Math-Net.Ru

Общероссийский математический портал

М. В. Бурнашев, Ш. Амари, Т. С. Хан, О некоторых задачах проверки гипотез с информационными ограничениями, *Теория вероятн. и ее примен.*, 2000, том 45, выпуск 4, 625–638

DOI: <https://doi.org/10.4213/tvp495>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 52.90.164.192

26 апреля 2023 г., 05:20:44

РОССИЙСКАЯ АКАДЕМИЯ НАУК

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИМЕНЕНИЯ

Основан в январе 1956 г.
Выходит 4 раза в год
Москва • Наука • ТВП

Том 45
Выпуск 4, 2000 год
октябрь, ноябрь, декабрь

© 1999 г. БУРНАШЕВ М. В.*, АМАРИ Ш.**, ХАН Т. С.***

О НЕКОТОРЫХ ЗАДАЧАХ ПРОВЕРКИ ГИПОТЕЗ С ИНФОРМАЦИОННЫМИ ОГРАНИЧЕНИЯМИ¹⁾

Рассматривается задача проверки гипотез, в которой часть данных не может наблюдаться непосредственно. Имеющийся помощник наблюдает эти данные и может передать нам некоторую ограниченную информацию о них. Какая именно информация позволит сделать наилучшие статистические выводы? В частности, какая минимальная информация достаточна для получения тех же результатов, как если бы мы непосредственно наблюдали все данные? Получены оценка сверху для величины этой минимальной информации и некоторые подобные результаты.

Ключевые слова и фразы: проверка гипотез, вероятность ошибки, скорость передачи информации, критическая скорость.

Введение. Общее описание проблемы

В данной работе мы хотели бы привлечь внимание специалистов к одному достаточно новому и в значительной степени неисследованному классу задач, а также представить некоторые частные результаты на эту тему. Хотя эти задачи впервые начали исследоваться в теории информации [1]–[5], они имеют статистическую сущность и для их понимания не требуется никаких знаний теории информации.

С этой целью мы рассмотрим некоторые интересные частные случаи следующей общей задачи.

*Институт проблем передачи информации РАН, Большой Каретный пер., 19, 101447 Москва, Россия; e-mail: burn@i.i.tp.ru

**RIKEN Brain Science Institute, Wako-shi, Hirosawa 2-1, Saitama 351-0198, Japan.

***Graduate School of Information Systems, University of Electro-Communications, Chfugaoka 1-5-1, Chofu, Tokyo 182, Japan.

¹⁾Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований, проект № 98-01-04108.

Пусть «статистику» необходимо сделать определенные статистические выводы о состоянии имеющейся системы (например, оценить неизвестный параметр, проверить некоторые гипотезы и т.д.). Для этого имеются два множества данных (наблюдений): множество $\mathcal{A}$ («available» — доступное) и множество $\mathcal{R}$ («remote» — удаленное).

Статистик непосредственно наблюдает все данные из множества $\mathcal{A}$.

Он не может непосредственно наблюдать данные из удаленного множества $\mathcal{R}$, однако его «помощник» (helper) может наблюдать их. Более того, помощнику разрешается передать статистику некоторый ограниченный объем информации об этих данных.

Задача состоит в следующем: какую информацию (ограниченную) об этих удаленных данных должен передать помощник, для того чтобы статистик (использующий также данные из множества $\mathcal{A}$) смог сделать наилучшие возможные статистические выводы (например, оценить неизвестный параметр с минимально возможной среднеквадратичной ошибкой и т.д.)?

Есть много практических ситуаций, где встречается подобная задача. Например, в некоторых приложениях данные из множества $\mathcal{R}$ представляют собой «мешающий шум», который «загрязнил» данные из множества $\mathcal{A}$, и поэтому нам хотелось бы «удалить» (по возможности) это «загрязнение», для того чтобы улучшить статистические выводы.

Далее мы будем рассматривать модели с дискретным временем, и, более того, под «ограниченным количеством информации» мы будем понимать, что помощник может посылать информацию статистику со скоростью передачи, не превышающей некоторую заданную величину $R > 0$ (см. далее пример 1). Конечно, если R велико настолько, что помощник может просто переслать статистику все данные из множества $\mathcal{R}$, то тогда мы приходим к традиционной статистической задаче (которая не является здесь предметом интереса для нас). По этой причине естественно предполагать, что R достаточно мало, чтобы избежать возможность примитивного пересылания данных.

Тем не менее, даже при таком предположении имеются случаи (иногда, возможно, естественные), когда оптимальное решение может быть получено достаточно просто. Например, так будет всегда, когда данные из множеств $\mathcal{A}$ и $\mathcal{R}$ представляет собой независимые наблюдения одного и того же явления. Следующий пример иллюстрирует такой случай.

Пример 1. Пусть все данные $\{x_1, \dots, x_n\}$ из множества $\mathcal{A}$ и все данные $\{x_{n+1}, \dots, x_{n+m}\}$ из множества $\mathcal{R}$ имеют вид

$$x_i = \theta + \xi_i, \quad i = 1, \dots, n + m, \quad -\infty < a < \theta < b < \infty,$$

где $\theta \in (a, b) \in \mathbf{R}^1$ — неизвестный параметр (подлежащий оценива-

нию) и ξ_i , $i = 1, \dots, n + m$, суть независимые одинаково распределенные случайные величины с известной плотностью распределения $f(z)$ относительно лебеговой меры на $\mathbf{R}^1$. Мы предполагаем, что функция $f(z)$, $z \in \mathbf{R}^1$, является достаточно гладкой, так что определена фишеровская информация $I(f)$, т.е.

$$\int_{-\infty}^{\infty} \frac{f'^2(z)}{f(z)} dz = I(f) < \infty.$$

Пусть $\hat{\theta}_{\mathcal{A}}(n)$ и $\hat{\theta}_{\mathcal{R}}(m)$ суть оценки максимального правдоподобия параметра θ , основанные на данных из $\mathcal{A}$ и $\mathcal{R}$ соответственно. Обозначим

$$\eta_{\mathcal{A}}(n) = (\hat{\theta}_{\mathcal{A}}(n) - \theta) \sqrt{nI}, \quad \eta_{\mathcal{R}}(m) = (\hat{\theta}_{\mathcal{R}}(m) - \theta) \sqrt{mI}.$$

Если теперь $n, m \rightarrow \infty$, то (при некоторых очень слабых дополнительных предположениях относительно $f(z)$) обе величины $\eta_{\mathcal{A}}(n)$ и $\eta_{\mathcal{R}}(m)$ являются асимптотически независимыми гауссовскими случайными величинами с параметрами $(0, 1)$ и, более того, все их моменты сходятся к соответствующим моментам этой гауссовской случайной величины [6]. Рассмотрим оценку

$$\hat{\theta}(n + m) = \frac{n\hat{\theta}_{\mathcal{A}}(n) + m\hat{\theta}_{\mathcal{R}}(m)}{n + m}$$

и обозначим $\eta(n + m) = (\hat{\theta}(n + m) - \theta) \sqrt{(n + m)I}$. Случайная величина $\eta(n + m)$ также является асимптотически $(0, 1)$ -гауссовской, и поэтому получаем $(n + m)I \mathbf{E}_{\theta}(\hat{\theta}(n + m) - \theta)^2 \rightarrow 1$, $n, m \rightarrow \infty$.

Известно (граница Крамера–Рао), что невозможно получить асимптотически меньшую среднеквадратичную ошибку. С другой стороны, для того чтобы построить оценку $\hat{\theta}(n + m)$, статистик, должен получить от помощника величину $\hat{\theta}_{\mathcal{R}}(m)$ с точностью $O((n + m)^{-1})$. Это означает, что он должен получить общее количество информации порядка $\ln((b - a)(n + m))$. Если $b - a$ конечно, то отсюда следует, что скорость передачи информации R имеет порядок $(\ln(n + m))/n$. Поэтому если $(\ln m)/n \rightarrow 0$, $n \rightarrow \infty$, то при больших n скорость передачи R может быть сделана произвольно малой и, следовательно, для любого $R > 0$ мы можем легко достичь асимптотически наилучший результат.

Ситуация становится значительно сложнее (и интереснее), когда имеется достаточно сильная зависимость между данными из множеств $\mathcal{A}$ и $\mathcal{R}$. Мы будем здесь в основном интересоваться случаем, когда ни статистик, ни помощник не могут сделать каких-либо хороших статистических выводов, основываясь только на собственных наблюдениях (другими словами, когда имеется очень сильная зависимость между данными из множеств $\mathcal{A}$ и $\mathcal{R}$).

Следующий пример (его модель будет для нас основной) иллюстрирует такой случай.

Пример 2. Рассмотрим устройство, называемое «двоичный симметричный канал» связи (ДСК) и представляющее наиболее интересную и популярную модель в теории информации [7], [8]. Его входной и выходной алфавиты являются двоичными, т.е. сигналы на входе и выходе канала принимают значения 0 и 1. Он характеризуется определенной переходной вероятностью $0 < p \leq \frac{1}{2}$. Это означает, что если x_i и y_i суть входной и выходной сигналы в произвольный момент времени i , то

$$\begin{aligned} \mathbf{P}\{y_i = 1 \mid x_i = 0\} &= \mathbf{P}\{y_i = 0 \mid x_i = 1\} = p, \\ \mathbf{P}\{y_i = 0 \mid x_i = 0\} &= \mathbf{P}\{y_i = 1 \mid x_i = 1\} = 1 - p = q. \end{aligned}$$

Более того, этот канал в различные моменты времени действует независимо, т.е. для любого набора моментов времени $i_1, \dots, i_n$ и любого набора $(x_{i_1}, \dots, x_{i_n}, y_{i_1}, \dots, y_{i_n})$ выполняется условие

$$\mathbf{P}\{y_{i_1}, \dots, y_{i_n} \mid x_{i_1}, \dots, x_{i_n}\} = \mathbf{P}\{y_{i_1} \mid x_{i_1}\} \cdots \mathbf{P}\{y_{i_n} \mid x_{i_n}\}.$$

(Такие каналы называют каналами без памяти.) В частности, если на вход канала поступает последовательность нулей, то на выходе канала наблюдается последовательность независимых бернуллиевских случайных величин с вероятностью успеха p .

Неизвестной величиной для нас здесь будет являться переходная вероятность $0 < p \leq \frac{1}{2}$, которую необходимо оценить или относительно которой необходимо проверить определенные гипотезы. При этом статистик наблюдает только выход канала $\mathcal{A} = (y_1, \dots, y_n)$, а его помощник наблюдает только вход канала $\mathcal{X} = (x_1, \dots, x_n)$. Мы будем предполагать, что относительно входного блока $(x_1, \dots, x_n)$ нет никакой априорной информации.

Ясно, что если статистик ничего не знает о входном блоке $x^n = (x_1, \dots, x_n)$, то он не сможет сделать никаких содержательных выводов о неизвестной величине p .

То, что помощник может посылать статистику информацию со скоростью, не превышающей заданную величину $R > 0$, означает, что им разрешается заранее разбить все входное пространство $E^n = \{0, 1\}^n$ на $N \leq 2^{Rn}$ произвольных частей $\{X_1, \dots, X_N\}$ и помощник сообщает статистику только, какой части X_i принадлежит входной блок x^n . Ясно, что при этом представляет интерес только случай $N < 2^n$, т.е. $R < 1$ (иначе помощник сможет просто переслать величину x^n).

Например, помощник может сообщить статистику точные значения первых Rn величин $x_1, \dots, x_{Rn}$ (но тогда ничего не сообщить о последующих величинах x_i). Однако такой простой способ разбиения входного пространства E^n (на цилиндрические множества $\{X_i\}$) не является, вообще говоря, оптимальным. С точки зрения статистика входные данные $(x_1, \dots, x_n)$ представляют собой очень сильный мешающий параметр.

Можно также сказать, что передача оптимальной ограниченной информации о блоке x^n означает оптимальное «сжатие» полной информации о блоке x^n . Конечно, это оптимальное «сжатие» зависит от имеющейся априорной информации о переходной вероятности p и используемого критерия качества.

З а м е ч а н и е. Ясно, что задача не изменится, если, наоборот, статистик наблюдает вход канала, а помощник наблюдает выход канала. В дальнейшем мы используем оба эти варианта постановки задачи.

Далее в работе для модели ДСК из примера 2 мы рассмотрим традиционную задачу проверки двух простых гипотез о параметре p , а во второй части работы (которая будет опубликована в следующих выпусках журнала) исследуем задачу оценивания параметра p .

В обеих задачах мы укажем некоторые разбиения $\{X_1, \dots, X_N\}$ и способы принятия решения (оценивания), которые асимптотически (при $n \rightarrow \infty$), возможно, близки к оптимальным. К сожалению, авторам пока не удалось показать, что лучше сделать нельзя, и это остается открытой проблемой.

Мы ограничиваемся здесь случаем ДСК (т.е. независимыми бернуллиевскими случайными величинами с неизвестным параметром p) по следующим причинам.

1. При достаточном знакомстве с теорией информации довольно ясно, что в интересных случаях некоторая функция, подобная *функции надежности* канала [7], [8], должна участвовать в решении задачи. С точки зрения этой функции надежности ДСК является очень показательным примером (т.е. он содержит все существенные проблемы; все другие каналы изучаются, основываясь на методах, развитых для ДСК; имеются только некоторые нижние и верхние границы для функции надежности ДСК и т.д.).

2. Все статистические величины и понятия (информация Кульбака–Лейблера, фишеровская информация и т.д.) имеют очень простой вид и геометрический смысл для ДСК. Поэтому в этом случае наиболее ясно будут видны основные трудности рассматриваемой проблемы и они не будут дополнительно усложняться более техническими вопросами.

Можно повторить также хорошо известное утверждение: «Покажите нам, как обращаться с ДСК (или бернуллиевскими распределениями), и мы покажем вам, как сделать то же самое для гораздо более широкого класса каналов (распределений)».

Всюду далее $\log x = \log_2 x$, $\exp_2 x = 2^x$. Для любого конечного множества A через $|A|$ обозначается его мощность. Для произвольной функции $f(x)$, $x \in A$, через $|f|$ обозначается мощность множества $f(A)$. Для различения входного и выходного пространств $E = \{0, 1\}$, мы будем обозначать их E_{in} и E_{out} соответственно.

2. Проверка двух простых гипотез

2.1. Постановка задачи и дуальной задачи. Мы рассматриваем ДСК с некоторой переходной вероятностью p , подлежащей проверке. Предположим, что p удовлетворяет одной из двух гипотез $H_0: p = p_0$ или $H_1: p = p_1$, где $0 < p_0 < p_1 \leq \frac{1}{2}$.

Через $\mathbf{P}$ и $\mathbf{Q}$ будем обозначать условные распределения на выходе ДСК для H_0 и H_1 соответственно. Тогда вероятности получить на выходе блок $y^n = (y_1, \dots, y_n)$ при условии, что на входе был блок $x^n = (x_1, \dots, x_n)$, даются соответственно формулами

$$\begin{aligned}\mathbf{P}\{y^n | x^n\} &= (1 - p_0)^{n-d(x^n, y^n)} p_0^{d(x^n, y^n)}, \\ \mathbf{Q}\{y^n | x^n\} &= (1 - p_1)^{n-d(x^n, y^n)} p_1^{d(x^n, y^n)},\end{aligned}$$

где $d(x^n, y^n)$ есть расстояние Хемминга между блоками x^n и y^n (т.е. число несовпадающих компонент этих векторов на всей длине n).

Нас интересует проверка этих гипотез, когда мы наблюдаем выход канала и получаем от помощника только некоторую ограниченную информацию о входном блоке. При этом мы будем рассматривать минимаксную постановку задачи.

Конкретно, мы предполагаем, что нам разрешается заранее разбить входное пространство E_{in}^n на N частей $\{X_1, \dots, X_N\}$. После этого мы наблюдаем выходной блок $y^n \in E_{\text{out}}^n$ и помощник сообщает нам только, какой части X_i принадлежал входной блок x^n . Основываясь на наблюдении y^n и номере части X_i , мы принимаем решение в пользу одной из гипотез H_0 или H_1 . Чтобы избежать излишнего усложнения, мы будем рассматривать только нерандомизованные правила принятия решения (при этом существо задачи и результаты сохраняются). Тогда общее правило принятия решения можно описать следующим образом. Для каждого элемента разбиения X_i мы выбираем некоторое множество $\mathcal{A}(X_i) \subset E_{\text{out}}^n$ и далее в зависимости от наблюдения y^n принимаем решение ($\mathcal{A}^c = E_{\text{out}}^n \setminus \mathcal{A}$):

$$y^n \in \mathcal{A}(X_i) \implies H_0, \quad y^n \in \mathcal{A}^c(X_i) \implies H_1.$$

Определим вероятности ошибок 1-го рода α_n и 2-го рода β_n как

$$\begin{aligned}\alpha_n &= \Pr(H_1 | H_0) = \max_{i=1, \dots, N} \max_{x^n \in X_i} \mathbf{P}\{\mathcal{A}^c(X_i) | x^n\}, \\ \beta_n &= \Pr(H_0 | H_1) = \max_{i=1, \dots, N} \max_{x^n \in X_i} \mathbf{Q}\{\mathcal{A}(X_i) | x^n\}.\end{aligned}$$

Пусть далее $\gamma > 0$ — некоторая заданная постоянная. Будем требовать, чтобы для вероятности ошибки 1-го рода выполнялось условие

$$\alpha_n = \Pr(H_1 | H_0) \leq 2^{-\gamma n}. \quad (1)$$

Нас интересует минимально возможная (по всем разбиениям входного множества и всем решениям) вероятность ошибки 2-го рода $\inf \beta_n$.

При этом мы будем исследовать асимптотическую ситуацию, когда $n \rightarrow \infty$ и $N = 2^{Rn}$, где $0 < R < 1$ — некоторая заданная постоянная²⁾. Тогда для наилучшего критерия обозначим

$$e(\gamma, R) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\inf \beta_n} > 0, \quad (2)$$

где $\inf$ берется по всем критериям, удовлетворяющим условию (1).

Нашей целью является нахождение (или получение хороших оценок) функции $e(\gamma, R)$.

Для нас будет полезно рассмотреть также следующую дуальную задачу (без помощника). Пусть задана некоторая величина $0 < r < 1$ и нам разрешается заранее выбрать любое множество $\mathcal{X} \subset E_{\text{in}}^n$, состоящее из $X = 2^{rn}$ входных блоков. Пусть известно также, что входным блоком может быть только блок из этого множества $\mathcal{X}$.

Теперь, зная множество $\mathcal{X}$, мы наблюдаем выход канала y^n и рассматриваем задачу проверки гипотез H_0 и H_1 . Далее для выбранного множества $\mathcal{A}$ в зависимости от наблюдения y^n принимается решение:

$$y^n \in \mathcal{A} \Rightarrow H_0, \quad y^n \in \mathcal{A}^c \Rightarrow H_1,$$

а вероятности ошибок 1-го рода и 2-го рода определяются как

$$\alpha_n = \max_{x^n \in \mathcal{X}} \mathbf{P} \{ \mathcal{A}^c \mid x^n \}, \quad \beta_n = \max_{x^n \in \mathcal{X}} \mathbf{Q} (\mathcal{A} \mid x^n).$$

Пусть теперь для вероятности ошибки 1-го рода выполняется условие (1) и мы хотим выбрать множество $\mathcal{X}$ мощности $X = 2^{rn}$ и правило принятия решения таким образом, чтобы достичь минимально возможной вероятности ошибки 2-го рода $\inf \beta_n$. Для такой дуальной задачи мы можем аналогично (2) определить функцию $e_2(\gamma, r)$. Следующий результат устанавливает простую связь между функциями $e(\gamma, R)$ и $e_2(\gamma, r)$.

Предложение 1. *Справедливо соотношение*

$$e(\gamma, 1 - R) = e_2(\gamma, R), \quad 0 \leq R \leq 1, \quad \gamma > 0. \quad (3)$$

Для доказательства предложения 1 нам потребуется простая лемма о «покрытиях» (конечно, известная).

Лемма. Пусть $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_X\} \subset E^n$ есть произвольное множество мощности X . Тогда существуют $K = \lceil n2^n/X \rceil$ «сдвигов» $\{\mathbf{y}_1, \dots, \mathbf{y}_K\} \subset E^n$ таких, что множества $\mathcal{X} + \mathbf{y}_i$, $i = 1, \dots, K$, покрывают все пространство E^n .

²⁾ Для упрощения формул здесь и далее мы не будем использовать знак целой части величины 2^{Rn} и аналогичных величин.

Доказательство. Выберем все K сдвигов случайным независимым образом (с возвращениями). Тогда для любого $K > n2^n \ln 2/X$ имеем

$$\begin{aligned} \text{Pr} \{ \text{найдется какая-либо непокрытая точка } x \in E^n \} \\ \leq 2^n \text{Pr} \{ \text{точка } 0 \text{ не покрыта} \} = 2^n (1 - X2^{-n})^K \\ \leq \exp\{-XK2^{-n} + n \ln 2\} < 1. \end{aligned}$$

Следовательно, среди таких случайно выбранных $K > n2^n \ln 2/X$ сдвигов существует набор, удовлетворяющий лемме.

Доказательство предложения 1. Пусть множество $\mathcal{X}$ мощности $\sim 2^{Rn}$ является наилучшим для дуальной задачи, т.е. оно обеспечивает вероятность ошибки 2-го рода $\sim 2^{-ne_2(\gamma, R)}$. В силу леммы все входное пространство E_{in}^n можно покрыть $N \sim 2^{(1-R)n}$ сдвинутыми версиями множества $\mathcal{X}$, каждая из которых имеет то же самое «качество различения». Уменьшая некоторые элементы этого покрытия, мы можем построить разбиение пространства E_{in}^n на $\sim 2^{(1-R)n}$ частей. При этом, так как рассматривается минимаксная постановка задачи, «качество различения» для каждой части будет не хуже, чем для исходного множества $\mathcal{X}$, откуда следует неравенство $e(\gamma, 1-R) \geq e_2(\gamma, R)$, $0 \leq R \leq 1$, $\gamma > 0$.

Наоборот, пусть в исходной задаче нам задано некоторое разбиение $\{X_1, \dots, X_N\}$ с $N \sim 2^{(1-R)n}$, обеспечивающее вероятность ошибки 2-го рода $\sim 2^{-ne(\gamma, 1-R)}$. Тогда существует некоторый элемент X_i этого разбиения мощности $\sim 2^{Rn}$, для которого в дуальной задаче вероятность ошибки 2-го рода также не превышает $2^{-ne(\gamma, 1-R)}$, откуда вытекает противоположное неравенство $e(\gamma, 1-R) \leq e_2(\gamma, R)$, $0 \leq R \leq 1$, $\gamma > 0$, что завершает доказательство предложения 1.

В силу предложения 1 достаточно исследовать функцию $e_2(\gamma, R)$, чем мы и будем заниматься далее. Однако прежде напомним некоторые результаты для известного входного блока.

2.2. Известный входной блок. Предположим сначала, что нам известен входной блок x^n и мы наблюдаем выходной блок y^n . Без потери общности можно считать, что x^n состоит только из нулей.

Ясно, что для оптимального критерия множество принятия решения в пользу p_0 есть шар $S(\tau n, 0)$ некоторого радиуса $\tau(\gamma)n \geq p_0 n$ с центром в нуле. Опираясь только экспоненциальными по n величинами, для коэффициента $\tau(\gamma)$ мы имеем условие

$$\begin{aligned} 2^{-\gamma n} \simeq \alpha_n \simeq \binom{n}{\tau n} (1-p_0)^{n-\tau n} p_0^{\tau n} \simeq 2^{nh(\tau)} (1-p_0)^{n-\tau n} p_0^{\tau n}, \\ h(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}, \end{aligned}$$

или

$$\gamma = \tau \log \frac{\tau}{p_0} + (1 - \tau) \log \frac{1 - \tau}{1 - p_0} \equiv D(\tau \| p_0). \quad (4)$$

Так как мы хотим иметь вероятность ошибки 2-го рода β_n также малой, то необходимо $p_0 \leq \tau \leq p_1$. Функция $D(\tau \| p_0)$ является U-выпуклой по τ и монотонно возрастающей при $\tau \geq p_0$. Поэтому γ должно удовлетворять условию $0 \leq \gamma \leq D(p_1 \| p_0)$. Для таких γ величина $\tau(\gamma)$ определяется как единственный корень (для $p_0 \leq \tau$) уравнения (4).

Для минимально возможной вероятности ошибки 2-го рода β_n имеем

$$\beta_n \simeq \binom{n}{\tau n} (1 - p_1)^{n - \tau n} p_1^{\tau n}$$

или

$$\frac{1}{n} \log \frac{1}{\beta_n} \simeq e(\gamma) = D(\tau \| p_1) = \gamma_2. \quad (5)$$

Удобно рассматривать $p_0 \leq \tau \leq p_1$ как параметр, через который могут быть выражены обе вероятности ошибки (см. (4) и (5)).

З а м е ч а н и е. Функция $D(x \| y)$ есть расхождение (divergence) для двух бернуллиевских случайных величин с параметрами x и y соответственно. Другими словами, она дает наилучшую экспоненту для вероятности ошибки 2-го рода при заданной вероятности ошибки 1-го рода (т.е. когда ее экспонента равна нулю) при проверке простой гипотезы $H_0: p = x$ против простой альтернативы $H_1: p = y$.

П р и м е р ы. 1) Пусть $\gamma = 0$, тогда $\tau = p_0$ и $\gamma_2 = D(p_0 \| p_1)$.

2) Пусть $\gamma_2 = 0$, тогда $\tau = p_1$ и $\gamma = D(p_1 \| p_0)$.

3) Пусть $\gamma = \gamma_2$, тогда τ есть единственный корень уравнения $D(\tau \| p_0) = D(\tau \| p_1)$, откуда следует

$$\tau = \log \frac{1 - p_0}{1 - p_1} \left(\log \frac{p_1(1 - p_0)}{p_0(1 - p_1)} \right)^{-1} \quad (6)$$

и $\gamma = \gamma_2 = D(\tau \| p_0) = D(\tau \| p_1)$.

2.3. Неизвестный входной блок и критическая скорость.

Как было показано выше, если мы знаем входной блок и $\alpha_n \leq 2^{-\gamma n}$, то наилучшая экспонента вероятности ошибки 2-го рода $e(\gamma)$ дается формулами (4)–(5).

Если же известно только, что входной блок принадлежит некоторому множеству $\mathcal{X}$ мощности $X \sim 2^{rn}$, то при наилучшем выборе такого множества $\mathcal{X}$ экспонента вероятности ошибки 2-го рода определяется функцией $e_2(\gamma, r)$. Ясно, что выполняется неравенство

$$e_2(\gamma, r) \leq e(\gamma), \quad \gamma \geq 0, \quad 0 \leq r \leq 1. \quad (7)$$

Функция $e_2(\gamma, r)$ монотонно не возрастает по r и $e_2(\gamma, 0) = e(\gamma)$. Поэтому относительно функции $e_2(\gamma, r)$ сразу возникает вопрос: существует ли $r > 0$, для которого в (7) выполняется равенство, и если да,

то какова максимальная такая скорость $r_{\text{crit}}(\gamma)$? Формально определим $r_{\text{crit}}(\gamma)$ как

$$r_{\text{crit}}(\gamma) = \sup \{r: e_2(\gamma, r) = e(\gamma)\}, \quad \gamma \geq 0. \quad (8)$$

Другими словами, какова наибольшая мощность 2^{rn} «наилучшего» множества $\mathcal{X}$, для которого мы можем достичь такой же асимптотической эффективности, как и при известном входном блоке (хотя мы и не знаем входной блок)?

Аналогично введем критическую скорость $R_{\text{crit}}(\gamma)$ для исходной задачи:

$$R_{\text{crit}}(\gamma) = \inf \{R: e(\gamma, R) = e(\gamma)\}, \quad \gamma \geq 0. \quad (9)$$

В силу предложения 1 имеем

$$R_{\text{crit}}(\gamma) = 1 - r_{\text{crit}}(\gamma), \quad \gamma \geq 0. \quad (10)$$

З а м е ч а н и е. Величина $r_{\text{crit}}(\gamma)$ вполне аналогична пропускной способности C канала связи, а функция $e_2(\gamma, r)$ аналогична функции надежности $E(r)$ в теории информации [7], [8]. При этом точный вид функции надежности $E(r)$ до сих пор неизвестен. Поэтому полное исследование функции $e_2(\gamma, r)$ (при $r > r_{\text{crit}}(\gamma)$) выглядит довольно трудной задачей.

3. Оценки для $r_{\text{crit}}(\gamma)$ и $e(\gamma, R)$

3.1. Граница снизу для $r_{\text{crit}}(\gamma)$ (случайный выбор множества $\mathcal{X}$). Как и ранее, пусть мера $\mathbf{P}$ соответствует p_0 , мера $\mathbf{Q}$ соответствует p_1 и $0 < p_0 < p_1 \leq \frac{1}{2}$. Рассмотрим все множества $\mathcal{X}$ мощности $X \sim 2^{Rn}$ на E_{in}^n . Пусть также выбрано некоторое правило принятия решения такое, что вероятность ошибки 1-го рода для каждого множества $\mathcal{X}$ не превышает заданной величины α_n . Тогда каждое $\mathcal{X}$ обладает своей вероятностью ошибки 2-го рода $\beta_n(\mathcal{X})$. Ясно, что существует множество $\mathcal{X}$, для которого величина $\beta_n(\mathcal{X})$ не превышает усредненной (по всем множествам $\mathcal{X}$) величины $\mathbf{E}\beta_n(\mathcal{X})$. Поэтому, если мы сумеем вычислить (или оценить сверху) величину $\mathbf{E}\beta_n(\mathcal{X})$, это даст определенную оценку снизу для функции $e_2(\gamma, r)$ и величины $r_{\text{crit}}(\gamma)$.

Такой метод случайного выбора (с возможными модификациями) представляет в теории информации наиболее универсальный способ получения различных теорем существования [7], [8]. Более того, во многих случаях результаты, полученные с его помощью, оказываются асимптотически оптимальными.

Для реализации этого подхода в качестве множества $\mathcal{X}$ мощности $X \sim 2^{Rn}$ выберем на E_{in}^n случайным равновероятным образом X различных точек $\{x_1, \dots, x_X\}$, и пусть y есть наше наблюдение.

В качестве области принятия решения $\mathcal{A}(\tau)$ в пользу p_0 используем

$$\mathcal{A}(\tau) = \left\{ y: \min_{i=1, \dots, X} d(y, x_i) \leq \tau n \right\},$$

где величина $p_0 \leq \tau \leq p_1$ будет выбрана позже.

Чтобы исследовать характеристики такого теста, можно без потери общности считать, что истинное значение блока x есть $x_1 = 0$.

Если справедлива гипотеза p_0 , то для вероятности ошибки 1-го рода имеем

$$\alpha_n \leq \mathbf{P}\{w(y) > \tau n \mid p_0, x_1\} \simeq \binom{n}{\tau n} (1-p_0)^{(1-\tau)n} p_0^{\tau n} \simeq \exp_2 \{-nD(\tau \| p_0)\}.$$

Пусть теперь справедлива гипотеза p_1 . Если $w(y) \leq \tau n$, то мы будем считать, что происходит ошибка при принятии решения. Если же $w(y) > \tau n$, то тогда мы можем совершить ошибку при принятии решения, только если в сфере радиуса τn с центром в y есть некоторая точка x_i . Тогда для усредненной вероятности ошибки 2-го рода $\mathbf{E}\beta_n$ имеем ($M = |E^n| = 2^n$, V — мощность шара радиуса τn в E_{in}^n)

$$\begin{aligned} \mathbf{E}\beta_n &\leq \mathbf{P}\{w(y) \leq \tau n \mid p_1, x_1\} + 1 - \frac{\binom{M-V}{X-1}}{\binom{M-1}{X-1}} \\ &= \exp_2 \{-nD(\tau \| p_1)\} + 1 - \prod_{i=1}^{V-1} \left[1 - \frac{X-1}{M-V-i} \right] \\ &\leq \exp_2 \{-nD(\tau \| p_1)\} + 1 - \left[1 - \frac{X-1}{M-2V+1} \right]^{V-1} \\ &\leq \exp_2 \{-nD(\tau \| p_1)\} + \frac{XV}{M-2V} \\ &\simeq \exp_2 \{-nD(\tau \| p_1)\} + \exp_2 \{-[1-h(\tau)-r]n\}. \end{aligned}$$

Следовательно, существует множество $\mathcal{X}$ мощности $X \sim 2^{\tau n}$, для которого при описанном выше правиле принятия решения выполняются неравенства

$$\alpha_n \leq \exp_2 \{-nD(\tau \| p_0)\}, \quad \beta_n \leq \exp_2 \left\{ -n \min \{D(\tau \| p_1), 1-h(\tau)-r\} \right\}.$$

Поэтому для функции $e_2(\gamma, r)$ справедлива граница снизу

$$e_2(\gamma, r) \geq \min \{D(\tau \| p_1), 1-h(\tau)-r\}, \quad 0 < r < 1, \quad (11)$$

где величина $p_0 \leq \tau \leq p_1$ при $0 \leq \gamma \leq D(p_1 \| p_0)$ определяется как единственный корень уравнения $\gamma = D(\tau \| p_0)$.

В частности, если мы хотим, чтобы было выполнено соотношение $\beta_n \simeq D(\tau \| p_1)$ (как для $X = 1$), то для этого достаточно иметь

$$r \leq 1 - h(\tau) - D(\tau \| p_1).$$

Последний результат можно сформулировать в следующем виде.

Предложение 2. Для критической скорости $r_{\text{crit}}(\gamma)$ в дуальной задаче справедлива оценка снизу

$$r_{\text{crit}}(\gamma) \geq 1 - h(\tau) - D(\tau \| p_1), \quad 0 \leq \gamma \leq D(p_1 \| p_0), \quad (12)$$

где функция $D(x \| y)$ определена в (4), а величина $p_0 \leq \tau \leq p_1$ есть единственный корень уравнения $\gamma = D(\tau \| p_0)$, $0 \leq \gamma \leq D(p_1 \| p_0)$.

З а м е ч а н и я. 1) Оценки (11), (12) остаются верными, даже когда мы проверяем сложную гипотезу $H_0: p \leq p_0$ против простой альтернативы $H_1: p = p_1$.

2) Пусть $\tau = p_1$, т.е. $\gamma = D(p_1 \| p_1)$. Тогда $D(\tau \| p_1) = 0$ и граница (12) принимает вид:

$$r_{\text{crit}}(D(p_1 \| p_1)) \geq 1 - h(p_1).$$

Эта граница определяется «сферической упаковкой» пространства E_{in}^n шарами радиуса $p_1 n$! Причина того, что мы, зная только множество $\mathcal{X}$ мощности $X \sim 2^{r^n}$, можем достичь такой же эффективности, как если бы мы знали входной блок x , заключается здесь в следующем. При хорошем выборе множества $\mathcal{X}$ (а при случайном выборе почти все $\mathcal{X}$ являются такими), зная выходной блок y и множество $\mathcal{X}$, можно с малой вероятностью ошибки определить, какой именно входной блок $x \in \mathcal{X}$ использовался (при любой гипотезе).

3.2. Случай $p_1 = \frac{1}{2}$. В специальном случае $p_1 = \frac{1}{2}$ оказывается возможным найти всю функцию $e_2(\gamma, r)$.

Если $p_1 = \frac{1}{2}$, то для любых x^n, y^n имеем $Q(y^n | x^n) = 2^{-n}$. Поэтому если в дуальной задаче $\mathcal{A}$ есть множество принятия решения в пользу p_0 , то для вероятности ошибки 2-го рода имеем $\beta_n = |\mathcal{A}| 2^{-n}$. В силу простоты этого выражения нам теперь будет удобнее задать экспоненту вероятности ошибки 2-го рода $0 \leq \gamma_2 \leq D(p_0 \| \frac{1}{2}) = 1 - h(p)$ (т.е. $\beta_n \simeq 2^{-\gamma_2 n}$) и исследовать наилучшую экспоненту вероятности ошибки 1-го рода $e_1(\gamma_2, r)$. При заданной величине γ_2 для мощности множества $\mathcal{A}$ мы имеем $|\mathcal{A}| \simeq 2^{(1-\gamma_2)n}$.

С другой стороны, для каждого входного блока $x_i \in \mathcal{X}$ оптимальная область принятия решения в пользу p_0 есть шар некоторого радиуса τn с центром в x_i . Поэтому для оптимального множества $\mathcal{X}$ область принятия $\mathcal{A}$ должна содержать «почти полностью» каждый из этих шаров. Ясно, что в таком множестве $\mathcal{X}$ все точки $\{x_i\}$ должны быть расположены максимально близко друг к другу, т.е. $\mathcal{X}$ является шаром, а $\mathcal{A}$ также является концентрическим с ним шаром. Отсюда следует

$$r_{\text{crit}}(\gamma) = 0, \quad \gamma \geq 0. \quad (13)$$

Пусть νn и μn суть радиусы множеств $\mathcal{X}$ и $\mathcal{A}$ соответственно. Так как мощность $|\mathcal{X}| \simeq 2^{r^n}$, то ν и μ определяются из соотношений

$$r = h(\nu), \quad 1 - \gamma_2 = h(\mu), \quad 0 \leq r < \frac{1}{2}, \quad p_0 \leq \mu < \frac{1}{2}. \quad (14)$$

Можно считать, что шары $\mathcal{X}$ и $\mathcal{Y}$ имеют центр в нуле. Если справедлива гипотеза p_0 и входной блок x имеет вес νn , то в силу закона больших чисел выходной блок y имеет (с вероятностью, близкой к 1) вес $(p_0 + \nu(1 - 2p_0))n$. Поэтому для того чтобы вероятность ошибки 1-го рода α_n была мала, необходимо, чтобы μ удовлетворяло условию

$$\mu \geq p_0 + \nu(1 - 2p_0). \quad (15)$$

Из этого условия и (14) следует

$$e_1(\gamma_2, r) = 0, \quad r \geq r_0(\gamma_2, p_0) = h \left(\frac{h^{-1}(1 - \gamma_2) - p_0}{1 - 2p_0} \right). \quad (16)$$

Теперь для $r < r_0(\gamma_2, p_0)$ оценим вероятность ошибки 1-го рода (она и определит функцию $e_1(\gamma, r)$). Ошибка 1-го рода происходит, если выходной блок имеет вес больше чем μn . Можно предполагать, что входной блок x имеет единицы на первых νn местах и нули на оставшихся $(1 - \nu)n$ местах. Пусть далее in означает число ошибок на первых νn местах, а jn означает число ошибок на оставшихся $(1 - \nu)n$ местах. Тогда ошибка при принятии решения происходит, если $\nu - i + j \geq \mu$. Поэтому обозначая $z = (1 - p_0)/p_0 > 1$, имеем

$$\begin{aligned} \alpha_n &\simeq (1 - p_0)^n \max \binom{\nu n}{in} \binom{(1 - \nu)n}{jn} z^{-(i+j)n} \\ &\simeq (1 - p_0)^n \exp_2 \left\{ n \max \left[\nu h \left(\frac{i}{\nu} \right) + (1 - \nu) h \left(\frac{j}{1 - \nu} \right) - (i + j) \log z \right] \right\}, \end{aligned} \quad (17)$$

где максимум берется по $0 \leq i \leq \nu$, $0 \leq j \leq 1 - \nu$, $\nu - i + j \geq \mu$.

Нетрудно проверить, что в точке максимума правой части (17) выполняется равенство $\nu - i + j = \mu$ (иначе нарушается условие (15)). Поэтому из (17) получаем

$$e_1(\gamma_2, r) = \log \frac{1}{1 - p_0} - \max_{0 \leq i \leq \nu} f(i), \quad (18)$$

$$f(i) = \nu h \left(\frac{i}{\nu} \right) + (1 - \nu) h \left(\frac{\mu - \nu + i}{1 - \nu} \right) - (2i + \mu - \nu) \log z, \quad z = \frac{1 - p_0}{p_0}.$$

Нетрудно проверить, что функция $f(i)$ является $\cap$ -выпуклой по i и достигает своего максимума внутри интервала $(0, \nu)$. Поэтому оптимальное значение i_0 является единственным корнем уравнения

$$\log \frac{\nu - i}{i} + \log \frac{1 - \mu - i}{\mu - \nu + i} = 2 \log z,$$

откуда, обозначая $u = (1 - 2p_0)/p_0^2 = z^2 - 1$, получаем

$$i_0 = \frac{\sqrt{[u(\mu - \nu) + 1]^2 + 4u\nu(1 - \mu)} - u(\mu - \nu) - 1}{2u}, \quad u = \frac{1 - 2p_0}{p_0^2}. \quad (19)$$

Полученные результаты можно сформулировать в следующем виде.

Предложение 3. Если $p_1 = \frac{1}{2}$, то $r_{\text{crit}}(\gamma) = 0$, $\gamma \geq 0$, а наилучшая экспонента $e_1(\gamma_2, r)$ вероятности ошибки 1-го рода дается формулой

$$e_1(\gamma_2, r) = \begin{cases} 0, & r \geq r_0(\gamma_2, p_0), \\ -\ln(1 - p_0) - f(i_0), & 0 \leq r < r_0(\gamma_2, p_0), \end{cases} \quad (20)$$

где $\mu, \nu, i_0, r_0(\gamma_2, p_0)$ определены в (14)–(16) и (19).

3.3. Полезный контрпример. К сожалению, нам не удалось пока получить хорошие оценки снизу для критической скорости $R_{\text{crit}}(\gamma)$ (или оценку сверху для $r_{\text{crit}}(\gamma)$).

Некоторые возникающие на этом пути трудности можно понять из следующего контрпримера.

Рассмотрим следующий вариант дуальной задачи. Пусть $0 < p_0 < p_1 < 1$ фиксированы и известно, что входной блок x^n принадлежит некоторому множеству $\mathcal{X} \subset E_{\text{in}}^n$ мощности $X \simeq 2^{rn}$. Вопрос: при какой максимальной скорости r_{max} роста мощности наилучшего множества $\mathcal{X}$ можно различить эти гипотезы, если требовать, только чтобы обе вероятности ошибки стремились к нулю? Ответ очень простой: $r_{\text{max}} = 1$.

Действительно, выберем в качестве множества $\mathcal{X} = \{x_1, \dots, x_X\}$ все точки на сфере некоторого радиуса $rn < n/2$. Тогда $X \simeq 2^{nh(r)}$. Так как входной блок имеет вес rn , то в силу закона больших чисел выходной блок с вероятностью, близкой к 1, будет иметь вес $(r + p(1 - 2r))n$. Поэтому при больших n мы сможем различить гипотезы $p = p_0$ и $p = p_1$ с малыми вероятностями ошибок для любого $r < \frac{1}{2}$, что и означает $r_{\text{max}} = 1$.

СПИСОК ЛИТЕРАТУРЫ

1. Ahlswede R., Csiszár I. Hypotheses testing with communication constraints. — IEEE Trans. Inform. Theory, 1986, v. 32, № 4, p. 533–542.
2. Zhang Z., Berger T. Estimation via compressed information. — IEEE Trans. Inform. Theory, 1988, v. 34, № 2, p. 198–211.
3. Han T. S., Kobayashi K. Exponential-type error probabilities for multiterminal hypothesis testing. — IEEE Trans. Inform. Theory, 1989, v. 35, № 1, p. 2–14.
4. Ahlswede R., Burnashev M. V. On minimax estimation in the presence of side information about remote data. — Ann. Statist., 1990, v. 18, № 1, p. 141–171.
5. Han T. S., Amari S. Parameter estimation with multiterminal data compression. — IEEE Trans. Inform. Theory, 1995, v. 41, № 6, p. 1802–1833.
6. Ибрагимов И. А., Хасъминский Р. З. Асимптотическая теория оценивания. М.: Наука, 1979, 528 с.
7. Фано Р. Передача информации. Статистическая теория связи. М.: Мир, 1965, 439 с.
8. Галлагер Р. Теория информации и надежная связь. М.: Советское радио, 1974, 720 с.