УДК 519.25

В.Є. Бахрушин, І.О. Дудко

ВИЗНАЧЕННЯ КРИТИЧНИХ ЗНАЧЕНЬ КРИТЕРІЮ ТИПУ КОЛМОГОРОВА-СМИРНОВА ДЛЯ ДЕЯКИХ ТИПІВ РОЗПОДІЛУ МЕТОДОМ МОНТЕ-КАРЛО

Анотація. В статті розглядається питання про оцінювання критичних значень критерію типу Колмогорова-Смирнова методом Монте-Карло для випадку коли параметри моделі розподілу визначають шляхом мінімізації розрахункового значення цього критерію.

Ключові слова: ідентифікація, функція розподілу, критерій Колмогорова-Смирнова, метод Монте-Карло.

Постановка проблеми. На практиці часто виникає необхідність ідентифікації моделей розподілу. При цьому начастіше ставлять завдання підібрати модель із заданого класу, яка б найкраще описувала наявні емпіричні дані. Ця процедура складається з трьох етапів [1,2]:

1) формування гіпотези про закон (модель) розподілу, що перевіряється;

2) оцінювання параметрів обраної моделі;

3) перевірка адекватності моделі за допомогою певних статистичних критеріїв.

На останньому етапі найчастіше використовують критерії типу Колмогорова-Смирнова, омега-квадрат, хі-квадрат та ін.

У [2] було показано, що модель розподілу можна ідентифікувати шляхом мінімізації розрахункового значення статистичного критерію згоди. Але існує проблема вибору критичних значень для прийняття чи відхилення гіпотези про адекватність моделі наявним даним.

Аналіз останніх досліджень і публікацій. Х. Ліллієфорсом [3] було оцінено критичні значення критерію типу Колмогорова-Смирнова для випадку, коли параметри моделі нормального розподілу оцінюють як середнє арифметичне та стандартне відхилення досліджуваної вибірки. В його роботі було досліджено по 1000 або більше вибірок обсягом від 4 до 30 елементів. Пізніше ці результати були уточнені П. Моліном та Г. Абді [4] за допомогою сучасних комп'ютерних технологій, що дало змогу збільшити кількість досліджених вибірок до 100000, а їх максимальний обсяг – до 50 елементів. В обох випадках використовувався метод статистичних випробувань Монте-Карло.

В [5] нами було доведено наявність кореляції та визначено рівняння зв'язку, що дає змогу встановити приблизну відповідність між критичними значеннями для випадку, коли ідентифікацію параметрів розподілу здійснюють шляхом мінімізації розрахункового значення критерію типу Колмогорова – Смирнова та відповідними значеннями критерію Ліллієфорса. В цій роботі розглядається інший підхід до визначення критичних значень критерію типу Колмогорова-Смирнова.

Постановка завдання. Методом Монте-Карло визначити критичні значення критерію типу Колмогорова-Смирнова для нормального, гамма, експоненціального розподілів, а також розподілу Вейбулла.

Виклад основного матеріалу дослідження. Серед непараметричних критеріїв перевірки відповідності емпіричних розподілів тим чи іншим теоретичним моделям $\{F(x;\Theta), \theta \in \Theta\}$ виокремлюють критерії типу Колмогорова-Смирнова [1]. Вони визначають ступінь відповідності за максимальним за абсолютною величиною відхиленням моделі від емпіричної функції розподілу. Тобто, для перевірки двосторонньої нульової гіпотези $H_0 : F_n(x) = F(x, \theta_0)$ розраховують максимальну відстань між емпіричною $F_n(x)$ і теоретичною $F(x)$ функціями розподілу: $D_n = \sup\limits_{|x|<\infty} \left|F_n(x) - F(x, \theta_0)\right|$. Розрахункове значення критерію Колмогорова – Смирнова визначають за формулою:

$$\lambda = \sqrt{n} D_n(\theta_0), \qquad (1)$$

де n – обсяг вибірки.

Формулу (1) застосовують, якщо значення параметрів $\theta = \theta_0$ теоретичної функції розподілу відомі. Але на практиці зазвичай трапляється, що ці параметри невідомі досліднику. Тому необхідно використовувати вектор вибіркових оцінок θ^*. Розрахункове значення критерію в цьому випадку визначають за формулою:

$$\lambda^* = \sqrt{n} D_n\left(\theta^*\right) \qquad (2)$$

При цьому змінюються критичні значення критерію.

ISSN 1562-9945

В результаті необхідно розв'язати одну з таких задач:

– оцінити відповідність моделі $F_n(x) = F(x, \theta)$ з відомими параметрами наявним емпіричним даним (критерій Колмогорова-Смирнова);

– оцінити відповідність наявним емпіричним даним моделі $F_n(x) = F(x, \theta^*)$ з параметрами, які визначено за вибірковими моментами (критерій Ліллієфорса);

– оцінити відповідність наявним емпіричним даним моделі $F_n(x) = F_{\min}(x, \theta^*)$ з параметрами, що визначають з умови мінімуму деякого показника якості моделі, наприклад розрахункового значення критерію типу Колмогорова-Смирнова. Задача такого типу розв'язується в цій роботі.

Методика дослідження аналогічна використаній в роботах [3, 4] і ґрунтувалась на методі Монте-Карло. Вона полягає у наступному. Генерується велика кількість вибірок із заданим законом розподілу. Для кожної вибірки визначаються вибіркові параметри (наприклад, для нормального розподілу – це вибіркові середнє і стандартне відхилення) та розрахункове значення D_n^*. Далі, використовуючи отримані вибіркові параметри як початкове наближення, уточнюємо параметри моделі шляхом мінімізації розрахункових значень D_n^* методом багатовимірної нелінійної оптимізації. Потім будуємо емпіричну функцію розподілу отриманих значень $D_n^*(\min)$. Квантилі відповідних рівнів цієї функції розподілу складають таблицю шуканих значень статистики типу Колмогорова-Смирнова.

В проведеному дослідженні спочатку генерували по 10000 вибірок обсягом N від 6 до 45 елементів згідно зі стандартним нормальним $N(0,1)$, гамма $\Gamma(1, 0.5)$, експоненціальним $Exp(0.8)$ розподілами та розподілом Вейбулла $W(1, 2)$.

З квантилів функцій розподілу відповідних статистик $D_n^*(\min)$ склали таблиці. Зокрема, в таблиці 1 наведено значення 0,05-квантилів $D_n^*(\min)$ для досліджених моделей розподілу і різних обсягів вибірок. Також тут наведено для порівняння аналогічні значення для критеріїв Ліллієфорса $N_L(\mu, \sigma^2)$ [3] та Колмогорова – Смирнова $N_{K-S}(\mu, \sigma^2)$ [4].

Таблиця 1

0,05 квантилі розподілу $D_n^*(\min)$

Обсяг вибірки, N	Квантилі $D_n^*(\min)$				Квантилі D_n	
	$N(\mu,\sigma^2)$	$\Gamma(k,\theta)$	$Exp(\lambda)$	$W(k,\lambda)$	$N_L(\mu,\sigma^2)$	$N_{K-S}(\mu,\sigma^2)$
6	0,226	0,245	0,291	0,313	0,319	0,324
7	0,212	0,230	0,275	0,290	0,300	0,304
8	0,202	0,218	0,265	0,272	0,285	0,288
9	0,192	0,203	0,253	0,263	0,271	0,274
10	0,183	0,193	0,242	0,248	0,258	0,262
11	0,175	0,183	0,234	0,236	0,249	0,251
12	0,169	0,175	0,227	0,228	0,242	0,243
13	0,163	0,169	0,217	0,219	0,234	0,234
14	0,157	0,162	0,212	0,214	0,227	0,226
15	0,152	0,157	0,205	0,205	0,220	0,220
16	0,149	0,153	0,200	0,200	0,213	0,213
17	0,144	0,148	0,193	0,194	0,206	0,207
18	0,140	0,145	0,189	0,187	0,200	0,202
19	0,137	0,140	0,182	0,183	0,195	0,197
20	0,134	0,137	0,180	0,179	0,190	0,192
25	0,120	0,122	0,161	0,161	0,180	0,173
30	0,111	0,112	0,147	0,148	0,161	0,159
35	0,103	0,105	0,139	0,136	0,150	0,155
40	0,096	0,097	0,129	0,129	0,140	0,148
41	0,096	0,096	0,128	0,126	0,138	0,139
42	0,094	0,095	0,126	0,126	0,137	0,135
43	0,093	0,095	0,125	0,123	0,135	0,134
44	0,092	0,093	0,123	0,123	0,134	0,132
45	0,091	0,093	0,121	0,122	0,132	0,131

На рис. 1 наведено графіки залежності значень вказаних квантилів від обсягу вибірки для аналізованих розподілів.

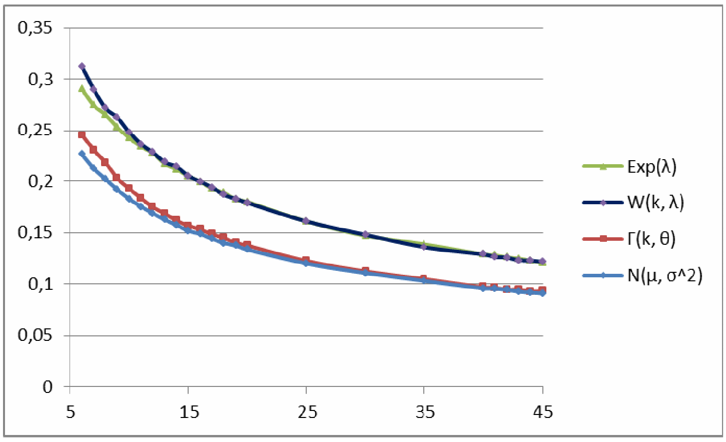

Рисунок 1 - Залежність 0,05 квантилів розподілу $D_n^*(\min)$ від обсягу вибірки

Для визначення граничних оцінок критичних значень критерію типу Колмогорова-Смирнова генерували по 10000 вибірок обсягом

 ISSN 1562-9945

1000 елементів і повторювали зазначену вище процедуру. Результати наведено у табл. 2.

Таблиця 2

Квантилі розподілу $D_n^*(min)$ для n = 1000

	Квантилі $D_n^*(\min)$				
Довірчий рівень, p	0,80	0,85	0,90	0,95	0,99
$N(\mu, \sigma^2)$	0,0171	0,0177	0,0184	0,0195	0,0216
$\Gamma(k, \theta)$	0,0170	0,0178	0,0187	0,0199	0,0227
$Exp(\lambda)$	0,0219	0,0227	0,0242	0,0259	0,0304
$W(k, \lambda)$	0,0222	0,0233	0,0245	0,0263	0,0297

Використовуючи формулу (2), отримали критичні значення $\lambda_{\min}^*$ критерію типу Колмогорова-Смирнова (табл. 3).

Таблиця 3

Критичні значення $\lambda_{\min}^*$ критерію типу Колмогорова-Смирнова

	Критичні значення $\lambda_{\min}^*$				
Довірчий рівень, p	0,80	0,85	0,90	0,95	0,99
$N(\mu, \sigma^2)$	0,5418	0,5605	0,5845	0,6171	0,6845
$\Gamma(k, \theta)$	0,5407	0,5630	0,5930	0,6296	0,7180
$Exp(\lambda)$	0,6925	0,7178	0,7652	0,8190	0,9613
$W(k, \lambda)$	0,7022	0,7382	0,7735	0,8311	0,9379
$N_L(\mu, \sigma^2)$	0,736	0,768	0,805	0,886	1.031
$N_{K-S}(\mu, \sigma^2)$	0,775	0,819	0,895	0,955	1,035

Висновки з проведеного дослідження. Показано можливість та оцінено методом Монте-Карло критичні значення критерію згоди типу Колмогорова – Смирнова для деяких типів розподілу у випадку, коли параметри моделі розподілу уточнюють шляхом мінімізації розрахункового значення цього критерію. У подальшому передбачається уточнити довірчі інтервали для отриманих оцінок і перевірити їх можливу залежність від інших факторів.

ЛІТЕРАТУРА

1. Бахрушин В.Є. Методи аналізу даних / В.Є. Бахрушин.– Запоріжжя:КПУ, 2011.–268 с.
2. Бахрушин В.Е. Проблемы идентификации моделей распределения случайных величин с применением современного программного обеспечения / В.Є. Бахрушин // Успехи современного естествознания. – 2011. – № 11. – С. 50 – 54.
3. Lilliefors H. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. Journal of the ASA, Vol. 62, No. 318 (Jun., 1967), 399-402.
4. Molin, P., Abdi H. (1998). New Tables and numerical approximation for the Kolmogorov- Smirnov/Lillierfors/Van Soest test of normality. Technical Report, University of Bourgogne (France) / University of Texas at Dallas (USA).
5. Бахрушин В.Є. Уточнення моделей нормального розподілу на основі мінімізації критерію Колмогорова – Смирнова / В.Є. Бахрушин, І.О. Дудко // Системні технології. Регіональний міжвузівський збірник наукових праць. – Випуск 5 (82). – Дніпропетровськ, 2012. – С. 95 –103.