Записки научных
семинаров ПОМИ
Том 320, 2004 г.

В. Н. Судаков, А. В. Судаков

ЗАВИСИМЫЕ ГАУССОВСКИЕ ВЫБОРКИ: ОЦЕНКИ РАЗБРОСА, СВОБОДНЫЕ ОТ ОБЪЕМА ВЫБОРКИ

В схеме не обязательно независимой гауссовской выборки объема n, интерпретируемой как выборка объема единица из некоторого центрированного гауссовского распределения γ на n-мерном координатном пространстве выборок $\mathbb{R}^n$, построена неулучшаемая по порядку равномерная по размерности n оценка степени разброса соответствующих выборочных распределений, зависящая только от спектрального радиуса ковариационной квадратичной формы q_γ рассматриваемой меры γ.

1. Под случайной "зависимой гауссовской выборкой объема n" мы здесь будем понимать пространство $\mathbb{R}^n$ всех мыслимых выборок x объема n, на котором задана некоторая гауссовская мера γ, управляющая игрой случая при получении выборки. При этом "выборочное распределение" $\mathbf{P}_{n,x}$ — элемент пространства вероятностных распределений на числовой прямой $\mathbb{R}$ — по обычному правилу детерминированным образом строится по выборке $x \in \mathbb{R}^n$ объема единица из распределения γ и не зависит от происхождения этой выборки. Если, например, $\gamma = N(0, I_n)$, где I_n (ковариационная матрица координатных функционалов относительно распределения γ) — это единичная $n \times n$-матрица, то γ-случайные элементы x соответствуют повторным независимым выборкам объема n из распределения $N(0,1)$.

Само пространство выборок $\mathbb{R}^n$ можно рассматривать как пространство функций на некотором множестве Ω, состоящем из n элементов: $\Omega = \{\omega_1, \ldots, \omega_n\}$, где элементы пространства Ω суть координатные функционалы на пространстве выборок, с равномерным вероятностным распределением $\mathbf{P} = \mathbf{P}_n$.

Здесь прослеживается аналогия "выборочного распределения" $\mathbf{P}_{n,x}$, соответствующего по известному правилу некоторой

Работа поддержана грантами РФФИ 02-01-00263, РФФИ-ННИО 04-01-04000, НШ-2258.200 и Программой фундаментальных исследований РАН "Современные проблемы теоретической математики".

выборке $x \in \mathbb{R}^n$, и "времени пребывания" $\mathbf{P} \circ x^{-1}$ для выборочной функции $[x\colon T \to \mathbb{R}]$ некоторого случайного процесса $X\colon T \to L^0(\Omega, \mathbf{P})$, если параметрическое множество T снабжено некоторой вероятностной мерой $\mathbf{P}$. Более общая ситуация с произвольным (бесконечномерным) вероятностным распределением $\mathbf{P}$ на множестве F линейных функционалов на выборочном гауссовском пространстве будет изучена позднее.

При исследовании вопросов, связанных с феноменом концентрации меры и существования типичных распределений (библиографию см. в [1]), в особенности в произвольной бесконечномерной ситуации, естественно, начиная анализ с конечномерных выборочных пространств, интерпретировать, как об этом говорилось выше, выборочные распределения как "времена пребывания" выборок, понимаемых как случайные функции на пространстве с вероятностной мерой, относительно некоторого равномерного на конечном множестве Ω вероятностного распределения, в терминах которого, независимо от размерности, и формулируется условие нормировки гауссовской меры γ.

2. Рассмотрим возникающую схему подробнее. Пусть, как и выше, $\Omega = \{\omega_1, \dots, \omega_n\}$ — некоторое множество, и $\mathbf{P}_n$ — равномерное распределение вероятностей на нем: $\mathbf{P}_n(\{\omega\}) = n^{-1}$ для каждого $\omega = \omega_1, \dots, \omega_n \in \Omega$. Пространство $\mathbb{R}^n$ функций X на множестве Ω естественно наделяется тем самым евклидовой (гильбертовой в более общей ситуации) структурой, в частности — $L^2(\mathbf{P}_n)$-нормой и соответствующим скалярным произведением: для $X = (X(\omega_1), \dots, X(\omega_n)) \in \mathbb{R}^n$

$$\|X\|^2 = \sum_{k=1}^{n} n^{-1}(X(\omega_k))^2. \tag{1}$$

(Здесь $X(\omega_k) = e_k(X)$, $k = 1, \dots, n$, $X \in \mathbb{R}^n$ — координатные функционалы на $\mathbb{R}^n$.)

Если теперь на евклидовом пространстве $(\mathbb{R}^n, \|\cdot\|)$ задать стандартную относительно этой нормы гауссовскую меру γ_0, то координатные функционалы e_k, $k = 1, \dots, n$ окажутся независимыми гауссовскими величинами с одинаковым распределением $\mathcal{L}(e_k) = N(0, n)$. Таким образом, координаты точек $\mathbb{R}^n$ — это носители нагрузок выборочных распределений для повторной независимой выборки объема n из распределения $N(0, n)$. Читатель, конечно, обратил внимание на то, что теперь, в отличие от обычной постановки, когда выборочные распределения с ростом n сближались с $N(0,1)$, наши в новом смысле “выборочные” распределения ведут себя иначе. Из дальнейших рассуждений будет видно, что стохастически эти “расплывающиеся” (со скоростью $n^{\frac{1}{2}}$) новые “выборочные распределения” с ростом n не “разбегаются друг от друга”, что на первый взгляд можно было бы предположить, хотя, как правило, и не концентрируются вокруг (как могло бы показаться в этом случае) распределения $N(0,n)$, своего для каждого n (термины будут уточнены). Главное же в том, что мы сможем полностью отказаться от предположения о независимости выборочных значений, ограничиваясь лишь требованием нормировки спектрального радиуса ковариационной квадратичной формы гауссовской меры γ (требование “субстандартности” гауссовской меры γ).

Мы намереваемся исследовать, при каком ограничении на гауссовскую (гауссовские) меры γ можно оценить в некотором разумном смысле степень концентрации “времен пребывания” наших конечных случайных гауссовских последовательностей. Простейшая оценка получается, если потребовать, чтобы сумма квадратов $\sum_{k=1}^{n}\sigma_k^2$ в L^2-норме длин полуосей эллипсоида рассеивания $\mathcal{E}_\gamma$ меры γ была ограничена фиксированной константой. Иначе говоря, чтобы этот эллипсоид был структурно ограничен (что тривиально для конечномерной модели, но не в общем случае) элементом пространства выборок, и его структурный супремум имел бы ограниченную фиксированной константой $L^2(\mathbf{P}_n)$-норму. В терминах этой константы (т.е. через квадрат гильберт-шмидтовской нормы ковариационного оператора меры γ) выражаются оценки концентрации самих гауссовских выборок, а с использованием этих последних и оценки концентрации выборочных распределений. Выполнение этого условия (ограниченность фиксированной константой следа ковариационного оператора, ограниченность суммы квадратов длин в $L^2(\mathbf{P})$ осей эллипсоида рассеивания $\mathcal{E}_\gamma$) необходимо и достаточно для того, чтобы можно было оценить снизу γ-меры центрированных шаров произвольного радиуса (теорема Минлоса–Сазонова), откуда и выводятся оценки для взаимной близости (степени концентрации) выборочных распределений.

Пафос настоящей работы как раз в том, что будут даны оценки взаимной близости выборочных распределений (не самих выборок по $L^2(\mathbf{P})$-норме, что не утверждается, и, конечно, тем более не по стандартной евклидовой норме в пространстве выборок $\mathbb{R}^n$), но в терминах только величины старшей полуоси эллипсоида $\mathcal{E}_\gamma$ в $L^2(\mathbf{P})$-норме, при произвольном поведении величин остальных полуосей. При этом условие нормировки гауссовской меры заключается только в том, что эллипсоид рассеивания $\mathcal{E}_\gamma$ рассматриваемой гауссовской меры на пространстве выборок (на пространстве $L^2(\mathbf{P})$) содержится в единичном шаре этой евклидовой L^2-нормы на пространстве $\mathbb{R}^n$, т.е. нормы, задаваемой определением (1).

Хотя в формуле (1) присутствует размерность n, возможность интерпретации такой нормы как $L^2(\mathbf{P})$-нормы позволяет устранить всякое упоминание о размерности того функционального пространства (пространства $L^2(\mathbf{P})$), на котором задается гауссовская мера γ, относительно которой и будут формулироваться утверждения о степени концентрации распределений $\mathcal{L}(X;\mathbf{P})$ случайных величин из $L^2(\mathbf{P})$. Это позволит позднее сформулировать и доказать теорему о степени концентрации распределений и для произвольного, в частности, бесконечномерного гильбертова пространства случайных величин.

3. Напомним или введем некоторые термины и обозначения, в некоторых случаях уже использовавшиеся выше и полезные для наших целей предположения и утверждения. (См. также [2].)

Все гауссовские пространства выборок предполагаются пространствами Лебега–Рохлина, так что ядро гауссовской меры — это всегда гильбертово пространство. Впрочем, для встречающихся в ближайшем тексте конечномерных выборочных гауссовских пространств это замечание несущественно.

Стандартной относительно некоторого вероятностного распределения $(\Omega,\mathbb{P})$ гауссовской мерой (или стандартным гауссовским слабым распределением) γ в пространстве $L^2(\mathbb{P})$ называется такая гауссовская мера (слабое распределение) γ_0, для которой ее эллипсоид рассеивания $\mathcal{E}_{\gamma_0}$ — это единичный шар в пространстве $L^2(\mathbb{P})$. Иначе говоря, ковариационный оператор такой меры γ_0 — единичный, а ее характеристический функционал, заданный на сопряженном с $L^2(\mathbf{P})$ пространстве F, имеет вид

$$[F \ni f \mapsto \exp(-\frac{1}{2}\|f\|^2)].$$

Центрированная гауссовская мера γ в пространстве $L^2(\mathbb{P})$ называется субстандартной гауссовской, если ее эллипсоид рассеивания содержится в единичном шаре этого пространства (т.е. если ее ковариационная квадратичная форма q_γ мажорируется квадратом евклидовой нормы в сопряженном ц $L^2(\mathbf{P})$ пространстве H: для каждого $f \in H$ выполняется неравенство

$$q_\gamma(f) := \|f\|^2_{L^2(\gamma)} \leqslant \|f\|^2_H).$$

Вероятностная мера μ на прямой называется сублапласовской, если она является образом $N(0,1) \circ \varphi^{-1}$ стандартного одномерного гауссовского распределения при отображении $\varphi\colon \mathbb{R} \to \mathbb{R}$, удовлетворяющем условию Липшица с константой единица ($\varphi \in \mathrm{Lip}\,1$). Среднее по сублапласовской мере расстояние точки на прямой до среднего этого сублапласовского распределения не превосходит 0.8.

Расстояние между вероятностными распределениями в метрическом пространстве S, в частности и в первую очередь на числовой прямой $\mathbb{R}$, будет измеряться в метрике Канторовича $\varkappa$. Напомним, что в пространстве M_1 вероятностных распределений с конечными средними на числовой прямой $\mathbb{R}$ для μ, $\nu \in M_1$

$$\varkappa(\mu,\nu) = \int |F_\mu(t) - F_\nu(t)|dt.$$

Закон распределения $\mathbf{P}\circ X^{-1}$ случайной величины $X \in L^0(\Omega,\mathbf{P})$ обозначается $\mathcal{L}(X;\mathbf{P})$ или просто $\mathcal{L}(X)$.

Множество всех монотонно неубывающих функций на отрезке $[0,1]$ будем обозначать $\mathcal{M}$, а подмножество, состоящее из монотонных функций, измеримых относительно разбиения ξ_n отрезка $[0,1]$ на n равных отрезков, будет обозначаться $\mathcal{M}_n$.

Вместе с каждой случайной величиной X (функцией на $(\Omega,\mathbf{P})$) мы будем рассматривать ее неубывающую перестановку $J(X) \in \mathcal{M}$, т.е. заданную на $[0,1]$ монотонно неубывающую функцию, для которой

$$\mathcal{L}(J(X);\mathrm{mes}) = \mathcal{L}(X;\mathbf{P}).$$

Здесь mes — мера Лебега на единичном отрезке. С естественными оговорками можно считать, что $J(X)(q) = F^{-1}(q)$, $q \in [0,1]$. В нашем случае, когда мера $\mathbf{P}$ — равномерное распределение на конечном множестве Ω, $\#\Omega = n$, монотонные функции $J(X)$ измеримы относительно разбиения ξ_n единичного отрезка на n равных частей.

Отображение J обладает следующим очевидным свойством: для любой функции f и любой случайной величины Z $\mathcal{L}(f(Z) = \mathcal{L}(f(J \circ Z))$

Предложение 1. *Для произвольных элементов* $X, Y \in L^1(\mathbf{P})$

$$\varkappa(\mathcal{L}(X;\mathbf{P}), \mathcal{L}(Y;\mathbf{P})) = \|J(X) - J(Y)\|_{L^1(\mathrm{mes})}.$$

В частности, для выборочных распределений, соответствующих двум выборкам $(x_k, k = 1, \dots, n)$ *и* $(y_k, k = 1, \dots, n)$ *одного объема, расстояние Канторовича между этими распределениями равно*

$$\varkappa(\mathcal{L}(x;\mathbf{P}_n), \mathcal{L}(y;\mathbf{P}_n)) = n^{-1} \sum_{k=1}^{n} |x_{(k)} - y_{(k)}|. \tag{2}$$

Участвуют разности соответствующих друг другу членов двух вариационных рядов.

Как и обычно, выборочным распределением $\mathbf{P}_n^x$, соответствующим элементу $x \in \mathbb{R}^n$, называется дискретное распределение

$$\mathcal{L}(x;\mathbf{P}_n) = \mathbf{P}_n \circ x^{-1}.$$

Здесь под x подразумевается произвольный элемент пространства $\mathbb{R}^n$, рассматриваемый как функция на Ω, т.е., если угодно, на множестве координатных функционалов e_k.

Основным инструментом, используемым в дальнейшем, будет служить изопериметрическое неравенство для гауссовских мер [3, 4] и его следствия. Следствием изопериметрического неравенства является следующее утверждение.

Предложение 2 (Теорема об образе гауссовской меры при липшицевом отображении).

Пусть (E, γ) *— векторное пространство с гауссовской мерой,* γ*,* $f\colon E \to \mathbb{R}$ *— отображение, удовлетворяющее условию Липшица с константой* 1 относительно гильбертовой нормы на ядре H_γ этой гауссовской меры. *Тогда образ* $\gamma \circ f^{-1}$ *меры* γ *при отображении* f *сублапласов.*

В частности, образ субстандартной гауссовской меры, заданной в евклидовом или гильбертовом пространстве E, при произвольном липшицевом с константой единица отображении является сублапласовым распределением на числовой прямой $\mathbb{R}$.

4. Переходим к формулировке и краткому изложению хода доказательства основной теоремы этой работы.

Теорема. *Существует такая абсолютная константа c, $c > 0$, что для любого натурального n какова бы ни была субстандартная гауссова мера γ на евклидовом пространстве $L^2(\mathbf{P}_n)$ всех выборок объема n, существует такое одномерное вероятностное распределение μ_γ на числовой прямой $\mathbb{R}$, что*

$$\mathbf{E}_\gamma \varkappa(\mathcal{L}(X;\mathbf{P}_n),\mu_\gamma) \leqslant c.$$

Иначе говоря,

$$\int\limits_{L^2(\mathbf{P}_n)} \varkappa(\mathcal{L}(X;\mathbf{P}_n),\mu_\gamma)\gamma(dX) \leqslant c.$$

Доказательство. Фиксируем n. Рассмотрим отображение $J\colon L^2(\mathbf{P}_n) \to \mathcal{M}_n \subset L^2[0,1]$. Отображение J является изоморфизмом на каждом конусе элементов $L^2(\mathbf{P}_n)$, элементы которого переводятся в конус M_n элементов $L^2(\mathbf{P}_n)$ со строго монотонно возрастающими элементами одной и той же перестановкой координат. Отображение J сохраняет распределения и на каждом описанном конусе является изометрией. Как следствие, в целом отображение J удовлетворяет условию Липшица с константой 1, и с помощью отображения J каждый аффинный функционал l на конусе $\mathcal{M}_n$ поднимается на всё $L^2(\mathbf{P}_n)$ до функционала, также обладающего свойством Липшица с константой 1.

Пусть $\mu_\gamma = \gamma \circ J^{-1}$ — образ меры γ при отображении J. Из липшицевости J следует теперь, что распределения аффинных функционалов на $(\mathcal{M},\mu_\gamma)$ имеют первый момент, и существует среднее меры μ — элемент $[\mu] \in \mathcal{M}$, $[\mu](t) = m_k$, если $t \in \Delta_k \in [0,1]/\xi_n$ — барицентр меры μ.

Среднее расстояние $E_\gamma \varkappa(\mathcal{L})X;\mathbf{P}_n),\mu_\gamma)$ равно средней величине L^1-разности элементов $JX \in \mathcal{M}_n$ и $[\mu]$. Элемент $[\mu]$ вычисляется покоординатно: для каждого элемента Δ_k разбиения ξ_n значение среднего определяется как среднее одномерного распределения, получающегося как ортогональная в $L^2(\mathrm{mes})$ проекция на одномерное координатное пространство. По липшивости проекторов, все одномерные распределения координатных функционалов сублапласовы, и средние этих одномерных распределений суть как раз значения функции μ на элементах разбиения ξ_n. Для каждого элемента Z конуса $\mathcal{M}_n$, $Z(t) = z_k$ для $t \in \Delta_k \in [0,1]/\xi_n$ его L^1-расстояние до элемента $[\mu]$ равно $\sum_{k=1}^n \cdot n^{-1}|z_k - m_k|$. Но L^1-норма разности на конусе $\mathcal{M}_n$ соответствует $\varkappa$-расстоянию распределений. Интегрируя это выражение и принимая во внимание сказанное выше о распределении отклонения сублапласовской случайной величины от своего среднего, мы приходим к заключению теоремы.

Утверждение, близкое к доказанному, в свое время высказывалось в качестве гипотезы Б. С. Цирельсоном [5] в связи с феноменом существования типичных распределений. Подробнее об этом будет говориться в дальнейших публикациях. Большую пользу также принесло неоднократное обсуждение проблемы с профессором Х. ф. Вайцзекером.

Литература

1. A. V. Sudakov, V. N. Sudakov, and H. v. Weizsaecker, *Typical Dustributions: Infinite-Dimensional Approaches. Asymptotic Methods in Probability and Statistics with Applications.* Birkhäuser, 2001, pp. 205–214.
2. V. Sudakov, *Gaussian Measures. A Brief Survey.* Rendicoti dell'istituto di matematica dell'universita di Trieste. Vol. XXVI, 1994, Supplemento.
3. 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.
4. B. S. Cirelson, I. A. Ibragimov and V. N. Sudakov, *Norms of Gaussian Sample Functions.* Lecture Notes in Mathematics, 550. 1976, pp. 20–41.
5. Б. С. Цирельсон, *Гипотеза.* Рукопись. Архив Лаборатории статистических методов ПОМИ РАН.

Sudakov V. N., Sudakov A. V. Dependent Gaussian samples: estimation of the scatter uniform in dimension.

An exact in order dimension-free estimation of the scatter of samples chosen according to an arbitrary Gaussian distribution is proposed. This estimation is only based on the spectral radius of the covariance quadratic form of γ.

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

Поступило 24 декабря 2004 г.