Теор. ймовірност. та матем. статист.
Вип. 61, 1999, стор. 126–130

Teor. Imovirnost. ta Matem. Statyst.
No. 61, 1999, pp. 126–130

ПОБУДОВА КРИТЕРІЮ ПЕРЕВІРКИ ГІПОТЕЗИ ПРО РОЗПОДІЛ ВИПАДКОВИХ ВЕКТОРІВ

УДК 519.21

В. МЕРГЕЛЬ

Резюме. В статті представлені ентропійні критерії перевірки гіпотез про розподіли Пірсон II та Пірсон VII випадкового вектора, а також про розподіл загальної помилки випадкової величини.

1. Вступ

Нехай $\vec{\xi}$ — випадковий вектор з m-вимірного евклідового простору з щільністю ймовірності $f(\vec{x})$. Задамо в цьому просторі відстань

$$\rho(\vec{x},\vec{y}) = \sqrt{\sum_{j=1}^{m}(x_j - y_j)^2},$$

де $\vec{x}^t = (x_1,\ldots,x_m) \in \mathbf{R}^m$ та $\vec{y}^t = (y_1,\ldots,y_m) \in \mathbf{R}^m$.

Нехай $v(\vec{y},r) = \{\vec{x} \in \mathbf{R}^m : \rho(\vec{x},\vec{y}) < r\}$ — відкрита куля радіуса $r > 0$ з центром в $\vec{y} \in \mathbf{R}^m$, і нехай її об'єм $|v(\vec{y},r)| = r^m c_1(m)$, де

$$c_1(m) = \frac{2\pi^{m/2}}{m\Gamma(m/2)}.$$

Ентропія випадкового вектора $\vec{\xi}$ по означенню дорівнює

$$H = -\int_{\mathbf{R}^m} f(\vec{x}) \ln f(\vec{x})\, d\vec{x}.$$

Нехай ми маємо незалежні однаково розподілені спостереження випадкового вектора $\vec{\xi}$: $\vec{X}_1,\vec{X}_2,\ldots,\vec{X}_N$, $N > 1$, і для фіксованого спостереження $\vec{X}_i$, $i \in \{1,\ldots,N\}$ та фіксованого $k \in \{1,\ldots,N-1\}$ визначимо випадкові величини $\rho_{i,k}$ таким чином:

$$\begin{aligned}
\rho_{i,1} &:= \min\left\{\rho\left(\vec{X}_i,\vec{X}_j\right), j \in \{1,\ldots,N\}\setminus\{i\}\right\} = \rho\left(\vec{X}_i,\vec{X}_{j_1}\right),\\
\rho_{i,2} &:= \min\left\{\rho\left(\vec{X}_i,\vec{X}_j\right), j \in \{1,\ldots,N\}\setminus\{i,j_1\}\right\} = \rho\left(\vec{X}_i,\vec{X}_{j_2}\right),\\
&\ldots\\
\rho_{i,k} &:= \min\left\{\rho\left(\vec{X}_i,\vec{X}_j\right), j \in \{1,\ldots,N\}\setminus\{i,j_1,\ldots,j_{k-1}\}\right\} = \rho\left(\vec{X}_i,\vec{X}_{j_k}\right),\\
&\ldots\\
\rho_{i,N-1} &:= \max\left\{\rho\left(\vec{X}_i,\vec{X}_j\right), j \in \{1,\ldots,N\}\setminus\{i\}\right\} = \rho\left(\vec{X}_i,\vec{X}_{j_{N-1}}\right).
\end{aligned}$$

Тоді з імовірністю 1 маємо $\rho(\vec{X}_i,\vec{X}_{j_1}) < \rho(\vec{X}_i,\vec{X}_{j_2}) < \cdots < \rho(\vec{X}_i,\vec{X}_{j_{N-1}})$.

1991 *AMS Mathematics Subject Classification.* Primary 62B10, 62F03.

Для фіксованого $k \in \{1, \dots, N-1\}$ покладемо

$$\overline{\rho}_k = \left\{ \prod_{i=1}^{N} \rho_{i,k} \right\}^{1/N}.$$

В роботі [2] показано, що для класу функцій щільності, що задовольняють умови

$$\int_{\mathbf{R}^m} |\ln f(\vec{x})|^{1+\varepsilon} f(\vec{x})\, d\vec{x} < \infty \tag{1}$$

та

$$\int_{\mathbf{R}^m} |\ln f(\vec{x})|^{2+\varepsilon} f(\vec{x})\, d\vec{x} < \infty \tag{2}$$

статистична оцінка ентропії H:

$$H_{k,N} = m \ln \overline{\rho}_k + \ln(N-1) - \psi(k) + \ln c_1(m),$$

(де $\psi(k) = (\frac{d}{dt}) \ln \Gamma(t)$ — дігамма функція) буде асимптотично незсуненою та конзистентною при $N \to \infty$ для будь-якого k.

Властивості оцінок такого типу були вперше досліджені в [5]. Пізніше умови, яким повинна задовільняти функція щільності розподілу випадкового вектора (див. [2]) були спрощені до умов (1) та (2), завдяки чому виявилося можливим значно розширити клас імовірносних розподілів і використати результати робіт [7] та [8] для побудови ентропійних критеріїв для розподілів Пірсон II та Пірсон VII та розподілу загальної помилки, які будуть представлені в наступних розділах.

2. Етропійний критерій перевірки гіпотези про розподіли Пірсон II та Пірсон VII

Нехай $\vec{X}^t = (X_1, \dots, X_m)$ — m-вимірний випадковий вектор. Розглянемо класс m-вимірних імовірносних функцій щільності $F = \{f(\vec{x})\colon \mathsf{E}_f[T_i(\vec{X})] = \alpha_i, i = 0, \dots, k\}$, де $T_i, i = 0, \dots, k$, абсолютно диференційовані функції по f та $T_0(\vec{X}) = \alpha_0 = 1$. Вважаємо надалі, що значення α_i та вигляд T_i, $i = 0, \dots, k$, відомі.

Принцип максимуму ентропії дає метод підбору невідомої щільності імовірності випадкового вектора $\vec{X}$, таким чином щоб ентропія була максимальна при обмеженнях, що визначають клас F.

Сформулюємо основні результати для характеризації розподілів Пірсон II та Пірсон VII згідно принципу максимума ентропії. Технічно, проблема полягає в тому, щоб означити класс щільностей імовірності так, щоб на розподілі Пірсон II або Пірсон VII з цього класу ентропія $H(f)$ досягала максимуму.

Як показано в [8] для розподілу Пірсон II з щільністю імовірності

$$f(\vec{x}) = \frac{\Gamma(\frac{m}{2} + p + 1)}{\pi^{m/2}\Gamma(p+1)} \left(1 - \vec{x}^t x\right)^p,$$

де $\vec{x} \in S = \{\vec{x} \in \mathbf{R}^m\colon \vec{x}^t x \le 1\}$ та $p > -1$, такий класс F визначається за допомогою такого обмеження на функцію щільності f:

$$\mathsf{E}_f \left[\ln\left(1 - \vec{X}^t \vec{X}\right)\right] = w_1\left(p+1; \frac{m}{2}\right),$$

де $w_1(p+1; m/2) = \Psi(p+1) - \Psi(m/2 + p + 1)$, а $\Psi(t) = (\frac{d}{dt}) \ln \Gamma(t)$.

Відповідно для розподілу Пірсон VII з щільністю імовірності

$$f(\vec{x}) = \frac{\Gamma(p)}{\pi^{m/2}\Gamma(p - m/2)} \left(1 + \vec{x}^t x\right)^p,$$

де $\vec{x} \in \mathbf{R}^m$ та $p > m/2$ обмеження на функцію щільності f буде таке:

$$\mathsf{E}_f \left[\ln \left(1 + \vec{X}^t X \right) \right] = w_2 \left(p; \frac{m}{2} \right),$$

де $w_2(p; m/2) = \Psi(p) - \Psi(p - m/2)$. Позначимо так визначений для розподілу Пірсон VII клас через F_1.

Тепер нехай ми маємо спостереження випадкового вектора $\vec{X}$: $\vec{X}_1, \ldots, \vec{X}_N$. Оцінимо за допомогою цієї виборки параметр p розподілів Пірсон II та VII.

Для оцінки параметра розподілу Пірсон II використаємо метод моментів для функції $g(\vec{x}) = x_1^2 + \cdots + x_m^2 = \vec{x}^t \vec{x}$, тоді

$$\mathsf{E}\, g(\vec{x}) = \frac{\Gamma(m/2 + p + 1)}{\pi^{m/2} \Gamma(p+1)} \int_S g(\vec{x}) \left(1 - \vec{x}^t \vec{x}\right)^p \, d\vec{x}.$$

Переходячи до сферичної системи координат і використовуючи те, що

$$\int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \sin^{m-2} \theta_1 \sin^{m-3} \theta_2 \ldots \sin \theta_{m-2} \, d\theta_1 \ldots \, d\theta_{p-2} \, d\theta_{p-1} = \frac{2\pi^{m/2}}{\Gamma(m/2)},$$

отримаємо

$$\mathsf{E}\, g(\vec{x}) = \frac{2\Gamma(m/2 + p + 1)}{\Gamma(m/2)\Gamma(p+1)} \int_0^1 \left(1 - r^2\right)^p r^{m+1} \, dr = \frac{m/2}{m/2 + p + 1}.$$

Звідси, поклавши $\overline{g} = N^{-1} \sum_{i=1}^N g(\vec{X}_i)$, отримаємо оцінку параметра p розподілу Пірсон II:

$$\overline{p} = \frac{m}{2\overline{g}} - \frac{m}{2} - 1.$$

Ця оцінка є конзистентною оцінкою параметра p (див. [1] с. 70).

Крім того, можна виразити (див. [8]) ентропію для розподілу Пірсон II

$$H(f) = \ln \left(\frac{\Gamma(m/2 + p + 1)}{\pi^m/2} \Gamma(p+1) \right) + p w_1 \left(p + 1; \frac{m}{2} \right).$$

Аналогічно отримаємо оцінку параметра p для розподілу Пірсон VII:

$$\mathsf{E}\, g(\vec{x}) = \frac{\Gamma(p)}{\pi^{m/2} \Gamma(p - m/2)} \int_{\mathbf{R}^m} g(\vec{x}) \left(1 + \vec{x}^t \vec{x}\right)^{-p} \, d\vec{x} = \frac{m/2}{p - 1 - m/2}.$$

Звідси отримаємо конзистентну оцінку параметра p розподілу Пірсон VII:

$$\widetilde{p} = \frac{m}{2\overline{g}} + \frac{m}{2} + 1.$$

Ентропія для розподілу Пірсон VII буде

$$H(f) = \ln \left(\frac{\Gamma(p)}{\pi^{m/2} \Gamma(p - m/2)} \right) - p w_2 \left(p; \frac{m}{2} \right).$$

Використовуючи все викладене вище, можна тепер сформулювати критерії перевірки гіпотез про розподіли Пірсона II та VII.

Теорема 1 (Ентропійний критерій перевірки гіпотези про розподіл Пірсона II). *Нехай $\vec{X}_1, \ldots, \vec{X}_N$ — спостереження випадкового вектора $\vec{X}$. Нехай гіпотеза H_0: $\vec{X}_1, \ldots, \vec{X}_N$ — вибірка з розподілом Пірсон II, а альтернатива H_1: $\vec{X}_1, \ldots, \vec{X}_N$ — вибірка з деяким іншим розподілом з класу F. Тоді гіпотезу H_0 треба прийняти лише у випадку, якщо*

$$H_{k,N} - \ln\left(\frac{\Gamma(m/2+\overline{p}+1)}{\pi^{m/2}\Gamma(\overline{p}+1)}\right) + \overline{p}w_1\left(\overline{p}+1; \frac{m}{2}\right) \to 0$$

по ймовірності при $N \to \infty$.

Теорема 2 (Ентропійний критерій перевірки гіпотези про розподіл Пірсона VII). *Нехай $\vec{X}_1, \ldots, \vec{X}_N$ — спостереження випадкового вектора $\vec{X}$. І нехай гіпотеза H_0: $\vec{X}_1, \ldots, \vec{X}_N$ — вибірка з розподілом Пірсон VII, а альтернатива H_1: $\vec{X}_1, \ldots, \vec{X}_N$ — вибірка з деяким іншим розподілом з класу F_1. Тоді гіпотезу H_0 треба прийняти лише у випадку, якщо*

$$H_{k,N} - \ln\left(\frac{\Gamma(\widetilde{p})}{\pi^{m/2}\Gamma(\widetilde{p}-m/2)}\right) - \widetilde{p}w_2\left(\widetilde{p}; \frac{m}{2}\right) \to 0$$

по ймовірності при $N \to \infty$.

3. Ентропійний критерій перевірки гіпотези про розподіл загальної помилки

Для одновимірного випадку оцінка ентропії $H_{k,N}$ має вигляд

$$H_{k,N} = \ln \bar{\rho}_k + \ln 2(N-1) - \psi(k),$$

де $\psi(k) = (\frac{d}{dt}) \ln \Gamma(t)$ — дігамма функція.

В роботі [6] була доведена наступна теорема:

Теорема 3. *Нехай K клас щільностей імовірності $f(x)$, $x \in \mathbf{R}^1$, заданих на носії* $\operatorname{supp}\{f\} = (-\infty, \infty)$, *які задовольняють умови (1) та (2). Відмітимо, що щільність*

$$f^*(x) = \frac{1}{2\sigma p^{1/p-1}\Gamma(1/p)} \exp\left(-\frac{1}{p}\left|\frac{x-\mu}{\sigma}\right|^p\right); \qquad p, \sigma > 0, \tag{3}$$

розподілу загальної помилки GED належить до цього класу. Тоді серед всіх щільностей f з класу K таких, що

$$\int_{-\infty}^{\infty} |x-\mu|^p f(x)\, dx = \sigma^p, \tag{4}$$

максимум ентропії досягається на розподілі загальної помилки.

Нехай тепер $X_1, X_2, \ldots X_N$, $N \geq 2$, незалежна вибірка випадкової величини з розподілом з класу K. Нехай оцінки $\bar{X}_N^{(p)}$ та $S_N^{(p)}$ задовольняють наступні рівняння:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\operatorname{sign}\left(X_i - \bar{X}_N^{(p)}\right)\right\}\left|X_i - \bar{X}_N^{(p)}\right|^{p-1} = 0$$

та

$$S_N^{(p)} = \left\{\frac{1}{N}\sum_{i=1}^{N}\left|X_i - \bar{X}_N^{(p)}\right|^p\right\}^{1/p}$$

для будь-якого $N \geq 2$ і для $p \neq 1$. Тоді (див. [7]) вибіркове середнє $\bar{X}_N^{(p)}$ та вибіркова дисперсія $S_N^{(p)}$ існують для будь-якого фіксованого $p > 0$ і збігаються відповідно з оцінками максимальної вірогідності параметрів μ та σ представленого в теоремі 3 розподілу загальної помилки. При гіпотезі H_0: $X_1, X_2, \dots X_N$ — вибірка з розподілу загальної помилки — ми маємо, що для кожного фіксованого $k \in \{1, \dots, N-1\}$

$$\zeta_N = \frac{1}{p} + \log\left(2S_N^{(p)} p^{1/p-1} \Gamma\left(\frac{1}{p}\right)\right) - H_{k,N} \to 0$$

по ймовірності при $N \to \infty$. Тоді ж як при альтернативі H_1: $X_1, X_2, \dots X_N$ — вибірка з деякого іншого розподілу з класу K — ми отримаємо, що $\zeta_N \to \text{const} \neq 0$ при $N \to \infty$ по ймовірності для кожного фіксованого $k \in \{1, 2, \dots, N-1\}$.

Література

1. М. Леоненко, Ю. Мішура, В. Пархоменко, М. Ядренко, *Теоретико-імовірнісні та статистичні методи в економетриці та фінансовій математиці* (1995), "Інформтехніка", Київ.
2. M. Goria, N. Leonenko, and V. Mergel, *Statistical estimators of entropy of random vector and its applications in testing of statistical hypotheses*, Australian and New Zealand Statistical Journal (подано до друку).
3. A. M. Kagan, Yu. V. Linnik, and C. B. Rao, *Characterization problems in mathematical statistics* (1973), Wiley, New York.
4. J. N. Kapur, *Maximum entropy models in science and engineering* (1989), Wiley, New York.
5. L. F. Kosachenko and N. N. Leonenko, *On statistical estimation of entropy of random vector*, Prob. Inf. Trans. **2** (1987), № 23, 9–16.
6. V. Mergel, *Maximum entropy principle in testing statistical hypotheses of goodness of fit*, Theory of Random Processes (подано до друку).
7. T. Taguchi, *On a generalization of Gaussian distribution*, Ann. Inst. Statist. Math. **30A** (1978), 211–242.
8. K. Zografos, *On maximum entropy characterization of Pearson's type II and VII Multivariate Distributions*, Jour. Multivariate An. **71** (1999), 67–75.

01033 Київ, Володимирська, 64, Київський університет ім. Тараса Шевченка, механіко-математичний факультет, кафедра теорії ймовірностей та математичної статистики
Електронна адреса: vik@g-auto.pul.kiev.ua

Надійшла 24/01/2000