УДК 519.23

Я.І. ЄЛЕЙКО
Львівський національний університет імені Івана Франка, Львів, Україна,
e-mail: *yikts@yahoo.com.*

О.А. ЯРОВА
Львівський національний університет імені Івана Франка, Львів, Україна,
e-mail: *oksana.yarova@lnu.edu.ua, oksanayarova93@gmail.com*.

СУМІШ РОЗПОДІЛІВ НА ОСНОВІ ЛАНЦЮГА МАРКОВА

Анотація. Розглянуто ланцюг Маркова, який перебуває під впливом даних зовнішнього середовища. На основі критерію Колмогорова знайдено суміш розподілів. Для вибірок нормальних розподілів побудовано емпіричні функції розподілів, знайдено ергодичний розподіл ланцюга Маркова та визначено суміш розподілів.

Ключові слова: ланцюг Маркова, матриця перехідних ймовірностей, критерій Колмогорова, нормальний розподіл, суміш розподілів.

Стани зовнішнього середовища описують за допомогою однорідного ланцюга Маркова $X(t)$ з дискретним часом $t = 0, 1, \ldots, n, \ldots$ і скінченною множиною станів $i = 1, 2, \ldots, m$ з перехідними ймовірностями

$$p_{ij} = P\{X(1) = j \mid X(0) = i\}.$$

Матриця перехідних ймовірностей

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

є нерозкладною. Отже, для неї існує єдиний ергодичний розподіл $p_1, \ldots, p_m$ такий, що

$$p_k = \sum p_j \cdot p_{jk}.$$

Нехай $p_{ij}^n = P\{X(n) = j \mid X(0) = i\}$ — перехідна ймовірність за n кроків. Тоді $x_1^i, \ldots, x_{k_1}^i, x_{k_1+1}^j, \ldots, x_{k_2}^j, x_{k_2+1}^s, \ldots, x_{k_3}^s \ldots$ — спостереження даних на траєкторії ланцюга Маркова.

Отже,

$$\frac{1}{N} \sum_{n=1}^{N} p_{ij}^n \to p_j.$$

Побудуємо емпіричну функцію розподілу. Позначимо

$$F_{em}^j(x) = \frac{\mu_i^j(x)}{n^j},$$

де $\mu_i^j(x)$ — кількість елементів вибірки $x_1^j(i), \ldots, x_{n_j}^j(i)$.

Зауважимо, що

$$\frac{n_j}{N} \to p_j.$$

Відомо, що $n_i^j = \sum_{s=1}^{n_i^j} p_{il}^s$ — середня кількість елементів вибірки $x_1^j(i), \ldots, x_{n_j}^j(i)$.

Тоді

$$\frac{n_i^j}{N} \xrightarrow[n \to \infty]{} \sum_{s=1}^{\infty} p_{ij}^s \to \infty \text{ для } N \to \infty.$$

Розглянемо вибірку $x_1, \ldots, x_N$. Тоді

$$F_{em}^N(x) = \frac{\mu^N(x)}{N}.$$

Застосуємо критерій Колмогорова.

Нехай існують теоретичні функції розподілу $F_i^j(x)$, $j = 1, 2, \ldots, m$, для генеральних сукупностей $x_1^j(i), \ldots, x_{n_j}^j(i)$, $j = 1, 2, \ldots, m$. Тоді стверджуємо, що за критерієм Колмогорова

$$F_i(x) = \sum p_j \cdot F_i^j(x)$$

— теоретична функція розподілу загальної вибірки $x_1, \ldots, x_N$.

Розглянемо таке співвідношення:

$$\sup_n \left| F_{em}^j(x) - \sum_{j=1}^m p_j \cdot F_i^j(x) \right| = \sup_n \left| \sum_{j=1}^m p_j \cdot F_{em(i)}^j(x) - \sum_{j=1}^m p_j \cdot F_i^j(x) \right| \le$$

$$\le \sup_n \sum_{j=1}^m \left| \frac{n^j}{N} \cdot F_{em(i)}^j(x) - p_j \cdot F_i^j(x) \right| \le$$

$$\le \sum_{j=1}^m \left(\sup_n \frac{n^j}{N} | F_{em(i)}^j(x) - F_i^j(x) | + \sup_n F_i^j(x) \left| \frac{n^j}{N} - p_j \right| \right).$$

Оскільки $\frac{n_j}{N} \to p_j$, то

$$\sum_{j=1}^m \left(\sup_n \frac{n^j}{N} | F_{em(i)}^j(x) - F_i^j(x) | + \sup_n F_i^j(x) \left| \frac{n^j}{N} - p_j \right| \right) \le$$

$$\le \sum_{j=1}^m \sup_n \frac{n^j}{N} \left| F_{em(i)}^j(x) - F_i^j(x) \right|.$$

Згідно з критерієм Колмогорова

$$F_{em(i)}^j(x) \to F_i^j(x).$$

Отже,

$$F_{em}^j(x) = \sum_{j=1}^m p_j \cdot F_i^j(x).$$

Розглянемо приклад. Задано ланцюг Маркова з трьома станами та матрицею перехідних ймовірностей

$$\begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}.$$

Знайдемо ергодичний розподіл ймовірностей:

$$\begin{cases} \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.7 & 0.2 & 0.1 \end{pmatrix} \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

і отримаємо систему рівнянь

$$\begin{cases} 0.2\pi_1 + 0.3\pi_2 + 0.5\pi_3 = \pi_1, \\ 0.4\pi_1 + 0.3\pi_2 + 0.3\pi_3 = \pi_2, \\ 0.7\pi_1 + 0.2\pi_2 + 0.1\pi_3 = \pi_3, \\ \pi_1 + \pi_2 + \pi_3 = 1. \end{cases}$$

Розв'язавши цю систему рівнянь, отримаємо ергодичний розподіл $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$.

Розглянемо три набори спостережень, кожен з яких має 50 значень на траєкторіях ланцюга Маркова:

1.653456763	1.388241180	1.370042048	2.889715749	1.386680711	1.452327426
0.726926330	1.418940873	0.621213689	0.972499233	1.085536897	0.239213415
3.051597033	0.823351461	-0.003737953	0.312557040	0.834027036	2.305121035
1.777297237	2.934088010	1.667866394	0.702201999	-0.560606781	-0.498425984
0.587292745	1.282017572	0.808885711	-0.605031245	1.281891494	0.555386384
0.508519323	1.870150940	0.725527304	0.953473163	1.110693803	1.178068450
1.007472174	0.750369263	1.676654023	-0.711910261	1.013610758	0.352698620
3.750503355	-0.196179477	2.46755622	1.657073255	0.505105343	-0.546637464
0.072670458	-0.286801798.				

Визначимо параметри вибірки. Середнє значення дорівнюватиме одиниці, а середнє квадратичне відхилення дорівнює 0.98.

За допомогою критерію Колмогорова–Смірнова перевіримо розподіл на нормальність у середовищі R. За результатом $D = 0.089$, $p = 0.78$.

Отже, цей розподіл статистично не відрізняється від нормального.

Побудуємо емпіричну функцію розподілу для цієї вибірки:

$$F_1(x) = \begin{cases} 0, & x \le -1, \\ 0.03, & -1 < x \le -0.1, \\ 0.16, & -0.1 < x \le 0.8, \\ 0.41, & 0.8 < x \le 1.7, \\ 0.71, & 1.7 < x \le 2.6, \\ 0.91, & 2.6 < x \le 3.5, \\ 1, & x > 3.5. \end{cases}$$

Перейдемо до другого набору даних:

2.6200790	1.4807486	2.6516347	3.0998681	2.6979984	3.9516891
1.8282090	2.0579951	2.1398216	1.3468950	3.9958547	1.6345657
1.9559110	0.4089037	2.3049905	3.6136520	1.5061578	0.4885126
1.9305657	1.1321377	0.8429866	2.0864765	0.8603961	1.1731040
1.2820280	2.1483502	3.1287606	0.4680255	2.6415658	3.2120756
0.5080343	2.4407163	3.2635186	1.3811768	2.2599562	2.1892540
2.5582583	1.5432355	3.3865899	1.6570116	1.0979401	1.1277084
2.1253112	1.7018780	1.6789929	1.6761571	1.3365631	0.5440983
1.4389285	0.9627656.				

Визначимо параметри вибірки. Середнє значення дорівнює 1.87, а середнє квадратичне відхилення дорівнює 1.1.

За допомогою критерію Колмогорова–Смірнова перевіримо розподіл на нормальність в середовищі R. За результатом маємо $D = 0.11$, $p = 0.46$.

Отже, розподіл статистично не відрізняється від нормального.

Побудуємо емпіричну функцію розподілу для цієї вибірки:

$$F_2(x) = \begin{cases} 0, & x \le 0.2, \\ 0.05, & 0.2 < x \le 1, \\ 0.18, & 1 < x \le 1.8, \\ 0.43, & 1.8 < x \le 2.6, \\ 0.72, & 2.6 < x \le 3.4, \\ 0.91, & 3.4 < x \le 4.2, \\ 1, & x > 4.2. \end{cases}$$

Розглянемо третій набір даних:

1.4582236	3.3613618	3.5988723	1.5311032	2.4145470	3.7649997
4.1848876	4.0056642	1.6345502	4.8907693	4.4812893	2.2148239
2.1035262	2.1683817	1.7803976	3.7949029	3.2922343	0.3651155
1.8472815	2.9907794	3.0557754	2.4519250	1.8266306	2.9583371
1.9086792	3.3957339	3.8498220	4.3758694	2.6504008	2.4593920
5.0544986	1.7427733	2.9673705	3.8026640	2.5616908	2.3349028
1.3042017	5.5171742	2.8226645	3.0785606	2.5294478	2.0869169
1.0487272	5.3957541	2.1235661	2.3550109	2.7280821	2.0473981
2.9800825	2.3875351.				

Визначимо параметри вибірки. Середнє значення дорівнює 2.8, а середнє квадратичне відхилення дорівнює 0.99.

За допомогою критерію Колмогорова–Смірнова перевіримо розподіл на нормальність в середовищі R. За результатом $D = 0.11$, $p = 0.45$.

Отже, розподіл статистично не відрізняється від нормального.

Побудуємо емпіричну функцію розподілу для цієї вибірки:

$$F_3(x) = \begin{cases} 0, & x \le 1.3, \\ 0.02, & 1.3 < x \le 2.1, \\ 0.14, & 2.1 < x \le 2.9, \\ 0.43, & 2.9 < x \le 3.7, \\ 0.77, & 3.7 < x \le 4.5, \\ 0.95, & 4.5 < x \le 5.2, \\ 1, & x > 5.2. \end{cases}$$

Побудуємо суміш розподілів:

$$F(x) = \frac{1}{3}\begin{cases} 0, & x \le -1, \\ 0.03, & -1 < x \le -0.1, \\ 0.16, & -0.1 < x \le 0.8, \\ 0.41, & 0.8 < x \le 1.7, \\ 0.71, & 1.7 < x \le 2.6, \\ 0.91, & 2.6 < x \le 3.5, \\ 1, & x > 3.5 \end{cases} + \frac{1}{3}\begin{cases} 0, & x \le 0.2, \\ 0.05, & 0.2 < x \le 1, \\ 0.18, & 1 < x \le 1.8, \\ 0.43, & 1.8 < x \le 2.6, \\ 0.72, & 2.6 < x \le 3.4, \\ 0.91, & 3.4 < x \le 4.2, \\ 1, & x > 4.2 \end{cases} + \frac{1}{3}\begin{cases} 0, & x \le 1.3, \\ 0.02, & 1.3 < x \le 2.1, \\ 0.14, & 2.1 < x \le 2.9, 0.43, \\ 0.43, & 2.9 < x \le 3.7, \\ 0.77, & 3.7 < x \le 4.5, \\ 0.95, & 4.5 < x \le 5.2, \\ 1, & x > 5.2. \end{cases}$$

Отже, знайдено суміш розподілів на основі ланцюга Маркова та визначено ергодичний розподіл ймовірностей.

СПИСОК ЛІТЕРАТУРИ

1. Ивченко Г.И., Медведев Ю.И. Математическая статистика. Москва: Высш. шк., 1984. 248 с.
2. Турчин В.М. Математична статистика в прикладах і задачах. Київ: НМК ВО, 1993. 164 с.
3. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. Москва: Высш. шк., 2010. 600 с.
4. Постовалов С.И., Чимитова Е.В., Карманов В.С. Математическая статистика. Новосибирск: НГТУ, 2012. 159 с.

Ya.I. Yeleyko, O.A. Yarova

MIXTURE OF DISTRIBUTIONS BASED ON THE MARKOV CHAIN

Abstract. A Markov chain under the influence of environmental data is considered. A mixture of distributions is found on the basis of Kolmogorov’s criterion. Empirical distribution functions are constructed for samples of normal distributions, the ergodic distribution of the Markov chain is found, and the mixture of distributions is determined.

Keywords: Markov chain, matrix of transition probabilities, Kolmogorov criterion, normal distribution, mixture of distributions.

Надійшла до редакції 25.01.2022