Журнал обчислювальної та прикладної математики | 2012, №4(110) | Journal of Computational & Applied Mathematics

УДК 519.71

НЕПАРАМЕТРИЧНІ МЕТОДИ СТАТИСТИЧНОГО РОЗПІЗНАВАННЯ ОБРАЗІВ (ЗА РОБОТАМИ НАУКОВОЇ ШКОЛИ Ю. І. ПЕТУНІНА)

Д. А. Клюшин

Резюме. В статті описуються результати, отримані науковою школою Ю. І. Петуніна в галузі статистичного розпізнавання образів: нерівність Височанського-Петуніна, p-статистика, критерії перевірки гіпотези про рівність розподілів, критерії перевірки гіпотези про рівність невідомих ймовірностей, критерії унімодальності вибірки.

1. Правило трьох сігм

Непараметричні критерії для перевірки гіпотези про рівність розподілів в статистичному розпізнаванні образів базуються на використанні довірчого інтервалу (a, b), що містить значення x генеральної сукупності G із гарантованою ймовірністю, тобто ймовірністю, що перевищує заданий довірчий рівень. Зокрема, в медичних та біологічних дослідженнях вважається стандартним довірчий рівень, що дорівнює 0,95:

$$p\left(x \in (a, b)\right) \geq 0,95.$$

Нерівність Чебишова-Б'єнаме стверджує, що для довільної випадкової величини x має місце співвідношення

$$p\left(\left|x - m\left(x\right)\right| \geq \lambda\right) \leq \frac{\sigma\left(x\right)}{\lambda^2},$$

де $m\left(x\right)$ — математичне сподівання випадкової величини x, $\sigma^2\left(x\right)$ — її дисперсія і $\lambda > 0$. З іншого боку нерівність Гаусса оцінює відхилення випадкової величини від моди: $p\left(\left|x - M\left(x\right)\right| \geq k\tau\right) \leq \frac{4}{9k^2}$, де $\tau^2 = \sigma\left(x\right) + \left(m(x) - M\left(x\right)\right)^2 > 0$ — другий момент розподілу $F(x)$ відносно моди M. За допомогою цієї нерівності було отримано строге доведення правила трьох сігм для випадкових величин із унімодальним розподілом.

Теорема 1. *(Нерівність Височанського-Петуніна [1], 1979). При всіх $k > 0$ для довільної випадкової величини x, що має унімодальний розподіл і скінченну дисперсію $\sigma^2\left(x\right) > 0$, має місце нерівність*

$$p\left(\left|x - m(x)\right| \geq k\sigma\left(x\right)\right) \leq \max\left\{\frac{4}{9k^2}, \frac{4 - k^2}{3k^2}\right\}.$$

Наслідок 1. *Для довільної випадкової величини x із унімодальним розподілом і скінченною дисперсією $\sigma^2(x) > 0$ при всіх $k \geq \sqrt{\frac{8}{3}}$ виконується нерівність*

$$p\left(|x - m(x)| \geq k\sigma(x)\right) \leq \frac{4}{9k^2}.$$

Зокрема, для $k = 3$ маємо $p\left(|x - m(x)| \geq 3\sigma(x)\right) \leq \frac{4}{81} \approx 0,049$.

Із нерівності Височанського-Петуніна випливає, що за довірчий інтервал, що містить основну розподілену масу значень генеральної сукупності G, можна взяти інтервал $\left(m\left(x\right) - 3\sigma\left(x\right), m\left(x\right) + 3\sigma\left(x\right)\right)$.

Інтервали такого типу використовуються для побудови непараметричних статистичних критеріїв для перевірки гіпотез про рівність розподілів. Ці критерії повинні мати якомога більшу потужність, отже, інтервали мають бути якнайкоротшими. Числа a, b довірчого інтервалу (a, b) називаються відповідно нижньою та верхньою довірчою межею для основної розподіленої маси значень генеральної сукупності G, а величина $p\left(x \notin (a, b)\right)$ — рівнем значущості інтервалу (a, b). Легко бачити, що рівень значущості інтервалу, побудованого за правилом трьох сігм, не перевищує 0,05.

Оскільки в статистичних дослідженнях і розпізнаванні образів математичне сподівання $m\left(x\right)$ і дисперсія $\sigma\left(x\right)$ оцінюються за вибіркою x_1, x_2, $\ldots$, x_n з генеральної сукупності G, одержаної шляхом простого випадкового вибору, в обчисленнях використовуються їх оцінки $m \approx \bar{x} = \frac{1}{n}\sum\limits_{k=1}^{n} x_k$ та $\sigma^2(x) \approx s^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(x_k - \bar{x})^2$. Тому довірчий інтервал $\bar{I}$, побудований за "правилом 3s"має вигляд $\bar{I} = \left(\bar{x} - 3s, \bar{x} + 3s\right)$.

2. Гіпотеза Хілла

В 1968 році американський статистик B.M.Hill сформулював твердження (Hill's assumption $A_{(n)}$) [2], згідно з яким, якщо для одержання вибірки $(x_1, x_2, \ldots, x_n) \in G$ використовується випадковий вибір, такий що вибіркові значення $x_1, x_2, \ldots, x_n$ є симетрично залежними випадковими величинами з однаковим абсолютно неперервним розподілом, то

$$P\left(x_{n+1} \in \left(x_{(i)}, x_{(j)}\right)\right) = \frac{j - i}{n + 1},$$

де x_{n+1} — наступне вибіркове значення, а $x_{(i)},\ x_{(j)}$ – порядкові статистики. Це твердження було доведено в роботах Ю. І. Петуніна та його учнів: для випадку незалежних однаково розподілених випадкових величин із неперервним розподілом — в 1982 році [3], а для симетрично залежних випадкових величин — в 2003 році [4].

3. Міра близькості між вибірками

В 1984 році Ю. І. Петунін увів у розгляд p-статистику — міру близькості між вибірками $x = (x_1, x_2, ..., x_n) \in G$ і $y = (y_1, y_2, ..., y_m) \in H$, що не залежить від їхнього розподілу [5]. Ця статистика обчислюється так. Припустимо, що $F_G(x) \equiv F_H(x)$ і побудуємо варіаційний ряд $x_{(1)} \le ... \le x_{(n)}$. Позначимо як A_{ij} подію, що полягає у тому, що вибіркове значення y_k потрапляє між порядковими статистиками $x_{(i)}$ і $x_{(j)}$: $A_{ij}^{(k)} = \left\{ x_k \in \left(x_{(i)}, x_{(j)} \right) \right\}$. За гіпотезою Хілла

$$P(A_{ij}^{(k)}) = P\left(x_k \in (x_{(i)}, x_{(j)}) \right) = p_{ij}^{(n)} = \frac{j-i}{n+1}.$$

Побудувавши довірчий інтервал для невідомої ймовірності події A_{ij} за формулами Вільсона

$$p_{ij}^{(1)} = \frac{h_{ij}^{(n,m)} m + g^2/2 - g\sqrt{h_{ij}^{(n,m)}(1-h_{ij}^{(n,m)})m + g^2/4}}{m+g^2},$$

$$p_{ij}^{(2)} = \frac{h_{ij}^{(n,m)} m + g^2/2 + g\sqrt{h_{ij}^{(n,m)}(1-h_{ij}^{(n,m)})m + g^2/4}}{m+g^2},$$

де $h_{ij}^{(n,m)}$ — частота появи події $A_{ij}^{(n)}$ в m випробуваннях, отримуємо довірчий інтервал $I_{ij} = \left(p_{ij}^{(1)}, p_{ij}^{(2)} \right)$ із рівнем, що визначається параметром g. Зокрема, при $g = 3$ рівень значущості цього інтервалу не перевищує 0,05. Обчислимо $N = n(n-1)/2$, де L — кількість інтервалів I_{ij}, які вміщують ймовірності p_{ij}, і $h = \rho(F_G, F_H) = \rho(x, y) = \frac{L}{N}$ — міра близькості між вибірками x та y, або p-статистика. Поклавши $h_{ij} = h, m = N$ та $g = 3$, отримуємо довірчий інтервал $I = \left(p^{(1)}, p^{(2)} \right)$ для ймовірності $p(B)$. Довірчі інтервали $I_{ij} = \left(p_{ij}^{(1)}, p_{ij}^{(2)} \right)$ та $I = \left(p^{(1)}, p^{(2)} \right)$ будемо називати інтервалами, які побудовані за правилом $3s$.

4. Непараметричний критерій еквівалентності генеральних сукупностей

Схема випробувань, в яких можуть з'являтися події $A_{ij}^{(k)}$, у випадку істинності гіпотези про рівність розподілів називається узагальненою схемою Бернуллі. Якщо ця гіпотеза хибна, то вказана схема випробувань називається модифікованою схемою Бернуллі. У загальному випадку, коли може бути істинною як гіпотеза H_0: $F_G(u) = F_H(u)$, так і H_1: $F_G(u) \ne F_H(u)$, схема випробувань називається МП-схемою [6–9]. Дослідження цих схем були предметом робіт Ю. І. Петуніна та його учнів у 1991–1992 рр. Непараметричний критерій еквівалентності генеральних сукупностей на підставі міри близькості між двома вибірками був запропонований в 2003 р. [10].

Зокрема, було доведено, що якщо в узагальненій схемі випробувань Бернуллі 1) $n = m$; 2) $0 < \lim\limits_{n\to\infty} p_{ij}^{(n)} = p_0 < 1$ та 3) $0 < \lim\limits_{n\to\infty} P\left(A_{ij} | H_0 \right) = p^* < 1$, то асимптотичний рівень значущості β послідовності довірчих інтервалів

$I_{ij}^{(n)}$ для ймовірностей $p_q^{(n)} = p_{ij}^{(n)} = p\left(A_{ij}^{(n)}\right)$, побудованих за правилом $3s$, не перевищує 0,05.

Нехай $B_1, B_2, ..., B_n, ...$ — послідовність подій, які можуть відбутися при проведенні випадкового експерименту E, $\lim\limits_{k\to\infty} p\left(B_k\right) = p^*$ і $h_{n_1}(B_1)$, $h_{n_2}(B_2)$, ..., $h_{n_k}(B_k), ...$ – послідовність частот подій $B_1, B_2, ..., B_n, ...$, відповідно, причому $\frac{k}{n_k} \to 0$ при $k \to \infty$.

Будемо називати експеримент Е посиленим випадковим експериментом, якщо $h_{n_k}\left(B_k\right) \to p^*$ при $k \to \infty$. Якщо за умов посиленого випадкового експерименту вибірки $x = (x_1, ..., x_n) \in G$ та $y = (y_1, ..., y_m) \in H$ мають однаковий об'єм, то асимптотичний рівень значущості інтервалу $I^{(n)} = \left(p^{(1)}, p^{(2)}\right)$, побудований за правилом $3s$ при $g = 3$ за допомогою формул Вільсона, не перевищує 0,05.

Критерій для перевірки основної гіпотези про рівність неперервних гіпотетичних функцій розподілу $F_G(u)$ і $F_H(u)$ з рівнем значущості, що приблизно дорівнює 0,05, формулюється так: якщо інтервал $I = \left(p^{(1)}, p^{(2)}\right)$ містить 0,95, то приймається основна гіпотеза, інакше приймається альтернативна гіпотеза.

5. Довірчі інтервали із гарантованим рівнем значущості

Як показано вище, рівень значущості непараметричного критерію про еквівалентність генеральних сукупностей на основі міри близькості залежить від рівня значущості довірчих інтервалів, що використовуються при обчисленні p-статистики. Бажано, щоб цей рвієнь значущості був гарантованим і не залежав від конкретних значень невідомої ймовірності і об'єму вибірки. Таким чином, постає задача дослідити довірчі інтервали із гарантованим рівнем значущості в класичній і узагальненій моделях Бернуллі. Перша задача формулюється так: побудувати довірчий інтервал для невідомої ймовірності p на основі частоти h в класичній моделі Бернуллі, яка складається із n випробувань.

Розглянемо дві функції, що залежать від $p \in [0,1] : \varphi(p) = |h - p|$ і $\psi(p) = \frac{1}{2n} + \frac{\lambda}{n}\tilde{\psi}(p)$, де $\tilde{\psi}(p) = \sqrt{np(1-p) + \frac{1}{12}}$, $p \in R^1$.

Графік функції $\tilde{\psi}(p)$, $p \in R^1$ являє собою верхню частину еліпса E із центром в точці $\left(\frac{1}{2}, 0\right)$, що проходить через точки $A = \left(\frac{1}{2n}\left(n + \sqrt{\frac{n}{3} + n^2}\right), 0\right)$, $B = \left(\frac{1}{2}, \sqrt{\frac{1}{12n} + \frac{1}{4}}\right)$, $C = \left(\frac{1}{2n}\left(n - \sqrt{\frac{n}{3} + n^2}\right), 0\right)$ і $D = \left(\frac{1}{2}, -\sqrt{\frac{1}{12n} + \frac{1}{4}}\right)$.

Графік функції $\psi(p)$, що не залежить від h, являє собою дугу еліпса Г, яка проходить через точки $(0, \psi(0))$, $\left(\frac{1}{2}, \psi\left(\frac{1}{2}\right)\right)$, $(1, \psi(1))$, так що функція $\psi(p)$ досягає свого мінімуму в точці $p = \frac{1}{2}$ і є симетричною відносно цієї точки.

Якщо $h > \psi(0) = \frac{1}{2n} + \frac{\lambda}{n\sqrt{12}}$, то нижня довірча межа p_1 є найменшим коренем рівняння

$$\left(1 + \frac{\lambda^2}{n}\right)p^2 - \left(\frac{\lambda^2}{n} - \frac{1}{n} + 2h\right)p + h^2 - \frac{h}{n} + \frac{1}{4n^2}\left(1 - \frac{\lambda^2}{3}\right) = 0.$$

Якщо $h \le \psi(0)$, то $p_1 = 0$.

Аналогічно, якщо $1-h > \psi(1)$, то верхня довірча межа p_2 є найбільшим коренем рівняння

$$\left(1+\frac{\lambda^2}{n}\right)p^2-\left(\frac{\lambda^2}{n}+\frac{1}{n}+2h\right)p+h^2+\frac{h}{n}+\frac{1}{4n^2}\left(1-\frac{\lambda^2}{3}\right)=0.$$

Якщо $1-h \le \psi(1)$, то $p_2 = 1$.

В узагальненій схемі Бернуллі якщо $h > \frac{1}{2m}+\frac{\lambda}{m\sqrt{12}} = \gamma$, то нижня довірча межа p_1 є найменшим коренем рівняння

$$\left(1+\tfrac{(m+n+1)\lambda^2}{(n+2)m}\right)p^2+\left(\tfrac{1}{m}-\tfrac{(m+n+1)\lambda^2}{(n+2)m}-2h\right)p+h^2-\tfrac{h}{m}+\tfrac{1}{4m^2}\left(1-\tfrac{\lambda^2}{3}\right)=0.$$

Якщо $h \le \gamma$, то $p_1 = 0$.

Якщо $1-h > \gamma$, то верхня довірча межа p_2 в узагальненій схемі Бернуллі є коренем рівняння

$$\left(1+\tfrac{(m+n+1)\lambda^2}{(n+2)m}\right)p^2-\left(\tfrac{1}{m}+\tfrac{(m+n+1)\lambda^2}{(n+2)m}+2h\right)p+h^2+\tfrac{h}{m}+\tfrac{1}{4m^2}\left(1-\tfrac{\lambda^2}{3}\right)=0.$$

Якщо $1-h \le \gamma$, то $p_2 = 1$.

Рівень значущості побудованого довірчого інтервалу не перевищує $\frac{4}{9}\frac{1}{\lambda^2}$ (зокрема, 0,05 для $\lambda = 3$).

Для порівняння двох ймовірностей розглянемо дві схеми Бернуллі B_1 і B_2 з випадковими експериментами E_n^1 і E_n^2, у яких подія A_1 відбувається з ймовірністю p_1, а подія A_2 — з ймовірністю p_2 відповідно. Нехай здійснюються n_1 випробувань у схемі B_1 і n_2 випробувань — у схемі B_2. Знайдемо частоти $h_1 = \frac{k_1}{n_1}$ і $h_2 = \frac{k_2}{n_2}$, що відповідають ймовірностям p_1 і p_2. Побудуємо статистичний критерій, що дозволяє по наявних частотах h_1 і h_2 зробити висновок про значуще відхилення ймовірностей p_1 і p_2, не використовуючи умову незалежності частот. Нехай B — схема Бернуллі, в якій випадкова подія A може відбутися з ймовірністю p, а частота h випадкової події A в n випробуваннях. Як відомо, $m(h) = p$, $\sigma(h) = \frac{pq}{n}$, де $q = 1-p$, зміщена оцінка $\sigma(h) : s_n^2 = \frac{h(1-h)}{n}$, незміщена оцінка $\sigma(h) : \hat{s}_n^2 = \frac{n}{n-1}s_n^2 = \frac{h(1-h)}{n-1}$.

В роботах [11], [12] показано, що при будь-якому натуральному $n \ge 2$ $\Delta\left(s_n^2\right) = m\left(s_n^2-\sigma(h)\right)^2 < \sigma\left(\hat{s}_n^2\right)$. Нехай $I_1 = \left(p_1^{(1)}, p_2^{(1)}\right)$ і $I_2 = \left(p_1^{(2)}, p_2^{(2)}\right)$ — довірчі інтервали для невідомих ймовірностей p_1 і p_2, відповідно, рівень значущості яких приблизно дорівнює 5%. Статистичні критерії для порівняння двох ймовірностей на основі довірчих інтервалів формулюються так.

Критерій 1. Основна гіпотеза про рівність ймовірностей p_1 і p_2 відхиляється, якщо довірчі інтервали I_1 і I_2 не перетинаються. Якщо основна гіпотеза є істинною, то рівень значущості критерію не перевищує 0,05.

Критерій 2. Основна гіпотеза відхиляється, якщо середина інтервалу I_2 не належить інтервалу I_1. Якщо основна гіпотеза є істинною, то рівень значущості критерію не перевищує 0,05.

6. Критерій гомогенності вибірок

Оскільки в основі викладеної вище теорії лежить припущення про унімодальність неперервних розподілів спостережуваних випадкових величин, виникає задача, яка полягає у визначенні кількості мод функції розподілу та їх параметрів на основі кусково-лінійної емпіричної функції розподілу. Нехай G — генеральна сукупність з неперервною строго зростаючою функцією розподілу $F(u)$. Тоді $F(u)$ має обернену функцію розподілу $F^{-1}(u)$. Носієм функції розподілу $F(u)$ називається мінімальний відрізок $[a,b]$, для якого $F(u) \equiv 0 \ \ \forall u \leq a$ і $F(u) \equiv 1 \ \ \forall u \geq b$. Якщо функція $F(u)$ є неперервною і строго зростаючою на носії $[a,b]$, то оберненою функцією розподілу є обернена функція до звуження $F_{[a,b]}(u)$ на носій $[a,b]$.

Розглянемо варіаційний ряд $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$, побудований по вибірці $x_1, x_2, ..., x_n$ і визначимо кусково-постійну емпіричну функцію розподілу $F_n^*(u)$. За теоремою Гливенка-Кантеллі

$$p\left(\lim_{n\to\infty} \max_u |F(u) - F_n^*(u)| = 0\right) = 1.$$

У кусково-постійної емпіричної функції розподілу немає оберненої функції (у її звичайному розумінні), тому її неможливо використовувати для оцінки оберненої функції розподілу. У зв'язку з цим виникає проблема побудови емпіричної функції розподілу, у якої була б обернена функція. Для неперервної монотонно зростаючої функції розподілу $F(u)$ з компактним носієм $[a,b]$ побудуємо кусково-лінійний сплайн

$$\tilde{F}_n^*(u) = \begin{cases} \frac{u}{n} + \frac{kx_{(k+1)} - (k+1)x_{(k)}}{n}, & x_{(k)} \leq u \leq x_{(k+1)}, \ x_{(0)} = a, \ k = 0, 1, ..., n; \\ 1, \quad x_{(n)} \leq u \leq b. \end{cases}$$

Кусково-лінійна емпірична функція розподілу $F_n^*(u)$, визначена на відрізку $[a,b]$, має на відрізку $\left[a, x_{(n)}\right]$ обернену функцію $(F_n^*(u))^{-1} \stackrel{\Delta}{=} \Psi_n^*(u)$, $0 \leq u \leq 1$. Функція $\Psi_n^*(u)$ є неперервною кусково-лінійною функцією, що проходить через точки $\left(\frac{k}{n}, x_{(k)}\right)$, де $x_{(k)}$– порядкові статистики. Легко помітити, що $\forall u \in R^1$

$$\left|F_n^*(u) - \tilde{F}_n^*(u)\right| < \frac{1}{n},$$

тому $\tilde{F}_n^*(u)$ є слушною оцінкою для гіпотетичної функції розподілу, і теорема Глівенка- Кантеллі залишається справедливою.

В роботі [13] доведені аналоги теореми Глівенка-Кантеллі. Якщо гіпотетична функція розподілу $F(u)$ генеральної сукупності є неперервною і має компактний носій $[a,b]$, то послідовність кусково-лінійних емпіричних функцій розподілу із ймовірністю одиниця збігається до функції розподілу в чебишовській метриці: $p\left(\lim_{n\to\infty} \max_u \left|F(u) - \tilde{F}_n^*(u)\right| = 0\right) = 1.$

Якщо гіпотетична функція розподілу $F(u)$ генеральної сукупності G є неперервною і має компактний носій $[a,b]$, то послідовність функцій $\{\Psi_n^*(u)\}$

збігається із ймовірністю одиниця до функції $F^{-1}(u)$, тобто

$$p\left\{\lim_{n\to\infty}\max_{a\le u\le b}\left|F^{-1}(u)-\Psi_n^*(u)\right|=0\right\}=1.$$

Генеральна сукупність називається гомогенною, якщо вона складається з однорідних об'єктів таких, що значення $x_1,\ldots,x_n$ кількісного показника x мають однакові маргінальні розподіли $F(u):F_k(u)=F(u), k=1,...,n$. На практиці, генеральну сукупність вважають гомогенною, якщо маргінальний розподіл $F(u)$ є унімодальним. Вибірку, що утворена конкатенацією двох вибірок $x^{(1)}=(x_1,...,x_k)\in G_1$ і $x^{(2)}=(x_{k+1},...,x_n)\in G_2$ з двох різних генеральних сукупностей G_1 і G_2 будемо називати складеною. Вибірка, що утворена об'єднанням двох вибірок $x^{(1)}=(x_1,...,x_k)\in G_1$ і $x^{(2)}=(x_{k+1},...,x_n)\in G_2$ з двох різних генеральних сукупностей G_1 і G_2 будемо називати змішаною. Розв'язок проблеми розпізнавання складених вибірок можна за допомогою міри близькості, описаної вище, а також проблеми розпізнавання змішаної вибірки шляхом аналізу точок розриву графіку варіаційного ряду і дослідження спейсингів викладено в роботах [14–16], а новий критерій одномодальності функцій розподілу — в роботі [17].

7. Застосування в медицині та біології

Результати в області непараметричного статистичного розпізнавання образів, отримані Ю. І. Петуніним та його учнями, а також їх застосування в медичних дослідженнях (зокрема, для діагностики раку) викладені в монографіях [18], [19]. Зокрема, в цих монографіях описано математичні основи процесу розпізнавання і діагностики раку молочної залози на підставі даних сканограм ДНК інтерфазних ядер клітин слизової оболонки порожнини рота; проведено ретроспективний аналіз ступеня значущості для прогнозу клінічних, цитогенетичних і морфологічних показників у хворих на злоякісну меланому; продемонстровано використання сплайнової регресії і модифікованого полігону розподілу для виявлення залежності ймовірності виникнення злоякісних новотворів в учасників ліквідації наслідків аварії на ЧАЕС від поглиненої дози опромінення і пошуку її точки переходу, після якої ймовірність виникнення онкологічних захворювань різко знижується; показана ефективність розроблених методів стратифікаційного аналізу генерацій клітин карциноми Герена, які піддалися рентгенівському опроміненню; доведена ефективність стратификаційного аналізу при дослідженні морфометричних показників клітин з інтактної популяції плоскоклітинного рака ротової порожнини людини, інтактної популяції клітин карциноми Герена у пацюків, також при аналізі впливу цисплатина на інтактні популяції ракових клітин в обох випадках.

8. Висновки

Для наукової школи Ю. І. Петуніна в галузі статистичного розпізнавання образів притаманно застосування строго обгрунтованих методів перевірки гіпотез, що не залежать від припущень щодо параметрів функції розподілу. До основних результатів, отриманих цією школою, належать нерівність Височанського-Петуніна, p-статистика, критерії перевірки гіпотези про рівність розподілів та багато інших. Ці методи виявили високу ефективність при розв'язанні практичних задач, що виникають в медицині та біології, зокрема, при діагностиці раку.

Література

1. Высочанский Д. Ф. Обоснование правила трех сигм для одномодальных распределений /Д. Ф. Высочанский, Ю. И. Петунин // Теория вероятностей и математическая статистика. — 1979. — Вып. 21. — С. 23–35.
2. Hill B. M. Posterior distribution of percentiles: Bayes' theorem for sampling from a population / B. M. Hill // Journal of the American Statistical Association. — 1968. — V. 63. — P. 677–691.
3. Мадреимов И. Характеризация равномерного распределения с помощью порядковых статистик / И. Мадреимов, Ю. И. Петунин // Теория вероятностей и математическая статистика. — 1982. — Вып. 27. — С. 92–102. .
4. Andrushkiw R. I. Construction of the bulk of general population in the case of exchangeable sample values / R. I. Andrushkiw, D. A. Klyushin, Yu. I. Petunin, V.V Lysyuk // In: Proceedings of the International Conference of Mathematics and Engineering Techniques in Medicine and Biological Science (METMBS'03). — Las Vegas, Nevada, USA, June 26-29, 2003. — P.486–489.
5. Ганина К. П. Количественные характеристики ядерного полиморфизма эпителиальных клеток при фиброаденоматозе и раке молочной железы / К. П. Ганина, Ю. И. Петунин, Я. Г. Тимошенко // Докл. АН УССР. — 1984. — 35, № 12. — С.1414–1449.
6. Матвейчук С. А. Обобщение схемы Бернулли, возникающее в вариационной статистике. I / С. А. Матвейчук, Ю. И. Петунин // Укр. матем. журнал. — 1991. — 42, № 4. — С. 518–528.
7. Матвейчук С. А. Обобщение схемы Бернулли, возникающее в вариационной статистике. II / С. А. Матвейчук, Ю. И. Петунин // Укр. матем. журнал. — 1991. — 48, № 6. — С. 779–785.
8. Бородянский Н. И. Статистические критерии для идентификации генеральной совокупности / Н. И. Бородянский, С. А. Матвейчук, Ю. И. Петунин // Кибернетика и системный анализ. — 1992. — № 1. — С. 148–153.
9. Johnson N. Some generalizations of Bernoulli and Polya-Eggenberger contagion models / / N. Johnson, S. Kotz // Statist. Paper. — 1991. — 32. — P. 1–17.
10. Клюшин Д. А. Непараметрический критерий эквивалентности генеральных совокупностей, основанный на мере близости между выборками / Д. А. Клюшин, Ю. И. Петунин // Укр. матем. журн. — 2003. — Т.5, № 2. — С. 147–163.
11. Клюшин Д. А. Точні довірчі межі для невідомої ймовірності в класичній та узагальненій моделях Бернуллі / Д. А. Клюшин, Ю. І. Петунін // Вісник Київ. Нац. Ун-ту, сер. фіз-мат наук. — 2005, №2. — С. 240–247.
12. Клюшин Д. А. Статистичний критерій для порівняння двох ймовірностей / Д. Клюшин, Ю. Петунін // Вісник Київського університету. Кібернетика. — 2005. — № 6. — С. 35–40.
13. Клюшин Д. А. Аналог теоремы Гливенко-Кантелли для обратных функций распределения / Д. А. Клюшин, Ю. І. Петунін, М. Ю. Савкіна // Вісник Київського університету. Кібернетика. — 2006. — № 7. — С. 31–34.

14. Bairamov I. G. Test of heterogeneity of general population / I. G. Bairamov, Yu. I. Petunin, D. A. Klyushin // Istatistik (Journal of the Turkish Statistical Assosiation). — 1999. — V. 1, № 3. — P. 19–29.
15. Петунин Ю. И. Стратификационный анализ морфологических показателей популяций раковых клеток с фенотипом лекарственной резистентности / Ю. Петунин, Д. Клюшин, Г. Кулик и др. // Кибернетика и системный анализ. — 2005. — № 6. — С. 158–167.
16. Петунин Ю. И. Оценка влияния величины дозы облучения на вероятность развития злокачественных новообразований на основе сплайновой регрессии / Ю. И. Петунин, Э. А. Демина, Д. А. Клюшин и др. // Кибернетика и системный анализ. — 2006. — № 3. — С. 168–176.
17. Клюшин Д. А. Новые критерии одномодальности функций распределения / Д. А. Клюшин // Журнал обчисл. та прикл. матем. — 2007. — № 1 (94). — С. 31–45.
18. Andrushkiw R. I. Computer-aided cytogenetic method of cancer diagnosis / R. I. Andrushkiw, N. V. Boroday, D. A. Klyushin, Yu. I. Petunin — New Jork: Nova Publishers. — 2007. — 303 p.
19. Клюшин Д. А. Доказательная медицина. Применение статистических методов / Д. А. Клюшин, Ю. И. Петунин — М.: Изд. дом Вильямс, 2007. — 320 с.

Факультет кібернетики, Київський національний університет імені Тараса Шевченка, вул. Володимирська, 64, Київ, 01601, Україна.

Надійшла 29.11.12