УДК 519.21

$\boxed{\text{І.Д.Квіт}}$, Є.В.Москвяк

ГІПОТЕЗА ПРО РОЗПОДІЛ

Нехай

$$x_{(\bar{1})} < x_{(\bar{2})} < \dots < x_{(\bar{j})} < \dots < x_{(\bar{k})} \qquad /1/$$

варіаційний ряд як завгодно зрізаної вибірки обсягом n незалежних спостережень над абсолютно неперервною випадковою змінною ξ, де $\bar{j}$ – середній ранг. Ряд /1/ може бути побудований на основі, наприклад, повної вибірки, підмножини повної вибірки, згрупованої вибірки, справа зрізаної вибірки, з двох боків зрізаної вибірки і т.п. Потрібно перевірити гіпотезу про те, що вибірку одержано з популяції, керованої функцією розподілу ймовірностей

$$F(x) = P\{\xi \leq x\}. \qquad /2/$$

Якщо гіпотеза /2/ істинна, то варіаційний ряд з елементами

$$r_{(\bar{j})} = F(x_{(\bar{j})}), \quad (j = 1, \dots, k) \qquad /3/$$

є значеннями рівномірного на інтервалі /0,1/ розподілу. Ряд /3/ перетворюється в повну випадкову стандартно рівномірну вибірку обсягу k зі значеннями $r'_1, r'_2, \dots, r'_k$, за формулами

$$r'_1 = B_{\bar{1},\, n+1-\bar{1}}\left(r_{(\bar{1})}\right), \quad r'_2 = B_{\bar{2}-\bar{1},\, n+1-\bar{2}}\left(\frac{r_{(\bar{2})} - r_{(\bar{1})}}{1 - r_{(\bar{1})}}\right), \dots,$$

$$r'_i = B_{\bar{i}-\overline{i-1},\, n+1-\bar{i}}\left(\frac{r_{(\bar{i})} - r_{(\overline{i-1})}}{1 - r_{(\overline{i-1})}}\right), \ (i = 1, \dots, k), \quad r_{(\bar{0})} \equiv 0, \dots \qquad /4/$$

$$r'_k = B_{\bar{k}-\overline{k-1},\, n+1-\bar{k}}\left(\frac{r_{(\bar{k})} - r_{(\overline{k-1})}}{1 - r_{(\overline{k-1})}}\right),$$

де

$$B_{(a,b)}(x) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1}\,dt, \quad 0<x<1, \ (a>0,\ b>0). \qquad /5/$$

Бета-розподіл /5/ табульований [2]. Він тісно пов"язаний з F - розподілом [1].

Зазначимо, що для повної вибірки формули /4/ набувають вигляду

$$r_1' = B_{1,n}(r_{(1)}) = 1-(1-r_{(1)})^n, \quad r_2' = B_{1,n-1}\left(\frac{r_{(2)}-r_{(1)}}{1-r_{(1)}}\right) = 1-\left(\frac{1-r_{(2)}}{1-r_{(1)}}\right)^{n-1}, \ldots,$$

$$r_i' = B_{1,n+1-i}\left(\frac{r_{(i)}-r_{(i-1)}}{1-r_{(i-1)}}\right) = 1-\left(\frac{1-r_{(i)}}{1-r_{(i-1)}}\right)^{n+1-i}, (i=1,\ldots,n),\ r_{(0)}\equiv 0, \ldots, \qquad /6/$$

$$r_n' = B_{1,1}\left(\frac{r_{(n)}-r_{(n-1)}}{1-r_{(n-1)}}\right) = 1-\frac{1-r_{(n)}}{1-r_{(n-1)}}.$$

Звідси

$$r_{(1)} = 1-(1-r_1')^{\frac{1}{n}}, \quad r_{(2)} = 1-(1-r_1')^{\frac{1}{n}}(1-r_2')^{\frac{1}{n-1}}, \ldots,$$

$$r_{(i)} = 1-\prod_{j=1}^{i}(1-r_j')^{\frac{1}{n+1-j}}, \quad (i=1,\ldots,n), \ldots, \qquad /7/$$

$$r_{(n)} = 1-(1-r_1')^{\frac{1}{n}}(1-r_2')^{\frac{1}{n-1}}\ldots(1-r_n').$$

Формули /6/ і /7/ взаємно обернені /див. [3] /. Формули /6/ переводять варіаційний ряд повної стандартно рівномірної вибірки у повну випадкову. Формули /7/ утворюють варіаційний ряд на основі повної випадкової стандартно рівномірної вибірки.

Використовуючи формули /7/, при $n=K$ дістаємо варіаційний ряд

$$r^*_{(1)} < r^*_{(2)} < \ldots < r^*_{(i)} < \ldots < r^*_{(K)} \qquad /8/$$

повної стандартно рівномірної вибірки обсягу K. Таким чином, перевірка гіпотези про те, що варіаційний ряд /1/ побудований на основі вибірки з популяції, керованої розподілом /2/, зводиться до перевірки гіпотези про те, що варіаційний ряд /8/ належить до повної стандартно рівномірної вибірки. Останню гіпотезу можемо перевірити, наприклад, за допомогою статистики Андерсона-Дарлінга /порівн. [3] :

$$A^2 = -\frac{1}{K}\sum_{i=1}^{K}(2i-1)\ln\left[r^*_{(i)}\left(1-r^*_{(K+1-i)}\right)\right] - K. \qquad /9/$$

5-2099

Розподіл статистики /9/ майже не залежить від k. Тому критерієм A^2 практично користуються починаючи від $k=3$. Критичні значення статистики A^2 при найчастіше вживаних рівнях значущості α такі

α	0,10	0,05	0,025	0,01
$A^2_{крит.}$	1,923	2,492	3,070	3,857

. /10/

Якщо при заданому рівні значущості α емпіричне значення статистики A^2 більше від критичного, то початкову гіпотезу про те, що варіаційний ряд /1/ одержано на основі вибірки з популяції, керованої розподілом /2/, відхиляємо. У протилежному випадку вважаємо, що гіпотеза /2/ не суперечить експериментальним даним /1/. Із двох допустимих розподілів перевагу надаємо тому, що для нього емпіричне значення статистики A^2 менше.

Приклади. 1. Після впорядкування вибірки обсягом $n = 15$ залишилося

$$X_{(3)} = 1928; \quad X_{(7)} = 5710; \quad X_{(10)} = 9939; \quad X_{(12)} = 14260.$$

Решта значень випадково втрачені або затерті. Перевірити гіпотезу про те, що розглянувана вибірка взята з популяції, керованої функцією розподілу ймовірностей

$$F(x) = 1 - e^{-\frac{x}{10000}}, \quad x > 0. \qquad /11/$$

За формулами /3/ і /11/

$r_{(3)} = 0,1753532$; $r_{(7)} = 0,4350399$; $r_{(10)} = 0,6298696$; $r_{(12)} = 0,7597319$.
Далі, за о формулою /4/

$$r_1' = B_{3,13}\,/0,1753532/ = 0,5234371;$$

$$r_2' = B_{1,9}\left/\frac{0,4350399-0,1753532}{1-0,1753532}\right/ = 0,5530561;$$

$$r_3' = B_{3,6}\left/\frac{0,6298696-0,4350399}{1-0,4350399}\right/ = 0,5601338;$$

$$r_4' = B_{2,4}\left/\frac{0,7597319-0,6298696}{1-0,6298696}\right/ = 0,571585.$$

Звідси, використовуючи формулу /7/ при $n = 4$, дістаємо
$r^*_{(1)} = 0,1691359$; $r^*_{(2)} = 0,3647456$; $r^*_{(3)} = 0,5786842$;
$r^*_{(4)} = 0,8195023$ варіаційний ряд повної стандартно рівномірної вибірки обсягу 4. На його основі обчислюємо емпіричне значення статистики /9/

$$A^2 = -\frac{1}{4}\sum_{i=1}^{4}(2i-1)\ln\left[r^*_{(i)}\left(1-r^*_{(5-i)}\right)\right]-4 = 0,2004875.$$

Це значення, згідно з таблицею /10/, менше від кожного з критичних значень відповідних заданим рівням значущості. Отже, гіпотеза про розподіл /11/ приймається.

2. Дано напрацювання в годинах до відмови F або зупинки S

200 S 511 F 674 F 700 S 900 S 1213 F 1536 F

Перевірити гіпотезу про те, що розглядувана вибірка взята з популяції, керованої функцією розподілу ймовірностей

$$F(x) = 1 - e^{-\left(\frac{x}{1000}\right)^2}, \quad x > 0. \qquad /12/$$

Середні ранги чотирьох відмов відповідно дорівнюють $\bar{1}=1,142857$; $\bar{2}=2,285714$; $\bar{3}=4,190476$; $\bar{4}=6,095238$.
За формулами /3/ і /12/

$r_{(1,142857)} = 0,229812$; $r_{(2,285714)} = 0,3650925$;

$r_{(4,190476)} = 0,770389$; $r_{(6,095238)} = 0,9055132$.

Далі, за формулою /4/

$$r'_1 = B_{1,142857;6,857143}/0,229812/ = 0,7951582;$$

$$r'_2 = B_{1,142857;5,714286}/0,1756458/ = 0,6298049;$$

$$r'_3 = B_{1,904762;3,809524}/0,6383551/ = 0,5885130;$$

$$r'_4 = B_{1,904762;1,904762}/0,5884918/ = 0,6101330.$$

Звідси, використовуючи формулу /7/ при $n = 4$, дістаємо
$r^*_{(1)} = 0,3272489$; $r^*_{(2)} = 0,5169424$; $r^*_{(3)} = 0,6901323$;
$r^*_{(4)} = 0,8791927$ варіаційний ряд повної стандартно рівномірної вибірки обсягу 4. Відповідне емпіричне значення статистики

/9/ дорівнює A^2 = 0,4733262 та потрапляє в область прийому гіпотези /I2/.

3. Дані про згруповані відмови та зупинки n = I00 однотипних технічних одиниць:

Напрацювання в інтервалі часу	Було відмов	Було зупинок
До t_1 = 35 тижнів	f_1 = 25	s_1 = I5
Від t_1 до t_2 = 70 тижнів	f_2 = I5	s_2 = 9
Від t_2 до t_3 = I05 тижнів	f_3 = 9	s_3 = 4
Від t_3 до t_4 = I40 тижнів	f_4 = 7	s_4 = 3
Після I40 тижнів	Не було	s_5 = I3

Потрібно перевірити гіпотезу про те, що дана вибірка з генеральної сукупності, керованої функцією розподілу ймовірностей

$$F(t) = 1 - e^{-\frac{t}{100}}, \quad t > 0. \qquad /I3/$$

У даному разі середні ранги 25-ї, 40-ї, 49-ї та 56-ї відмови відповідно дорівнюють

$\bar{1}$=29,36046; $\bar{2}$=50,0257I; $\bar{3}$=63,92778; $\bar{4}$=76,285I8.

За формулами /3/ і /I3/

$r_{(\bar{1})}$ = 0,2953II9; $r_{(\bar{2})}$ = 0,5034I46; $r_{(\bar{3})}$ = 0,650062I;
$r_{(\bar{4})}$ = 0,753403I.

Далі, за формулою /4/

$r_1' = B_{29,36046;7I,63954}$/0,2953II9/ = 0,4072094;

$r_2' = B_{20,66525;50,97429}$/0,2953II7/ = 0,40I2985;

$r_3' = B_{I3,90207;37,07222}$/0,2953II7/ = 0,68I9728;

$r_4' = B_{I2,3574;24,74482}$/0,2953I23/ = 0,347360I.

Звідси, використовуючи формулу /7/ при n = 4, дістаємо
$r^*_{(1)}$ = 0,I225445; $r^*_{(2)}$ = 0,2604602; $r^*_{(3)}$ = 0,582945;
$r^*_{(4)}$ = 0,7278I4 варіаційний ряд повної стандартно рівномірної вибірки обсягу /4/. Оскільки емпіричне значення статистики /9/ дорівнює A^2 = 0,35I53 < 2,492, то гіпотеза /I3/ не суперечить експериментальним даним.

1. Мардия К., Земроч П., Таблицы F-распределений. М., 1984. 2. Пирсон К. Таблицы неполной бета-функции. М., 1974. 3. O'Reilly F.J., Stephens M.A., Transforming Censored Samples for Testing Fit // Technometrics. 1988. Vol. 30, p. 79-86.

Стаття надійшла до редколегії 17.04.90

УДК 519.21

О.П.Гнатишин, Є.В.Москвяк

МЕТОД МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ ДЛЯ ЗРІЗАНОЇ ВИБІРКИ

Нехай абсолютно неперервна випадкова змінна t з функцією розподілу ймовірностей $F(t;\theta)$, залежною від s невідомих параметрів $\theta=(\theta_1,\ldots,\theta_s)$, $(s=1,2)$, описує напрацювання технічної одиниці до відмови. Тоді надійність $R(t;\theta)$, густина ймовірностей $f(t;\theta)$, інтенсивність відмов $h(t;\theta)$ і риск $H(t;\theta)$ відповідно дорівнюють

$$R=1-F=e^{-H},\quad f=F',\quad h=\frac{f}{R}=-\frac{d}{dt}\ln R,\quad H(t;\theta)=\int_0^t h(x;\theta)\,dx. \qquad /1/$$

1. Незгруповані відмови та зупинки. Нехай k з n однотипних технічних одиниць працювали до відмови F, а решта $n-k$ до зупинки S. Позначимо моменти відмов через t_{iF} $(i=1,\ldots,k)$, а моменти зупинок через t_{jS}, $(j=1,\ldots,n-k)$. У припущенні, що напрацювання до відмови описується розподілом $F(t;\theta)$, потрібно оцінити невідомі параметри $\theta=(\theta_1,\ldots,\theta_s)$.

Згідно з методом максимальної правдоподібності /ММП/ записуємо функцію правдоподібності

$$L=\prod_{i=1}^{k} f(t_{iF};\theta)\prod_{j=1}^{n-k} R(t_{jS};\theta).$$

Звідси

$$\ln L=\sum_{i=1}^{k}\ln f(t_{iF};\theta)+\sum_{j=1}^{n-k}\ln R(t_{jS};\theta), \qquad /2/$$

логарифм правдоподібності виражено через густину та надійність.