Trochę o statystyce (2016) *

Antoni Kościelski

1 Podstawowe pojęcia statystyczne

1.1 Próba i przestrzeń prób

Próba to coś, co zostało zaobserwowane, na przykład wynik serii doświadczeń. Często pojęcie próby formalizujemy jako skończony ciąg liczb rzeczystych $\vec{x} = (x_1, x_2, \ldots, x_n) \in R^n$.

Przestrzeń prób $\mathcal{X}$ to zbiór wszystkich spodziewanych obserwacji lub wszystkich możliwych wyników serii doświadczeń. Zwykle $\mathcal{X}$ jest podzbiorem przestrzeni R^n.

1.2 Przestrzenie i hipotezy statystyczne

Przestrzeń statystyczna to zbiór prób $\mathcal{X}$ i parametryczna rodzina $\{P_\theta : \theta \in \Theta\}$ rozkładów prawdopodobieństwa na $\mathcal{X}$. Parametry są zwykle liczbami rzeczywistymi, rozkład P_θ może jednak zależeć od jednego lub kilku parametrów. Tak więc $\Theta \subseteq R^m$. Zbiór Θ może się składać z wszystkich dopuszczalnych z matematycznego punktu widzenia układów liczb rzeczywistych lub z aktualnie rozważanych.

Hipoteza statystyczna to jakieś stwierdzenie o rozważanych rozkładach. Można ją sformalizować jako zbiór parametrów (wyznaczających rozkłady, dla których zachodzi to stwierdzenie). Zwykle rozważa się dwie hipotezy: hipotezę zerową $H_0 \subseteq \Theta$ oraz hipotezę alternatywną $H_1 = \Theta \setminus H_0$. Obie hipotezy w sumie definiują zbiór Θ.

1.3 Przykład

Możemy rozważać serie n-krotnych rzutów monetą. Dwa możliwe zdarzenia: wypadnięcie orła i reszki, możemy kodować za pomocą liczb naturalnych, odpowiednio 0 i 1. Próbą może być dowolny ciąg zero- jedynkowy. Przestrzeń prób w tym przypadku to $\mathcal{X} = \{0, 1\}^n \subseteq R^n$.

Prawdopodobieństwo (nieznane) wyrzucenia w jednym rzucie orła ($x = 0$), a także reszki ($x = 1$) można wyrazić jednolitym wzorem $\theta^x(1-\theta)^{1-x}$ (reszka wypada z prawdopodobieństwem θ, orzeł – $1 - \theta$). Z matematycznego punktu widzenia θ musi być liczbą rzeczywistą z przedziału $(0, 1)$.

Przy niezależnych rzutach monetą można przyjąć, że prawdopodobieństwo uzyskania wyniku $(x_1, x_2, \ldots, x_n)$ jest dane wzorem

$$P_\theta(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \theta^{x_i}(1 - \theta)^{1-x_i}.$$

*Niniejszy tekst powstał w oparciu o podręczniki *Wnioskowanie statystyczne* S.D. Silvey'a (PWN 1978) oraz *Statystyka matematyczna* Mirosława Krzyśki (WN UAM 2004)

Teraz możemy rozważać różne hipotezy dotyczące tego modelu, na przykład $H_0 = \{1/2\}$ i $H_1 = (0,1) \setminus \{1/2\}$. Statystycy zajmując się taką hipotezą powiedzieliby i zapisaliby raczej, że zajmują się weryfikacją hipotezy $H_0 : \theta = 1/2$ przeciwko hipotezie $H1 : \theta \neq 1/2$. W tym przypadku należy przyjąć, że $\Theta = (0,1)$. Możnaby też przyjąć, że $\Theta = \{1/3, 1/4\}$, $H_0 = \{1/3\}$ i $H_1 = \{1/4\}$. Wtedy byłaby weryfikowana hipoteza $H_0 : \theta = 1/3$ przeciwko hipotezie $H_1 : \theta \neq 1/4$, czyli stwierdzenie, że z dwóch podanych liczb lepiej opisuje rozważane doświadczenie liczba $\theta = 1/3$.

1.4 Test, zbiór odrzucenia, statystyka testowa

Testem hipotezy H_0 (przeciwko hipotezie H_1) nazywamy dowolną funkcję $\varphi : \mathcal{X} \to \{0,1\}$. Przyjmujemy, że $\varphi(\vec{x}) = 1$ oznacza, że odrzucamy hipotezę H_0, a jeżeli $\varphi(\vec{x}) = 0$, to nie widzimy powodu odrzucenia H_0. Raczej nie powinniśmy stwierdzać, że zachodzi hipoteza H_0. Wskazane są słabsze sformułowania, że hipoteza ta wydaje się prawdziwa, rekomendujemy jej przyjęcie, lub coś podobnego.

Zbiór $B = \{\vec{x} \in \mathcal{X} : \varphi(\vec{x}) = 1\}$ nazywamy zbiorem odrzucenia lub obszarem krytycznym. Często zbiór B można zdefiniować tak, by spełniał równość postaci $B = \{\vec{x} \in \mathcal{X} : t(\vec{x}) \in A\}$ dla pewnej funkcji $t : \mathcal{X} \to R$. Taką funkcję nazywamy statystyką testową.

1.5 Błędy, moc testu

Mówimy, że popełniamy błąd I rodzaju, jeżeli niesłusznie odrzucamy hipotezę H_0. Błąd II rodzaju polega na niesłusznym nieodrzuceniu hipotezy H_0.

Przypuśćmy, że mamy próbę $\vec{x}$ (znalezioną w wyniku przeprowadzenia stosownych badań). Mocą testu nazywamy funkcję $\beta : \Theta \to [0,1]$ zdefiniowaną wzorem

$$\beta(\theta) = P_\theta(\vec{x} \in B).$$

Zauważmy, że jeżeli prawdziwy rozkład ma parametry θ i $\theta \in H_0$, to

$$\beta(\theta) = P_\theta(\vec{x} \in B) = \text{ prawdopodobieństwo popełnienia błędu I rodzaju,}$$

a gdy $\theta \in H_1$, to

$$\beta(\theta) = P_\theta(\vec{x} \in B) = 1 - \text{ prawdopodobieństwo błędu II rodzaju,}$$

Nietrudno zauważyć, że byłoby najlepiej, gdyby udało się znaleźć test, dla którego prawdopodobieństwa popełnienia błędów obu rodzajów byłyby równe 0, a więc taki, że $\beta(\theta) = 0$ dla $\theta \in H_0$ oraz $\beta(\theta) = 1$ w pozostałych przypadkach.

1.6 Poziom istotności i rozmiar testu

Mówimy, że test jest testem o poziomie istotności α, jeżeli

$$sup_{\theta \in H_0} \beta(\theta) \leqslant \alpha.$$

Test ma rozmiar α, jeżeli

$$sup_{\theta \in H_0} \beta(\theta) = \alpha.$$

2 Testy oparte na ilorazie wiarygodności

(Cały czas są zachowywane wprowadzone oznaczenia.)

2.1 Funkcja i iloraz wiarygodności

Zamiast rozkładu P_θ będziemy teraz posługiwać się jego gęstością L_θ. Przyjmijmy, że $L(\theta, \vec{x}) = L_\theta(\vec{x})$. Funkcja L jest funkcją dwóch zmiennych. Ustalając pierwszą zmienną na θ przekształcamy L w gęstość rozkładu P_θ. Jeżeli ustalimy drugą zmienną na $\vec{x}$, to otrzymujemy funkcję wiarygodności $\vec{x}$.

Ilorazem wiarygodności będziemy nazywać funkcję $\lambda : \mathcal{X} \to R$ daną wzorem

$$\lambda(\vec{x}) = \frac{sup_{\,\theta \in H_0} L(\theta, \vec{x})}{sup_{\,\theta \in \Theta} L(\theta, \vec{x})}.$$

2.2 Test oparty o iloraz wiarygodności

Testem opartym o iloraz wiarygodności nazywamy test na poziomie istotności α hipotezy H_0 przeciwko hipotezie alternatywnej H_1, który ma obszar odrzucenia zdefiniowany nierównością

$$\lambda(\vec{x}) \leqslant \lambda_0,$$

gdzie λ_0 jest tak dobrane, aby $P(\lambda(\vec{x}) \leqslant \lambda_0 | H_0) = \alpha$.

Zauważmy od razu, że przeprowadzenie tego testu wymaga założenia o skończoności $sup_{\,\theta \in \Theta}\, L(\theta, \vec{x})$.

Sensowność takiego testu, jego ewentualne własności to osobne zagadnienie, zwykle wykraczające poza rozważania matematyczne.

3 Pierwszy przykład

Przypuśćmy, że z jakiś powodów wyliczyliśmy m wartości $X = (x_1, \ldots, x_m)$ zmiennych losowych o tym samym rozkładzie $N(\mu_1, \sigma^2)$ i n wartości $Y = (y_1, \ldots, Y_n)$ zmiennych losowych o (tym samym) rozkładzie $N(\mu_2, \sigma^2)$. Chcemy skonstruować test weryfikujący hipotezę $H_0 : \mu_1 = \mu_2$ przeciwko hipotezie alternatywnej $H_1 : \mu_1 \neq \mu_2$ (parametry μ_1 i μ_2 nie są znane, testujemy ich równość).

Oczywiście przyjmujemy, że

$$\mathcal{X} = R^{m+n}, \quad \Theta = \{(m_1, m_2, v) \in R^3 : v > 0\},$$

$$H_0 = \{(m, m, v) \in R^3 : v > 0\}, \quad H_1 = \{(m_1, m_2, v) \in R^3 : m_1 \neq m_2 \wedge v > 0\},$$

$$L(\mu_1, \mu_2, \sigma^2, X, Y) = \frac{1}{\sqrt{(2\pi\sigma^2)^{m+n}}} e^{-\frac{\sum_{i=1}^{m}(x_i - \mu_1)^2 + \sum_{i=1}^{n}(y_i - \mu_2)^2}{2\sigma^2}}.$$

3.1 Obliczanie $sup_{\,\theta \in \Theta} L(\theta, \vec{x})$

Obliczanie tego supremum sprowadza się do znalezienia metodą największej wiarygodności estymatorów parametrów rozkładu. Możemy więc postępować podobnie. Najpierw logarytmujemy gęstość, następnie obliczamy pochodne cząstowe logarytmu:

$$\log L(\mu_1, \mu_2, \sigma^2, X, Y) = -\frac{m+n}{2} \log(2\pi\sigma^2) - \frac{\sum_{i=1}^{m}(x_i - \mu_1)^2 + \sum_{i=1}^{n}(y_i - \mu_2)^2}{2\sigma^2},$$

$$\frac{\partial \log L(\mu_1,\mu_2,\sigma^2,X,Y)}{\partial \mu_1} = \frac{\sum_{i=1}^{m}(x_i-\mu_1)}{\sigma^2},$$

$$\frac{\partial \log L(\mu_1,\mu_2,\sigma^2,X,Y)}{\partial \mu_2} = \frac{\sum_{i=1}^{n}(y_i-\mu_2)}{\sigma^2},$$

$$\frac{\partial \log L(\mu_1,\mu_2,\sigma^2,X,Y)}{\partial \sigma^2} = -\frac{m+n}{2\sigma^2} + \frac{\sum_{i=1}^{m}(x_i-\mu_1)^2+\sum_{i=1}^{n}(y_i-\mu_2)^2}{2\sigma^4}.$$

Licząc ostatnią pochodną różniczkujemy nie po σ, a po zmiennej, za którą podstawiliśmy σ^2. Wobec tego układ równań

$$\frac{\partial \log L(\mu_1,\mu_2,\sigma^2,X,Y)}{\partial \mu_1} = \frac{\partial \log L(\mu_1,\mu_2,\sigma^2,X,Y)}{\partial \mu_2} = \frac{\partial \log L(\mu_1,\mu_2,\sigma^2,X,Y)}{\partial \sigma^2} = 0$$

ma rozwiązanie

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{m} x_i}{m} = \bar{X},\ \ \hat{\mu}_2 = \bar{Y},\ \ \hat{\sigma}^2 = \frac{1}{m+n}\left(\sum_{i=1}^{m}(x_i-\bar{X})^2+\sum_{i=1}^{n}(y_i-\bar{Y})^2\right).$$

Dalej powinniśmy sprawdzić, że znaleźliśmy globalne maksimum interesującej nas funkcji. Jeżeli tak rzeczywiście jest, to logarytm interesującego nas supremum $sup_{\,\theta\in\Theta}L(\theta,\vec{x})$ jest dany wzorem

$$\log sup_{\,\Theta}L(\mu_1,\mu_2,\sigma^2,X,Y) = \log L(\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}^2,X,Y) = -\frac{m+n}{2}\log(2\pi\hat{\sigma}^2) - \frac{m+n}{2}.$$

3.2 Obliczanie $sup_{\,\theta\in H_0}L(\theta,\vec{x})$

Analogicznie obliczamy drugie z dwóch potrzebnych supremów. Po uwzględnieniu równości $\mu_1=\mu_2$ zachodzącej dla parametrów z H_0 logarytmujemy gęstość, następnie obliczamy pochodne cząstowe logarytmu:

$$\log L(\mu,\sigma^2,X,Y) = -\frac{m+n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^{m}(x_i-\mu)^2+\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2},$$

$$\frac{\partial \log L(\mu,\sigma^2,X,Y)}{\partial \mu} = \frac{\sum_{i=1}^{m}(x_i-\mu)+\sum_{i=1}^{n}(y_i-\mu)}{\sigma^2},$$

$$\frac{\partial \log L(\mu,\sigma^2,X,Y)}{\partial \sigma^2} = -\frac{m+n}{2\sigma^2} + \frac{\sum_{i=1}^{m}(x_i-\mu)^2+\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^4}.$$

Tym razem układ równań

$$\frac{\partial \log L(\mu_1,\sigma^2,X,Y)}{\partial \mu} = \frac{\partial \log L(\mu,\sigma^2,X,Y)}{\partial \sigma^2} = 0$$

ma rozwiązanie

$$\tilde{\mu} = \frac{\sum_{i=1}^m x_i + \sum_{i=1}^n y_i}{m+n} = \frac{m\bar{X} + n\bar{Y}}{m+n}, \quad \tilde{\sigma}^2 = \frac{1}{m+n}\left(\sum_{i=1}^m (x_i - \tilde{\mu})^2 + \sum_{i=1}^n (y_i - \tilde{\mu})^2\right).$$

Po sprawdzeniu, że mamy globalne maksimum interesującej nas funkcji otrzymujemy, że logarytm interesującego nas supremum $sup_{\,\theta\in\Theta} L(\theta, \vec{x})$ jest dany wzorem

$$\log sup_{\,\Theta} L(\mu, \mu, \sigma^2, X, Y) = \log L(\tilde{\mu}, \tilde{\mu}, \tilde{\sigma}^2, X, Y) = -\frac{m+n}{2}\log(2\pi\tilde{\sigma}^2) - \frac{m+n}{2}.$$

3.3 Statystyka testowa

Test oparty o iloraz wiarygodności rozstrzyga hipotezę analizując wartość ilorazu wiarygodności

$$\lambda(\vec{x}) = \frac{sup_{\,\theta\in H_0} L(\theta, \vec{x})}{sup_{\,\theta\in\Theta} L(\theta, \vec{x})}.$$

W naszym przypadku będziemy analizować wartość funkcji

$$T_1(X,Y) = \frac{sup_{\,(\mu,\mu,\sigma^2)\in H_0} L(\mu, \mu, \sigma^2, X, Y)}{sup_{\,(\mu_1,\mu_2,\sigma^2)\in\Theta} L(\mu_1, \mu_2, \sigma^2, X, Y)}.$$

Z przeprowadzonych obliczeń wynika, że logarytm tej funkcji jest dany wzorem

$$\log T_1(X,Y) = \frac{m+n}{2}\log(2\pi\hat{\sigma}^2) - \frac{m+n}{2}\log(2\pi\tilde{\sigma}^2) = \log\left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{\frac{m+n}{2}}.$$

Ostatecznie otrzymujemy, że

$$T_1(X,Y) = \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{\frac{m+n}{2}}.$$

Wyliczona właśnie funkcja T_1 pełni w rozważanym teście rolę jednej z możliwych statystyk testowych: wynik testu otrzymujemy porównując wartość $T_1(X,Y)$ z pewną liczbą λ_0.

3.4 Pewien wzór

Na siódmej liście zadań pojawił się następujący wzór:

$$\sum_{i=1}^m (x_i - \mu)^2 = \sum_{i=1}^m (x_i - \bar{X})^2 + m(\bar{X} - \mu)^2.$$

Korzystając z tego wzoru można pokazać także wzór następujący:

$$\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} = 1 + \frac{mn}{(m+n)^2}\cdot\frac{(\bar{X} - \bar{Y})^2}{\hat{\sigma}^2}.$$

Jeżeli teraz przyjmiemy, że

$$T_2(X,Y) = \frac{\sqrt{mn}}{m+n}\cdot\frac{\bar{X} - \bar{Y}}{\hat{\sigma}},$$

to będzie prawdziwa zależność

$$T_1(X,Y) = (1 + T_2^2(X,Y))^{-\frac{m+n}{2}}.$$

W tej sytuacji warunki

$$T_1(X,Y) \leqslant (a^2+1)^{-\frac{m+n}{2}} \quad \text{oraz} \quad |\,T_2(X,Y)\,| \geqslant a$$

są równoważne. Dalej będziemy starać się zdefiniować odpowiednie a.

3.5 Trochę o rozkładach różnych zmiennych

Przypuśćmy, że mamy niezależne zmienne losowe $X_1, \ldots, X_m$ oraz $Y_1, \ldots, Y_n$, wszystkie o rozkładzie normalnym $N(\mu, \sigma^2)$.

Znamy wiele faktów dotyczących rozkładów zmiennych, w szczególności

1) jeżeli $X \sim N(\mu, \sigma^2)$, to $cX \sim N(c\mu, c^2\sigma^2)$,

2) jeżeli $X \sim N(\mu_1, \sigma_1^2)$ i $Y \sim N(\mu_2, \sigma_2^2)$, to $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

3) jeżeli $X_i \sim N(\mu, \sigma^2)$ dla $i = 1, \ldots, m$, to $\bar{X} \sim N(\mu, \frac{\sigma^2}{m})$.

W szczególności, dla rozważanych zmiennych, z powyższych faktów, z zadania 5 z listy 7 i z definicji odpowiednich rozkładów otrzymujemy

1)
$$\sqrt{\frac{mn}{m+n}} \cdot \frac{\bar{X} - \bar{Y}}{\sigma} \sim N(0, 1),$$

2)
$$\frac{\sum_{i=1}^{m}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(m-1)$$

i analogicznie dla zmiennych Y_i (patrz zadanie 5),

3)
$$\frac{\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2(m+n-2).$$

Z definicji rozkładu Studenta[1] otrzymujemy teraz, że zmienna

$$T_3(X, Y) = \sqrt{m+n-2} \cdot T_2(X, Y) = \sqrt{m+n-2} \cdot \frac{\sqrt{mn}}{m+n} \cdot \frac{\bar{X} - \bar{Y}}{\hat{\sigma}} =$$

$$= \frac{\sqrt{\frac{mn}{m+n}} \cdot \frac{\bar{X} - \bar{Y}}{\sigma}}{\sqrt{\frac{(m+n)\hat{\sigma}^2}{(m+n-2)\sigma^2}}} = \frac{\sqrt{\frac{mn}{m+n}} \cdot \frac{\bar{X} - \bar{Y}}{\sigma}}{\sqrt{\frac{1}{m+n-2} \frac{\sum_{i=1}^{m}(x_i - \bar{X})^2 + \sum_{i=1}^{n}(y_i - \bar{Y})^2}{\sigma^2}}}$$

ma właśnie taki rozkład $t(m+n-2)$ z $m+n-2$ stopniami swobody.

Tak więc dowiedliśmy, że zmienna losowa T_3 ma ściśle określony rozkład. Wymagało to jednak założenia, wszystkie $m+n$ analizowane liczby były wartościami zmiennych losowych o tym samym rozkładzie.

[1] Rozkład Studenta $t(k)$ z k stopniami swobody można zdefiniować jako rozkład zmiennej losowej $\frac{U}{\sqrt{V/k}}$, gdzie zmienne U i V mają odpowiednio rozkłady $N(0, 1)$ i $\chi^2(k)$ i są niezależne.

3.6 Zapewnianie poziomu istotności testu

W rozważanym teście statystykami testowymi mogą być funkcje T_1, T_2 i T_3.

Zauważmy, że jeżeli prawdziwa jest hipoteza H_0, to warunki

$$|\,T_3(X,Y)\,| \geqslant b, \quad |\,T_2(X,Y)\,| \geqslant \frac{b}{\sqrt{m+n-2}} = a \text{ oraz } T_1(X,Y) \leqslant (a^2+1)^{-\frac{m+n}{2}}$$

są równoważne, patrz str. 3.4. Co więcej, łatwo jest kontrolować prawdopodobieństwo zachodzenia pierwszej nierówność. Jeżeli będzie zachodziła z pewnym prawdopodobieństwem, to z tym samym prawdopodobieństwem będą zachodziły pozostałe nierówności.

Weźmy teraz liczbę α, $0 < \alpha < 1$. Zagwarantujemy, że pierwsza z tych nierówności zachodzi z prawdopodobieństwem α.

Niech F oznacza dystrybuantę rozkładu Studenta z $m+n-2$ stopniami swobody. Ponieważ dystrybuanta rozkładu Studenta jest funkcją ciągła, więc dla pewnej liczby b zachodzi równość $F(b) = 1 - \frac{\alpha}{2}$. Ponieważ rozkład Studenta jest symetryczny, więc $F(0) = \frac{1}{2}$ oraz $b > 0$. Oczywiście,

$$P(T_3(X,Y) \geqslant b) = 1 - F(b) = 1 - (1 - \frac{\alpha}{2}) = \frac{\alpha}{2}.$$

Z symetryczności rozkładu otrzymujemy, że

$$P(|\,T_3(X,Y)\,| \geqslant b) = \alpha.$$

4 Prostszy przykład, teoria Neumana - Pearsona

Sposób testowania wykorzystany w poprzednim przykładzie zastosujemy teraz do dowolnych rozkładów w sytuacji, gdy rozważamy tylko dwa parametry θ_0 oraz θ_1. Będziemy testować hipotezę $H_0 : \theta = \theta_0$ przeciwko hipotezie $H_1 : \theta = \theta_1$. W tym przypadku iloraz wiarygodności ma postać

$$\lambda(X) = \frac{L(\theta_0, X)}{\max\{L(\theta_0, X), L(\theta_1, X)\}}.$$

Poza tym $\lambda(X) \in [0,1]$. Wobec tego jest sens rozważać tylko zbiory odrzucenia postaci $\lambda(X) < \lambda_0$ dla $\lambda_0 < 1$. Wtedy jednak warunek $\lambda(X) < \lambda_0$ można równoważnie wyrazić w postaci

$$L(\theta_0, X) < \lambda_0 L(\theta_0, X) \vee L(\theta_0, X) < \lambda_0 L(\theta_1, X) \;\text{ oraz }\; L(\theta_0, X) < \lambda_0 L(\theta_1, X).$$

Powinniśmy zajmować się więc zbiorami odrzucenia postaci

$$B = \{X \in \mathcal{X} : L(\theta_0, X) < \lambda_0 L(\theta_1, X)\}.$$

4.1 Intuicyjne uzasadnienie metody

W przypadku dyskretnym, a z pewnymi zastrzeżeniami także w ciągłym, gęstość $L(\theta, X)$ to prawdopodobieństwo uzyskania wyniku X podczas losowych doświadczeń podlegających rozkładowi z parametrem θ. Liczba ta jest na ogół mała, zwłaszcza jeżeli jest możliwe bardzo wiele różnych wyników. Interesuje nas jednak, czy liczba ta jest relatywnie mała w porównaniu z prawdopodobieństwami wyliczonymi dla innych parametrów. Prawdopodobieństwo wyliczone zgodnie z właściwym rozkładem powinno dać wynik typowy dla rozważanej sytuacji. Jeżeli wydaje się być bardzo małe i sprawia wrażenie zaniżonego, to może być wyliczone zgodnie z rozkładem z niepoprawnym parametrem.

4.2 Test i jego poziom istotności

Po pierwsze zauważmy, że w tym bardzo prostym przykładzie poziom istotności testu powinien szacować tylko dla jednego rozkładu, z parametrem θ_0, prawdopodobieństwo odrzucenia hipotezy.

Tym razem będzie nas interesować test określonego rozmiaru α. Taki test powinien istnieć dla rozkładów ciągłych: dla danego α powinno dać się dobrać λ_0 takie, że

$$P_{\theta_0}(\lambda(X) < \lambda_0) = \alpha.$$

Dla rozkładów dyskretnych może być jednak różnie. Wtedy, ustępując nieco z wymagań dotycząch α, może uda nam się zapewnić prawdziwość powyższej równości.

Jeżeli tak będzie, to test ze zbiorem odrzucenia B, polegający na sprawdzeniu, czy $\lambda(X) < \lambda_0$, będzie testem rozmiaru α. Dalej będziemy zajmować się tym testem.

4.3 Lemat Neumana - Pearsona

Mówimy, że test jest jednostajnie najmocniejszy na poziomie istotności α, jeżeli w tym teście prawdopodobieństwo popełnienia błędu II rodzaju jest możliwie małe, czyli nie przekracza prawdopodobieństwa popełnienia błędu II rodzaju dla innych testów rozważanej hipotezy o poziomie istotności α.

Dla rozważanych teraz hipotez prawdopodobieństwo popełnienia błędu II rodzaju wyraża się wzorem

$$P_{\theta_1}(\mathcal{X} \setminus B) = 1 - P_{\theta_1}(B).$$

Jest ono możliwie małe wtedy, gdy prawdopodobieństwo $P_{\theta_1}(B)$ jest możliwie duże i przekracza analogiczne prawdopodobieństwo dla innych testów tej samej hipotezy na poziomie istotności α.

Lemat 4.1 (Neumana - Pearsona) *Przypuśćmy, że α i B są takie, jak wyżej, oraz B' jest zbiorem odrzucenia pewnego testu hipotezy $H_0 : \theta = \theta_0$ przeciwko hipotezie $H_1 : \theta = \theta_1$ na poziomie istotności α. Wtedy $P_{\theta_1}(B) \geqslant P_{\theta_1}(B')$.*

Dowód. Najpierw zauważmy, że z nierówności $L(\theta_0, X) < \lambda_0 L(\theta_1, X)$ słusznej dla $X \in B$ wynika, że

$$P_{\theta_1}(B \setminus B') = \int_{B\setminus B'} L(\theta_1, X)\, dX \geqslant \frac{1}{\lambda_0} \int_{B\setminus B'} L(\theta_0, X)\, dX = \frac{1}{\lambda_0} P_{\theta_0}(B \setminus B').$$

Analogicznie, z nierówności $L(\theta_0, X) \geqslant \lambda_0 L(\theta_1, X)$ zachodzącej dla $X \notin B$ otrzymujemy, że

$$P_{\theta_1}(B' \setminus B) \leqslant \frac{1}{\lambda_0} P_{\theta_0}(B' \setminus B)).$$

Łącząc te dwie nierówności otrzymujemy, że

$$P_{\theta_1}(B \setminus B') - P_{\theta_1}(B' \setminus B) \geqslant \frac{1}{\lambda_0}(P_{\theta_0}(B \setminus B') - P_{\theta_0}(B' \setminus B).$$

Łatwo zauważyć, że

$$P_\theta(B) - P_\theta(B') = P_\theta(B \setminus B') - P_\theta(B' \setminus B)$$

dla wszelkich rozkładów. Łącząc dwa ostatnie fakty dostajemy tezę lematu

$$P_{\theta_1}(B) - P_{\theta_1}(B') \geqslant \frac{1}{\lambda_0}(P_{\theta_0}(B) - P_{\theta_0}(B')) = \frac{1}{\lambda_0}(\alpha - P_{\theta_0}(B')) \geqslant 0. \ \square$$

Jak widać, lemat Neumana - Pearsona stwierdza bezpośrednio, że w przypadku hipotezy $H_0 : \theta = \theta_0$ przeciwko hipotezie $H_1 : \theta = \theta_1$ testy oparte na ilorazie wiarygodności mają dobre własności: są jednostajnie najmocniejsze.