УДК 519. 21

В. И. Степахно, М. Я. Ружило (Нац. аграр. ун-т, Київ)

О ВЫБОРКАХ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕКТОРОВ В ПРОСТРАНСТВАХ НЕОГРАНИЧЕННО ВОЗРАСТАЮЩЕЙ РАЗМЕРНОСТИ

We study the asymptotic behavior of a set of random vectors ξ_i, $i = 1, \ldots, m$ whose coordinates are independent and equally distributed in a space of infinitely increasing dimension. We investigate the asymptotics of distribution of the random vectors, the consistency of sets $M_m^{(n)} = \{\xi_1, \ldots, \xi_m\}$ and $X_n^\lambda = \{x \in X_n : \rho(x) \le \lambda n\}$, and the mutual disposition of vector pairs.

Вивчається асимптотична поведінка множини випадкових векторів ξ_i, $i = 1, \ldots, m$, координати яких незалежні і однаково розподілені у просторі необмежено зростаючої розмірності. Досліджується асимптотика розподілу випадкових векторів, сумісність множин $M_m^{(n)} = \{\xi_1, \ldots, \xi_m\}$ і $X_n^\lambda = \{x \in X_n : \rho(x) \le \lambda n\}$, взаємне розташування пар векторів.

1. Рассматриваются независимые одинаково распределенные случайные векторы $\{\xi_k, k = 1, \ldots, m\}$ в пространстве R^n. Предполагается, что координаты вектора $\xi_k = \left(\xi_k^1, \ldots, \xi_k^n\right)$ независимы и одинаково распределены. Обозначим через $M_m^{(n)}$ случайное множество в R^n: $M_m^{(n)} = (\xi_k, \ldots, \xi_m)$.

В настоящей статье изучается асимптотическое поведение этого множества, когда $n \to \infty$ и $m \to \infty$. Такая задача рассматривалась в работах А. В. Скорохода и В. И. Степахно [1, 2] для случая, когда ξ_1^1 имеет гауссовское распределение или распределение некоторого специального вида.

Здесь изучается случай, когда $P\left\{\xi_1^1 = \pm 1\right\} = 1/2$. Это простейшее дискретное распределение и в этом случае возможно более полное исследование поведения множества $M_m^{(n)}$. Отметим, что применяемый метод тесно связан с идеями „больших уклонений", впервые рассмотренными H. Chernoff [3] (см. дальнейшее развитие этих исследований в монографии R. S. Ellis [4].)

2. В указанных выше работах А. В. Скорохода и В. И. Степахно основные результаты относились к исследованию множества на прямой $\{\|\xi_k\|, k = 1, \ldots, m\}$. В нашем случае $P\{\|\xi_k\| = \sqrt{n}\} = 1$ и такие исследования теряют смысл. Тем не менее можно получить содержательные теоремы (утверждения), характеризующие множество $M_m^{(n)}$.

Обозначим через X_n множество векторов X_i в пространстве R^n, для которых $x^i \in \{-1; 1\}, i = 1, \ldots, n$, x^i — координаты X.

Очевидно, $M_m^{(n)} \subset X_n$. Для характеристики $M_m^{(n)}$ будем использовать функции $\varphi : X_n \to R$, для которых $\varphi(\Pi X) = \varphi(X)$ для любого оператора Π, переставляющего координаты вектора X. Т. е. φ есть симметричная функция координат. Она зависит только от числа единиц среди координат (остальные будут –1). Поэтому любая такая функция есть некоторая функция от

$$\rho(x) = \sum_{i=1}^{n} x_i.$$

Эта функция и будет использоваться для описания множества $M_m^{(n)}$. Предварительно исследуем асимптотику распределения $\rho(\xi)$.

Лемма 1. *Пусть* $|k| \le n$, $\frac{n-k}{2} \in Z_+$, *тогда*

$$\log P\{\rho(\xi) = k\} = n g\left(\frac{|k|}{n}\right) + o(n), \tag{1}$$

где $g(\lambda)$: $[0;1] \to R'$ *определяется формулой*

$$g(\lambda) = \frac{1+\lambda}{2}\log\frac{1}{1+\lambda} + \frac{1-\lambda}{2}\log\frac{1}{1-\lambda}. \tag{2}$$

Доказательство. Из формулы Бернулли следует

$$P\{\rho(\xi) = k\} = C_n^{(n-k)/2} \cdot 2^{-n} = \frac{n!\cdot 2^{-n}}{\left(\frac{n+k}{2}\right)!\left(\frac{n-k}{2}\right)!},$$

$$\log P\{\rho(\xi) = k\} = \log \frac{n!\cdot 2^{-n}}{\left(\frac{n+k}{2}\right)!\left(\frac{n-k}{2}\right)!} =$$

$$= \log n! - n\log 2 - \log\left(\frac{n+k}{2}\right)! - \log\left(\frac{n-k}{2}\right)! =$$

$$= n\log n - n - n\log 2 - \left(\frac{n+k}{2}\right)\log\frac{n+k}{2} + \frac{n+k}{2} -$$

$$- \frac{n-k}{2}\log\frac{n-k}{2} + \frac{n-k}{2} + o(n) =$$

$$= n\left(\frac{n+k}{2n}\log\frac{n}{n+k} + \frac{n-k}{2n}\log\frac{n}{n-k}\right) + o(n) = n g\left(\frac{|k|}{n}\right) + o(n).$$

При этом использован вывод из формулы Стирлинга

$$\log n! = n\log n - n + o(n).$$

Значит,

$$\log P\{\rho(\xi) = k\} = n g\left(\frac{|k|}{n}\right) + o(n).$$

Замечание 1. Функция $g(\lambda)$ на отрезке $[0;1]$ строго убывает и выпукла вверх, $g(0) = 0$, $g(1) = -\log 2$.

Теорема 1. *Пусть* $X_n = \{x \in X_n\colon \rho(x) \le \lambda n\}$, $m \sim c^n$, *где* $c > 1$. *Тогда справедливы следующие утверждения:*

1) *если* $1 < c < 2$ *и* α — *решение уравнения* $g(\alpha) = -\log c$, *то*

$$\lim_{n\to\infty} P\left\{M_m^{(n)} \subset X_n^{\lambda}\right\} = \begin{cases} 0 & \textit{при } \lambda < \alpha; \\ 1 & \textit{при } \lambda > \alpha; \end{cases}$$

2) *если* $c > 2$, *то* $\lim\limits_{n\to\infty} P\left\{M_m^{(n)} = X_n\right\} = 1.$

Доказательство. Пусть $\lambda > \alpha$, тогда

$$P\{|\rho(x)| > n\lambda\} = \sum_{|k| > n\lambda} C_n^{(n-k)/2}\cdot 2^{-n} \le$$

$$\le n \sup_{|k| > n\lambda} C_n^{(n-k)/2}\cdot 2^{-n} \le e^{n(g(\lambda)+\varepsilon)+o(n)} \quad \forall\, \varepsilon > 0.$$

Следовательно,

$$P\left\{X_n^\lambda \setminus M_m^{(n)} \neq 0\right\} \leq P\left\{\sup_{k \leq m} |\rho(\xi)| > n\lambda\right\} \leq$$

$$\leq me^{n(g(\lambda)+\varepsilon)+o(n)} = C^n e^{n(g(\lambda)+\varepsilon)+o(n)} =$$

$$= e^{n\log c} e^{n(g(\lambda)+\varepsilon)} = e^{n(g(\lambda)+\log c+\varepsilon)} = 0$$

для $n \to \infty$ и достаточно малого $\varepsilon > 0$, поскольку

$$g(\lambda) + \log c + \varepsilon = g(\lambda) - g(\alpha) + \varepsilon < 0$$

вследствие того, что $g(\lambda)$ и $g(\alpha)$ — функции с отрицательными значениями и $g(\lambda) < g(\alpha)$ ($g(\lambda)$ — убывающая функция). Значит, второе соотношение утверждения 1 доказано.

Пусть $\lambda < \alpha$, тогда $P\{|\rho(\xi)| > n\lambda\} \leq e^{n(g(\lambda)+\varepsilon)+o(n)}$ $\forall\ \varepsilon > 0$. Следовательно,

$$P\left\{M_m^{(n)} \subset X_n^\lambda\right\} = P\left\{\sup_{k \leq m} |\rho(\xi_k)| \leq \lambda n\right\} =$$

$$= (P\{|\rho(\xi)| \leq \lambda n\})^m = (1 - P\{|\rho(\xi)| > \lambda n\})^m =$$

$$= \left(1 - e^{n(g(\lambda)+\varepsilon)+o(n)}\right)^m \leq \exp\left\{-m \exp\left\{n(g(\lambda)+\varepsilon)+o(n)\right\}\right\} =$$

$$= \exp\left\{-\exp\left\{n(g(\lambda)+\log C+\varepsilon)\right\}\right\} = 0$$

для достаточно малого $\varepsilon > 0$ и $n \to \infty$, поскольку $g(\lambda) + \log C + \varepsilon = g(\lambda) - g(\alpha) + \varepsilon > 0$ вледствие того, что $g(\lambda) > g(\alpha)$, когда $\lambda < \alpha$. Первое утверждение доказано.

Для $x \in X_n$ положим

$$\nu_m^{(n)}(x) = \sum_{k \leq m} \mathbf{1}_{\xi_k = x}, \qquad \nu_m^{(n)} = \inf_{x \in X_n} \nu_m^{(n)}(x).$$

Покажем, что $\lim\limits_{n \to \infty} P\left\{\nu_m^{(n)} \geq 1\right\} = 1$ для всех l.

Пусть m изменяется от 1 до C^n. Будем трактовать его как время и обозначим через ζ_m число различных элементов множества $M_m^{(n)}$. Заметим, что $\zeta_{m+1} = \zeta_m + 1_{\zeta_{m+1} \in X_n \setminus M^{(n)}}$ и ζ_{m+1} не зависит ни от множества $M_m^{(n)}$, ни от ζ_m. Поэтому ζ_m есть цепь Маркова, для которой

$$P\{\zeta_{m+1} - \zeta_m = 0 / \zeta_m = k\} = \frac{k}{2^n},$$

$$P\{\zeta_{m+1} - \zeta_m = 1 / \zeta_m = k\} = \frac{2^n - k}{2^n}.$$

Эта цепь Маркова приводится в состояние k в момент времени τ_k, которое имеет геометрическое распределение

$$P\{\tau_k = l\} = \left(\frac{k}{2^n}\right)^{l-1} \frac{2^n - k}{2^n}, \qquad l = 1, 2, \ldots;$$

при этом величины τ_k независимы для разных k.

Обозначим $\theta^{(n)} = \tau_1 + \ldots + \tau_{2^n - 1}$, $m = \theta^{(n)}$ — время, необходимое для того, чтобы $M_m^{(n)} = X_n$. Значит, при $m = C^m$

$$P\left\{M_m^{(n)} = X_n\right\} = P\left\{\sum_{k=1}^{2^n-1} \tau_k \le C^n\right\} =$$

$$= 1 - P\left\{\sum_{k=1}^{2^n-1} \tau_k > C^n\right\} \le 1 - \frac{1}{C^n} E \sum_{k=1}^{2^n-1} \tau_k = 1,$$

поскольку $E\tau_k = \dfrac{2^n}{2^n - k}$,

$$\sum_{k=1}^{2^n-1} \frac{2^n}{2^n - k} = 2^n \sum_{k=1}^{2^n-1} \frac{1}{2^n - k} \sim 2^n \log 2^n,$$

а $\lim\limits_{n\to\infty} \dfrac{1}{C^n} 2^n \log 2^n = 0$ при $C > 2$. Итак, второе утверждение доказано.

Замечание 2. Пусть выполнены условия теоремы. Мы можем построить случайные времена $\theta_1^{(n)}, \theta_2^{(n)}, \ldots$ так, чтобы

$$\left\{\theta_1^{(n)} + \ldots + \theta_l^{(n)} + i,\ i = 1, \ldots, \theta_{l+1}^{(n)}\right\} = X_n.$$

Тогда при $m \ge \theta_1^{(n)} + \ldots + \theta_l^{(n)}$ будет $\nu_m^{(n)} \ge l$ и, следовательно,

$$P\left\{\nu_m^{(n)} \ge l\right\} \le 1 - P\left\{\theta_1^{(n)} + \ldots + \theta_l^{(n)}\right\} \le$$

$$\le 1 - \frac{l M\theta^{(n)}}{m}, \quad \text{а} \quad \frac{l M\theta^{(n)}}{m} \to 0 \quad \text{для любого } l.$$

Заметим, что можно выбрать $l = ln$, лишь бы $\dfrac{\ln M\theta^{(n)}}{m} \to 0$.

3. Будем теперь изучать взаимное расположение пар векторов. Для этого будем рассматривать функции $f(x, \tilde{x})\colon X_n^2 \to R$, удовлетворяющие условию $f(x, \tilde{x}) = f(\Pi x, \Pi \tilde{x})$. Можно показать, что такая функция есть скалярное произведение от $(x, \tilde{x})$.

Для характеристики множества $\left\{f(\xi_i, \xi_j)\colon i, j \in \overline{1, m}\right\}$ достаточно изучить множество $\left\{(\xi_i, \xi_j)\colon i, j \in \overline{1, m}\right\}$. Нам потребуются некоторые вспомогательные результаты.

Лемма 2. *Пусть* $x \in X_n$. *Тогда для* $|k| \le n$

$$P\{(\xi_1, x) = k\} = P\{\rho(\xi_1) = k\}.$$

Доказательство. Положим $\tilde{\xi} = \left(\tilde{\xi}^1, \ldots, \tilde{\xi}^n\right)$, где $\tilde{\xi}^i = \xi_1^i x^i$, $(x^1, \ldots, x^n) = X$. Очевидно, $P\left\{\tilde{\xi}^i = \pm 1\right\} = 1/2$, так что $\tilde{\xi}$ имеет такое же распределение, как и ξ_1. Так как $(\xi_1, x) = \rho(\xi)$, то

$$P\{(\xi_1, x) = k\} = P\{\rho(\tilde{\xi}) = k\} = P\{\rho(\xi_1) = k\}.$$

Следствие.

1) $P\{(\xi_1, \xi_2) = k / \xi_1\} = P\{\rho(\xi_2) = k\}$;

2) $P\{(\xi_1, \xi_2) = k, (\xi_1, \xi_3) = l\} = P\{\rho(\xi_1) = k\} P\{\rho(\xi_1) = l\}$.

Рассмотрим случайную величину

$$\eta_k = \sum_{i \ne j} 1_{(\xi_i, \xi_j) = k}.$$

Ниже вычислены ее первые два момента. Пусть $\alpha_k = P\{\rho(\xi_1) = k\}$. Очевидно,

$$M\eta_k = (m^2 - m)\alpha_k, \tag{3}$$

$$M\eta_k^2 = \sum_{i \neq j} \sum_{i_1 \neq j_1} P\{(\xi_i, \xi_j) = k, (\xi_{i_1}, \xi_{j_1}) = k\}.$$

Заметим, что из следствия 2) вытекает соотношение

$$P\{(\xi_i, \xi_j) = k, (\xi_{i_1}, \xi_{j_1}) = k\} = \begin{cases} \alpha_k, & \text{если } i = i_1,\ j = j_1 \text{ или } i = j_1,\ j = i_1; \\ \alpha_k^2 & \text{— во всех остальных случаях.} \end{cases}$$

Поэтому

$$M\eta_k^2 = 2(m^2 - m)\alpha_k + \left[(m^2 - m)^2 - 2(m^2 - m)\right]\alpha_k^2. \tag{4}$$

В частности,

$$D\eta_k = M\eta_k^2 - (M\eta_k)^2 = 2(m^2 - m)(\alpha_k - \alpha_k^2). \tag{5}$$

Теорема 2. *Пусть* $m \sim C^n$, $2\log C > -g(\lambda)$. *Тогда*

$$\min\left\{\eta_k; g\left(\frac{|k|}{n}\right) \geq g(\lambda)\right\} \to \infty$$

по вероятности, если $n \to \infty$.

Доказательство. Пусть $1 > \delta > 0$. Тогда

$$P\{\eta_k \leq (1 - \delta)M\eta_k\} \leq P\{|\eta_k - M\eta_k| > \delta M\eta_k\} \leq \frac{D\eta_k}{\delta^2 (M\eta_k)^2} =$$

$$= \frac{2(m^2 - m)(\alpha_k - \alpha_k^2)}{\delta^2 (m^2 - m)^2 \alpha_k^2} \leq \frac{2}{\delta^2 (m^2 - m)\alpha_k} \leq \frac{4}{\delta^2 m^2 \alpha_k}.$$

Следовательно,

$$P\left\{\min_{g\left(\frac{|k|}{n}\right) \geq g(\lambda)} \frac{\eta_k}{M\eta_k} \geq 1 - \delta\right\} \leq$$

$$\leq \sum_{|k| \leq g(\lambda) n} P\{\eta_k \leq (1 - \delta)M\eta_k\} \leq \sum_{|k| \leq g(\lambda) n} \frac{4}{\delta^2 m^2 \alpha_k}.$$

Заметим, что из формулы (1) следует соотношение

$$\frac{1}{m^2 \alpha_k} \sim \exp\left(-2n \log c - ng\left(\frac{|k|}{n}\right) + o(n)\right) \leq \exp\{-n\varepsilon + o(n)\},$$

где $0 < \varepsilon < 2\log C + g(\lambda)$; такое $\varepsilon > 0$ существует в силу условия теоремы. Значит,

$$P\left\{\min_{g\left(\frac{|k|}{n}\right) \geq g(\lambda)} \eta_k \geq (1 - \delta) \min M\eta_k\right\} \leq 2ne^{-n\varepsilon + o(n)} \to 0$$

при $g\left(\frac{|k|}{n}\right) \geq g(\lambda)$. Остается заметить, что в тех же обозначениях

$$M\eta_k \sim e^{n\varepsilon + o(n)} \to \infty.$$

Теорема доказана.

Замечание 3. Пусть $2\log c < -g(\bar{\lambda})$. Тогда

$$\left\{\sup \eta_k : g\left(\frac{|k|}{n}\right) \le g(\bar{\lambda})\right\} \to 0$$

по вероятности при $n \to \infty$.

Это вытекает из соотношения

$$M\eta_k = \exp\left\{n\left(2\log c + g\left(\frac{|k|}{n}\right)\right) + o(n)\right\} \le \exp\{-n\varepsilon + o(n)\},$$

где $2\log c + g(\bar{\lambda}) \le -\varepsilon \ \forall \varepsilon > 0$.

Теорема 3. *Пусть* $m \sim c^n$, $c > 1$ *и* λ_0 *есть решение уравнения* $2\log c + g(\lambda_0) = 0$. *Тогда* $\min\left\{\frac{1}{n}\left|\xi_i - \xi_j\right| : i \ne j\right\} \to 2(1 - \lambda_0)$ *сходится по вероятности.*

Доказательство. Имеем $\left|\xi_i - \xi_j\right| = 2n\left(1 - \frac{(\xi_i, \xi_j)}{n}\right)$. Пусть $\lambda_1 < \lambda_0 < \lambda_2$. Из теоремы 2 вытекает

$$\lim_{n \to \infty} P\left\{\max_{i \ne j} (\xi_i, \xi_j) > 2n(1 - \lambda_2)\right\} = 1,$$

а из замечания к теореме —

$$\lim_{n \to \infty} P\left\{\max_{i \ne j} (\xi_i, \xi_j) < 2n(1 - \lambda_1)\right\} = 0.$$

Теорема доказана.

1. *Скороход А. В., Степахно В. И.* О некоторых эмпирических характеристиках многомерного нормального распределения // Теория вероятностей и ее применения. – 1991. – **36**, № 2. – С. 386 – 395.
2. *Скороход А. В., Степахно В. И.* Об одном классе распределений и связанных с ним теоремах о больших уклонениях // Докл. АН УССР. Сер. А. – 1991. – № 7. – С. 34 – 38.
3. *Chernoff H.* A assured asymptotic efficiency for test of a hypothesis based on the sum of ofservations // Ann. Math. Statist. – 1952. – **23**. – P. 493 – 507.
4. *Ellis R. S.* Entnopy large deviations and statistical mananics. – New York etc.: Springer-Verlag, 1985. – P. 38 – 50.

Получено 19.12.96