Записки научных
семинаров ПОМИ
Том 328, 2005 г.

В. Н. Судаков

ГАУССОВСКАЯ КОНЦЕНТРАЦИЯ В МЕТРИКЕ КАНТОРОВИЧА РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН И ФУНКЦИИ КВАНТИЛЕЙ

Пусть $X(t, \omega)$ — гауссовский случайный процесс. Для его параметрического множества (например, для отрезка $[0, 1]$ с мерой Лебега) нас будут интересовать (случайные) времена пребывания траекторий этого процесса, т.е. вероятностные меры, возникающие как образы $\mathrm{mes} \circ X^{-1}$ меры Лебега на параметрическом множестве $[0, 1]$ при отображении, осуществляемом (случайной) выборочной функцией. Нас в первую очередь будут интересовать гауссовские случайные процессы, т.е. такие модели, когда времена пребывания возникают как законы распределения элементов пространства случайных величин $L^2[0, 1]$, выбираемых по некоторой центрированной гауссовской мере γ, сосредоточенной на банаховом сепарабельном пространстве $L^2[0, 1]$, т.е. образующей пространство Лебега–Рохлина, что позволяет говорить об условных мерах на элементах измеримых разбиений.

Легко представить себе, что чем „меньше" ковариационная квадратичная форма гауссовской меры на пространстве (почти) линейных функционалов над ней, т.е. чем больше сопряженная с ней квадратичная форма концентрации Q_γ, описывающая, в частности, эллипсоид рассеивания $\mathcal{E}_\gamma$ меры γ в линейном пространстве с этой гауссовой мерой, тем меньше разброс времен пребывания вокруг некоторого среднего их значения — элемента аффинного конуса M вероятностных мер на прямой. Однако, сравнительную „малость" или „величину" квадратичной формы можно измерять по-разному, например, суммой квадратов экстремальных значений этой квадратичной формы относительно квадрата гильбертовой нормы или же, как это выяснилось теперь, более тонким инструментом — супремумом множества всех экстремальных значений, т.е. минимальным радиусом шара, со-

Настоящая работа частично поддерживалась грантами НШ 2258.2003.1 и РФФИ-ННИО 04-01-04000.

держащего эллипсоид рассеивания. На первый взгляд это может показаться удивительным, ведь след такой матрицы может быть выбран как угодно большим. Например, произведение большого количества одномерных стандартных гауссовских распределений сконцентрировано, в основном, вне шара $C\mathcal{E}_\gamma$, где множитель C может быть сделан как угодно большим. Тем самым, выборочные функции будут иметь как угодно большие средние, и не очевидно, что их распределения не будут при этом неограниченно расплываться по пространству M_2 всех одномерных вероятностных распределений с дисперсиями, снабженному какой-то подходящей метрикой. (Мы будем использовать метрику Канторовича (Монжа–Канторовича–Добрушина) $\varkappa$.) Будет показано, что в описанной ситуации существует такая абсолютная константа c, что если эллипсоид рассеивания $\mathcal{E}_\gamma$ содержится в единичном шаре пространства $L^1([0,1],\mathrm{mes})$, то существует такое распределение $\bar{\mathbf{P}}_\gamma$, что средняя величина $\varkappa$-расстояния нашего (случайного) времени пребывания до распределения $\bar{\mathbf{P}}_\gamma$ не превосходит 0,8. (Получена и точная константа.)

Введем некоторые обозначения.

Пусть $L := L^1([0,1],\mathrm{mes})$ (mes — мера Лебега).

$(\Omega,\mathcal{F},\mathbf{P})$ — вероятностное пространство, $X:\Omega\to\mathbb{R}$ — случайная величина.

Распределение вероятностей значений случайной величины X обозначим $\mathcal{L}(X;\mathbf{P})$.

Символ γ обозначает центрированную гауссовскую „субстандартную“ вероятностную меру в пространстве $L^2(\Omega,\mathcal{F},\mathbf{P})$, т.е. такую гауссовскую меру, что ее эллипсоид рассеивания $\mathcal{E}_\gamma$ (единичный шар ядра H_γ меры γ, иначе — пространства Камерона–Мартина) содержится в единичном шаре $V_2\subset L^2(\Omega,\mathcal{F},\mathbf{P})$.

Символ $\varkappa$ обозначает расстояние Канторовича в пространстве M вероятностных мер λ на прямой $\mathbb{R}$, обладающих барицентром $b(\lambda)$.

$\mathfrak{M}\subset L$ — конус монотонно неубывающих элементов z пространства L, $z\colon[0,1]\to\mathbb{R}$ (кроме не более, чем счетного множества). „Функция квантилей“ $q(X;\alpha)$, где $\alpha\in[0,1]$ — это квантиль порядка α случайной величины X. (Если для некоторого X распределение $\mathcal{L}(X;\mathbf{P})$ имеет пустые интервалы, т.е. некоторые квантили не определены, можно считать, что значением функции q в точке α служит множество всех точек такого незаполненного интервала. Таких α не более счетного множества.) Функция квантилей $q(X,\,\cdot\,)\colon [0,1]\to\mathbb{R}$ может быть определена и как как функция, обратная для функции распределения F_X случайной величины X (с той же оговоркой). Буква q используется и для обозначения канонического отображения $q\colon L^1(\Omega,\mathcal{F},\mathbf{P})\to\mathfrak{M}$, сопоставляющего каждой случайной величине X ее неубывающую перестановку. График каждого элемента z при фиксированном X представляет собой график сохраняющего меру отображения пространства с мерой $([0,1],\mathrm{mes})$ в числовую прямую, снабженную мерой $\mathcal{L}(X;\mathbf{P})$, и можно рассматривать на этом графике вероятностную меру μ_z, канонически проектирующуюся в каждую из этих двух мер. Пространство функций квантилей — это семейство функций аргумента α: $X\mapsto q(X,\alpha)$, $\alpha\in[0,1]$.

Символ μ или $\mu(\gamma)$ обозначает вероятностную меру на $\mathfrak{M}$, являющуюся образом меры γ при отображении q, т.е. $\mu=\gamma\circ q^{-1}$. Эта мера может быть интерпретирована как смесь упомянутых мер μ_z на графиках элементов z, где смешивающей мерой служит мера γ на „параметрическом множестве“ $L^2(\mathbf{P})$.

Теперь перечислим для дальнейших ссылок несколько известных утверждений или простых следствий из них.

A. Пусть на векторном пространстве E задана гауссовская мера γ, так что пространство (E,γ) является гауссовским пространством Лебега–Рохлина (ядро H_γ гауссовской меры γ — гильбертово пространство). Пусть на этом пространстве задана некоторая функция f, удовлетворяющая условию $\mathrm{Lip}\,1$ относительно канонической гильбертовой нормы на пространстве H_γ. Тогда распределение $\gamma\circ f^{-1}$ является образом одномерной стандартной ($N(0,1)$) гауссовской меры при некотором отображении $\mathbb{R}\to\mathbb{R}$ класса $\mathrm{Lip}\,1$. (Будем называть такие меры сублапласовскими.)

Это утверждение — одна из формулировок изопериметрической теоремы для гауссовских мер [5].

B. Если в условиях предыдущего пункта E — сепарабельное банахово пространство, и эллипсоид рассеивания $\mathcal{E}_\gamma$ меры γ содержится в единичном шаре V пространства E, а отображение g — нестрого сжимающее (класса $\mathrm{Lip}\,1$, не обязательно линейное) отображение $E\to F$, где F — другое банахово пространство, то любой функционал h класса $\mathrm{Lip}\,1$ на пространстве F переводит $\gamma\circ g^{-1}$ в сублапласову меру $\gamma\circ g^{-1}\circ h^{-1}$, т.е. имеет сублапласово распределение. Аналогичное утверждение справедливо и для более длинных цепочек пространств и сжимающих отображений.

C. Пусть X и Y — произвольные случайные величины. Тогда расстояние Канторовича $\varkappa(\mathcal{L}(X), \mathcal{L}(Y))$ между их распределениями (определямое как минимальная „работа" по транспортировке одной меры в другую), равно L^1-норме разности их неубывающих перестановок („расстояние Добрушина"). Заметим, что если сами случайные величины X и Y имеют средние значения, что гарантирует конечность расстояния Канторовича, то их неубывающие перестановки принадлежат конусу $\mathfrak{M} \subset L$ с одномерной вершиной — подпространством констант.

D. Как отмечалось в работе [2], если в пространстве $L^2(\mathbf{P})$, где $\mathbf{P}$ — равномерное вероятностное распределение на n точках, задана гауссовская центрированная мера, эллипсоид рассеивания $\mathcal{E}_\gamma$ которой содержится в единичном шаре этого пространства, то непосредственно по определению координатные функционалы обладают свойством Lip 1 относительно гильбертовой (см. также A) нормы в ядре H_γ меры γ — а с ними и все функции квантилей. Аналогично и в общей ситуации, когда $\mathbf{P}$ — неатомическая мера, и мы рассматриваем пространство $\mathfrak{M}$ монотонно неубывающих функций на единичном отрезке, координатные функционалы на пространстве L не могут быть определены, но определяются (mes-почти все) координатные функционалы e_α, $\alpha \in [0,1]$ на конусе $\mathfrak{M}$. Для таких координатных функционалов, которые можно определить, рассматривая разбиения ζ_n отрезка $[0,1]$ на 2^n равных отрезков δ_{nk} и факторпространства по соотношению эквивалентности $z' \overset{n,k}{\sim} z'' \Leftrightarrow z'(\alpha) = z''(\alpha)$ для $\alpha \in \Delta_{n,k}$. Из B вытекает, что образ меры μ при каноническом отображении $\mathfrak{M} \to M/\zeta_n$ — субстандартное распределение. Увеличивая n и переходя к пределу согласно B и C, мы получаем, что определенные координатные функционалы на $\mathfrak{M}$ имеют каждый сублапласово распределение $\mathcal{L}(z(\alpha);\mu) = \mu_\alpha$.

E. Пусть Y — случайная величина с сублапласовским распределением. Тогда минимум по всем $a \in \mathbb{R}$ величины $\mathbf{E}(X-a)^2$ достигается при $a = b(X)$ ($b(X)$ — барицентр распределения случайной величины X) и не превосходит $2^{\frac{1}{2}}\pi^{-\frac{1}{2}}$).

Определим функцию $\bar{z}_\mu \colon \alpha \mapsto b(\mu_a lpha)$. Получаем некоторую неубывающую функцию (элемент $\mathfrak{M}$), и ей соответствует некоторое распределение $\bar{\mathbf{P}}_\mu$. Это средняя траектория z, γ-среднее распределений элементов пространства $L^2(\mathbf{P})$, вокруг которого и группируются другие распределения элементов $L^2(\mathbf{P})$.

Расстояние $\varkappa(\mathcal{L}(X), \bar{\mathbf{P}}_\gamma)$ равно $\|z_X(\alpha) - \bar{z}_\mu(\alpha)\|$. Это случайная величина, равная расстоянию Канторовича между $\mathcal{L}(X)$ и $\bar{\mathbf{P}}$. Выбор наудачу элемента X по мере γ равнозначен выбору по мере μ_γ элемента z из $\mathfrak{M}$.

Окончательно получаем:

$$\begin{aligned}
\mathbf{E}\varkappa(\mathcal{L}(X;\mathbf{P}), \bar{\mathbf{P}}_\gamma) &= \mathbf{E}\mu\|z_X(\alpha) - \bar{z}_\mu(\alpha)\|_L = \\
&= \int\limits_{\mathfrak{M}} \mu(dz) \int\limits_0^1 |z_X(\alpha) - \bar{z}_\mu(\alpha)| d\alpha = \\
&= \int\limits_0^1 d\alpha \int\limits_{\mathfrak{M}} |z_X(\alpha) - \bar{z}_\mu(\alpha)| d\mu \leqslant \\
&\leqslant \int\limits_0^1 d\alpha \int\limits_{\mathfrak{M}} 2^{\frac{1}{2}}\pi^{-\frac{1}{2}} d\mu = 2^{\frac{1}{2}}\pi^{-\frac{1}{2}} \leqslant 0{,}798
\end{aligned}$$

Формулировка результата. Пусть эллипсоид рассеивания $\mathcal{E}_\gamma$ гауссовской меры γ в пространстве $L^2(\Omega, \mathcal{F}, \mathbf{P})$ содержится в единичном шаре V этого пространства. В силу пункта C мы приходим к выводу о том, что в описанной ситуации найдется такое вероятностное распределение $\bar{\mathbf{P}}(\gamma)$ („типичное по гауссовской мере γ распределение" элементов пространства $L^1(\Omega, \mathcal{F}, \mathbf{P})$), что среднее расстояние Канторовича распределения $\mathcal{L}(X;\mathbf{P})$ случайно по мере γ выбранного элемента этого пространства до этого типичного распределения не превосходит $2^{\frac{1}{2}}\pi^{-\frac{1}{2}} < 0{,}8$.

Это утверждение (в чуть более слабой форме, без числовых оценок) было в качестве гипотезы высказано Б. С. Цирельсоном. Оно существенно перекрывает свой дискретный вариант [2] (где, кроме того, оценка константы содержала замеченную М. А. Лифшицем ошибку). Из него также вытекает анонсированное в [3] утверждение 14.1.3 на стр. 210. В работе [3] можно найти библиографию, в частности, ссылку на работу Х. ф. Вайцзекера, обсуждение всей проблематики с которым было весьма полезно. А. В. Судаков построил несколько важных контрпримеров к различным ранним гипотезам. Всем этим математикам я чрезвычайно обязан.

Результаты, цитированные в пункте A, можно найти в работах [4] и [5].

Литература

1. V. Sudakov, *Gaussian measures. A brief survey.* — Rendiconti dell'Istituto di Matematica dell'Univ. di Trieste **26** (1994, supplemento), 289–325.
2. В. Н. Судаков, А. В. Судаков, *Зависимые гауссовские выборки: оценки разброса, свободные от объема выборки.* — В кн.: Записки научн. семин. ПОМИ. т. 320, 2004, с. 166–173.
3. A. V. Sudakov, V. N. Sudakov, and H. v. Weizsaecker, *Typical Distributions: Infinite-Dimensional Approaches.* Asymptotic Methods in Probability and Statistics. Birkhäuser, 2001, pp. 205–214.
4. В. Н. Судаков, Б. С. Цирельсон, *Экстремальные свойства полупространств для сферически инвариантных мер.* — В кн.: Записки научн. семин. ЛОМИ. т. 41, 1974, с. 14–24.
5. B. S. Cirel'son, I. A. Ibragimov, and V. N. Sudakov, *Norms of Gaussians Sample Functions.* Lecture Notes in Mathematics. vol. 550, 1976, pp. 20–41.

Sudakov V. N. Gaussian concentration in the Kantorovich metric of distributions of random variables and the quantile functions.

A sketch of the proof of the following theorem. Let the unit ball of the kernel space H_γ of a centered Gaussian measure γ in the space L^2 is a subspace of the unit ball of this space. There exists a ("typical") univariate distribution $\bar{\mathbf{P}}_\gamma$ such that the expectation with respect to γ of the Kantorovich distance between the distribution of an element of L^2 chosen at random and this typical distribution is less than 0.8.

С.-Петербургское отделение
Математического института
им. В. А. Стеклова РАН

E-mail: sudakov@pdmi.ras.ru

Поступило 15 декабря 2005 г.