

КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

Кафедра математической статистики

И.Н. Володин

ОПТИМАЛЬНЫЕ СТАТИСТИЧЕСКИЕ РЕШЕНИЯ

КАЗАНЬ – 2012

УДК 519.22

ББК 22.172

В 68

*Печатается по решению методической комиссии Института ВМ и ИТ
Казанского (Приволжского) федерального университета*

Протокол № 2 от 01 октября 2012 г.

Научный редактор

кандидат физ.-мат. наук, доцент С.Г. Халиуллин

Рецензенты:

кандидат физ.-мат. наук, доцент КФУ С.В. Симушкин

кандидат тех. наук, доцент КГТУ И.И. Хамдеев

Володин И.Н.

Оптимальные статистические решения Учебное пособие / И.Н. Володин. – Казань: Казанский (Приволжский) федеральный университет, 2012. – 182 с.

Учебное пособие представляет углубленный курс математической статистики с упором на построение оптимальных процедур статистического вывода. Затрагиваются основные проблемы данной теории: аксиоматика математической статистики, достаточные статистики, информационные неравенства, байесовские и минимаксные процедуры, оценки с минимальным риском, равномерно наиболее мощные критерии, инвариантность, планирование объема испытаний. Существенно новым является изложение d -апостериорного подхода к проблеме гарантийности статистического вывода – основной научной тематики кафедры математической статистики Казанского университета. Изложение ведется в духе монографий Э.Лемана с привлечением ряда конструкций из монографии Ш.Закса и работ преподавателей кафедры. Звездочкой отмечен материал, предназначенный для семинарских занятий. Разработки соответствуют семестровому курсу лекций по современным проблемам математической статистики. Учебное пособие по теории оптимального статистического вывода предназначается для студентов старших курсов и аспирантов, специализирующихся в области теории вероятностей и математической статистики.

©Казанский (Приволжский) федеральный университет, 2012

©Володин И.Н., 2012

Литература

1. Леман. Э. *Проверка статистических гипотез*/ Э. Леман. – М.: Наука, 1979.– 408 с.
2. Леман. Э. *Теория точечного оценивания*/ Э. Леман. – М.: Наука, 1991. – 444 с.
3. Закс. Ш. *Теория статистических выводов*/ Ш. Закс. – М.: Мир, 1975. – 776 с.
4. Володин И. Н. *Лекции по теории вероятностей и математической статистике* / И.Н. Володин. – Казань: Казан. гос. ун-т, 2006. – 272 с.
5. Володин И. Н. *Гарантийный статистический контроль качества: апостериорный подход* / И.Н. Володин, А.А. Новиков, С.В. Симушкин // *Обзорные прикладной и промышленной математики*. – 1994. – Т. 1. – № 2. – С. 148–178.
6. Володин И. Н. *Нижние границы для среднего объема наблюдений в гарантийных процедурах статистического вывода*/ И.Н. Володин// *Исследования по прикладной математике*. – Казань: Изд-во Казан. ун-та 2011. – Вып. 27. – С. 70-116.

Оглавление

§1.	Проблема статистического вывода	5
§2.	Статистические структуры, обладающие достаточными статисти- ками	20
§3.	Меры информации и нижние границы для среднего объема вы- борки	34
§4.	Байесовские решения	46
§5.	Минимаксные решения	62
§6.	Статистические оценки с равномерно минимальным риском . . .	70
§7.	Равномерно наиболее мощные критерии для статистических мо- делей с монотонным отношением правдоподобия	83
§8.	Равномерно наиболее точные доверительные границы	98
§9.	Равномерно наиболее мощные несмещенные критерии	105
§10.	Инвариантный статистический вывод	121
§11.	Статистический критерий, минимизирующий d-риск	139
§12.	Гарантийный статистический вывод	154
	Предметный указатель	178

§ 1. Проблема статистического вывода

1.1. Проблема решения и вероятностная модель. Статистическое исследование начинается с описания объекта статистического эксперимента, формализации *пространства* \mathcal{D} *решений* d и первоначального, достаточно грубого задания потерь $L(d', d)$, на прямом произведении $\mathcal{D} \times \mathcal{D}$ со значениями в \mathbb{R}_+^k . Принятием некоторого решения $d \in \mathcal{D}$ завершается статистическое исследование, и значение $L(d', d)$ соответствует потерям (в определенных единицах измерения), которые обусловлены принятием решения d , когда в действительности “правильным” решением является d' ; правильным считается такое решение, которое не приводит к каким-либо потерям: $L(d, d) = 0$ при любом $d \in \mathcal{D}$.

Решение d принимается на основе результатов наблюдений некоторых характеристик исследуемого объекта. Всевозможные состояния объекта в момент наблюдения формализуются введением пространства Ω элементарных исходов ω , однозначно определяющих значения наблюдаемых характеристик. Предполагается, что на измеримом пространстве (Ω, \mathcal{A}) состояний объекта существует вероятность $P(A)$, $A \in \mathcal{A}$.

Вероятностное пространство (Ω, \mathcal{A}, P) , при всей его значимости как единого пространства, гарантирующего существование распределений наблюдаемых характеристик объекта, вряд ли может претендовать на составляющую вероятностной модели с точки зрения практического применения к вычислению величины средних потерь. Дело в том, что только в крайне редких случаях можно точно определить пространство элементарных исходов, не говоря уж о задании распределений на \mathcal{A} . Обычно статистик оперирует с распределениями наблюдаемых характеристик объекта, и такими могут быть измеримые действительные или векторные функции от ω или траектории случайных процессов, категориальные величины и т.п. При таком разнообразии типов наблюдаемых характеристик разумнее назвать их *случайными элементами*. Далее, величина средних потерь может существенно зависеть от того, какой именно случайный элемент наблюдается в статистическом эксперименте на каждом его шаге, и если статистик располагает набором случайных элементов, то он может управлять наблю-

дениями в процессе последовательного случайного выбора, наблюдая на разных этапах выбора разные случайные элементы. Итак, задание вероятностной модели начинается с выбора семейства $\Upsilon = \{\xi_i = \xi_i(\omega), i \in \mathcal{I}\}$ случайных элементов, каждый из которых претендует быть наблюдаемым на некотором этапе статистического эксперимента; затем определяется по возможности узкое семейство $\mathcal{F}_i = \{F_i(\cdot | \theta), \theta \in \Theta\}$ распределений каждого случайного элемента ξ_i на измеримом пространстве (X_i, \mathcal{A}_i) его значений, $i \in \mathcal{I}$.

Множество Θ называется *параметрическим пространством*; в общем случае это абстрактное пространство индексов; выбор некоторого θ из Θ означает фиксирование распределений $\{F_i\}$ случайных элементов $\{\xi_i\}$.

В классической (небайесовской) статистике вероятностная модель обычно определяется как класс семейств

$$\mathcal{F} = \{\mathcal{F}_i, i \in \mathcal{I}\} = \{\{F_i(\cdot | \theta), \theta \in \Theta\}, i \in \mathcal{I}\}.$$

Однако существует довольно много статистических проблем (например, статистический контроль качества, статистическая диагностика, классификация и т.п.), где задание только класса \mathcal{F} не является достаточным для корректного с точки зрения существа проблемы определения величины средних потерь. Дело в том, что значение (неизвестное) параметра θ в момент проведения статистического эксперимента может быть реализацией некоторого случайного элемента ϑ , и возникает необходимость в определении семейства $\mathcal{G} = \{G_\lambda, \lambda \in \Lambda\}$ возможных распределений ϑ на измеримом пространстве (Θ, \mathcal{B}) его значений. Естественно, если $\vartheta = \text{const}$, то \mathcal{G} определяется как семейство вырожденных (сосредоточенных в каждой точке θ параметрического пространства Θ) распределений.

Пара $(\mathcal{F}, \mathcal{G})$ называется *вероятностной моделью*.

Сделаем несколько ограничений на введенные выше пространства и семейства распределений. Так, пространства \mathcal{D}, Θ , и все $X_i, i \in \mathcal{I}$, будут считаться *польскими* (полными, сепарабельными, метрическими пространствами). Множество индексов \mathcal{I} не должно быть более чем счетным (ограничение, не слишком обременительное с точки зрения практических

приложений, но значительно упрощающее некоторые доказательства). Таким образом, все наблюдаемые случайные величины ξ_i занумерованы: $i = 1, 2, \dots$. Каждое из семейств вероятностных мер $\mathcal{F}_i = \{F_i(\cdot | \theta), \theta \in \Theta\}$ предполагается доминированным соответствующей σ -конечной положительной мерой μ_i , $i \in \mathcal{I}$, а семейство априорных распределений $\mathcal{G} = \{G_\lambda, \lambda \in \Lambda\}$ доминированным σ -конечной положительной мерой χ . Следовательно, существуют производные Радона–Никодима

$$f_i(x | \theta) = \frac{dF_i}{d\mu_i}(x | \theta), \quad x \in \mathcal{X}_i, \quad i \in \mathcal{I}, \quad \theta \in \Theta; \quad g_\lambda(\theta) = \frac{dG_\lambda}{d\chi}(\theta), \quad \theta \in \Theta, \quad \lambda \in \Lambda,$$

где $f_i(x | \theta)$ трактуется как плотность условного распределения ξ_i относительно случайного элемента ϑ , $i \in \mathcal{I}$, а $g_\lambda(\theta)$ – как плотность маргинального распределения ϑ при каждом фиксированном $\lambda \in \Lambda$. Естественно, такая трактовка возможна лишь при \mathcal{B} -измеримости функций $F_i(A | \theta)$, $\theta \in \Theta$, при любых $A \in \mathcal{A}_i$ и $i \in \mathcal{I}$, то есть $F_i(\cdot | \cdot)$ должна быть переходной вероятностью.

Доминируемость семейств распределений, определяющих вероятностную модель, позволяет задать ее посредством пары семейств функций плотности $\{f_i\}$ и $\{g_\lambda\}$ или семейством совместных распределений $H_{\lambda,i}(A, B)$, $A \in \mathcal{A}_i$, $B \in \mathcal{B}$, случайных элементов ξ_i и θ с плотностями

$$h_{\lambda,i}(x, \theta) = f_i(x | \theta) g_\lambda(\theta), \quad x \in \mathcal{X}_i, \quad \theta \in \Theta, \quad \lambda \in \Lambda,$$

по прямому произведению мер μ_i и χ , $i \in \mathcal{I}$.

Введение параметра θ связано с особой спецификой статистического вывода: если известно истинное значение θ (распределение вероятностей в момент наблюдения), то статистик всегда может принять правильное решение $d = d_\theta$. Следовательно, проблему принятия решения всегда можно интерпретировать в терминах значений параметра θ , и эта особенность статистического вывода позволяет задать *функцию потерь* не на $\mathcal{D} \times \mathcal{D}$, а на произведении пространств $\Theta \times \mathcal{D}$. Сохраняя прежнее обозначение (L) для функции потерь, положим $L(\theta, d) \stackrel{df}{=} L(d_\theta, d)$, $\theta \in \Theta$, $d \in \mathcal{D}$.

Следует особо отметить, что после построения вероятностной модели функция потерь обычно уточняется (или вообще вводится изначально).

Например, в задаче различения двух сложных гипотез, касающихся значения параметра θ , первоначальная функция потерь типа 1-0 иногда видоизменяется и становится зависящей от расстояния между θ и границей, разделяющей параметрические гипотезы.

Задание пространства решений \mathcal{D} и функции потерь $L(\theta, d)$, $\theta \in \Theta$, $d \in \mathcal{D}$, определяет *проблему статистического вывода*.

1.2. Статистический эксперимент. Построением вероятностной модели заканчивается та часть работы статистика, которая предшествует постановке статистического эксперимента – наблюдению копий случайных элементов из класса Υ .

Статистический эксперимент состоит из некоторого числа шагов. Каждому шагу присваивается номер $k = 0, 1, 2, \dots$. На любом шаге с номером $k \neq 0$ наблюдается только одна независимая копия (с точки зрения совпадения распределений) X_{i_k} случайного элемента $\xi_{i_k} \in \Upsilon$ и фиксируется результат x_{i_k} ее наблюдения, после чего решается вопрос о выполнении одного из альтернативных действий: a_c – продолжить наблюдения или a_s – остановить эксперимент. Если выбрано действие a_c , то статистик определяет индекс i_{k+1} случайного элемента класса Υ , независимая копия которого будет наблюдаться на следующем $k + 1$ -м шаге эксперимента, после чего переходит к наблюдениям на этом шаге. Если выбрано действие a_s , то эксперимент считается завершенным и статистик принимает некоторое решение $d \in \mathcal{D}$. Этот последний этап статистического исследования рассматривается отдельно – он не относится к работе статистика, связанной с постановкой статистического эксперимента.

Особо определяется нулевой шаг ($k = 0$) эксперимента. На этом шаге не наблюдается какой-либо случайный элемент класса Υ ; статистик только выбирает одно из альтернативных действий a_s и a_c : или отказаться от проведения наблюдений вообще, или начать эксперимент и выбрать случайный элемент, подлежащий наблюдению на первом шаге. Для того чтобы не рассматривать каждый раз отдельно нулевой шаг эксперимента, будем считать, что на этом шаге осуществляется “наблюдение” фиктивного случайного элемента $X_{i_0}(= X_0)$, который принимает единственное “зна-

чение” x_{i_0} ; это значение можно трактовать как совокупность априорных сведений об исследуемом объекте, позволяющих статистику решать вопрос о целесообразности проведения статистического эксперимента. Измеримое пространство “значений” X_{i_0} будет обозначаться $(X_{i_0}, \mathcal{A}_{i_0})$, где $X_{i_0} = \{x_{i_0}\}$ – одноточечное множество, а $\mathcal{A}_{i_0} = \{x_{i_0}, \emptyset\}$. “Распределение” X_{i_0} определяется формальным равенством $F_{i_0}(X_{i_0} | \theta) = f_{i_0}(x_{i_0} | \theta) = 1$ при любом $\theta \in \Theta$.

Наблюдение любого случайного элемента ξ_i из класса Υ сопряжено с определенными затратами c_i , $i \in \mathcal{I}$, причем стоимость “наблюдения” X_0 , естественно, равна нулю. Не ограничивая общности, предположим, что $c_i \leq 1$ при любом $i \in \mathcal{I}$. Если все $c_i = 1$, $i \in \mathcal{I}$, то затраты на постановку статистического эксперимента равны числу случайных элементов, наблюдаемых в эксперименте, то есть объему выборки. В дальнейшем будут рассматриваться только такие проблемы, где затраты на постановку статистического эксперимента определяются только числом шагов (объемом наблюдений), – обобщение на случай зависимости цены не только от индекса наблюдаемого случайного элемента, но и от номера шага является самостоятельной и довольно сложной задачей.

Решение об остановке эксперимента и выбор случайных элементов, подлежащих наблюдению на каждом шаге, в общем случае зависят от результатов проведенных наблюдений x_{i_0}, x_{i_1}, \dots и осуществляются введением специальных статистических правил, которые будут определены ниже. Однако сам процесс наблюдения копий случайных величин из класса Υ осуществляется в рамках так называемого *простого случайного выбора* с управлением наблюдениями. Для того чтобы формализовать это понятие, рассмотрим сначала статистический эксперимент с фиксированной схемой выбора, в которой индексы $i_0, i_1, i_2, \dots, i_n$ и момент остановки n эксперимента фиксированы а priori.

Пусть $t_n = (i_0, i_1, \dots, i_n)$ – набор индексов из \mathcal{I} с пространством значений \mathcal{I}^n , $n = 0, 1, \dots$, $\mathcal{I}^0 = \{i_0\}$, и $\mathcal{T} = \sum_{n=0}^{\infty} \mathcal{I}^n$. Элемент t пространства \mathcal{T} определяется значением $n (= 0, 1, \dots)$ и набором индексов t_n . На этом пространстве введем частичный порядок: $t_m < t_n$, если

$m < n$ и первые m компонент вектора t_n совпадают с компонентами вектора t_m . Далее, на прямом произведении вероятностных пространств $(\Omega^n, \mathcal{A}^n, P^n)$ определим вектор из независимых случайных элементов $\mathbf{X}^{(t_n)} = (X_{i_0}, X_{i_1}, \dots, X_{i_n})$ – независимых копий соответствующих случайных элементов из класса Υ , подлежащих наблюдению в эксперименте. Вектор $\mathbf{X}^{(t_n)}$ с измеримым пространством значений

$$(\mathcal{X}_{t_n}, \mathcal{A}_{t_n}) = \left(\prod_{k=0}^n \mathcal{X}_{i_k}, \bigotimes_{k=0}^n \mathcal{A}_{i_k} \right)$$

называется *случайной выборкой фиксированного объема n* . Распределение вектора $\mathbf{X}^{(t_n)}$ при каждом фиксированном значении $\theta \in \Theta$ определяется функцией плотности

$$p_{t_n}(x^{(n)} | \theta) = \prod_{k=0}^n f_{i_k}(x_k | \theta), \quad x^{(n)} \in \mathcal{X}_{t_n},$$

по мере $\mu_{t_n} = \mu_{i_1} \times \dots \times \mu_{i_n}$.

По аналогии с введением пространства \mathcal{T} определим измеримое пространство значений всевозможных выборок

$$(\mathcal{X}, \mathcal{A}) = \sum_{t \in \mathcal{T}} (\mathcal{X}_t, \mathcal{A}_t) = \sum_{n=0}^{\infty} \sum_{t_n \in \mathcal{I}^n} (\mathcal{X}_{t_n}, \mathcal{A}_{t_n}).$$

Элемент $\mathbf{x}^{(t_n)} = (x_{i_0}, x_{i_1}, \dots, x_{i_n})$ пространства \mathcal{X} представляет *выборочные данные*, полученные в рамках фиксированной схемы простого случайного выбора. Введенные понятия и пространства позволяют ввести более сложную схему получения выборочных данных.

1.3. Управление статистическим экспериментом и статистическая структура. Процесс наблюдения независимых копий случайных элементов из класса Υ осуществляется посредством последовательного применения двух правил.

Правило остановки $\varphi_s = \{\varphi_s(a | \mathbf{X}^{(t_n)}), a = a_s, a_c; t_n < t_{n+1}, n = 0, 1, \dots\}$ определяется как последовательность переходных вероятностей с выборочных пространств $\{\mathcal{X}_{t_n}, t_n < t_{n+1}, n = 0, 1, \dots, \}$ на алгебру,

порожденную двумя событиями: a_s – остановка эксперимента (прекращение наблюдений) и a_c – продолжение наблюдений. После получения результата $\mathbf{x}^{(t_n)}$ на n -м шаге эксперимента с вероятностью $\varphi_s(a_s | \mathbf{x}^{(t_n)}) = 1 - \varphi_s(a_c | \mathbf{x}^{(t_n)})$ наблюдения прекращаются.

Правило выбора $\varphi_c = \{\varphi_c(i | \mathbf{X}^{(t_n)}), t_n < t_{n+1}, n = 0, 1, \dots, i \in \mathcal{I}\}$ определяется как последовательность переходных вероятностей с выборочных пространств $\{\mathcal{X}_{t_n}, t_n < t_{n+1}, n = 0, 1, \dots, \}$ на фазовую алгебру подмножеств пространства индексов \mathcal{I} . После получения результата $\mathbf{x}^{(t_n)}$ на n -м шаге эксперимента с вероятностью $\varphi_c(i | \mathbf{x}^{(t_n)})$ выбирается индекс i случайного элемента ξ_i , независимая копия X_i которого будет наблюдаться на $(n + 1)$ -м шаге.

В случае *рандомизированных* правил φ_s и φ_c , когда переходные вероятности отличны от нуля и единицы, выбор соответствующих действий в эксперименте осуществляется применением процедуры рандомизации.

Естественно, эти правила можно объединить в одно *правило управления* ρ статистическим экспериментом, определив последовательность переходных вероятностей с упорядоченных выборочных пространств на фазовую алгебру событий пространства $\bar{\mathcal{I}} = \{\mathcal{I}, a_s\}$. Каждая переходная вероятность этого правила определяется как

$$\begin{aligned}\rho(i | \mathbf{X}^{(t_n)}) &= \varphi_c(i | \mathbf{X}^{(t_n)}) \cdot \varphi_s(a_c | \mathbf{X}^{(t_n)}), \quad i \in \mathcal{I}; \\ \rho(a_s | \mathbf{X}^{(t_n)}) &= \varphi_s(a_s | \mathbf{X}^{(t_n)}).\end{aligned}$$

Пара правил $\rho = (\varphi_s, \varphi_c)$ или общее правило ρ называется *управлением статистическим экспериментом*.

С правилами φ_s и φ_c связываются две характеристики управления: *момент остановки* ν и *управляющая переменная* $\tau_\nu = (\iota_1, \dots, \iota_\nu)$. Реализация n момента остановки ν определяет *объем выборки*, а реализация $t_n = (i_1, \dots, i_n)$ управляющей переменной τ_ν – индексы случайных элементов, наблюдаемых в статистическом эксперименте. Управляющая переменная определяет объемы выборок

$$\nu_i, \quad i \in \mathcal{I}, \quad \sum_{i \in \mathcal{I}} \nu_i = \nu,$$

из соответствующих распределений $F_i, i \in \mathcal{I}$.

Распределения ν и τ_ν , согласованных с соответствующими последовательностями сигма-алгебр, определяются условными вероятностями

$$\mathbf{P} \left\{ \nu = n \mid \mathbf{X}^{(t_n)} \right\} = \varphi_s \left(a_s \mid \mathbf{X}^{(t_n)} \right) \prod_{k=1}^{n-1} \varphi_s \left(a_c \mid \mathbf{X}^{(t_k)} \right),$$

$$\mathbf{P} \left\{ \tau_\nu = t_n \mid \mathbf{X}^{(t_n)} \right\} = \varphi_s \left(a_s \mid \mathbf{X}^{(t_n)} \right) \prod_{k=1}^{n-1} \varphi_s \left(a_c \mid \mathbf{X}^{(t_k)} \right) \varphi_c \left(i_{k+1} \mid \mathbf{X}^{(t_k)} \right),$$

$$t_n \in \mathcal{I}^n, \quad n = 0, 1, \dots,$$

которые определяют их распределения при каждом значении параметра $\theta \in \Theta$ посредством их усреднения по распределениям наблюдаемых случайных элементов:

$$\mathbf{P}_\theta(\nu = n) = \sum_{t_n \in \mathcal{I}^n} \int_{\mathcal{X}_{t_n}} \varphi_s \left(a_s \mid x^{(n)} \right) \left[\prod_{k=1}^{n-1} \varphi_s \left(a_c \mid x^{(k)} \right) \right] p_{t_n} \left(x^{(n)} \mid \theta \right) d\mu_{t_n} \left(x^{(n)} \right),$$

$$\mathbf{P}_\theta(\tau_\nu = t_n) = \int_{\mathcal{X}_{t_n}} \varphi_s \left(a_s \mid x^{(n)} \right) \left[\prod_{k=1}^{n-1} \varphi_s \left(a_c \mid x^{(k)} \right) \varphi_c \left(i_{k+1} \mid x^{(k)} \right) \right] p_{t_n} \left(x^{(n)} \mid \theta \right) d\mu_{t_n} \left(x^{(n)} \right),$$

$$t_n \in \mathcal{I}^n, \quad n = 0, 1, \dots$$

Таким образом, статистический эксперимент состоит в наблюдении случайного вектора (*случайной выборки*) $\mathbf{X} = \mathbf{X}^{(\tau_\nu)} = (X_{t_1}, \dots, X_{t_\nu})$ с измеримым пространством значений (*выборочным пространством*) $(\mathcal{X}, \mathcal{A})$. Распределение \mathbf{X} определяется функциями плотности

$$p_{\rho, t_n} \left(x^{(n)} \mid \theta \right) = \varphi_s \left(a_s \mid x^{(n)} \right) \left[\prod_{k=0}^{n-1} \varphi_s \left(a_c \mid x^{(k)} \right) \varphi_c \left(i_{k+1} \mid x^{(k)} \right) \right] p_{t_n} \left(x^{(n)} \mid \theta \right),$$

$$x^{(n)} \in \mathcal{X}_{t_n}, \quad t_n \in \mathcal{I}^n, \quad n = 0, 1, \dots; \quad \theta \in \Theta,$$

по мере μ_{t_n} . Соответствующий этим плотностям класс

$$\mathcal{P}_\rho = \{ \mathcal{P}_{\rho, t}, t \in \mathcal{T} \} = \{ \{ P_{\rho, t}(\cdot \mid \theta), \theta \in \Theta \}, t \in \mathcal{T} \}$$

семейств (по параметру $\theta \in \Theta$) распределений на (X, \mathcal{A}) обычно называется *статистическим экспериментом*, однако в данном контексте это может привести к недоразумениям с понятием статистического эксперимента как последовательности определенных действий по наблюдениям случайных элементов. Класс \mathcal{P}_ρ более естественно назвать *статистической моделью*. Тогда пара семейств $(\mathcal{P}_\rho, \mathcal{G})$ будет называться *статистической структурой*.

Если априорные распределения $G_\lambda \in \mathcal{G}$ не вырождены, то в рамках статистической структуры особо важную роль играет маргинальное распределение случайной выборки \mathbf{X} , определяемое функциями плотности

$$p_{\lambda, t_n}(x^{(n)}) = \int_{\Theta} p_{\rho, t_n}(x^{(n)} | \theta) g_\lambda(\theta) d\chi(\theta), \quad x^{(n)} \in \mathcal{X}_{t_n}, \quad t_n \in \mathcal{I}^n,$$

а также апостериорное распределение ϑ с функцией плотности

$$h_\lambda(\theta | \mathbf{X}) = \frac{p_{\rho, \tau_\nu}(\mathbf{X} | \theta) g_\lambda(\theta)}{p_{\lambda, \tau_\nu}(\mathbf{X})}.$$

1.4. Правило принятия решения и стратегия. После остановки эксперимента статистик приступает к принятию определенного решения $d \in \mathcal{D}$.

Определим *правило принятия решения* φ_d как переходную вероятность $\varphi_d(D | \mathbf{X})$, $D \in \mathcal{C}$, с выборочного пространства \mathcal{X} на фазовую сигма-алгебру \mathcal{C} подмножеств пространства решений \mathcal{D} .

Правило φ_d при каждом фиксированном результате $\mathbf{x}^{(t_n)}$ статистического эксперимента определяет на \mathcal{C} распределение вероятностей $\varphi_d(\cdot | \mathbf{x}^{(t_n)})$, так что практическая реализация φ_d осуществляется наблюдением случайного элемента $\delta_{t_n} = \delta_{t_n}(\mathbf{x}^{(t_n)})$, имеющего данное распределение. Статистика $\delta_{\tau_\nu} = \delta_{\tau_\nu}(\mathbf{X})$ называется *решающей функцией*.

Таким образом, статистическое исследование осуществляется посредством применения тройки правил $\varphi = (\varphi_s, \varphi_c, \varphi_d)$, и данный триплет обычно называется *стратегией* или *процедурой статистического вывода*.

При каждом фиксированном значении $\theta \in \Theta$ правило φ_d порождает стохастическое отображение из $(\mathcal{X}, \mathcal{A}, F_{\rho, \tau_\nu}(\cdot | \theta))$ в $(\mathcal{D}, \mathcal{C}, \Psi(\cdot | \theta))$, где

$$\Psi(D | \theta) =$$

$$\mathbf{E}_\theta \varphi_d(D | \mathbf{X}) = \sum_{n=1}^{\infty} \sum_{t_n \in \mathcal{I}^n} \int_{\mathcal{X}_{t_n}} \varphi_d(D | x^{(n)}) p_{\rho, t_n}(x^{(n)} | \theta) d\mu_{t_n}(x^{(n)}), \quad D \in \mathfrak{C}.$$

Семейство вероятностных мер $\{\Psi(\cdot | \theta), \theta \in \Theta\}$ на измеримом пространстве $(\mathcal{D}, \mathfrak{C})$ называется *образом* стратегии φ .

Поскольку доминированность вероятностной модели определяется только свойствами параметрического пространства Θ , то $\{\Psi(\cdot | \theta), \theta \in \Theta\}$ также доминировано некоторой сигма-конечной положительной мерой γ на $(\mathcal{D}, \mathfrak{C})$. Следуя А.Вальду назовем семейство плотностей $\{\psi(\cdot | \theta), \theta \in \Theta\}$, где

$$\psi(a | \theta) = \frac{d\Psi}{d\gamma}(a | \theta), \quad a \in \mathcal{D},$$

оперативной характеристикой стратегии φ .

В случае невырожденных априорных распределений ϑ важную роль будут играть такие характеристики стратегии φ , как *априорный образ*

$$\Psi_\lambda(D) = \int_{\Theta} \Psi(D | \theta) g_\lambda(\theta) d\chi(\theta), \quad D \in \mathfrak{C}, \lambda \in \Lambda,$$

и *d-апостериорный образ* (*d-апостериорная оперативная характеристика* φ), определяемый функцией плотности

$$\pi_\lambda(\theta | d) = \frac{\psi(d | \theta) g_\lambda(\theta)}{\psi_\lambda(d)}, \quad \theta \in \Theta, d \in \mathcal{D},$$

где $\psi_\lambda(d)$, $d \in \mathcal{D}$, – функция плотности распределения $\Psi_\lambda(\cdot)$ по мере γ . Естественно, такая запись функции плотности $\pi(\cdot | \cdot)$ оправдывается высказанным ранее утверждением о \mathcal{B} -измеримости распределений наблюдаемых случайных элементов и она определена только для тех $d \in \mathcal{D}$, где $\psi_\lambda(d) > 0$.

1.5. Риск процедуры и ее гарантийность. Задание различных распределений и их плотностей, определяемых статистической структурой и стратегией φ , необходимо для вычисления основной характеристики φ – величины средних потерь “в длинном ряду статистических экспериментов”. Несмотря на всю эвристичность последней, взятой в кавычки, фразы,

именно такое осмысление величины средних потерь приводит к корректному, отвечающему существу дела определению гарантийности процедур статистического вывода. Средние потери, обусловленные многократным применением стратегии φ , обычно называются *риском* данной стратегии, и существует несколько определений этой основной характеристики стратегии.

Исторически первое определение соотносится с экспериментами, в которых θ не является реализацией случайной величины. Неизвестное статистическое значение θ является некоторой “мировой” константой, например скоростью света в пустоте, зарядом электрона, вероятностью наследования доминантного признака и т.п., хотя область использования такого подхода к определению риска была несравненно шире. Величина средних потерь определяется посредством формального рассмотрения последовательности независимых статистических экспериментов, в которых значение θ одно и то же. Среднее значение потерь

$$R(\varphi | \theta) = \mathbf{E}_\theta L \left(\theta, \delta_{\tau_\nu} \left(\mathbf{X}^{(\tau_\nu)} \right) \right) = \int_{\mathcal{D}} L(\theta, a) \psi(a | \theta) d\gamma(a), \quad \theta \in \Theta,$$

называется *функцией риска*. Чтобы отметить в данном определении особую роль фиксированного значения θ , в дальнейшем $R(\varphi | \theta)$, $\theta \in \Theta$, будет называться также *θ -риском*.

Иной подход к риску специфичен для ситуаций, когда существует реальная последовательность статистических экспериментов, в которых значения параметра θ допускают интерпретацию как последовательность реализаций случайных величин. Это задачи контроля качества выпускаемой продукции, медицинской диагностики, социологических исследований по так называемым малым областям и т.п. Так, например, в задачах контроля качества производитель проводит контроль качества каждой партии выпускаемой продукции, принимая решение о ее реализации (поставки потребителю), или решение отклонить партию от приемки с особым решением о ее дальнейшей судьбе. Величина θ -риска при функции потерь типа 0–1 состоит из двух компонент: риска изготовителя и риска поставщика. Первая компонента определяет вероятность отклонения кондиционного продукта

при его контроле, а вторая – вероятность поставки потребителю не того, чего бы он хотел. С точки зрения, скажем, потребителя, второй риск ему не слишком интересен – его в большой степени волнует доля некондиционной продукции, которую он получает от завода–изготовителя. Следовательно, как потребителю, так и изготовителю важно знать величину средних потерь среди тех экспериментов, которые закончились принятием только одного из решений: принятием продукции для потребителя и ее отклонением для поставщика. Так как в этой и аналогичных проблемах значения $\theta_1, \theta_2, \dots$ в последовательности статистических экспериментов по контролю периодически выпускаемой продукции можно трактовать как результаты реализации простого случайного выбора из априорного распределения G , то средние потери определяются величиной d -риска – значением $R_\lambda(\varphi | d)$, $d \in \mathcal{D}$, условного математического ожидания

$$R_\lambda(\varphi | \delta) = \mathbf{E} \{L(\vartheta, \delta) | \delta\} = \int_{\Theta} L(\theta, \delta) \pi_\lambda(\theta | \delta) d\chi(\theta)$$

функции L (как функции случайного элемента ϑ) относительно сигма-алгебры, порожденной решающей функцией $\delta = \delta_{\tau_\nu}(\mathbf{X}^{(\tau_\nu)})$.

Следует особо отметить, что d -риск зависит также от параметра $\lambda \in \Lambda$, определяющего распределение ϑ . Именно наличие реальной последовательности исходов статистических экспериментов поставяет информацию об истинном “значении” λ в конкретной статистической проблеме и позволяет определить (хотя бы приближенно) априорное распределение.

Наконец, введем определение риска стратегии φ , которое обычно используется в так называемой *байесовской статистике*. Это интегральная характеристика, в которой потери усредняются как по значениям параметра, так и по принятым решениям, называется *априорным риском* стратегии φ :

$$R_\varphi(\lambda) = \mathbf{E} L(\vartheta, \delta) = \int_{\Theta} \int_{\mathcal{D}} L(\theta, a) \psi(a | \theta) g_\lambda(\theta) d\chi(\theta) d\gamma(a), \quad \lambda \in \Lambda.$$

Данная характеристика стратегии используется в проблемах минимизации θ -риска, когда априорное распределение G трактуется как некоторая

весовая функция и в общем случае G не обязательно является распределением вероятностей. С точки же зрения d-риска априорный риск является некоторой огрубленной характеристикой стратегии φ .

В проблемах минимизации d-риска и априорного риска основную роль играет *апостериорный риск* – условное среднее потерь по ϑ относительно сигма-алгебры, порожденной случайной выборкой:

$$\mathfrak{R}_\lambda(d | \mathbf{X}) = \mathbf{E}\{L(\vartheta, d) | \mathbf{X}\} = \int_{\Theta} L(\theta, d) h_\lambda(\theta | \mathbf{X}) d\chi(\theta), \quad d \in \mathcal{D}.$$

Легко видеть, что *априорный риск равен среднему значению апостериорного риска по маргинальному распределению \mathbf{X} , а d-риск равен условному математическому ожиданию апостериорного риска относительно решающей функции δ* , что, собственно, и является важным при построении стратегий, минимизирующих величину риска.

Стратегия φ называется *гарантийной*, если ее риск ограничен сверху некоторой функцией $r(\cdot)$ от, соответственно, θ , d или λ , – разным определениям риска соответствуют разные понятия гарантийности статистического вывода. В связи с этим введем три вида нормированных функций потерь $L(\theta, d)/r(\theta)$, $L(\theta, d)/r(d)$, $L(\theta, d)/r(\lambda)$ и соответствующие им нормированные риски $\bar{R}(\varphi | \theta)$, $\bar{R}_\lambda(\varphi | d)$, $\bar{R}_\varphi(\lambda)$.

Стратегия φ называется

θ -гарантийной, если $\bar{R}(\varphi | \theta) \leq 1$, $\forall \theta \in \Theta$;

d -гарантийной, если $\bar{R}_\lambda(\varphi | d) \leq 1$, $\forall d \in \{d : \psi_\lambda(d) > 0, d \in \mathcal{D}\}$, $\forall \lambda \in \Lambda$;

λ -гарантийной, если $\bar{R}_\varphi(\lambda) \leq 1$, $\forall \lambda \in \Lambda$.

С введенными определениями риска стратегии и ее гарантийности связаны две основные задачи математической статистики.

(I) При фиксированном управлении $\rho = (\varphi_s, \varphi_c)$ статистическим экспериментом в рамках функции потерь определенного вида и в заданном классе правил принятия решения φ_d строится правило φ_d^* , которое доставляет минимум риска при любом значении его аргумента: равномерно по всем $\theta \in \Theta$ минимизирует величину θ -риска, доставляет минимум значению d-риска при любых $d \in \mathcal{D}$ и любых $\lambda \in \Lambda$, минимизирует априорный

риск (λ -риск) при каждом $\lambda \in \Lambda$. Один из вариантов этой задачи – построение минимаксных правил φ_d , минимизирующих наибольшее из возможных значений риска.

(II) Построение гарантийных процедур статистического вывода как таковых и отыскание среди них стратегий, минимизирующих определенные функционалы от средних затрат на постановку статистического эксперимента, в частности, минимизирующих функционал от среднего значения объема выборки (момента останова ν статистического эксперимента). В таких задачах основную роль играют такие характеристики управления ρ статистическим экспериментом, как

$$\mathbf{C}(\theta) = \sum_{i \in \mathcal{I}} \mathbf{E}_{\theta} \nu_i, \theta \in \Theta; \quad \mathbf{C}_{\lambda} = \mathbf{E}_{\lambda} \mathbf{C}(\vartheta), \lambda \in \Lambda;$$

$$\mathbf{C}_{\lambda}(d) = \sum_{i \in \mathcal{I}} \mathbf{E}_{\lambda} \{\nu_i \mid \delta(X) = d\}, d \in \mathcal{D}.$$

Примеры задачи (I): построение несмещенных оценок параметра θ с равномерно минимальным θ -риском при выпуклых функциях потерь и оценок с равномерно минимальным значением d -риска; построение равномерно наиболее мощных критериев в классе критериев заданного уровня α и аналогичная задача для d -риска; байесовские решения и их эмпирические аналоги равномерной по $\lambda \in \Lambda$ минимизации априорного риска в асимптотическом смысле (при больших архивах данных).

Пример задачи (II) – определение необходимого объема выборки: в классе гарантийных стратегий с фиксированными объемами наблюдений $n (= 0, 1, \dots)$ найти стратегию с минимальным значением n . Более сложная задача – минимизация функционалов от средних затрат $\mathbf{C}(\theta)$ или \mathbf{C}_{λ} на постановку статистического эксперимента.

В дальнейшем нас будут интересовать, в основном, задачи последнего типа, то есть проблемы планирования объема наблюдений (или, более общо, построение правила останова статистического эксперимента), обеспечивающего существование гарантийной процедуры статистического вывода.

ЗАЧЕТНЫЕ ЗАДАНИЯ

Выборка из нормального (θ, σ^2) распределения проводится в два этапа. Назначаются две константы c_1 и c_2 , $c_1 < c_2$. Сначала берется выборка объема $n_1 = 2$. Если результаты (x_1, x_2) наблюдений (X_1, X_2) оказались больше c_2 , то наблюдения прекращаются. Если $c_1 < \min\{x_1, x_2\} \leq c_2$, то наблюдения продолжаются с вероятностью 0,5, если же $\min\{x_1, x_2\} \leq c_1$, то – с вероятностью единица. В случае принятия решения о продолжения наблюдений на втором этапе берется выборка объема 4, то есть наблюдаются независимые случайные величины (X_3, \dots, X_6) и статистический эксперимент завершается.

1. Определить правило остановки φ_s статистического эксперимента.
2. Найти распределение момента остановки ν .
3. Определить функцию плотности случайной выборки $X^{(\nu)}$.
4. Найти распределение выборочного среднего $\bar{X} = \frac{1}{\nu} \sum_{k=1}^{\nu} X_k$.
5. Рассматривая \bar{X} как нерандомизированную оценку значения θ , определить правило φ_d принятия решения и найти его θ -риск при квадратичной функции потерь $L(\theta, d) = (\theta - d)^2$.
- 6*. Пусть θ есть реализация случайной величины ϑ с нормальным (μ, τ^2) распределением. Вычислить d-риск \bar{X} при квадратичной функции потерь.

§ 2. Статистические модели, обладающие достаточными статистиками

2.1. Достаточность, полнота и подчиненность. В предыдущем параграфе мы рассматривали статистический эксперимент, который состоит в наблюдении случайного вектора (*случайной выборки*) $\mathbf{X} = \mathbf{X}^{(\tau_\nu)} = (X_{t_1}, \dots, X_{t_\nu})$ с измеримым пространством значений (*выборочным пространством*) $(\mathcal{X}, \mathcal{A})$. Распределение \mathbf{X} определяется функциями плотности

$$p_{\rho, t_n} \left(x^{(n)} | \theta \right) = \varphi_s \left(a_s | x^{(n)} \right) \left[\prod_{k=0}^{n-1} \varphi_s \left(a_c | x^{(k)} \right) \varphi_c \left(i_{k+1} | x^{(k)} \right) \right] p_{t_n} \left(x^{(n)} | \theta \right),$$

$$x^{(n)} \in \mathcal{X}_{t_n}, t_n \in \mathcal{I}^n, n = 0, 1, \dots; \theta \in \Theta,$$

по мере μ_{t_n} . Соответствующий этой плотности класс

$$\mathcal{P}_\rho = \{ \mathcal{P}_{\rho, t}, t \in \mathcal{T} \} = \{ \{ P_{\rho, t}(\cdot | \theta), \theta \in \Theta \}, t \in \mathcal{T} \}$$

семейств (по параметру $\theta \in \Theta$) распределений на $(\mathcal{X}, \mathcal{A})$ обычно называется *статистическим экспериментом*, но, чтобы не было недоразумений с понятием статистического эксперимента как последовательности определенных действий по наблюдениям случайных элементов, мы договорились называть класс \mathcal{P}_ρ *статистической моделью*. Тогда пара семейств $(\mathcal{P}_\rho, \mathcal{G})$ будет называться *статистической структурой*.

В данном параграфе, за исключением некоторых иллюстративных примеров, не будет играть особой роли управление ρ статистическим экспериментом. Поэтому мы упростим обозначения: пусть \mathbf{X} означает случайную выборку, $p(x | \theta)$ – ее функцию плотности по мере μ , $P(A | \theta)$, $A \in \mathcal{A}$, – распределение \mathbf{X} на измеримом выборочном пространстве $(\mathcal{X}, \mathcal{A})$, а $\mathcal{P} = \{ P(\cdot | \theta), \theta \in \Theta \}$ – статистическую модель.

Сейчас мы изучим особый класс статистических моделей, который допускает значительную редукцию выборочных данных без потери информации путем принятия решения только по значению некоторой статистики $T(\mathbf{X})$ – измеримой функции от случайной выборки \mathbf{X} .

Определение 2.1. Статистика $T = T(\mathbf{X})$ называется *достаточной* для семейства распределений (статистической модели) $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$, если условное распределение случайной выборки \mathbf{X} относительно статистики T (точнее, относительно σ -подалгебры $\mathcal{A}_T \subset \mathcal{A}$, порожденной статистикой T) не зависит от параметра $\theta (\in \Theta)$.

Теорема 2.1. Пусть $\varphi_d = \varphi_d(D | \mathbf{X})$, $D \in \mathcal{C}$, – некоторое правило принятия решения в статистической проблеме с измеримым пространством решений $(\mathcal{D}, \mathcal{C})$. Если статистическая модель \mathcal{P} обладает достаточной статистикой T , то существует правило принятия решения $\tilde{\varphi}_d = \tilde{\varphi}_d(D | T)$, $D \in \mathcal{C}$, зависящее от выборочных данных только через значение статистики T и имеющее тот же образ (распределение решающей функции) $\{\Psi(\cdot | \theta), \theta \in \Theta\}$, что и правило φ_d .

Доказательство. Указанное в формулировке теоремы правило $\tilde{\varphi}_d$ находится с помощью вычисления условного математического ожидания φ_d относительно статистики T :

$$\tilde{\varphi}_d(D | T(\mathbf{X})) = \mathbf{E} \{ \varphi_d(D | \mathbf{X}) | T(\mathbf{X}) \}.$$

Это действительно правило принятия решения, поскольку усреднение не изменяет область значений $[0, 1]$ правила φ_d и, в силу досточности T , правило $\tilde{\varphi}_d$ не зависит от θ . Оно имеет тот же образ, поскольку

$$\begin{aligned} \tilde{\Psi}(\cdot | \theta) &= \mathbf{E}_\theta \tilde{\varphi}_d(\cdot | T(\mathbf{X})) = \\ &= \mathbf{E}_\theta [\mathbf{E} \{ \varphi_d(\cdot | \mathbf{X}) | T(\mathbf{X}) \}] = \mathbf{E}_\theta \varphi_d(\cdot | \mathbf{X}) = \Psi(\cdot | \theta) \end{aligned}$$

(как известно, математическое ожидание от условного среднего равно безусловному среднему). \square

Заметим, что совпадение образов правил φ_d и $\tilde{\varphi}_d$ влечет совпадение их θ -, λ - и d -рисков.

Найти достаточную статистику, если она существует, можно с помощью следующего *критерия факторизации* Неймана.

Теорема 2.2. Для того чтобы статистика T была достаточной для семейства \mathcal{P} , доминированного σ -конечной мерой μ , необходимо и достаточно

существования таких неотрицательных функций g_θ и h , чтобы плотность $p(x|\theta)$ допускала представление

$$p(x|\theta) = g_\theta(T(x))h(x) \quad (2.1)$$

почти всюду по мере μ .

Доказательство теоремы мы проведем только для дискретного семейства распределений, когда функция плотности выборки $p(x|\theta) = \mathbf{P}_\theta(\mathbf{X} = x)$. В случае непрерывного распределения схема доказательства та же, но придется делать замену в n -кратном интеграле. Доказательство теоремы для произвольных семейств распределений достаточно сложно и приводится в учебнике А.А. Боровкова “Математическая статистика”, М.: Наука, 1984, с. 431-434.

Достаточность. Пусть выполняется факторизационное представление (2.1); требуется показать, что условное распределение \mathbf{X} относительно T не зависит от θ . Воспользуемся формулой условной вероятности для вычисления условной плотности \mathbf{X} относительно T :

$$\mathbf{P}_\theta(\mathbf{X} = x | T(\mathbf{X}) = t) = \frac{\mathbf{P}_\theta(\{\mathbf{X} = x\} \cap \{T(\mathbf{X}) = t\})}{\mathbf{P}_\theta(T(\mathbf{X}) = t)}.$$

События, стоящие в числителе, будут несовместными, если $T(\mathbf{X}) \neq t$, и в этом случае условная вероятность равна нулю (не зависит от θ). Если же $T(\mathbf{X}) = t$, то первое по порядку событие в числителе влечет второе, и поэтому формула для вычисления условной вероятности упрощается:

$$\mathbf{P}_\theta(\mathbf{X} = x | T(\mathbf{X}) = t) = \frac{\mathbf{P}_\theta(\mathbf{X} = x)}{\mathbf{P}_\theta(T(\mathbf{X}) = t)}.$$

Так как

$$\mathbf{P}_\theta(\mathbf{X} = x) = p(x|\theta),$$

то используя представление (2.1), получаем, что (напомним, $T(x) = t$)

$$\mathbf{P}_\theta(\mathbf{X} = x | T(\mathbf{X}) = t) =$$

$$\frac{g_\theta(T(x))h(x)}{\sum_{y: T(y)=t} g_\theta(T(y))h(y)} = \frac{h(x)}{\sum_{y: T(y)=t} h(y)}.$$

Таким образом, условное распределение не зависит от θ , и поэтому статистика T достаточна для \mathcal{P} .

Необходимость. Пусть T – достаточная статистика, так что условное распределение

$$\mathbf{P}_\theta(\mathbf{X} = x | T() = t) = K(x, t),$$

где функция K не зависит от θ . Требуется показать, что в этом случае для функции плотности выборки справедливо представление (2.1).

Имеем

$$\begin{aligned} p(x | \theta) &= \mathbf{P}_\theta(\mathbf{X} = x) = \\ &= \mathbf{P}_\theta\left(\{\mathbf{X} = x\} \cap \{T(\mathbf{X}) = T(x)\}\right) = \\ &= \mathbf{P}_\theta(T(\mathbf{X}) = T(x)) \cdot \mathbf{P}_\theta(\mathbf{X} = x | T(\mathbf{X}) = T(x)). \end{aligned}$$

Мы получили представление (2.1) с

$$g_\theta(T(x)) = \mathbf{P}_\theta(T(\mathbf{X}) = T(x)), \quad h(x) = K(x, T(x)). \quad \square$$

Одно существенное замечание к этой теореме. Если \mathcal{P} – статистическая модель с управлением наблюдениями, для которой имеет место факторизационное тождество (2.1), то достаточной статистикой будет триплет $(\nu, \tau_\nu, T(\mathbf{X}^{(\tau_\nu)}))$, состоящий из момента остановки ν , управляющей переменной τ_ν и статистики $T(\mathbf{X}^{(\tau_\nu)})$.

Критерий факторизации (2.1) неоднозначно определяет достаточную статистику. Например, $\sum_1^n X_k$ – достаточная статистика при выборе в схеме Бернулли с вероятностью успешного испытания θ . Из представления (2.1) следует, что двумерная статистика $(\sum_1^m X_k, \sum_{m+1}^n X_k)$ будет также достаточной, но эта статистика осуществляет меньшую редукцию данных. Более общо, если T достаточна и $T = h(S)$, то S также достаточна, – знание значения S влечет знание T . Кроме того, T обеспечивает бóльшую редукцию данных, чем S , если только h не есть взаимно однозначное отображение,

и тогда T и S эквивалентны с точки зрения поступающей из наблюдений информации. Естественно, среди всех достаточных статистик предпочтение должно отдаваться той, которая осуществляет наиболее полную редукцию данных.

Определение 2.2. Достаточная статистика T называется *минимальной*, если для любой достаточной статистики S существует функция h , такая, что $T = h(S)$.

Обычно найти минимальную достаточную статистику довольно легко, если все распределения из семейства \mathcal{P} имеют общий носитель. Метод ее построения указывают следующие просто устанавливаемые утверждения.

Предложение 2.1. Пусть \mathcal{P} – конечное семейство с плотностями p_0, p_1, \dots, p_m , имеющими один и тот же носитель. Тогда m -мерная статистика отношения правдоподобия

$$T(\mathbf{X}) = \left(\frac{p_1(\mathbf{X})}{p_0(\mathbf{X})}, \dots, \frac{p_m(\mathbf{X})}{p_0(\mathbf{X})} \right)$$

будет минимальной достаточной статистикой.

Доказательство. Из факторизационного тождества (2.1) следует, что T есть функция любой другой достаточной статистики и, следовательно, является минимальной. \square

Предложение 2.2. Пусть \mathcal{P} – семейство распределений с общим носителем и \mathcal{P}_0 – подсемейство \mathcal{P} . Если T есть минимальная достаточная статистика для \mathcal{P}_0 и достаточная статистика для \mathcal{P} , то T – минимальная достаточная статистика и для \mathcal{P} .

Доказательство. Если S – любая достаточная статистика для \mathcal{P} , то она, в силу критерия факторизации, достаточна и для более узкого семейства \mathcal{P}_0 . Так как T минимальная достаточная статистика для \mathcal{P}_0 , то существует такая функция h , что $T = h(S)$. Следовательно, статистика T , будучи достаточной для всего семейства \mathcal{P} , является функцией любой достаточной для этого семейства статистики, то есть является минимальной достаточной для \mathcal{P} . \square

Общий подход к использованию этих предложений для нахождения минимальной достаточной статистики виден на следующем примере.

ПРИМЕР 2.1. *Минимальная достаточная статистика для статистической модели гамма-распределения.* Пусть $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$ – семейство распределений случайной выборки $X = (X_1, \dots, X_n)$ фиксированного объема n из двухпараметрического гамма-распределения с плотностью

$$f(x | \theta) = \frac{1}{a^\lambda \Gamma(\lambda)} x^{\lambda-1} \exp \left\{ -\frac{x}{a} \right\}, \quad x > 0; \quad \theta = (a, \lambda) \in \Theta = \mathbb{R}_+^2.$$

Для отыскания минимальной достаточной статистики рассмотрим подсемейство, состоящее из трех распределений, параметры которых фиксированы и равны некоторым $\theta_0, \theta_1, \theta_2$. Статистика отношения правдоподобия из предложения 2.1 в данном конкретном случае состоит из двух компонент: $T = (T_1, T_2)$, где

$$T_1(X) = \frac{a_0^{n\lambda_0} \Gamma^n(\lambda_0)}{a_1^{n\lambda_1} \Gamma^n(\lambda_1)} \left[\prod_{k=1}^n X_k \right]^{\lambda_1 - \lambda_0} \exp \left\{ -\left(\frac{1}{a_1} - \frac{1}{a_0} \right) \sum_{k=1}^n X_k \right\},$$

$$T_2(X) = \frac{a_0^{n\lambda_0} \Gamma^n(\lambda_0)}{a_2^{n\lambda_2} \Gamma^n(\lambda_2)} \left[\prod_{k=1}^n X_k \right]^{\lambda_2 - \lambda_0} \exp \left\{ -\left(\frac{1}{a_2} - \frac{1}{a_0} \right) \sum_{k=1}^n X_k \right\}.$$

Легко видеть, что статистика T эквивалентна (порождает ту же σ -подалгебру) статистике

$$\left(\prod_{k=1}^n X_k, \quad \sum_{k=1}^n X_k \right),$$

которая в силу предложения 2.2 является минимальной достаточной.

Кроме минимальности достаточная статистика должна обладать еще одним, не менее важным свойством, необходимым в построении оптимальных процедур статистического вывода.

Определение 2.3. Семейство вероятностных мер $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$ статистики $T = T(\mathbf{X})$ называется *полным* (в дальнейшем короче – статистика $T = T(\mathbf{X})$ называется *полной*), если для любой измеримой функции g равенство $\mathbf{E}_\theta g(T(\mathbf{X})) = 0$ при всех $\theta \in \Theta$ влечет $g(t) = 0$ почти всюду по любой мере из семейства \mathcal{P} .

Как будет видно позже, полнота приносит с собой значительное упрощение статистической ситуации. Полные достаточные статистики особенно эффективны при редукции данных, поскольку справедливо следующее утверждение *полная достаточная статистика всегда минимальна* или (более точная формулировка), *если минимальная достаточная статистика существует, то для того, чтобы достаточная статистика была полной, необходимо, чтобы она была минимальной*. Идею простого доказательства этого утверждения можно найти на с. 68 книги Э.Лемана “Теория точечного оценивания”.

Приведем несколько статистических моделей, обладающих полными достаточными статистиками.

ПРИМЕР 2.2. *Полнота достаточной статистики в схеме испытаний Бернулли.* Пусть $X = (X_1, \dots, X_n)$ случайная выборка из двухточечного распределения: каждое X_i , $i = 1, \dots, n$, принимает значение 1 с вероятностью θ и значение 0 с вероятностью $1 - \theta$. Семейство \mathcal{P} распределений X определяется функцией плотности

$$p(x | \theta) = f_n(x^{(n)} | \theta) = \theta^{\sum_1^n x_i} (1 - \theta)^{n - \sum_1^n x_i},$$

а семейство распределений достаточной статистики $T = \sum_1^n X_i$ – функцией плотности биномиального распределения

$$h(t | \theta) = C_n^t \theta^t (1 - \theta)^{n-t}, \quad t = 0, 1, \dots, n, \quad \theta \in (0; 1).$$

По аналогии с построениями примера 2.1 легко показать, что T – минимальная достаточная статистика. Докажем, что T полная достаточная статистика: из справедливости равенства

$$\sum_{t=0}^n g(t) C_n^t \theta^t (1 - \theta)^{n-t} = 0 \tag{2.2}$$

для всех $\theta \in (0; 1)$ следует $g(t) = 0$ в точках $t = 0, 1, \dots, n$ носителя распределения статистики T .

Положив $z = \theta/(1 - \theta)$, представим (2.2) в эквивалентном виде

$$\sum_{t=0}^n g(t) C_n^t z^t = 0.$$

Правая часть этого равенства представляет многочлен от переменной $z \in \mathbb{R}_+$ степени n . Равенство нулю многочлена при всевозможных значениях переменной z из интервала $(0, \infty)$ возможно лишь при нулевых значениях его коэффициентов, то есть когда $g(t) = 0$ при любом $t = 0, 1, \dots, n$. Следовательно, T – полная достаточная статистика.

ПРИМЕР 2.3. *Полнота достаточной статистики для семейства распределений случайной выборки из показательного распределения.* Пусть $X = (X_1, \dots, X_n)$ – случайная выборка из показательного распределения с плотностью $f(x | \theta) = \theta \exp\{-\theta x\}$, $x \geq 0$, $\theta > 0$. Семейство \mathcal{P} распределений случайной выборки X определяется n -мерной функцией плотности

$$p(x^{(n)} | \theta) = \theta^n \exp\left\{-\theta \sum_{i=1}^n x_i\right\},$$

так что, в силу теоремы факторизации, $T = \sum_{i=1}^n X_i$ – достаточная статистика. Минимальность T доказывается по аналогии с примером 2.1. Установим полноту T , используя непосредственно определение полной статистики.

Статистика T имеет гамма-распределение с плотностью

$$h(t | \theta) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t}.$$

Равенство

$$\frac{\theta^n}{\Gamma(n)} \int_0^{\infty} g(t) t^{n-1} e^{-\theta t} dt = 0 \tag{2.3}$$

при всех $\theta \in \mathbb{R}_+$ влечет $g(t) = 0$ для почти всех по мере Лебега $t > 0$, поскольку интеграл в правой части (2.4) представляет преобразование Лапласа, для которого, как и для преобразования Фурье, существует формула обращения, показывающая, что прообразом функции, равной нулю, может быть только функция, также равная нулю.

ПРИМЕР 2.4. *Полнота достаточной статистики для семейства распределений случайной выборки из равномерного распределения.* Пусть $X = (X_1, \dots, X_n)$ – случайная выборка из равномерного на интервале $(0, \theta)$, $\theta > 0$, распределения с плотностью $f(x | \theta) = \mathbf{I}_{(0, \theta)}(x)/\theta$, $x \in \mathbb{R}_+$, где $\mathbf{I}_{(0, \theta)}(x)$ – индикаторная функция, так что f отлична от нуля и равна

$1/\theta$ только на интервале $(0, \theta)$. Случайная выборка X имеет n -мерную функцию плотности

$$p(x^{(n)} | \theta) = \frac{\prod_{i=1}^n \mathbf{I}_{(0, \theta)}(x_i)}{\theta^n},$$

которая отлична от нуля только в том случае, когда крайний член вариационного ряда $X_{(n)} \leq \theta$. В силу теоремы факторизации $X_{(n)}$ является достаточной статистикой. Установим полноту этой статистики.

Нетрудно убедиться, что статистика $T = X_{(n)}$ имеет распределение, сосредоточенное на интервале $(0, \theta)$, с плотностью $h(t | \theta) = n t^{n-1} / \theta^n$. Равенство

$$n \theta^{-n} \int_0^{\theta} g(t) t^{n-1} dt = 0$$

при всех $\theta > 0$ влечет, что для любого интервала $(a, b) \subset (0, \infty)$

$$\int_a^b g(t) t^{n-1} dt = 0. \quad (2.4)$$

Пусть $g(t) = g^+(t) - g^-(t)$, где $g^+(t)$ и $-g^-(t)$ обозначают положительную и отрицательную части $g(t)$ соответственно. Подставляя это разложение в правую часть (2.4), получаем, что для любого интервала (a, b)

$$\int_a^b g^+(t) t^{n-1} dt = \int_a^b g^-(t) t^{n-1} dt.$$

Из построений при определении интеграла Лебега следует, что аналогичное равенство справедливо для любого борелевского множества A на положительной прямой. Следовательно, две положительные меры

$$\mu^+(A) = \int_A g^+(t) t^{n-1} dt \quad \text{и} \quad \mu^-(A) = \int_A g^-(t) t^{n-1} dt$$

совпадают на всех борелевских множествах, что с необходимостью влечет $g^+(t) = g^-(t)$ почти всюду по мере Лебега. Поэтому $g(t) = 0$ для почти всех t .

Итак, $X_{(n)}$ — полная достаточная статистика и, следовательно, минимальная.

ПРИМЕР 2.5 *неполной достаточной статистики.* Пусть X_1, \dots, X_n и Y_1, \dots, Y_m – случайные выборки из нормальных распределений с параметрами (μ, σ_1) и (μ, σ_2) соответственно (средние значения распределений совпадают, а дисперсии разные). Статистическая модель определяется $(n + m)$ -мерной функцией плотности

$$f_{n+m}(x^{(n)}, y^{(m)} | \theta) = C(\theta) \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_1^n x_i^2 + \frac{\mu}{\sigma_1^2} \sum_1^n x_i - \frac{1}{2\sigma_2^2} \sum_1^m y_i^2 + \frac{\mu}{\sigma_2^2} \sum_1^m y_i \right\},$$

$$\theta = (\mu, \sigma_1, \sigma_2).$$

В силу теоремы факторизации статистика

$$T = \left(\sum_1^n X_i, \sum_1^n X_i^2, \sum_1^m Y_i, \sum_1^m Y_i^2, \right)$$

является достаточной. Однако она неполна, поскольку

$$\mathbf{E}_\theta \left(\frac{1}{n} \sum_1^n X_i - \frac{1}{m} \sum_1^m Y_i, \right) = 0$$

при всех значениях трехмерного параметра θ .

Рассмотрим еще один класс статистик, который приводит к весьма интересным результатам, когда существуют полные достаточные статистики.

Определение 2.4. Статистика $V = V(X)$ называется *подчиненной*, если ее распределение не зависит от параметра θ , индексирующего статистическую модель.

Например, статистическая модель выборки из нормального (θ, σ^2) распределения при фиксированном значении дисперсии σ^2 обладает полной (следовательно, минимальной) достаточной статистикой \bar{X} (выборочное среднее). Распределение выборочной дисперсии S^2 не зависит от θ , поэтому S^2 – подчиненная статистика.

Как было показано в общем курсе математической статистики, \bar{X} и S^2 – независимые статистики. Этот результат следует из следующего утверждения, носящего более общий характер.

Теорема 2.3. (теорема Басу) Если T – полная достаточная статистика для семейства \mathcal{P} , то любая подчиненная статистика V не зависит от T .

Доказательство. Если V – подчиненная статистика, то вероятность $p_A = \mathbf{P}_\theta(V \in A)$ не зависит от θ для всех A . Пусть $\eta_A(T) = \mathbf{P}(V \in A | T)$. Тогда $\mathbf{E}_\theta \eta_A(T) = p_A$ и, следовательно, в силу полноты статистики T равенство $\mathbf{E}_\theta [\eta_A(T) - p_A] = 0$ для всех $\theta \in \Theta$ влечет

$$\eta_A(T) = \mathbf{P}(V \in A | T) = p_A = \mathbf{P}(V \in A).$$

Последнее равенство есть определяет независимость статистик T и V . □

2.2. Экспоненциальное семейство статистических моделей.

Многочисленные примеры статистических моделей, обладающих достаточными статистиками, были даны в общем курсе математической статистики. Все они принадлежали одному замечательному классу семейств, которое имеет следующий вид.

Определение 2.5. Семейство распределений (статистических моделей) $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$, образует k -параметрическое экспоненциальное семейство, если распределения $P(\cdot | \theta)$ имеют плотности вида

$$p(x | \theta) = C(\theta) \exp \left[\sum_{i=1}^k Q_i(\theta) T_i(x) \right] h(x) \quad (2.5)$$

относительно некоторой общей меры μ .

В представлении (2.5) Q_i и C суть вещественнозначные функции от параметров, T_i , $i = 1, \dots, k$, – вещественнозначные статистики.

К экспоненциальным семействам принадлежат такие распределения, как биномиальное, пуассоновское, мультиномиальное, нормальное (включая многомерное), гамма и ряд других. Известен также следующий замечательный результат, принадлежащий Купману и Дармуа: если носитель распределения $P(\cdot | \theta)$ случайной выборки фиксированного объема n не зависит от θ , то при определенных условиях регулярности существование достаточной статистики размерности $k < n$ влечет экспоненциальность

семейства $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$. Пример семейства, которое обладает достаточной статистикой, но не является экспоненциальным, – семейство распределений случайной выборки фиксированного объема n из равномерного на интервале (θ_1, θ_2) распределения, для которого существует двумерная достаточная статистика $(X_{(1)}, X_{(n)})$ – крайние члены вариационного ряда.

Экспоненциальным семействам, кроме того, что они обладают достаточными статистиками, присущ ряд замечательных свойств, которые будут существенно использоваться в дальнейшем. Доказательства этих свойств не сложны и их можно найти в книге Э.Лемана “Проверка статистических гипотез”, § 7, глава 2.

Предложение 2.3. Пусть φ – ограниченная измеримая функция на выборочном пространстве $(\mathcal{X}, \mathcal{A})$. Тогда интеграл

$$\int_{\mathcal{X}} \varphi(x) \exp \left[\sum_{i=1}^k \theta_i T_i(x) \right] d\mu(x),$$

рассматриваемый как функция переменных $\theta_1, \dots, \theta_k$, обладает производными любого порядка и эти производные можно вычислять дифференцированием под знаком интеграла.

Предложение 2.4. Пусть X имеет распределение из экспоненциального семейства

$$dP(x | \theta, \vartheta) = B(\theta, \vartheta) \exp \left[\sum_{i=1}^r \theta_i U_i(x) + \sum_{j=1}^s \vartheta_j T_j(x) \right] d\mu(x).$$

Тогда существуют меры $\lambda_\theta, \theta \in \Theta$, и вероятностные меры $\nu_t, t \in \mathcal{T}$, на s и r -мерном евклидовых пространствах соответственно, такие, что

(I) распределение векторной статистики $T = (T_1, \dots, T_s)$ принадлежит экспоненциальному семейству

$$dP^T(t | \theta, \vartheta) = B(\theta, \vartheta) \exp \left[\sum_{j=1}^s \vartheta_j t_j \right] d\lambda_\theta(t),$$

(II) условное распределение векторной статистики $U = (U_1, \dots, U_r)$ относительно статистики T принадлежит экспоненциальному семейству

вида

$$dP_{\theta}^{U|T}(u|t) = B_t(\theta) \exp \left[\sum_{i=1}^r \theta_i u_i \right] d\nu_t(u),$$

то есть, в частности, не зависит от ϑ .

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. Найти минимальную достаточную статистику для статистической модели (выборка фиксированного объема n) двухпараметрического распределения Парето с плотностью

$$f(x|\theta) = \frac{a}{\lambda} \left(\frac{a}{x} \right)^{\lambda+1}, \quad x > a; \quad \theta = (a, \lambda), \quad a > 0, \quad \lambda > 0.$$

2. Пусть $h(x)$, $x > 0$, – положительная интегрируемая функция, которая определяет следующую функцию плотности:

$$f(x|\theta) = \begin{cases} c(\theta) h(x), & \text{если } 0 < x < \theta, \\ 0, & \text{если } x \notin [0, \theta]. \end{cases}$$

Докажите, что $X_{(n)} = \max_{1 \leq k \leq n} X_k$ – достаточная статистика для статистической модели случайной выборки фиксированного объема n из распределения с плотностью $f(\cdot|\cdot)$.

3. Пусть $X^{(n)} = (X_1, \dots, X_n)$, $n \geq 2$, – случайная выборка из нормального (θ, θ^2) распределения, в котором среднее совпадает со стандартным отклонением. Найдите минимальную достаточную статистику и докажите, что она не является полной.

- 4*. Докажите, что для статистической модели распределения Коши с плотностью

$$f(x|\theta) = \frac{1}{\pi [1 + (x - \theta)^2]}, \quad x \in \mathbb{R}, \quad \theta \in \mathbb{R},$$

минимальной достаточной статистикой является вариационный ряд $(X_{(1)}, \dots, X_{(n)})$. *Указание:* воспользуйтесь доказательством аналогичного утверждения для логистического распределения на стр. 47 книги Э.Лемана “Теория точечного оценивания”.

5. Пусть $X^{(n)} = (X_1, \dots, X_n)$, $n \geq 2$, – случайная выборка из гамма-распределения с параметром формы λ и параметром масштаба θ . Используя теорему Басу докажите, что статистики

$$T_1(X^{(n)}) = \sum_{k=1}^n X_k \quad \text{и} \quad T_2(X^{(n)}) = \sum_{k=1}^n \ln X_k - n \ln \left(\sum_{k=1}^n X_k \right)$$

независимы.

§ 3. Меры информации и нижние границы для среднего объема наблюдений в гарантийных процедурах

3.1. Меры информации. Для того, чтобы ввести понятие информации, которую поставляют данные статистического эксперимента, рассмотрим простейшую статистическую структуру, состоящую из двух распределений $\{P_\theta, \theta = \theta_1, \theta_2\}$ на выборочном пространстве (X, \mathcal{A}) и априорной вероятности $g_i = \mathbf{P}(\vartheta = \theta_i)$, $i = 1, 2$, $g_1 + g_2 = 1$. Пусть $p(x | \theta)$, $x \in X$, – функция плотности распределения P_θ , $\theta = \theta_1, \theta_2$, по мере μ .

До проведения наблюдений случайной выборки \mathbf{X} априорные шансы в пользу P_{θ_1} против P_{θ_2} естественно определить как отношение g_1 / g_2 , которое показывает насколько априори более вероятно, что выбор будет происходить именно из распределения P_{θ_1} . Пусть \mathbf{x} – результат наблюдения \mathbf{X} . В силу формулы Байеса апостериорная вероятность

$$\mathbf{P}\{\vartheta = \theta_i | \mathbf{X} = \mathbf{x}\} = \frac{g_i p(\mathbf{x} | \theta_i)}{g_1 p(\mathbf{x} | \theta_1) + g_2 p(\mathbf{x} | \theta_2)}, \quad i = 1, 2.$$

Апостериорные шансы в пользу P_{θ_1} против P_{θ_2} становятся равными

$$\frac{\mathbf{P}\{\vartheta = \theta_1 | \mathbf{X} = \mathbf{x}\}}{\mathbf{P}\{\vartheta = \theta_2 | \mathbf{X} = \mathbf{x}\}} = \frac{g_1}{g_2} \cdot \frac{p(\mathbf{x} | \theta_1)}{p(\mathbf{x} | \theta_2)},$$

то есть в $p(\mathbf{x} | \theta_1) / p(\mathbf{x} | \theta_2)$ раз больше (или меньше) априорных шансов g_1 / g_2 .

В теории информации логарифм отношения плотностей $\ln p(\mathbf{x} | \theta_1) - \ln p(\mathbf{x} | \theta_2)$ (логарифм отношения правдоподобия) обычно называют *информацией*, которую поставляет результат статистического эксперимента \mathbf{x} , в пользу $\vartheta = \theta_1$ против $\vartheta = \theta_2$. Среднее значение информации в пользу $\vartheta = \theta_1$ против $\vartheta = \theta_2$ равно

$$I(\theta_1, \theta_2 | \mathbf{X}) = \int p(x | \theta_1) \ln \frac{p(x | \theta_1)}{p(x | \theta_2)} d\mu(x).$$

Здесь и далее пределы интегрирования будут опускаться, если интегрирование ведется по всевозможным значениям X случайной выборки X . Интеграл полагается равным нулю на множестве

$$\{x : p(x | \theta_1) = 0\} \cup \{x : p(x | \theta_1) = p(x | \theta_2)\},$$

и информация полагается равной бесконечности, если мера μ множества $\{x : p(x | \theta_1) \neq 0, p(x | \theta_2) = 0\}$ отлична от нуля.

Определение 3.1. $\mathbf{I}(\theta_1, \theta_2 | \mathbf{X})$ называется *информацией по Кульбаку–Лейблеру* или *различающей информацией*.

В рамках общей статистической модели $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$, доминированной положительной сигма-конечной мерой μ , определим различающую (компоненты статистической модели) информацию как функцию на Θ^2 , равную

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}) = \int p(x | \theta) \ln \frac{p(x | \theta)}{p(x | \vartheta)} d\mu(x), \quad \theta, \vartheta \in \Theta.$$

(Чтобы избежать в дальнейшем отдельного рассмотрения тривиальных ситуаций, касающихся измерения информации, предположим, что носители всех распределений статистической модели одинаковы.)

Нашей ближайшей задачей будет исследование асимптотического поведения различающей информации $\mathbf{I}(\theta, \vartheta | \mathbf{X})$, когда $\vartheta \rightarrow \theta$, в предположении, что $\Theta \subset \mathbb{R}$. Такое исследование имеет целью ввести понятие не только различающей компоненты статистической модели информации, но также и информации, которое содержится в результатах статистического эксперимента относительно возможного значения параметра θ , то есть *точечной* информации. Для простоты изложения рассмотрим только случай $\Theta \in \mathbb{R}$ (θ – скалярный параметр).

Выберем некоторое фиксированное значение θ параметра и пусть $p(\cdot | \theta)$ – функция плотности случайной выборки \mathcal{X} . Введем семейство плотностей $\mathfrak{P}_\Delta = \{p(\cdot | \theta + \delta), \delta \in [-\Delta; \Delta]\}$ и потребуем, чтобы элементы этого семейства удовлетворяли следующим условиям регулярности.

(A1) Для почти всех $x [\mu]$ и при любом $\delta \in [-\Delta; \Delta]$ существуют частные производные $\partial^i p(x | \theta + \delta) / \partial \delta^i$, $i = 1, 2, 3$.

(A2) Для любого $\delta \in [-\Delta; \Delta]$ частные производные функции плотности удовлетворяют неравенствам $|\partial^i p(\mathbf{x} | \theta + \delta) / \partial \delta^i| \leq H_i(x)$, $i = 1, 2$, где функции H_1 и H_2 интегрируемы по мере μ ;

(A3) Для любого $\delta \in [-\Delta; \Delta]$ частная производная третьего порядка от логарифма функции плотности удовлетворяет неравенству $|\partial^3 \ln p(\mathbf{x} | \theta + \delta) / \partial \delta^3| \leq H(x)$, причем $\int p(x | \theta) H(x) d\mu(x) \leq M < \infty$.

Предложение 3.1. Если выполняются условия (A1)–(A3), то

$$\lim_{\Delta \rightarrow 0} \frac{2 \mathbf{I}(\theta, \theta + \Delta | \mathbf{X})}{\Delta^2} = \int \left[\frac{\partial \ln p(x | \theta)}{\partial \theta} \right]^2 p(x | \theta) d\mu(x). \quad (3.1)$$

Доказательство. Согласно формуле Тейлора имеем следующее представление подынтегрального логарифма:

$$\ln \frac{p(x | \theta + \Delta)}{p(x | \theta)} = \Delta \frac{p'(x | \theta)}{p(x | \theta)} + \frac{\Delta^2}{2} \left[\frac{p''(x | \theta)}{p(x | \theta)} - \left(\frac{p'(x | \theta)}{p(x | \theta)} \right)^2 \right] + \frac{\Delta^3}{3!} \frac{\partial^3 \ln p(x | \theta + \delta)}{\partial \delta^3} \Bigg|_{\delta=\lambda\Delta},$$

где $0 < \lambda < 1$, $p'(x | \theta)$ и $p''(x | \theta)$ – частные производные по δ от функции плотности $p(x | \theta + \delta)$ в точке $\delta = 0$.

В силу условий регулярности (A1) и (A2) интегралы от производных

$$\int \frac{p'(x | \theta)}{p(x | \theta)} p(x | \theta) d\mu(x) = \frac{d}{d\theta} \int p(x | \theta) d\mu(x) = 0,$$

$$\int \frac{p''(x | \theta)}{p(x | \theta)} p(x | \theta) d\mu(x) = \frac{d^2}{d\theta^2} \int p(x | \theta) d\mu(x) = 0,$$

а в силу условия (A3)

$$\Delta \int \frac{\partial^3 \ln p(x | \theta + \delta)}{\partial \delta^3} \Bigg|_{\delta=\lambda\Delta} p(x | \theta) d\mu(x) \rightarrow 0.$$

Таким образом, после интегрирования разложения логарифма, деления полученного выражения на Δ^2 и вычисления предела по $\Delta \rightarrow 0$ полученного выражения получаем утверждение предложения (3.1).

□

Определение 3.2. Интеграл

$$\mathbf{i}(\theta | \mathbf{X}) = \int \left[\frac{\partial \ln p(x | \theta)}{\partial \theta} \right]^2 p(x | \theta) d\mu(x), \quad \theta \in \Theta,$$

называется *информацией по Фишеру* или *точечной информацией*.

3.2. Свойства мер информации. Сформулируем и докажем несколько предложений, посвященных замечательным свойствам введенных мер информации (напомним, что носители всех распределений статистической модели полагаются одинаковыми).

Предложение 3.1. *Функции $\mathbf{I}(\theta, \vartheta | \mathbf{X}) \geq 0$ и $\mathbf{i}(\theta | \mathbf{X}) \geq 0$ при любых $\theta, \vartheta \in \Theta$. Различающая информация $\mathbf{I}(\theta, \vartheta | \mathbf{X}) = 0$ тогда и только тогда, когда $p(x | \theta) = p(x | \vartheta)$ для почти всех x по мере μ . Фишеровская информация $\mathbf{i}(\theta | \mathbf{X}) = 0$ только в том случае, когда существует такая окрестность $U(\theta)$ точки $\vartheta = \theta$, что $p(x | \vartheta)$ принимает одно и то же значение $C(x)$ при любом $\vartheta \in U(\theta)$ с возможным нарушением этого условия только при значениях x , имеющих в совокупности нулевую μ меру.*

Доказательство. Справедливость сформулированных утверждений, относящихся к фишеровской информации, очевидна.

Для различающей информации всё следует из неравенства Йенсена, примененного к среднему значению по распределению $P(\cdot | \vartheta)$ к выпуклой функции $\eta \ln \eta$, где $\eta = p(\mathbf{X} | \theta) / p(\mathbf{X} | \vartheta)$. Действительно,

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}) = \int \left[\frac{p(x | \theta)}{p(x | \vartheta)} \ln \frac{p(x | \theta)}{p(x | \vartheta)} \right] p(x | \vartheta) d\mu(x) =$$

$$\mathbf{E}_{\vartheta} \eta \ln \eta \geq \mathbf{E}_{\vartheta} \eta \ln \mathbf{E}_{\vartheta} \eta = 1 \ln 1 = 0,$$

и равенство в неравенстве Йенсена достигается тогда и только тогда, когда

$$\eta = p(\mathbf{X} | \theta) / p(\mathbf{X} | \vartheta) = \text{const.}$$

для почти всех x по мере μ . □

Предложение 3.2. *Пусть $T = T(\mathbf{X})$ – статистика на $(\mathcal{X}, \mathcal{A})$ с измеримым пространством значений $(\mathcal{T}, \mathcal{C})$. Различающая информация $\mathbf{I}(\theta, \vartheta | \mathbf{X}) \geq \mathbf{I}(\theta, \vartheta | T)$, причем знак равенства достигается тогда и только тогда, когда T – достаточная статистика для статистической модели \mathcal{P} . При выполнении условий (A1)–(A3) как для семейства $\mathfrak{P}_{\Delta} = \{p(\cdot | \theta + \delta), \delta \in [-\Delta; \Delta]\}$, так и для соответствующего семейства*

$\mathfrak{G}_\Delta = \{g(\cdot | \theta + \delta), \delta \in [-\Delta; \Delta]\}$ распределений статистики T , аналогичное утверждение справедливо для фишеровской информации: $\mathbf{i}(\vartheta | \mathbf{X}) \geq \mathbf{i}(\vartheta | T)$ при всех ϑ , принадлежащих некоторой окрестности $U(\theta)$ точки $\vartheta = \theta$.

Доказательство. Достаточно доказать справедливость предложения только для различающей информации, ибо для фишеровской информации аналогичное утверждение следует из Леммы 3.1. Действительно,

$$\mathbf{i}(\vartheta | \mathbf{X}) = \lim_{\Delta \rightarrow 0} \frac{2 \mathbf{I}(\theta, \theta + \Delta | \mathbf{X})}{\Delta^2} \geq \lim_{\Delta \rightarrow 0} \frac{2 \mathbf{I}(\theta, \theta + \Delta | T)}{\Delta^2} = \mathbf{i}(\vartheta | T).$$

Пусть $q(t)$, $t \in \mathcal{T}$, – функция плотности статистики T по некоторой мере γ на \mathcal{E} . Требуется показать, что

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}) - \mathbf{I}(\theta, \vartheta | T) = \int p(x | \theta) \ln \frac{p(x | \theta) q(T(x) | \vartheta)}{p(x | \vartheta) q(T(x) | \theta)} d\mu(x) \geq 0. \quad (3.4)$$

В обозначениях

$$\eta(x) = \frac{p(x | \theta) q(T(x) | \vartheta)}{p(x | \vartheta) q(T(x) | \theta)}, \quad h(x) = \frac{p(x | \vartheta) q(T(x) | \theta)}{q(T(x) | \vartheta)}$$

разность информации принимает вид

$$\int [\eta(x) \ln \eta(x)] h(x) d\mu(x).$$

Покажем, что $h(x)$, $x \in \mathcal{X}$, – функция плотности по мере μ . Очевидно, $h(x) \geq 0$, и остается только показать, что интеграл от h по мере μ равен единице. Представляя этот интеграл в терминах функции плотности q статистики T , получаем

$$\int h(x) d\mu(x) = \int_{\mathcal{T}} \frac{q(t | \theta)}{q(t | \vartheta)} \cdot q(t | \vartheta) d\gamma(t) = 1.$$

Итак, интеграл в (3.4) есть математическое ожидание $\mathbf{E} [\eta(\mathbf{X}) \ln \eta(\mathbf{X})]$ по распределению с плотностью h . Используя неравенство Йенсена для выпуклой функции $\eta \ln \eta$ и учитывая, что

$$\mathbf{E} \eta = \int \frac{p(x | \theta) q(T(x) | \vartheta)}{p(x | \vartheta) q(T(x) | \theta)} \cdot \frac{p(x | \vartheta) q(T(x) | \theta)}{q(T(x) | \vartheta)} d\mu(x) = 1,$$

получаем неравенство (3.4).

Знак равенства в неравенстве Йенсена достигается тогда и только тогда, когда случайная величина η равна константе, то есть

$$\frac{p(x|\theta)q(T(x)|\vartheta)}{p(x|\vartheta)q(T(x)|\theta)} = C(\theta, \vartheta)$$

почти всюду по мере μ . При каждом фиксированном значении ϑ это равенство есть факторизационное тождество Неймана для плотности p :

$$p(x|\theta) = C(\theta, \vartheta) \frac{q(T(x)|\theta)}{q(T(x)|\vartheta)} \cdot p(x|\vartheta), \quad \theta \in \Theta,$$

откуда следует, что T – достаточная статистика. \square

Следующее предложение устанавливает свойство аддитивности рассматриваемых мер информации. Для простоты изложения будет рассматриваться лишь случай одной наблюдаемой случайной величины ξ , когда статистический эксперимент управляется только правилом остановки φ_s и наблюдается случайная выборка $\mathbf{X}^{(\nu)} = (X_1, \dots, X_n)$ – вектор независимых копий ξ .

Предложение 3.3. Пусть $\mathbf{X} = \mathbf{X}^{(\nu)}$ – случайная выборка из распределения $P(\cdot|\theta)$ в статистическом эксперименте с моментом остановки ν среднее значение которого конечно: $\mathbf{E}_\theta \nu < \infty$. Тогда

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}^{(\nu)}) = \mathbf{E}_\theta \nu \cdot \mathbf{I}(\theta, \vartheta | \xi). \quad (3.5)$$

Если у любой плотности семейства $\mathfrak{P}_\Delta = \{p(\cdot|\theta + \delta), \delta \in [-\Delta; \Delta]\}$ существует частная производная по δ , которая по модулю равномерно ограничена для всех $\delta \in [-\Delta; \Delta]$ некоторой интегрируемой по мере μ функцией (см. аналогичные условия в (A1)–(A2)), то

$$\mathbf{i}(\theta, | \mathbf{X}^{(\nu)}) = \mathbf{E}_\theta \nu \cdot \mathbf{i}(\theta, | \xi). \quad (3.6)$$

Доказательство проведем только для различающей информации. Что касается информации по Фишеру, то ее свойство аддитивности устанавливается аналогичным образом и предлагается в качестве упражнения.

Для эксперимента с моментом остановки ν

$$p(\mathbf{X} | \theta) = \varphi_s(a_s | \mathbf{X}^{(\nu)}) \prod_{k=0}^{\nu-1} \varphi_s(a_c | \mathbf{X}^{(k)}) \cdot \prod_{k=0}^{\nu} f(X_k | \theta).$$

Следовательно, различающая информация

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}^{(\nu)}) = \mathbf{E}_\theta \ln \frac{p(\mathbf{X} | \theta)}{p(\mathbf{X} | \vartheta)} = \mathbf{E}_\theta \sum_{k=0}^{\nu} \ln \frac{f(X_k | \theta)}{f(X_k | \vartheta)}.$$

Определим последовательность индикаторных функций $\{I_n(\mathbf{x}^{(n-1)}), n = 1, 2, \dots\}$, полагая $I_n(\mathbf{x}^{(n-1)}) = 1$, если эксперимент не был остановлен до $(n-1)$ -го шага включительно с результатом $\mathbf{x}^{(n-1)} = (x_1, \dots, x_{n-1})$, и полагая $I_n(\mathbf{x}^{(n-1)}) = 0$ в противном случае. В терминах индикаторных функций χ различающая информация принимает вид

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}^{(\nu)}) = \sum_{n=1}^{\infty} \mathbf{E}_\theta I_n(\mathbf{x}^{(n-1)}) \ln \frac{f(X_n | \theta)}{f(X_n | \vartheta)}.$$

Поскольку индикатор $\mathbf{I}_n(\mathbf{x}^{(n-1)})$ зависит только от $n-1$ первых компонент выборки, то он не зависит от n -й компоненты X_n , и поэтому

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}^{(\nu)}) = \sum_{n=1}^{\infty} \mathbf{E}_\theta I_n(\mathbf{x}^{(n-1)}) \mathbf{E}_\theta \ln \frac{f(X_n | \theta)}{f(X_n | \vartheta)}.$$

Так как X_n имеет тоже распределение, что и ξ , то

$$\mathbf{E}_\theta \ln \frac{f(X_n | \theta)}{f(X_n | \vartheta)} = \mathbf{I}(\theta, \vartheta | \xi).$$

Кроме этого $\mathbf{E}_\theta I_n(\mathbf{x}^{(n-1)}) = \mathbf{P}_\theta(\nu \geq n)$, так что

$$\mathbf{I}(\theta, \vartheta | \mathbf{X}^{(\nu)}) = \mathbf{I}(\theta, \vartheta | \xi) \sum_{n=1}^{\infty} \mathbf{P}_\theta(\nu \geq n).$$

Наконец, в силу конечности среднего значения момента остановки

$$\sum_{n=1}^{\infty} \mathbf{P}_\theta(\nu \geq n) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbf{P}_\theta(\nu = k) = \sum_{k=0}^{\infty} k \mathbf{P}_\theta(\nu = k) = \mathbf{E}_\theta \nu$$

откуда следует равенство(3.5). □

3.3. Нижние границы для среднего объема выборки. Из доказанных предложений 3.1–3.3 вытекает ряд замечательных неравенств, устанавливающих нижнюю границу для среднего объема наблюдений любой гарантийной процедуры статистического вывода.

Пусть Φ – класс гарантийных (по θ , d или λ) стратегий $\varphi = (\varphi_s, \varphi_d)$ с решающими функциями $\delta_\varphi = \delta_\varphi(\mathbf{X})$.

Теорема 3.1. *Для любой гарантийной стратегии с моментом остановки ν при любом $\theta \in \Theta$ справедливо неравенство*

$$\mathbf{E}_\theta \nu \geq \inf_{\varphi \in \Phi} \sup_{\vartheta \in \Theta} \frac{\mathbf{I}(\theta, \vartheta | \delta_\varphi)}{\mathbf{I}(\theta, \vartheta | \xi)}. \quad (3.7)$$

Если $\Theta \subseteq \mathbb{R}$ и выполняются условия (A1)–(A3), то

$$\mathbf{E}_\theta \nu \geq \inf_{\varphi \in \Phi} \frac{\mathbf{i}(\theta | \delta_\varphi)}{\mathbf{i}(\theta | \xi)}, \quad (3.8)$$

и если $\delta_\varphi(\mathbf{X})$ есть несмещенная оценка параметрической функции $\gamma(\theta)$, то

$$\mathbf{E}_\theta \nu \geq \inf_{\varphi \in \Phi} \frac{[d\gamma(\theta)/d\theta]^2}{\mathbf{i}(\theta | \xi) \cdot \mathbf{D}_\theta \delta_\varphi}. \quad (3.9)$$

Для достижения нижних границ (3.7)–(3.8) необходимо, чтобы $(\nu, \delta_\varphi(\mathbf{X}^{(\nu)}))$ была достаточной статистикой для семейства распределений $\{P(\cdot | \theta), \theta \in \Theta\}$. Нижняя граница (3.9) достигается тогда и только тогда, когда существует такая параметрическая функция $C_\nu(\theta)$, что

$$\delta_\varphi(\mathbf{X}^{(\nu)}) - \gamma(\theta) = C_\nu(\theta) \sum_{k=1}^{\nu} \frac{\partial \ln f(X_k)}{\partial \theta}.$$

Неравенство (3.8) является ослаблением неравенства (3.7), а неравенство (3.9) есть ослабление неравенства (3.8).

Доказательство. Неравенство (3.7)–(3.8) следуют немедленно из свойств мер информации, доказанных в Предложениях 3.1–3.3. Ослабление (3.8) этого неравенства получают применением Леммы 3.1 к отношению информации в правой части (3.7). Наконец, неравенство (3.9) есть обобщение известного в теории оценивания неравенства Рао–Крамера на случай

оценки параметрической функции по данным последовательно планируемого эксперимента. Оно доказывается применением неравенства Шварца к следующим математическим ожиданиям:

$$[d\gamma(\theta)/d\theta]^2 = \left[\mathbf{E}_\theta \left(\delta_\varphi(\mathbf{X}^{(\nu)}) - \gamma(\theta) \right) \cdot \sum_{k=1}^{\nu} \frac{\partial \ln f(X_k)}{\partial \theta} \right]^2 \leq \\ \mathbf{E}_\theta \left(\delta_\varphi(\mathbf{X}^{(\nu)}) - \gamma(\theta) \right)^2 \cdot \mathbf{E}_\theta \left(\sum_{k=1}^{\nu} \frac{\partial \ln f(X_k)}{\partial \theta} \right)^2 = \mathbf{D}_\theta \delta_\varphi \cdot \mathbf{E}_\theta \nu \cdot \mathbf{i}(\theta | \xi).$$

□

Полученные в Теореме 3.1 границы для среднего объема выборки могут, в первую очередь, служить эталоном для исследования эффективности процедур статистического вывода с точки зрения стоимости проведения наблюдений в статистическом эксперименте. Если рассматривать риск конкретной процедуры как некоторое ограничение на риски всевозможных стратегий, дающих решение данной статистической проблемы, то отношение полученной при этих ограничениях одной из границ (3.7)–(3.9) к среднему объему выборки этой процедуры естественно принять за меру ее эффективности. По крайней мере именно так определяется эффективность процедур с точки зрения границ для риска.

Другое, более существенное в аспекте практического использования применение полученных границ состоит в их трактовке как некоторого критерия недостаточности объема проведенных наблюдений для построения гарантийной процедуры. Располагая собранным статистическим материалом, на основе которого статистик предполагает сделать гарантийный статистический вывод с заданными ограничениями на риск, он может сравнивать объем проведенных наблюдений с его возможной нижней границей и, таким образом, в случае неблагоприятного для него исхода искать процедуру с более слабыми ограничениями на риск. Естественно, построение гарантийных стратегий со средним объемом выборки, равным нижней границе, является одной из основных проблем теории статистического вывода.

Таким образом, решение задач на экстремум, связанных с вычислением нижних границ для среднего объема выборки, является первым шагом в

их практическом использовании. Сложность задач на экстремум, содержащихся в формулировке Теоремы 3.1, во многом определяется спецификой пространства решений \mathcal{D} . Приведем один из простейших примеров на решение данной задачи .

Гарантийное различие двух гипотез. Рассмотрим задачу гарантийного различия двух гипотез $H_0 : \theta \in \Theta_0$ и $H_1 : \theta \in \Theta_1$ по данным последовательно планируемого эксперимента с моментом остановки ν . Гарантийность статистического вывода о справедливости той или иной гипотезы понимается в смысле выполнения заданных ограничений α и β , $\alpha + \beta < 1$ на соответствующие вероятности ошибок первого и второго рода: вероятность отклонить нулевую гипотезу H_0 , когда она верна (выбор идет из распределения $P(\cdot | \theta)$ с $\theta \in \Theta_0$) должна не превосходить α , а вероятность принять нулевую гипотезу H_0 , когда справедлива альтернатива H_1 (выбор идет из распределения $P(\cdot | \theta)$ с $\theta \in \Theta_1$), должна не превосходить β . Естественно, построить гарантийный критерий $\varphi = (\varphi_s, \varphi_d)$ можно только в том случае, если параметрические пространства Θ_0 и Θ_1 не имеют общей границы, а разделены так называемой “областью безразличия”.

В терминах оперативной характеристики

$$\mathbf{P}_\theta (\delta_\varphi(\mathbf{X}) = d) = \psi(d | \theta), \quad d = (d_0, d_1), \quad \theta \in \Theta,$$

критерия φ его гарантийность означает выполнение следующих неравенств:

$$\psi(d_1 | \theta) \leq \alpha, \quad \text{если } \theta \in \Theta_0, \quad \psi(d_0 | \theta) \leq \beta, \quad \text{если } \theta \in \Theta_1.$$

Эти неравенства определяют класс Φ гарантийных критериев, по которому вычисляется infimum в правой части неравенства (3.7). Различающая информация, содержащаяся в решающей функции δ_φ критерия φ выражается через оперативную характеристику следующим образом:

$$\mathbf{I}(\theta, \vartheta | \delta_\varphi) = \mathbf{E}_\theta \ln \frac{\psi(\delta_\varphi(\mathbf{X} | \theta))}{\psi(\delta_\varphi(\mathbf{X} | \vartheta))} = \psi(d_0 | \theta) \ln \frac{\psi(d_0 | \theta)}{\psi(d_0 | \vartheta)} + \psi(d_1 | \theta) \ln \frac{\psi(d_1 | \theta)}{\psi(d_1 | \vartheta)}. \quad (3.10)$$

Пусть \mathbf{E}_θ оценивается снизу при $\theta \in \Theta_0$. Чтобы избежать раскрытия неопределенностей типа 0/0, будем вычислять supremum в правой части

(3.7) только по $\vartheta \in \Theta_1$, то есть ослабим неравенство (3.7). Тогда в терминах оперативной характеристики ослабленное неравенство (3.7) принимает следующий вид:

$$\mathbf{E}_\theta \nu \geq \inf_{\varphi \in \Phi} \sup_{\vartheta \in \Theta_1} \frac{\mathbf{I}(\theta, \vartheta | \delta_\varphi)}{\mathbf{I}(\theta, \vartheta | \xi)}, \quad \theta \in \Theta_0.$$

Для разнесенных значений θ и ϑ по соответствующим различаемым гипотезам подобластям Θ_0 и Θ_1 параметрического пространства Θ вероятности ошибки первого рода соответствуют значения оперативной характеристики $\psi(d_1 | \theta) = 1 - \psi(d_0 | \theta)$, $\theta \in \Theta_0$, а вероятности ошибки второго рода – значения $\psi(d_0 | \vartheta) = 1 - \psi(d_1 | \vartheta)$, $\vartheta \in \Theta_1$. Поскольку именно на эти функции накладываются ограничения α и β , представим различающую информацию (см. (3.10) в следующем виде:

$$\mathbf{I}(\theta, \vartheta | \delta_\varphi) = (1 - \psi(d_1 | \theta)) \ln \frac{1 - \psi(d_1 | \theta)}{\psi(d_0 | \vartheta)} + \psi(d_1 | \theta) \ln \frac{\psi(d_1 | \theta)}{1 - \psi(d_0 | \vartheta)}.$$

Правая часть этого равенства представляет значение функции Вальда

$$w(x, y) = x \ln \frac{x}{1 - y} + (1 - x) \ln \frac{1 - x}{y}, \quad x > 0, y > 0, \quad x + y < 1,$$

при $x = \psi(d_1 | \theta)$, $y = \psi(d_1 | \vartheta)$. Сделаем еще одно ослабление неравенства (3.7), переставив местами *infimum* и *supremum* и запишем $\mathbf{I}(\theta, \vartheta | \delta_\varphi)$ в терминах функции Вальда:

$$\mathbf{E}_\theta \nu \geq \sup_{\vartheta \in \Theta_1} \inf_{x \leq \alpha, y \leq \beta} \frac{w(x, y)}{\mathbf{I}(\theta, \vartheta | \xi)}, \quad \theta \in \Theta_0.$$

Дифференцированием функции Вальда по x при фиксированном y и по y при фиксированном x , нетрудно убедиться, что $w(x, y)$ есть убывающая функция в области $x > 0, y > 0, x + y < 1$. Следовательно, *infimum* достигается в точке $x = \alpha, y = \beta$ и нижняя граница для среднего объема наблюдений при справедливости нулевой гипотезы приобретает вид

$$\mathbf{E}_\theta \nu \geq \frac{w(\alpha, \beta)}{\inf_{\vartheta \in \Theta_1} \mathbf{I}(\theta, \vartheta | \xi)}, \quad \theta \in \Theta_0. \quad (3.11)$$

С помощью аналогичных рассуждений и ослаблений неравенства (3.7), нетрудно получить аналогичную оценку снизу для среднего объема

наблюдений при справедливости альтернативной гипотезы H_1 :

$$\mathbf{E}_\theta \nu \geq \frac{w(\beta, \alpha)}{\inf_{\vartheta \in \Theta_0} \mathbf{I}(\vartheta, \theta | \xi)}, \quad \theta \in \Theta_1. \quad (3.12)$$

В главе 12 будет показано, что для определенных семейств распределений наблюдаемой случайной величины ξ существует последовательный критерий, на котором достигаются границы (3.11)–(3.12).

ЗАЧЕТНЫЕ ЗАДАНИЯ

- 1.* Доказать Предложение 3.3 об аддитивности мер информации для информации по Фишеру $\mathbf{i}(\theta, | \mathbf{X}^{(\nu)})$.
2. Вычислить нижние границы (3.11)–(3.12) при различении гипотез $H_0 : \theta \leq \theta_0$ и $H_1 : \theta \geq \theta_1$, $\theta_0 < \theta_1$, о параметре θ нормального $N(\theta, \sigma^2)$ распределения при известном значении σ .
3. Вычислить нижние границы (3.11)–(3.12) при различении гипотез $H_0 : \theta \leq \theta_0$ и $H_1 : \theta \geq \theta_1$, $\theta_0 < \theta_1$, о параметре масштаба θ показательного распределения $E(\theta)$.

§ 4. Байесовские решения

4.1. Байесовское правило принятия решения. Мы приступаем к решению задачи построения правила принятия φ_d , минимизирующего риск процедуры $\varphi = (\varphi_\rho, \varphi_d)$ при фиксированном управлении $\varphi_\rho = (\varphi_s, \varphi_c)$ экспериментом: предполагается, что правила остановки φ_s и выбора φ_c заданы. Начнем с проблемы минимизации априорного риска, когда априорное распределение G определено полностью.

Правило принятия решения, на котором достигается минимум априорного риска, называется *байесовским*, а соответствующий ему априорный риск – *байесовским риском*. Естественно, могут возникать ситуации, когда нижняя грань априорного риска не достижима. В таком случае используется так называемое ε -байесовское правило, доставляющее минимум априорному риску приближенно, с заданной степенью точности.

В дальнейшем будет показано, как на практике построить аналог байесовского правила, оценивая неизвестное априорное распределение по имеющемуся архиву данных, состоящему из совокупности выборок, в которых значения параметров представляют последовательность реализаций независимых и одинаково распределенных по априорному распределению G случайных элементов.

Предполагая φ_ρ заданным, мы можем упростить обозначения. Вместо φ_d будем писать просто φ , поскольку задание процедуры сводится к определению только правила φ_d . Случайная выборка будет обозначаться буквой X , ее функция плотности записываться как $p(x|\theta)$, $x \in \mathcal{X}$, $\theta \in \Theta$. Пусть μ – доминирующая статистический эксперимент \mathcal{P} мера, а $g(\theta)$, $\theta \in \Theta$, – функция плотности априорного распределения G по мере χ .

Априорный риск $R(\varphi)$ процедуры, которую можно считать заданной, если задана решающая функция $\delta = \delta(X)$, определяется как математическое ожидание $\mathbf{E} L(\vartheta, \delta(X))$ от функции потерь по совместному распределению ϑ и X с плотностью $p(x|\theta)g(\theta)$ по мере $\mu \times \chi$. Интегральное представление априорного риска имеет вид

$$R(\varphi) = \int_{\Theta} \left[\int_{\mathcal{X}} \left[\int_{\mathcal{D}} L(\theta, a) \varphi(da|x) \right] p(x|\theta) d\mu(x) \right] g(\theta) d\chi(\theta).$$

Переставляя порядок интегрирования по θ и x , запишем априорный риск с помощью маргинального распределения случайной выборки X с плотностью

$$p_G(x) = \int_{\Theta} p(x|\theta) g(t) d\chi(t)$$

и апостериорного распределения ϑ с плотностью

$$h(\theta|X) = \frac{p(X|\theta)g(\theta)}{p_G(X)},$$

то есть представим его в виде

$$R(\varphi) = \int_{\mathcal{X}} \left[\int_{\Theta} \left[\int_{\mathcal{D}} L(\theta, a) \varphi(da|x) \right] h(\theta|x) d\chi(\theta) \right] p_G(x) d\mu(x).$$

Внутренний интеграл в этом представлении априорного риска

$$\mathfrak{R}(\varphi|X) = \int_{\Theta} \left[\int_{\mathcal{D}} L(\theta, a) \varphi(da|X) \right] h(\theta|X) d\chi(\theta)$$

называется *апостериорным риском* процедуры (правила) φ , и этот риск играет ключевую роль в построении байесовских правил принятия решения. Пусть φ_G – правило принятия решения, доставляющее наименьшее значение апостериорному риску, то есть

$$\mathfrak{R}(\varphi_G|X) = \inf_{\varphi} \mathfrak{R}(\varphi|X)$$

при естественном предположении, что нижняя грань в правой части достигается на некотором правиле принятия решения. Предполагая конечность всех интегралов в записи априорного риска, убедимся, что справедлива

Теорема 4.1. *Правило φ_G является байесовским, то есть доставляет наименьшее значение априорному риску $R(\varphi)$.*

Доказательство. Априорный риск есть среднее значение апостериорного риска:

$$R(\varphi) = \int_{\mathcal{X}} \mathfrak{R}(\varphi|x) p_G(x) d\mu(x).$$

Следующая очевидная цепочка неравенств доказывает данную теорему:

$$\inf_{\varphi} R(\varphi) \geq \int_{\mathcal{X}} \inf_{\varphi} \mathfrak{R}(\varphi|x) p_G(x) d\mu(x) =$$

$$\int_{\mathcal{X}} \mathfrak{R}(\varphi_G | x) p_G(x) d\mu(x) = R(\varphi_G) \geq \inf_{\varphi} R(\varphi). \quad \square$$

Таким образом, построение байесовского решения сводится к отысканию правила решения φ , доставляющего минимум апостериорному риску. В случае нерандомизированных правил данная проблема решается особенно просто, ибо состоит в отыскании минимума функции на пространстве решений \mathcal{D} .

Следует отметить одно из замечательных свойств байесовских правил – они не зависят от управления статистическим экспериментом. Это объясняется тем, что множители в функции плотности случайной выборки, которые содержат компоненты правил управления, зависят только от $x^{(n)}$, входят в числитель и знаменатель апостериорного распределения и выносятся за знак интеграла, то есть их можно просто сократить. Таким образом, компоненты $\varphi_g(\cdot | X^{(n)})$ байесовского правила принятия решения определяются из апостериорного риска для каждого шага $n = 1, 2, \dots$ статистического эксперимента и не зависят от φ_s и φ_c .

Обращаясь к теореме факторизации, легко видеть, что в случае существования достаточной статистики $T(X)$, апостериорное распределение (следовательно, и байесовское решение) зависит от X только через значение T .

ПРИМЕР 4.1. *Байесовский критерий различения двух гипотез.* Пусть X – случайная выборка из распределения с функцией плотности $f(x | \theta)$, и при наблюдении X параметр θ является реализацией случайного элемента ϑ с априорным распределением G . Статистическая проблема состоит в выборе одной из двух гипотез, касающихся значения θ : $H_0 : \theta \in \Theta_0$ и $H_1 : \theta \in \Theta_1 = \Theta^c$, так что пространство решений состоит из двух точек d_0 и d_1 , соответствующих проверяемым гипотезам. В данной проблеме функция потерь обычно выбирается в виде

$$L(\theta, d_i) = \begin{cases} 1, & \text{если } \theta \in \Theta_{1-i}, \\ 0, & \text{если } \theta \in \Theta_i, \end{cases} \quad i = 0, 1,$$

и называется функцией потерь типа 1–0.

Правило принятия решения φ определяется вероятностью принятия решения d_0 : $\varphi(d_0 | X) = 1 - \varphi(d_1 | X)$. При каждом фиксированном результате x наблюдения X средние потери равны

$$\begin{aligned} \int_{\mathcal{D}} L(\theta, a) \varphi(da | X) &= L(\theta, d_0) \varphi(d_0 | X) + L(\theta, d_1) \varphi(d_1 | X) = \\ &= \begin{cases} \varphi(d_0 | X), & \text{если } \theta \in \Theta_1, \\ 1 - \varphi(d_0 | X), & \text{если } \theta \in \Theta_0, \end{cases} \end{aligned}$$

так что апостериорный риск

$$\begin{aligned} \mathfrak{R}(\varphi | X) &= \int_{\Theta} \left[\int_{\mathcal{D}} L(\theta, a) \varphi(da | X) \right] h(\theta | X) d\chi(\theta) = \\ &= \varphi(d_0 | X) \mathbf{P}(\vartheta \in \Theta_1 | X) + [1 - \varphi(d_0 | X)] \mathbf{P}(\vartheta \in \Theta_0 | X). \end{aligned}$$

Апостериорный риск есть линейная функция от $\varphi(d_0 | X)$, и его минимум достигается при $\varphi(d_0 | X) = 1$, если апостериорная вероятность $\mathbf{P}(\vartheta \in \Theta_0 | X) > \mathbf{P}(\vartheta \in \Theta_1 | X)$, и при $\varphi(d_0 | X) = 0$, когда $\mathbf{P}(\vartheta \in \Theta_0 | X) < \mathbf{P}(\vartheta \in \Theta_1 | X)$. В случае равенства апостериорных вероятностей различаемых гипотез правило принятия решения определяется произвольным образом – можно принимать любую из гипотез с вероятностью единица или использовать некоторое рандомизированное правило.

Таким образом, байесовское правило φ_G принимает гипотезу, имеющую наибольшее значение апостериорной вероятности ее справедливости. Байесовский риск равен минимуму из априорных вероятностей различаемых гипотез:

$$R(\varphi_G) = \min \{ G(\Theta_0), G(\Theta_1) \}.$$

Нетрудно видеть, что результаты этого примера непосредственно переносятся на случай различения произвольного числа $m \geq 2$ гипотез: байесовское правило принимает гипотезу, которая имеет наибольшую апостериорную вероятность ее справедливости.

ПРИМЕР 4.2. *Оценка значения параметра при квадратичной функции потерь.* По результатам наблюдения случайной выборки X из распределения с плотностью $f(x | \theta)$, $x \in \mathcal{X}$, $\theta \in \Theta \subseteq \mathbb{R}$, оценивается скалярный параметр θ при квадратичной функции потерь $L(\theta, d) = (\theta - d)^2$.

Как и в предыдущем примере, значение θ при наблюдении X есть реализация случайной величины ϑ с известным априорным распределением G . В этой статистической проблеме пространство решений \mathcal{D} совпадает с параметрическим пространством Θ ; любое правило принятия решения $\varphi(D|X)$, $D \in \mathcal{B}$, является переходной вероятностью с выборочного пространства X на борелевскую σ -алгебру \mathcal{B} , порожденную интервалами на прямой \mathbb{R} .

При каждом фиксированном результате x наблюдения X средние потери равны

$$\int_{\mathbb{R}} (\theta - a)^2 \varphi(da | X) = \mathbf{E}\{(\theta - \delta(X))^2 | X\},$$

где $\delta = \delta(X)$ – рандомизированная оценка параметра θ (решающая функция). В силу неравенства Йенсена для выпуклой по d функции $(\theta - d)^2$ для условного математического ожидания справедливо неравенство

$$\mathbf{E}\{(\theta - \delta(X))^2 | X\} \geq \left(\theta - \int_{\mathbb{R}} a \varphi(da | X) \right)^2,$$

так что любому рандомизированному правилу φ соответствует нерадомизированная оценка

$$\hat{\theta}(X) = \int_{\mathbb{R}} a \varphi(da | X),$$

риск которой меньше (или равен) рандомизированной оценке δ .

Итак, при построении байесовской оценки $\hat{\theta}_G$ можно ограничиться рассмотрением только нерандомизированных оценок, определяя $\hat{\theta}_G$ как точку достижения по аргументу d минимума апостериорного риска

$$\int_{\mathbb{R}} (t - d)^2 h(t | X) dt.$$

Используя снова неравенство Йенсена, получаем, что этот интеграл достигает своего минимума в точке

$$\hat{\theta}_G(X) = \int_{\mathbb{R}} t h(t | X) dt.$$

Таким образом, байесовская оценка $\hat{\theta}_G = \hat{\theta}_G(X)$ равна апостериорному среднему значению, а байесовский риск – априорной дисперсии этой

оценки:

$$R(\hat{\theta}_G) = \int_{\Theta} \int_{\mathbb{R}} (\theta - \hat{\theta}_G(x))^2 p(x | \theta) d\mu(x) dG(\theta).$$

В практических приложениях довольно часто встречаются следующие вероятностные модели $(\mathcal{F}, \mathcal{G}, \cdot)$, которые поэтому целесообразно использовать для конкретизации статистических правил, полученных в примерах 4.1–4.2.

Модель N-N. В статистическом эксперименте наблюдается последовательность независимых копий случайной величины ξ , имеющей нормальное (θ, σ^2) распределение; значение дисперсии σ^2 известно, а неизвестное значение среднего θ составляет предмет статистического вывода. Таким образом, семейство возможных распределений ξ определяется как $\mathcal{F} = \{F(\cdot | \theta), \theta \in \Theta = \mathbb{R}\}$, где

$$F(A | \theta) = \frac{1}{\sqrt{2\pi}\sigma} \int_A \exp\left\{-\frac{(x - \theta)^2}{2\sigma^2}\right\} dx,$$

для любого борелевского множества A на действительной прямой \mathbb{R} .

Пусть изучение природы исследуемого объекта привело к заключению, что значение θ при наблюдении случайной выборки $X = (X_1, \dots, X_\nu)$ является реализацией случайной величины ϑ , распределенной также по нормальному закону с параметрами (μ, τ^2) . Предположим, что значения параметров μ и τ статистике известны, так что вторая составляющая вероятностной модели состоит из единственного нормального распределения

$$G(B) = \frac{1}{\sqrt{2\pi}\tau} \int_B \exp\left\{-\frac{(\theta - \mu)^2}{2\tau^2}\right\} d\theta, \quad B \in \mathcal{B}.$$

Как было замечено выше, байесовское решение не зависит от управления статистическим экспериментом, поэтому можно ограничиться рассмотрением только выборки фиксированного объема n . В таком случае семейство распределений случайной выборки \mathcal{P} редуцируется к семейству распределений достаточной статистики \bar{X} (выборочному среднему), нормальное распределение которой имеет функцию плотности

$$p(x | \theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{n(x - \theta)^2}{2\sigma^2}\right\}.$$

Функция плотности маргинального распределения \bar{X} вычисляется путем интегрирования по θ совместной функции плотности \bar{X} и θ :

$$p_G(x) = \int_{-\infty}^{+\infty} p(x|\theta) g(\theta) dt =$$

$$\int_{-\infty}^{+\infty} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{n(x-\theta)^2}{2\sigma^2}\right\} \cdot \frac{1}{\sqrt{2\pi}\tau} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\} d\theta =$$

$$\frac{1}{\sqrt{2\pi}(\sigma^2/n + \tau^2)} \exp\left\{-\frac{(x-\mu)^2}{2(\sigma^2/n + \tau^2)}\right\}.$$

Заметим, что интеграл вычислять не надо, поскольку он представляет свертку нормального $(0, \sigma^2/n)$ и нормального (μ, τ^2) распределений, то есть функцию плотности суммы двух независимых случайных величин, имеющих нормальное распределение, для которого справедлива теорема сложения.

Функция плотности апостериорного распределения находится приведением квадратичной формы, стоящей под экспонентой (см. ниже) к виду $(\theta - \mathcal{M})^2/2s^2$:

$$h(\theta|\bar{X}) = p(\bar{X}|\theta) g(\theta)/p_G(\bar{X}) =$$

$$\frac{\sqrt{\sigma^2/n + \tau^2}}{\sqrt{2\pi}\sigma\tau} \exp\left\{-\frac{n(\bar{X}-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2} + \frac{(\bar{X}-\mu)^2}{2(\sigma^2/n + \tau^2)}\right\} =$$

$$\frac{1}{\sqrt{2\pi}s} \exp\left\{-\frac{(\theta-\mathcal{M})^2}{2s^2}\right\},$$

где апостериорные среднее и дисперсия равны соответственно

$$\mathcal{M} = \frac{n\bar{X}\sigma^{-2} + \mu\tau^{-2}}{n\sigma^{-2} + \tau^{-2}} \quad \text{и} \quad s^2 = \frac{1}{n\sigma^{-2} + \tau^{-2}}.$$

Таким образом, апостериорное распределение ϑ есть нормальное распределение с параметрами \mathcal{M} и s^2 , причем апостериорное среднее \mathcal{M} есть выпуклая комбинация выборочного среднего \bar{X} и априорного среднего μ , а апостериорная дисперсия s^2 не зависит от выборочных данных.

Обратимся к примеру 4.1 и определим параметрические подмножества, соответствующие различаемым гипотезам, как $\Theta_0 = (-\infty, \theta_0]$, $\Theta_1 =$

$(\theta_0, +\infty)$. В таком случае апостериорные вероятности справедливости гипотез равны

$$\mathbf{P}(\theta \leq \theta_0 | \bar{X}) = \Phi\left(\frac{\theta_0 - \mathcal{M}}{s}\right) \quad \text{и} \quad \mathbf{P}(\theta > \theta_0 | \bar{X}) = 1 - \Phi\left(\frac{\theta_0 - \mathcal{M}}{s}\right).$$

Байесовское правило принимает гипотезу $H_0 : \theta \leq \theta_0$, если

$$\Phi\left(\frac{\theta_0 - \mathcal{M}}{s}\right) \leq \frac{1}{2},$$

то есть когда апостериорное среднее $\mathcal{M} \leq \theta_0$, и что может быть справедливее, чем это правило!

Что же касается примера 4.2, где рассматривается проблема оценки скалярного параметра при квадратичной функции потерь, то здесь байесовское правило определяется статистикой (оценкой), равной апостериорному среднему. Следовательно,

$$\hat{\theta}_G(\bar{X}) = \mathcal{M} = \frac{n\bar{X}\sigma^{-2} + \mu\tau^{-2}}{n\sigma^{-2} + \tau^{-2}}.$$

Модель В-В. Пусть случайный выбор осуществляется в рамках схемы испытаний Бернулли с вероятностью успешного испытания, равной θ . При фиксированном объеме испытаний n достаточной статистикой является $T = \sum_1^n X_k$, реализующая общее количество успехов в проведенных испытаниях. Следовательно, статистический эксперимент \mathcal{P} редуцируется к семейству биномиальных распределений с функциями плотности по считающей мере

$$p(t|\theta) = C_n^t \theta^t (1-\theta)^{n-t}, \quad t = 0, 1, \dots, n, \quad \theta \in (0, 1).$$

В том случае, когда вероятность успеха θ в наблюдаемых исходах схемы Бернулли является реализацией случайной величины ϑ , на практике обычно используется априорное бета-распределение с функцией плотности

$$g(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}, \quad \theta \in (0, 1), \quad \alpha > 0, \beta > 0.$$

Такой выбор априорного распределения диктуется, в основном, не столько природой исследуемого объекта, сколько исключительным разнообразием

форм функции плотности бета-распределения. Она может быть симметричной, скошенной влево, скошенной вправо, U-образной, иметь отрицательный или положительный эксцесс.

Не вычисляя маргинальное распределение статистики T , докажите, что апостериорное распределение ϑ есть снова бета-распределение с параметрами $\alpha + T$ и $n + \beta - T$, а байесовская оценка значения θ при квадратичных потерях

$$\hat{\theta}_G = \frac{\alpha + T}{n + \alpha + \beta}.$$

Модель P-G. В этой модели выбор происходит из распределения Пуассона с функцией плотности по считающей мере

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, \dots, \quad \theta \in \Theta = \mathbf{R}_+.$$

Для семейства распределений случайной выборки из распределения Пуассона существует достаточная статистика $T = \sum_1^n X_k$, имеющая также распределение Пуассона с параметром $n\theta$. Следовательно, статистический эксперимент определяется функцией плотности по считающей мере

$$p(t|\theta) = \frac{n^t \theta^t e^{-n\theta}}{t!}, \quad t = 0, 1, \dots, \quad \theta \in \Theta = \mathbf{R}_+. \quad (4.1)$$

Предположим, что исследование физической природы параметра интенсивности θ распределения Пуассона показало, что значение θ есть реализация случайной величины ϑ , имеющей гамма-распределение с функцией плотности

$$g(\theta; \lambda, a) = \frac{a^\lambda}{\Gamma(\lambda)} \theta^{\lambda-1} e^{-a\theta}, \quad \theta > 0, \quad \lambda > 0, \quad a > 0.$$

Не вычисляя маргинальное распределение статистики T , докажите, что апостериорное распределение ϑ есть снова гамма-распределение с параметрами $\lambda + T$ и $n + a$, а байесовская оценка значения θ при квадратичных потерях

$$\hat{\theta}_G = \frac{\lambda + T}{n + a}.$$

4.2. Эмпирический байесовский подход к принятию решения. Практическое применение байесовского подхода к принятию решения связано, как правило, со статистическим анализом последовательности идентичных объектов, распределения наблюдаемой характеристики которых (случайного элемента ξ с функцией плотности $f(x|\theta)$) имеют одну и ту же форму и отличаются только значениями параметра θ . Эти неизвестные статистику значения и представляют предмет статистического решения. Пусть было обследовано N объектов, и принятие решения относительно значения θ_k параметра k -го объекта производилось по выборочным данным $x_{k,1}, \dots, x_{k,n}$ – результатам наблюдений независимых копий ξ , где $k = 1, \dots, N$. Таким образом, статистик располагает *архивом* данных

$$\mathbf{x}^{[N]} = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \cdot & \dots & \cdot \\ x_{N1} & \dots & x_{Nn} \end{bmatrix}$$

Строка $\mathbf{x}_k = (x_{k1}, \dots, x_{kn})$ архива $\mathbf{x}^{[N]}$ с номером $k (= 1, \dots, N)$ составляет результаты наблюдения случайного вектора $\mathbf{X}_k = (X_{k1}, \dots, X_{kn})$, функция плотности которого

$$p(\mathbf{x}|\theta_k) = \prod_{i=1}^n f(x_i|\theta_k), \quad k = 1, \dots, N.$$

Случайные векторы $\mathbf{X}_1, \dots, \mathbf{X}_N$ независимы в совокупности, а вектор $(\theta_1, \dots, \theta_N)$ представляет реализацию случайного вектора $(\vartheta_1, \dots, \vartheta_N)$ с независимыми одинаково распределенными компонентами – случайную выборку из априорного распределения G . Следовательно, $p(\mathbf{x}|\vartheta_k)$ есть плотность условного распределения вектора \mathbf{X}_k относительно случайного элемента ϑ_k , и маргинальное распределение этого вектора определяется функцией плотности $p_G(\mathbf{x})$ при любом $k = 1, \dots, N$. Таким образом, “случайный аналог”

$$\mathbf{X}^{[N]} = \begin{bmatrix} X_{11} & \dots & X_{1n} \\ \cdot & \dots & \cdot \\ X_{N1} & \dots & X_{Nn} \end{bmatrix}$$

архива $\mathbf{x}^{[N]}$ имеет распределение с функцией плотности

$$p_{G,n}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{k=1}^N p_G(\mathbf{x}_k).$$

Если $\mathcal{G} = \{G_\lambda, \lambda \in \Lambda\}$ есть параметрическое семейство априорных распределений, то значение параметра λ можно оценить, например, используя метод максимального правдоподобия, основанный на функции плотности $p_{G_\lambda,n}(\mathbf{x}^{[N]})$, в которой вместо аргумента подставляются архивные данные. Проиллюстрируем этот метод оценки λ на примере модели N-N.

ПРИМЕР 4.3. *Эмпирический байесовский вывод в модели N-N.* Рассматривается ситуация, когда значения параметров μ и τ априорного распределения не известны. Что касается значения дисперсии σ^2 , которое не изменяется от выборки к выборке, то ее значение, как правило, также не известно, и проблема его оценки включается в общую задачу оценки параметров априорного распределения, если трактовать σ^2 как значение случайной величины с вырожденным (сосредоточенным в одной точке) априорным распределением.

Поскольку статистический эксперимент нормального распределения обладает достаточной статистикой (\bar{X}, S^2) , то архив данных редуцируется к вектору $\mathbf{X}^{[N]} = (\bar{X}_1, S_1^2), \dots, (\bar{X}_N, S_N^2)$ с двумерными компонентами, состоящими из выборочного среднего \bar{X} и выборочной дисперсии S^2 . Отметим одну немаловажную деталь при редукции архива к достаточным статистикам, которая обсуждалась также в общем курсе математической статистики. Следует использовать несмещенные оценки

$$S_k^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ki} - \bar{X}_k)^2, \quad k = 1, \dots, N,$$

параметра σ^2 по каждой выборке архива, ибо, в противном случае, общая оценка σ^2 будет несостоятельной и при малых объемах n (например, $n = 2$) каждой выборки будет выдавать заниженное значение дисперсии.

Как утверждает известная теорема Фишера, статистики \bar{X} и S^2 независимы, \bar{X} имеет нормальное распределение с параметрами $(\theta, \sigma^2/n)$, а

$(n-1)S^2/\sigma^2$ распределена по закону хи-квадрат с $(n-1)$ -й степенью свободы. В примере 4.1 было установлено, что маргинальное распределение \bar{X} также нормально с параметрами μ и $\theta + \sigma^2/n$; распределение S^2 остается прежним. Следовательно, функция правдоподобия архива $\mathbf{x}^{[N]}$ равна

$$L\left(\mu, \tau^2, \sigma^2 \mid \mathbf{x}^{[N]}\right) = \frac{1}{(2\pi(\tau^2 + \sigma^2/n))^{N/2}} \exp\left\{-\frac{1}{2(\tau^2 + \sigma^2/n)} \sum_{k=1}^N (\bar{X}_k - \mu)^2\right\} \times \frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2} \sigma^{n-1} \Gamma((n-1)/2)} \prod_{k=1}^N S_k^{n-3} \exp\left\{-\frac{n-1}{\sigma^2} \sum_{k=1}^N S_k^2\right\}.$$

Точка $(\hat{\mu}, \hat{\tau}^2, \hat{\sigma}^2)$ (оценка максимального правдоподобия трехмерного параметра) достижения максимума этой функции осуществляется стандартным методом математического анализа нахождения экстремума функции многих переменных. Опуская тривиальные алгебраические выкладки, приведем окончательные формулы для вычисления оценок параметров априорного распределения:

$$\hat{\mu} = \bar{\mathbf{X}} = \frac{1}{N} \sum_{k=1}^n \bar{X}_k, \quad \hat{\tau}^2 = \frac{1}{N} \sum_{k=1}^n [(\bar{X}_k - \bar{\mathbf{X}})^2 - S_k^2], \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^n S_k^2.$$

Заканчивая этот пример, отметим, что для моделей В-В и Р-Г оценки максимального правдоподобия, к сожалению, не находятся в явном виде. Можно, конечно, получить явные оценки по методу моментов или воспользоваться вычислительными методами нахождения экстремума функции многих переменных.

Рассмотрим теперь случай, когда априорное распределение полностью не известно. Здесь построение эмпирического аналога байесовской процедуры возможно далеко не всегда. Г. Роббинс в 1955 году впервые построил эмпирическую байесовскую оценку для параметра пуассоновского распределения при квадратичной функции потерь. В последующем этот пример был распространен на задачу оценки в рамках однопараметрического экспоненциального семейства при квадратичных потерях и проверки односторонних гипотез при специальной (линейной по параметру) функции

потерь. Решение было получено благодаря тому, что апостериорный риск можно было выразить в терминах маргинальной функции плотности достаточной статистики, которая затем оценивалась по архиву данных. Мы рассмотрим два примера на построение непараметрического аналога байесовской оценки для дискретного (процедура Роббинса) и непрерывного распределений.

ПРИМЕР 4.4. *Эмпирическая байесовская оценка параметра распределения Пуассона при неизвестном априорном распределении.* Рассматривается вероятностная модель P-G, где выбор происходит из распределения Пуассона, а статистическая проблема состоит в байесовской оценке параметра интенсивности θ при квадратичной функции потерь.

Как было замечено выше, для семейства распределений случайной выборки из распределения Пуассона существует достаточная статистика $T = \sum_1^n X_k$, имеющая также распределение Пуассона с параметром $n\theta$; функция плотности $p(t|\theta)$ статистики T определялась формулой (4.1). Пусть G – априорное распределение ϑ ,

$$p_G(t) = \int_0^\infty p(t|\theta) dG(\theta) -$$

функция плотности маргинального распределения статистики T .

При квадратичной функции потерь байесовская оценка равна апостериорному среднему, которое для данной модели можно представить в следующем виде:

$$\begin{aligned} \hat{\theta}_G &= n^T \int_0^\infty \theta \cdot \theta^T e^{-n\theta} dG(\theta) / [T! \cdot p_G(T)] = \\ &= \frac{(T+1)}{n} \cdot n^{T+1} \int_0^\infty \theta^{T+1} e^{-n\theta} dG(\theta) / [(T+1)! \cdot p_G(T)] = \\ &= \frac{(T+1)}{n} \cdot \frac{p_G(T+1)}{p_G(T)}. \end{aligned}$$

Пусть t – значение статистики T в текущем эксперименте, по наблюдениям в котором оценивается параметр θ . Чтобы вычислить значение байесовской оценки $\hat{\theta}_G$, найдем оценки вероятностей $p_G(t)$ и $p_G(t+1)$ по архиву t_1, \dots, t_N предыдущих текущему эксперименту значений статистики T . Пусть N_t, N_{t+1} – количества значений статистики T в архиве,

которые равны t и $t + 1$ соответственно. Поскольку $p_G(t) = P(T = t)$, то имеем частотные оценки

$$\hat{p}_{G,N}(t) = \frac{N_t}{N}, \quad \hat{p}_{G,N}(t + 1) = \frac{N_{t+1}}{N},$$

которые определяют значение эмпирической байесовской оценки

$$\hat{\theta}_{G,N} = \frac{(T + 1)}{n} \cdot \frac{N_{t+1}}{N_t}.$$

Небольшое замечание: для того, чтобы избежать “деления на нуль”, если в архиве данных не найдется значение статистики T , равное t , то достаточно включить это значение в архив, объем которого станет равным $N + 1$.

ПРИМЕР 4.5. *Эмпирическая байесовская оценка среднего значения нормального распределения при неизвестном априорном распределении.* Обратимся к проблеме, рассмотренной в примере 4.3, когда априорное распределение G среднего значения ϑ наблюдаемой случайной величины, имеющей нормальное (θ, σ^2) распределение, полностью не известно. Как было показано в примере 4.3, оценка неизвестного, общего для всех экспериментов значения дисперсии σ^2 достаточно просто оценивается по архиву данных вне зависимости от априорного распределения ϑ .

Байесовская оценка θ при квадратичной функции потерь есть апостериорное среднее, которое в данном случае имеет вид

$$\hat{\theta}_G(\bar{X}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \theta \exp \left\{ -\frac{n(\bar{X} - \theta)^2}{2\sigma^2} \right\} dG(\theta) / p_G(\bar{X}),$$

где маргинальная плотность выборочного среднего \bar{X} вычисляется по формуле

$$p_G(x) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{n(x - \theta)^2}{2\sigma^2} \right\} dG(\theta).$$

Поскольку производная маргинальной плотности

$$\begin{aligned} p'_G(x) &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \frac{n(\theta - x)}{\sigma^2} \exp \left\{ -\frac{n(x - \theta)^2}{2\sigma^2} \right\} dG(\theta) = \\ &= \frac{n}{\sigma^2} p_G(x) \left[\hat{\theta}_G(x) - x \right], \end{aligned}$$

то значение байесовской оценки, когда результат вычисления выборочного среднего \bar{X} в текущем эксперименте равен x , приобретает вид

$$\hat{\theta}_G(x) = \frac{\sigma^2}{n} \frac{d}{dx} \ln p_G(x) + x.$$

Таким образом, проблема построения эмпирической байесовской оценки θ сводится к достаточно гладкой оценке функции плотности $p_G(x)$ маргинального распределения \bar{X} по результатам x_1, \dots, x_N наблюдений этой статистики в предыдущих N экспериментах, то есть по архиву $\bar{X}_1, \dots, \bar{X}_N$. Обычно используется так называемая ядерная оценка плотности

$$\hat{p}_{G,N}(x) = \frac{1}{N h} \sum_{k=1}^N K\left(\frac{x - \bar{X}_k}{h}\right),$$

где h специально подбираемое число, а $K(\cdot)$ – ядро оценки, которое также требует специальной конкретизации. Оценка функции плотности составляет отдельную, достаточно сложную задачу, которой посвящаются большие монографии.

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. В рамках модели В-В, не вычисляя маргинальное распределение достаточной статистики T , докажите, что апостериорное распределение ϑ есть бета-распределение с параметрами $\alpha + T$ и $n + \beta - T$. Найдите байесовскую оценку значения θ при квадратичных потерях.
2. В рамках модели Р-Г, не вычисляя маргинальное распределение достаточной статистики T , докажите, что апостериорное распределение ϑ есть гамма-распределение с параметрами $\lambda + T$ и $n + a$. Найдите байесовскую оценку значения θ при квадратичных потерях.
3. Для модели Р-Г найдите байесовскую оценку параметра θ при функции потерь

$$L(\theta, d) = \begin{cases} 1, & \text{если } |\theta - d|/\theta > \Delta, \\ 0, & \text{если } |\theta - d|/\theta \leq \Delta. \end{cases}$$

4. Пусть X имеет геометрическое распределение:

$$\mathbf{P}(X = x) = (1 - \theta)^{x-1}\theta, \quad x = 1, 2, \dots, \quad \theta \in (0, 1).$$

При наблюдении X значение θ есть реализация случайной величины ϑ , имеющей бета распределение с параметрами α и β . Найдите байесовскую оценку θ при квадратичных потерях.

5*. Пусть $X^{(n)} = (X_1, \dots, X_n)$ – случайная выборка из показательного распределения с плотностью

$$f(x | \theta) = \theta \exp\{-\theta x\}, \quad x \geq 0, \quad \theta > 0.$$

В предположении, что при наблюдении $X^{(n)}$ значение θ есть реализация случайной величины ϑ , априорное распределение G которой полностью не известно, предложите метод построения эмпирической байесовской оценки θ , аналогичный методу примера 4.5.

§ 5. Минимаксные решения

Методы минимизации априорного риска, которые были представлены в предыдущем параграфе, служат основным инструментом в решении довольно большого круга задач на построение оптимальных процедур статистического вывода. Во многом это объясняется тем, что байесовские решения и их “слабые” пределы, (значения параметров априорного распределения устремляются к определенным пределам, чаще всего, к бесконечности) образуют класс так называемых *допустимых* решений: не существует правил принятия решения, функция риска которых равномерно лучше байесовского. Конечно, это утверждение, доказанное А.Вальдом в первой половине прошлого века, справедливо при определенных условиях на параметрическое пространство, пространство решений и функцию потерь, но эти условия практически всегда выполнимы в конкретных статистических исследованиях. В этом параграфе будут разрабатываться методы построения правил принятия решения, которые минимизируют наибольшее значение функции риска. Специальный выбор априорного распределения, не связанный со спецификацией вероятностной модели, является основой для решения таких задач.

Рассмотрим общую статистическую проблему с пространством решений \mathcal{D} , функцией потерь $L(\theta, d)$ и семейством $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$, возможных распределений случайной выборки X для заданного управления $\rho = (\varphi_s, \varphi_c)$ статистическим экспериментом. Пусть $R(\varphi | \theta) = \mathbf{E}_\theta L(\theta, \delta(X))$, $\theta \in \Theta$, – функция риска процедуры φ , и $\delta = \delta(X)$ – соответствующая правилу принятия решения φ_d решающая функция. Как и в предыдущем параграфе, управление ρ будет считаться фиксированным (заданным), а правило принятия решения φ_d будет определяться как решение некоторой задачи на экстремум. В связи с этим выбор правила φ_d отождествляется с заданием всей процедуры $\varphi = (\rho, \varphi_d)$ статистического выбора, так что индекс d в записи правила φ_d будет опускаться.

Определение 5.1. Правило принятия решения φ^* называется *ми-*

нимаксным, если

$$\sup_{\theta \in \Theta} R(\varphi^* | \theta) \leq \sup_{\theta \in \Theta} R(\varphi | \theta),$$

каково бы ни было правило φ .

Таким образом, на минимаксном правиле φ^* достигается наименьшее значение максимума функции риска.

Предложение 5.1. Пусть \mathcal{G} – класс всевозможных априорных распределений на измеримом параметрическом пространстве (Θ, \mathcal{B}) , включающий класс \mathcal{H} , вырожденных в каждой точке θ параметрического пространства Θ априорных распределений. Тогда

$$\sup_{G \in \mathcal{G}} R_G(\varphi) = \sup_{\theta \in \Theta} R(\varphi | \theta), \quad (5.1)$$

где $R_G(\varphi)$ – априорный риск правила φ .

Доказательство. Поскольку

$$R_G(\varphi) = \int_{\Theta} R(\varphi | \theta) dG(\theta) \leq \sup_{\theta \in \Theta} R(\varphi | \theta),$$

то

$$\sup_{G \in \mathcal{G}} R_G(\varphi) \leq \sup_{\theta \in \Theta} R(\varphi | \theta).$$

Справедливость обратного неравенства, откуда будет следовать утверждение Предложения, доказывается столь же просто:

$$\sup_{\theta \in \Theta} R(\varphi | \theta) = \sup_{G \in \mathcal{H}} R_G(\varphi) \geq \sup_{G \in \mathcal{G}} R_G(\varphi). \quad \square$$

Равенство (5.1) говорит о том, что минимаксная оценка, возможно, будет байесовской при некотором “наихудшем” из всех априорных распределений. В связи с этим введем следующее

Определение 5.2. Априорное распределение G^* называется *наименее благоприятным*, если соответствующий ему байесовский риск не меньше байесовского риска при любом другом априорном распределении, то есть $R_{G^*}(\varphi_{G^*}) \geq R_G(\varphi_G)$.

Теорема 5.1. *Если для байесовского риска, соответствующего некоторому априорному распределению G , имеет место равенство*

$$R_G(\varphi_G) = \sup_{\theta \in \Theta} R(\varphi_G | \theta), \quad (5.2)$$

то (i) правило φ_G является минимаксным, (ii) G есть наименее благоприятное распределение.

Доказательство. (i) Пусть φ – любое правило принятия решения. Тогда

$$\begin{aligned} \sup_{\theta \in \Theta} R(\varphi | \theta) &\geq \int_{\Theta} R(\varphi | \theta) dG(\theta) \geq \int_{\Theta} R(\varphi_G | \theta) dG(\theta) = \\ &R_G(\varphi_G) = \sup_{\theta \in \Theta} R(\varphi_G | \theta) \end{aligned}$$

(см. равенство (5.2)). Согласно определению 5.1 φ_G есть минимаксное правило.

(ii) Если G' – некоторое другое априорное распределение, то

$$\begin{aligned} R_{G'}(\varphi_{G'}) &= \int_{\Theta} R(\varphi_{G'} | \theta) dG'(\theta) \leq \int_{\Theta} R(\varphi_G | \theta) dG'(\theta) \leq \\ &\sup_{\theta \in \Theta} R(\varphi_G | \theta) = R_G(\varphi_G), \end{aligned}$$

так что по определению 5.2 G – наименее благоприятное распределение. \square

Равенство (5.2) утверждает, что среднее значение $R(\varphi_G | \vartheta)$ по априорному распределению G равно максимуму функции риска $R(\varphi_G | \theta)$. Это выполняется, когда функция риска постоянна или, более общим образом, наименее благоприятное распределение G приписывает вероятность единица множеству параметров, на котором функция риска достигает своего наибольшего значения. Следовательно, справедливо

Следствие 5.1. *Если некоторое байесовское правило φ_G имеет постоянную функцию риска, то оно является минимаксным.*

ПРИМЕР 5.1. *Минимаксная оценка вероятности успешного испытания в схеме Бернулли.* Пусть $X = (X_1, \dots, X_n)$ – последовательность индикаторов успешного испытания, которое проводится в рамках схемы Бернулли с вероятностью успеха θ . Найдем минимаксную оценку этого параметра

при нормированной квадратичной функции потерь вида

$$L(\theta, d) = \frac{(\theta - d)^2}{\theta(1 - \theta)}.$$

Выберем в качестве априорного распределения φ равномерное распределение на интервале $(0, 1)$ и построим байесовское правило принятия решения относительно значения параметра θ .

Так как функция потерь выпукла по d , то байесовское правило не будет рандомизированным и будет иметь вид некоторой статистики $\hat{\theta} = \hat{\theta}(T)$ – функции от достаточной статистики $T = \sum_1^n X_i$, которая доставляет минимум апостериорному риску

$$\mathfrak{R}(d|T) = \frac{1}{p_G(T)} \int_0^1 L(\theta, d) p(T|\theta) g(\theta) d\theta,$$

где плотность $g(\theta)$, $\theta \in (0, 1)$, равномерного априорного распределения отлична от нуля только на интервале $(0, 1)$ и равна 1 на этом интервале, а маргинальная плотность статистики T

$$p_G(t) = P(T = t) = \int_0^1 p(t|\theta) g(\theta) d\theta = \int_0^1 C_n^t \theta^t (1 - \theta)^{n-t} d\theta = \\ C_n^t \cdot B(t + 1, n - t + 1).$$

Следовательно, апостериорное распределение ϑ есть бета-распределение с параметрами $t + 1$ и $n - t + 1$, так что апостериорный риск

$$\mathfrak{R}(d|T) = \frac{1}{B(t + 1, n - t + 1)} \int_0^1 \frac{(\theta - d)^2}{\theta(1 - \theta)} \theta^t (1 - \theta)^{n-t} d\theta = \\ \frac{B(t, n - t)}{B(t + 1, n - t + 1)} \left[\frac{1}{B(t, n - t)} \int_0^1 (\theta - d)^2 \theta^{t-1} (1 - \theta)^{n-t-1} d\theta \right].$$

Интеграл в квадратных скобках достигает минимума при d , равном среднему значению бета-распределения с параметрами t и $n - t$. Это среднее

легко вычисляется и доставляет байесовскую оценку вероятности θ успешного испытания:

$$\hat{\theta}_G(t) = \frac{1}{B(t, n-t)} \int_0^1 \theta \cdot \theta^{t-1} (1-\theta)^{n-t-1} d\theta = \frac{B(t+1, n-t)}{B(t, n-t)} = \frac{t}{n}.$$

Итак, байесовская оценка при равномерном априорном распределении совпадает с оценкой максимального правдоподобия и ее значение равно относительной частоте успехов в n испытаниях Бернулли:

$$\hat{\theta}_G(T) = \frac{1}{n} \sum_{i=1}^n X_i.$$

Дисперсия статистики T/n равна $\theta(1-\theta)/n$, так что функция риска этой оценки

$$R(\hat{\theta}_G | \theta) = \frac{E(\theta - \hat{\theta}_G(T))^2}{\theta(1-\theta)} = \frac{\theta(1-\theta)}{n\theta(1-\theta)} = \frac{1}{n}.$$

Поскольку функция риска не зависит от θ , то T/n – минимаксная оценка.

Если в качестве функции потерь взять обычную квадратичную функцию $L(\theta, d) = (\theta - d)^2$, то минимаксная оценка будет иметь вид

$$\hat{\theta}(T) = \frac{T + \sqrt{n}/2}{n + \sqrt{n}/2}.$$

Для того чтобы доказать это, возьмите в качестве априорного распределения бета-распределение с одинаковыми параметрами, равными $\sqrt{n}/2$, и сделайте вычисления, аналогичные приведенным выше.

Метод нахождения минимаксных оценок, который дает утверждение Теоремы 5.1, требует существования наименее благоприятного априорного распределения, и если такового не существует, то метод не применим. Рассмотрим, например, проблему минимаксной оценки среднего значения θ нормального (θ, σ^2) распределения. Интуитивно мы осознаем, что для параметра сдвига не стоит отдавать предпочтение какому-либо отдельному значению θ , в то время как в примере 5.1 в большей степени ощущалась необходимость концентрирования априорного распределения около значения вероятности успеха $\theta = 0,5$, – при этом значении θ биномиальное распределение имеет наибольшую дисперсию. Собственно, эта смутная догадка

получила свое подтверждение в задаче, которая предложена для самостоятельного решения в конце примера 5.1. В случае же параметра сдвига равномерность наименее благоприятного априорного распределения становится почти очевидной. Но область возможных значений параметра сдвига есть вся числовая прямая, и она не может быть носителем собственного равномерного распределения. Поэтому следует обобщить метод Теоремы 5.1 на случай несобственных наименее благоприятных распределений, определяя их как пределы вероятностных мер.

Пусть $\{G_n, n = 1, 2, \dots\}$ – последовательность собственных априорных распределений и $R_n = R_{G_n}(\varphi_{G_n})$ – байесовский риск, соответствующий G_n . Предположим, что последовательность байесовских рисков при $n \rightarrow \infty$ имеет некоторый конечный предел R .

Определение 5.3. Последовательность априорных распределений $\{G_n, n = 1, 2, \dots\}$ называется *наименее благоприятной*, если байесовский риск R_G , соответствующий любому априорному распределению G , удовлетворяет неравенству $R_G \leq R$.

Теорема 5.2. Пусть $\{G_n\}$ – последовательность собственных априорных распределений, для которых байесовские риски $\{R_n\}$ имеют конечный предел R . Если для функции риска некоторого правила принятия решения φ имеет место равенство

$$\sup_{\theta \in \Theta} R(\varphi | \theta) = R, \quad (5.3)$$

то (i) правило φ_G является минимаксным, (ii) последовательность G_n наименее благоприятна.

Доказательство. (i) Пусть φ' – любое другое правило принятия решения. Тогда

$$\sup_{\theta \in \Theta} R(\varphi' | \theta) \geq \int_{\Theta} R(\varphi' | \theta) dG_n(\theta) \geq \int_{\Theta} R(\varphi_{G_n} | \theta) dG_n(\theta) = R_n.$$

Поскольку это неравенство справедливо при любом $n = 1, 2, \dots$, то в силу условия (5.3)

$$\sup_{\theta \in \Theta} R(\varphi' | \theta) \geq \lim_{n \rightarrow \infty} R_n = R = \sup_{\theta \in \Theta} R(\varphi | \theta).$$

Следовательно, φ – минимаксное правило.

(ii) Пусть G – любое априорное распределение. Тогда в силу условия (5.3)

$$R_G = \int_{\Theta} R(\varphi_G | \theta) dG(\theta) \leq \int_{\Theta} R(\varphi | \theta) dG(\theta) \leq \sup_{\theta \in \Theta} R(\varphi | \theta) = R,$$

так что по определению 5.3 последовательность G_n наименее благоприятна. \square

ПРИМЕР 5.2. *Минимаксная оценка среднего значения нормального распределения.* Пусть $X = (X_1, \dots, X_n)$ – случайная выборка фиксированного объема n из нормального (θ, σ^2) распределения. Докажем, что выборочное среднее \bar{X} есть минимаксная оценка θ при функции потерь типа 1–0:

$$L(\theta, d) = \begin{cases} 0, & \text{если } |\theta - d| \leq \Delta, \\ 1, & \text{если } |\theta - d| > \Delta. \end{cases}$$

Поскольку выборочное среднее имеет нормальное распределение с параметрами $(\theta, \sigma^2/n)$, то функция риска оценки \bar{X} при выбранной функции потерь равна

$$R(\bar{X} | \theta) = P(|\bar{X} - \theta| > \Delta) = 2 \left[1 - \Phi \left(\frac{\Delta \sqrt{n}}{\sigma} \right) \right]. \quad (5.4)$$

Введем последовательность нормальных априорных распределений с одинаковыми средними $\mu_m = 0$ и дисперсиями $\sigma_m = m$, $m = 1, 2, \dots$. В § 4 при изучении модели N–N было показано, что апостериорное распределение ϑ есть нормальное распределение с параметрами

$$\mathcal{M} = \frac{n \bar{X} \sigma^{-2} + \mu \tau^{-2}}{n \sigma^{-2} + \tau^{-2}} \quad \text{и} \quad \mathcal{S}^2 = \frac{1}{n \sigma^{-2} + \tau^{-2}},$$

которые для частного случая нормального $(0, m)$ априорного распределения принимают вид

$$\mathcal{M} = \frac{\bar{X}}{1 + \sigma^2/(n m)} \quad \text{и} \quad \mathcal{S}^2 = \frac{\sigma^2}{n + \sigma^2/m}.$$

Апостериорный риск при функции потерь 1–0 равен апостериорной вероятности события $|\vartheta - d| > \Delta$. Поскольку функция плотности нормального апостериорного распределения симметрична относительно апостериорного среднего \mathcal{M} , то, как и в случае квадратичных потерь, апостериорное среднее \mathcal{M} есть байесовская оценка θ , апостериорный риск которой

$$\mathfrak{R}(\mathcal{M} | \bar{X}) = P(|\vartheta - \mathcal{M}| > \Delta | \bar{X}) = 2 \left[1 - \Phi \left(\frac{\Delta}{s} \right) \right].$$

Так как апостериорный риск не зависит от \bar{X} , то он равен априорному риску, а поскольку апостериорная дисперсия при $m \rightarrow \infty$ сходится к дисперсии σ^2/n выборочного среднего \bar{X} , то априорный риск при $m \rightarrow \infty$ сходится к

$$2 \left[1 - \Phi \left(\frac{\Delta \sqrt{n}}{\sigma} \right) \right] -$$

функции риска оценки \bar{X} , не зависящей от θ (см. формулу (5.4)). Следовательно, супремум по θ функции риска равен пределу априорного риска, равенство (5.3) в условии Теоремы 5.2 выполняется и \bar{X} есть минимаксная оценка.

Следует заметить, что найти минимаксные оценки в большинстве практически значимых задач нелегко. Например, наименее благоприятное распределение в задаче оценки вероятности успешного испытания в схеме Бернулли при функции потерь типа 1–0 сосредоточено в конечном числе точек интервала $(0, 1)$, нахождение координат которых представляет сложную вычислительную задачу. Аналогичная ситуация с оценкой среднего значения нормального распределения, если известно, что его значение принадлежит некоторому интервалу на прямой конечной длины.

ЗАЧЕТНЫЕ ЗАДАНИЯ

- 1*. Найдите минимаксную оценку параметра θ распределения Пуассона при функции потерь

$$L(\theta, d) = \begin{cases} 1, & \text{если } |\theta - d|/\theta > \Delta, \\ 0, & \text{если } |\theta - d|/\theta \leq \Delta. \end{cases}$$

§ 6. Статистические оценки равномерно минимальным риском

6.1. Несмещенные оценки с равномерно минимальной функцией θ -риска. Оптимальность правил принятия решения, методы построения которых рассматривались в двух предыдущих параграфах, носила “глобальный” характер – минимизировались определенные функционалы от функции риска для правила принятия решения в любой статистической проблеме, определяемой пространством решений, функцией потерь и вероятностной моделью. Если же обратиться к задаче равномерной минимизации функций θ - и d -риска по их аргументам, то здесь приходится рассматривать статистические проблемы оценки параметров, проверки гипотез и т.д. по отдельности. Данный параграф посвящен равномерной (по всем значениям аргумента) минимизации функции риска для правил оценки скалярного параметра, индексирующего семейство распределений наблюдаемого случайного элемента.

Итак, пусть X – случайная выборка, $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$ – статистическая модель (семейство возможных распределений X), доминируемая мерой μ , и $p(x | \theta)$ – функция плотности распределения $P(\cdot | \theta)$ по этой мере. Рассмотрим задачу оценки параметра θ , когда $\Theta \subset \mathbb{R}$, то есть θ – скалярный параметр, а пространство решений $\mathcal{D} = \Theta$. Естественно, рассматривается задача построения оптимальной оценки: для заданной функции потерь $L(\theta, d)$ требуется построить правило φ оценки значения θ , которое минимизирует величину средних потерь $R(\varphi | \theta) = \mathbf{E}_\theta L(\theta, \delta(X))$ (функцию риска) одновременно по всем значениям аргумента θ в предположении, что правила остановки и выбора заданы до проведения наблюдений.

Как известно из общего курса математической статистики, такая задача не имеет смысла, если не накладывать определенные ограничения на класс правил принятия решения. К сожалению, приходится накладывать также существенные ограничения на функцию потерь L и статистический эксперимент \mathcal{P} . Задача будет решаться при следующих условиях.

А. Функция потерь $L(\theta, d)$, $\theta \in \Theta \subset \mathbb{R}$, $d \in \mathcal{D} \subset \mathbb{R}$ есть строго выпуклая

(вниз) функция аргумента $d \in \mathbb{R}$.

В. Семейство распределений $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$ обладает полной достаточной статистикой $T = T(X)$.

С. Класс правил φ оценки θ обладает свойством несмещенности:

$$\mathbf{E}_\theta \delta(X) = \theta \quad \text{для всех } \theta \in \Theta,$$

где $\delta = \delta(X)$ – решающая функция (рандомизированная оценка) правила φ .

Поиск оптимальной оценки начнем с еще большего сужения класса правил φ .

Лемма 6.1. При выполнении условия А оценка, доставляющая наименьшее значение функции риска, является нерандомизированной.

Доказательство немедленно следует из неравенства Йенсена для математического ожидания от выпуклых функций. При каждом фиксированном результате x наблюдения случайной выборки X величина средних потерь от применения некоторого правила φ

$$\int_{\mathbb{R}} L(\theta, a) \varphi(da | x) \geq L\left(\theta, \int_{\mathbb{R}} a \varphi(da | x)\right).$$

Следовательно, нерандомизированная оценка

$$\hat{\theta}(X) = \int_{\mathbb{R}} a \varphi(da | X)$$

не хуже рандомизированного правила φ . □

Для построения оптимальных оценок потребуется вычислять их условное математическое ожидание относительно достаточной статистики.

Определение 6.1. Оценка $\theta^*(T) = \mathbf{E}\{\hat{\theta}(X) | T\}$ называется *проекцией* оценки $\hat{\theta}(X)$ на достаточную статистику $T = T(X)$.

Следует особо отметить роль достаточности T в определении проекции оценки. Если T не является достаточной статистикой, то условное

математическое ожидание будет зависеть от параметра θ , значение которого не известно, так что проекция не может служить оценкой этого значения.

Лемма 6.2. Если $\hat{\theta}(X)$ – несмещенная оценка, то ее проекция $\theta^*(T)$ на достаточную статистику T также обладает свойством несмещенности.

Доказательство. Используя свойство условного математического ожидания: “среднее от условного среднего равно безусловному среднему”, получаем утверждение Леммы:

$$\mathbf{E}_\theta \theta^*(T) = \mathbf{E}_\theta \left[\mathbf{E} \{ \hat{\theta}(X) | \Theta \} \right] = \mathbf{E}_\theta \hat{\theta}(X) = \theta$$

для всех $\theta \in \Theta$. □

Лемма 6.3. При выполнении условия *A* операция проектирования оценки не приводит к увеличению риска: $R(\theta^* | \theta) \leq R(\hat{\theta} | \theta)$ для всех $\theta \in \Theta$.

Доказательство. Используя неравенство Йенсена для условного математического ожидания, а также свойство условного математического ожидания, которое только что использовалось при доказательстве предыдущей Леммы, получаем требуемое неравенство:

$$R(\theta^* | \theta) = \mathbf{E}_\theta L(\theta, E \{ \hat{\theta}(X) | \Theta \}) \leq E_\theta \left[\mathbf{E} \{ L(\theta, \hat{\theta}(X)) | \Theta \} \right] = \mathbf{E}_\theta L(\theta, \hat{\theta}(X)) = R(\hat{\theta} | \theta). \quad \square$$

Лемма 6.4. При выполнении условия *B* (достаточная статистика T является полной) проектирование любой несмещенной оценки дает одну и ту же оценку.

Доказательство. Пусть $\hat{\theta}_1(X)$ и $\hat{\theta}_2(X)$ – несмещенные оценки θ , то есть

$$\mathbf{E}_\theta \left[\hat{\theta}_1(X) - \hat{\theta}_2(X) \right] = \theta - \theta = 0.$$

Используя снова свойство условного математического ожидания, находим, что

$$\mathbf{E}_\theta [\theta_1^*(X) - \theta_2^*(X)] = \mathbf{E}_\theta \mathbf{E} \left\{ \left[\hat{\theta}_1(X) - \hat{\theta}_2(X) \right] \middle| T \right\} =$$

$$\mathbf{E}_\theta \left[\hat{\theta}_1(X) - \hat{\theta}_2(X) \right] = 0$$

при всех $\theta \in \Theta$. Так как T – полная достаточная статистика, то равенство нулю при всех $\theta \in \Theta$ математического ожидания от статистики $\theta_1^*(X) - \theta_2^*(X)$ влечет, что эта статистика почти наверное равна нулю, то есть $\theta_1^*(X) = \theta_2^*(X)$. \square

Теорема 6.1 (Рао–Блекуэлла–Колмогорова). *Если выполняются условия А–С, то несмещенная оценка $\theta^*(T)$, зависящая только от полной достаточной статистики T , доставляет минимум функции риска равномерно по всем значениям аргумента $\theta \in \Theta$; общий вид такой оценки:*

$$\theta^*(T) = \mathbf{E} \{ \hat{\theta}(X) | T \}, \quad (6.1)$$

где $\hat{\theta}(x)$ – любая несмещенная оценка параметра θ .

Доказательство. То, что равенство (6.1) определяет общий вид оптимальной оценки, непосредственно следует из Лемм 6.4 и 6.3. Если же некоторая несмещенная оценка $\hat{\theta}(T)$ является функцией только достаточной статистики T , то ее проекция на T совпадает с оценкой $\hat{\theta}(T)$, что доказывает первое утверждение теоремы. \square

Утверждение Теоремы 6.1 доставляет два метода построения несмещенных оценок с минимальным риском (в дальнейшем используется аббревиатура НОМР). Первый состоит в том, подбирается подходящая функция от достаточной статистики, которая обладает свойством несмещенности. Если это удалось сделать, то мы располагаем искомой оценкой параметра θ . Например, выборочное среднее \bar{X} есть несмещенная оценка математического ожидания наблюдаемой случайной величины, то есть параметрической функции $h(\theta) = \mathbf{E}_\theta \xi$. Следовательно, $\hat{\theta} = \bar{X}$ есть НОМР для среднего значения нормального распределения, вероятности успеха в испытаниях Бернулли, параметра интенсивности распределения Пуассона, масштабного параметра показательного распределения. Исправленная на смещение выборочная дисперсия есть НОМР для дисперсии нормального распределения.

Оценка более сложных параметрических функций осуществляется вторым методом – проектированием некоторой, достаточно простой оценки на полную достаточную статистику.

ПРИМЕР 6.1. *Оценка надежности при показательном распределении долговечности.* Пусть $X = (X_1, \dots, X_n)$ – случайная выборка из показательного распределения и $H(\theta) = \exp\{-t_0/\theta\}$ – надежность (вероятность безотказной работы), соответствующая гарантийному сроку службы t_0 . Построим НОМР для параметрической функции $H(\theta)$. Заметим, что оценка $H(\bar{X})$, основанная на подстановке НОРМ для θ , является смещенной (попробуйте показать это). Приходится искать другую (попроще) несмещенную оценку $H(\theta)$.

Предлагается использовать совсем “никудашную”, но зато “архипростейшую” оценку, основанную на первой компоненте X_1 случайной выборки $n \geq 2$:

$$\hat{H}(X_1) = \begin{cases} 1, & \text{если } X_1 \geq t_0, \\ 0, & \text{если } X_1 < t_0. \end{cases}$$

Поскольку $\mathbf{E}_\theta \hat{H}(X_1) = P(X_1 \geq t_0) = \exp\{-t_0/\theta\} = H(\theta)$, то данная оценка является несмещенной; спроектируем ее на достаточную статистику $T = \sum_1^n X_k$. Для этого надо сначала найти совместную функцию плотности X_1 и T , затем маргинальную функцию плотности $g^T(t)$ статистики T и, наконец, условную функцию плотности X_1 относительно T : $g^{X_1|T}(x|T) = g(x, T)/g^T(T)$. Тогда искомая НОМР (см. (6.1))

$$H^*(T) = \mathbf{E}\{\hat{H}(X_1) | T\} = P\{X_1 \geq t_0 | T\} = \int_{t_0}^{\infty} g^{X_1|T}(x|T) dx. \quad (6.2)$$

Известно, что статистика T имеет гамма распределение с функцией плотности

$$g^T(t) = \frac{1}{\theta^n \Gamma(n)} t^{n-1} \exp\left\{-\frac{t}{\theta}\right\}.$$

Чтобы найти совместную функцию плотности $g(x, t)$, установим сначала совместную функцию плотности $h(x, y)$ независимых статистик $X = X_1$ и $Y = \sum_2^n X_k$. Понятно, что она равна произведению функций плотностей

показательного и гамма распределений:

$$h(x, y) = \frac{1}{\theta} \exp \left\{ -\frac{x}{\theta} \right\} \cdot \frac{1}{\theta^{n-1} \Gamma(n)} y^{n-2} \exp \left\{ -\frac{y}{\theta} \right\} =$$

$$\frac{y^{n-2}}{\theta^n \Gamma(n-1)} \exp \left\{ -\frac{x+y}{\theta} \right\}.$$

Тогда совместная функция распределения статистик $X = X_1$ и $T = X + Y$

$$G(x, t) = P(X < x, X + Y < t) = \int_0^x du \int_0^{t-u} h(u, v) dv,$$

при условии $0 < x < t$. Совместная функция плотности этих статистик

$$g(x, t) = \frac{\partial^2 G(x, t)}{\partial x \partial t} = \frac{d}{dx} \int_0^x h(u, t-u) du = h(x, t-x) =$$

$$\frac{1}{\theta^n \Gamma(n)} (t-x)^{n-2} \exp \left\{ -\frac{t}{\theta} \right\},$$

которая отлична от нуля только в области $0 < x < t$.

Наконец, функция плотности условного распределения X_1 относительно статистики T

$$g^{X_1|T}(x|T) = \frac{g(x, T)}{g^T(T)} = \frac{n-1}{T} \left(1 - \frac{x}{T}\right)^{n-2},$$

которая не равна нулю только при условии $x < T$. Искомая оценка (см. (6.2))

$$H^*(T) = \frac{n-1}{T} \int_{t_0}^{\max\{t_0, T\}} \left(1 - \frac{x}{T}\right)^{n-2} dx = \begin{cases} (1 - t_0/T)^{n-1}, & \text{если } T \geq t_0, \\ 0, & \text{если } T < t_0. \end{cases}$$

Отметим, что при больших объемах выборок эта оценка асимптотически ($n \rightarrow \infty$) эквивалентна оценке

$$H(\bar{X}) = \exp \left\{ -\frac{n t_0}{T} \right\} = \exp \left\{ -\frac{t_0}{\bar{X}} \right\},$$

основанной на подстановке НОМР для θ .

6.2*. Статистические оценки с равномерно минимальной функцией d -риска. Мы снова возвращаемся к вероятностным моделям,

в которых значение параметра при проведении наблюдений в статистическом эксперименте является случайным элементом с заданным априорным распределением G . Функция d -риска определялась в первом параграфе как условное математическое ожидание функции потерь относительно решающей функции:

$$R_G(\varphi | \delta) = \mathbf{E} \{L(\vartheta, \delta) | \delta\}.$$

Таким образом, как и любое условное математическое ожидание, d -риск является случайной величиной. При существовании регулярных условных вероятностей его можно представить как функцию на пространстве решений \mathcal{D} :

$$R_G(\varphi | d) = \mathbf{E} \{L(\vartheta, d) | \delta(X) = d\}, \quad d \in \mathcal{D}.$$

В практических приложениях такая функция при каждом фиксированном d интерпретируется как величина средних потерь среди тех статистических экспериментов, которые завершились принятием решения d .

Рассмотрим задачу минимизации $R_G(\varphi | d)$ равномерно по всем $d \in \mathcal{D}$. Удивительно, но решение этой задачи не требует столь обременительных условий А–С, которые накладывались при равномерной минимизации функции θ -риска. По-существу определяющим условием существования оценок с равномерно минимальным d -риском является существование достаточных статистик, и только. Основная идея построения таких оценок, как и байесовских, состоит в представлении d -риска через апостериорный риск: функция потерь сначала усредняется по апостериорному распределению ϑ , а потом по условному распределению X относительно статистики (решающей функции) $\delta = \delta(X)$:

$$R_G(\varphi | d) = \mathbf{E} \{ \mathfrak{R}(\varphi | X) | \delta(X) = d \}. \quad (6, 3)$$

Напомним некоторые определения, которые вводились в § 1 и относились к априорным и d -апостериорным характеристикам правил принятия решения φ . В случае невырожденных априорных распределений ϑ *априорный образ* правила φ (в общем случае, когда управление статистическим экспериментом не задано, то стратегии φ) определялся как

$$\Psi_G(D) = \int_{\Theta} \Psi(D | \theta) dG(\theta), \quad D \in \mathcal{C},$$

где $\Psi(D|\theta)$, $D \in \mathfrak{C}$, $\theta \in \Theta$, – образ φ . Априорному образу соответствовала *априорная оперативная характеристика* $\psi_G(d)$, $d \in \mathfrak{D}$, – функция плотности распределения $\Psi_G(\cdot)$ по мере γ на измеримом пространстве решений $(\mathfrak{D}, \mathfrak{C})$. Определялись также *d-апостериорный образ* и соответствующая ему *d-апостериорная оперативная характеристика* φ :

$$\psi_G(\theta|d) = \frac{\psi(d|\theta)g(\theta)}{\psi_G(d)}, \quad \theta \in \Theta, d \in \mathfrak{D}.$$

Определение 6.2. Правило принятия решения φ^* называется *правиллом, равномерно минимизирующим d-риск* (в дальнейшем коротко: *U-правиллом*), если на носителе $\mathfrak{D}_\varphi = \{d : \psi_G(d) > 0, d \in \mathfrak{D}\}$ образа $\Psi_G(D)$, $D \in \mathfrak{C}$, любого правила φ *d-риск* φ^* не больше *d-риска* φ : $R_G(\varphi^*|d) \leq R_G(\varphi|d)$ для всех $d \in \mathfrak{D}_\varphi$.

В § 4 апостериорный риск определялся как

$$\mathfrak{R}(\varphi|X) = \int_{\Theta} \left[\int_{\mathfrak{D}} L(\theta, a) \varphi(da|X) \right] h(\theta|X) d\chi(\theta).$$

Для построения *U-правилл* будет удобнее использовать другой вариант апостериорного риска, в котором по сравнению с $\mathfrak{R}(\varphi|X)$ не осуществляется усреднение потерь по распределению $\varphi(\cdot|X)$:

$$\mathfrak{R}(d|X) = \int_{\Theta} L(\theta, d) h(\theta|X) d\chi(\theta).$$

Это более традиционное определение апостериорного риска как случайной функции аргумента $d \in \mathfrak{D}$. При фиксированном значении $X = x$ апостериорный риск $\mathfrak{R}(d|x)$ можно трактовать как величину средних потерь, если решение d принимается без учета полученных в эксперименте данных x .

Приступим теперь к построению *U-правилл*. Зафиксируем некоторое $d \in \mathfrak{D}$ и пусть $\mathfrak{X}(d)$ – множество точек $x \in \mathfrak{X}$, на котором достигается минимум апостериорного риска $\mathfrak{R}(d|x)$. Введем класс $\mathfrak{K}(d)$ правил φ , принимающих решение d лишь в том случае, когда выборочные данные $x \in \mathfrak{X}(d)$. Таким образом, апостериорный риск $\mathfrak{R}(d|x)$ постоянен на множестве $\mathfrak{X}(d)$, которое содержит множество $\{x : \delta(x) = d\} = \{x : \varphi(D|x) > 0; \forall D \in \mathfrak{C}, D \supseteq d\}$, где $\delta(X)$ – решающая функция правила $\varphi \in \mathfrak{K}(d)$. Поэтому *d-апостериорный риск* такого правила

$$R_G(\varphi|d) = \min_{x \in \mathfrak{X}} \mathfrak{R}(d|x) \tag{6.4}$$

для любого $d \in \mathcal{D}_\varphi$.

Покажем, что $\varphi \in \mathcal{K}(d)$ является необходимым и достаточным условием минимальности функции d -риска φ в точке d .

Теорема 6.2. (i) Если $\varphi^* \in \mathcal{K}(d)$, то для любого правила φ и любого $d \in \mathcal{D}_\varphi$ выполняется неравенство

$$R_G(\varphi^* | d) \leq R_G(\varphi | d).$$

(ii) Если d -риск правила $\tilde{\varphi}$ удовлетворяет неравенству

$$R_G(\tilde{\varphi} | d) \leq R_G(\varphi | d) \tag{6.5}$$

при любом φ и $d \in \mathcal{D}_\varphi$, то существует такое правило $\varphi^* \in \mathcal{K}(d)$, что

$$P_G \{ \tilde{\varphi}(D | X) \neq \varphi^*(D | X) \} = 0$$

при любом $D \in \mathcal{E}$, где P_G – маргинальное распределение X .

Доказательство. (i) То что условие $\varphi^* \in \mathcal{K}(d)$ является достаточным для минимальности d -риска этого правила в любой точке d , для которой $\psi_G^*(d) = 0$, вытекает из равенства (по определению) $R_G(\varphi^* | d) = 0$. Если же $\psi_G^*(d) > 0$, то (см. (6.3) и (6.4))

$$R_G(\varphi | d) = \mathbf{E} \{ \mathfrak{R}(d | X) | \delta(X) = d \} \geq \min_{x \in \mathcal{X}} \mathfrak{R}(d | x) = R_G(\varphi^* | d).$$

(ii) Пусть $\tilde{\varphi}$ – правило, удовлетворяющее неравенству (6.5). Так как $\mathcal{X}(d) = \{ x : \delta^*(x) = d \}$, то для доказательства необходимости требуется установить равенство вероятностей (по маргинальному распределению P_G случайной выборки X) принадлежности X множествам $\tilde{\mathcal{X}}(d) = \{ x : \tilde{\delta}(x) = d \}$ и $\mathcal{X}(d)$ или, что то же, $P_G(Y(d)) = 0$, где $Y(d) = \tilde{\mathcal{X}}(d) \setminus \mathcal{X}(d)$.

Допустим противное: $P_G(Y(d)) > 0$. Покажем, что тогда выполняется строгое неравенство

$$R_G(\tilde{\varphi} | d) > \min_{x \in \mathcal{X}} \mathfrak{R}(d | x),$$

откуда получаем противоречие: правило $\tilde{\varphi}$ не удовлетворяет (6.4), что противоречит (6.5). Действительно, в терминах индикаторной функции $\mathbf{I}_A(x)$, $A \in \mathcal{A}$, $x \in \mathcal{X}$, события $A = \mathcal{X}(d)$ d -апостериорный риск

$$R_G(\tilde{\varphi} | d) = \mathbf{E} \left\{ \mathfrak{R}(d | X) \mathbf{I}_A(x) | \tilde{\delta}(X) = d \right\} +$$

$$\mathbf{E} \left\{ \mathfrak{R}(d | X) \mathbf{I}_{A^c}(x) \mid \tilde{\delta}(X) = d \right\} >$$

$$\min_{x \in \mathcal{X}} \mathfrak{R}(d | x) \left[\mathbf{E} \left\{ \mathbf{I}_A(x) \mid \tilde{\delta}(X) = d \right\} + \mathbf{E} \left\{ \mathbf{I}_{A^c}(x) \mid \tilde{\delta}(X) = d \right\} \right] =$$

$$\min_{x \in \mathcal{X}} \mathfrak{R}(d | x),$$

поскольку

$$\mathbf{E} \left\{ \mathbf{I}_{A^c}(x) \mid \tilde{\delta}(X) = d \right\} = \frac{P_G(Y(d))}{\tilde{\psi}_G(d)}. \quad \square$$

Таким образом, U -правила являются представителями класса

$$\mathcal{K} = \bigcap_{d \in \mathcal{D}} \mathcal{K}(d),$$

и их построение осуществляется по следующей схеме. Для каждого $d \in \mathcal{D}$ находится множество $\mathcal{X}(d)$, точки x которого доставляют минимум апостериорному риску $\mathfrak{R}(d | x)$. Решение d принимается лишь в том случае, когда выборочные данные $x \in \mathcal{X}(d)$. Может оказаться, что существует подмножество $D \subseteq \mathcal{D}$, для которого

$$Z(D) = \bigcap_{d \in D} \mathcal{X}(d) \neq \emptyset.$$

В таком случае при $x \in Z$ принимается любое из решений $d \in \mathcal{D}$ в соответствии с произвольным распределением (рандомизированным правилом), носителем которого является множество Z . Если

$$\mathcal{Y} = \bigcup_{d \in \mathcal{D}} \mathcal{X}(d) = \mathcal{X}$$

(объединение множеств $\mathcal{X}(d)$ совпадает с выборочным пространством \mathcal{X}), то, следуя указанному выше построению, всегда можно найти U -правило; в противном случае U -правил не существует.

Утверждения Теоремы 6.2 относятся к любой проблеме статистического вывода. Однако в задачах с конечным пространством решений \mathcal{D} (проверка гипотез), как правило, за исключением патологических, практически не интересных, статистических моделей, множество \mathcal{Y} является пустым и U -правил не существует. Однако в задачах оценки параметров

при наличии достаточных статистик U -правила существуют и приведенные ниже примеры иллюстрируют общий метод их построения.

ПРИМЕР 6.2. *U -оценки при квадратичной функции потерь.* Рассмотрим задачу оценки скалярного параметра θ в статистической модели $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$ при квадратичной функции потерь $L(\theta, d) = (\theta - d)^2$ и заданном априорном распределении G случайного параметра ϑ .

Пусть $M(X) = E\{\vartheta | X\}$ – апостериорное среднее параметра. Представим апостериорный риск в следующем виде:

$$\begin{aligned} \mathfrak{R}(d | X) &= \mathbf{E}\{(\vartheta - d)^2 | X\} = \\ &= \mathbf{E}\{(\vartheta - M(X))^2 + 2(\vartheta - M(X))(M(X) - d) + (d - M(X))^2 | X\} = \\ &= \mathbf{E}\{(\vartheta - M(X))^2 | X\} + (d - M(X))^2. \end{aligned} \quad (6.6)$$

Такое разбиение апостериорного риска на два слагаемых более наглядно объясняет, почему байесовская оценка θ (точка достижения минимума $\mathfrak{R}(d | X)$ по d) есть апостериорное среднее: $\hat{\theta}_G = M(X)$, а байесовский риск равен априорной дисперсии этой оценки: $R_G(\hat{\theta}_G) = \mathbf{E}_G D\{\vartheta | X\}$ (математическое ожидание от апостериорной дисперсии берется по маргинальному распределению P_G случайной выборки X). При построении U -оценки приходится искать точку $x = X(d)$ достижения минимума $\mathfrak{R}(d | x)$ по переменной x . Если апостериорная дисперсия не зависит от X , то U -оценка, очевидно, совпадает с байесовской. Именно такой является U -оценка в рамках модели $N-N$. В противном случае U -оценка может существенно отличаться от байесовской. Покажем это на примере оценки вероятности θ успешного испытания в схеме Бернулли при равномерном априорном распределении ϑ на интервале $(0, 1)$.

В примере 5.1 было установлено, что апостериорное распределение ϑ есть бета-распределение с параметрами $\alpha = T+1$ и $\beta = n-T+1$. Поскольку для бета-распределения среднее и дисперсия равны соответственно

$$\frac{\alpha}{\alpha + \beta} \quad \text{и} \quad \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)},$$

то байесовская оценка θ при квадратичных потерях $\hat{\theta}_G = (T + 1)/(n + 2)$.

Построение U -оценки сопряжено с поиском точки $T(d)$ достижения минимума апостериорного риска по переменной T . Представление (6. 6) апостериорного риска в данном случае имеет вид

$$\mathfrak{R}(d|T) = \frac{(T+1)(n-T+1)}{(n+2)^2(n+3)} + \left(\frac{T}{n} - d\right)^2.$$

Дифференцируя это выражение по T и приравнявая полученную производную нулю, получаем, что $T(d)$ определяется решением уравнения

$$\frac{n-2T}{(n+2)^2(n+3)} + \frac{2T}{n^2} + \frac{2d}{n} = 0.$$

Решение этого уравнения относительно d дает U -оценку:

$$\delta_U(T) = \frac{n^2}{2(n+2)^2(n+3)} + T \left[\frac{1}{n} - \frac{n}{(n+2)^2(n+3)} \right].$$

Легко видеть, что при объеме наблюдений $n \rightarrow \infty$ обе оценки с точностью до слагаемых порядка $O(n^{-1})$ эквивалентны оценке максимального правдоподобия $\hat{\theta}_n = T/n$.

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. Найдите НОМР для дисперсии $\theta(1-\theta)$ двухточечного распределения (дисперсии случайной величины, наблюдаемой в испытаниях Бернулли).
2. Найдите НОМР для параметра формы гамма-распределения при известном параметре масштаба.
3. Найдите НОРМ для параметра λ распределения Парето

$$F(x|\lambda) = 1 - (a/x)^\lambda, \quad x \geq a,$$

при известном значении параметра положения a .

4. Найдите НОМР для параметра θ равномерного на интервале $(0, \theta)$ распределения.

5*. По аналогии с примером 6.1 найдите НОРМ для параметрической функции $e^{-\theta}$, где θ – параметр распределения Пуассона, то есть для вероятности $\mathbf{P}_\theta(\xi = 0)$.

6*. Для модели P–G найдите U -оценку параметра θ при функции потерь

$$L(\theta, d) = \begin{cases} 1, & \text{если } |\theta - d|/\theta > \Delta, \\ 0, & \text{если } |\theta - d|/\theta \leq \Delta. \end{cases}$$

Сравните d -риски U -оценки и байесовской (см. задание 3 в § 4).

§ 7. Равномерно наиболее мощные критерии для статистических моделей с монотонным отношением правдоподобия

В приложениях математической статистики существует обширный класс задач, в которых требуется проверить истинность некоторого высказывания относительно исследуемого объекта или выбрать одно из альтернативных решений, которое определит дальнейшее поведение статистика по отношению к этому объекту. Например, при аттестации партии дизельного топлива по общему содержанию серы мы должны не только дать точечную оценку данной характеристики топлива, но и принять решение о качестве выпускаемого продукта, которое повлечет за собой одно из следующих действий – или отослать топливо потребителю, или произвести дополнительную очистку топлива от вредных примесей. В исследованиях, подобных опытам Менделя, часто надо проверить гипотезу относительно предполагаемого значения вероятности наследования доминантного признака. Селекционер, работающий над получением нового вида пшеницы, должен подкрепить свое заключение о превосходстве нового вида над тем, который уже используется в сельскохозяйственной практике, с помощью сопоставления данных об урожайности этих видов. И так далее, и тому подобное, – вы сами можете привести примеры таких задач по выбору одного из ряда альтернативных решений.

В дальнейшем будут рассматриваться задачи, связанные только с выбором одного из двух решений. Пусть мы высказываем некоторое суждение (или предпринимаем действие) об исследуемом объекте, и пусть d_0 – решение об истинности этого суждения, в то время как d_1 – решение о его ложности. Таким образом, пространство решений \mathcal{D} в данной статистической проблеме состоит из двух точек: $\mathcal{D} = \{d_0, d_1\}$.

Для выбора одного из решений наблюдается случайная выборка X из некоторого распределения $P(\cdot | \theta)$, значение параметра θ которого не известно. Пусть Θ – параметрическое пространство (область возможных значений θ). В соответствии с принятой нами в §1 идеологией статисти-

ческого вывода мы сопоставляем каждому решению $d \in \mathcal{D}$ определенное подмножество Θ_d пространства Θ , то есть интерпретируем каждое решение в терминах высказываний об истинном значении параметра θ . В нашей статистической проблеме выбора одного из двух решений положим $\Theta_i = \Theta_{d_i}$, $i = 0, 1$, и введем ряд понятий и определений, используемых при решении этой проблемы.

Утверждение $H_0 : \theta \in \Theta_0$ называется *нулевой гипотезой*, а утверждение $H_1 : \theta \in \Theta_1$ – *альтернативной гипотезой* или (коротко) *альтернативой*. Гипотеза H_i называется *простой*, если соответствующее Θ_i состоит из одной точки параметрического пространства Θ ; в противном случае H_i называется *сложной гипотезой*; $i = 0, 1$.

Правило, по которому принимается или отвергается нулевая гипотеза H_0 , называется *критерием*. Иногда добавляется – *критерий согласия* (с нулевой гипотезой), особенно, когда альтернатива H_1 определена не совсем четко и под H_1 подразумевается “все остальное”. В случае полного равноправия гипотез говорят о критерии *различения гипотез*. Критерий определяется заданием переходной вероятности $\varphi(X) = \varphi(d_1 | X)$: при результате x статистического эксперимента $\varphi(x)$ определяет вероятность, с которой отвергается нулевая гипотеза H_0 (принимается альтернатива H_1) с помощью проведения соответствующей процедуры рандомизации. Функция $\varphi(x)$, $x \in \mathcal{X}$, поэтому называется *критической функцией* или, просто, *критерием* ($\varphi(x)$ “критикует” нулевую гипотезу). Если правило φ нерандомизировано, то есть $\varphi(x)$ принимает только значения 0 или 1, то подмножество $S = \{x : \varphi(x) = 1\}$ выборочного пространства \mathcal{X} называется *критической областью*: если выборочные данные x попадают в эту область, то нулевая гипотеза H_0 отклоняется и принимается альтернативное решение – справедлива H_1 . Область $A = S^c = \mathcal{X} \setminus S$ называется *областью принятия* нулевой гипотезы.

В рассматриваемой статистической проблеме величина риска, связанная с отклонением верной гипотезы, обычно соотносится с функцией потерь типа 0–1: потери считаются равными 1, если принята гипотеза H_i , а в действительности $\theta \in \Theta_{1-i}$, $i = 0, 1$; если же принята H_i и $\theta \in \Theta_i$, $i = 0, 1$, то

потери полагаются равными нулю. Легко видеть, что величина риска при любом значении параметра θ может быть определена с помощью функции $m(\theta) = \mathbf{E}_\theta \varphi(X)$, которая называется *функцией мощности* критерия φ . Эта функция указывает, как часто мы отклоняем нулевую гипотезу, когда θ – истинное значение параметра, и хорошим следует считать тот критерий, у которого функция $m(\theta)$ принимает близкие к нулю значения в области Θ_0 и близкие к единице – в области Θ_1 . В связи с этим вводятся две компоненты функции риска: $\alpha(\theta) = m(\theta)$ при $\theta \in \Theta_0$ и $\beta(\theta) = 1 - m(\theta)$ при $\theta \in \Theta_1$. Функция $\alpha(\theta)$, $\theta \in \Theta_0$ называется *вероятностью ошибки первого рода* – она указывает относительную частоту отклонения гипотезы H_0 , когда она в действительности верна ($\theta \in \Theta_0$). Функция $\beta(\theta)$, $\theta \in \Theta_1$ называется *вероятностью ошибки второго рода* – она указывает относительную частоту принятия гипотезы H_0 , когда она ложна (верна альтернативная гипотеза $H_1 : \theta \in \Theta_1$). Заметим, что функция мощности $m(\theta)$ в области Θ_1 трактуется как вероятность отклонения гипотезы H_0 , когда в действительности выбор идет из распределения с альтернативным значением $\theta \in \Theta_1$, и поэтому часть $m(\theta)$ при $\theta \in \Theta_1$ называется *мощностью* критерия φ и обозначается $\mathbf{m}(\theta)$, $\theta \in \Theta_1$.

Легко понять, что при фиксированном объеме наблюдений невозможно одновременно минимизировать вероятности обеих ошибок. Это особенно ясно в случае нерандомизированного критерия: для уменьшения вероятности ошибки первого рода $\alpha(\theta) = \mathbf{P}_\theta(X \in S)$, $\theta \in \Theta_0$, необходимо уменьшить критическую область S , что приведет к увеличению области A принятия нулевой гипотезы и, следовательно, к увеличению вероятности ошибки второго рода $\beta(\theta) = \mathbf{P}_\theta(X \in A)$, $\theta \in \Theta_1$. Здесь возникает такая же ситуация, что и в проблеме построения оценки параметра θ с равномерно минимальным риском, – такие оценки существуют только в определенном классе статистических правил, например в классе несмещенных оценок. Однако даже и помимо задачи проверки гипотез с минимальной вероятностью ошибки, и намного раньше создания общей теории наиболее мощных критериев в статистической практике сложился следующий подход к управлению риском критерия.

Предположим, что отклонение гипотезы H_0 , когда она в действительности верна, приводит к более тяжким последствиям, чем ее принятие при справедливости альтернативы. В таком случае мы заинтересованы в первую очередь контролировать вероятность ошибки первого рода. С этой целью заранее фиксируется (выбирается) некоторый уровень α , выше которого вероятность ошибки первого рода не допустима, и критическая область S (критерий φ) определяется таким образом, что $\alpha(\theta) \leq \alpha$, каково бы ни было $\theta \in \Theta_0$. Это ограничение α на вероятность ошибки первого рода называется *уровнем значимости*, а сам критерий φ , для которого выполняется это ограничение, – *критерием уровня α* . Наибольшее значение вероятности ошибки первого рода

$$\bar{\alpha} = \sup_{\theta \in \Theta_0} \alpha(\theta)$$

называется *размером* критерия φ , и если $\bar{\alpha} = \alpha$, то говорят о *критерии φ размера α* .

В этом выборе ограничения именно на вероятность ошибки первого, а не второго рода проявляется типичная асимметрия в практической ценности гипотезы и альтернативы. Например, если проверяется эффективность нового лекарственного препарата, то нулевой гипотезе должно соответствовать решение о его неэффективности, ибо, отклонив эту гипотезу, когда она верна, мы внедрим в лечебную практику бесполезное или вредное лекарство, что приведет к более тяжким последствиям, чем отклонение в действительности эффективного препарата. Но если мы ищем золото, анализируя состав кернов при бурении предполагаемого месторождения, то естественно принять за нулевую гипотезу утверждение о наличии золота, ибо отклонив ее, когда она верна, мы потеряем намного больше, чем стоимость нескольких дополнительных анализов, удостоверяющих, что золото в разбуренной местности отсутствует.

Следует также обратить особое внимание на общую методологию проверки гипотез, отражаемую в выборе малого значения уровня α . Если наши выборочные данные попадают в область S с исключительно малой вероятностью, то естественно предположить, что то утверждение, которое привело к этому маловероятному событию, не соответствует истине, и отклонить

его. Поступая таким образом, мы будем терять в действительности верную гипотезу H_0 крайне редко – не более, чем в $100\alpha\%$ случаев.

Основная задача теории проверки гипотез состоит в построении критерия заданного уровня α , который равномерно по всем $\theta \in \Theta$ доставляет максимум мощности $\mathbf{m}(\theta)$, $\theta \in \Theta_1$, в классе всех критериев уровня α . Такой критерий φ^* , для которого выполняются неравенства $\mathbf{E}_\theta \varphi^*(X) \leq \alpha$ для всех $\theta \in \Theta_0$ и $\mathbf{E}_\theta \varphi^*(X) \geq \mathbf{E}_\theta \varphi(X)$ для всех $\theta \in \Theta_1$ и любого критерия φ с $\mathbf{E}_\theta \varphi(X) \leq \alpha$ для всех $\theta \in \Theta_0$, называется *равномерно наиболее мощным* (в дальнейшем РНМ критерием).

Основная идея метода построения РНМ критериев раскрывается в утверждении, которое обычно называется “фундаментальной леммой теории проверки гипотез” или “фундаментальной леммой Неймана–Пирсона”. В этой лемме рассматривается критерий проверки простой гипотезы H_0 : выборочный вектор X имеет функцию плотности $p_0(x)$, $x \in X$ при также простой альтернативе H_1 : распределение X имеет функцию плотности $p_1(x)$, $x \in X$. Критерий строится на основе статистики *отношения правдоподобия*

$$L = L(X) = \frac{p_1(X)}{p_0(X)}.$$

Лемма 7.1. Пусть P_0 и P_1 – распределения вероятностей, обладающие плотностями p_0 и p_1 соответственно, по отношению к некоторой мере μ и определяющие соответствующие гипотезы H_0 и H_1

(i) *Существование.* Для проверки гипотезы H_0 при альтернативе H_1 найдутся критерий φ и константа k такие, что для любого фиксированного уровня значимости α , $0 < \alpha < 1$,

$$\mathbf{E}_0 \varphi(X) = \alpha \tag{7.1}$$

и

$$\varphi(x) = \begin{cases} 1, & \text{если } p_1(x) > k p_0(x), \\ 0, & \text{если } p_1(x) < k p_0(x). \end{cases} \tag{7.2}$$

(ii) *Достаточное условие для критерия наибольшей мощности.* Если критерий удовлетворяет требованиям (7.1) и (7.2) при некотором k , то он является наиболее мощным критерием уровня α для проверки H_0 против H_1 .

(iii) Необходимое условие для критерия наибольшей мощности. Если φ – наиболее мощный критерий уровня α для проверки H_0 против H_1 , то при некотором k он удовлетворяет (7.2) почти всюду по мере μ . Он также удовлетворяет (7.1), кроме случая, когда существует критерий размера строго меньше α и мощности 1.

Доказательство. (i) Рассмотрим функцию

$$\alpha(c) = \mathbf{P}_0(p_1(X) > c p_0(X)).$$

При вычислении P_0 вероятностей можно ограничиться только множествами, принадлежащими носителю распределения P_0 , убирая из них точки выборочного пространства \mathcal{X} , где функция плотности $p_0(x) = 0$. В силу этого замечания $1 - \alpha(c)$, $c \in \mathbf{R}$, является функцией распределения статистики отношения правдоподобия $L(X) = p_1(X)/p_0(X)$, так что функция $\alpha(c)$, $c \in \mathbf{R}$, не возрастает, непрерывна справа и

$$\mathbf{P}_0(L(X) = c) = \alpha(c - 0) - \alpha(c), \quad \alpha(-\infty) = 1, \quad \alpha(+\infty) = 0.$$

Для заданного уровня значимости α , $0 < \alpha < 1$, определим c_0 из соотношений $\alpha(c_0) \leq \alpha \leq \alpha(c_0 - 0)$ и введем критерий

$$\varphi(x) = \begin{cases} 1, & \text{если } p_1(x) > c_0 p_0(x), \\ \gamma, & \text{если } p_1(x) = c_0 p_0(x), \\ 0, & \text{если } p_1(x) < c_0 p_0(x), \end{cases} \quad (7.3)$$

где

$$\gamma = \frac{\alpha - \alpha(c_0)}{\alpha(c_0 - 0) - \alpha(c_0)}.$$

Постоянная рандомизации γ определена всюду, кроме случая $\alpha(c_0 - 0) - \alpha(c_0) = \mathbf{P}_0(p_1(X) = c_0 p_0(X)) = 0$, так что критерий φ определен почти всюду. Размер этого критерия

$$\mathbf{E}_0 \varphi(X) = P_0(L(X) > c_0) + \frac{\alpha - \alpha(c_0)}{\alpha(c_0 - 0) - \alpha(c_0)} \mathbf{P}_0(L(X) = c_0) = \alpha.$$

Таким образом, критерий вида (7.2) заданного размера α (критерий, удовлетворяющий равенству (7.1)) существует, если положить $k = c_0$.

Заметим, что значение *критической константы* c_0 определяется, по существу, единственным образом, кроме случая, когда $\alpha(c) = \alpha$ для целого интервала значений c . Если (c', c'') такой интервал, то \mathbf{P}_0 вероятность события $X \in C$, где $C = \{x : p_0(x) > 0 \text{ и } c' < L(x) < c''\}$, равна нулю: $\mathbf{P}_0(C) = \alpha(c') - \alpha(c'' - 0) = 0$. Так как $p_0(x) > 0$, когда $x \in C$, то последнее равенство возможно лишь в случае $\mu(C) = 0$. Но вероятность \mathbf{P}_1 также непрерывна относительно меры μ , откуда следует, что $\mathbf{P}_1(C) = 0$. Таким образом, множества, отвечающие различным значениям c , для которых $\alpha(c) = \alpha$, имеют нулевую вероятность по обоим тестируемым распределениям, и поэтому вообще могут быть исключены из выборочного пространства.

(ii) Пусть φ – критерий вида (7.2) и размера α (выполняется равенство (7.2)), а φ^* – другой критерий уровня α , то есть $\mathbf{E}_0 \varphi^*(X) \leq \alpha$. Требуется показать, что $\mathbf{E}_1 \varphi(X) \geq \mathbf{E}_1 \varphi^*(X)$.

Введем два подмножества выборочного пространства \mathcal{X} :

$$S^+ = \{x : \varphi(x) - \varphi^*(x) > 0\} \quad \text{и} \quad S^- = \{x : \varphi(x) - \varphi^*(x) < 0\}.$$

Если $x \in S^+$, то $\varphi(x) > \varphi^*(x) \geq 0$, и из вида (7.2) критерия φ следует, что тогда и $p_1(x) - k p_0(x) > 0$. Если же $x \in S^-$, то $\varphi(x) < \varphi^*(x) \leq 1$, и аналогичные рассуждения показывают, что тогда $p_1(x) - k p_0(x) \leq 0$. Следовательно, произведение $(\varphi(x) - \varphi^*(x))(p_1(x) - k p_0(x))$ всегда неотрицательно, так что

$$\int_{\mathcal{X}} (\varphi - \varphi^*) (p_1 - k p_0) d\mu = \int_{S^+ \cup S^-} (\varphi - \varphi^*) (p_1 - k p_0) d\mu \geq 0,$$

и для разности мощностей φ и φ^* получаем оценку

$$\int_{\mathcal{X}} (\varphi - \varphi^*) p_1 d\mu \geq \int_{\mathcal{X}} (\varphi - \varphi^*) k p_0 d\mu = \alpha - \mathbf{E}_0 \varphi^*(X) \geq 0.$$

Таким образом, мощность критерия φ^* не больше мощности φ , что и требовалось доказать.

(iii) Пусть φ^* – наиболее мощный критерий уровня α и φ – критерий, для которого выполняются условия (7.1) и (7.2). Требуется показать, что

$$\mu(\{x : \varphi(x) \neq \varphi^*(x)\}) = 0.$$

Рассмотрим множество

$$S = (S^+ \cup S^-) \cap \{x : p_1(x) \neq kp_0(x)\}.$$

Достаточно показать, что $\mu(S) = 0$. Допустим противное: $\mu(S) > 0$.

Так как произведение $(\varphi(x) - \varphi^*(x))(p_1(x) - kp_0(x)) > 0$, если $x \in S$, то

$$\int_{S^+ \cup S^-} (\varphi - \varphi^*)(p_1 - kp_0) d\mu = \int_S (\varphi - \varphi^*)(p_1 - kp_0) d\mu > 0,$$

то есть φ оказывается более мощным, чем φ^* , и мы приходим к противоречию. Следовательно, $\mu(S) = 0$ и необходимость условий (7.1) и (7.2) для наибольшей мощности критерия φ^* доказана.

Если φ^* имеет размер строго меньший заданного уровня α и мощность, меньшую 1, то имеется возможность включить в критическую область $\{x : \varphi^*(x) = 1\}$ дополнительные точки (или “части” точек, используя процедуру рандомизации). Такая операция приведет или к увеличению мощности (вплоть до значения 1), или размер критерия станет равным заданному уровню α . Таким образом, или $\mathbf{E}_0 \varphi^*(X) = \alpha$, или $\mathbf{E}_0 \varphi^*(X) = 1$. \square

Следствие 7.1. Пусть \mathbf{m} – мощность наиболее мощного критерия уровня α для проверки простой гипотезы H_0 : распределение X есть P_0 при альтернативе H_1 : распределение X есть P_1 . Тогда $\alpha < \mathbf{m}$, за исключением случая $P_0 = P_1$.

Доказательство. Так как критерий $\varphi \equiv \alpha$ имеет мощность, равную α , то, очевидно, $\alpha \leq \mathbf{m}$. Если $\alpha = \mathbf{m} < 1$, то критерий $\varphi \equiv \alpha$ оказывается наиболее мощным, и в силу утверждения (iii) Леммы 7.1 он должен удовлетворять (7.2). Но тогда $p_0(x) = kp_1(x)$ для почти всех x по мере μ . Поскольку

$$1 = \int_{\mathcal{X}} p_0 d\mu = k \int_{\mathcal{X}} p_1 d\mu = k,$$

то $p_0(x) = p_1(x)$ для почти всех x по мере μ . Следовательно, $P_0 = P_1$. \square

Лемма Неймана–Пирсона и следствие 7.1 дают метод построения равномерно наиболее мощных критериев. Рассмотрим сначала случай проверки сложной гипотезы $H_0 : \theta \leq \theta_0$ при аналогичной “односторонней” альтернативе $H_1 : \theta > \theta_0$ относительно скалярного параметра $\theta \in \Theta \subseteq \mathbb{R}$. Однако даже в этом простейшем случае построить РНМ критерий удастся только для частного случая статистических моделей, обладающих достаточными статистиками.

Определение 7.1. Семейство $\mathcal{P} = \{P(\cdot | \theta), \theta \in \mathbb{R}\}$ называется семейством с монотонным отношением относительно статистики $T = T(X)$ отношением правдоподобия, если для любых θ', θ'' с $\theta'' > \theta'$ отношение функций плотности $p(x | \theta'')/p(x | \theta')$ есть неубывающая функция статистики $T(x)$.

Теорема 7.1. Пусть θ – действительный параметр и семейство \mathcal{P} распределений случайной выборки X обладает монотонным отношением правдоподобия относительно некоторой статистики $T(X)$. Тогда

(i) Для проверки $H_0 : \theta \leq \theta_0$ при альтернативе $H_1 : \theta > \theta_0$ существует РНМ критерий

$$\varphi(x) = \begin{cases} 1, & \text{если } T(x) > C, \\ \gamma, & \text{если } T(x) = C, \\ 0, & \text{если } T(x) < C, \end{cases} \quad (7.4)$$

где C и γ выбираются так, чтобы выполнялось равенство

$$E_{\theta_0} \varphi(X) = \alpha. \quad (7.5)$$

(ii) Функция мощности $m(\theta) = E_{\theta} \varphi(X)$ этого критерия строго возрастает во всех точках, где $m(\theta) < 1$.

(iii) Для любого $\theta < \theta_0$ критерий (7.3) минимизирует вероятность ошибки первого рода $\alpha(\theta) = m(\theta)$, $\theta < \theta_0$ в классе всех критериев, удовлетворяющих (7.5).

Доказательство (i) и (ii). Рассмотрим сначала задачу проверки простой гипотезы $H'_0 : \theta = \theta_0$ при простой альтернативе $H'_1 : \theta = \theta_1 (> \theta_0)$.

Согласно Лемме 7.1, для данной задачи существует наиболее мощный критерий (7.3), который в силу того, что функция $p_1(x)/p_0(x)$ есть неубывающая функция $T(x)$, можно представить в виде (7.4). Постоянные C и γ в новой записи, естественно, приобретают другой вид, но для дальнейших целей их конкретный вид не играет роли, поскольку из утверждения (i) Леммы 7.1 следует, что существуют такие постоянные C и γ , для которых (7.4) и (7.5) выполняются.

Замечателен тот факт, что критерий (7.4) является наиболее мощным при любой альтернативе $\theta_1 > \theta_0$, то есть он является равномерно наиболее мощным критерием проверки простой гипотезы H'_0 при сложной альтернативе $H_1 : \theta > \theta_0$, именно той альтернативе, которая выдвигается в исходной постановке задачи.

Далее, если θ' и θ'' – два произвольных значения параметра, причем $\theta' < \theta''$, то из утверждения (ii) Леммы 7.1 следует, что критерий (7.4) является наиболее мощным критерием проверки простой гипотезы $\theta = \theta'$ при простой альтернативе $\theta = \theta''$ при уровне $\alpha' = m(\theta')$. Но тогда в силу Следствия 7.1 $m(\theta') < m(\theta'')$, что, во-первых, доказывает утверждение (ii) данной теоремы о неубывании функции мощности $m(\theta)$ при всех $\theta \in \mathbb{R}$, а во-вторых, показывает, что для критерия (7.4) справедливо неравенство

$$\mathbf{E}_\theta \varphi(X) \leq \mathbf{E}_{\theta_0} \varphi(X) = \alpha \quad (7.6)$$

для всех $\theta \leq \theta_0$. Следовательно, критерий (7.4) имеет заданный уровень α как критерий проверки исходной гипотезы $H_0 : \theta \leq \theta_0$ при альтернативе $H_1 : \theta > \theta_0$. Но класс критериев, для которых верно неравенство (7.6) содержится в более широком классе критериев с $\mathbf{E}_{\theta_0} \varphi(X) \leq \alpha$, в котором (7.4) максимизирует $\mathbf{m}(\theta)$ при всех $\theta > \theta_0$. Следовательно, (7.4) есть РНМ критерий и справедливо утверждение (i) данной теоремы.

(iii) следует из того факта, что критерий, *минимизирующий* мощность в задаче проверки простой гипотезы при простой альтернативе, получается из Леммы 7.1, если в ней все неравенства заменить на обратные. \square

Очевидные изменения в формулировке теоремы 7.1 дают решение проблемы проверки гипотезы $H_0 : \theta \geq \theta_0$ при альтернативе $H_1 : \theta < \theta_0$

Понятно, что статистические модели экспоненциальных распределений (см. определение 2.5) при определенных условиях на функцию $Q(\theta)$ обладают монотонным отношением правдоподобия относительно достаточной статистики T . В силу этого замечания из теоремы 7.1 немедленно вытекает

Следствие 7.2. Пусть случайная выборка X имеет распределение с плотностью (по некоторой мере μ)

$$p(x|\theta) = B(\theta) \exp\{Q(\theta)T(x)\}h(x),$$

где $Q(\theta)$ – строго возрастающая функция действительного параметра θ . Тогда формулы (7.4) и (7.5) Теоремы 7.1 определяют РНМ критерий для проверки гипотезы $H_0 : \theta \leq \theta_0$ при альтернативе $H_1 : \theta > \theta_0$. Если Q убывает, то неравенства в (7.4) заменяются противоположными.

ПРИМЕР 7.1. РНМ критерий для вероятности успешного испытания в схеме Бернулли. Мы неоднократно рассматривали статистическую модель для наблюдений в схеме Бернулли. Распределение случайной выборки $X = (X_1, \dots, X_n)$ имеет функцию плотности

$$p_n(x^{(n)}|\theta) = \theta^{\sum_1^n x_k} (1-\theta)^{n-\sum_1^n x_k},$$

где $\theta = \mathbf{P}(X_k = 1)$ (вероятность успешного испытания). Нетрудно проверить, что данная статистическая модель обладает монотонным отношением достаточной статистики $T = \sum_{k=1}^n X_k$ отношением правдоподобия. Функция плотности статистики T по считающей мере отлична от нуля только в целочисленных точках $t = 0, 1, \dots, n$ и равна

$$p(t|\theta) = C_n^t \theta^t (1-\theta)^{n-t}.$$

Рассмотрим задачу проверки гипотезы $H_0 : \theta \leq \theta_0$ при альтернативе $H_1 : \theta > \theta_0$. В силу теоремы 7.1 РНМ критерий уровня α определяется критической функцией

$$\varphi(T) = \begin{cases} 1, & \text{если } T > C, \\ \gamma, & \text{если } T = C, \\ 0, & \text{если } T < C, \end{cases} \quad (7.7)$$

где C и γ выбираются так, чтобы выполнялось равенство

$$\mathbf{E}_{\theta_0} \varphi(T(X)) = \alpha.$$

Определим критическую константу $C = C_\alpha$ как наименьшее целое число, для которого

$$h(C) = \sum_{t=C+1}^n p(t | \theta_0) \leq \alpha, \quad (7.8)$$

Если $C_\alpha + 1 \leq n$ и в (7.8) имеет место строгое неравенство, то постоянная рандомизации γ вычисляется по формуле

$$\gamma = \frac{\alpha - h(C_\alpha)}{p(C_\alpha | \theta_0)}.$$

Если $C_\alpha + 1 \leq n$ и в (7.8) имеет место равенство, то H_0 принимается, если $T \leq C_\alpha$, и отвергается, когда $T > C_\alpha$, – в этом случае постоянная рандомизации $\gamma = 0$. Если, наконец, $p(n | \theta_0) > \alpha$ (критическая константа C_α не определяется), то гипотеза H_0 отвергается с вероятностью $\gamma = \alpha / p(n | \theta_0)$ только в случае $T = n$; при $T < n$ нулевая гипотеза H_0 всегда принимается.

Интересно отметить, что в этом примере РНМ критерий можно трактовать как нерандомизированный, если включить процедуру рандомизации в статистический эксперимент. Когда $T = C_\alpha$ или n , то наблюдается случайная величина U , не зависящая от X и равномерно распределенная на интервале $[0, 1]$. Гипотеза H_0 отвергается, если результат u наблюдения U удовлетворяет неравенству $u > 1 - \gamma$. Будем теперь трактовать случайный вектор (X_1, \dots, X_n, U) как случайную выборку. Тогда нерандомизированный критерий с критической областью $T + U > C_\alpha + 1 - \gamma$, очевидно, эквивалентен РНМ критерию (7.7).

РНМ критерии существуют не только для односторонних, но также и для некоторых двусторонних гипотез вида $H_0 : \theta \leq \theta_1$ или $\theta \geq \theta_2$, где θ_1, θ_2 . Проблемы подобного рода могут возникать, например, если мы желаем определить, удовлетворяются ли заданные требования относительно пропорции ингредиента в каком-либо медикаменте. Следующая теорема указывает вид РНМ критерия и поведение его функции мощности.

Теорема 7.2. (i) Если распределение выборки X принадлежит однопараметрическому экспоненциальному семейству с функцией плотности по мере μ

$$p(x|\theta) = B(\theta) \exp\{Q(\theta)T(x)\}h(x),$$

где Q – строго возрастающая функция, то для проверки гипотезы $H_0: \theta \leq \theta_1$ или $\theta \geq \theta_2$, (θ_1, θ_2) при альтернативе $H_1: \theta_1 < \theta < \theta_2$ существует РНМ критерий, имеющий вид

$$\varphi(T) = \begin{cases} 1, & \text{если } C_1 < T < C_2 \quad (C_1 < C_2), \\ \gamma_i, & \text{если } T = C_i, \quad i = 1, 2, \\ 0, & \text{если } T < C_1 \text{ или } T > C_2, \end{cases} \quad (7.9)$$

где C_i и γ_i определяются из условий

$$\mathbf{E}_{\theta_1} \varphi(T(X)) = \mathbf{E}_{\theta_2} \varphi(T(X)) = \alpha. \quad (7.10)$$

(ii) Критерий (7.9) минимизирует $\mathbf{E}_{\theta} \varphi(X)$ при условии (7.10) при всех $\theta < \theta_1$ и $\theta > \theta_2$.

(iii) Функция мощности этого критерия имеет максимум в точке $\theta_0 \in (\theta_1, \theta_2)$ и строго убывает при удалении θ от θ_0 вправо или влево. При этом исключается случай, когда распределение статистики T сосредоточено всего в двух точках.

Отметим, что РНМ критерии для проверки гипотезы $H_0: \theta_1 \leq \theta \leq \theta_2$ или для гипотезы $H_0: \theta = \theta_0$ могут не существовать, – необходимо наложить дополнительное условие несмещенности на класс рассматриваемых критериев.

Доказательство данной теоремы достаточно сложно и требует предварительного обобщения Леммы 7.1 на случай различения более чем двух простых гипотез. Приведем только общую схему доказательства, остановившись более подробно на том, что РНМ критерий имеет вид (7.9).

Рассмотрим сначала задачу построения критерия $\varphi(X) = \psi(T(X))$, который удовлетворяет (7.10) и доставляет максимум $\mathbf{E}_{\theta_0}\varphi(X)$ в промежуточной точке $\theta_0 : \theta_1 < \theta_0 < \theta_2$. Воспользуемся методом неопределенных множителей Лагранжа, согласно которому данная задача эквивалентна отысканию максимума по φ функционала

$$\mathbf{E}_{\theta_0}\varphi(X) - \lambda_1\mathbf{E}_{\theta_1}\varphi(X) - \lambda_2\mathbf{E}_{\theta_2}\varphi(X) = \int_{\mathcal{X}} [p(x|\theta_0) - \lambda_1p(x|\theta_1) - \lambda_2p(x|\theta_2)] \varphi(x) d\mu(x).$$

Поскольку $0 \leq \varphi \leq 1$, то максимум, очевидно, достигается на функции $\varphi^*(x)$, равной 1, если $p(x|\theta_0) - \lambda_1p(x|\theta_1) - \lambda_2p(x|\theta_2) > 0$, и $\varphi^*(x) = 0$ в случае противоположного неравенства.

Для экспоненциального семейства, указанного в начале формулировки теоремы, $\varphi^*(x) = \psi^*(T(X)) = 1$, если

$$\lambda_1 B(\theta_1) \exp\{Q(\theta_1)T(x)\} + \lambda_2 B(\theta_2) \exp\{Q(\theta_2)T(x)\} < B(\theta_0) \exp\{Q(\theta_0)T(x)\}.$$

Используя более компактную форму этого неравенства, получаем, что $\psi(t) = 1$, когда

$$a_1 \exp\{b_1 t\} + a_2 \exp\{b_2 t\} < 1,$$

где $b_1 < 0 < b_2$ в силу строго возрастания функции $Q(\cdot)$.

Числа a_1 и a_2 не одновременно быть отрицательными, ибо тогда критерий всегда (при любом результате наблюдения $T(X)$) будет отклонять нулевую гипотезу. Если одно из a отрицательно, а другое положительно, то легко убедиться, что $w(t) = a_1 \exp\{b_1 t\} + a_2 \exp\{b_2 t\}$ есть строго монотонная функция, так что ψ^* не может удовлетворять (7.10). И только в случае положительности a_1 и a_2 функция $w(t)$ выпукла вниз и постоянные a_1, a_2, b_1, b_2 можно выбрать так, что критерий ψ^* будет удовлетворять (7.10). Поскольку ψ^* не зависит от θ_0 , то это, по существу, означает, что он РНМ критерий. Доказательство всех остальных свойств этого критерия, указанных в формулировке Теоремы 7.2, можно найти в § 7 главы 3 книги Э.Лемана “Проверка статистических гипотез”.

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. Для среднего значения θ нормального (θ, σ^2) распределения при известном значении σ^2 постройте РНМ критерии для проверки гипотез, рассмотренных в теоремах 7.1 и 7.2. Нарисуйте эскизы графиков функций мощности для этих критериев.
2. Для проверки гипотезы $\theta \leq \theta_0$ при альтернативе $\theta > \theta_0$ о параметре положения равномерного на интервале $(0, \theta)$ распределения постройте РНМ критерий, вычислите его мощность и нарисуйте эскиз графика функции мощности этого критерия.
3. По аналогии с примером 7.1 постройте РНМ критерий для проверки гипотезы $\theta \leq \theta_0$ при альтернативе $\theta > \theta_0$ о параметре θ распределения Пуассона. Используя асимптотическую нормальность тестовой статистики, аппроксимируйте функцию мощности критерия посредством функции нормального распределения; приведите приближенную формулу для расчета критической константы по заданному уровню значимости.
4. Для параметра формы λ гамма-распределения при известном значении $a = 1$ масштабного параметра постройте РНМ критерий заданного уровня α .
- 5*. Докажите, что для проверки гипотезы $H_0 : \theta_1 \leq \theta \leq \theta_2$ при альтернативе $H_1 : \theta < \theta_1$ или $\theta > \theta_2$ не существует РНМ критерия в классе всевозможных критериев заданного уровня α .

§ 8. Равномерно наиболее точные доверительные границы

В проблеме точечной оценки $\hat{\theta}(X)$ параметра θ при функции потерь типа 1–0 рассматривался класс оценок, гарантирующих заданную точность Δ . В этом классе искалась оценка, которая доставляла максимум некоторым функционалам от надежности $\mathbf{P}_\theta(|\hat{\theta}(X) - \theta| \leq \Delta)$ процедуры оценивания, например, максимизировалось наименьшее значение надежности по $\theta \in \Theta$ или среднее значение надежности по некоторому априорному распределению на параметрическом пространстве. При малых объемах наблюдений трудно ожидать большую величину надежности даже для оптимальных оценок и не слишком жестких требований к точности. Поэтому на практике чаще пытаются контролировать надежность оценки, накладывая ограничения на точность, которая может зависеть от результатов наблюдений. Это так называемый метод *доверительной* или *интервальной* оценки θ , который будет изучаться в данном параграфе. Но сначала дадим строгое определение объекта дальнейших исследований в общем случае абстрактного параметрического пространства Θ .

Определение 8.1. Подмножество $\Theta(X)$ параметрического пространства Θ , размеры и конфигурация которого определяются значениями выборочных данных, называется $(1 - \alpha)$ -*доверительным множеством*, если вероятность того, что $\Theta(X)$ накроет истинное значение параметра θ больше или равна заданному *доверительному уровню* $1 - \alpha$ при любом значении $\theta \in \Theta$:

$$\inf_{\theta \in \Theta} \mathbf{P}_\theta (\Theta(X) \ni \theta) \geq 1 - \alpha.$$

Если $\Theta \subset \mathbf{R}$, а доверительное множество имеет вид интервала на прямой \mathbf{R} : $\Theta(X) = (\underline{\theta}(X), \bar{\theta}(X))$, то $\Theta(X)$ называется $1 - \alpha$ -*доверительным интервалом*, в котором различаются *нижний* $\underline{\theta}(X)$ и *верхний доверительные пределы*. Если $\Theta(X) = (\underline{\theta}(X), \infty)$, то статистика $\underline{\theta}(X)$ называется *нижней доверительной границей*, а если $\Theta(X) = (-\infty, \bar{\theta}(X))$, то статистика $\bar{\theta}(X)$ называется *верхней доверительной границей*.

В общем курсе математической статистики строились доверительные

интервалы для параметров нормального распределения и асимптотически (объем наблюдений $n \rightarrow \infty$) доверительные интервалы для вероятности успешного испытания в схеме Бернулли и параметра распределения Пуассона. Возникает естественный вопрос, насколько хороши эти интервалы и, вообще, как ввести понятие оптимального доверительного множества? Естественно желание строить доверительные множества наименьшего (в среднем) диаметра или объема, но такие множества существуют очень редко и при весьма специфических ограничениях на статистическую модель. Следующее определение оптимального доверительного множества, методам построения которого посвящено дальнейшее содержание данного параграфа, принадлежит Дж.Нейману.

Определение 8.2. $(1 - \alpha)$ -доверительное множество $\Theta^*(X)$ называется *равномерно наиболее точным* (РНТ) доверительным множеством на $B \subset \Theta \times \Theta$, если для любых $(\theta, \theta') \in B$ и любого $(1 - \alpha)$ -доверительного множества $\Theta(X)$ справедливо неравенство

$$\mathbf{P}_\theta (\Theta^*(X) \ni \theta') \leq \mathbf{P}_\theta (\Theta(X) \ni \theta').$$

Таким образом, РНТ доверительное множество с предписанной вероятностью $1 - \alpha$ покрывает истинное значение параметра θ и с наименьшей (по сравнению с остальными доверительными множествами) – другое значение параметра θ' , которое связано с θ соотношением $(\theta, \theta') \in B$. Выбор множества B зависит, естественно, от конфигурации доверительного множества. Так, если строится верхняя доверительная граница $\bar{\theta}(X)$, то минимизируется вероятность накрытия значений $\theta' > \theta$, следовательно $B = \{ (\theta, \theta') : \theta' > \theta \}$. Если строится доверительный интервал, то следует минимизировать все точки, лежащие по обе стороны от истинного значения θ , так что $B = \{ (\theta, \theta') : \theta' \neq \theta \}$.

Основным инструментом построения РНТ доверительных множеств служат РНМ критерии при специальном выборе нулевой и альтернативной гипотез. Суть этого метода, о котором говорилось в общем курсе математической статистики, раскрывает следующая Теорема 8.1. В ее формулировке присутствуют нерандомизированные критерии с заданной областью принятия нулевой гипотезы. Тем не менее, она остается справедливой и для ран-

доминированных критериев, если процедуру рандомизации осуществлять с помощью наблюдения равномерно распределенной на интервале $[0, 1]$ случайной величины U и включить результат u ее наблюдения в выборочные данные x . Тогда любой рандомизированный критерий с критической функцией $\varphi(X)$ эквивалентен (имеет ту же функцию мощности) нерандомизированному критерию с критической областью $\{(x, u) : \varphi(x) \geq u\}$.

Теорема 8.1. (i) Рассмотрим для каждого $\theta_0 \in \Theta$ какой-либо критерий уровня α для проверки гипотезы $H(\theta_0) : \theta = \theta_0$. Пусть $A(\theta_0) \subset \mathcal{X}$ – область принятия $H(\theta_0)$. При каждом результате x наблюдения случайной выборки X определим подмножество $\Theta(x)$ значений параметра, полагая

$$\Theta(x) = \{\theta : x \in A(\theta), \theta \in \Theta\}.$$

Тогда $\Theta(X)$ является доверительным множеством для θ с доверительным уровнем $1 - \alpha$.

(ii) Если $A(\theta_0)$ является РНМ критерием уровня α для проверки гипотезы $H(\theta_0)$ при альтернативе $K(\theta_0)$, то $\Theta(X)$ минимизирует вероятность

$$\mathbf{P}_\theta(\Theta(X) \ni \theta') \text{ для всех } \theta \in K(\theta')$$

в классе всех доверительных множеств для θ с доверительным уровнем $1 - \alpha$.

Доказательство (i) По определению $\Theta(x)$ включение $\theta \in \Theta(x)$ выполняется тогда и только тогда, когда $x \in A(\theta)$, и, следовательно,

$$\mathbf{P}_\theta(\theta \in \Theta(X)) = \mathbf{P}_\theta(X \in A(\theta)) \geq 1 - \alpha.$$

(ii) Если $\Theta^*(X)$ – любое другое доверительное множество уровня $1 - \alpha$ и если $A^*(\theta) = \{x : \theta \in \Theta^*(X)\}$, то

$$\mathbf{P}_\theta(X \in A^*(\theta)) = \mathbf{P}_\theta(\Theta^*(X) \ni \theta) \geq 1 - \alpha,$$

так что $A^*(\theta_0)$ является областью принятия для критерия проверки гипотезы $H(\theta_0)$ с уровнем α . Так как $A(\theta_0)$ – область принятия РНМ критерия, то вероятность принадлежности выборочных данных $A(\theta_0)$, когда $\theta \in K(\theta_0)$

(вероятность ошибки второго рода) принимает наименьшее значение. Следовательно, при любом $\theta \in K(\theta_0)$

$$\mathbf{P}_\theta(X \in A^*(\theta_0)) \geq \mathbf{P}_\theta(X \in A(\theta_0)),$$

откуда

$$\mathbf{P}_\theta(\Theta^*(X) \ni \theta_0) \geq \mathbf{P}_\theta(\Theta(X) \ni \theta_0). \quad \square$$

Таким образом, $\Theta^*(X)$ является РНТ доверительным множеством на $B = \{(\theta, \theta') : \theta \in K(\theta')\}$.

В предыдущем параграфе был найден РНМ критерий для проверки односторонних гипотез при аналогичных односторонних альтернативах. Применим этот критерий в построении РНТ верхних и нижних доверительных границ для значения действительного параметра.

Теорема 8.2. Пусть статистическая модель обладает монотонным относительно некоторой статистики $T = T(X)$ отношением правдоподобия и функция распределения $G(t | \theta)$ статистики T непрерывна по аргументу $t \in \mathbb{R}$ при каждом фиксированном значении $\theta \in \Theta \subset \mathbb{R}$.

(i) При каждом значении $1 - \alpha$ доверительного уровня существует РНТ нижняя доверительная граница $\underline{\theta}(X)$ для θ .

(ii) Пусть $t = T(x)$ – значение статистики T , когда результат статистического эксперимента равен x . Если уравнение

$$G(t | \theta) = 1 - \alpha$$

имеет решение $\theta = \hat{\theta}(t) \in \Theta$, то решение единственно и $\underline{\theta}(X) = \hat{\theta}(T)$.

Доказательство (i) Поскольку функция распределения статистики T непрерывна, то для каждого $\theta_0 \in \Theta$ РНМ критерий проверки гипотезы $\theta = \theta_0$ при альтернативе $\theta > \theta_0$ будет нерандомизированным (см. Теорему 7.1) с критической областью $T(x) > C(\theta_0)$, где критическая константа $C(\theta_0)$ удовлетворяет уравнению $\mathbf{P}_{\theta_0}(T(X) > C(\theta_0)) = \alpha$.

При любом $\theta \in \Theta$ для области $A(\theta) = \{x : T(x) \leq C(\theta)\}$ принятия РНМ критерия сопоставим (см. Теорему 8.1) $(1 - \alpha)$ -доверительное множество $\Theta(x) = \{\theta : x \in A(\theta)\}$. Если функция $C(\theta)$ монотонно возрастает

с ростом θ , то РНТ нижняя доверительная граница $\underline{\theta}(X) = C^{-1}(T(X))$, где $C^{-1}(\cdot)$ – функция, обратная к $C(\cdot)$. Осталось показать, что функции $C(\cdot)$ строго возрастает.

Пусть $\theta_1 > \theta_0$ – некоторое другое значение параметра θ из области альтернатив и пусть $C(\theta_1)$ удовлетворяет уравнению $P_{\theta_1}(T(X) > C(\theta_1)) = \alpha$. Поскольку РНМ критерий является несмещенным (см. Следствие 7.1), то

$$\mathbf{P}_{\theta_1}(T(X) > C(\theta_1)) = \alpha = \mathbf{P}_{\theta_0}(T(X) > C(\theta_0)) < \mathbf{P}_{\theta_1}(T(X) > C(\theta_0)).$$

Следовательно, $C(\theta_0) < C(\theta_1)$ при любых $\theta_0 < \theta_1$, так что $C(\theta)$, $\theta \in \Theta$, – строго возрастающая функция на параметрическом пространстве.

(ii) Из следствия 7.1 вытекает, что $G(t|\theta)$ является строго убывающей функцией θ в каждой точке t , в которой $0 < G(t|\theta) < 1$. Следовательно, уравнение $G(t|\theta) = 1 - \alpha$ имеет не более одного решения. При каждом $\theta \in \Theta$ значение $C(\theta)$ удовлетворяет уравнению $\mathbf{P}_{\theta}(T(X) < C(\theta)) = G(C(\theta)|\theta) = 1 - \alpha$. Естественно, это уравнение остается в силе и при $\theta = \underline{\theta} = \underline{\theta}(X)$, то есть $G(C(\underline{\theta})|\underline{\theta}) = 1 - \alpha$. Так как $C(\underline{\theta}) = C(C^{-1}(T)) = T$, то получаем искомое уравнение $G(T|\underline{\theta}) = 1 - \alpha$. \square

При тех же предположениях соответствующая верхняя доверительная граница является решением уравнения $G(t|\theta) = \alpha$.

Доказанная теорема применима и в случае дискретных распределений статистики T . Пара (T, U) (см. замечание, предшествующее формулировке Теоремы 8.1) имеет особенно простое представление, когда T принимает целочисленные значения. В этом случае статистика $S = T + U$ эквивалентна паре (T, U) , так как с вероятностью единица $T = [S]$, $U = S - [S]$, где $[S]$ есть наибольшее целое, не превосходящее S . Распределение S непрерывно и доверительные границы можно строить, отправляясь от этой статистики.

ПРИМЕР 8.1. РНТ нижняя доверительная граница для вероятности успешного испытания в схеме Бернулли. Статистическая модель испытаний в схеме Бернулли обладает монотонным относительно достаточной статистики $T = \sum_{k=1}^n X_k$ отношением правдоподобия. Функция плотности

статистики T по считающей мере отлична от нуля только в целочисленных точках $t = 0, 1, \dots, n$ и равна

$$p(t|\theta) = C_n^t \theta^t (1 - \theta)^{n-t}.$$

Пусть случайная величина U , имеющая равномерное распределение на интервале $[0, 1]$, используется в процедуре рандомизации для РНМ критерия проверки гипотезы $\theta = \theta_0$ при альтернативе $\theta > \theta_0$. Согласно предшествующему данному примеру замечанию, для построения РНТ нижней доверительной границы с помощью Теоремы 8.2 введем статистику $S = T + U$ и покажем, что она имеет непрерывное распределение с функцией плотности

$$g(s|\theta) = C_n^{[s]} \theta^{[s]} (1 - \theta)^{n-[s]}, \quad (8.1)$$

отличной от нуля только в области $0 \leq s < n + 1$.

Функция распределения статистики S

$$G(s|\theta) = \mathbf{P}(S < s|\theta) = \mathbf{P}(S < [s]|\theta) + \mathbf{P}(S = [s], 0 \leq U < s - [s]|\theta) =$$

$$\sum_{k=0}^{[s]-1} p(k|\theta) + C_n^{[s]} \theta^{[s]} (1 - \theta)^{n-[s]}(s - [s]),$$

поскольку U не зависит от T . Это непрерывная функция, которая является нижней огибающей функции биномиального распределения. Нетрудно видеть, что ее производная определяется формулой (8.1). Таким образом, нижняя доверительная граница $\underline{\theta}(S)$ определяется решением уравнения $G(S|\theta) = 1 - \alpha$, а верхняя $\bar{\theta}(S)$ – решением уравнения $G(S|\theta) = \alpha$.

Понятно, что $(\underline{\theta}(S), \bar{\theta}(S))$ будет $(1 - 2\alpha)$ -доверительным интервалом для θ , но поскольку мы не имеем РНМ критерия для проверки гипотезы $\theta = \theta_0$ при альтернативе $\theta \neq \theta_0$, то мы ничего не можем сказать об оптимальности такого доверительного интервала. В следующем параграфе будут строиться РНМ несмещенные критерии для проверки такого рода гипотез, которым будут соответствовать РНТ несмещенные доверительные интервалы.

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. Для среднего значения θ нормального (θ, σ^2) распределения при известном значении σ^2 постройте верхнюю РНТ доверительную границу, используя критерий, полученный при выполнении задания 1 из § 7.
2. Для параметр положения θ равномерного на интервале $(0, \theta)$ распределения найдите верхнюю РНТ доверительную границу, используя критерий, полученный при выполнении задания 2 из § 7.
3. По аналогии с примером 8.1 постройте РНТ нижнюю доверительную границу для параметра θ распределения Пуассона. Используя асимптотическую нормальность тестовой статистики, полученную при выполнении задания 3 из § 7, приведите приближенную формулу для расчета этой границы.
4. Для параметра формы λ гамма-распределения при известном значении $a = 1$ масштабного параметра постройте РНТ верхнюю доверительную границу, используя критерий, построенный в задании 4 из § 7.

§ 9. Равномерно наиболее мощные несмещенные критерии

9.1. Равномерно наиболее мощные несмещенные критерии для однопараметрических экспоненциальных семейств. Рассмотрим общую проблему проверки гипотезы $H_0 : \theta \in \Theta_0$ при альтернативе $H_1 : \theta \in \Theta_1$ и пусть φ – критерий заданного уровня α для проверки H_0 . Если мощность этого критерия при некоторых альтернативных значениях $\theta \in \Theta_1$ окажется меньше уровня α , то при этих значениях вероятность принятия нулевой гипотезы окажется больше, чем когда она верна. Естественно, это нежелательное свойство критерия, и если при построении РНМ критериев наложить дополнительное ограничение на класс конкурирующих процедур проверки, устраняющее такого рода патологии, то это благотворно отразится на практическом использовании критериев.

Определение 9.1. Критерий φ , для функции мощности которого выполняются неравенства

$$m(\theta) = \mathbf{E}_\theta \varphi(X) \leq \alpha, \text{ если } \theta \in \Theta_0, \quad \text{и} \quad m(\theta) \geq \alpha, \text{ если } \theta \in \Theta_1,$$

называется *несмещенным*.

Ограничение несмещенности на класс рассматриваемых критериев оказывается полезным не только с практической точки зрения – в широком классе проблем, где РНМ критерии не существуют, тем не менее существуют РНМ несмещенные критерии. В частности, сюда включаются некоторые гипотезы вида $\theta_1 \leq \theta \leq \theta_2$ или $\theta = \theta_0$, когда распределение случайной выборки зависит кроме θ также и от других (*мешающих*) параметров.

Если $m(\theta)$ – непрерывная функция на параметрическом пространстве Θ , то несмещенность влечет $m(\theta) = \alpha$ для всех $\theta \in \Gamma$, где Γ – общая граница Θ_0 и $\Theta_1 = \Theta_0^c$, то есть множество точек θ предельных как для Θ_0 , так и для Θ_1 . Критерии, удовлетворяющие последнему условию, называются *подобными на границе* Θ_0 и Θ_1 . Так как оперировать с условием подобия на границе удобнее, чем с условием несмещенности, то следующая Лемма играет важную роль в отыскании РНМ несмещенных критериев.

Лемма 9.1. Если распределения $P(\cdot | \theta)$, $\theta \in \Theta$, случайной выборки X таковы, что функция мощности любого критерия непрерывна, и если φ^* является РНМ среди всех критериев, подобных на границе, то φ^* – РНМ несмещенный критерий.

Доказательство. Класс критериев, подобных на границе, содержит в себе класс несмещенных критериев. Следовательно, подобный на границе РНМ критерий φ^* равномерно не менее мощен, чем любой несмещенный критерий. С другой стороны, критерий φ^* является несмещенным, так как он равномерно не менее мощен, чем критерий $\varphi(x) \equiv \alpha$. \square

Дальнейшие результаты данного параграфа будут относиться к экспоненциальным семействам распределений случайной выборки X . Эти распределения были определены в § 2 и имели плотность вида

$$p(x | \theta) = B(\theta) \exp \left[\sum_{i=1}^k Q_i(\theta) T_i(x) \right] h(x).$$

Используя более естественную параметризацию и включая положительный множитель $h(x)$ в доминирующую меру μ , придадим экспоненциальному семейству форму

$$p(x | \theta) = B(\theta) \exp \left[\sum_{i=1}^k \theta_i T_i(x) \right] h(x).$$

Определим теперь вид РНМ критериев для проверки двухсторонних гипотез, когда редукция класса рассматриваемых критериев к классу несмещенных приводит к существованию РНМ критериев.

Теорема 9.1. Пусть распределение случайной выборки X принадлежит однопараметрическому экспоненциальному семейству с плотностью

$$p(x | \theta) = B(\theta) \exp\{\theta T(x)\} h(x). \quad (9.1)$$

Тогда для проверки гипотезы $H_0 : \theta_1 \leq \theta \leq \theta_2$ при альтернативе $H_1 : \theta < \theta_1$ или $\theta > \theta_2$ существует РНМ несмещенный критерий, определяемый равенствами

$$\varphi(x) = \begin{cases} 1, & \text{если } T(x) < C_1 \text{ или } T(x) > C_2, \\ \gamma_i, & \text{если } T(x) = C_i, \quad i = 1, 2, \\ 0, & \text{если } C_1 < T(x) < C_2, \end{cases} \quad (9.2)$$

где C_i и γ_i находятся из уравнений

$$E_{\theta_1} \varphi(X) = E_{\theta_2} \varphi(X) = \alpha. \quad (9.3)$$

Доказательство. В силу Леммы 9.2 функция мощности любого критерия для экспоненциального семейства непрерывна, поэтому применима Лемма 9.1 – несмещенность заменяется требованием подобия на границе, так что РНМ критерий уровня α должен удовлетворять уравнениям (9.3).

Множество граничных точек $\Gamma = \{\theta_1, \theta_2\}$, поэтому рассмотрим сначала задачу максимизации $E_{\theta} \varphi(X)$ для значения θ , лежащего вне интервала $[\theta_1, \theta_2]$ с учетом ограничений (9.3). Если задачу сформулировать в терминах $1 - \varphi(x)$, то из части (ii) Теоремы 7.2 вытекает, что ее решение дается формулами (9.2) и (9.3). Следовательно, критерий (9.2) есть РНМ критерий в классе всех несмещенных критериев. Как видно из части (iii) Теоремы 7.2, функция мощности этого критерия имеет минимум в точке, расположенной между θ_1 и θ_2 , и строго возрастает при удалении от точки минимума влево или вправо. \square

Теорема 9.2. Пусть распределение случайной выборки X принадлежит однопараметрическому экспоненциальному семейству с плотностью (9.1). Тогда для проверки гипотезы $H_0 : \theta = \theta_0$ при альтернативе $H_1 : \theta \neq \theta_0$ существует РНМ несмещенный критерий, определяемый равенствами (9.2). Постоянные C и γ в этом случае находятся из уравнений

$$E_{\theta_0} \varphi(X) = \alpha. \quad (9.4)$$

$$E_{\theta_0} [T(X) \varphi(X)] = \alpha E_{\theta_0} T(X). \quad (9.5)$$

Доказательство того, что искомый РНМ несмещенный критерий имеет вид (9.2), не столь сложно, сколько требует нескольких предварительных лемм и, в первую очередь, обобщения фундаментальной леммы Неймана–Пирсона на случай различения более чем двух гипотез. Поэтому доказательство этой части теоремы мы опускаем, но то, что постоянные C и γ определяются из уравнений (9.4) и (9.5), доказать просто необходимо для понимания дальнейших результатов данного параграфа.

Экспоненциальное семейство (9.1) обладает достаточной статистикой $T = T(X)$, поэтому в силу Теоремы 2.1 можно ограничиться критериями,

зависящими от x только через значения достаточной статистики: $\varphi = \varphi(T)$. Из утверждения (I) Предложения 2.4 следует, что функцию мощности критерия $\varphi(T)$ можно представить в виде интеграла

$$m(\theta) = B(\theta) \int_{\mathbf{R}} e^{\theta t} \varphi(t) d\nu(t) . \quad (9.6)$$

Несмещенность критерия влечет необходимость выполнения равенства (9.4), откуда и из существования производных у функции мощности (Предложения 2.3) вытекает, что θ_0 есть точка минимума функции мощности и производная в этой точке $m'(\theta_0) = 0$. Дифференцируя (9.6) под знаком интеграла, получаем следующее представление для производной от функции мощности:

$$m'(\theta) = \mathbf{E}_{\theta} [T \varphi(T)] + \frac{B'(\theta)}{B(\theta)} \mathbf{E}_{\theta} \varphi(T) . \quad (9.7)$$

Для критерия $\varphi(T) \equiv \alpha$ последнее уравнение превращается в

$$0 = \mathbf{E}_{\theta} T + \frac{B'(\theta)}{B(\theta)} .$$

Подстановка в (9.7) $B'(\theta)/B(\theta) = -\mathbf{E}_{\theta} T$ дает следующее выражение для производной от функции мощности:

$$m'(\theta) = \mathbf{E}_{\theta} [T \varphi(T)] - \mathbf{E}_{\theta} T \mathbf{E}_{\theta} \varphi(T) .$$

Последнее уравнение при $\theta = \theta_0$, когда $m'(\theta_0) = 0$, совпадает с (9.5). Таким образом, несмещенность, в дополнение к (9.4), влечет и (9.5). \square

ЗАМЕЧАНИЕ 9.1. Определение РНМ несмещенного критерия для проверки гипотезы $\theta = \theta_0$ по формулам (9.2), (9.4) и (9.5) значительно упрощается, когда распределение статистики T при справедливости нулевой гипотезы ($\theta = \theta_0$) является симметричным относительно некоторой точки m , то есть $\mathbf{P}(T < m - t | \theta_0) = \mathbf{P}(T > m + t | \theta_0)$. В этом случае РНМ критерий будет симметричным: $\varphi(m - T) = \varphi(m + T)$, поскольку такого вида критерии удовлетворяют (9.4), если положить $\gamma_2 = \gamma_1$, $C_2 = 2m - C_1$, а константы C_1 и γ_1 находить из уравнения

$$\mathbf{P}(T < C_1 | \theta_0) + \gamma_1 \mathbf{P}(T = C_1 | \theta_0) = \alpha/2 .$$

Тогда $\mathbf{E}_{\theta_0} \varphi(T) = \alpha$, то есть выполняется условие (9.4), причем это условие влечет (9.5). Действительно, функция $(t - m) \varphi(t)$, $t \in \mathbf{R}$, в этом случае является нечетной, $\mathbf{E}_{\theta_0} T = m$, и поэтому

$$\mathbf{E}_{\theta_0} [T \varphi(T)] = \mathbf{E}_{\theta_0} [(T - m) \varphi(T)] + m \mathbf{E}_{\theta_0} \varphi(T) = m \alpha = \alpha \mathbf{E}_{\theta_0} T.$$

Например, такие симметричные РНМ критерии существуют при проверке гипотезы $\theta = 1/2$ о вероятности успешного испытания в схеме Бернулли или при проверке гипотезы $\theta = \theta_0$ о среднем значении нормального распределения. При больших объемах испытаний n распределения часто используемых на практике тестовых статистик T аппроксимируется нормальным распределением, и в таком случае резонно также ограничиться симметричными решающими функциями. \square

9.2. Равномерно наиболее мощные несмещенные критерии при наличии мешающих параметров (многопараметрические экспоненциальные семейства). В § 7 и в разделе 9.1 настоящего параграфа были решены задачи построения РНМ критериев заданного уровня α для проверки следующих гипотез и соответствующих им альтернатив относительно скалярного параметра θ экспоненциального семейства (9.1):

$$\begin{array}{ll} (A) & H_0 : \theta \leq \theta_0 & H_1 : \theta > \theta_0 \\ (B) & H_0 : \theta \leq \theta_1 \text{ или } \theta \geq \theta_2 & H_1 : \theta_1 < \theta < \theta_2 \\ (C) & H_0 : \theta_1 \leq \theta \leq \theta_2 & H_1 : \theta < \theta_1 \text{ или } \theta > \theta_2 \\ (D) & H_0 : \theta = \theta_0 & H_1 : \theta \neq \theta_0 . \end{array}$$

Для проверки гипотез (А) и (В) РНМ критерии в классе всех критериев заданного уровня α определялись, соответственно, соотношениями (7.4)–(7.5) и (7.9)–(7.10). Что же касается проверки гипотез (С) и (D), то здесь РНМ критерии уровня α существовали только в классе несмещенных критериев. Общий вид критерия для проблем (С) и (D) определялся формулой (9.2), однако постоянные критерия определялись по-разному: для (С) критические константы и вероятности процедуры рандомизации определялись из уравнений (9.3), а для (D) – из уравнений (9.4) и (9.5).

В этом разделе будут рассматриваться аналогичные задачи для параметра θ многопараметрического экспоненциального семейства

$$dP(x | \theta, \vartheta) = B(\theta, \vartheta) \exp \left[\theta U(x) + \sum_{i=1}^k \vartheta_i T_i(x) \right] d\mu(x), \quad (\theta, \vartheta) \in \Theta, \quad (9.8)$$

где $\vartheta = (\vartheta_1, \dots, \vartheta_k)$ – вектор так называемых *мешающих* параметров, для которых $T = (T_1, \dots, T_k)$ – векторная достаточная статистика; предполагается, что параметрическое пространство Θ выпукло и имеет размерность $k + 1$. Поскольку для семейства распределений (9.8) существует $(k + 1)$ -мерная достаточная статистика (U, T) , то статистическая модель (9.8) редуцируется к семейству распределений этой статистики, распределение которой в силу Предложения 2.4 определяется формулой

$$dP^{U,T}(u, t | \theta, \vartheta) = B(\theta, \vartheta) \exp \left[\theta u + \sum_{i=1}^k \vartheta_i t_i \right] d\nu(u, t), \quad (\theta, \vartheta) \in \Theta,$$

а условное распределение U относительно T принадлежит экспоненциальному семейству

$$dP_{\theta}^{U|T}(u | t) = B_t(\theta) \exp\{\theta u\} d\nu_t(u).$$

Поскольку это семейство не зависит от мешающего параметра ϑ , то естественно рассматривать его как новую статистическую модель и строить РНМ критерий аналогии с решениями проблем (А–(D)), максимизируя мощность критерия на каждой поверхности $T(X) = t$ выборочного пространства \mathcal{X} . Статистика U будет лежать в основании критической функции, а постоянные C и γ будут зависеть от наблюдаемого значения t статистики T .

Итак, для проверки гипотез (А) предлагается критерий

$$\varphi_A(u, t) = \begin{cases} 1, & \text{если } u > C(t), \\ \gamma, & \text{если } u = C(t), \\ 0, & \text{если } u < C(t), \end{cases} \quad (9.9)$$

где $C(t)$ и $\gamma(t)$ выбираются так, чтобы выполнялось равенство

$$\mathbf{E}_{\theta_0} \{ \varphi_A(U, T) | T = t \} = \alpha. \quad (9.10)$$

Гипотезы (В) предлагается различать с помощью критерия

$$\varphi_B(u, t) = \begin{cases} 1, & \text{если } C_1(t) < u < C_2(t), \\ \gamma_i, & \text{если } u = C_i(t), \quad i = 1, 2, \\ 0, & \text{если } u < C_1(t) \text{ или } u > C_2(t), \end{cases} \quad (9.11)$$

где $C_i(t)$ и $\gamma_i(t)$, $i = 1, 2$, определяются из условий

$$\mathbf{E}_{\theta_1} \{ \varphi_B(U, T) | T = t \} = \mathbf{E}_{\theta_2} \{ \varphi_B(U, T) | T = t \} = \alpha. \quad (9.12)$$

Наконец, для решения проблемы проверки гипотез (С) и (D) предлагается общий критерий

$$\varphi(u, t) = \begin{cases} 1, & \text{если } u < C_1(t) \text{ или } u > C_2(t), \\ \gamma_i, & \text{если } u = C_i(t), \quad i = 1, 2, \\ 0, & \text{если } C_1(t) < u < C_2(t), \end{cases} \quad (9.13)$$

где C_i и γ_i , $i = 1, 2$, в случае задачи (С) находятся из уравнений

$$\mathbf{E}_{\theta_1} \{ \varphi(U, T) | T = t \} = \mathbf{E}_{\theta_2} \{ \varphi(U, T) | T = t \} = \alpha, \quad (9.14)$$

а для задачи (D) определяются из соотношений

$$\mathbf{E}_{\theta_0} \{ \varphi(U, T) | T = t \} = \alpha, \quad (9.15)$$

$$\mathbf{E}_{\theta_0} \{ U \varphi(U, T) | T = t \} = \alpha \mathbf{E}_{\theta_0} \{ U | T = t \}. \quad (9.16)$$

Равенства (9.10), (9.12), (9.14) – (9.16) должны выполняться при всех значениях t статистики $T(X)$.

Естественно, можно показать, что эти критерии являются РНМ мощными в классе всех несмещенных критериев уровня α , но только при дополнительном условии полноты семейства распределений достаточной статистики (U, T) .

Существо данной проблемы заключается в характеристизации *подобных* критериев на границе различаемых гипотез, из которой будет следовать несмещенность РНМ критериев (см. Лемму 9.1). В проблеме проверки гипотез (А)–(D) с мешающими параметрами такая граница $\Gamma = \Gamma_\theta \times \Gamma_\vartheta$, где Γ_θ – граница, относящаяся к тестируемому параметру θ и состоящая

не более чем из двух точек, а Γ_{ϑ} есть область всевозможных значений мешающего параметра $\vartheta = (\vartheta_1, \dots, \vartheta_k)$, то есть подмножество параметрического пространства Θ размерности k .

Рассмотрим данную проблему в более общем виде. Пусть T – достаточная статистика для подсемейства $\mathcal{P}^X = \{P^X(\cdot | \theta), \theta \in \Gamma\}$ статистической модели \mathcal{P} и $\mathcal{P}^T = \{P^T(\cdot | \theta), \theta \in \Gamma\}$ – соответствующее подсемейство распределений статистики T , когда параметр $\theta \in \Gamma \subset \Theta$. Требуется характеризовать класс критериев φ , для которых $\mathbf{E}_{\theta} \varphi(X) = \alpha$ при всех распределениях X , принадлежащих заданному семейству \mathcal{P}^X , то есть класс критериев, подобных по отношению к \mathcal{P}^X или Γ .

Любой критерий, для которого $\mathbf{E} \{ \varphi(X) | T(X) = t \} = \alpha$ для почти всех значений t статистики T по любому распределению из семейства \mathcal{P}^T оказывается подобным по отношению к \mathcal{P}^X . Доказательство тривиально:

$$\mathbf{E}_{\theta} \varphi(X) = \mathbf{E}_{\theta} E \{ \varphi(X) | T(X) = t \} = \alpha.$$

О критериях, удовлетворяющих

$$\mathbf{E} \{ \varphi(X) | T(X) = t \} = \alpha \quad \mathcal{P}^T\text{-почти всюду,} \quad (9.17)$$

говорят, что они имеют *неймановскую структуру* относительно статистики T . Для таких критериев вероятность отклонения нулевой гипотезы при ее справедливости равна α на каждой из поверхностей $S_t = \{x : T(x) = t\} \subset \mathcal{X}$. Так как распределение на каждой S_t не зависит от значений $\theta \in \Gamma$, то условие (9.17), по существу, сводит проблему к проверке простой гипотезы при каждом значении t . Среди критериев, обладающих неймановской структурой, часто без труда находится наиболее мощный, для чего проблема оптимизации решается на каждой поверхности S_t отдельно. Собственно, именно этот метод был использован при построении критериев (9.9)–(9.16). Но вся проблема состоит в том, что *построенный таким образом критерий оказывается наиболее мощным среди всех подобных критериев (и несмещенных критериев) только в том случае, если каждый подобный критерий обладает неймановской структурой*. Условие, при котором верно это утверждение, формулируется в терминах полноты семейства распределений статистики T . Понятие полноты было введено в § 2, напомним, что

это означает.

Определение 2.3. Статистика $T = T(X)$ называется *полной* для семейства вероятностных мер $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta\}$, если для любой измеримой функции g равенство $\mathbf{E}_\theta g(T(X)) = 0$ при всех $\theta \in \Theta$ влечет $g(t) = 0$ почти всюду по любой мере из семейства \mathcal{P} .

Лемма 9.2. Пусть X – случайная выборка и пусть T – достаточная статистика для статистической модели \mathcal{P} . Если семейство $\mathcal{P}^T = \{P^T(\cdot | \theta), \theta \in \Gamma\}$ распределений T полно, то все подобные критерии имеют неймановскую структуру (удовлетворяют (9.17)).

Доказательство. Пусть \mathcal{P}^T – полное семейство распределений и $\varphi(X)$ – критерий, подобный относительно $\mathcal{P}^X = \{P^X(\cdot | \theta), \theta \in \Gamma\}$. Тогда $\mathbf{E}_\theta [\varphi(X) - \alpha] = 0$ при всех $\theta \in \Gamma$. Если $\psi(t) = \mathbf{E} \{\varphi(X) - \alpha | T(X) = t\}$, то $\mathbf{E}_\theta \psi(T) = 0$ при всех $\theta \in \Gamma$. Из полноты семейства \mathcal{P}^T следует, что $\psi(t) = 0$, откуда

$$\mathbf{E} \{\varphi(X) | T(X) = t\} = \alpha \quad \mathcal{P}^T\text{-почти всюду,}$$

то есть любой подобный критерий имеет неймановскую структуру. \square

Интересно заметить, что верно и обратное утверждение: если подобные критерии имеют неймановскую структуру, то семейство распределений достаточной статистики *ограниченно полно* – в определении полноты любая измеримая функция g ограничена. Мы не будем доказывать это утверждение, поскольку оно используется только при доказательстве нижеследующей Теоремы 9.3. Поскольку доказать эту теорему мы также не успеем из-за нехватки времени, то зачем, спрашивается, доказывать вспомогательные утверждения, хотя, возможно, оно имеет самостоятельный интерес. Обратимся лучше к формулировке условий, при которых экспоненциальное семейство обладает полными достаточными статистиками.

Лемма 9.3. Пусть X – случайная выборка с распределением вероятностей

$$dP(x | \theta) = B(\theta) \exp \left[\sum_{i=1}^k \theta_i T_i(x) \right] d\mu(x)$$

и пусть $\mathcal{P}^T = \{P^T(\cdot|\theta), \theta \in \Gamma\}$ – семейство распределений векторной статистики $T = (T_1, \dots, T_k)$. Тогда, если подмножество Γ параметрического пространства Θ содержит k -мерный прямоугольник, то \mathcal{P}^T полно.

Сформулируем основной результат, который вытекает из всех предыдущих построений данного раздела.

Теорема 9.3. *Критерии, определяемые формулами (9.9)–(9.16), являются РНМ несмещенными критериями в классе всех несмещенных критериев заданного уровня α для проверки соответствующих гипотез (A)–(D).*

Наиболее любопытные и любознательные могут прочитать доказательства недоказанных утверждений в параграфах 3 и 4 главы 4 книги Э.Лемана “Проверка статистических гипотез”.

9.3 Примеры на построение равномерно наиболее мощных несмещенных критериев.

ПРИМЕР 9.1. *Сравнение параметров интенсивности двух распределений Пуассона.* Практически значимое содержание данного примера можно интерпретировать в терминах сравнения двух источников радиоактивности. Пусть построение вероятностной модели позволяет статистику трактовать результаты наблюдений x и y над каждым объектом радиоактивного излучения как реализации независимых случайных величин X и Y , имеющих распределения Пуассона с параметрами λ и μ соответственно.

Представим совместную функцию плотности X и Y в виде (9.8), удобном для применения Теоремы (9.3):

$$p(x, y | \lambda, \mu) = \frac{\lambda^x}{x!} e^{-\lambda} \cdot \frac{\mu^y}{y!} e^{-\mu} =$$

$$\frac{\exp\{-(\lambda + \mu)\}}{x! y!} \exp\left\{y \ln \frac{\mu}{\lambda} + (x + y) \ln \lambda\right\}.$$

В таком представлении тестируемый параметр $\theta = \ln(\mu/\lambda)$ с соответствующей ему достаточной статистикой $U = Y$, а мешающий параметр $\vartheta = \ln \lambda$

с достаточной статистикой $T = X + Y$. Условное распределение U относительно T определяется условной функцией плотности

$$p^{Y|(X+Y)}(u|t) = \frac{\mathbf{P}(Y = u, X + Y = t)}{\mathbf{P}(X + Y = t)}.$$

Вычислим вероятности в числителе и знаменателе этой формулы, используя независимость X и Y , а также теорему сложения для распределения Пуассона,

$$\mathbf{P}(Y = u, X + Y = t) = \mathbf{P}(X = t - u, Y = u) = \frac{\lambda^{t-u}}{(t-u)!} e^{-\lambda} \cdot \frac{\mu^u}{u!} e^{-\mu},$$

$$\mathbf{P}(X + Y = t) = \frac{(\lambda + \mu)^t}{t!} e^{-(\lambda + \mu)}.$$

Таким образом, условная функция плотности, по которой строятся критерии проверки гипотез (A)–(D),

$$p^{Y|(X+Y)}(u|t) = C_t^u \left(\frac{\mu}{\lambda + \mu} \right)^u \left(\frac{\lambda}{\lambda + \mu} \right)^{t-u} \quad u = 0, 1, \dots, t.$$

Это функция плотности биномиального распределения с t испытаниями в схеме Бернулли и вероятностью $p = \mu/(\lambda + \mu)$ успешного исхода. Для проверки, например, гипотезы $H_0 : \mu \leq \lambda$ можно воспользоваться формулами примера 7.1, где параметр $\theta = \mu/(\lambda + \mu)$, а нулевая гипотеза $H_0 : \theta \leq \theta_0 = 0,5$.

В данном примере статистический эксперимент состоял в наблюдении двух случайных выборок каждая объема $n = 1$. Если выборки состоят из n и, соответственно, m наблюдений, то следует использовать достаточные статистики

$$X = \sum_{k=1}^n X_k, \quad \text{и} \quad Y = \sum_{k=1}^m Y_k,$$

которые распределены по закону Пуассона с параметрами $n\lambda$ и $m\mu$ соответственно.

ПРИМЕР 9.2. Критерий знаков. Чтобы определить, какой из двух товаров предпочитает потребитель, опрашивают n лиц. Результат отмечается знаком плюс, когда предпочитают товар A_1 , и знаком минус, если предпочитается товар A_2 . Полное число плюсов в результате опроса представляет

реализацию достаточной статистики T , имеющей биномиальное распределение с параметрами n и p . Рассмотрим проблему проверки гипотезы $p = 1/2$ об отсутствии разницы в спросе при альтернативе $p \neq 1/2$. РНМ несмещенный критерий, который в данном случае называется *критерием знаков*, отвергает нулевую гипотезу, если значение $|T - n/2|$ достаточно велико.

Допустим, что потребители могут воздержаться от ответа. Обозначим θ_1, θ_2 и θ_0 вероятности предпочесть A_1 , предпочесть A_2 и воздержаться соответственно; $\theta_1 + \theta_2 + \theta_0 = 1$. Тогда наблюдения статистик T_1, T_2, T_0 , $T_1 + T_2 + T_0 = n$, соответствуют количествам потребителей, для которых осуществляются указанные возможности. Эти статистики имеют полиномиальное распределение с функцией плотности

$$p(t_0, t_1, t_2 | \theta) = \frac{n!}{t_0! t_1! t_2!} \theta_0^{t_0} \theta_1^{t_1} \theta_2^{t_2}, \quad t_0 + t_1 + t_2 = n,$$

и гипотеза, которую надлежит проверить имеет вид $H_0 : \theta_1 = \theta_2$.

Представим совместную функцию плотности статистик T_0, T_1, T_2 в форме

$$\frac{n!}{t_0! t_1! t_2!} \left(\frac{\theta_1}{1 - \theta_0 - \theta_1} \right)^{t_1} \left(\frac{\theta_0}{1 - \theta_0 - \theta_1} \right)^{t_0} (1 - \theta_0 - \theta_1)^n,$$

из которой видно, что это – экспоненциальное семейство с

$$U = T_1, \quad T = T_0, \quad \theta = \ln \frac{\theta_1}{1 - \theta_0 - \theta_1}, \quad \vartheta = \ln \frac{\theta_0}{1 - \theta_0 - \theta_1}.$$

Гипотеза $H_0 : \theta_1 = \theta_2 = 1 - \theta_0 - \theta_1$ эквивалентна гипотезе $\theta = 0$.

Согласно теореме 9.3, для проверки этой гипотезы при альтернативе $H : \theta \neq 0$ существует РНМ несмещенный критерий, для построения которого требуется найти условное распределение T_1 относительно T_0 . При фиксированном значении $T_0 = t$ полиномиальное распределение превращается в биномиальное с параметрами $n - t$ и $p = \theta_1 / (\theta_1 + \theta_2)$ (докажите это!). Проблема проверки гипотезы $\theta_1 = \theta_2$ свелась к проблеме проверки гипотезы $p = 1/2$ в схеме испытаний Бернулли с объемом испытаний $n - t$ (напомним, t – число “ничьих”) и вероятностью успеха p . Критическая область такого критерия (с точностью до рандомизации) имеет вид

$$\left| T_1 - \frac{n - T_0}{2} \right| > C(T_0).$$

Таким образом, РНМ несмещенный критерий отбрасывает случаи, когда предпочтение не было высказано, и сводится к применению критерия знаков к оставшимся данным.

Мощность критерия сильно зависит от вероятности θ_0 , определяющей распределение статистики T_0 (число ничьих). Если θ_0 слишком велико, то мы вынуждены применять критерий знаков на основе достаточно скудного объема данных. В качестве альтернативного способа обращения с ничейными исходами иногда предлагают приписывать каждой ничьей знак (+) или (-) наудачу, с вероятностью $1/2$, а потом уже применять критерий знаков. Покажите, что тогда полное число плюсов T'_1 будет иметь биномиальное распределение с параметрами n и $p = \theta_1 + \theta_0/2$. Исходная гипотеза H_0 заменяется на гипотезу $p = 1/2$, которая отклоняется, когда $|T'_1 - n/2| > C$. Так как такой критерий включает рандомизацию не только на границе критической области, то он менее мощен, чем РНМ несмещенный критерий, игнорирующий ничьи. Следовательно, случайное “расщепление” ничьих приводит только к потере мощности и не рекомендуется к практическому использованию.

ПРИМЕР 9.3. *Проверка гипотез о параметрах нормального распределения.* При построении РНМ несмещенных критериев для параметров нормального (μ, σ^2) распределения основную роль играет независимость выборочного среднего \bar{X} и выборочной дисперсии S^2 , а также независимость статистики Стьюдента от \bar{X} и S^2 при определенном значении μ . Благодаря этому свойству нормального распределения РНМ несмещенные критерии не являются условными (теряют неймановскую структуру), так что константы, определяющие критерии, не зависят от значений каких-либо статистик и вычисляются до проведения наблюдений по заданному уровню значимости α . Чтобы убедиться в этом, напомним некоторые понятия и результаты из § 2.

Определение 2.4. Статистика $V = V(X)$ называется *подчиненной*, если ее распределение не зависит от параметра θ , индексирующего статистическую модель.

Статистическая модель выборки из нормального (μ, σ^2) распределения при фиксированном значении дисперсии σ^2 обладает полной достаточной статистикой \bar{X} . Распределение выборочной дисперсии S^2 не зависит от μ , поэтому S^2 – подчиненная статистика.

Независимость статистик \bar{X} и S^2 следует из следующего утверждения, носящего более общий характер.

Теорема 2.3. (теорема Басу) Если T – полная достаточная статистика для семейства \mathcal{P} , то любая подчиненная статистика V не зависит от T .

Как и в предыдущих примерах, для построения РНМ несмещенных критериев представим функцию плотности выборочного вектора нормального распределения в форме экспоненциального распределения (9.8):

$$\begin{aligned} p(x | \mu, \sigma^2) &= \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} = \\ &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i \right\}. \end{aligned} \quad (9.18)$$

При построении РНМ критериев в задачах (А)–(D) для параметра σ^2 сравнение (9.8) и (9.18) показывает, что

$$\theta = -\frac{1}{2\sigma^2}, \quad U(X) = \sum_{i=1}^n X_i^2; \quad \vartheta = \frac{n\mu}{\sigma^2}, \quad T(X) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Рассмотрим, например, задачу (А), остальные решаются аналогично. Нулевая гипотеза $H_0 : \sigma^2 \leq \sigma_0^2$ в терминах параметра θ имеет аналогичный вид $H_0 : \theta \leq \theta_0$. В силу непрерывности распределения достаточных статистик РНМ критерий нерандомизирован и имеет критическую область (см. (9.9))

$$U(X) = \sum_{i=1}^n X_i^2 > C(\bar{X}),$$

где критическая константа $C(\bar{x})$ определяется из уравнения

$$\mathbf{P}_{\sigma_0} \left\{ \sum_{i=1}^n X_i^2 > C(\bar{X}) \mid \bar{X} = \bar{x} \right\} = \alpha.$$

Это уравнение можно переписать в виде

$$\mathbf{P}_{\sigma_0} \left\{ \sum_{i=1}^n X_i^2 - n \bar{X}^2 > C'(\bar{X}) \mid \bar{X} = \bar{x} \right\} = \alpha,$$

где $C'(\bar{x}) = C(\bar{x}) - n \bar{x}^2$ такая же, по сути дела, критическая константа, как и $C(\bar{x})$. Поскольку

$$\sum_{i=1}^n X_i^2 - n \bar{X}^2 = nS^2,$$

а S^2 не зависит от \bar{X} , то и критическая константа $C'(\bar{x})$ не зависит от \bar{x} и определяется известным нам из общего курса способом с помощью хи-квадрат распределения с $(n - 1)$ -й степенью свободы:

$$\mathbf{P}_{\sigma_0} \left(\sum_{i=1}^n (X_i - \bar{X})^2 > C \right) = \mathbf{P}_{\sigma_0} \left(\frac{nS^2}{\sigma_0^2} > \frac{C}{\sigma_0^2} \right) = 1 - K(C/\sigma_0^2) = \alpha,$$

откуда $C = C(\alpha) = \sigma_0^2 K^{-1}(1 - \alpha)$.

Таким образом, знакомый нам из общего курса критерий, тестирующий значение дисперсии нормального распределения, есть РНМ несмещенный критерий.

Покажем теперь, что и критерий Стьюдента для проверки гипотезы $H_0 : \mu \leq \mu_0$ при альтернативе $H_1 : \mu > \mu_0$ также является РНМ несмещенным критерием. Поскольку μ – параметр сдвига, то можно в целях простоты выкладок рассмотреть сначала случай $\mu_0 = 0$, а потом просто сделать замену $x \rightarrow x + \mu_0$.

Функция плотности (9.18) превращается в (9.8), если положить

$$\theta = \frac{n\mu}{\sigma^2}, \quad U(X) = \bar{X}; \quad \vartheta = -\frac{1}{2\sigma^2}, \quad T(X) = \sum_{i=1}^n X_i^2.$$

РНМ несмещенный критерий является нерандомизированным и определяется критической областью $U(X) = \bar{X} > C(T)$, где критическая константа $C(T)$ определяется из уравнения (напомним, $\mu_0 = 0$)

$$\mathbf{P}_{\mu_0} \{ \bar{X} > C(T) \mid T = t \} = \alpha.$$

Простым исследованием производной легко убедиться, что функция $y = x/\sqrt{t - nx^2}$ есть возрастающая функция аргумента x . Следовательно,

уравнение для определения критической константы можно переписать в виде

$$\mathbf{P}_{\mu_0} \left\{ \frac{\bar{X}}{\sqrt{T - n\bar{X}^2}} \sqrt{n-1} > C'(T) \mid T = t \right\} = \alpha.$$

Распределение статистики Стьюдента

$$T_{n-1} = \frac{\bar{X}}{\sqrt{T - n\bar{X}^2}} \sqrt{n-1},$$

когда выборка берется из нормального $(0, \sigma^2)$ распределения, не зависит от дисперсии σ^2 , то есть это подчиненная статистика. В силу Теоремы 2.3 она не зависит от достаточной статистики T . Поэтому критическая константа $C'(T)$ не зависит от T и определяется посредством квантили распределения Стьюдента с $(n-1)$ -й степенью свободы.

ЗАЧЕТНЫЕ ЗАДАНИЯ

- 1*. По аналогии с примером 9.1 постройте РНМ несмещенный критерий проверки гипотезы о равенстве вероятностей успешного испытания в двух независимых схемах Бернулли.
- 2*. Пусть X_1, \dots, X_n – случайная выборка из распределения с плотностью $f(x \mid \theta, \vartheta) = \theta^{-1} \exp\{-(x-\vartheta)/\theta\}$, $x > \vartheta$. Покажите, что для проверки гипотезы $\theta = 1$ существует РНМ несмещенный критерий с областью принятия

$$C_1 \leq 2 \sum_{k=1}^n (X_k - X_{(1)}) \leq C_2.$$

- 3*. Пусть X_1, \dots, X_n – случайная выборка из распределения с плотностью $f(x \mid \theta, \vartheta) = \vartheta^{-1} \exp\{-(x-\theta)/\vartheta\}$, $x > \theta$. Покажите, что для проверки гипотезы $\theta = 0$ существует РНМ несмещенный критерий с областью принятия

$$0 \leq n \frac{X_{(1)}}{\sum_{k=1}^n (X_k - X_{(1)})} \leq C.$$

§ 10. Инвариантный статистический вывод

10.1. Группы преобразований выборочного пространства; максимальные инварианты. Существует достаточно обширный класс проблем статистического вывода, где определенное преобразование выборочных данных не должно приводить к принятию решения, которое бы противоречило решению, принятому на основе исходных данных. Например, если выборочные данные представляют замеры количества вредных примесей в дизельном топливе, то решение о его кондиционности не должно зависеть от того, в каких единицах производятся замеры – в граммах, миллиграммах или в процентом отношении к весу всей партии топлива. Следовательно, статистический вывод должен обладать свойством инвариантности по отношению к масштабным преобразованиям выборочных данных. Если статистическая проблема состоит в оценке функции распределения (типичная задача при оценке надежности или принятия решения о гарантийном сроке службы выпускаемых изделий), то оценка не должна зависеть от того, в каком порядке поступают выборочные данные, так что статистический вывод должен обладать свойством инвариантности относительно перестановок компонент выборки. Таких примеров с требованием инвариантности статистического вывода относительно выбора системы координат при проведении наблюдений, инвариантности относительно ортогональных преобразований и тому подобное можно привести сколько угодно.

Общим математическим выражением такого рода требованию “беспристрастности” статистического вывода является его инвариантность относительно подходящей группы преобразований выборочного пространства. Напомним некоторые определения и термины из теории групп.

Определение 10.1. Множество G элементов называется *группой*, если выполняются следующие условия.

- (A) Определена операция *группового умножения*, которая любым двум элементам $a, b \in G$ ставит в соответствие элемент $c \in G$. Элемент c называется *произведением* элементов a и b и обозначается ab .

(B) Групповое умножение *ассоциативно*: $(ab)c = a(bc)$.

(C) Существует элемент $e \in G$, называемый *единицей*, такой, что $ae = ea$ для всех $a \in G$.

(D) Для каждого элемента $a \in G$ существует ему *обратный* элемент $a^{-1} \in G$ такой, что $aa^{-1} = a^{-1}a = e$.

Если элементы G являются отображением некоторого пространства X на себя, причем групповое произведение g_2g_1 определяется как результат выполнения преобразования с помощью элемента g_1 и последующего применения элемента g_2 , то G называется *группой преобразований*.

Точка $x \in X$ пробегает *траекторию*, когда к ней применяются все преобразования g из G ; это означает, что траектория, содержащая точку x , состоит из множества точек $\{gx : g \in G\}$. Множество траекторий группы G образует *разбиение* пространства X .

Функция $\varphi(x)$, $x \in X$, называется *инвариантной относительно группы G* , если $\varphi(gx) = \varphi(x)$ для всех $x \in X$ и всех $g \in G$. Из этого определения следует, что функция инвариантна тогда и только тогда, когда она постоянна на каждой траектории.

Функция $T = T(x)$, $x \in X$, называется *максимальным инвариантом*, если она инвариантна и если выполняется условие: из равенства $T(x_1) = T(x_2)$ следует, что $x_2 = gx_1$ для некоторого $g \in G$, то есть если $T(\cdot)$ постоянна на каждой траектории и на различных траекториях принимает различные значения.

Лемма 10.1. Пусть T – максимальный инвариант относительно группы G . Тогда необходимое и достаточное условие для инвариантности функции φ состоит в том, чтобы φ зависела от $x \in X$ только через $T(x)$, то есть чтобы существовала такая функция $h(t)$, $t \in \mathcal{T}$, что $\varphi(x) = h(T(x))$ для всех $x \in X$.

Доказательство. Необходимость. Если $\varphi(x) = h(T(x))$ для всех $x \in X$, тогда $\varphi(gx) = h(T(gx)) = \varphi(x)$, то есть φ инвариантна.

Достаточность. Зависимость φ от x только через $T(x)$ означает, что функция φ постоянна на тех подмножествах X , на которых постоянна

T . Однако, если φ инвариантна и если $T(x_1) = T(x_2)$, то $x_2 = gx_1$ для некоторого $g \in G$, и поэтому $\varphi(x_2) = \varphi(gx_1) = \varphi(x_1)$ в силу инвариантности функции φ . \square

Рассмотрим несколько примеров на построение максимальных инвариантов для групп преобразований выборочного пространства \mathcal{X} – пространства значений случайной выборки $X = (X_1, \dots, X_n)$ фиксированного объема n .

ПРИМЕР 10.1. Группа сдвигов. Каждый элемент g группы сдвигов G определяется выбором числа $a \in \mathbb{R}$, которое прибавляется к каждой компоненте $x = (x_1, \dots, x_n)$ выборочного вектора: $gx = x + a = (x_1 + a, \dots, x_n + a)$. Вектор разностей $T(x) = (x_1 - x_n, \dots, x_{n-1} - x_n)$ размерности $n - 1$ инвариантен относительно группы сдвигов и является максимальным инвариантом. Чтобы доказать это, надо найти такое число c , что $x' = x + c$, когда $T(x) = T(x')$. Пусть $x_i - x_n = x'_i - x'_n$, $i = 1, \dots, n - 1$. Выбирая $c = x'_n - x_n$, получаем, что $x'_i = x_i + c$ для всех $i = 1, \dots, n - 1$, что и требовалось доказать.

Естественно, существуют и другие максимальные инварианты, например, $T(x) = (x_2 - x_1, \dots, x_n - x_1)$. В частном случае, когда $n = 1$, не существует нетривиальных инвариантных функций. Все пространство состоит из одной траектории, так что любые две точки могут быть переведены одна в другую выбором сдвига $g \in G$. В таком случае группа преобразований называется *транзитивной*. Инвариантными функциями могут быть только константы $\varphi(x) = c$.

ПРИМЕР 10.2. Группа масштабных преобразований. Каждый элемент g группы G определяется выбором числа $a \neq 0$, на которое умножается каждая компонента выборочного вектора: $gx = ax = (ax_1, \dots, ax_n)$. В предположении, что начало координат имеет нулевую вероятность и, тем самым, может быть исключено из рассмотрения, максимальным инвариантом группы масштабных преобразований является функция $T(x) = (x_1/x_n, \dots, x_{n-1}/x_n)$. Докажите это по аналогии с примером 10.1.

ПРИМЕР 10.3. Группа линейных преобразований. Каждый элемент g этой группы определяется выбором двух чисел: $a \neq 0$ и $b \in \mathbb{R}$. Выборочный

вектор x преобразуется в $gx = ax + b = (ax_1 + b, \dots, ax_n + b)$, Докажите, что инвариантная функция

$$T(x) = \frac{x_2 - x_1}{x_n - x_1}, \dots, \frac{x_{n-1} - x_1}{x_n - x_1}$$

является максимальным инвариантом.

ПРИМЕР 10.4. Группа ортогональных преобразований. Пусть $x = (x_1, \dots, x_n)$ выборочный вектор m -мерного распределения, то есть каждая компонента x_i выборки x является вектором $x_i = (x_{i1}, \dots, x_{im})$. Если G – группа ортогональных преобразований \mathbf{R}^m , то максимальным инвариантом будет функция

$$T(x) = \left(\sum_{k=1}^m x_{1k}^2, \dots, \sum_{k=1}^m x_{nk}^2 \right).$$

10.2. Инвариантная статистическая проблема. Обратимся снова к рассуждениям о необходимости введения инвариантных статистических правил, которые были приведены в начале предыдущего пункта. Легко понять, что эти соображения должны быть положены в построение вероятностной модели, выбор функции потерь и повлечь ограничения на класс правил принятия решения.

Группа G оставляет статистическую проблему инвариантной, если выполняются следующие условия.

(I) Она оставляет инвариантной статистическую модель $\mathcal{P} = \{P(\cdot, |\theta), \theta \in \Theta\}$, то есть для любого возможного распределения $P(A|\theta)$, $A \in \mathcal{A}$, случайной выборки X распределение gX , скажем, $P(\cdot|\theta')$, также принадлежит \mathcal{P} . Точка θ' параметрического пространства Θ , связанная с θ указанным образом, будет обозначаться $\bar{g}\theta$, так что $\mathbf{P}(gX \in A|\theta) = \mathbf{P}(X \in A|\bar{g}\theta)$. Более общим образом это условие можно сформулировать в терминах математического ожидания от любой статистики T :

$$\mathbf{E}_\theta T(gX) = \mathbf{E}_{\bar{g}\theta} T(X). \quad (10.1)$$

Преобразования \bar{g} образуют группу \bar{G} преобразований параметрического пространства Θ . Параметрическое пространство остается инвариантным относительно группы \bar{G} , если $\bar{g}\theta \in \Theta$ для всех $\theta \in \Theta$ и всех $\bar{g} \in \bar{G}$

и если, дополнительно, для каждого $\theta' \in \Theta$ существует такое $\theta \in \Theta$, что $\bar{g}\theta = \theta'$. Эти два условия можно выразить в форме равенства $\bar{g}\Theta = \Theta$.

(II) Для каждого $g \in G$ существует такое отображение g^* пространства решений \mathcal{D} на себя, что функция потерь не меняется при этом отображении, то есть $L(\bar{g}\theta, g^*d) = L(\theta, d)$. Это условие порождает ограничение на выбор правил принятия решения, который в терминах решающих функций $\delta(x)$ принимает вид

$$\delta(gx) = g^* \delta(x) \quad \text{для всех } x \in X \text{ и } g \in G. \quad (10.2)$$

Таким образом, группа G преобразований выборочного пространства X порождает группы преобразований \bar{G} и G^* параметрического пространства Θ и пространства решений \mathcal{D} соответственно.

При этих предположениях преобразованная задача в терминах $X' = gX$, $\theta' = \bar{g}\theta$ и $d' = g^*d$ формально идентична первоначальной задаче в терминах X , θ и d . Решающая функция δ первоначальной задачи остается пригодной и после преобразования. Интерпретируя преобразование как замену координатной системы, мы выбрали бы, наблюдая x' , решение, которое в новой системе записывается как $\delta(x')$, а в старой – как $g^{*-1}\delta(x')$. Если принимаемое решение не должно зависеть от выбора системы координат, то последнее решение должно совпадать с первоначальным δ , то есть решающее правило должно удовлетворять условию *инвариантности*, которое в терминах решающих функций записывается как соотношение (10.2).

Одно существенное замечание терминологического характера. Поскольку решающая функция, как таковая, не является инвариантной функцией ($\delta(gx) \neq \delta(x)$, вообще говоря), то δ , которая удовлетворяет равенству (10.2), называют *эквивариантной* решающей функцией.

Теорема 10.1. *Если δ – эквивариантная решающая функция в задаче, которая инвариантна относительно группы преобразований G , то функция риска соответствующего δ правила φ принятия решения удовлетворяет равенству*

$$R(\varphi | \bar{g}\theta) = R(\varphi | \theta) \quad \text{для всех } \theta \in \Theta. \quad (10.3)$$

Доказательство. По определению функции риска

$$R(\varphi | \bar{g}\theta) = E_{\bar{g}\theta} L(\bar{g}\theta, \delta(X)). \quad (10.4)$$

Из (10.1) следует, что правая часть (10.4) равна

$$\mathbf{E}_\theta L(\bar{g}\theta, \delta(gX)) = \mathbf{E}_\theta L(\bar{g}\theta, g^*\delta(X)) = \mathbf{E}_\theta L(\theta, \delta(X)) = R(\varphi | \theta). \quad \square$$

Следствие 10.1. В предположениях теоремы 10.1, если \bar{G} есть транзитивная группа преобразований параметрического пространства Θ , то функция риска постоянна (не зависит от $\theta \in \Theta$).

Доказательство. По определению транзитивности для любых двух точек $\theta', \theta \in \Theta$ существует такое $\bar{g} \in \bar{G}$, что $\theta' = \bar{g}\theta$. В силу формулы (10.3)

$$R(\varphi | \theta') = R(\varphi | \bar{g}\theta) = R(\varphi | \theta) \quad \text{для всех } \theta', \theta \in \Theta,$$

то есть $R(\varphi | \theta)$ не зависит от $\theta \in \Theta$. □

Это свойство эквивариантных решающих функций позволяет найти общий вид эквивариантной оценки для параметра сдвига или масштаба, которые доставляют минимум функции риска.

10.3. Эквивариантная оценка параметра сдвига с минимальным риском. Метод построения таких оценок (в дальнейшем коротко - ЭМР оценок) аналогичен тому, что использовался в § 6 при построении несмещенных оценок с равномерно минимальной функцией риска, но в задаче оптимальной эквивариантной оценки не требуется выпуклости функции потерь по аргументу $d \in \mathcal{D}$. Общая теория эквивариантного оценивания содержится в главе 3 книги Э.Лемана “Теория точечного оценивания”, куда и отсылается любознательный студент. Мы рассмотрим только частную проблему ЭМР нерандомизированной оценки параметра сдвига, в которой функция плотности по мере Лебега случайной выборки $X = (X_1, \dots, X_n)$ зависит от параметра сдвига θ в форме

$$f_n(x - \theta) = \prod_{i=1}^n f(x_i - \theta), \quad x \in \mathcal{X} = \mathbb{R}^n, \quad \theta \in \Theta = \mathbb{R},$$

и функция потерь имеет вид $L(d - \theta)$.

Поскольку группа сдвигов является транзитивной, то в силу следствия 10.1 функция риска любой ЭМР оценки не зависит от θ . Общий вид такой оценки устанавливает

Лемма 10.2. Пусть $\hat{\theta}_0(X)$ – некоторая (любая) эквивариантная оценка. Для того, чтобы оценка $\hat{\theta}(X)$ была эквивариантной необходимо и достаточно, чтобы существовала такая инвариантная функция $U(x)$, $x \in \mathcal{X}$, что $\hat{\theta}(X) = \hat{\theta}_0(X) - U(X)$.

Доказательство. Достаточность. Если существует функция U , определяющая оценку $\hat{\theta}(X)$, то

$$\hat{\theta}(X + a) = \hat{\theta}_0(X + a) - U(X + a) = \hat{\theta}(X) = \hat{\theta}_0(X) - U(X)$$

при любых $a \in \mathbb{R}^n$. Следовательно, $\hat{\theta}(X)$ – эквивариантная оценка.

Необходимость. Если $\hat{\theta}(X)$ – эквивариантная оценка, то положим $U(x) = \hat{\theta}_0(x) - \hat{\theta}(x)$. Тогда

$$U(x + a) = \hat{\theta}_0(x + a) - \hat{\theta}(x + a) = \hat{\theta}_0(x) + a - \hat{\theta}(x) - a = U(x),$$

так что $U(x)$ есть искомая инвариантная функция. □

Замечание 10.1. В силу Леммы 10.1 функция $U(x)$ зависит от $x = (x_1, \dots, x_n)$ только через значения максимального инварианта $T(x) = (x_1 - x_n, \dots, x_{n-1} - x_n)$, когда $n \geq 2$. При $n = 1$ функция $U(x) = \text{const}$. Таким образом, $U(x) = V(T(x))$.

Метод вычисления ЭМР дает

Теорема 10.2. Пусть $\hat{\theta}_0(X)$ – некоторая эквивариантная оценка с конечным риском. Допустим, что для каждого значения $t = T(x)$ максимального инварианта $T(X)$ существует число $V(t) = V^*(t)$, доставляющая минимум условному математическому ожиданию

$$\mathbf{E} \left\{ L \left(\hat{\theta}_0(X) - V(t) \right) \mid T(X) = t \right\}.$$

Тогда эквивариантная относительно сдвига нерандомизированная оценка $\hat{\theta}^*(X)$ для параметра θ с минимальным риском существует и вычисляется по формуле

$$\hat{\theta}^*(X) = \hat{\theta}_0(X) - V^*(T(X)).$$

Доказательство. В силу Леммы 10.2 и замечания 10.1 ЭМР оценку можно найти, определив функцию $V(\cdot)$, которая минимизирует функцию риска

$$R(\widehat{\theta} | \theta) = \mathbf{E}_\theta L \left(\widehat{\theta}_0(X) - V(t) - \theta \right).$$

Так как в силу Следствия 10.1 риск эквивариантной оценки не зависит от θ , то достаточно минимизировать математическое ожидание при $\theta = 0$:

$$R(\widehat{\theta} | 0) = \mathbf{E}_0 L \left(\widehat{\theta}_0(X) - V(t) \right) = \int_{\mathcal{T}} \mathbf{E}_0 \left\{ L \left(\widehat{\theta}_0(X) - V(t) \right) \mid T(X) = t \right\} dW_0(t),$$

где $(\mathcal{T}, \mathcal{C})$ – измеримое пространство значений максимального инварианта T , а W_θ – распределение вероятностей на сигма-алгебре \mathcal{C} . Здесь мы воспользовались известным свойством условного математического ожидания: “среднее от условного среднего равно безусловному среднему”. Как и в построении байесовского решения минимизация математического ожидания эквивалента минимизации условного математического ожидания, которое в силу условия теоремы конечно, так что минимизация имеет смысл. \square

Для некоторых наиболее употребительных функций потерь ЭВМ достаточно просто выражается через характеристики распределения условного распределения X относительно $T(X)$.

Следствие 10.2. Если $L(d - \theta) = (d - \theta)^2$, то

$$V^*(T(X)) = \mathbf{E} \left\{ \widehat{\theta}_0(X) \mid T(X) \right\},$$

а если $L(d - \theta) = |d - \theta|$, то $V^*(T(X))$ есть любая медиана условного распределения $\widehat{\theta}_0(X)$ относительно $T(X)$.

Доказательство этого утверждения предоставляется читателям в качестве упражнения. \square

Найдем ЭМР при квадратичной функции потерь.

Теорема 10.3. Пусть наблюдаемая случайная величина ξ имеет конечное математическое ожидание $\mathbf{E}_0 \xi$. При выполнении условий Теоремы

10.2 с $L(d - \theta) = (d - \theta)^2$ ЭМР оценка θ вычисляется по формуле

$$\hat{\theta}^*(X) = \frac{\int_{-\infty}^{\infty} u \prod_{i=1}^n f(X_i - u) du}{\int_{-\infty}^{\infty} \prod_{i=1}^n f(X_i - u) du} \quad (10.5)$$

и в этой форме известна как оценка Питмена для θ .

Доказательство. Согласно Теореме 10.2 ЭМР оценка при квадратичной функции потерь

$$\hat{\theta}^*(X) = \hat{\theta}_0(X) - \mathbf{E}_0 \left\{ \hat{\theta}_0(X) \mid T(X) \right\},$$

где $\hat{\theta}_0$ –любая эквивариантная оценка. Положим $\hat{\theta}_0(X) = X_n$. Чтобы вычислить условное математическое ожидание $\mathbf{E}_0 \{ X_n \mid T(X) \}$, найдем условную функцию плотности X_n относительно $T(X)$.

В общем курсе теории вероятностей условная функция случайной величины Y относительно случайной величины Z определялась как случайная функция

$$f^{Y|Z}(y | Z) = \frac{f^{Y,Z}(y, Z)}{f^Z(Z)},$$

где $f^{Y,Z}(y, z)$ – функция плотности совместного распределения Y и Z , а $f^Z(z)$ – маргинальная функция распределения Z . Применим это определение к вычислению условной функции плотности X_n относительно $T(X)$, полагая $Z = T(X)$ и $Y = X_n$.

Так как

$$X_n = x, \quad X_1 - X_n = t_1, \dots, X_{n-1} - X_n = t_{n-1}$$

влечет

$$X_1 = x + t_1, \dots, X_{n-1} = x + t_{n-1}, \quad X_n = x,$$

то совместная функция плотности X_n и $T(X)$

$$w^{X_n, T}(x, t_1, \dots, t_n) = f_n(x + t_1, \dots, x + t_{n-1}, x) = f(x) \prod_{i=1}^{n-1} f(t_i + x).$$

Функция распределения статистики $T(X) = (X_1 - X_n, \dots, X_{n-1} - X_n)$

$$W^T(t_1, \dots, t_{n-1}) = \mathbf{P}(X_1 < t_1 + X_n, \dots, X_{n-1} < t_{n-1} + X_n) =$$

$$\int_{-\infty}^{\infty} f(x_n) dx_n \int_0^{t_1+x_n} f(x_1) dx_1 \cdots \int_{-\infty}^{t_{n-1}+x_n} f(x_{n-1}) dx_{n-1},$$

а ее функция плотности

$$w^T(t_1, \dots, t_{n-1}) = \frac{\partial^{n-1} W^T}{\partial t_1 \dots \partial t_{n-1}} = \int_0^{\infty} f(x) \prod_{i=1}^{n-1} f(t_i + x) dx.$$

Таким образом, условная функция плотности X_n относительно $T(X)$

$$w^{X_n|T}(x|T(X)) = \frac{f(x) \prod_{i=1}^{n-1} f(X_i - X_n + x)}{\int_{-\infty}^{\infty} f(x) \prod_{i=1}^{n-1} f(X_i - X_n + x) dx}$$

и условное математическое ожидание X_n относительно $T(X)$

$$\begin{aligned} \mathbf{E}_0\{X_n|T(X)\} &= \int_{-\infty}^{\infty} x w^{X_n|T}(x|T(X)) dx = \\ &= \frac{\int_{-\infty}^{\infty} x f(x) \prod_{i=1}^{n-1} f(X_i - X_n + x) dx}{\int_{-\infty}^{\infty} f(x) \prod_{i=1}^{n-1} f(X_i - X_n + x) dx}. \end{aligned}$$

Замена $u = X_n - x$ дает запись условного математического ожидания, из которого следует утверждение теорем:

$$\mathbf{E}_0\{X_n|T(X)\} = X_n - \frac{\int_{-\infty}^{\infty} u \prod_{i=1}^n f(X_i - u) du}{\int_{-\infty}^{\infty} \prod_{i=1}^n f(X_i - u) du}. \quad \square$$

ПРИМЕР 10.5. ЭМР оценка среднего значения θ нормального распределения $N(\theta, \sigma^2)$. В статистическом эксперименте наблюдаются независимые копии случайной величины ξ с функцией плотности

$$f(x - \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \theta)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}, \quad \theta \in \mathbb{R}.$$

Произведение плотностей, стоящее в числителе и знаменателе ЭМР оценки (10.5)

$$\prod_{i=1}^n f(X_i - u) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ - \sum_{i=1}^n \frac{(X_i - u)^2}{2\sigma^2} \right\} =$$

$$\frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ - \frac{nS^2}{2\sigma^2} \right\} \exp \left\{ - \frac{n}{2\sigma^2} (u - \bar{X})^2 \right\},$$

где

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 -$$

выборочное среднее и выборочная дисперсия соответственно.

Вычисление интегралов в (10.5) дает ЭМР оценку $\hat{\theta}^*(X) = \bar{X}$. Таким образом, выборочное среднее \bar{X} как оценка среднего значения θ нормального распределения $N(\theta, \sigma^2)$ является не только минимаксной оценкой и несмещенной оценкой с равномерно минимальным риском, но и эквивалентной оценкой с минимальным риском, каково бы ни было значение дисперсии σ^2 .

ПРИМЕР 10.6. ЭМР оценка параметра положения θ равномерного $U(\theta - b, \theta + b)$ распределения. В статистическом эксперименте наблюдаются независимые копии случайной величины ξ с функцией плотности

$$f(x - \theta) = \frac{I(-b \leq x - \theta) I(b \geq x - \theta)}{2b}, \quad x \in \mathbb{R}, \quad \theta \in \mathbb{R},$$

где $I(A)$ – индикаторная функция события $A \in \mathcal{A}$.

Произведение плотностей, стоящее в числителе и знаменателе ЭМР оценки (10.5)

$$\prod_{i=1}^n f(X_i - u) = \frac{I(-b \leq X_{(1)} - u) I(b \geq X_{(n)} - u)}{(2b)^n},$$

где $X_{(1)} = \min\{X_1, \dots, X_n\}$, $X_{(n)} = \max\{X_1, \dots, X_n\}$ – крайние члены вариационного ряда. Следовательно, ЭМР θ

$$\hat{\theta}^*(X) = \frac{\int_{X_{(n)}-b}^{X_{(1)}+b} u \, du}{\int_{X_{(n)}-b}^{X_{(1)}+b} du} = \frac{X_{(1)} + X_{(n)}}{2}.$$

10.4. Наиболее мощные инвариантные критерии. Обратимся к построению инвариантных критериев проверки гипотез при наличии мешающего параметра ϑ , когда распределение наблюдаемого случайного элемента ξ зависит от ϑ “групповым образом”. Например, ξ есть действительная случайная величина, функция распределения которой имеет вид $F_\lambda(bx + a, \cdot)$, $b > 0$, $a \in \mathbb{R}$. Относительно значения параметра λ проверяется некоторая гипотеза при мешающем параметре $\theta = (a, b)$, который определяет группу линейных преобразований выборочного пространства \mathcal{X} .

Для общей группы преобразований G задача проверки гипотезы $H_0 : \theta \in \Theta_0$ при альтернативе $H_1 : \theta \in \Theta_1$ остается инвариантной относительно G , если элементы \bar{g} группы \bar{G} сохраняют гипотетические подпространства Θ_0 и Θ_1 , то есть в дополнение к $\bar{g}\Theta = \Theta$ требуется, чтобы и $\bar{g}\Theta_0 = \Theta_0$.

Критерий φ , удовлетворяющий равенствам $\varphi(gx) = \varphi(x)$ для всех $x \in \mathcal{X}$ и всех $g \in G$ называется *инвариантным относительно группы G* . В силу Леммы 10.1 любой инвариантный критерий зависит от $x \in \mathcal{X}$ только через значения $T(x)$ максимального инварианта T . Таким образом, при построении наиболее мощного критерия с помощью леммы Неймана–Пирсона мы сталкиваемся, в первую очередь, с необходимостью вывода распределений максимальных инвариантов. Рассмотрим несколько примеров на решение данной задачи.

ПРИМЕР 10.7. *Различение двух простых гипотез с мешающим масштабным параметром.* Проверяется простая гипотеза H_0 : выборка $X = (X_1, \dots, X_n)$ берется из непрерывного распределения с функцией плотности $f_0 = \theta f_0(\theta x)$, $x > 0$, $\theta > 0$, при альтернативе H_1 , которая предполагает, что распределение каждого X_i , $i = 1, \dots, n$, имеет плотность $f_1 = \theta f_1(\theta x)$ с тем же носителем \mathbb{R}_+ и таким же пространством значений мешающего параметра θ . Как было отмечено в Примере 10.2, максимальным инвариантом относительно группы масштабных преобразований может служить статистика

$$T(X) = \left(\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n} \right).$$

Найдем распределение этой $(n - 1)$ -мерной статистики, предполагая, что

выбор происходит из распределения, сосредоточенного на положительной полуоси с плотностью $f(x)$, $x > 0$.

Функция распределения статистики T

$$H(t_1, \dots, t_{n-1}) = \mathbf{P}(X_1 < t_1 X_n, \dots, X_{n-1} < t_{n-1} X_n) = \int_0^\infty f(x_n) dx_n \int_0^{t_1 x_n} f(x_1) dx_1 \cdots \int_0^{t_{n-1} x_n} f(x_{n-1}) dx_{n-1};$$

ее функция плотности

$$h(t_1, \dots, t_{n-1}) = \frac{\partial^{n-1} H}{\partial t_1 \dots \partial t_{n-1}} = \int_0^\infty x^{n-1} f(x) \prod_{i=1}^{n-1} f(t_i x) dx.$$

Критерий Неймана–Пирсона для проверки H_0 при альтернативе H_1 основан на статистике отношения правдоподобия. Правдоподобие H_i , соответствующее наблюдению статистики T , равно

$$L_i(T) = h\left(\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}\right) = \int_0^\infty x^{n-1} f_i(x) \prod_{k=1}^{n-1} f_i\left(\frac{X_k}{X_n} x\right) dx, \quad i = 0, 1.$$

Замена $x = t X_n$ приводит правдоподобие H_i к виду

$$L_i(T) = X_n^n \int_0^\infty t^{n-1} \prod_{k=1}^n f_i(X_k t) dt, \quad i = 0, 1. \quad (10.6)$$

В силу непрерывности распределения, из которого берется выборка, наиболее мощный критерий не рандомизирован и имеет критическую область $L_1(T)/L_0(T) > C$, где постоянная $C = C(\alpha)$ определяется по заданному уровню значимости α из уравнения

$$\mathbf{P}_0\left(\frac{L_1(T)}{L_0(T)} > C\right) = \alpha.$$

Если отношение правдоподобия $L_1(T)/L_0(T)$ является монотонной функцией некоторой статистики $U(T)$, а гипотезы H_0 и H_1 вида (А)–(D) (см. пункт 9.2) составляют утверждения о параметре λ распределения с функцией плотности $\theta f_\lambda(\theta x)$, то методы § 9 позволяют строить РНМ инвариантные несмещенные критерии проверки таких гипотез.

ПРИМЕР 10.8. Проверка гипотезы о дисперсии нормального распределения (максимальные инварианты при наличии достаточных статистик). Если статистическая модель случайной выборки $X = (X_1, \dots, X_n)$ обладает достаточной статистикой, то статистический вывод осуществляется с использованием только семейства распределений этой статистики. Так, если X – выборка из нормального (θ, σ^2) распределения, то для проверки гипотезы $H_0 : \sigma \leq \sigma_0$ при альтернативе $H_1 : \sigma > \sigma_0$ следует опираться только на статистики $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, $S^2 = \sum_{k=1}^n (X_k - \bar{X})^2$.

Гипотеза H_0 остается инвариантной относительно группы сдвигов $X'_k = X_k + a$, $k = 1, \dots, n$, $a \in \mathbb{R}$. В терминах достаточных статистик \bar{X} и S^2 это преобразование переходит в $\bar{X}' = \bar{X} + a$, $S'^2 = S^2$, и максимальным инвариантом является статистика S^2 . Поэтому все инвариантные критерии должны зависеть только от S^2 . Как следует из результатов § 9, существует РНМ несмещенный инвариантный критерий, отклоняющий гипотезу H_0 , когда $S^2 > C$, и этот критерий совпадает с РНМ несмещенным критерием (см. пример (9.3)).

Рассмотрим еще одну задачу на построение РНМ инвариантного критерия, представляющую особый интерес при тестировании вероятностных моделей в испытаниях на долговечность (теория надежности).

ПРИМЕР 10.9. РНМ инвариантный критерий проверки вероятностной модели “отсутствие последствия” в испытаниях на долговечность. Пусть выборка $X = (X_1, \dots, X_n)$ берется из гамма распределения с плотностью

$$\frac{1}{\theta} f_{\lambda} \left(\frac{x}{\theta} \right) = \frac{1}{\theta^{\lambda} \Gamma(\lambda)} x^{\lambda-1} \exp \left\{ -\frac{x}{\theta} \right\}, \quad x > 0; \theta > 0, \lambda > 0.$$

Проверяется гипотеза $H_0 : \lambda = 1$ при альтернативе $H_1 : \lambda > 1$. В терминах испытаний на долговечность, когда выборочные данные представляют сроки службы n испытуемых объектов, гипотеза H_0 соответствует вероятностной модели “отсутствие последствия” (показательное распределение долговечности), а альтернатива H_1 – вероятностной модели износа (старения) объекта при воздействии пиковых нагрузок.

Гипотеза H_0 инвариантна относительно группы масштабных преобразований, поэтому ее проверка должна осуществляться на основе максимального инварианта

$$T(X) = \left(\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n} \right),$$

функция правдоподобия которого (см. (10.6),

$$L(\lambda | T) = X_n^n \int_0^\infty t^{n-1} \prod_{k=1}^n f_\lambda(X_k t) dt = \\ \frac{X_n^n}{\Gamma^n(\lambda)} \prod_{k=1}^n X_k^{\lambda-1} \int_0^\infty t^{n\lambda-1} \exp \left\{ -t \sum_{k=1}^n X_k \right\} dt = X_n^n \frac{\Gamma(n\lambda)}{\Gamma^n(\lambda)} \frac{\prod_{k=1}^n X_k^{\lambda-1}}{(\sum_{k=1}^n X_k)^{n\lambda}}.$$

Нетрудно заметить, что логарифмическое отношение правдоподобия $\ln L(\lambda'' | T) - \ln L(\lambda' | T)$ при $\lambda'' > \lambda'$ есть монотонно возрастающая функция статистики

$$U(T) = \sum_{k=1}^n \ln X_k - n \ln \left(\sum_{k=1}^n X_k \right),$$

так что РНМ инвариантный критерий является нерандомизированным и отвергает модель “отсутствие последействия” в пользу модели “старение,” если $U(T) > C$.

Естественно, этот же критерий можно получить с помощью отношения правдоподобия условных распределений выборочного вектора относительно достаточной (для мешающего параметра) статистики $\sum_{k=1}^n X_k$. Однако в любом случае остается проблема вычисления критической константы C по заданному уровню значимости α . К сожалению, точное распределение статистики $U(T)$ имеет совершенно “нерабочий вид” типа двойного несобственного интеграла. При больших объемах n предпочтительнее воспользоваться асимптотическим распределением этой статистики. Для упрощения вывода асимптотического распределения $U(T)$ введем статистику $V(T) = U(T)/n$.

Теорема 10.4. *Статистика $V = V(T)$ асимптотически ($n \rightarrow \infty$) нормальна с параметрами*

$$\mu(\lambda) = \mathbf{E}_\lambda V = \psi(\lambda) - \psi(n\lambda), \quad \sigma^2(\lambda) = \mathbf{D}_\lambda V = \psi'(\lambda)/n - \psi'(n\lambda), \quad (10.7)$$

где $\psi(x) = d \ln \Gamma(x)/dx$ – так называемая пси-функция Эйлера, а $\psi'(x)$ – производная этой функции. Для функция мощности критерия $V > C$ имеет место асимптотическое представление

$$m(\lambda) = 1 - \Phi \left(\frac{C - \mu(\lambda)}{\sigma(\lambda)} \right) + O \left(\frac{1}{\sqrt{n}} \right),$$

где критическая константа $C = C(\alpha) = \mu(1) + \sigma(1) \Phi^{-1}(1 - \alpha)$.

Доказательство. Асимптотическая нормальность статистики

$$V = \frac{1}{n} \sum_{k=1}^n \ln X_k - \ln \left(\sum_{k=1}^n X_k \right)$$

устанавливается точно так же, как мы доказывали асимптотическую нормальность оценки параметра по методу моментов, разлагая функцию от выборочного среднего \bar{X} в ряд Тейлора в окрестности $\mathbf{E} \bar{X}$ и сохраняя только линейные члены. В нашем случае

$$V = \frac{1}{n} \sum_{k=1}^n \ln X_k - \ln n - \ln \bar{X}.$$

Так как $\mathbf{E}_\lambda \bar{X} = \lambda$, $\mathbf{E}_\lambda \ln X = \psi(\lambda)$ (вспомните, как в общем курсе теории вероятностей вычислялись моменты гамма-распределения), то

$$\begin{aligned} \sqrt{n}V &= \frac{1}{\sqrt{n}} \sum_{k=1}^n [\ln X_k - \psi(\lambda)] + \sqrt{n}\psi(\lambda) - \sqrt{n} \ln n - \sqrt{n} \ln \bar{X} = \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^n [\ln X_k - \psi(\lambda)] + \sqrt{n}\psi(\lambda) - \\ &= \sqrt{n} \ln n - \frac{\sqrt{n}(\bar{X} - \lambda)}{\lambda} + \frac{\sqrt{n}(\bar{X} - \lambda)^2}{2[\lambda + \gamma(\bar{X} - \lambda)]^2}, \end{aligned} \quad (10.8)$$

где $0 < \gamma < 1$, остаточный член (последнее слагаемое в (10.8)) сходится по вероятности к нулю со скоростью $O(1/\sqrt{n})$, а линейные члены разложения, будучи суммами независимых одинаково распределенных случайных величин с конечным вторым моментом, обеспечивают асимптотическую нормальность в силу центральной предельной теоремы.

Естественно, разложение (10.8) позволяет вычислить параметры асимптотической нормальности, но в данном случае можно найти точное

значение математического ожидания и дисперсии U , используя ее характеристическую функцию, которая также вычисляется в явном виде:

$$\begin{aligned} \varphi(t) &= \mathbf{E}_\lambda e^{itU} = \frac{1}{\Gamma^n(\lambda)} \int_0^\infty \cdots \int_0^\infty \prod_{k=1}^n x_k^{\lambda-1} \times \\ &\exp \left\{ it \left[\sum_{k=1}^n \ln X_k - n \ln \left(\sum_{k=1}^n X_k \right) \right] - \sum_{k=1}^n x_k \right\} dx_1 \cdots dx_n = \\ &\frac{1}{\Gamma^n(\lambda)} \int_0^\infty \cdots \int_0^\infty \left(\sum_{k=1}^n x_k \right)^{-nit} \prod_{k=1}^n x_k^{\lambda+it-1} \exp \left\{ - \sum_{k=1}^n x_k \right\} dx_1 \cdots dx_n. \end{aligned}$$

Последний n -кратный интеграл с точностью до нормировочной константы есть математическое ожидание $\sum_{k=1}^n X_k$, когда X_1, \dots, X_n независимы и одинаково гамма-распределены с параметром формы $\lambda + it$. Естественно, это просто вольная интерпретация, облегчающая вычисление кратного интеграла. Итак,

$$\begin{aligned} \varphi(t) &= \frac{\Gamma^n(\lambda + it)}{\Gamma^n(\lambda)} \frac{1}{\Gamma(n(\lambda + it))} \int_0^\infty x^{-nit} x^{n(\lambda+it)-1} e^{-x} dx = \\ &= \frac{\Gamma(n\lambda) \Gamma^n(\lambda + it)}{\Gamma^n(\lambda) \Gamma(n(\lambda + it))}. \end{aligned}$$

Разлагая $\ln \varphi(t)$ в ряд Тейлора по степеням it , находим среднее значение U (коэффициент при it) и дисперсию V (коэффициент при $(it)^2/2$). Формулы (10.7) для среднего и дисперсии статистики V получаются делением $\mathbf{E}_\lambda U$ на n и делением $\mathbf{D}_\lambda U$ на n^2 соответственно. \square

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. По аналогии с примером 9.3, где находились РНМ несмещенные критерии проверки гипотез о параметрах нормального распределения, докажите, что критерий в примере 10.9 для тестирования показательного распределения при альтернативном гамма-распределении является РНМ несмещенным критерием

2. Найдите функцию плотности максимального инварианта

$$T(X^{(n)}) = \left(\frac{X_1 - X_{n-1}}{X_n - X_{n-1}}, \dots, \frac{X_{n-2} - X_{n-1}}{X_n - X_{n-1}} \right),$$

когда выборка $X^{(n)}$ берется из распределения с плотностью

$$\sigma^{-1} f((x - \mu)/\sigma), \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+.$$

Найдите явный вид функции плотности статистики T , когда f – плотность нормального (μ, σ^2) распределения.

3.* По аналогии с примером п.10.9 постройте РНМ инвариантный критерий проверки гипотезы $H_0 : \lambda = 0$ при альтернативе $H_1 : \lambda > 0$ о параметре λ распределения

$$F_\lambda \left(\frac{x}{\theta} \right) = \left(\frac{x}{\theta} \right)^{\lambda+1}, \quad 0 \leq x \leq \theta,$$

при мешающем параметре θ .

4.** По аналогии с примером п.10.5 постройте РНМ инвариантный критерий проверки гипотезы $H_0 : \lambda = 0$ при альтернативе $H_1 : \lambda > 0$ о параметре формы λ обратного гауссовского распределения с функцией плотности

$$\frac{1}{\theta} f_\lambda \left(\frac{x}{\theta} \right) = \frac{1}{\theta \sqrt{2\pi}} \left(\frac{\theta}{x} \right)^{3/2} \exp \left\{ -\frac{1}{2} \left(\sqrt{\frac{\theta}{x}} - \lambda \sqrt{\frac{x}{\theta}} \right)^2 \right\}, \quad x > 0,$$

при мешающем масштабном параметре θ .

§ 11*. Статистический критерий, минимизирующий d-риск

11.1. Основные характеристики статистического контроля качества; задача минимизации d-риска. В § 1 при определении величины средних потерь, наряду с понятием θ -риска (функции риска $R(\varphi | \theta) = \mathbf{E}_\theta L(\theta, \delta(X))$), приводились примеры статистических проблем, где θ -риск не отражает по-существу ту величину средних потерь, которая необходима для утверждения гарантийности статистического вывода, понятия оптимальной процедуры, планирования объема испытаний и т.п. Усреднение потерь в таких задачах, как статистический контроль качества и медицинская диагностика, следует осуществлять по “случайному параметру” ϑ среди тех статистических экспериментов, которые завершились принятием одного и того же решения $d \in \mathcal{D}$, то есть рассматривать функцию d-риска $R_G(\varphi | d) = \mathbf{E}^{\vartheta|\delta} \{ L(\vartheta, d) | \delta(X) = d \}$, $d \in \mathcal{D}$. При проведении наблюдений над такого рода объектами значение параметра θ , определяющего распределение случайной выборки, является реализацией случайного элемента ϑ с априорным распределением G . Мы будем предполагать, как и при решении байесовских задач, что распределение G полностью известно; в противном случае следует производить его оценку по накопленному архиву данных – результатов наблюдений в предыдущих экспериментах. Чтобы подчеркнуть необходимость такого определения величины риска, напомним один пример из общего курса теории вероятностей, который иллюстрировал применение формулы Байеса.

ПРИМЕР 11.1. *Вычисление уровня выходного качества и уровня контроля при статистическом контроле качества.* Производитель продукта должен выполнять определенные договорные обязательства перед потребителем, которые, так или иначе, сводятся к ограничениям на долю некондиционной продукции, поставляемой потребителю, или, что то же, доля кондиционной продукции должна быть достаточно высокой. Обеспечение этих ограничений достигается с помощью контроля (как правило, выборочного) производимой продукции. Пусть Q_{in} – доля кондиционной продукции сре-

ди изготавливаемой предприятием. Обычно эта доля называется *входным уровнем качества*, и необходимость контроля продукции обуславливается невысоким значением Q_{in} , которое не удовлетворяет потребителя. Если контроль продукции производится на основе обследования только ее части (так называемый *выборочный* или *статистический* контроль качества), то возникает вероятность принятия ошибочного решения о качестве контролируемого продукта: с некоторой вероятностью β процедура контроля может пропустить некондиционный продукт или, наоборот, с вероятностью α отклонить кондиционный. Вероятность β называется *риском потребителя*, а вероятность α – *риском изготовителя*. Эти риски есть не что иное, как вероятности ошибок первого (α) и второго (β) рода. Зная значения Q_{in} , α и β , можно, используя формулу Байеса, вычислить *выходной уровень качества* Q_{out} – долю кондиционной продукции, среди отсылаемой потребителю после контроля.

Пусть B_1 – событие, состоящее в том, что поступивший на контроль продукт кондиционен, а $B_2 = B_1^c$ – продукт “плохой”. В таких обозначениях $P(B_1) = Q_{in}$. Пусть, далее, A – утверждение о кондиционности продукта после его контроля. Тогда $Q_{out} = P(B_1 | A)$ – вероятность кондиционности продукта при условии, что он прошел контроль. Наконец, $P(A | B_1) = 1 - \alpha$ и $P(A | B_2) = \beta$. По формуле Байеса

$$Q_{out} = P(B_1 | A) = \frac{P(A | B_1)P(B_1)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2)} = \frac{(1 - \alpha)Q_{in}}{(1 - \alpha)Q_{in} + \beta(1 - Q_{in})}.$$

Проиллюстрируем расчеты, производимые по этой формуле, на основе конкретных числовых данных. Пусть предприятие работает из рук вон плохо: $Q_{in} = 0.1$ (90% выпускаемой продукции не удовлетворяет нормам качества), но на предприятии существует довольно жесткий контроль, в котором риск потребителя $\beta = 0.01$, а риск изготовителя $\alpha = 0.1$. Тогда выходной уровень качества

$$Q_{out} = \frac{0.9 \cdot 0.1}{0.9 \cdot 0.1 + 0.01 \cdot 0.9} = \frac{10}{11} \approx 0.91,$$

и это совсем неплохо по сравнению с тем, что было до контроля.

Однако существует еще один аспект данной проблемы, который важен для изготовителя продукции. Его, естественно, интересует доля некондиционной продукции, среди той что была отклонена при контроле. Это так называемый *уровень контроля* Q_{cont} , который также вычисляется по формуле Байеса. Пусть A^c – решение о некондиционности продукта после его контроля. Тогда

$$Q_{cont} = P(B_2 | A^c) = \frac{P(A^c | B_2)P(B_2)}{P(A^c | B_1)P(B_1) + P(A^c | B_2)P(B_2)} = \frac{(1 - \beta)(1 - Q_{in})}{\alpha Q_{in} + (1 - \beta)(1 - Q_{in})}.$$

При том же входном уровне качества $Q_{in} = 0.1$ и тех же рисках потребителя $\beta = 0.01$ и изготовителя $\alpha = 0.1$ уровень контроля

$$Q_{cont} = \frac{0.99 \cdot 0.9}{0.1 \cdot 0.1 + 0.99 \cdot 0.9} \approx 0.99.$$

Следовательно, и уровень контроля, как и уровень выходного качества, достаточно высоки.

Возникает естественный вопрос, зачем определять процедуру статистического контроля по заданным ограничениям α и β на риски изготовителя и потребителя, то есть вводить гарантийный критерий различения двух гипотез, соответствующих кондиционности и некондиционности выпускаемой продукции. Если нас интересуют только уровни выходного качества и контроля, то естественнее строить критерий с ограничениями снизу на эти уровни или с ограничениями сверху β_0 и β_1 на долю некондиционной продукции среди принятой и кондиционной среди отклоненной соответственно. Естественно, при фиксированном объеме инспекции выпускаемой продукции (фиксированном объеме испытаний) невозможно контролировать оба риска, поэтому возникает необходимость в построении аналога наиболее мощного критерия, когда один из d-рисков ограничивается “уровнем значимости” β_0 , а другой минимален. Рассмотрим эту задачу в рамках общего подхода к проблеме проверки гипотез.

Пусть, как обычно, X – случайная выборка, функция плотности которой $p(x | \theta)$, $x \in \mathcal{X}$, зависит от параметра $\theta (\in \Theta)$, значение которого не

известно. По результату x наблюдения X требуется выбрать одно из двух альтернативных утверждений $H_0 : \theta \in \Theta_0$ или $H_1 : \theta \in \Theta_1 = \Theta_0^c$ о значении параметра θ . Данной проблеме соответствует пространство решений \mathcal{D} , состоящее всего из двух точек: d_0 – решение о справедливости гипотезы H_0 и d_1 – решение об истинности альтернативы.

Решение о справедливости той или иной гипотезы принимается с помощью рандомизированного правила (критической функции) $\varphi(x)$, указывающего, с какой вероятностью следует отвергать гипотезу H_0 (принимать решение d_1 о справедливости альтернативы), если x – результат статистического эксперимента. Правилу φ соответствует решающая функция $\delta(x)$, которая каждому наблюденному значению x ставит в соответствие решение d_0 или решение d_1 по результату u наблюдения случайной величины U с равномерным распределением на отрезке $[0, 1]$: $\delta(x) = d_0$, если $\varphi(x) > u$.

Напомним, что вероятность принятия решения d_0 при фиксированном значении параметра θ

$$\Psi(d_0 | \theta) = \mathbf{P}_\theta\{\delta(X) = d_0\} = \mathbf{E}_\theta\{1 - \varphi(X)\}$$

вместе с вероятностью $\Psi(d_1 | \theta) = 1 - \Psi(d_0 | \theta)$ принятия решения d_1 называется *образом* решающего правила (критерия) φ .

Рассматривается класс статистических проблем, в которых значение параметра θ при проведении наблюдений есть реализация некоторого случайного элемента ϑ и пусть P – совместное распределение вектора (X, ϑ) на пространстве $\mathcal{X} \times \Theta$, функция плотности которого по мере $\mu \times \chi$ равна $h(x, \theta) = p(x | \theta) g(\theta)$, где g – функция плотности априорного распределения G “случайного параметра” ϑ . Нижний индекс θ у знака вероятности \mathbf{P}_θ , как всегда, будет означать, что эта вероятность вычисляется при заданном значении параметра θ . Верхний индекс у знака вероятности используется для конкретизации случайного элемента, относительно которой эта вероятность вычисляется, если только это не видно из контекста записи. Аналогичные соглашения справедливы и для знака математического ожидания \mathbf{E} .

Безусловные вероятности принятия решений d_0 и d_1 (априорный об-

раз φ) вычисляются по формулам

$$\Psi(d_1) = \mathbf{P}(\delta(X) = d_1) = \mathbf{E}[\Psi(d_1 | \vartheta)] = \mathbf{E}^\vartheta[\mathbf{E}_{\vartheta} \varphi(X)], \quad \Psi(d_0) = 1 - \Psi(d_1).$$

Индекс λ или G у Ψ мы опускаем, чтобы избежать громоздкой записи формул.

Пусть $\mathbf{I}_i(\theta)$ – индикаторная функция множества Θ_i , то есть $\mathbf{I}_i(\theta) = 1$ тогда и только тогда, когда $\theta \in \Theta_i$, $i = 0, 1$. Качество критерия φ описывается двумя величинами

$$\mathcal{R}(d_0 | \varphi) = \mathbf{P}(\vartheta \in \Theta_1 | \delta(X) = d_0) = \frac{\mathbf{E}[\mathbf{I}_1(\vartheta)(1 - \varphi(X))]}{\Psi(d_0)}, \quad (11.1)$$

$$\mathcal{R}(d_1 | \varphi) = \mathbf{P}(\vartheta \in \Theta_0 | \delta(X) = d_1) = \frac{\mathbf{E}[(\mathbf{I}_0(\vartheta))\varphi(X)]}{\Psi(d_1)}, \quad (11.2)$$

представляющими собой условные вероятности принятия ошибочных решений и называемыми *d-апостериорными рисками (коротко, d-рисками) 1-го и 2-го рода*. Если соответствующие условия имеют нулевую вероятность, то риски полагаются равными нулю.

ОСНОВНАЯ ПРОБЛЕМА. *Требуется построить критерий φ^* , удовлетворяющий заданному ограничению на величину d-риска 1-го рода:*

$$\mathcal{R}(d_0 | \varphi^*) \leq \beta_0,$$

и минимизирующий d-риск 2-го рода: $\mathcal{R}(d_1 | \varphi^) \rightarrow \min$.*

11.2. Оптимальный критерий. Данный параграф представляет дословное изложение рукописи С.В. Симушкина и содержит центральный и основополагающий результат проверки гипотез в рамках d-апостериорного подхода к проблеме гарантийности – Лемма Симушкина является аналогом фундаментальной Леммы Неймана-Пирсона о структуре наиболее мощном критерии проверки гипотез. Критерий, минимизирующий d-риск второго рода при заданных ограничениях на d-риск первого рода был получен Сергеем Владимировичем еще в начале 80-х годов прошлого столетия и похоронен (по другому не скажешь) в рукописях ВИНТИ.

Определение величины средних потерь среди тех экспериментов, которые завершились принятием одного и того же решения d , приводит, если

проводить параллели с θ -риском, к замене в формулах, определяющих θ -оптимальные статистические правила, переменных d и x на θ и наоборот. Так, если РНМ критерий определялся через отношение правдоподобия, то критерий, минимизирующий d -риск, следует строить, используя отношение апостериорных вероятностей гипотез: $p(x|\theta)$ заменяется на $h_G(\theta|x)$. Поскольку апостериорная вероятность $\mathbf{P}(\vartheta \in \Theta_1 | X) = 1 - \mathbf{P}(\vartheta \in \Theta_0 | X)$, то отношение апостериорных вероятностей есть монотонная функция $\mathbf{P}(\vartheta \in \Theta_1 | X)$. Следуя сформулированной выше эвристической параллели между двумя подходами к определению риска, критерий φ^* , минимизирующий d -риск второго рода, должен определяться апостериорной вероятностью $\mathbf{P}(\vartheta \in \Theta_1 | X)$. Ниже будет показано, что оптимальный критерий φ^* отвергает нулевую гипотезу, если ее апостериорная вероятность достаточно мала, или, что то же, апостериорная вероятность $\mathbf{P}(\vartheta \in \Theta_1 | X) > C$, где C определяется по заданным ограничениям β_0 на d -риск первого рода с возможной рандомизацией на границе $\mathbf{P}(\vartheta \in \Theta_1 | X) = C$ критической области, если эта граница имеет ненулевую вероятность по распределению \mathbf{P} .

Предвосхищая дальнейшие пути в доказательстве оптимальности такого критерия, сразу же расскажем, в чем состоит заключительный этап доказательства. Если критерий φ , не совпадающий с φ^* , имеет d -риск в некоторой точке $d (= d_0$ или $d_1)$, для которого справедливо неравенство $\mathcal{R}(d|\varphi) \leq \mathcal{R}(d|\varphi^*)$, то для него вероятность принятия этого решения $\mathbf{P}(\delta = d) < \mathbf{P}(\delta^* = d)$. Поскольку сумма вероятностей $\mathbf{P}(\delta = d_0) + \mathbf{P}(\delta = d_1) = 1$, то неравенства

$$\mathcal{R}(d_0|\varphi) \leq \mathcal{R}(d_0|\varphi^*) \quad \text{и} \quad \mathcal{R}(d_1|\varphi) \leq \mathcal{R}(d_1|\varphi^*)$$

не могут выполняться одновременно, если только критерии φ и φ^* не совпадают.

Докажем сначала два вспомогательных утверждения, в которых апостериорная вероятность $R_i(x) = \mathbf{P}(\vartheta \in \Theta_{1-i} | X = x)$ трактуется (с точки зрения функции потерь 1–0) как апостериорный риск от принятия решения d_i , $i = 0, 1$, при любом результате x наблюдения случайной выборки X ,

то есть принятие решения игнорирует результат статистического эксперимента. В терминах апостериорных рисков претендент φ^* на оптимальное правило принимает вид

$$\varphi^*(x) = \begin{cases} 1, & \text{если } R_0(x) > C, \\ 0, & \text{если } R_0(x) < C. \end{cases} \quad (11.3)$$

Лемма 11.1 (i) Для любого критерия φ d -риски

$$\mathcal{R}(d_i | \varphi) = \mathbf{E}\{R_i(X) | \delta(X) = d_i\}, \quad i = 0, 1,$$

и для вычисления условных математических ожиданий справедливы формулы (11.1) и (11.2).

(ii) Если найдется такая константа A , что $\mathbf{P}(R_i(X) > A) = 1$, то для любого “невырожденного” в точке d_i критерия φ d -риски $\mathcal{R}(d_i | \varphi) > A$, $i = 0, 1$.

(iii) Для решающего правила φ^* вида (11.3)

$$\mathcal{R}(d_0 | \varphi^*) \leq C, \quad \mathcal{R}(d_1 | \varphi^*) \leq 1 - C.$$

Доказательство. Первое утверждение леммы есть следствие основного свойства условного математического ожидания, по которому операцию вычисления математического ожидания по более грубому разбиению ($\delta(X) = d$) пространства значений случайного элемента можно предварить операцией вычисления условного математического ожидания по более мелкому разбиению этого пространства ($X = x$). В связи с этим смотрите формулы (11.1) и (11.2).

Второе утверждение леммы является прямым следствием первого утверждения, поскольку замена $R_i(X)$ на меньшую величину A приводит к уменьшению условного математического ожидания.

Третье утверждение леммы доказывается аналогичным образом:

$$\begin{aligned} \mathcal{R}(d_0 | \varphi^*) &= \mathbf{E}\{R_0(X) | \delta(X) = d_0\} = \mathbf{E}\{R_0(X) | R_0 < C\} \leq C, \\ \mathcal{R}(d_1 | \varphi^*) &= \mathbf{E}\{R_1(X) | R_0 > C\} = \mathbf{E}\{R_1(X) | R_1 < 1 - C\} \leq 1 - C. \quad \square \end{aligned}$$

Лемма 11.2. Если d -риски критерия φ и критерия φ^* в некоторой точке d удовлетворяют неравенству

$$\mathcal{R}(d | \varphi) \leq \mathcal{R}(d | \varphi^*),$$

то для безусловных вероятностей принятия решения d этими критериями справедливо неравенство

$$\Psi(d) = \mathbf{P}(\delta(X) = d) \leq \mathbf{P}(\delta^*(X) = d) = \Psi^*(d).$$

Доказательство леммы проведем только для решения $d = d_0$. Сначала заметим, для критериев φ и φ^* функция

$$H(x) = [\varphi^*(x) - \varphi(x)][C - R_0(x)] \leq 0, \quad \forall x \in \mathcal{X}.$$

Доказательство этого неравенства совпадает с доказательством аналогичного неравенства в лемме Неймана–Пирсона. Действительно, если $C - R_0(x) > 0$, то $\varphi^*(x) = 0$ и $\varphi^*(x) - \varphi(x) = -\varphi(x) \leq 0$. И наоборот, если $C - R_0(x) < 0$, то $\varphi^*(x) = 1$ и $\varphi^*(x) - \varphi(x) = 1 - \varphi(x) \geq 0$. Следовательно, математическое ожидание

$$\mathbf{E}[\varphi^*(X) - \varphi(X)][C - R_0(X)] \leq 0, \quad (11.4)$$

со строгим знаком неравенства только в том случае, если с положительной вероятностью на множестве $R_0(x) \neq C$ φ и φ^* различаются.

Путем очевидных выкладок неравенство (11.4) преобразуется к виду

$$C \mathbf{E}[\varphi^*(X) - \varphi(X)] \leq \mathbf{E} R_0(X)[\varphi^*(X) - \varphi(X)].$$

Левая часть этого неравенства равна

$$C[\Psi^*(d_0) - \Psi(d_0)] = C[\Psi(d_0) - \Psi^*(d_0)].$$

Правую часть в силу утверждения (i) леммы 11.1 и формулы (11.1) можно записать как

$$\mathbf{E} R_0(X)[((1 - \varphi(X)) - (1 - \varphi^*(X)))] = \mathcal{R}(d_0 | \varphi) \Psi(d_0) - \mathcal{R}(d_0 | \varphi^*) \Psi^*(d_0).$$

По предположению настоящей леммы эта разность меньше или равна

$$\mathcal{R}(d_0 | \varphi^*) [\Psi(d_0) - \Psi^*(d_0)].$$

Таким образом, справедливо неравенство

$$[C - \mathcal{R}(d_0 | \varphi^*)][\Psi(d_0) - \Psi^*(d_0)] \leq 0,$$

которое в силу третьего утверждения Леммы 11.1 влечет доказываемое неравенство для вероятностей принятия решения d_0 . Для другого решения доказательство проводится аналогично (с заменой C на $1 - C$). \square

Теорема 11.1 (лемма Симушкина). Если d -риск 1-го рода статистического критерия φ

$$\mathcal{R}(d_0 | \varphi) \leq \mathcal{R}(d_0 | \varphi^*)$$

и φ не совпадает с φ^* , то d -риск 2-го рода

$$\mathcal{R}(d_1 | \varphi) > \mathcal{R}(d_1 | \varphi^*).$$

Доказательство. Поскольку сумма вероятностей $\Psi(d) + \Psi(d_0) = 1$ для любого критерия, то d -риск никакого правила не может быть меньше d -риска правила φ^* сразу для обоих решений. \square

Таким образом, если подобрать константу $C = C(\beta_0)$ так, чтобы d -риск первого рода $\mathcal{R}(d_0 | \varphi^*) = \beta_0$, то критерий φ^* становится оптимальным среди всех критериев, удовлетворяющих заданным ограничениям на d -риск первого рода. Это, естественно, можно сделать по аналогии с критерием Неймана–Пирсона, вводя рандомизацию на границе критической области критерия φ^* , но только при не слишком малых и не слишком больших значениях β_0 . Для того, чтобы понять, в чем суть проблемы, рассмотрим критерий

$$\varphi_c(x) = \begin{cases} 1, & \text{если } R_0(x) \geq c, \\ 0, & \text{если } R_0(x) < c, \end{cases}$$

который по форме совпадает с φ^* , но определен при всех значениях $x \in \mathcal{X}$. Пусть

$$r(c) = \mathcal{R}(d_0 | \varphi_c) = \mathbf{E}\{\vartheta \in \Theta_1 | R_0(x) < c\} -$$

d -риск первого рода правила φ_c , рассматриваемый как функция аргумента $c \in [0, 1]$, и c_0 – наименьшее значение апостериорной вероятности $R_0(x)$ в предположении, что $\mathbf{P}(R_0(X) < c) > 0$ для всех $c > c_0$.

Лемма 11.3. Функция $r(c)$

(i) не убывает,

(ii) непрерывна слева,

(iii) принимает значения 0 или $c_0 \leq r(c) \leq G_1 = \mathbf{P}(\vartheta \in \Theta_1)$.

Доказательство. (i) Так как с ростом c множество $\{x : R_0(x) < c\}$ расширяется, то безусловная вероятность этого события (безусловная вероятность принятия решения d_0) увеличивается. Из утверждения Леммы 11.2 следует, что тогда d -риск первого рода не может уменьшиться.

(ii) Непрерывность слева устанавливается так же, как доказывалась непрерывность слева функции распределения (см. общий курс ТВ и МС).

(iii) Правило φ' , принимающее решение d_0 при любых $x \in \mathcal{X}$, имеет d -риск первого рода, равный безусловной вероятности справедливости альтернативы

$$\mathcal{R}(d_0 | \varphi') = P(\vartheta \in \Theta_1) = G_1.$$

Поскольку безусловная вероятность принятия решения d_0 критерием φ_c при любом c меньше соответствующей вероятности для φ' , то из Леммы 11.2 вытекает, что $\mathcal{R}(d_0 | \varphi_c) \leq \mathcal{R}(d_0 | \varphi') = G_1$ при любых $c \in [0, 1]$.

С другой стороны, для любого $c > c_0$ вероятность $\mathbf{P}(R_0(X) < c) > 0$ (см. определение c_0), поэтому критерий φ_c не вырожден и его d -риск первого рода, в силу утверждения (ii) Леммы 11.1), $c_0 \leq \mathcal{R}(d_0 | \varphi_c) < c$.

Если $c_0 > 0$, тогда при любых $c < c_0$ вероятность принятия решения d_0 по критерию φ_c равна нулю и по построению $\mathcal{R}(d_0 | \varphi_c) = 0$ \square

Из доказанной леммы можно сделать следующие выводы.

1). Если ограничение $\beta_0 < \gamma_0$, тогда не существует невырожденных критериев φ , у которых d -риск $\mathcal{R}(d_0 | \varphi) \leq \beta_0$.

2). Если ограничение $\beta_0 \geq G_1$, тогда решающее правило, принимающее решение d_0 при любых результатах наблюдений (то есть, по-существу, без проведения наблюдений) удовлетворяет ограничению β_0 на величину d -риска первого рода.

Итак, пусть $c_0 < \beta_0 < G_1$. Как отмечалось выше, если найдется константа c^* , для которой $r(c^*) = \beta_0$, тогда критерий φ_{c^*} и будет оптимальным.

Поскольку функция $r(c)$ изменяется (не убывая) от c_0 до G_1 , то при сделанных предположениях относительно β_0 существует константа

$$c^* = \sup\{c : r(c) \leq \beta_0\}.$$

Так как функция $r(c)$ непрерывна слева, то и $r(c^*) \leq \beta_0$. Нам осталось изучить случай, когда функция $r(c)$ терпит разрыв в точке c^* , причем $r(c^*) < \beta_0$.

Рассмотрим функцию $\tilde{r}(c) = \mathcal{R}(d_0 | \tilde{\varphi}_c)$, где критерий

$$\tilde{\varphi}_c(x) = \begin{cases} 1, & \text{если } R_0(x) > c, \\ 0, & \text{если } R_0(x) \leq c, \end{cases}$$

Эта функция так же, как и $r(c)$, не убывает и изменяется от γ_0 до G_1 . Единственное её отличие от $r(c)$ состоит в том, что она непрерывна справа. Кроме того, поскольку при $c < c'$ вероятность $\mathbf{P}(R_0(X) < c) \leq \mathbf{P}(R_0(X) \leq c) \leq \mathbf{P}(R_0(X) < c')$, то

$$r(c) \leq \tilde{r}(c) \leq r(c').$$

Если предположить, что в точке $c = c^*$ функция $\tilde{r}(c^*) < \beta_0$, тогда в силу её непрерывности справа найдется точка $c' > c^*$, в которой также $\tilde{r}(c') < \beta_0$. В этом случае для любой точки c'' такой, что $c^* < c'' < c < c'$, функция $r(c'') \leq r(c) \leq \tilde{r}(c) \leq \tilde{r}(c') < \beta_0$, то есть существует такое $c'' > c^*$, для которого также $r(c'') < \beta_0$, что противоречит выбору c^* как максимальной точки с таким свойством.

Теорема 11.2. *Если $c_0 < \beta_0 < G_1$, тогда существуют такие константы C и γ , что для критерия*

$$\varphi^*(x) = \begin{cases} 1, & \text{если } R_0(x) > C, \\ \gamma, & \text{если } R_0(x) = C, \\ 0, & \text{если } R_0(x) < C \end{cases} \quad (11.5)$$

- (i) d -риск 1-го рода $\mathcal{R}(d_0 | \varphi^*) = \beta_0$;
- (ii) d -риск 2-го рода $\mathcal{R}(d_1 | \varphi^*)$ минимален среди всех критериев φ , удовлетворяющих ограничению $\mathcal{R}(d_0 | \varphi) \leq \beta_0$ на величину d -риска первого рода.

Доказательство. Рассмотрим критерий

$$\varphi_q^*(x) = \begin{cases} 1, & \text{если } R_0(x) > c^*, \\ q, & \text{если } R_0(x) = c^*, \\ 0, & \text{если } R_0(x) < c^*. \end{cases}$$

Его d-риск первого рода равен (см. формулу (11.1))

$$\begin{aligned} \mathcal{R}(d_0 | \varphi_q^*) &= \frac{\mathbf{E}[\mathbf{I}_1(\vartheta) (1 - \varphi_q^*(X))]}{\mathbf{E}(1 - \varphi_q^*(X))} = \\ &= \frac{\mathbf{P}(\{\vartheta \in \Theta_1\} \cap \{R_0(X) < c^*\}) + q \mathbf{P}(\{\vartheta \in \Theta_1\} \cap \{R_0(X) = c^*\})}{\mathbf{P}(R_0(X) < c^*) + q \mathbf{P}(R_0(X) = c^*)}. \end{aligned}$$

Очевидно, относительно переменной q этот риск представляет собой непрерывную, монотонную дробно-линейную функцию.

При $q = 0$ риск $\mathcal{R}(d_0 | \varphi_0^*) = r(c^*) < \beta_0$, а при $q = 1$ риск $\mathcal{R}(d_0 | \varphi_1^*) = \tilde{r}(c^*) \geq \beta_0$. Следовательно, существует (единственная) константа γ , при которой $\mathcal{R}(d_0 | \varphi_\gamma^*) = \beta_0$. Таким образом, критерий $\varphi^*(x)$ с константой $C = c^*$ и константой γ , определяемой как решение уравнения $\mathcal{R}(d_0 | \varphi_q^*) = \beta_0$ относительно переменной q , является оптимальным критерием. \square

Замечание. Из доказательства теоремы видно, что необходимость в рандомизации возникает только в случае, когда $\mathbf{P}(R_0(X) = c^*) > 0$.

Вернемся к нашему начальному примеру с контролем качества. Предположим, что качество выпускаемой продукции характеризуется процентным содержанием вредной примеси, количество которой определяет значение параметра θ . Продукт считается кондиционным, если $\theta \leq \theta_0$, и пусть эта область Θ_0 параметрического пространства $\Theta = [0, 100]$ соответствует нулевой гипотезе H_0 . Некондиционность продукта определяется значениями θ , превышающими допустимое значение θ_0 , то есть альтернативная гипотеза H_1 определяется интервалом $\Theta_1 = \Theta_0^c = (\theta_0, 100]$. Подобного рода пример мы рассматривали в общем курсе ТВ и МС, когда вводили модель нормального распределения $N(\theta, \sigma^2)$, – определялось общее содержание серы в дизельном топливе. Естественно, те же доводы, что и при изучении метода, которым определяется содержание общей серы, справедливы и для спецификации семейства априорных распределений – это

снова нормальное распределение с некоторыми параметрами (μ, τ^2) . При такой формализации проблемы, контроль качества, осуществляемый с помощью оптимального критерия (11.5), будет гарантировать заданную долю β_0 некондиционного ($\theta \in \Theta_1$) продукта, поставляемого потребителю, ибо вероятность “нежелательного” события $\vartheta \in \Theta_1$ при условии, что продукт прошел контроль ($\delta(x) = d_0$), не превосходит β_0 . В тоже время, критерий (11.5) минимизирует долю β_1 кондиционной продукции, среди отклоненной при контроле.

Мы еще вернемся к детальному построению критерия (11.5) в рамках модели N–N, а пока докажем одно утверждение, которое значительно упрощает построение этого критерия для рассматриваемых нами моделей N–N, В–В и P–G в § 4.

Теорема 11.3. Пусть θ – действительный параметр, относительно значений которого выдвигается нулевая гипотеза $H_0 : \theta \leq \theta_0$, при альтернативе $H_1 : \theta > \theta_0$. Если статистическая модель $\mathcal{P} = \{P(\cdot | \theta), \theta \in \Theta \subset \mathbb{R}\}$ обладает монотонным относительно некоторой статистики $T = T(X)$ отношением правдоподобия, то оптимальный критерий (11.5) принимает вид

$$\varphi^*(x) = \begin{cases} 1, & \text{если } T(x) > C, \\ \gamma, & \text{если } T(x) = C, \\ 0, & \text{если } T(x) < C, \end{cases}$$

где постоянные $C = C(\beta_0)$ и $\gamma = \gamma(\beta_0)$ определяются решением уравнения

$$\frac{\mathbf{P}(\{\vartheta \in \Theta_1\} \cap \{T(X) < C\}) + \gamma \mathbf{P}(\{\vartheta \in \Theta_1\} \cap \{T(X) = C\})}{\mathbf{P}(T(X) < C) + \gamma \mathbf{P}(T(X) = C)} = \beta_0. \quad (11.6)$$

Доказательство. Критерий (11.5) не меняет своих оптимальных свойств, если вместо статистики $R_0(X) = \mathbf{P}(\vartheta \in \Theta_1 | X)$ использовать отношение апостериорных вероятностей $R_0(X)/R(X)$ – монотонно возрастающую функцию $t/(1-t)$ аргумента $t = \mathbf{P}(\vartheta \in \Theta_1 | X)$. Напомним, что именно эта статистика служила эвристическим доводом в пользу оптимальности критерия, основанного на отношении апостериорных вероятностей различаемых гипотез. Для односторонних гипотез, представленных в

формулировке данной теоремы, статистика отношения апостериорных вероятностей имеет вид отношения двух интегралов по непересекающимся интервалам, которые соответствуют гипотезам H_1 и H_0 :

$$L(X) = \int_{-\infty}^{\theta_0} p(X|\theta) dG(\theta) / \int_{\theta_0}^{\infty} p(X|\theta) dG(\theta).$$

По определению семейства распределений, обладающих монотонным отношением статистики T отношением правдоподобия, отношение плотностей $p(x|\theta)/p(x|\theta_0)$ зависит от x только через значение статистики $T(x)$, не убывает, если $\theta > \theta_0$, и не возрастает, если $\theta < \theta_0$. Следовательно, запись статистики $L(X)$ в эквивалентной форме

$$L(X) = \int_{-\infty}^{\theta_0} \frac{p(X|\theta)}{p(X|\theta_0)} dG(\theta) / \int_{\theta_0}^{\infty} \frac{p(X|\theta)}{p(X|\theta_0)} dG(\theta)$$

показывает, что $L(X)$ есть неубывающая функция $T(X)$. \square

ПРИМЕР 11.2. *Статистический контроль качества в рамках модели N–N с гарантированным уровнем выходного качества.* Обратимся снова к примеру с контролем качества выпускаемой продукции по содержанию вредной примеси. Мы рассматривали его перед формулировкой и доказательством Теоремы 11.3, которая дает инструмент для построения процедуры контроля с гарантированным уровнем $1 - \beta_0$ выходного качества в рамках вероятностной модели N–N.

Статистическая модель N–N (семейство распределений случайной выборки) обладает монотонным отношением статистики $T(X) = \bar{X}$ отношением правдоподобия. Поскольку нормальное распределение со средним θ и дисперсией σ^2/n статистики T принадлежит непрерывному типу, то *критерий, гарантирующий выходной уровень качества, является нерандомизированным и отвергает гипотезу кондиционности продукта, если $\bar{X} > C$, где C определяется из уравнения (см. (11.6) при $\gamma = 0$)*

$$\mathbf{P}(\vartheta \in \Theta_1 | \bar{X} < C) = \frac{\mathbf{P}(\{\vartheta > \theta_0\} \cap \{\bar{X} < C\})}{\mathbf{P}(\bar{X} < C)} = \beta_0. \quad (11.7)$$

Условное распределение \bar{X} относительно ϑ есть нормальное распределение со средним ϑ и дисперсией σ^2/n , распределение $\vartheta \sim N(\mu, \tau^2)$,

маргинальное распределение $\bar{X} \sim N(\mu, \tau^2 + \sigma^2/n)$. Таким образом, уравнение (11.7) для определения критической константы $C = C(\beta_0)$ имеет вид

$$\frac{1}{\sqrt{2\pi} \tau} \int_{\theta_0}^{\infty} \Phi \left(\frac{C - \theta}{\sigma} \sqrt{n} \right) \exp \left\{ -\frac{(\theta - \mu)^2}{2\tau^2} \right\} d\theta = \beta_0 \Phi \left(\frac{C - \mu}{\sqrt{\tau^2 + \sigma^2/n}} \right).$$

Решить такое уравнение, естественно, можно только с помощью компьютера, используя пакеты прикладных программ численного интегрирования.

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. По аналогии с примером 11.2 постройте статистический контроль качества с гарантированным выходным уровнем качества в рамках модели В–В (так называемый приемочный контроль по качественному признаку), пренебрегая процедурой рандомизации.
2. По выборке фиксированного объема n из распределения с плотностью $f(x) = \theta \exp\{-\theta x\}$, $x > 0$, где θ – реализация случайной величины ϑ с априорным гамма-распределением, постройте статистический контроль качества с гарантированным выходным уровнем надежности: проверяется гипотеза $\exp\{-\theta x_0\} > P$.

§ 12. ГАРАНТИЙНЫЙ СТАТИСТИЧЕСКИЙ ВЫВОД.

Мы приступаем к решению второй из основных проблем теории статистических выводов – построению статистических процедур, гарантирующих заданные ограничения на риск и минимизирующих при этом объем наблюдений.

В задачах θ -гарантийности такие ограничения $r(\theta)$ накладываются на функцию риска: процедура $\varphi = (\varphi_s, \varphi_c, \varphi_d)$ называется θ -гарантийной, если $R(\varphi | \theta) \leq r(\theta)$. При этом условии ставится задача на оптимизацию управления статистическим экспериментом $\rho = (\varphi_s, \varphi_c)$ – соответствующий ρ момент остановки ν должен минимизировать некоторый функционал от его распределения, например, $\sup_{\theta \in \Theta} \mathbf{E}_\theta \nu$.

При построении оптимальных по объему наблюдений d -гарантийных процедур ограничения $r(d)$ накладываются на функцию d -риска: $R_G(\varphi | d) \leq r(d)$, и при этом условии минимизируется, скажем, безусловное среднее значение $\mathbf{E}_G \nu$ объема наблюдений или $\sup_{d \in \mathcal{D}} \mathbf{E} \{ \nu | \delta(X^{(\nu)}) = d \}$.

Как отмечалось в первом параграфе, если в оптимизации объема наблюдений θ -гарантийных процедур имеется ряд достижений в задачах оценки и проверки гипотез, то для d -гарантийных процедур решение проблемы оптимизации известно для тех же задач только при управлении ρ , компоненты которого не зависят от результатов наблюдений на каждом шаге статистического эксперимента, то есть в классе моментов остановки вида $\nu = n (= 0, 1, \dots)$.

Решение такого рода задач несколько упрощается, если найти правило принятия решения φ_d , соответствующее оптимальному моменту остановки. Очевидно, оно должно быть минимаксным по отношению к нормированному на заданные ограничения риску, то есть минимизировать $\sup_{\theta \in \Theta} R(\varphi | \theta) / r(\theta)$ или, соответственно, $\sup_{d \in \mathcal{D}} R_G(\varphi | d) / r(d)$ при каждом фиксированном управлении ρ статистическим экспериментом. Поскольку минимаксные решения являются байесовскими (или пределами байесовских) при наименее благоприятных априорных распределениях, а байесовские решения не зависят от управления, то такая задача часто имеет решение. Все это мы рассмотрим ниже на примере традиционных задач

математической статистики.

12.1. Статистические оценки с гарантированной точностью и надежностью. В рамках общей проблемы θ -гарантийных процедур оценки параметра θ рассматриваемая задача состоит в построении доверительной области фиксированного диаметра с оптимизацией требуемого для выполнения таких гарантий объема наблюдений. Мы не будем касаться проблемы оценки многомерного параметра и рассмотрим только задачу гарантийной оценки $\delta(X^{(\nu)})$ по результатам наблюдения одной выборки $X^{(\nu)} = (X_1, \dots, X_\nu)$ действительного параметра θ . Задача состоит в минимизации $\sup_{\theta \in \Theta} \mathbf{E}_\theta \nu$ с ограничениями $\mathbf{P}_\theta(|\delta(X^{(\nu)}) - \theta| \leq \Delta) \geq 1 - \alpha$ при любом $\theta \in \Theta \subset \mathbb{R}$, где Δ – требуемая точность оценки, а $1 - \alpha$ – ее минимально допустимая надежность.

Можно доказать, и это не так сложно, как громоздко, что в случае оценки параметра сдвига, то есть когда функция плотности наблюдаемой случайной величины ξ имеет вид $f(x - \theta)$, оптимальное правило остановки φ_s^* не зависит от результатов наблюдений. Это следствие Теоремы 10.1 о постоянстве функции θ -риска эквивариантной оценки при транзитивной группе преобразований параметрического пространства Θ . Рандомизация момента остановки, которую можно провести до проведения наблюдений, незначительно сокращает средний объем наблюдений при практически используемых (малых) значениях α , и ею можно пренебречь. Таким образом, практически осмысленная задача оптимального планирования объема испытаний состоит в определении минимального объема выборки n , при котором существует оценка $\hat{\theta}(X^{(n)})$ с заданными ограничениями на ее точность (Δ) и надежность ($1 - \alpha$). Такой объем наблюдений называется *необходимым объемом выборки*. Естественно, без спецификации $f(x - \theta)$ здесь не обойдешься, поэтому решение данной проблемы иллюстрируется на примерах.

ПРИМЕР 12.1. *Необходимый объем выборки при оценке среднего значения нормального распределения.* Рассмотрим сначала случай, когда значение дисперсии σ^2 нормального распределения известно, а оценке подлежит среднее значение θ . В такой ситуации решение задачи по оптимизации

объема наблюдений не вызывает затруднений. Всё архипросто: выборочное среднее \bar{X} является минимаксной оценкой, получаемой (см. § 5) как предел байесовской в рамках модели N–N, к тому же это эквивариантная оценка относительно группы сдвигов. Распределение \bar{X} нормально с параметрами $(\theta, \sigma^2/n)$. Следовательно, необходимый объем выборки определяется как наименьшее n , удовлетворяющее неравенству

$$\mathbf{P}(|\bar{X} - \theta| \leq \Delta) = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1 \geq 1 - \alpha$$

или, что то же, неравенству

$$n \geq \left[\frac{\Phi(1 - \alpha/2)}{\Delta} \sigma \right]^2.$$

(Эту формулу мы выводили в общем курсе математической статистики без обсуждения проблемы оптимизации объема наблюдений.)

В качестве задания на дом предлагается решить задачу оптимальной рандомизации объема наблюдений: момент остановки ν , не зависящий от результатов наблюдений, должен быть выбран так, чтобы выполнялось равенство

$$\mathbf{E} \left[2\Phi\left(\frac{\Delta}{\sigma}\sqrt{\nu}\right) - 1 \right] = 1 - \alpha$$

и при этом достигался минимум

$$\mathbf{E} \nu = \sum_{k=1}^{\infty} k p_k$$

по всем последовательностям вероятностей

$$\left\{ p_k, k = 0, 1, \dots, \sum_{i=0}^{\infty} p_i = 1 \right\}.$$

Всё значительно усложняется, если значение σ не известно.

Теорема 12.1. В задаче оценивания среднего значения θ нормального распределения с гарантированной точностью Δ и гарантированной надежностью $1 - \alpha$ по выборке $X^{(n)}$ фиксированного объема n не существует оценки $\hat{\theta}(X^{(n)})$, которая является гарантийной при каком-либо конечном объеме наблюдений n .

Доказательство. Проблема состоит в построении такой оценки, чтобы выполнялось неравенство $\mathbf{P}_{\theta,\sigma}(|\hat{\theta}(X^{(n)}) - \theta| \leq \Delta) \geq 1 - \alpha$ при любых значениях θ и σ . Однако

$$\sup_{\hat{\theta}} \inf_{\theta,\sigma} \mathbf{P}_{\theta,\sigma}(|\hat{\theta}(X^{(n)}) - \theta| \leq \Delta) \leq \inf_{\theta,\sigma} \sup_{\hat{\theta}} \mathbf{P}_{\theta,\sigma}(|\hat{\theta}(X^{(n)}) - \theta| \leq \Delta) \leq \liminf_{\sigma \rightarrow \infty} \inf_{\theta} \sup_{\hat{\theta}} \mathbf{P}_{\theta,\sigma}(|\hat{\theta}(X^{(n)}) - \theta| \leq \Delta).$$

При каждом фиксированном значении σ наибольшее значение надежности

$$\mathbf{P}_{\theta,\sigma}(|\hat{\theta}(X^{(n)}) - \theta| \leq \Delta)$$

достигается на оценке $\hat{\theta} = \bar{X}$ – минимаксной оценки θ при известном значении σ . Следовательно,

$$\begin{aligned} \liminf_{\sigma \rightarrow \infty} \inf_{\theta} \sup_{\hat{\theta}} \mathbf{P}_{\theta,\sigma}(|\hat{\theta}(X^{(n)}) - \theta| \leq \Delta) = \\ \liminf_{\sigma \rightarrow \infty} \left[2\Phi\left(\frac{\Delta}{\sigma} \sqrt{n}\right) - 1 \right] = 0, \end{aligned}$$

то есть требуемая надежность $1 - \alpha$ при любом конечном фиксированном объеме наблюдений не достижима. \square

Однако если обратиться к последовательному планированию статистического эксперимента, то гарантийная оценка существует и решение проблемы гарантийности дает так называемая *двухступенчатая процедура Стейна*.

Процедура Стейна предусматривает два этапа в проведении статистического эксперимента. На первом этапе берется выборка фиксированного объема $n_0 \geq 2$, по результатам наблюдений вычисляются выборочное среднее \bar{X}_{n_0} , несмещенная оценка

$$S_{n_0}^2 = \frac{1}{n_0 - 1} \sum_{k=1}^{n_0} (X_k - \bar{X}_{n_0})^2$$

дисперсии σ^2 и определяется момент остановки статистического эксперимента

$$\nu = \max \left\{ n_0, \left[\left(\frac{t_\alpha S_{n_0}}{\Delta} \right)^2 \right] + 1 \right\},$$

где $t_\alpha = S_{n_0-1}^{-1}(1 - \alpha/2)$ – квантиль распределения Стьюдента с $n_0 - 1$ степенями свободы. Наблюдения прекращаются на первом этапе, если $\nu = n_0$. Если же $\nu > n_0$, то статистик проводит (второй этап эксперимента) еще $\nu - n_0$ наблюдений.

Теорема 12.2. *Выборочное среднее \bar{X}_ν является оценкой θ , гарантирующей заданные точность Δ и надежность $1 - \alpha$ процедуры оценивания.*

Доказательство. Надежность оценки \bar{X}_ν можно представить как среднее значение условного математического ожидания надежности относительно момента остановки ν :

$$\mathbf{P}_{\theta,\sigma}(|\bar{X}_\nu - \theta| \leq \Delta) = \mathbf{E}^\nu \{ \mathbf{P}_{\theta,\sigma}(|\bar{X}_\nu - \theta| \leq \Delta) | \nu \}.$$

Легко видеть, что статистики \bar{X}_ν и ν суть независимые случайные величины. Действительно, условная вероятность $\mathbf{P}(\bar{X}_n < x | \nu = n) = \mathbf{P}(\bar{X}_n < x)$, ибо первые n_0 слагаемых в \bar{X} не зависят от выборочной дисперсии $S_{n_0}^2$ (напомним, момент остановки зависит только от $S_{n_0}^2$) в силу теоремы Фишера, а оставшиеся $n - n_0$ слагаемых не зависят от X_1, \dots, X_{n_0} . Следовательно,

$$\begin{aligned} \mathbf{P}_{\theta,\sigma}(|\bar{X}_\nu - \theta| \leq \Delta) &= 2 \mathbf{E} \Phi \left(\frac{\Delta}{\sigma} \sqrt{\nu} \right) - 1 \geq 2 \mathbf{E} \Phi \left(\frac{\Delta}{\sigma} \frac{t_\alpha S_{n_0}}{\Delta} \right) - 1 = \\ &= 2 \mathbf{E} \Phi \left(\frac{S_{n_0}}{\sigma} t_\alpha \right) - 1 = 2 \mathbf{E} \Phi \left(\frac{\chi_{n_0-1}}{\sqrt{n_0-1}} t_\alpha \right) - 1, \end{aligned}$$

где χ_{n_0-1} – случайная величина, квадрат которой имеет хи-квадрат распределение с $n_0 - 1$ степенью свободы. Осталось только показать, что

$$2 \mathbf{E} \Phi \left(\frac{\chi_{n_0-1}}{\sqrt{n_0-1}} t_\alpha \right) - 1 = 1 - \alpha.$$

Введем обозначение $m = (n_0 - 1)/2$ и вычислим математическое ожидание:

$$\mathbf{E} \Phi \left(\frac{\chi_{n_0-1}}{\sqrt{n_0-1}} t_\alpha \right) = C_m \int_{-\infty}^{t_\alpha} dx \int_0^\infty \sqrt{t} t^{m-1} \exp \left\{ -\frac{1}{2} \left(\frac{tx^2}{n_0-1} + t \right) \right\} dt,$$

где

$$C_m = \frac{1}{\sqrt{2\pi(n_0-1)}} \frac{1}{2^m \Gamma(m)}.$$

Интеграл по dt

$$\int_0^{\infty} t^{m-1/2} \exp \left\{ -\frac{t}{2} \left(\frac{x^2}{n_0 - 1} + 1 \right) \right\} dt = \frac{2^{m+1/2} \Gamma(m + 1/2)}{(1 + x^2/(n_0 - 1))^{m+1/2}}.$$

Таким образом,

$$\mathbf{E} \Phi \left(\frac{\chi_{n_0-1}}{\sqrt{n_0-1}} t_\alpha \right) = \frac{1}{\sqrt{\pi(n_0-1)}} \frac{\Gamma(n_0/2)}{\Gamma((n_0-1)/2)} \int_{-\infty}^{t_\alpha} \left(1 + \frac{x^2}{n_0-1} \right)^{-n_0/2} dx.$$

Правая сторона этого равенства представляет значение функции распределения Стьюдента с $n_0 - 1$ степенями свободы в точке t_α и, следовательно, равна $1 - \alpha/2$, а общее значение надежности оценки не меньше чем $1 - \alpha$. \square

Обсудив проблему оптимизации объема наблюдений при гарантийной оценке среднего значения θ , рассмотрим аналогичную задачу для оценки σ .

ПРИМЕР 12.2. *Оптимальный объем выборки при оценке стандартного отклонения нормального распределения с гарантированной относительной точностью и надежностью.* Рассмотрим проблему гарантийной оценки с минимальным средним объемом наблюдений для параметра σ нормального (μ, σ^2) распределения. С практической точки зрения это более осмысленная задача, чем оценка дисперсии, ибо именно σ , а не σ^2 используется при характеристике рассеяния распределений и выборочных данных. Поскольку σ – параметр масштаба, то ошибка в ее оценивании обычно измеряется в отношении к истинному значению, то есть функция потерь типа 1–0 должна иметь вид

$$L(\sigma, d) = \begin{cases} 1, & \text{если } |d/\sigma - 1| > \Delta, \\ 0, & \text{если } |d/\sigma - 1| \leq \Delta. \end{cases}$$

Итак, для минимизации среднего объема выборки необходимо найти минимаксную оценку σ . Естественно, такая оценка должна зависеть только от достаточных статистик \bar{X} и S^2 , но насколько оправдано участие \bar{X} в данной задаче? Действительно, \bar{X} можно игнорировать, потому что данная

статистическая проблема инвариантна относительно группы масштабных преобразований, эквивариантная оценка должна зависеть от максимального инварианта, коим при редукции данных к достаточным статистикам является S^2 , и существует замечательная теорема (увы, слишком сложная в доказательстве), утверждающая, что минимаксная эквивариантная оценка параметра масштаба является минимаксной в классе всех оценок. Итак, проблема свелась к нахождению минимаксной оценки, являющейся только функцией максимального инварианта S^2 .

Обратимся к байесовскому методу построения минимаксной оценки. Поскольку апостериорное распределение не зависит от правила останова, то такой метод даст универсальную минимаксную оценку, которая соответствует минимальному среднему объему наблюдений.

Введем “случайный” параметра $\vartheta = \sigma^{-2}$, положим $m = (n - 1)/2$ и запишем в этих обозначениях условную плотность статистики $T = \sum_{k=1}^n (X_k - \bar{X})^2$ относительно ϑ :

$$p(t | \vartheta) = \frac{\vartheta^m}{2^m \Gamma(m)} t^{m-1} \exp \left\{ -\frac{t \vartheta}{2} \right\}, \quad t > 0.$$

Выберем в качестве априорного распределения ϑ гамма-распределение с плотностью

$$g(\theta) = \frac{a^\lambda}{\Gamma(\lambda)} \theta^{\lambda-1} \exp\{-a\theta\}, \quad t > 0; a > 0, \lambda > 0.$$

Найдем плотность апостериорного распределения ϑ , которая равна совместной плотности T и ϑ , поделенной на маргинальную плотность T . Однако уже из вида совместной плотности T и ϑ

$$p(t | \theta)g(\theta) = C(t; a \lambda) \theta^{m+\lambda-1} \exp\{-\theta(a + t/2)\},$$

где

$$C(t; a \lambda) = \frac{a^\lambda t^{m-1}}{2^m \Gamma(m) \Gamma(\lambda)},$$

можно сделать вывод, что апостериорное распределение есть гамма-распределение с плотностью

$$h(\theta | T) = \frac{A^\gamma}{\Gamma(\gamma)} \theta^{\gamma-1} \exp\{-A\theta\},$$

где $A = a + T$, $\gamma = m + \lambda$.

При построении байесовской оценки для функции потерь типа 1–0 удобнее оперировать не апостериорным риском, а апостериорной надежностью

$$\mathcal{H}(d | T) = \int_{|d\sqrt{\theta}-1| \leq \Delta} h(\theta | T) dt = \frac{A^\gamma}{\Gamma(\gamma)} \int_{(1-\Delta)/d^2}^{((1+\Delta)/d)^2} \theta^{\gamma-1} \exp\{-A\theta\} dt.$$

Вычисляя от этой функции производную по аргументу d и приравнивая ее нулю, находим точку достижения максимума апостериорной надежности, а вместе с тем и байесовскую оценку σ :

$$\hat{\sigma}_G = \sqrt{\frac{2\Delta A}{\gamma \ln((1+\Delta)/(1-\Delta))}}.$$

Легко проверить, что при стремлении обеих параметров a и λ априорного распределения к нулю мы получаем минимаксную оценку σ , равную

$$\hat{\sigma}_n = S \sqrt{\frac{2\Delta}{\ln((1+\Delta)/(1-\Delta))}},$$

где

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2.$$

Дело в том, что предел апостериорной надежности не зависит от T , следовательно, совпадает с пределом априорной надежности, наконец, предел априорной надежности совпадает с надежностью оценки $\hat{\sigma}_n$, которая не зависит от σ и равна

$$H(n, \Delta) = K_{n-1} \left((n-1) b(\Delta) (1+\Delta)^2 \right) - K_{n-1} \left((n-1) b(\Delta) (1-\Delta)^2 \right),$$

где

$$b(\Delta) = \frac{1}{2\Delta} \ln \frac{1+\Delta}{1-\Delta}.$$

Ссылка на Теорему 5.2 о минимаксе завершает построение минимаксной оценки. Постоянство ее риска влечет независимость правила остановки от результатов статистического эксперимента. Следовательно, пренебрегая тривиальной рандомизацией при определении объема наблюдений, находим минимальный объем n^* наблюдений, обеспечивающий относительную

точность Δ и надежность $1 - \alpha$ процедуры оценивания, как минимальное целое n , удовлетворяющее неравенству $H(n, \Delta) \geq 1 - \alpha$.

Заметим, что последнее неравенство имеет асимптотическое решение, если воспользоваться известной нормальной аппроксимацией хи-квадрат распределения, имеющей точность $O(1/\sqrt{n})$. При $\Delta \rightarrow 0$ минимальный объем n^* имеет асимптотику:

$$n^* \sim \left(\frac{\Phi^{-1}(1 - \alpha/2)}{\Delta \sqrt{2}} \right)^2.$$

12.2. Оптимальный последовательный критерий, гарантирующий заданные ограничения на вероятности ошибок. При различении двух простых гипотез (H_0 : выбор происходит из распределения с плотностью f_0 и H_1 : плотность наблюдаемой случайной величины f_1) правило принятия решения (критерий) Неймана–Пирсона является минимаксным при любом правиле останова. Оптимальное по объему наблюдений правило останова было найдено Вальдом и оно так же, как и критерий Неймана–Пирсона, определяется отношением правдоподобия. Вся процедура проведения наблюдений и принятия решения называется *последовательным критерием отношения вероятностей* (сокращенно ПКОВ) и определяется следующим образом.

Для каждого шага $n = 1, 2, \dots$ статистического эксперимента введем статистику отношения правдоподобия

$$L_n(X^{(n)}) = \prod_{k=1}^n \frac{f_1(X_k)}{f_0(X_k)},$$

выберем две неотрицательные константы A_0 и A_1 , $A_0 < 1 < A_1$, (правило их выбора будет указано ниже) и определим нерандомизированное правило останова статистического эксперимента посредством момента останова

$$\nu = \min\{n : L_n(X^{(n)}) \leq A_0 \text{ или } L_n(X^{(n)}) > A_1\}. \quad (12.1)$$

Правило принятия решения определяется следующим образом: если после останова эксперимента оказалось, что $L_n(X^{(n)}) \leq A_0$, то принимается гипотеза H_0 , если же $L_n(X^{(n)}) > A_1$, то принимается H_1 .

В конце сороковых годов прошлого века Вальд и Вольфовиц показали, что ПКОВ обладает следующим оптимальным свойством.

Теорема 12.3. Среди всех процедур различения гипотез H_0 и H_1 , вероятности ошибок первого и второго рода которых не превосходят соответствующих вероятностей последовательного критерия отношения вероятностей, последний обладает минимальными значениями $\mathbf{E}_0 \nu$ и $\mathbf{E}_1 \nu$.

Доказательство этой теоремы более чем сложно и мы не будем им заниматься. Заметим только, что если можно выбрать константы A_0 и A_1 таким образом, чтобы вероятности ошибок первого и второго рода ПКОВ были равны заданным значениям α_0 и α_1 , то ПКОВ минимизирует средний объем наблюдений при справедливости обеих гипотез среди всех (α_0, α_1) -гарантийных процедур проверки простых гипотез. На этот счет имеется теорема, что в случае непрерывного распределения статистики $L_n(X^{(n)})$ при каждом фиксированном $n (= 1, 2, \dots)$ существуют константы $A_0 = A_0(\alpha_0, \alpha_1)$ и $A_1 = A_1(\alpha_0, \alpha_1)$, на которых вероятности ошибок ПКОВ равны предписанным значениям α_0 и α_1 . В случае же дискретного распределения $L_n(X^{(n)})$ этого можно достичь, осуществляя рандомизацию в принятии решения на границе области продолжения эксперимента, а также допуская принятие решения о справедливости той или иной гипотезы без проведения наблюдений (естественно, не с вероятностью единица). Как это сделать и останется ли после этого ПКОВ оптимальным по объему наблюдений критерием, науке не известно.

Тем не менее, существует довольно простой способ приближенного определения $A_0 = A_0(\alpha_0, \alpha_1)$ и $A_1 = A_1(\alpha_0, \alpha_1)$, причем теоретические изыскания в этом направлении показывают, что полученные таким образом границы асимптотически точны при $\alpha_0, \alpha_1 \rightarrow 0$. Изложенный ниже метод определения границ основан на *конечности* момента остановки ν , задаваемого формулой (12.1). Положим

$$Z_k = \ln \frac{f_1(X_k)}{f_0(X_k)}, \quad k = 1, 2, \dots,$$

и докажем, что справедливо

Предложение 12.1. Если распределение наблюдаемой случайной величины X_1 таково, что $\mathbf{P}(Z_1 = 0) < 1$, то существуют такие $0 < \delta < 1$ и $C > 0$, что

$$\mathbf{P}(\nu \geq n) \leq C\delta^n.$$

Доказательство. Пусть $c = \ln A_1 - \ln A_0$ и допустим сначала, что $\mathbf{P}(|Z_1| \leq c) = p < 1$. Событие $\nu \geq n$ означает, что

$$\ln A_0 < \sum_{i=1}^k Z_i < \ln A_0$$

при всех $k = 1, \dots, n-1$ и, следовательно, $|Z_k| \leq c$ при всех $k = 1, \dots, n-1$. Таким образом, $\mathbf{P}(\nu \geq n) \leq p^{n-1}$ и утверждение предложения выполняется с $C = p^{-1}$ и $\delta = p$. Если же $\mathbf{P}(|Z_1| \leq c) = 1$, то существует такое, возможно, достаточно большое, но конечное r , что

$$\mathbf{P}\left(\left|\sum_{i=1}^r Z_i\right| \leq c\right) = p < 1.$$

Разбивая всю последовательность $\{Z_i, i \geq 1\}$ на куски длины r , убеждаемся (по аналогии с предыдущим, где Z_k заменяется на сумму из r аналогичных слагаемых), что $\mathbf{P}(\nu \geq rm) \leq p^{m-1}$. Следовательно,

$$\mathbf{P}(\nu \geq n) \leq p^{\lceil n/r \rceil - 1} \leq p^{n/r - 2},$$

и мы снова получаем утверждение предложения с $C = p^{-2}$ и $\delta = p^{1/r}$. \square

Обратимся теперь к методу приближенного определения границ $A_0 = A_0(\alpha_0, \alpha_1)$ и $A_1 = A_1(\alpha_0, \alpha_1)$. Пусть

$$p_{in} = p_{in}(x^{(n)}) = \prod_{k=1}^n f_i(x_k), \quad i = 0, 1,$$

μ_n – мера, по которой вычисляются данные n -мерные функции плотности, R_n – часть выборочного пространства, определяемая неравенствами

$$A_0 < \frac{p_{1k}}{p_{0k}} < A_1 \quad \text{при} \quad k = 1, \dots, n-1 \quad \text{и} \quad A_1 \leq \frac{p_{1n}}{p_{0n}}.$$

Это множество тех выборочных данных x_1, \dots, x_n , для которых статистический эксперимент заканчивается на шаге с номером n , после чего принимается гипотеза H_1 . Тогда вероятность ошибки первого рода для ПКОВ

$$a_0 = \sum_{n=1}^{\infty} \int_{R_n} p_{0n} d\mu_n \leq \frac{1}{A_1} \sum_{n=1}^{\infty} \int_{R_n} p_{1n} d\mu_n = \frac{1 - a_1}{A_1},$$

где a_1 – вероятность ошибки второго рода. Аналогично, если S_n есть часть выборочного пространства, для которой $\nu = n$ и принимается гипотеза H_0 , то

$$1 - a_0 = \sum_{n=1}^{\infty} \int_{S_n} p_{0n} d\mu_n \geq \frac{a_1}{A_0}.$$

Естественно, все эти неравенства справедливы лишь при конечности момента остановки ν , что было установлено в Предложении 12.1:

$$\sum_{n=1}^{\infty} \mathbf{P}_i(\nu = n) = \sum_{n=1}^{\infty} \int_{R_n \cup S_n} p_{in} d\mu_n = 1 \quad \text{для } i = 0, 1.$$

Неравенства

$$A_0 \geq \frac{a_1}{1 - a_0}, \quad A_1 \leq \frac{1 - a_1}{a_0}$$

наводят на мысль об аппроксимации границ A_0 и A_1 , соответствующих заданным α_0 и α_1 , величинами

$$A'_0 = \frac{\alpha_1}{1 - \alpha_0}, \quad A'_1 = \frac{1 - \alpha_1}{\alpha_0}.$$

Именно такие границы используются в практических применениях ПКОВ, и как показывают опыт и результаты статистического моделирования, это практически приемлемые приближения при α_0 и α_1 порядка 0,01–0,05.

Понятно, что на практике лучше использовать логарифмический вариант статистики отношения правдоподобия, то есть определять область продолжения наблюдений неравенствами

$$\ln \frac{\alpha_1}{1 - \alpha_0} < \sum_{k=1}^n \ln \frac{f_1(x_k)}{f_0(x_k)} < \ln \frac{1 - \alpha_1}{\alpha_0}. \quad (12.2)$$

Таким образом, приближенно гарантийная процедура последовательного различения простых гипотез останавливает статистический эксперимент,

когда впервые нарушаются неравенства (12.2). Если в момент остановки эксперимента ($\nu = n$) логарифмическое отношение правдоподобия

$$\mathcal{L}_n = \mathcal{L}_n(x^{(n)}) = \sum_{k=1}^n \ln \frac{f_1(x_k)}{f_0(x_k)}$$

оказалось на/или ниже нижней границы области продолжения наблюдений, то принимается гипотеза H_0 , если же на/или выше верхней, то – H_1 .

Вычислим средние значения момента остановки $E_i \nu$, $i = 0, 1$, предполагая, что в момент остановки эксперимента значение статистики \mathcal{L}_n в точности равно одной из границ, то есть получим приближенное значение для среднего объема наблюдений, когда пренебрегаем величиной перескока статистикой \mathcal{L}_n границ области (12.2) в момент остановки.

Используя Предложение 3.3, устанавливающее свойство аддитивности различающей информации в последовательно планируемых экспериментах, находим, что

$$\mathbf{E}_0 \mathcal{L}_\nu(X^{(\nu)}) = -I(0, 1 | \xi) \mathbf{E}_0 \nu, \quad \mathbf{E}_1 \mathcal{L}_\nu(X^{(\nu)}) = I(1, 0 | \xi) \mathbf{E}_1 \nu,$$

где, как и в § 3,

$$I(i, 1 - i | \xi) = \mathbf{E}_i \ln \frac{f_i(\xi)}{f_{1-i}(\xi)}, \quad i = 0, 1, -$$

информация по Кульбаку–Лейблеру (различающая информация), содержащая в наблюдении случайной величины ξ в пользу гипотезы H_i против H_{1-i} , $i = 0, 1$.

С другой стороны, поскольку по предположению статистика \mathcal{L}_ν может принимать только два значения, равные правой и левой частям неравенства (12.2), с вероятностями, определяемыми заданными α_0 и α_1 , то

$$\begin{aligned} \mathbf{E}_0 \mathcal{L}_\nu(X^{(\nu)}) &= (1 - \alpha_0) \ln \frac{\alpha_1}{1 - \alpha_0} + \alpha_0 \ln \frac{1 - \alpha_1}{\alpha_0} \\ \mathbf{E}_1 \mathcal{L}_\nu(X^{(\nu)}) &= \alpha_1 \ln \frac{\alpha_1}{1 - \alpha_0} + (1 - \alpha_1) \ln \frac{1 - \alpha_1}{\alpha_0}. \end{aligned}$$

Следовательно,

$$\mathbf{E}_0 \nu = \frac{(1 - \alpha_0) \ln [(1 - \alpha_0)/\alpha_1] + \alpha_0 \ln [\alpha_0/(1 - \alpha_1)]}{I(0, 1 | \xi)},$$

$$\mathbf{E}_1 \nu = \frac{\alpha_1 \ln [\alpha_1]/(1 - \alpha_0] + (1 - \alpha_1) \ln [(1 - \alpha_1)/\alpha_0]}{I(1, 0 | \xi)}.$$

Полученные приближенные значения для среднего объема выборки совпадают с нижними границами (3.11)–(3.12) для любой гарантийной процедуры различения двух простых гипотез. Таким образом, мы получили нестрогое подтверждение оптимальности ПКОВ с точки зрения минимальности объема наблюдений.

Рассмотрим теперь задачу различения сложных гипотез $H_0 : \theta \leq \theta_0$ и $H_1 : \theta > \theta_0$ с точки зрения возможности применить ПКОВ в этом случае. Итак, предполагается, что плотность наблюдаемой случайной величины зависит от действительного параметра θ . Заметим, что на практике обычно не интересуются мощностью критерия при альтернативах $\theta (> \theta_0)$, близких к границе θ_0 , но желают контролировать вероятность обнаружения альтернатив, достаточно удаленных от θ_0 . Критерий, следовательно, должен удовлетворять в терминах его функции мощности $m(\theta)$, $\theta \in \Theta \subset \mathbb{R}$, условиям $m(\theta) \leq \alpha_0$ при $\theta \leq \theta_0$ и $m(\theta) \geq 1 - \alpha_1$, когда $\theta \geq \theta_1$, где $\theta_1 (> \theta_0)$ – некоторое фиксированное альтернативное значение параметра θ . Интервал (θ_0, θ_1) обычно называется *областью безразличия* – предполагается, что когда истинное значение θ принадлежит этому интервалу, то статистик не несет существенных потерь, принимая гипотезу $\theta \leq \theta_0$.

Оказывается, и это можно доказать, в случае статистических моделей с монотонным относительно некоторой статистики $T_n = T_n(X^{(n)})$ отношением правдоподобия ПКОВ, как критерий проверки простой гипотезы $H_0 : \theta = \theta_0$ при простой альтернативе $H_1 : \theta = \theta_1$, имеет монотонно возрастающую функцию мощности, для которой, как и для среднего объема наблюдений, существуют достаточно простые приближенные формулы. Таким образом, ПКОВ с вероятностями ошибок α_0 и α_1 решает поставленную задачу гарантийного различения сложных гипотез, минимизируя средний объем наблюдений, когда истинное значение параметра θ равно θ_0 или θ_1 . Но теперь нас интересует $\mathbf{E}_\theta \nu$ при всех $\theta \in \Theta$. Обычно функция $\mathbf{E}_\theta \nu$, $\theta \in \Theta$, имеет максимум между θ_0 и θ_1 и убывает, когда θ убывает от точки максимума в любом направлении. К сожалению, максималь-

ное значение $\mathbf{E}_\theta \nu$ бывает больше, чем наименьший фиксированный объем наблюдений, при котором существует (α_0, α_1) -гарантийный критерий с тривиальным моментом остановки $\nu = n$. Важная проблема построения (α_0, α_1) -гарантийного критерия, который минимизировал бы $\sup_{\theta \in \Theta} \mathbf{E}_\theta \nu$, до сих пор не получила надлежащего решения.

В заключение отметим, что все полученные выше результаты для характеристик ПКОВ носят асимптотический характер: границы для области продолжения наблюдений и средние значения объема наблюдений асимптотически точны, когда $\theta_1 \rightarrow \theta_0$ или $\alpha_0, \alpha_1 \rightarrow 0$.

12.3. Асимптотика минимального объема выборки, при котором существует θ -гарантийный критерий. Средний объем наблюдений в последовательном гарантийном критерии отношения вероятностей обычно сравнивают с так называемым *необходимым объемом выборки* – минимальным числом наблюдений n^* , обеспечивающим существование (α_0, α_1) -гарантийного критерия в классе всех критериев с моментом остановки, не зависящим от результатов статистического эксперимента.

Как и в конце предыдущего пункта 12.2, рассмотрим задачу различения сложных гипотез $H_0 : \theta \leq \theta_0$ и $H_1 : \theta > \theta_0$ в предположении, что плотность наблюдаемой случайной величины ξ зависит от действительного параметра θ и что *статистическая модель обладает монотонным отношением правдоподобия*. Рассматривается класс критериев, основанных на выборках фиксированного объема $n = 1, 2, \dots$. Требуется определить наименьшее число наблюдений $n^* = n^*(\theta_0, \theta_1)$, при котором существует критерий из заданного класса, функция мощности которого удовлетворяет условиям $m(\theta) \leq \alpha_0$ при $\theta \leq \theta_0$ и $m(\theta) \geq 1 - \alpha_1$, когда $\theta \geq \theta_1$, где $\theta_1 (> \theta_0)$ – некоторое фиксированное альтернативное значение параметра θ .

Асимптотика n^* будет исследоваться при выполнении следующих условий регулярности.

(А) *Статистический эксперимент регулярен в точке θ_0 , то есть функция плотности $f(x | \theta)$ удовлетворяет условиям (А1)–(А3) из § 3 в некоторой окрестности $U(\theta_0)$ точки θ_0 (эти условия использовались при определении информации по Фишеру $i(\theta | \xi)$).*

(В) Случайная величина $\Lambda(\xi) = \ln [f(\xi | \theta_1)/f(\xi | \theta_0)]$ обладает конечным абсолютным моментом третьего порядка $\mathbf{E}_\theta |\Lambda(\xi)|^3$ при всех $\theta \in U(\theta_0)$.

Теорема 12.3. Пусть

$$\tilde{n} = \tilde{n}(\theta_0, \theta_1) = \frac{[\Phi^{-1}(1 - \alpha_0) + \Phi^{-1}(1 - \alpha_1)]^2}{(\theta_1 - \theta_0)^2 i(\theta_0 | \xi)}. \quad (12.3)$$

Тогда при выполнении условий (А) и (В)

$$\lim_{\theta_1 \rightarrow \theta_0} \frac{n^*(\theta_0, \theta_1)}{\tilde{n}(\theta_0, \theta_1)} = 1.$$

Доказательству теоремы предположим лемму технического характера, устанавливающую асимптотическую ($n \rightarrow \infty$) нормальность статистики отношения правдоподобия

$$\mathcal{L}_n = \mathcal{L}_n(X^{(n)}) = \sum_{k=1}^n \ln \frac{f(X_k | \theta_1)}{f_0(X_k | \theta_0)}$$

и асимптотику при $\theta_1 \rightarrow \theta_0$ ее среднего значения $\mathbf{E}_\theta \mathcal{L}_n = n\mu(\theta) = n \mathbf{E}_\theta \Lambda(\xi)$ и дисперсии $\mathbf{D}_\theta \mathcal{L}_n = n\sigma^2(\theta) = n \mathbf{D}_\theta \Lambda(\xi)$, когда значение параметра $\theta = \theta_i$, $i = 0, 1$.

Лемма 12.1. При выполнении условия (В) для функции распределения

$$F_{\theta,n}(z) = P \left(\frac{\mathcal{L}_n - n\mu(\theta)}{\sigma(\theta)\sqrt{n}} < z \right)$$

нормированной статистики логарифмического отношения правдоподобия имеет место представление

$$F_{\theta,n}(z) = \Phi(z) + C(n, \theta)/\sqrt{n}, \quad (12.4)$$

где $C(n, \theta)$ – ограниченная функция целочисленного аргумента n и $\theta \in U(\theta_0)$.

При выполнении условия (А) и $\theta_1 \rightarrow \theta_0$ справедливы асимптотические формулы

$$\mu(\theta_0) = -\frac{\Delta^2}{2} i(\theta_0 | \xi) + o(\Delta^2), \quad \mu(\theta_1) = \frac{\Delta^2}{2} i(\theta_0 | \xi) + o(\Delta^2),$$

$$\sigma^2(\theta_0) \sim \sigma^2(\theta_1) = \Delta^2 i(\theta_0 | \xi) + o(\Delta^2),$$

где $\Delta = \theta_1 - \theta_0$.

Доказательство. Представление (12.4) есть следствие известной предельной теоремы Берри–Эссена для сумм независимых одинаково распределенных случайных величин. Наличие третьего момента у слагаемых позволяет представить распределение суммы в виде (12.4) при любых конечных n , хотя остаточный член в этом представлении характеризует равномерную скорость сходимости F_n к нормальному закону при $n \rightarrow \infty$.

Асимптотические формулы для средних значений $\mu(\theta_0)$ и $\mu(\theta_1)$ есть частный случай Леммы 3.1, устанавливающей связь между различающей и точечной информацией. Асимптотические формулы для дисперсий доказываются применением формулы Тейлора к функции $\ln(1+x)$:

$$\begin{aligned} \sigma^2(\theta_0) &= \mathbf{E}_{\theta_0} \Lambda^2(\xi) + O(\Delta^4) = \mathbf{E}_{\theta_0} \left[\ln \left(1 - \left(1 - \frac{f(\xi | \theta_1)}{f(\xi | \theta_0)} \right) \right) \right]^2 + O(\Delta^4) = \\ &= \mathbf{E}_{\theta_0} \left[B(\xi) + \frac{B^2(\xi)}{2(1 - \lambda \Lambda(\xi))^2} \right]^2 + O(\Delta^4), \end{aligned}$$

где $0 < \lambda < 1$ и в силу условия (A) (сравните с доказательством Леммы 3.1)

$$B(\xi) = \frac{f(\xi | \theta_1) - f(\xi | \theta_0)}{f(\xi | \theta_0)} = \Delta \left[\frac{\partial \ln f(\xi | \theta)}{\partial \theta} \right]_{\theta=\theta_0} + o(\Delta).$$

Асимптотическая формула для $\sigma^2(\theta_1)$ выводится аналогичным образом.

□

Доказательство Теоремы 12.3. Во избежание тривиальных затруднений будем предполагать, что F_n – непрерывная, строго возрастающая функция аргумента z при любом $\theta \in U(\theta_0)$. Это предположение влечет существование нерандомизированного критерия отношения правдоподобия для различения простых гипотез $H'_0 : \theta = \theta_0$ и $H'_1 : \theta = \theta_1$, размер которого равен заданному уровню значимости α_0 . Можно убедиться, что дискретность распределения статистики \mathcal{L}_n не влияет на справедливость утверждения теоремы.

Поскольку рассматриваются только статистические модели, обладающие монотонным отношением правдоподобия, то задача гарантийного различения сложных гипотез $H_0 : \theta \leq \theta_0$ и $H_1 : \theta > \theta_0$ сводится к различению простых гипотез $H'_0 : \theta = \theta_0$ и $H'_1 : \theta = \theta_1$, для которых критерий отношения правдоподобия с критической областью $\mathcal{L}_n > C$ является наиболее мощным, и, следовательно, именно на нем достигается минимальный, обеспечивающий гарантийность объем наблюдений.

Критерий определяется заданием двух величин C и n , которые должны быть выбраны так, чтобы вероятности ошибок не превосходили заданных значений α_0 и α_1 :

$$\mathbf{P}_{\theta_0}(\mathcal{L}_n \leq C) = F_{\theta_0, n} \left(\frac{C - n\mu(\theta_0)}{\sigma(\theta_0)\sqrt{n}} \right) \geq 1 - \alpha_0,$$

$$\mathbf{P}_{\theta_1}(\mathcal{L}_n \leq C) = F_{\theta_1, n} \left(\frac{C - n\mu(\theta_1)}{\sigma(\theta_1)\sqrt{n}} \right) \leq \alpha_1 .$$

Если пара (C, n) , удовлетворяет этим условиям, то в силу монотонности функций распределения $C_n(\alpha_0) \leq C \leq C_n(\alpha_1)$, где

$$C_n(\alpha_0) = n\mu(\theta_0) + \sigma(\theta_0)F_{\theta_0, n}^{-1}(1 - \alpha_0)\sqrt{n}, \quad C_n(\alpha_1) = n\mu(\theta_1) + \sigma(\theta_1)F_{\theta_1, n}^{-1}(\alpha_1)\sqrt{n}.$$

Более того, если $C_n(\alpha_0) \leq C' \leq C_n(\alpha_1)$, то пара чисел (C', n) также соответствует гарантийному критерию.

Если при заданном n гарантийный критерий существует, то это означает, что $C_n(\alpha_0) \leq C_n(\alpha_1)$, то есть

$$\sqrt{n}(\mu(\theta_1) - \mu(\theta_0)) \geq \sigma(\theta_0) F_{\theta_0, n}^{-1}(1 - \alpha_0) - \sigma(\theta_1) F_{\theta_1, n}^{-1}(\alpha_1) .$$

Несуществование гарантийного критерия равносильно противоположному неравенству. Так как критерий Неймана–Пирсона наиболее мощный, то минимальное n , которое обеспечивает существование гарантийного критерия, должно удовлетворять также неравенству

$$\sqrt{n-1}(\mu(\theta_1) - \mu(\theta_0)) < \sigma(\theta_0) F_{\theta_0, n-1}^{-1}(1 - \alpha_0) - \sigma(\theta_1) F_{\theta_1, n-1}^{-1}(\alpha_1) .$$

Но $\mu(\theta_1) - \mu(\theta_0) > 0$, и поэтому необходимый объем выборки n^* определяется неравенствами

$$\left(\frac{\sigma(\theta_0) F_{\theta_0, n^*}^{-1}(1 - \alpha_0) - \sigma(\theta_1) F_{\theta_1, n^*}^{-1}(\alpha_1)}{\mu(\theta_1) - \mu(\theta_0)} \right)^2 \leq n^* <$$

$$1 + \left(\frac{\sigma(\theta_0) F_{\theta_0, n^*-1}^{-1}(1 - \alpha_0) - \sigma(\theta_1) F_{\theta_1, n^*-1}^{-1}(\alpha_1)}{\mu(\theta_1) - \mu(\theta_0)} \right)^2.$$

Если $\Delta = \theta_1 - \theta_0 \rightarrow 0$, то в силу Леммы 12.1 эти неравенства приобретают следующий асимптотический вид (напомним, $\Phi^{-1}(p) = -\Phi^{-1}(1 - p)$):

$$\frac{[\Phi^{-1}(1 - \alpha_0) + \Phi^{-1}(1 - \alpha_1) + C_1/\sqrt{n^*}]^2}{\Delta^2 i(\theta_0 | \xi)} \leq n^* < 1 + \frac{[\Phi^{-1}(1 - \alpha_0) + \Phi^{-1}(1 - \alpha_1) + C_2/\sqrt{n^*}]^2}{\Delta^2 i(\theta_0 | \xi)}, \quad (12.5)$$

где $C_i, i = 1, 2$, – ограниченные функции θ_1 и n .

То, что $n^* \sim \tilde{n}$, когда $\Delta \rightarrow 0$, устанавливается теперь достаточно просто. Подставим в (12.5) $n^* = \tilde{n}(1 + \varepsilon)$, где \tilde{n} заменяется правой частью (12.3). Тривиальные выкладки, связанные с такой подстановкой, приводят к замене неравенств (12.5) на неравенство $C_3\Delta \leq \varepsilon < 1 + C_4\Delta$, где C_3 и C_4 – ограниченные функции. Последнее неравенство, очевидно, влечет утверждение теоремы. \square

Представляет интерес сравнение асимптотики необходимого объема выборки n^* со средними значениями объема наблюдений в последовательном критерии отношения вероятностей. Для проблемы различения гипотез $H'_0 : \theta = \theta_0$ и $H'_1 : \theta = \theta_1$, когда $\Delta \rightarrow 0$, формулы для среднего объема наблюдений в ПКОВ приобретают вид

$$\begin{aligned} \mathbf{E}_{\theta_0} \nu &= \frac{(1 - \alpha_0) \ln [(1 - \alpha_0)/\alpha_1] + \alpha_0 \ln [\alpha_0/(1 - \alpha_1)]}{I(\theta_0, \theta_1 | \xi)} \sim \\ &2 \frac{(1 - \alpha_0) \ln [(1 - \alpha_0)/\alpha_1] + \alpha_0 \ln [\alpha_0/(1 - \alpha_1)]}{(\theta_1 - \theta_0)^2 i(\theta_0 | \xi)}, \\ \mathbf{E}_{\theta_1} \nu &= \frac{\alpha_1 \ln [\alpha_1]/(1 - \alpha_0) + (1 - \alpha_1) \ln [(1 - \alpha_1)/\alpha_0]}{I(\theta_1, \theta_0 | \xi)} \sim \\ &2 \frac{\alpha_1 \ln [\alpha_1]/(1 - \alpha_0) + (1 - \alpha_1) \ln [(1 - \alpha_1)/\alpha_0]}{(\theta_1 - \theta_0)^2 i(\theta_0 | \xi)}. \end{aligned}$$

Таким образом, асимптотическая ($\Delta \rightarrow 0$) эффективность ПКОВ по отношению к необходимому объему наблюдений определяется величинами

$$\frac{n^*}{\mathbf{E}_{\theta_0} \nu} = \frac{1}{2} \frac{[\Phi^{-1}(1 - \alpha_0) + \Phi^{-1}(1 - \alpha_1)]^2}{(1 - \alpha_0) \ln [(1 - \alpha_0)/\alpha_1] + \alpha_0 \ln [\alpha_0/(1 - \alpha_1)]},$$

$$\frac{n^*}{\mathbf{E}_{\theta_1} \nu} = \frac{1}{2} \frac{[\Phi^{-1}(1 - \alpha_0) + \Phi^{-1}(1 - \alpha_1)]^2}{\alpha_1 \ln[\alpha_1/(1 - \alpha_0)] + (1 - \alpha_1) \ln[(1 - \alpha_1)/\alpha_0]}.$$

Численные расчеты для α_0 и α_1 порядка нескольких сотых показывают, что использование ПКОВ вместо критерия Неймана–Пирсона сокращает объем наблюдений в полтора–два раза.

12.4. D-гарантийные процедуры статистического вывода. Во введении к данному параграфу говорилось, что при построении оптимальных по объему наблюдений d-гарантийных процедур ограничения $r(d)$ накладываются на функцию d-риска: $R_G(\varphi | d) \leq r(d)$, и при этом условии минимизируется, скажем, безусловное среднее значение $\mathbf{E}_G \nu$ объема наблюдений или $\sup_{d \in \mathcal{D}} \mathbf{E} \{ \nu | \delta(X^{(\nu)}) = d \}$. Однако данная проблема до сих пор не решена. Для d-гарантийных критериев проверки гипотез $H_0 : \theta \leq \theta_0$ и $H_1 : \theta > \theta_0$, основанных на фиксированном числе наблюдений, получены асимптотические формулы для необходимого объема выборки, причем асимптотики проводятся в двух случаях: (1) априорное распределение стягивается в точку θ_0 , разграничивающую гипотезы, и (2) ограничения β_0, β_1 на d-риски стремятся к нулю. Вывод таких асимптотических формул сопряжен со значительными трудностями как в математическом, так и в концептуальном плане. Однако существует удивительно простой способ построения d-гарантийных процедур практически для любой проблемы статистического вывода. Суть его состоит в следующем.

Любой процедуре $\varphi = (\varphi_s, \varphi_c, \varphi_d)$ статистического вывода можно сопоставить ее d-гарантийную модификацию, изменив только правило остановки φ_s . Сделать это можно следующим образом.

Пусть $r(d)$, $d \in \mathcal{D}$, – заданные ограничения на d-риск. Напомним, процедура φ называется d-гарантийной, если нормированный d-риск (функция d-риска, поделенная на $r(d)$)

$$\bar{R}_G(\varphi | d) \leq 1, \quad \forall d \in \{d : \psi_G(d) > 0, d \in \mathcal{D}\}.$$

Нормированный d-риск есть условное математическое ожидание относительно решающей функции $\delta(X^{(\nu)})$ от нормированного апостериорного рис-

ка, который мы запишем в виде

$$\bar{\mathfrak{R}}(\delta(X^{(\nu)}) | X^{(\nu)}) = \int_{\Theta} L(\theta, \delta(X^{(\nu)})) p(X^{(\nu)} | \theta) dG(\theta) / \left[r(\delta(X^{(\nu)})) p_G(X^{(\nu)}) \right].$$

Такой вид апостериорного риска без усреднения потерь по решающему правилу φ_d мы использовали при построении байесовских решений, только вместо $\delta(X^{(\nu)})$ писалось просто d . На каждом шаге статистического эксперимента с номером $n = 1, 2, \dots$ после получения результата $x^{(n)}$ наблюдения $X^{(n)}$ мы можем применить правило φ_d и выдать некоторое решение $d_n = \delta(x^{(n)})$, проводя, если это необходимо, рандомизацию. Определим теперь новый (нерандомизированный) момент остановки

$$\bar{\nu} = \min\{ n : \bar{\mathfrak{R}}(d_n | x^{(n)}) \leq 1 \},$$

то есть будем проводить наблюдения до тех пор, пока нормированный апостериорный риск при заданных результатах наблюдений и решением, выданным правилом φ_d , не станет меньше 1. Очевидно, *новая процедура* $\bar{\varphi} = (\bar{\varphi}_s, \varphi_c, \varphi_d)$, где $\bar{\varphi}_s$ – *правило остановки с моментом остановки $\bar{\nu}$ является d -гарантийной*.

Естественно, если еще заменить правила φ_c и φ_d на байесовские при нормированной на $r(d)$ функции потерь, то есть уменьшить нормированный апостериорный риск на каждом шаге эксперимента, то мы получим момент остановки, который прекращает наблюдения на более ранних шагах эксперимента с сохранением d -гарантийности статистического вывода. Такую процедуру проведения наблюдений и принятия решений будем называть *процедурой первого перескока*. Посмотрим, как выглядит эта процедура для классических задач оценки и проверки гипотез в рамках модели N–N.

ПРИМЕР 12.3. *D-гарантийная оценка среднего значения нормального распределения.* Пусть наблюдаемая последовательность X_1, X_2, \dots состоит из независимых, одинаково распределенных $N(\theta, \sigma^2)$ случайных величин. Задача состоит в оценке среднего значения θ нормального распределения с известной дисперсией σ^2 , когда не известное значение θ есть реализация

случайной величины $\vartheta \sim N(\mu, \tau^2)$. Оценка $\hat{\theta}_\nu = \hat{\theta}(X^{(\nu)})$ должна быть d-гарантийной – удовлетворять заданным требованиям к точности Δ и “d-надежности,” то есть должно выполняться неравенство

$$\mathbf{P} \left(|\hat{\theta}_\nu - \vartheta| \leq \Delta \mid \hat{\theta}(X^{(\nu)}) = d \right) \geq 1 - \alpha, \quad \forall d \in \mathcal{D} = \Theta = \mathbb{R}.$$

Таким образом, мы имеем проблему оценки действительного параметра при функции потерь 1–0 в рамках модели N–N, которую мы рассматривали в § 4 с байесовской точки зрения. Ограничения $r(d) = \alpha$ не зависят от $d \in \Theta$, поэтому нормировать апостериорный риск не стоит, к тому же удобнее оперировать не риском, а надежностью оценки. Поскольку апостериорное распределение ϑ есть нормальное с параметрами (см. § 4)

$$\mathcal{M} = \frac{n\bar{X}\sigma^{-2} + \mu\tau^{-2}}{n\sigma^{-2} + \tau^{-2}} \quad \text{и} \quad \mathcal{S}^2 = \frac{1}{n\sigma^{-2} + \tau^{-2}},$$

а байесовская оценка θ есть апостериорное среднее \mathcal{M} , то апостериорная надежность не зависит от результатов наблюдений и, следовательно, совпадает с d-надежностью.

Итак, процедура первого перескока в данной статистической проблеме оказалась основанной на фиксированном числе наблюдений n^* , которое определяется как наименьшее целое число, удовлетворяющее неравенству $2\Phi(\Delta/\mathcal{S}) - 1 \geq 1 - \alpha$, откуда

$$n^* = \left[\left(\frac{\Phi^{-1}(1 - \alpha/2)}{\Delta} \sigma \right)^2 - \frac{\sigma^2}{\tau^2} \right] + 1.$$

Получается то, что очень похоже на оценку с гарантией малости θ -риска, однако необходимый для обеспечения d-гарантийности объем наблюдений становится меньше, что особенно ощутимо при малых значениях дисперсии τ^2 априорного распределения. Действительно, если τ^2 очень мало, то мы априори знаем, что θ практически совпадает с μ . Более того, если выражение в квадратных скобках становится отрицательным, то следует принимать решение $\theta = \mu$ без проведения каких-либо наблюдений.

ПРИМЕР 12.4. D-гарантийная процедура первого перескока в проблеме различения двух сложных гипотез о среднем значении нормального

распределения. Рассматривается та же задача, что и в примере 11.2 предыдущего параграфа, но теперь ставится задача контролирования d-риска не только первого рода, но и второго.

Итак, проверяется гипотеза $H_0 : \theta \leq \theta_0$ при альтернативе $H_1 : \theta > \theta_0$, по выборке из нормального (θ, σ^2) распределения, когда θ есть реализация случайной величины ϑ , имеющей нормальное распределение с параметрами (μ, τ^2) . Требуется построить критерий первого перескока, гарантирующий заданные ограничения (β_0, β_1) на соответствующие d-риски.

Очевидно, что такой критерий будет останавливать эксперимент, когда впервые или апостериорная вероятность гипотезы H_0 становится меньше β_0 или H_1 – меньше β_1 . В первом случае принимается гипотеза H_1 , во втором – H_0 . Мы только что в предыдущем примере напоминали, что апостериорное распределение ϑ есть нормальное с параметрами $(\mathcal{M}, \mathcal{S}^2)$. Следовательно, момент остановки

$$\bar{\nu} = \min \left\{ n : \Phi \left(\frac{\theta_0 - \mathcal{M}}{\mathcal{S}} \right) \leq \beta_0 \text{ или } \Phi \left(\frac{\theta_0 - \mathcal{M}}{\mathcal{S}} \right) \geq 1 - \beta_1 \right\} ,$$

так что область продолжения наблюдений имеет следующий вид:

$$\theta_0 - \sqrt{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \Phi^{-1}(1 - \beta_1) < \frac{n\bar{X}_n\sigma^{-2} + \mu\tau - 2}{n\sigma^{-2} + \tau^{-2}} < \theta_0 + \sqrt{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \Phi^{-1}(1 - \beta_0) .$$

Таким образом, в задаче d-гарантийного различения гипотез процедура первого перескока оказывается строго последовательной процедурой, момент остановки которой определяется выходом выборочного среднего \bar{X}_n из некоторого интервала.

Результаты статистического моделирования показывают, что процедура первого перескока обладает большой вероятностью остановки на первых шагах эксперимента даже в том случае, когда априорное среднее совпадает с границей различаемых гипотез ($\mu = \theta_0$), но, по всей вероятности, в этом случае $E\nu = \infty$ (гипотеза!). Конечно, если μ сильно отличается от θ_0 , то остановка, как правило, будет происходить после первого наблюдения. Свою лепту в такого рода комментариев вносит и отношение дисперсий σ^2/τ^2 , и вы сами можете судить о воздействии этого отношения на поведение момента остановки ν .

ЗАЧЕТНЫЕ ЗАДАНИЯ

1. Предложите метод для нахождения необходимого объема наблюдений при проверке гипотезы $\theta < \theta_0$ с заданным уровнем значимости α_0 о масштабном параметре θ показательного распределения, когда гарантируется требуемая мощность $1 - \alpha_1$ критерия в области $\theta \geq \theta_1 (> \theta_0)$.
- 2*. В рамках модели N–N найдите явную формулу для определения необходимого объема наблюдений в задаче проверки гипотезы $\theta \leq 0$ при альтернативе $\theta > 0$, когда параметры априорного распределения $\sigma = \tau = 1, \mu = 0$.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Апостериорный риск $\mathcal{R}_\lambda(d|\mathbf{X})$, $d \in \mathcal{D}$ (posterior risk), 17
- Априорный образ $\Psi_\lambda(\cdot)$, $\lambda \in \Lambda$ (prior image), 14
- Априорный риск $R_\varphi(\lambda)$, $\lambda \in \Lambda$ (prior risk), 16
- Асимптотика необходимого объема выборки
(necessary sample size asymptotic), 168
- Байесовская оценка $\hat{\theta}_G$ (Bayes estimator), 49
- Байесовский критерий (Bayes test), 48
- Байесовский риск (Bayes risk), 46
- Байесовское правило (Bayes rule), 46
- Басу теорема (Basu theorem), 30
- Вероятностная модель $(\mathcal{F}, \mathcal{G})$ (probability model), 6
- Выборочное пространство $(\mathcal{X}, \mathcal{A})$ (sample space), 12
- Выходной уровень качества Q_{out} (output quality level), 139
- Гарантийное оценивание (guaranteed estimation), 155
- Гарантийность статистического вывода
(statistical inference guarantee), 17
- Гипотеза (hypothesis), 84
- Группа линейных преобразований (location-scale group), 123
- Группа масштабных преобразований (scale group), 123
- Группа преобразований выборочного пространства
(transformation group of sample space), 122
- Группа сдвигов (location group), 123
- Двухступенчатая процедура Стейна (Stein two-stage procedure), 157
- D-гарантийное оценивание (d-guaranteed estimation), 174
- Доверительная граница (confidence limit/bound), 98
- Доверительное множество (confidence set), 98
- Доверительный интервал (confidence interval), 98
- Доверительный уровень (confidence level), 98
- Достаточная статистика (sufficient statistic), 21
- Достаточная статистика минимальная (minimal sufficient statistic), 24

D-риск первого и второго рода (the first and the second kind d-risk), 143

Инварриантная статистическая проблема

(invariant statistical problem), 124

Инвариантная функция (invariant function), 122

Инвариантный критерий (invariant test), 132

Информация по Кульбаку–Лейблеру I, $(\theta, \vartheta | \mathbf{X})$, $\theta, \vartheta \in \Theta$,
(Kullback-Leibler information), 35

Информация по Фишеру $i(\theta | \mathbf{X})$, $\theta \in \Theta$ (Fisher information), 36

Критерий достаточности – факторизационное тождество

(factorization identity), 21

Критерий с минимальным d-риском (minimal d-risk test), 149

Критерий проверки гипотезы (test), 84

Критерий Стьюдента (Student test), 120

Критерий Фишера (Fisher test), 119

Критическая область (critical region), 84

Критическая функция (critical function), 84

Максимальный инвариант (maximal invariant), 122

Минимаксное решение (minimax decision), 62

Модель В-В (В-В model), 53

Модель N-N (N-N model), 51

Модель P-G (P-G model), 54

Момент остановки – объем выборки ν

(stopping time – sample size), 11

Монотонное отношение правдоподобия

(monotone likelihood ratio), 91

Наиболее мощный инвариантный критерий

(invariant most power test), 132

Наименее благоприятное распределение

(least favorable distribution), 63

Неймановская структура (Neyman structure), 112

Необходимый объем выборки (necessary sample size), 155

Несмещенная оценка (unbiased estimator), 71

Несмещенный критерий (unbiased test), 105

Нижние границы для среднего объема выборки
(lower bounds for mean sample size), 41

Образ стратегии $\{\Psi(\cdot | \theta), \theta \in \Theta\}$ (strategy image), 14

Оперативная характеристика $\{\psi(\cdot | \theta), \theta \in \Theta\}$
(operating characteristic), 14

Оценка надежности оптимальная (optimal estimation of reliability), 74

Оценка с равномерно минимальным d-риском
(uniform minimum d-risk estimator), 75

Оценка с равномерно минимальным риском
(uniform minimum risk estimator), 77

Ошибки первого и второго рода
(the first and the second kind errors), 85

Параметрическое пространство Θ (parametric space), 6

Подчиненная статистика (ancillary statistic), 29

Полная статистика (complete statistic), 25

Полное семейство распределений (complete family of distributions), 25

Последовательный критерий отношения вероятностей
(ratio probability sequential test), 162

Правило выбора φ_c (choice rule), 11

Правило остановки φ_s (stopping rule), 10

Правило принятия решения φ_d (decision rule), 13

Правило управления $\rho = (\varphi_s, \varphi_c)$ (control rule), 11

Проективная оценка (projective estimator), 71

Пространство решений \mathcal{D} (decision space), 5

Процедура первого перескока (the first jump over procedure), 174

Процедура статистического вывода $\varphi = (\varphi_s, \varphi_c, \varphi_d)$
(statistical procedure), 13

Равномерно наиболее мощный критерий
(uniformly most powerful test), 87

Равномерно наиболее мощный несмещенный критерий

(uniformly most powerful unbiased test), 105

Равномерно наиболее точная нижняя/верхняя доверительная граница
(uniformly most accurate lower/upper confidence bound), 101

Равномерно наиболее точное доверительное множество
(uniformly most accurate confidence set), 99

Различающая информация – информация по Кульбаку–Лейблеру
 $I(\theta, \vartheta | \mathbf{X})$ $\theta, \vartheta \in \Theta$ (distinguishing information), 35

Размер критерия (size of test), 86

Рао–Блекуэлла–Колмогорова теорема
(Rao-Blackwell-Kolmogorov theorem), 73

Регулярный статистический эксперимент
(regular statistical experiment), 35

Решающая функция $\delta_\nu(X)$ (decision function), 13

Риск изготовителя α (manufacture risk), 140

Риск потребителя β (consumer risk), 140

Случайная выборка $\mathbf{X} = \mathbf{X}^{(\tau_\nu)}$ (random sample), 12

Статистика отношения правдоподобия (likelihood ratio statistic), 87

Статистическая модель (статистический эксперимент) \mathcal{P}_ρ
(statistical model), 12

Статистическая структура (statistical structure) $(\mathcal{P}_\rho, \mathcal{G})$, 13

Статистический контроль с гарантированным выходным качеством
(statistical control with guaranteed output quality), 152

Статистический эксперимент (statistical experiment), 8

Статистическое U-правило (U-rule), 77

Стратегия $\varphi = (\varphi_s, \varphi_c, \varphi_d)$ (strategy), 13

Теорема о минимаксе (minimax theorem), 67

Точечная информация – информация по Фишеру $i(\theta | \mathbf{X})$, $\theta \in \Theta$
(Fisher information), 36

Траектория группы (orbit of group), 122

Транзитивная группа преобразований (transitive group), 123

Управляющая переменная τ_ν (control variable), 11

Уровень значимости (significance level), 86

Уровень контроля Q_{cont} (control level), 139

Факторизационное тождество (factorization identity), 21

Фундаментальная лемма Неймана–Пирсона
(Neuman-Pearson lemma), 87

Функция d-риска (d-риск) $R_\lambda(\varphi | d)$, $d \in \mathcal{D}$ (d-risk function), 16

Функция мощности (power function), 85

Функция потерь $L(\theta, d)$ (loss function), 7

Функция риска (θ -риск) $R(\varphi | \theta)$, $\theta \in \Theta$ (risk function), 15

Эквивариантная оценка
(equivariant estimator), 126

Эквивариантная решающая функция
(equivariant decision function), 125

Экспоненциальное семейство распределений
(exponential family), 30

Эмпирический байесовский подход
(empirical Bayes approach), 55