ISSN 2078-3744. Вісник Львів. ун-ту. Серія мех.-мат. 2021. Випуск 91. С. 99–104
Visnyk of the Lviv Univ. Series Mech. Math. 2021. Issue 91. P. 99–104
http://publications.lnu.edu.ua/bulletins/index.php/mmf
doi: http://dx.doi.org/10.30970/vmm.2021.91.099-104

УДК 519.2

ПОБУДОВА СТАТИСТИЧНИХ КРИТЕРІЇВ З УРАХУВАННЯМ ВПЛИВУ ЗОВНІШНЬОГО СЕРЕДОВИЩА

Ярослав ЄЛЕЙКО, Оксана ЯРОВА, Святослав ГОЛОВАТИЙ

Львівський національний університет імені Івана Франка,
вул. Університетська, 1, 79000, м. Львів
e-mail: oksanayarova93@gmail.com

Досліджено вплив зовнішнього середовища на випадкові події та їхні функції розподілу. Розглянуто вибірку випадкових величин, яка залежить від випадкового середовища й описується за допомогою повної групи подій із заданими ймовірностями для кожної події. Основну вибірку розділено на k вибірок, залежно від подій, якими вони описуються. Для кожної з вибірок побудовано емпіричну функцію розподілу та застосовано критерій Колмогорова. В результаті отримано суміш розподілів. Також, розглянуто приклад застосування цієї теорії. Для вибірки знайдено суміш розподілу та параметри отриманих функцій розподілів.

Ключові слова: критерій Колмогорова, функція розподілу, суміш розподілів, показниковий розподіл.

Розглянемо вибірку $X_1, X_2, \ldots, X_n$, яка залежить від зовнішнього середовища, що описується за допомогою повної групи подій $A_1, A_2, \ldots, A_k$ з ймовірностями $p_1, p_2, \ldots, p_k$, відповідно. При чому

$$\sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{k} p_i = 1.$$

Сортуємо вибірку за подіями так:
$x_{11}, x_{12}, \ldots, x_{1r_1}$ для події A_1, за умови, що функція розподілу $F_1(x)$;
$x_{21}, x_{22}, \ldots, x_{2r_2}$ для події A_2, за умови, що функція розподілу $F_2(x)$;
. . . . ,
$x_{k1}, x_{k2}, \ldots, x_{kr_k}$ для події A_k, за умови, що функція розподілу $F_k(x)$.

2020 Mathematics Subject Classification: 60J25

Тоді вибірка $x_1, x_2, \ldots, x_n$ буде описуватись функцією розподілу, яка визначається за допомогою суміші розподілів

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + \cdots + p_k F_k(x).$$

Побудуємо емпіричні функції розподілів для кожної з вибірок за критерієм Колмогорова. Найперше, впорядкуємо вибірку $x_{11}, x_{12}, \ldots, x_{1r_1}$, у підсумку отримаємо $x_{i_1}, x_{i_2}, \ldots, x_{i_n}$. Кількість елементів вибірки позначимо n_1. Далі, розділимо вибірку на $\sqrt{n_1}$ інтервалів і знайдемо кількість елементів у кожному з інтервалів. Позначимо n_{1i} — кількість елементів в i-му інтервалі. Отже, емпірична функція розподілу для першої вибірки матиме вигляд:

$$F^1_{em}(x) = \begin{cases} 0, & x \leqslant x_{n_{11}} \\ \frac{n_{11}}{n_1}, & x_{n_{11}} < x \leqslant x_{n_{12}} \\ \frac{n_{11}+n_{12}}{n_1}, & x_{n_{12}} < x \leqslant x_{n_{13}} \\ \ldots & \ldots \\ 1, & x > x_{n_{1n}} \end{cases}$$

Аналогічно будуємо всі емпіричні функції розподілів.

Застосуємо критерій Колмогорова. Для цього розглянемо таку різницю:

$$\left|F_{em}(x) - F(x)\right| =$$
$$= \left|\frac{n_1}{n} F^1_{em}(x) + \frac{n_2}{n} F^2_{em}(x) + \cdots + \frac{n_k}{n} F^k_{em}(x) - (p_1 F_1(x) + p_2 F_2(x) + \cdots + p_k F_k(x))\right|$$

Оцінимо такий вираз

$$\left|\frac{n_i}{n} F^i_{em}(x) - p_i F_i(x)\right| \leqslant \left|\frac{n_i}{n} F^i_{em}(x) - \frac{n_i}{n} F_i(x)\right| + \left|\frac{n_i}{n} F_i(x) - p_i F_i(x)\right| \leqslant$$
$$\leqslant \frac{n_i}{n}\left|F^i_{em}(x) - F_i(x)\right| + |F_i(x)| \cdot \left|\frac{n_i}{n} - p_i\right|.$$

$F_i(x)$ — функція розподілу, тому $|F_i(x)| \leqslant 1$, $\frac{n_i}{n} \to p_i$, тому

$$\left|\frac{n_i}{n} F^i_{em}(x) - p_i F_i(x)\right| \leqslant \frac{n_i}{n}\left|F^i_{em}(x) - F_i(x)\right|.$$

Розглянемо критерій Колмогорова з наступними гіпотезами

$$H_0 : F^i_{em}(x) = F_i(x) H_1 : F^i_{em}(x) \neq F_i(x).$$

Позначимо

$$D_{n_i} = \sup_j \left|F^{in_i}_{em}(x) - F_{in_i}(x)\right|.$$

Побудуємо критичну область

$$\tau_\alpha = \{t \geqslant t_\alpha\}.$$

Тоді

$$P\{D_{n_i} \in \tau_\alpha | H_0\} = P\left\{\sqrt{n_i} D_{n_i} \geqslant \lambda_\alpha | H_0\right\} \approx 1 - K(\lambda_\alpha) = \alpha.$$

За великих $n_i, n_i \geqslant 20$ та вибраному рівні значущості α число λ_α визначається співвідношенням

$$K(\lambda_\alpha) = 1 - \alpha$$

нульова гіпотеза приймається, якщо

$$D_{n_i} \leqslant \frac{\lambda_\alpha}{\sqrt{n_i}}.$$

Отже,

$$|F^i_{em}(x) - F_i(x)| \leqslant \frac{\lambda_\alpha}{\sqrt{n_i}}.$$

Тоді

$$\frac{n_i}{n}|F^i_{em}(x) - F_i(x)| \leqslant \frac{n_i}{n}\frac{\lambda_\alpha}{\sqrt{n_i}} = \frac{\sqrt{n_i}}{n}\lambda_\alpha.$$

Вибираємо такі $N_{01}, \ldots, N_{0k}$, що $n_{11} > N_{01}, \ldots, n_{1k} > N_{0k}$.

Нехай $\widetilde{N_0} = \max\{N_{01}, N_{02}, \ldots, N_{0k}\}$. Позначимо $\dfrac{\sqrt{\widetilde{N_0}}}{n}\lambda_\alpha = \dfrac{\varepsilon}{k}$, $\forall\varepsilon > 0$. Тоді

$$|F_{em}(x) - F(x)| < k \cdot \frac{\varepsilon}{k} = \varepsilon.$$

Отож

$$F_{em}(x) \to F(x).$$

Тому сформулюємо теорему.

Теорема 1. *Нехай вибірка $X_1, X_2, \ldots, X_n$ залежить від зовнішнього середовища, що описується за допомогою повної групи подій $A_1, A_2, \ldots, A_k$. Тоді, $x_{i_1}, \ldots, x_{i_k}$, що відбуваються за умови A_i, мають функцію розподілу F_i.*

Розглянемо вибірку з 150 даних 0,889 0,340 0,398 0,060 0,084 0,381 0,145 1,517 0,396 0,237 0,274 0,014 0,322 0,260 0,041 0,185 0,277 0,325 0,097 0,902 0,064 0,263 0,849 0,082 0,889 0,090 0,166 0,239 0,801 0,612 0,066 0,033 0,189 1,084 0,054 0,308 0,327 0,174 0,160 0,273 0,096 0,317 2,047 0,380 0,162 0,235 0,030 0,539 0,526 0,399 2,058 0,401 1,457 0,912 0,199 0,562 0,664 0,060 0,125 1,255 0,008 1,049 0,080 0,204 0,486 0,099 0,927 0,421 0,398 0,647 0,438 0,070 0,729 0,461 0,493 0,097 0,354 0,370 0,062 0,080 0,149 0,549 6,172 1,094 0,354 2,214 0,168 2,042 0,563 0,867 0,375 3,624 1,703 1,127 0,048 0,597 1,476 2,759 0,669 0,636 0,261 1,840 1,750 1,389 1,333 0,004 0,962 5,455 0,137 0,634 4,362 0,034 1,664 0,065 1,464 0,656 1,178 3,615 0,435 1,655 0,566 1,137 0,736 3,114 0,008 0,102 0,651 0,186 0,697 0,324 2,164 0,063 0,931 0,540 0,710 0,263 3,619 0,984 0,877 1,574 3,736 0,908 0,547 0,819 0,759 0,026 0,937 2,617 0,440 0,078

Сортуємо цю вибірку щодо двох подій A_1 та A_2. Подія A_1 визначається такими випадковими величинами
0,008 0,014 0,030 0,033 0,041 0,054 0,060 0,060 0,062 0,064 0,066 0,070 0,078 0,080 0,082 0,084 0,090 0,096 0,097 0,097 0,099 0,125 0,145 0,160 0,162 0,166 0,174 0,185 0,189 0,199 0,204 0,235 0,237 0,239 0,260 0,263 0,273 0,274 0,277 0,308 0,317 0,322 0,325 0,327 0,340 0,354 0,370 0,380 0,381 0,396 0,398 0,398 0,399 0,401 0,421 0,438 0,461 0,486 0,493 0,526 0,539 0,562 0,612 0,647 0,664 0,729 0,801 0,849 0,889 0,889 0,902 0,912 0,927 1,049 1,084 1,255 1,457 1,517 2,047 2,058

Обсяг вибірки 80. Подію A_2 описує вибірка обсягом 70 випадкових величин
0,004 0,008 0,026 0,034 0,048 0,063 0,065 0,080 0,102 0,137 0,149 0,168 0,186 0,261 0,263 0,324 0,354 0,375 0,435 0,440 0,540 0,547 0,549 0,563 0,566 0,597 0,634 0,636 0,651 0,656

0,669 0,697 0,710 0,736 0,759 0,819 0,867 0,877 0,908 0,931 0,937 0,962 0,984 1,094 1,127 1,137 1,178 1,333 1,389 1,464 1,476 1,574 1,655 1,664 1,703 1,750 1,840 2,042 2,164 2,214 2,617 2,759 3,114 3,615 3,619 3,624 3,736 4,362 5,455 6,172

Побудуємо емпіричну функцію розподілу для першої вибірки.

$$F^1_{em}(x) = \begin{cases} 0, & x \leqslant 0,27; \\ 0,45, & 0,27 < x \leqslant 0,53; \\ 0,75, & 0,53 < x \leqslant 0,8; \\ 0,825, & 0,8 < x \leqslant 1,06; \\ 0,925, & 1,07 < x \leqslant 1,33; \\ 0,95, & 1,33 < x \leqslant 1,6; \\ 0,975, & 1,6 < x \leqslant 2,1; \\ 1, & x > 2,1. \end{cases}$$

Проведемо критерій Колмогорова для перевірки гіпотези
H_0 : емпіричний розподіл статистично не відрізняється від показникового.
Застосувавши критерій, отримаємо $D(80) = 0,032$, $\lambda_{0,05} = 1,3581$.
Отже, на рівні значущості 0,05 розподіл приймаємо нульову гіпотезу.
Знайдемо параметр розподілу.
За методом моментів для показникового розподілу

$$\hat{\lambda} = \frac{1}{EX} = \frac{n_1}{\sum_{i=1}^{n_1} n_i} = 2,4.$$

Побудуємо емпіричну функцію розподілу для другої вибірки.

$$F^2_{em}(x) = \begin{cases} 0, & x \leqslant 0,78; \\ 0,5, & 0,78 < x \leqslant 1,56; \\ 0,728, & 1,56 < x \leqslant 2,3; \\ 0,857, & 2,3 < x \leqslant 3,1; \\ 0,886, & 3,1 < x \leqslant 3,9; \\ 0,957, & 3,9 < x \leqslant 4,7; \\ 0,971, & 4,7 < x \leqslant 6,2; \\ 1, & x > 6,2. \end{cases}$$

Проведемо критерій Колмогорова для перевірки гіпотези
H_0 : емпіричний розподіл статистично не відрізняється від показникового.
Застосувавши критерій, отримаємо $D(70) = 0,034$, $\lambda_{0,05} = 1,3581$.
Тому на рівні значущості 0,05 розподіл приймаємо нульову гіпотезу.
Знайдемо параметр розподілу.
За методом моментів

$$\hat{\lambda} = 0,81.$$

Знайдемо суміш розподілів

$$F(x) = \frac{8}{15}\begin{cases} 0, & x \leqslant 0,27 \\ 0,45, & 0,27 < x \leqslant 0,53 \\ 0,75, & 0,53 < x \leqslant 0,8 \\ 0,825, & 0,8 < x \leqslant 1,06 \\ 0,925, & 1,07 < x \leqslant 1,33 \\ 0,95, & 1,33 < x \leqslant 1,6 \\ 0,975, & 1,6 < x \leqslant 2,1 \\ 1, & x > 2,1 \end{cases} + \frac{7}{15}\begin{cases} 0, & x \leqslant 0,78 \\ 0,5, & 0,78 < x \leqslant 1,56 \\ 0,728, & 1,56 < x \leqslant 2,3 \\ 0,857, & 2,3 < x \leqslant 3,1 \\ 0,886, & 3,1 < x \leqslant 3,9 \\ 0,957, & 3,9 < x \leqslant 4,7 \\ 0,971, & 4,7 < x \leqslant 6,2 \\ 1, & x > 6,2. \end{cases}$$

Визначимо розподіл суміші за допомогою критерію Колмогорова з гіпотезою H_0 : емпіричний розподіл статистично не відрізняється від гамма-розподілу.
Значення статистики $D(150) = 0,038$.
Отже, на рівні значущості 0,05 розподіл приймаємо гіпотезу H_0.
За методом моментів для гамма-розподілу

$$\hat{\alpha} = \frac{(\bar{x})^2}{\bar{x^2} - (\bar{x})^2} = 0,6$$

$$\hat{\beta} = \frac{\bar{x^2} - \bar{x})^2}{\bar{x}} = 1,3.$$

Побудуємо емпіричну функцію розподілу

$$F_{em}(x) = \begin{cases} 0, & x \leqslant 0,58 \\ 0,58, & 0,58 < x \leqslant 1,17 \\ 0,8, & 1,17 < x \leqslant 1,75 \\ 0,89, & 1,75 < x \leqslant 2,3 \\ 0,93, & 2,3 < x \leqslant 2,9 \\ 0,94, & 2,9 < x \leqslant 3,5 \\ 0,95, & 3,5 < x \leqslant 4 \\ 0,98, & 4 < x \leqslant 4,7 \\ 0,986, & 4,7 < x \leqslant 5,8 \\ 0,99, & 5,8 < x \leqslant 6,2 \\ 1, & x > 6,2 \end{cases}$$

На підставі цієї праці побудовано суміш розподілу для вибірки випадкових величин, яка залежить від впливу зовнішнього середовища.

Список використаної літератури

1. Г. И. Ивченко, Ю. И. Медведев, *Введение в математическую статистику*, Москва, ЛКИ, 2010, 600с.
2. Г. И. Ивченко, Ю. И. Медведев, А. В. Чистяков, *Сборник задач по матиматической статистике*, Москва, Высшая школа, 1989, 255с.

3. А. В. Скороход, *Лекції з теорії випадкових процесів*, Навч. посібн., Київ, Либідь, 1990, 168с.
4. В. М. Руденко, *Математична статистика*, Навч. посібн., Київ, Центр учбової літератури, 2012, 304с.

Стаття: надійшла до редколегії 21.04.2021
прийнята до друку 18.05.2021

CONSTRUCTION OF STATISTICAL CRITERIA TAKING INTO ACCOUNT THE INFLUENCE OF THE EXTREMAL ENVIRONMENT

Yaroslav YELEYKO, Oksana YAROVA, Sviatoslav HOLOVATYY

Ivan Franko Lviv National University of Lviv,
Universytetska Str., 1, 79000, Lviv, Ukraine
e-mail: oksanayarova93@gmail.com

The paper is devoted to the study of the influence of the external environment on random events and their distribution functions. A sample of random variables is considered, which depends on the random environment and is described using a complete group of events with given probabilities for each event. The main sample is divided into k samples depending on the events they describe. An empirical distribution function was constructed for each of the samples and the Kolmogorov criterion was applied. The result is a mixture of distributions. Also, an example of application of this theory is considered. A mixture of distributions and parameters of the obtained distribution functions are found for the sample.

Key words: Kolmogorov criterion, distribution function, mixture of distributions, exponential distribution.