

1 Wykład 6

Przykład 1.1 Podczas rozprawy sądowej, wykorzystując zebrane dowody i zeznania świadków, sędzia musi odpowiedzieć na pytanie: Czy prawdą jest, że podsądny jest winien? Zadanie to można przedstawić w języku testowania hipotez:

$$\begin{aligned}H_0 &: \text{Podsądny jest niewinny} \\ H_1 &: \text{Podsądny jest winien}\end{aligned}$$

Decyzja, podejmowana przez sąd, polega na wyborze hipotezy, która jego zdaniem jest prawdziwa. Sąd może jednak popełnić jeden z dwóch możliwych błędów:

H_0 jest prawdziwa, a sąd twierdzi, że H_1 jest prawdziwa
(sąd skazał niewinnego)

lub też

H_1 jest prawdziwa, a sąd twierdzi, że H_0 jest prawdziwa
(sąd uwolnił winnego)

Skutki tych błędów są różne. W prawodawstwie wielu krajów przyjmuje się, że pierwszy błąd jest bardziej dotkliwy - sądy skazują gdy dysponują mocnymi argumentami przeciw podsądnemu. Ten sposób postępowania nazywa się zasadą domniemania niewinności. ✗

Jeżeli θ jest parametrem modelu statystycznego, układ dwóch hipotez może być sformułowany jako:

$$\begin{aligned}H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1\end{aligned},$$

oraz $\Theta_0 \cap \Theta_1 = \emptyset$. W postępowaniu, prowadzącym do rozstrzygnięcia, która z tych dwóch hipotez jest prawdziwa, możemy popełnić dwa błędy:

- błąd pierwszego rodzaju:

twierdzimy, że H_1 jest prawdziwa gdy H_0 jest prawdziwa

- błąd drugiego rodzaju:

twierdzimy, że H_0 jest prawdziwa gdy H_1 jest prawdziwa

Optymalne postępowanie powinno prowadzić do zmniejszenia ryzyka popełnienia błędu. Niestety, nie można uniknąć popełnienia obu błędów jednocześnie. Zazwyczaj, przy próbie zminimalizowania błędu pierwszego rodzaju zwiększa się szansa popełnienia błędu drugiego rodzaju. Podobnie jest z błędem drugiego rodzaju. Aby uniknąć kłopotów postępuje się w statystyce podobnie jak w orzecznictwie sądowym: obowiązuje zasada domniemania prawdziwości hipotezy zerowej, często nazywaną zasadą konserwatyzmu.

Zasada domniemania prawdziwości hipotezy zerowej oznacza, że chcemy mieć gwarancję, że błąd pierwszego rodzaju¹ nie wystąpi zbyt często. Musimy

¹a więc błąd polegający na tym, że odrzucamy hipotezę zerową, gdy jest ona prawdziwa

więc określić dopuszczalną granicę błędu I rodzaju. Spośród wszystkich znanych nam sposobów podjęcia decyzji odrzucenia hipotezy zerowej, spełniających ten warunek, wybieramy ten, który daje najmniejszy błąd II rodzaju. Jeżeli jako hipotezę zerową wybierzemy stan wiedzy *przed* przeprowadzeniem eksperymentu to zasada domniemania prawdziwości hipotezy zerowej oznacza, że przyjmujemy nowe fakty² jedynie wtedy, gdy jej odrzucenie będzie dobrze udokumentowane. Podobnie, jak w postępowaniu sądowym, gdzie odrzucenie hipotezy o niewinności musi być dobrze udokumentowane (wątpliwości są uwzględniane z korzyścią dla oskarżonego).

Definicja 1.1 *Hipotezę nazywamy prostą jeżeli zbiór parametrów ją kreślających jest jednoelementowy. Hipoteza jest złożona jeśli nie jest prosta.*

Przykład 1.2 *Rzucamy monetą i chcielibyśmy dowiedzieć się, czy moneta jest symetryczna czy niesymetryczna. Układ dwóch hipotez w tym przypadku byłby następujący:*

$$\begin{aligned} H_0 &: \theta = \frac{1}{2}, \\ H_1 &: \theta \neq \frac{1}{2} \end{aligned}$$

Hipoteza zerowa jest prosta, hipoteza konkurencyjna jest złożona.✂

Przykład 1.3 *Obserwujemy w doświadczeniu wartości pewnej zmiennej o dystrybucji F , pochodzącej z rodziny dystrybuant $\mathcal{F}$. Chcemy rozstrzygnąć, czy ma ona rozkład normalny. Układ dwóch hipotez w tym przypadku jest następujący:*

$$\begin{aligned} H_0 &: F \in \mathcal{N} = \{\mathcal{N}(\mu, \sigma) : \mu \in \mathcal{R}, \sigma \in (0, \infty]\} \\ H_1 &: F \in \mathcal{F} - \mathcal{N} \end{aligned}$$

W tym przykładzie zarówno hipoteza zerowa jak i konkurencyjna są złożone.✂

Definicja 1.2 *Hipoteza, w której zbiór parametrów jest podzbiorem $\mathcal{R}^k$ nazywa się hipotezą parametryczną, w przeciwnym przypadku mówimy o hipotezie nieparametrycznej. Dodatkowo, jeśli w hipotezie zerowej jest mowa o typie rozkładu, to mówimy o hipotezie zgodności.*

Hipotezy w przykładzie 1.2 są parametryczne, natomiast w przykładzie 1.3 są nieparametrycznymi hipotezami zgodności.

Decyzja o wyborze hipotezy powinna być oparta na obserwacjach i musi wskazywać, jakie warunki muszą one spełniać aby można było odrzucać hipotezę zerową. A więc reguła podjęcia decyzji oznacza wskazanie zbioru $C \subset \mathcal{R}^n$, zwanego *zbiorem krytycznym*. Ustalenie zbioru krytycznego nazywa się też *wyborem testu statystycznego*. Gdy obserwacja $X \in C$ to będziemy odrzucać hipotezę zerową³. W przeciwnym przypadku hipotezy zerowej nie będziemy odrzucać. W postępowaniu sądowym zbiór krytyczny oznaczałby określenie, jakie dowody wystarczą do udowodnienia winy oskarżonemu.

Błąd I rodzaju oznacza zdarzenie $X \in C$ pod warunkiem, że hipoteza zerowa jest prawdziwa. Miarą szansy zajścia tego zdarzenia jest jego prawdopodobieństwo.

²a więc odrzucimy hipotezę zerową

³co oznacza przyjęcie hipotezy konkurencyjnej

Definicja 1.3 *Rozmiarem testu nazywamy prawdopodobieństwo*

$$\sup \{P(X \in C | \theta) : \theta \in \Theta_0\}$$

Rozmiar testu jest więc największym prawdopodobieństwem popełnienia błędu I rodzaju. Podobnie moglibyśmy określić prawdopodobieństwo błędu II rodzaju:

$$\sup \{P(X \notin C | \theta) : \theta \in \Theta_1\}$$

Zasada domniemania prawdziwości hipotezy zerowej oznacza, że rozmiar testu nie przekroczy pewnej wartości α zwanej *poziomą istotnością*:

$$\sup \{P(X \in C | \theta) : \theta \in \Theta_0\} \leq \alpha$$

Mówimy wtedy, że *test jest na poziomie istotności α* .

W praktyce przyjmuje się dwa standardowe poziomy istotności: $\alpha = 0.05$ i $\alpha = 0.01$.

Przykład 1.4 *Rzucamy $n = 5$ razy monetą, w której prawdopodobieństwo wyrzucenia orła wynosi θ . Chcemy rozstrzygnąć hipotezę*

$$H_0 : \theta = \frac{1}{2}$$

przeciwko hipotezie

$$H_1 : \theta = \frac{1}{4}$$

Niech X oznacza liczbę orłów uzyskanych w tych 5 rzutach. Jako zbiór krytyczny wybierzemy przedział $C = [0, 2]$. Rozmiar tego testu wynosi:

$$\begin{aligned} \sup \{P(X \in C | \theta) : \theta \in \Theta_0\} &= P(X \leq 2 | \theta = \frac{1}{2}) = \\ &= \sum_{i=0}^2 \binom{5}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{5-i} = \frac{1}{2} \end{aligned}$$

natomiast błąd drugiego rodzaju wynosi

$$\begin{aligned} \sup \{P(X \notin C | \theta) : \theta \in \Theta_1\} &= P(X > 2 | \theta = \frac{1}{4}) = \\ &= 1 - \sum_{i=0}^2 \binom{5}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{5-i} = \frac{53}{512} \approx .104 \end{aligned}$$

✗

Chcielibyśmy umieć konstruować możliwie najlepsze testy na zadanym poziomie ufności. Słowo "najlepszy" oznacza, że wśród *wszystkich* testów na zadanym poziomie ufności ten test ma najmniejsze prawdopodobieństwo błędu II rodzaju. Zagadnienie to w szczególnym przypadku, rozwiązali w 1928, Jerzy Neyman i Egon Pearson⁴.

Konstrukcja testów według Neymana i Pearsona opiera się na pojęciu ilorazu wiarygodności hipotez.

⁴Jerzy Sława-Neyman (1984-1981), matematyk polski; Egon Sharpe Pearson (1895-1980) matematyk angielski

Definicja 1.4 Iloraz wiarygodności hipotez H_0, H_1 jest ilorazem

$$L_x(H_0, H_1) = \frac{L_x(H_1)}{L_x(H_0)}$$

gdzie

$$L_x(H) = \sup \{ \theta \in \Theta_H : f_X(x|\theta) \}$$

oraz gdzie Θ_H jest zbiorem parametrów opisujących hipotezę H a $f_X(x|\theta)$ jest gęstością (rozkładem w przypadku zmiennej dyskretnej) cechy X w punkcie x . Funkcję $L_x(H)$ nazywamy wiarygodnością hipotezy H dla obserwacji x .

Iloraz wiarygodności podaje, ile razy częściej wystąpiłby wynik x gdyby prawdziwa była hipoteza H_1 niż gdyby prawdziwa była hipoteza H_0 . Skłonni byłibyśmy odrzucać hipotezę H_0 na rzecz hipotezy H_1 na podstawie wyniku x gdyby iloraz wiarygodności $L_x(H_0, H_1)$ miał dużą wartość.

Przykład 1.5 (cd przykładu (1.4)) Iloraz wiarygodności hipotez gdy uzyskamy x orłów w 5 rzutach wynosi:

$$L_x(H_0, H_1) = \frac{P\{X=x|\theta=\frac{1}{4}\}}{P\{X=x|\theta=\frac{1}{2}\}} = \frac{\binom{5}{x} (\frac{1}{4})^x (\frac{3}{4})^{5-x}}{\binom{5}{x} (\frac{1}{2})^x (\frac{1}{2})^{5-x}} = \frac{243}{32} 3^{-x}$$

Widać, że im mniejsza liczba wyrzuconych orłów, tym bardziej byłibyśmy skłonni odrzucać hipotezę H_0 na rzecz hipotezy H_1 , co jest zgodne z naszą intuicją. Gdy, na przykład, w wyniku 5 rzutów wypadną 2 orły, to iloraz wiarygodności w tym przypadku wynosi

$$\frac{243}{32} 3^{-2} \approx .844$$

co sugeruje, że hipoteza H_1 jest słabsza od hipotezy H_0 . ✗

Będziemy rozważać zbiory krytyczne postaci $C = \{x : L_x(H_0, H_1) > k\}$. Poniższe twierdzenie, znane pod nazwą lematu Neymana-Pearsona podaje konstrukcję testów optymalnych, gdy testujemy prostą hipotezę przeciwko prostej hipotezie alternatywnej.

Lemat 1.1 (Neymana - Pearsona) Niech

$$\begin{aligned} H_0 &: f = f_0, \\ H_1 &: f = f_1 \end{aligned}$$

i f_0 oraz f_1 będą gęstościami prawdopodobieństwa, ciągłymi i dodatnimi na tym samym zbiorze.

Wtedy wśród wszystkich testów o rozmiarze nie przekraczającym α test o najmniejszym prawdopodobieństwie błędów II rodzaju dany jest wzorem:

$$C = \{x : L_x(H_0, H_1) > k\},$$

gdzie k spełnia warunek

$$\alpha = P(X \in C | H_0) = \int_C f_0(x) dx$$

Dowód. Niech D będzie dowolnym testem na poziomie $\leq \alpha$. Oznacza to, że

$$P(X \in D | H_0) \leq \alpha.$$

Niech

$$\phi_A(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{gdy } x \in A \\ 0 & \text{gdy } x \notin A \end{cases}$$

Wtedy dla każdego x

$$0 \leq (\phi_C(x) - \phi_D(x))(f_1(x) - kf_0(x))$$

oraz

$$\begin{aligned} 0 &\leq \int_C (\phi_C(x) - \phi_D(x))(f_1(x) - kf_0(x)) dx = \\ &= P(X \in C | H_1) - P(X \in D | H_1) - k(P(X \in C | H_0) - P(X \in D | H_0)) = \\ &= P(X \in C | H_1) - P(X \in D | H_1) - k(\alpha - P(X \in D | H_0)) \leq \\ &\leq P(X \in C | H_1) - P(X \in D | H_1) \end{aligned}$$

Tę ostatnią nierówność można zapisać w innej postaci:

$$P(X \notin C | H_1) \leq P(X \notin D | H_1)$$

co oznacza, że test C ma mniejszy błąd II rodzaju niż test D . ■

Przykład 1.6 Niech $X^T = (x_1, x_2, \dots, x_n)$ będzie próbą prostą z rozkładu normalnego $\mathcal{N}(\mu, \sigma)$ i odchylenie standardowe σ jest znane. Chcemy testować hipotezy

$$\begin{aligned} H_0 &: \mu = \mu_0, \\ H_1 &: \mu = \mu_1 \end{aligned}$$

gdzie $\mu_1 > \mu_0$. Z lematu Neymana - Pearsona wynika, że obszar krytyczny musi być skonstruowany za pomocą ilorazu wiarygodności

$$\begin{aligned} L_x(H_0, H_1) &= \frac{\prod_{i=1}^n 2\pi\sigma^2 \exp\left\{-\frac{(x_i - \mu_1)^2}{2\sigma^2}\right\}}{\left(\prod_{i=1}^n 2\pi\sigma^2\right) \exp\left\{-\frac{(x_i - \mu_0)^2}{2\sigma^2}\right\}} = \\ &= \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \mu_1)^2}{2\sigma^2}\right\} = \\ &= \exp\left\{-\frac{\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\mu_0^2 - \sum_{i=1}^n x_i^2 + 2\bar{x} \sum_{i=1}^n x_i - n\mu_1^2}{2\sigma^2}\right\} \end{aligned}$$

Iloraz wiarygodności jest rosnącą funkcją $\bar{x}$ tak więc rozwiązywanie nierówności $L_x(H_0, H_1) > k$ jest równoważne rozwiązaniu nierówności $\bar{x} > k^*$. Tak więc będziemy odrzucać hipotezę H_0 gdy obserwacje $X^T = (x_1, x_2, \dots, x_n)$ spełniają

$$\bar{x} > k^*$$

gdzie stała k^* jest rozwiązaniem równania

$$P(\bar{x} > k^* | H_0) = \alpha$$

To ostatnie równanie da się łatwo wyrazić przez dystrybuantę Φ standardowego rozkładu normalnego, gdyż zmienna losowa

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

ma standardowy rozkład normalny. Korzystając z tego faktu,

$$\begin{aligned} \alpha &= P(\bar{x} > k^* | H_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{k^* - \mu_0}{\sigma/\sqrt{n}} \mid H_0\right) = \\ &= 1 - \Phi\left(\frac{k^* - \mu_0}{\sigma/\sqrt{n}}\right) \end{aligned}$$

Jeżeli wprowadzimy oznaczenie $z_\alpha \stackrel{\text{def}}{=} \Phi^{-1}(1 - \alpha)$ to $k^* = z_\alpha \sigma / \sqrt{n} + \mu_0$. Na przykład, gdy $\alpha = 0,05$ to $z_\alpha = 1.65$, gdy $\alpha = 0,01$ to $z_\alpha = 2.33$