ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 3**

УДК 621.391.1 : 519.713 : 517.977.5

© 2020 г. **А.В. Колногоров**

ГАУССОВСКИЙ ДВУРУКИЙ БАНДИТ: ПРЕДЕЛЬНОЕ ОПИСАНИЕ[1]

Для гауссовского двурукого бандита, который возникает при анализе пакетной обработки данных, изучается предельное поведение минимаксного риска, если горизонт управления N неограниченно растет. Минимаксный риск ищется как байесовский, вычисленный относительно наихудшего априорного распределения. Показано, что наиболее высокие требования к управлению предъявляются в области "близких" распределений, где математические ожидания доходов различаются на величину порядка $N^{-1/2}$. В области "близких" распределений получены рекуррентное интегро-разностное уравнение для нахождения байесовского риска относительно наихудшего априорного распределения в инвариантной форме с горизонтом управления, равным единице, и дифференциальное уравнение в частных производных второго порядка в предельном случае. Результаты позволяют оценить качество пакетной обработки. Например, минимаксный риск, соответствующий пакетной обработке данных, разбитых на 50 пакетов, может быть лишь на 2% выше своего предельного значения, если число пакетов неограниченно растет. В случае бернуллиевского двурукого бандита показано, что оптимальная обработка данных по одному не является более эффективной, чем пакетная, если N неограниченно растет.

Ключевые слова: гауссовский двурукий бандит, минимаксный и байесовский подходы, пакетная обработка, асимптотическая минимаксная теорема.

DOI: 10.31857/S0555292320030055

§ 1. Введение

Статья продолжает работу [1], в которой дан обзор некоторых других подходов к рассматриваемой задаче и библиография. Опишем кратко результаты [1]. Гауссовский двурукий бандит – это управляемый случайный процесс ξ_n, $n = 1, \ldots, N$, значения которого интерпретируются как доходы, зависят только от выбираемых в текущие моменты времени действий y_n и имеют нормальные распределения с плотностями $f_{D_\ell}(x \,|\, m_\ell)$, если $y_n = \ell$ $(\ell = 1, 2)$, где $f_D(x \,|\, m) = (2\pi D)^{-1/2} \exp\{-(x-m)^2/(2D)\}$. Предполагается, что D_1, D_2 – априори известные дисперсии, а m_1, m_2 – неизвестные математические ожидания. Такой двурукий бандит описывается параметром $\theta = (m_1, m_2)$. Допустимое множество параметров имеет вид $\Theta = \{\theta : \ |m_1 - m_2| \leqslant \leqslant 2C\}$, где $0 < C < \infty$.

Гауссовский двурукий бандит возникает, если одинаковые действия применяются к пакетам данных, а для управления используются суммарные доходы в пакетах. В силу центральной предельной теоремы для широкого класса процессов распределения суммарных доходов в пакетах данных близки к нормальным, если размеры пакетов достаточно велики. При обработке начальных пакетов могут быть получены

[1] Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 20-01-00062).

оценки дисперсий одношаговых доходов D_1, D_2. Их можно использовать при дальнейшей обработке, так как минимаксный риск мало меняется при малом изменении дисперсий (см. лемму 8 этой статьи). Важное свойство пакетной обработки, установленное в [1], состоит в том, что она практически не увеличивает минимаксный риск, если число пакетов, на которые разбиты данные, достаточно велико.

Управляющая стратегия σ в момент времени $n+1$ определяет выбор действия y_{n+1} в зависимости от известной предыстории (X_1, n_1, X_2, n_2), где n_1, n_2 – текущие полные количества применений обоих действий $(n_1+n_2=n)$, X_1, X_2 – соответствующие полные доходы. Таким образом $\sigma_\ell(X_1, n_1, X_2, n_2) = \Pr(y_{n+1} = \ell \,|\, X_1, n_1, X_2, n_2)$, $\ell = 1, 2$.

Для формулировки цели управления определим функцию потерь

$$L_N(\sigma, \theta) = N(m_1 \vee m_2) - \mathbf{E}_{\sigma,\theta}\left(\sum_{n=1}^{N} \xi_n\right), \tag{1.1}$$

которая описывает математическое ожидание потерь полного дохода относительно максимально возможной величины вследствие неполноты информации. Здесь $(m_1 \vee m_2)$ – максимум из m_1, m_2, а $\mathbf{E}_{\sigma,\theta}$ обозначает знак математического ожидания, вычисленного по мере, порожденной стратегией σ и параметром θ. Обозначим через $\lambda(\theta)$ априорную плотность распределения параметра на множестве Θ. По функции потерь определяются минимаксный и байесовский риски

$$R_N^M(\Theta) = \inf_{\{\sigma\}} \sup_{\Theta} L_N(\sigma, \theta), \tag{1.2}$$

$$R_N^B(\lambda) = \inf_{\{\sigma\}} \int_{\Theta} L_N(\sigma, \theta)\lambda(\theta)\, d\theta, \tag{1.3}$$

соответствующие оптимальные стратегии называются минимаксной и байесовской стратегиями. Преимущество минимаксного подхода состоит в его независимости от априорного распределения, для выбора которого, как правило, нет ясных критериев. Однако прямых методов нахождения минимаксных стратегии и риска не существует. С другой стороны, байесовский подход дает возможность написать рекуррентное уравнение, позволяющее найти байесовские стратегию и риск методом динамического программирования. Объединяет минимаксный и байесовский подходы основная теорема теории игр, согласно которой при широких предположениях выполнено равенство

$$R_N^M(\Theta) = \sup_{\lambda} R_N^B(\lambda), \tag{1.4}$$

т.е. минимаксный риск (1.2) равен байесовскому риску (1.3), вычисленному относительно наихудшего априорного распределения, на котором байесовский риск максимален, а минимаксная стратегия совпадает с соответствующей байесовской. Именно этот подход для нахождения минимаксных стратегии и риска рассмотрен в [1]. Отметим, что другой подход к пакетной обработке, предполагающий использование небольшого числа пакетов возрастающего объема, в близких постановках независимо рассмотрен в [2, 3], где установлены, по существу, одинаковые оценки минимаксного риска в этом случае.

В данной статье изучаются предельные свойства минимаксных стратегии и риска, если число обрабатываемых пакетов неограниченно растет. Установлено, что при $N \to \infty$ существует предел для нормированного минимаксного риска $N^{-1/2} R_N^M(\Theta)$, что позволяет оценить качество пакетной обработки. Например, в случае разбиения данных на 50 пакетов нормированный минимаксный риск может быть лишь на 2% выше своего предельного значения. Особая роль в исследовании отводится обла-

сти "близких" распределений, удовлетворяющих условию $|m_1 - m_2| \leqslant 2cN^{-1/2}$, где $c > 0$ – достаточно большое фиксированное число. Наибольшие потери достигаются именно в этой области и, соответственно, в ней же предъявляются наиболее высокие требования к управлению.

Структура статьи следующая. В § 2 обсуждается роль "близких" распределений и предлагается класс стратегий, позволяющих отделить не "близкие" распределения за счет определения для них лучшего действия на коротком начальном этапе управления. В § 3 уравнение для вычисления байесовских стратегии и риска относительно класса распределений, к которому принадлежит наихудшее, дается в инвариантной форме с горизонтом управления, равным единице. Такое представление справедливо именно в области "близких" распределений. В § 4 показано, что если количество обрабатываемых пакетов неограниченно растет, то существует непрерывный предел у решения этого уравнения, а само инвариантное рекуррентное уравнение переходит в дифференциальное уравнение в частных производных второго порядка. Строгое доказательство этого результата требует некоторых дополнительных оценок гладкости предельного решения, которые пока не получены. Но численные эксперименты, представленные в § 5, подтверждают близость решений рекуррентного и дифференциального уравнений. Наконец, в § 6 показано, что в случае бернуллиевского двурукого бандита обработка данных по одному не позволяет уменьшить величину минимаксного риска в сравнении с пакетной обработкой, если количество данных неограниченно растет, поскольку предельные описания обработок пакетами и по одному даются одинаковыми дифференциальными уравнениями. Заключение содержится в § 7.

§ 2. "Близкие" распределения и их роль в управлении

Под "близкими" будем понимать распределения, определяемые множеством параметров $\Theta_{cN^{-1/2}} = \{\theta : |m_1 - m_2| \leqslant 2cN^{-1/2}\}$, где $c > 0$ – достаточно большая фиксированная константа. Ниже устанавливается, что при больших N для анализа максимальных потерь достаточно ограничиться только "близкими" распределениями. В частности, приводится известный результат о том, что максимальные потери, имеющие порядок $N^{1/2}$, достигаются именно в области "близких" распределений. Далее предложен класс стратегий, способных отделить "удаленные" распределения за счет определения для них на коротком начальном этапе лучшего действия и применения его затем до конца управления. Нормированный минимаксный риск $N^{-1/2}R_N^M(\Theta_C)$, обеспечиваемый этими стратегиями на множестве $\Theta_C = \{\theta : |m_1 - m_2| \leqslant 2C\}$, путем выбора подходящих параметров стратегии может быть сделан при больших N сколь угодно близким к нормированному риску $N^{-1/2}R_N^M(\Theta_{cN^{-1/2}})$ для некоторого $c > 0$.

Начнем с оценок максимальных потерь. Справедлива

Теорема 1. *В области "близких" распределений порядок максимальных потерь не ниже* $N^{1/2}$. *Если же известно, что* $|m_1 - m_2| \geqslant \delta > 0$, *то порядок максимальных потерь не превосходит* $\ln(N)$.

Доказательство. Для оценки снизу потерь в области "близких" распределений рассмотрим поведение гауссовского двурукого бандита, характеризуемого дисперсиями D_1, D_2 и более узким множеством параметров $\{\theta_1 = (D_1^{1/2}m, -D_2^{1/2}m), \theta_2 = (-D_1^{1/2}m, D_2^{1/2}m)\}$. Управление будем осуществлять на основе наблюдений процесса z_n, $n = 1, \ldots, N$, где $z_n = \xi_n D_1^{-1/2}$, если $y_n = 1$, и $z_n = -\xi_n D_2^{-1/2}$, если $y_n = 2$. Легко видеть, что при любой стратегии управления $\{z_n\}$ – независимые нормально распределенные случайные величины с единичной дисперсией и математическим ожиданием, равным m, если $\theta = \theta_1$, и $-m$, если $\theta = \theta_2$. Положим $Z_n = \sum\limits_{i=1}^{n} z_i$. То-

гда согласно лемме Неймана – Пирсона на n-м шаге принимаются гипотезы $m > 0$ и $m < 0$, если $Z_n > 0$ и $Z_n < 0$ соответственно.

Пусть для определенности $\theta = \theta_1$, причем $m = xN^{-1/2}$, $x > 0$. Обозначим через $\Phi(x) = \int\limits_{-\infty}^{x} f_1(t\,|\,0)\,dt$ функцию стандартного нормального распределения, и пусть $\widetilde{D}^{1/2} = 0{,}5(D_1^{1/2} + D_2^{1/2})$. Тогда одношаговые потери на n-м шаге равны

$$2m\widetilde{D}^{1/2}\Pr(Z_n < 0) = (\widetilde{D}/N)^{1/2}2x\Phi(-xN^{-1/2}n/n^{1/2}) = (\widetilde{D}N)^{1/2}2x\Phi(-xt^{1/2})\Delta t,$$

где $t = n/N$, $\Delta t = N^{-1}$. А потери на всем горизонте управления равны

$$(\widetilde{D}N)^{1/2}\left(2x\sum_{n=1}^{N}\Phi(-xt^{1/2})\Delta t\right) \sim (\widetilde{D}N)^{1/2}\left(2x\int\limits_{0}^{1}\Phi(-xt^{1/2})\,dt\right).$$

Максимальное значение выражения в скобках достигается при $x \approx 1{,}22$ и приблизительно равно 0,53. Поэтому порядок максимальных потерь в области "близких" распределений не ниже $N^{1/2}$.

Предположим теперь, что $|m_1 - m_2| \geqslant \delta > 0$, и рассмотрим стратегию, которая сначала применяет каждое из действий по $k = \varkappa\ln(N)$ раз, получая за это полные доходы X_1 и X_2, а затем оставшиеся $N - 2k$ раз применяет только то действие, которому соответствует больший из доходов X_1, X_2. Отметим, что $X_1 - X_2$ имеет нормальное распределение с параметрами $\mathbf{E}(X_1 - X_2) = k(m_1 - m_2)$, $\mathbf{D}(X_1 - X_2) = 2kD$, где $D = 0{,}5(D_1 + D_2)$. Если $m_1 > m_2$, то потери равны

$$L_N(m_1, m_2) = (m_1 - m_2)\,(k + \Pr(X_1 - X_2 < 0)(N - 2k)) =$$
$$= (m_1 - m_2)\left(k + \Phi\left(-(m_1 - m_2)(0{,}5k/D)^{1/2}\right)(N - 2k)\right).$$

С использованием оценки (см. [4, с. 43])

$$\Phi(-x) < (2\pi)^{-1/2}x^{-1}\exp(-x^2/2) < \exp(-x^2/2) \quad \text{при } x > 1 \tag{2.1}$$

получаем, что

$$L_N(m_1, m_2) < (m_1 - m_2)\left(\varkappa\ln(N) + N^{1-0{,}25(m_1-m_2)^2\varkappa D^{-1}}\right) =$$
$$= (m_1 - m_2)\varkappa\ln(N) + \alpha_N,$$

где $\alpha_N \to 0$ при $N \to \infty$, если $0{,}25\varkappa\delta^2 D^{-1} > 1$. ▲

Отметим, что в случае равных дисперсий порядок оценки $N^{1/2}$ дан ранее в [5,6], а оценки $\ln(N)$ – в [7]. Логарифмическая оценка порядка роста максимальных потерь справедлива и для ряда других стратегий (см., например, [8,9]).

Установим, что выбором подходящей стратегии можно перевести управление в область "близких" распределений. Для некоторой стратегии σ, действующей на горизонте управления $[2n_0 + 1, N]$, где $2n_0 < N$, определим стратегию σ^*, действующую на всем горизонте $[1, N]$. Эта стратегия на начальном этапе управления $[1, 2n_0]$ применяет действия по очереди до тех пор, пока абсолютная разность полных доходов за их применение не превысит величины порога $\widetilde{\alpha}N^{1/2}$ или не закончится начальный этап управления ($\widetilde{\alpha} = \alpha D^{1/2}$, $D = 0{,}5(D_1 + D_2)$, $\alpha > 0$). Если величина порога превышена на начальном этапе в момент времени $2k_0$ ($2k_0 \leqslant 2n_0$), то на заключительном этапе $[2k_0 + 1, N]$ применяется только то действие, которому соответствует бóльшая величина полного дохода в момент времени $2k_0$. Если же до конца начального этапа порог не был достигнут, то на заключительном этапе управления $[2n_0 + 1, N]$ применяется стратегия управления σ.

Покажем, что выбором подходящих n_0 и α эта стратегия позволяет перевести управление в область "близких" распределений. Нам потребуются оценки, представленные ниже в леммах 1–3.

Лемма 1. *Пусть* $m_1 - m_2 = b$, $0 < b < \infty$, *величины порогов равны* $\pm a$, $0 < a < \infty$. *Тогда вероятность достижения порога в точке* $-a$ *не превосходит величины*

$$P_e(a,b) \leqslant \frac{e^{-ab/D}}{1+e^{-ab/D}}. \tag{2.2}$$

Доказательство. Обозначим через $z_k = (\xi_{2k-1} - \xi_{2k})$ текущие разности доходов при применении пороговой стратегии, распределенные с плотностями $(4\pi D)^{-1/2}\exp(-(z_k-b)^2/(4D))$, $k = 1,\dots,n_0$. Следуя [5], обозначим $u_k = \sum\limits_{i=1}^{k} z_i$ и введем подмножества

$$A_{1,k} = \{|u_j| < a \text{ при } 1 \leqslant j < k \text{ и } u_k \geqslant a\},$$
$$A_{2,k} = \{|u_j| < a \text{ при } 1 \leqslant j < k \text{ и } u_k \leqslant -a\}, \quad B = \{|u_j| < a \text{ при } 1 \leqslant j \leqslant n_0\}.$$

Принадлежность процесса подмножествам $A_{1,k}$, $A_{2,k}$ означает, что пороги a и $-a$ впервые были достигнуты в момент времени $2k$, попадание в подмножество B означает, что ни один из этих порогов не был достигнут до конца начального этапа. Ясно, что

$$\Pr(A_{1,k}) = \int\limits_{A_{1,k}} (4\pi D)^{-k/2} \prod_{i=1}^{k} e^{-\frac{(z_i-b)^2}{4D}} dz^k =$$
$$= \int\limits_{A_{1,k}} (4\pi D)^{-k/2} e^{\frac{bu_k}{2D}} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k \geqslant e^{\frac{ab}{2D}} \int\limits_{A_{1,k}} (4\pi D)^{-k/2} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k. \tag{2.3}$$

Здесь $dz^k = dz_1 \dots dz_k$. Аналогично,

$$\Pr(A_{2,k}) =$$
$$= \int\limits_{A_{2,k}} (4\pi D)^{-k/2} e^{\frac{bu_k}{2D}} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k \leqslant e^{-\frac{ab}{2D}} \int\limits_{A_{2,k}} (4\pi D)^{-k/2} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k. \tag{2.4}$$

Так как подмножества $A_{1,k}$, $A_{2,k}$ переходят одно в другое при преобразовании координат $z_i \leftarrow -z_i$, $i = 1,\dots,k$, то значения интегралов в (2.3), (2.4) одинаковы, и следовательно,

$$\frac{\Pr(A_{2,k})}{\Pr(A_{1,k})} \leqslant e^{-ab/D}. \tag{2.5}$$

Обозначим через $A_1 = \bigcup\limits_{k=1}^{n_0} A_{1,k}$, $A_2 = \bigcup\limits_{k=1}^{n_0} A_{2,k}$ события, состоящие в том, что были достигнуты пороги a и $-a$ соответственно. Так как все события $A_{1,k}$, $A_{2,k}$, $k = 1,\dots,n_0$, попарно несовместны, то из (2.5) следует, что

$$\frac{\Pr(A_2)}{\Pr(A_1)} \leqslant e^{-ab/D}.$$

Так как $P_e(a,b) = \Pr(A_2)$ и $\Pr(A_1) + \Pr(A_2) = 1 - \Pr(B) < 1$, то формула (2.2) справедлива. ▲

Лемма 2. *Пусть* $m_1 - m_2 = 2\beta(D/N)^{1/2}$, $0 < c < \infty$, $\Pr(\overline{B}) = 1 - \Pr(B)$. *Справедливы оценки*

$$L_N(\sigma^*, \theta) \leqslant 2c \Pr(\overline{B})(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \Pr(B) \tag{2.6}$$

при $|\beta| \leqslant c$ *и*

$$L_N(\sigma^*, \theta) \leqslant \left(\alpha + \max_{\beta \geqslant c}(2\beta e^{-2\alpha\beta}) + o(1)\right)(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \Pr(B) \tag{2.7}$$

при $|\beta| \geqslant c$.

Доказательство. Оценка (2.6) следует из того, что применение одного и того же действия на оставшемся горизонте управления после достижения одного из порогов, которое происходит с вероятностью $\Pr(\overline{B})$, не может дать потери, бóльшие чем $2c(DN)^{1/2}$. При этом применение стратегии σ на заключительном этапе происходит с вероятностью $\Pr(B)$.

Докажем оценку (2.7). Положим $a = \widetilde{\alpha}N^{1/2}$, $b = 2\widetilde{\beta}N^{-1/2}$, где $\widetilde{\beta} = \beta D^{1/2}$ и без ограничения общности будем считать, что $\beta > 0$. Через $k^* = \min(k : |u_k| \geqslant \widetilde{\alpha}N^{1/2})$ обозначим момент достижения одного из порогов и положим $k_0 = \min(k^*, n_0)$. Тогда ожидаемые потери допускают оценку

$$L_N(\sigma^*, \theta) \leqslant L_N^{(1)}(\alpha, \beta) + L_N^{(2)}(\alpha, \beta) + L_{N-2n_0}(\sigma, \theta) \Pr(B), \tag{2.8}$$

где $L_N^{(1)}(\alpha, \beta), L_N^{(2)}(\alpha, \beta)$ возникают в случае достижения одного из порогов и равны

$$\begin{aligned} &L_N^{(1)}(\alpha, \beta) = \\ &= \mathbf{E}\left(\sum_{n=1}^{2k_0}(m_1 - \xi_n)\right) = 2\,\mathbf{E}\left(\sum_{k=1}^{k_0}(m_1 - \xi_{2k-1})\right) + \mathbf{E}\left(\sum_{k=1}^{k_0}(\xi_{2k-1} - \xi_{2k})\right), \\ &L_N^{(2)}(\alpha, \beta) = \mathbf{E}\left(\sum_{n=2k_0+1}^{N}(m_1 - \xi_n)\right). \end{aligned} \tag{2.9}$$

Так как $\sum\limits_{k=1}^{n}(m_1 - \xi_{2k-1})$, $n = 1, 2, \ldots,$ – мартингал, а k_0 – момент остановки, то

$$\mathbf{E}\left(\sum_{k=1}^{k_0}(m_1 - \xi_{2k-1})\right) = 0,$$

поэтому из первого равенства в (2.9) следует, что

$$L_N^{(1)}(\alpha, \beta) = \mathbf{E}\left(\sum_{k=1}^{k_0}(\xi_{2k-1} - \xi_{2k})\right) \leqslant (\widetilde{\alpha} + o(1))N^{1/2}, \tag{2.10}$$

где $o(1)$ вызвано превышением порога на последнем шаге. Из (2.2) и второго равенства в (2.9) следует оценка

$$L_N^{(2)}(\alpha, \beta) \leqslant (2\widetilde{\beta})N^{-1/2}\,\mathbf{E}(N - 2k_0)P_e(a, b) \leqslant \max_{\beta \geqslant c}(2\beta e^{-2\alpha\beta})(DN)^{1/2}. \tag{2.11}$$

Из (2.8), (2.10), (2.11) следует справедливость оценки (2.7). ▲

Сделаем несколько замечаний. Пороговая стратегия, действующая на всем горизонте управления, т.е. при $2n_0 = N$, впервые предложена в [5,6] для бернуллиевского

двурукого бандита. Лемма 1 использует идею, предложенную в [5], остальные представленные здесь оценки оригинальны.

Оценка (2.7) при $2n_0 = N$ и $c = 0$ соответствует тому, что пороговая стратегия применяется ко всем двуруким бандитам на всем горизонте управления. В этом случае максимальные потери не превосходят $\Big(\alpha + \max\limits_{\beta \geqslant 0}(2\beta e^{-2\alpha\beta}) + o(1)\Big)(DN)^{1/2} \leqslant$ $\leqslant (\alpha + (\alpha e)^{-1} + o(1))(DN)^{1/2}$ и достигаются при $2\alpha\beta = 1$, т.е. в области "близких" распределений. При $\alpha = e^{-1/2}$ максимальные потери асимптотически не превосходят величины $2e^{-1/2}(DN)^{1/2} \approx 1{,}213(DN)^{1/2}$. В [5, 6] дана более точная оценка множителя 0,752 (при $D = 0{,}25$), но представленные здесь выкладки значительно короче.

Наконец отметим, что если c достаточно велико, то $\max\limits_{\beta \geqslant c}(2\beta e^{-2\alpha\beta}) = 2ce^{-2\alpha c}$.

Для дальнейшего нам потребуется оценка вероятности $\Pr(\overline{B})$, где $\overline{B}$ означает дополнение события B. Здесь также используется идея из [5].

Лемма 3. *Пусть* $a = \alpha(DN)^{1/2}$, $\alpha > 0$, *и* $m_1 - m_2 = 2\beta(D/N)^{1/2}$, $|\beta| \leqslant c < \infty$. *Пусть оценка вероятности* $\Pr(\overline{B})$ *выполняется на момент времени* $2tN$. *Тогда при* $N \to \infty$ *справедлива оценка*

$$\Pr(\overline{B}) \leqslant \Phi\left(\frac{-2\beta t - \alpha}{(2t)^{1/2}}\right)(1 + e^{2\alpha\beta}) + \Phi\left(\frac{2\beta t - \alpha}{(2t)^{1/2}}\right)(1 + e^{-2\alpha\beta}). \tag{2.12}$$

Доказательство. Последовательность случайных величин

$$s_t = N^{-1/2}\sum_{\nu=1}^{tN}(\xi_{2\nu-1} - \xi_{2\nu}), \quad \text{где } t = nN^{-1},\ n = 1, 2, \ldots,$$

описывает одномерное случайное блуждание с дисперсией шага $2DN^{-1}$ и сносом $2\widetilde{\beta}N^{-1}$, где $\widetilde{\beta} = \beta D^{1/2}$. Рассмотрим случайное блуждание, начинающееся из точки $x = 0$ с поглощающим экраном в точке $x = -\widetilde{\alpha}$, где $\widetilde{\alpha} = \alpha D^{1/2}$. При $N \to \infty$ плотность распределения s_t слабо сходится к плотности $f(x, t)$, $0 \leqslant t \leqslant 0{,}5$, $x \geqslant -\widetilde{\alpha}$, удовлетворяющей уравнению Фоккера – Планка – Колмогорова

$$f'_t = -2\widetilde{\beta}f'_x + Df''_{xx}$$

с начальным условием $f(x, 0) = \delta(x)$ и граничным условием $f(-\widetilde{\alpha}, t) = 0$, $0 \leqslant t \leqslant 0{,}5$, где $\delta(x)$ – дельта-функция Дирака. Решение этого уравнения имеет вид

$$f(x,t) = \frac{1}{(4\pi Dt)^{1/2}}\exp\left(-\frac{(x - 2\widetilde{\beta}t)^2}{4Dt}\right) - \frac{e^{-2\widetilde{\alpha}\widetilde{\beta}/D}}{(4\pi Dt)^{1/2}}\exp\left(-\frac{(x + 2\widetilde{\alpha} - 2\widetilde{\beta}t)^2}{4Dt}\right).$$

Поэтому вероятность того, что случайное блуждание не было поглощено до момента времени t, есть

$$P_1(\alpha,\beta) = \int\limits_{-\widetilde{\alpha}}^{\infty} f(x,t)\,dx = \Phi\left(\frac{2\widetilde{\beta}t + \widetilde{\alpha}}{(2Dt)^{1/2}}\right) - e^{-2\widetilde{\alpha}\widetilde{\beta}/D}\Phi\left(\frac{2\widetilde{\beta}t - \widetilde{\alpha}}{(2Dt)^{1/2}}\right).$$

Так как $\Phi(x) + \Phi(-x) = 1$ при всех x, то вероятность $P_2(\alpha, \beta)$ поглощения до момента времени t равна

$$P_2(\alpha,\beta) = 1 - P_1(\alpha,\beta) = \Phi\left(\frac{-2\widetilde{\beta}t - \widetilde{\alpha}}{(2Dt)^{1/2}}\right) + e^{-2\widetilde{\alpha}\widetilde{\beta}/D}\Phi\left(\frac{2\widetilde{\beta}t - \widetilde{\alpha}}{(2Dt)^{1/2}}\right).$$

Ясно, что для случайного блуждания, начинающегося из точки $x = 0$, с двумя поглощающими экранами в точках $x = -\widetilde{\alpha}$ и $x = \widetilde{\alpha}$ вероятность поглощения в точке $x = -\widetilde{\alpha}$ не превосходит $P_2(\alpha, \beta)$. В силу симметрии вероятность поглощения в точке $x = \widetilde{\alpha}$ не превосходит $P_2(\alpha, -\beta)$. Поэтому $\Pr(\overline{B}) \leqslant P_2(\alpha, \beta) + P_2(\alpha, -\beta)$, откуда следует (2.12). ▲

Теорема 2. *Положим $n_0 = \varepsilon_0 N$ и выберем параметры стратегии σ^* следующим образом. По заданным малым $\alpha > 0$ и $\delta > 0$ определим достаточно большое $c > 0$ так, чтобы $\max\limits_{\beta \geqslant c} 2\beta e^{-2\alpha\beta} = 2ce^{-2\alpha c} = \delta$. После этого определим ε_0 из условия $2c\varepsilon_0 = 0{,}1\alpha$. Тогда при $N \to \infty$ справедливы оценки*

$$L_N(\sigma^*, \theta) \leqslant (2c)^{-1}\delta^2 (DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \tag{2.13}$$

при $|\beta| \leqslant c$,

$$L_N(\sigma^*, \theta) \leqslant (\alpha + \delta + o(1))\,(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \tag{2.14}$$

при $c \leqslant |\beta| \leqslant 20c$ и

$$L_N(\sigma^*, \theta) \leqslant \left(\alpha + 20\delta^{20}(2c)^{-19} + 20\delta^{5/2}(2c)^{-3/2} + o(1)\right)(DN)^{1/2} \tag{2.15}$$

при $|\beta| \geqslant 20c$.

Доказательство. Без ограничения общности считаем, что $\beta > 0$. Докажем оценку (2.13). Рассмотрим оценку (2.6), в которой $\Pr(\overline{B})$ оценивается из (2.12) при $t = \varepsilon_0$. Отметим, что $\exp(-(\alpha + 2\beta\varepsilon_0)^2/(4\varepsilon_0))e^{2\alpha\beta} = \exp(-(\alpha - 2\beta\varepsilon_0)^2/(4\varepsilon_0))$, причем при выбранных параметрах $(\alpha - 2\beta\varepsilon_0)(2\varepsilon_0)^{-1/2} \gg 1$. С учетом (2.1), (2.12) отсюда следует, что $\Pr(\overline{B}) < \exp(-(\alpha - 2\beta\varepsilon_0)^2/(4\varepsilon_0))$. Отметим, что $\alpha - 2\beta\varepsilon_0 > 0{,}9\alpha$ и $\alpha = 20c\varepsilon_0$. Поэтому $\Pr(\overline{B}) < \exp(-0{,}81\alpha^2/(4\varepsilon_0)) = \exp(-0{,}81 \times 5\alpha c) < \exp(-4\alpha c)$. Из (2.6) следует, что в этом случае $L_N(\sigma^*, \theta) \leqslant 2ce^{-4\alpha c}(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta)$. Так как $e^{-2\alpha c} = (2c)^{-1}\delta$, то оценка (2.14) выполнена.

Пусть теперь $c \leqslant \beta \leqslant 20c$. Оценка (2.14) следует из (2.7), так как $\Pr(B) \leqslant 1$.

Чтобы установить (2.15), используем (2.7) при $\beta \geqslant 20c$, при этом $\max\limits_{\beta \geqslant 20c} 2\beta e^{-2\alpha\beta} = 40ce^{-40\alpha c} = 20\delta(2c)^{-19}\delta^{19}$. Требуется дополнительно оценить $L_{N-2n_0}(\sigma, \theta)\Pr(B)$. При выбранных параметрах $\Pr(B) < \Phi(-(2\beta\varepsilon_0 - \alpha)/(2\varepsilon_0)^{1/2})$, причем $\beta\varepsilon_0 > \alpha$. С учетом (2.1) получаем оценку $\Pr(B) < \exp\left(-(2\beta\varepsilon_0 - \alpha)^2/(4\varepsilon_0)\right) < \exp\left(-\beta^2\varepsilon_0/4\right) < e^{-\alpha\beta/4}$. Так как всегда верно $L_{N-2n_0}(\sigma, \theta) \leqslant 2\beta(DN)^{1/2}$, то $L_{N-2n_0}(\sigma, \theta)\Pr(B) \leqslant (DN)^{1/2} \max\limits_{\beta \geqslant 20c}(2\beta e^{-\alpha\beta/4})$. При выбранных параметрах имеем $\max\limits_{\beta \geqslant 20c}(2\beta e^{-\alpha\beta/4}) = 40ce^{-5\alpha c} = 20\delta(2c)^{-3/2}\delta^{3/2}$. ▲

Следствие 1. *Обозначим $R_N^{*M}(\Theta) = \inf\limits_{\{\sigma^*\}} \sup\limits_{\Theta} L_N(\sigma, \theta)$. Ясно, что $R_N^M(\Theta_C) \leqslant R_N^{*M}(\Theta_C)$. Из теоремы следует, что для любого $\varepsilon > 0$ с помощью выбора подходящих α, c и ε_0 можно при больших N обеспечить выполнение неравенства $R_N^{*M}(\Theta_C) \leqslant R_{N-2n_0}^M(\Theta_{cN^{-1/2}}) + \varepsilon(DN)^{1/2}$. Так как всегда верно $R_{N-2n_0}^M(\Theta_{cN^{-1/2}}) \leqslant R_N^M(\Theta_{cN^{-1/2}})$, а $R_N^M(\Theta_{cN^{-1/2}}) \leqslant R_N^M(\Theta_C)$ при $cN^{-1/2} < C$, то из теоремы следует, что*

$$\limsup_{N\to\infty} (DN)^{-1/2} \left| R_N^M\left(\Theta_{cN^{-1/2}}\right) - R_N^M\left(\Theta_C\right)\right| \leqslant \varepsilon.$$

Поэтому в следующих параграфах анализ ведется в области "близких" распределений. Приведем примеры выбора значений α, c, ε_0. Пусть $\alpha = \delta = 0{,}01$. Тогда $2c \approx 1170$, $\varepsilon_0 \approx 0{,}85 \times 10^{-6}$. Если $\alpha = \delta = 0{,}001$, то $2c \approx 16700$, $\varepsilon_0 \approx 0{,}6 \times 10^{-8}$.

Это говорит о том, что оценки (2.13)–(2.15) носят скорее теоретический характер. Практически приемлемые параметры можно выбрать с помощью математического моделирования, один такой пример приведен в [10].

Таким образом, при больших N имеет место парадоксальная ситуация: наиболее высокие требования к управлению надо обеспечить там, где, казалось бы, оно не требуется совсем, т.е. в области "близких" распределений. Покажем теперь, что использование пакетной обработки позволяет выйти из области "близких" распределений. Рассмотрим двурукий бандит с доходами ζ_t, $t = 1, \dots, T$, характеризуемыми математическими ожиданиями m_1, m_2 и дисперсиями D_1, D_2 в случае применения первого и второго действий, причем $m_1 - m_2 = xT^{-1/2}$. Пусть $T = NM$. Будем применять одинаковые действия к пакетам из M данных, для управления используем значения процесса $\xi_n = M^{-1/2} \sum\limits_{t=(n-1)M+1}^{nM} \zeta_t$, $n = 1, \dots, N$. Тогда $m'_\ell = \mathbf{E}(\xi_n \,|\, y_n = \ell) = M^{1/2} m_\ell$, $D'_\ell = \mathbf{D}(\xi_n \,|\, y_n = \ell) = D_\ell$, $\ell = 1, 2$, и следовательно, $m'_1 - m'_2 = xN^{-1/2}$. Если число пакетов и полное число данных связаны соотношением $N \ll T$, то распределения доходов в пакетах уже не будут "близкими" относительно T. При этом в [1] установлено, что пакетная обработка практически не приводит к увеличению минимаксного риска, если количество пакетов достаточно велико.

§ 3. Инвариантное описание

Далее будем рассматривать управление в области "близких" распределений и на всем горизонте $[1, N]$. Получим рекуррентное уравнение для нахождения байесовских стратегии и риска относительно класса априорных распределений, к которому принадлежит наихудшее, в инвариантной форме с горизонтом управления, равным единице. Для вычислений удобно поменять параметризацию следующим образом: $m_1 = m + v$, $m_2 = m - v$, тогда множество "близких" распределений принимает вид $\Theta_{cN^{-1/2}} = \{\theta : |v| \leqslant cN^{-1/2}\}$. В [1] установлено, что асимптотически наихудшая априорная плотность распределения может быть выбрана в виде

$$\nu_a(m, v) = \varkappa_a(m)\rho(v), \tag{3.1}$$

где $\varkappa_a(m) = (2a)^{-1}$ – плотность равномерного распределения на отрезке $|m| \leqslant a$, причем $a \to \infty$. Плотность $\rho(v)$, соответствующая наихудшему распределению вида (3.1), может быть найдена из условия (1.4). В случае $D_1 = D_2$ плотность $\rho(v)$ может быть выбрана симметричной, т.е. $\rho(v) = \rho(-v)$.

Рассмотрим стратегии пакетной обработки, которые сначала применяют действия поровну к $2M_0$ данным, а затем управляют оптимально, применяя действия к пакетам из M данных. Пусть n_1, n_2 – полные количества применений обоих действий к моменту времени $n = n_1 + n_2$, а X_1, X_2 – полные доходы за применение обоих действий. Обозначим $n^*_\ell = n_\ell D_\ell$, $n'_\ell = n_\ell / D_\ell$, $M^*_\ell = MD_\ell$, $M'_\ell = M/D_\ell$, $\ell = 1, 2$, а также $U = (X_1 n_2 - X_2 n_1)/n'$, где $n' = n'_1 + n'_2$. В [1] получено рекуррентное уравнение (см. [1, теорема 2, формулы (3.8)–(3.11)]), позволяющее найти байесовские стратегию и риск, соответствующие априорной плотности вида (3.1) с помощью рекуррентного вычисления рисков $R_M(U, n_1, n_2)$. Для получения рекуррентного уравнения в инвариантной форме с горизонтом управления, равным единице, положим

$$\begin{aligned}
&C = cN^{-1/2}, \quad w = N^{1/2}v, \quad \varrho(w) = N^{-1/2}\rho(v), \quad u = UN^{-1/2},\\
&t_\ell = n_\ell N^{-1}, \quad t^*_\ell = n^*_\ell N^{-1}, \quad t'_\ell = n'_\ell N^{-1}, \quad t = t_1 + t_2, \quad t' = t'_1 + t'_2,\\
&r_\varepsilon(u, t_1, t_2) = N^{-1/2} R_M(U, n_1, n_2), \quad r^{(\ell)}_\varepsilon(u, t_1, t_2) = N^{-1/2} R^{(\ell)}_M(U, n_1, n_2),\\
&\varepsilon_0 = M_0 N^{-1}, \quad \varepsilon = MN^{-1}, \quad \varepsilon^*_\ell = M^*_\ell N^{-1}, \quad \varepsilon'_\ell = M'_\ell N^{-1}, \quad \ell = 1, 2.
\end{aligned} \tag{3.2}$$

Кроме того, обозначим $f_D(u) = f_D(u\,|\,0)$, $D_g^2 = D_1 D_2$, $D_h^{-1} = 0{,}5(D_1^{-1} + D_2^{-1})$. Справедлива

Теорема 3. *Пусть априорная плотность распределения имеет вид* (3.1) *и* $a \to \infty$. *Байесовские стратегия и риск могут быть найдены в результате решения рекуррентного уравнения динамического программирования*

$$r_\varepsilon(u, t_1, t_2) = \min_{\ell=1,2} r_\varepsilon^{(\ell)}(u, t_1, t_2), \tag{3.3}$$

где $r_\varepsilon^{(1)}(u, t_1, t_2) = r_\varepsilon^{(2)}(u, t_1, t_2) = 0$ *при* $t_1 + t_2 = 1$, *и далее*

$$\begin{aligned} r_\varepsilon^{(1)}(u, t_1, t_2) &= \varepsilon g^{(1)}(u, t_1, t_2) + r_\varepsilon(u, t_1 + \varepsilon, t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u), \\ r_\varepsilon^{(2)}(u, t_1, t_2) &= \varepsilon g^{(2)}(u, t_1, t_2) + r_\varepsilon(u, t_1, t_2 + \varepsilon) * f_{\varepsilon_2^* t_1^2 (t')^{-1} (t'+\varepsilon_2')^{-1}}(u) \end{aligned} \tag{3.4}$$

при $2\varepsilon_0 \leqslant t_1 + t_2 < 1$. *Здесь* $r(u) * f(u) = \int\limits_{-\infty}^{\infty} r(u - x) f(x)\, dx$ *означает свертку функций,*

$$\begin{aligned} g^{(1)}(u, t_1, t_2) &= \int\limits_{-c}^{0} 2|w| g(w; u, t_1, t_2) \varrho(w)\, dw, \\ g^{(2)}(u, t_1, t_2) &= \int\limits_{0}^{c} 2w g(w; u, t_1, t_2) \varrho(w)\, dw, \\ g(w; u, t_1, t_2) &= \exp\left(2D_g^{-2}(uw - w^2 t_1 t_2 (t')^{-1})\right). \end{aligned} \tag{3.5}$$

Байесовская стратегия при $t \leqslant 2\varepsilon_0$ $(n \leqslant 2M_0)$ *применяет действия по очереди. При* $t > 2\varepsilon_0$ $(n > 2M_0)$ *на промежутке времени* $(t, t + \varepsilon]$ (*иначе говоря,* $tN < n \leqslant tN + M$) *применяется то действие, которому соответствует меньшая величина* $r_\varepsilon^{(\ell)}(u, t_1, t_2)$; *при равенстве* $r_\varepsilon^{(1)}(u, t_1, t_2) = r_\varepsilon^{(2)}(u, t_1, t_2)$ *выбор может быть произвольным. Тогда байесовский риск, соответствующий* (3.1), *вычисляется по формуле*

$$N^{-1/2} R_N^B(\rho(v)) = l(\varrho(w)) + \int\limits_{-\infty}^{\infty} f_{0,5\varepsilon_0 D_g^2 D_h}(u) r_\varepsilon(u, \varepsilon_0, \varepsilon_0)\, du, \tag{3.6}$$

где

$$l(\varrho(w)) = \varepsilon_0 \int\limits_{-c}^{c} 2|w| \varrho(w)\, dw.$$

Доказательство. Теорема доказывается путем выполнения замены переменных (3.2) в уравнениях (3.8)–(3.11) из [1]. ▲

Отметим, что при различных горизонтах управления N уравнение в инвариантной форме может оказаться одинаковым, так как оно определяется не горизонтом управления и размерами начального и последующих пакетов M_0, M, а относительными размерами пакетов ε_0, ε.

Из теоремы 3 следует, что оптимальная стратегия всегда выбирает либо первое, либо второе действие. Важным свойством стратегии является ее пороговый характер. Положим $\Delta r_\varepsilon(u, t_1, t_2) = r_\varepsilon^{(1)}(u, t_1, t_2) - r_\varepsilon^{(2)}(u, t_1, t_2)$. Ясно, что критериями выбора первого и второго действий являются условия $\Delta r_\varepsilon(u, t_1, t_2) < 0$ и $\Delta r_\varepsilon(u, t_1, t_2) > 0$. Обозначим $G^+(\cdot) = \max(G(\cdot), 0)$, $G^-(\cdot) = \max(-G(\cdot), 0)$.

Теорема 4. *Справедливо представление* $\Delta r_\varepsilon(u,t_1,t_2) = \varepsilon G_\varepsilon(u,t_1,t_2)$, *где* $G_\varepsilon(u,t_1,t_2) = G_\varepsilon^{(1)}(u,t_1,t_2) - G_\varepsilon^{(2)}(u,t_1,t_2)$. *При этом* $G_\varepsilon^{(\ell)}(u,t_1,t_2) = g^{(\ell)}(u,t_1,t_2)$ *при* $t_1+t_2 = 1-\varepsilon$ *и далее* $G_\varepsilon^{(1)}(u,t_1,t_2) = G_\varepsilon^{+}(u,t_1,t_2+\varepsilon) * f_{\varepsilon_2^* t_1^2 (t')^{-1}(t'+\varepsilon_2')^{-1}}(u)$, $G_\varepsilon^{(2)}(u,t_1,t_2) = G_\varepsilon^{-}(u,t_1+\varepsilon,t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1}(t'+\varepsilon_1')^{-1}}(u)$ *при* $t_1+t_2 < 1-\varepsilon$, $\ell = 1,2$. *Поэтому* $\Delta r_\varepsilon(u,t_1,t_2)$ *монотонно убывает по* u *так, что* $\Delta r_\varepsilon(-\infty,t_1,t_2) = +\infty$, $\Delta r_\varepsilon(+\infty,t_1,t_2) = -\infty$. *Следовательно, существуют пороги* $\{T_\varepsilon(t_1,t_2)\}$, *такие что* $\Delta r_\varepsilon(u,t_1,t_2) < 0$ *при* $u > T_\varepsilon(t_1,t_2)$, $\Delta r_\varepsilon(u,t_1,t_2) > 0$ *при* $u < T_\varepsilon(t_1,t_2)$.

В случае $D_1 = D_2 = 1$ эквивалентная теорема доказана в [11] (см. [11, теорема 2]). Доказательство легко переносится на рассматриваемый случай.

Замечание 1. Из (3.3)–(3.5) следует, что все риски $\{r_\varepsilon(u,t_1,t_2)\}$ непрерывны по u, следовательно, все $\{r_\varepsilon^{(\ell)}(u,t_1,t_2)\}$ – бесконечно дифференцируемы по u. Из теоремы 4 дополнительно следует, что первые производные $\{(r_\varepsilon)'_u(u,t_1,t_2)\}$ имеют ровно по одному скачку порядка ε.

Наконец, дадим инвариантное рекуррентное уравнение для вычисления потерь. Ограничимся стратегиями σ, которые к первым $2M_0$ данным применяют оба действия равное количество раз, а затем одинаковые действия применяют к пакетам из M данных в зависимости от текущей предыстории U, n_1, n_2, так как именно к этому классу принадлежит оптимальная стратегия.

Теорема 5. *Пусть* $\sigma_\ell(u,t_1,t_2)$ *обозначает стратегию, т.е.* $\sigma_\ell(u,t_1,t_2) = \Pr(y_{(t,t+\varepsilon]} = \ell \,|\, u,t_1,t_2)$, $\ell = 1,2$. *В условиях теоремы* 3 *ожидаемые потери вычисляются в соответствии с уравнением*

$$l_\varepsilon(u,t_1,t_2) = \sigma_1(u,t_1,t_2) l_\varepsilon^{(1)}(u,t_1,t_2) + \sigma_2(u,t_1,t_2) l_\varepsilon^{(2)}(u,t_1,t_2), \tag{3.7}$$

где $l_\varepsilon^{(1)}(u,t_1,t_2) = l_\varepsilon^{(2)}(u,t_1,t_2) = 0$ *при* $t_1+t_2 = 1$,

$$\begin{aligned} l_\varepsilon^{(1)}(u,t_1,t_2) &= \varepsilon g^{(1)}(u,t_1,t_2) + l_\varepsilon(u,t_1+\varepsilon,t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1}(t'+\varepsilon_1')^{-1}}(u), \\ l_\varepsilon^{(2)}(u,t_1,t_2) &= \varepsilon g^{(2)}(u,t_1,t_2) + l_\varepsilon(u,t_1,t_2+\varepsilon) * f_{\varepsilon_2^* t_1^2 (t')^{-1}(t'+\varepsilon_2')^{-1}}(u), \end{aligned} \tag{3.8}$$

если $t_1+t_2 < 1$, $t_1 \geqslant \varepsilon_0$ *и* $t_2 \geqslant \varepsilon_0$. *Ожидаемые потери вычисляются в соответствии с формулой*

$$N^{-1/2} L_N(\sigma;\rho(v)) = l(\varrho(w)) + \int\limits_{-\infty}^{\infty} f_{0,5\varepsilon_0 D_g^2 D_h}(u) l_\varepsilon(u,\varepsilon_0,\varepsilon_0)\, du. \tag{3.9}$$

Доказательство. Теорема доказывается путем выполнения замены переменных (3.2) в уравнениях (3.15)–(3.17) из [1]. Дополнительно выполняются замены $l_\varepsilon(u,t_1,t_2) = N^{-1/2} L_M(U,n_1,n_2)$, $l_\varepsilon^{(\ell)}(u,t_1,t_2) = N^{-1/2} L_M^{(\ell)}(U,n_1,n_2)$, $\ell = 1,2$. ▲

§ 4. Предельное описание

В этом параграфе устанавливаются условия Липшица для риска $r_\varepsilon(u,t_1,t_2)$ по аргументам u,t_1,t_2 и его предельное описание дифференциальным уравнением в частных производных второго порядка. Кроме того, устанавливается упомянутый в § 1 результат о том, что малое изменение дисперсий приводит к малому изменению потерь. Далее без ограничения общности считаем, что $\max(D_1,D_2) = 1$, $\min(D_1,D_2) = \alpha > 0$ (см. [1, теорема 4]).

Положим $c_1 = c\alpha^{-1}$, $k(u) = \min(k^{(1)}(u), k^{(2)}(u))$, где

$$k^{(1)}(u) = \int\limits_{-c}^{0} 2|w| \exp\left(2D_g^{-2}uw\right) \varrho(w)\, dw, \quad k^{(2)}(u) = \int\limits_{0}^{c} 2w \exp\left(2D_g^{-2}uw\right) \varrho(w)\, dw.$$

Очевидно, что $k(u)$ – ограниченная функция, причем $k(u) \to +0$ при $u \to \pm\infty$, а $g^{(\ell)}(u, t_1, t_2) \leqslant k^{(\ell)}(u)$ при всех допустимых u, t_1, t_2 и $\ell = 1, 2$.

Лемма 4. *Справедливы оценки*

$$\begin{aligned} &k(u) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u) \leqslant e^{2cc_1\varepsilon} k(u), \\ &k(u) * f_{\varepsilon_2^* t_1^2 (t')^{-1} (t'+\varepsilon_2')^{-1}}(u) \leqslant e^{2cc_1\varepsilon} k(u). \end{aligned} \tag{4.1}$$

Доказательство. Непосредственно проверяется равенство $\exp(Cu) * f_\varepsilon(u) = \exp(Cu) \times \exp(0{,}5C^2\varepsilon)$, где C – функция, не зависящая от u. Проверим первую оценку в (4.1). Действительно,

$$\begin{aligned} &k^{(\ell)}(u) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u) \leqslant \exp\left(0{,}5 \cdot 4 (D_1 D_2)^{-2} c^2 \varepsilon D_1 D_2^2\right) k^{(\ell)}(u) \leqslant \\ &\leqslant e^{2cc_1\varepsilon} k^{(\ell)}(u). \end{aligned} \tag{4.2}$$

Далее, так как $k(u) \leqslant k^{(\ell)}(u)$, то $k(u) * f_\varepsilon(u) \leqslant k^{(\ell)}(u) * f_\varepsilon(u)$, $\ell = 1, 2$, и следовательно, $k(u) * f_\varepsilon(u) \leqslant \min\left(k^{(1)}(u) * f_\varepsilon(u), k^{(2)}(u) * f_\varepsilon(u)\right)$. Поэтому из (4.2) следует первая оценка в (4.1). Вторая оценка в (4.1) проверяется аналогично. ▲

Установим равномерные ограниченность и непрерывность $r_\varepsilon(u, t_1, t_2)$.

Лемма 5. *Справедлива оценка*

$$r_\varepsilon(u, t_1, t_2) \leqslant (1 - t) e^{2(1-t-\varepsilon)cc_1} k(u). \tag{4.3}$$

Доказательство проводится по индукции. При $t = 1 - \varepsilon$ имеем $r^{(\ell)}(u, t_1, t_2) \leqslant \varepsilon k^{(\ell)}(u)$, $\ell = 1, 2$, и следовательно, оценка справедлива. Пусть она справедлива при $t_1 + t_2 = t + \varepsilon$. С учетом (3.4), предположения индукции и (4.1) получаем

$$\begin{aligned} &r_\varepsilon^{(1)}(u, t_1, t_2) \leqslant \varepsilon k^{(1)}(u) + r_\varepsilon(u, t_1 + \varepsilon, t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u) \leqslant \\ &\leqslant \varepsilon k^{(1)}(u) + (1 - t - \varepsilon) e^{2(1-t-2\varepsilon)cc_1} e^{2\varepsilon cc_1} k(u) \leqslant (1 - t) e^{2(1-t-\varepsilon)cc_1} k^{(1)}(u). \end{aligned}$$

Точно так же $r_\varepsilon^{(2)}(u, t_1, t_2) \leqslant (1 - t) e^{2(1-t-\varepsilon)cc_1} k^{(2)}(u)$. Отсюда следует (4.3). ▲

Лемма 6. *Справедлива оценка*

$$|(r_\varepsilon)'_u(u, t_1, t_2)| \leqslant 2c_1 (1 - t) e^{2(1-t-\varepsilon)cc_1} k(u). \tag{4.4}$$

Доказательство. Отметим, что справедливо представление

$$r_\varepsilon(u, t_1, t_2) = \sigma_1(u, t_1, t_2) r_\varepsilon^{(1)}(u, t_1 + \varepsilon, t_2) + \sigma_2(u, t_1, t_2) r_\varepsilon^{(2)}(u, t_1, t_2 + \varepsilon), \tag{4.5}$$

где $\sigma_\ell(u, t_1, t_2)$ – байесовская стратегия, которая является кусочно-постоянной и принимает только значения 0 и 1, причем ее нужно считать неизменной при дифференцировании (4.5) по u. Поэтому

$$\begin{aligned} &|(r_\varepsilon)'_u(u, t_1, t_2)| \leqslant \\ &\leqslant \sigma_1(u, t_1, t_2) |(r_\varepsilon^{(1)})'_u(u, t_1 + \varepsilon, t_2)| + \sigma_2(u, t_1, t_2) |(r_\varepsilon^{(2)})'_u(u, t_1, t_2 + \varepsilon)|. \end{aligned} \tag{4.6}$$

Кроме того, из (3.4) следуют оценки

$$
\begin{aligned}
&|(r_\varepsilon^{(1)})'_u(u,t_1,t_2)| \leqslant \\
&\leqslant \varepsilon|(g^{(1)})'_u(u,t_1,t_2)| + |(r_\varepsilon)'_u(u,t_1+\varepsilon,t_2)| * f_{\varepsilon_1^* t_2^2 (t')^{-1}(t'+\varepsilon_1')^{-1}}(u), \\
&|(r_\varepsilon^{(2)})'_u(u,t_1,t_2)| \leqslant \\
&\leqslant \varepsilon|(g^{(2)})'_u(u,t_1,t_2)| + |(r_\varepsilon)'_u(u,t_1,t_2+\varepsilon)| * f_{\varepsilon_2^* t_1^2 (t')^{-1}(t'+\varepsilon_2')^{-1}}(u).
\end{aligned}
\tag{4.7}
$$

Отметим, что $|(g^{(\ell)})'_u(u,t_1,t_2)| \leqslant 2cD_g^{-2}k^{(\ell)}(u) = 2c_1k^{(\ell)}(u)$ при всех допустимых u,t_1,t_2 и $\ell = 1,2$. Дальнейшие рассуждения выполняются по индукции аналогично приведенным в лемме 5 с учетом (4.6), (4.7) и того, что $\sigma_\ell(u,t_1,t_2)$ – стратегия, обеспечивающая нахождение $r(u,t_1,t_2)$. ▲

Установим условия Липшица для $r_\varepsilon(u,t_1,t_2)$ по t_1,t_2. Ограничимся оценками по t_1.

Лемма 7. *Пусть* $\delta = K\varepsilon$, $\widetilde{\varepsilon}_i = \varepsilon_1^* t_2^2 (t' + (K-i)\varepsilon_1')^{-1}(t' + (K-i+1)\varepsilon_1')^{-1}$, $\widetilde{\delta}_i = \sum\limits_{j=1}^{i} \widetilde{\varepsilon}_j$. *Тогда справедливо неравенство*

$$
|r_\varepsilon(u,t_1,t_2) - r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u)| \leqslant \delta k^{(1)}(u) e^{2cc_1\delta}. \tag{4.8}
$$

Доказательство. Обозначим через $r_\varepsilon^{(1,K)}(u,t_1,t_2)$ риск, соответствующий тому, что сначала K раз применялось первое действие, а затем управление выполнялось оптимально. Тогда при $i = 1,\ldots,K$ выполнено равенство

$$
\begin{aligned}
&r_\varepsilon^{(1,i)}(u,t_1+(K-i)\varepsilon,t_2) = \\
&= \varepsilon g^{(1)}(u,t_1+(K-i)\varepsilon,t_2) + r_\varepsilon^{(1,i-1)}(u,t_1+(K-i+1)\varepsilon,t_2) * f_{\widetilde{\varepsilon}_i}(u),
\end{aligned}
$$

где $r_\varepsilon^{(1,0)}(\cdot) = r_\varepsilon(\cdot)$. Так как $f_{\widetilde{\varepsilon}_i}(u) * f_{\widetilde{\varepsilon}_j}(u) = f_{\widetilde{\varepsilon}_i+\widetilde{\varepsilon}_j}(u)$, а $\widetilde{\delta}_i \leqslant D_1\delta(t_2/t')^2$, то отсюда следует, что

$$
\begin{aligned}
&r_\varepsilon^{(1,K)}(u,t_1,t_2) \leqslant \varepsilon\sum_{i=1}^{K} g^{(1)}(u,t_1+(K-i)\varepsilon,t_2) * f_{\widetilde{\delta}_i}(u) + r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u) \leqslant \\
&\leqslant \delta k^{(1)}(u) e^{2cc_1\delta} + r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u).
\end{aligned}
$$

Здесь оценка первого слагаемого аналогична оценкам леммы 4. Отметим, что $r_\varepsilon(u,t_1,t_2) \leqslant r_\varepsilon^{(1,K)}(u,t_1,t_2)$. С другой стороны, справедлива оценка $r_\varepsilon(u,t_1,t_2) \geqslant \geqslant r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u)$, которая следует из того, что байесовский риск на меньшем горизонте управления $[t+\delta,1]$ при наличии дополнительной информации, обусловленной K-кратным применением первого действия, всегда не превосходит байесовского риска на большем горизонте управления $[t,1]$. Отсюда следует (4.8). ▲

Следствие 2. *Так как* $|(r_\varepsilon)'_u(\cdot)| \leqslant k = \max\limits_u 2c_1e^{2(1-\varepsilon)cc_1}k(u)$ *равномерно по* u, *то* $|r_\varepsilon(u,t_1+\delta,t_2) - r_\varepsilon(u-x,t_1+\delta,t_2)| \leqslant k|x|$, *поэтому*

$$
|r_\varepsilon(u,t_1+\delta,t_2) - r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u)| \leqslant k\int\limits_{-\infty}^{\infty} |x| f_{\widetilde{\delta}_K}(x)\,dx = k(2/\pi)^{1/2}\delta_K^{1/2}. \tag{4.9}
$$

Из (4.8), (4.9) *следует, что при малых* δ *разность* $|r_\varepsilon(u,t_1,t_2) - r_\varepsilon(u,t_1+\delta,t_2)|$ *равномерно ограничена величиной порядка* $\delta^{1/2}$. *С использованием замечания* 1 *нетрудно*

установить, что $|r_\varepsilon(u,t_1,t_2)-r_\varepsilon(u,t_1+\delta,t_2)|$ *ограничена величиной порядка* δ, *однако соответствующая равномерная оценка пока не получена.*

Для дальнейшего сделаем несколько замечаний. Во-первых, для наихудшей априорной плотности $\varrho(w)$ найдется такое $w_0>0$, что $\varrho(w)=0$ при $|w|<w_0$. Это следует из того, что соответствующая байесовская стратегия является уравнивающей, т.е. обеспечивает одинаковые потери, равные минимаксному риску, для всех w, на которых $\varrho(w)\neq 0$, а потери, соответствующие w, не превосходят $2|w|$. Поэтому $k(u)\leqslant 2c\exp(-\alpha^{-1}w_0|u|)$, и в частности, $k(u)\leqslant\varepsilon|\delta|$ при $|u|\geqslant U_1=$ $=\alpha w_0^{-1}\left(\ln(2c)-\ln(\varepsilon)-\ln|\delta|\right)$.

Во-вторых, если для вычисления потерь использовать одинаковую стратегию, но различные функции $g^{(\ell)}(u,t_1,t_2)$ и $\widehat{g}^{(\ell)}(u,t_1,t_2)$, удовлетворяющие условию $|g^{(\ell)}(u,t_1,t_2)-\widehat{g}^{(\ell)}(u,t_1,t_2)|\leqslant\varepsilon|\delta|$, то для соответствующих им потерь справедлива оценка $|l(u,t_1,t_2)-\widehat{l}(u,t_1,t_2)|\leqslant|\delta|(1-t)$, которая устанавливается по индукции.

В-третьих, обозначим через U_2 максимальное абсолютное значение порога переключения стратегии σ. Ясно, что достаточно определить $g^{(1)}(u,t_1,t_2)$ при $u\geqslant -U_2$, а $g^{(2)}(u,t_1,t_2)$ при $u\leqslant U_2$. Обозначим $U_3=\max(U_1,U_2)$. Из сделанных замечаний следует, что если положить $\widehat{g}^{(\ell)}(u,t_1,t_2)=g^{(\ell)}(u,t_1,t_2)$ при $|u|\leqslant U_3$ и $\widehat{g}^{(\ell)}(u,t_1,t_2)=$ $=0$ при $|u|>U_3$, то для соответствующих потерь будет выполнена оценка $|l(u,t_1,t_2)-\widehat{l}(u,t_1,t_2)|\leqslant|\delta|(1-t)$.

Установим теперь, что малые изменения дисперсий приводят к малым изменениям потерь. Достаточно рассмотреть изменения D_1 и потери $\widehat{l}_\varepsilon(u,t_1,t_2)$. Справедлива

Лемма 8. *Пусть* D_2 *фиксирована, а* D_1 *может меняться. Обозначим через* $\widehat{l}_\varepsilon(D_1;\sigma;\cdot)$ *потери, в которых явно указана зависимость от* D_1 *и* σ *и использованы функции* $\{\widehat{g}^{(\ell)}(u,t_1,t_2)\}$. *Тогда для достаточно малых* $|\delta|$ *справедлива оценка*

$$|\widehat{l}_\varepsilon(D_1;\sigma;u,t_1,t_2)-\widehat{l}_\varepsilon(D_1+\delta;\sigma;u,t_1,t_2)|\leqslant 3|\delta|\alpha^{-2}(U_3c+c^2)k(u), \tag{4.10}$$

где $\sigma=\sigma_\ell(D_1;u,t_1,t_2)$ – *байесовская стратегия, соответствующая* D_1,u,t_1,t_2, *а* $U_3=\max(U_1,U_2)$ *и при фиксированном* ε *имеет порядок* $\ln|\delta|$.

Доказательство. Так как $|uw-w^2t_1t_2(t')^{-1}|\leqslant(U_3c+c^2)$ при $|u|\leqslant U_3$, а $D_1^{-1}-$ $-(D_1+\delta)^{-1}=D_1^{-2}(\delta+o(\delta))$, то

$$\begin{aligned}
&|g^{(\ell)}(D_1,u,t_1,t_2)-g^{(\ell)}(D_1+\delta,u,t_1,t_2)|\leqslant\\
&\leqslant g^{(\ell)}(D_1,u,t_1,t_2)\times\left(\exp\left\{2D_2^{-1}(U_3c+c^2)\left|D_1^{-1}-(D_1+\delta)^{-1}\right|\right\}-1\right)\leqslant\\
&\leqslant g^{(\ell)}(D_1,u,t_1,t_2)\left(\exp\left\{2(|\delta|+o(\delta))\alpha^{-2}(U_3c+c^2)\right\}-1\right)
\end{aligned}$$

при $|u|\leqslant U_3$ и для достаточно малых $|\delta|$

$$|\widehat{g}^{(\ell)}(D_1;u,t_1,t_2)-\widehat{g}^{(\ell)}(D_1+\delta;u,t_1,t_2)|\leqslant 3|\delta|\alpha^{-2}(U_3c+c^2)g^{(\ell)}(D_1;u,t_1,t_2).$$

Обозначим

$$\begin{aligned}
&\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)=|\widehat{l}_\varepsilon(D_1,\sigma;u,t_1,t_2)-\widehat{l}_\varepsilon(D_1+\delta;\sigma;u,t_1,t_2)|,\\
&\Delta\widehat{g}^{(\ell)}(D_1;u,t_1,t_2)=|\widehat{g}^{(\ell)}(D_1;u,t_1,t_2)-\widehat{g}^{(\ell)}(D_1+\delta;u,t_1,t_2)|.
\end{aligned}$$

Так как $\sigma=\sigma_\ell(D_1;u,t_1,t_2)$ – оптимальная стратегия, соответствующая D_1,u,t_1,t_2, то

$$\begin{aligned}
&\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)\leqslant\sigma_1(D_1;u,t_1,t_2)\Delta\widehat{l}_\varepsilon^{(1)}(D_1;u,t_1,t_2)+\\
&+\sigma_2(D_1;u,t_1,t_2)\Delta\widehat{l}_\varepsilon^{(2)}(D_1;u,t_1,t_2),
\end{aligned}$$

где

$$\Delta\widehat{l}_\varepsilon^{(1)}(D_1;u,t_1,t_2)\leqslant\varepsilon\Delta\widehat{g}^{(1)}(D_1;u,t_1,t_2)+\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)*f_{\varepsilon_1^*t_2^2(t')^{-1}(t'+\varepsilon_1')^{-1}}(u),$$
$$\Delta\widehat{l}_\varepsilon^{(2)}(D_1;u,t_1,t_2)\leqslant\varepsilon\Delta\widehat{g}^{(2)}(D_1;u,t_1,t_2)+\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)*f_{\varepsilon_2^*t_1^2(t')^{-1}(t'+\varepsilon_2')^{-1}}(u).$$

Далее оценки выполняются по индукции аналогично приведенным в леммах 5, 6. ▲

Установим существование непрерывного предела $r_\varepsilon(u,t_1,t_2)$ при $\varepsilon\to 0$.

Теорема 6. *При всех u,t_1,t_2, при которых определено решение уравнения* (3.3)–(3.5), *существует непрерывный предел $r(u,t_1,t_2)=\lim\limits_{\varepsilon\to+0}r_\varepsilon(u,t_1,t_2)$, который можно доопределить по непрерывности на все допустимые u,t_1,t_2 ($t_1\geqslant\varepsilon_0$, $t_2\geqslant\varepsilon_0$, $t_1+t_2\leqslant 1$). Этот предел равномерно ограничен, имеет равномерно ограниченную производную по u и равномерно непрерывен по t_1,t_2 с константами, указанными в* (4.3), (4.4), (4.8) *и* (4.9) *соответственно. При $\rho_N(v)=N^{-1/2}\varrho(w)$, $w=N^{1/2}v$ и $\varepsilon_0\to+0$ для байесовского риска* (3.6) *справедлива асимптотическая оценка*

$$\lim_{N\to\infty}N^{-1/2}R_N^B(\rho_N(v))=\lim_{\varepsilon_0\to 0}r(\varrho;0,\varepsilon_0,\varepsilon_0),\tag{4.11}$$

где в обозначении $r(\varrho;0,\varepsilon_0,\varepsilon_0)$ явно указана зависимость от $\varrho(w)$.

Доказательство. Пусть N – исходный горизонт управления. Тогда решение уравнения (3.3)–(3.5) определено при $t=2\varepsilon_0+n\varepsilon$, где $\varepsilon_0=M_0/N$ и $\varepsilon=M/N$ – относительные размеры начального и последующего пакетов. Если горизонт управления N фиксирован, то уменьшение величины ε соответствует тому, что действия можно менять чаще. Ясно, что при этом байесовские риски не возрастают. Поэтому $r_\varepsilon(u,t_1,t_2)$ является неотрицательной неубывающей функцией ε, и следовательно, имеет предел при $\varepsilon\to 0$. Непрерывность предела устанавливается предельным переходом в (4.3), (4.4), (4.8) и (4.9). Формула (4.11) следует из (3.6) и непрерывности $r(\varrho;u,t_1,t_2)$. ▲

Приведем дифференциальное уравнение в частных производных второго порядка для вычисления $r(u,t_1,t_2)$. Строгое обоснование этого вывода требует равномерных оценок производных $r_\varepsilon(u,t_1,t_2)$, поэтому ограничимся нестрогими рассуждениями, дополнив их результатами численного моделирования в § 5. Если $r_\varepsilon(u,\cdot\,)$ – достаточно гладкая функция u, то $r_\varepsilon(u-x,\cdot\,)=r_\varepsilon(u,\cdot\,)-(r_\varepsilon)'_u(u,\cdot\,)x+0{,}5(r_\varepsilon)''_{uu}(u,\cdot\,)x^2+o(x^2)$. Так как

$$\int\limits_{-\infty}^{\infty}f_{\widetilde{\varepsilon}}(x)\,dx=1,\quad\int\limits_{-\infty}^{\infty}xf_{\widetilde{\varepsilon}}(x)\,dx=0,\quad\int\limits_{-\infty}^{\infty}x^2f_{\widetilde{\varepsilon}}(x)\,dx=\widetilde{\varepsilon},$$

то из (3.4) при малых ε следуют уравнения

$$r_\varepsilon^{(1)}(u,t_1,t_2)=\varepsilon g^{(1)}(u,t_1,t_2)+r_\varepsilon(u,t_1+\varepsilon,t_2)+$$
$$+0{,}5\varepsilon_1^*t_2^2(t')^{-1}(t'+\varepsilon_1')^{-1}(r_\varepsilon)''_{uu}(u,t_1+\varepsilon,t_2)+o(\varepsilon),$$
$$r_\varepsilon^{(2)}(u,t_1,t_2)=\varepsilon g^{(2)}(u,t_1,t_2)+r_\varepsilon(u,t_1,t_2+\varepsilon)+$$
$$+0{,}5\varepsilon_2^*t_1^2(t')^{-1}(t'+\varepsilon_2')^{-1}(r_\varepsilon)''_{uu}(u,t_1,t_2+\varepsilon)+o(\varepsilon).$$

Дополняя их уравнением (3.3), записанным в виде

$$\min_{\ell=1,2}\Big(r_\varepsilon^{(\ell)}(u,t_1,t_2)-r_\varepsilon(u,t_1,t_2)\Big)=0,$$

в пределе при $\varepsilon \to 0$ получаем уравнение для $r = r(u, t_1, t_2)$

$$\min_{\ell=1,2}\left(\frac{\partial r}{\partial t_\ell} + \frac{D_\ell t_\ell^2}{2(t')^2} \times \frac{\partial^2 r}{\partial u^2} + g^{(\ell)}(u, t_1, t_2)\right) = 0, \tag{4.12}$$

где $\overline{\ell} = 3 - \ell$, с начальным условием

$$r(u, t_1, t_2) = 0 \quad \text{при } t_1 + t_2 = 1. \tag{4.13}$$

При этом байесовская стратегия предписывает выбирать ℓ-е действие, если меньше ℓ-е выражение в левой части (4.12). Для численного определения $r(u, t_1, t_2)$ в соответствии с (4.12) следует использовать разностное уравнение

$$\begin{aligned}
&r(u, t_1, t_2) = \min\bigl(r^{(1)}(u, t_1, t_2), r^{(2)}(u, t_1, t_2)\bigr),\\
&r^{(1)}(u, t_1, t_2) = r(u, t_1 + \Delta t, t_2) +\\
&+ \Delta t\bigl(g^{(1)}(u, t_1, t_2) + 0{,}5 D_1 (t_2/t')^2 D^2 r(u, t_1 + \Delta t, t_2)\bigr),\\
&r^{(2)}(u, t_1, t_2) = r(u, t_1, t_2 + \Delta t) +\\
&+ \Delta t\bigl(g^{(2)}(u, t_1, t_2) + 0{,}5 D_2 (t_1/t')^2 D^2 r(u, t_1, t_2 + \Delta t)\bigr),
\end{aligned} \tag{4.14}$$

где $D^2 r(u, \cdot) = \Delta u^{-2}\,(r(u + \Delta u, \cdot) - 2r(u, \cdot) + r(u - \Delta u, \cdot))$, с начальным условием (4.13). Аналогично для вычисления потерь, соответствующих стратегии σ в предельном случае, следует использовать разностное уравнение

$$\begin{aligned}
&l(u, t_1, t_2) = \sigma_1(u, t_1, t_2) l^{(1)}(u, t_1, t_2) + \sigma_2(u, t_1, t_2) l^{(2)}(u, t_1, t_2),\\
&l^{(1)}(u, t_1, t_2) = l(u, t_1 + \Delta t, t_2) +\\
&+ \Delta t\bigl(g^{(1)}(u, t_1, t_2) + 0{,}5 D_1 (t_2/t')^2 D^2 l(u, t_1 + \Delta t, t_2)\bigr),\\
&l^{(2)}(u, t_1, t_2) = l(u, t_1, t_2 + \Delta t) +\\
&+ \Delta t\bigl(g^{(2)}(u, t_1, t_2) + 0{,}5 D_2 (t_1/t')^2 D^2 l(u, t_1, t_2 + \Delta t)\bigr),
\end{aligned} \tag{4.15}$$

где $D^2 l(u, \cdot) = \Delta u^{-2}\,(l(u + \Delta u, \cdot) - 2l(u, \cdot) + l(u - \Delta u, \cdot))$, с начальным условием $l(u, t_1, t_2) = 0$ при $t_1 + t_2 = 1$.

§ 5. Численные эксперименты

На рис. 1 представлены результаты вычисления рисков $\{r_\varepsilon(u, t_1, t_2)\}$ как минимумов из $\{r_\varepsilon^{(1)}(u, t_1, t_2), r_\varepsilon^{(2)}(u, t_1, t_2)\}$ для $(t_1, t_2) = (0{,}4; 0{,}4), (0{,}2; 0{,}2), (0{,}1; 0{,}1)$. Жирными линиями представлены риски $\{r_\varepsilon^{(1)}(u, t_1, t_2)\}$, тонкими – риски $\{r_\varepsilon^{(2)}(u, t_1, t_2)\}$. При этом меньшим значениям $t_1 + t_2$ соответствуют бóльшие риски $r_\varepsilon^{(1)}(u, t_1, t_2)$, $r_\varepsilon^{(2)}(u, t_1, t_2)$.

Рис. 2 демонстрирует непрерывность рисков по u и их близость при малых изменениях t_2. Жирными линиями представлены четыре группы рисков $\{r(u, t_1, t_2)\}$, полученных в результате решения разностного уравнения (4.14), причем чем меньше $t_1 + t_2$, тем выше соответствующая кривая. Снизу вверх четыре пары кривых представлены для $(t_1, t_2) = (0{,}4; 0{,}44), (0{,}4; 0{,}4)$, $(t_1, t_2) = (0{,}2; 0{,}24), (0{,}2; 0{,}2)$, $(t_1, t_2) = (0{,}1; 0{,}14), (0{,}1; 0{,}1)$, $(t_1, t_2) = (0{,}001; 0{,}001), (0{,}001; 0{,}041)$. Последнюю можно рассматривать как предельную пару кривых $(t_1, t_2) = (\varepsilon_0; \varepsilon_0), (\varepsilon_0; \varepsilon_0 + 0{,}04)$ при $\varepsilon_0 \to +0$, так как кривые, вычисленные при $\varepsilon_0 = 0{,}001; 0{,}002; 0{,}004$, практически совпадают. Для первых трех пар также были вычислены риски $\{r_\varepsilon(u, t_1, t_2)\}$ с использованием уравнения (3.3)–(3.4), они представлены тонкими линиями и совсем немного превышают $\{r(u, t_1, t_2)\}$. Отметим, что кривые на рис. 1 и 2 вычислены при

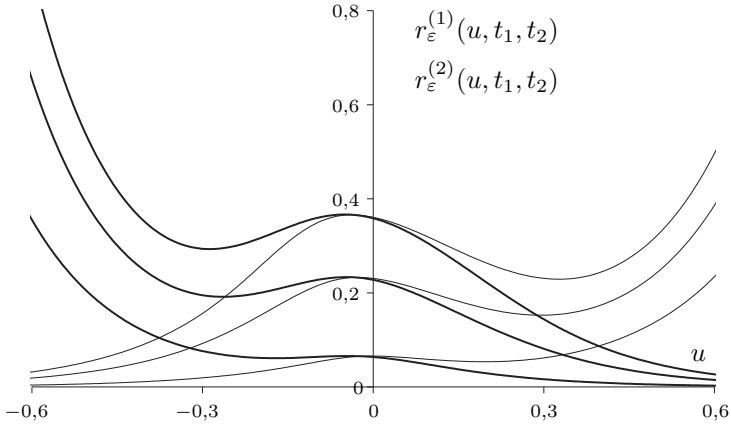

Рис. 1. Риски $r_\varepsilon^{(1)}(u,t_1,t_2)$, $r_\varepsilon^{(2)}(u,t_1,t_2)$ при $(t_1,t_2)=(0{,}4;0{,}4),(0{,}2;0{,}2),(0{,}1;0{,}1)$

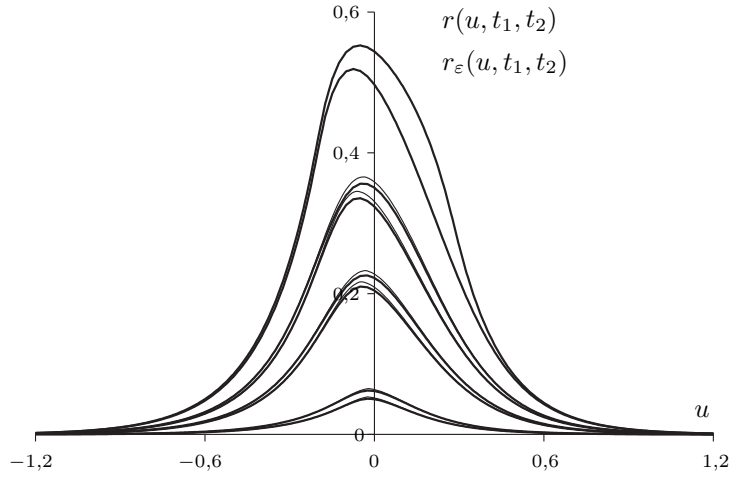

Рис. 2. Риски $r(u,t_1,t_2)$, $r_\varepsilon(u,t_1,t_2)$

$D_1=1$, $D_2=0{,}5$, плотность $\varrho(w)$ сосредоточена в точках $d_1=1{,}49$ и $d_2=-1{,}53$ с вероятностями 0,54 и 0,46 (это распределение близко к наихудшему при выбранных параметрах), $\varepsilon=0{,}02$, $\Delta t=2000^{-1}$. Вычисления выполнялись в области $|u|\leqslant 2{,}5$, шаг Δu равен 0,005 при выполнении численного интегрирования и 0,025 при решении уравнения (4.14). Отметим также, что выбор области и шага численного интегрирования влияет только на точность результата, в то время как при численном решении уравнения (4.14) для обеспечения устойчивости требуется выполнение условия $\Delta t/\Delta u^2<1$ (см., например, [12]).

На рис. 3 представлены риски и потери, вычисленные при $D_1=D_2=1$, если $d_1=-d_2=d$, $\varrho(w)=0{,}5\,(\delta(w-d_1)+\delta(w-d_2))$ и $\delta(\cdot)$ – дельта-функция Дирака. Если наихудшее априорное распределение сосредоточено в двух точках, то оно соответствует d, при котором байесовский риск достигает максимума. Жирные линии *1* и *2* получены в результате решения уравнения (3.3)–(3.5), причем линия *2* характеризует значение байесовского риска без потерь на начальном этапе, равных $2d\varepsilon_0$. Максимум байесовского риска $r_\varepsilon(d)$ приблизительно равен 0,652 при $d\approx 1{,}6$. Жирные линии *3* и *4* соответствуют потерям $l_\varepsilon(d)$ (полным и без потерь на начальном

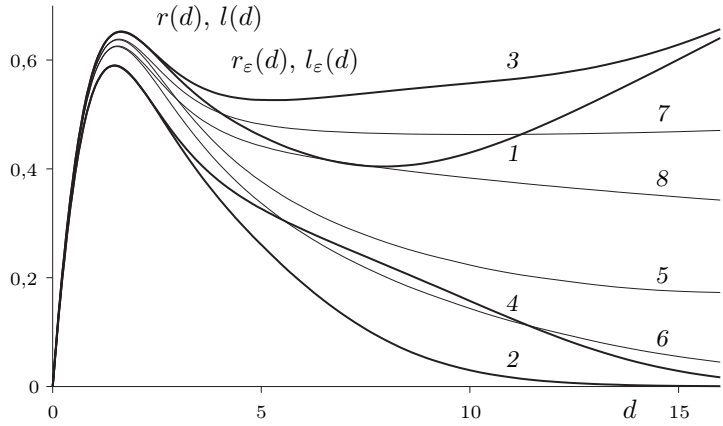

Рис. 3. Риски и потери как функции априорного распределения

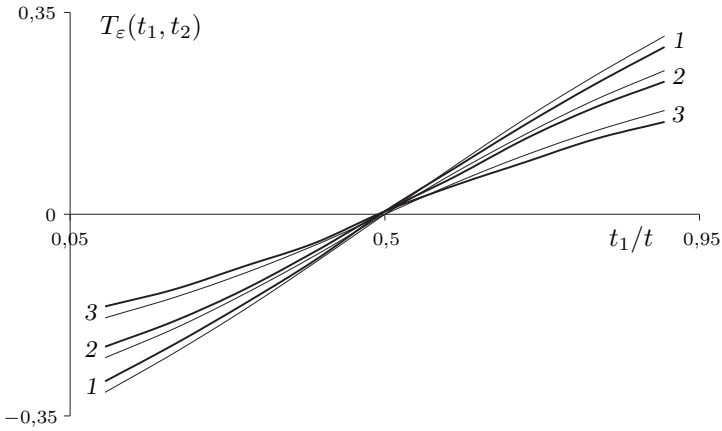

Рис. 4. Значения порогов как функции от t_1/t

этапе), которые вычислены в соответствии с уравнением (3.7)–(3.8) для найденной байесовской стратегии. Они также имеют единственный максимум при $d \approx 1{,}6$. Здесь $\varepsilon = \varepsilon_0 = 0{,}02$, $\Delta u = 0{,}005$, $|u| \leqslant 2{,}5$, т.е. кривые соответствуют пакетной обработке данных, разбитых на 50 пакетов.

Тонкие линии *5* и *6* соответствуют байесовским рискам $r(d)$ (полному и без начальных потерь), найденным в результате решения уравнения (4.14). Максимум байесовского риска при $d \approx 1{,}6$ приблизительно равен 0,637, т.е. меньше, чем для пакетной обработки, приблизительно на 2%. Тонкие линии *7* и *8* соответствуют потерям $l(d)$ (полным и без потерь на начальном этапе), полученным для найденной стратегии в результате решения уравнения (4.15). Здесь $\Delta t = 2000^{-1}$, $\varepsilon_0 = 0{,}004$, $\Delta u = 0{,}025$, $|u| \leqslant 2{,}5$.

Наконец, на рис. 4 представлены значения порогов байесовской стратегии, определенные численными методами при решении уравнения (3.3)–(3.5). Пары линий *1*, *2*, *3* соответствуют значениям $t = 0{,}2; 0{,}6; 0{,}8$, причем жирные линии получены при $\varepsilon = 0{,}02$, а тонкие при $\varepsilon = 0{,}01$. Здесь $D_1 = D_2 = 1$, $d_1 = -d_2 = 1{,}6$, $\varrho(w) = 0{,}5\,(\delta(w - d_1) + \delta(w - d_2))$. Эти результаты показывают, что можно ожидать сходимости стратегии при $\varepsilon \to +0$.

§ 6. Обработка данных по одному. Бернуллиевский двурукий бандит

В этом параграфе показано, что для бернуллиевского двурукого бандита максимальные потери являются такими же, как и для гауссовского, если $N \to \infty$. Это следует из того, что в этом случае уравнение для вычисления байесовского риска совпадает с (4.12), а минимаксный риск не меньше любого байесовского. Доходы бернуллиевского двурукого бандита ξ_n, $n = 1, 2, \ldots, N$, зависят от текущих выбираемых действий y_n следующим образом:

$$\Pr(\xi_n = 1 \,|\, y_n = \ell) = p_\ell, \quad \Pr(\xi_n = 0 \,|\, y_n = \ell) = q_\ell, \quad \ell = 1, 2.$$

Бернуллиевский двурукий бандит описывается векторным параметром $\theta = (p_1, p_2)$. Допустимое множество параметров Θ пока считаем произвольным.

Стратегия управления σ в момент времени $n+1$ описывает выбор действия в зависимости от предыстории (X_1, n_1, X_2, n_2), где n_1, n_2 – текущие полные количества применений обоих действий ($n_1 + n_2 = n$), X_1, X_2 – соответствующие полные доходы. Таким образом, $\sigma_\ell(X_1, n_1, X_2, n_2) = \Pr(y_{n+1} = \ell \,|\, X_1, n_1, X_2, n_2)$, $\ell = 1, 2$. Будем предполагать, что в начале управления стратегия применяет оба действия по n_0 раз. Если $n_0 \ll N$, то это практически не влияет на полные потери. Функция потерь имеет вид

$$L_N(\sigma, \theta) = N(p_1 \vee p_2) - \mathbf{E}_{\sigma,\theta}\left(\sum_{n=1}^{N} \xi_n\right),$$

а байесовский риск относительно априорной плотности распределения $\lambda(p_1, p_2)$ равен

$$R_N^B(\lambda) = \min_{\{\sigma\}} \iint\limits_{\Theta} L_N(\sigma, \theta)\, \lambda(p_1, p_2)\, dp_1\, dp_2. \tag{6.1}$$

Запишем стандартное уравнение динамического программирования для вычисления байесовского риска (6.1). Апостериорная плотность распределения, соответствующая предыстории процесса (X_1, n_1, X_2, n_2), равна

$$\lambda(p_1, p_2 \,|\, X_1, n_1, X_2, n_2) = \frac{B(X_1, n_1 \,|\, p_1) B(X_2, n_2 \,|\, p_2) \lambda(p_1, p_2)}{P(X_1, n_1, X_2, n_2)}, \tag{6.2}$$

$$\text{где} \quad P(X_1, n_1, X_2, n_2) = \iint\limits_{\Theta} B(X_1, n_1 \,|\, p_1) B(X_2, n_2 \,|\, p_2)\, \lambda(p_1, p_2)\, dp_1\, dp_2 \tag{6.3}$$

и

$$B(X, n \,|\, p) = \binom{n}{X} p^X (1-p)^{n-X}.$$

Если дополнительно положить $B(X, n \,|\, p) = 1$ при $n = 0$, $X = 0$, то формула (6.2) сохранится и при $n_1 = 0$ и/или $n_2 = 0$.

Обозначим через $R_{N-n}^B(X_1, n_1, X_2, n_2)$ байесовский риск на последних $(N - n)$ шагах, вычисленный относительно апостериорной плотности распределения (6.2). В дальнейшем будем для сокращения записи обозначать через $\_\ell$ предысторию X_ℓ, n_ℓ, если она одинакова во всем выражении. Для нахождения байесовского риска (6.1) следует решать стандартное уравнение динамического программирования

$$R_{N-n}^B(X_1, n_1, X_2, n_2) = \min\bigl(R_{N-n}^{B(1)}(X_1, n_1, X_2, n_2), R_{N-n}^{B(2)}(X_1, n_1, X_2, n_2)\bigr), \tag{6.4}$$

где $R^{B(1)}_{N-n}(X_1,n_1,X_2,n_2)=R^{B(2)}_{N-n}(X_1,n_1,X_2,n_2)=0$ при $n=N$, и далее

$$
\begin{aligned}
&R^{B(1)}_{N-n}(X_1,n_1,\__2)=\iint\limits_{\Theta}\lambda(p_1,p_2\,|\,X_1,n_1,\__2)\times\\
&\times\Big((p_2-p_1)^+ + \mathbf{E}^{(1)}_x\,R^{B}_{N-n-1}(X_1+x,n_1+1,\__2)\Big)\,dp_1\,dp_2,\\
&R^{B(2)}_{N-n}(\__1,X_2,n_2)=\iint\limits_{\Theta}\lambda(p_1,p_2\,|\,\__1,X_2,n_2)\times\\
&\times\Big((p_1-p_2)^+ + \mathbf{E}^{(2)}_x\,R^{B}_{N-n-1}(\__1,X_2+x,n_2+1)\Big)\,dp_1\,dp_2
\end{aligned}
\tag{6.5}
$$

при $2n_0\leqslant n<N$. Здесь $\mathbf{E}^{(\ell)}_x\,R(x)=q_\ell R(0)+p_\ell R(1)$, $q_\ell=1-p_\ell$.

В уравнениях (6.4), (6.5) риск $R^{B(\ell)}_{N-n}(\cdot)$ равен математическому ожиданию полных потерь на оставшемся горизонте управления длины $N-n$, если сначала было применено ℓ-е действие, а затем управление осуществлялось оптимально ($\ell=1,2$). Байесовская стратегия предписывает выбирать действие, соответствующее меньшей текущей величине $R^{B(\ell)}_{N-n}(\cdot)$, $\ell=1,2$; в случае их равенства выбор может быть произвольным. Байесовский риск (6.1) равен

$$
\begin{aligned}
&R^{B}_{N}(\lambda)=n_0\iint\limits_{\Theta}|p_1-p_2|\lambda(p_1,p_2)\,dp_1\,dp_2+\\
&+\sum_{X_1=0}^{n_0}\sum_{X_2=0}^{n_0}R^{B}_{N-2n_0}(X_1,n_0,X_2,n_0)P(X_1,n_0,X_2,n_0).
\end{aligned}
\tag{6.6}
$$

Обозначим

$$
\begin{aligned}
&\widetilde{R}(X_1,n_1,X_2,n_2)=R^{B}_{N-n}(X_1,n_1,X_2,n_2)P(X_1,n_1,X_2,n_2),\\
&\widetilde{R}^{(\ell)}(X_1,n_1,X_2,n_2)=R^{B(\ell)}_{N-n}(X_1,n_1,X_2,n_2)P(X_1,n_1,X_2,n_2),\quad \ell=1,2,
\end{aligned}
$$

где $P(X_1,n_1,X_2,n_2)$ определена в (6.3). Справедлива

Т е о р е м а 7. *Рассмотрим рекуррентное уравнение*

$$
\widetilde{R}(X_1,n_1,X_2,n_2)=\min\big(\widetilde{R}^{(1)}(X_1,n_1,X_2,n_2),\widetilde{R}^{(2)}(X_1,n_1,X_2,n_2)\big),
\tag{6.7}
$$

где $\widetilde{R}^{(1)}(X_1,n_1,X_2,n_2)=\widetilde{R}^{(2)}(X_1,n_1,X_2,n_2)=0$ *при* $n=N$, *и далее*

$$
\begin{aligned}
&\widetilde{R}^{(1)}(X_1,n_1,\__2)=\widetilde{g}^{(1)}(X_1,n_1,\__2)+\sum_{x=0}^{1}\widetilde{R}(X_1+x,n_1+1,\__2)h(X_1,n_1,x),\\
&\widetilde{R}^{(2)}(\__1,X_2,n_2)=\widetilde{g}^{(2)}(\__1,X_2,n_2)+\sum_{x=0}^{1}\widetilde{R}(\__1,X_2+x,n_2+1)h(X_2,n_2,x)
\end{aligned}
\tag{6.8}
$$

при $2n_0\leqslant n<N$. *Здесь*

$$
\begin{aligned}
&\widetilde{g}^{(1)}(X_1,n_1,X_2,n_2)=\iint\limits_{\Theta}(p_2-p_1)^+B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\lambda(p_1,p_2)\,dp_1\,dp_2,\\
&\widetilde{g}^{(2)}(X_1,n_1,X_2,n_2)=\iint\limits_{\Theta}(p_1-p_2)^+B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\lambda(p_1,p_2)\,dp_1\,dp_2
\end{aligned}
\tag{6.9}
$$

и

$$h(X,n,0)=\frac{n+1-X}{n+1},\quad h(X,n,1)=\frac{X+1}{n+1}. \tag{6.10}$$

Байесовская стратегия предписывает выбирать действие, которое соответствует меньшей текущей величине $\widetilde{R}^{(\ell)}(\cdot)$, $\ell=1,2$; *в случае их равенства выбор может быть произвольным. Байесовский риск* (6.1) *равен*

$$R_N^B(\lambda)=n_0\iint\limits_{\Theta}|p_1-p_2|\lambda(p_1,p_2)\,dp_1\,dp_2+\sum_{X_1=0}^{n_0}\sum_{X_2=0}^{n_0}\widetilde{R}(X_1,n_0,X_2,n_0). \tag{6.11}$$

Доказательство. Формулы (6.7)–(6.9), (6.11) следуют из (6.4)–(6.6). В проверке нуждается (6.10). Достаточно рассмотреть $h(X_1,n_1,x)$. Действительно,

$$\begin{aligned}h(X_1,n_1,x)&=\frac{\iint\limits_{\Theta}p_1^xq_1^{1-x}B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\,\lambda(p_1,p_2)\,dp_1\,dp_2}{P(X_1+x,n_1+1,X_2,n_2)}=\\&=\frac{\iint\limits_{\Theta}p_1^xq_1^{1-x}B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\,\lambda(p_1,p_2)\,dp_1\,dp_2}{\iint\limits_{\Theta}B(X_1+x,n_1+1\,|\,p_1)B(X_2,n_2\,|\,p_2)\,\lambda(p_1,p_2)\,dp_1\,dp_2}=\\&=\frac{p_1^xq_1^{1-x}B(X_1,n_1\,|\,p_1)}{B(X_1+x,n_1+1\,|\,p_1)}=\frac{\binom{n_1}{X_1}}{\binom{n_1+1}{X_1+x}}.\end{aligned}$$

Непосредственно проверяется, что это соответствует выражению в (6.10). ▲

Выберем некоторое $0<p<1$ и рассмотрим следующую замену переменных: $p_1=p+(\mu+w)N^{-1/2}$, $p_2=p+(\mu-w)N^{-1/2}$, $t=nN^{-1}$, $X_\ell=n_\ell p+x_\ell N^{1/2}$, $t_\ell=n_\ell N^{-1}$, $\ell=1,2$. В качестве множества параметров выберем следующее: $\Theta=$ $=\{(p+(\mu+w)N^{-1/2},p+(\mu-w)N^{-1/2}):|w|\leqslant c,\,|\mu|\leqslant a_N\}$, где $c>0$ – достаточно большое фиксированное число, $a_N=N^\alpha$, $0<\alpha<1/2$, и N достаточно велико. Геометрически Θ – это узкая полоса с центром в точке (p,p), параллельная главной диагонали квадрата $\{(p_1,p_2):\ p_\ell\in[0,1],\ \ell=1,2\}$. Отношение длин меньшей и большей сторон этой полосы стремится к нулю с ростом N, так же как и сами длины обеих сторон.

Далее удобно изменить параметризацию и от (p_1,p_2) перейти к (μ,w). Априорную плотность распределения выберем в виде $N\varkappa_a(\mu)\varrho(w)$, где $\varkappa_a(\mu)$ – плотность равномерного распределения при $|\mu|\leqslant a_N$, а $\varrho(w)$ – произвольная плотность при $|w|\leqslant c$. Отметим, что близкая к данной априорная плотность распределения выбрана для оценки минимаксного риска снизу в [13].

Положим $D_1=D_2=pq$, где $q=1-p$, $n_\ell^*=D_\ell n_\ell$, $n'_\ell=n_\ell/D_\ell$, $t_\ell^*=n_\ell^*N^{-1}$, $t'_\ell=n'_\ell N^{-1}$, $t'=t'_1+t'_2$, $\varepsilon_0=n_0N^{-1}$, $\varepsilon=N^{-1}$, $\delta=N^{-1/2}$ и $\widetilde{R}(X_1,n_1,X_2,n_2)=$ $=N^{-1/2}\widetilde{r}(x_1,t_1,x_2,t_2)$. Будем писать $x_N\sim y_N$, если $\lim\limits_{N\to\infty}x_N/y_N=1$, и $x_N\lesssim y_N$, если $\lim\limits_{N\to\infty}x_N/y_N\leqslant 1$. Если n_0 достаточно велико, то при $n_\ell>n_0$ справедливы оценки

$$\begin{aligned}&B(X_1,n_1\,|\,p_1)\sim N^{-1/2}f_{t_1^*}(x_1\,|\,(\mu+w)t_1),\\&B(X_2,n_2\,|\,p_2)\sim N^{-1/2}f_{t_2^*}(x_2\,|\,(\mu-w)t_2).\end{aligned} \tag{6.12}$$

Действительно, в силу локальной предельной теоремы имеем $B(X_\ell, n_\ell \,|\, p_\ell) \sim$ $\sim f_{n_\ell^*}(X_\ell \,|\, n_\ell p_\ell) \sim f_{n_\ell^*}(X_\ell - n_\ell p \,|\, n_\ell p_\ell - n_\ell p)$, откуда с учетом сделанной замены переменных следует (6.12).

Пусть последовательность $\{b_N\}$ такова, что $b_N \to +\infty$, $b_N/a_N \to 0$ при $N \to \infty$, и пусть $y = (\overline{x}_1 t_1' + \overline{x}_2 t_2')/t'$, $z = x_1 t_2 - x_2 t_1$, где $\overline{x}_\ell = x_\ell / t_\ell$. Справедливы оценки

$$\begin{aligned}
&\widetilde{g}^{(\ell)}(X_1, n_1, X_2, n_2) \sim N^{-3/2}(2a_N)^{-1}\widetilde{g}^{(\ell)}(z, t_1, t_2) \quad \text{при } |y| \leqslant a_N - b_N,\\
&\widetilde{g}^{(\ell)}(X_1, n_1, X_2, n_2) \lesssim N^{-3/2}(2a_N)^{-1}\widetilde{g}^{(\ell)}(z, t_1, t_2)\\
&\text{при } a_N - b_N < |y| \leqslant a_N + b_N,\\
&\widetilde{g}^{(\ell)}(X_1, n_1, X_2, n_2) = N^{-3/2}(2a_N)^{-1} o(e^{-\gamma(|y|-a_N)^2})\\
&\text{при } |y| > a_N + b_N,\ \gamma > 0,
\end{aligned} \tag{6.13}$$

где

$$\begin{aligned}
\widetilde{g}^{(1)}(z, t_1, t_2) &= \int\limits_{-c}^{0} 2|w| f_{t_1^* t_2^* t'}(z - 2wt_1t_2)\varrho(w)\,dw,\\
\widetilde{g}^{(2)}(z, t_1, t_2) &= \int\limits_{0}^{c} 2|w| f_{t_1^* t_2^* t'}(z - 2wt_1t_2)\varrho(w)\,dw.
\end{aligned} \tag{6.14}$$

Достаточно проверить (6.13), (6.14) при $\ell = 1$. Из (6.9), (6.12) следует, что

$$\begin{aligned}
&\widetilde{g}^{(1)}(X_1, n_1, X_2, n_2) \sim N^{-3/2}(2a_N)^{-1} \times\\
&\times \int\limits_{-c}^{0} \left(\int\limits_{-a_N}^{a_N} f_{t_1^*}(x_1 \,|\, (\mu + w)t_1) f_{t_2^*}(x_2 \,|\, (\mu - w)t_2)\,d\mu \right) 2|w|\varrho(w)\,dw.
\end{aligned}$$

Отсюда (6.13), (6.14) легко следуют, если для оценки внутреннего интеграла использовать непосредственно проверяемое равенство

$$\begin{aligned}
&f_{t_1^*}(x_1 \,|\, (\mu + w)t_1) f_{t_2^s}(x_2 \,|\, (\mu - w)t_2) =\\
&= f_{(t')^{-1}}\left(y + (t')^{-1}(t_2' - t_1')w - \mu\right) f_{t_1^* t_2^* t'}\left(z - 2wt_1t_2\right).
\end{aligned}$$

Аналогично с учетом (6.3), (6.12) выполняется аппроксимация

$$\begin{aligned}
&P(X_1, n_1, X_2, n_2) \sim\\
&\sim N^{-1}(2a_N)^{-1} \int\limits_{-c}^{c} \left(\int\limits_{-a_N}^{a_N} f_{t_1^*}(x_1 \,|\, (\mu + w)t_1) f_{t_2^*}(x_2 \,|\, (\mu - w)t_2)\,d\mu \right) \varrho(w)\,dw.
\end{aligned}$$

Так как $R^B_{N-n}(\cdot) \leqslant 2cN^{1/2}$, то с учетом определения $\widetilde{r}(x_1, t_1, x_2, t_2)$ отсюда можно показать, что

$$\begin{aligned}
&\widetilde{r}(x_1, t_1, x_2, t_2) = (2a_N)^{-1}O(1) \quad \text{при } |y| \leqslant a_N + b_N,\\
&\widetilde{r}(x_1, t_1, x_2, t_2) = (2a_N)^{-1}o(e^{-\gamma(|y|-a_N)^2}) \quad \text{при } |y| > a_N + b_N,\ \gamma > 0.
\end{aligned} \tag{6.15}$$

Покажем, что при $|y| \leqslant a_N - b_N$ и $N \to \infty$ из (6.7)–(6.11) следует дифференциальное уравнение в частных производных второго порядка

$$\min_{\ell=1,2}\left(\widetilde{r}_{t_\ell}' + t_\ell^{-1}\widetilde{r} + zt_\ell^{-1}\widetilde{r}_z' + 0{,}5D_\ell t_\ell^2 \widetilde{r}_{zz}'' + \widetilde{g}^{(\ell)}(z, t_1, t_2)\right) = 0 \tag{6.16}$$

с начальным условием

$$\widetilde{r}(z,t_1,t_2)=0 \quad \text{при } t_1+t_2=1. \tag{6.17}$$

При этом байесовский риск равен

$$R_N^B(\lambda)\sim N^{1/2}\left(\varepsilon_0\int\limits_{-c}^{c}2|w|\varrho(w)\,dw+\int\limits_{-\infty}^{\infty}\widetilde{r}(z,\varepsilon_0,\varepsilon_0)\,dz\right). \tag{6.18}$$

Замечание 2. В нашем случае D_1 и D_2 одинаковы, однако для гауссовского двурукого бандита с различными дисперсиями D_1 и D_2 дифференциальное уравнение и байесовский риск записываются именно в виде (6.16)–(6.18), если за основу взять формулы (3.2)–(3.5) из [1].

Чтобы записать уравнения (6.8) с использованием переменных x_1,t_1,x_2,t_2, заметим, что паре переменных (X_ℓ,n_ℓ) соответствует пара (x_ℓ,t_ℓ) по определению, а парам $(X_\ell,n_\ell+1)$ и $(X_\ell+1,n_\ell+1)$ соответствуют пары $(x_\ell-p\delta,t_\ell+\varepsilon)$ и $(x_\ell+q\delta,t_\ell+\varepsilon)$. Например, для $(X_\ell,n_\ell+1)$ имеем $x_\ell\leftarrow(X_\ell-(n_\ell+1)p)N^{-1/2}=x_\ell-p\delta$, $t_\ell\leftarrow$ $\leftarrow(n_\ell+1)N^{-1}=t_\ell+\varepsilon$. Далее, с учетом (6.10)

$$h(X_\ell,n_\ell,0)=\frac{n_\ell+1-n_\ell p-x_\ell N^{1/2}}{n_\ell+1}=q+pt_\ell^{-1}\varepsilon-x_\ell t_\ell^{-1}\delta+o(\varepsilon),$$
$$h(X_\ell,n_\ell,1)=\frac{n_\ell p+x_\ell N^{1/2}+1}{n_\ell+1}=p+qt_\ell^{-1}\varepsilon+x_\ell t_\ell^{-1}\delta+o(\varepsilon).$$

При $\ell=1$ уравнение (6.8) примет вид

$$\widetilde{r}^{(1)}(x_1,t_1,x_2,t_2)=\varepsilon(2a_N)^{-1}\widetilde{g}^{(1)}(z,t_1,t_2)+\widetilde{r}(x_1-p\delta,t_1+\varepsilon,x_2,t_2)\times$$
$$\times(q+pt_1^{-1}\varepsilon-x_1t_1^{-1}\delta)+\widetilde{r}(x_1+q\delta,t_1+\varepsilon,x_2,n_2)(p+qt_1^{-1}\varepsilon+x_1t_1^{-1}\delta)+o(\varepsilon).$$

Предполагая, что $\widetilde{r}(x_1,\cdot\,)$ является достаточно гладкой функцией, и выполняя замены $\widetilde{r}(x_1-p\delta,\cdot\,)=\widetilde{r}(x_1,\cdot\,)-p\delta\widetilde{r}'_{x_1}(x_1,\cdot\,)+0{,}5p^2\varepsilon\widetilde{r}''_{x_1x_1}(x_1,\cdot\,)+o(\varepsilon)$ и $\widetilde{r}(x_1+q\delta,\cdot\,)=$ $=\widetilde{r}(x_1,\cdot\,)+q\delta\widetilde{r}'_{x_1}(x_1,\cdot\,)+0{,}5q^2\varepsilon\widetilde{r}''_{x_1x_1}(x_1,\cdot\,)+o(\varepsilon)$, получаем

$$\widetilde{r}^{(1)}(x_1,t_1,x_2,t_2)=\varepsilon(2a_N)^{-1}\widetilde{g}^{(1)}(z,t_1,t_2)+\widetilde{r}(x_1,t_1+\varepsilon,x_2,t_2)(1+\varepsilon t_1^{-1})+$$
$$+\varepsilon\widetilde{r}'_{x_1}(x_1,t_1+\varepsilon,x_2,t_2)x_1t_1^{-1}+\varepsilon 0{,}5D_1\widetilde{r}''_{x_1x_1}(x_1,t_1+\varepsilon,x_2,t_2)+o(\varepsilon), \tag{6.19}$$

Положим $\widetilde{r}^{(1)}(x_1,t_1,x_2,t_2)=(2a_N)^{-1}\widetilde{r}^{(1)}(z,t_1,t_2)$, $\widetilde{r}(x_1,t_1,x_2,t_2)=(2a_N)^{-1}\widetilde{r}(z,t_1,t_2)$. Заметим, что $\widetilde{r}(x_1,t_1+\varepsilon,x_2,t_2)=(2a_N)^{-1}\widetilde{r}(z-x_2\varepsilon,t_1+\varepsilon,t_2)=(2a_N)^{-1}(\widetilde{r}(z,t_1+\varepsilon,t_2)-$ $-x_2\varepsilon\widetilde{r}'_z(z,t_1+\varepsilon,t_2)+o(\varepsilon))$, $\widetilde{r}'_{x_1}(x_1,t_1+\varepsilon,x_2,t_2)=(2a_N)^{-1}t_2\widetilde{r}'_z(z-x_2\varepsilon,t_1+\varepsilon,t_2)$, $\widetilde{r}''_{x_1x_1}(x_1,t_1+\varepsilon,x_2,t_2)=(2a_N)^{-1}t_2^2\widetilde{r}''_{zz}(z-x_2\varepsilon,t_1+\varepsilon,t_2)$. Поэтому (6.19) принимает вид

$$\widetilde{r}^{(1)}(z,t_1,t_2)=\varepsilon\widetilde{g}^{(1)}(z,t_1,t_2)+\widetilde{r}(z,t_1+\varepsilon,t_2)(1+\varepsilon t_1^{-1})+$$
$$+\varepsilon\widetilde{r}'_z(z,t_1+\varepsilon,t_2)zt_1^{-1}+\varepsilon 0{,}5D_1t_2^2\widetilde{r}''_{zz}(z,t_1+\varepsilon,t_2)+o(\varepsilon). \tag{6.20}$$

Аналогично при $\ell=2$ имеем

$$\widetilde{r}^{(2)}(z,t_1,t_2)=\varepsilon\widetilde{g}^{(2)}(z,t_1,t_2)+\widetilde{r}(z,t_1,t_2+\varepsilon)(1+\varepsilon t_2^{-1})+$$
$$+\varepsilon\widetilde{r}'_z(z,t_1,t_2+\varepsilon)zt_2^{-1}+\varepsilon 0{,}5D_2t_1^2\widetilde{r}''_{zz}(z,t_1,t_2+\varepsilon)+o(\varepsilon). \tag{6.21}$$

Дополняя (6.20), (6.21) уравнением $\min\limits_{\ell=1,2}(\widetilde{r}(z,t_1,t_2)-\widetilde{r}^{(\ell)}(z,t_1,t_2))=0$, соответствующим уравнению (6.7), в пределе при $\varepsilon\to0$ получаем (6.16). Далее, с учетом выбран-

ных множества Θ и априорной плотности распределения байесовский риск (6.11) равен

$$R_N^B(\lambda) = N^{1/2}\left(\varepsilon_0 \int\limits_{-c}^{c} 2|w|\varrho(w)\,dw + \sum_{x_1=-\varepsilon_0 pN^{1/2}}^{\varepsilon_0 qN^{1/2}} \sum_{x_2=-\varepsilon_0 pN^{1/2}}^{\varepsilon_0 qN^{1/2}} \widetilde{r}(x_1,\varepsilon_0,x_2,\varepsilon_0)\delta^2\right),$$

где δ характеризует шаг по x_1, x_2, поэтому при $\varepsilon \to 0$

$$R_N^B(\lambda) \sim N^{1/2}\left(\varepsilon_0 \int\limits_{-c}^{c} 2|w|\varrho(w)\,dw + \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \widetilde{r}(x_1,\varepsilon_0,x_2,\varepsilon_0)\,dx_1\,dx_2\right). \tag{6.22}$$

Снова преобразуем переменные: $y = (t_1'\overline{x}_1 + t_2'\overline{x}_2)t'^{-1}$, $z = x_1t_2 - x_2t_1$, где $t_1 = t_2 = \varepsilon_0$. Якобиан преобразования равен $|\partial(x_1,x_2)/\partial(y,z)| = 1$. С учетом (6.15) двойной интеграл в (6.22) равен

$$(2a_N)^{-1}\int\limits_{-\infty}^{\infty}\left(\int\limits_{-(a_N-b_N)}^{a_N-b_N}\widetilde{r}(z,\varepsilon_0,\varepsilon_0)\,dy\right)dz \sim \left(\int\limits_{-\infty}^{\infty}\widetilde{r}(z,\varepsilon_0,\varepsilon_0)\,dz\right).$$

Отсюда и из (6.22) следует (6.18). Наконец, справедлива

Теорема 8. *Пусть $\widetilde{r}(z,t_1,t_2)$ удовлетворяет дифференциальному уравнению* (6.16) *с начальным условием* (6.17). *Положим $u = z/t'$, $\widetilde{r}(z,t_1,t_2) = r(z/t',t_1,t_2)f_{t_1^*t_2^*t'}(z)$. Тогда $r(u,t_1,t_2)$ удовлетворяет дифференциальному уравнению* (4.12) *с начальным условием* (4.13).

Доказательство. Покажем сначала, что справедливо представление вида $\widetilde{r}(z,t_1,t_2) = \mathrm{r}(z,t_1,t_2)f_{t_1^*t_2^*t'}(z)$, где функция $\mathrm{r}(z,t_1,t_2)$ удовлетворяет уравнению

$$\min_{\ell=1,2}\left(\mathrm{r}'_{t_\ell} + \frac{z}{D_\ell t'}\cdot\mathrm{r}'_z + \frac{D_\ell t_\ell^2}{2}\cdot\mathrm{r}''_{zz} + \mathrm{g}^{(\ell)}(z,t_1,t_2)\right) = 0, \tag{6.23}$$

где $\mathrm{g}^{(\ell)}(z,t_1,t_2) = \widetilde{g}^{(\ell)}(z,t_1,t_2)f^{-1}_{t_1^*t_2^*t'}(z)$, а $\{\widetilde{g}^{(\ell)}(z,t_1,t_2)\}$ определены в (6.14). Для сокращения выкладок опускаем ниже зависимость функций $f, \mathrm{r}, \widetilde{r}, \mathrm{g}, \widetilde{g}$ от z, t_1, t_2. При этом f'_{t_1}, f'_z, f''_{zz} означают соответствующие частные производные. Справедливы равенства

$$f'_{t_1} = \frac{t_1 + D_1t'}{2t_1^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)f, \quad f'_z = -\frac{z}{t_1^*t_2^*t'}f, \quad f''_{zz} = \frac{1}{t_1^*t_2^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)f.$$

Полагая $\widetilde{r} = \mathrm{r}f$ и подставляя в (6.16) при $\ell = 1$, получаем

$$\begin{aligned}
&(\mathrm{r}'_{t_1}f + \mathrm{r}f'_{t_1}) + \frac{\mathrm{r}f}{t_1} + \frac{z}{t_1}(\mathrm{r}'_zf + \mathrm{r}f'_z) + D_1\frac{t_2^2}{2}(\mathrm{r}''_{zz}f + 2\mathrm{r}'_zf'_z + \mathrm{r}f''_{zz}) + \mathrm{g}^{(1)}f = \\
&= \left(\mathrm{r}'_{t_1} + \frac{t_1 + D_1t'}{2t_1^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)\mathrm{r}\right)f + \frac{\mathrm{r}}{t_1}f + \frac{z}{t_1}\cdot\left(\mathrm{r}'_z - \frac{z}{t_1^*t_2^*t'}\cdot\mathrm{r}\right)f + \\
&+ D_1\frac{t_2^2}{2}\cdot\left(\mathrm{r}''_{zz} - \frac{2z}{t_1^*t_2^*t'}\mathrm{r}'_z + \frac{1}{t_1^*t_2^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)\mathrm{r}\right)f + \mathrm{g}^{(1)}f = \\
&= \left(\mathrm{r}'_{t_1} + \frac{z}{D_1t'}\cdot\mathrm{r}'_z + \frac{D_1t_2^2}{2}\cdot\mathrm{r}''_{zz} + \mathrm{g}^{(1)}\right)f.
\end{aligned}$$

Поэтому выражение, соответствующее $\ell = 1$ в (6.16), обращается в нуль тогда и только тогда, когда обращается в нуль выражение, соответствующее $\ell = 1$ в (6.23). Далее положим $r(u, t_1, t_2) = \mathrm{r}(ut', t_1, t_2)$. Тогда $\mathrm{g}^{(\ell)}(ut', t_1, t_2) = g^{(\ell)}(u, t_1, t_2)$, $r'_{t_1}(u, t_1, t_2) = \mathrm{r}'_z(ut', t_1, t_2)D_1^{-1}u + \mathrm{r}'_{t_1}(ut', t_1, t_2)$, $r'_u(u, t_1, t_2) = \mathrm{r}'_z(ut', t_1, t_2)t'$, $r''_{uu}(u, t_1, t_2) = \mathrm{r}''_{zz}(ut', t_1, t_2)(t')^2$. Поэтому

$$\mathrm{r}'_{t_1}(ut', t_1, t_2) + \frac{z}{D_1 t'} \cdot \mathrm{r}'_z(ut', t_1, t_2) + \frac{D_1 t_2^2}{2} \cdot \mathrm{r}''_{zz}(ut', t_1, t_2) + \mathrm{g}^{(\ell)}(ut', t_1, t_2) =$$
$$= r'_{t_1}(u, t_1, t_2) + \frac{D_1 t_2^2}{2(t')^2} \cdot r''_{uu}(u, t_1, t_2) + g^{(\ell)}(u, t_1, t_2).$$

Поэтому выражение, соответствующее $\ell = 1$ в (6.23), обращается в нуль тогда и только тогда, когда обращается в нуль выражение, соответствующее $\ell = 1$ в (4.12). Для выражения, соответствующего $\ell = 2$, проверка выполняется аналогично. Поэтому уравнение (6.16) с начальным условием (6.17) эквивалентно уравнению (4.12) с начальным условием (4.13). ▲

Наконец, подставляя $\widetilde{r}(z, t_1, t_2) = r(u, t_1, t_2) f_{t_1^* t_2^* t'}(ut')$, $z = ut'$ при $t_1 = t_2 = \varepsilon_0$ в (6.18) и предполагая, что $r(u, t_1, t_2)$ является непрерывной функцией u, t_1, t_2, получаем, что байесовский риск при $N \to \infty$ и $\varepsilon_0 \to +0$ определяется формулой (4.11).

§ 7. Заключение

Для гауссовского двурукого бандита, характеризующего пакетную обработку данных, получено асимптотическое описание байесовского риска относительно наихудшего априорного распределения, т.е. эти результаты описывают пакетную обработку больших данных. В силу основной теоремы теории игр это позволяет находить минимаксные стратегию и риск, которые определяются полным количеством данных N и дисперсиями одношаговых доходов D_1, D_2. Поскольку при пакетной обработке одинаковые действия применяются к группам данных, можно ожидать, что эти результаты останутся справедливыми для широкого класса процессов, у которых одношаговые доходы подчиняются центральной предельной теореме. При этом дисперсии доходов можно считать известными, так как их можно оценить на коротком начальном отрезке управления, а минимаксный риск мало меняется при малом изменении дисперсий. Наибольшие потери достигаются в области "близких" распределений, для которых математические ожидания доходов различаются на величину порядка $N^{-1/2}$. Кроме того, оказалось, что в случае бернуллиевского двурукого бандита максимальные потери при оптимальной обработке больших данных по одному не могут быть сделаны меньше потерь, соответствующих оптимальной пакетной обработке. Можно ожидать, что этот результат сохранится и для широкого класса процессов с другими распределениями одношаговых доходов.

Автор выражает глубокую признательность рецензенту за внимание к статье и полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Колногоров А.В.* Гауссовский двурукий бандит и оптимизация групповой обработки данных // Пробл. передачи информ. 2018. Т. 54. № 1. С. 93–111.
2. *Perchet V., Rigollet P., Chassang S., Snowberg E.* Batched Bandit Problems // Ann. Statist. 2016. V. 44. № 2. P. 660–681.
3. *Колногоров А.В.* Задача о двуруком бандите для систем с параллельной обработкой данных // Пробл. передачи информ. 2012. Т. 48. № 1. С. 83–95.
4. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей. М.: Наука, 1987.

5. *Vogel W.* A Sequential Design for the Two-Armed Bandit // Ann. Math. Statist. 1960. V. 31. № 2. P. 430–443.
6. *Vogel W.* An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem // Ann. Math. Statist. 1960. V. 31. № 2. P. 444–451.
7. *Lai T.L., Levin B., Robbins H., Siegmund D.* Sequential Medical Trials // Proc. Natl. Acad. Sci. USA. 1980. V. 77. № 6. P. 3135–3138.
8. *Lai T.L., Robbins H.* Asymptotically Efficient Adaptive Allocation Rules // Adv. in Appl. Math. 1985. V. 6. № 1. P. 4–22.
9. *Kaufmann E.* On Bayesian Index Policies for Sequential Resource Allocation // Ann. Statist. 2018. V. 46. № 2. P. 842–865.
10. *Колногоров А.В.* Робастное параллельное управление в случайной среде (задача о двуруком бандите) // АиТ. 2012. № 4. С. 114–130.
11. *Колногоров А.В.* К предельному описанию робастного параллельного управления в случайной среде // АиТ. 2015. № 7. С. 111–126.
12. *Самарский А.А.* Теория разностных схем. М.: Наука, 1989.
13. *Bather J.A.* The Minimax Risk for the Two-Armed Bandit Problem // Mathematical Learning Models — Theory and Algorithms. Lect. Notes Statist. V. 20. New York: Springer-Verlag, 1983. P. 1–11.

Колногоров Александр Валерианович
Новгородский государственный университет
им. Ярослава Мудрого, кафедра
прикладной математики и информатики
kolnogorov53@mail.ru

Поступила в редакцию
06.04.2020
После доработки
02.06.2020
Принята к публикации
02.06.2020