УДК 519.21

В. И. Степахно, канд. физ.-мат. наук
(Ин-т математики АН Украины, Киев)

О множествах независимых векторов в пространствах неограниченно возрастающей размерности

Изучается область пространства, в которой сосредоточены независимые наблюдения случайных векторов, если размерность пространства и число наблюдений стремятся к бесконечности, а функция распределения компонент ξ_k^i наблюдаемых векторов не зависит от n и m и удовлетворяет условиям: 1) она непрерывна, симметрична: $F(x) = 1 - F(-x)$; 2) ее хвосты — медленно меняющиеся функции.

Вивчається область простору, де розташовані незалежні спостереження випадкових векторів, якщо розмірність простору і число спостережень необмежено зростають, а функція розподілу компонент ξ_i векторів, що спостерігаються, не залежить від n і m і задовольняє умови: 1) вона неперервна, симетрична: $F(x) = 1 - F(-x)$; 2) її хвости — повільно змінні функції.

1. Рассмотрим независимые случайные векторы $\xi_1, \dots, \xi_m$ в пространстве R^n, $\xi_i = (\xi_i^1, \dots, \xi_i^n)$. Предполагается, что ξ_i^k, $k = 1, \dots, n$, независимы и одинаково распределены. Изучаются некоторые геометрические свойства множества $\{\xi_1, \dots, \xi_m\} \subset R^n$ в предположении, что $n \to \infty$ и $m \to \infty$.

Такая задача рассматривалась для нормально распределенных величин в [1], для величин с устойчивым распределением в [2]. В настоящей работе предполагается, что функция распределения $F(x)$ величин ξ_i^k не зависит от n и m и удовлетворяет условиям; 1) она непрерывна, симметрична: $F(x) =$

$= 1 - F(-x)$; 2) ее хвосты — медленно меняющиеся функции [3], т. е. функция $L(x) = F(-x) + 1 - F(x)$ удовлетворяет условию для всех $k > 0$

$$\lim_{x\to\infty} \frac{L(kx)}{L(x)} = 1;$$

3) $L(x)$ абсолютно непрерывна и

$$L'(x) = -\frac{\Psi(x)}{x}, \tag{1}$$

где $\Psi(x)$ — непрерывная, убывающая к нулю, медленно меняющаяся функция.

Нас будет в основном интересовать случай, когда $m \sim a^n$ при $a > 1$. При этом будем исследовать характер расположения множества $M_m^{(n)} = \{\xi_1, \ldots, \xi_m\}$ в различных сферических слоях R^n.

2. Как показано в [1], при указанных m в нормальном случае $M_m^{(n)}$ лежит в некотором неслучайном сферическом слое, который при $a \to \infty$ превращается в шар, при этом векторы при соответствующей нормировке его плотно заполняют. Для устойчивых распределений [2] такая картина наблюдается для «нижнего» слоя, а в более высоких слоях векторы сосредоточены возле координатных гиперплоскостей и наконец возле координатных осей. В общем такая же картина будет и в рассматриваемом случае. Однако тот метод исследования, который использовался в [1, 2], основанный на явной асимптотике распределений $|\xi_k|$ здесь неприменим, поскольку такой асимптотики нет. Нам приходится обходиться лишь представлением (1) и свойствами медленно меняющихся функций.

3. Выделим векторы, образующие нижний слой в рассматриваемом случае.

Теорема 1. *Пусть c_n удовлетворяют условию $\Psi(c_n)\, n \to \infty$, а величины $\xi_1', \ldots, \xi_n'$ в R^n независимы и имеют распределение*

$$P\{\xi_r' < x\} = \frac{F(x) - F(-c_n)}{F(c_n) - F(-c_n)}, \quad -c_n \leqslant x \leqslant c_n.$$

Тогда $s_n' = \xi_1' + \ldots + \xi_n'$ асимптотически нормально.

Доказательство. Имеем $M\xi_n' = 0$,

$$[1 - L(c_n)]\, M(\xi_n')^2 = -\int_0^{c_n} x^2 L'(x)\, dx = \int_0^{c_n} x\Psi(x)\, dx \sim$$

$$\sim \Psi(c_n) \int_0^{c_n} x dx = \frac{1}{2} c_n^2 \Psi(c_n),$$

$$[1 - L(c_n)]\, M|\xi_n'|^3 = \int_0^{c_n} x^2 \Psi(x)\, dx = \frac{c_n^3}{4} \Psi(c_n).$$

Поэтому

$$\sum_{k=1}^{n} M|\xi_k'|^3 \Big/ \Big(\sum_{k=1}^{n} D\xi_k'\Big)^{3/2} \sim \frac{\frac{n}{3} c_n^3 \Psi(c_n)}{\left[\frac{n}{2} c_n^2 \Psi(c_n)\right]^{3/2}} = \frac{2^{3/2}}{3\sqrt{n\Psi(c_n)}} \to 0.$$

Значит, выполнено условие Ляпунова. Теорема доказана.

4. Разобьем теперь все величины $\{\xi_i^k,\ k \leqslant n,\ i \leqslant m\}$ на два множества. К первому отнесем те из них, для которых $|\xi_i^k| \leqslant c_n$; ко второму —

остальные. Будем называть первое множество множеством нормальных координат, а векторы ξ_i, у которых все координаты нормальны,— нормальными векторами, их множество обозначим через V_N. Число Card V_N случайно, оно имеет биномиальное распределение с параметрами $([1 - L(c_n)]^n, m)$, т. е.

$$P\{\operatorname{card} V_N = s\} = C_m^s [1 - L(c_n)]^{ns} \{1 - [1 - L(c_n)]^n\}^{m-s},$$

при этом $\dfrac{\operatorname{card} V_N}{M \operatorname{card} V_N} \to 1$ по вероятности при $n \to \infty$, а $M \operatorname{card} V_N = m \times \times [1 - L(c_n)]^n$. Векторы из V_N образуют нижний слой. Все они асимптотически нормальны и можно ожидать, что их поведение похоже на поведение $M_m^{(n)}$ в случае нормального распределения. Хотя здесь нужно проводить дополнительные исследования распределения $|\xi_i|$ и $\xi_i/|\xi_i|$. Они будут асимптотически такие же, как и нормально распределенные векторы. Но нам нужны оценки вероятностей событий, если эти вероятности имеют порядок $1/m$.

Обозначим множество всех координат $\{\xi_i^k,\ k \leqslant n,\ i \leqslant m\}$ через K, $\overline{N} = K \setminus N$. Величины из $\overline{N}$ будем называть «большими координатами». Сколько больших координат может иметь вектор ξ_i? Число больших координат также имеет биномиальное распределение с параметрами $(L(c_n,)\ n)$,

т. е. если $\bar{\nu}_i = \sum\limits_{k=1}^{n} I_{\{\xi_i^k \in \overline{N}\}}$, то $P\{\bar{\nu}_i = t\} = C_n^t L^t(c_n) [1 - L(c_n)]^{n-t}$;

среднее значение числа больших координат $M\nu_i = nL(c_n)$.

Лемма 1. *Пусть величины ν_n имеют биномиальное распределение с параметрами (p_n, n), $np_n \to \infty$, $p_n \to 0$. Тогда для всякого $s > 1$ и $\varepsilon > 0$*

$$P\{\nu_n > -sn \ln^{-1} p_n\} = O(e^{-n\gamma}),$$

где $\gamma = s - \varepsilon$.

Доказательство. При $k > np + 1$

$$P\{\nu_n = k + 1\} \geqslant P\{\nu_n = k\}.$$

Поэтому

$$P\{\nu_n \geqslant k\} \leqslant nP\{\nu_n = k\}, \quad \frac{\ln P\{\nu_n \geqslant k\}}{n} \leqslant \frac{\ln n}{n} + \frac{\ln P\{\nu_n = k\}}{n} =$$

$$= \frac{\ln n! - \ln k! - \ln (n-k)!}{n} + \frac{k \ln p_n}{n} + \frac{(n-k)}{n} \ln(1 - p_n).$$

Если $k \sim sn \ln^{-1} \dfrac{1}{p_n}$, то

$$\frac{k \ln p_n}{n} \to -s, \quad \frac{n-k}{n} \to 1, \quad \ln(1 - p_n) \to 0.$$

Таким образом,

$$\frac{\ln n! - \ln k! - \ln(n-k)!}{n} \sim \frac{n \ln n - k \ln k - (n-k)\ln(n-k)}{n} =$$

$$= \frac{k}{n} \ln \frac{k}{n} + \frac{n-k}{n} \ln\left(1 - \frac{k}{n}\right).$$

Так как $\dfrac{n-k}{n} \ln\left(1 - \dfrac{k}{n}\right) \to 0$, то и

$$\frac{k}{n} \ln \frac{k}{n} = \frac{s}{\ln \frac{1}{p_n}} \ln\left(\frac{s}{\ln \frac{1}{p_n}}\right) \to 0.$$

Значит, для всякого $\varepsilon > 0$ и для достаточно больших n $\ln P\{\nu_n \geqslant -sn \times \times \ln^{-1} p_n\} \leqslant -(s-\varepsilon)n$. Лемма доказана.

Следствие 1. *Каково бы ни было a ($m \sim a^n$) найдется такое $s > 0$, что*

$$mP\left\{\bar{\nu}_i \geqslant s \cdot n \ln^{-1} \frac{1}{p_n}\right\} \to 0.$$

Таким образом, с вероятностью, которая стремится к нулю, среди всех m векторов не будет таких, у которых число «больших» координат превышает $(sn)/\ln \dfrac{1}{L(c_n)}$. Количество больших координат ведет себя как $O\left(\ln^{-1} \dfrac{1}{L(c_n)}\right)$.

5. Рассмотрим теперь поведение длин векторов, содержащих «очень большие» координаты. (Они будут указаны ниже.) Оказывается, в этом случае можно получить асимптотику как для распределения длины, так и плотности этого распределения.

Лемма 2. *Пусть B_n таковы, что $n\,\Psi(B_n) \log n \to 0$. Тогда*

1) *для всех $r \leqslant n$*

$$F_{B_n}^{(r)}(x) = F_{B_n}^{r}(x)(1+o(1)), \quad 1 - F_{Bn}^{(r)}(x) = r(1 - F_{B_n}(x))(1+o(1)),$$

$o(1) \to 0$, если первый множитель стремится к 0;

2) *для $t \geqslant nB_n$ плотность $\varphi_{B_n}^{(r)}(t)$ распределения $F_{B_n}^{(r)}$ удовлетворяет соотношению*

$$\varphi_{B_n}^{(r)}(t) = \frac{r\psi(t)}{t}(1+o(1)),$$

где $o(1) \to 0$ при $n \to \infty$ равномерно по t.

Здесь

$$F_{B_n}(x) = \frac{L(B_n) - L(\sqrt{x})}{L(B_n)} = P\{\xi_1^2 < x / \xi_1^2 > B_n^2\}, \quad x \geqslant B_n^2,$$

$F_{B_n}^{(r)}(x) = F_{B_n}^{*r}(x)$ (*r — кратная свертка*).

Доказательство. Используем неравенства

$$P^r\left(\xi_1' < \sqrt{\frac{t}{r}}\right\} \leqslant P\left\{\sum_{k=1}^{r} (\xi_k')^2 < t\right\} \leqslant P^r\{\xi_1' \leqslant \sqrt{t}\},$$

$$1 - P^r\left\{\xi_1' < \sqrt{\frac{t}{r}}\right\} \leqslant P\left\{\sum_{k=1}^{r} (\xi_k')^2 > t\right\} \geqslant 1 - P^r\{\xi_1' < \sqrt{t}\},$$

где $(\xi_k')^2$ имеет распределение $F_{B_n}(x)$, причем они независимы при различных k. Для доказательства первого равенства в утверждении 1 леммы 2 достаточно доказать, что

$$r\frac{P\{\xi_1' < \sqrt{t}\} - P\{\xi_1' < \sqrt{t/r}\}}{P\{\xi_1' < t\}} = \frac{r}{L(B_n) - L(\sqrt{t})} \times$$

$$\times \int_{\sqrt{t/r}}^{\sqrt{t}} \frac{\Psi(x)}{x} dx \leqslant \frac{r}{L(B_n) - L(\sqrt{t})} \Psi(\sqrt{t}) \ln \sqrt{r}.$$

Если $P^r\left\{\xi_1' < \sqrt{\dfrac{t}{r}}\right\} \to 1$, то

$$r(1 - P\{\xi_1' < \sqrt{t/r}\}) \geqslant P\left\{\sum_{k=1}^{n} (\xi_1')^2 > t\right\} \geqslant r(1 - P\{\xi_1' < t\})\, P^{r-1}\{\xi_1' < t\},$$

$$1-P\{\xi_1'<\sqrt{t/r}\}=\frac{L(\sqrt{t/r})}{L(B_n)},\quad 1-P\{\xi_1'<\sqrt{t}\}=\frac{L\sqrt{t}}{L(B_n)}.$$

Поэтому

$$\frac{1-P\{\xi_1'<\sqrt{t/r}\}}{1-P\{\xi_1'<\sqrt{t}\}}=\frac{L(\sqrt{t/r})}{L(\sqrt{t})}=1+\frac{L(\sqrt{t/r})-L(\sqrt{t})}{L(\sqrt{t})}=$$

$$=1+\frac{1}{L(\sqrt{t})}\int\limits_{\sqrt{t/r}}^{\sqrt{t}}\frac{\Psi(x)}{x}\,dx=1+O\left(\frac{\Psi(\sqrt{t/r})}{L(\sqrt{t})}\right).$$

Лемма доказана.

Следствие 2. *Пусть h_n и t_n таковы, что* $\lim\limits_{n\to\infty}\sqrt[n]{t_n/h_n}=\bar{a}$. *Тогда в слое $\{x: x\in R^n,\ t_n<|x|^2<t_n+h_n\}$ будет при $m=a^n$ с вероятностью $\to 1$ бесконечное число векторов при $a>\bar{a}$ и ни одного при $a<\bar{a}$. Это вытекает из того, что это число эквивалентно*

$$a^n\varphi_{B_n}^{(r)}(t_n)\,h_n\sim r\psi(t_n)\left(\frac{a}{\bar{a}}\right)^n.$$

6. Рассмотрим теперь распределение векторов внутри такого слоя.

Пусть теперь A_n таково, что $\Psi(nA_n)/\Psi(A_n)\to 1$. Рассмотрим независимые случайные величины $\xi_{A_n}^{(i)}$, $i=1,2,\ldots,r_n$, с распределением $F_{A_n}=\dfrac{R(A_n)-R(x)}{R(A_n)}$. Пусть $r_n\to\infty$ выбрано так, что $\dfrac{r_n}{n}\ln\dfrac{1}{R(A_n)}\to\infty$. Тогда (см. предыдущую лемму), каково бы ни было $a>1$, для $m=O(a^n)$

$$P\left\{\sup_{k\leqslant m}\sum_{i=1}^{n}I_{\{|\xi_k^i|>A_n\}}>r_n\right\}\to 0\quad\text{при}\quad n\to\infty.$$

Это означает, что для достаточно больших n все векторы ξ_k имеют не более r_n координат, превышающих A_n.

Рассмотрим условное совместное распределение $\{\xi_{A_n}^{(i)},\ i=1,2,\ldots,r_n\}$ при условии, что $\Sigma(\xi_{A_n}^{(i)})^2=t_n$, где t_n выбрано так, что $\dfrac{t_n^2}{r_nA_n^2}\to\infty$. Обозначим его $g_{t_n}(x_1,\ldots,x_{r_n})$. Нас будут интересовать стационарные точки плотности на сфере $\sum\limits_{i=1}^{r_n}x_i^2=t_n$ (условные экстремумы), в частности абсолютный максимум и абсолютный минимум.

Лемма 3. *Стационарными точками плотности на сфере $\sum\limits_{i=1}^{r_n}x_i^2=t_n$ являются точки с координатами $\{\bar{x}_i,\ i=1,\ldots,r_n\}$, которые при некотором $1\leqslant s\leqslant r_n$ удовлетворяют условиям*

$$\bar{x}_{i_1}=\bar{x}_{i_2}=\ldots=\bar{x}_{i_s}=A_n,$$

$$\bar{x}_j=\sqrt{\frac{t-sA_n^2}{r-s}},\quad j\bar{\in}\{i_1,\ldots,i_s\},\quad j\leqslant r_n.$$

Значения плотности в стационарных точках

$$g(s,t_n)=\left[\frac{\Psi\left(\sqrt{\dfrac{t_n-sA_n^2}{r-s}}\right)}{\left(\dfrac{t_n-sA_n^2}{r-s}\right)^{1/2}}\right]^{r_n}\left[\frac{\Psi(A_n)\sqrt{\dfrac{t_n-sA_n^2}{r_n-s}}}{A_n\Psi\left(\sqrt{\dfrac{t_n-sA_n^2}{r_n-s}}\right)}\right]^s.$$

При этом максимальное и минимальное значения будут соответственно

$$g_{\max}(t_n) = \left(\frac{\Psi(A_n)}{A_n}\right)^{r_n-1} \frac{\Psi(\sqrt{t_n-(r_n-1)A_n^2})}{\sqrt{t_n-(r_n-1)A_n^2}} \sim$$

$$\sim \frac{[\Psi(A_n)]^{r_n-1}}{A_n^{r_n-1}} \frac{\Psi(\sqrt{t_n})}{\sqrt{t_n}}, \quad g_{\min}(t_n) = \left[\frac{\Psi(\sqrt{t_n/r_n})}{\sqrt{t_n/r_n}}\right]^{r_n}.$$

Пусть выполняются условия: а) $A_n^{r_n/n} \to a_0$, б) $t_n^{1,n} \to b_0^2$.

Теорема 2. *Пусть $a_0 > 1$, $b_0 > 1$. Тогда для $m \asymp a^n$, $a = 1$, в слое $t_n < \sum_1^{r_n} x_i^2 < t_n + h_n$ будут выполняться с вероятностью, стремящейся к 1, следующие условия:*

1) *при $a < a_0 b_0$ в нем нет наблюдения;*

2) *при $a_0 b_0^k < a < a_0 b_0^{k+1}$ в нем бесконечное число наблюдений и все они лежат в области V_k;*

3) *если $a = a_n$, $a_n \to \infty$ и $a_n/a_0 \in (b_0^{k_n}, b_0^{k_n+1})$, где $k_n < r_n$, то все наблюдения лежат в V_{k_n}.*

Здесь $V_k = \{x \in R^{r_n} : (x_1, \ldots, x_{r_n})^{1/n} \in (a_0 b_0, a_0 b_0^{k+1})\}$.

Доказательство. При выполнении условий а) и б) будет

$$\lim_{n\to\infty} \sqrt[n]{g_{\max}(t_n)} = \frac{1}{a_0 b_0}, \quad \lim_{n\to\infty} \sqrt[n]{\frac{g(s+1, t_n)}{g(s, t_n)}} = b_0.$$

Так как $g_{\max}(t_n) = g(r_n - 1, t_n)$, то $\sqrt[n]{g(r_n - k, t_n)} \sim \frac{1}{a_0 b_0^k}$. Заметим теперь, что число векторов в единице объема слоя равно $m g_{t_n}(x_1, \ldots, x_{r_n}) \asymp \asymp \left(\frac{a}{a_0 b_0^k}\right)^n$ и $\frac{a}{a_0 b_0^s} < 1$ при $x \overline{\in} V_k$.

Теорема 3. *Пусть $a_0 = 1$, $b_0 = 1$ и существует $\lim_{n\to\infty} t_n^{r_n/n} = b_1$. Тогда в слое $t_n < \sum_{i=1}^{r_n} x_i^2 \leqslant t_n + h_n$ выполняются с вероятностью, стремящейся к 1, следующие условия:*

1) *при $1 < a < b_1^\alpha$, $0 < \alpha < 1$ все наблюдения лежат в области V_k, где $k > (\alpha + \varepsilon) r_n$, каково бы ни было $\varepsilon > 0$;*

2) *при $a > b_1$ для всякого $\delta > 0$, сфера радиуса δt_n с центром на сфере радиуса t_n содержит бесконечное число наблюдений.*

Доказательство аналогично доказательству теоремы 2.

Замечание. В условиях теоремы 2 наблюдения с ростом a концентрируются сначала возле осей координат, потом возле точек, у которых несколько координат принимают максимальные значения. В условиях теоремы 2 сразу бесконечное число координат (определенная доля) принимают максимальные значения, а затем векторы «равномерно» заполняют r -мерные сечения слоя.

1. *Скороход А. В., Степахно В. И.* О некоторых эмпирических характеристиках многомерного нормального распределения // Теория вероятностей и ее применения.— 1991.— 41, вып. 1.— С. 117—124.
2. *Степахно В. И.* О больших выборках наблюдений случайных векторов большой размерности //Укр. мат. журн.— 1991.— 47, № 3.— С. 347—358.
3. *Секета Е.* Правильно меняющиеся функции.— М. : Наука, 1985.— 144 с.
4. *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов.— М. : Наука, 1965.— 656 с.

Получено 01.02.91