УДК 519.21

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

# О больших выборках наблюдений случайных векторов большой размерности

Изучается область пространства, в которой сосредоточены независимые наблюдения случайных векторов, если размерность пространства и число наблюдений стремятся к бесконечности, а компоненты наблюдаемых векторов независимы и одинаково распределены с устойчивым распределением с параметром $0 < \alpha < 2$.

Вивчається область у просторі, де зосереджені незалежні спостереження випадкових векторів, якщо розмірність простору і число спостережень необмежено зростають, а компоненти векторів, що спостерігаються, незалежні і однаково розподілені відповідно із стійким розподілом з параметром $0 < \alpha < 2$.

Пусть $\xi_1, \ldots, \xi_m$ — независимые одинаково распределенные векторы в $R^n$. Нас интересуют свойства случайного множества $M_m^{(n)} = \{\xi_1, \ldots, \xi_m\}$ при $n \to \infty$ и $m \to \infty$. Для случая, когда $\xi_i$ — гауссов вектор в $R^n$ со средним 0 и единичной корреляционной матрицей, некоторые результаты имеются в [1]. Здесь рассматриваются векторы $\xi_i$ с независимыми одинаково распределенными компонентами, имеющими устойчивое распределение с параметром $0 < \alpha < 2$. Оказывается, что в этом случае поведение $M_m^{(n)}$ существенно отличается от того, что установлено в [1].

1. Исследуем сначала величины $\rho_- = \inf\limits_{1 \leqslant k \leqslant m} |\xi_k|$. Положим

$$\rho_k^2 = |\xi_k|^2 = \sum_{l=1}^{n} (\xi_k^l)^2. \tag{1}$$

Лемма 1. *Величины*

$$\eta_n^{(k)} = \frac{1}{n^{2/\alpha}} \sum_{l=1}^{n} (\xi_k^l)^2 \tag{2}$$

*имеют предельное устойчивое распределение. Существуют такие $\gamma > 0$, $c_1$, $c_2$, что при $x > 0$*

$$P\{\eta_n^{(k)} < x\} = c_1 x^{-\gamma} \exp\{-c_2 x^{-\frac{\alpha}{2-\gamma}}\}[1 + \beta_n(x)], \tag{3}$$

*где $\beta_n(x) \to 0$ при $n \to \infty$, а $x \to 0$.*

Доказательство можно получить, используя асимптотическое представление для устойчивых законов распределения (см. [2, с. 121], теорема 2.5.2).

Пусть теперь $\rho_m^* < \rho_{m-1}^* < \ldots < \rho_1^*$ — вариационный ряд для величин $\rho_k$. Заметим, что $\rho_m^* = \rho_-$.

Теорема 1. *При $n \to \infty$ и $m \to \infty$*

$$\frac{(\ln m)^{\frac{2-\alpha}{2\alpha}} \rho_-}{c_2^{\frac{2-\alpha}{2\alpha}} n^{1/\alpha}} \to 1$$

*по вероятности.*

Доказательство. Имеем

$$P\left\{\rho_- < \frac{c_2^{\frac{2-\alpha}{2\alpha}} n^{1/\alpha}}{(\ln m)^{\frac{2-\alpha}{2\alpha}}}(1+\varepsilon)\right\} = P\left\{\rho_-^2 < \frac{c_2^{\frac{2-\alpha}{\alpha}} n^{\frac{2}{\alpha}}}{(\ln m)^{\frac{2-\alpha}{\alpha}}}(1+\varepsilon)^2\right\} =$$

$$= P\left\{\inf_{1 \leqslant k \leqslant m} \eta_n^{(k)} < \frac{c_2^{\frac{2-\alpha}{\alpha}}(1+\varepsilon)^2}{(\ln m)^{\frac{2-\alpha}{\alpha}}}\right\} = 1 - \left[1 - P\left\{\eta_n^{(1)} < \frac{c_2^{\frac{2-\alpha}{\alpha}}(1+\varepsilon)^2}{(\ln m)^{\frac{2-\alpha}{\alpha}}}\right\}\right]^m. \tag{4}$$

На основании формулы (3)

$$P\left\{\eta_n^{(1)} < \frac{c_2^{\frac{2-\alpha}{\alpha}}(1+\varepsilon)^2}{(\ln m)^{\frac{2-\alpha}{\alpha}}}\right\} = c_1\left[\frac{c_2^{\frac{2-\alpha}{\alpha}}(1+\varepsilon)^2}{(\ln m)^{\frac{2-\alpha}{\alpha}}}\right]^{-\gamma} m^{-(1+\varepsilon)^{\frac{2\alpha}{2-\alpha}}}.$$

Поэтому

$$\lim_{m\to\infty} mP\left\{\eta_n^{(1)} < \frac{c_2^{\frac{2-\alpha}{\alpha}}(1+\varepsilon)^2}{(\ln m)^{\frac{2-\alpha}{\alpha}}}\right\} = +\infty.$$

Значит, левая часть равенства (4) стремится к единице. Поменяв знак перед $\varepsilon$ с плюса на минус, убеждаемся, что

$$\lim_{m\to\infty} mP\left\{\eta_n^{(1)} < \frac{c_2^{\frac{2-\alpha}{\alpha}}(1+\varepsilon)^2}{(\ln m)^{\frac{2-\alpha}{\alpha}}}\right\} = 0$$

и, следовательно,

$$\lim_{m\to\infty} P\left\{\rho_- < \frac{c_2^{\frac{2-\alpha}{2\alpha}} n^{1/\alpha}}{(\ln m)^{\frac{2-\alpha}{2\alpha}}}(1-\varepsilon)\right\} = 0.$$

Теорема доказана.

Замечание 1. Если формально положить $\alpha = 2$, то получим $\rho_-/\sqrt{n} \to 1$ по вероятности. В [1] это доказано для нормальных векторов в предположении, что $\ln m = o(n)$.

Замечание 2. Если существует $\lim\limits_{n\to\infty} \frac{\ln m}{n^{\frac{2}{2-\alpha}}} \neq 0$, то $\rho_-$ сходится по вероятности к постоянной.

Замечание 3. Для того чтобы $\rho_- \to \infty$ по вероятности, необходимо и достаточно, чтобы

$$\lim_{n\to\infty} \frac{\ln m}{n^{\frac{2}{2-\alpha}}} = 0.$$

Рассмотрим теперь предельное распределение $\rho_k^*$, $k = 1, 2, \ldots, l$, где $l$ — любое натуральное число.

Т е о р е м а 2. *Совместное предельное распределение величин при* $n\to\infty$ *и* $m\to\infty$

$$\left\{\frac{1}{(nm)^{2/\alpha}}(\rho_k^*)^2,\quad k=1,2,\ldots,l\right\} \tag{5}$$

*сходится к совместному распределению величин* $\eta_1,\ldots,\eta_l$, *которые образуют однородную цепь Маркова с вероятностью перехода*

$$P\{0<y<\eta_{k+1}<x\mid\eta_k=x\}=c_3\left(\frac{1}{y^{2/\alpha}}-\frac{1}{x^{2/\alpha}}\right)\exp\left\{c_3\left(\frac{1}{x^{2/\alpha}}-\frac{1}{y^{2/\alpha}}\right)\right\}, \tag{6}$$

*причем распределение* $\eta_1$ *определяется равенством*

$$P\{\eta_1<x\}=\exp\{-c_3x^{-2/\alpha}\}, \tag{7}$$

*где*

$$c_3=\lim_{x\to\infty}x^{1/\alpha}P\{|\xi_1^{(1)}|>x\}. \tag{8}$$

Д о к а з а т е л ь с т в о. Заметим, что существование предела (8) вытекает из того, что $\xi_1^{(1)}$ имеет устойчивое распределение с показателем $\alpha$. Тогда и

$$\lim_{x\to\infty}x^{2/\alpha}P\{\eta_n^{(1)}>x\}=c_3.$$

Имеем

$$\lim_{\substack{n\to\infty\\ m\to\infty}}P\left\{\frac{1}{(nm)^{2/\alpha}}(\rho_1^*)^2<x\right\}=$$

$$=\lim_{\substack{n\to\infty\\ m\to\infty}}P\{\sup_{1\leqslant k\leqslant m}\eta_n^{(k)}<m^{2/\alpha}x\}=\lim_{\substack{n\to\infty\\ m\to\infty}}[1-P\{\eta_n^{(1)}>m^{2/\alpha}x\}]^m=$$

$$=\exp\{-\lim_{\substack{n\to\infty\\ m\to\infty}}mP\{\eta_n^{(1)}>m^{2/\alpha}x\}\}=e^{-c_3/x^{\alpha}}.$$

Этим установлена формула (7).

2. Рассмотрим теперь случайный процесс

$$\eta_n(t)=\frac{1}{m^{2/\alpha}}\sum_{k<tm^{2/\alpha}}\eta_n^{(k)}.$$

На основании теоремы 2 [3, с. 620] распределение процесса $\eta_n(t)$ в $\mathcal{D}_{[0,1]}$ (пространстве числовых функций без разрывов второго рода, непрерывных справа) слабо сходится в топологии $J$ к распределению процесса $\eta(t)$ — однородного от устойчивого процесса с независимыми приращениями, для которого

$$Me^{is\eta(t)}=\exp\{-tC(is)^{\alpha/2}\}, \tag{9}$$

где $C>0$ — некоторая постоянная. Поэтому совместное распределение величин $\rho_1^*,\ldots,\rho_l^*$, которые являются максимальными и следующими по величине скачками процесса $\eta_n(t)$, сходится к совместному распределению первых $l$ по величине скачков процесса $\eta(t)$, расположенных в порядке убывания. Заметим, что

$$\eta(t)=\int_0^t\int_0^\infty x\nu(t,dx).$$

Здесь $\nu(t,dx)$ — пуассоновская мера с независимыми значениями. При этом

$$M\nu(t,A)=ta\int_A\frac{dx}{x^{1+2/\alpha}},$$

где $a$ — некоторая постоянная. Если $\eta_1$ — максимальный скачок $\eta(t)$ на $[0, 1]$, то для $z > 0$

$$\{\eta_1 < z\} = \{\nu(1, [z, \infty)) = 0\}$$

и

$$P\{\eta_1 < z\} = \exp\{-M\nu(1, [z, \infty))\} = \exp\left\{-a\int\limits_z^\infty \frac{dx}{x^{1+\alpha/2}}\right\} =$$

$$= \exp\left\{-\frac{2a}{\alpha x^{\alpha/2}}\right\}.$$

Поэтому $2a/\alpha = c_3$. Если $\eta_1, \ldots, \eta_l$ — первые $l$ скачков процесса $\eta(t)$, расположенных в порядке убывания, то при $x_0 > x_1 > \ldots > x_{l-1} > x_l$

$$P\{x_0 > \eta_1 > x_1 > \eta_2 > \ldots > x_{l-1} > \eta_l > x_l\} =$$

$$= P\{\nu(1, [x_0, \infty)) = 0, \nu(1, [x_1, x_0]) = 1, \ldots, \nu(1, [x_l, x_{l-1}] = 1\} =$$

$$= \exp\left\{-\frac{c_3}{x_0^{\alpha/2}}\right\}\prod_{1\leqslant k\leqslant m}\left[\left(\frac{c_3}{x_k^{\alpha/2}} - \frac{c_3}{x_{k-1}^{\alpha/2}}\right)\exp\left\{\frac{c_3}{x_{k-1}^{\alpha/2}} - \frac{c_3}{x_k^{\alpha/2}}\right\}\right].$$

Отсюда вытекает, что $\eta_1, \ldots, \eta_l$ — цепь Маркова и формула (6) для вероятности перехода этой цепи.

Замечание 4. Утверждение теоремы остается справедливым и при фиксированном $n$ и $m \to \infty$.

Следствие 1. *Цепь Маркова* $\{\eta_1, \ldots, \eta_l, \ldots\}$ *удовлетворяет условию*

$$M(\eta_{k+1} \mid \eta_k) \leqslant \frac{\alpha}{2}\eta_k, \tag{10}$$

*следовательно,* $(\alpha/2)^{-k}\eta_k$ *есть супермартингал, и поэтому ограничена с вероятностью* 1.

Соотношение (10) можно получить, используя равенство (6):

$$M(\eta_{k+1} \mid \eta_k = x) = \int\limits_0^x c_3(y^{-2/\alpha} - x^{-2/\alpha})\exp c_3(x^{-2/\alpha} - y^{2/\alpha})\,dy =$$

$$= \frac{\alpha}{2}\int\limits_0^\infty \frac{t}{c_3}e^{-t}\left(\frac{t}{c_3} + x^{-2/\alpha}\right)^{-\alpha/2-1}dt \leqslant \frac{\alpha}{2}\int\limits_0^\infty e^{-t}\left(\frac{t}{c_3} + x^{-2/\alpha}\right)^{-\alpha/2}dt \leqslant$$

$$\leqslant \frac{\alpha}{2}\int\limits_0^\infty e^{-t}(x^{-2/\alpha})^{-\alpha/2}dt = \frac{\alpha}{2}x$$

(была сделана замена переменной интегрирования $t = c_3(y^{-2/\alpha} - x^{-2/\alpha})$).

Замечание 5. Так как $(\rho_1^*)^2$ — величина порядка $(nm)^{2/\alpha}$, то для всех $l$ и $\varepsilon > 0$ существует такое $b$, что

$$P\left\{(\rho_l^*)^2 < b(nm)^{2/\alpha}\left(\frac{\alpha}{2}\right)^l\right\} \geqslant 1 - \varepsilon \tag{11}$$

для достаточно больших $n$ и $m$. Более того, если $\alpha/2 < q < 1$, то неравенство (11) будет справедливым, если $\alpha/2$ заменить на $q$, для $l = O(\ln m)$.

3. Рассмотрим теперь распределение центральных проекций векторов $\xi_k$ на единичную сферу пространства $R^n - S_1^{(n)}(0)$, т. е. величин $\xi_k/|\xi_k|$. Такое распределение зависит от «слоя», в котором расположен вектор $\xi_k$, т. е. от $|\xi_k|$. Будем различать три слоя. «нижний» — этот те $\xi_k$, для которых $|\xi_k|n^{-1/\alpha} \to 0$ при $n \to \infty$, «верхний» — те, для которых $|\xi_k|n^{-1/\alpha} \to +\infty$ по вероятности, остальные образуют «средний» слой. Очевидно, подавляющее число векторов находится в среднем слое.

Теорема 3. *Пусть $\varepsilon_n \to 0$, тогда*

$$\lim_{n\to\infty} M\left[e^{i(z,\xi_1)} \mid |z_1| \leqslant \varepsilon_n n^{1/\alpha}\right] = e^{-\frac{1}{2}(z,z)}. \tag{12}$$

Доказательство. Запишем совместную характеристическую функцию $\frac{\xi_1}{\varepsilon_n n^{1/\alpha}}$ и $|\xi_1|^2$:

$$Me^{i\alpha_n \frac{(z,\xi_1)}{\varepsilon_n n^{1/\alpha}} + it|\xi_1^{(k)}|^2} = \prod_{k=1}^{n} Me^{i\alpha_n (z_k, \xi_1^{(k)}/\varepsilon_n n^{1/\alpha}) + it(\xi_1(k))^2}.$$

Условная характеристическая функция под знаком предела в левой части (12) может быть представлена в виде

$$\frac{\displaystyle\int \prod_{k=1}^{n} M \exp\left\{i\left(z_k \frac{\xi_1^{(k)}}{\varepsilon_n n^{1/\alpha}}\right) + it(\xi_1^{(k)})^2\right\} \frac{\sin \varepsilon_n n^{1/\alpha} t}{t}\, dt}{\displaystyle\int \prod_{k=1}^{n} M\exp\{it(\xi_1^{(k)})^2\} \frac{\sin \varepsilon_n n^{1/\alpha} t}{t}\, dt}.$$

Знаменатель этой дроби на основании леммы 1 эквивалентен

$$c_1 \varepsilon_n^{-\gamma} \exp\{-c_2 \varepsilon_n^{\overline{2-\alpha}}\},$$

$$M\exp\left\{iz_k \frac{\xi_1^{(k)}}{\varepsilon_n n^{1/\alpha}} + it|\xi_1^{(k)}|^2\right\} = M\exp\left\{iz_k \frac{\xi_1^{(k)}}{\varepsilon_n n^{1/\alpha}} + \sqrt{2it}\,\theta_k \xi_1^{(k)}\right\} =$$

$$= Mf\left(\frac{z_k}{\varepsilon_n n^{1/\alpha}} + \sqrt{-2it}\,\theta_k\right),$$

где $f$ — характеристическая функция устойчивого закона, $\theta_k$ — нормальная величина $M\theta_k = 0$, $D\theta_k = 1$. Таким образом

$$M\left[e^{i\left(z, \frac{\xi_1}{\varepsilon_n n^{1/\alpha}}\right)} \mid |\xi_1| \leqslant \varepsilon_n n^{1/\alpha}\right] \sim$$

$$\sim c_1^{-1} \varepsilon_n^{\gamma} \exp\{c_2 \varepsilon_n^{\overline{2-\alpha}}\}\, M \int \prod_{k=1}^{n} f\left(\frac{z_k}{\varepsilon_n n^{1/\alpha}} + \sqrt{-2it}\,\theta_k\right) \frac{\sin \varepsilon_n n^{1/\alpha} t}{t}\, dt.$$

Использовав выражение для $\ln f$ и сделав замену в интеграле $t' = \varepsilon_n n^{1/\alpha} t$, можем получить требуемое, разложив $\ln f$ в ряд Тейлора в точке $\sqrt{-2it}\,\theta_k$.

Следствие 2. *Векторы в нижнем слое ведут себя так же, как и в случае нормально распределенных векторов.*

Поведение векторов в верхнем слое характеризует следующая теорема.

Теорема 4. *Пусть $\alpha_n \to \infty$, тогда для всякого $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left\{\frac{|\xi_1|}{\max\limits_{1\leqslant i\leqslant n} |\xi_1^{(i)}|} > 1 - \varepsilon \mid |\xi_1| > \alpha_n n^{1/\alpha}\right\} = 1.$$

Доказательство. Рассмотрим совокупность одинаково распределенных независимых случайных величин $\{|\xi_k^{(i)}|, 1 \leqslant k \leqslant m, 1 \leqslant i \leqslant n\}$. Запишем соответствующий вариационный ряд $\zeta_1^* > \xi_2^* > \ldots > \zeta_{nm}^*$. Тогда для любых $l$ $\zeta_1^*, \zeta_2^*, \ldots, \zeta_l^*$ по теореме 2 будут иметь такое же совместное предельное распределение, как и $\rho_1^*, \ldots, \rho_l^*$. Отсюда вытекает, что $\zeta_k^* / \rho_k^* \to 1$ по вероятности. Очевидно, это возможно лишь в случае, когда $\zeta_1^* = \xi_{r_1}^{(i_1)}$, $\rho_k^* = |\xi_{r_1}|, \ldots, \zeta_l^* = \xi_{r_l}^{(i_l)}, \rho_l^* = |\xi_{r_l}|$, все числа $r_1, r_2, \ldots, r_l$ различны. Это будет выполняться, когда $\rho_l^*/n^{1/\alpha} \to \infty$ по вероятности. Из соотношения $\xi_k^*/\rho_k^* \to 1$ и последнего замечания вытекает доказательство теоремы

Следствие 3. *Направление векторов в верхнем слое асимптотически совпадает с направлением одной из координатных осей.*

Теорема 5. *Пусть* $\xi = \{\xi^{(1)}, \ldots, \xi^{(n)}\}$ — *случайный вектор в* $R^n$, *имеющий такое же распределение, как и* $\xi_1$. *Пусть, кроме того,* $|\xi^{(1)*}| > \ldots > |\xi^{(n)*}|$ — *вариационный ряд, построенный по* $\{\xi^{(k)}, k = 1, 2, \ldots\}$. *Тогда для любых* $\varepsilon > 0$ *и* $\delta > 0$ *можно указать такое* $l$, *что*

$$\varlimsup_{n\to\infty} P\left\{\sum_{k=l+1}^{n} |\xi^{(k)*}|^2 > \delta \sum_{k=1}^{n} |\xi^{(k)*}|^2\right\} < \varepsilon.$$

Доказательство. Воспользуемся уже упоминавшимся результатом о сходимости совместного распределения величин

$$\frac{1}{n^{1/\alpha}} \sum_{k=1}^{n} |\xi^{(k)*}|^2, \quad \frac{1}{n^{1/\alpha}} |\xi^{(1)*}|^2, \ldots, \frac{1}{n^{1/\alpha}} |\xi^{(l)*}|^2$$

к совместному распределению величин $\eta^{(1)}, \eta_1, \ldots, \eta_l$, где $\eta(t)$ — процесс с характеристической функцией (9), $\eta_k$ — его скачки. Доказательство теоремы вытекает из равенства

$$\eta_1 = \sum_{k=1}^{\infty} \eta_k.$$

Следствие 4. *Направления векторов в среднем слое асимптотически приближаются к направлениям конечномерных плоскостей (фиксированной размерности).*


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. *Золотарев В. М.* Одномерные устойчивые распределения.—М.: Наука, 1983.— 304 с.
3. *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов.— М. : Наука, 1965.— 656 с.



Получено 20.09.90