

УДК 519:233:2

О.І. Білобродько, І.В. Мозгова

*Дніпропетровський національний університет ім. О. Гончара*

## **ВІДТВОРЕННЯ СПЛАЙН-РОЗПОДІЛУ ВЕЙБУЛЛА МЕТОДОМ КВАНТИЛІВ ЗА ЦЕНЗУРОВАНИМИ ДАНИМИ**

Наведено теоретичні засади та практичні розрахунки застосування методу квантилів для знаходження параметрів сплайн-розподілу Вейбулла з одним вузлом за цензурованими даними.

**Постановка проблеми.** У задачах аналізу статистичних даних мають місце випадки, коли певна частина систем, що випробовуються або контролюються, не дала результатів за час спостереження, а час реалізації іншої частини точно не відомий. У цьому разі виникає необхідність проведення статистичного аналізу на основі специфічних вибірових даних, особливою яких є відсутність відомостей про моменти реалізації частини спостережень, що досліджуються. Це явище носить назву цензурування даних, а отримані вибірки – цензурованими вибірками (ЦВ) [2]. Так, наприклад, у теорії

---

© О.І. Білобродько, І.В. Мозгова, 2008

надійності, даними ЦВ є наробітки об'єктів, що відмовили, наробітки об'єктів, що не відмовили, а також інтервали часу, впродовж яких об'єкт відмовив, але момент відмови точно невідомий. Первинний аналіз та обробка таких даних мають певні особливості, що достатньо фундаментально описані в [1 – 2].

Слід зазначити, що характер певних вибірових даних, у тому числі й цензурованих, не завжди є однорідним. Тому актуальною задачею є розробка алгоритмів відтворення розподілів у вигляді сумішей [3] та сплайн-розподілів [5]. Для класичних вибірових даних існують чисельні публікації в цьому напрямку, водночас питанню ефективності використання моделей сумішей та сплайн-розподілів на основі цензурованих даних не приділено належної уваги.

**Аналіз останніх досягнень.** Широке розповсюдження у практиці статистичної обробки даних отримав розподіл Вейбулла, який традиційно використовується для опису часу відмов приладів, що ставляться на випробовування.

Процедури відтворення чистого розподілу Вейбулла за цензурованими даними на основі класичних методів математичної статистики є досить відомими [1; 2; 6]. У [4] наведено результати щодо застосування модифікованих методів до відтворення суміші двох розподілів Вейбулла за цензурованими даними. Однак невирішеним залишається питання використання методів математичної статистики, розробки алгоритмів та програмного забезпечення для відтворення сплайн-розподілів за зазначеними даними.

**Постановка задачі.** Розглянемо застосування методу квантилів до відтворення сплайн-розподілу Вейбулла з одним вузлом [5], що має функцію розподілу

$$F(t, \alpha_1, \beta_1, \beta_2, t_0) = \begin{cases} 1 - \exp\left(-\frac{t^{\beta_1}}{\alpha_1}\right), & 0 \leq t \leq t_0; \\ 1 - \exp\left(-\frac{t_0^{\beta_1}}{\alpha_1} \left(\frac{t}{t_0}\right)^{\beta_2}\right), & t_0 \leq t < \infty. \end{cases} \quad (1)$$

Метод квантилів полягає в тому, що висувають гіпотезу про тип функції розподілу  $F(t, \vec{\Theta})$  і розв'язують систему рівнянь

$$F\left(t_i, \vec{\Theta}\right) = F_n(t_i), \quad i = \overline{1, k}, \quad (2)$$

де  $n$  – кількість спостережень,  $k$  – кількість параметрів розподілу.

Момент  $t_i$  появи  $l_i$ -го спостереження можна розглядати як емпіричну квантиль, що відповідає рівню  $q_i = \frac{l_i}{n}$ ,  $0 < q_i < 1$ . Вибір  $t_i$ , а отже й  $F_n(t_i)$ , визначається відповідно до рівняння  $F_n(t_i) = q_i$ . У випадку цензурованих вибірових даних відомими є лише наближені значення квантилів. Розглянемо випадок, коли вузол  $t_0 = t_i, i = \overline{1, m}$ , де  $m < n$  – кількість спостережень після цензурування. Замінюючи в рівняннях (2) значення квантилів їхніми оцінками для сплайн-розподілу Вейбулла з одним вузлом отримаємо систему

$$F(t_i, \hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{t}_0) = \frac{i}{n}, \quad i = \overline{1, 3}.$$

Оцінки параметрів  $\vec{\hat{\Theta}} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{t}_0)$  можна одержати з умови

$$\rho_0 = \min_{v=2, m-2} \max_{i=1, m} \left| F_n(t_i) - F(t_i, \vec{\hat{\Theta}}_v) \right|,$$

де  $n$  – загальна кількість спостережень,  $m$  – кількість спостережень після цензурування. Не обмежуючи загальності дослідження розглянемо випадок цензурування праворуч. Позначимо момент цензурування як  $T_2$ . Нехай  $t_2 < T_1 < t_0$ ,  $T_1 < T_0 < t_0$ ,  $t_0 < T_2 < t_{n-2}$ . Для сплайн-розподілу Вейбулла з одним вузлом маємо наступну систему рівнянь

$$\begin{cases} 1 - e^{-\frac{T_1^{\hat{\beta}_1}}{\hat{\alpha}_1}} = F_n(T_1) \\ 1 - e^{-\frac{T_0^{\hat{\beta}_1}}{\hat{\alpha}_1}} = F_n(T_0) \\ 1 - e^{-\frac{T_2^{\hat{\beta}_2}}{\hat{\alpha}_1 T_0^{\hat{\beta}_2 - \hat{\beta}_1}}} = F_n(T_2) \end{cases} \quad (3)$$

**Основна частина дослідження.** У результаті перетворень одержимо систему (3) у вигляді:

$$\begin{cases} \ln(1 - F_n(T_1)) = -\frac{T_1^{\hat{\beta}_1}}{\hat{\alpha}_1} \\ \ln(1 - F_n(T_0)) = -\frac{T_0^{\hat{\beta}_1}}{\hat{\alpha}_1} \\ \ln(1 - F_n(T_2)) = -\frac{T_2^{\hat{\beta}_2}}{\hat{\alpha}_1 T_0^{\hat{\beta}_2 - \hat{\beta}_1}}. \end{cases} \quad (4)$$

Звідки отримаємо значення оцінок параметрів:

$$\begin{aligned} \hat{\alpha}_1 &= -\frac{T_1^{\hat{\beta}_1}}{\ln(1 - F_n(T_1))}; \\ \hat{\beta}_1 &= \frac{\ln(\ln(1 - F_n(T_0))) - \ln(\ln(1 - F_n(T_1)))}{\ln(T_0) - \ln(T_1)}. \end{aligned} \quad (5)$$

Розв'яжемо третє рівняння з системи рівнянь (4) відносно  $\hat{\beta}_2$

$$\ln(1 - F_n(T_2)) = -\frac{T_0^{\hat{\beta}_1}}{\hat{\alpha}_1} \left( \frac{T_2}{T_0} \right)^{\hat{\beta}_2}.$$

Підставимо замість  $\alpha_1$  вираз з (5), отримаємо

$$\ln(1 - F_n(T_2)) = \left( \frac{T_0}{T_1} \right)^{\hat{\beta}_1} \left( \frac{T_2}{T_0} \right)^{\hat{\beta}_2} \ln(1 - F_n(T_1)).$$

Виконаємо наступні перетворення:

$$\begin{aligned} \frac{\ln(1 - F_n(T_2))}{\ln(1 - F_n(T_1))} &= \left( \frac{T_0}{T_1} \right)^{\hat{\beta}_1} \left( \frac{T_2}{T_0} \right)^{\hat{\beta}_2}; \\ \ln \left( \frac{\ln(1 - F_n(T_2))}{\ln(1 - F_n(T_1))} \right) &= \hat{\beta}_1 \ln \left( \frac{T_0}{T_1} \right) + \hat{\beta}_2 \ln \left( \frac{T_2}{T_0} \right); \\ \hat{\beta}_2 \ln \left( \frac{T_2}{T_0} \right) &= \ln \left( \frac{\ln(1 - F_n(T_2))}{\ln(1 - F_n(T_1))} \right) - \hat{\beta}_1 \ln \left( \frac{T_0}{T_1} \right); \end{aligned}$$

$$\hat{\beta}_2 = \frac{\ln\left(\frac{\ln(1-F_n(T_2))}{\ln(1-F_n(T_1))}\right) - \hat{\beta}_1 \ln\left(\frac{T_0}{T_1}\right)}{\ln\left(\frac{T_2}{T_0}\right)}.$$

Використовуючи замість  $\hat{\beta}_1$  вираз з (5), отримаємо:

$$\hat{\beta}_2 = \frac{\ln\left(\frac{\ln(1-F_n(T_2))}{\ln(1-F_n(T_1))}\right) - \frac{\ln(\ln(1-F_n(T_0))) - \ln(\ln(1-F_n(T_1)))}{\ln(T_0) - \ln(T_1)} \ln\left(\frac{T_0}{T_1}\right)}{\ln\left(\frac{T_2}{T_0}\right)}.$$

Виконаємо перетворення:

$$\hat{\beta}_2 = \frac{\ln\left(\frac{\ln(1-F_n(T_2))}{\ln(1-F_n(T_1))}\right) - (\ln(\ln(1-F_n(T_0))) - \ln(\ln(1-F_n(T_1))))}{\ln(T_2) - \ln(T_0)}.$$

Отримаємо оцінку параметра  $\beta_2$

$$\hat{\beta}_2 = \frac{\ln(\ln(1-F_n(T_2))) - \ln(\ln(1-F_n(T_0)))}{\ln(T_2) - \ln(T_0)}. \quad (6)$$

З метою знаходження дисперсій оцінок параметрів скористаємося методом розкладання в ряд Тейлора. Зауважимо, що функція (1) має неперервні частинні похідні по  $t$  та параметрам  $\alpha_1, \beta_1, \beta_2$  якщо  $t \neq t_0$ .

Введемо позначення

$$q_i = \frac{l_i}{n}, \quad \delta t_i = t_{l_i} - t_{q_i}, \quad t_{l_i} = F^{-1}\left(q_i, \bar{\Theta}\right), \quad i = \overline{1,3}; \quad .$$

$$\delta \alpha_1 = \hat{\alpha}_1 - \alpha_1, \quad \delta \beta_1 = \hat{\beta}_1 - \beta_1, \quad \delta \beta_2 = \hat{\beta}_2 - \beta_2.$$

З точністю до безкінечно малих отримаємо систему [1]

$$\begin{aligned} & \frac{\partial F(t_{q_i}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{t}_0)}{\partial t} \delta t_i + \frac{\partial F(t_{q_i}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{t}_0)}{\partial \alpha_1} \delta \alpha_1 + \\ & + \frac{\partial F(t_{q_i}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{t}_0)}{\partial \beta_1} \delta \beta_1 + \frac{\partial F(t_{q_i}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{t}_0)}{\partial \beta_2} \delta \beta_2 = 0, \quad i = \overline{1,3}. \end{aligned} \quad (7)$$

Розв'язуючи систему лінійних рівнянь (7) відносно похибок  $\delta \alpha_1, \delta \beta_1, \delta \beta_2$ , отримаємо

$$\begin{aligned}
\delta\alpha_1 &= A_{11}\delta T_1 + A_{10}\delta T_0; \\
\delta\beta_1 &= A_{21}\delta T_1 + A_{20}\delta T_0; \\
\delta\beta_2 &= A_{31}\delta T_1 + A_{30}\delta T_0 + A_{32}\delta T_2,
\end{aligned} \tag{8}$$

де

$$\begin{aligned}
A_{11} &= \frac{\hat{\alpha}_1 \hat{\beta}_1 \ln T_0}{T_1 (\ln T_1 - \ln T_0)}; \quad A_{10} = -\frac{\hat{\alpha}_1 \hat{\beta}_1 \ln T_1}{T_0 (\ln T_1 - \ln T_0)}; \\
A_{21} &= -\frac{\hat{\beta}_1}{T_1 (\ln T_1 - \ln T_0)}; \quad A_{20} = \frac{\hat{\beta}_1}{T_0 (\ln T_1 - \ln T_0)}; \\
A_{31} &= -\frac{\hat{\beta}_1 (1 + \hat{\alpha}_1 \ln T_0)}{T_1 (\ln T_2 - \ln T_0)}; \quad A_{30} = \frac{\hat{\alpha}_1 \hat{\beta}_1 \ln T_1}{T_0 (\ln T_2 - \ln T_0)}; \\
A_{32} &= \frac{\hat{\beta}_2}{T_2 (\ln T_2 - \ln T_0)}.
\end{aligned}$$

Дисперсії оцінок параметрів знайдемо з рівнянь:

$$\begin{cases} D(\hat{\alpha}_1) = A_{11}^2 M(\delta T_1)^2 + 2A_{11}A_{10}M(T_1)M(T_0) + A_{10}^2 M(\delta T_0)^2; \\ D(\hat{\beta}_1) = A_{21}^2 M(\delta T_1)^2 + 2A_{21}A_{20}M(T_1)M(T_0) + A_{20}^2 M(\delta T_0)^2; \\ D(\hat{\beta}_2) = A_{31}^2 M(\delta T_1)^2 + 2A_{31}A_{30}M(T_1)M(T_0) + A_{30}^2 M(\delta T_0)^2 + \\ + 2A_{31}A_{32}M(T_1)M(T_2) + 2A_{30}A_{32}M(T_0)M(T_2) + A_{32}^2 M(\delta T_2)^2. \end{cases} \tag{9}$$

Для перевірки гіпотези про достовірність оцінок скористаємося

$$\text{статистиками } U_{\alpha_1} = \frac{|\alpha_1 - \hat{\alpha}_1|}{\sqrt{D(\hat{\alpha}_1)}}, \quad U_{\beta_1} = \frac{|\beta_1 - \hat{\beta}_1|}{\sqrt{D(\hat{\beta}_1)}}, \quad U_{\beta_2} = \frac{|\beta_2 - \hat{\beta}_2|}{\sqrt{D(\hat{\beta}_2)}}.$$

Таким чином, маємо наступний алгоритм відтворення сплайн-розподілу Вейбулла з одним вузлом для цензурованих даних:

1. Застосувати процедури первинного статистичного аналізу до вибірових даних з урахуванням характеру даних.
2. Визначити  $T_2$  – момент цензурування праворуч. Тут і надалі обрати в якості  $T_j$   $j$ -й елемент вибірки.
3. Обрати в якості  $T_0$   $j$ -й елемент згенерованої вибірки. Вважати вузел  $\hat{t}_0 = t_{j+1}$ .
4. Обрати  $T_1$  з умови  $T_1 < T_0, i = 1$ .

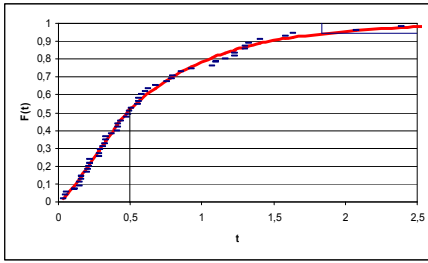
5. За виразом (6) обчислити оцінки  $\hat{\alpha}_1$  та  $\hat{\beta}_1$ , за виразом (6) оцінку  $\hat{\beta}_2$ .
  6. Обчислити  $\max_{i=\overline{1,n}} \left| F_n(t_i) - F\left(t_i, \vec{\Theta}_v\right) \right|$ , де  $v$  – номер поточної ітерації.
  7. Обчислити  $\rho_0 = \min_{v=2, m-2} \max_{i=\overline{1,m}} \left| F_n(t_i) - F\left(t_i, \vec{\Theta}_v\right) \right|$ , де  $n$  – загальна кількість спостережень,  $m$  – кількість спостережень після цензурування.
  8. Обрати наступне значення  $T_1$  з умови  $i < j$  та перейти на крок 7.
  9. Обрати наступний вузол  $T_0, j < n$  та перейти на крок 6.
  10. Обчислити довірчі інтервали оцінок згідно (9). Перевірити достовірність отриманих оцінок.
  11. Виконати перевірку вірогідності щодо збігу емпіричного розподілу з теоретичним за уточненим критерієм згоди Колмогорова.
- У таблиці 1 та на рисунку 1 наведено результати застосування алгоритму до відтворення сплайн-розподілу Вейбулла за згенерованими даними та цензуруванням праворуч.

Таблиця 1

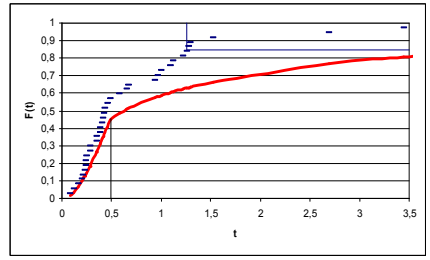
**Параметри генерації та результати відтворення**

Параметри генерації	Результат відтворення	$\Delta K$	$U$	Рис.
$n = 50, m = 40, \alpha_1 = 0.4,$ $\beta_1 = 1.5, \beta_2 = 1.1,$ $t_0 = 0.5$	$\hat{\alpha}_1 = 0.52,$ $\hat{\beta}_1 = 1.38, \beta_2 = 1.05,$ $t_0 = 0.5049$	0.86	$U_{\alpha_1} = 1.095$ $U_{\beta_1} = 0.93$ $U_{\beta_2} = 2.23$	1.а
$n = 50, m = 40, \alpha_1 = 0.4,$ $\beta_1 = 2, \beta_2 = 0.5,$ $t_0 = 0.5$	$\hat{\alpha}_1 = 0.46,$ $\hat{\beta}_1 = 2.17, \beta_2 = 0.47,$ $t_0 = 0.487$	0.32	$U_{\alpha_1} = 0.635$ $U_{\beta_1} = 1.35$ $U_{\beta_2} = 2.095$	1.б
$n = 50, m = 30, \alpha_1 = 2,$ $\beta_1 = 0.5, \beta_2 = 2,$ $t_0 = 0.5$	$\hat{\alpha}_1 = 1.96,$ $\hat{\beta}_1 = 0.44, \beta_2 = 2.47,$ $t_0 = 0.5053$	0.51	$U_{\alpha_1} = 0.17$ $U_{\beta_1} = 0.468$ $U_{\beta_2} = 1.85$	1.в

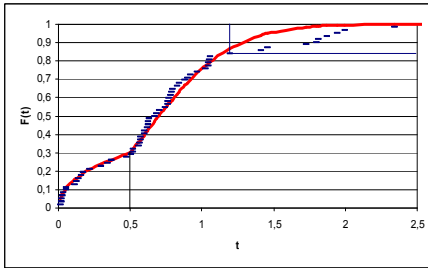
$\Delta K$  – вірогідність за уточненим критерієм Колмогорова



а



б



в

- а)  $\beta_1 \approx \beta_2 > 0$  ;
- б)  $\beta_1 > 1$ ;  $\beta_2 < 1$ ;
- в)  $\beta_1 < 1$ ;  $\beta_2 > 1$ .

**Рис. 1. Результати роботи алгоритмів генерації та відтворення сплайн-розподілу Вейбулла з одним вузлом**

**Висновки та перспективи.** Очевидно, що методи аналізу цензурованих вибірок відрізняються від методів обробки повних вибірок: цензурована вибірка містить менше інформації, що безпосередньо впливає на точність оцінювання параметрів апроксимуючого закону розподілу. З іншого боку розглянутий в роботі метод квантилів є одним з найбільш універсальних відносно типу цензурування. При цьому досить висока точність оцінок параметрів досягається шляхом вибору найбільш оптимальних квантилів. Проведені експериментальні дослідження показали безпосередній вплив цензурування на достовірність отриманих оцінок. У випадку великих дисперсій отримані методом квантилів оцінки параметрів можуть бути використаними в якості першого наближення з метою подальшого уточнення.



## Бібліографічні посилання

1. **Гнеденко Б.В.** Математические методы в теории надежности / Б.В. Гнеденко, Ю.К. Беляев, А.Д. Соловьев – М., 1965. – 542 с.
2. **Скрипник В.М.** Анализ надежности технических систем по цензурированным выборкам / В.М. Скрипник, А.Е. Назин, Ю.Г. Приходько, Благовещенский Ю.Н. – М., 1988. – 184 с.
3. **Миленький А.В.** Классификация сигналов в условиях неопределенности. Стат. методы самообучения и распознавания образов. – М., 1975. – 328 с.
4. **Мозговая И.В.** Восстановление смеси распределений Вейбулла по цензурированным данным /И.В. Мозговая // Актуальные проблемы автоматизации и информационных технологий. – Д., 999 – Т.2. С. 80 – 88.
5. **Приставка А.Ф.** Сплайн-распределения в статистическом анализе / А.Ф. Приставка // – ,1996. – 50 с.
6. ГОСТ 27.504–84. Надежность в технике. Методы оценки показателей надежности по цензурированным выборкам. – М., 1984.

*Надійшла до редколегії 06.10.08*