § 2), что необходимым объемом $n(\alpha_0,\alpha_1)$ выборки называется минимальное целое n, для которого существует решающая функция $\delta = \delta(X^{(n)})$ с заданными ограничениями на риски

$$\mathbf{P}\{\vartheta > \theta_0 \mid \delta = d^0\} \leqslant \alpha_0, \qquad \mathbf{P}\{\vartheta \leqslant \theta_0 \mid \delta = d^1\} \leqslant \alpha_1. \tag{4.2}$$

Из теоремы 2.3 вытекает, что необходимому объему выборки соответствует решающая функция вида (2.1), так что в силу (4.2) в основу определения $n(\alpha_0,\alpha_1)$ следует положить неравенства

$$\frac{\mathbf{P}\{\vartheta > \theta_0, L_n < C\} + \gamma\mathbf{P}\{\vartheta > \theta_0, L_n = C\}}{\mathbf{P}(L_n < C) + \gamma\mathbf{P}(L_n = C)} \leqslant \alpha_0,$$

$$\frac{\mathbf{P}\{\vartheta \leqslant \theta_0, L_n > C\} + (1-\gamma)\mathbf{P}\{\vartheta \leqslant \theta_0, L_n = C\}}{\mathbf{P}(L_n > C) + (1-\gamma)\mathbf{P}(L_n = C)} \leqslant \alpha_1,$$

в которых

$$L_n = \frac{\int_0^\infty f_n(x^{(n)} \mid \theta_0 + u\tau) g(u,\tau)\, du}{\int_{-\infty}^0 f_n(x^{(n)} \mid \theta_0 + u\tau) g(u,\tau)\, du}.$$

При выполнении (4.1) и стандартных условий регулярности (см. [11, с. 93]) для семейства распределений наблюдаемой случайной величины X справедлива

Теорема 4.1 [5]. *Для необходимого объема выборки $n(\alpha_0, \alpha_1)$ и соответствующей ему критической константы $C = C(\alpha_0, \alpha_1)$ справедливы соотношения*

$$n(\alpha_0, \alpha_1) = t\tau^{-2} + o(\tau^{-2}), \tag{4.3}$$

$$C(\alpha_0, \alpha_1) = \frac{\int_0^\infty \exp\{sut - u^2 t \mathrm{i}(\theta_0)/2\} g(u)\, du}{\int_{-\infty}^0 \exp\{sut - u^2 t \mathrm{i}(\theta_0)/2\} g(u)\, du} + o(1),$$

где $\mathrm{i}(\theta) = \mathbf{E}_\theta[\partial \ln f(X \mid \theta)/\partial\theta]^2$ — информация по Фишеру в точке θ; постоянные s и t определяются как решение (единственное) системы уравнений

$$\begin{gathered} \int_{-\infty}^0 g(u)\Phi(-(s - u\mathrm{i}(\theta_0))\sqrt{t/\mathrm{i}(\theta_0)})\, du = \frac{\alpha_1(p_1 - \alpha_0)}{1 - \alpha_0 - \alpha_1}, \\ \int_0^\infty g(u)\Phi((s - u\mathrm{i}(\theta_0))\sqrt{t/\mathrm{i}(\theta_0)})\, du = \frac{\alpha_0(p_0 - \alpha_1)}{1 - \alpha_0 - \alpha_1}, \end{gathered} \tag{4.4}$$

где $p_i = G(\Theta_i)$, $i = 0, 1$.

Проиллюстрируем использование результатов этой теоремы в примере п. 2 введения, где рассматривалась задача гарантийного контроля качества по альтернативному признаку в рамках модели Bi–Be.

Пусть априорное бета-распределение имеет среднее значение, равное θ_0, а его дисперсия τ^2 стремится к нулю. В терминах параметров λ_1 и λ_2 априорного распределения это соответствует случаю $\lambda_1 \to \infty$, $\lambda_2 \to \infty$, $\theta_0 = \lambda_1/(\lambda_1 + \lambda_2)$, $\tau^2 = \theta_0(1-\theta_0)/(1+\lambda_1+\lambda_2)$. Предельная плотность для $g(u,\tau) = B^{-1}(\lambda_1, \lambda_2)(\theta_0 + u\tau)^{\lambda_1 - 1}(1 - \theta_0 - u\tau)^{\lambda_2 - 1}$ есть плотность стандартного нормального распределения: $g(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$.

В случае $\alpha_0 = \alpha_1$ систему уравнений (4.4) удается разрешить в явном виде и получить формулу для необходимого объема выборки в соответствии с главным членом асимптотики (4.3). В статье [5] приводится таблица, иллюстрирующая точность асимптотических формул теоремы 4.1. При α_0 и α_1 порядка 0,05–0,10 отношение $n(\alpha_0, \alpha_1)$ к главному члену в асимптотике (4.3) равно 0,7–0,9, когда τ имеет порядок 0,1–0,2.

Рассмотрим теперь более общую ситуацию, когда момент остановки зависит от результатов испытаний. Замечательным результатом апостериорного подхода к проблеме гарантийности является тот факт, что здесь для любой статистической проблемы, будь то задача проверки гипотез, оценки параметров и пр., всегда существует простая стратегия (напомним, что априорное распределение предполагается известным), гарантирующая любые заданные ограничения на риски. Применительно к задаче контроля качества, связанной с различением гипотез H_0: $\theta \leqslant \theta_0$ и H_1: $\theta > \theta_0$, эта процедура определяется следующим образом. Рассмотрим момент остановки

$$\nu = \min\{n\colon\ L_n \leqslant A \text{ или } L_n \geqslant B\} \tag{4.5}$$

и решающую функцию

$$\delta = \begin{cases} d^0, & \text{если} \quad L_n \leqslant A, \\ d^1, & \text{если} \quad L_n \geqslant B, \end{cases} \tag{4.6}$$

где $L_n = L_n(x^{(n)}) = f_n(x^{(n)} \mid \Theta_1)/f_n(x^{(n)} \mid \Theta_0)$ (см. (2.2)),

$$A = \frac{p_0\alpha_0}{p_1(1-\alpha_0)}, \quad B = \frac{p_0(1-\alpha_1)}{p_1\alpha_1}, \qquad p_i = \mathbf{P}\{\vartheta \in \Theta_i\}, \quad i = 0, 1.$$

Нетрудно заметить, что данная процедура останавливает эксперимент, когда одна из апостериорных вероятностей событий $\{\vartheta \in \Theta_0\}$ или $\{\vartheta \in \Theta_1\}$ впервые «перескакивает» соответствующее ограничение α_0 или α_1 на риск, и при этом принимается решение, отвечающее гипотезе с наибольшей апостериорной вероятностью. Это так называемая процедура «первого перескока», хорошо известная в теории последовательных байесовских процедур статистического вывода (см., например, [10, § 9.10]). Так как (см. (2.3)) d-риск есть условное математическое ожидание апостериорного риска относительно решающей функции δ, то процедура (4.5)–(4.6) является гарантийной (подробнее об этой процедуре см. в [4], [9], [15], [16], [21]).

Приведем один результат об оптимальности этой процедуры в схеме стягивающихся априорных распределений. Для формализации понятия оптимальности введем следующее определение.

Для любого момента остановки ν статистического эксперимента назовем количеством информации в наблюдении $X^{(\nu)}$ в пользу гипотезы H_0 против H_1 величину

$$J_\varphi(\Theta_0, \Theta_1; X^{(\nu)}) = \mathbf{E}\left\{\varphi\left(\frac{f_\nu(X^\nu \mid \Theta_1)}{f_\nu(X^\nu \mid \Theta_0)}\right) \Bigm| \vartheta \in \Theta_0\right\},$$

где φ: $\mathbf{R}_+ \to \mathbf{R}$ есть некоторая выпуклая функция. Заметим, что при $\varphi(x) = -\ln x$ введенная мера информации совпадает с известным количеством информации по Кульбаку–Лейблеру в пользу простой гипотезы $f(\cdot \mid \Theta_0)$ против $f(\cdot \mid \Theta_1)$.

Лемма 4.1 [15]. *Если* $R(d^i;\{\nu,\delta\}) \leqslant \alpha_i$, $i=0,1$, *то*

$$\begin{aligned} J_\varphi\bigl(\Theta_0,\Theta_1;X^{(\nu)}\bigr) &\geqslant \frac{\alpha_1}{p_0}\,\frac{p_1-\alpha_0}{1-\alpha_0-\alpha_1}\,\varphi\biggl(\frac{p_0(1-\alpha_1)}{p_1\alpha_1}\biggr) \\ &\quad+\frac{(1-\alpha_0)(p_0-\alpha_1)}{p_0(1-\alpha_0-\alpha_1)}\,\varphi\biggl(\frac{p_0\alpha_0}{p_1(1-\alpha_0)}\biggr), \\ J_\varphi\bigl(\Theta_1,\Theta_0;X^{(\nu)}\bigr) &\geqslant \frac{p_0-\alpha_1}{p_1}\,\frac{\alpha_0}{1-\alpha_0-\alpha_1}\,\varphi\biggl(\frac{p_1(1-\alpha_0)}{p_0\alpha_0}\biggr) \\ &\quad+\frac{(1-\alpha_1)(p_1-\alpha_0)}{p_1(1-\alpha_0-\alpha_1)}\,\varphi\biggl(\frac{p_1\alpha_1}{p_0(1-\alpha_1)}\biggr). \end{aligned} \tag{4.7}$$

Неравенства (4.7), связывающие момент остановки ν с ограничениями α_0, α_1 на риски, позволяют ввести понятие оптимальности процедуры $\langle\nu,\delta\rangle$ с точки зрения достижимости правых частей неравенств (4.7): вместо достаточного объема наблюдений требуется достаточность количества информации, поставляемой наблюдением $X^{(\nu)}$.

Сформулируем теперь предположения относительно вероятностной модели, при которых будет установлена асимптотическая при $\tau\to 0$ оптимальность процедуры первого перескока (ср. с известным результатом об асимптотической оптимальности критерия Вальда при контигуальных альтернативах).

(1) Функция плотности $f(x\mid\theta)$ наблюдаемой случайной величины X дважды непрерывно дифференцируема по θ.

(2) Семейство распределений $\{F(\cdot\mid\theta),\theta\in\Theta\}$ удовлетворяет стандартным условиям регулярности (см. [11, с. 93]).

(3) Для любого $\theta\in\Theta$ информационное количество Фишера $\mathrm{i}(\theta)$ положительно.

(4)

$$\lim_{C\to\infty}\overline{\lim_{u\to\theta_0}}\mathbf{E}_u|H(X\mid u)\mid\chi(H(X\mid u)>C)=0,$$

$$\lim_{\substack{u\to\theta_0\\ h\to 0}}\mathbf{E}_u\sup_{|v-u|\leqslant h}|H(X\mid u)-H(X\mid v)|=0,$$

где $H(X\mid u)=\frac{\partial^2}{\partial u^2}\ln f(X\mid u)$, $\chi(A)$ — индикатор множества A.

(5) Априорное распределение G абсолютно непрерывно или дискретно.

(6) Случайная величина ϑ с распределением G допускает представление $\vartheta=\theta_0+\tau\xi$, где τ — «малый» параметр, ξ — некоторая невырожденная случайная величина (постулируется схема стягивающихся при $\tau\to 0$ априорных распределений).

(7) Интегралы в определении информации $J_\varphi(\Theta_i,\Theta_{1-i};X^\nu)$ сходятся равномерно (условие на $\varphi(\cdot)$).

Теорема 4.2 [15]. *Пусть выполнены условия* (1)–(7) *и* $0<\alpha_0<\mathbf{P}\{\vartheta\leqslant\theta_0\}$, $0<\alpha_1<\mathbf{P}\{\vartheta>\theta_0\}$. *Тогда процедура, определяемая соотношениями* (4.5)–(4.6), *обладает следующими свойствами*:

(i) *распределение случайной величины* $\tau^2\nu$ *при* $\tau \to 0$ *слабо сходится к некоторому собственному распределению*;

(ii) $\lim_{\tau\to 0} R(d^i; \{\nu, \delta\}) = \alpha_i$, $i = 0, 1$;

(iii) $\lim_{\tau\to 0} J_\varphi(\Theta_i, \Theta_{1-i}; X^{(\nu)}) = w_i$, $i = 0, 1$, *где* w_1 *и* w_0 — *нижние границы в правых частях* (4.7).

Замечание. Если априорное распределение G сосредоточено в конечном числе точек, то в дополнение к (i) можно утверждать, что $\tau^2\mathrm{E}\{\nu \mid \vartheta \in \Theta_i\} \to \mathrm{const}$, когда $\tau \to 0$.

Пример 4.1. Обратимся к модели $\mathcal{N}$–$\mathcal{N}$ и рассмотрим задачу различения гипотез H_0: $\theta \geqslant \theta_0$ и H_1: $\theta < \theta_0$. Пусть априорное распределение таково, что $\mu = \theta_0$ и дисперсия τ^2 стремится к нулю. В этом случае процедура первого перескока определяется последовательностью статистик

$$T_n = \frac{n(\overline{X} - \mu)}{\sigma\sqrt{n\sigma^{-2} + \tau^{-2}}}, \qquad n = 1, 2, \ldots,$$

и имеет вид

$$\begin{aligned} \nu_\tau &= \min\{n\colon T_n \notin (\Phi^{-1}(\alpha_1), \Phi^{-1}(1-\alpha_0))\}, \\ \delta_\tau &= \begin{cases} d^0, \text{ если } T_n \geqslant \Phi^{-1}(1-\alpha_0), \\ d^1, \text{ если } T_n \leqslant \Phi^{-1}(\alpha_1). \end{cases} \end{aligned} \tag{4.8}$$

Таким образом, теорема 4.2 устанавливает некоторое свойство асимптотической оптимальности процедуры (4.8). Более подробно свойства процедуры первого перескока не изучены. Неизвестно, например, конечен ли средний объем наблюдений $\mathrm{E}\,\nu_\tau$ в процедуре (4.8). Результаты статистического моделирования указывают на большую вероятность остановки эксперимента после проведения первого испытания (на первом шаге эксперимента), и в то же время имели место единичные случаи, когда эксперимент состоял из фантастически большого числа шагов — десятки и сотни тысяч наблюдений.

В общем случае различения сложных гипотез остается открытым вопрос об оптимальности процедуры первого перескока в смысле минимальности среднего объема испытаний.

Естественно, возникает вопрос, насколько точны границы (4.7). Когда $\varphi(x) = -\ln x$ и априорное распределение сосредоточено в двух точках, границы (4.7) превращаются в границы Вальда для среднего объема последовательной выборки, в которых вероятности ошибок первого и второго рода выражены через d-риски. Следовательно, для процедуры (4.5)–(4.6) границы достигаются асимптотически при сближении различаемых простых гипотез. Существенно иная ситуация возникает при различении сложных гипотез (например, $\theta \geqslant \theta_0$, $\theta < \theta_0$). Если ограничиться процедурами с фиксированным объемом испытаний, то нетрудно показать, что неравенства (4.7) при $\alpha_0 = \alpha_1 = \alpha$ дают нижнюю границу для n порядка $-\ln\alpha$, хотя, как видно из примера § 2 (см. формулу (2.5)), оптимальная процедура имеет n порядка α^{-2}. Следовательно, в случае процедур с фиксированным объемом испытаний границы (4.7) оказываются грубыми (см. [21]).

§ 5. Оптимальные доверительные семейства: аттестация качества

В этом параграфе обсуждается вопрос аттестации качества выпускаемой продукции, который в терминах математической статистики формулируется как задача построения доверительного множества для неизвестного значения параметра θ. Наша концепция доверительного оценивания исходит из естественной практики использования доверительного утверждения как некоторого универсального способа одновременной проверки нескольких гипотез. Так, если в паспортных данных на плавку металла указана нижняя граница $\underline{\theta}$ для ее прочности, то эта плавка может быть приобретена любым из потребителей, у которого требования к прочности плавки имеют вид $[\theta_0, \infty)$ с нормой $\theta_0 \leqslant \underline{\theta}$. Эта же плавка будет, очевидно, отвергнута потребителями с более высокой нормой $\theta_0 > \underline{\theta}$ и, кроме того, потребителями, требования которых имеют вид $[\theta_0; \theta_1]$, $\theta_0 < \theta_1$. Поэтому любое доверительное утверждение о θ должно, во-первых, опираться на класс $\mathcal{B}$ возможных требований к параметру θ и, во-вторых, содержать в себе совокупность всех тех подмножеств B из $\mathcal{B}$, для которых риск от принятия гипотезы $\vartheta \in B$ не превосходит заданный уровень α.

Точное определение доверительного множества и некоторые утверждения об оптимальности приводятся в статье [8]. В задачах аттестации качества обычно используются требования интервального вида, поэтому здесь мы ограничимся рассмотрением только задачи построения верхней доверительной границы; нижняя граница строится по аналогии с верхней. Оптимизация доверительной оценки будет проводиться в классе простых стратегий (процедур) с фиксированным неслучайным числом наблюдений n. Таким образом, предполагается, что априорное распределение G полностью известно, и мы рассматриваем только задачи начального этапа роббинсовской схемы построения оптимальных стратегий доверительного оценивания — дальнейшие построения должны проводиться в соответствии с концепциями § 3. Вопрос о последовательных процедурах доверительного оценивания остается открытым — мы не располагаем результатами в этой области.

Итак, предположим, что класс требований $\mathcal{B}$, налагаемых на значение параметра θ, состоит из односторонних интервалов $(-\infty, \theta]$, $\theta \in \mathbf{R}$, и пусть вероятностная модель имеет монотонное относительно некоторой статистики $T_n = T_n(X^{(n)})$ отношение правдоподобия. Статистику $\overline{\theta}_n = \overline{\theta}_n(X^{(n)})$ назовем *верхней* $(1-\alpha)$-*доверительной границей*, если для всех $\theta \in \mathbf{R}$ условная вероятность

$$\mathbf{P}\{\vartheta \leqslant \theta \mid \overline{\theta}_n \leqslant \theta\} \geqslant 1 - \alpha. \tag{5.1}$$

Верхняя $(1-\alpha)$-доверительная граница $\overline{\theta}_n^*$ называется ***наиболее точной***, если для любой другой верхней $(1-\alpha)$-доверительной границы $\overline{\theta}_n$ и любого $\theta \in \mathbf{R}$ выполняется неравенство

$$\mathbf{P}\{\vartheta > \theta \mid \overline{\theta}_n^* > \theta\} \geqslant \mathbf{P}\{\vartheta > \theta \mid \overline{\theta}_n > \theta\}.$$

Таким образом, если гипотеза H_0: $\theta \leqslant \theta_0$ принимается, когда $\overline{\theta}_n \leqslant \theta_0$, то среди всех таких исходов средняя доля ошибочных решений не будет превосходить заданный уровень α. Наиболее точная верхняя граница доставляет минимум средней доли ошибок при отвержении H_0.

Заметим, что байесовская верхняя граница $\bar{b}_n$, определяемая как квантиль $\bar{b}_n = \min\{\theta\colon \mathbf{P}\{\vartheta \leqslant \theta \mid x^{(n)}\} \geqslant 1-\alpha\}$ апостериорного распределения, будет также верхней $(1-\alpha)$-доверительной границей в смысле (5.1), однако, как будет показано ниже, ее точностные характеристики далеки от оптимальных.

Как и в классической теории Неймана–Пирсона, здесь существует связь между задачей построения наиболее точной границы и задачей (B) построения стратегии, гарантирующей заданный уровень α для доли ошибочных решений d^0 и минимизирующей долю ошибочных решений d^1 (см. решение этой задачи в § 2).

Теорема 5.1 [8]. *Если при заданном объеме выборки n вероятностная модель имеет монотонное относительно статистики $T_n = T_n(X^{(n)})$ отношение правдоподобия, то наиболее точная верхняя $(1-\alpha)$-доверительная граница равна минимальному $\theta \in \mathbf{R}$, для которого $\mathbf{P}\{\vartheta \leqslant \theta \mid T_n(X^{(n)}) \leqslant T_n(x^{(n)})\} \geqslant 1-\alpha$, где $T_n(x^{(n)})$ — значение статистики $T_n(X^{(n)})$, полученное в эксперименте.*

Например, для модели $\mathcal{N}$–$\mathcal{N}$ наиболее точная верхняя граница определяется из уравнения

$$\frac{1}{\sqrt{2\pi}\tau}\int_{-\infty}^{\overline{\theta}_n^*} \Phi\left(\frac{z-\overline{x}_n}{\sigma}\sqrt{n}\right)\exp\left\{-\frac{(z-\mu)^2}{2\tau^2}\right\}dz = (1-\alpha)\Phi\left(\frac{\overline{x}_n-\mu}{\sqrt{\tau^2+\sigma^2/n}}\right).$$

Подстановка вместо μ, τ и σ их оценок по архиву данных приводит к эмпирическому аналогу оптимальной доверительной границы.

Исследуем поведение $\overline{\theta}_n^*$ с ростом числа наблюдений n. Для простоты изложения предположим, что при всех n статистика T_n имеет непрерывную функцию распределения и функция G априорного распределения строго монотонна. Пусть, кроме того, последовательность статистик $\{T_n, n \geqslant 1\}$ выбрана так, что для всех фиксированных значений параметра $\theta \in \mathbf{R}$ существует почти наверное $\lim_{n\to\infty} T_n = C(\theta)$, где $C(\theta)$ — строго возрастающая функция θ. В таком случае имеет место следующий результат.

Теорема 5.2 [8]. (i) *Наиболее точная верхняя $(1-\alpha)$-доверительная граница $\overline{\theta}_n^*$ равна $G^{-1}((1-\alpha)G(\theta)) + r_n$, где $\lim_{n\to\infty} r_n = 0\,(\mathbf{P}_\theta$-п.н.).*

(ii) *Если* $\widehat{\theta}_n$ — *состоятельная оценка* θ, *то статистика* $\widetilde{\theta}_n = G^{-1}((1-\alpha)G(\widehat{\theta}_n))$ *является асимптотической верхней* $(1-\alpha)$-*доверительной границей*: $\lim_{n\to\infty} \mathbf{P}\{\vartheta \leqslant \theta \mid \widetilde{\theta}_n \leqslant \theta\} = 1-\alpha$ $(\forall\theta \leqslant G^{-1}(1-\alpha))$.

З а м е ч а н и е 1. Из утверждения (i) этой теоремы следует, в частности, что наиболее точная верхняя граница, по крайней мере при больших n, расположена левее (!) истинного значения параметра θ. На первый взгляд это кажется парадоксальным. Однако вспомним, что в примере п. 1 введения было получено приемочное число меньше нормы. Кроме того, следующий «экстремальный» пример показывает, что если уровень $1-\alpha$ фиксирован, то ничего другого ожидать не приходится.

Предположим, что в эксперименте непосредственно, без каких-либо ошибок, наблюдается значение параметра θ. Тогда «естественная» верхняя граница $\overline{\theta} = \theta$, где θ — истинное значение параметра, будет $(1-\alpha)$-доверительной границей:

$$\mathbf{P}\{\vartheta \leqslant \theta \mid \overline{\theta} \leqslant \theta\} = 1 > 1-\alpha \qquad (\forall\theta \in \mathbf{R}), \tag{5.2}$$

и в то же время для этой границы

$$\mathbf{P}\{\vartheta > \theta \mid \overline{\theta} \leqslant \theta\} = 0 \qquad (\forall\theta \in \mathbf{R}). \tag{5.3}$$

Однако легко видеть, что граница $\overline{\theta}^* = G^{-1}((1-\alpha)G(\theta))$ удовлетворяет соотношениям (5.2) и (5.3) и в то же время $\overline{\theta}^* < \overline{\theta} = \theta$.

З а м е ч а н и е 2. Хорошо известно, что при весьма общих условиях апостериорное распределение ϑ с ростом объема выборки n стягивается к истинной точке θ. Поэтому байесовская граница $\overline{b}_n$ является состоятельной оценкой θ, что вместе с первым утверждением теоремы 5.2 свидетельствует о неоптимальности $\overline{b}_n$.

Для модели $\mathcal{N}$–$\mathcal{N}$ можно привести более точные утверждения относительно асимптотического поведения $\overline{\theta}_n^*$.

Теорема 5.3 [7]. *В рамках модели* $\mathcal{N}$–$\mathcal{N}$ *наиболее точная верхняя* $(1-\alpha)$-*доверительная граница для* θ *имеет вид*:

$$\overline{\theta}_n^* = H(\overline{x}_n) - (\overline{x}_n - \mu)\frac{(1-\alpha)\sigma^2}{2n\tau^3}\exp\left\{-\frac{(\overline{x}_n-\mu)^2 - (H(\overline{x}_n)-\mu)^2}{2\tau^2}\right\} + \frac{\Delta_n}{n^2},$$

где $H(y) = \mu + \tau\Phi^{-1}((1-\alpha)\Phi((y-\mu)/\tau))$ *и при любом* $\theta \in \mathbf{R}$ *почти наверное* $\overline{\lim}_{n\to\infty}|\Delta_n| < \infty$.

Более содержательные результаты имеют место в случае островершинных априорных распределений, рассмотренных в § 4, когда с ростом n априорное распределение сосредоточивается в точке $\theta_0 = \mathbf{E}\vartheta = \mu$. Оговорим, что такое предположение не следует трактовать буквально, — априорное распределение фиксировано и не может меняться в процессе эксперимента — это просто схема асимптотического анализа. В такой схеме естественно ограничиться требованиями вида $(-\infty, \theta_n)$, где

$\theta_n = \mu + a\tau$, a — некоторая постоянная, τ — малый параметр, а априорное распределение имеет плотность вида $\frac{1}{\tau}g\left(\frac{\theta-\mu}{\tau}, \tau\right)$. Нетрудно понять, что в регулярном случае, определяемом условиями (1)–(7) из § 4, τ следует положить равным $t/\sqrt{n}$, где t — некоторая постоянная, определяемая степенью концентрации априорного распределения около точки $\theta_0 = \mu$ (см. пример с моделью Bi–Be в § 4). Определим $\overline{a}_n = \overline{a}_n(y)$, $y \in \mathbf{R}$, как наименьшее из всех a, удовлетворяющих неравенству

$$\int_{-\infty}^{a} g(u;0)\Phi\left((y-u)t\sqrt{\mathrm{i}(\mu)}\right)du \geqslant (1-\alpha)\int_{-\infty}^{\infty} g(u;0)\Phi\left((y-u)t\sqrt{\mathrm{i}(\mu)}\right)du.$$

Таким образом, $\overline{a}_n$ есть наиболее точная верхняя $(1-\alpha)$-доверительная граница, построенная по одному наблюдению нормальной $(0, 1/t^2\mathrm{i}(\mu))$ случайной величины с априорной плотностью $g(\theta;0)$.

Теорема 5.4 [8]. *Пусть $\widehat{\theta}_n$ — асимптотически нормальная $(t_n, 1/n\mathrm{i}(t_n))$ оценка параметра θ по выборке из распределения $F(\cdot \mid t_n)$ и $t_n \to \mu$ при $n \to \infty$. При выполнении условий регулярности теоремы* 4.1 *и требований θ_n вида $\theta_n = \mu + at/\sqrt{n}$ граница $\overline{\theta}_n = \mu + \overline{a}(y)t/\sqrt{n}$ с $y = (\widehat{\theta}_n - \mu)\sqrt{n}/t$*

(i) *имеет асимптотически доверительный уровень* $1-\alpha$:

$$\lim_{n\to\infty} \mathbf{P}\{\vartheta \leqslant \theta_n \mid \overline{\theta}_n \leqslant \theta_n\} = 1-\alpha;$$

(ii) *асимптотически эквивалентна оптимальной границе $\overline{\theta}_n^*$*:

$$\lim_{n\to\infty} \frac{\mathbf{P}\{\vartheta \leqslant \theta_n \mid \overline{\theta}_n > \theta_n\}}{\mathbf{P}\{\vartheta \leqslant \theta_n \mid \overline{\theta}_n^* > \theta_n\}} = 1.$$

Численные иллюстрации $\overline{\theta}_n^*$ и $\overline{\theta}_n$ в сравнении с байесовской границей $\overline{b}_n$ для ряда моделей байесовского вывода приведены в статье [8].

СПИСОК ЛИТЕРАТУРЫ

1. *Беляев Ю. К.* Вероятностные методы выборочного контроля. М.: Наука, 1975, 407 с.
2. *Бернштейн С. Н.* О «доверительных» вероятностях Фишера. — Изв. АН СССР, сер. матем., 1941, т. 5, с. 85–94.
3. *Браунли К. А.* Статистическая теория и методология в науке и технике. М.: Наука, 1977, 407 с.
4. *Володин И. Н.* Гарантийные процедуры статистического вывода (определение объема выборки). — В сб.: Исследования по прикладной математике. Казань: Изд-во Казанского ун-та, 1984, в. 10, с. 13–53.

5. *Володин И. Н., Новиков А. А.* Асимптотика необходимого объема выборки при d-гарантийном различении двух гипотез. — Изв. вузов, сер. матем., 1983, № 11, с. 59–66.
6. *Володин И. Н., Новиков Ан. А.* Статистическая оценка с асимптотически минимальным d-риском. — Теория вероятн. и ее примен., 1993, т. 38, № 1, с. 20–32.
7. *Володин И. Н., Симушкин С. В.* Несмещенность и байесовость. — Изв. вузов, сер. матем., 1987, № 1, с. 3–7.
8. *Володин И. Н., Симушкин С. В.* Доверительное оценивание в d-апостериорном подходе. — Теория вероятн. и ее примен., 1990, т. 35, № 2, с. 242–254.
9. *Володин И. Н., Симушкин С. В.* О d-апостериорном подходе к проблеме статистического вывода. — Труды III международной конференции по теории вероятностей и математической статистике. Вильнюс: ИМК АН ЛитССР, 1981, т. 1, с. 101–102.
10. *Закс Ш.* Теория статистических выводов. М.: Мир, 1975, 776 с.
11. *Ибрагимов И. А., Хасьминский Р. З.* Асимптотическая теория оценивания. М.: Наука, 1973, 523 с.
12. *Королев В. Ю.* Неклассические задачи проверки простой гипотезы против простой альтернативы. — В сб.: Численные методы в математической физике. М.: МГУ, 1986, с. 100–102.
13. *Крамер Г.* Случайные величины и распределения вероятностей. М.: ИЛ, 1947, 144 с.
14. *Круопис Ю.* Минимизация целевых функций некоторых систем контроля качества. — В сб.: Оптимизация систем контроля качества. Вильнюс: ИМК АН ЛитССР, 1981, с. 12–27.
15. *Новиков Ан. А.* Асимптотическая оптимальность последовательного d-гарантийного критерия. — Теория вероятн. и ее примен., 1987, т. XXXII, № 2, с. 387–391.
16. *Симушкин С. В.* Оптимальный объем выборки при d-гарантийном различении гипотез. — Изв. вузов., сер. матем., 1982, № 5, с. 47–52.
17. *Симушкин С. В.* Эмпирический d-апостериорный подход к проблеме гарантийности статистического вывода. — Изв. вузов., сер. матем., 1983, № 11, с. 42–58.
18. *Hill J. R.* A general framework for model-based statistics. — Biometrika, 1990, v. 77, № 1, p. 115–126.
19. *Novikov A. A., Volodin I. N.* Pointed priors and asymptotics of the a posteriori risk. — In: Proceedings of Fourth International Vilnius Conference on Probability Theory and Mathematical Statistics. Vilnius/Utrecht: Mokslas/VNU, 1987, p. 421–431.
20. *Simushkin S. V., Volodin I. N.* Statistical inference with a minimal d-risk. — Lect. Notes Math., 1983, v. 1021, p. 629–636.
21. *Volodin I. N.* Necessary sample size for guaranteed (by Bol'shev) discrimination of hypotheses. — In: Fifth International Vilnius Conference on Probability Theory and Mathematical Statistics: Abstracts of communications. Vilnius: Inst. Math. Cybern., 1989, vol. II, p. 218–219.