WYKŁADY Z TEORII TESTOWANIA HIPOTEZ

Tadeusz Inglot

WYKŁADY Z TEORII TESTOWANIA HIPOTEZ

Oficyna Wydawnicza GiS
Wrocław 2021

Tadeusz Inglot
Wydział Matematyki
Politechnika Wrocławska
tadeusz.inglot@pwr.edu.pl

Projekt okładki:
IMPRESJA Studio Grafiki Reklamowej

Skład książki w systemie LaTeX wykonał autor.

ISBN 978-83–62780-90–7

Wydanie I, Wrocław 2021
Oficyna Wydawnicza GiS, s.c., www.gis.wroc.pl
Druk i oprawa: Drukarnia I-BiS Bierońscy, Sp. kom.

Spis treści

Przedmowa

Niniejszy podręcznik jest zapisem wykładów z *Teorii testowania hipotez*, które prowadziłem w latach 2005–2015 dla studentów matematyki specjalności *statystyka matematyczna* na 2 lub 3 semestrze studiów magisterskich (8 lub 9 semestr jednolitych pięcioletnich studiów magisterskich). Przedmiot obejmował 30 godzin wykładu oraz 30 godzin laboratorium z ćwiczeniami audytoryjnymi. Książka nie jest więc całościowym przeglądem bardzo obszernego współcześnie działu statystyki, a zawiera raczej pewien subiektywny, ograniczony czasem wykładu, wybór materiału. Starałem się pokazać najważniejsze wyniki klasyczne, a równocześnie nie rezygnować z odniesień do aktualnych, nowych rozwiązań. Na przykładzie wybranych zagadnień i twierdzeń, przedstawiłem ogólne idee, które przewijają się w różnych zagadnieniach testowania oraz te fakty z teorii prawdopodobieństwa, które stoją za rozważaniami statystycznymi. Uzupełnieniem wykładu są propozycje zadań i ćwiczeń laboratoryjnych. Z jednej strony ilustrują wykład, a z drugiej rozszerzają jego zakres poprzez wprowadzenie do innych problemów testowania i zapozanie czytelnika z nieobjętymi wykładem testami znanymi z literatury ostatnich kilkudziesięciu lat.

Rozdział 7.3 zawiera podstawowe wiadomości o efektywności Kallenberga, które nie były dotąd prezentowane w literaturze w języku polskim poza artykułem Ledwiny (2005). Twierdzenie 12 i jego dowód w rozdziale 8.2, oparty na monografii [**Va**], są nowe. Rozdział 8 wykracza poza ramy wykładu i został dodany dla kompletności niniejszego opracowania. Rozdział 4 można wyłożyć przed rozdziałami 2 i 3.

Od czytelnika wymaga się znajomości standardowych wykładów z analizy, algebry, teorii miary, analizy funkcjonalnej, rachunku prawdopodobieństwa oraz statystyki matematycznej przewidzianych w programach studiów matematycznych. Do podstawowych pojęć i twierdzeń z tych przedmiotów odwołuję się bez bliższych wyjaśnień lub odsyłania do podręczników.

Wrocław, wrzesień 2021 r. Tadeusz Inglot

1. Wstęp

Niech X będzie zmienną losową o wartościach w przestrzeni mierzalnej $(\mathcal{X}, \mathcal{A})$ oraz $\mathcal{P}$ pewną rodziną rozkładów na $\mathcal{X}$ i niech $X_1, ..., X_n$ będzie próbą prostą z rozkładu $P \in \mathcal{P}$. Produkt $(\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}^n) = (\mathcal{X}, \mathcal{A}, \mathcal{P})^n$ nazywamy modelem statystycznym. W dalszym ciągu wykładu, mówiąc o takim modelu, będziemy używać wygodnego terminu roboczego *model produktowy.*

Pojęcie modelu statystycznego definiuje się ogólniej, dopuszczając obserwacje o różnych rozkładach, a nawet zależne (co tutaj pomijamy). Niech $X_1, ..., X_n$ będą niezależnymi zmiennymi losowymi, gdzie X_i mają wartości w przestrzeniach $(\mathcal{X}_i, \mathcal{A}_i, \mathcal{P}_i),\ i = 1, ..., n$, a $\mathcal{P}_i$ są pewnymi rodzinami rozkładów. Wtedy przestrzeń produktową

$$(\mathcal{X}_1 \times ... \times \mathcal{X}_n, \mathcal{A}_1 \times ... \times \mathcal{A}_n, \mathcal{P}_1 \times ... \times \mathcal{P}_n)$$

nazywamy modelem statystycznym. Przypuśćmy, że $\widetilde{\mathcal{P}}_0 \subset \widetilde{\mathcal{P}} = \mathcal{P}_1 \times ... \times \mathcal{P}_n$ jest ustalonym podzbiorem $\widetilde{\mathcal{P}}$ oraz $\widetilde{\mathcal{P}}_1 = \widetilde{\mathcal{P}} \setminus \widetilde{\mathcal{P}}_0$. Rozważamy zagadnienie decyzyjne polegające na rozstrzygnięciu czy $\widetilde{P} \in \widetilde{\mathcal{P}}_0$ czy $\widetilde{P} \in \widetilde{\mathcal{P}}_1$ na podstawie próby $X_1, ..., X_n$, gdzie $\widetilde{P}$ jest rozkładem tej próby (w modelu produktowym $\widetilde{P} = P^n$). Oznaczmy dwie konkurencyjne hipotezy $H_j : \widetilde{P} \in \widetilde{\mathcal{P}}_j,\ j = 0, 1$. Rozważamy więc problem testowania H_0 (hipoteza zerowa) przeciwko H_1 (hipoteza alternatywna). Rozstrzygnięcia dokonuje się za pomocą testu statystycznego.

Bez straty ogólności możemy przyjąć, że rodzina $\mathcal{P}$ (w modelu produktowym) jest sparametryzowana, tzn. $\mathcal{P} = \{P_\gamma : \gamma \in \Gamma\}$, gdzie Γ jest pewnym zbiorem parametrów. Wtedy, oznaczając $\Gamma_0 = \{\gamma : P_\gamma \in \mathcal{P}_0\}$, równoważnie $H_0 : \gamma \in \Gamma_0$ przeciwko $H_1 : \gamma \in \Gamma_1 = \Gamma \setminus \Gamma_0$. Jeśli Γ jest podzbiorem $\mathbb{R}^k,\ k \geqslant 1$, to mówimy o problemie parametrycznym, w przeciwnym wypadku nieparametrycznym. Często Γ ma strukturę produktową $\Gamma = \Theta \times E$. Wtedy parametr oznaczamy $\gamma = (\vartheta, \eta)$. Taka sytuacja ma miejsce przy testowaniu hipotezy $H_0 : \gamma \in \Theta_0 \times E$ przeciwko $H_1 : \gamma \in \Theta_1 \times E$, gdzie Θ_0 jest pewnym podzbiorem Θ, a $\Theta_1 = \Theta \setminus \Theta_0$.

Nieokreślony przez hipotezę parametr $\eta \in E$ nazywamy parametrem zakłócającym (ang. nuisance parameter). Jeśli $\Theta \subset \mathbb{R}^k$, $k \geqslant 1$, oraz E jest podzbiorem przestrzeni nieskończenie wymiarowej, to mówimy o problemie semiparametrycznym.

Problemy nieparametryczne są ważnymi zagadnieniami wnioskowania statystycznego. Prawie cały nasz wykład poświęcimy tym zagadnieniom i metodom konstrukcji testów nieparametrycznych. Dla prostoty ograniczymy się do testów prawostronnych niezrandomizowanych (co nie jest istotnym ograniczeniem). Niech $T_n = T(X_1, ..., X_n)$ będzie statystyką o wartościach rzeczywistych.

Definicja 1. Testem prawostronnym hipotezy H_0 przeciwko H_1 (w modelu produktowym) określonym przez statystykę testową T_n na poziomie istotności $\alpha \in (0, 1)$ nazywamy funkcję $\mathbf{1}_{C_{t_\alpha}}$, gdzie

$$C_t = \{(x_1, ..., x_n) \in \mathcal{X}^n : T(x_1, ..., x_n) > t\}, \ \ t \in \mathbb{R},$$

$\mathbf{1}_A$ oznacza funkcję charakterystyczną zbioru A, a wartość krytyczna t_α jest równa

$$t_\alpha = \inf\{t \in \mathbb{R} : \sup_{P \in \mathcal{P}_0} P^n(C_t) \leqslant \alpha\}.$$

Jeśli $\mathbf{1}_{C_{t_\alpha}} = 1$, to H_0 odrzucamy, w przeciwnym wypadku H_0 przyjmujemy. Liczbę $\sup_{P \in \mathcal{P}_0} P^n(C_{t_\alpha})$ nazywamy rozmiarem testu T (czyli określonego przez statystykę T_n). Definicję 1 można przeformułować na dowolne modele, wprowadzając oczywiste zmiany.

W wielu klasycznych problemach nieparametrycznych hipotezę zerową można sprowadzić do znikania pewnej funkcji. Zobaczymy to na kilku ważnych przykładach.

Przykład 1. (*testowanie zgodności na prostej*) Niech $\mathcal{X} = \mathbb{R}$, $\mathcal{P}$ będzie rodziną rozkładów na $\mathbb{R}$, a $X_1, ..., X_n$ próbą z rozkładu $P \in \mathcal{P}$ o dystrybuancie $F(x)$. Dla ustalonego $P_0 \in \mathcal{P}$ o ciągłej dystrybuancie $F_0(x)$ rozważamy testowanie hipotezy prostej $H_0 : P = P_0$ przeciwko $H_1 : P \neq P_0$. Oznaczmy $g(x) = F(x) - F_0(x)$. Wtedy równoważnie możemy napisać $H_0 : g = 0$. Zwykle rozważa się próbę transformowaną na odcinek $[0, 1]$ przez nałożenie dystrybuanty hipotetycznej $F_0(x)$. Jeśli więc $U_i = F_0(X_i)$, $i = 1, ..., n$, to zmienne losowe U_i mają, przy prawdziwości hipotezy H_0, rozkład jednostajny na odcinku $[0, 1]$ (por.

lemat na początku rozdziału 4) i zagadnienie sprowadza się do testowania jednostajności i rozstrzygnięcia czy transformowana funkcja g, czyli $g \circ F_0^{-1}(t) = F \circ F_0^{-1}(t) - t,\ \ t \in (0,1)$, znika. Temu problemowi poświęcony jest rozdział 4.

Przykład 2. (*testowanie niezależności*) Niech $\mathcal{X} = \mathbb{R}^2$, a $\mathcal{P}$ będzie rodziną rozkładów na płaszczyźnie o ciągłych dystrybuantach brzegowych oznaczanych przez $F(x)$ oraz $G(y)$ i niech $(X_1, Y_1), ..., (X_n, Y_n)$ będzie próbą z rozkładu $P \in \mathcal{P}$ o dystrybuancie $H(x,y)$. Rozważamy testowanie hipotezy o niezależności rozkładów brzegowych $H_0 : H(x,y) = F(x)G(y),\ \ x, y \in \mathbb{R}$. Oznaczmy $g(x,y) = H(x,y) - F(x)G(y)$. Wtedy równoważnie możemy napisać $H_0 : g = 0$. Analogicznie jak w poprzednim przykładzie, nałożenie dystrybuant brzegowych na obserwacje prowadzi do próby transformowanej $(U_i, V_i) = (F(X_i), G(Y_i)),\ \ i = 1, ..., n$, i problem sprowadza się do testowania niezależności na kwadracie jednostkowym. Równoważnie możemy napisać $H_0 : g(F^{-1}(s), G^{-1}(t)) = C(s,t) - st \equiv 0$, gdzie $C(s,t) = H(F^{-1}(s), G^{-1}(t))$ jest funkcją łącznikową (ang. copula function). Nieokreślone przez hipotezę dystrybuanty brzegowe F, G są parametrami zakłócającymi w rozważanym problemie. Zagadnieniu testowania niezależności poświęcone są zad. 44-46 oraz ćwiczenie laboratoryjne 7.

Przykład 3. (*testowanie symetrii*) Niech $\mathcal{X} = \mathbb{R}$, a $\mathcal{P}$ będzie rodziną rozkładów P na prostej o ciągłych dystrybuantach oznaczanych przez $F(x)$. Testujemy hipotezę H_0 : *P jest symetryczny względem 0.* Oznaczmy $F_s(x) = \frac{1}{2}(F(x) + 1 - F(-x))$. Rozkład P jest symetryczny względem 0 wtedy i tylko wtedy, gdy $F - F_s = 0$. Podobnie jak w poprzednich przykładach, transformujemy dane na odcinek jednostkowy przez nałożenie dystrybuanty $F_s(x)$. Wtedy $F \circ F_s^{-1} = A$ jest funkcją absolutnie ciągłą na $[0,1]$ o pochodnej oznaczanej przez a. Zatem równoważnie testujemy hipotezę $H_0 : g = 0$, gdzie $g(t) = a(t) - 1$. Nieokreślona przez hipotezę H_0 dystrybuanta $F_s(x)$ jest parametrem zakłócającym.

Dwa następne przykłady dotyczą modeli, w których obserwacje mogą mieć różne rozkłady.

Przykład 4. (*testowanie stałości regresji*) Niech $r(t),\ \ t \in [0,1]$, będzie nieznaną ciągłą funkcją (regresji) o całce równej 0. Obserwujemy proces stochastyczny $X_t = m + r(t) + \sigma\varepsilon_t,\ t \in [0,1]$, gdzie ε_t jest szumem losowym o średniej 0 i wariancji 1 i rozkładach jednowymiarowych o tej samej dystrybuancie $F(x)$. Wybieramy różne punkty $t_1, ..., t_n$ z odcinka jednostkowego i otrzymujemy próbę $X_{t_1}, ..., X_{t_n}$. Zmienne losowe X_{t_i} mają na ogół rozkłady o różnych dystrybuantach $F((x - m - r(t_i))/\sigma)$. Testujemy hipotezę $H_0 : r = 0$. Temu zagadnieniu jest poświęcone ćwiczenie laboratoryjne 5.

Przykład 5. (*testowanie jednorodności, problem dwóch prób*) Niech $n_1 = n_1(n), n_2 = n_2(n)$ będą ciągami liczb naturalnych, takimi że $n_1 + n_2 = n$ i niech P, Q będą dwoma rozkładami na prostej o ciągłych dystrybuantach $F(x), G(x)$. Rozważamy model $(\mathbb{R}^n, \mathcal{R}^n, P^{n_1} \times Q^{n_2})$, próbę $X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}$ i hipotezę $H_0 : P = Q$. Oznaczmy $H_n(x) = (n_1F(x) + n_2G(x))/n$. Po transformacji próby na odcinek $(0,1)$ za pomocą dystrybuanty $H_n(x)$ różnica dystrybuant $F(x) - G(x)$ przyjmuje postać $g(t) = F \circ H_n^{-1}(t) - G \circ H_n^{-1}(t),\ t \in (0,1)$, i testujemy hipotezę $H_0 : g = 0$. Temu zagadnieniu są poświęcone zad. 33-42 oraz ćwiczenie laboratoryjne 6.

Powyższe przykłady problemów testowania i sprowadzenia ich do znikania pewnej funkcji tłumaczą podobieństwa konstrukcji testów w wielu istotnie różnych zagadnieniach. Jednak każde z nich ma swoją własną specyfikę wymagającą oddzielnego rozpatrzenia. Aby przedstawić możliwie szeroko ogólne idee konstrukcji testów, powtarzające się w różnych problemach, uczynimy to skupiając uwagę na testach zgodności.

W dalszym ciągu wykładu rozważamy wyłącznie modele produktowe.

2. Testowanie hipotez prostych

Rozważamy model $(\mathcal{X}, \mathcal{A}, \mathcal{P})^n$, gdzie $\mathcal{P} = \{P_0, P_1\}$ i testowanie hipotez prostych $H_0 : P = P_0$ przeciwko $H_1 : P = P_1$. Niech λ będzie miarą dominującą rozkłady P_0 i P_1, np. $\lambda = P_0 + P_1$. Oznaczmy $p_0 = dP_0/d\lambda$, $p_1 = dP_1/d\lambda$ gęstości obu konkurujących rozkładów. Wówczas statystykę (logarytm ilorazu funkcji wiarogodności)

$$V_n = V(X_1, ..., X_n) = \sum_{i=1}^{n} \log \frac{p_1(X_i)}{p_0(X_i)}$$

nazywamy statystyką Neymana-Pearsona, a test (prawostronny) oparty na niej testem Neymana-Pearsona (NP).

Lemat Neymana-Pearsona. Dla każdego $\alpha \in (0, 1)$ test Neymana-Pearsona o rozmiarze α jest najmocniejszym testem (niezrandomizowanym) H_0 przeciwko H_1, tzn. dla każdego testu T o tym samym rozmiarze α moc $\beta_V(P_1)$ testu V jest nie mniejsza niż moc $\beta_T(P_1)$ testu T, czyli $\beta_V(P_1) \geqslant \beta_T(P_1)$.

Dowód tego twierdzenia jest podany poniżej w ogólniejszej wersji jako twierdzenie 3. Można zapytać, jak duża jest ta największa moc $\beta_V(P_1) = \beta_V(P_1, n)$. Aby odpowiedzieć na to pytanie, określimy odległość (informację) Kullbacka-Leiblera rozkładu P_0 od P_1 wzorem

$$D(P_0||P_1) = D(p_0||p_1) = \int_{\mathcal{X}} p_0 \log \frac{p_0}{p_1} d\lambda \in [0, \infty],$$

gdy P_0 jest absolutnie ciągła względem P_1, tj. $P_0 \ll P_1$, oraz $+\infty$ w przeciwnym wypadku. Następujące twierdzenie zostało udowodnione przez Chernoffa (1956).

Twierdzenie 1 (Lemat Steina). Dla dowolnego $\alpha \in (0, 1)$ i testu Neymana-Pearsona na poziomie istotności α mamy

$$\lim_{n\to\infty}\frac{1}{n}\log(1-\beta_V(P_1,n))=-D(P_0||P_1).$$

Innymi słowy, $\beta_V(P_1,n)=1-\exp\{-nD(P_0||P_1)+o(n)\}$, czyli moc dąży wykładniczo do 1 i $D(P_0||P_1)$ określa tempo wykładniczego zanikania prawdopodobieństwa błędu drugiego rodzaju. Człon $o(n)$ zależy od przyjętego poziomu istotności i dla konkretnych wartości n wcale nie musi być mały w porównaniu z głównym członem.

Dowód. Oznaczmy przez $t_{\alpha n}$ dokładną wartość krytyczną testu NP, $A=\{V_n>t_{\alpha n}\}$ oraz dla dowolnego $\varepsilon>0$

$$B_\varepsilon=\{-nD(P_0||P_1)-n\varepsilon\leqslant V_n\leqslant -nD(P_0||P_1)+n\varepsilon\}.$$

Ponadto, dla $j=0,1$ oznaczmy przez $p_{jn}(x)=\prod_{i=1}^n p_j(x_i)$ gęstość łączną próby z rozkładu P_j. Wtedy $V_n=\log(p_{1n}/p_{0n})$ i

$$E_0V_n=n\int\limits_{\mathcal{X}}p_0\log\frac{p_1}{p_0}d\lambda=-nD(P_0||P_1).$$

Stąd i z prawa wielkich liczb Chinczyna mamy $P_0^n(B_\varepsilon^c)\to 0$. Ponieważ $\alpha\in(0,1)$, to $t_{\alpha n}\leqslant -nD(P_0||P_1)+n\varepsilon$ dla dostatecznie dużych n. Z nierówności Markowa

$$\begin{aligned}1-\beta_V(P_1,n)&=P_1^n(A^c)=P_1^n\left(\frac{p_{1n}}{p_{0n}}\leqslant e^{t_{\alpha n}}\right)=P_1^n\left(\frac{p_{0n}}{p_{1n}}\geqslant e^{-t_{\alpha n}}\right)\\&\leqslant e^{t_{\alpha n}}\int\limits_{\mathcal{X}^n}\frac{p_{0n}}{p_{1n}}p_{1n}d\lambda^n=e^{t_{\alpha n}},\end{aligned}$$

co łącznie z poprzednim oszacowaniem daje $\log(1-\beta_V(P_1))\leqslant t_{\alpha n}\leqslant -nD(P_0||P_1)+n\varepsilon$.

Z drugiej strony na zbiorze B_ε mamy $p_{1n}/p_{0n}=e^{V_n}\geqslant e^{-nD(P_0||P_1)-n\varepsilon}$ i stąd

$$\begin{aligned}1-\beta_V(P_1,n)&\geqslant P_1^n(A^c\cap B_\varepsilon)=\int\limits_{A^c\cap B_\varepsilon}p_{1n}d\lambda^n=\int\limits_{A^c\cap B_\varepsilon}\frac{p_{1n}}{p_{0n}}p_{0n}d\lambda^n\\&=e^{-nD(P_0||P_1)-n\varepsilon}P_0^n(A^c\cap B_\varepsilon)\geqslant e^{-nD(P_0||P_1)-n\varepsilon}(P_0^n(A^c)-P_0^n(B_\varepsilon^c))\\&\geqslant e^{-nD(P_0||P_1)-n\varepsilon}(1-\alpha-o(1)),\end{aligned}$$

co daje $\log(1-\beta_V(P_1))\geqslant -nD(P_0||P_1)-n\varepsilon+\log(1-\alpha-o(1))$. Z dowolności ε i powyższego oszacowania z góry wynika teza lematu. $\square$

Dla zobrazowania wielkości prawdopodobieństwa błędu II rodzaju weźmiemy jako P_0 rozkład jednostajny na [0,1], a jako P_1 rozkład beta z parametrami $p = q = 1.5$. Zatem $p_0(t) \equiv 1$ oraz $p_1(t) = \left(8\sqrt{t(1-t)}\right)/\pi$, $t \in [0,1]$. Wtedy $D(p_0||p_1) = \int_0^1 \log(\pi/\sqrt{64t(1-t)})dt \approx 0.0653$. W tabeli poniżej podane są empiryczne prawdopodobieństwa błędów II rodzaju (w procentach) dla kilku poziomów istotności α i kilku liczebności prób n, a w ostatnim wierszu wartości $e^{-nD(P_0||P_1)}$. Zauważmy, że wartości w ostatnim wierszu nie zależą od poziomu istotności, a rzeczywiste prawdopodobieństwo błędu istotnie od niego zależy.

Prawdopodobieństwa błędów II rodzaju dla testu NP w %

n	10	20	30	40	50	60	70	80	90	100
$\alpha = 0.3$	30	16	9	4.9	2.7	1.5	0.8	0.4	0.25	0.13
$\alpha = 0.25$	37	22	12	6.8	3.7	2.0	1.1	0.7	0.32	0.16
$\alpha = 0.2$	45	28	17	9.4	5.5	3.4	1.9	1.4	0.68	0.43
$\alpha = 0.1$	65	45	30	21.7	14.6	9.5	6.4	4.1	2.62	1.71
$\alpha = 0.05$	79	61	50	36.3	26.3	18.9	13.6	9.2	6.56	4.41
$e^{-nD(P_0\|\|P_1)}$	52	27	14	7.3	3.8	2.0	1.0	0.5	0.28	0.15

Widać, że ostatni wiersz dość dobrze odpowiada prawdopodobieństwom błędów II rodzaju dla poziomu istotności 0.25, dużo większego niż stosowany w praktyce.

Ze względu na zamienność roli P_0 i P_1 oraz poziomu istotności i mocy testu prawdziwa jest dualna wersja lematu Steina.

Twierdzenie 2. Dla dowolnego $\beta \in (0,1)$ niech $\alpha_V(P_0) = \alpha_V(P_0, n)$ będzie poziomem istotności testu NP, którego moc jest równa β. Wtedy

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_V(P_0, n) = -D(P_1||P_0).$$

Wróćmy teraz do dowodu lematu Neymana-Pearsona w ogólniejszej wersji. Niech $\mathcal{P} = \{P_\gamma : \gamma \in \Gamma\}$ będzie pewną (niepustą) rodziną rozkładów na $(\mathcal{X}, \mathcal{A})$, przy czym $P_0 \notin \mathcal{P}$. Rozważamy testowanie hipotezy $H_0 : P = P_0$ przeciwko $H_1 : P \in \mathcal{P}$. Przypuśćmy, że Γ ma strukturę miarową $(\Gamma, \mathcal{B}, \omega)$, gdzie $\mathcal{B}$ jest σ-ciałem podzbiorów Γ, a ω rozkładem a priori na $(\Gamma, \mathcal{B})$. Załóżmy ponadto, że dla pewnej miary dominującej λ mamy $P_0 \ll \lambda$ oraz $P_\gamma \ll \lambda$ dla wszystkich $\gamma \in \Gamma$. Oznaczmy $p_\gamma = dP_\gamma/d\lambda$ oraz $p_{\gamma n}(x) = \prod_{i=1}^n p_\gamma(x_i)$. Rozważmy statystykę

$$V_n^* = \int\limits_{\Gamma} \frac{p_{\gamma n}}{p_{0n}} d\omega(\gamma).$$

Test prawostronny V^* nazywamy optymalnym testem bayesowskim dla rozważanego problemu, gdyż prawdziwe jest następujące twierdzenie.

Twierdzenie 3. Dla dowolnego $t > 0$ niech $\alpha_t^* = P_0^n(V_n^* > t)$, a $\overline{\beta}^*_{\alpha_t^*} = \int_\Gamma P_\gamma^n(V_n^* > t)d\omega(\gamma)$ będzie mocą średnią testu V^* o rozmiarze α_t^*. Wtedy dla dowolnego testu T hipotezy H_0 przeciwko H_1 o rozmiarze $\alpha \in (0,1)$ zachodzi relacja

$$t(\alpha - \alpha_t^*) + (\overline{\beta}^*_{\alpha_t^*} - \overline{\beta}_\alpha) \geqslant 0,$$

gdzie $\overline{\beta}_\alpha = \int_\Gamma P_\gamma^n(T_n > t_{\alpha n})d\omega(\gamma)$ jest mocą średnią testu T. W szczególności dla α i t takich, że $\alpha_t^* = \alpha$ mamy $\overline{\beta}^*_\alpha \geqslant \overline{\beta}_\alpha$, tzn. dla dowolnego ustalonego rozkładu a priori test V^* ma największą moc średnią wśród wszystkich testów o tym samym rozmiarze α_t^*.

Zauważmy, że twierdzenie 1 jest szczególnym przypadkiem twierdzenia 3, gdy wybierzemy rozkład a priori ω skoncentrowany na ustalonym punkcie zbioru Γ.

Dowód. Oznaczmy $A_t^* = \{V_n^* > t\}$ oraz $A = \{T_n > t_{\alpha n}\}$. Wtedy mamy

$$\begin{aligned} &t(\alpha - \alpha_t^*) + (\overline{\beta}^*_{\alpha_t^*} - \overline{\beta}_\alpha) \\ &= t\int\limits_A p_{0n} d\lambda^n - t\int\limits_{A_t^*} p_{0n} d\lambda^n + \int\limits_{A_t^*}\int\limits_\Gamma p_{\gamma n} d\omega(\gamma) d\lambda^n - \int\limits_A\int\limits_\Gamma p_{\gamma n} d\omega(\gamma) d\lambda^n \\ &= \int\limits_{A_t^*}\int\limits_\Gamma \left(\frac{p_{\gamma n}}{p_{0n}} - t\right) p_{0n} d\omega(\gamma) d\lambda^n - \int\limits_A\int\limits_\Gamma \left(\frac{p_{\gamma n}}{p_{0n}} - t\right) p_{0n} d\omega(\gamma) d\lambda^n = \end{aligned}$$

$$\int\limits_{A_t^* \cap A^c} \left(\int\limits_\Gamma \frac{p_{\gamma n}}{p_{0n}} d\omega(\gamma) - t\right) p_{0n} d\lambda^n - \int\limits_{A_t^{*c} \cap A} \left(\int\limits_\Gamma \frac{p_{\gamma n}}{p_{0n}} d\omega(\gamma) - t\right) p_{0n} d\lambda^n \geqslant 0,$$

gdyż, z określenia V_n^*, na zbiorze A_t^* wyrażenie w nawiasie w pierwszej całce jest dodatnie, a w drugiej całce na zbiorze A_t^{*c} niedodatnie. To kończy dowód. □

Dla optymalnego testu bayesowskiego, przy odpowiedniej regularności modelu (szczegółowe sformułowanie pomijamy), prawdziwy jest odpowiednik lematu Steina w postaci $\log(1-\overline{\beta}^*_\alpha) = -D(p_{0n}||f_n) + O(1)$, gdzie $f_n(x) = \int_\Gamma p_{\gamma n}(x) d\omega(\gamma)$ (Clarke i Barron, 1990).

Można zapytać, jak duża jest różnica (strata) między średnią mocą optymalnego testu bayesowskiego i mocą testu NP skonstruowanego dla ustalonego rozkładu $P_1 \in \mathcal{P}$. Aby nieco przybliżyć odpowiedź na to pytanie, rozważmy skończoną rodzinę rozkładów $\mathcal{P} = \{P_1, ..., P_k\}$ o gęstościach $p_j,\ j = 1, ..., k$, takich że dla odpowiednio dobranego c funkcje $c(p_j - p_0)$ tworzą układ ortonormalny względem P_0. Wtedy, mówiąc nie całkiem precyzyjnie, strata mocy w środkowym zakresie wyraża się przez entropię Shannona rozkładu a priori $H(\omega) = -\sum_{j=1}^k w_j \log_2 w_j$, gdzie $w_j = \omega(\{j\})$. Dokładniej, przy mocy testu NP w przedziale $[0.5, 0.9]$ (tj. przy odpowiednio dobranym do liczebności próby współczynniku c), zgrubne oszacowanie tej różnicy wynosi $0.1 H(\omega)$. Przyjmując $w_j = 1/k,\ j = 1, ..., k$, czyli jednostajny rozkład a priori, strata średniej mocy do mocy testu NP wynosi ok. $0.1 \log_2 k$.