Math-Net.Ru

Общероссийский математический портал

В. И. Бахтин, Спектральный потенциал, действие Кульбака и большие отклонения эмпирических мер на измеримых пространствах, *Теория вероятн. и ее примен.*, 2014, том 59, выпуск 4, 625–638

DOI: <https://doi.org/10.4213/tvp4589>

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: 52.87.193.239

26 апреля 2023 г., 05:42:53

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И ЕЕ ПРИМЕНЕНИЯ

Основан в январе 1956 г.

Выходит 4 раза в год

Москва • Наука

Том 59

Выпуск 4, 2014 год

октябрь, ноябрь, декабрь

© 2014 г.

БАХТИН В. И.*

СПЕКТРАЛЬНЫЙ ПОТЕНЦИАЛ, ДЕЙСТВИЕ КУЛЬБАКА И
БОЛЬШИЕ УКЛОНЕНИЯ ЭМПИРИЧЕСКИХ МЕР
НА ИЗМЕРИМЫХ ПРОСТРАНСТВАХ¹⁾

В статье известный принцип больших уклонений для эмпирических мер обосновывается в случае, когда независимые одинаково распределенные случайные величины принимают значения не в топологическом пространстве, а лишь в измеримом. При этом область определения информационной функции Кульбака–Лейблера расширяется таким образом, что теорема Макмиллана–Бреймана становится частным случаем принципа больших уклонений.

Ключевые слова и фразы: принцип больших уклонений, эмпирическая мера, спектральный потенциал, функция уклонений, расстояние Кульбака–Лейблера, действие Кульбака.

Из закона больших чисел вытекает, что эмпирические меры, порожденные независимыми случайными величинами с общим распределением μ , по вероятности сходятся к μ (в слабой топологии). Более того, известно, что вероятность попадания эмпирической меры в малую окрестность $O(\nu)$ любого другого распределения ν , отличного от μ , стремится к нулю с экспоненциальной скоростью. Грубо говоря, принцип больших уклонений состоит в существовании такой неотрицательной функции $\rho(\nu, \mu)$, что эта вероятность в первом приближении эквивалентна $e^{-n\rho(\nu, \mu)}$, где n — количество случайных величин, при помощи которых строится эмпирическая мера.

В настоящее время существует довольно обширная литература о принципе больших уклонений. В ней разные авторы называют функцию

*Белорусский государственный университет, Минск, Беларусь; John Paul II Catholic University of Lublin, Poland; e-mail: bakhtin@tut.by

¹⁾ Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (проект № Ф12Р-061).

$\rho(\nu, \mu)$ разными именами: относительной энтропией [2], функцией уклонений [4], [5], [7], функционалом действия [8], информационной функцией Кульбака–Лейблера [3], расстоянием Кульбака–Лейблера [6] (хотя, конечно, она не является расстоянием в общепринятом смысле, поскольку не симметрична). По-видимому, впервые явная формула для функции $\rho(\nu, \mu)$ была получена И. Н. Сановым [9].

В абсолютном большинстве работ по принципу больших уклонений рассматриваются случайные величины со значениями в польском (т.е. полном сепарабельном метрическом) пространстве и в единичных случаях рассматриваются случайные величины со значениями в топологическом пространстве (см., например, [3]) или таком измеримом пространстве, в котором σ -алгебра порождается открытыми шарами и не обязательно является борелевской (см. [7, раздел 7]). Данная работа отличается от других тем, что в ней изначально вообще не требуется никакой топологии, а на множестве мер рассматривается слабая топология, порожденная ограниченными измеримыми функциями. Другое (менее существенное) отличие состоит в том, что вместо вероятностных мер ν рассматриваются произвольные линейные функционалы на пространстве ограниченных измеримых функций и не требуется, чтобы мера μ была нормирована. Функцию уклонений с такой расширенной областью определения мы называем действием Кульбака. Это расширение, наряду с последовательным использованием двойственности по Лежандру, позволяет существенно упростить доказательство принципа больших уклонений и отказаться от топологии на множестве значений случайных величин. Кроме того, оно позволяет представить обычную информационную энтропию как частный случай действия Кульбака, а классическую теорему Макмиллана — как частный случай принципа больших уклонений.

Важнейшую роль в обосновании принципа больших уклонений играет функционал, двойственный по Лежандру к действию Кульбака. Станным образом он не получил никакого собственного имени в литературе. Мы будем называть его в статье спектральным потенциалом (потому что его прямое обобщение совпадает с логарифмом спектрального радиуса оператора взвешенного сдвига, см., например, [1]). Начнем с изучения свойств спектрального потенциала.

1. Спектральный потенциал. Будем рассматривать произвольное множество X , на котором фиксирована некоторая σ -алгебра $\mathfrak{A}$ его подмножеств. Введем следующие обозначения:

$M_1(X)$ — совокупность всех вероятностных мер на пространстве $(X, \mathfrak{A})$;

$M(X)$ — множество всех ненулевых (счетно-аддитивных, неотрицательных и конечных) мер на $(X, \mathfrak{A})$;

$B(X)$ — пространство всех вещественнозначных ограниченных измеримых функций на $(X, \mathfrak{A})$, снабженное равномерной нормой;

$B^*(X)$ — пространство непрерывных линейных функционалов на $B(X)$;

$M'(X)$ — совокупность положительных линейных функционалов на $B(X)$ (принимаящих неотрицательные значения на неотрицательных функциях).

Каждая мера $\mu \in M(X)$ порождает линейный функционал на $B(X)$ вида

$$\mu[f] = \int_X f d\mu. \quad (1)$$

Очевидно, что этот функционал положителен, а в случае вероятностной меры μ нормирован (принимает единичное значение на единичной функции). Таким образом, меры $\mu \in M(X)$ отождествляются с элементами пространства $B^*(X)$. С другой стороны, элементы $\nu \in M'(X)$ можно рассматривать как конечно-аддитивные меры на $(X, \mathfrak{A})$ (при этом мера множества $A \in \mathfrak{A}$ определяется равенством $\nu(A) = \nu[\chi_A]$, где χ_A — характеристическая функция). С учетом этих отождествлений имеют место вложения

$$M_1(X) \subset M(X) \subset M'(X) \subset B^*(X).$$

С целью сокращения записи условимся использовать ниже обозначение (1) не только для функций $f \in B(X)$, но и для всех функций $f \in L^1(X, \mu)$.

Зафиксируем какую-нибудь меру $\mu \in M(X)$. Определим для нее множество измеримых функций

$$D(\mu) = \{\varphi: (X, \mathfrak{A}) \rightarrow \mathbf{R} \mid \mu[e^\varphi] < +\infty\}.$$

Очевидно, что если к функции $\varphi \in D(\mu)$ добавить любую функцию $\psi \in B(X)$, то их сумма тоже будет лежать в $D(\mu)$. В силу неравенства Гёльдера для всяких $\varphi, \psi \in D(\mu)$ выполняется соотношение

$$\mu[e^{(1-t)\varphi+t\psi}] \leq (\mu[e^\varphi])^{1-t}(\mu[e^\psi])^t, \quad t \in [0, 1]. \quad (2)$$

Из него следует, что множество $D(\mu)$ выпукло.

Спектральным потенциалом мы будем называть нелинейный функционал

$$\lambda(\varphi, \mu) = \ln \mu[e^\varphi], \quad \mu \in M(X), \quad \varphi: (X, \mathfrak{A}) \rightarrow \mathbf{R}. \quad (3)$$

Он обладает рядом замечательных свойств. Рассмотрим их по порядку.

Во-первых, спектральный потенциал конечен, если $\varphi \in D(\mu)$, и равен $+\infty$ в противном случае. Во-вторых, он монотонен: если $\varphi \geq \psi$,

то $\lambda(\varphi, \mu) \geq \lambda(\psi, \mu)$. В-третьих, он аддитивно однороден: $\lambda(\varphi + t, \mu) = \lambda(\varphi, \mu) + t$ для любой константы t . В-четвертых, из неравенства (2) вытекает, что этот функционал является выпуклым по φ :

$$\ln \mu[e^{(1-t)\varphi + t\psi}] \leq (1-t) \ln \mu[e^\varphi] + t \ln \mu[e^\psi], \quad t \in [0, 1].$$

В-пятых, в силу вогнутости логарифма он вогнут по отношению к аргументу μ :

$$\ln \{(1-t)\mu[e^\varphi] + t\nu[e^\varphi]\} \geq (1-t) \ln \mu[e^\varphi] + t \ln \nu[e^\varphi], \quad \mu, \nu \in M(X).$$

И в-шестых, можно заметить (хотя в дальнейшем это нам не понадобится), что при всех $\varphi \in D(\mu)$ и $\psi \in B(X)$ выражение $\lambda(\varphi + \psi, \mu)$ аналитически зависит от аргумента ψ .

Для всякой функции $\varphi \in D(\mu)$ определим вероятностную меру $\mu_\varphi \in M_1(X)$ с помощью формулы

$$\mu_\varphi[f] = \frac{\mu[e^\varphi f]}{\mu[e^\varphi]}, \quad f \in B(X).$$

Эта мера эквивалентна μ и имеет по отношению к μ плотность $e^{\varphi - \lambda(\varphi, \mu)}$. С другой стороны, из теоремы Радона–Никодима вытекает, что любая вероятностная мера ν , эквивалентная μ , может быть представлена в виде $\nu = \mu_\varphi$, где $\varphi = \ln\{d\nu/d\mu\}$. (Две меры называются эквивалентными, если они абсолютно непрерывны друг относительно друга.)

Вычислим первые две производные от спектрального потенциала по отношению к аргументу φ . Предположим, что измеримые функции φ, f удовлетворяют условию $\varphi + tf \in D(\mu)$ при всех достаточно малых вещественных t . В таком случае введем обозначение

$$\lambda'(\varphi, \mu)[f] = \left. \frac{d\lambda(\varphi + tf, \mu)}{dt} \right|_{t=0}.$$

Это есть не что иное, как производная от спектрального потенциала по направлению f в точке φ . Элементарные выкладки показывают, что эта производная существует, конечна и имеет вид

$$\lambda'(\varphi, \mu)[f] = \left. \frac{d \ln \mu[e^{\varphi + tf}]}{dt} \right|_{t=0} = \frac{\mu[e^\varphi f]}{\mu[e^\varphi]} = \mu_\varphi[f]. \quad (4)$$

В частности, формула (4) верна для всех функций $\varphi \in D(\mu)$ и $f \in B(X)$.

Пусть теперь измеримые функции φ, f, g удовлетворяют условию $\varphi + tf + sg \in D(\mu)$ при всех достаточно малых t, s . В силу выпуклости множества $D(\mu)$ это условие равносильно тому, что по отдельности $\varphi + tf \in D(\mu)$ и $\varphi + tg \in D(\mu)$ при достаточно малых t . Положим

$$\lambda''(\varphi, \mu)[f, g] = \left. \frac{\partial^2 \lambda(\varphi + tf + sg, \mu)}{\partial s \partial t} \right|_{s, t=0}.$$

Вычислим последнюю производную в явном виде с помощью уже имеющейся формулы (4):

$$\begin{aligned}\lambda''(\varphi, \mu)[f, g] &= \frac{\partial}{\partial s} \left(\frac{\mu[e^{\varphi+sg} f]}{\mu[e^{\varphi+sg}]} \right) \Big|_{s=0} = \frac{\mu[e^{\varphi} f g]}{\mu[e^{\varphi}]} - \frac{\mu[e^{\varphi} f] \mu[e^{\varphi} g]}{(\mu[e^{\varphi}])^2} \\ &= \mu_{\varphi}[fg] - \mu_{\varphi}[f] \mu_{\varphi}[g].\end{aligned}\quad (5)$$

В теории вероятностей выражение $\mu_{\varphi}[f]$ называют *математическим ожиданием* случайной величины f по отношению к вероятностной мере μ_{φ} , а выражение $\mu_{\varphi}[fg] - \mu_{\varphi}[f] \mu_{\varphi}[g]$ называют *ковариацией* случайных величин f и g . В частности, значение квадратичной формы

$$\lambda''(\varphi, \mu)[f, f] = \frac{d^2 \lambda(\varphi + tf, \mu)}{dt^2} \Big|_{t=0} = \mu_{\varphi}[f^2] - (\mu_{\varphi}[f])^2 \quad (6)$$

называется дисперсией случайной величины f (по отношению к распределению вероятностей μ_{φ}).

Завершая этот пункт, отметим, что при выводе формул (4) и (5) мы никак не обосновывали законность дифференцирования под знаком интеграла. Это несложно сделать, применяя теорему Лебега о мажорируемой сходимости к семействам $\mu_{\varphi+tf}[f]$ и $\mu_{\varphi+tf+sg}[fg]$.

2. Действие Кульбака. Всюду ниже термин «мера» будет использоваться для обозначения только *неотрицательных* (конечно- или счетно-аддитивных) мер.

Будем говорить, что (конечно-аддитивная) мера ν абсолютно непрерывна по отношению к (конечно-аддитивной) мере μ , если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что из условия $\mu(A) < \delta$ следует $\nu(A) < \varepsilon$. Нетрудно видеть, что в случае счетно-аддитивных мер это определение равносильно общепринятому (т.е. импликации $\mu(A) = 0 \Rightarrow \nu(A) = 0$). Отметим еще, что всякая конечно-аддитивная мера ν , абсолютно непрерывная по отношению к некоторой счетно-аддитивной мере μ , автоматически будет счетно-аддитивной. Действительно, счетная аддитивность меры равносильна ее непрерывности, а непрерывность меры μ влечет непрерывность ν (непрерывность меры μ означает, что из условий $A_1 \supset A_2 \supset \dots$ и $\bigcap A_n = \emptyset$ вытекает $\mu(A_n) \rightarrow 0$).

Теорема Радона–Никодима утверждает, что если мера $\nu \in M(X)$ абсолютно непрерывна по отношению к мере $\mu \in M(X)$, то она представляется в виде $\nu = \varphi \mu$, где φ — некоторая неотрицательная измеримая функция (плотность меры ν по отношению к μ). Эта функция определяется однозначно с точностью до множества нулевой меры μ .

Действием Кульбака мы будем называть функционал от двух аргументов $\nu \in B^*(X)$ и $\mu \in M(X)$ вида

$$\rho(\nu, \mu) = \begin{cases} \mu[\varphi \ln \varphi] = \nu[\ln \varphi], & \text{если } \nu \in M_1(X) \text{ и } \nu = \varphi\mu; \\ +\infty & \text{во всех остальных случаях.} \end{cases} \quad (7)$$

Выражаясь более конкретно, «все остальные случаи» попадают как минимум в одну из следующих трех категорий: не абсолютно непрерывные относительно μ конечно-аддитивные (в том числе и счетно-аддитивные) положительные нормированные меры ν ; ненормированные функционалы ν ; неположительные функционалы ν .

В известной автору литературе этот функционал определяют лишь для вероятностных мер ν и μ при помощи формулы

$$\rho(\nu, \mu) = \int_X \ln \frac{d\nu}{d\mu} d\nu, \quad (8)$$

которая является более привычной формой записи первого случая формулы (7). Однако полезно рассматривать действие Кульбака не только для вероятностных мер. Например, если множество X конечно и $\mu(x) = 1$ для всех $x \in X$, то оно с точностью до знака совпадает с энтропией Шеннона:

$$H(\nu) = - \sum_{x \in X} \nu(x) \ln \nu(x) = -\rho(\nu, \mu). \quad (9)$$

Приведем еще один аргумент в пользу расширения области определения действия Кульбака. Во всех работах по принципу больших отклонений доказывается (и затем используется) утверждение, что для любой положительной константы c множество $\{\nu \in M_1(X) \mid \rho(\nu, \mu) \leq c\}$ слабо компактно. Обычно этот факт доказывают, проверяя, что указанное семейство мер обладает свойством плотности. Оказывается, если определить функционал $\rho(\cdot, \mu)$ не только на $M_1(X)$, а на всем пространстве $B^*(X)$, это доказательство становится совсем элементарным (см. следствие 2.2).

В случае вероятностной меры μ действие Кульбака неотрицательно и обращается в нуль только при $\nu = \mu$. Действительно, если функционал ν не является абсолютно непрерывной по отношению к μ вероятностной мерой, то тогда $\rho(\nu, \mu) = +\infty$. А если ν — вероятностная мера вида $\nu = \varphi\mu$, то из неравенства Йенсена и выпуклости функции $f(x) = x \ln x$ вытекает, что

$$\rho(\nu, \mu) = \mu[f(\varphi)] \geq f(\mu[\varphi]) = 0$$

(поскольку $\mu[\varphi] = \nu[1] = 1$), причем равенство $\rho(\nu, \mu) = 0$ возможно лишь тогда, когда функция φ почти всюду постоянна и, следовательно, ν совпадает с μ .

Произвольная мера $\mu \in M(X)$ может быть записана в виде $\mu = c\mu_1$, где $c = \mu[1]$ и $\mu_1 \in M_1(X)$. В случае $\nu \in M_1(X)$ из определения действия Кульбака следует, что

$$\rho(\nu, \mu) = \rho(\nu, \mu_1) - \ln c \geq -\ln \mu[1].$$

При $\nu \notin M_1(X)$ это неравенство и подавно верно, поскольку действие Кульбака бесконечно.

Важнейшее свойство действия Кульбака состоит в том, что оно двойственно по Лежандру спектральному потенциалу. Чтобы установить это свойство, мы вначале докажем *неравенство Юнга*

$$\rho(\nu, \mu) \geq \nu[\psi] - \lambda(\psi, \mu), \quad \nu, \mu \in M(X). \quad (10)$$

Теорема 2.1. *Неравенство Юнга верно для всех функций ψ , для которых определен интеграл $\nu[\psi] = \int_X \psi d\nu$ (т.е. при условии $\psi \in L^1(X, \nu)$). Оно обращается в равенство тогда и только тогда, когда $\nu = \mu_\psi$.*

Доказательство. Если $\rho(\nu, \mu) = +\infty$ или $\lambda(\psi, \mu) = +\infty$, то доказываемое неравенство тривиально. Если же обе эти величины конечны, то из определения действия Кульбака вытекает, что мера ν вероятностная и имеет вид $\nu = \varphi\mu$, где φ — неотрицательная измеримая функция. В этом случае

$$\begin{aligned} \lambda(\psi, \mu) &= \ln \mu[e^\psi] \geq \ln \int_{\varphi>0} e^\psi d\mu = \ln \int_{\varphi>0} e^{\psi - \ln \varphi} d\nu \\ &= \ln \nu[e^{\psi - \ln \varphi}] \geq \nu[\psi - \ln \varphi] \end{aligned}$$

(на последнем шаге мы использовали неравенство Йенсена и вогнутость логарифма). А поскольку $\rho(\nu, \mu) = \nu[\ln \varphi]$, эта выкладка влечет за собой неравенство (10).

Напомним, что $\mu_\psi = e^{\psi - \lambda(\psi, \mu)}\mu$. И если $\nu = \mu_\psi$, то по определению

$$\rho(\nu, \mu) = \nu[\psi - \lambda(\psi, \mu)] = \nu[\psi] - \lambda(\psi, \mu).$$

Наоборот, предположим, что $\rho(\nu, \mu) = \nu[\psi] - \lambda(\psi, \mu)$. Тогда вычтем это равенство из неравенства Юнга $\rho(\nu, \mu) \geq \nu[\varphi] - \lambda(\varphi, \mu)$. У нас получится

$$\lambda(\varphi, \mu) - \lambda(\psi, \mu) \geq \nu[\varphi - \psi].$$

Значит, мера ν является субградиентом спектрального потенциала в точке ψ . С другой стороны, в силу (4) производная от спектрального потенциала по любому направлению $f \in B(X)$ в этой точке равна $\mu_\psi[f]$. Отсюда следует, что ν совпадает с μ_ψ . Теорема 2.1 доказана.

Теорема 2.2. Действие Кульбака $\rho(\nu, \mu)$ по отношению к аргументу ν является преобразованием Лежандра от спектрального потенциала:

$$\rho(\nu, \mu) = \sup_{\psi \in B(X)} \{\nu[\psi] - \lambda(\psi, \mu)\}, \quad \nu \in B^*(X), \quad \mu \in M(X). \quad (11)$$

Доказательство. Для всех функций $\psi \in B(X)$ и мер $\nu \in M(X)$ выполняется неравенство Юнга. А если $\nu \notin M(X)$, то оно выполняется и подавно, поскольку в этом случае по определению действия Кульбака $\rho(\nu, \mu) = +\infty$. Поэтому, чтобы доказать теорему, нам достаточно для каждого функционала $\nu \in B^*(X)$ предъявить такое семейство функций $\psi_t \in B(X)$, зависящее от вещественного параметра t , на котором достигается супремум в (11).

Вначале предположим, что ν — абсолютно непрерывная (относительно μ) вероятностная мера. Тогда она имеет вид $\nu = \varphi\mu$, где φ — неотрицательная плотность. Рассмотрим семейство функций

$$\psi_t = \begin{cases} \ln \varphi, & \text{если } e^{-t} \leq \varphi \leq e^t, \\ t, & \text{если } \varphi > e^t, \\ -t, & \text{если } \varphi < e^{-t}, \end{cases}$$

где $t > 0$. Для него при $t \rightarrow +\infty$ выполняются соотношения

$$\begin{aligned} \mu[e^{\psi_t}] &= \int_{e^{-t} \leq \varphi \leq e^t} \varphi d\mu + \int_{\varphi > e^t} e^t d\mu + \int_{\varphi < e^{-t}} e^{-t} d\mu \rightarrow \int_X \varphi d\mu = 1, \\ \nu[\psi_t] &= \int_{e^{-t} \leq \varphi \leq e^t} \varphi \ln \varphi d\mu + \int_{\varphi > e^t} t\varphi d\mu + \int_{\varphi < e^{-t}} -t\varphi d\mu \rightarrow \mu[\varphi \ln \varphi], \\ \nu[\psi_t] - \lambda(\psi_t, \mu) &= \nu[\psi_t] - \ln \mu[e^{\psi_t}] \rightarrow \mu[\varphi \ln \varphi] = \rho(\nu, \mu). \end{aligned}$$

Тем самым, равенство (11) доказано.

Во всех остальных случаях, когда ν не является абсолютно непрерывной вероятностной мерой, по определению $\rho(\nu, \mu) = +\infty$. Разберем эти случаи по порядку.

Если функционал ν положителен и нормирован, но при этом не абсолютно непрерывен по отношению к мере μ , то существует такое число $\varepsilon > 0$, что для любого $\delta > 0$ найдется множество $A_\delta \in \mathfrak{A}$, для которого $\mu(A_\delta) < \delta$ и одновременно $\nu(A_\delta) > \varepsilon$. В этом случае рассмотрим семейство функций

$$\psi_\delta(x) = \begin{cases} -\ln \delta, & \text{если } x \in A_\delta; \\ 0, & \text{если } x \notin A_\delta. \end{cases}$$

Нетрудно видеть, что для него

$$\nu[\psi_\delta] - \lambda(\psi_\delta, \mu) \geq -\varepsilon \ln \delta - \ln \mu[e^{\psi_\delta}] \geq -\varepsilon \ln \delta - \ln(1 + \mu[1]).$$

Последнее выражение стремится к $+\infty$ при уменьшении δ , и равенство (11) опять выполняется.

Если функционал ν не нормирован, то положим $\psi_t = t$. Тогда $\lambda(\psi_t, \mu) = \ln \mu[e^t] = t + \ln \mu[1]$. Поэтому выражение

$$\nu[\psi_t] - \lambda(\psi_t, \mu) = t(\nu[1] - 1) - \ln \mu[1]$$

не ограничено сверху, и равенство (11) снова верно.

Наконец, если функционал ν не положителен, то существует такая неотрицательная функция $\varphi \in B(X)$, что $\nu[\varphi] < 0$. Рассмотрим семейство $\psi_t = -t\varphi$, где $t > 0$. Для него

$$\nu[\psi_t] - \lambda(\psi_t, \mu) \geq -t\nu[\varphi] - \ln \mu[1] \rightarrow +\infty \quad \text{при } t \rightarrow +\infty,$$

и равенство (11) остается в силе. Теорема 2.2 доказана.

Теорема 2.3. *Если X — метрическое пространство с борелевской σ -алгеброй, то*

$$\rho(\nu, \mu) = \sup_{\psi \in C(X)} \{\nu[\psi] - \lambda(\psi, \mu)\}, \quad \nu \in B^*(X), \quad \mu \in M(X),$$

где $C(X)$ — пространство ограниченных непрерывных функций на X .

Доказательство. Достаточно во всех предыдущих рассуждениях заменить функцию ψ_t на такую непрерывную функцию ξ_t , для которой значение $\nu[\xi_t]$ аппроксимирует $\nu[\psi_t]$ и одновременно $\mu[e^{\xi_t}]$ аппроксимирует $\mu[e^{\psi_t}]$. Теорема 2.3 доказана.

Следствие 2.1. *Функционал $\rho(\cdot, \mu)$ выпукл и полунепрерывен снизу на пространстве $B^*(X)$.*

Здесь имеется в виду полунепрерывность по отношению к слабой топологии на $B^*(X)$, порожденной функциями из $B(X)$, а в случае метрического пространства X также и по отношению к слабой топологии, порожденной ограниченными непрерывными функциями.

Доказательство. Это свойства преобразования Лежандра.

Следствие 2.2. *Для любого числа c множество $\{\nu \mid \rho(\nu, \mu) \leq c\}$ слабо компактно.*

Доказательство. В силу полунепрерывности функционала $\rho(\cdot, \mu)$ это множество замкнуто в $B^*(X)$. Кроме того, оно состоит только из вероятностных мер (поскольку если ν не является вероятностной мерой, то $\rho(\nu, \mu) = +\infty$). Значит, оно целиком лежит в единичном шаре в $B^*(X)$, который слабо компактен (по теореме Алаоглу). Следствие 2.2 доказано.

Из теоремы 2.1 вытекает так называемый *вариационный принцип*

$$\lambda(\varphi, \mu) = \sup_{\nu \in M_1(X)} \{\nu[\varphi] - \rho(\nu, \mu)\}, \quad \varphi \in B(X), \quad \mu \in M(X). \quad (12)$$

Он означает, что при фиксированной мере μ сужение функционала $\lambda(\cdot, \mu)$ на пространство $B(X)$ является преобразованием Лежандра от действия Кульбака $\rho(\cdot, \mu)$.

Без доказательства отметим еще, что вариационный принцип (12) останется в силе, если в нем рассматривать произвольные измеримые функции φ , а супремум брать только по таким мерам ν , для которых определен интеграл $\nu[\varphi] = \int_X \varphi d\nu$. Однако этот супремум уже не обязательно будет максимумом.

3. Принцип больших уклонений. Пусть, как и раньше, на произвольном измеримом пространстве $(X, \mathfrak{A})$ определены множества всех положительных мер $M(X)$ и вероятностных мер $M_1(X)$, пространство ограниченных измеримых функций $B(X)$, а также сопряженное к нему пространство $B^*(X)$.

Каждой конечной последовательности $x = (x_1, \dots, x_n) \in X^n$ поставим в соответствие *эмпирическую меру* $\delta_{x,n} \in M_1(X)$, которая сосредоточена на множестве $\{x_1, \dots, x_n\}$ и приписывает каждой точке x_i меру $1/n$. Интеграл от любой функции f по эмпирической мере имеет вид

$$\delta_{x,n}[f] = \frac{f(x_1) + \dots + f(x_n)}{n}.$$

Фиксируем любую вероятностную меру $\mu \in M_1(X)$. Если рассматривать точки $x_i \in X$ как независимые случайные величины с одинаковым распределением вероятностей μ , то эмпирическая мера $\delta_{x,n}$ сама становится случайной величиной (со значениями в $M_1(X)$). Нас будет интересовать асимптотика ее распределения. Оказывается, в первом приближении эта асимптотика описывается с помощью действия Кульбака.

Ниже нам потребуются две топологии на множествах $B^*(X)$ и $M(X)$. Первая из них — это слабая топология, порожденная функциями из $B(X)$. Она замечательна тем, что по отношению к ней единичный шар в пространстве $B^*(X)$ компактен. Вторую топологию на множестве $M(X)$ определим при помощи системы окрестностей

$$O(\mu) = \left\{ \nu \in M(X) : \left| \int_X f_i d\nu - \int_X f_i d\mu \right| < \varepsilon, \quad i = 1, \dots, n \right\}, \\ \mu \in M(X),$$

где $f_1, \dots, f_n \in L^1(X, \mu)$ и $\varepsilon > 0$. Мы будем называть ее *тонкой топологией*. Она полезна тем, что позволяет придать обычному закону больших чисел следующую форму: для любой вероятностной меры $\mu \in M_1(X)$ последовательность эмпирических мер $\delta_{x,n}$ сходится (по вероятности) к μ в тонкой топологии.

Обозначим через μ^n декартову степень меры $\mu \in M(X)$, определенную на X^n . В соответствии с терминологией А. А. Боровкова и А. А. Могульского [7], следующую теорему естественно назвать локальным принципом больших уклонений.

Теорема 3.1. *Для любой меры $\mu \in M(X)$, любого функционала $\nu \in B^*(X)$ и числа $\varepsilon > 0$ найдется такая слабая окрестность $O(\nu) \subset B^*(X)$, что*

$$\mu^n \{x = (x_1, \dots, x_n) \in X^n \mid \delta_{x,n} \in O(\nu)\} \leq e^{-n(\rho(\nu, \mu) - \varepsilon)}. \quad (13)$$

С другой стороны, для любых двух мер $\mu, \nu \in M(X)$, положительного числа ε и тонкой окрестности $O(\nu) \subset M(X)$ имеет место асимптотическая оценка

$$\mu^n \{x = (x_1, \dots, x_n) \in X^n \mid \delta_{x,n} \in O(\nu)\} \geq e^{-n(\rho(\nu, \mu) + \varepsilon)}, \quad n \rightarrow \infty. \quad (14)$$

В случае метрического пространства X с борелевской σ -алгеброй окрестность $O(\nu)$ в (13) можно также выбрать из слабой топологии, порожденной ограниченными непрерывными функциями.

З а м е ч а н и е. Если $\rho(\nu, \mu) = +\infty$, то под разностью $\rho(\nu, \mu) - \varepsilon$ в оценке (13) подразумевается произвольное положительное число.

Д о к а з а т е л ь с т в о. По теореме 2.2 для любого $\varepsilon > 0$ найдется такая функция $\psi \in B(X)$, что

$$\rho(\nu, \mu) - \frac{\varepsilon}{2} < \nu[\psi] - \lambda(\psi, \mu). \quad (15)$$

Рассмотрим вероятностную меру $\mu_\psi = e^{\psi - \lambda(\psi, \mu)} \mu$. Очевидно, что

$$\frac{d\mu^n(x)}{d\mu_\psi^n(x)} = \prod_{i=1}^n \frac{d\mu(x_i)}{d\mu_\psi(x_i)} = \prod_{i=1}^n e^{\lambda(\psi, \mu) - \psi(x_i)} = e^{n(\lambda(\psi, \mu) - \delta_{x,n}[\psi])}. \quad (16)$$

Определим окрестность функционала ν следующим образом:

$$O(\nu) = \left\{ \delta \in B^*(X) \mid \delta[\psi] > \nu[\psi] - \frac{\varepsilon}{2} \right\}.$$

Тогда при условии $\delta_{x,n} \in O(\nu)$ из (16) и (15) вытекают неравенства

$$\frac{d\mu^n(x)}{d\mu_\psi^n(x)} < e^{n(\lambda(\psi, \mu) - \nu[\psi] + \varepsilon/2)} < e^{n(-\rho(\nu, \mu) + \varepsilon)}.$$

Следовательно,

$$\begin{aligned} \mu^n \{x \in X^n \mid \delta_{x,n} \in O(\nu)\} &= \int_{\delta_{x,n} \in O(\nu)} d\mu^n(x) \leq \int_{\delta_{x,n} \in O(\nu)} e^{n(-\rho(\nu, \mu) + \varepsilon)} d\mu_\psi^n(x) \\ &\leq e^{n(-\rho(\nu, \mu) + \varepsilon)}. \end{aligned}$$

Тем самым, первая часть теоремы доказана.

Оценка (14) тривиально выполняется, если $\rho(\nu, \mu) = +\infty$. Поэтому достаточно доказать ее для случая, когда ν является вероятностной мерой вида $\nu = \varphi\mu$ и действие Кульбака $\rho(\nu, \mu) = \nu[\ln \varphi]$ конечно. Зафиксируем любое число $\varepsilon > 0$ и любую тонкую окрестность $O(\nu) \subset M(X)$. Определим последовательность множеств

$$Y_n = \left\{ x \in X^n \mid \delta_{x,n} \in O(\nu), |\delta_{x,n}[\ln \varphi] - \nu[\ln \varphi]| < \frac{\varepsilon}{2} \right\}$$

(в последнем неравенстве подразумевается, что во всех точках последовательности $x = (x_1, \dots, x_n)$ выполняется условие $\varphi(x_i) > 0$). Заметим, что для $x \in Y_n$

$$\frac{d\mu^n(x)}{d\nu^n(x)} = \prod_{i=1}^n \frac{d\mu(x_i)}{d\nu(x_i)} = \prod_{i=1}^n \frac{1}{\varphi(x_i)} = e^{-n\delta_{x,n}[\ln \varphi]} > e^{-n(\nu[\ln \varphi] + \varepsilon/2)}.$$

Следовательно,

$$\mu^n(Y_n) = \int_{Y_n} d\mu^n(x) \geq \int_{Y_n} e^{-n(\nu[\ln \varphi] + \varepsilon/2)} d\nu^n(x) = e^{-n\rho(\nu, \mu) - n\varepsilon/2} \nu^n(Y_n). \quad (17)$$

В силу закона больших чисел $\nu^n(Y_n) \rightarrow 1$. Поэтому из (17) вытекает оценка (14).

В случае метрического пространства X доказательство остается тем же, только в (15) нужно выбирать не измеримую функцию ψ , а непрерывную (с помощью теоремы 2.3). Теорема 3.1 доказана.

В качестве примера выведем из локального принципа больших отклонений классическую теорему Макмиллана.

Теорема 3.2 (Макмиллан). *Для любой вероятностной меры P на конечном множестве X и любого числа $\varepsilon > 0$ найдется такая окрестность $O(P)$, что*

$$\#\{x = (x_1, \dots, x_n) \in X^n \mid \delta_{x,n} \in O(P)\} \leq e^{n(H(P) + \varepsilon)}.$$

С другой стороны, для любой окрестности $O(P)$ и любого числа $\varepsilon > 0$

$$\#\{x = (x_1, \dots, x_n) \in X^n \mid \delta_{x,n} \in O(P)\} \geq e^{n(H(P) - \varepsilon)}, \quad n \rightarrow \infty.$$

Здесь символ $\#$ обозначает мощность стоящего после него множества, а $H(\cdot)$ обозначает энтропию, определяемую формулой (9).

Доказательство. Это вытекает из равенств (9) и предыдущей теоремы, если в ней положить $\nu = P$ и $\mu(x) = 1$ для всех $x \in X$.

Теорема 3.3 (принцип больших уклонений). Пусть $\mu \in M(X)$. Если множество $U \subset M(X)$ открыто в тонкой топологии, то

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mu^n \{x \in X^n \mid \delta_{x,n} \in U\} \geq - \inf_{\nu \in U} \rho(\nu, \mu), \quad (18)$$

а если множество $V \subset B^*(X)$ замкнуто в слабой топологии, то

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \ln \mu^n \{x \in X^n \mid \delta_{x,n} \in V\} \leq - \inf_{\nu \in V} \rho(\nu, \mu). \quad (19)$$

Доказательство. Оценка (18) сразу вытекает из (14). Напомним, что эмпирические меры $\delta_{x,n}$ являются вероятностными и, следовательно, принадлежат единичному шару в $B^*(X)$. Поэтому без ограничения общности можно считать, что множество V в (19) слабо компактно. Зафиксируем любое число $\varepsilon > 0$. Для каждого функционала $\nu \in V$ найдем такую слабую окрестность $O(\nu)$, для которой выполняется оценка (13). Выберем из всех этих окрестностей конечное покрытие множества V . Суммируя по выбранным окрестностям оценки (13), убеждаемся, что

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \ln \mu^n \{x \in X^n \mid \delta_{x,n} \in V\} \leq - \inf_{\nu \in V} \rho(\nu, \mu) + \varepsilon.$$

В силу произвольности ε отсюда вытекает (19). Теорема 3.3 доказана.

Теорема 3.1 позволяет привести статистическое обоснование для вариационного принципа (12). Рассмотрим величину

$$\begin{aligned} e^{n\lambda(\varphi, \mu)} &= (\mu[e^\varphi])^n = \int_{X^n} e^{\varphi(x_1) + \dots + \varphi(x_n)} d\mu(x_1) \dots d\mu(x_n) \\ &= \int_{X^n} e^{n\delta_{x,n}[\varphi]} d\mu^n(x). \end{aligned} \quad (20)$$

Для каждой вероятностной меры ν существует такая малая окрестность $O(\nu)$, что при всех достаточно больших n выполняются приближенные равенства

$$\int_{x: \delta_{x,n} \in O(\nu)} e^{n\delta_{x,n}[\varphi]} d\mu^n(x) \approx e^{n\nu[\varphi]} \mu^n \{x \mid \delta_{x,n} \in O(\nu)\} \approx e^{n\nu[\varphi]} e^{-n\rho(\nu, \mu)} \quad (21)$$

(с точностью до множителя, находящегося в промежутке от $e^{-n\varepsilon}$ до $e^{n\varepsilon}$). Сравнение формул (20) и (21) показывает, что $\lambda(\varphi, \mu) \geq \nu[\varphi] - \rho(\nu, \mu)$. С другой стороны, все множество вероятностных мер можно покрыть конечным числом окрестностей указанного вида. Тогда интеграл, стоящий в правой части формулы (20), будет оцениваться сверху конечной суммой интегралов вида (21), откуда можно заключить, что $\lambda(\varphi, \mu) \leq \sup_\nu \{\nu[\varphi] - \rho(\nu, \mu)\}$.

СПИСОК ЛИТЕРАТУРЫ

1. *Antonevich A. B., Bakhtin V. I., Lebedev A. V.* On t -entropy and variational principle for the spectral radii of transfer and weighted shift operators. — *Ergodic Theory Dynam. Systems*, 2011, v. 31, № 4, p. 995–1042.
2. *Deuschel J.-D., Stroock D. W.* Large Deviations. Boston: Academic Press, 1989, 307 p. (Pure Appl. Math., v. 137.)
3. *Groeneboom P., Oosterhoff J., Ruymgaart F. H.* Large deviation theorems for empirical probability measures. — *Ann. Probab.*, 1979, v. 7, № 4, p. 553–586.
4. *Jain N.* An introduction to large deviations. — *Lecture Notes in Math.*, 1985, v. 1153, p. 273–296.
5. *Varadhan S. R. S.* Large deviations. — *Ann. Probab.*, 2008, v. 36, № 2, p. 397–419.
6. *Боровков А. А.* Математическая статистика: Оценка параметров. Проверка гипотез. М.: Наука, 1984, 472 с.
7. *Боровков А. А., Мозульский А. А.* О принципах больших отклонений в метрических пространствах. — *Сиб. матем. журн.*, 2010, т. 51, № 6, с. 1251–1269.
8. *Вентцель А. Д., Фрейдлин М. И.* Флуктуации в динамических системах под действием малых случайных возмущений. М.: Наука, 1979, 424 с.
9. *Санов И. Н.* О вероятности больших отклонений случайных величин. — *Матем. сб.*, 1957, т. 42, № 1, с. 11–44.

Поступила в редакцию

9.II.2013

Исправленный вариант

10.III.2014