Wykład 6: Metody wyznaczania testów statystycznych

Jak już wspomnieliśmy, dobry test to taki, dla którego zarówno prawdopodobieństwo błędu I-go jak i II-go rodzaju jest małe. Jedną z metod zdefiniowania dobrego testu jest ustalenie poziomu istotności $\alpha \in (0,1)$, czyli najmniejszego górnego ograniczenia na prawdopodobieństwo błędu I-go rodzaju, a następnie żądanie by prawdopodobieństwo błędu II-go rodzaju było jak najmniejsze.

Sytuacja jest prosta, gdy zarówno hipoteza zerowa H_0 jak i hipoteza alternatywna H_1 to hipotezy proste, tzn., gdy $\Theta = \{\theta_0, \theta_1\}$ i testujemy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$.

Przykład 6. 1. Niech $\Theta = \{\theta_0, \theta_1\}$. Rozważmy dwa testy służące do weryfikacji $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$: test φ_1 i test φ_2. Załóżmy, że oba te testy mają poziom istotności α, tzn.

- $P(\text{błędu I-go rodzaju dla } \varphi_1) = P(\text{test } \varphi_1 \text{ odrzuci } H_0|H_0) = P(\text{test } \varphi_1 \text{ odrzuci } H_0|\theta_0) = \alpha$;
- $P(\text{błędu I-go rodzaju dla } \varphi_2) = P(\text{test } \varphi_2 \text{ odrzuci } H_0|\theta_0) = \alpha$.

Ponadto załóżmy, że

- $P(\text{błędu II-go rodzaju dla } \varphi_1) = P(\text{test } \varphi_1 \text{ przyjmie } H_0|H_1) = P(\text{test } \varphi_1 \text{ przyjmie } H_0|\theta_1) = 0,7$ (równoważnie $P(\text{test } \varphi_1 \text{ odrzuci } H_0|\theta_1) = 0,3$);
- $P(\text{błędu II-go rodzaju dla } \varphi_2) = P(\text{test } \varphi_2 \text{ przyjmie } H_0|\theta_1) = 0,2$ (równoważnie $P(\text{test } \varphi_2 \text{ odrzuci } H_0|\theta_1) = 0,8$).

Który z testów φ_1 i φ_2 jest lepszy? Oczywiście test φ_2.

Definicja. Niech $\Theta = \{\theta_0, \theta_1\}$. *Najmocniejszym testem na poziomie istotności α dla $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$* nazywamy test φ^*, który spełnia poniższe dwa warunki:

1. ma poziom istotności α (tzn. $P(\text{test } \varphi^* \text{ odrzuci } H_0|\theta_0) = \alpha$);
2. dla każdego testu φ służącego do weryfikacji H_0 przeciwko H_1 i posiadającego poziom istotności $\leqslant \alpha$ mamy

$$P(\text{test } \varphi \text{ odrzuci } H_0|\theta_1) \leqslant P(\text{test } \varphi^* \text{ odrzuci } H_0|\theta_1).$$

Innymi słowy, gdy $\Theta = \{\theta_0, \theta_1\}$, to najmocniejszy test na poziomie istotności α dla $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ ma tą pożądaną własność, że spośród wszystkich testów na poziomie istotności α lub mniejszym z największym prawdopodobieństwem odrzuca hipotezę H_0 wtedy, kiedy jest ona fałszywa.

Lemat Neymana-Pearsona podaje sposób konstrukcji takich testów.

Twierdzenie 6.1 (Lemat Neymana-Pearsona)**.** Niech $X_1, X_2, \ldots, X_n$ będzie próbą losową z rozkładu ciągłego o gęstości $f_\theta(x)$ lub dyskretnego o funkcji masy prawdopodobieństwa $p_\theta(x)$, gdzie $\theta \in \{\theta_0, \theta_1\}$. Ustalmy $\alpha \in (0,1)$. Wtedy test, który odrzuca H_0, gdy

$$\frac{L(\theta_0; X_1, X_2, \ldots, X_n)}{L(\theta_1; X_1, X_2, \ldots, X_n)} \leqslant k^*,$$

gdzie L oznacza funkcję wiarygodności zaś k^* spełnia

$$P\left(\frac{L(\theta_0; X_1, X_2, \ldots, X_n)}{L(\theta_1; X_1, X_2, \ldots, X_n)} \leqslant k^*|\theta_0\right) = \alpha \tag{1}$$

jest najmocniejszym testem na poziomie istotności α dla $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$.

Uwaga 6.1. W przypadku, gdy $X_1, X_2, \ldots, X_n$ mają rozkłady dyskretne nie dla każdej $\alpha \in (0,1)$ znajdziemy k^* spełniające (1). Można ten problem rozwiązać rozważając testy zrandomizowane. Nie będziemy się jednak takimi testami zajmować, bo w praktyce w takiej sytuacji α zastępuje się przez największe możliwe $\alpha^* < \alpha$, dla którego istnieje k^* spełniające (1). W ten sposób otrzymujemy najmocniejszy test o poziomie istotności $\alpha^* < \alpha$.

Przykład 6.2. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu wykładniczego $Exp(\lambda)$, czyli rozkładu o gęstości $f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$. Wyznaczymy najmocniejszy test na poziomie istotności α do weryfikacji hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda = \lambda_1$, gdzie

(a) $0 < \lambda_0 < \lambda_1$;

(b) $0 < \lambda_1 < \lambda_0$.

Rozwiązanie przykładu 6.2:
(a) Funkcja wiarygodności przyjmuje postać dla $x_1, \ldots, x_n > 0$

$$L(\lambda; x_1, \ldots, x_n) = f_\lambda(x_1) f_\lambda(x_2) \cdots f_\lambda(x_n) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

Stąd na mocy lematu Neymana-Pearsona, szukany test to taki, który odrzuca H_0*, gdy*

$$\frac{L(\theta_0; X_1, X_2, \ldots, X_n)}{L(\theta_1; X_1, X_2, \ldots, X_n)} = \frac{\lambda_0^n e^{-\lambda_0 \sum_{i=1}^n X_i}}{\lambda_1^n e^{-\lambda_1 \sum_{i=1}^n X_i}} = \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{-(\lambda_0 - \lambda_1) \sum_{i=1}^n X_i} \leqslant k^*,$$

co równoważnie możemy zapisać jako

$$e^{-(\lambda_0 - \lambda_1) \sum_{i=1}^n X_i} \leqslant \left(\frac{\lambda_1}{\lambda_0}\right)^n k^*$$

$$-(\lambda_0 - \lambda_1) \sum_{i=1}^n X_i \leqslant \ln\left[\left(\frac{\lambda_1}{\lambda_0}\right)^n k^*\right] \tag{2}$$

$$\sum_{i=1}^n X_i \leqslant \frac{1}{\lambda_1 - \lambda_0} \ln\left[\left(\frac{\lambda_1}{\lambda_0}\right)^n k^*\right]. \tag{3}$$

Oznaczając $\tilde{k} = \frac{1}{\lambda_1 - \lambda_0} \ln\left[\left(\frac{\lambda_1}{\lambda_0}\right)^n k^*\right]$*, sprowadzamy zadanie do wyznaczenia* $\tilde{k}$ *takiego, że*

$$P\left(\sum_{i=1}^n X_i \leqslant \tilde{k} | \lambda_0\right) = \alpha,$$

czyli $\tilde{k}$ *takiego, że*

$$P\left(\sum_{i=1}^n Y_i \leqslant \tilde{k}\right) = \alpha, \tag{4}$$

gdzie $Y_i, i = 1, 2, \ldots, n$*, są niezależne o tym samym rozkładzie* $Exp(\lambda_0)$*. Z twierdzenia 6.2.3 wiemy, że rozkład* $Exp(\lambda_0)$ *to to samo co rozkład* $Gamma(1, \lambda_0)$*. Stąd* Y_i*,* $i = 1, 2, \ldots, n$*, to niezależne zmienne losowe o tym samym rozkładzie* $Gamma(1, \lambda_0)$ *i używając twierdzenia 6.2.1(b) otrzymujemy*

$$Z = \sum_{i=1}^n Y_i \text{ ma rozkład } Gamma(n, \lambda_0).$$

Zatem warunek (4) sprowadza się do warunku

$$P(Z \leqslant \tilde{k}) = \alpha,$$

co oznacza, że $\tilde{k}$ *to kwantyl rzędu* α *rozkładu* $Gamma(n, \lambda_0)$*.*

Reasumując, szukany test to taki, który odrzuca H_0*, gdy* $\sum_{i=1}^n X_i \leqslant \tilde{k}$*, gdzie* $\tilde{k}$ *to kwantyl rzędu* α *rozkładu* $Gamma(n, \lambda_0)$*. Innymi słowy statystyka testowa szukanego testu to* $G = \sum_{i=1}^n X_i$ *a zbiór krytyczny to* $W = (0, \tilde{k}]$*, gdzie* $\tilde{k}$ *to kwantyl rzędu* α *rozkładu* $Gamma(n, \lambda_0)$*.*

(b) Początek rozwiązania jest taki sam jak w punkcie (a). Pierwsza różnica pojawi się przy przejściu z (2) do (3): teraz obie strony nierówności (2) będziemy dzielić przez liczbę ujemną, więc (3) przyjmie postać

$$\sum_{i=1}^{n} X_i \geqslant \frac{1}{\lambda_1 - \lambda_0} \ln\left[\left(\frac{\lambda_1}{\lambda_0}\right)^n k^*\right] \stackrel{ozn.}{=} \bar{k}$$

i będziemy szukać $\bar{k}$ takiego, że

$$P\left(\sum_{i=1}^{n} X_i \geqslant \bar{k}|\lambda_0\right) = \alpha. \tag{5}$$

Skoro $P\left(\sum_{i=1}^{n} X_i \geqslant \bar{k}|\lambda_0\right) = 1 - P\left(\sum_{i=1}^{n} X_i < \bar{k}|\lambda_0\right) = 1 - P\left(\sum_{i=1}^{n} X_i \leqslant \bar{k}|\lambda_0\right)$, gdzie ostatnia równość wynika stąd, że $\sum_{i=1}^{n} X_i$ ma rozkład ciągły, to (5) można przepisać jako

$$P\left(\sum_{i=1}^{n} X_i \leqslant \bar{k}|\lambda_0\right) = 1 - \alpha$$

i powtarzając rozumowanie przeprowadzone w punkcie (a) otrzymujemy, że $\bar{k}$ to kwantyl rzędu $1-\alpha$ rozkładu $Gamma(n, \lambda_0)$.

Reasumując, szukany test to taki, który odrzuca H_0, gdy $\sum_{i=1}^{n} X_i \geqslant \bar{k}$, gdzie $\bar{k}$ to kwantyl rzędu $1-\alpha$ rozkładu $Gamma(n, \lambda_0)$. Innymi słowy statystyka testowa szukanego testu to $G = \sum_{i=1}^{n} X_i$ a zbiór krytyczny to $W = [\bar{k}, \infty)$, gdzie $\bar{k}$ to kwantyl rzędu $1-\alpha$ rozkładu $Gamma(n, \lambda_0)$.

Teraz przejdziemy do przypadku, gdy hipotezy H_0 i H_1 nie są koniecznie hipotezami prostymi, tzn., gdy $H_0 : \theta \in \Theta_0$ i $H_1 : \theta \in \Theta \setminus \Theta_0$. Wtedy sytuacja jest znacznie trudniejsza.

Przykład 6.3. Niech $\Theta = \{\theta_0, \theta_a, \theta_b, \theta_c\}$. Rozważmy dwa testy służące do weryfikacji $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \in \{\theta_a, \theta_b, \theta_c\}$: test φ_1 i test φ_2. Załóżmy, że oba te testy mają poziom istotności α, tzn.

$$P(\text{test } \varphi_1 \text{ odrzuci } H_0|\theta_0) = P(\text{test } \varphi_2 \text{ odrzuci } H_0|\theta_0) = \alpha.$$

Ponadto załóżmy, że

$$\begin{aligned} P(\text{test } \varphi_1 \text{ odrzuci } H_0|\theta_a) &= 0,7, \qquad & P(\text{test } \varphi_2 \text{ odrzuci } H_0|\theta_a) &= 0,6, \\ P(\text{test } \varphi_1 \text{ odrzuci } H_0|\theta_b) &= 0,8, \qquad & P(\text{test } \varphi_2 \text{ odrzuci } H_0|\theta_b) &= 0,6, \\ P(\text{test } \varphi_1 \text{ odrzuci } H_0|\theta_c) &= 0,6, \qquad & P(\text{test } \varphi_2 \text{ odrzuci } H_0|\theta_c) &= 0,9. \end{aligned}$$

Który z tych dwóch testów jest lepszy? Trudno jednoznacznie to stwierdzić, bo test φ_1 zachowuje się lepiej, gdy $\theta = \theta_a$ lub $\theta = \theta_b$ (wtedy test φ_1 z większym prawdopodobieństwem niż test φ_2 odrzuca H_0 w sytuacji, gdy jest ona fałszywa), zaś test φ_2 zachowuje się lepiej, gdy $\theta = \theta_c$ (wtedy test φ_2 z większym prawdopodobieństwem niż test φ_1 odrzuca H_0 w sytuacji, gdy jest ona fałszywa).

Załóżmy, że jest trzeci test φ_3 służący do weryfikacji $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \in \{\theta_a, \theta_b, \theta_c\}$, mający poziom istotności α i taki, że

$$\begin{aligned} P(\text{test } \varphi_3 \text{ odrzuci } H_0|\theta_a) &= 0,85, \\ P(\text{test } \varphi_3 \text{ odrzuci } H_0|\theta_b) &= 0,8, \\ P(\text{test } \varphi_3 \text{ odrzuci } H_0|\theta_c) &= 0,9. \end{aligned}$$

Test φ_3 jest najlepszym testem spośród φ_1, φ_2 i φ_3, bo z największym prawdopodobieństwem odrzuca H_0 w sytuacji, gdy jest ona fałszywa, tzn. dla każdego $i \in \{1, 2, 3\}$ mamy

$$P(\text{test } \varphi_i \text{ odrzuci } H_0|\theta_1) \leqslant P(\text{test } \varphi_3 \text{ odrzuci } H_0|\theta_1) \text{ dla każdego } \theta_1 \in \{\theta_a, \theta_b, \theta_c\}.$$

Test o poziomie istotności α, który ze wszystkich możliwych testów służących do weryfikacji $H_0 : \theta \in \Theta_0$ przeciwko $H_1 : \theta \in \Theta \setminus \Theta_0$ i posiadających poziom istotności $\leqslant \alpha$, z największym prawdopodobieństwem odrzuca H_0 w sytuacji, gdy jest ona fałszywa, nazywamy testem jednostajnie najmocniejszym na poziomie istotności α.

Definicja. Niech $H_0 : \theta \in \Theta_0$ i $H_1 : \theta \in \Theta \setminus \Theta_0$. *Jednostajnie najmocniejszym testem na poziomie istotności α dla H_0 przeciwko H_1* nazywamy test φ^*, który spełnia poniższe dwa warunki:

1. ma poziom istotności α;

2. dla każdego testu φ służącego do weryfikacji H_0 przeciwko H_1 i posiadającego poziom istotności $\leqslant \alpha$ mamy

$$P(\text{test } \varphi \text{ odrzuci } H_0|\theta_1) \leqslant P(\text{test } \varphi^* \text{ odrzuci } H_0|\theta_1) \text{ dla każdego } \theta_1 \in \Theta \setminus \Theta_0.$$

Niestety nie dla każdego problemu testowania statytycznego istnieje jednostajnie najmocniejszy test. Dla niektórych problemów nie tylko taki test istnieje, ale można go znaleźć używając lematu Neymana-Pearsona, nawet wtedy, gdy H_1 (ewentualnie też H_0) nie jest hipotezą prostą i szukamy testu jednostajnie najmocniejszego.

Przykład 6.4. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu wykładniczego $Exp(\lambda)$, gdzie $\lambda \geqslant \lambda_0 > 0$. Wyznaczymy jednostajnie najmocniejszy test na poziomie istotności α do weryfikacji hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda > \lambda_0$.

Rozwiązanie przykładu 6.4:
Do weryfikacji hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda > \lambda_0$ użyjemy tego samego testu co w przykładzie 6.2(a), bo test ten jest taki sam dla wszystkich $\lambda_1 > \lambda_0$. W konsekwencji dla każdego $\lambda_1 > \lambda_0$ mamy ten sam najmocniejszy test, co oznacza, że jest on jednostajnie najmocniejszy.

Znanych jest wiele innych metod wyznaczania dobrych testów. Jedna z nich polega na zaproponowaniu funkcji próby $s(X_1, X_2, \ldots, X_n)$, która nie zawiera żadnych nieznanych parametrów (funkcja taka zwana jest *statystyką*) i takiej, że $s(X_1, X_2, \ldots, X_n)$ zachowuje się wyraźnie inaczej, gdy prawdziwa jest H_0 i gdy prawdziwa jest H_1. Za $s(X_1, X_2, \ldots, X_n)$ można przyjąć dobry estymator parametru θ, którego dotyczą hipotezy H_0 i H_1 (np. estymator największej wiarygodności). Statystyka ta, po ewentualnych przekształceniach (które znowu nie mogą zależeć od nieznanych parametrów) prowadzących do uproszczenia jej rozkładu, staje się statystyką testową szukanego testu.

Twierdzenie 6.2 (Fakty pomocnicze)**.**

1. Fakt dotyczący sumowania niezależnych zmiennych losowych

Jeśli $X_1, X_2, \ldots, X_n$ to niezależne zmienne losowe i

(a) X_i ma rozkład normalny $\mathcal{N}(\mu_i, \sigma_i^2)$ dla $i = 1, 2, \ldots, n$, to

$$X_1 + X_2 + \cdots + X_n \text{ ma rozkład normalny } \mathcal{N}(\mu_1 + \mu_2 + \ldots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2);$$

(b) X_i, $i = 1, 2, \ldots, n$, ma rozkład gamma $Gamma(a_i, \beta)$, tzn. X_i ma gęstość

$$f_{a_i,\beta}(x) = \begin{cases} \frac{\beta^{a_i}}{\Gamma(a_i)} x^{a_i - 1} \exp(-\beta x) & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}, \quad a_i > 0, \beta > 0,$$

to $X_1 + X_2 + \cdots + X_n$ ma rozkład gamma $Gamma(a_1 + a_2 + \cdots + a_n, \beta)$;

(c) X_i, $i = 1, 2, \ldots, n$, ma rozkład dwupunktowy o prawdopodobieństwie sukcesu θ z funkcją masy prawdopodobieństwa $P(X_i = x) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$, to

$$X_1 + X_2 + \cdots + X_n \text{ ma rozkład dwumianowy } binom(n, \theta);$$

ogólniej, jeśli X_i, $i = 1, 2, \ldots, n$, ma rozkład dwumianowy $binom(m_i, \theta)$, to

$$X_1 + X_2 + \cdots + X_n \text{ ma rozkład dwumianowy } binom(m_1 + m_2 + \cdots + m_n, \theta).$$

2. Fakt dotyczący rozkładu beta

Jeśli X ma rozkład beta o parametreach $a > 0$ i 1, tzn. X ma gęstość

$$f_a(x) = \begin{cases} a x^{a-1} & \text{dla } x \in (0,1) \\ 0 & \text{dla } x \notin (0,1) \end{cases}, \quad \text{gdzie } a > 0,$$

to $Y = -\ln X$ ma rozkład wykładniczy $Exp(a)$, tzn. Y ma gęstość $f_a(x) = \begin{cases} a e^{-ax}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$.

3. Związek pomiędzy rozkładem wykładniczym a rozkładem gamma

Rozkład wykładniczy $Exp(\lambda)$, $\lambda > 0$, to rozkładu o gęstości $f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$, zaś rozkład gamma $Gamma(a, \beta)$, $a > 0, \beta > 0$, to rozkładu o gęstości

$$f_{a,\beta}(x) = \begin{cases} \frac{\beta^a}{\Gamma(a)} x^{a-1} \exp(-\beta x) & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}.$$

Stąd widać, że rozkład wykładniczy $Exp(\lambda)$ to to samo co rozkład gamma $Gamma(1, \lambda)$.

Przypomnienie. Kwantyl rzędu $\alpha \in (0,1)$ ciągłej zmiennej losowej X to liczba q_α spełniająca

$$P(X \leqslant q_\alpha) = \alpha.$$

Funkcje w pakiecie R służące do liczenia kwantyli:

rozkład	kantyl rzędu $\alpha \in (0,1)$
$Exp(\lambda)$	qexp(p= α, rate= λ)
$Gamma(a, \beta)$	qgamma(p= α, shape=a, rate= β)
$\mathcal{N}(\mu, \sigma^2)$	qnorm(p= α, mean= μ, sd= σ)