Записки научных
семинаров ПОМИ
Том 260, 1999 г.

А. В. Судаков, В. Н. Судаков

ЭМПИРИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ И РАВНОМЕРНЫЕ ПО РАЗМЕРНОСТИ ОЦЕНКИ РАЗБРОСА В ЗАДАЧЕ О ТИПИЧНЫХ РАСПРЕДЕЛЕНИЯХ

Пусть E — конечномерное пространство большой размерности и пусть $\mathbf{P}$ — некоторая борелевская вероятностная мера второго порядка на этом пространстве, т.е. такая вероятностная мера, для которой линейные на E функционалы X — элементы сопряженного с E пространства $F = E'$ — имеют относительно этой меры распределения $\mathcal{L}(X, \mathbf{P})$ с конечными вторыми моментами. Здесь $\mathcal{L}(X, \mathbf{P}) = \mathcal{L}(X) = \mathbf{P} \circ X^{-1}$ — распределение элемента $X \in F$, рассматриваемого как случайная величина относительно вероятностной меры $\mathbf{P}$. Проблематика типичных распределений (см., например, [1]) — это в первую очередь выяснение условий, при которых существует такое одномерное распределение $\overline{\mathbf{P}}$, которое в том или другом смысле «типично» для распределений $\mathcal{L}(X, \mathbf{P})$ элементов пространства F. Под типичностью же понимается типичность относительно некоторого вероятностного распределения μ на пространстве F случайных величин, которое в каком-то смысле представляется «естественным» для этого пространства. Если, например, пространство F снабжено структурой евклидова пространства (какой-то априорной или индуцированной вложением $F \subset L^2(E, \mathbf{P})$), то в качестве такой естественной вероятностной меры μ в пространстве F могут быть рассмотрены равномерное распределение m_F на единичной сфере (границе единичного шара) пространства F, гауссовская мера с теми же корреляционными характеристиками, что и мера m_F, или стандартная гауссовская мера γ_F° в евклидовом пространстве F. Свойство типичности некоторого одномерного распределения $\overline{\mathbf{P}}$ означает концентрацию (в том или ином смысле) меры $\mu \circ \mathcal{L}^{-1}(\cdot, \mathbf{P})$ вокруг этого распределения. Таким образом, понятие концентрации, типичности некоторого распределения связано с

Работа поддержана грантами РФФИ 99-01-00109 и 96-15-96199.

фиксацией некоторой метрики ρ на множестве $\mathcal{P}_2(\mathbb{R}^1)$ вероятностных мер второго порядка на вещественной прямой $\mathbb{R}$, причем вся ситуация вокруг рассматриваемой проблемы может существенно меняться при замене одной метрики на другую. Так, бывает удобно вместо поиска конкретного типичного распределения $\overline{\mathbf{P}}$ доказывать малость среднего расстояния

$$\overline{\boldsymbol{\rho}}(\mathbf{P}, \mu) = \mathbf{E}_{\mu\otimes\mu}(\rho(\mathcal{L}(X_1, \mathbf{P}), \mathcal{L}(X_2, \mathbf{P}))),$$

где X_1, X_2 — независимо выбранные по мере μ случайные величины из пространства F. При этом, очевидно, малость величины $\overline{\boldsymbol{\rho}}$ означает высокую концентрацию меры $\mu\circ\mathcal{L}^{-1}(\cdot, \mathbf{P})$ и существование искомого распределения $\overline{\mathbf{P}}$ (при этом задача оценивания степени типичности этого распределения связана со свойствами пространства $\mathcal{P}_2(\mathbb{R}^1), \rho$, т.е. со свойствами метрики ρ). Так, для метрики

$$\rho_2(\alpha, \beta) = \int |F_\alpha(t) - F_\beta(t)|^2 dt,$$

использованной в подобных целях С. В. Нагаевым [2], среднее расстояние

$$\overline{\boldsymbol{\rho}}(\mathbf{P}, \mu) = \iint\limits_{E\times E} \rho(\mathcal{L}(X_1, \mathbf{P}), \mathcal{L}(X_2, \mathbf{P}))(\mu\otimes\mu)\, d(X_1, X_2)$$

в классе гауссовских центрированных мер в пространстве F, ковариационные операторы которых мажорируются единичным (т.е. в классе гауссовских мер, мажорируемых стандартной гауссовской мерой), достигает минимума по μ на одномерном гауссовском распределении, в то время как для более сильных естественных метрик (например, для метрики Канторовича), положение дел гораздо сложнее, см. [3].

Особое значение имеет задача получения оценок, не зависящих явно от размерности конечномерного пространства случайных величин F, на принципиальную важность чего в свое время обратил внимание одного из авторов Б. С. Цирельсон. Мы свяжем существование таких оценок со свойствами метрики ρ, проявляющимися в связи с рассмотрением эмпирических мер для повторных независимых выборок большого объема.

Пусть F — евклидово пространство размерности n, $\mu = \gamma_0^n$ — стандартная гауссовская мера в этом пространстве. Пусть

$\{\omega_{n,1},\ldots,\omega_{n,n}\}$ — какой-нибудь ортонормированный базис в F. Пусть $\mathbf{P}_n^0$ обозначает равномерное распределение на множестве

$$S(\mathbf{P}_n^0)=\{\,n^{\frac{1}{2}}\omega_{n,k}:k=1,\ldots,n\,\}.$$

Пространство F реализуем как пространство наборов коэффициентов разложений его элементов по базису, сопряженному с базисом $\{\omega_{n,1},\ldots,\omega_{n,n}\}$, превратив, таким образом, F в пространство l_n^2. Нетрудно убедиться, что евклидова структура этого пространства совпадает с евклидовой структурой, индуцированной уже упоминавшимся вложением $F\subset L^2(E,\mathbf{P}_n^0)$.

Пусть $X\in F$ выбирается наудачу в соответствии со стандартной гауссовской мерой $\mu=\gamma_F^{\circ}$. При этом X как элемент пространства l_n^2 можно рассматривать как независимую повторную выборку из стандартного нормального закона $N(0,1)$. Мы, однако, в первую очередь будем рассматривать значения функции X на элементах пространства E, на которых сосредоточена мера $\mathbf{P}=\mathbf{P}_n^0$ (т.е. воспринимая X прежде всего как элемент пространства $L^2(\mathbf{P})$, не обращая внимание на размерность пространства E). Если как функция X на элементах ортонормированного базиса ассоциировалась для нас с повторной независимой выборкой из закона $N(0,1)$, то теперь значения той же функции на элементах $S(\mathbf{P}_n^0)$ будут выборочными значениями для подобной выборки из закона $N(0,n)$. Выборочное распределение $\theta(X)$, отвечающее этой выборке, — это мера из пространства $\mathcal{P}(\mathbb{R})$:

$$\theta(X)=\mathbf{P}\circ X^{-1}\in\mathcal{P}(\mathbb{R}).$$

Будем в дальнейшем рассматривать метрики ρ на подпространствах пространства $\mathcal{P}(\mathbb{R})$, порождаемые следующим образом. Пусть $\mathcal{X}$ — нормированное с нормой $\|\cdot\|$ пространство измеримых функций на пространстве $([0,1],\mathfrak{B},\mathrm{mes})$, причем для любого $X\in\mathcal{X}$ выполнено $\|X\|=\|X\circ T\|$, где T — произвольный автоморфизм отрезка $[0,1]$ с мерой Лебега mes. Тогда для $\alpha,\beta\in\mathcal{X}$ выполнено равенство $\rho(\alpha,\beta)=\|If-Ig\|$, где $\mathcal{L}(f)=\alpha$, $\mathcal{L}(g)=\beta$ и I — неубывающая перестановка $f\in\mathcal{X}$.

Пусть теперь некоторая метрика ρ на пространстве $\mathcal{P}(\mathbb{R})$ такова, что для некоторой абсолютной константы C для любого натурального n среднее по распределению $\gamma_F^{\circ}\otimes\gamma_F^{\circ}$ среднее значение расстояния между двумя независимо полученными эмпирическими мерами не превосходит C:

$$\mathbf{E}\rho(\mathbf{P}_n^0 \circ X_1^{-1}, \mathbf{P}_n^0 \circ X_2^{-1}) \leqslant C. \qquad (*)$$

(Вместо расстояния между двумя независимо полученными эмпирическими мерами можно рассматривать расстояние между эмпирической мерой и «теоретическим» распределением, т.е. $N(0,n)$, но мы интересуемся формулировкой, но содержащей размерности). Пусть M — класс таких метрик. Для метрики $\rho \in M$, по сказанному выше, мы приходим к следующему заключению. Это, по существу, переформулировка свойства $(*)$ в терминах не эмпирических мер, а распределений случайных элементов пространства E.

Предложение 1. *Пусть в векторном n-мерном пространстве E вероятностная мера $\mathbf{P}$ представляет собой n атомов, несущих равные нагрузки и сосредоточенных на линейно-независимых элементах $\omega_1, \ldots, \omega_n$ пространства E. Пусть γ — стандартная гауссовская мера на линейном пространстве $E = F'$ с евклидовой структурой, определяемой вложением $E \subset L^2(\mathbf{P})$. Для метрик ρ из класса M среднее значение случайной величины*

$$\rho(\mathbf{P} \circ X_1^{-1}, \mathbf{P} \circ X_2^{-1}), (X_1, X_2) \in (E \times E, \gamma \otimes \gamma)$$

не превосходит $C = C(\rho)$.

Хотя свойство M относится только к эмпирическим мерам для независимых выборок, Предложение 1 может быть обобщено на случай произвольных мер $\mathbf{P}$ второго порядка. Действительно, поскольку оценка среднего расстояния константой $C = C(\rho)$ для равномерных распределений на конечных линейно-независимых подмножествах любой мощности не зависит от размерности, а для произвольной меры второго порядка существует аппроксимирующая эту меру в требуемом смысле последовательность мер такого типа (с увеличением размерности пространств их задания), мы приходим к следующей теореме.

Теорема 1. *Пусть в векторном конечномерном пространстве E задана вероятностная мера $\mathbf{P}$ второго порядка. Пусть γ — стандартная гауссовская мера на линейном пространстве $E = F'$ с евклидовой структурой, определяемой вложением $E \subset L^2(\mathbf{P})$. Для метрик ρ из класса M среднее значение случайной величины*

$$\rho(\mathbf{P} \circ X_1^{-1}, \mathbf{P} \circ X_2^{-1}), (X_1, X_2) \in (E \times E, \gamma \otimes \gamma)$$

не превосходит $C = C(\rho)$.

Наконец, несколько слов о возможном методе проверки свойства M.

Теорема 2. *Пусть метрика* ρ *такова, что для любой меры второго порядка* $\mathbf{P}$ *функция*

$$\rho((X_1, X_2)) = \rho(\mathbf{P} \circ X_1^{-1}, \mathbf{P} \circ X_2^{-1})$$

липшицева с константой единица относительно евклидовой метрики на пространстве $E \oplus E$*, где на каждом прямом слагаемом евклидова метрика порождается, как обычно, вложением в соответствующее пространство* $L^2(\mathbf{P})$*. Если для некоторых абсолютных положительных констант* a *и* b *выполняется условие*

$$(\gamma_F \otimes \gamma_F)(\{\, (X_1, X_2) : \rho(\mathbf{P} \circ X_1^{-1}, \mathbf{P} \circ X_2^{-1}) \leqslant a \,\}) \geqslant b,$$

то метрика ρ *обладает свойством* M.

Литература

1. V. N. Sudakov, H. v. Weizsäcker, *Typical distributions: the infinite-dimensional approach*. Abstracts of communications of the International conference "Asymptotic Methods in Probability and Mathematical Statistics", St. Petersburg, June 24–28, 1998.
2. С. В. Нагаев, *О распределении линейных функционалов в конечномерных пространствах высокой размерности.* — ДАН СССР **263**, № 2 (1982), 295–296.
3. А. В. Судаков, *Контрпример к гипотезе о монотонности интеграла по гауссовской мере.* — ЗНС ПОМИ, **260** (1999).

Sudakov A. V., Sudakov V. N. Empirical distributions and dimension-invariant estimations in the problem of existence of typical distributions.

A connection between empirical measures and dimension-invariant estimations in the problem of existence of typical distributions for finite-dimensional spaces of random variables is established.

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

Поступило 20 декабря 1999 г.