

Testowanie zgodności z rozkładem Weibulla

Tomasz Gładki

Promotor: dr Grzegorz Wyłupek

2025

Wprowadzenie

Rozważamy przypadek hipotezy złożonej, w której parametry opisujące rozkład są nieznane. Dla ścisłości niech

$$\mathcal{W} = \{W(\lambda, k) \mid \lambda, k > 0\}$$

będzie rodziną rozkładów Weibulla, gdzie $W(\lambda, k)$ jest rozkładem Weibulla definiowanym przez następującą funkcję gęstości:

$$f(x, \lambda, k) = \frac{k}{\lambda^k} x^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right), \quad x \geq 0,$$

gdzie $\lambda > 0$ jest parametrem skali, a $k > 0$ parametrem kształtu.

Problem testowania

Kontynuując niech $X_1, \dots, X_n$ będą niezależnymi kopiami o jednakowym rozkładzie zmiennej losowej X z rozkładem $\mathbb{P}^X$. Będziemy badać założenie, że $\mathbb{P}^X$ należy do rodziny rozkładów Weibulla lub równoważnie weryfikować złożoną hipotezę zerową

$$H_0 : \mathbb{P}^X \in \mathcal{W}, \quad (1)$$

przeciwko globalnej alternatywie.

Lemat Steina

Pod odpowiednimi warunkami, zmienna losowa X opisana na liczbach rzeczywistych, pochodzi z rozkładu o różniczkowalnej gęstości f wtedy i tylko wtedy, gdy

$$\mathbb{E} \left[p'(X) + \frac{f'(X)}{f(X)} p(X) \right] = 0 \quad (2)$$

dla każdej funkcji p z odpowiednio dużej klasy funkcji testowych. Tutaj f' oznacza pochodną funkcji f . Jeśli f jest gęstością standardowego rozkładu normalnego, wtedy (2) jest znaną charakterystyką standardowego rozkładu normalnego w lemacie Steina.

Twierdzenie 1

Niech $\lambda, k > 0$ oraz X będzie dodatnią zmienną losową z różniczkowalną gęstością f i transformatą Laplace'a $\mathcal{L}_X$ spełniającą warunek $\mathbb{E}[X(d/dxf(x)|_X)/f(X)] < \infty$. Zmienna losowa X ma rozkład Weibulla $W(\lambda, k)$ wtedy i tylko wtedy, gdy

$$t\mathcal{L}_X(t) = \mathbb{E}\left[\frac{1}{X}\left(k\left(\frac{X}{\lambda}\right)^k - k + 1\right)(1 - e^{-tX})\right] \quad (3)$$

dla każdego $t > 0$.

Statystyka testowa

Oznaczając estymatory λ, k przez $\hat{\lambda}_n$ oraz $\hat{k}_n$ i zastąpimy transformatę Laplace'a przez wersję empiryczną $\hat{\mathcal{L}}_X(t) = n^{-1} \sum_{j=1}^n e^{-tX_j}$, to naszą statystyką testową będzie ważona odległość w normie L^2 przedstawiona poniższym wzorem

$$T_n = n \int_0^\infty \left| \frac{1}{n} \sum_{j=1}^n \frac{1}{X_j} \left(\hat{k}_n \left(\frac{X_j}{\hat{\lambda}_n} \right)^{\hat{k}_n} - \hat{k}_n + 1 \right) (1 - e^{-tX_j}) - \frac{t}{n} \sum_{j=1}^n e^{-tX_j} \right| w(t) dt, \quad (4)$$

gdzie $w : [0, \infty) \rightarrow (0, \infty)$ jest dodatnią funkcją wagi spełniającą warunek

$$\int_0^\infty (t^4 + 1) w(t) dt < \infty. \quad (5)$$

Hipotezę zerową odrzucamy dla dużych wartości statystyki T_n .

Statystyka testowa

Jeśli wybraną funkcją wagi w z (4) jest $w_a^{(1)}(t) = e^{-a|t|}$ lub $w_a^{(2)}(t) = e^{-at^2}$, $t \in [0, \infty)$ gdzie $a > 0$ jest jakimś parametrem dopasowania, to statystyka testowa T_n może być przedstawiona w formie prostszej do obliczeń.

Stąd oznaczając

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt, \quad x \geq 0,$$

jako funkcję błędu oraz dalej oznaczając

$$r(i, \hat{\lambda}_n, \hat{k}_n) = X_i^{-1} \left(\hat{k}_n \left(X_i / \hat{\lambda}_n \right)^{\hat{k}_n} - \hat{k}_n + 1 \right),$$

otrzymujemy

$$T_{n,a}^{(1)} = \frac{1}{n} \sum_{i,j=1}^n \left\{ r(i, \hat{\lambda}_n, \hat{k}_n) r(j, \hat{\lambda}_n, \hat{k}_n) \left(\frac{1}{a} - \frac{1}{a + X_i} \right. \right. \\ \left. \left. - \frac{1}{a + X_j} + \frac{1}{a + X_i + X_j} \right) \right. \\ \left. - 2r(j, \hat{\lambda}_n, \hat{k}_n) \left(\frac{1}{(a + X_i)^2} - \frac{1}{(a + X_i + X_j)^2} \right) + \frac{2}{(a + X_i + X_j)^3} \right\}$$

Statystyka testowa

oraz

$$\begin{aligned} T_{n,a}^{(2)} = & \frac{1}{n} \sum_{i,j=1}^n \left\{ r(i, \hat{\lambda}_n, \hat{k}_n) r(j, \hat{\lambda}_n, \hat{k}_n) \frac{1}{2} \sqrt{\frac{\pi}{a}} \left(1 - e^{X_i^2/(4a)} \left(1 - \operatorname{erf} \left(\frac{X_i}{2\sqrt{a}} \right) \right) \right. \right. \\ & - e^{X_j^2/(4a)} \left(1 - \operatorname{erf} \left(\frac{X_j}{2\sqrt{a}} \right) \right) + \exp \left(\frac{(X_i + X_j)^2}{4a} \right) \left(1 - \operatorname{erf} \left(\frac{X_i + X_j}{2\sqrt{a}} \right) \right) \Bigg) \\ & - \frac{r(j, \hat{\lambda}_n, \hat{k}_n) \sqrt{\pi}}{2(\sqrt{a})^3} \left(\exp \left(\frac{(X_i + X_j)^2}{4a} \right) (X_i + X_j) \left(1 - \operatorname{erf} \left(\frac{X_i + X_j}{2\sqrt{a}} \right) \right) \right. \\ & \left. \left. - e^{X_i^2/(4a)} X_i \left(1 - \operatorname{erf} \left(\frac{X_i}{2\sqrt{a}} \right) \right) \right) + \frac{1}{8(\sqrt{a})^5} \left(-2\sqrt{a}(X_i + X_j) \right. \right. \\ & \left. \left. + \exp \left(\frac{(X_i + X_j)^2}{4a} \right) \sqrt{\pi} (2a + (X_i + X_j)^2) \left(1 - \operatorname{erf} \left(\frac{X_i + X_j}{2\sqrt{a}} \right) \right) \right) \right\}. \end{aligned}$$

Teoria asymptotyczna

Rozważmy macierz trójkątną zmiennych losowych $X_{n,1}, \dots, X_{n,n}$, $n \in \mathbb{N}$ w następnych rzędach, zdefiniowaną na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$, gdzie

$$X_{n,1} \sim W(\lambda_n, k_n), \quad k_n, \lambda_n > 0,$$

oraz

$$\lim_{n \rightarrow \infty} \lambda_n = \lambda_0 > 0, \quad \lim_{n \rightarrow \infty} k_n = k_0 > 0. \quad (6)$$

Dalej niech zmienna losowa X ma rozkład Weibulla $W(\lambda_0, k_0)$.

Estymatory Największej Wiarygodności

Zakładamy, że estymatory $\hat{\lambda}_n, \hat{k}_n$ pozwalają się przedstawić za pomocą poniższej reprezentacji liniowej

$$\sqrt{n}(\hat{\lambda}_n - \lambda_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \Psi_1(X_{n,j}, \lambda_n, k_n) + o_{\mathbb{P}}(1), \quad (7)$$

$$\sqrt{n}(\hat{k}_n - k_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \Psi_2(X_{n,j}, \lambda_n, k_n) + o_{\mathbb{P}}(1), \quad (8)$$

gdzie Ψ_1 oraz Ψ_2 są funkcjami mierzalnymi, spełniającymi warunki

$$\mathbb{E}[\Psi_1(X_{n,1}, \lambda_n, k_n)] = 0, \quad \mathbb{E}[\Psi_2(X_{n,1}, \lambda_n, k_n)] = 0, \quad (9)$$

$$\mathbb{E}[\Psi_1^2(X_{n,1}, \lambda_n, k_n)] < \infty, \quad \mathbb{E}[\Psi_2^2(X_{n,1}, \lambda_n, k_n)] < \infty, \quad (10)$$

oraz

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[\Psi_1^2(X_{n,1}, \lambda_n, k_n)] &= \mathbb{E}[\Psi_1^2(X, \lambda_0, k_0)], \\ \lim_{n \rightarrow \infty} \mathbb{E}[\Psi_2^2(X_{n,1}, \lambda_n, k_n)] &= \mathbb{E}[\Psi_2^2(X, \lambda_0, k_0)]. \end{aligned} \quad (11)$$

Estymatory NW

$$\hat{\lambda}_n = \left(\frac{1}{n} \sum_{i=1}^n X_{n,i}^{\hat{k}_n} \right)^{1/\hat{k}_n},$$

$$\frac{n}{\hat{k}_n} + \sum_{i=1}^n \log X_{n,i} = \frac{n}{\sum_{i=1}^n X_{n,i}^{\hat{k}_n}} \sum_{i=1}^n X_{n,i}^{\hat{k}_n} \log X_{n,i}.$$