

RYSZARD ZIELIŃSKI

Siedem wykładów
wprowadzających do
statystyki
matematycznej

Zadania zweryfikowała
oraz wskazówkami i rozwiązaniami uzupełniła
Agata Boratyńska

WARSZAWA 2004

Siedem wykładów
wprowadzających do
statystyki
matematycznej

www.impan.gov.pl/~rziel/7ALL.pdf

PRZEDMOWA

Książka jest przeznaczona dla matematyków, ale napisana jest w aspekcie zastosowań.

Fakt, że jest przeznaczona dla matematyków wyraża się w sposobie prowadzenia wykładu i w założonym poziomie wiedzy matematycznej Czytelnika. Zakładam mianowicie, że Czytelnik pamięta analizę, algebrę liniową, podstawy teorii funkcji i teorii prawdopodobieństwa mniej więcej w zakresie pierwszych trzech lat studiów uniwersyteckich. W zakresie statystyki matematycznej książka jest w pełni samowystarczalna.

Fakt, że jest napisana w aspekcie zastosowań wyraża się w głębszej prezentacji pozamatematycznych motywacji rozważanych zagadnień statystycznych. Z dużym naciskiem staram się wyjaśnić przede wszystkim to, "o co chodzi", wierząc, że z potrzebną techniką matematyk potrafi sobie poradzić. Służą temu zarówno liczne przykłady, jak i sposób prowadzenia wykładu. Na przykład w wykładzie o weryfikacji hipotez statystycznych staram się dokładnie przedstawić logiczne podstawy testów. Demonstruję więc przede wszystkim testy istotności, dopiero później wprowadzam porządek w rodzinie testów ("test mocniejszy") i w tym dopiero kontekście prezentuję teorię Neymana–Pearsona. Pomijam pasjonujące czystych matematyków teorie asymptotyczne.

Książka w znacznym stopniu opiera się na notatkach z wykładów, jakie prowadziłem na kierunku zastosowań na Wydziale Matematyki Informatyki i Mechaniki Uniwersytetu Warszawskiego dla studentów czwartego roku. Zarówno sam wykład (prowadzony w wymiarze 30 godzin wykładu i 30 godzin ćwiczeń), jak i ta książka jest w dosłownym sensie wprowadzeniem w problematykę statystyki matematycznej i w żadnym wypadku nie może zastąpić bardziej kompletnej monografii takich jak Bartoszewicza(1989), Lehmana(1983,1986) lub Barry(1982).

Ryszard Zieliński

Warszawa, w listopadzie 1990

=====
Nowa wersja została zamknięta 10 maja 2004 roku. W stosunku do wersji książkowej PWN 1990, została wzbogacona o wskazówki i rozwiązania do trudniejszych zadań. Zostały również poprawione drobne pomyłki redakcyjne i typograficzne. Książka jest dostępna pod adresem www.impan.gov.pl/~rziel/7ALL.pdf. Kolejna (niniejsza) wersja została zamknięta 2 listopada 2009 roku: dodano zadanie V.9.

Wykład I

MODEL STATYSTYCZNY

1. Przykłady wprowadzające

Najogólniej mówiąc, statystyką nazywamy kolekcjonowanie danych liczbowych i wnioskowanie z nich. Można wyróżnić dwa rodzaje sytuacji, w których zajmujemy się statystyką.

1. Nie mamy żadnej wiedzy *a priori* o przedmiocie naszych zainteresowań i na podstawie zbieranych danych chcemy dopiero zorientować się w problemie oraz sformułować jakieś wstępne teorie o badanych zjawiskach. Z takimi sytuacjami mamy do czynienia na przykład w przypadku zupełnie nowych i zaskakujących wykopalisk archeologicznych, w przypadku danych z wypraw kosmicznych odbywanych z zamiarem odkrywania nowych światów, ale również np. w przypadku nowej choroby, takiej jak AIDS, kiedy to kolekcjonowanie wszelkich dostępnych informacji o środowiskach, z których pochodzą chorzy, jak również o samym przebiegu choroby, ma posłużyć do sformułowania wstępnych teorii na temat mechanizmów jej powstawania i rozprzestrzeniania się. Tym działem statystyki, zwanym *statystyczną analizą danych*, nie będziemy się zajmowali.

2. Wiedza *a priori* o przedmiocie badań jest już sformułowana w postaci pewnych teorii lub hipotez i zadanie statystyki polega na tym, żeby na podstawie nowych, odpowiednio zbieranych danych uzupełnić tę teorię lub zweryfikować odpowiednie hipotezy. Na przykład istnieje teoria, według której gdzieś między Merkurym a Słońcem krąży jakaś planeta i potrzebne są dane, które tę teorię zdyskwalifikują lub wyznaczają miejsce i czas, gdzie tę planetę można zaobserwować.

Statystyka matematyczna zajmuje się metodami kolekcjonowania danych i wnioskowania z nich, gdy wiedza *a priori* jest sformułowana w postaci pewnych *modeli probabilistycznych*. Ta probabilistyka może w naturalny sposób tkwić w samych badanych zjawiskach, ale może być również wprowadzana przez badacza.

Oto dwa przykłady.

Przykład 1. *Przedmiotem badania jest zbiór składający się z, powiedzmy, N elementów i zawierający pewną liczbę, powiedzmy M , elementów wyróżnionych. Interesuje nas przypadek, gdy N jest ustalone i znane, M nie jest znane i chcemy dowiedzieć się coś na temat "jak duże jest M ".*

Przykłady z życia to pytanie o liczbę ludzi w Polsce, którzy spędzają przed telewizorem co najmniej 10 godzin tygodniowo, pytanie o liczbę sztuk wadliwych w dużej partii produktów, itp.

Jeżeli N jest na tyle duże, że obejrzenie wszystkich elementów i zliczenie liczby elementów wyróżnionych nie jest możliwe lub nie jest opłacalne, można postąpić (i tak właśnie się często robi) w następujący sposób.

Z badanego N -elementowego zbioru losujemy n -elementowy podzbiór. Jeżeli wykonujemy to w taki sposób, że każdy n -elementowy podzbiór może być wylosowany z takim samym prawdopodobieństwem, to prawdopodobieństwo, że w wybranym n -elementowym podzbiorku znajdzie się x elementów wyróżnionych wynosi

$$(1) \quad p(x; N, M, n) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}.$$

Oznaczmy przez X liczbę wyróżnionych elementów w wylosowanym podzbiorku. Jest to zmienna losowa o rozkładzie hipergeometrycznym:

$$P\{X = x\} = p(x; N, M, n), \quad x = 0, 1, \dots, \min\{n, M\}.$$

Zadanie polega na tym, żeby na podstawie obserwacji zmiennej losowej X odpowiedzieć na interesujące nas pytania dotyczące nieznannej liczby M , na przykład "ile wynosi M " albo "czy $M > M_0$ " dla pewnej ustalonej liczby M_0 , lub tp. ■

Będziemy posługiwali się następującą, historycznie ukształtowaną terminologią: badany zbiór będziemy nazywali *populacją*, a losowany podzbiór — *próbą* lub *próbą losową*. Używany jest również termin *próbka*.

Przykład 2. Dokonujemy pomiaru pewnej nieznannej wielkości μ (np. długości, masy, wydajności procesu technologicznego). Pomiar zwykle jest obciążony pewnym błędem — oznaczmy ten błąd przez ϵ — tak, że wynikiem pomiaru jest $X = \mu + \epsilon$. Na podstawie wyniku pomiaru X lub na podstawie serii takich pomiarów $X_i = \mu + \epsilon_i$, $i = 1, 2, \dots, n$, mamy udzielić odpowiednich informacji o nieznannej wielkości μ .

Jeżeli przyjmujemy, że błąd ϵ jest wielkością losową, wchodzimy w dziedzinę statystyki matematycznej. Różne, i coraz bardziej szczegółowe, założenia o probabilistycznej naturze zmiennej losowej ϵ prowadzą do różnych, coraz węższych, statystycznych modeli pomiaru. Zwykle zakłada się, że ϵ jest zmienną losową, której rozkład nie zależy od μ . O samym rozkładzie zakłada się, że jest rozkładem symetrycznym względem zera. Jeżeli wykonuje się serię pomiarów X_1, X_2, \dots, X_n , to najczęściej zakłada się, że $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie. W metrologii uzasadnia się, że za ten rozkład można przyjąć pewien rozkład normalny $N(0, \sigma^2)$ o wariancji σ^2 , której wielkość jest związana z klasą dokładności przyrządu pomiarowego; wtedy gęstość łącznego rozkładu pomiarów X_1, X_2, \dots, X_n wyraża się wzorem

$$(2) \quad f_{\mu, \sigma}(x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2\right\}.$$

Jak w przykładzie 1, na podstawie obserwacji (wektorowej) zmiennej losowej $\mathbf{X} = (X_1, X_2, \dots, X_n)$ o rozkładzie z gęstością (2) należy sformułować pewne wnioski o nieznannej wartości parametru μ tego rozkładu. ■

2. Model statystyczny

Punktem wyjścia w naszych rozważaniach będzie zawsze pewien element losowy X (zmienna losowa, skończony lub nieskończony ciąg zmiennych losowych); będziemy często o nim mówili: *wynik eksperymentu*, *wynik pomiaru*, *wynik obserwacji* lub po prostu *obserwacja*. Zbiór wartości elementu losowego X będziemy oznaczali przez \mathcal{X} i nazywaliśmy *przestrzenią próby*. We wszystkich naszych wykładach \mathcal{X} będzie zbiorem skończonym lub przeliczalnym, albo pewnym obszarem w skończenie wymiarowej przestrzeni \mathbf{R}^n . Niech $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ będzie rodziną rozkładów prawdopodobieństwa na przestrzeni prób \mathcal{X} , indeksowaną pewnym parametrem θ przebiegającym zbiór Θ . Dokładniej, \mathcal{P} jest rodziną rozkładów prawdopodobieństwa na odpowiednim σ -ciele zdarzeń losowych w \mathcal{X} , ale wobec naszego ograniczenia się do wyżej wymienionych przypadków będzie to zawsze albo σ -ciało wszystkich podzbiorów, albo σ -ciało podzbiorów borelowskich, więc nie będziemy tego specjalnie podkreślali.

Przestrzeń próby wraz z rodziną rozkładów \mathcal{P} , tzn. obiekt

$$(\mathcal{X}, \{P_\theta : \theta \in \Theta\}),$$

nazywamy *modelem statystycznym* (używa się również nazwy *przestrzeń statystyczna*). Odwzorowania z \mathcal{X} w \mathbf{R}^k nazywamy *statystykami* lub, jeżeli zależy nam na takim podkreśleniu, *k-wymiarowymi statystykami*.

Jeżeli $X = (X_1, X_2, \dots, X_n)$, przy czym X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie, to będziemy stosowali również oznaczenie

$$(\mathcal{X}, \{P_\theta : \theta \in \Theta\})^n,$$

w którym \mathcal{X} jest zbiorem wartości zmiennej losowej X_1 (a więc każdej ze zmiennych X_1, X_2, \dots, X_n) oraz P_θ jest rozkładem tej zmiennej losowej. Używa się wtedy również terminologii: X_1, X_2, \dots, X_n jest *próbą z rozkładu P_θ* lub *próbą z populacji P_θ* dla pewnego $\theta \in \Theta$.

Będziemy zawsze zakładali, że jeżeli $\theta_1 \neq \theta_2$, to $P_{\theta_1} \neq P_{\theta_2}$ (o takich modelach mówimy, że są identyfikowalne: znając rozkład P_θ , znamy wartość parametru θ). Wprowadzenie parametru θ do rozważań ułatwia sformułowania wielu problemów, a dopóki nie wprowadzamy ograniczeń na zbiór Θ , odbywa się to bez straty ogólności rozważań, bo każdą rodzinę \mathcal{P} rozkładów prawdopodobieństwa możemy "sparametryzować", przyjmując za parametr θ rozkładu P sam rozkład P .

Przykład 1 (cd.). W przykładzie 1 ustalonymi i znanymi wielkościami są liczność populacji N i liczność próby n . Nieznany parametrem jest $M \in \{0, 1, \dots, N\}$. Przestrzenią próby jest zbiór $\{0, 1, 2, \dots, n\}$. Rodziną rozkładów prawdopodobieństwa na przestrzeni próby jest rodzina rozkładów hipergeometrycznych (1) indeksowana parametrem M . O wyniku obserwacji, tzn. o zmiennej losowej X wiemy, że ma pewien rozkład z tej rodziny, ale nie wiemy który z nich. ■

Przykład 2 (cd.). W przykładzie 2 mamy do czynienia z modelem statystycznym

$$(\mathbf{R}^1, \{f_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]: \mu \in \mathbf{R}^1, \sigma > 0\})^n,$$

tzn. z modelem

$$(\mathbf{R}^n, \{f_{\mu, \sigma}(x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} \exp[-\frac{1}{2} \sum_{i=1}^n (\frac{x_i - \mu}{\sigma})^2]: \mu \in \mathbf{R}^1, \sigma > 0\}).$$

W rozważanej sytuacji wiemy, że zmienna losowa X ma pewien rozkład z rodziny $\{f_{\mu, \sigma}(x): \mu \in \mathbf{R}^1, \sigma > 0\}$, ale nie wiemy, który z nich. Zadanie polega na tym, żeby na podstawie obserwacji X_1, X_2, \dots, X_n sformułować odpowiednie wnioski o tym nieznanym rozkładzie (”zidentyfikować” ten rozkład). ■

3. Podstawowe problemy statystyki matematycznej

Dany jest model statystyczny $(\mathcal{X}, \{P_\theta: \theta \in \Theta\})$ i obserwacja X o wartościach w \mathcal{X} ; o tej obserwacji wiadomo, że jest zmienną losową o pewnym rozkładzie P_θ . Najogólniej mówiąc, zadanie polega na tym, żeby na podstawie obserwacji X odpowiedzieć na pewne pytania na temat nieznanego θ .

Jeżeli pytanie brzmi po prostu ”ile wynosi θ ?”, mówimy o problemie *estymacji*. Formalnie: zadanie polega wtedy na skonstruowaniu takiego odwzorowania $\hat{\theta}: \mathcal{X} \rightarrow \Theta$, żeby wielkość $\hat{\theta}(X)$ można było traktować jako ”dobre przybliżenie” nieznannej wartości θ . Wszystko, czym zajmuje się teoria estymacji, zależy od tego, co rozumiemy przez ”dobre przybliżenie”. Można sobie na przykład wyobrazić, że w Θ jest określona odległość d i że chcemy znaleźć taką funkcję $\theta^*: \mathcal{X} \rightarrow \Theta$, żeby $E_\theta d(\theta^*(X), \theta) \leq E_\theta d(\hat{\theta}(X), \theta)$ dla wszystkich odwzorowań $\hat{\theta}: \mathcal{X} \rightarrow \Theta$, jednostajnie względem $\theta \in \Theta$. Takie *optymalne estymatory* θ^* rzadko udaje się skonstruować; pewne podejście przedstawimy szczegółowo w trzecim wykładzie. Częściej postępuje się w ten sposób, że na drodze różnych rozważań heurystycznych dochodzi się do wniosku, iż odpowiednim estymatorem będzie, powiedzmy, $\tilde{\theta}: \mathcal{X} \rightarrow \Theta$; zadanie statystyki matematycznej polega wtedy na tym, żeby zbadać własności tego estymatora i sformułować wnioski na temat jego dokładności. Dwie najbardziej znane metody tego typu, a mianowicie metodę opartą na koncepcji *wiarogodności* i *metodę najmniejszych kwadratów*, przedstawimy w wykładach piątym i szóstym.

Problem estymacji formułuje się czasami w inny sposób (ilustrujemy to dla przypadku $\Theta = \mathbf{R}^1$): skonstruować takie dwie funkcje $\underline{\theta}(X)$ i $\bar{\theta}(X)$, żeby z zadanym z góry, bliskim jedności, prawdopodobieństwem γ , zachodziło $P_\theta\{\underline{\theta}(X) \leq \theta \leq \bar{\theta}(X)\} \geq \gamma$ dla każdego $\theta \in \Theta$. W takiej sytuacji mówimy o *estymacji przedziałowej* (w odróżnieniu od *estymacji punktowej*, o której była mowa wyżej), a przedział $(\underline{\theta}(X), \bar{\theta}(X))$ — jest to oczywiście przedział losowy — nazywa się *przedziałem ufności* na poziomie ufności γ . Czasami

postuluje się ponadto, żeby różnica $\bar{\theta}(X) - \theta(X)$ nie przekraczała pewnej z góry zadanej wielkości: mówimy wtedy o *estymacji zadaną precyzją*. Całej tej problematyki w naszych wykładach nie będziemy rozwijali.

Inne problemy statystyki matematycznej są związane z następującym postawieniem zagadnienia: w przestrzeni Θ wyróżniony jest pewien podzbiór Θ_0 i pytamy, czy $\theta \in \Theta_0$. W takiej sytuacji zdanie " $\theta \in \Theta_0$ " nazywa się *hipotezą statystyczną*, a cała problematyka nosi nazwę *teorii weryfikacji hipotez statystycznych*. Typowy "przykład z życia" to porównywanie dwóch leków i pytanie, czy jeden z nich jest skuteczniejszy od drugiego. Tym zagadnieniom poświęcamy wykład czwarty.

Dwa wymienione wyżej działy: *teoria estymacji* i *teoria weryfikacji hipotez statystycznych*, składają się na *klasyczną statystykę matematyczną*. W naszych wykładach nie wychodzimy (z wyjątkiem wykładu siódmego) poza ten przedmiot. Wykład siódmy jest poświęcony *teorii decyzji statystycznych*. Jest to bardzo duży rozdział współczesnej statystyki matematycznej i jej praktycznych zastosowań.

4. Podstawowe twierdzenie statystyki matematycznej

Dany jest model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$ obserwacji X . Jak to już mówiliśmy, zadanie polega na tym, żeby na podstawie obserwacji X w jakimś sensie odtworzyć nieznaną rozkład P_θ , z którego pochodzi ta obserwacja. Jak to jest w ogóle możliwe?

Niech naszą obserwacją będzie próba losowa X_1, X_2, \dots, X_n — ciąg niezależnych (rzeczywistych) zmiennych losowych o jednakowym rozkładzie z dystrybuantą F . Niech $F_n(t)$, $t \in \mathbf{R}^1$, będzie *dystrybuantą empiryczną* z próby X_1, X_2, \dots, X_n , tzn. niech

$$(3) \quad F_n(t) = \frac{\#\{1 \leq j \leq n: X_j \leq t\}}{n} .$$

Następujące trzy lematy i twierdzenie wyjaśniają, w jakim sensie próba X_1, X_2, \dots, X_n odtwarza rozkład, z którego pochodzi.

Dla danej funkcji $\psi(X_1, \dots, X_n)$ obserwacji X_1, X_2, \dots, X_n z rozkładu prawdopodobieństwa o dystrybuancie F , niech $E_F \psi(X_1, \dots, X_n)$ oznacza wartość oczekiwaną tej funkcji.

LEMAT 1. Dla każdego $t \in \mathbf{R}^1$ mamy $E_F F_n(t) = F(t)$. □

LEMAT 2. Dla każdego $t \in \mathbf{R}^1$ mamy $P_F \{\lim_{n \rightarrow \infty} F_n(t) = F(t)\} = 1$. □

LEMAT 3. Jeżeli próba losowa X_1, X_2, \dots, X_n pochodzi z rozkładu o dystrybuancie F , to dla każdego $t \in \mathbf{R}^1$ rozkład zmiennej losowej

$$\sqrt{n} \frac{F_n(t) - F(t)}{\sqrt{F(t)[1 - F(t)]}}$$

dąży do rozkładu normalnego $N(0, 1)$, gdy $n \rightarrow \infty$. □

Zauważmy, że lemat 2 i lemat 3 formułują po prostu mocne prawo wielkich liczb i centralne twierdzenie graniczne dla schematu Bernoulliego.

TWIERDZENIE 1 (podstawowe twierdzenie statystyki matematycznej). *Niech*

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|.$$

Jeżeli próba X_1, X_2, \dots, X_n pochodzi z rozkładu o dystrybuancie F , to $D_n \rightarrow 0$ z prawdopodobieństwem 1, gdy $n \rightarrow \infty$.

D o w ó d. Niech próba X_1, X_2, \dots, X_n pochodzi z rozkładu o dystrybuancie F . Mówiąc dalej "z prawdopodobieństwem 1" lub krótko "z p.1", mamy na myśli rozkład prawdopodobieństwa o dystrybuancie F .

Ustalmy dowolnie liczbę naturalną M . Dla $k = 1, 2, \dots, M - 1$ niech

$$x_{k,M} = \inf\{x: F(x-0) \leq \frac{k}{M} \leq F(x)\}.$$

Wtedy

$$(-\infty, x_{1,M}), [x_{1,M}, x_{2,M}), \dots, [x_{M-1,M}, +\infty)$$

jest rozbiem prostej \mathbf{R}^1 .

Oznaczając $x_{0,M} = -\infty$ oraz $x_{M,M} = +\infty$ i uwzględniając to, że $F_n(x_{0,M}) = F(x_{0,M}) = 0$ oraz $F_n(x_{M,M} - 0) = F(x_{M,M} - 0) = 1$, dla x należącego do k -tego ($k = 0, 1, \dots, M - 1$) przedziału rozbitcia mamy

$$F_n(x_{k,M}) \leq F_n(x) \leq F_n(x_{k+1,M} - 0),$$

$$F(x_{k,M}) \leq F(x) \leq F(x_{k+1,M} - 0),$$

przy czym

$$0 \leq F(x_{k+1,M} - 0) - F(x_{k,M}) \leq \frac{1}{M}.$$

Zatem

$$F_n(x) - F(x) \leq F_n(x_{k+1,M} - 0) - F(x_{k,M}) \leq F_n(x_{k+1,M} - 0) - F(x_{k+1,M} - 0) + \frac{1}{M}$$

oraz

$$F_n(x) - F(x) \geq F_n(x_{k,M}) - F(x_{k+1,M} - 0) \geq F_n(x_{k,M}) - F(x_{k,M}) - \frac{1}{M},$$

czyli

$$(4) \quad |F_n(x) - F(x)| \leq \max\{|F_n(x_{k,M}) - F(x_{k,M})|, |F_n(x_{k+1,M} - 0) - F(x_{k+1,M} - 0)|\} + \frac{1}{M}.$$

Oznaczając

$$\Delta_{M,n}^{(1)} = \max_{0 \leq k \leq M-1} |F_n(x_{k,M}) - F(x_{k,M})|,$$

$$\Delta_{M,n}^{(2)} = \max_{0 \leq k \leq M-1} |F_n(x_{k+1,M} - 0) - F(x_{k+1,M} - 0)|,$$

otrzymujemy oszacowanie

$$(5) \quad D_n \leq \max\{\Delta_{M,n}^{(1)}, \Delta_{M,n}^{(2)}\} + \frac{1}{M}.$$

Na mocy lematu 2, dla każdego k mamy z prawdopodobieństwem 1

$$F_n(x_{k,M}) - F(x_{k,M}) \rightarrow 0,$$

$$F_n(x_{k+1,M} - 0) - F(x_{k+1,M} - 0) \rightarrow 0,$$

więc (skończona liczba różnych k) również $\Delta_{M,n}^{(1)} \rightarrow 0$ oraz $\Delta_{M,n}^{(2)} \rightarrow 0$ z p.1, czyli także

$$\max\{\Delta_{M,n}^{(1)}, \Delta_{M,n}^{(2)}\} \rightarrow 0 \quad \text{z p.1.}$$

Zatem

$$\limsup_{n \rightarrow \infty} D_n \leq \frac{1}{M} \quad \text{z p.1.}$$

Ponieważ M jest dowolną liczbą naturalną, otrzymujemy tezę twierdzenia. \square

Powyższe twierdzenie 1 jest znane w literaturze również jako *twierdzenie Gliwienki-Cantelliego*.

5. Zadania

1. Wykonujemy n doświadczeń losowych, z których każde kończy się sukcesem z prawdopodobieństwem θ . Wiadomo, że $\theta \in [\theta_1, \theta_2]$, gdzie $\theta_1, \theta_2 \in [0, 1]$ są ustalone. Sformułować model statystyczny tego eksperymentu.

2. Pewne urządzenie techniczne pracuje dopóty, dopóki nie uszkodzi się któryś z k elementów typu A lub któryś z l elementów typu B . Czas życia elementów typu A jest zmienną losową o rozkładzie wykładniczym z gęstością $f_\alpha(x) = \alpha^{-1} \exp(-x/\alpha)$, a czas życia elementów typu B jest zmienną losową o rozkładzie wykładniczym z gęstością $f_\beta(x) = \beta^{-1} \exp(-x/\beta)$ i wszystkie te zmienne losowe są niezależne. Obserwuje się czas życia T całego urządzenia. Sformułować model statystyczny tej obserwacji. Jak wygląda przestrzeń statystyczna w tym zadaniu gdy nie zakłada się niezależności czasów bezawaryjnej pracy poszczególnych elementów?

3. Wykonujemy ciąg niezależnych doświadczeń, z których każde kończy się sukcesem z nieznanym prawdopodobieństwem θ lub porażką z prawdopodobieństwem $1 - \theta$. Doświadczenia wykonujemy dopóty, dopóki nie uzyskamy m sukcesów. Sformułować model statystyczny przy założeniu, że wyniki poszczególnych eksperymentów są niezależnymi zmiennymi losowymi.

4. Przeprowadza się $n = \sum_{j=1}^k n_j$ eksperymentów w taki sposób, że n_j eksperymentów wykonuje się na poziomie x_j , $j = 1, 2, \dots, k$. Prawdopodobieństwo sukcesu w eksperymencie przeprowadzanym na poziomie x jest równe

$$p(x) = \frac{1}{1 + e^{-(\alpha + \beta x)}}, \quad \alpha \in \mathbf{R}^1, \beta > 0,$$

gdzie (α, β) jest nieznanym parametrem. Sformułować model statystyczny tego eksperymentu.

Następujące zadania przypominają te fakty z teorii prawdopodobieństwa, z których będziemy korzystali w dalszych wykładach. W celu łatwiejszego powoływania się na nie, formułujemy je w postaci zadań. Krótką tabelkę podstawowych rozkładów prawdopodobieństwa, o których mówimy w naszych wykładach, podajemy na końcu książki.

5. Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $\Gamma(\alpha, \lambda)$, to $\sum_{i=1}^n X_i$ ma rozkład $\Gamma(n\alpha, \lambda)$.

6. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi o rozkładzie wykładniczym $E(\theta, \beta)$ i niech

$$Y_1 = nX_{1:n}, \quad Y_j = (n - j + 1)(X_{j:n} - X_{j-1:n}), \quad j = 2, 3, \dots, n.$$

Wykazać, że zmienne losowe Y_1, Y_2, \dots, Y_n są niezależne i wyznaczyć ich rozkład. Wykazać, że zmienne losowe $X_{1:n}$ oraz $\sum_{j=1}^n (X_j - X_{1:n})$ są niezależne i wyznaczyć ich rozkład.

7. Jeżeli zmienna losowa X ma rozkład $N(0, \sigma^2)$, to zmienna losowa X^2 ma rozkład $\Gamma(\frac{1}{2}, 2\sigma^2)$. (Rozkład $\Gamma(\frac{n}{2}, 2)$ nazywa się rozkładem chi-kwadrat o n stopniach swobody).

8. Mówimy, że wektor losowy lub punkt losowy \mathbf{X} w \mathbf{R}^n ma n -wymiarowy rozkład normalny i piszemy $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{C})$, jeżeli gęstość rozkładu prawdopodobieństwa tego wektora (istnieje i) wyraża się wzorem

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{C}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\},$$

gdzie $\boldsymbol{\mu} = E\mathbf{X}$ jest pewnym wektorem oraz $\mathbf{C} = \text{Var}\mathbf{X}$ jest macierzą dodatnio określoną.

Niech $\mathbf{Y} = \mathbf{A}(\mathbf{X} - \boldsymbol{\mu})$, gdzie \mathbf{A} jest pewną macierzą nieosobliwą.

Niech $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ oraz $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$.

Sprawdzić, że

(a) Jeżeli $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{C})$, to $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{B})$. Wyznaczyć macierz \mathbf{B} .

(b) Jeżeli macierz \mathbf{A} jest ortonormalna oraz $\boldsymbol{\mu} = \mathbf{0}$, to $\sum_{j=1}^n X_j^2 = \sum_{j=1}^n Y_j^2$.

(c) Jeżeli ponadto X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, \sigma^2)$, to również Y_1, Y_2, \dots, Y_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, \sigma^2)$.

9. Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, 1)$, to $\sum_{i=1}^n X_i^2$ ma rozkład $\Gamma(\frac{n}{2}, 2)$.

10. Sprawdzić, że macierz $\mathbf{W} = (w_{i,j})_{i,j=1,2,\dots,n}$, określona wzorami

$$w_{1,j} = \frac{1}{\sqrt{n}}, \quad j = 1, 2, \dots, n,$$

$$w_{i,j} = \frac{1}{\sqrt{i(i-1)}}, \quad i = 2, 3, \dots, n; \quad j < i,$$

$$w_{i,i} = -\sqrt{\frac{i-1}{i}}, \quad i = 2, 3, \dots, n,$$

$$w_{i,j} = 0, \quad j > i,$$

jest macierzą ortonormalną (przekształcenie Helmerta).

Niech $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$, $\bar{X} = \sum_{i=1}^n X_i/n$ oraz $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$.

Wykazać, że

- (i) jeżeli $\mathbf{Y} = \mathbf{W}\mathbf{X}$, to $Y_1 = \sqrt{n}\bar{X}$, $Y_2^2 + Y_3^2 + \dots + Y_n^2 = S^2$;
 (ii) jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(\mu, \sigma^2)$, to \bar{X} i S^2 są niezależnymi zmiennymi losowymi.

11. Niech \mathbf{X} będzie n -wymiarową zmienną losową o rozkładzie normalnym $N(\mathbf{0}, \mathbf{I})$. Niech \mathbf{P} będzie symetryczną macierzą idempotentną rzędu $r < n$. Wykazać, że $\mathbf{X}^T \mathbf{P} \mathbf{X}$ oraz $\mathbf{X}^T (\mathbf{I} - \mathbf{P}) \mathbf{X}$ są niezależnymi zmiennymi losowymi o rozkładach chi-kwadrat.

Ogólniej, niech $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_k$ będą takimi symetrycznymi macierzami idempotentnymi, że $\mathbf{P}_1 + \mathbf{P}_2 + \dots + \mathbf{P}_k = \mathbf{I}$. Wykazać, że zmienne losowe $\mathbf{X}^T \mathbf{P}_i \mathbf{X}$, $i = 1, 2, \dots, k$, są niezależnymi zmiennymi losowymi o rozkładach chi-kwadrat.

12. Jeżeli zmienna losowa X ma rozkład normalny $N(0, 1)$, zmienna losowa Y ma rozkład chi-kwadrat o n stopniach swobody i te zmienne losowe są niezależne, to rozkład zmiennej losowej $t = \frac{X}{\sqrt{Y/n}}$ nazywa się rozkładem t Studenta o n stopniach swobody.

Wyznaczyć gęstość prawdopodobieństwa tego rozkładu i naszkicować jej wykres dla kilku różnych wartości naturalnych n .

13. Jeżeli zmienna losowa X ma rozkład chi-kwadrat o n stopniach swobody, zmienna losowa Y ma rozkład chi-kwadrat o m stopniach swobody i te zmienne losowe są niezależne, to rozkład zmiennej losowej $F = \frac{X/n}{Y/m}$ nazywa się rozkładem F (lub rozkładem F Snedecora). Wyznaczyć gęstość prawdopodobieństwa tego rozkładu i naszkicować jej wykres dla kilku różnych wartości naturalnych n i m .

Wykład II

STATYSTYKI DOSTATECZNE

1. Preliminaria

W całym wykładzie będziemy często w istotny sposób korzystali z pojęcia *rozkładu warunkowego* i *warunkowej wartości oczekiwanej*. Nie będziemy wykorzystywali tych pojęć w ich pełnej ogólności: przedstawimy tu dokładnie tylko to, co nam będzie dalej potrzebne.

W bieżącym paragrafie rozważamy przestrzeń probabilistyczną (Ω, \mathcal{F}, P) i zmienne losowe X, Y, Z, \dots , określone na tej przestrzeni.

Niech najpierw X i Y będą dyskretnymi zmiennymi losowymi, to znaczy niech $X(\Omega) = \{x_1, x_2, \dots\}$ oraz $Y(\Omega) = \{y_1, y_2, \dots\}$. Zakładamy, że $P\{Y = y_j\} > 0$ dla każdego $j = 1, 2, \dots, i$ (jak w elementarnym rachunku prawdopodobieństwa) definiujemy *warunkowy rozkład zmiennej losowej X , gdy $Y = y_j$* wzorem

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}}, \quad i = 1, 2, \dots$$

Wielkość

$$E(X | Y = y_j) = \sum_i x_i P\{X = x_i | Y = y_j\}$$

nazywamy *warunkową wartością oczekiwaną zmiennej losowej X , gdy $Y = y_j$* .

Niech teraz X i Y będą zmiennymi losowymi "typu ciągłego" na (Ω, \mathcal{F}, P) , tzn. takimi zmiennymi losowymi, których rozkłady mają gęstości względem miary Lebesgue'a. Oznaczmy te gęstości przez $f_{X,Y}(x, y)$ — gęstość łącznego rozkładu zmiennych losowych X i Y oraz $f_X(x), f_Y(y)$ — gęstości rozkładów brzegowych zmiennych losowych X i Y . Zakładamy, że $f_Y(y) > 0$. Mamy wtedy

$$P\{X \leq x, Y \leq y\} = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) dt ds,$$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, t) dt,$$

$$P\{X \leq x\} = \int_{-\infty}^x f_X(s) ds, \quad \text{itp.}$$

Definiujemy *rozkład warunkowy zmiennej losowej X , gdy $Y = y$* , poprzez jego gęstość

$$f_{X|y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Wielkość

$$E(X|Y = y) = \int_{-\infty}^{+\infty} x f_{X|y}(x) dx$$

nazywamy *warunkową wartością oczekiwaną zmiennej losowej X , gdy $Y = y$* .

Dalej będziemy stosowali jednolite oznaczenia $f_{X,Y}(x,y)$, $f_X(x)$, $f_Y(y)$, rozumiejąc, że w przypadku rozkładów dyskretnych chodzi tu o gęstość względem miary liczącej.

Zauważmy, że $E(X|Y = y)$ jest pewną funkcją argumentu y . W wielu zastosowaniach wygodniej jest rozważać warunkową wartość oczekiwaną zmiennej losowej X "pod warunkiem zmiennej losowej Y " jako funkcję na Ω (tzn. jako zmienną losową); tę funkcję oznaczamy przez $E(X|Y)$ i definiujemy wzorem

$$(1) \quad E(X|Y)(\omega) = E(X|Y = y), \quad \text{gdy } Y(\omega) = y.$$

W szczególności, prawdopodobieństwo warunkowe zdarzenia $\{X \in A\}$ "pod warunkiem zmiennej losowej Y " traktujemy, przy ustalonym A , jako zmienną losową $E(\mathbf{1}_A|Y)$ i oznaczamy przez $P\{X \in A|Y\}$. Mamy więc

$$P\{X \in A|Y\}(\omega) = \int_A f_{X|y}(t) dt, \quad \text{gdy } Y(\omega) = y.$$

Zwracamy tu uwagę na pewien dualizm pojęcia warunkowego rozkładu zmiennej losowej X , mianowicie przy ustalonym zbiorze A wielkość $P\{X \in A|Y\}$ jest zmienną losową na (Ω, \mathcal{F}) , natomiast przy ustalonym $y \in Y(\Omega)$ funkcja $P\{\cdot | Y = y\}$ jest rozkładem prawdopodobieństwa zmiennej losowej X .

Istotne jest, żebyśmy zdawali sobie sprawę ze struktury zmiennej losowej $E(X|Y)$, a w szczególności zmiennej losowej $P\{X \in A|Y\}$. Zmienna losowa Y — z samej definicji zmiennej losowej — jest funkcją rzeczywistą na Ω , mierzalną względem σ -ciała \mathcal{F} . Niech \mathcal{B} będzie σ -ciałem zbiorów borelowskich na prostej i niech

$$\sigma(Y) = \{Y^{-1}(B) : B \in \mathcal{B}\}$$

będzie σ -ciałem generowanym przez zmienną losową Y . Otóż $E(X|Y)$ jest zmienną losową na (Ω, \mathcal{F}) , mierzalną względem σ -ciała $\sigma(Y)$. W szczególności, $E(X|Y)$ jest stałą na warstwicach funkcji Y , tzn. na zbiorach $\{\omega : Y(\omega) = y\}$, $y \in \mathbf{R}^1$. Jeżeli dwie różne zmienne losowe Y i Z generują takie same σ -ciała, tzn. jeżeli $\sigma(Y) = \sigma(Z)$, to oczywiście $E(X|Y) = E(X|Z)$. Możemy więc wspinać się na jeszcze jeden szczebel abstrakcji i rozpatrywać pod- σ -ciało \mathcal{A} σ -ciała \mathcal{F} i *warunkową wartość oczekiwaną zmiennej losowej X względem σ -ciała \mathcal{A}* . Piszemy wtedy $E(X|\mathcal{A})$. Będziemy dalej korzystali z intuicji z tym związanej i czasami z tych ogólniejszych oznaczeń, ale nie będziemy rozwijali tego zagadnienia w pełnej ogólności, gdyż pojawiają się tu trudności związane z tym, że nie dla każdego σ -ciała istnieje zmienna losowa generująca to σ -ciało. Zmienne losowe Y i Z generujące to samo σ -ciało będziemy nazywali *równoważnymi zmiennymi losowymi*. Oczywiście zmienne losowe Y i Z są równoważne, jeżeli istnieją takie funkcje g i h , że $Y = g(Z)$ oraz $Z = h(Y)$.

Odnotujmy następujące własności warunkowych wartości oczekiwanych; dla wygodniejszego powoływania się na nie, sformułujemy je w postaci lematu.

LEMAT 1. Jeżeli odpowiednie wartości oczekiwane istnieją, to

- (i) $E(E(X|Y)) = EX$;
- (ii) Zmienna losowa $E(X|Y)$ jest stała na zbiorach $\{\omega: Y(\omega) = \text{const}\}$. Jeżeli Z jest pewną funkcją zmiennej losowej Y , to $E(X \cdot Z|Y) = Z \cdot E(X|Y)$.
- (iii) $E(X|Y)$ ma wszystkie własności "zwykłej" wartości oczekiwanej zmiennej losowej X , np. dla stałych c_1, c_2 i zmiennych losowych X_1, X_2 mamy $E(c_1X_1 + c_2X_2|Y) = c_1E(X_1|Y) + c_2E(X_2|Y)$.
- (iv) $\text{Var } X = E\text{Var}(X|Y) + \text{Var } E(X|Y)$

D o w ó d. Dowód pozostawiamy jako ćwiczenie. Dla przykładu pokażemy tylko, jak dowieść własności (iv).

Na mocy (i) mamy

$$\text{Var } X = E(X - EX)^2 = E(E[(X - EX)^2|Y]).$$

Zapiszmy warunkową wartość oczekiwaną z ostatniego wyrażenia w postaci

$$E[(X - EX)^2|Y] = E[(X - E(X|Y) + E(X|Y) - EX)^2|Y].$$

Wielkość $E[(X - E(X|Y))^2|Y]$ jest wariancją zmiennej losowej X względem rozkładu warunkowego przy danym Y . Oznaczmy tę wielkość przez $\text{Var}(X|Y)$. Wartość oczekiwana tej zmiennej losowej tworzy pierwszy wyraz prawej strony wzoru (iv).

Wielkość $E[(E(X|Y) - EX)^2|Y]$ jest, po uśrednieniu, wariancją zmiennej losowej $E(X|Y)$ i tworzy drugi składnik po prawej stronie wzoru (iv).

Wielkość $E[(X - E(X|Y))(E(X|Y) - EX)|Y]$ jest równa zeru. \square

Jako wniosek z lematu 1(iv) otrzymujemy, że zawsze

$$(2) \quad \text{Var } E(X|Y) \leq \text{Var } X.$$

Dla bardziej pedantycznego Czytelnika odnotujemy, że — jak zawsze w teorii prawdopodobieństwa — wszystkie relacje między zmiennymi losowymi, które wyżej rozważaliśmy, powinny być rozumiane jako relacje zachodzące z prawdopodobieństwem 1. Sam jednak w całym wykładzie, kładąc nacisk na aplikacyjny aspekt rozważanych zagadnień, nie będę bardzo pedantyczny w demonstrowaniu różnych konstrukcji teoretycznych.

2. Przykład wprowadzający

Weźmy pod uwagę model statystyczny $(\{0, 1\}, \{P_\theta\{X = 1\} = \theta: 0 \leq \theta \leq 1\})^n$. Rozkład prawdopodobieństwa na przestrzeni proby $\mathcal{X} = \{0, 1\}^n$ ma postać

$$\begin{aligned} P_\theta\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} &= \\ &= \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}, \quad x = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n. \end{aligned}$$

Określmy statystykę T wzorem

$$T = \sum_{i=1}^n X_i$$

("liczba sukcesów w schemacie Bernoulliego"). Rozkład tej statystyki jest dobrze znanym rozkładem dwumianowym:

$$P_\theta\{T = t\} = \binom{n}{t} \theta^t (1 - \theta)^{n-t}, \quad t = 0, 1, \dots, n.$$

Łatwo sprawdzamy, że rozkład warunkowy próby losowej X_1, X_2, \dots, X_n , gdy $T = t$, ma postać

$$P_\theta\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T = t\} = \begin{cases} \binom{n}{t}^{-1}, & \text{gdy } \sum_{i=1}^n x_i = t. \\ 0 & \text{w p.p.} \end{cases}$$

Wynika stąd, że rozkład warunkowy $P_\theta\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T = t\}$ nie zależy od parametru θ . Możemy to interpretować w następujący sposób: gdy wiemy, że $T = t$, wtedy wiadomość o tym, który z $\binom{n}{t}$ punktów przestrzeni próby faktycznie się zrealizował, nie wnosi żadnych informacji o parametrze θ . Inaczej: jeżeli znamy łączną liczbę sukcesów w ciągu doświadczeń Bernoulliego, to informacja o kolejności, w jakiej się one pojawiały, nie wnosi nic nowego do naszej możliwości wnioskowania o wartości prawdopodobieństwa sukcesu θ .

Ten fakt jest od tak dawna i tak głęboko zakodowany w świadomości statystyków, że w omawianej sytuacji zwykle od razu rozważają model statystyczny prób Bernoulliego

$$(\{0, 1, 2, \dots, n\}, \{P_\theta\{T = t\} = \binom{n}{t} \theta^t (1 - \theta)^{n-t} : 0 \leq \theta \leq 1\})$$

zamiast naszego wyjściowego modelu.

To co wyżej powiedzieliśmy uzasadnia nazwanie T *statystyką dostateczną dla parametru θ* (lub: *statystyką dostateczną dla rozważanej rodziny rozkładów $\{P_\theta : \theta \in \Theta\}$*).

3. Definicja statystyki dostatecznej. Przykłady

Rozważamy ogólny model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$ z przestrzenią próby \mathcal{X} i rodziną rozkładów prawdopodobieństwa $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Niech T będzie statystyką.

DEFINICJA 1. Statystyka T nazywa się *statystyką dostateczną* (*statystyką dostateczną dla \mathcal{P}* lub *statystyką dostateczną dla θ*), jeżeli dla każdej wartości t tej statystyki rozkład warunkowy $P_\theta\{\cdot | T = t\}$ nie zależy od θ . \square

Z tej definicji wynika, że jeżeli statystyka T jest dostateczna i jeżeli statystyki T i S są równoważne, to również statystyka S jest dostateczna.

Przykład 1. Jeżeli X_1, X_2, \dots, X_n jest próbą losową, to dla każdego zdarzenia losowego A oraz dla każdego punktu x_1, x_2, \dots, x_n z przestrzeni próby mamy

$$P_\theta\{(X_1, X_2, \dots, X_n) \in A | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \mathbf{1}_A(x_1, x_2, \dots, x_n).$$

Ponieważ to prawdopodobieństwo nie zależy od θ , więc próba jest zawsze statystyką dostateczną. \blacksquare

Przykład 2. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego o gęstości

$$f_\sigma(x) = (\sigma\sqrt{2\pi})^{-1} \exp\left[-\frac{x^2}{2\sigma^2}\right], \quad \sigma > 0.$$

Weźmy pod uwagę statystykę $T = \sum_{i=1}^n X_i^2$. Udowodnimy, że jest to statystyka dostateczna dla σ .

Gęstość rozkładu prawdopodobieństwa próby wyraża się wzorem

$$f_\sigma(x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right\}.$$

Rozpatrzmy następujące wzajemnie jednoznaczne przekształcenie \mathbf{R}^n na siebie:

$$(3) \quad \begin{aligned} x_1 &= t \cos \varphi_1 \cos \varphi_2 \dots \cos \varphi_{n-1}, \\ x_2 &= t \sin \varphi_1 \cos \varphi_2 \dots \cos \varphi_{n-1}, \\ x_3 &= t \sin \varphi_2 \dots \cos \varphi_{n-1}, \\ &\dots \\ x_n &= t \sin \varphi_{n-1}, \end{aligned}$$

gdzie $0 < t < \infty$, $0 \leq \varphi_1 \leq 2\pi$, $-\frac{\pi}{2} < \varphi_2, \varphi_3, \dots, \varphi_{n-1} < \frac{\pi}{2}$.

Jakobian tego przekształcenia jest równy $t^{n-1} \cos \varphi_2 \cos^2 \varphi_3 \dots \cos^{n-2} \varphi_{n-1}$.

Niech $(T, \Phi_1, \Phi_2, \dots, \Phi_{n-1})$ będzie zmienną losową otrzymaną w wyniku przekształcenia (3) zmiennej losowej (X_1, X_2, \dots, X_n) . Zauważmy, że jeżeli wartość statystyki T jest ustalona, to próba (X_1, X_2, \dots, X_n) zmienia się wtedy i tylko wtedy, gdy zmienia się zmienna losowa $(\Phi_1, \Phi_2, \dots, \Phi_{n-1})$. Zatem statystyka T jest dostateczna dla σ wtedy i tylko wtedy, gdy dla każdej wartości t rozkład warunkowy zmiennej losowej $(\Phi_1, \Phi_2, \dots, \Phi_{n-1})$, pod warunkiem $T = t$, nie zależy od σ .

Oznaczmy przez g_σ gęstość zmiennej losowej $(T, \Phi_1, \Phi_2, \dots, \Phi_{n-1})$. Mamy

$$\begin{aligned} g_\sigma(t, \varphi_1, \varphi_2, \dots, \varphi_{n-1}) &= \\ &= (\sigma\sqrt{2\pi})^{-n/2} \exp\left[-\frac{t^2}{2\sigma^2}\right] t^{n-1} \cos \varphi_2 \cos^2 \varphi_3 \dots \cos^{n-2} \varphi_{n-1}, \end{aligned}$$

więc gęstość rozkładu warunkowego zmiennej losowej $(\Phi_1, \Phi_2, \dots, \Phi_{n-1})$, pod warunkiem $T = t$, jest równa $\text{const} \cdot \cos \varphi_2 \cos^2 \varphi_3 \dots \cos^{n-2} \varphi_{n-1}$, co nie zależy od σ . ■

Podkreślamy, że statystyka dostateczna T nie musi być statystyką jednowymiarową, tzn. odwzorowaniem przestrzeni próby \mathcal{X} w \mathbf{R}^1 . W przykładzie 1 mieliśmy n -wymiarową statystykę dostateczną (była to mianowicie cała próba). Z sytuacją, gdy T jest statystyką jednowymiarową, spotkaliśmy się w przykładzie wprowadzającym w paragrafie 1 oraz w ostatnim przykładzie.

W typowych sytuacjach można skonstruować k -wymiarowe statystyki dostateczne dla k dużo mniejszego niż wielkość próby n . Jest to bardzo istotne dla praktycznych zastosowań, dlatego że za pomocą statystyki dostatecznej uzyskujemy redukcję danych bez jakiegokolwiek straty informacji potrzebnej do wnioskowania o nieznanym rozkładzie.

4. Kryterium faktoryzacji

Prosty sposób rozpoznawania, czy dana statystyka T jest dostateczna i konstruowania statystyk dostatecznych daje następujące twierdzenie.

TWIERDZENIE 1 (kryterium faktoryzacji). *Statystyka T jest dostateczna wtedy i tylko wtedy, gdy gęstość rozkładu prawdopodobieństwa próby X_1, X_2, \dots, X_n można przedstawić w postaci*

$$(4) \quad f_{\theta}(x_1, x_2, \dots, x_n) = g_{\theta}(T(x_1, x_2, \dots, x_n))h(x_1, x_2, \dots, x_n),$$

gdzie funkcja h nie zależy od θ , a funkcja g_{θ} , zależna od θ , zależy od x_1, x_2, \dots, x_n tylko poprzez wartość statystyki T .

D o w ó d. Podamy dowód tego twierdzenia tylko dla dwóch najprostszych przypadków: rozkładów dyskretnych i rozkładów absolutnie ciągłych.

1) Przypadek rozkładów dyskretnych.

Przypuśćmy, że zachodzi (4). Ustalmy $\mathbf{x} = (x_1, x_2, \dots, x_n)$ oraz t . Jeżeli $\mathbf{x} \in T^{-1}(t)$, to

$$\begin{aligned} P_{\theta}\{\mathbf{X} = \mathbf{x} | T = t\} &= \frac{P_{\theta}\{\mathbf{X} = \mathbf{x}, T = t\}}{P_{\theta}\{T = t\}} = \frac{P_{\theta}\{\mathbf{X} = \mathbf{x}\}}{P_{\theta}\{T = t\}} = \\ &= \frac{g_{\theta}(T(\mathbf{x}))h(\mathbf{x})}{\sum_{\mathbf{x}T(\mathbf{x})=t} g_{\theta}(T(\mathbf{x}))h(\mathbf{x})} = \frac{g_{\theta}(t)h(\mathbf{x})}{\sum_{\mathbf{x}T(\mathbf{x})=t} g_{\theta}(t)h(\mathbf{x})} = \frac{h(\mathbf{x})}{\sum_{\mathbf{x}T(\mathbf{x})=t} h(\mathbf{x})}, \end{aligned}$$

co nie zależy od θ .

Jeżeli $\mathbf{x} \notin T^{-1}(t)$, to $P_{\theta}\{\mathbf{X} = \mathbf{x} | T = t\} = 0$, co znowu nie zależy od θ .

Przypuśćmy, że statystyka T jest dostateczna, tzn. że

$$P_{\theta}\{\mathbf{X} = \mathbf{x} | T = t\} = k(\mathbf{x}, t)$$

nie zależy od θ . Wtedy, dla $\mathbf{x} \in T^{-1}(t)$, na mocy równości

$$P_{\theta}\{\mathbf{X} = \mathbf{x} | T = t\} = \frac{P_{\theta}\{\mathbf{X} = \mathbf{x}\}}{P_{\theta}\{T = t\}}$$

otrzymujemy

$$P_{\theta}\{\mathbf{X} = \mathbf{x}\} = k(\mathbf{x}, t)P_{\theta}\{T = t\},$$

czyli faktoryzację (4).

2) Przypadek rozkładów ciągłych.

Niech $\mathbf{X} = (X_1, X_2, \dots, X_n)$ będzie daną próbą i niech $f_{\theta}^{\mathbf{X}}(\mathbf{x})$ będzie gęstością jej rozkładu. Weźmy pod uwagę r -wymiarową statystykę $T = (T_1, T_2, \dots, T_r)$, $r < n$. Niech $Y = (Y_1, Y_2, \dots, Y_{n-r})$, gdzie Y_j są takimi funkcjami próby, że odwzorowanie

$$\Psi(X_1, X_2, \dots, X_n) = (T_1, T_2, \dots, T_r, Y_1, Y_2, \dots, Y_{n-r})$$

jest wzajemnie jednoznacznym odwzorowaniem \mathbf{R}^n w siebie. Wtedy gęstość $f_{\theta}^{\mathbf{X}}(\mathbf{x})$ zmiennej losowej \mathbf{X} i gęstość $f_{\theta}^{T,Y}(t, y)$ zmiennej losowej (T, Y) są związane wzorem

$$(5) \quad f_{\theta}^{\mathbf{X}}(\mathbf{x}) = f_{\theta}^{T,Y}(T(\mathbf{x}), Y(\mathbf{x})) |J|,$$

gdzie $|J|$ jest jakobianem danego przekształcenia. Gęstość rozkładu warunkowego zmiennej losowej Y , gdy $T = t$, wyraża się zatem wzorem

$$(6) \quad f_{\theta}^{Y|t}(y) = \frac{f_{\theta}^{T,Y}(t, y)}{\int f_{\theta}^{T,Y}(t, s) ds}.$$

Mamy dowieść (por. przykład 2), że ta gęstość nie zależy od θ wtedy i tylko wtedy, gdy spełnione jest (4).

Przypuśćmy, że zachodzi (4), tzn. że $f_{\theta}^{\mathbf{X}}(\mathbf{x}) = g_{\theta}(T(\mathbf{x}))h(\mathbf{x})$. Na mocy (5)

$$f_{\theta}^{T,Y}(t, y) = f_{\theta}^{\mathbf{X}}(\Psi^{-1}(t, y)) |J^{-1}|,$$

co z kolei na mocy (4) jest równe $g_{\theta}(t) h(\Psi^{-1}(t, y)) |J^{-1}|$. Na mocy (6) otrzymujemy więc

$$\begin{aligned} f_{\theta}^{Y|t}(y) &= \frac{g_{\theta}(t) h(\Psi^{-1}(t, y)) |J^{-1}|}{\int g_{\theta}(t) h(\Psi^{-1}(t, s)) |J^{-1}| ds} \\ &= \frac{h(\Psi^{-1}(t, y)) |J^{-1}|}{\int h(\Psi^{-1}(t, s)) |J^{-1}| ds}, \end{aligned}$$

co nie zależy od θ .

Przypuśćmy teraz, że $f_{\theta}^{Y|t}(y)$ nie zależy od θ i oznaczmy tę wielkość przez $k(t, y)$. Wtedy, na mocy (6),

$$f_{\theta}^{T,Y}(t, y) = g_{\theta}(t) k(t, y),$$

gdzie

$$g_{\theta}(t) = \int f_{\theta}^{T,Y}(t, s) ds.$$

Na mocy (5) otrzymujemy więc

$$f_{\theta}^{\mathbf{X}}(\mathbf{x}) = f_{\theta}^{T,Y}(T(\mathbf{x}), Y(\mathbf{x})) |J| = g_{\theta}(T(\mathbf{x})) k(T(\mathbf{x}), Y(\mathbf{x})) |J|$$

i, kładąc $h(\mathbf{x}) = k(T(\mathbf{x}), Y(\mathbf{x})) |J|$, otrzymujemy faktoryzację (4). □

Przykład 3. Gęstość (względem miary liczącej) rozkładu próby X_1, X_2, \dots, X_n Bernoulliego wyraża się wzorem

$$P_\theta\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}.$$

Kładąc $T = \sum X_i$, $g_\theta(t) = \theta^t (1 - \theta)^{n-t}$ oraz $h(x) = 1$, na mocy kryterium faktoryzacji stwierdzamy, że T jest statystyką dostateczną. ■

Przykład 4. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu jednostajnego na przedziale $(0, \theta)$, $\theta > 0$, tzn. próbą z rozkładu o gęstości $f_\theta(x) = \theta^{-1} \mathbf{1}_{(0, \theta)}(x)$. Gęstość rozkładu prawdopodobieństwa próby możemy zapisać w postaci

$$f_\theta(x_1, x_2, \dots, x_n) = \theta^{-n} \mathbf{1}_{(0, \theta)}(x_{n:n}) \mathbf{1}_{(0, \infty)}(x_{1:n}).$$

Na mocy kryterium faktoryzacji $X_{n:n}$ jest statystyką dostateczną. ■

5. Minimalne statystyki dostateczne

Dla ustalenia uwagi, wszystkie rodziny $\{P_\theta : \theta \in \Theta\}$ rozważane w tym paragrafie, są rodzinami rozkładów na prostej. Rozkłady rozważanej rodziny są albo wszystkie dyskretne ("absolutnie ciągle względem miary liczącej"), albo wszystkie ciągle ("absolutnie ciągle względem miary Lebesgue'a").

Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(0, \sigma^2)$, $\sigma^2 > 0$. Z przykładu 1 wiemy, że cała próba X_1, X_2, \dots, X_n jest statystyką dostateczną. Wiemy również (por. zadanie 3), że statystyka pozycyjna $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ jest statystyką dostateczną. Jest to oczywiście "mniejsza" statystyka w tym sensie, że σ -ciało generowane przez statystykę pozycyjną jest pod- σ -ciałem σ -ciała generowanego przez próbę. Inaczej: statystyka pozycyjna $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ jest pewną funkcją próby X_1, X_2, \dots, X_n , ale nie odwrotnie: każdej wartości statystyki pozycyjnej $(x_{1:n}, x_{2:n}, \dots, x_{n:n})$ odpowiada $n!$ prób, z których taka wartość może pochodzić. Z przykładu 2 wiemy, że w rozważanym problemie statystyka $\sum_{i=1}^n X_i^2$ jest również dostateczna; jest to jeszcze mniejsza statystyka.

DEFINICJA 2. Statystykę dostateczną S nazywamy *minimalną statystyką dostateczną*, jeżeli dla każdej statystyki dostatecznej T istnieje funkcja h taka, że $S = h(T)$. □

Równoważnie: statystyka dostateczna S jest minimalną statystyką dostateczną, jeżeli dla każdej statystyki dostatecznej T mamy $\sigma(S) \subset \sigma(T)$. To sformułowanie bardziej poglądowo wyjaśnia użycie tu przymiotnika "minimalna".

Powstaje naturalne pytanie o minimalną statystykę dostateczną w danym problemie statystycznym $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$. Czy taka statystyka istnieje? Jak ją skonstruować?

Ogólna odpowiedź na pierwsze pytanie, dla wszystkich problemów rozważanych w naszych wykładach, jest pozytywna, ale dowód wymaga bogatszych narzędzi matematycznych niż te, którymi się tutaj posługujemy.

Drugie pytanie ma kapitalne znaczenie dla zastosowań gdyż, po pierwsze, w istocie rzeczy dotyczy maksymalnej redukcji danych bez straty informacji dla wnioskowania o nieznanym rozkładzie prawdopodobieństwa i, po drugie, ma bezpośredni związek z konstrukcją optymalnych reguł wnioskowania statystycznego.

Następujące dwa lematy pozwalają na efektywne skonstruowanie minimalnej statystyki dostatecznej w większości problemów, z którymi spotykamy się w praktycznych zastosowaniach.

LEMAT 2. Niech $\mathcal{P} = \{P_i : i = 1, 2, \dots\}$ będzie skończoną lub przeliczalną rodziną rozkładów o gęstościach p_i , $i = 1, 2, \dots$. Niech Λ będzie dowolnie ustalonym takim rozkładem prawdopodobieństwa na zbiorze $\{1, 2, \dots\}$, że $\lambda_i = \Lambda(\{i\}) > 0$ dla każdego $i = 1, 2, \dots$, i niech $P_\Lambda = \sum_i \lambda_i P_i$. Wtedy

$$(7) \quad S(X) = \left(\frac{p_1(X)}{p_\Lambda(X)}, \frac{p_2(X)}{p_\Lambda(X)}, \dots \right)$$

jest minimalną statystyką dostateczną.

Jeżeli $\mathcal{P} = \{P_i : i = 0, 1, 2, \dots\}$ jest rodziną rozkładów o wspólnym nośniku i o gęstościach $p_i : i = 0, 1, 2, \dots$, to

$$S(X) = \left(\frac{p_1(X)}{p_0(X)}, \frac{p_2(X)}{p_0(X)}, \dots \right)$$

jest minimalną statystyką dostateczną.

D o w ó d. Jeżeli $T = T(X)$ jest dowolną statystyką dostateczną dla \mathcal{P} , to na mocy twierdzenia o faktoryzacji każdy iloraz $p_i(x)/p_\Lambda(x)$ zależy od x tylko poprzez wartość $T(x)$. Stąd wynika, że statystyka (7) jest funkcją każdej statystyki dostatecznej. Statystyka S sama jest dostateczna dla \mathcal{P} znowu z kryterium faktoryzacji, bo przyjmując $u_j = g_j(u_1, u_2, \dots)$, mamy

$$p_j(x) = g_j(S(x)) p_\Lambda(x).$$

Zatem $S(X)$ jest minimalną statystyką dostateczną.

Dowód drugiej części twierdzenia jest analogiczny. \square

Następny ważny i łatwy w zastosowaniach lemat 3 wymaga w dowodzie dokładniejszego rozumowania: zadowolamy się skonstruowaniem pewnej relacji z dokładnością do zbiorów zerowych.

DEFINICJA 3. Mówimy, że rodziny rozkładów prawdopodobieństwa \mathcal{Q} i \mathcal{P} są równoważne, jeżeli dla każdego zdarzenia A mamy $Q(A) = 0$ ($\forall Q \in \mathcal{Q}$) wtedy i tylko wtedy, gdy $P(A) = 0$ ($\forall P \in \mathcal{P}$). Zbiór A taki, że $P(A) = 0$ ($\forall P \in \mathcal{P}$) nazywa się zbiorem zerowym w \mathcal{P} . \square

LEMAT 3. Niech $\mathcal{P}_0 \subset \mathcal{P}$ będzie podrodziną rodziny \mathcal{P} , równoważną z rodziną \mathcal{P} . Jeżeli statystyka S jest minimalną statystyką dostateczną dla \mathcal{P}_0 i dostateczną dla \mathcal{P} , to jest minimalną statystyką dostateczną dla \mathcal{P} .

D o w ó d. Niech T będzie dowolną statystyką dostateczną dla \mathcal{P} . Zatem T jest również dostateczna dla \mathcal{P}_0 . Ale S jest minimalną statystyką dostateczną dla \mathcal{P}_0 , więc istnieje taka funkcja h , że $S = h(T)$ z dokładnością do zbiorów zerowych w \mathcal{P}_0 , a więc również z dokładnością do zbiorów zerowych w \mathcal{P} , czyli S jest minimalną statystyką dostateczną w \mathcal{P} . \square

Przykład 5. Niech X_1, X_2, \dots, X_n będzie próbą z pewnego rozkładu z rodziny

$$\mathcal{P} = \{E(0, \theta) : \theta > 0\},$$

gdzie $E(0, \theta)$ jest rozkładem wykładniczym o gęstości $p_\theta(x) = \theta^{-1} \exp[-x/\theta] \mathbf{1}_{(0, \infty)}(x)$. Wtedy

$$p_\theta(x_1, x_2, \dots, x_n) = \theta^{-n} \exp\left\{-\sum_{i=1}^n x_i/\theta\right\}.$$

Dwuelementowa rodzina $\mathcal{P}_0 = \{p_{\theta_1}, p_{\theta_2}\}$, $\theta_1 \neq \theta_2$ jest równoważna z rodziną \mathcal{P} . Statystyka

$$S(X_1, X_2, \dots, X_n) = \frac{p_{\theta_2}(X_1, X_2, \dots, X_n)}{p_{\theta_1}(X_1, X_2, \dots, X_n)} = \frac{\theta_1^n}{\theta_2^n} \exp\left\{-\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right) \sum_{i=1}^n X_i\right\}$$

jest minimalną statystyką dostateczną dla \mathcal{P}_0 . Jest to statystyka równoważna ze statystyką $T = \sum_{i=1}^n X_i$. Na mocy twierdzenia o faktoryzacji jest to statystyka dostateczna dla \mathcal{P} , więc T jest minimalną statystyką dostateczną dla \mathcal{P} . \blacksquare

Przykład 6. Niech $\mathcal{P} = \{U(\theta - \frac{1}{2}, \theta + \frac{1}{2}) : \theta \in \mathbf{R}^1\}$ będzie rodziną rozkładów jednostajnych na przedziałach $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Niech $\mathcal{P}_0 = \{U(w_i - \frac{1}{2}, w_i + \frac{1}{2}) : i = 1, 2, \dots\}$, gdzie (w_1, w_2, \dots) jest ciągiem wszystkich liczb wymiernych. Podrodzina \mathcal{P}_0 jest równoważna z rodziną \mathcal{P} . Niech Λ będzie dowolnie ustalonym rozkładem z lematu 2. Wtedy, dla próby losowej X_1, X_2, \dots, X_n i dla każdego $i = 1, 2, \dots$, mamy

$$\begin{aligned} s_i(X_1, X_2, \dots, X_n) &= \frac{p_{w_i}(X_1, X_2, \dots, X_n)}{p_\Lambda(X_1, X_2, \dots, X_n)} \\ &= \frac{\mathbf{1}_{(w_i - \frac{1}{2}, w_i + \frac{1}{2})}(X_{1:n}) \mathbf{1}_{(w_i - \frac{1}{2}, w_i + \frac{1}{2})}(X_{n:n})}{\sum_{i=1}^n \lambda_i \mathbf{1}_{(w_i - \frac{1}{2}, w_i + \frac{1}{2})}(X_{1:n}) \mathbf{1}_{(w_i - \frac{1}{2}, w_i + \frac{1}{2})}(X_{n:n})}. \end{aligned}$$

Statystyka $S(X_1, X_2, \dots, X_n) = (s_1(X_1, X_2, \dots, X_n), s_2(X_1, X_2, \dots, X_n), \dots)$, określona wzorem (7) w lemacie 2, jest równoważna ze statystyką $(X_{1:n}, X_{n:n})$, bo odwzorowanie $(X_{1:n}, X_{n:n}) \rightarrow S(X_1, X_2, \dots, X_n)$ jest wzajemnie jednoznaczne: wystarczy zauważyć, że

$$\begin{aligned} x_{1:n} &= \sup\{w_i : s_i(x_1, x_2, \dots, x_n) > 0\} - \frac{1}{2}, \\ x_{n:n} &= \inf\{w_i : s_i(x_1, x_2, \dots, x_n) > 0\} + \frac{1}{2}. \end{aligned}$$

Zatem statystyka $(X_{1:n}, X_{n:n})$ jest minimalną statystyką dostateczną dla \mathcal{P}_0 , a ponieważ (z kryterium faktoryzacji) jest statystyką dostateczną dla \mathcal{P} , więc jest minimalną statystyką dostateczną w rozważanym modelu. \blacksquare

Pewien inny dogodny sposób konstruowania minimalnych statystyk dostatecznych związany jest z następującym rozumowaniem. Jeżeli T jest statystyką dostateczną, to z twierdzenia o faktoryzacji mamy

$$\frac{f_\theta(x)}{f_\theta(x')} = \frac{g_\theta(T(x)) h(x)}{g_\theta(T(x')) h(x')}$$

i stąd wynika, że iloraz $f_\theta(x)/f_\theta(x')$ nie zależy od θ , gdy x i x' należą do tej samej warstwy statystyki T , tzn. gdy $T(x) = T(x')$. Jeżeli S jest minimalną statystyką dostateczną, to $T(x) = T(x')$ implikuje, że $S(x) = S(x')$. Zatem S generuje najgrubsze rozbitcie przestrzeni próby o tej własności, że jeżeli x i x' przebiegają ten sam zbiór rozbitcia, to $f_\theta(x)/f_\theta(x')$ nie zależy od θ . W konkluzji: S jest minimalną statystyką dostateczną jeżeli $S(x) = S(x')$ wtedy i tylko wtedy, gdy iloraz $f_\theta(x)/f_\theta(x')$ nie zależy od θ .

Przykład 7. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu Cauchy'ego $C(\theta, 1)$ o gęstości

$$f_\theta(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \theta)^2}, \quad \theta \in \mathbf{R}^1.$$

Gęstość rozkładu próby wyraża się wzorem

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{\pi^n} \prod_{i=1}^n \frac{1}{1 + (x_i - \theta)^2}.$$

Weźmy pod uwagę iloraz

$$\frac{f_\theta(x_1, x_2, \dots, x_n)}{f_\theta(x'_1, x'_2, \dots, x'_n)} = \prod_{i=1}^n \frac{1 + (x'_i - \theta)^2}{1 + (x_i - \theta)^2}.$$

Jest to iloraz dwóch wielomianów stopnia $2n$ względem parametru $\theta \in \mathbf{R}^1$, o współczynniku przy θ^{2n} równym jedności. Nie zależy on od θ wtedy i tylko wtedy, gdy współczynniki przy tych samych potęgach θ w liczniku i mianowniku są równe. Tak jest wtedy i tylko wtedy, gdy ciągi liczb (x_1, x_2, \dots, x_n) oraz $(x'_1, x'_2, \dots, x'_n)$ różnią się tylko porządkiem. Zatem minimalną statystyką dostateczną jest statystyka pozycyjna $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$. ■

6. Statystyki swobodne. Statystyki zupełne. Twierdzenie Basu

DEFINICJA 4. Statystykę $V = V(X)$ nazywamy *statystyką swobodną*, jeżeli jej rozkład nie zależy od θ . Statystykę $V = V(X)$ nazywamy *statystyką swobodną pierwszego rzędu*, gdy wartość oczekiwana $E_\theta V(X)$ nie zależy od θ . □

Intuicyjnie można się spodziewać, że maksymalna redukcja danych do statystyki dostatecznej T zachodzi wtedy, gdy nie istnieje funkcja h , różna od stałej, taka, żeby rozkład zmiennej losowej $h(T)$ nie zależał od θ . W tak szerokim ujęciu ta koncepcja nie jest eksploatowana, ale w sensie swobody pierwszego rzędu odgrywa ważną rolę.

DEFINICJA 5. Mówimy, że rodzina rozkładów \mathcal{P} jest *zupełna*, jeżeli

$$\int X dP = 0 \quad (\forall P \in \mathcal{P}) \quad \text{implikuje} \quad X \equiv 0 \quad (\mathcal{P} - p.w.).$$

Mówimy, że statystyka T jest *zupełna*, jeżeli rodzina jej rozkładów jest rodziną zupełną, tzn. jeżeli z faktu, że $E_\theta h(T) = 0$ ($\forall \theta \in \Theta$) wynika, iż $h \equiv 0$ ($\mathcal{P} - p.w.$). \square

Jest to formalne ujęcie własności statystyki T polegającej na tym, że nie istnieje funkcja h tej statystyki, różna od stałej, która by miała wartość oczekiwaną niezależną od θ .

Okazuje się jednak, że nawet redukcja do minimalnej statystyki dostatecznej nie musi w tym sensie być zupełna (istnieją minimalne statystyki dostateczne, z których można jeszcze "wycisnąć" coś, co nie zależy od θ).

Przykład 8. Pokazaliśmy (por. przykład 7), że w rodzinie $\{C(\theta, 1) : \theta \in \mathbf{R}^1\}$ rozkładów Cauchy'ego statystyka $T = (X_{1:n}, X_{2:n}, \dots, X_{n:n})$ jest minimalną statystyką dostateczną.

Rozważana rodzina rozkładów jest "rodziną z parametrem położenia": dla dystrybucyj $F_\theta(x)$ mamy $F_\theta(x) = F(x - \theta)$, przy ustalonej dystrybucji $F = F_0$. Dla takich rodzin statystyka $X_{n:n} - X_{1:n}$ ma rozkład niezależny od θ , bo

$$P_\theta\{X_{n:n} - X_{1:n} \leq t\} = P_\theta\{(X_{n:n} - \theta) - (X_{1:n} - \theta) \leq t\}$$

ale jeżeli X_1, X_2, \dots, X_n pochodzi z rozkładu F_θ , to $X_1 - \theta, X_2 - \theta, \dots, X_n - \theta$ pochodzi z rozkładu $F = F_0$, czyli $P_\theta\{X_{n:n} - X_{1:n} \leq t\} = P_0\{X_{n:n} - X_{1:n} \leq t\}$, co nie zależy od θ . Minimalna statystyka dostateczna $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ nie jest więc w rozważanej rodzinie statystyką zupełną. \blacksquare

Ogólny związek między zupełnością i dostatecznością podaje następujące twierdzenie.

TWIERDZENIE 2. Jeżeli T jest statystyką dostateczną zupełną, to jest minimalną statystyką dostateczną.

D o w ó d. Nie podajemy dowodu w pełnej ogólności, choć twierdzenie jest ogólnie prawdziwe (pomijamy dyskusję na temat istnienia minimalnej statystyki dostatecznej).

Niech U będzie minimalną statystyką dostateczną. Wykażemy, że jeżeli T jest statystyką dostateczną zupełną, to statystyki T i U są równoważne.

Z definicji minimalnej statystyki dostatecznej istnieje funkcja g taka, że $U = g(T)$. Z drugiej strony, mamy zawsze $E_\theta(E_\theta(T|U)) = E_\theta T$ ($\forall \theta$), czyli $E_\theta[E_\theta(T|U) - T] = 0$ ($\forall \theta$). Ale $E_\theta(T|U)$ jest funkcją statystyki U , która z kolei jest funkcją statystyki T , więc $E_\theta(T|U) - T$ jest funkcją statystyki T . Statystyka T jest zupełna, więc $T = E_\theta(T|U)$, czyli $T = h(U)$ dla pewnej funkcji h . \square

Zakończymy ten paragraf następującym bardzo ważnym i pożytecznym twierdzeniem.

TWIERDZENIE 3 (twierdzenie Basu). Jeżeli T jest statystyką dostateczną zupełną w rodzinie $\{P_\theta : \theta \in \Theta\}$ i jeżeli V jest statystyką swobodną, to statystyki T i V są niezależne.

D o w ó d. Mamy wykazać, że dla każdego $\theta \in \Theta$ i dla każdego zdarzenia losowego A

$$P_\theta\{V \in A | T\} = P_\theta\{V \in A\}.$$

Ponieważ V jest statystyką swobodną, więc $P_\theta\{V \in A\}$ nie zależy od θ ; oznaczmy tę wielkość przez p_A .

Z drugiej strony, zawsze mamy

$$E_\theta[P_\theta\{V \in A | T\}] = P_\theta\{V \in A\},$$

więc dla każdego θ mamy $E_\theta[P_\theta\{V \in A | T\}] = p_A$, czyli dla każdego θ zachodzi $E_\theta[P_\theta\{V \in A | T\} - p_A] = 0$. Ponieważ $P_\theta\{V \in A | T\} - p_A$ jest funkcją statystyki T i T jest zupełna, więc $P_\theta\{V \in A | T\} - p_A \equiv 0$. \square

7. Rodziny wykładnicze rozkładów

W całym wykładzie rezygnujemy z prezentacji rodzin wykładniczych w pełnej ogólności; wszędzie dalej pojęcie rodziny wykładniczej jest zawężone do tzw. regularnych rodzin wykładniczych.

DEFINICJA 6. Rodzina rozkładów prawdopodobieństwa $\{P_\theta : \theta \in \Theta\}$ nazywa się *rodziną wykładniczą*, jeżeli każdy rozkład P_θ ma gęstość p_θ (względem tej samej miary) i ta gęstość ma postać

$$p_\theta(x) = \exp\left\{\sum_{j=1}^k c_j(\theta)T_j(x) - b(\theta)\right\} \cdot h(x),$$

gdzie $T_1(x), T_2(x), \dots, T_k(x)$ są funkcjami liniowo niezależnymi oraz

$$\{(c_1(\theta), c_2(\theta), \dots, c_k(\theta)) : \theta \in \Theta\}$$

jest pewnym k -wymiarowym zbiorem w \mathbf{R}^k . \square

Przykład 9. Rozkład dwupunktowy $P_\theta\{X = 1\} = \theta = 1 - P_\theta\{X = 0\}$ można zapisać w postaci

$$p_\theta(x) = \exp\left\{x \log \frac{\theta}{1-\theta} + \log(1-\theta)\right\}, \quad x = 0, 1.$$

Rodzina $\{P_\theta : \theta \in (0, 1)\}$ tych rozkładów jest więc rodziną wykładniczą. \blacksquare

Przykład 10. Gęstość rozkładu normalnego można przedstawić w postaci

$$f_{\mu, \sigma}(x) = \exp\left\{-\frac{1}{2\sigma^2} \cdot x^2 + \frac{\mu}{\sigma^2} \cdot x - \left(\frac{\mu^2}{2\sigma^2} + \log(\sigma\sqrt{2\pi})\right)\right\}.$$

Rodzina rozkładów normalnych $\{N(\mu, \sigma^2) : \mu \in \mathbf{R}^1, \sigma > 0\}$ jest rodziną wykładniczą. \blacksquare

Bez straty ogólności możemy rozkłady z rodziny wykładniczej zapisać w "naturalnej" parametryzacji:

$$p_{\theta}(x) = \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) - b(\theta) \right\}, \quad (\theta_1, \theta_2, \dots, \theta_k) \in \Theta,$$

gdzie Θ jest pewnym k -wymiarowym zbiorem w \mathbf{R}^k .

TWIERDZENIE 4. *Jeżeli $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$, $\Theta \subset \mathbf{R}^k$, jest rodziną wykładniczą rozkładów z gęstościami*

$$p_{\theta}(x) = \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) - b(\theta) \right\},$$

to $(T_1(X), T_2(X), \dots, T_k(X))$ jest (k -wymiarową) minimalną statystyką dostateczną.

D o w ó d. Bez straty ogólności przyjmujemy, że moc zbioru Θ jest większa od k , bo jeżeli tak nie jest, to minimalną dostateczność dowodzi się tak jak dla skończonej rodziny (por. wniosek 1).

Dostateczność wynika z kryterium faktoryzacji.

Dla dowodu minimalności wybierzmy $k+1$ punktów w Θ

$$\theta^l = (\theta_1^l, \theta_2^l, \dots, \theta_k^l), \quad l = 0, 1, \dots, k,$$

takich, że macierz (stopnia $k \times k$)

$$(8) \quad [(\theta_j^l - \theta_j^0)]_{j,l=1,2,\dots,k}$$

jest nieosobliwa. Weźmy pod uwagę rodzinę $\mathcal{P}_0 = \{p_{\theta^0}, p_{\theta^1}, \dots, p_{\theta^k}\}$. W tej rodzinie minimalną statystyką dostateczną jest

$$\left(\sum_{j=1}^k (\theta_j^1 - \theta_j^0) T_j(X), \sum_{j=1}^k (\theta_j^2 - \theta_j^0) T_j(X), \dots, \sum_{j=1}^k (\theta_j^k - \theta_j^0) T_j(X) \right),$$

czyli, na mocy nieosobliwości macierzy (8), statystyka $(T_1(X), T_2(X), \dots, T_k(X))$.

Rodzina $\mathcal{P}_0 \subset \mathcal{P}$ jest równoważna z rodziną \mathcal{P} , bo wszystkie rozkłady rodziny wykładniczej mają ten sam nośnik. Na mocy lematów 2 i 3 otrzymujemy tezę twierdzenia. \square

TWIERDZENIE 5. *Jeżeli $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$ jest rodziną wykładniczą z gęstościami $p_{\theta}(x) = \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) - b(\theta) \right\}$, to $(T_1(X), T_2(X), \dots, T_k(X))$ jest statystyką dostateczną zupełną.* \square

Nie podajemy szczegółowego dowodu tego twierdzenia. Ogólna idea dowodu jest następująca. Rozszerzamy przestrzeń parametrów Θ w taki sposób, że $\theta_1, \theta_2, \dots, \theta_k$ traktujemy jako liczby zespolone. Wtedy dla każdej funkcji φ całki $\int \varphi p_\theta$, jeżeli istnieją i są skończone, są funkcjami analitycznymi parametru zespolonego θ . Jeżeli dla każdego rzeczywistego θ mamy $\int \varphi(T(x))p_\theta(x)dx = 0$, to tak jest na całym Θ . Stąd wnioskujemy, że musi być $\varphi \equiv 0$, co dowodzi zupełności.

Również bez dowodu (tym razem dowód pozostawiamy jako zadanie 11) podajemy następujące pożyteczne twierdzenie.

Twierdzenie 6. *Jeżeli X_1, X_2, \dots, X_n jest próbą z pewnego rozkładu $P_\theta \in \mathcal{P}$ z rodziny wykładniczej \mathcal{P} , to*

$$\left(\sum_{i=1}^n T_1(X_i), \sum_{i=1}^n T_2(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

jest minimalną i zupełną statystyką dostateczną. □

Przykład 11. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu gamma $\Gamma(\alpha, \lambda)$ z parametrem kształtu $\alpha > 0$ i parametrem skali $\lambda > 0$, o gęstości

$$f_{\alpha, \lambda}(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\lambda} \mathbf{1}_{[0, \infty)}(x).$$

Rodzina rozkładów gamma $\{\Gamma(\alpha, \lambda) : \alpha > 0, \lambda > 0\}$ jest rodziną wykładniczą:

$$f_{\alpha, \lambda}(x) = \exp \left\{ -\frac{1}{\lambda} x + (\alpha - 1) \log x - \log [\lambda^\alpha \Gamma(\alpha)] \right\} \cdot \mathbf{1}_{[0, \infty)}(x).$$

Zatem statystyka $T = (T_1, T_2)$, gdzie

$$T_1 = \sum_{i=1}^n X_i, \quad T_2 = \sum_{i=1}^n \log X_i,$$

jest minimalną i zupełną statystyką dostateczną. ■

Przykład 12. Rozpatrzmy rodzinę rozkładów normalnych $N(\mu, \sigma^2)$ z gęstościami $f_{\mu, \sigma}(x) = (\sigma\sqrt{2\pi})^{-1} \exp\{-(x-\mu)^2/2\sigma^2\}$, z dodatnią średnią μ i z odchyleniem standardowym σ proporcjonalnym do średniej, ze znanym współczynnikiem proporcjonalności κ , tzn. niech $\mu > 0$ oraz $\sigma = \kappa\mu$. Tak jak w przykładzie 10 mamy

$$f_{\mu, \sigma}(x) = \exp \left\{ -\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\sigma^2} x - \left(\frac{\mu^2}{2\sigma^2} + \log(\sigma\sqrt{2\pi}) \right) \right\},$$

ale teraz

$$\left\{ \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2} \right) : \sigma = \kappa\mu, \mu > 0 \right\}$$

nie jest zbiorem dwuwymiarowym, więc rozważana rodzina $\{N(\mu, \sigma^2) : \sigma = \kappa\mu, \mu > 0\}$ nie jest rodziną wykładniczą. W szczególności, nie dają się teraz zastosować twierdzenia 4, 5 i 6 o minimalności i zupełności statystyki $(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ z próby X_1, X_2, \dots, X_n . (por. zadanie 12). ■

8. Zadania

1. Niech X_1, X_2, \dots, X_n będzie daną próbą. Niech

$$T = (X_{1:n}, X_{2:n}, \dots, X_{n:n})$$

będzie statystyką pozycyjną z próby X_1, X_2, \dots, X_n i niech $U = (U_1, U_2, \dots, U_n)$ oraz $S = (S_1, S_2, \dots, S_n)$ będą statystykami określonymi wzorami

$$U_1 = \sum_i X_i, \quad U_2 = \sum_{i \neq j} X_i X_j, \quad \dots, \quad U_n = X_1 X_2 \dots X_n,$$

$$S_k = X_1^k + X_2^k + \dots + X_n^k, \quad k = 1, 2, \dots, n.$$

Udowodnić równoważność tych statystyk.

2. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu Poissona o średniej $\theta > 0$. Wyznaczyć rozkład warunkowy próby pod warunkiem, że $T = t$, gdzie $T = X_1 + X_2 + \dots + X_n$. Wykazać, że T jest statystyką dostateczną.

3. Niech \mathcal{F} będzie rodziną wszystkich rozkładów na prostej o ciągłych dystrybuantach i niech X_1, X_2, \dots, X_n będzie próbą z rozkładu $F \in \mathcal{F}$. Wykazać, że statystyka pozycyjna jest statystyką dostateczną.

Ogólniej: niech \mathcal{P} będzie rodziną rozkładów prawdopodobieństwa P takich, że

$$P\{(X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)}) \in A\} = P\{(X_1, X_2, \dots, X_n) \in A\}$$

dla wszystkich zbiorów borelowskich A i wszystkich permutacji π zbioru $\{1, 2, \dots, n\}$. Wykazać, że statystyka pozycyjna jest dostateczna.

4. Wyznaczyć statystykę dostateczną z próby X_1, X_2, \dots, X_n dla rodziny rozkładów $\{U(\theta - \frac{1}{2}, \theta + \frac{1}{2}) : \theta \in \mathbf{R}^1\}$.

5. Rozważamy rodzinę rozkładów wykładniczych $E(\theta, \beta)$ o gęstościach

$$f_{\theta, \beta}(x) = \beta^{-1} \exp[-(x - \theta)/\beta] \mathbf{1}_{[\theta, \infty)}(x).$$

Niech X_1, X_2, \dots, X_n będzie próbą z tego rozkładu. Wykazać, że statystyka $(X_{1:n}, \sum(X_i - X_{1:n}))$ jest minimalną statystyką dostateczną.

6. Wykazać, że w rodzinie rozkładów logistycznych $L(\theta, 1)$ o gęstościach

$$f_{\theta}(x) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2}, \quad \theta \in \mathbf{R}^1,$$

statystyka pozycyjna jest minimalną statystyką dostateczną.

7. Niech $\mathcal{P} = \{C(0, \lambda) : \lambda > 0\}$ będzie rodziną rozkładów Cauchy'ego z parametrem skali λ , o gęstościach

$$f_{\lambda}(x) = \frac{\lambda}{\pi} \frac{1}{\lambda^2 + x^2},$$

a X_1, X_2, \dots, X_n niech będzie próbą z pewnego rozkładu tej rodziny. Niech $Y_i = X_i^2$ i niech $S = (Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})$ będzie statystyką pozycyjną obserwacji Y_1, Y_2, \dots, Y_n . Wykazać, że S jest minimalną statystyką dostateczną.

8. Statystyka $(X_{1:n}, X_{n:n})$ jest minimalną statystyką dostateczną dla rodziny rozkładów równomiernych $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathbf{R}^1$ (por. przykład 6). Wykazać, że nie jest to statystyka zupełna.

9. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu $U(\theta, \tau)$, $-\infty < \theta < \tau < \infty$, równomiernego na przedziale (θ, τ) .

(a) Wykazać, że $(X_{1:n}, X_{n:n})$ jest minimalną statystyką dostateczną.

(b) Sprawdzić, czy jest to statystyka zupełna.

10. Niech \mathcal{P}_0 i \mathcal{P}_1 będą dwiema rodzinami rozkładów prawdopodobieństwa, takimi że $\mathcal{P}_0 \subset \mathcal{P}_1$ i każdy zbiór zerowy w \mathcal{P}_0 jest zbiorem zerowym w \mathcal{P}_1 . Wykazać, że jeżeli \mathcal{P}_0 jest rodziną zupełną, to również \mathcal{P}_1 jest rodziną zupełną. Sprawdzić, że rodzina rozkładów dwumianowych $\mathcal{P}_0 = \{P_\theta\{X = k\} = \binom{n}{k}\theta^k(1-\theta)^{n-k} : 0 \leq \theta \leq 1\}$ jest zupełna i że rodzina $\mathcal{P}_1 = \mathcal{P}_0 \cup \{Q\}$, gdzie Q jest rozkładem Poissona o wartości oczekiwanej równej jedności, nie jest zupełna.

11. Udowodnić twierdzenie 6.

12. Sprawdzić, że statystyka $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ w przykładzie 12 jest statystyką dostateczną, ale nie jest statystyką zupełną.

13. Niech \mathcal{P} będzie rodziną wszystkich rozkładów prawdopodobieństwa na prostej, które mają gęstości (względem miary Lebesgue'a). Niech X_1, X_2, \dots, X_n będzie próbą z pewnego rozkładu tej rodziny. Wykazać, że statystyka pozycyjna $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ jest dostateczna i zupełna, a więc również minimalna.

Wykład III

ESTYMATORY NIEOBciążONE O MINIMALNEJ WARIANCJI

1. Sformułowanie problemu

Niech $\theta \in \Theta$ będzie nieznanym parametrem, który mamy oszacować i niech X_1, X_2, \dots, X_n będzie próbą z rozkładu P_θ . Zadanie polega na skonstruowaniu funkcji $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$ obserwacji X_1, X_2, \dots, X_n , takiej żeby $\hat{\theta}_n$ było bliskie θ .

Sformułujemy to dokładniej dla pewnej uproszczonej sytuacji. Rozważamy model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$. Niech $g : \Theta \rightarrow \mathbf{R}^1$ będzie daną funkcją rzeczywistą i przypuśćmy, że zadanie polega na oszacowaniu nieznannej wartości $g(\theta)$. Jeżeli dokonamy tego oszacowania za pomocą funkcji rzeczywistej $\delta(X_1, X_2, \dots, X_n)$, to

$$\delta(X_1, X_2, \dots, X_n) - g(\theta)$$

będzie błędem tego oszacowania. Jest to oczywiście zmienna losowa, a odpowiedni wybór funkcji δ ma uczynić tę zmienną losową w jakimś sensie małą.

Typowe postępowanie polega na wyborze δ w taki sposób, żeby minimalizować *błąd średniokwadratowy* oszacowania δ , a mianowicie

$$(1) \quad R_\delta(\theta) = E_\theta(\delta(X_1, X_2, \dots, X_n) - g(\theta))^2.$$

Wielkość R_δ nosi nazwę *funkcji ryzyka* lub krótko *ryzyka* estymatora δ .

Oczywiście najlepszym estymatorem byłaby funkcja δ_0 , taka że

$$(2) \quad R_{\delta_0}(\theta) \leq R_\delta(\theta) \quad (\forall \theta \in \Theta) \quad (\forall \delta \in \mathcal{D}),$$

gdzie \mathcal{D} jest rodziną rozważanych estymatorów.

Jeżeli \mathcal{D} jest rodziną wszystkich estymatorów, to taki optymalny estymator nie istnieje. Wybierając bowiem $\delta \equiv g(\theta_0)$ dla pewnego ustalonego $\theta_0 \in \Theta$, możemy uzyskać $R_\delta(\theta_0) = 0$, a więc dla jednostajnie najlepszego estymatora δ_0 musielibyśmy mieć $R_{\delta_0}(\theta) \equiv 0$, a tak mogłoby być tylko wtedy, gdyby $\delta_0(X_1, X_2, \dots, X_n) \equiv g(\theta)$ dla każdego $\theta \in \Theta$, co jest oczywiście niemożliwe.

Ponieważ klasa wszystkich estymatorów zawiera tak "bezsensowne" estymatory jak stałe, ograniczenie się do rozważania węższych klas estymatorów jest całkiem naturalne.

Jedną z interesujących klas tworzą tzw. *estymatory nieobciążone*. Są to estymatory spełniające warunek

$$(3) \quad E_{\theta} \delta(X_1, X_2, \dots, X_n) = g(\theta) \quad (\forall \theta \in \Theta).$$

Takie estymatory "średnio estymują tak jak należy".

Inną klasę tworzą *estymatory ekwiwariantne*. Zamiast ogólnej definicji, z której i tak nie korzystaliśmy w naszych wykładach, podamy prosty przykład takiej klasy. Weźmy pod uwagę model statystyczny $(\mathbf{R}^1, \{F_{\theta} : \theta \in \mathbf{R}^1\})$, gdzie F_{θ} jest dystrybuantą, taką że $F_{\theta}(x) = F(x - \theta)$ dla pewnej ustalonej dystrybuanty F . Jeżeli obserwacja X ma rozkład F_{θ} , to – dla każdej stałej c – obserwacja $X + c$ ma rozkład $F_{\theta+c}$. W takiej sytuacji, jeżeli $\delta(X)$ ma być estymatorem θ , to $\delta(X+c)$ powinno być estymatorem $\theta+c$. Zatem estymator δ powinien spełniać warunek $\delta(X+c) = \delta(X) + c \quad (\forall c \in \mathbf{R}^1)$.

Czasami wyróżnia się klasy estymatorów nie – jak wyżej – przez postulowanie ich własności, lecz przez postulowanie ich "kształtu". W taki właśnie sposób wyróżnia się klasę *estymatorów liniowych* (tzn. liniowych funkcji $\delta: \mathcal{X} \rightarrow \mathbf{R}^1$). Innym przykładem tego typu jest klasa *estymatorów rekurencyjnych*: jeżeli obserwacje X_1, X_2, \dots prowadzi się w taki sposób, że chciałoby się uzyskiwać na bieżąco estymatory $\delta_1(X_1), \delta_2(X_1, X_2), \dots$, to wygodnie byłoby, szczególnie dla dużych n , gdyby wartość estymatora δ_{n+1} na $(n+1)$ -szym kroku zależała tylko od wartości δ_n tego estymatora na n -tym kroku i od nowej obserwacji X_{n+1} .

W naszym wykładzie ograniczamy się tylko do klasy estymatorów nieobciążonych. Zauważmy, że jeżeli δ jest estymatorem nieobciążonym, to błąd średniokwadratowy (1) jest po prostu wariancją tego estymatora. Estymator δ_0 spełniający (2) w klasie estymatorów spełniających (3) nazywa się *estymatorem nieobciążonym o minimalnej wariancji*. Będziemy dla takich estymatorów używali skrótu *ENMW*, a jeżeli będziemy chcieli podkreślić, że jest to estymator nieobciążony o minimalnej wariancji dla wielkości $g(\theta)$, będziemy pisali *ENMW* $[g(\theta)]$.

W wykładzie podamy dwa twierdzenia stanowiące podstawę dla efektywnej konstrukcji *ENMW*, skonstruujemy takie estymatory dla pewnych funkcji g w jednopróbkowym modelu gaussowskim i w końcu powiemy o pewnych kłopotach związanych z tym podejściem.

2. Konstrukcja

Efektywna konstrukcja *ENMW* w wielu modelach statystycznych opiera się na dwóch następujących twierdzeniach.

TWIERDZENIE 1 (Rao–Blackwella). *Niech $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$ będzie rodziną rozkładów na przestrzeni próby \mathcal{X} i niech $g: \Theta \rightarrow \mathbf{R}^1$ będzie daną funkcją. Jeżeli \hat{g} jest estymatorem nieobciążonym funkcji g i jeżeli T jest statystyką dostateczną dla \mathcal{P} , to $E_{\theta}(\hat{g}|T)$ jest również estymatorem nieobciążonym, a wariancja tego estymatora jest jednostajnie nie większa od wariancji estymatora \hat{g} .*

D o w ó d. Ponieważ T jest statystyką dostateczną, więc $E_{\theta}(\hat{g}|T)$ nie zależy od θ , jest więc statystyką. Na mocy lematu 2.1(i) i nieobciążoności estymatora \hat{g} mamy

$$E_{\theta}(E_{\theta}(\hat{g}|T)) = E_{\theta}\hat{g} = g(\theta),$$

więc $E_\theta(\hat{g}|T)$ jest estymatorem nieobciążonym. Na mocy wzoru (2.2) mamy

$$\text{Var}_\theta E_\theta(\hat{g}|T) \leq \text{Var}_\theta \hat{g}$$

dla każdego $\theta \in \Theta$. □

TWIERDZENIE 2. *Jeżeli T jest statystyką dostateczną zupełną i jeżeli dla danej funkcji g istnieje funkcja \hat{g} taka, że*

$$(4) \quad E_\theta \hat{g}(T) = g(\theta) \quad (\forall \theta \in \Theta),$$

to $\hat{g}(T)$ jest $ENMW[g(\theta)]$.

K o m e n t a r z. To twierdzenie wypowiada się czasami w następujący sposób: jeżeli T jest statystyką dostateczną zupełną, to $\hat{g}(T)$ jest $ENMW$ swojej wartości oczekiwanej.

D o w ó d. Niech $\hat{g}(T)$ spełnia założenia twierdzenia i niech $\hat{g}_1(T)$ będzie dowolną funkcją spełniającą założenia twierdzenia. Wtedy

$$E_\theta(\hat{g}_1(T) - \hat{g}(T)) = 0 \quad (\forall \theta \in \Theta)$$

i z zupełności statystyki dostatecznej T wynika, że $\hat{g}_1(T) = \hat{g}(T)$. Zatem $\hat{g}(T)$ jest jedynym estymatorem nieobciążonym funkcji $g(\theta)$, opartym na statystyce T , a więc również estymatorem nieobciążonym o minimalnej wariancji w klasie estymatorów nieobciążonych. □

Zauważmy, że jeżeli T jest statystyką dostateczną zupełną i jeżeli $\hat{g}_1(T)$ oraz $\hat{g}_2(T)$ są estymatorami nieobciążonymi funkcji $g(\theta)$, to $\hat{g}_1(T) = \hat{g}_2(T)$. W szczególności, jeżeli $S_1(X)$ oraz $S_2(X)$ są dwoma estymatorami nieobciążonymi, to $E(S_1(X)|T) = E(S_2(X)|T)$. Zatem dla dowolnego estymatora nieobciążonego $S(X)$, estymator $E(S(X)|T)$ jest estymatorem nieobciążonym o minimalnej wariancji. Ten wynik jest czasami cytowany jako *twierdzenie Lehmana-Scheffégo*.

Dwa podane wyżej twierdzenia stanowią podstawę teoretyczną dla dwóch następujących metod konstrukcji $ENMW$.

M e t o d a 1 jest bezpośrednią konsekwencją twierdzenia 2: jeżeli mamy estymować $g(\theta)$ i jeżeli T jest statystyką dostateczną zupełną, to dla wyznaczenia $ENMW$ wystarczy znaleźć taką funkcję \hat{g} , żeby zachodziło (4).

Przykład 1. *Przypuśćmy, że zadanie polega na oszacowaniu wariancji $\theta(1 - \theta)$ rozkładu dwupunktowego $P_\theta\{X = 1\} = \theta = 1 - P_\theta\{X = 0\}$ na podstawie próby X_1, X_2, \dots, X_n z tego rozkładu. Statystyka $T = \sum_{i=1}^n X_i$ jest statystyką dostateczną zupełną. Jej rozkład jest rozkładem dwumianowym*

$$P_\theta\{T = t\} = \binom{n}{t} \theta^t (1 - \theta)^{n-t}.$$

Problem wyznaczenia $ENMW$ dla $g(\theta) = \theta(1 - \theta)$ sprowadza się do wyznaczenia funkcji \hat{g} takiej, żeby

$$(5) \quad \sum_{t=0}^n \hat{g}(t) \binom{n}{t} \theta^t (1 - \theta)^{n-t} = \theta(1 - \theta) \quad (\forall \theta \in \Theta).$$

Wprowadzając oznaczenie $v = \theta/(1 - \theta)$ i zapisując $v(1 + v)^{n-2}$ w postaci $\sum_{t=1}^{n-1} \binom{n-2}{t-1} v^t$, sprowadzamy (5) do postaci

$$\sum_{t=0}^n \hat{g}(t) \binom{n}{t} v^t = \sum_{t=1}^{n-1} \binom{n-2}{t-1} v^t \quad (\forall v \in (0, \infty)).$$

Porównując współczynniki obu wielomianów, ostatecznie otrzymujemy estymator

$$\hat{g}(T) = \frac{T(n-T)}{n(n-1)}.$$

■

M e t o d a 2 jest oparta na następującym wniosku z twierdzeń 1 i 2: jeżeli \tilde{g} jest dowolnym estymatorem nieobciążonym funkcji $g(\theta)$ i jeżeli T jest statystyką dostateczną zupełną, to

$$\hat{g}(T) = E_{\theta}(\tilde{g}|T)$$

jest $ENMW[g(\theta)]$.

Przykład 2. Przypuśćmy, że zadanie polega na oszacowaniu parametru $\lambda = e^{-\theta} = P_{\theta}\{X = 0\}$ dla zmiennej losowej X o rozkładzie Poissona $P(\theta)$, na podstawie próby X_1, X_2, \dots, X_n z tego rozkładu.

(To zadanie pojawia się w kontekście różnych zastosowań: jeżeli rozważa się procesy stochastyczne Poissona, to λ może być interpretowane jako prawdopodobieństwo zdarzenia polegającego na tym, że w odcinku czasu o danej długości nie pojawi się żaden "sygnał".)

Skonstruowanie "jakiegoś" estymatora nieobciążonego w tym problemie jest bardzo łatwe: oznaczając $Y_j = \mathbf{1}_{\{0\}}(X_j)$, możemy wziąć

$$(6) \quad \hat{\lambda} = \frac{1}{n} \sum_{j=1}^n Y_j.$$

Dla wariancji tego estymatora mamy

$$(7) \quad \text{Var}_{\theta} \hat{\lambda} = \frac{\lambda(1-\lambda)}{n}.$$

Statystyka $T = \sum_{j=1}^n X_j$ jest statystyką dostateczną zupełną, więc $\lambda^* = E_{\theta}(\hat{\lambda}|T)$ jest $ENMW[\lambda]$. Wyznaczymy tę statystykę w przypadku, gdy $n \geq 2$ (dla $n = 1$ rozwiązaniem jest oczywiście Y_1).

Kolejne rachunki dają

$$\begin{aligned} E_{\theta}(\hat{\lambda}|T = t) &= E_{\theta}\left(\frac{1}{n} \sum_{j=1}^n Y_j \middle| T = t\right) \\ &= E_{\theta}(Y_1|T = t) \\ &= P_{\theta}\{X_1 = 0|T = t\} \\ &= \sum_{\substack{x_2 + \dots + x_n = t \\ 0 \leq x_2, \dots, x_n \leq t}} P_{\theta}\{X_1 = 0, X_2 = x_2, \dots, X_n = x_n|T = t\} \end{aligned}$$

Ale

$$P_\theta\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T = t\} = \begin{cases} \frac{t!}{n^t x_1! x_2! \dots x_n!}, & \text{gd}y \sum x_i = t, \\ 0, & \text{w p.p.} \end{cases}$$

oraz

$$\sum_{x_1+x_2+\dots+x_n=t} \frac{t!}{x_1! x_2! \dots x_n!} = \underbrace{\left(1 + 1 + \dots + 1\right)}_{n \text{ razy}}^t = n^t,$$

więc

$$E_\theta(\hat{\lambda} | T = t) = \left(1 - \frac{1}{n}\right)^t$$

i ostatecznie

$$\lambda^* = \left(1 - \frac{1}{n}\right)^T.$$

Porównanie wariancji estymatorów $\hat{\lambda}$ i λ^* pozostawiamy jako zadanie 1. ■

3. ENMW w jednopróbkowym modelu gaussowskim

3.1. Statystyki

Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, \sigma^2)$. Przez Φ oznaczamy dystrybuantę rozkładu $N(0, 1)$. Niech $\bar{X} = \sum_{i=1}^n X_i/n$ będzie średnią z próby. Niech

$$(9) \quad S^2 = \begin{cases} \sum_{i=1}^n (X_i - \mu)^2, & \text{gd}y \mu \text{ jest znane,} \\ \sum_{i=1}^n (X_i - \bar{X})^2, & \text{gd}y \mu \text{ nie jest znane.} \end{cases}$$

Zmienna losowa S^2/σ^2 ma rozkład chi-kwadrat o ν stopniach swobody, gdzie $\nu = n$, gdy μ jest znane, oraz $\nu = n - 1$ w przeciwnym przypadku; przypominamy, że rozkład chi-kwadrat o ν stopniach swobody ma gęstość

$$(10) \quad g_\nu(x) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{[0, \infty)}(x).$$

Łatwo sprawdzamy, że jeżeli $\alpha + \nu > 0$, to

$$(11) \quad E_{\mu, \sigma} S^\alpha = \frac{\sigma^\alpha}{K_{\nu, \alpha}},$$

gdzie

$$(12) \quad K_{\nu, \alpha} = \frac{\Gamma(\frac{\nu}{2})}{2^{\frac{\alpha}{2}} \Gamma(\frac{\nu+\alpha}{2})}.$$

(Gdy $\alpha + \nu \leq 0$, mamy $E_{\mu, \sigma} S^\alpha = +\infty$.)

Wynika stąd, że jeżeli $\alpha + \nu > 0$, to $K_{\alpha,\nu}S^\alpha$ jest estymatorem nieobciążonym parametru σ^α .

Zmienne losowe \bar{X} i S^2 są niezależne.

Materiał przedstawiony w tym paragrafie był przedmiotem zadań w wykładzie 1.

3.2. Estymacja μ , gdy σ jest znane

Jeżeli σ jest znane, to \bar{X} jest statystyką dostateczną zupełną dla rodziny rozkładów $\{N(\mu, \sigma^2) : \mu \in \mathbf{R}^1\}$. Ponieważ dla każdego $\mu \in \mathbf{R}^1$ mamy $E_{\mu,\sigma}\bar{X} = \mu$, więc \bar{X} jest estymatorem nieobciążonym o minimalnej wariancji parametru μ .

Weźmy pod uwagę funkcję $g(\mu) = \mu^2$. Ponieważ $E_{\mu,\sigma}\bar{X}^2 = \mu^2 + \sigma^2/n$, więc $\bar{X}^2 - \sigma^2/n$ jest $ENMW[\mu^2]$. Analogicznie można skonstruować $ENMW[\mu^k]$ dla $k > 2$.

3.3. Estymacja σ^α , gdy μ jest znane

Jeżeli μ jest znane, to S^2 jest statystyką dostateczną zupełną. Zatem $K_{n,\alpha}S^\alpha$ jest $ENMW[\sigma^\alpha]$. W zastosowaniach ważną rolę odgrywają dwa następujące przypadki:

$$(13) \quad \frac{\Gamma(\frac{n}{2})}{\sqrt{2}\Gamma(\frac{n+1}{2})}S \quad \text{jest} \quad ENMW[\sigma],$$

$$(14) \quad \frac{1}{n}S^2 \quad \text{jest} \quad ENMW[\sigma^2].$$

3.4. Przypadek, gdy μ oraz σ nie są znane

Średnia z próby \bar{X} jest $ENMW[\mu]$. Statystyka $S^2/(n-1)$ jest $ENMW[\sigma^2]$. Statystyka $\Gamma(\frac{n-1}{2})S/\sqrt{2}\Gamma(\frac{n}{2})$ jest $ENMW[\sigma]$. Te proste stwierdzenia pozostawiamy do sprawdzenia.

Chwili zastanowienia wymaga zauważenie, że

$$\frac{\sqrt{2}\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2}-1)}\frac{\bar{X}}{S} \quad \text{jest} \quad ENMW\left[\frac{\mu}{\sigma}\right].$$

Dla dowodu tego faktu trzeba skorzystać z tego, że $K_{n-1,-1}S^{-1}$ jest $ENMW[\sigma^{-1}]$, \bar{X} jest $ENMW[\mu]$ i że \bar{X} oraz S są niezależne.

3.5. Estymacja kwantyla rozkładu $N(\mu, \sigma^2)$

Niech $p \in (0, 1)$ będzie ustaloną liczbą i niech u_p będzie p -tym kwantylem rozkładu $N(\mu, \sigma^2)$, tzn. niech u_p będzie rozwiązaniem (względem u) równania

$$P_{\mu, \sigma}\{X \leq u\} = p.$$

Mamy więc $u_p = \mu + \sigma \cdot \Phi^{-1}(p)$. Ponieważ (\bar{X}, S) jest statystyką dostateczną zupełną, więc $\bar{X} + K_{n-1,1}S\Phi^{-1}(p)$ jest ENMW[u_p].

3.6. Estymacja prawdopodobieństwa $P_{\mu, \sigma}\{X \leq u\}$

Niech $u \in \mathbf{R}^1$ będzie ustaloną liczbą i niech $p = P_{\mu, \sigma}\{X \leq u\}$. Statystyka z próby X_1, X_2, \dots, X_n , określona wzorem

$$Y_1 = \mathbf{1}_{(-\infty, u]}(X_1),$$

jest estymatorem nieobciążonym prawdopodobieństwa p . Ponieważ (\bar{X}, S) jest statystyką dostateczną zupełną, więc

$$(15) \quad \hat{p} = E_{\mu, \sigma}(Y_1 | \bar{X}, S)$$

jest ENMW[p]. Pokażemy jak obliczać ten estymator.

Oznaczając

$$T = \frac{X_1 - \bar{X}}{S},$$

mamy

$$(16) \quad E_{\mu, \sigma}(Y_1 | \bar{X} = \bar{x}, S = s) = P_{\mu, \sigma}\left\{T \leq \frac{u - \bar{X}}{S} \mid \bar{X} = \bar{x}, S = s\right\}.$$

LEMAT 1. *Statystyki T oraz (\bar{X}, S) są niezależne.*

D o w ó d.

Zauważmy, że T jest statystyką swobodną, a ponieważ (\bar{X}, S) jest statystyką dostateczną zupełną, więc na mocy twierdzenia Basu otrzymujemy, że T oraz (\bar{X}, S) są niezależne. \square

Niech $B(\alpha, \beta)$ oznacza rozkład beta o gęstości

$$f_{\alpha, \beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{(0,1)}(x).$$

LEMAT 2. *Jeżeli Z_1, Z_2, \dots, Z_r są niezależnymi zmiennymi losowymi o jednakowym rozkładzie gamma $\Gamma(\alpha, 1/b)$, to zmienna losowa $W = Z_1/(Z_1 + Z_2 + \dots + Z_r)$ ma rozkład beta $B(\alpha, (r-1)\alpha)$.*

D o w ó d.

Rozpatrzmy wzajemnie jednoznaczne przekształcenie

$$\begin{aligned} W &= \frac{Z_1}{Z_1 + Z_2 + \dots + Z_r}, \\ W_1 &= \frac{Z_1 + Z_2}{Z_1 + Z_2 + \dots + Z_r}, \\ &\dots \\ W_{r-2} &= \frac{Z_1 + Z_2 + \dots + Z_{r-1}}{Z_1 + Z_2 + \dots + Z_r}, \\ W_{r-1} &= Z_1 + Z_2 + \dots + Z_r \end{aligned}$$

o jacobianie równym W_{r-1}^{r-1} . Gęstość łącznego rozkładu tych nowych zmiennych losowych wyraża się wzorem

$$\begin{aligned} f(w, w_1, \dots, w_{r-2}, w_{r-1}) &= \\ &= \frac{b^{r\alpha}}{\Gamma^r(\alpha)} w^{\alpha-1} (w_1 - w)^{\alpha-1} \dots (w_{r-2} - w_{r-3})^{\alpha-1} (1 - w_{r-2})^{\alpha-1} w_{r-1}^{r\alpha-1} e^{-bw_{r-1}} \end{aligned}$$

dla $0 < w < w_1 < \dots < w_{r-2} < 1, \quad 0 < w_{r-1} < \infty$.

Całkując, przy wykorzystaniu wzoru

$$\int_A^B (x-A)^{\alpha-1} (B-x)^{\beta-1} dx = (B-A)^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

kolejno względem $w_{r-1}, w_{r-2}, \dots, w_1$, otrzymujemy gęstość rozkładu brzegowego zmiennej losowej W

$$f(w) = \frac{\Gamma(r\alpha)}{\Gamma(\alpha)\Gamma((r-1)\alpha)} w^{\alpha-1} (1-w)^{(r-1)\alpha-1} \mathbf{1}_{(0,1)}(w),$$

co kończy dowód lematu. □

W celu efektywnego wyznaczenia estymatora \hat{p} rozumiemy w następujący sposób. Na mocy lematu 1 statystyki T oraz (\bar{X}, S) są niezależne, więc

$$(17) \quad P_{\mu, \sigma}\left\{T \leq \frac{u - \bar{X}}{S} \mid \bar{X} = \bar{x}, S = s\right\} = P_{\mu, \sigma}\left\{T \leq \frac{u - \bar{x}}{s}\right\}.$$

W dowodzie lematu 1 pokazaliśmy, że rozkład zmiennej losowej T nie zależy od parametru (μ, σ) , więc dla dalszych rachunków możemy przyjąć, że rozważana próba X_1, X_2, \dots, X_n pochodzi z rozkładu normalnego $N(0, 1)$. Weźmy pod uwagę dowolne ortonormalne przekształcenie próby (X_1, X_2, \dots, X_n) w $(\xi_1, \xi_2, \dots, \xi_n)$ takie, że

$$\xi_1 = \sqrt{n}\bar{X},$$

$$\xi_n = \sqrt{\frac{n}{n-1}}(\bar{X} - X_1).$$

Wtedy (por. zadanie 1.10, gdzie używaliśmy pewnego innego przekształcenia) $\xi_1, \xi_2, \dots, \xi_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie normalnym $N(0, 1)$, przy czym

$$\sum_{j=2}^n \xi_j^2 = S^2 \quad \text{oraz} \quad \sqrt{\frac{n}{n-1}}(\bar{X} - X_1) = \xi_n.$$

Zatem

$$\frac{\xi_n^2}{\xi_2^2 + \xi_3^2 + \dots + \xi_n^2} = \frac{n}{n-1} T^2.$$

Na mocy lematu 2, zmienna losowa $\frac{n}{n-1} T^2$ ma rozkład beta $B\left(\frac{1}{2}, \frac{n-2}{2}\right)$. Łatwo sprawdzamy, że gęstość rozkładu prawdopodobieństwa zmiennej losowej T wyraża się wzorem

$$(18) \quad h(t) = \sqrt{\frac{n}{n-1}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)} \left(1 - \frac{nt^2}{n-1}\right)^{\frac{n}{2}-2} \mathbf{1}_{\left(-\sqrt{\frac{n-1}{n}}, \sqrt{\frac{n-1}{n}}\right)}(t),$$

więc ostatecznie otrzymujemy

$$(19) \quad \hat{p} = \int_{-\sqrt{\frac{n-1}{n}}}^{\frac{u-\bar{x}}{s}} h(t) dt.$$

Praktyczne stosowanie tego estymatora wymaga oczywiście odpowiednich tablic lub wspomaganie komputerowego.

4. Kłopoty z ENMW

Przypomnijmy, iż ograniczenie się do klasy estymatorów nieobciążonych było spowodowane tym, że w klasie wszystkich estymatorów nie istnieją estymatory jednostajnie minimalizujące błąd oszacowania. Redukcja przez nieobciążoność stwarza jednak czasami pewne niedogodności.

1. *Estymatory nieobciążone nie zawsze istnieją.* Przypuśćmy na przykład, że w rozkładzie dwumianowym $P_\theta\{X = x\} = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, $\theta \in (0, 1)$, mamy oszacować wielkość $g(\theta) = \theta^{-1}$. Estymator nieobciążony $\hat{g}(X)$ musiałby spełniać warunek

$$\sum_{x=0}^n \hat{g}(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{1}{\theta} \quad \forall \theta \in (0, 1),$$

tzn. warunek

$$v \sum_{x=0}^n \hat{g}(x) \binom{n}{x} v^x = (1+v)^{n+1} \quad \forall v \in (0, +\infty),$$

co jest niemożliwe, bo lewa strona dąży do zera, a prawa do 1, gdy $v \rightarrow 0$.

2. *ENMW może nie istnieć, chociaż istnieją estymatory nieobciążone* (por zad. 5).

3. *ENMW może okazać się "gorszy" od estymatora obciążonego.* Przypomnijmy, że punktem wyjścia dla oceny estymatora $\delta(X)$ był jego błąd średniokwadratowy

$$R_\delta(\theta) = E_\theta(\delta(X) - g(\theta))^2.$$

Przypuśćmy, że zadanie polega na oszacowaniu $g(\mu, \sigma) = \sigma^2$ w rozkładzie normalnym $N(\mu, \sigma^2)$. Wiemy, że estymatorem nieobciążonym o minimalnej wariancji jest

$$\frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2.$$

Weźmy pod uwagę następującą klasę estymatorów, zawierającą ten estymator,

$$(20) \quad \hat{\sigma}_c^2 = c \cdot \sum_{j=1}^n (X_j - \bar{X})^2, \quad c > 0.$$

Ryzyko estymatora $\hat{\sigma}_c^2$ jest równe

$$\sigma^4 \cdot [c^2(n^2 - 1) - 2c(n - 1) + 1].$$

Oczywiste jest, że estymator (obciążony!) ze stałą $c = (n + 1)^{-1}$ ma jednostajnie najmniejsze ryzyko w tej klasie.

4. *ENMW może być zupełnie bezsensowny.* Przypuśćmy, że w rozkładzie

$$P_\theta\{X = x\} = \frac{\theta^x e^{-\theta}}{x!(1 - e^{-\theta})} \quad x = 1, 2, \dots,$$

mamy, na podstawie jednej obserwacji X , oszacować $g(\theta) = e^{-\theta}$. Oczywiście X jest statystyką dostateczną zupełną, więc estymatorem nieobciążonym o minimalnej wariancji będzie $T(X)$ spełniające warunek

$$\sum_{x=1}^{\infty} T(x) \frac{\theta^x e^{-\theta}}{x!(1 - e^{-\theta})} = e^{-\theta} \quad \forall \theta > 0.$$

Wynika stąd, że

$$(21) \quad T(x) = (-1)^{x+1}.$$

Oczywiście $e^{-\theta} \in (0, 1)$. Tymczasem estymator T szacuje tę wielkość za pomocą liczby $+1$ (gdy x jest nieparzyste) lub -1 (gdy x jest parzyste). Ryzyko tego estymatora jest równe $R_T(\theta) = 1 - e^{-2\theta}$.

Weźmy pod uwagę "trochę bardziej naturalny" estymator

$$(22) \quad S(x) = \begin{cases} 0, & \text{gdy } x \text{ jest parzyste,} \\ 1, & \text{gdy } x \text{ jest nieparzyste.} \end{cases}$$

Dla tego estymatora mamy ryzyko $R_S(\theta) = \frac{1}{2}(1 - e^{-\theta})$. Wobec tego, że $\frac{1}{2}(1 - e^{-\theta}) < (1 + e^{-\theta})(1 - e^{-\theta}) = 1 - e^{-2\theta}$, estymator S (również niezbyt wyrafinowany) jest zawsze lepszy od estymatora nieobciążonego o minimalnej wariancji T .

Sformułowane wyżej kłopoty nie są kłopotami specyficznymi dla estymatorów nieobciążonych i dla kryterium "błąd średniokwadratowy". Jest to sytuacja raczej typowa w statystyce: ograniczenie się do klasy estymatorów, w której można sensownie sformułować, a następnie rozwiązać problem optymalizacji, pozostawia poza rozważaną klasą estymatory, które według zaakceptowanego kryterium mogą okazać się lepsze od najlepszych estymatorów w rozważanej klasie.

5. Zadania

1. Na podstawie próby X_1, X_2, \dots, X_n z rozkładu Poissona $P(\theta)$ estymujemy parametr $\lambda = e^{-\theta}$. Zbadać estymatory

$$\hat{\lambda} = \frac{1}{n} \sum_{j=1}^n Y_j \quad \text{oraz} \quad \lambda^* = \left(1 - \frac{1}{n}\right)^T,$$

gdzie $Y_j = \mathbf{1}_{\{0\}}(X_j)$ oraz $T = \sum_{j=1}^n X_j$. Wyznaczyć błąd średniokwadratowy tych estymatorów. Porównać ich rozkłady dla $n = 4$ i $\theta = \frac{1}{2}$ ($\lambda = 0.60653$).

2. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, \sigma^2)$, gdzie σ^2 jest znaną stałą. Skonstruować $ENMW[\mu^k]$.

3. Sprawdzić, że statystyki podane w punkcie 3.4 są estymatorami nieobciążonymi o minimalnej wariancji dla, odpowiednio, μ , σ^2 oraz σ .

4. Uzupełnić punkt 3.6 wszystkimi szczegółami technicznymi.

5. Zmienna losowa X ma rozkład jednostajny na zbiorze $\{\theta-1, \theta, \theta+1\}$, przy czym θ jest nieznaną liczbą całkowitą. Wykazać, że $ENMW[\theta]$ nie istnieje chociaż istnieją estymatory nieobciążone.

6. Obliczyć błąd średniokwadratowy estymatora $\hat{\sigma}_c^2$ określonego wzorem (20).

7. Wyznaczyć ryzyko estymatorów T (wzór (21)) i S (wzór (22)).

8. Zmienna losowa X ma "ucięty" rozkład Poissona

$$P_\theta\{X = x\} = \frac{\theta^x e^{-\theta}}{x!(1 - e^{-\theta})}, \quad x = 1, 2, \dots; \quad \theta > 0.$$

Próba losowa X_1, X_2, \dots, X_N z tego rozkładu zawiera n_r obserwacji o wartości r . Wykazać, że

$$\theta^* = \frac{1}{N} \sum_{r=2}^{\infty} r \cdot n_r$$

jest estymatorem nieobciążonym parametru θ .

9. Niech X będzie zmienną losową o rozkładzie dwumianowym

$$P_\theta\{X = x\} = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, 2, \dots, n; \quad \theta \in (0, 1).$$

Wyznaczyć estymator nieobciążony dla $g(\theta) = \theta^m$, gdzie m jest ustaloną liczbą całkowitą. Przedyskutować, dla jakich m taki estymator istnieje.

10. Rozważamy ciąg niezależnych doświadczeń z prawdopodobieństwem sukcesu θ w każdym doświadczeniu. Niech X będzie liczbą porażek zaobserwowanych do chwili otrzymania m sukcesów (m jest ustaloną liczbą naturalną). Skonstruować estymator nieobciążony dla $g(\theta) = \theta^{-1}$. (Jak wiadomo — por. zadanie 3.9 — w przypadku z góry ustalonej liczby doświadczeń taki estymator nie istnieje).

11. W populacji składającej się z N elementów znajduje się pewna nieznaną liczbą M elementów wyróżnionych. Losujemy bez zwracania n elementów i obserwujemy liczbę X elementów wyróżnionych w wylosowanej próbie. Skonstruować $ENMW[M]$.

12. Niech T_1 i T_2 będą dwoma estymatorami nieobciążonymi o jednostajnie minimalnej wariancji. Wykazać, że $T_1 = T_2$.

13. Zmienne losowe X_1, X_2, \dots, X_n są niezależne i mają jednakowy rozkład o gęstości $\theta e^{-\theta x} \mathbf{1}_{(0, \infty)}(x)$, $\theta > 0$. Wykazać, że zmienna losowa przyjmująca wartość 1, gdy $X_1 \geq k$, i wartość 0, gdy $X_1 < k$, jest estymatorem nieobciążonym dla $g(\theta) = e^{-k\theta}$, gdzie k jest ustaloną liczbą. Na tej podstawie wykazać, że, przy odpowiednim wyborze statystyki T , zmienna losowa

$$\hat{g}(T) = \begin{cases} 0, & \text{gdy } T < k, \\ \left(1 - \frac{k}{T}\right)^{n-1}, & \text{gdy } T \geq k, \end{cases}$$

jest $ENMW[g(\theta)]$.

Wykład IV

TESTOWANIE HIPOTEZ STATYSTYCZNYCH

1. Wprowadzenie

Badacz zjawisk fizycznych, przyrodniczych, ekonomicznych, itp. postępuje zwykle w następujący sposób. Na podstawie całej dotychczasowej i znanej mu wiedzy o danym zjawisku i na podstawie własnej kontemplacji buduje pewną teorię rozważanego zjawiska. Takie teorie mogą czasami brzmieć niezwykle fantastycznie, ale tak czy inaczej ostatecznym ich sprawdzianem jest doświadczenie. Uczciwy badacz szczerze poszukuje faktów negujących jego teorię i w jakimś stopniu ją potwierdza, gdy coraz to nowe eksperymenty i obserwacje nie dostarczają takich faktów. Oczywiście może współistnieć kilka różnych teorii danego zjawiska; doświadczenie w końcu eliminuje niektóre z nich, nowe fakty z doświadczeń prowadzą do formułowania nowych teorii i ten proces poznawania otaczającego nas świata postępuje "w nieskończoność".

Weryfikacja teorii przebiega zwykle według schematu: jeżeli dana teoria jest prawdziwa, to określone zdarzenie Z jest niemożliwe; zaobserwowanie zdarzenia Z dyskwalifikuje ("falsyfikuje") teorię. W tak ogólnym ujęciu nie będziemy się oczywiście tym zajmowali – jest to problem ogólnych zasad rozwijania wiedzy o otaczającym nas świecie i logiki poznania naukowego. Statystyka matematyczna wkracza w tę dziedzinę w sytuacji, gdy teoria, którą należy zweryfikować, formułuje pewien probabilistyczny model badanego zjawiska. W takiej sytuacji schemat: "jeżeli dana teoria jest prawdziwa, to zdarzenie Z jest niemożliwe" rzadko udaje się zastosować. Trudność tkwi w wyspecyfikowaniu zdarzeń niemożliwych przy założeniu, że teoria jest prawdziwa. Można natomiast zwykle wyspecyfikować zdarzenia "mało prawdopodobne", tak mało prawdopodobne, że aż "praktycznie niemożliwe". Zaobserwowanie takiego zdarzenia jest argumentem przeciwko weryfikowanej teorii, tym mocniejszym, im takie zdarzenie jest mniej prawdopodobne. Statystyczna teoria weryfikacji hipotez ma dostarczać badaczom różnych dziedzin wiedzy właśnie tego typu argumentów.

Formalnie: rozważamy, odpowiedni dla danej teorii, model probabilistyczny $H = (\Omega, \mathcal{A}, P)$ i — odpowiednio do doświadczenia, które ma weryfikować ten model — specyfikujemy zmienną losową X i zbiór jej wartości ("przestrzeń próby") \mathcal{X} . Poszukujemy takiego rozbicia przestrzeni próby \mathcal{X} na dwa zbiory \mathbf{K} i \mathbf{A} , żeby zdarzenie " $X \in \mathbf{K}$ " było "praktycznie niemożliwe" i "świadczyło przeciwko modelowi H ". Jeżeli w wyniku odpowiedniego eksperymentu zaobserwujemy zdarzenie " $X \in \mathbf{K}$ ", kwestionujemy weryfikowaną teorię.

Praktycznie robi się to w ten sposób, że poszukuje się takiej statystyki $T = T(X)$, której duże wartości bardziej przemawiają przeciwko weryfikowanej hipotezie H niż małe i za "obszar krytyczny" \mathbf{K} testu statystycznego przyjmuje się zbiór postaci $\{T > t_\alpha\}$, gdzie t_α jest "odpowiednio wybraną liczbą". Tę "odpowiednio wybraną liczbę" konstruuje się w następujący sposób. Umawiamy się, że "zdarzenia tak mało prawdopodobne, że aż praktycznie niemożliwe" to takie zdarzenia, których prawdopodobieństwo jest nie większe od ustalonej, małej liczby $\alpha \in (0, 1)$. W zastosowaniach najczęściej $\alpha = 0.01$ lub $\alpha = 0.05$ – nie chodzi przy tym o sympatię do takich okrągłych liczb tylko o to, że na coś trzeba się zdecydować, gdy dochodzi do budowania praktycznych przepisów testowania i konstrukcji odpowiednich tablic i procedur komputerowych. Liczbę α nazywa się *poziomą istotnością* testu. Liczbę t_α nazywa się *wartością krytyczną*. Jest to najmniejsza liczba t taka, że $P\{T > t\} \leq \alpha$. W konsekwencji uważamy, że zdarzenie $\{T > t_\alpha\}$ "przeczy weryfikowanej hipotezie". Liczbę $P\{T > t_\alpha\}$ nazywamy *rozmiarem testu*.

Zauważmy, że przy takim podejściu hipoteza, iż teoria jest poprawna, zamienia się w hipotezę, że obserwowana zmienna losowa X ma dany rozkład prawdopodobieństwa; mówimy wtedy o *hipotezie statystycznej* i o *weryfikacji hipotez statystycznych*. Postępowanie testowe zilustrujemy w naszym wykładzie kilkoma przykładami.

Już na pierwszy rzut oka widać, że istnieje pewna dowolność w wyborze statystyki T i obszaru krytycznego \mathbf{K} , czyli testu statystycznego; jest to poniekąd konsekwencją pewnej dowolności w wyborze tych doświadczeń, które mają falsyfikować rozważaną teorię. Statystyka matematyczna wykorzystuje tę dowolność w ten sposób, że buduje testy "jak najlepsze". Dokładne sformułowanie tego problemu i sugestie jego rozwiązania w ramach teorii Neymana–Pearsona przedstawimy w paragrafie 6. Wybór statystyki T i konstrukcję zdarzeń $\{T > t_\alpha\}$ dla pewnych najczęściej spotykanych w zastosowaniach sytuacji prezentujemy w paragrafach 2–5. Ponieważ weryfikowaną hipotezę "odrzuca się", gdy $T > t_\alpha$, co formułuje się zwykle słowami "gdy T jest istotnie duże", odpowiednie procedury postępowania nazywa się *testami istotności*.

Model probabilistyczny $H = (\Omega, \mathcal{A}, P)$, o którym wyżej mówiliśmy, specyfikuje pewną przestrzeń probabilistyczną. Mówimy wtedy o *hipotezie prostej*. Przykładem jest hipoteza, że w danym ciągu doświadczeń Bernoulliego prawdopodobieństwo sukcesu jest równe pewnej ustalonej liczbie. Mówimy o *hipotezie złożonej*, gdy weryfikowana teoria specyfikuje rodzinę modeli. Na przykład hipotezą złożoną jest hipoteza, według której prawdopodobieństwo sukcesu w danym ciągu doświadczeń Bernoulliego nie przekracza danej liczby z przedziału $(0, 1)$. Innym przykładem hipotezy złożonej jest hipoteza głosząca, że dany proces fizyczny (np. proces rozpadu promieniotwórczego) jest jakimś procesem losowym Poissona. W naszym wykładzie mówimy o testowaniu zarówno hipotez prostych, jak i hipotez złożonych.

2. Test zgodności Kołmogorowa

2.1. Oznaczenia

Oznaczamy przez \mathcal{F} rodzinę wszystkich ciągłych dystrybuant na prostej oraz przez P_F rozkład prawdopodobieństwa o dystrybuancie F . Rozważamy model statystyczny $(\mathbf{R}^1, \{P_F : F \in \mathcal{F}\})^n$. Będziemy korzystali z następujących faktów:

- jeżeli zmienna losowa X ma rozkład P_F , to zmienna losowa $F(X)$ ma rozkład jednostajny $U(0, 1)$;
- jeżeli X_1, X_2, \dots, X_n jest próbą z rozkładu P_F , to

$$P_F\{X_{j:n} \leq x\} = \sum_{i=j}^n \binom{n}{i} F^i(x) [1 - F(x)]^{n-i};$$

- jeżeli $X_{j:n}$ jest statystyką pozycyjną z próby X_1, X_2, \dots, X_n z rozkładu P_F , to $F(X_{j:n})$ jest statystyką pozycyjną z próby U_1, U_2, \dots, U_n z rozkładu $U(0, 1)$.

Mówiąc ogólnie i na razie niezbyt precyzyjnie, *test zgodności Kołmogorowa* (krótko: *test Kołmogorowa*) służy do weryfikacji hipotezy, że rozważana zmienna losowa X ma rozkład o danej ciągłej dystrybuancie F , przy czym statystyka testu jest oparta na różnicy między hipotetyczną dystrybuantą F a dystrybuantą empiryczną z próby X_1, X_2, \dots, X_n . W zależności od sposobu sformułowania weryfikowanej hipotezy, wyróżnia się kilka wersji tego testu.

W całym wykładzie zakładamy, że próba X_1, X_2, \dots, X_n pochodzi z pewnego rozkładu P_G o nieznannej dystrybuancie $G \in \mathcal{F}$. Dystrybuantę empiryczną oznaczamy przez G_n .

2.2. Hipoteza prosta

Niech $F \in \mathcal{F}$ będzie ustaloną dystrybuantą. Zadanie polega na zweryfikowaniu hipotezy

$$(1) \quad H: G = F.$$

Weźmy pod uwagę *statystykę Kołmogorowa*

$$(2) \quad D_n = \sup_{-\infty < x < \infty} |G_n(x) - F(x)|.$$

Jest oczywiste, że duże wartości tej statystyki przemawiają przeciwko hipotezie H . Wobec tego konstrukcja testu sprowadza się do wyznaczenia, dla zadanego poziomu istotności α , takiej (najmniejszej) liczby $D_n(\alpha)$, dla której

$$P_F\{D_n > D_n(\alpha)\} \leq \alpha.$$

Wartość krytyczna $D_n(\alpha)$ jest więc wybrana tak, że "jeżeli hipoteza H jest prawdziwa" (tzn. jeżeli próba X_1, X_2, \dots, X_n rzeczywiście pochodzi z rozkładu P_F), to prawdopodobieństwo odrzucenia tej hipotezy nie przekracza z góry zadanej (małej) liczby α (jeżeli hipoteza H jest prawdziwa, to zdarzenie losowe $\{D_n > D_n(\alpha)\}$ jest "praktycznie niemożliwe").

Obliczenie, dla danych α , n oraz F , wartości krytycznych $D_n(\alpha)$ nie następuje za pomocą trudności. Ponadto okazuje się, że $D_n(\alpha)$ może być ustalone uniwersalnie, niezależnie od hipotetycznego rozkładu F . Wynika to stąd, że

$$\begin{aligned}
P_F\{D_n > D_n(\alpha)\} &= P_F\left\{\sup_{-\infty < x < \infty} |G_n(x) - F(x)| > D_n(\alpha)\right\} \\
&= P_F\left\{\sup_{-\infty < x < \infty} \left|\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{[X_{j:n}, \infty)}(x) - F(x)\right| > D_n(\alpha)\right\} \\
&= P_F\left\{\sup_{0 < u < 1} \left|\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{[X_{j:n}, \infty)}(F^{-1}(u)) - u\right| > D_n(\alpha)\right\} \\
&= P_F\left\{\sup_{0 < u < 1} \left|\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{[F(X_{j:n}), 1)}(u) - u\right| > D_n(\alpha)\right\} \\
&= P_{U(0,1)}\left\{\sup_{0 < u < 1} \left|\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{[U_{j:n}, 1)}(u) - u\right| > D_n(\alpha)\right\} \\
&= P_{U(0,1)}\left\{\sup_{0 < u < 1} |G_n(u) - u| > D_n(\alpha)\right\}.
\end{aligned}$$

Wartości krytyczne $D_n(\alpha)$ zostały stabilizowane, a obszerne tablice są łatwo dostępne. Praktyczne obliczenia wartości statystyki D_n z próby X_1, X_2, \dots, X_n opierają się na spostrzeżeniu, że $\sup_{-\infty < x < \infty} |G_n(x) - F(x)|$ realizuje się w jednym z punktów skoku dystrybuanty empirycznej G_n .

Ponieważ D_n nie zmienia się przy monotonicznych przekształceniach argumentu x , możemy wykonać te obliczenia według następujących łatwych wzorów:

$$(3) \quad D_n^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - z_i\right),$$

$$(4) \quad D_n^- = \max_{1 \leq i \leq n} \left(z_i - \frac{i-1}{n}\right),$$

$$(5) \quad D_n = \max\{D_n^+, D_n^-\},$$

gdzie

$$(6) \quad z_i = F(X_{i:n}).$$

2.3. Hipoteza złożona

2.3.1. Uwagi ogólne

Przypomnijmy, że hipoteza nazywa się hipotezą złożoną, gdy wyróżnia rodzinę rozkładów, większą niż jednoelementową. Przedstawimy zagadnienie weryfikacji dwóch rodzajów hipotez złożonych; w każdej z rozważanych sytuacji hipoteza wyróżnia pewien podzbiór w rodzinie \mathcal{F} .

2.3.2. Hipoteza $H^- : G \leq F$

Niech, jak poprzednio, $F \in \mathcal{F}$ będzie ustaloną dystrybuantą i niech G będzie nieznaną dystrybuantą zmiennej losowej X . W licznych zastosowaniach praktycznych zadanie polega na weryfikacji hipotezy

$$(7) \quad H^- : G \leq F.$$

Teraz duże wartości statystyki

$$(8) \quad D_n^+ = \sup_{-\infty < x < \infty} (G_n(x) - F(x))$$

przemawiają przeciwko weryfikowanej hipotezie H^- . Konstrukcja testu sprowadza się, jak poprzednio, do wyznaczenia najmniejszej liczby $D_n^+(\alpha)$ (wartości krytycznej testu na poziomie istotności α) takiej, żeby

$$(9) \quad P_G\{D_n^+ > D_n^+(\alpha)\} \leq \alpha \quad (\forall G \leq F).$$

Znowu, jak poprzednio, warunek (9) oznacza, że prawdopodobieństwo odrzucenia weryfikowanej hipotezy H^- , gdy jest ona prawdziwa (tzn. gdy $G \leq F$), ma być nie większe niż z góry założona liczba α .

Zauważmy, że jeżeli $G \leq F$, to

$$\begin{aligned} \sup_{-\infty < x < \infty} (G_n(x) - F(x)) &\leq \sup_{-\infty < x < \infty} (G_n(x) - G(x)) + \sup_{-\infty < x < \infty} (G(x) - F(x)) \\ &\leq \sup_{-\infty < x < \infty} (G_n(x) - G(x)). \end{aligned}$$

A więc dla każdej liczby $y > 0$ oraz dla każdej dystrybuanty $G \leq F$ mamy

$$\begin{aligned} P_G\{D_n^+ > y\} &= P_G\left\{\sup_{-\infty < x < \infty} (G_n(x) - F(x)) > y\right\} \\ &\leq P_G\left\{\sup_{-\infty < x < \infty} (G_n(x) - G(x)) > y\right\} \\ &= P_F\left\{\sup_{-\infty < x < \infty} (G_n(x) - F(x)) > y\right\} \\ &= P_F\{D_n^+ > y\}. \end{aligned}$$

Nierówność

$$P_G\{D_n^+ > y\} \leq P_F\{D_n^+ > y\} \quad (\forall G \leq F)$$

jest dokładna.

Z powyższych rozważań wynika, że zadanie wyznaczenia wartości krytycznej testu redukuje się do wyznaczenia (najmniejszej) liczby $D_n^+(\alpha)$ takiej, że

$$(10) \quad P_F\{D_n^+ > D_n^+(\alpha)\} \leq \alpha.$$

Wartości te, podobnie jak w przypadku rozważanej poprzednio hipotezy prostej i statystyki D_n , nie zależą od F . W praktycznych zastosowaniach przyjmuje się przybliżenie

$$D_n(\alpha) \approx D_n^+\left(\frac{\alpha}{2}\right).$$

Dokładność tego przybliżenia jest tym lepsza, im α jest mniejsze; np. dla $\alpha \leq 0.2$ różnica obu wielkości nie przekracza 5×10^{-4} , a dla $\alpha \leq 0.1$ nie przekracza 5×10^{-5} .

Takie samo rozumowanie prowadzi do konstrukcji testu hipotezy $H^+ : G \geq F$. Odpowiednią statystyką jest tu $D_n^- = -\inf_{-\infty < x < \infty} (G_n(x) - F(x))$. Łatwo można sprawdzić, że jeżeli próba X_1, X_2, \dots, X_n pochodzi z rozkładu o dystrybuancie F , to statystyka D_n^- ma taki sam rozkład jak statystyka D_n^+ . Wynika stąd, że $D_n^-(\alpha) = D_n^+(\alpha)$.

2.3.3. Hipoteza o normalności rozkładu

Wiele zagadnień praktycznych prowadzi do weryfikacji hipotezy

$$(11) \quad H : \text{zmienna losowa } X \text{ ma rozkład normalny}.$$

Oczywiście w zastosowaniach zamiast "rozkład normalny" może pojawić się tu "rozkład wykładniczy", "rozkład Poissona" lub jakikolwiek inny rozkład. Rzecz w tym, że hipoteza specyfikuje tylko "kształt" rozkładu i że nie chodzi w niej o jakiś konkretny rozkład normalny $N(\mu, \sigma^2)$ z ustalonym parametrem (μ, σ^2) . Równoważnie hipoteza mogłaby brzmieć

$$H : \text{istnieją takie } \mu \in \mathbf{R}^1 \text{ oraz } \sigma > 0, \text{ że } X \text{ ma rozkład normalny } N(\mu, \sigma^2).$$

Jest to oczywiście hipoteza złożona. Test oparty na statystyce Kołmogorowa D_n powinien być zbudowany w taki sposób, żeby odrzucać hipotezę, gdy $D_n > D_n^N(\alpha)$, gdzie $D_n^N(\alpha)$ jest liczbą wyznaczoną w taki sposób, że

$$(12) \quad P_G\{D_n > D_n^N(\alpha)\} \leq \alpha \text{ dla każdego rozkładu normalnego } G.$$

Nie jest znane rozwiązanie tego zadania. W praktyce postępuje się w następujący sposób. Z próby X_1, X_2, \dots, X_n oblicza się średnią $\bar{X} = \sum_{j=1}^n X_j/n$ i wariancję $s^2 = \sum_{j=1}^n (X_j - \bar{X})^2/(n-1)$. Wartość statystyki D_n oblicza się według wzorów (3), (4), (5), przyjmując

$$z_i = \Phi\left(\frac{X_{i:n} - \bar{X}}{s}\right),$$

gdzie Φ jest dystrybuantą rozkładu normalnego $N(0, 1)$. Wartością krytyczną (na poziomie istotności α) dla statystyki $(\sqrt{n} - 0.01 + 0.85/\sqrt{n})D_n$ jest $D^N(\alpha)$ podane w następującej tabelce:

α	0.15	0.10	0.05	0.025	0.01
$D^N(\alpha)$	0.775	0.819	0.895	0.995	1.035

Te wartości krytyczne zostały obliczone metodą Monte Carlo. Również współczynnik $(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$ został wyznaczony empirycznie.

To co wyżej przedstawiliśmy ilustruje pewien sposób postępowania prowadzących do praktycznych rozwiązań trudnych i jeszcze teoretycznie nie w pełni rozwiązanych problemów konstrukcji procedur statystycznych.

3. Porównywanie średnich dwóch rozkładów normalnych

3.1. Sformułowanie zagadnienia

Rozpatrujemy następujące zagadnienie: na podstawie dwóch niezależnych prób losowych X_1, X_2, \dots, X_m oraz Y_1, Y_2, \dots, Y_n , pochodzących, odpowiednio, z rozkładów o dystrybuantach F i G i o średnich $E_F X = \mu$ oraz $E_G Y = \nu$, chcemy zweryfikować hipotezę $H: \mu = \nu$. Równolegle będziemy rozważali zagadnienie weryfikacji hipotezy $H^+: \mu \geq \nu$.

Rozpatrzmy kilka prostych modeli statystycznych, w których F i G są dystrybuantami rozkładów normalnych.

Jak zwykle przez \bar{X} i \bar{Y} oznaczamy, odpowiednio, średnie z prób X_1, X_2, \dots, X_m i Y_1, Y_2, \dots, Y_n .

3.2. Przypadek rozkładów normalnych o jednakowej wariancji

Jest naturalne, że duże wartości różnicy $|\bar{X} - \bar{Y}|$ przemawiają przeciwko hipotezie H , a duże wartości różnicy $\bar{Y} - \bar{X}$ przeciwko hipotezie H^+ . Ta uwaga może być punktem wyjścia do konstrukcji odpowiednich testów.

Rozpatrzmy najpierw przypadek, gdy wspólna wariancja σ^2 wszystkich zmiennych losowych jest znana. Wtedy \bar{X} ma rozkład normalny $N(\mu, \sigma^2/m)$, \bar{Y} ma rozkład normalny $N(\nu, \sigma^2/n)$ oraz $\bar{Y} - \bar{X}$ ma rozkład normalny $N(\nu - \mu, (\frac{1}{m} + \frac{1}{n})\sigma^2)$. Jeżeli hipoteza H jest prawdziwa, to $\bar{Y} - \bar{X}$ ma rozkład normalny $N(0, (\frac{1}{m} + \frac{1}{n})\sigma^2)$, więc obszarem krytycznym testu na poziomie istotności α jest

$$\{(\bar{X}, \bar{Y}): |\bar{X} - \bar{Y}| > \sigma \sqrt{\frac{1}{m} + \frac{1}{n}} \Phi^{-1}(1 - \frac{\alpha}{2})\}.$$

W przypadku hipotezy H^+ obszar krytyczny ma postać

$$\{(\bar{X}, \bar{Y}): \bar{Y} - \bar{X} > \sigma \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot z_\alpha\},$$

gdzie z_α jest wybrane tak, żeby

$$(13) \quad P_{\mu, \nu}\{\bar{Y} - \bar{X} > \sigma \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot z_\alpha\} \leq \alpha, \quad \text{gdy } \mu \geq \nu,$$

przy czym $P_{\mu, \nu}$ jest łącznym rozkładem prób $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$, gdy pierw-

sza próba pochodzi z rozkładu $N(\mu, \sigma^2)$, a druga z rozkładu $N(\nu, \sigma^2)$. Ponieważ

$$\begin{aligned} P_{\mu, \nu} \{ \bar{Y} - \bar{X} > \sigma \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot z_\alpha \} &= P_{\mu, \nu} \left\{ \frac{(\bar{Y} - \bar{X}) - (\nu - \mu)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} > z_\alpha - \frac{\nu - \mu}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \right\} \\ &= P_{\mu, \nu} \left\{ \xi > z_\alpha - \frac{\nu - \mu}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \right\}, \end{aligned}$$

gdzie ξ jest zmienną losową o rozkładzie normalnym $N(0, 1)$ przy każdym (μ, ν) , więc

$$\sup_{\mu, \nu: \mu \geq \nu} P_{\mu, \nu} \{ \bar{Y} - \bar{X} > \sigma \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot z_\alpha \} = P_{N(0,1)} \{ \xi > z_\alpha \}$$

i warunek (13) będzie spełniony, gdy przyjmiemy $z_\alpha = \Phi^{-1}(1 - \alpha)$.

Gdy wspólna wariancja σ^2 nie jest znana, analogiczna konstrukcja testów opiera się na fakcie, że w przypadku gdy hipoteza H jest prawdziwa, statystyka

$$(14) \quad \frac{\bar{Y} - \bar{X}}{\sqrt{\sum (X_j - \bar{X})^2 + \sum (Y_j - \bar{Y})^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

ma rozkład t Studenta o $m+n-2$ stopniach swobody.

3.3. Przypadek dowolnych rozkładów normalnych

Przypadek, gdy wariancje rozkładów, z których pochodzą próby X_1, X_2, \dots, X_m oraz Y_1, Y_2, \dots, Y_n , są znane, chociaż być może różne, jest trywialny: wystarczy zauważyć, że jeżeli X_1, X_2, \dots, X_m jest próbą z rozkładu $N(\mu, \sigma^2)$ oraz Y_1, Y_2, \dots, Y_n jest próbą z rozkładu $N(\nu, \tau^2)$, to różnica $\bar{Y} - \bar{X}$ ma rozkład normalny $N(\nu - \mu, \frac{\sigma^2}{m} + \frac{\tau^2}{n})$.

Przypadek, gdy wariancje nie są znane, jest skomplikowany (w literaturze znany jest jako *zagadnienie Behrensa-Fishera*); problem konstrukcji odpowiednich testów dla hipotezy H lub H^+ nie jest w pełni rozwiązany. Typowy sposób weryfikacji hipotezy H przebiega według następującego algorytmu. Niech

$$s_x^2 = \sum_{j=1}^m (X_j - \bar{X})^2 / (m-1), \quad s_y^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2 / (n-1) \quad \text{oraz} \quad c = \frac{s_x^2/m}{s_x^2/m + s_y^2/n}.$$

Hipotezę $H : \mu = \nu$ odrzuca się, gdy

$$\left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \right| > V(c, m-1, n-1, \alpha),$$

gdzie α jest poziomem istotności testu oraz $V(c, m-1, n-1, \alpha)$ są wartościami krytycznymi, stabilizowanymi tak, żeby prawdopodobieństwo odrzucenia H , gdy jest ona prawdziwa, nie przekraczało α . Nie będziemy tego zagadnienia tutaj szerzej rozwijali. (Tablice wartości krytycznych $V(c, m-1, n-1, \alpha)$ są łatwo dostępne.)

4. Hipoteza o parametrze położenia

Celem tej części wykładu jest przedstawienie *testów nieparametrycznych* na przykładzie najsłynniejszego chyba przedstawiciela tej grupy testów, a mianowicie na przykładzie *testu Manna–Whitneya–Wilcozona (MWW)*. Ogólnie mówiąc, *nieparametrycznym modelem statystycznym* nazywamy taki model statystyczny, w którym nie istnieje skończenie wymiarowa parametryzacja rodziny rozkładów prawdopodobieństwa, tzn. parametryzacja za pomocą pewnego $\theta \in \Theta \subset \mathbf{R}^k$ dla pewnego naturalnego k . Na przykład modelem nieparametrycznym jest model statystyczny $(\mathbf{R}^1, \{P_F : F \in \mathcal{F}\})$, w którym P_F jest rozkładem prawdopodobieństwa o dystrybuancie F oraz \mathcal{F} jest rodziną dystrybuant ciągłych.

Prezentowany dalej test *MWW* jest jednocześnie *testem permutacyjnym* lub ogólniej *testem kombinatorycznym*. Oto prosty przykład znakomicie ilustrujący ideę. Niech X_1, X_2, \dots, X_n będzie ciągiem zmiennych losowych o jednakowym rozkładzie z pewną ciągłą (nieznaną) dystrybuantą F . Testujemy hipotezę H , że jest to ciąg **niezależnych** zmiennych losowych. Jeżeli testowana hipoteza jest prawdziwa, to statystyka pozycyjna jest minimalną statystyką dostateczną. Rozkład warunkowy próby, pod warunkiem tej statystyki, nie zależy od nieznannej dystrybuanty F : każda permutacja statystyki pozycyjnej jest jednakowo prawdopodobna i jej prawdopodobieństwo (warunkowe) jest równe $(n!)^{-1}$. Jeżeli w eksperymencie losowym otrzymamy taki wynik X_1, X_2, \dots, X_n , że $X_1 < X_2 < \dots < X_n$, to skłonni będziemy zakwestionować hipotezę H , skłonni tym bardziej, im większe będzie n . Jeżeli $(n!)^{-1} \leq \alpha$, to po zaobserwowaniu takiego wyniku doświadczenia, na poziomie istotności α odrzucimy weryfikowaną hipotezę H .

Po tych wstępnych uwagach przechodzimy do prezentacji testu *MWW* dla weryfikacji hipotezy o parametrze położenia w dwóch próbach.

Przypuśćmy, że próba X_1, X_2, \dots, X_m pochodzi z rozkładu o dystrybuancie F_μ oraz próba Y_1, Y_2, \dots, Y_n z rozkładu o dystrybuancie F_ν , przy czym

$$F_\mu(x) = F(x - \mu), \quad F_\nu(y) = F(y - \nu),$$

gdzie $F \in \mathcal{F}$ jest pewną (nieznaną) ciągłą dystrybuantą. Rozważamy zagadnienie weryfikacji hipotezy $H: \mu = \nu$ lub, odpowiednio, hipotezy $H^+: \mu \geq \nu$.

Oparcie testu na różnicy średnich z prób jest teraz nie tylko teoretycznie nieuzasadnione (bo, nie zakładając, że rozkład F ma w ogóle wartość oczekiwaną, nie wiadomo, co takie średnie reprezentują), ale również niepraktyczne, bo rozkład różnicy $\bar{X} - \bar{Y}$ musi zależeć od rozkładu F (a więc również wartość krytyczna testu zależałaby od F), a ten rozkład jest z założenia nie znany.

Pewien pomysł weryfikacji rozważanych hipotez jest zawarty w tzw. *teście Manna–Whitneya–Wilcozona (MWW)* (zwanym również *testem Wilcozona* lub, w pewnej odmianie, *testem Manna–Whitneya*).

Uporządkujmy obie próby X_1, X_2, \dots, X_m oraz Y_1, Y_2, \dots, Y_n w jeden ciąg niemalejący (ze względu na założenie o ciągłości F , z prawdopodobieństwem 1 jest to ciąg ściśle rosnący). Niech R_1, R_2, \dots, R_m oraz S_1, S_2, \dots, S_n będą *rangami* (kolejnymi numerami), odpowiednio, obserwacji X_1, X_2, \dots, X_m oraz Y_1, Y_2, \dots, Y_n w tej połączonej próbie. Jest oczywiste, że duże wartości statystyki $W = \sum_{j=1}^n S_j$ (jest to tzw. "suma rang Y -ków") świadczą przeciwko hipotezie H^+ .

Rozpatrzmy najpierw przypadek testowania hipotezy H^+ . Zgodnie z ogólnymi zasadami, zadanie polega na wyznaczeniu takiej wartości krytycznej $w_{m,n}(\alpha)$, żeby

$$(15) \quad P_{\mu,\nu}\{W > w_{m,n}(\alpha)\} \leq \alpha \quad \text{dla wszystkich } (\mu, \nu) \text{ takich, że } \mu \geq \nu.$$

Zauważmy, że

— dla każdej liczby $c \in \mathbf{R}^1$

$$\begin{aligned} W(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n) &= \\ &= W(X_1 + c, X_2 + c, \dots, X_m + c, Y_1 + c, Y_2 + c, \dots, Y_n + c); \end{aligned}$$

— dla każdej liczby $c \geq 0$

$$\begin{aligned} W(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n) &\leq \\ &\leq W(X_1, X_2, \dots, X_m, Y_1 + c, Y_2 + c, \dots, Y_n + c). \end{aligned}$$

Zatem, jeżeli $\mu \geq \nu$, to

$$\begin{aligned} P_{\mu,\nu}\{W(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n) > w\} &\leq \\ &\leq P_{\mu,\nu}\{W(X_1, X_2, \dots, X_m, Y_1 + (\mu - \nu), Y_2 + (\mu - \nu), \dots, Y_n + (\mu - \nu)) > w\}. \end{aligned}$$

Ale jeżeli Y_1, Y_2, \dots, Y_n jest próbą z rozkładu F_ν , to $Y_1 + (\mu - \nu), Y_2 + (\mu - \nu), \dots, Y_n + (\mu - \nu)$ jest próbą z rozkładu F_μ , czyli

$$P_{\mu,\nu}\{W > w\} \leq P_{\mu,\mu}\{W > w\} \quad \text{dla wszystkich } (\mu, \nu) \text{ takich, że } \mu \geq \nu.$$

Zatem warunek (15) można przepisać w postaci

$$(16) \quad P_{\mu,\mu}\{W > w\} \leq \alpha \quad \text{dla wszystkich } \mu \in \mathbf{R}^1.$$

Wyrażenie po lewej stronie nierówności (16) jest prawdopodobieństwem zdarzenia losowego $\{W > w\}$, gdy obie próby pochodzą z takiego samego rozkładu o pewnej ciągłej dystrybucji F ; oznaczmy to prawdopodobieństwo krótko przez P_F . Jest oczywiste, że najmniejsza liczba $w_{m,n}(\alpha)$, taka że

$$(17) \quad P_F\{W > w_{m,n}(\alpha)\} \leq \alpha,$$

jest wartością krytyczną testu hipotezy H^+ , na poziomie istotności α .

Wyznaczenie wartości $w_{m,n}(\alpha)$ we wzorze (17) nie nastęrcza trudności. Przede wszystkim zauważmy, że jeżeli obie próby X_1, X_2, \dots, X_m oraz Y_1, Y_2, \dots, Y_n pochodzą z tego samego rozkładu o ciągłej dystrybucji F , to rozkład statystyki W nie zależy od wyboru tego rozkładu F : każda kombinacja m obserwacji X_1, X_2, \dots, X_m i n obserwacji Y_1, Y_2, \dots, Y_n jest jednakowo prawdopodobna, a wszystkich takich kombinacji jest $\binom{m+n}{m}$. Jeżeli wśród nich jest $k(w)$ kombinacji takich, na których wartość statystyki W jest większa od w , to

$$(18) \quad P_F\{W > w\} = \frac{k(w)}{\binom{m+n}{m}}.$$

Ponieważ zasada konstrukcji wartości krytycznej $w_{m,n}(\alpha)$ jest wspólna dla wszystkich tzw. *testów kombinatorycznych* i ponieważ dla dokładnego uzyskania założonego poziomu istotności potrzebna jest dodatkowa *randomizacja*, wyjaśnimy wszystkie te kwestie przeprowadzając szczegółowo konstrukcję testu *MWW* dla przypadku $m = 4$, $n = 2$ oraz $\alpha = 0.2$. Wszystkich możliwych, jednakowo prawdopodobnych kombinacji mamy teraz 15; oto one (symbol "x" oznacza "jakąś obserwację z próby X_1, X_2, \dots, X_m ", symbol "y" oznacza jakąś obserwację z próby Y_1, Y_2, \dots, Y_n ", a ponadto przy każdej kombinacji podajemy wartość statystyki W):

1)	x	x	x	x	y	y	11
2)	x	x	x	y	x	y	10
3)	x	x	x	y	y	x	9
4)	x	x	y	x	x	y	9
5)	x	x	y	x	y	x	8
6)	x	y	x	x	x	y	8
7)	x	x	y	y	x	x	7
8)	x	y	x	x	y	x	7
9)	y	x	x	x	x	y	7
10)	x	y	x	y	x	x	6
11)	y	x	x	x	y	x	6
12)	x	y	y	x	x	x	5
13)	y	x	x	y	x	x	5
14)	y	x	y	x	x	x	4
15)	y	y	x	x	x	x	3

Mamy więc, na przykład, $P_F\{W = 6\} = \frac{2}{15}$, $P_F\{W \leq 3\} = \frac{1}{15}$, itp. Najmniejszą wartością $w_{4,2}(0.2)$, spełniającą warunek (17), jest 9; mamy bowiem $P_F\{W > 9\} = \frac{2}{15} = 0.13$, ale $P_F\{W > 8\} = \frac{4}{15} = 0.26$. Testem na poziomie istotności $\alpha = 0.2$ jest więc test, odrzucający hipotezę H^+ , gdy statystyka W przyjmie wartość większą od 9. Rozmiar tego testu wynosi 0.13. Obszar krytyczny tego testu składa się z takich prób X_1, X_2, \dots, X_m i Y_1, Y_2, \dots, Y_n , dla których $W \in \{10, 11\}$.

W praktyce wartości krytyczne odczytuje się z odpowiednich tablic lub pakietów komputerowych.

Test o rozmiarze dokładnie równym założonemu poziomowi istotności możemy skonstruować w ten sposób, że do obszaru krytycznego zaliczymy, oprócz punktów z przestrzeni próby, takich że $W \in \{10, 11\}$, także "częściowo" punkty, dla których $W = 9$. W tym celu wyznaczamy liczbę λ , taką że

$$P_F\{W > 9\} + \lambda \cdot P_F\{W = 9\} = 0.20.$$

W rozważanym przypadku jest to liczba $\lambda = 0.5$. Testowanie przebiega w następujący sposób: odrzucić H^+ , gdy $W = 10$ lub $W = 11$; gdy $W = 9$, rzucić monetą i zakwestionować H^+ , gdy w wyniku tego rzutu otrzymamy orła. Jest to przykład *testu randomizowanego*.

5. Porównanie k średnich (analiza wariancji)

Problem jest następujący. Na podstawie k prób

$$\begin{array}{cccc} X_{1,1}, & X_{1,2}, & \dots, & X_{1,n_1}, \\ X_{2,1}, & X_{2,2}, & \dots, & X_{2,n_2}, \\ \dots & \dots & \dots & \dots \\ X_{k,1}, & X_{k,2}, & \dots, & X_{k,n_k}, \end{array}$$

pochozących, odpowiednio, z rozkładów normalnych $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$, ..., $N(\mu_k, \sigma^2)$, należy zweryfikować hipotezę

$$(19) \quad H: \mu_1 = \mu_2 = \dots = \mu_k.$$

Taki problem pojawia się w zastosowaniach na przykład wtedy, gdy weryfikujemy hipotezę, że poziom jakiegoś wyróżnionego czynnika nie ma wpływu na poziom badanego zjawiska.

Niech

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}, \quad i = 1, 2, \dots, k,$$

będą średnimi z poszczególnych prób. Gdyby hipoteza H była prawdziwa, wszystkie średnie \bar{X}_i , $i = 1, 2, \dots, k$, byłyby mniej więcej takie same. Wydaje się więc rozsądnym przyjąć za statystykę testu jakąś miarę zróżnicowania tych średnich, np. ich wariancję proporcjonalną do $\sum_{i=1}^k (\bar{X}_i - \bar{X})^2$, gdzie \bar{X} jest np. średnią wszystkich średnich \bar{X}_i lub średnią wszystkich obserwacji. Statystyka testowa mogłaby mieć również postać $\sum_{i=1}^k |\bar{X}_i - \bar{X}|$, $\sum_{i=1}^k |\bar{X}_i - \text{mediana średnich } \bar{X}_i|$, $\max\{\bar{X}_i : i = 1, 2, \dots, k\} - \min\{\bar{X}_i : i = 1, 2, \dots, k\}$, itp. Za wyborem pierwszej z tych statystyk przemawia to, że jest to pewna forma kwadratowa obserwacji, więc jej rozkład powinien być jakimś rozkładem chi-kwadrat; gdyby taki fakt rzeczywiście miał miejsce, ułatwiłoby to operowanie proponowaną statystyką.

Następujące twierdzenie stanowi podstawę teoretyczną dla konstrukcji odpowiedniego testu (przez $\text{rz}\mathbf{A}$ lub $\text{rz}(\mathbf{A})$ oznaczamy rząd macierzy \mathbf{A}).

TWIERDZENIE 1 (Cochrana–Fishera). Niech $\mathbf{Z} = (Z_1, Z_2, \dots, Z_N)^T$ będzie wektorem losowym o rozkładzie normalnym $N(\mathbf{0}, \mathbf{I})$, gdzie \mathbf{I} jest macierzą identycznościową. Niech, dla $i = 1, 2, \dots, k$, $\mathbf{Z}^T \mathbf{A}_i \mathbf{Z}$ będą formami kwadratowymi, takimi że $\text{rz}(\mathbf{A}_i) = n_i$ i niech $\mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^k \mathbf{Z}^T \mathbf{A}_i \mathbf{Z}$. Wówczas: zmienne losowe $\mathbf{Z}^T \mathbf{A}_i \mathbf{Z}$, $i = 1, 2, \dots, k$, są niezależne i mają, odpowiednio, rozkłady chi-kwadrat o n_i stopniach swobody wtedy i tylko wtedy, gdy $\sum_{i=1}^k n_i = N$.

D o w ó d. Jeżeli zmienne losowe $\mathbf{Z}^T \mathbf{A}_i \mathbf{Z}$, $i = 1, 2, \dots, k$, są niezależne i mają, odpowiednio, rozkłady chi-kwadrat o n_i stopniach swobody, to ich suma ma rozkład chi-kwadrat o $\sum_{i=1}^k n_i$ stopniach swobody. Ponieważ z założenia $\sum_{i=1}^k \mathbf{Z}^T \mathbf{A}_i \mathbf{Z} = \mathbf{Z}^T \mathbf{Z}$, a zmienna losowa $\mathbf{Z}^T \mathbf{Z}$ ma rozkład chi-kwadrat o N stopniach swobody, więc $\sum_{i=1}^k n_i = N$.

Przypuśćmy, że $\sum_{i=1}^k n_i = N$.

Ponieważ $\text{rz} \mathbf{A}_i = n_i$, więc istnieje n_i niezależnych funkcji liniowych

$$b_{1,1}Z_1 + \dots + b_{1,N}Z_N$$

...

$$b_{n_i,1}Z_1 + \dots + b_{n_i,N}Z_N$$

takich, że

$$\mathbf{Z}^T \mathbf{A}_i \mathbf{Z} = \pm (b_{1,1}Z_1 + \dots + b_{1,N}Z_N)^2 \pm \dots \pm (b_{n_i,1}Z_1 + \dots + b_{n_i,N}Z_N)^2.$$

Niech $\mathbf{B} = (b_{i,j})_{i,j=1,2,\dots,N}$. Mamy

$$\sum_{i=1}^k \mathbf{Z}^T \mathbf{A}_i \mathbf{Z} = \mathbf{Z}^T \mathbf{B}^T \mathbf{\Delta} \mathbf{B} \mathbf{Z},$$

gdzie $\mathbf{\Delta}$ jest macierzą diagonalną o wartościach $+1$ lub -1 na przekątnej. Z założenia mamy

$$\mathbf{Z}^T \mathbf{Z} = \mathbf{Z}^T \mathbf{B}^T \mathbf{\Delta} \mathbf{B} \mathbf{Z},$$

gdzie wektor \mathbf{Z} ma rozkład normalny $N(\mathbf{0}, \mathbf{I})$, a więc równość

$$\mathbf{z}^T \mathbf{z} = \mathbf{z}^T \mathbf{B}^T \mathbf{\Delta} \mathbf{B} \mathbf{z}$$

zachodzi dla (prawie) wszystkich wektorów $\mathbf{z} \in \mathbf{R}^N$. Może tak być wtedy i tylko wtedy, gdy

$$\mathbf{B}^T \mathbf{\Delta} \mathbf{B} = \mathbf{I}.$$

Stąd wynika, że $\text{rz}(\mathbf{B}) = N$. Zatem $\mathbf{\Delta} = (\mathbf{B}^T)^{-1} \mathbf{B}^{-1} = (\mathbf{B} \mathbf{B}^T)^{-1}$. Macierz $\mathbf{B} \mathbf{B}^T$, rzędu N , jest dodatnio określona, więc wszystkie elementy diagonalne macierzy $\mathbf{\Delta}$ są równe $+1$, czyli $\mathbf{B} \mathbf{B}^T = \mathbf{I}$, skąd wynika, że \mathbf{B} jest macierzą ortonormalną. Zatem składowe V_1, V_2, \dots, V_N wektora $\mathbf{V} = \mathbf{B} \mathbf{X}$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, 1)$. Ponieważ z konstrukcji

$$\mathbf{Z}^T \mathbf{A}_1 \mathbf{Z} = V_1^2 + \dots + V_{n_1}^2$$

$$\mathbf{Z}^T \mathbf{A}_2 \mathbf{Z} = V_{n_1+1}^2 + \dots + V_{n_1+n_2}^2$$

...

więc $\mathbf{Z}^T \mathbf{A}_i \mathbf{Z}$, $i = 1, 2, \dots, k$, są niezależne i mają, odpowiednio, rozkłady chi-kwadrat o n_i stopniach swobody. \square

Wykorzystamy teraz to twierdzenie w rozważanym przez nas problemie porównania k średnich.

Niech $Y_{i,j} = (X_{i,j} - \mu_i)/\sigma$ i niech $\mathbf{Y} = (Y_{1,1}, Y_{1,2}, \dots, Y_{1,n_1}, Y_{2,1}, \dots, Y_{k,n_k})^T$. Niech $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{i,j}$, $i = 1, 2, \dots, k$. Niech $\bar{Y} = \sum_{i=1}^k n_i \bar{Y}_i / N = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{i,j} / N$. Oczywiście \mathbf{Y} ma rozkład normalny $N(\mathbf{0}, \mathbf{I})$. Ponieważ

$$\mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{i,j}^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i + \bar{Y}_i - \bar{Y} + \bar{Y})^2,$$

więc

$$(20) \quad \mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2 + \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 + N \cdot \bar{Y}^2.$$

Mamy

$$\sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2 = \mathbf{Y}^T \mathbf{B}_i^T \mathbf{B}_i \mathbf{Y},$$

gdzie \mathbf{B}_i jest macierzą stopnia $N \times N$ o wyrazach, które dla $p, q = 1, 2, \dots, n_i$ są równe

$$b_{n_1+n_2+\dots+n_{i-1}+p, n_1+n_2+\dots+n_{i-1}+q} = \begin{cases} 1 - \frac{1}{n_i}, & \text{gdy } p = q, \\ -\frac{1}{n_i}, & \text{gdy } p \neq q, \end{cases}$$

i są równe zero poza tym. Łatwo obliczamy, że $\text{rz}(\mathbf{B}_i) = n_i - 1$. Zatem pierwszy składnik po prawej stronie wzoru (20) jest sumą k form kwadratowych rzędu, odpowiednio, $n_1 - 1, n_2 - 1, \dots, n_k - 1$.

Drugi składnik po prawej stronie wzoru (20) jest kwadratem długości wektora

$$(\sqrt{n_1}(\bar{Y}_1 - \bar{Y}), \sqrt{n_2}(\bar{Y}_2 - \bar{Y}), \dots, \sqrt{n_k}(\bar{Y}_k - \bar{Y}))^T = \mathbf{C} \mathbf{Y},$$

gdzie \mathbf{C} jest macierzą stopnia $k \times N$ o wyrazach

$$c_{p,q} = \begin{cases} \sqrt{n_p} \left(\frac{1}{n_p} - \frac{1}{N} \right), & \text{gdy } p = 1, 2, \dots, k; q = N_p + 1, N_p + 2, \dots, N_p + n_p, \\ -\frac{\sqrt{n_p}}{N}, & \text{w p.p.,} \end{cases}$$

przy czym $N_p = 0$ dla $p = 1$, $N_p = n_1 + n_2 + \dots + n_{p-1}$ dla $p = 2, 3, \dots, k$. Łatwo obliczamy, że $\text{rz}(\mathbf{C}) = k - 1$. Forma kwadratowa $N \cdot \bar{Y}^2$ jest oczywiście rzędu 1.

Otrzymaliśmy więc, że $\mathbf{Y}^T \mathbf{Y}$ (por. wzór (20)) jest sumą k form kwadratowych rzędu $n_1 - 1, n_2 - 1, \dots, n_k - 1$, formy kwadratowej rzędu $k - 1$ i formy kwadratowej rzędu 1. Ponieważ $\sum_{i=1}^k (n_i - 1) + (k - 1) + 1 = N$, na mocy twierdzenia Cochran–Fishera wnioskujemy, że te formy kwadratowe są niezależnymi zmiennymi losowymi i mają rozkłady chi-kwadrat z liczbami stopni swobody równymi ich rzędom. W szczególności

$$(21) \quad \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 = \frac{1}{\sigma^2} \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$$

ma rozkład chi-kwadrat o $k - 1$ stopniach swobody.

Jeżeli wariancja σ^2 jest znana, to wielkość po prawej stronie wzoru (21) jest statystyką z próby. Duże wartości tej statystyki świadczą przeciwko weryfikowanej hipotezie. Odpowiednią dla założonego poziomu istotności $\alpha \in (0, 1)$ wartość krytyczną tej statystyki znajdujemy w tablicach rozkładu chi-kwadrat o $k-1$ stopniach swobody.

Jeżeli wariancja σ^2 nie jest znana, wielkość (21) nie jest statystyką z próby. Zauważmy jednak, że – gdy weryfikowana hipoteza jest prawdziwa – pierwszy składnik po prawej stronie wzoru (20), a mianowicie

$$(22) \quad \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2 = \frac{1}{\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{i,j} - \bar{X})^2$$

jest zmienną losową o rozkładzie chi-kwadrat o $N-k$ stopniach swobody, niezależną od zmiennej losowej (21). Iloraz tych dwóch zmiennych losowych

$$(23) \quad F_{k-1, N-k} = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2 / (N-k)}$$

nie zależy od żadnego nieznanego parametru i jest statystyką o rozkładzie F Snedecora o (p, q) stopniach swobody (por. zadanie 13 w wykładzie 1). Test hipotezy H , na poziomie istotności α , odrzuca H , gdy $F_{k-1, N-k} > F(\alpha; k-1, N-k)$, gdzie $F(\alpha; p, q)$ oznacza kwantyl rzędu $(1-\alpha)$ rozkładu F Snedecora (tablice kwantyli tego rozkładu są łatwo dostępne).

Opisana wyżej procedura testowa jest pewnym szczególnym przypadkiem procedur rozważanych w tzw. *analizie wariancji*. Nazwa pochodzi stąd, że we wzorze (20) wariancja wszystkich obserwacji $X_{1,1}, X_{1,2}, \dots, X_{k,n_k}$, równa $\frac{1}{N} \mathbf{X}^T \mathbf{X} - \bar{X}^2$, zostaje rozłożona ("analiza") na sumę wariancji "wewnątrzpróbkowej" $\frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2$ i wariancji "międzypróbkowej" $\frac{1}{N} \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$. Z tego punktu widzenia skonstruowany test może być interpretowany również w następujący sposób: hipoteza H o równości średnich zostaje odrzucona, gdy "wariancja międzypróbkowa jest duża na tle wariancji wewnątrzpróbkowej".

6. Porównywanie testów. Teoria Neymana–Pearsona

6.1. Wprowadzenie

Rozpatrzmy następujące zadanie. Pewna zmienna losowa X ma rozkład normalny $N(\mu, 1)$ o nieznannej średniej μ i chcemy zweryfikować hipotezę $H: \mu = 0$. Przyjmijmy poziom istotności $\alpha = 0.05$.

Oto trzy różne testy tej hipotezy.

1. Weźmy pod uwagę statystykę $T_1(X) = X$. Duże dodatnie wartości tej statystyki oczywiście przeczą hipotezie. Łatwo sprawdzamy, że odpowiednia wartość krytyczna jest równa $t_1(\alpha) = \Phi^{-1}(1-\alpha) = 1.645$. Test oparty na statystyce T_1 odrzuca hipotezę H , gdy $T_1(X) > t_1(\alpha)$.

2. Weźmy pod uwagę statystykę $T_2(X) = |X|$. Teraz również jest oczywiste, że duże wartości tej statystyki kwestionują hipotezę H . Test odrzuca tę hipotezę, gdy $T_2(X) > t_2(\alpha)$, gdzie $t_2(\alpha) = \Phi^{-1}(1 - \frac{\alpha}{2}) = 1.960$.

3. Wykonajmy rzut regularną kostką dwudziestościanową o ścianach ponumerowanych kolejnymi liczbami $1, 2, \dots, 20$ i niech T_3 będzie liczbą zaobserwowaną w wyniku tego rzutu. Umówmy się, że odrzucamy H , gdy $T_3 = 1$ (w tym teście nie wykorzystujemy obserwacji X).

Wszystkie trzy testy są testami na poziomie istotności $\alpha = 0.05$, więc każdy z nich rozwiązuje postawione zadanie. Który z nich robi to lepiej?

Jest oczywiste, że bez wprowadzenia nowych elementów do naszych rozważań nie potrafimy odpowiedzieć na to pytanie. Oto jeden ze sposobów wprowadzenia tych nowych elementów.

Przypuśćmy, że rozważana zmienna losowa ma rozkład normalny $N(\mu, 1)$ z pewną średnią $\mu \neq 0$. Wtedy hipoteza H jest fałszywa. Każdy z testów T_1, T_2, T_3 odrzuca hipotezę H , gdy jest ona prawdziwa, z prawdopodobieństwem $\alpha = 0.05$. Naturalne byłoby uznać za lepszy ten z nich, który z większym prawdopodobieństwem odrzuca H , gdy $\mu \neq 0$ (tzn. gdy H jest fałszywa). Odpowiednie prawdopodobieństwa odrzucenia H przez testy T_1, T_2, T_3 w przypadku $\mu = 2$ wynoszą $0.639, 0.515$ oraz 0.05 , więc gdybyśmy chcieli z możliwie dużym prawdopodobieństwem odrzucać H , gdy $\mu = 2$, powinniśmy za najlepszy spośród trzech rozważanych testów uznać test oparty na statystyce T_1 . Powstaje pytanie, czy można skonstruować jeszcze lepszy test?

Odpowiednie prawdopodobieństwa dla $\mu = -2$ wynoszą $0.0013, 0.516$ oraz 0.05 . Teraz nasz wybór padłby na test T_2 . Czy można skonstruować test, który byłby jednocześnie najlepszy (tzn. najlepszy jednocześnie dla wszystkich $\mu \neq 0$)?

Sformułujmy to dokładniej i ogólniej.

Weryfikowana hipoteza H specyfikuje pewną rodzinę rozkładów $\{P_\theta : \theta \in \Theta_H\}$ zmiennej losowej X ; przestrzeń próby oznaczamy, jak zwykle, przez \mathcal{X} .

Test hipotezy H będziemy utożsamiali z funkcją $\phi : \mathcal{X} \rightarrow [0, 1]$ przy następującej interpretacji: jeżeli $\phi(x) = 1$, to zaobserwowanie wartości $x \in \mathcal{X}$ pociąga za sobą odrzucenie weryfikowanej hipotezy H ; zaobserwowanie wartości $x \in \mathcal{X}$ takiej, że $\phi(x) = 0$ nie daje podstaw do kwestionowania hipotezy; jeżeli $\phi(x) \in (0, 1)$, uruchamiamy dodatkowy mechanizm losowy, niezależny od X , i z prawdopodobieństwem $\phi(x)$ odrzucamy H . Tak skonstruowany test ϕ nazywamy *testem randomizowanym*. Test ϕ o wartościach w zbiorze $\{0, 1\}$ nazywamy *testem nierandomizowanym* lub po prostu *testem*; wtedy zbiór $\{x \in \mathcal{X} : \phi(x) = 1\}$ jest *zbiorem krytycznym* (lub *obszarem krytycznym*) testu.

Ustalmy liczbę $\alpha \in (0, 1)$. Testem hipotezy H na poziomie istotności α nazywamy każdy test ϕ taki, że

$$(24) \quad E_\theta \phi(X) \leq \alpha \quad (\forall \theta \in \Theta_H).$$

W przypadku testu nierandomizowanego $E_\theta \phi(X)$ jest po prostu prawdopodobieństwem odrzucenia hipotezy H , a warunek (24) orzeka, że to prawdopodobieństwo ma być nie większe niż α , "gdy hipoteza H jest prawdziwa". Liczba $\sup\{E_\theta \phi(X) : \theta \in \Theta_H\}$ nazywa się *rozmiarem testu*.

Błąd polegający na odrzuceniu hipotezy H , gdy jest ona prawdziwa, nazywa się *błędem pierwszego rodzaju*. Funkcja $E_\theta \phi(X)$ przypisuje każdemu $\theta \in \Theta_H$ prawdopodobieństwo błędu pierwszego rodzaju.

Weźmy pod uwagę jeszcze jedną rodzinę rozkładów $\{P_\theta : \theta \in \Theta_K\}$ na przestrzeni próby \mathcal{X} . Hipotezę $K : \theta \in \Theta_K$ będziemy nazywali *hipotezą konkurencyjną* lub *hipotezą alternatywną*.

Niech ϕ i ψ będą dwoma testami na poziomie istotności α . Powiemy, że test ϕ jest *mocniejszy* niż test ψ , jeżeli

$$(25) \quad \begin{cases} E_\theta \phi(X) \geq E_\theta \psi(X) & \forall \theta \in \Theta_K, \\ E_\theta \phi(X) > E_\theta \psi(X) & \text{dla pewnego } \theta \in \Theta_K. \end{cases}$$

Jest oczywiste, co przy takim porządkowaniu testów oznacza pojęcie *test najmocniejszy* lub *test jednostajnie najmocniejszy*. Dla testu jednostajnie najmocniejszego będziemy używali skrótu *test JNM*.

Funkcję $\Theta_K \ni \theta \rightarrow E_\theta \phi(X)$ nazywamy *funkcją mocy* lub krócej *mocą* testu ϕ . Wielkość $E_\theta \phi(X)$ opisuje prawdopodobieństwo odrzucenia weryfikowanej hipotezy H , gdy rozważana zmienna losowa ma faktycznie rozkład P_θ , $\theta \in \Theta_K$, a więc gdy hipoteza H jest fałszywa. Błąd polegający na nieodrżuceniu weryfikowanej hipotezy H , gdy jest ona fałszywa, nazywa się *błędem drugiego rodzaju*, a prawdopodobieństwo tego błędu jest równe $1 - E_\theta \phi(X)$, $\theta \in \Theta_K$.

Problem konstrukcji testu *JNM* może być efektywnie rozwiązany tylko w niewielu sytuacjach. Podstawowy lemat Neymana–Pearsona (paragraf 5.2) dotyczy sytuacji, gdy H i K są hipotezami prostymi. Stosunkowo łatwo można skonstruować test *JNM* dla tzw. hipotez jednostronnych w modelu statystycznym z rodziną rozkładów $\{P_\theta : \theta \in \mathbf{R}^1\}$ o monotonicznym ilorazie wiarygodności (paragraf 5.3). Czasami udaje się skonstruować testy *JNM* w pewnych węższych klasach testów, ale tego zagadnienia nie będziemy rozwijali.

W bardziej realistycznych (a więc i matematycznie bardziej złożonych) sytuacjach konstruuje się (już prawie od stu lat) różne testy na drodze rozważań heurystycznych (por. testy istotności w poprzednich paragrafach 2, 3 i 4). Takie testy poddaje się różnego rodzaju badaniom (np. symulacyjnym) i w końcu dla konkretnego zagadnienia wybiera się test, który wydaje się najlepszy. O konstrukcji testów opartych na koncepcji wiarygodności powiemy w wykładzie piątym.

6.2. Podstawowy lemat Neymana–Pearsona

Rozważamy najpierw przypadek testowania prostej hipotezy H przeciwko prostej hipotezie alternatywnej K .

TWIERDZENIE 2 (podstawowy lemat Neymana–Pearsona). *Niech P_0 i P_1 będą rozkładami prawdopodobieństwa i niech f_0 i f_1 będą gęstościami tych rozkładów (względem pewnej ustalonej miary μ). Niech $\alpha \in (0, 1)$ będzie ustaloną liczbą.*

(a) (Istnienie testu). *Istnieją stałe t i γ takie, że*

$$(26) \quad \phi(x) = \begin{cases} 1, & \text{gdy } f_1(x) > t f_0(x), \\ \gamma, & \text{gdy } f_1(x) = t f_0(x), \\ 0, & \text{gdy } f_1(x) < t f_0(x), \end{cases}$$

jest testem hipotezy $H : P_0$ przeciwko hipotezie $K : P_1$, na poziomie istotności α , tzn. testem takim, że

$$(27) \quad E_0\phi(X) = \alpha.$$

(b) (Dostateczność). Jeżeli test ϕ spełnia warunek (27) i dla pewnego t warunek

$$(28) \quad \phi(x) = \begin{cases} 1, & \text{gdy } f_1(x) > tf_0(x), \\ 0, & \text{gdy } f_1(x) < tf_0(x), \end{cases}$$

to ϕ jest testem najmocniejszym dla testowania H przeciwko K na poziomie istotności α .

(c) (Koniczność). Jeżeli ϕ jest testem najmocniejszym na poziomie istotności α dla testowania H przeciwko K , to dla pewnego t spełnia on warunek (28).

D o w ó d.

Ad (a). Zdefiniujmy funkcję

$$T(x) = \begin{cases} \frac{f_1(x)}{f_0(x)}, & \text{gdy } f_0(x) > 0, \\ +\infty, & \text{gdy } f_0(x) = 0, \end{cases}$$

i weźmy pod uwagę ogon dystrybuanty statystyki $T(X)$, gdy X ma rozkład P_0 :

$$\alpha(t) = P_0\{T(X) > t\}.$$

Z definicji zmiennej losowej T i z własności dystrybuanty wynika, że dla każdego $\alpha \in (0, 1)$ istnieje t^* takie, że

$$\alpha(t^*) \leq \alpha \leq \alpha(t^* - 0).$$

Ponieważ $\alpha(t^* - 0) = \alpha(t^*)$ wtedy i tylko wtedy, gdy $P_0\{T(X) = t^*\} = 0$, tzn. wtedy i tylko wtedy, gdy $P_0\{f_1(X) = t^*f_0(X)\} = 0$, więc możemy zdefiniować

$$\phi(x) = \begin{cases} 1, & \text{gdy } f_1(x) > t^*f_0(x), \\ \frac{\alpha - \alpha(t^*)}{\alpha(t^* - 0) - \alpha(t^*)}, & \text{gdy } f_1(x) = t^*f_0(x), \\ 0, & \text{gdy } f_1(x) < t^*f_0(x). \end{cases}$$

Ponieważ wtedy

$$\begin{aligned} E_0\phi(X) &= P_0\{f_1(X) > t^*f_0(X)\} + \frac{\alpha - \alpha(t^*)}{\alpha(t^* - 0) - \alpha(t^*)} P_0\{f_1(X) = t^*f_0(X)\} \\ &= \alpha(t^*) + \frac{\alpha - \alpha(t^*)}{\alpha(t^* - 0) - \alpha(t^*)} (\alpha(t^* - 0) - \alpha(t^*)) \\ &= \alpha \end{aligned}$$

zatem tak skonstruowany test ϕ spełnia warunki (26) i (27).

Przypomnijmy, że wartość t^* jest określona z warunku $\alpha(t^*) \leq \alpha \leq \alpha(t^* - 0)$. Jest ona określona jednoznacznie poza przypadkiem, gdy $\alpha(t) = \alpha$ na pewnym przedziale $[t', t'']$. Taki przedział można jednak wyłączyć z naszych rozważań z uzasadnieniem, że oba rozważane rozkłady P_0 i P_1 przypisują mu zerowe prawdopodobieństwo. Wynika to z następującego rozumowania. Zapiszmy warunek

$$P_0\{t' \leq T(X) \leq t''\} = 0$$

w postaci

$$\int_{\{x: t' \leq T(x) \leq t''\}} f_0(x) \mu(dx) = 0.$$

Funkcja $f_0(x)$ nie może być równa zeru na zbiorze całkowania, bo dla $f_0(x) = 0$ mamy $T(x) = +\infty$, więc przedział $[t', t'']$ musiałby mieć postać $[t', +\infty)$, a $\alpha(t)$ na takim przedziale, gdyby było stałe, musiałoby być równe zeru. Skoro $f_0(x) > 0$ na zbiorze całkowania, to $\mu\{x: t' \leq T(x) \leq t''\} = 0$, czyli również $P_1\{x: t' \leq T(x) \leq t''\} = 0$.

Ad (b). Niech ϕ będzie testem spełniającym (27) i (28) i niech ϕ^* będzie innym testem takim, że

$$E_0\phi^*(X) \leq \alpha.$$

Mamy

$$\begin{aligned} (29) \quad E_1\phi(X) - E_1\phi^*(X) &= \int (\phi(x) - \phi^*(x)) f_1(x) \mu(dx) = \\ &= \int_{\{x: \phi(x) > \phi^*(x)\}} (\phi(x) - \phi^*(x)) f_1(x) \mu(dx) + \\ &\quad + \int_{\{x: \phi(x) < \phi^*(x)\}} (\phi(x) - \phi^*(x)) f_1(x) \mu(dx). \end{aligned}$$

Jeżeli $\phi(x) > \phi^*(x)$, to musi być $\phi(x) > 0$, ale wtedy z definicji testu ϕ musi być $f_1(x) \geq t f_0(x)$, więc pierwsza całka po prawej stronie wzoru (29) jest nie mniejsza od

$$t \int_{\{x: \phi(x) > \phi^*(x)\}} (\phi(x) - \phi^*(x)) f_0(x) \mu(dx).$$

Analogicznie, na zbiorze $\{x: \phi(x) < \phi^*(x)\}$ musi być $\phi(x) < 1$, czyli $f_1(x) \leq t f_0(x)$. Wynika stąd, że druga z tych całek jest nie mniejsza od

$$t \int_{\{x: \phi(x) < \phi^*(x)\}} (\phi(x) - \phi^*(x)) f_0(x) \mu(dx),$$

czyli

$$\begin{aligned} E_1\phi(X) - E_1\phi^*(X) &\geq t \cdot \int (\phi(x) - \phi^*(x)) f_0(x) \mu(dx) \\ &= t \cdot (E_0\phi(X) - E_0\phi^*(X)) \\ &= t \cdot (\alpha - E_0\phi^*(X)) \geq 0. \end{aligned}$$

Zatem test ϕ jest co najmniej tak mocny, jak dowolny test ϕ^* .

Ad (c). Niech ϕ będzie testem najmocniejszym na poziomie α i niech ϕ^* będzie testem o tym samym rozmiarze, spełniającym warunek (28). Weźmy pod uwagę zbiór

$$C = \{x : \phi(x) \neq \phi^*(x), f_1(x) \neq tf_0(x)\}.$$

W celu udowodnienia tezy mamy wykazać, że $\mu(C) = 0$.

Rozumując jak w dowodzie tezy (b), stwierdzamy, że na zbiorze C iloczyn

$$(\phi^*(x) - \phi(x))(f_1(x) - tf_0(x))$$

jest dodatni (bo gdy $f_1 > tf_0$, wtedy z założenia mamy $\phi^* = 1$, więc musi być $\phi^*(x) - \phi(x) > 0$, a gdy $f_1 < tf_0$, wtedy $\phi^* = 0$ i musi być $\phi^*(x) - \phi(x) > 0$), czyli

$$\begin{aligned} \int (\phi^*(x) - \phi(x))(f_1(x) - tf_0(x))\mu(dx) &= \\ &= \int_C (\phi^*(x) - \phi(x))(f_1(x) - tf_0(x))\mu(dx) > 0, \quad \text{gdy } \mu(C) > 0. \end{aligned}$$

Ale wtedy

$$\int (\phi^*(x) - \phi(x))f_1(x)\mu(dx) > t \int (\phi^*(x) - \phi(x))f_0(x)\mu(dx) = 0,$$

skąd wynika, że ϕ^* jest mocniejszy od ϕ , wbrew założeniu, czyli musi być $\mu(C) = 0$. \square

Lemat Neymana–Pearsona sugeruje, że konstrukcja testu najmocniejszego powinna przebiegać w ten sposób, aby do obszaru krytycznego tego testu zaliczać punkty o największej wartości ilorazu $f_1(x)/f_0(x)$ dopóty, dopóki prawdopodobieństwo P_0 tego obszaru nie przekroczy założonego poziomu istotności. Ilustrujemy to następującym przykładem.

Przykład 1. W następującej tabelce podano gęstości dwóch rozkładów: rozkładu dwumianowego $B(10; 0.1)$ i Poissona $P(1)$:

x	$B(10; 0.1)$	$P(1)$	$\frac{P(1)}{B(10; 0.1)}$
0	0.34868	0.36788	1.05506
1	0.38742	0.36788	0.94956
2	0.19371	0.18394	0.94956
3	0.05739	0.06131	1.06830
4	0.01116	0.01533	1.37366
5	0.00149	0.00307	2.06040
6	0.00014	0.00051	3.64286
7	0.00001	0.00007	7
8	0.00000	0.00001	$+\infty$
...			

W celu weryfikacji hipotezy $H : B(10; 0.1)$ wobec hipotezy alternatywnej $K : P(1)$, do obszaru krytycznego bez wątpienia należy zaliczyć $\{8, 9, \dots\}$. Włączając kolejno dalsze punkty o największych wartościach ilorazu $P(1)/B(10; 0.1)$, otrzymujemy

obszar krytyczny \mathbf{K}	$P_H\{X \in \mathbf{K}\}$	$P_K\{X \in \mathbf{K}\}$
$\{x : x \geq 8\}$	0	0.00001
$\{x : x \geq 7\}$	0.00001	0.00008
$\{x : x \geq 6\}$	0.00015	0.00059
$\{x : x \geq 5\}$	0.00164	0.00366
$\{x : x \geq 4\}$	0.01280	0.01899
$\{x : x \geq 3\}$	0.07019	0.08030
$\{x : x \geq 3\} \cup \{x : x = 0\}$	0.41887	0.44818
\mathcal{X}	1	1

Test nierandomizowany na poziomie istotności $\alpha = 0.05$ ma postać

$$\phi(x) = \begin{cases} 1, & \text{gdy } x \geq 4, \\ 0, & \text{gdy } x < 4. \end{cases}$$

Rozmiar tego testu wynosi 0.01280.

Dobierając γ tak, żeby

$$P_H\{X \geq 4\} + \gamma \cdot P_H\{X = 3\} = 0.05,$$

czyli $\gamma = 0.6482$, otrzymujemy test na poziomie istotności 0.05 i o rozmiarze 0.05:

$$\phi(x) = \begin{cases} 1, & \text{gdy } x \geq 4, \\ 0.6482, & \text{gdy } x = 3, \\ 0, & \text{gdy } x \leq 2. \end{cases}$$

Jest to test najmocniejszy na poziomie istotności $\alpha = 0.05$, ale jego moc jest równa tylko 0.05873. Oznacza to, że prawdopodobieństwo nie odrzucenia weryfikowanej hipotezy $H : B(10; 0.1)$, gdy prawdziwa jest hipoteza alternatywna $K : P(1)$, wynosi aż 0.94127. ■

Ten przykład poucza, żeby pojęcia "test najmocniejszy" nie traktować zbyt optymistycznie: "najmocniejszy" nie musi bowiem oznaczać tego, co w potocznych zastosowaniach chcielibyśmy rozumieć jako "mocny" lub choćby "zadowolająco mocny".

Odnotujmy następujący wniosek z lematu Neymana–Pearsona.

WNIOSEK 1. Jeżeli β jest mocą najmocniejszego testu na poziomie $\alpha \in (0, 1)$ dla testowania $H : P_0$ wobec $K : P_1$, to $\beta > \alpha$, chyba że $P_0 = P_1$.

D o ó d. Test $\phi(x) \equiv \alpha$ jest testem na poziomie istotności α i moc tego testu jest równa α , a więc moc β testu najmocniejszego musi spełniać nierówność $\beta \geq \alpha$. Przypuśćmy, że $\beta = \alpha$. Wtedy test $\phi(x) \equiv \alpha$ jest testem najmocniejszym, a ponieważ jako test najmocniejszy musi spełniać warunek (28) więc, wobec tego że $\alpha \in (0, 1)$, musi być $f_0(x) = t f_1(x)$ dla pewnego t . Ale skoro f_0 i f_1 są gęstościami, to musi być $t = 1$, czyli $P_0 = P_1$. □

6.3. Testy JNM w modelach z monotonicznym ilorazem wiarygodności

Rozpatrujemy model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$, w którym Θ jest pewnym przedziałem na prostej. Zakładamy, że rozkłady P_θ mają gęstości p_θ (względem pewnej ustalonej miary μ). Mówimy, że $\{P_\theta : \theta \in \Theta\}$ jest *rodziną rozkładów z monotonicznym ilorazem wiarygodności*, jeżeli istnieje taka funkcja $T(x)$, że jeżeli $\theta' > \theta$, to iloraz $p_{\theta'}(x)/p_\theta(x)$ jest niemalejącą funkcją $T(x)$. Pojęcie "wiarygodność" komentujemy w wykładzie piątym, a na razie umówmy się, że "tak się mówi".

Przykład 2. Rodzina $\{p_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} : 0 \leq \theta \leq 1\}$ jest rodziną z monotonicznym ilorazem wiarygodności. ■

Przykład 3. Rodzina rozkładów jednostajnych $\{U(0, \theta) : \theta > 0\}$ jest rodziną z monotonicznym ilorazem wiarygodności. ■

Przykład 4. Ważną i obszerną klasę rodzin z monotonicznym ilorazem wiarygodności stanowią jednoparametrowe rodziny wykładnicze z gęstościami

$$p_\theta(x) = \exp\{c(\theta)T(x) - b(\theta)\} \cdot h(x). \quad \blacksquare$$

W modelach statystycznych z monotonicznym ilorazem wiarygodności istnieją testy *JNM* dla weryfikacji jednostronnych hipotez złożonych postaci $H : \theta \leq \theta_0$ wobec (złożonych) alternatyw $K : \theta > \theta_0$; charakteryzację takich testów podaje następujące twierdzenie.

TWIERDZENIE 3. Niech θ będzie parametrem rzeczywistym, a zmienna losowa X niech ma gęstość prawdopodobieństwa $p_\theta(x)$. Niech $\{p_\theta(x) : \theta \in \mathbf{R}^1\}$ będzie rodziną rozkładów z monotonicznym ilorazem wiarygodności względem $T(x)$. Niech $\alpha \in (0, 1)$ będzie ustalonym poziomem istotności.

(a) Dla testowania hipotezy $H : \theta \leq \theta_0$, przy alternatywie $K : \theta > \theta_0$, istnieje test *JNM*, określony w następujący sposób:

$$(30) \quad \phi(x) = \begin{cases} 1, & \text{gdy } T(x) > C, \\ \gamma, & \text{gdy } T(x) = C, \\ 0, & \text{gdy } T(x) < C, \end{cases}$$

gdzie stałe C i γ są wyznaczone z warunku

$$(31) \quad E_{\theta_0} \phi(X) = \alpha.$$

(b) Funkcja

$$(32) \quad \beta(\theta) = E_\theta \phi(X)$$

jest ściśle rosnąca w każdym punkcie θ , w którym $\beta(\theta) < 1$.

(c) Dla każdego θ' test określony warunkami (30) i (31) jest testem JNM dla testowania $H':\theta \leq \theta'$ przy $K':\theta > \theta'$, na poziomie istotności $\alpha' = \beta(\theta')$.

(d) Dla dowolnego $\theta < \theta_0$ test ten minimalizuje $\beta(\theta)$ (prawdopodobieństwo błędu pierwszego rodzaju) wśród wszystkich testów spełniających warunek (31).

D o w ó d.

Ad (a) i (b). Rozważmy najpierw prostą hipotezę $H_0:\theta = \theta_0$ i prostą alternatywę $H_1:\theta = \theta_1$, dla pewnego ustalonego $\theta_1 > \theta_0$.

Niech ϕ będzie testem najmocniejszym dla weryfikacji H_0 wobec H_1 . Na mocy podstawowego lematu Neymana–Pearsona ten test ma postać (26), gdzie $f_0 = p_{\theta_0}$ oraz $f_1 = p_{\theta_1}$. W rodzinie rozkładów z monotonicznym ilorazem wiarygodności warunek $p_{\theta_1}(x)/p_{\theta_0}(x) > t$ jest równoważny z warunkiem $T(x) > C$, przy odpowiedniej stałej C , a więc test ϕ ma postać (30) i spełnia warunek (31).

Weźmy dowolne $\theta', \theta'' \in \mathbf{R}^1$ ($\theta' < \theta''$) i niech $\alpha' = E_{\theta'}\phi(X)$. Z tezy (b) podstawowego lematu Neymana–Pearsona wynika, że ϕ jest również testem najmocniejszym dla testowania hipotezy $H':\theta = \theta'$ wobec $K':\theta = \theta''$ na poziomie istotności α' . Teraz teza (b) wynika z wniosku 1.

Oznaczmy przez \mathcal{K}_{θ_0} klasę wszystkich testów ψ , takich że

$$E_{\theta_0}\psi = \alpha,$$

oraz przez $\mathcal{K}_{\leq\theta_0}$ klasę wszystkich testów ψ , takich że

$$E_{\theta}\psi \leq \alpha \quad \text{dla } \theta \leq \theta_0.$$

Ze względu na monotoniczność funkcji mocy mamy

$$E_{\theta}\phi \leq E_{\theta_0}\phi \quad \text{dla } \theta \leq \theta_0,$$

więc test ϕ jest testem na poziomie istotności α także dla rozszerzonej hipotezy $H:\theta \leq \theta_0$. Z konstrukcji ϕ jest testem najmocniejszym dla weryfikacji $H_0:\theta = \theta_0$ wobec $H_1:\theta = \theta_1$, więc

$$E_{\theta_1}\phi \geq E_{\theta_1}\psi \quad \forall \psi \in \mathcal{K}_{\theta_0},$$

ale wobec tego, że $\mathcal{K}_{\leq\theta_0} \subset \mathcal{K}_{\theta_0}$, mamy również

$$E_{\theta_1}\phi \geq E_{\theta_1}\psi \quad \forall \psi \in \mathcal{K}_{\leq\theta_0},$$

więc ϕ jest testem najmocniejszym dla $H:\theta \leq \theta_0$ wobec $H_1:\theta = \theta_1$. Ponieważ ϕ nie zależy od wyboru $\theta_1 > \theta_0$, jest to test JNM dla $H:\theta \leq \theta_0$ wobec $K:\theta > \theta_0$.

Ad (c). Tezę (c) dowodzi się analogicznie, biorąc za punkt wyjścia hipotezę $\theta \leq \theta'$ wobec $\theta > \theta'$ i jako poziom istotności moc skonstruowanego wyżej testu w punkcie $\theta = \theta'$.

Ad (d). Wystarczy zauważyć, że $1 - \beta(\theta)$ jest mocą testu JNM hipotezy $H:\theta \geq \theta_0$ wobec $K:\theta < \theta_0$, na poziomie istotności $1 - \alpha$. \square

WNIOSEK 2. Jeżeli $\{P_\theta: \theta \in \Theta\}$ jest rodziną wykładniczą o gęstościach

$$p_\theta(x) = \exp\{c(\theta)T(x) - b(\theta)\}h(x)$$

i jeżeli $c(\theta)$ jest funkcją ściśle rosnącą, to test JNM hipotezy $H: \theta \leq \theta_0$ wobec $K: \theta > \theta_0$ ma postać

$$\phi(x) = \begin{cases} 1, & \text{gdy } T(x) > C, \\ \gamma, & \text{gdy } T(x) = C, \\ 0, & \text{gdy } T(x) < C, \end{cases}$$

gdzie C i γ są wyznaczone z warunku, że $E_{\theta_0}\phi(X) = \alpha$.

Jeżeli $c(\theta)$ jest funkcją ściśle malejącą, to w definicji testu ϕ znaki nierówności zmieniają się na przeciwnie. \square

Przykład 5. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, 1)$. Weryfikujemy hipotezę $H: \mu \leq 0$ wobec hipotezy alternatywnej $K: \mu > 0$, na poziomie istotności $\alpha \in (0, 1)$.

Ponieważ

$$f_\mu(x_1, x_2, \dots, x_n) = (2\pi)^{-\frac{n}{2}} \exp\left\{\mu \cdot \sum_{i=1}^n x_i - \frac{1}{2}\mu^2\right\} \cdot e^{-\sum x_i^2/2},$$

więc mamy do czynienia z jednoparametrową rodziną wykładniczą z funkcją $c(\mu) = \mu$ oraz $T(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i$. Zatem test JNM ma postać

$$\phi(x) = \begin{cases} 1, & \text{gdy } \sum_{i=1}^n x_i > C, \\ \gamma, & \text{gdy } \sum_{i=1}^n x_i = C, \\ 0, & \text{gdy } \sum_{i=1}^n x_i < C. \end{cases}$$

Ponieważ $P_\mu\{\sum_{i=1}^n X_i = C\} = 0$, test redukuje się do postaci

$$\phi(x) = \begin{cases} 1, & \text{gdy } \sum_{i=1}^n x_i \geq C, \\ 0, & \text{gdy } \sum_{i=1}^n x_i < C, \end{cases}$$

gdzie C wyznacza się z warunku, żeby $E_0\phi(X) = \alpha$.

Ponieważ $\sum_{i=1}^n X_i$ ma rozkład normalny $N(n\mu, n)$, więc

$$E_\mu\phi(X) = P_\mu\left\{\sum_{i=1}^n X_i > C\right\} = 1 - \Phi\left(\frac{C - n\mu}{\sqrt{n}}\right).$$

Z warunku $E_0\phi(X) = \alpha$ otrzymujemy $C = \sqrt{n}\Phi^{-1}(1 - \alpha)$. Dla mocy testu ϕ mamy

$$\beta(\mu) = 1 - \Phi(\Phi^{-1}(1 - \alpha) - \mu\sqrt{n}).$$

Jest to funkcja ściśle rosnąca, przy czym $\beta(0) = \alpha$, $\beta(\mu) \rightarrow 0$, gdy $\mu \rightarrow -\infty$, oraz $\beta(\mu) \rightarrow 1$, gdy $\mu \rightarrow +\infty$. \blacksquare

6.4. Przykład, gdy test JNM nie istnieje

Niech zmienna losowa X ma rozkład normalny $N(\mu, 1)$, $\mu \in \mathbf{R}^1$, i niech zadanie polega na weryfikacji hipotezy $H: \mu = 0$ wobec hipotezy alternatywnej $K: \mu \neq 0$. Rodzina $\{N(\mu, 1) : \mu \in \mathbf{R}^1\}$ jest rodziną z monotonicznym ilorazem wiarygodności. Test najmocniejszy na pewnej hipotezie alternatywnej $\mu > 0$ ma moc, która na każdej hipotezie alternatywnej $\mu < 0$ ma moc mniejszą od α , więc jest na każdej takiej hipotezie dominowany np. przez test stały $\phi \equiv \alpha$.

7. Zadania

1. Próba X_1, X_2, \dots, X_6 pochodzi z pewnego rozkładu o ciągłej dystrybuancie G . Testem Kołmogorowa weryfikujemy hipotezę, że G jest rozkładem wykładniczym o gęstości $f(x) = 2e^{-2x} \mathbf{1}_{(0, \infty)}(x)$. Podać dokładnie algorytm testowania tej hipotezy na poziomie istotności $\alpha = 0.01$ (wartość krytyczną testu odczytać w odpowiedniej tablicy).

2. Zmienna losowa X ma rozkład jednostajny $U(0, \theta)$, $\theta > 0$. Dla weryfikacji hipotezy $H: \theta \leq \theta_0$ skonstruować test istotności na poziomie istotności α , oparty na minimalnej statystyce dostatecznej z próby X_1, X_2, \dots, X_n .

3. Próba X_1, X_2, \dots, X_m pochodzi z pewnego rozkładu o dystrybuancie $F_\delta(x) = F(x - \delta)$, gdzie F jest pewną dystrybuantą ciągłą, a próba Y_1, Y_2, \dots, Y_n pochodzi z rozkładu o dystrybuancie F . Próby są niezależne. Hipotezę $H: \delta \leq 0$ weryfikujemy w następujący sposób. Zliczamy liczbę S elementów w próbie X_1, X_2, \dots, X_m , większych od wszystkich elementów próby Y_1, Y_2, \dots, Y_n . Hipotezę H kwestionujemy, gdy $S \geq s$, gdzie s jest wartością krytyczną testu wybraną tak, żeby test miał założony poziom istotności α . Wyznaczyć s . Podać test dla przypadku $m = n = 5$ oraz $\alpha = 0.01$.

4. Na podstawie pojedynczej obserwacji X weryfikuje się hipotezę $H: "X$ ma rozkład normalny $N(0, 1)"$ przeciwko hipotezie alternatywnej $K: "X$ ma rozkład Laplace'a o gęstości $\frac{1}{2}e^{-|x|}$, $x \in \mathbf{R}^1"$.

Skonstruować najmocniejszy test na poziomie istotności $\alpha \in (0, 1)$.

5. W celu zweryfikowania hipotezy, że nieznanne prawdopodobieństwo sukcesu jest mniejsze od $\frac{1}{2}$, wykonuje się 20 niezależnych prób i hipotezę odrzuca się, gdy liczba sukcesów jest większa lub równa 12. Wyznaczyć funkcję prawdopodobieństwa błędu pierwszego rodzaju i funkcję prawdopodobieństwa błędu drugiego rodzaju.

6. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, 1)$. Weryfikuje się hipotezę, że $\mu = 0$, za pomocą testu z obszarem krytycznym $\{(x_1, x_2, \dots, x_n) : \sqrt{n} |\frac{1}{n} \sum_{i=1}^n x_i| > 2\}$. Obliczyć rozmiar testu. Naszkicować wykres funkcji mocy tego testu dla $\mu \in \mathbf{R}^1$.

7. Jeżeli przestrzeń próby jest zbiorem euklidesowym, a rozkłady P_0 i P_1 mają gęstości względem miary Lebesgue'a, to przy każdym poziomie istotności $\alpha \in (0, 1)$, dla testowania $H: P_0$ wobec $K: P_1$ istnieje najmocniejszy test nierandomizowany. (Skorzystać z następującego lematu Halmosa: niech $f \geq 0$, $\int_A f(x) dx = a$; wtedy dla każdego $b \in [0, a]$ istnieje zbiór $B \subset A$ taki, że $\int_B f(x) dx = b$.)

8. Niech X_1, X_2, \dots, X_n będzie ciągiem niezależnych zmiennych losowych o rozkładach $N(\mu_i, 1)$, $\mu_i \in \mathbf{R}^1$, $i = 1, 2, \dots, n$. Weryfikuje się hipotezę H , że wszystkie μ_i są równe zero przy hipotezie alternatywnej, że dla pewnego $r \in \{1, 2, \dots, n\}$ zachodzi $\mu_i = \frac{1}{2}$, $i = 1, 2, \dots, r$ oraz $\mu_i = -\frac{1}{2}$, $i = r + 1, r + 2, \dots, n$. Wykazać, że najmocniejszy test na poziomie istotności $\alpha = 0.05$ ma obszar krytyczny

$$\{(x_1, x_2, \dots, x_n) : \sum_{i=1}^r x_i - \sum_{i=r+1}^n x_i > 1.645\sqrt{n}\}.$$

Jak duże musi być n , żeby moc tego testu była równa co najmniej 0.9?

9. Zmienna losowa X ma rozkład jednostajny $U(0, \theta)$. Skonstruować najmocniejszy test na poziomie istotności α dla weryfikacji hipotezy $H : \theta = \theta_0$ wobec hipotezy alternatywnej $K : \theta = \theta_1$ dla pewnych θ_0, θ_1 takich, że $\theta_0 < \theta_1$. Obliczyć moc tego testu. Czy jest to test JNM dla testowania $H : \theta \leq \theta_0$ przeciwko $K : \theta > \theta_0$?

10. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu jednostajnego $U(0, \theta)$. Wyznaczyć test JNM dla weryfikacji hipotezy $H : \theta = \theta_0$ wobec hipotezy alternatywnej $K : \theta \neq \theta_0$, na poziomie istotności $\alpha \in (0, 1)$.

11. Próba X_1, X_2, \dots, X_n pochodzi z pewnego rozkładu wykładniczego o gęstości $f_\theta(x) = \exp\{-(x-\theta)\} \mathbf{1}_{(\theta, \infty)}(x)$. Hipotezę $H : \theta \leq 1$ przy hipotezie alternatywnej $K : \theta > 1$ weryfikuje się za pomocą testu z obszarem krytycznym postaci $\{(x_1, x_2, \dots, x_n) : x_{1:n} > c\}$, $c \in \mathbf{R}^1$.

Wyznaczyć stałą c tak, żeby test miał rozmiar α . Sprawdzić, czy jest to test JNM. Naszkicować wykres funkcji mocy testu.

12. Niech X_1, X_2, \dots, X_n będzie próbą z pewnego rozkładu o gęstości $f_{a,b}(x) = ae^{-a(x-b)} \mathbf{1}_{(b, \infty)}(x)$, $a > 0, b \in \mathbf{R}^1$.

(a) Zakładając, że a jest daną stałą, skonstruować test JNM na poziomie istotności α dla weryfikowania hipotezy $H : b = b_0$ przeciwko $H : b \neq b_0$. (Wskazówka: jeżeli X ma rozkład o gęstości $f_{a,b}(x)$, to $Y = e^{-aX}$ ma rozkład jednostajny $U(0, e^{-ab})$.)

(b) Wyznaczyć test JNM na poziomie istotności α dla weryfikowania hipotezy $H : a = a_0, b = b_0$ przy alternatywie $K : a > a_0, b < b_0$. Wykonać szczegółową konstrukcję testu dla przypadku $n = 10$, $a_0 = 1$, $b_0 = 1$, $\alpha = 0.01$.

(Istnienie testu JNM w modelu dwuparametrowym jest czymś wyjątkowym.)

13. Niech X_1, X_2, \dots, X_n będzie próbą z pewnego rozkładu o gęstości $f_\theta(x) = (2\theta)^{-1} e^{-x/2\theta} \mathbf{1}_{(0, \infty)}(x)$, $\theta > 0$. Niech $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ będą statystykami pozycyjnymi z tej próby.

Przypuśćmy, że w wyniku obserwacji otrzymujemy najpierw $X_{1:n}$, następnie $X_{2:n}$, itd. i że obserwacje prowadzimy do momentu zaobserwowania $X_{r:n}$, gdzie r jest ustaloną liczbą. Na podstawie obserwacji $X_{1:n}, X_{2:n}, \dots, X_{r:n}$ weryfikujemy hipotezę $H : \theta \geq \theta_0$ przy hipotezie alternatywnej $K : \theta < \theta_0$, na poziomie istotności $\alpha \in (0, 1)$.

Niech $\theta_0 = 1000$ oraz $\alpha = 0.05$.

(a) Wyznaczyć obszar krytyczny testu przy $r = 4$ i znaleźć moc tego testu przy alternatywie $\theta_1 = 500$.

(b) Znaleźć wartość r potrzebną do uzyskania mocy ≥ 0.95 przy tej alternatywie.

Wykład V

WIAROGODNOŚĆ

1. Koncepcja

Rozpatrujemy model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$, w którym wszystkie rozkłady prawdopodobieństwa mają gęstości względem pewnej ustalonej miary. Przez $p_\theta(x)$ oznaczamy gęstość rozkładu P_θ .

Zauważmy, że $p_\theta(x)$ traktujemy tu jak funkcję argumentu $x \in \mathcal{X}$, przy ustalonej (nieznanej) wartości parametru $\theta \in \Theta$.

DEFINICJA 1. Dla ustalonego $x \in \mathcal{X}$, wielkość

$$L(\theta; x) = p_\theta(x), \quad \theta \in \Theta,$$

nazywamy *wiarogodnością* θ , gdy zaobserwowano x . □

Następujący przykład ilustruje intuicję związane z tym pojęciem.

Przykład 1. Niech $\mathcal{X} = \{0, 1, 2, \dots, n\}$ i niech $p_\theta(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$, $x \in \mathcal{X}$, $\theta \in [0, 1]$. Jest to model statystyczny ciągu n doświadczeń Bernoulliego z (nieznanym) prawdopodobieństwem sukcesu θ . Przy ustalonym $x \in \mathcal{X}$, funkcję

$$L(\theta; x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad \theta \in [0, 1],$$

można interpretować w następujący sposób. Przypuśćmy, że w wyniku powyższego doświadczenia Bernoulliego zaobserwowano x sukcesów. Gdyby (nieznana) wartość parametru θ była równa θ_1 , prawdopodobieństwo otrzymania takiego wyniku byłoby równe $L(\theta_1; x)$. Gdyby dla pewnej innej wartości θ_2 parametru θ było $L(\theta_2; x) > L(\theta_1; x)$, oznaczałoby to, że (zaobserwowany) wynik x jest przy $\theta = \theta_2$ bardziej prawdopodobny niż przy $\theta = \theta_1$. W tym sensie "bardziej wiarogodne" jest to, że nieznana wartość parametru θ , przy której zaobserwowano x , była równa θ_2 niż to, że była ona równa θ_1 . ■

Koncepcja wiarogodności była, w szczególnych przypadkach, używana już przez Gaussa. Ogólnie idea została sformułowana i rozwinięta w pierwszych dziesięcioleciach dwudziestego wieku przez R.A.Fishera.

2. Estymatory największej wiarygodności

2.1. Konstrukcja

Zasada największej wiarygodności głosi, że jeżeli wynikiem obserwacji jest X , to za estymator nieznannej wartości parametru θ należy przyjąć tę wartość θ , która maksymalizuje wiarygodność $L(\theta; X)$.

DEFINICJA 2. Jeżeli przy każdym ustalonym $x \in \mathcal{X}$, istnieje $\hat{\theta} = \hat{\theta}(x) \in \Theta$, takie że

$$L(\hat{\theta}; x) \geq L(\theta; x) \quad (\forall \theta \in \Theta),$$

to odwzorowanie $\hat{\theta}: \mathcal{X} \rightarrow \Theta$ nazywamy *estymatorem największej wiarygodności parametru θ* . □

Zamiast "estymator największej wiarygodności" będziemy pisali krótko *ENW*, a zamiast "estymator największej wiarygodności parametru θ " będziemy pisali *ENW* $[\theta]$.

Zauważmy, że jeżeli $h: \Theta \rightarrow \Theta$ jest odwzorowaniem wzajemnie jednoznaczny i jeżeli $\hat{\theta}$ jest *ENW* $[\theta]$, to $h(\hat{\theta})$ jest *ENW* $[h(\theta)]$. W związku z tym dla **dowolnego** odwzorowania h określonego na Θ **definiujemy** *ENW* $[h(\theta)]$ jako $h(\hat{\theta})$, gdzie $\hat{\theta}$ jest *ENW* $[\theta]$. Wszędzie dalej pojęcia *ENW* używamy w tak rozszerzonym sensie.

Przykład 2. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, \sigma^2)$ o nieznanym parametrze (μ, σ^2) . Wtedy

$$L(\mu, \sigma^2; x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n \left(\frac{x_j - \mu}{\sigma} \right)^2 \right\}$$

i *ENW* $[(\mu, \sigma^2)]$ jest $(\bar{X}, \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2)$. ■

Uwaga techniczna. Maksymalizacja L jest równoważna z maksymalizacją $l = \log L$. Uwzględnienie tego faktu często znakomicie upraszcza wyznaczanie *ENW*.

2.2. Błąd średniokwadratowy ENW

W naszych wykładach, za kryterium oceny jakości estymatorów przyjęliśmy błąd średniokwadratowy. Stwierdziliśmy, że w klasie wszystkich estymatorów nie istnieje estymator o jednostajnie minimalnym błędzie średniokwadratowym. W wykładzie trzecim ograniczyliśmy się do klasy estymatorów nieobciążonych i podaliśmy metody konstrukcji estymatorów o jednostajnie minimalnym ryzyku w tej klasie.

Estymatory największej wiarygodności pojawiają się w wyniku przyjęcia pewnej arbitralnej zasady, którą sformułowaliśmy na początku punktu 2.1. To, że estymatory największej wiarygodności nie muszą być nieobciążone, jest oczywiste: jeżeli $\hat{\theta}$ jest nieobciążonym

$ENW[\theta]$ oraz h jest dowolnym odwzorowaniem na Θ , to $h(\hat{\theta})$ jest $ENW[h(\theta)]$, ale $h(\hat{\theta})$ nie musi być estymatorem nieobciążonym wielkości $h(\theta)$.

Powstaje naturalne pytanie o to, jak dobre estymatory uzyskuje się, przyjmując zasadę największej wiarygodności. Okazuje się, że błąd średniokwadratowy ENW może być jednostajnie mniejszy od błędu $ENMW$, może być jednostajnie większy lub że błędy tych estymatorów mogą być nieporównywalne. Zilustrujemy te wszystkie przypadki przykładami.

Przykład 3. W wykładzie trzecim rozważaliśmy klasę estymatorów

$$\hat{\sigma}_c^2 = c \cdot \sum_{i=1}^n (X_i - \bar{X})^2, \quad c > 0.$$

Były to estymatory wariancji σ^2 w rozkładzie normalnym $N(\mu, \sigma^2)$. Estymatorem nieobciążonym o minimalnej wariancji w tej klasie jest $\hat{\sigma}_{n-1}^2$, estymatorem największej wiarygodności jest $\hat{\sigma}_n^2$, więc błąd średniokwadratowy ENW jest jednostajnie mniejszy od błędu średniokwadratowego $ENMW$. (Oba te estymatory są dominowane przez estymator $\hat{\sigma}_{n+1}^2$). ■

Przykład 4. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, \sigma^2)$, gdzie σ^2 jest znaną wariancją, i niech zadanie polega na estymacji $g(\mu, \sigma^2) = \mu^2$.

$ENMW$ jest $\bar{X}^2 - \sigma^2/n$. ENW jest \bar{X}^2 . Ponieważ $E_{\mu, \sigma^2} \bar{X}^2 = \mu^2 + \sigma^2/n$, więc

$$R_{\bar{X}^2 - \sigma^2/n}(\mu, \sigma^2) = E_{\mu, \sigma^2} (\bar{X}^2 - \frac{\sigma^2}{n} - \mu^2)^2 \leq E_{\mu, \sigma^2} (\bar{X}^2 - \mu^2)^2 = R_{\bar{X}^2}(\mu, \sigma^2) \quad \forall (\mu, \sigma^2).$$

Zatem $ENMW$ jest jednostajnie nie gorszy od ENW . Dokładniejsze rachunki prowadzą do wyniku

$$R_{\bar{X}^2 - \sigma^2/n}(\mu, \sigma^2) = R_{\bar{X}^2}(\mu, \sigma^2) - \left(\frac{\sigma^2}{n}\right)^2 < R_{\bar{X}^2}(\mu, \sigma^2),$$

więc w rozważanym problemie $ENMW$ jest jednostajnie lepszy od ENW .

Odnotujmy jednak, że estymowana wielkość μ^2 jest nieujemna; $ENW[\mu^2]$ jest również zawsze nieujemny, natomiast $ENMW[\mu^2]$ może przyjmować wartości ujemne. ■

Przykład 5. W wykładzie trzecim pokazaliśmy, że jeżeli X_1, X_2, \dots, X_n jest próbą z rozkładu Poissona $P(\theta)$, $\theta > 0$ oraz $T = \sum_{i=1}^n X_i$, to $ENMW[e^{-\theta}]$ jest $\lambda^* = (1 - \frac{1}{n})^T$.

Dla logarytmu funkcji wiarygodności mamy $l(\theta; x_1, x_2, \dots, x_n) = T \cdot \log \theta - n\theta + \text{const}$, więc $ENW[\theta]$ jest $\hat{\theta} = T/n$. Zatem $ENW[e^{-\theta}]$ jest $\tilde{\lambda} = e^{-T/n}$. Oznaczmy $\lambda = e^{-\theta}$. Błędy średniokwadratowe tych estymatorów są równe

$$R_{\lambda^*}(\lambda) = \lambda^{2 - \frac{1}{n}} - \lambda^2$$

$$R_{\tilde{\lambda}}(\lambda) = \lambda^{n(1 - e^{-2/n})} - 2\lambda^{1 + n(1 - e^{-1/n})} + \lambda^2.$$

Okazuje się, że dla pewnych wartości λ estymator największej wiarygodności jest lepszy, a dla innych gorszy od estymatora nieobciążonego o minimalnej wariancji. ■

Wyznaczenie ENW w danym problemie wymaga wyznaczenia funkcji wiarygodności (a więc odpowiedniej funkcji gęstości prawdopodobieństwa) i znalezienia wartości argumentu tej funkcji, przy której osiąga ona swoją wartość największą.

Pierwszy krok jest pewnym zadaniem z teorii prawdopodobieństwa, na ogół niezbyt trudnym nawet wtedy, gdy kolejne obserwacje X_1, X_2, \dots, X_n nie są niezależnymi zmiennymi losowymi o jednakowym rozkładzie prawdopodobieństwa; zilustrujemy to dwoma przykładami w następnym punkcie. O kłopotach związanych z drugim krokiem mówimy w punkcie 2.4.

2.3. ENW w złożonych doświadczeniach

Przykład 6. Mierzy się pewną własność liczbową (ciężar, długość, lub tp) każdego z n różnych obiektów. Oznaczmy te mierzone wielkości przez $\mu_1, \mu_2, \dots, \mu_n$. Każdy obiekt jest mierzony k razy. Błąd każdego pomiaru jest zmienną losową o rozkładzie normalnym $N(0, \sigma^2)$ z nieznaną wariancją σ^2 , a wyniki poszczególnych pomiarów są niezależnymi zmiennymi losowymi. Należy oszacować $\mu_1, \mu_2, \dots, \mu_n$ oraz σ^2 .

Niech $X_{i,j}$ ($i = 1, 2, \dots, k; j = 1, 2, \dots, n$) będzie wynikiem i -tego pomiaru j -tego obiektu. Na mocy powyższych założeń, $X_{i,j}$ jest zmienną losową o rozkładzie normalnym $N(\mu_j, \sigma^2)$. Przy każdym ustalonym j , oszacowanie μ_j nie następuje żadnymi trudnościami: $ENMW$ i jednocześnie ENW jest średnią z próby $X_{1,j}, X_{2,j}, \dots, X_{k,j}$.

Oszacowanie wariancji σ^2 jest bardziej złożone, bo wymaga jakiegoś złożenia informacji o tym parametrze zawartych w k próbach pochodzących z k różnych rozkładów normalnych.

Na mocy niezależności pomiarów $X_{i,j}$ otrzymujemy funkcję wiarygodności

$$\begin{aligned} L(\mu_1, \mu_2, \dots, \mu_n, \sigma^2; x_{1,1}, x_{1,2}, \dots, x_{k,n}) &= \\ &= \prod_{i=1}^k f_{\mu_1, \sigma^2}(x_{i,1}) \cdot \prod_{i=1}^k f_{\mu_2, \sigma^2}(x_{i,2}) \cdot \dots \cdot \prod_{i=1}^k f_{\mu_n, \sigma^2}(x_{i,n}), \end{aligned}$$

gdzie $f_{\mu, \sigma^2}(x)$ jest gęstością rozkładu normalnego $N(\mu, \sigma^2)$. Funkcja $l = \log L$ przyjmuje postać

$$\begin{aligned} l(\mu_1, \mu_2, \dots, \mu_n, \sigma^2; x_{1,1}, x_{1,2}, \dots, x_{k,n}) &= \\ &= -kn \cdot \log \sigma - kn \cdot \log \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{j=1}^n \sum_{i=1}^k (x_{i,j} - \mu_j)^2, \end{aligned}$$

więc estymatorami dla μ_j ($j = 1, 2, \dots, n$) oraz σ^2 są

$$\begin{aligned} \hat{\mu}_j &= \frac{1}{k} \sum_{i=1}^k X_{i,j}, \quad j = 1, 2, \dots, k, \\ \hat{\sigma}^2 &= \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k (X_{i,j} - \hat{\mu}_j)^2. \end{aligned}$$

Chociaż nie wszystkie obserwacje $X_{i,j}$ miały taki sam rozkład prawdopodobieństwa, skonstruowanie $ENW[(\mu_1, \mu_2, \dots, \mu_n, \sigma^2)]$ nie sprawiło tu żadnych trudności. ■

Przykład 7. Czas życia T pewnych obiektów (organizmów, urządzeń) mierzymy liczbą naturalną. Niech $\{T = k\}$ oznacza, że obiekt "zmarł" w przedziale czasu $(k-1, k]$, $k = 1, 2, \dots$. Przypuśćmy, że T jest zmienną losową o rozkładzie geometrycznym:

$$P_\theta\{T = k\} = (1 - \theta)\theta^{k-1}, \quad k = 1, 2, \dots,$$

gdzie $\theta \in (0, 1)$ jest nieznanym parametrem. Zadanie polega na tym, żeby na podstawie próby T_1, T_2, \dots, T_n z tego rozkładu oszacować θ .

Przypuśćmy, że badania prowadzimy przez r okresów i parametr θ mamy oszacować natychmiast po upływie tego czasu. Jest to model doświadczenia z częściową obserwacją, bo gdy pewien obiekt żyje dłużej niż r okresów, czas jego życia nie jest znany.

Ponieważ $P_\theta\{T > r\} = \theta^r$, więc funkcja wiarygodności przyjmuje postać

$$L(\theta; t_1, t_2, \dots, t_n) = \prod_{i=1}^n [(1 - \theta)\theta^{t_i-1} \mathbf{1}_{\{1, 2, \dots, r\}}(t_i) + \theta^r \mathbf{1}_{\{r+1, r+2, \dots\}}(t_i)].$$

Oznaczając przez S sumę czasów życia tych obiektów, które przeżyły co najwyżej r okresów oraz przez M liczbę pozostałych obiektów, otrzymujemy

$$L(\theta; t_1, t_2, \dots, t_n) = \theta^{(r+1)M+S-n} (1 - \theta)^{n-M},$$

a więc estymatorem największej wiarygodności parametru θ jest

$$\hat{\theta} = \frac{(r+1)M + S - n}{rM + S}.$$

■

2.4. Kłopoty z ENW

Przed wszystkim odnotujmy następujące ogólne spostrzeżenie: estymatory największej wiarygodności wywodzą się z pewnej arbitralnie przyjętej zasady, a nie — jak na przykład estymatory o jednostajnie minimalnym ryzyku — z rozwiązania pewnego dobrze postawionego zadania optymalizacji (to, czy rozwiązywane zadanie optymalizacji opisuje to, o co w praktycznych zastosowaniach rzeczywiście chodzi jest zupełnie innym zagadnieniem — wybór kwadratowej funkcji do opisu błędu jest również w pełni arbitralny). Wiemy już, że zasada największej wiarygodności czasami prowadzi do lepszych rezultatów niż rozwiązania optymalne w teorii estymatorów nieobciążonych o minimalnej wariancji, ale czasami prowadzi do rozwiązań gorszych. Samo pojęcie "lepszy" lub "gorszy" jest przy tym również arbitralne. Warto to wszystko mieć na uwadze, gdy w rozważanym modelu statystycznym uda nam się uzyskać elegancki ENW.

Inne problemy związane z ENW mają raczej charakter techniczny.

1. *ENW nie zawsze istnieją.* Prosty, ale raczej patologiczny przykład jest następujący: wiarogodność (dwuwymiarowego) parametru (μ, σ^2) w rozkładzie normalnym, gdy dokonujemy jednej obserwacji x , wynosi

$$L(\mu, \sigma^2; x) = (\sigma\sqrt{2\pi})^{-1} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Tu $ENW[(\mu, \sigma^2)]$ nie istnieje, bo kładąc $\mu = x$ oraz wybierając σ dowolnie małe, możemy uzyskać dowolnie dużą wartość L .

Oto bardziej naturalny przykład.

Dokonyjemy pomiaru pewnej wielkości μ przyrządem, którego błąd "w warunkach normalnych" ma rozkład normalny ze znaną wariancją, powiedzmy rozkład $N(0, 1)$. Na skutek "grubszej pomyłki" może jednak pojawić się wynik o pewnej innej, nieznannej wariancji σ^2 . Przyjmijmy, że prawdopodobieństwo takiej pomyłki jest małe i wynosi ϵ . Wtedy wynik X pomiaru jest zmienną losową o rozkładzie z gęstością

$$f_{\mu, \sigma^2}(x) = \frac{1-\epsilon}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} + \frac{\epsilon}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Jeżeli dokonamy n takich pomiarów X_1, X_2, \dots, X_n , to wiarogodność parametru (μ, σ^2) ma postać

$$L(\mu, \sigma^2; x_1, x_2, \dots, x_n) = \prod_{j=1}^n \left[\frac{1-\epsilon}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_j-\mu)^2\right\} + \frac{\epsilon}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x_j-\mu}{\sigma}\right)^2\right\} \right].$$

Znowu okazuje się, że dla każdej dowolnie dużej liczby M można znaleźć takie (μ, σ^2) , żeby $L > M$. W tym interesującym i ważnym dla zastosowań przypadku ENW nie istnieje.

2. *ENW może nie być określony jednoznacznie.* Inaczej mówiąc: może istnieć dużo różnych ENW i wtedy, pozostając na gruncie zasady największej wiarogodności, nie wiadomo, który z nich wybrać.

Oto przykład. Jeżeli X_1, X_2, \dots, X_n jest próbą z rozkładu $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, to

$$f_{\theta}(x_1, x_2, \dots, x_n) = \begin{cases} 1, & \text{gdy } \theta - \frac{1}{2} \leq x_1, x_2, \dots, x_n \leq \theta + \frac{1}{2}, \\ 0, & \text{poza tym} \end{cases}$$

czyli

$$L(\theta; x_1, x_2, \dots, x_n) = \begin{cases} 1, & \text{gdy } x_{n:n} - \frac{1}{2} \leq \theta \leq x_{1:n} + \frac{1}{2}, \\ 0, & \text{poza tym.} \end{cases}$$

Wynika stąd, że każda liczba z przedziału $[X_{n:n} - \frac{1}{2}, X_{1:n} + \frac{1}{2}]$ jest $ENW[\theta]$.

3. *Efektywne wyznaczenie wartości ENW może być bardzo trudne.* Wypisanie funkcji wiarogodności w jawnej postaci na ogół nie sprawia trudności, ale wyznaczenie punktu maksimum globalnego tej funkcji, gdy istnieje więcej niż jedno maksimum lokalne lub gdy przestrzeń parametrów Θ jest bardziej złożonym zbiorem, może stanowić niebagatelny problem teoretyczny lub/i numeryczny.

3. Testy oparte na ilorazie wiarygodności

3.1. Konstrukcja

Niech hipoteza H wyróżnia w przestrzeni parametrów Θ modelu statystycznego $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$ podzbiór Θ_H . Dla hipotezy alternatywnej K niech $\Theta_K = \Theta \setminus \Theta_H$.

DEFINICJA 3. Wiarygodnością hipotezy H , gdy zaobserwowano x , nazywamy

$$L_H(x) = \sup_{\theta \in \Theta_H} L(\theta; x).$$

□

Jeżeli weryfikujemy hipotezę H wobec hipotezy alternatywnej K , to *zasada największej wiarygodności* mówi, że należy zakwestionować H , gdy hipoteza K okazuje się bardziej wiarygodna niż H , a dokładniej, gdy

$$(1) \quad \frac{L_K(x)}{L_H(x)} > \lambda_0 \geq 1,$$

gdzie λ_0 jest stałą dobraną tak, aby test miał zadany poziom istotności α , tzn. tak, aby

$$P_\theta \left\{ \frac{L_K(X)}{L_H(X)} > \lambda_0 \right\} \leq \alpha, \quad \text{gdzie } \theta \in \Theta_H.$$

Zwróćmy uwagę na analogię pomiędzy (1) a odpowiednią wielkością służącą do konstrukcji testów najmocniejszych w teorii Neymana–Pearsona.

Niech $\hat{\theta}$ będzie taką wartością parametru $\theta \in \Theta_H \subset \Theta$, że

$$L(\hat{\theta}; x) = \sup_{\theta \in \Theta_H} L(\theta; x)$$

i niech $\hat{\theta}$ będzie $ENW[\theta]$, tzn. taką wartością parametru θ , że

$$L(\hat{\theta}; x) = \sup_{\theta \in \Theta} L(\theta; x).$$

Zauważmy, że dla $\lambda_0 \geq 1$ mamy

$$\frac{L_K(x)}{L_H(x)} > \lambda_0 \quad \text{wtedy i tylko wtedy, gdy} \quad \frac{L(\hat{\theta}; x)}{L(\hat{\theta}; x)} > \lambda_0$$

Sugeruje to, aby obszar krytyczny testu zdefiniować jako zbiór tych $x \in \mathcal{X}$, dla których

$$\frac{L(\hat{\theta}; x)}{L(\hat{\theta}; x)} > \lambda_0,$$

przy odpowiedniej stałej λ_0 .

Wielkość po lewej stronie tej nierówności będziemy oznaczali przez $\lambda(x)$.

DEFINICJA 4. Test o obszarze krytycznym postaci

$$\{x: \lambda(x) > \lambda_0\}$$

nazywamy *testem opartym na ilorazie wiarygodności*. \square

Chociaż zasada największej wiarygodności nie jest oparta na żadnych ściśle sprecyzowanych rozważaniach optymalizacyjnych, to jednak okazała się ona skuteczna w uzyskiwaniu zadowalających postępowań w wielu konkretnych zagadnieniach statystycznych. Pewnym argumentem na korzyść tej heurystycznej zasady jest to, że w sformułowaniu (1), w przypadku prostych hipotez H i K oraz w przypadku rodzin rozkładów o monotonicznym ilorazie wiarygodności, prowadzi do takiego samego obszaru krytycznego, jaki daje teoria Neymana–Pearsona. Zasada największej wiarygodności okazuje się przydatna szczególnie wtedy, gdy testy JNM nie istnieją lub gdy jest trudno je skonstruować, a także — jak w problemach estymacji — gdy obserwacje X_1, X_2, \dots, X_n nie są niezależnymi zmiennymi losowymi o jednakowym rozkładzie. Czasami jednak prowadzi do ewidentnie nonsensownych wyników (patrz niżej przykład 10).

3.2. Przykłady

Przykład 8. Próba X_1, X_2, \dots, X_m pochodzi z rozkładu normalnego $N(\mu_1, \sigma^2)$, próba Y_1, Y_2, \dots, Y_n pochodzi z rozkładu $N(\mu_2, \sigma^2)$ i próby są niezależne. Zadanie polega na tym, żeby dla danego $\alpha \in (0, 1)$ skonstruować test na poziomie istotności α dla weryfikacji hipotezy $H: \mu_1 = \mu_2$ wobec hipotezy alternatywnej $K: \mu_1 \neq \mu_2$.

Jak już wiemy, test JNM w tym problemie nie istnieje.

Dla logarytmu funkcji wiarygodności mamy

$$(2) \quad l(\mu_1, \mu_2, \sigma^2; x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = \\ = -\frac{m+n}{2} \log \sigma^2 - (m+n) \log \sqrt{2\pi} - \frac{1}{2\sigma^2} \left[\sum_{i=1}^m (x_i - \mu_1)^2 + \sum_{j=1}^n (y_j - \mu_2)^2 \right].$$

Dla estymatora największej wiarygodności $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2)$ parametru (μ_1, μ_2, σ^2) otrzymujemy

$$\hat{\mu}_1 = \bar{X},$$

$$\hat{\mu}_2 = \bar{Y},$$

$$\hat{\sigma}^2 = \frac{1}{m+n} \left[\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right]$$

oraz

$$L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2; x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = (2\pi e)^{-(m+n)/2} \hat{\sigma}^{-(m+n)}.$$

Funkcja wiarygodności (2), obcięta do zbioru $\{(\mu_1, \mu_2, \sigma^2) : \mu_1 = \mu_2\}$, osiąga swoją wartość największą w punkcie

$$\dot{\mu}_1 = \dot{\mu}_2 = \frac{m\bar{X} + n\bar{Y}}{m+n},$$

$$\hat{\sigma}^2 = \frac{1}{m+n} \left[\sum_{i=1}^m (X_i - \hat{\mu}_1)^2 + \sum_{j=1}^n (Y_j - \hat{\mu}_2)^2 \right].$$

Wtedy

$$L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2; x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = (2\pi e)^{-(m+n)/2} \hat{\sigma}^{-(m+n)}.$$

Otrzymujemy więc

$$\lambda(x) = \left(\frac{\hat{\sigma}^2}{\sigma^2} \right)^{\frac{m+n}{2}} = \left[1 + \frac{mn}{(m+n)^2} \frac{(\bar{Y} - \bar{X})^2}{\hat{\sigma}^2} \right]^{\frac{m+n}{2}}.$$

Oczywiście $\lambda(x) > \lambda_0$ wtedy i tylko wtedy, gdy

$$\frac{|\bar{Y} - \bar{X}|}{\sqrt{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2}} > c,$$

przy odpowiedniej stałej c . Zatem obszar krytyczny testu opartego na ilorazie wiarygodności ma postać

$$(3) \quad \left\{ \frac{|\bar{y} - \bar{x}|}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{j=1}^n (y_j - \bar{y})^2}} > c \right\}.$$

Stałą c wyznaczamy odpowiednio do wybranego poziomu istotności $\alpha \in (0, 1)$ testu. Przeprowadza się to w następujący sposób.

Jeżeli weryfikowana hipoteza H jest prawdziwa, to, dla pewnego μ , średnia \bar{X} ma rozkład normalny $N(\mu, \sigma^2/m)$, \bar{Y} ma rozkład $N(\mu, \sigma^2/n)$, $\sigma^{-2} \sum_{i=1}^m (X_i - \bar{X})^2$ ma rozkład chi-kwadrat o $(m-1)$ stopniach swobody, $\sigma^{-2} \sum_{j=1}^n (Y_j - \bar{Y})^2$ ma rozkład chi-kwadrat o $(n-1)$ stopniach swobody i wszystkie te zmienne losowe są niezależne. Stąd wynika, że zmienna losowa

$$\frac{\bar{Y} - \bar{X}}{\sigma} \sqrt{\frac{mn}{m+n}}$$

ma rozkład normalny $N(0, 1)$, zmienna losowa $\sigma^{-2} (\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2)$ ma rozkład chi-kwadrat o $(m+n-2)$ stopniach swobody, a więc zmienna losowa

$$\frac{\bar{Y} - \bar{X}}{\sqrt{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

ma rozkład t Studenta o $(m+n-2)$ stopniach swobody. Wartością krytyczną c testu (3) na poziomie istotności α jest więc liczba $t(\alpha, m+n-2)$ wyznaczona tak, że jeżeli t_{m+n-2} jest zmienną losową o rozkładzie t Studenta o $(m+n-2)$ stopniach swobody, to

$$P\{|t_{m+n-2}| > t(\alpha, m+n-2)\} = \alpha.$$

W zastosowaniach wartości $t(\alpha, m+n-2)$ odczytuje się w łatwo dostępnych tablicach lub z łatwo dostępnych standardowych procedur komputerowych. ■

Przykład 9. Próba X_1, X_2, \dots, X_n , $n \geq 2$, pochodzi z pewnego rozkładu normalnego $N(\mu, \sigma^2)$, $\mu \in \mathbf{R}^1$, $\sigma > 0$. Dla danego $\alpha \in (0, 1)$, skonstruować test na poziomie istotności α dla weryfikacji hipotezy $H: \sigma^2 = \sigma_0^2$ wobec hipotezy alternatywnej $K: \sigma^2 \neq \sigma_0^2$, gdzie σ_0^2 jest ustaloną liczbą dodatnią.

Dla wiarygodności parametru (μ, σ^2) mamy

$$L(\mu, \sigma^2; x_1, x_2, \dots, x_n) = (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}.$$

Funkcja wiarygodności osiąga maksimum na zbiorze $\{(\mu, \sigma^2): \mu \in \mathbf{R}^1, \sigma^2 > 0\}$ w punkcie $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2/n$, i to maksimum wynosi

$$L(\hat{\mu}, \hat{\sigma}^2; x_1, x_2, \dots, x_n) = (\hat{\sigma}\sqrt{2\pi})^{-n} e^{-n/2}.$$

Na zbiorze $\{(\mu, \sigma^2): \mu \in \mathbf{R}^1, \sigma^2 = \sigma_0^2\}$ maksimum tej funkcji jest osiąganę w punkcie $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \sigma_0^2$ i wynosi

$$(\sigma_0\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x})^2\right\}.$$

Zatem

$$\lambda(x_1, x_2, \dots, x_n) = \left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{-\frac{n}{2}} \exp\left\{\frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma_0^2} - 1\right)\right\}$$

i obszar krytyczny $\{\lambda > \lambda_0\}$ przyjmuje postać

$$\left\{\frac{\hat{\sigma}^2}{\sigma_0^2} < \lambda'_0\right\} \cup \left\{\frac{\hat{\sigma}^2}{\sigma_0^2} > \lambda''_0\right\},$$

gdzie λ'_0, λ''_0 są liczbami wybranymi tak, żeby

$$P_{\sigma_0}\left\{\left\{\frac{\hat{\sigma}^2}{\sigma_0^2} < \lambda'_0\right\} \cup \left\{\frac{\hat{\sigma}^2}{\sigma_0^2} > \lambda''_0\right\}\right\} = \alpha.$$

Dokładną konstrukcję testu na poziomie istotności α pozostawiamy jako zadanie. ■

Przykład 10. W tym przykładzie pokazujemy, że test oparty na ilorazie wiarygodności może być zupełnie bezużyteczny.

Niech zmienna losowa X ma pewien rozkład $P \in \mathcal{P}$, skupiony na skończonym zbiorze $\mathcal{X} = \{O, S_1, S_2, \dots, S_n, Q_1, Q_2, \dots, Q_n\}$. Niech

$$\mathcal{P} = \{P_0\} \cup \{P_\theta: \theta = (\theta_1, \theta_2, \dots, \theta_n), \sum_{i=1}^n \theta_i = 1, \theta_i \geq 0\},$$

gdzie

$$P_0\{X = O\} = \alpha, \quad P_0\{X = S_i\} = \frac{\alpha}{n}, \quad P_0\{X = Q_i\} = \frac{1 - 2\alpha}{n},$$

$$P_\theta\{X = 0\} = 1 - \frac{1}{n}, \quad P_\theta\{X = S_i\} = \frac{\theta_i}{n}, \quad P_\theta\{X = Q_i\} = 0,$$

przy czym $\alpha \in (0, 1/2)$ jest ustaloną liczbą. Weryfikujemy hipotezę prostą $H : P = P_0$ wobec złożonej hipotezy alternatywnej $K : P \in \mathcal{P} \setminus \{P_0\}$.

Dla ustalonego $x \in \mathcal{X}$ mamy

$$P_0\{X = x\} = \begin{cases} \alpha, & \text{gdy } x = O, \\ \frac{\alpha}{n}, & \text{gdy } x \in \{S_1, S_2, \dots, S_n\}, \\ \frac{1 - 2\alpha}{n}, & \text{gdy } x \in \{Q_1, Q_2, \dots, Q_n\}, \end{cases}$$

$$\sup_{P \in \mathcal{P} \setminus \{P_0\}} P_\theta\{X = x\} = \begin{cases} 1 - \frac{1}{n}, & \text{gdy } x = O, \\ \frac{1}{n}, & \text{gdy } x \in \{S_1, S_2, \dots, S_n\}, \\ 0, & \text{gdy } x \in \{Q_1, Q_2, \dots, Q_n\}. \end{cases}$$

Zatem

$$\lambda(x) = \frac{\sup_{P \in \mathcal{P} \setminus \{P_0\}} P_\theta\{X = x\}}{P_0\{X = x\}} = \begin{cases} (1 - \frac{1}{n}) \cdot \frac{1}{\alpha}, & \text{gdy } x = O, \\ \frac{1}{\alpha}, & \text{gdy } x \in \{S_1, S_2, \dots, S_n\}, \\ 0, & \text{gdy } x \in \{Q_1, Q_2, \dots, Q_n\}. \end{cases}$$

Oczywiście $(1 - \frac{1}{n}) \frac{1}{\alpha} < \frac{1}{\alpha}$ oraz $P_0\{X \in \{S_1, S_2, \dots, S_n\}\} = \alpha$, więc test na poziomie istotności α ma obszar krytyczny $\{S_1, S_2, \dots, S_n\}$. Moc tego testu jest stała i wynosi $\frac{1}{n}$. Zatem, jeżeli $\frac{1}{n} < \alpha$, skonstruowany test jest gorszy od testu, który bez patrzenia na wynik obserwacji po prostu odrzuca H z prawdopodobieństwem α . ■

4. Zadania

1. Porównać ryzyka $ENMW$ i ENW w przykładzie 5.
2. Wyznaczyć $ENW[M]$ w rozkładzie $P_M\{X = x\} = \binom{M}{x} \binom{N-M}{n-x} / \binom{N}{n}$ (rozkład hipergeometryczny), gdzie N oraz n są ustalonymi liczbami naturalnymi.
3. Wykonuje się niezależne doświadczenia, z prawdopodobieństwem sukcesu θ w każdym doświadczeniu, dopóty, dopóki nie zaobserwuje się k sukcesów ($k \geq 1$ jest ustaloną liczbą). Wyznaczyć $ENW[\theta]$.

4. Mówimy, że zmienna losowa Y ma rozkład logarytmonormalny o parametrze (μ, σ) , jeżeli zmienna losowa $X = \log Y$ ma rozkład normalny $N(\mu, \sigma^2)$. Wtedy $E_{\mu, \sigma} Y = \exp\{\mu + \frac{1}{2}\sigma^2\}$.

Niech Y_1, Y_2, \dots, Y_n będzie próbą z rozkładu logarytmonormalnego o parametrze (μ, σ) . Skonstruować estymator największej wiarygodności wartości oczekiwanej zmiennej losowej Y . Wyznaczyć wartość oczekiwaną i wariancję tego estymatora.

5. Jeżeli T jest statystyką dostateczną dla $\theta \in \Theta$ oraz $\hat{\theta}$ jest jedynym $ENW[\theta]$, to $\hat{\theta}$ zależy od obserwacji X tylko poprzez $T(X)$.

6. Naszycować funkcję mocy testu z przykładu 8.

7. Przeprowadzić dokładną konstrukcję testu z przykładu 9. Podać jawną postać testu dla $\sigma_0^2 = 1$, $n = 10$ oraz $\alpha = 0.01$.

8. Wynikiem X obserwacji jest liczba sukcesów w n niezależnych doświadczeniach, przy czym prawdopodobieństwo θ sukcesu w pojedynczym doświadczeniu nie jest znane. Weryfikuje się hipotezę $H: \theta \leq \theta_0$ wobec hipotezy alternatywnej $K: \theta > \theta_0$, na poziomie istotności α . Wyznaczyć test oparty na ilorazie wiarygodności. Czy jest to test JNM?

9. Niech X_1, X_2 będzie dwuelementową próbą z rozkładu Cauchy'ego $C(\theta, 1)$. Wykazać, że jeżeli $|X_1 - X_2| > 2$, to istnieją dwa różne $ENW[\theta]$. Autorem zadania jest Piotr Szymański.

Wykład VI

METODA NAJMNIEJSZYCH KWADRATÓW MODELE LINIOWE

1. Przykłady wprowadzające

Wprowadźcie *metoda najmniejszych kwadratów* (MNK) jest pewną ogólną metodą estymacji, a *estymatory uzyskiwane metodą najmniejszych kwadratów* (EMNK) są typowym przykładem *estymatorów najmniejszej odległości*, w naszym wykładzie ograniczymy się do przedstawienia związanej z tym tematyki tylko w kontekście *analizy regresji*, a szczególnie w kontekście *modeli liniowych*.

Rozpatrywane tu modele statystyczne różnią się nieco od modeli dotychczas rozważanych: pojawia się w nich pewien nowy parametr, którym na ogół możemy manipulować (tymi manipulacjami zajmuje się *teoria planowania eksperymentu statystycznego*). Nie będę jednak w wykładzie formułował explicite tych modeli; pozostawiam to jako zadanie. Oto kilka przykładów ilustrujących rozważane tu modele.

Przykład 1. Ciało o jednostkowej masie umieszczono w polu działania pewnej siły o nieznannej wartości F . Obserwuje się położenia Y_1, Y_2, \dots, Y_n ciała w wybranych chwilach t_1, t_2, \dots, t_n . Obserwacje obarczone są błędami losowymi $\epsilon_1, \epsilon_2, \dots, \epsilon_n$. Zadanie polega na oszacowaniu F .

Jak wiadomo, droga przebyta w czasie t przez ciało o jednostkowej masie pod wpływem siły F wynosi $\frac{1}{2}Ft^2$. Zatem obserwujemy zmienne losowe

$$Y_i = \frac{1}{2}Ft_i^2 + \epsilon_i, \quad i = 1, 2, \dots, n,$$

i na podstawie tych obserwacji mamy skonstruować estymator \hat{F} wielkości F . ■

Przykład 2. Pewien produkt chemiczny można wytwarzać bez użycia katalizatorów, ale przypuszcza się, że w obecności katalizatorów wydajność procesu będzie większa. Dysponujemy dwoma katalizatorami A i B , których możemy użyć w (dowolnych) dawkach $x \geq 0, y \geq 0$. Spodziewamy się, że wtedy wydajność procesu będzie równa $\mu + \alpha x + \beta y$, gdzie α i β są nieznanymi "efektywnościami" katalizatorów A i B , a μ jest wydajnością procesu prowadzonego bez użycia katalizatorów. W celu oszacowania wielkości μ, α i β , rejestrujemy wydajność procesu przy wybranych poziomach $x_i, i = 1, 2, \dots, n$, oraz $y_j, j = 1, 2, \dots, m$, katalizatorów A i B i otrzymujemy wyniki

$$Y_{i,j} = \mu + \alpha x_i + \beta y_j + \epsilon_{i,j}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m,$$

gdzie $\epsilon_{i,j}$ są błędami losowymi. ■

Przykład 3. W ekonometrii rozważa się następującą funkcję (jest to tzw. funkcja Cobba–Douglasa) opisującą wielkość produktu x w zależności od nakładu pracy n i nakładu kapitału k :

$$x = An^\alpha k^\beta,$$

gdzie A, α i β są pewnymi stałymi (stałe α i β noszą nazwy: elastyczność produktu względem nakładu pracy i elastyczność produktu względem nakładu kapitału). Obserwacje wielkości produktu końcowego x , przy danych nakładach pracy n i nakładach kapitału k , obarczone są błędem losowym. Ekonomisci interesują się wielkościami α i β . ■

Przykład 4. Bada się skuteczność pewnego preparatu służącego do zwalczania pewnego gatunku owadów. Można spodziewać się, że frakcja $\phi(x)$ owadów, które nie przeżyją dawki preparatu w stężeniu x , jest niemalejącą funkcją o wartościach w przedziale $[0, 1]$, taką że $\phi(0) = 0$ i $\phi(+\infty) = 1$. Często przyjmuje się ustalony kształt krzywej ϕ i na podstawie odpowiedniego eksperymentu szacuje się jej parametry. Na przykład

$$\phi(x) = \frac{1}{1 - \exp[-\beta_1 - \beta_2 x]}, \quad \beta_1 \in \mathbf{R}^1, \beta_2 > 0$$

proceedzi do tzw. modeli logistycznych. ■

2. Idea MNK

Obserwujemy zmienne losowe Y_1, Y_2, \dots, Y_n , o których wiemy, że

$$EY_i = g_i(\theta), \quad i = 1, 2, \dots, n,$$

gdzie $g_i: \Theta \rightarrow \mathbf{R}^1$ są danymi funkcjami oraz $\theta \in \Theta \subset \mathbf{R}^k$.

Jeżeli parametr θ przebiega zbiór Θ , to punkt $(g_1(\theta), g_2(\theta), \dots, g_n(\theta))$ przebiega pewien zbiór $\Gamma \subset \mathbf{R}^n$. Zaobserwowany punkt $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ również leży w \mathbf{R}^n .

Idea metody najmniejszych kwadratów polega na tym, żeby

1) w zbiorze Γ znaleźć punkt $\gamma = \gamma(Y_1, Y_2, \dots, Y_n)$ najbliższy punktowi $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ (nazwa metody pochodzi stąd, że punkt $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$ jest takim punktem w zbiorze Γ , który minimalizuje odległość euklidesową punktu \mathbf{Y} od zbioru Γ , tzn. sumę kwadratów $\sum_{i=1}^n (Y_i - \gamma_i)^2$),

2) za oszacowanie parametru θ przyjąć taki punkt $\hat{\theta} \in \Theta$, któremu "odpowiada" wyznaczony punkt γ , tzn. punkt $\hat{\theta}$ taki, że

$$(g_1(\hat{\theta}), g_2(\hat{\theta}), \dots, g_n(\hat{\theta})) = \gamma(Y_1, Y_2, \dots, Y_n).$$

Zwykle oba te etapy łączy się w jeden i za *EMNK* przyjmuje się θ minimalizującą wielkość

$$S(\theta) = \sum_{i=1}^n (Y_i - g_i(\theta))^2.$$

Na przykład, jeżeli w pierwszym przykładzie w chwilach t_1, t_2, \dots, t_n zarejestrowano położenia Y_1, Y_2, \dots, Y_n , to $EMNK[F]$ jest \hat{F} minimalizujące wielkość

$$S(F) = \sum_{i=1}^n \left(Y_i - \frac{1}{2} F t_i^2 \right)^2,$$

tzn. $\hat{F} = 2 \sum Y_i t_i^2 / \sum t_i^4$.

Z ogólnej konstrukcji wynika, że jeżeli zbiór Γ jest domknięty, to odpowiedni punkt $\gamma \in \Gamma$ zawsze istnieje, a więc zawsze istnieje, chociaż być może nie tylko jeden, estymator. Efektywne wyznaczenie wartości tego estymatora może jednak być numerycznie skomplikowane. Na przykład w czwartym przykładzie trzeba wyznaczyć $(\hat{\beta}_1, \hat{\beta}_2)$ minimalizujące wielkość

$$\sum_{i=1}^n \left(\frac{Z_i}{n_i} - \frac{1}{1 - \exp[-\beta_1 - \beta_2 x_i]} \right)^2,$$

gdzie n_i jest liczbą owadów, którym podano dawkę x_i preparatu, a Z_i jest liczbą owadów, które tej dawki nie przeżyły. W licznych praktycznych poradnikach można znaleźć różne pomysły ułatwiające rozwiązanie, np. linearyzacja funkcji g_i , schematy iteracyjne, itp.

Zgodnie z duchem naszego wykładu przyjrzymy się dokładniej $EMNK$ z punktu widzenia ich błędu średniokwadratowego. Pewne konkretne wyniki udaje się tu jednak uzyskać tylko dla tzw. modeli liniowych i w szczególności dla gaussowskich modeli liniowych.

3. EMNK w modelach liniowych

3.1. Ogólna postać modelu liniowego

Modelem liniowym nazywamy taki model statystyczny, w którym obserwacje Y_1, Y_2, \dots, Y_n mają postać

$$(1) \quad Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \dots + x_{i,k}\beta_k + \epsilon_i, \quad i = 1, 2, \dots, n,$$

gdzie $x_{i,j}$ są ustalonymi liczbami, ϵ_i są "błędami losowymi", a β_j , $j = 1, 2, \dots, k$, są nieznanymi stałymi (parametrami modelu).

Przykład 5. Jeżeli zadanie polega na oszacowaniu nieznanej wielkości μ za pomocą n jej pomiarów Y_i , $i = 1, 2, \dots, n$, obciążonych addytywnymi błędami losowymi ϵ_i , to odpowiedni model jest liniowy i ma postać

$$Y_i = \mu + \epsilon_i, \quad i = 1, 2, \dots, n.$$

■

Przykład 6 (*regresja drugiego stopnia*). Jeżeli wartość oczekiwana zmiennej losowej Y jest funkcją

$$EY = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

pewnej zmiennej t i zadanie polega na wnioskowaniu o współczynnikach $\alpha_0, \alpha_1, \alpha_2$ na podstawie obserwacji zmiennej losowej Y dla wybranych wartości t , to zadanie można sformułować w postaci modelu liniowego (1), w którym $k = 3$, $x_{i,1} = 1$, $x_{i,2} = t_i$, $x_{i,3} = t_i^2$, $\beta_1 = \alpha_0$, $\beta_2 = \alpha_1$ oraz $\beta_3 = \alpha_2$. ■

Przykład 7 (*porównanie dwóch technologii, zabiegów, leków, itp.*). Jeżeli μ_1 jest średnią wartością badanej cechy (wydajności, zużycia surowca, stanu zdrowia) dla pewnej technologii, zaś μ_2 taką samą wielkością dla pewnej innej, konkurencyjnej technologii i gdy na podstawie n_1 obiektów wykonanych według pierwszej technologii oraz n_2 obiektów wykonanych według drugiej mamy oszacować lub porównać μ_1 i μ_2 , to model obserwacji ma postać

$$Y_i = x_{i,1}\mu_1 + x_{i,2}\mu_2 + \epsilon_i,$$

gdzie $x_{i,1} = 1, x_{i,2} = 0$, gdy Y_i pochodzi z pierwszej technologii, a $x_{i,1} = 0, x_{i,2} = 1$, gdy Y_i pochodzi z drugiej technologii. W takich i podobnych sytuacjach może nas interesować nie tyle oszacowanie wielkości μ_1 i μ_2 , ile oszacowanie ich różnicy $\mu_1 - \mu_2$. ■

Oznaczmy przez \mathbf{X} macierz o n wierszach i k kolumnach:

$$\mathbf{X} = (x_{i,j})_{i=1,2,\dots,n; j=1,2,\dots,k},$$

przez $\boldsymbol{\beta}$ wektor kolumnowy $(\beta_1, \beta_2, \dots, \beta_k)^T$, przez \mathbf{Y} wektor kolumnowy obserwacji $(Y_1, Y_2, \dots, Y_n)^T$ i przez $\boldsymbol{\epsilon}$ wektor kolumnowy błędów $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$. Model liniowy (1) będziemy zapisywali w postaci

$$(2) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

O błędach losowych $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ zakładamy, że są niezależnymi zmiennymi losowymi o rozkładach takich, że $E\epsilon_i = 0, E\epsilon_i^2 = \sigma^2$, tzn. że $E\boldsymbol{\epsilon} = \mathbf{0}$ oraz $\text{Var}\boldsymbol{\epsilon} = E\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T = \sigma^2\mathbf{I}$.

3.2. EMNK w modelu liniowym. Twierdzenie Gaussa–Markowa

Zadanie oszacowania wektora $\boldsymbol{\beta}$ metodą najmniejszych kwadratów polega na wyznaczeniu $\boldsymbol{\beta}$ minimalizującego, przy oczywistych oznaczeniach,

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{i,j}\beta_j \right)^2 = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Jeżeli $\mathbf{Y} = \mathbf{0}$, to rozwiązaniem jest $\boldsymbol{\beta} = \mathbf{0}$. Załóżmy, że $\mathbf{Y} \neq \mathbf{0}$. Norma $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|$ jest minimalizowana przez takie $\boldsymbol{\beta} \neq \mathbf{0}$, że $\mathbf{X}\boldsymbol{\beta}$ jest rzutem ortogonalnym wektora \mathbf{Y} na podprzestrzeń $\{\mathbf{X}\boldsymbol{\beta}: \boldsymbol{\beta} \in \mathbf{R}^k\}$, tzn. przez $\boldsymbol{\beta}$ spełniające warunek

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{X} = \mathbf{0},$$

czyli warunek

$$(3) \quad \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{Y}.$$

Równanie (3) ma zawsze rozwiązania, a więc $EMNK[\boldsymbol{\beta}]$ zawsze istnieje. Wynika to stąd, że $\text{Im} \mathbf{X}^T = \text{Im} \mathbf{X}^T \mathbf{X}$, gdzie $\text{Im} \mathbf{A}$ oznacza obraz macierzy \mathbf{A} (por. zadanie 2).

Jeżeli macierz $\mathbf{X}^T \mathbf{X}$ jest nieosobliwa, to estymatorem parametru wektorowego $\boldsymbol{\beta}$, uzyskanym metodą najmniejszych kwadratów, jest

$$(4) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

W celu oceny jakości skonstruowanego estymatora, zbadamy jego własności z punktu widzenia estymacji nieobciążonej z jednostajnie minimalną wariancją.

Zacniemy od przypadku, gdy $\text{rz} \mathbf{X} = k \leq n$.

TWIERDZENIE 1. *Jeżeli $\text{rz} \mathbf{X} = k \leq n$, to $\hat{\boldsymbol{\beta}}$ jest estymatorem nieobciążonym, a jego wariancja wyraża się wzorem $\text{Var}_{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$.*

Dowód tego twierdzenia pozostawiamy jako zadanie. \square

Jeżeli estymator o wartościach liczbowych był nieobciążony, to jego błąd średniokwadratowy był równy po prostu jego wariancji. Tutaj mamy do czynienia z estymatorem o wartościach wektorowych, więc jego "wariancją" jest macierz — jest to tzw. macierz kowariancji. W takiej sytuacji nieobciążony estymator $\hat{\boldsymbol{\beta}}_1$ będziemy uważali za nie gorszy od nieobciążonego estymatora $\hat{\boldsymbol{\beta}}_2$, gdy dla każdego $\boldsymbol{\beta}$ różnica $\text{Var}_{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}_2 - \text{Var}_{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}_1$ jest macierzą nieujemnie określoną. Równoważne określenie jest następujące: estymator $\hat{\boldsymbol{\beta}}_1$ jest nie gorszy od estymatora $\hat{\boldsymbol{\beta}}_2$, jeżeli dla każdej funkcji liniowej $\mathbf{c}^T \boldsymbol{\beta}$

$$\text{Var}_{\boldsymbol{\beta}} (\mathbf{c}^T \hat{\boldsymbol{\beta}}_1) \leq \text{Var}_{\boldsymbol{\beta}} (\mathbf{c}^T \hat{\boldsymbol{\beta}}_2) \quad \forall \boldsymbol{\beta} \in \mathbf{R}^k.$$

Odpowiednikiem estymatora nieobciążonego o jednostajnie minimalnej wariancji jest teraz estymator najlepszy w sensie tak określonego porządku.

TWIERDZENIE 2. *Jeżeli błędy losowe $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ mają wartość oczekiwaną równą zeru, taką samą wariancję σ^2 i są nieskorelowane, to dla każdego $\mathbf{c} \in \mathbf{R}^k$ i dla każdego liniowego nieobciążonego estymatora $\hat{\boldsymbol{\phi}}$ funkcji parametrycznej $\mathbf{c}^T \boldsymbol{\beta}$ zachodzi*

$$\text{Var}_{\boldsymbol{\beta}} \mathbf{c}^T \hat{\boldsymbol{\beta}} \leq \text{Var}_{\boldsymbol{\beta}} \hat{\boldsymbol{\phi}} \quad \forall \boldsymbol{\beta} \in \mathbf{R}^k.$$

\square

Dowodu twierdzenia nie podaję, bo dotyczy ono szczególnego przypadku, gdy \mathbf{X} jest macierzą "pełnego rzędu". Przed sformulowaniem ogólnego twierdzenia i podaniem jego dowodu sformułuję kilka istotnych uwag na temat przypadku, gdy macierz $\mathbf{X}^T \mathbf{X}$ jest osobliwa.

Przede wszystkim odnotujmy, że jeżeli $rz\mathbf{X}^T\mathbf{X} = rz\mathbf{X} < k$, to wprawdzie istnieją $EMNK[\beta]$ i są to estymatory liniowe, to jednak może nie istnieć liniowy estymator nieobciążony wektora β . Przyjrzyjmy się temu dokładniej.

Rozpatrzmy liniowe funkcje $\mathbf{c}^T\beta$ parametru β ("liniowe funkcje parametryczne"). Jest oczywiste, że istnieje nieobciążony estymator wektora β wtedy i tylko wtedy, gdy dla każdego \mathbf{c} istnieje estymator nieobciążony funkcji parametrycznej $\mathbf{c}^T\beta$. Scharakteryzujemy te funkcje, dla których taki estymator istnieje – są to tak zwane "funkcje estymowalne" (używa się też nazwy "nieobciążenie estymowalne").

Przypuśćmy, że istnieje nieobciążony estymator liniowy $\mathbf{b}^T\mathbf{Y}$ funkcji parametrycznej $\mathbf{c}^T\beta$, czyli że

$$E_{\beta} \mathbf{b}^T\mathbf{Y} = \mathbf{c}^T\beta \quad \forall \beta \in \mathbf{R}^k,$$

co jest równoważne z tym, że

$$\mathbf{b}^T\mathbf{X}\beta = \mathbf{c}^T\beta \quad \forall \beta \in \mathbf{R}^k.$$

Ten warunek jest spełniony wtedy i tylko wtedy, gdy

$$(5) \quad \mathbf{c} = \mathbf{X}^T\mathbf{b},$$

tnz. wtedy i tylko wtedy, gdy $\mathbf{c} \in \text{Im}\mathbf{X}^T$.

Jeżeli $\hat{\beta}$ jest dowolnym rozwiązaniem równania (1), to za $EMNK[\mathbf{c}^T\beta]$ przyjmujemy $\mathbf{c}^T\hat{\beta}$.

TWIERDZENIE 3 (twierdzenie Gaussa–Markowa). *Jeżeli $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ mają wartość oczekiwaną zero, taką samą wariancję σ^2 i są nieskorelowane, to dla każdej nieobciążonej estymowalnej funkcji parametrycznej $\mathbf{c}^T\beta$ i dla każdego nieobciążonego estymatora liniowego $\mathbf{b}^T\mathbf{Y}$ tej funkcji zachodzi*

$$\text{Var}_{\beta} \mathbf{c}^T\hat{\beta} \leq \text{Var}_{\beta} \mathbf{b}^T\mathbf{Y} \quad \forall \beta \in \mathbf{R}^k.$$

D o w ó d. Niech $\mathbf{b}^T\mathbf{Y}$ będzie estymatorem nieobciążonym funkcji parametrycznej $\mathbf{c}^T\beta$ i niech \mathbf{a} będzie rzutem ortogonalnym \mathbf{b} na $\text{Im}\mathbf{X}$. Ponieważ $\mathbf{a}^T(\mathbf{b} - \mathbf{a}) = 0$, otrzymujemy

$$\begin{aligned} \text{Var}_{\beta} \mathbf{b}^T\mathbf{Y} &= \sigma^2 \mathbf{b}^T\mathbf{b} \\ &= \sigma^2 \left((\mathbf{a} + \mathbf{b} - \mathbf{a})^T (\mathbf{a} + \mathbf{b} - \mathbf{a}) \right) \\ &= \sigma^2 \left(\mathbf{a}^T\mathbf{a} + (\mathbf{b} - \mathbf{a})^T(\mathbf{b} - \mathbf{a}) \right), \end{aligned}$$

więc

$$(6) \quad \text{Var}_{\beta} \mathbf{b}^T\mathbf{Y} \geq \sigma^2 \mathbf{a}^T\mathbf{a}.$$

Z drugiej strony, na mocy wzoru (5) mamy $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \mathbf{b}^T \mathbf{X} \hat{\boldsymbol{\beta}} = (\mathbf{a} + \mathbf{b} - \mathbf{a})^T \mathbf{X} \hat{\boldsymbol{\beta}}$. Ponieważ, z definicji wektora \mathbf{a} , wektor $(\mathbf{b} - \mathbf{a})$ jest ortogonalny do $\text{Im} \mathbf{X}$, czyli $(\mathbf{b} - \mathbf{a})^T \mathbf{X} = 0$, otrzymujemy, że $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \mathbf{a}^T \mathbf{X} \hat{\boldsymbol{\beta}}$, a ponieważ $\mathbf{a} \in \text{Im} \mathbf{X}$, więc dla pewnego $\boldsymbol{\alpha}$ mamy $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}$. Korzystając teraz z tego, że $\hat{\boldsymbol{\beta}}$ jest rozwiązaniem równania (3), otrzymujemy, że $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{Y}$, czyli $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \mathbf{a}^T \mathbf{Y}$. Zatem $\text{Var}_{\boldsymbol{\beta}} \mathbf{c}^T \hat{\boldsymbol{\beta}} = \text{Var}_{\boldsymbol{\beta}} \mathbf{a}^T \mathbf{Y} = \sigma^2 \mathbf{a}^T \mathbf{a}$, i ostatecznie na mocy (6) otrzymujemy, że $\text{Var}_{\boldsymbol{\beta}} \mathbf{c}^T \hat{\boldsymbol{\beta}} \leq \text{Var}_{\boldsymbol{\beta}} \mathbf{b}^T \mathbf{Y}$. \square

Do tej pory skupialiśmy naszą uwagę na estymacji parametru $\boldsymbol{\beta}$. W wyjściowym modelu jest jednak jeszcze jeden, najczęściej nieznan, parametr σ^2 .

Jeżeli już wyznaczyliśmy estymator $\hat{\boldsymbol{\beta}}$, to można spodziewać się, że tak zwana *resztowa suma kwadratów*

$$(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$$

(jest to "odległość" obserwacji \mathbf{Y} od wyestymowanej wartości średniej $E\mathbf{Y}$) będzie tym większa, im większa jest wariancja σ^2 . Rozumujemy w następujący sposób.

Wektor błędu $\boldsymbol{\epsilon}$ można zapisać w postaci

$$\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) + \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Wektory $(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$ i $\mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ są ortogonalne. Wektor $\mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ leży w k -wymiarowej podprzestrzeni $\text{Im} \mathbf{X} \subset \mathbf{R}^n$, więc wektor $(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$ leży w $(n - k)$ -wymiarowej podprzestrzeni, ortogonalnej do $\text{Im} \mathbf{X}$. Można znaleźć taką nową bazę ortonormalną w \mathbf{R}^n , żeby pierwsze k wektorów bazy należało do $\text{Im} \mathbf{X}$ oraz pozostałe $(n - k)$ do $(\text{Im} \mathbf{X})^\perp$. Oznaczając przez $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$ wektor błędów w tej nowej bazie, otrzymujemy

$$\begin{aligned} \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\eta_1, \dots, \eta_k, 0, \dots, 0)^T, \\ (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) &= (0, \dots, 0, \eta_{k+1}, \dots, \eta_n)^T, \end{aligned}$$

więc

$$(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \eta_{k+1}^2 + \eta_{k+2}^2 + \dots + \eta_n^2.$$

Ponieważ przy przekształceniu ortogonalnym wektor losowy $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ o nieskorelowanych współrzędnych takich, że $E\epsilon_i = 0$, $E\epsilon_i^2 = \sigma^2$ ($i = 1, 2, \dots, n$) przechodzi na wektor $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$ o tych samych własnościach, więc

$$E_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \sum_{i=k+1}^n E_{\boldsymbol{\beta}} \eta_i^2 = (n - k) \sigma^2.$$

Zatem wielkość

$$s^2 = \frac{1}{n-k} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

jest estymatorem nieobciążonym wariancji σ^2 .

3.3. EMNK w gaussowskim modelu liniowym

Jeżeli błędy losowe $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ mają rozkłady normalne $N(0, \sigma^2)$, to obserwacje Y_1, Y_2, \dots, Y_n mają rozkłady normalne, a gęstość łącznego rozkładu tych obserwacji ma postać

$$f(y_1, y_2, \dots, y_n; \beta_1, \beta_2, \dots, \beta_k, \sigma) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\},$$

gdzie $\mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbf{R}^n$. Ponieważ

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta},$$

więc oznaczając j -tą składową wektora $\mathbf{X}^T \mathbf{Y}$ przez T_j , $j = 1, 2, \dots, k$, i kładąc $T_{k+1} = \mathbf{Y}^T \mathbf{Y}$, możemy przepisać tę gęstość w postaci

$$\exp \left\{ \frac{2}{\sigma^2} \sum_{j=1}^k \beta_j T_j - \frac{1}{2\sigma^2} T_{k+1} - b(\boldsymbol{\beta}, \sigma) \right\},$$

gdzie $b(\boldsymbol{\beta}, \sigma)$ jest funkcją parametru $(\boldsymbol{\beta}, \sigma)$, niezależną od obserwacji.

Jeżeli a priori nie nakładamy żadnych ograniczeń na parametr $(\boldsymbol{\beta}, \sigma)$, to przestrzeń tego parametru jest $(k+1)$ -wymiarowa, a rodzina rozkładów obserwacji \mathbf{Y} jest rodziną wykładniczą. Zatem $\mathbf{T} = (T_1, T_2, \dots, T_k, T_{k+1})$ jest statystyką dostateczną zupełną.

Ponieważ estymator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_k \end{pmatrix}$$

jest funkcją statystyki dostatecznej zupełnej i jest nieobciążony, więc jest to $ENMW[\boldsymbol{\beta}]$.

Podobnie

$$\begin{aligned} (n-k)s^2 &= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{Y}^T \mathbf{Y} - 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= \mathbf{Y}^T \mathbf{Y} - 2 \left((\mathbf{X}^T \mathbf{X})^{-1} \right) \mathbf{X}^T \mathbf{Y} + (\mathbf{X}^T \mathbf{Y})^T (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}). \end{aligned}$$

Tu $\mathbf{Y}^T \mathbf{Y} = T_{k+1}$ oraz $\mathbf{X}^T \mathbf{Y} = (T_1, T_2, \dots, T_k)^T$, więc s^2 jest $ENMW[\sigma^2]$.

Wynika stąd, że w przypadku gaussowskim estymator $\hat{\beta}$, skonstruowany metodą najmniejszych kwadratów, jest estymatorem o minimalnej wariancji w klasie wszystkich estymatorów nieobciążonych, a nie tylko w klasie nieobciążonych estymatorów liniowych, jak to ma miejsce bez założenia o normalności rozkładu błędów.

Przypomnijmy, że $\hat{\beta} = EMNK[\beta]$ to taki estymator, który minimalizuje

$$(\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta).$$

Ponieważ w modelu gaussowskim wiarygodność parametru (β, σ^2) przy obserwacjach $\mathbf{y} = (y_1, y_2, \dots, y_n)$ ma postać

$$l(\beta, \sigma; \mathbf{y}) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \right\},$$

więc $\hat{\beta}$ maksymalizuje $l(\beta, \sigma; \mathbf{y})$ przy każdym ustalonym σ^2 . Zatem $\hat{\beta}$ jest estymatorem największej wiarygodności parametru β . Ponadto, przy ustalonym $\hat{\beta}$, s^2 maksymalizuje $l(\hat{\beta}, \sigma; \mathbf{y})$ ze względu na σ , więc s^2 jest estymatorem największej wiarygodności parametru σ^2 .

Wynika stąd, że w gaussowskim modelu liniowym teoria estymatorów nieobciążonych o minimalnej wariancji, zasada największej wiarygodności oraz metoda najmniejszych kwadratów prowadzą do tego samego rezultatu.

4. Zadania

1. Sformułować modele statystyczne zagadnień w przykładach 1, 2, 3 i 4.
2. Sprawdzić, że dla każdej macierzy \mathbf{X} zachodzi równość $\text{Im}\mathbf{X}^T = \text{Im}\mathbf{X}^T\mathbf{X}$.
3. Udowodnić twierdzenie 1.
4. Podać $EMNK[\mu]$ w przykładzie 5.
5. Skonstruować $EMNK[\mu_1 - \mu_2]$ w przykładzie 7.
6. Przypuśćmy, że obserwacje x_1, x_2, \dots, x_n mogą być przedstawione w postaci

$$x_i = \beta_0 + \beta_1 a_i + \epsilon_i, \quad i = 1, 2, \dots, n,$$

gdzie a_1, a_2, \dots, a_n są znanymi wartościami pewnej zmiennej towarzyszącej oraz $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ są nieskorelowanymi błędami o wspólnej wariancji σ^2 . Sprawdzić, że oba parametry β_0 i β_1 są estymowalne wtedy i tylko wtedy, gdy nie wszystkie wartości a_i są sobie równe. Przedyskutować intuicyjną treść tego wyniku sporządzając odpowiedni wykres punktów (a_i, x_i) . Wykazać, że gdy nie wszystkie a_i są sobie równe, estymatory $\hat{\beta}_0$ i $\hat{\beta}_1$, otrzymane metodą najmniejszych kwadratów, mają postać

$$\hat{\beta}_1 = \frac{\sum (a_i - \bar{a}) x_i}{\sum (a_i - \bar{a})^2}, \quad \hat{\beta}_0 = \bar{x} - \bar{a} \hat{\beta}_1.$$

Zauważyć, że w modelu gaussowskim estymatory $\hat{\beta}_0$ i $\hat{\beta}_1$ są zmiennymi losowymi o rozkładzie normalnym. Wyznaczyć macierz kowariancji tych estymatorów.

7. Wyniki pewnych obserwacji x_1, x_2, \dots, x_n mogą być przedstawione w postaci

$$x_i = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \epsilon_i \quad i = 1, 2, \dots, n$$

gdzie a_i są wartościami zmiennej towarzyszącej oraz $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ są wielkościami nieskorelowanymi o jednakowej wariancji. Pokazać, że parametr $\beta = (\beta_0, \beta_1, \beta_2)$ jest estymowalny wtedy i tylko wtedy, gdy wśród a_1, a_2, \dots, a_n znajdują się co najmniej trzy różne wartości. Wyznaczyć EMNK $[\beta_0, \beta_1, \beta_2]$.

8. Niech x_1, x_2, x_3, x_4 będą wynikami lotniczych pomiarów kątów $\theta_1, \theta_2, \theta_3, \theta_4$ pewnego czworokąta na powierzchni ziemi. Założyć, że te obserwacje są obciążone błędami losowymi, które są niezależne, mają wartość oczekiwaną równą zero i wspólną wariancję σ^2 i przy tych założeniach wyznaczyć estymatory wielkości kątów metodą najmniejszych kwadratów. Wyznaczyć nieobciążony estymator wariancji σ^2 .

Przyjąć, że wiadomo, że dany czworokąt jest równoległobokiem, takim że $\theta_1 = \theta_3$ oraz $\theta_2 = \theta_4$. Jaką postać mają wtedy estymatory kątów i jak można by szacować σ^2 ?

Wykład VII

TEORIA DECYZJI STATYSTYCZNYCH

1. Sformułowanie problemu

Rozpatrujemy model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$, a wraz z nim dwa dodatkowe obiekty: zbiór \mathcal{A} i nieujemną funkcję $L: \Theta \times \mathcal{A} \rightarrow \mathbf{R}^1$. Zbiór \mathcal{A} nazywamy *zbiorem decyzji* (lub *zbiorem akcji*), jego elementy $a \in \mathcal{A}$ nazywamy *decyzjami* (lub *akcjami*), a funkcję L nazywamy *funkcją straty*. Wartość $L(\theta, a)$ tej funkcji interpretujemy jako wielkość straty, jaką ponosimy wtedy, gdy obserwacja X ma rozkład P_θ (gdy "prawdziwym" rozkładem jest P_θ), a podjęta zostaje decyzja a .

Decyzje są podejmowane na podstawie obserwacji X . Każda funkcja $\delta: \mathcal{X} \rightarrow \mathcal{A}$ określa następującą regułę postępowania: jeżeli zaobserwowano $x \in \mathcal{X}$, należy podjąć decyzję $\delta(x)$. Funkcję δ nazywamy *regułą decyzyjną*, a zbiór rozważanych reguł decyzyjnych oznaczamy przez \mathcal{D} .

Ogólniej, niech \mathcal{A}^* będzie rodziną wszystkich rozkładów prawdopodobieństwa na \mathcal{A} i niech $\delta^*: \mathcal{X} \rightarrow \mathcal{A}^*$. Funkcję δ^* nazywamy *randomizowaną regułą decyzyjną*. Randomizowana reguła decyzyjna określa następujący algorytm postępowania: jeżeli zaobserwowano $x \in \mathcal{X}$, należy decyzję wylosować według rozkładu prawdopodobieństwa $\delta^*(x)$. Choć w wielu zagadnieniach optymalnymi regułami decyzyjnymi okazują się być randomizowane reguły decyzyjne, w naszym wykładzie ograniczamy się do prezentacji teorii decyzji statystycznych tylko w stopniu pozwalającym na pozostanie w zbiorze \mathcal{D} nierandomizowanych reguł decyzyjnych.

Wielkość $E_\theta L(\theta, \delta(X))$ jest oczekiwaną stratą, gdy obserwacja X ma rozkład P_θ , a postępujemy według reguły decyzyjnej δ . Funkcja

$$(1) \quad R(\theta, \delta) = E_\theta L(\theta, \delta(X)), \quad \theta \in \Theta,$$

nazywa się *ryzykiem* lub *funkcją ryzyka* reguły decyzyjnej δ .

Teoria decyzji statystycznych zajmuje się konstruowaniem najlepszych reguł decyzyjnych. Oczywiście zależnie od tego, jaki sens nadaje się zdaniu "reguła decyzyjna δ_1 jest lepsza od reguły decyzyjnej δ_2 ", otrzymuje się różne "najlepsze" reguły decyzyjne. W wykładzie podajemy pewne formalizacje tego zagadnienia, ale dalsze rozważania poprzedzimy dwoma przykładami.

Przykład 1. Rozważamy model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$. Niech $g: \Theta \rightarrow \mathbf{R}^1$ będzie daną funkcją. Niech $\mathcal{A} = \mathbf{R}^1$ i niech $L(\theta, a) = (g(\theta) - a)^2$. Niech $\hat{g}: \mathcal{X} \rightarrow \mathbf{R}^1$ będzie

danym odwzorowaniem. Wielkość $L(\theta, \hat{g}(X))$ interpretujemy jako stratę, którą ponosimy wtedy, gdy obserwacja X ma rozkład P_θ , a my "podejmujemy decyzję" $\hat{g}(X)$. Jeżeli ta decyzja polega na orzeczeniu, że nieznaną wartość parametru $\theta \in \Theta$ jest równa $\hat{g}(X)$, znajdujemy się na gruncie teorii estymacji z kwadratową funkcją straty. Problem wyboru optymalnej reguły decyzyjnej \hat{g} jest tu dobrze nam już znanym problemem konstrukcji estymatora o minimalnym błędzie średniokwadratowym (por. wykład 3). ■

Przykład 2. W modelu statystycznym $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$ niech $\Theta = \{\theta_1, \theta_2\}$, $\mathcal{A} = \{a_1, a_2\}$ i niech

$$L(\theta_i, a_j) = \begin{cases} 0, & \text{gdy } i = j, \\ 1, & \text{gdy } i \neq j. \end{cases}$$

Funkcja ryzyka reguły decyzyjnej $\varphi: \mathcal{X} \rightarrow \mathcal{A}$ wyraża się wzorem

$$R(\theta, \varphi) = \begin{cases} P_{\theta_1}\{\varphi(X) = a_2\}, & \text{gdy } \theta = \theta_1, \\ P_{\theta_2}\{\varphi(X) = a_1\}, & \text{gdy } \theta = \theta_2. \end{cases}$$

Jeżeli ustalimy liczbę $\alpha \in (0, 1)$, ograniczymy się do rozważania reguł decyzyjnych φ , dla których $P_{\theta_1}\{\varphi(X) = a_2\} \leq \alpha$ i jeżeli za najlepszą w tej klasie uznamy tę regułę, która minimalizuje $P_{\theta_2}\{\varphi(X) = a_1\}$, to zadanie wyznaczenia najlepszej reguły decyzyjnej będziemy mogli rozważać jako zadanie wyznaczenia najmocniejszego testu na poziomie istotności α dla weryfikacji hipotezy $H: \theta = \theta_1$ wobec hipotezy alternatywnej $K: \theta = \theta_2$ (por. wykład 4). ■

Dwa powyższe przykłady pokazują, że teoria decyzji statystycznych może być traktowana jako uogólnienie teorii optymalnej estymacji i teorii optymalnych testów hipotez statystycznych. Teoria decyzji statystycznych pozwala na spojrzenie "z wyższego stanowiska" na to, czym się dotychczas zajmowaliśmy i jednocześnie pozwala na głębsze wniknięcie w naturę rozważanych poprzednio problemów. Pozwala również na istotne rozszerzenie repertuaru rozważanych problemów, choćby o zagadnienia dyskryminacji, o których powiemy w paragrafie 3.

2. Optymalne reguły decyzyjne

2.1. Wprowadzenie

Mówimy, że reguła decyzyjna δ_1 jest nie gorsza od reguły decyzyjnej δ_2 , jeżeli

$$R(\theta, \delta_1) \leq R(\theta, \delta_2) \quad \forall \theta \in \Theta.$$

Mówimy, że reguła decyzyjna δ_1 jest lepsza od reguły decyzyjnej δ_2 , jeżeli δ_1 jest nie gorsza od δ_2 i ponadto $R(\theta, \delta_1) < R(\theta, \delta_2)$ dla pewnego $\theta \in \Theta$.

Przy ustalonym zbiorze \mathcal{D} rozważanych reguł decyzyjnych mówimy, że reguła decyzyjna δ jest *dopuszczalna*, jeżeli w zbiorze \mathcal{D} nie ma reguł lepszych od δ .

Jest oczywiste, że we wszystkich rozważaniach na temat optymalnych reguł decyzyjnych możemy ograniczyć się do zbioru reguł dopuszczalnych. Niestety, pojawiają się tu pewne trudności.

Jeden rodzaj trudności jest związany z dobrym, przejrzystym opisem zbioru reguł dopuszczalnych. To, co się czasami udaje uzyskać, to parametryzacja tego zbioru za pomocą rozkładów prawdopodobieństwa na przestrzeni parametrów Θ : w pewnych modelach statystycznych istnieje wzajemnie jednoznaczne odwzorowanie pomiędzy zbiorem wszystkich reguł dopuszczalnych i zbiorem wszystkich rozkładów prawdopodobieństwa na Θ . W punkcie 2.2 naszego wykładu mówimy o pewnej technice prowadzącej do redukcji zbioru \mathcal{D} do pewnego mniejszego zbioru zawierającego wszystkie reguły dopuszczalne, ale — być może — zawierającego także reguły niedopuszczalne.

Drugi rodzaj trudności jest związany z zastosowaniami: praktyk wolałby otrzymać od statystyka jakąś jedną, dobrze określoną ”optymalną” regułę decyzyjną zamiast całego zbioru reguł dopuszczalnych, w którym, w końcu, i tak musi dokonać pewnego wyboru. Pokonanie tej trudności wymaga jednak wprowadzenia pewnych dodatkowych elementów do naszych rozważań: mówimy o tym w punkcie 2.3 (reguły bayesowskie) i w punkcie 2.4 (reguły minimaksowe).

2.2. Redukcja przez dostateczność

Następujące twierdzenie jest uogólnieniem znanego nam już z teorii estymacji twierdzenia Rao–Blackwella.

TWIERDZENIE 1. *Niech $\mathcal{A} \subset \mathcal{R}^k$ będzie zbiorem wypukłym i niech, dla każdego $\theta \in \Theta$, funkcja $L(\theta, \cdot)$ będzie wypukła.*

Jeżeli $\delta \in \mathcal{D}$ jest regułą decyzyjną i T jest statystyką dostateczną, to reguła decyzyjna

$$\tilde{\delta} = E(\delta(X)|T = t)$$

jest nie gorsza od reguły decyzyjnej δ .

D o w ó d. Dowód wynika natychmiast z nierówności Jensena, jeżeli zapiszemy ją w postaci

$$E(L(\theta, \delta(X))|T = t) \geq L(\theta, E(\delta(X)|T = t)).$$

i zauważymy, że prawa strona tej nierówności to $L(\theta, \tilde{\delta}(t))$. Obliczając wartości oczekiwane E_θ obu stron, otrzymujemy tezę twierdzenia. \square

Przykład 3. *Rozpatrzmy model statystyczny $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})^n$, gdzie P_θ jest rozkładem normalnym $N(\theta, 1)$, $\theta \in \mathbf{R}^1$. Niech $\mathcal{A} = \mathbf{R}^1$ i niech $L(\theta, a) = |\theta - a|$.*

Postać funkcji straty (por. zad. 2) i symetria rozkładu względem θ sugerują, że estymatorem tego parametru mogłaby być mediana $\text{med}(X)$ (wartość środkowa) z próby

X_1, X_2, \dots, X_n , tzn. statystyka pozycyjna $X_{k+1:n}$, gdy $n = 2k + 1$ jest liczbą nieparzystą lub $(\bar{X}_{k:n} + X_{k+1:n})/2$, gdy $n = 2k$ jest liczbą parzystą. W ogólnym przypadku oczywiście jest to dobry estymator, ale w rozważanym tu przypadku rozkładu normalnego średnia \bar{X} z próby jest statystyką dostateczną, więc na mocy twierdzenia 1 estymator $E(\text{med}(X)|\bar{X})$ nie jest gorszy. Ponieważ $E(\text{med}(X)|\bar{X}) = \bar{X}$ (por. zad. 3), średnia z próby nie jest gorsza od mediany. Odnotujmy, że na mocy twierdzenia 1 ten wynik pozostaje w mocy dla dowolnej funkcji straty $L(\theta, a)$, która dla każdego $\theta \in \Theta$ jest wypukłą funkcją argumentu a . ■

2.3. Bayesowskie reguły decyzyjne

Niech τ będzie rozkładem prawdopodobieństwa na zbiorze Θ . Ze względu na tradycję, rozkład τ nazywamy *rozkładem a priori* parametru θ .

Niech δ będzie daną regułą decyzyjną i niech $R(\theta, \delta)$ będzie jej funkcją ryzyka.

Wielkość

$$(2) \quad r(\tau, \delta) = \int R(\theta, \delta) \tau(d\theta)$$

nazywamy *ryzykiem bayesowskim reguły decyzyjnej δ przy rozkładzie a priori τ* .

Przy ustalonym rozkładzie a priori τ mówimy, że reguła decyzyjna δ_1 jest nie gorsza od reguły decyzyjnej δ_2 , gdy $r(\tau, \delta_1) \leq r(\tau, \delta_2)$, i że jest lepsza, gdy $r(\tau, \delta_1) < r(\tau, \delta_2)$. Tak określona relacja w zbiorze \mathcal{D} reguł decyzyjnych porządkuje liniowo ten zbiór.

Jeżeli, przy ustalonym rozkładzie a priori τ , w zbiorze \mathcal{D} istnieje reguła decyzyjna δ_τ taka, że

$$r(\tau, \delta_\tau) \leq r(\tau, \delta) \quad \forall \delta \in \mathcal{D},$$

to regułę decyzyjną δ_τ nazywamy *regułą bayesowską przy rozkładzie a priori τ* .

Efektywne wyznaczenie reguły bayesowskiej nie nastęrcza większych trudności. Konstrukcja opiera się na twierdzeniu Fubiniego

$$(3) \quad \int \left(\int L(\theta, \delta(x)) P_\theta(dx) \right) \tau(d\theta) = \int \left(\int L(\theta, \delta(x)) \tau_x(d\theta) \right) P(dx),$$

gdzie τ_x oraz P są odpowiednimi rozkładami prawdopodobieństwa na Θ oraz X takimi, że

$$(4) \quad P_\theta(dx) \tau(d\theta) = \tau_x(d\theta) P(dx)$$

(por. elementarny wzór Bayesa). Rozkład τ_x nazywa się *rozkładem a posteriori*.

Na mocy wzoru (3), efektywne wyznaczenie bayesowskiej reguły decyzyjnej sprowadza się do wyznaczenia, przy każdym ustalonym $x \in \mathcal{X}$, takiej wartości $a \in \mathcal{A}$, która minimalizuje $\int L(\theta, a) \tau_x(d\theta)$. W szczególności, w problemie estymacji parametru $\theta \in \mathbf{R}^1$ przy kwadratowej funkcji straty $L(\theta, a) = (\theta - a)^2$ estymatorem bayesowskim jest wartość oczekiwana rozkładu a posteriori (por. zad. 1), a przy funkcji straty $L(\theta, a) = |\theta - a|$ estymatorem bayesowskim jest mediana tego rozkładu (por. zad. 2).

Przykład 4. Niech X będzie obserwacją o rozkładzie dwumianowym $b(\theta, n)$. Niech $B(\alpha, \beta)$ będzie rozkładem *a priori* parametru $\theta \in (0, 1)$. Estymujemy parametr θ przy kwadratowej funkcji straty.

Rozkład *a posteriori* parametru θ ma, na mocy wzoru (4), gęstość proporcjonalną do $\theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$, więc jest to rozkład beta $B(\alpha+x, \beta+n-x)$. Wartość oczekiwana zmiennej losowej o rozkładzie beta $B(p, q)$ jest równa $p/(p+q)$, więc estymatorem bayesowskim parametru θ jest $(X+\alpha)/(\alpha+\beta+n)$. ■

Wokół koncepcji rozkładu *a priori* i w ogóle wokół "podejścia bayesowskiego" rozwinęła się poważna kontrowersja między statystykami. Początkowo rozkład *a priori* został wprowadzony jako pewien sposób prezentacji przekonań *a priori* (tzn. przed eksperymentem) co do możliwych wartości nieznanego parametru. Ta wiedza aprioryczna jest w wyniku eksperymentu statystycznego modyfikowana do wiedzy aposteriorycznej, opisywanej za pomocą rozkładu *a posteriori*. Ponieważ rozkład *a posteriori*, a więc i optymalna reguła decyzyjna zależą od wyboru rozkładu *a priori*, racjonalnemu wyborowi tego ostatniego poświęcono wiele prac i wiele dyskusji. Cała ta problematyka jest raczej problematyką epistemologiczną niż tylko statystyczną, chociaż nie zawsze jest to wyraźnie demonstrowane.

Niezależnie od tych dyskusji, bayesowskie reguły decyzyjne odgrywają w konstrukcji dopuszczalnych reguł decyzyjnych taką samą rolę, jak rozwiązania uzyskiwane za pomocą minimalizacji różnych funkcjonatów w technice skalaryzacji przy konstrukcji rozwiązań w wielokryterialnych problemach optymalizacyjnych. Następujące twierdzenie pokazuje ten związek.

TWIERDZENIE 2. *Jeżeli reguła decyzyjna δ jest jedyną regułą bayesowską przy pewnym rozkładzie *a priori*, to δ jest regułą dopuszczalną.*

D o w ó d. Przypuśćmy, że δ jest jedyną regułą bayesowską przy pewnym rozkładzie *a priori* i że δ' jest regułą, która ją dominuje, tzn. regułą taką, że $R(\theta, \delta') \leq R(\theta, \delta) \forall \theta \in \Theta$. Wtedy $r(\tau, \delta') \leq r(\tau, \delta)$ skąd wynika, że $\delta' = \delta$. □

Przykład 5. Niech X będzie zmienną losową o rozkładzie dwumianowym $b(\theta, n)$. Estymujemy θ przy kwadratowej funkcji straty $L(\theta, a) = (\theta - a)^2$. Estymator

$$\theta^* = \frac{X + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}}$$

jest estymatorem dopuszczalnym. Wynika to stąd, że jest to jedyny estymator bayesowski przy rozkładzie *a priori* beta $B(\frac{1}{2}\sqrt{n}, \frac{1}{2}\sqrt{n})$. ■

A oto pewne inne wykorzystanie techniki bayesowskiej: pokażemy, że średnia \bar{X} z próby X_1, X_2, \dots, X_n z rozkładu normalnego $N(\mu, 1)$ jest, przy kwadratowej funkcji strat, estymatorem dopuszczalnym parametru μ .

Ryzyko estymatora \bar{X} jest stałe i równe $1/n$. Przypuśćmy, że \bar{X} nie jest estymatorem dopuszczalnym. Wtedy istnieje estymator $\tilde{\mu}$ taki, że

$$R(\mu, \tilde{\mu}) \leq \frac{1}{n} \quad (\forall \mu \in \mathbf{R}^1),$$

$$R(\mu, \tilde{\mu}) < \frac{1}{n} \quad \text{dla pewnego } \mu \in \mathbf{R}^1.$$

W rozważanym tu problemie, dla każdego estymatora $\tilde{\mu}$ ryzyko $R(\mu, \tilde{\mu})$ jest funkcją ciągłą argumentu $\mu \in \mathbf{R}^1$. Zatem istnieje $\epsilon > 0$ i istnieją liczby μ_1 i μ_2 takie, że

$$R(\mu, \tilde{\mu}) < \frac{1}{n} - \epsilon \quad \forall \mu \in (\mu_1, \mu_2).$$

Weźmy pod uwagę rozkład *a priori* $N(0, \tau^2)$ parametru μ . Wtedy

$$r(\tau, \tilde{\mu}) = \frac{1}{\tau\sqrt{2\pi}} \int_{-\infty}^{+\infty} R(\mu, \tilde{\mu}) e^{-\frac{1}{2}(\frac{\mu}{\tau})^2} d\mu$$

jest ryzykiem bayesowskim estymatora $\tilde{\mu}$ przy tym rozkładzie *a priori* oraz

$$r_\tau = \frac{\tau^2}{1 + n\tau^2}$$

jest ryzykiem bayesowskim estymatora bayesowskiego przy tym rozkładzie *a priori* (por. zad. 4). Weźmy pod uwagę iloraz

$$\begin{aligned} \frac{\frac{1}{n} - r(\tau, \tilde{\mu})}{\frac{1}{n} - r_\tau} &= \frac{\frac{1}{\tau\sqrt{2\pi}} \int_{-\infty}^{+\infty} [\frac{1}{n} - R(\mu, \tilde{\mu})] \exp[-\frac{1}{2}(\frac{\mu}{\tau})^2] d\mu}{\frac{1}{n} - \frac{\tau^2}{1 + n\tau^2}} \\ &\geq \frac{n(1 + n\tau^2)\epsilon}{\tau\sqrt{2\pi}} \int_{\mu_1}^{\mu_2} \exp[-\frac{1}{2}(\frac{\mu}{\tau})^2] d\mu. \end{aligned}$$

Prawa strona tej nierówności dąży do $+\infty$, gdy $\tau \rightarrow +\infty$, więc istnieje takie τ_0 , że $r(\tau_0, \tilde{\mu}) < r_{\tau_0}$, co przeczy temu, że r_{τ_0} jest ryzykiem bayesowskim estymatora bayesowskiego przy rozkładzie *a priori* $N(0, \tau_0^2)$.

Przedstawiony tu wynik nie jest jednak tak naturalny, jak by to się wydawało: okazuje się, że średnia z próby z k -wymiarowego ($k > 2$) rozkładu normalnego $N(\boldsymbol{\mu}, \mathbf{I})$ nie jest estymatorem dopuszczalnym parametru $\boldsymbol{\mu}$ (por. zad. 5).

2.4. Minimaksowe reguły decyzyjne

Reguła decyzyjna δ^* nazywa się *regułą minimaksową*, jeżeli

$$(5) \quad \sup_{\theta \in \Theta} R(\theta, \delta^*) = \inf_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta).$$

W zastosowaniach, np. w ekonomii, reguły minimaksowe są traktowane jako reguły "asekuracyjne": nawet w najgorszej sytuacji mają prowadzić do możliwie dobrego wyniku.

Następujące dwa twierdzenia pozwalają czasami wykazać, że rozważana reguła decyzyjna jest regułą minimaksową.

Twierdzenie 3. *Jeżeli reguła decyzyjna δ jest regułą bayesowską względem pewnego rozkładu a priori τ oraz $R(\theta, \delta) \leq r(\tau, \delta)$ dla każdego θ , to δ jest regułą minimaksową. W szczególności, jeżeli reguła bayesowska ma stałe ryzyko, to jest ona regułą minimaksową.*

D o w ó d. Por. zadanie 6. □

Twierdzenie 4. *Niech $(\delta_n, n \geq 1)$ oraz $(\tau_n, n \geq 1)$ będą ciągami reguł decyzyjnych i rozkładów a priori takich, że δ_n jest regułą bayesowską przy rozkładzie a priori τ_n , $n \geq 1$. Jeżeli $r(\tau_n, \delta_n) \rightarrow C$, gdy $n \rightarrow \infty$, i pewna reguła decyzyjna δ spełnia warunek $R(\theta, \delta) \leq C$, to δ jest regułą minimaksową.*

D o w ó d. Por. zadanie 7. □

Druga część twierdzenia 3 pozwala czasami efektywnie skonstruować minimaksową regułę decyzyjną: wystarczy znaleźć regułę bayesowską o stałym ryzyku.

Przykład 6. Niech X będzie obserwacją o rozkładzie dwumianowym $b(\theta, n)$. Estymujemy θ przy kwadratowej funkcji strat.

Jeżeli rozkład a priori parametru θ jest rozkładem beta $B(\alpha, \beta)$, to (por. przykład 4) estymatorem bayesowskim jest $(X + \alpha)/(\alpha + \beta + n)$. Ryzyko tego estymatora jest równe wariancji rozkładu a posteriori

$$R\left(\theta, \frac{X + \alpha}{\alpha + \beta + n}\right) = \frac{n\theta(1 - \theta) + [\alpha(1 - \theta) - \beta\theta]^2}{(\alpha + \beta + n)^2}.$$

To ryzyko jest stałą funkcją argumentu $\theta \in (0, 1)$, jeżeli

$$(\alpha + \beta)^2 = n \quad \text{oraz} \quad 2\alpha(\alpha + \beta) = n,$$

czyli jeżeli $\alpha = \beta = \frac{1}{2}\sqrt{n}$. Zatem estymatorem minimaksowym w rozważanym problemie jest $\theta^* = (X + \frac{1}{2}\sqrt{n})/(n + \sqrt{n})$ — por. przykład 5. ■

Skorzystamy z twierdzenia 4 dla dowodu ważnego faktu, że — przy kwadratowej funkcji straty — średnia z próby z rozkładu normalnego $N(\mu, \sigma^2)$ o znanej wariancji σ^2 jest estymatorem minimaksowym parametru μ .

Przykład 7. Niech X_1, X_2, \dots, X_n będzie próbą z rozkładu normalnego $N(\mu, \sigma^2)$, gdzie σ^2 jest znaną wariancją; bez straty na ogólności rozważań założymy, że $\sigma^2 = n$. Estymujemy μ przy kwadratowej funkcji straty $L(\mu, a) = (\mu - a)^2$.

Niech $N(0, \tau^2)$ będzie rozkładem *a priori* parametru μ . Wtedy estymatorem bayesowskim jest $\tau^2 \bar{X} / (1 + \tau^2)$, a ryzyko tego estymatora jest równe $\tau^2 / (1 + \tau^2)$ (por. zad. 4).

Ponieważ z jednej strony $\lim_{\tau \rightarrow \infty} \frac{\tau^2}{1 + \tau^2} = 1$ oraz z drugiej strony, ryzyko estymatora \bar{X} jest stałe i równe 1, więc na mocy twierdzenia 4, \bar{X} jest estymatorem minimaxowym. ■

Uogólnienie tego wyniku na przypadek próby z k -wymiarowego ($k \geq 1$) rozkładu normalnego pozostawiamy jako zadanie 8.

3. Zastosowania w dyskryminacji

Teoria dyskryminacji zajmuje się przede wszystkim takimi modelami statystycznymi $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$, w których zbiór $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ jest zbiorem skończonym. Zagadnienie dyskryminacji polega — jak sama nazwa wskazuje — na rozstrzygnięciu, z którego z rozkładów P_{θ_i} pochodzi dana obserwacja X .

W typowych problemach dyskryminacji w powyższym modelu, przestrzeń decyzji \mathcal{A} jest zbiorem $\{1, 2, \dots, k\}$. Jeżeli δ jest regułą decyzyjną, to $\delta(x) = i$ oznacza wskazanie na rozkład P_{θ_i} jako na rozkład, z którego pochodzi dana obserwacja $X = x$.

W bardziej ogólnych sytuacjach (por. dalej przykład 9) rozpatruje się dowolny zbiór Θ i jego rozbicie $\{\Theta_1, \Theta_2, \dots, \Theta_k\}$. Wtedy problem dyskryminacji polega na wskazaniu podzbioru Θ_i , zawierającego to θ , które indeksuje rozkład P_θ danej obserwacji X .

Na zagadnienia dyskryminacji można spojrzeć jak na zagadnienia estymacji parametru $\theta \in \Theta$. Jednak natura problemów dyskryminacji narzuca funkcje straty, nie zawsze odpowiednie dla problemu estymacji.

Na zagadnienie dyskryminacji można również spojrzeć jak na zagadnienie jednoczesnej weryfikacji k hipotez statystycznych (prostych lub, w ogólniejszym przypadku, o którym była przed chwilą mowa, złożonych). Podejście do tego zagadnienia w duchu teorii Neymana–Pearsona wprowadza jednak pewną asymetrię w traktowaniu różnych hipotez, co z punktu widzenia zastosowań nie zawsze jest do zaakceptowania.

Z punktu widzenia teorii decyzji statystycznych, teoria dyskryminacji stanowi pewien szczególny przypadek. Specyfikę tej teorii z punktu widzenia zastosowań zilustrujemy dwoma przykładami.

Przykład 8. Zmienna losowa X ma rozkład normalny $P_\mu = N(\mu, \sigma^2)$ lub rozkład normalny $P_\nu = N(\nu, \sigma^2)$. Na podstawie obserwacji X należy rozstrzygnąć, który z nich. Parametry μ, ν oraz σ^2 są znane; bez straty na ogólności rozważań założymy, że $\mu < \nu$.

Konstrukcja reguły decyzyjnej polega na rozbiciu przestrzeni próby \mathbf{R}^1 na dwa zbiory A_μ i A_ν takie, że jeżeli $X \in A_\mu$, rozstrzygamy na korzyść rozkładu P_μ , a jeżeli $X \in A_\nu$, rozstrzygamy na rzecz rozkładu P_ν .

Prawdopodobieństwa błędnych rozstrzygnięć wynoszą, odpowiednio, $P_\mu\{X \in A_\nu\}$ oraz $P_\nu\{X \in A_\mu\}$. Przyjmijmy te prawdopodobieństwa za kryterium porównywania reguł dyskryminacyjnych. Bardziej formalnie: weźmy pod uwagę funkcję straty

$$L(\theta, a) = \begin{cases} 0, & \text{gdy } \theta = a, \\ 1, & \text{gdy } \theta \neq a. \end{cases}$$

Wtedy ryzyko reguły dyskryminacyjnej δ , wyznaczonej przez zbiory A_μ, A_ν , wyraża się wzorem

$$R(\theta, \delta) = \begin{cases} P_\mu\{X \in A_\nu\}, & \text{gdy } \theta = \mu, \\ P_\nu\{X \in A_\mu\}, & \text{gdy } \theta = \nu. \end{cases}$$

Jeżeli ustalimy liczbę $\alpha \in (0, 1)$, to w klasie wszystkich reguł dyskryminacyjnych, dla których $P_\mu\{X \in A_\nu\} = \alpha$, regułą minimalizującą $P_\nu\{X \in A_\mu\}$ jest reguła δ_m , w której zbiory A_μ i A_ν mają postać

$$A_\mu = \{x: x \leq m\}, \quad A_\nu = \{x: x > m\},$$

przy odpowiednio (dla danego α) wybranej liczbie m (por. teoria testów najmocniejszych). Oczywiście jest, że dwie reguły dyskryminacyjne δ_m oraz $\delta_{m'}$ są nieporównywalne, gdy $m \neq m'$. W konsekwencji stwierdzamy, że $\{\delta_m: m \in \mathbf{R}^1\}$ jest zbiorem wszystkich dopuszczalnych reguł dyskryminacyjnych.

Każda reguła dyskryminacyjna $\delta_m, m \in \mathbf{R}^1$, jest regułą bayesowską (por. zad. 9). Zatem reguła dyskryminacyjna, dla której $P_\mu\{X \in A_\nu\} = P_\nu\{X \in A_\mu\}$, jest regułą minimaxową. ■

Wszystkie powyższe rezultaty otrzymaliśmy jako proste wnioski z ogólnej teorii decyzji statystycznych. W zastosowaniach dyskryminacji pojawiają się jednak pewne dalsze problemy. Zilustrujemy je, jak również pewne ich rozwiązania, konkretnymi przykładami liczbowymi.

Przypuśćmy, że $\mu = 0$, $\nu = 1$ oraz $\sigma^2 = 1$. Minimaxową regułą dyskryminacyjną jest reguła δ_m dla $m = \frac{1}{2}$. Mamy dla niej

$$P_\mu\{X \geq \frac{1}{2}\} = P_\nu\{X \leq \frac{1}{2}\} = 0.31.$$

Trudno jest sobie wyobrazić, żeby reguła dyskryminacyjna z tak dużymi prawdopodobieństwami błędów, choć optymalna, mogła znaleźć praktyczne zastosowanie. (Zauważmy, że dla każdej innej reguły dyskryminacyjnej przynajmniej jedno z tych prawdopodobieństw jest większe od 0.31.)

Oto dwie różne propozycje praktycznych rozwiązań w takiej sytuacji.

Jeżeli daną małą liczbę $\alpha \in (0, 1)$, powiedzmy $\alpha = 0.01$, uznamy za dopuszczalną wartość każdego z błędów, możemy wyznaczyć dwie liczby m_1 i m_2 takie, że

$$P_\mu\{X > m_2\} = P_\nu\{X < m_1\} = \alpha$$

i rozstrzygnąć na rzecz rozkładu P_μ , gdy $X < m_1$, lub na rzecz rozkładu P_ν , gdy $X > m_2$. Teraz jednak $m_1 < m_2$ (dla $\alpha = 0.01$ mamy $m_1 = -1.326$ oraz $m_2 = 2.326$), więc w celu

utrzymania prawdopodobieństw błędnych decyzji na zadanym poziomie α , w przypadku gdy $X \in (m_1, m_2)$, należało by wstrzymać się od wydawania werdyktu. Konkluzja "wynik eksperymentu nie pozwala na rozstrzygnięcie" nie jest czymś niezwykłym.

Inne rozwiązanie możemy zaproponować wtedy, gdy istnieje możliwość przeprowadzenia kilku, powiedzmy n , niezależnych obserwacji X_1, X_2, \dots, X_n zmiennej losowej X . Ponieważ średnia \bar{X} z próby ma rozkład normalny $N(\mu, \sigma^2/n)$ lub $N(\nu, \sigma^2/n)$, gdy X ma, odpowiednio, rozkład $N(\mu, \sigma^2)$ lub $N(\nu, \sigma^2)$, więc dla każdego $\alpha \in (0, 1)$ możemy wyznaczyć takie $n \geq 1$, żeby

$$P_\mu\{X \geq \frac{\mu + \nu}{2}\} = P_\nu\{X \leq \frac{\mu + \nu}{2}\} \leq \alpha.$$

Mamy

$$P_\mu\{X \geq \frac{\mu + \nu}{2}\} = P_\nu\{X \leq \frac{\mu + \nu}{2}\} = \Phi\left(\frac{\mu - \nu}{2\sigma} \sqrt{n}\right),$$

więc wystarczy wybrać

$$n \geq \frac{4\sigma^2}{(\mu - \nu)^2} [\Phi^{-1}(\alpha)]^2.$$

Na przykład, w rozważanym przez nas przypadku otrzymujemy, że dla utrzymania prawdopodobieństw obu błędów na poziomie $\alpha \leq 0.01$ wystarczą $n \geq 22$ obserwacje.

W następnym "przykładzie z życia" mówimy o dyskryminacji w modelu statystycznym $(\mathcal{X}, \{P_\theta : \theta \in \Theta\})$ z nieskończoną przestrzenią parametrów $\Theta = (0, 1)$, w której wyróżniono dwa rozłączne podzbiory i na podstawie obserwacji X należy rozstrzygnąć, do którego z nich należy nieznanne θ .

Przykład 9 (statystyczna kontrola jakości). *W dużej partii produktów znajduje się 100w procent sztuk wadliwych. Nabywca tej partii poddaje jej jakość kontroli wrywkowej, polegającej na tym, że wybiera z niej na chybił-trafił n sztuk i zlicza liczbę X sztuk wadliwych w tej próbce. Jeżeli $X \leq m$, uznaje partię za dobrą, jeżeli $X > m$, kwestionuje jakość partii. Nabywca w zasadzie zgadza się na zaakceptowanie partii jako dobrej, jeżeli jej wadliwość w nie przekracza danej liczby w_0 (w_0 nazywa się wadliwością dopuszczalną), i zdecydowany jest odrzucać partię (wraz z wszystkimi ekonomicznymi konsekwencjami tego aktu), których wadliwość jest większa od danej liczby w_1 (w_1 nazywa się wadliwością dyskwalifikującą).*

Zadanie polega na wyznaczeniu liczb n i m w taki sposób, żeby

$$(6) \quad \begin{aligned} P_w\{X > m\} &\leq \alpha, & \text{gdy } w \leq w_0, \\ P_w\{X \leq m\} &\leq \beta, & \text{gdy } w \geq w_1, \end{aligned}$$

gdzie α i β są stałymi kwantyfikującymi skłonność nabywcy do odrzucania partii dobrej i akceptowania partii wadliwej.

Na mocy twierdzenia 3(b) z wykładu IV, nierówności (6) są równoważne z nierównościami

$$(7) \quad \begin{aligned} P_{w_0}\{X > m\} &\leq \alpha, \\ P_{w_1}\{X \leq m\} &\leq \beta. \end{aligned}$$

Układ nierówności (7) ma nieskończenie wiele rozwiązań. W zastosowaniach wybiera się jednoznaczne rozwiązanie z najmniejszą liczbą n ("najtańsza kontrola jakości"). W Polskich Normach podających przepisy postępowania przy ocenie jakości partii przyjmuje się $\alpha = \beta = 0.05$. Dla wybranych wartości wadliwości dopuszczalnych i wadliwości dyskwalifikujących odpowiednie licznosci prób n i wartości krytyczne m można odczytać w podanych tam tablicach. Na przykład dla $w_0 = 1\%$ i $w_1 = 6.3\%$ mamy $n = 100, m = 2$. ■

4. Zadania

1. Niech X będzie zmienną losową o rozkładzie P takim, że $E_P X^2 < \infty$. Wykazać, że funkcja $E_P(X - t)^2$, $t \in \mathbf{R}^1$, osiąga minimum dla $t = E_P X$.

2. Niech X będzie zmienną losową o rozkładzie P takim, że $E_P |X| < \infty$. Wykazać, że funkcja $E_P |X - t|$, $t \in \mathbf{R}^1$ osiąga minimum dla $t = \text{med}(P)$.

Przypominamy, że medianą $\text{med}(P)$ rozkładu P zmiennej losowej X jest każda liczba m taka, że $P\{X \leq m\} \geq 1/2$ oraz $P\{X \geq m\} \geq 1/2$.

3. Przy założeniach z przykładu 3 wykazać, że $E(\text{med}(X)|T) = \bar{X}$, gdzie $T = \sum_{i=1}^n X_i$.

4. Niech X będzie obserwacją o rozkładzie normalnym $N(\mu, \sigma^2)$ z nieznaną wartością oczekiwaną μ i znaną wariancją σ^2 . Rozważamy zagadnienie estymacji parametru μ przy kwadratowej funkcji straty i przy rozkładzie normalnym $N(\nu, \tau^2)$ parametru μ jako rozkładzie *a priori*, gdzie $(\nu, \tau^2) \in \mathbf{R}^1 \times \mathbf{R}_+^1$ jest ustalone. Wyznaczyć rozkład *a posteriori*, estymator bayesowski i ryzyko tego estymatora.

5. Wykazać, że średnia z próby z k -wymiarowego ($k > 2$) rozkładu normalnego $N(\boldsymbol{\mu}, \mathbf{I})$ nie jest estymatorem dopuszczalnym parametru $\boldsymbol{\mu}$. (Wskazówka. Rozpatrzyć estymator

$$\hat{\boldsymbol{\mu}} = \left(1 - \frac{k-2}{n\|\bar{\mathbf{X}}\|^2}\right) \bar{\mathbf{X}}$$

i wykazać, że różnica $R(\boldsymbol{\mu}, \bar{\mathbf{X}}) - R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) > 0$. W rachunkach pojawiają się całki postaci

$$\int \frac{x_i(x_i - \mu_i)}{\|x\|^2} \exp[-\frac{1}{2}(x_i - \mu_i)^2] dx_i, \text{ które po całkowaniu przez części można zapisać jako}$$

$$\int \left(\frac{1}{\|x\|^2} - \frac{2x_i^2}{\|x\|^4} \right) \exp[-\frac{1}{2}(x_i - \mu_i)^2] dx_i.$$

6. Udowodnić twierdzenie 3.

7. Udowodnić twierdzenie 4.

8. Niech $\bar{\mathbf{X}}$ będzie średnią z próby z k -wymiarowego rozkładu normalnego $N(\boldsymbol{\mu}, \mathbf{I})$. Estymujemy $\boldsymbol{\mu}$ przy kwadratowej funkcji straty $\|\boldsymbol{\mu} - \mathbf{a}\|^2$. Wykazać, że $\bar{\mathbf{X}}$ jest estymatorem minimaxowym.

9. Niech $m \in \mathbf{R}^1$ będzie daną liczbą. Przy założeniach przykładu 8 wykazać, że reguła dyskryminacyjna δ_m jest regułą bayesowską przy pewnym rozkładzie *a priori*. Wyznaczyć ten rozkład. Wyznaczyć regułę minimaxową.

10. Wyznaczyć reguły bayesowskie i reguły minimaxowe dla dyskryminacji

(a) dwóch rozkładów normalnych $N(\mu, \sigma^2)$ i $N(\nu, \tau^2)$;

(b) dwóch k -wymiarowych rozkładów normalnych $N(\boldsymbol{\mu}, \mathbf{I})$ i $N(\boldsymbol{\nu}, \mathbf{I})$.

Podać ogólną metodę konstrukcji reguł decyzyjnych dla dyskryminacji dowolnych dwóch rozkładów prawdopodobieństwa.

11. Rozpatrzyć uogólnienie zagadnienia z przykładu 8 na przypadek k rozkładów normalnych. Podać postać reguł bayesowskich i reguły minimaxowej.

WSKAZÓWKI

Wykład II. Statystyki dostateczne

1. Wykorzystać zasadnicze twierdzenie teorii wielomianów symetrycznych, patrz np. Mostowski A. i Stark M. *Elementy algebry wyższej*, PWN.

2. Statystyka $\sum_{i=1}^n X_i$ ma rozkład Poissona z parametrem $n\theta$.

3. Zauważyć, że

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | X_{1:n} = x_1, X_{2:n} = x_2, \dots, X_{n:n} = x_n\} = \frac{1}{n!}.$$

5. Skorzystać z kryterium faktoryzacji oraz lematów 2 i 3 str. 24 i 25.

6, 7. Skorzystać z kryterium faktoryzacji oraz przykładu 7 str. 26.

8. Rozważyć funkcję $T(X_{1:n}, X_{n:n}) = X_{n:n} - X_{1:n}$.

9. a) Patrz przykład 6 str. 25

b) Statystyka $(X_{1:n}, X_{n:n})$ ma rozkład o gęstości

$$f_{\theta, \tau}(x, y) = \begin{cases} \frac{n(n-1)}{(\tau-\theta)^n} (y-x)^{n-2}, & \text{gd } \theta < x < y < \tau, \\ 0, & \text{w pozostałych przypadkach.} \end{cases}$$

12. Rozważyć funkcje postaci

$$h\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right) = a\left(\sum_{i=1}^n X_i\right)^2 + b\sum_{i=1}^n X_i^2.$$

13. Skorzystać z zadań 1 i 10, przyjąć

$$\mathcal{P}_0 = \left\{ p_\theta(x) = c(\theta) \exp\left(-\sum_{i=1}^n \theta_i x^i\right) \mathbf{1}_{(0, +\infty)}(x) : \theta = (\theta_1, \theta_2, \dots, \theta_n) \in \Theta \right\},$$

gdzie Θ jest pewnym n -wymiarowym zbiorem w \mathbf{R}^n tych wartości parametru, dla których funkcja gęstości jest całkowalna i $c(\theta)$ jest stałą normującą. Zauważyć, że rodzina \mathcal{P}_0 jest rodziną wykładniczą i skorzystać z twierdzeń 5 i 6 str. 29 i 30.

Wykład III. Estymatory nieobciążone o minimalnej wariancji

1. Oba estymatory są nieobciążone, aby wyznaczyć błąd średniokwadratowy wystarczy policzyć wariancje estymatorów.

2. Średnia z próby jest statystyką dostateczną i zupełną oraz $E_\mu(\bar{X})^2 = \mu^2 + \sigma^2/n$, $E_\mu(\bar{X})^3 = \mu^3 + 3\mu\sigma^2/n$, i.t.d.

- 3.** Policzyc wartości oczekiwane, skorzystać z punktu 3.1 str. 37.
- 5.** Pokazać, że dowolny estymator T nieobciążony spełnia $T(\theta-1)+T(\theta)+T(\theta+1) = 3\theta$ i policzyć wariancję.
- 6.** Skorzystać z tego, że $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2$ ma rozkład chi-kwadrat z $n-1$ stopniami swobody.
- 8.** Zauważyć, że

$$\frac{1}{N} \sum_{r=2}^{+\infty} r \cdot n_r = \frac{1}{N} \sum_{i=1}^N X_i \cdot \mathbf{1}_{[2,+\infty)}(X_i)$$

- 9.** Zmienna losowa X jest statystyką dostateczną i zupełną w tym modelu (rodzina wykładnicza), należy więc wyznaczyć funkcję h taką, żeby

$$\sum_{i=1}^n h(i) \binom{n}{i} \theta^i (1-\theta)^{n-i} = \theta^m.$$

- 10, 11.** Pokazać, że X jest statystyką dostateczną i zupełną oraz wyznaczyć funkcję $h(X)$, której wartość oczekiwana jest równa estymowanemu parametrowi.
- 12.** Skorzystać z następujących nierówności

$$\text{Var} \left(\frac{1}{2} (T_1 + T_2) \right) \geq \text{Var}(T_1) = \text{Var}(T_2),$$

$$\text{Cov}^2(T_1, T_2) \leq \text{Var}(T_1)\text{Var}(T_2).$$

- 13.** Wykazać, że $T = \sum_{i=1}^n X_i$ jest statystyką dostateczną i zupełną. Zauważyć, że T ma rozkład gamma $\Gamma(n, 1/\theta)$. Obliczyć $E(Y|T) = P(X_1 \geq k|T)$, gdzie Y jest podaną w treści zadania zmienną losową przyjmującą wartości 0 i 1.

Wykład IV. Testowanie hipotez statystycznych

- 2.** Rozpatrzyć obszar krytyczny postaci $\{X_{n:n} > t_\alpha\}$.
- 3.** Zauważyć, że jeśli $X \sim F_\delta$, to $X - \delta \sim F$ i pokazać, że $P_\delta\{S \geq s\} \leq P_0\{S \geq s\}$ dla $\delta \leq 0$.
- 4.** Zastosować lemat Neymana-Pearsona.

8. Zastosować lemat Neymana-Pearsona i pokazać, że

$$\frac{\exp\left(-\sum_{i=1}^r(x_i - \frac{1}{2})^2 - \sum_{i=r+1}^n(x_i + \frac{1}{2})^2\right)}{\exp\left(-\sum_{i=1}^n x_i^2\right)} > t \iff \sum_{i=1}^r x_i - \sum_{i=r+1}^n x_i > t_1.$$

9. Pokazać, że każdy test o obszarze krytycznym postaci $Z = A \cup [\theta_0, +\infty)$, gdzie $A \subseteq (0, \theta_0)$ spełnia warunek $P_0\{X \in A\} = \alpha$ i P_0 oznacza prawdopodobieństwo przy rozkładzie $U(0, \theta_0)$, jest testem najmocniejszym dla weryfikacji hipotezy $H : \theta = \theta_0$ przy alternatywie $K : \theta = \theta_1, \theta_0 < \theta_1$.

10. Rozważyć hipotezę $H : \theta = \theta_0$ przy alternatywie $K_1 : \theta > \theta_0$, a następnie hipotezę $H : \theta = \theta_0$ przy alternatywie $K_2 : \theta < \theta_0$. Porównać postać zbiorów krytycznych.

11. Zauważyć, że rodzina rozważanych rozkładów jest rodziną z monotonicznym ilorzem wiarygodności.

12. Wykorzystać wskazówkę podaną w treści zadania i zadanie 10 str. 70.

13. Wyznaczyć rodzinę rozkładów obserwowanej zmiennej losowej, pokazać, że jest to rodzina z monotonicznym ilorzem wiarygodności, skonstruować test JNM . Pokazać, że statystyka testowa jest sumą niezależnych zmiennych losowych o tym samym rozkładzie wykładniczym (skorzystać z zadania 6 str. 14).

Wykład V. Wiarygodność

2. Rozważyć nierówność $P_{M+1}\{X = x\}/P_M\{X = x\} \geq 1$.

4. Wyznaczyć estymatory największej wiarygodności parametrów μ i σ w modelu normalnym na podstawie próby (X_1, X_2, \dots, X_n) , gdzie $X_i = \log Y_i$.

5. Skorzystać z kryterium faktoryzacji.

6. Jeżeli $Y_j \sim N(\mu_2, \sigma^2)$, $j = 1, 2, \dots, n$ i $X_i \sim N(\mu_1, \sigma^2)$, $i = 1, 2, \dots, m$, to statystyka testowa t_{m+n-2} ma niecentralny rozkład t-Studenta z $m+n-2$ stopniami swobody i parametrem niecentralności $\sigma^{-1}(\mu_2 - \mu_1)/\sqrt{1/n + 1/m}$.

Wykład VI. Metoda najmniejszych kwadratów. Modele liniowe

2. Skorzystać z tego, że $\text{Im}X$ jest ortogonalnym dopełnieniem jądra przekształcenia X^T .

4, 5, 6, 7, 8. Wykorzystać wzór 4 str. 87 w odpowiednich modelach.

Wykład VII. Teoria decyzji statystycznych

1. Zróżniczkować $f(t) = E_P(X - t)^2$ względem t .
3. Dla każdego ustalonego $j = 1, 2, \dots, [n/2]$, zmienne losowe

$$\text{med}(X) - X_{j:n} \quad \text{oraz} \quad X_{n-j+1:n} - \text{med}(X)$$

mają taki sam rozkład.

4. Rozkład *a posteriori* ma gęstość proporcjonalną do iloczynu gęstości rozkładu *a priori* i obserwowanej zmiennej losowej, stąd zauważyć, że jest to rozkład normalny.
8. Wyznaczyć ryzyko i zastosować twierdzenie 4 str. 99.
- 9, 10. Rozważyć dwupunktowe rozkłady *a priori* i wyznaczyć reguły bayesowskie.
11. Rozważyć k -punktowe rozkłady *a priori* i wyznaczyć reguły bayesowskie jak w zadaniu 9.

ROZWIĄZANIA

Wykład II. Statystyki dostateczne

1. Równoważność statystyk T i U . Weźmy pod uwagę wielomian

$$W_n(t) = \prod_{i=1}^n (t - X_i).$$

W tym wielomianie U_1, U_2, \dots, U_n (z odpowiednimi znakami) są współczynnikami przy $t^{n-1}, t^{n-2}, \dots, t^0$ (współczynnik przy t^n jest równy 1), natomiast X_1, X_2, \dots, X_n są pierwiastkami.

Równoważność statystyk U i S . Na mocy wzorów Newtona (teoria wielomianów symetrycznych) mamy

$$S_k - U_1 S_{k-1} + U_2 S_{k-2} + \dots + (-1)^{k-1} U_{k-1} S^1 + (-1)^k U_k = 0,$$

gdzie $1 \leq k \leq n$. Dla $k = 1$ mamy $S_1 - U_1 = 0$, dla $k = 2$ mamy $S_2 - U_1 S_1 + U_2 = 0$ itd, więc (U_1, U_2, \dots, U_n) jednoznacznie wyznaczają (S_1, S_2, \dots, S_n) i odwrotnie.

2. Rozkład warunkowy próby pod warunkiem, że $T = t$ jest następujący

$$P_\theta\{X_1 = x_1, \dots, X_n = x_n | T = t\} = \begin{cases} \frac{t!}{n^t x_1! x_2! \dots x_n!}, & \text{gdzie } t = \sum_{i=1}^n x_i, \\ 0, & \text{gdzie } t \neq \sum_{i=1}^n x_i, \end{cases}$$

gdzie x_i są liczbami całkowitymi nieujemnymi. Rozkład ten nie zależy od parametru θ .

4. Gęstość zmiennej losowej (X_1, X_2, \dots, X_n) jest równa

$$p_\theta(x_1, x_2, \dots, x_n) = \mathbf{1}_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x_{1:n}) \mathbf{1}_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x_{n:n}),$$

zatem z kryterium faktoryzacji statystyka $T = (X_{1:n}, X_{n:n})$ jest dostateczna.

5. Statystyki $(X_{1:n}, \sum_{i=1}^n (X_i - X_{1:n}))$ i $T = (X_{1:n}, \sum_{i=1}^n X_i)$ są równoważne. Pokażemy, że T jest minimalną statystyką dostateczną. Gęstość zmiennej losowej (X_1, X_2, \dots, X_n) jest równa

$$f_{\theta, \beta}(x_1, x_2, \dots, x_n) = \beta^{-n} \exp(n\theta) \exp(-\beta \sum_{i=1}^n x_i) \mathbf{1}_{[\theta, \infty)}(x_{1:n}),$$

zatem z kryterium faktoryzacji statystyka T jest dostateczna. Aby pokazać, że T jest minimalną statystyką dostateczną rozważamy podrodzinę rozkładów $\mathcal{P}_0 = \{f_{w_i, j} : w_i \in Q \wedge j = 1, 2\}$ (Q oznacza zbiór liczb wymiernych) i rozkład $f_{\lambda, 1} = \sum_{i=1}^{\infty} \lambda_i f_{w_i, 1}$, gdzie $\lambda_i > 0$ i $\sum_{i=1}^{\infty} \lambda_i = 1$. Podrodzina \mathcal{P}_0 jest równoważna z rodziną z treści zadania, a korzystając z lematu 2 str. 24 statystyka

$$S = (s_1(X_1, X_2, \dots, X_n), s_2(X_1, X_2, \dots, X_n), \dots),$$

gdzie

$$s_i = \frac{f_{w_i,2}(X_1, X_2, \dots, X_n)}{f_{\Lambda,1}(X_1, X_2, \dots, X_n)},$$

jest minimalną statystyką dostateczną. Statystyki T i S są równoważne (wystarczy za-
uważyć, że $x_{1:n} = \sup\{w_i : s_i(x_1, x_2, \dots, x_n) > 0\}$ oraz

$$\sum_{i=1}^n x_i = -\ln \left(e^{-nw_k} \sum_{i=1}^{\infty} \lambda_i e^{nw_i} s_k(x_1, x_2, \dots, x_n) \right)$$

dla $s_k > 0$).

8. Dystrybuanta rozkładu statystyki $T(X_{1:n}, X_{n:n}) = X_{n:n} - X_{1:n}$ nie zależy od parametru θ , ponieważ

$$P_{\theta}\{X_{n:n} - X_{1:n} < x\} = P_{\theta}\{(X_{n:n} - \theta) - (X_{1:n} - \theta) < x\} = P_0\{X_{n:n} - X_{1:n} < x\}.$$

Zatem dla każdej wartości θ wartość oczekiwana $E_{\theta}T - c = 0$, gdzie $c = ET$, ale funkcja $h(X_{1:n}, X_{n:n}) = X_{n:n} - X_{1:n} - c$ nie jest równa zero.

9. Należy wykazać, że jeżeli

$$(w) \quad \forall \theta, \tau; \theta < \tau \quad \int_{\theta}^{\tau} \int_{\theta}^y f(x, y)(y-x)^m dx dy = 0,$$

to $f(x, y) = 0$ (pomijamy tu sprawy postaci " $f(x, y) = 0$ dla prawie wszystkich x, y przy mierze Lebesgue'a").

Weźmy pod uwagę funkcję

$$g_{\theta}(y) = \int_{\theta}^y f(x, y)(y-x)^m dx.$$

Warunek (w) ma postać

$$\forall \tau > \theta \quad \int_{\theta}^{\tau} g_{\theta}(y) dy = 0.$$

Przy ustalonym θ , tak jest tylko wtedy, gdy $g_{\theta}(y) = 0$ dla każdego y . Zatem

$$\forall y \quad \int_{\theta}^{\tau} f(x, y)(y-x)^m dx = 0,$$

więc $f(x, y) = 0$ dla każdego x i y .

10. Niech h będzie funkcją spełniającą warunek

$$\forall P \in \mathcal{P}_1 \quad E_P h(X) = 0.$$

Wtedy

$$\forall P \in \mathcal{P}_0 \quad E_P h(X) = 0.$$

Rodzina \mathcal{P}_0 jest zupełna i każdy zbiór zerowy w \mathcal{P}_0 jest zbiorem zerowym w \mathcal{P}_1 , zatem

$$h \equiv 0(\mathcal{P}_0 - p.w.) \implies h \equiv 0(\mathcal{P}_1 - p.w.),$$

co dowodzi zupełności rodziny \mathcal{P}_1 .

Rodzina rozkładów dwumianowych jest zupełna jako rodzina wykładnicza. Aby pokazać, że rodzina $\mathcal{P}_1 = \mathcal{P}_0 \cup \{Q\}$ nie jest zupełna rozważyc funkcję

$$h(x) = \begin{cases} 0 & \text{gd}y \ x = 0, 1, 2, \dots, n, \\ (n+1)!e & \text{gd}y \ x = n+1, \\ -(n+2)!e & \text{gd}y \ x = n+2, \\ 0 & \text{gd}y \ x > n+2. \end{cases}$$

11. Rodzina rozkładów wektora (X_1, X_2, \dots, X_n) jest rodziną wykładniczą, korzystamy więc z twierdzenia 5.

12. Jeżeli (X_1, X_2, \dots, X_n) jest próbą losową z rozkładu $N(\mu, \kappa^2 \mu^2)$, to

$$E_\mu \left(\sum_{i=1}^n X_i \right)^2 = (n\kappa^2 + n^2)\mu^2 \quad \text{oraz} \quad E_\mu \left(\sum_{i=1}^n X_i^2 \right) = n(\kappa^2 + 1)\mu^2.$$

Niech

$$h \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right) = \left(\sum_{i=1}^n X_i \right)^2 - \frac{\kappa^2 + n}{\kappa^2 + 1} \sum_{i=1}^n X_i^2,$$

wtedy

$$\forall \mu > 0 \quad E_\mu h \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right) = 0,$$

ale h nie jest funkcją równą zero.

Wykład III. Estymatory nieobciążone o minimalnej wariancji

1.

$$\text{Var} \hat{\lambda} = \frac{1}{n} \lambda(1-\lambda),$$

$$\text{Var} \lambda^* = \lambda^{2-\frac{1}{n}} - \lambda^2.$$

Zauważyć, że $\text{Var} \lambda^* / \text{Var} \hat{\lambda} < 1$ dla każdego λ oraz n .

2.

$$\text{ENMW}(\mu^2) = \bar{X}^2 - \frac{\sigma^2}{n},$$

$$\text{ENMW}(\mu^3) = \bar{X}^3 - 3\bar{X} \frac{\sigma^2}{n},$$

$$ENMW(\mu^4) = \bar{X}^4 - \binom{4}{2} ENMW(\mu^2) \frac{\sigma^2}{n} - 3 \left(\frac{\sigma^2}{n} \right)^2.$$

Ogólnie: dla k parzystych

$$ENMW(\mu^k) = \bar{X}^k - \binom{k}{2} ENMW(\mu^{k-2}) \frac{\sigma^2}{n} - 3 \binom{k}{4} ENMW(\mu^{k-4}) \left(\frac{\sigma^2}{n} \right)^2 - \dots - \binom{k}{k} (k-1)!! \left(\frac{\sigma^2}{n} \right)^{\frac{k}{2}},$$

a dla k nieparzystych

$$ENMW(\mu^k) = \bar{X}^k - \binom{k}{2} ENMW(\mu^{k-2}) \frac{\sigma^2}{n} - 3 \binom{k}{4} ENMW(\mu^{k-4}) \left(\frac{\sigma^2}{n} \right)^2 - \dots - (k-2)!! \binom{k}{k-1} \left(\frac{\sigma^2}{n} \right)^{\frac{k-1}{2}}.$$

3. Statystyka $T = (\bar{X}, S^2)$ jest statystyką dostateczną i zupełną w rozważanym modelu. Statystyki podane w punkcie 3.4 str. 38 są estymatorami nieobciążonymi odpowiednich parametrów i funkcjami statystyki T .

5. Niech $T(\theta - 1) = x$, $T(\theta) = y$ i $T(\theta + 1) = z$. Funkcja $Var_{\theta} T = \frac{1}{3}(x^2 + y^2 + z^2) - \theta^2$ osiąga minimum przy warunku $x + y + z = 3\theta$, gdy $x = y = z = \theta$, a to dowodzi braku $ENMW(\theta)$.

6. Niech $S_c^2 = c \sum_{i=1}^n (X_i - \bar{X})^2$. Wtedy

$$\frac{1}{c\sigma^2} S_c^2 \sim \chi_{n-1}^2 \implies E_{\sigma^2} S_c^2 = c(n-1)\sigma^2 \quad \wedge \quad Var_{\sigma^2} S_c^2 = 2c^2(n-1)\sigma^4.$$

Błąd średniokwadratowy estymatora S_c^2 jest równy

$$E_{\sigma^2} (S_c^2 - \sigma^2)^2 = Var_{\sigma^2} S_c^2 + (E_{\sigma^2} S_c^2 - \sigma^2)^2 = \sigma^4 [c^2(n^2 - 1) - 2c(n-1) + 1].$$

8. Otrzymujemy

$$\begin{aligned} E_{\theta} \theta^* &= E_{\theta} \left(\frac{1}{N} \sum_{i=1}^N X_i \cdot \mathbf{1}_{[2, +\infty)}(X_i) \right) \\ &= E_{\theta} (X_1 \cdot \mathbf{1}_{[2, +\infty)}(X_1)) = \frac{e^{-\theta}}{1 - e^{-\theta}} \sum_{x=2}^{+\infty} x \frac{\theta^x}{x!} = \frac{e^{-\theta} \theta}{1 - e^{-\theta}} \sum_{x=1}^{+\infty} \frac{\theta^x}{x!} = \theta. \end{aligned}$$

9. $ENMW(\theta^m)$ istnieje, jeśli $0 \leq m \leq n$ oraz

$$ENMW(\theta^m)(x) = \begin{cases} 0, & \text{gdy } 0 \leq x < m, \\ \frac{1}{\binom{n}{m}}, & \text{gdy } x = m, \\ \frac{\binom{x}{m}}{\binom{n}{m}}, & \text{gdy } m < x \leq n. \end{cases}$$

10. Szukamy funkcji $h(x)$ określonej dla $x = 0, 1, 2, \dots$ takiej, że $E_\theta h(X) = \theta^{-1}$. Zatem

$$\theta E_\theta h(X) = \sum_{x=0}^{+\infty} h(x) \binom{m+x-1}{x} \theta^{m+1} (1-\theta)^x = 1.$$

Równość zachodzi, jeśli

$$h(x) \binom{m+x-1}{x} = \binom{m+x}{x},$$

stąd

$$h(x) = 1 + \frac{x}{m}.$$

11. Dowodzimy, że rodzina rozważanych rozkładów jest zupełna. Niech g spełnia

$$\forall M = 0, 1, 2, \dots, N \quad E_M g(X) = 0.$$

Wtedy

$$E_0 g(X) = g(0) \cdot 1 = 0 \implies g(0) = 0,$$

$$E_1 g(X) = g(0) \frac{\binom{N-1}{n}}{\binom{N}{n}} + g(1) \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = 0 \implies g(1) = 0$$

i analogicznie dla pozostałych wartości zmiennej losowej X . Wartość oczekiwana $E_M X = nM/N$, stąd $ENMW(M) = NX/n$.

12. Statystyka $T = \frac{1}{2}(T_1 + T_2)$ jest również estymatorem nieobciążonym, zatem

$$\text{Var} \left(\frac{1}{2}(T_1 + T_2) \right) = \frac{1}{4} (\text{Var}T_1 + \text{Var}T_2) + \frac{1}{2} \text{Cov}(T_1, T_2) \geq \text{Var}T_1 = \text{Var}T_2.$$

Stąd

$$\text{Cov}(T_1, T_2) \geq \text{Var}T_1.$$

Z własności kowariancji

$$\text{Cov}^2(T_1, T_2) \leq \text{Var}T_1 \cdot \text{Var}T_2.$$

Z dwóch ostatnich nierówności otrzymujemy, że

$$\text{Cov}(T_1, T_2) = \text{Var}T_1 = \text{Var}T_2,$$

stąd $T_1 = aT_2 + b$ i $a > 0$, ale $ET_1 = ET_2$ i $\text{Var}T_1 = \text{Var}T_2$ zatem $a = 1$ i $b = 0$.

13. Niech

$$Y = \begin{cases} 1 & \text{gd}y X_1 \geq k, \\ 0 & \text{gd}y X_1 < k. \end{cases}$$

Wtedy $E_\theta Y = P_\theta(X_1 \geq k) = e^{-k\theta}$. Statystyka $T = X_1 + X_2 + \dots + X_n$ jest statystyką dostateczną i zupełną (wykładnicza rodzina rozkładów) i T ma rozkład gamma $\Gamma(n, 1/\theta)$ (por. zadanie 5 str.14). Z twierdzeń 1 i 2 (str. 34 i 35) wynika, że $Z = E(Y|T)$ jest $ENMW(e^{-k\theta})$.

Pokażemy, że $Z = \hat{g}(T)$. Mamy

$$E(Y|T = t) = P\{X_1 \geq k|T = t\}.$$

Dystrybuanta łącznego rozkładu zmiennej losowej (X_1, T) jest równa

$$\begin{aligned} F_\theta(x, t) &= P_\theta \left\{ X_1 \leq x \wedge \sum_{i=2}^n X_i \leq t - X_1 \right\} \\ &= \int_0^x \int_0^{t-y} \frac{\theta^n}{\Gamma(n-1)} z^{n-2} e^{-\theta z} e^{-\theta y} dz dy. \end{aligned}$$

Różniczkując otrzymujemy gęstość rozkładu zmiennej (X_1, T)

$$f_\theta(x, t) = \frac{\theta^n}{\Gamma(n-1)} (t-x)^{n-2} e^{-\theta t} \quad \text{gd}y \quad 0 < x \leq t < +\infty.$$

Zatem gęstość rozkładu zmiennej X_1 pod warunkiem $T = t$ jest równa

$$f(x|t) = (n-1) \frac{(t-x)^{n-2}}{t^{n-1}} \quad \text{dla} \quad 0 < x \leq t$$

i stąd

$$E(Y|T = t) = \begin{cases} 0, & \text{gd}y \quad t < k, \\ \frac{n-1}{t^{n-1}} \int_k^t (t-x)^{n-2} dx = \left(1 - \frac{k}{t}\right)^{n-1}, & \text{gd}y \quad t \geq k. \end{cases}$$

Wykład IV. Testowanie hipotez statystycznych

1. Uporządkować obserwacje w ciąg rosnący $x_{1:6}, x_{2:6}, \dots, x_{6:6}$.

Obliczyć $z_i = F(x_{i:6}) = 1 - \exp(-2x_{i:6})$, $i = 1, 2, \dots, 6$.

Obliczyć $D^+ = \max_{i=1,2,\dots,6} (\frac{i}{6} - z_i)$ oraz $D^- = \max_{i=1,2,\dots,6} (z_i - \frac{i-1}{6})$.

Odczytać z tablic wartość krytyczną $D_6(0.01) = 0.61661$.

Jeżeli $\max(D^+, D^-) > D_6(0.01)$, to odrzucić hipotezę z treści zadania.

2. Obszar krytyczny testu jest postaci

$$K = \{(x_1, x_2, \dots, x_n) : x_{n:n} > t_\alpha\},$$

gdzie $x_{n:n} = \max\{x_1, x_2, \dots, x_n\}$ i t_α spełnia warunek $P_{\theta_0}(K) = \alpha$. Zmienna losowa $X_{n:n}$ ma rozkład o gęstości

$$f_\theta(x) = \frac{n}{\theta^n} x^{n-1} \mathbf{1}_{(0,\theta)}(x),$$

zatem $t_\alpha = \theta_0(1 - \alpha)^{1/n}$.

3. Niech $\delta < 0$. Wtedy

$$S(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n) \leq S(X_1 - \delta, X_2 - \delta, \dots, X_m - \delta, Y_1, \dots, Y_n),$$

stąd

$$P_\delta\{S \geq s\} \leq P_0\{S \geq s\}.$$

Zatem s wyliczamy z warunku $P_0\{S \geq s\} \leq \alpha$.

Jeżeli $\delta = 0$, to zmienne losowe $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ są niezależne o tym samym rozkładzie i

$$P_0\{S \geq s\} = \frac{\binom{m+n-s}{m-s}}{\binom{m+n}{n}}.$$

Przy $m = n = 5$ i $\alpha = 0.01$ otrzymujemy

$$P_0\{S \geq 5\} = \frac{1}{252} < 0.01 \quad \text{i} \quad P_0\{S \geq 4\} = \frac{1}{42} > 0.01,$$

Aby otrzymać test o rozmiarze $\alpha = 0.01$ dokonujemy randomizacji otrzymując test

$$\phi(S) = \begin{cases} 1 & \text{gd}y \ S = 5, \\ \frac{38}{125} & \text{gd}y \ S = 4, \\ 0 & \text{gd}y \ S < 4. \end{cases}$$

4. Prawdziwa jest równoważność

$$\frac{f_1(x)}{f_0(x)} > c \iff ||x| - 1| > t \iff \begin{cases} |x| > t + 1, & \text{gd}y \ t \geq 1, \\ |x| > t + 1 \vee |x| < 1 - t, & \text{gd}y \ 0 < t < 1. \end{cases}$$

Jeżeli $X \sim N(0, 1)$, to $P\{|X| > 2\} = 0.0455$.

Rozważamy dwa przypadki.

Jeżeli $\alpha \leq 0.0455$, to obszar krytyczny testu najmocniejszego jest postaci

$$W = \{x : |x| > t + 1\}, \quad \text{gdzie} \quad t + 1 = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Jeżeli $\alpha > 0.0455$, to

$$W = \{x : |x| > 1 + t \vee |x| < 1 - t\}, \quad \text{gdzie} \quad \Phi(1 + t) - \Phi(1 - t) = \frac{1 - \alpha}{2}.$$

5. Funkcja prawdopodobieństwa błędu pierwszego rodzaju (gdzie $\theta \leq \frac{1}{2}$):

$$f_1(\theta) = P_\theta\{X \geq 12\} = \sum_{x=12}^{20} \binom{20}{x} \theta^x (1 - \theta)^{20-x};$$

funkcja prawdopodobieństwa błędu drugiego rodzaju (gdzie $\theta > \frac{1}{2}$):

$$f_2(\theta) = \sum_{x=0}^{11} \binom{20}{x} \theta^x (1 - \theta)^{20-x}.$$

6. Statystyka $T = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ ma rozkład $N(\mu\sqrt{n}, 1)$, stąd rozmiar testu α jest równy

$$\alpha = P_0\{|T| > 2\} = 2(1 - \Phi(2)) = 0.0455,$$

a funkcja mocy $\beta(\mu)$ ma postać

$$\beta(\mu) = P_\mu\{|T| > 2\} = \Phi(-2 - \mu\sqrt{n}) + 1 - \Phi(2 - \mu\sqrt{n}).$$

7. Z lematu Neymana-Pearsona istnieje test najmocniejszy postaci

$$\phi(x) = \begin{cases} 1 & \text{gdy } f_1(x) > tf_0(x), \\ \gamma & \text{gdy } f_1(x) = tf_0(x), \\ 0 & \text{gdy } f_1(x) < tf_0(x), \end{cases}$$

gdzie f_1, f_0 oznaczają gęstości rozkładów P_1 i P_0 .

Jeżeli $P_0\{x : f_1(x) > tf_0(x)\} = \alpha$, to $\gamma = 0$ i powyższy test jest testem niezrandomizowanym.

Jeżeli $P_0\{x : f_1(x) > tf_0(x)\} = b < \alpha$ i $P_0\{x : f_1(x) = tf_0(x)\} = a \geq \alpha - b$, to z lematu Halmosa istnieje zbiór $B \subseteq \{x : f_1(x) = tf_0(x)\}$ taki, że $P_0(B) = \alpha - b$. Test o obszarze krytycznym

$$W = \{x : f_1(x) > tf_0(x)\} \cup B$$

jest testem niezrandomizowanym i spełnia warunek dostateczny dla testu najmocniejszego w lemacie Neymana-Pearsona.

8. Korzystając ze wskazówki obszar krytyczny testu jest postaci

$$W = \left\{ (x_1, x_2, \dots, x_n) : \sum_{i=1}^r x_i - \sum_{i=r+1}^n x_i > t_1 \right\}.$$

Zmienna losowa $\sum_{i=1}^r X_i - \sum_{i=r+1}^n X_i$ przy prawdziwości hipotezy H ma rozkład $N(0, n)$, stąd

$$P_0(W) = 1 - \Phi\left(\frac{t_1}{\sqrt{n}}\right) = 0.05 \implies t_1 = 1.645\sqrt{n}.$$

Przy alternatywie zmienna losowa $\sum_{i=1}^r X_i - \sum_{i=r+1}^n X_i$ ma rozkład $N(n/2, n)$, stąd moc testu

$$\beta = 1 - \Phi\left(1.645 - \frac{\sqrt{n}}{2}\right).$$

Aby wyznaczyć liczebność n rozwiązujemy nierówność $1 - \Phi\left(1.645 - \frac{\sqrt{n}}{2}\right) > 0.9$ i otrzymujemy $n \geq 35$.

9. Iloraz gęstości

$$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = \begin{cases} \frac{\theta_0}{\theta_1} & \text{gdy } x \in (0, \theta_0), \\ +\infty & \text{gdy } x \geq \theta_0. \end{cases}$$

Niech A będzie podzbiorem zbioru $(0, \theta_0)$, takim że $P_{\theta_0}\{X \in A\} = \alpha$. Wtedy zbiór $Z = A \cup [\theta_0, +\infty)$ jest zbiorem krytycznym testu najmocniejszego (spełnia warunek

wystarczający dla testu najmocniejszego w lemacie Neymana-Pearsona). Moc testu jest równa

$$\beta = P_{\theta_1}\{X \in A\} + P_{\theta_1}\{X \geq \theta_0\} = \frac{\alpha\theta_0}{\theta_1} + \frac{\theta_1 - \theta_0}{\theta_1} = 1 - (1 - \alpha)\frac{\theta_0}{\theta_1}.$$

Rodzina rozkładów jednostajnych jest rodziną z monotonicznym ilorazem wiarygodności, stąd obszar krytyczny testu *JNM* dla testowania $H : \theta \leq \theta_0$ przeciwko $K : \theta > \theta_0$ jest postaci

$$W = \{x : x > c\},$$

gdzie c spełnia równanie

$$P_{\theta_0}(W) = \int_c^{\theta_0} \frac{1}{\theta_0} dx = 1 - \frac{c}{\theta_0} = \alpha,$$

stąd $c = \theta_0(1 - \alpha)$. Zatem wśród testów najmocniejszych dla testowania hipotezy $H : \theta = \theta_0$ przeciwko $K : \theta = \theta_1$ jest test *JNM* dla testowania $H : \theta \leq \theta_0$ przeciwko $K : \theta > \theta_0$.

10. Iloraz gęstości przy $\theta_1 > \theta_0$ jest postaci

$$\frac{p_{\theta_1}(x_1, x_2, \dots, x_n)}{p_{\theta_0}(x_1, x_2, \dots, x_n)} = \begin{cases} \left(\frac{\theta_0}{\theta_1}\right)^n & \text{gdy } x_{n:n} \in (0, \theta_0), \\ +\infty & \text{gdy } x_{n:n} \geq \theta_0, \end{cases}$$

a przy $\theta_1 < \theta_0$ jest postaci

$$\frac{p_{\theta_1}(x_1, x_2, \dots, x_n)}{p_{\theta_0}(x_1, x_2, \dots, x_n)} = \begin{cases} \left(\frac{\theta_0}{\theta_1}\right)^n & \text{gdy } x_{n:n} \in (0, \theta_1), \\ 0 & \text{gdy } x_{n:n} \in [\theta_1, \theta_0]. \end{cases}$$

Przeprowadzając rozumowanie analogiczne jak w zadaniu 9, dla weryfikacji hipotezy $H : \theta = \theta_0$ wobec hipotezy alternatywnej $K : \theta > \theta_0$ otrzymujemy, że test o obszarze krytycznym $Z = \{(x_1, x_2, \dots, x_n) : x_{n:n} \in A \cup [\theta_0, +\infty)\}$, gdzie $A \subseteq (0, \theta_0)$ spełnia warunek $P_{\theta_0}\{X_{n:n} \in A\} = \alpha$, jest testem *JNM* na poziomie istotności α .

Dla weryfikacji hipotezy $H : \theta = \theta_0$ wobec hipotezy alternatywnej $K : \theta < \theta_0$ otrzymujemy, że test o obszarze krytycznym $W = \{(x_1, x_2, \dots, x_n) : x_{n:n} \in (0, c) \cup [\theta_0, +\infty)\}$, gdzie c spełnia warunek $P_{\theta_0}\{X_{n:n} < c\} = \alpha$, jest testem *JNM* na poziomie istotności α (test o obszarze krytycznym W spełnia warunek wystarczający z lematu Neymana-Pearsona dla każdej alternatywy postaci $K : \theta = \theta_1$, gdzie $\theta_1 < \theta_0$ i nie zależy od wyboru θ_1).

Wybierając za zbiór A przedział $(0, c)$, gdzie $c = \theta_0\alpha^{1/n}$, otrzymujemy $Z = W$. Zatem test o obszarze krytycznym W jest testem *JNM* dla weryfikacji hipotezy $H : \theta = \theta_0$ wobec hipotezy alternatywnej $K : \theta \neq \theta_0$ na poziomie istotności α .

11. Zmienna losowa $X_{1:n}$ ma rozkład o gęstości

$$g_{\theta}(x) = n \exp(-n(x - \theta)) \mathbf{1}_{(\theta, +\infty)}(x),$$

stąd

$$P_{\theta}\{X_{1:n} > c\} = \begin{cases} \exp(-n(c - \theta)), & \text{gdy } \theta < c, \\ 1, & \text{gdy } \theta \geq c \end{cases}$$

jest rosnącą funkcją zmiennej θ . Wartość c wyliczamy z równania $\exp(-n(c-1)) = \alpha$ otrzymując $c = 1 - \frac{1}{n} \ln \alpha$.

Niech $\theta_1 > \theta_0$, wtedy

$$\frac{f_{\theta_1}(x_1, x_2, \dots, x_n)}{f_{\theta_0}(x_1, x_2, \dots, x_n)} = \begin{cases} 0, & \text{gd } x_{1:n} \leq \theta_1, \\ \exp(n(\theta_1 - \theta_0)), & \text{gd } x_{1:n} > \theta_1, \end{cases}$$

zatem rozważana rodzina rozkładów jest rodziną z monotonicznym ilorazem wiarygodności względem statystyki $X_{1:n}$ i z twierdzenia 3 ze str. 66 otrzymujemy, że skonstruowany test jest testem *JNM*.

12. a) Korzystając ze wskazówki, zadanie sprowadza się do wyznaczenia testu *JNM* dla weryfikowania hipotezy $H : \theta = \theta_0$ przy alternatywie $K : \theta \neq \theta_0$, gdzie $\theta = e^{-ab}$ i $\theta_0 = e^{-a_0 b_0}$, na podstawie próby losowej Y_1, Y_2, \dots, Y_n z rozkładu jednostajnego $U(0, \theta)$, gdzie $Y_i = e^{-aX_i}$. Korzystając z zadania 10 otrzymujemy test o obszarze krytycznym

$$W = \{(y_1, y_2, \dots, y_n) : y_{n:n} < c \vee y_{n:n} \geq \theta_0\},$$

gdzie $c = \theta_0 \alpha^{\frac{1}{n}}$. Zauważmy, że $Y_{n:n} = e^{-aX_{1:n}}$, zatem

$$W = \{(x_1, x_2, \dots, x_n) : x_{1:n} \leq b_0 \vee x_{1:n} > b_0 - \frac{\ln \alpha}{na}\}.$$

b) Dla weryfikacji hipotezy $H : a = a_0, b = b_0$ przy alternatywie $K_1 : a = a_0, b < b_0$, korzystając z zadania 10, obszarem krytycznym testu *JNM* na poziomie istotności α jest zbiór

$$Z = \{x = (x_1, x_2, \dots, x_n) : x \in A \vee x_{1:n} \leq b_0\},$$

gdzie $A \subseteq (b_0, +\infty)^n$ spełnia warunek $P_{a_0, b_0}\{X \in A\} = \alpha$ i P_{a_0, b_0} oznacza rozkład zmiennej losowej $X = (X_1, X_2, \dots, X_n)$ przy prawdziwości hipotezy H .

Rodzina rozkładów o gęstościach

$$f_{a,b_0}(x_1, x_2, \dots, x_n) = a^n \exp\left(-a \sum_{i=1}^n x_i + nab_0\right) \mathbf{1}_{(b_0, +\infty)}(x_{1:n}),$$

gdzie $a > 0$ jest parametrem, jest rodziną z monotonicznym ilorazem wiarygodności względem statystyki $T = \sum_{i=1}^n X_i$, zatem korzystając z twierdzenia 3 str. 66, testem *JNM* dla weryfikacji hipotezy $H : a = a_0, b = b_0$ przy alternatywie $K_1 : a > a_0$ jest test o obszarze krytycznym $W = \{(x_1, x_2, \dots, x_n) : \sum_{i=1}^n x_i > t\}$ i $P_{a_0, b_0}\{T > t\} = \alpha$. Jeżeli X ma rozkład o gęstości $f_{a,b}(x) = a_0 e^{-a(x-b)} \mathbf{1}_{(b, +\infty)}(x)$, to $2a(X-b) \sim \chi_1^2$. Zatem jeżeli H jest prawdziwa, to $2a_0(T - nb_0) \sim \chi_{2n}^2$, a stąd $2a_0(t - nb_0)$ jest kwantylem rzędu $1 - \alpha$ w rozkładzie χ_{2n}^2 .

Z postaci zbiorów krytycznych W i Z otrzymujemy test *JNM* dla weryfikacji hipotezy $H : a = a_0, b = b_0$ przy alternatywie $K_2 : a > a_0, b < b_0$. Jest to test o obszarze krytycznym

$$W_1 = \left\{ (x_1, x_2, \dots, x_n) : \sum_{i=1}^n x_i > t \vee x_{1:n} \leq b_0 \right\}.$$

13. Rozkład obserwowanej zmiennej losowej (Y_1, Y_2, \dots, Y_r) ma gęstość

$$p_\theta(y_1, \dots, y_r) = \begin{cases} \frac{\binom{n}{r} r!}{(2\theta)^r} \exp\left(-\frac{\sum_{i=1}^r y_i + (n-r)y_r}{2\theta}\right), & \text{gdy } 0 < y_1 < \dots < y_r, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Rodzina rozkładów o tej gęstości jest rodziną z monotonicznym ilorazem wiarygodności względem

$$T(y_1, y_2, \dots, y_r) = \sum_{i=1}^r y_i + (n-r)y_r.$$

Stąd (twierdzenie 3 str. 66) test *JNM* dla weryfikowania hipotezy $H : \theta \geq \theta_0$ przy hipotezie alternatywnej $K : \theta < \theta_0$ ma obszar krytyczny postaci

$$W = \{(y_1, y_2, \dots, y_r) : T(y_1, y_2, \dots, y_r) < c\},$$

gdzie c spełnia warunek $P_{\theta_0}(W) = \alpha$. Niech $Z_i = (n-i+1)(X_{i:n} - X_{i-1:n})$, $i = 1, 2, \dots, n$. Wtedy $T = \sum_{i=1}^r Z_i$. Zmienne Z_i , $i = 1, 2, \dots, n$, są niezależne o tym samym rozkładzie wykładniczym o gęstości

$$f_\theta(x) = \frac{1}{2\theta} e^{-x/2\theta} \mathbf{1}_{(0,\infty)}(x),$$

zatem statystyka T ma rozkład gamma $\Gamma(r, 2\theta)$ (por. zadania 5 i 6 str. 14) i zmienna losowa T/θ ma rozkład $\Gamma(r, 2) = \chi_{2r}^2$. Stąd

$$P_{\theta_0}(W) = P_{\theta_0}\left(\frac{T}{\theta_0} < c_1\right) = F(c_1) = \alpha,$$

gdzie F jest dystrybuantą rozkładu χ_{2r}^2 . Przy $r = 4$, $\theta_0 = 1000$ i $\alpha = 0.05$ otrzymujemy $c_1 = 2.73$. Moc testu przy alternatywie $\theta = 500$ jest równa

$$P_{500}\left(\frac{T}{\theta_0} < 2.73\right) = P_{500}\left(\frac{T}{500} < 2 \cdot 2.73\right) = 0.29.$$

Aby moc testu przy tej alternatywie była ≥ 0.95 potrzeba $r \geq 23$.

Wykład V. Wiarygodność

2. Nierówność $P_{M+1}\{X = x\}/P_M\{X = x\} \geq 1$ jest spełniona wtedy i tylko wtedy, gdy $M \leq x(N+1)/n - 1$, zatem $ENW(M) = \lfloor x(N+1)/n \rfloor$.

3. Funkcja wiarygodności jest równa

$$L(\theta; x) = \binom{x+k-1}{x} \theta^k (1-\theta)^x.$$

Wyznaczając maksimum względem θ , otrzymujemy $ENW(\theta) = k/(k+x)$.

4. $ENW(\exp(\mu + \sigma^2/2)) = \exp(\bar{X} + S^2/2)$, gdzie $\bar{X} = \sum_{i=1}^n X_i/n$, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ oraz $X_i = \ln Y_i$, $i = 1, 2, \dots, n$. Zmienne \bar{X} i S^2 są niezależne, stąd

$$E\left(\exp\left(\bar{X} + \frac{1}{2}S^2\right)\right) = E(\exp(\bar{X}))E\left(\exp\left(\frac{1}{2}S^2\right)\right) = \frac{\exp\left(\mu + \frac{1}{2}\sigma^2\right)}{\left(1 - \frac{\sigma^2}{n}\right)^{2n-2}}.$$

7. Numeryczne rozwiązanie względem $(\lambda'_0, \lambda''_0)$ układu równań

$$\int_{\lambda'_0}^{\lambda''_0} f_{n-1}(x) dx = 1 - \alpha.$$

$$\lambda'^{-n/2} \exp\left\{\frac{n}{2}(\lambda'_0 - 1)\right\} = \lambda''^{-n/2} \exp\left\{\frac{n}{2}(\lambda''_0 - 1)\right\}$$

dla $n = 10$, $\alpha = 0.01$ daje wynik $\lambda'_0 = 0.20394$, $\lambda''_0 = 2.8364$.

8. Niech

$$\Lambda(x) = \frac{\sup_{\theta \in (0,1)} \theta^x (1-\theta)^{n-x}}{\sup_{\theta \in (0,\theta_0)} \theta^x (1-\theta)^{n-x}}.$$

Jeżeli $x/n \leq \theta_0$, to $\Lambda(x) = 1$.

Jeżeli $x/n > \theta_0$, to

$$\Lambda(x) = \frac{\left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x}}{\theta_0^x (1 - \theta_0)^{n-x}}.$$

Funkcja Λ jest rosnącą funkcją zmiennej x , więc obszar krytyczny jest postaci

$$W = \{x : x > k\}$$

gdzie k dobrane jest tak, aby $\sup_{\theta \in (0,\theta_0]} P_\theta\{x > k\} \leq \alpha$. Ze względu na dyskretność rozkładu zmiennej X może się zdarzyć, że w powyższej nierówności nie zachodzi równość. Wtedy otrzymany test nie jest testem JNM .

Wykład VI. Metoda najmniejszych kwadratów. Modele liniowe

1. *Ad.1.* Obserwowana zmienna losowa: $Y = (Y_1, Y_2, \dots, Y_n)$ o wartościach w \mathbf{R}^n , rodzina rozkładów: $\mathcal{P} = \{P_F : E_{P_F} Y_i = \frac{1}{2} F t_i^2, F \in \mathbf{R}\}$ (F jest nieznanym parametrem).

Ad.2. Obserwowana zmienna losowa:

$$Y = (Y_{1,1}, Y_{1,2}, \dots, Y_{1,m}, Y_{2,1}, Y_{2,2}, \dots, Y_{2,m}, \dots, Y_{n,m})$$

o wartościach w \mathbf{R}^{nm} , rodzina rozkładów:

$$\mathcal{P} = \{P_{\mu,\alpha,\beta} : E_{P_{\mu,\alpha,\beta}} Y_{i,j} = \mu + \alpha x_i + \beta y_j, \mu, \alpha, \beta \in \mathbf{R}\}.$$

Ad.3. Obserwowana zmienna losowa:

$$X = (X_{1,1}, X_{1,2}, \dots, X_{1,m}, X_{2,1}, X_{2,2}, \dots, X_{2,m}, \dots, X_{n,m})$$

o wartościach w \mathbf{R}^{nm} , rodzina rozkładów:

$$\mathcal{P} = \{P_{A,\alpha,\beta} : E_{P_{A,\alpha,\beta}} X_{i,j} = An_i^\alpha k_j^\beta, A, \alpha, \beta \in \mathbf{R}\}.$$

Ad.4. Obserwowana zmienna losowa: $Y = (Y_1, Y_2, \dots, Y_m)$, gdzie Y_i oznacza liczbę owadów, które nie przeżywają dawki x_i preparatu i $Y_i \in \{0, 1, 2, \dots, n_i\}$, rodzina rozkładów:

$$\mathcal{P} = \left\{ P_{\beta_1, \beta_2} : P_{\beta_1, \beta_2}(y_1, y_2, \dots, y_m) = \prod_{i=1}^m \binom{n_i}{y_i} \phi(x_i)^{y_i} (1 - \phi(x_i))^{n_i - y_i}, \beta_1, \beta_2 > 0 \right\},$$

gdzie $\phi(x) = 1/(1 - \exp(-\beta_1 - \beta_2 x))$.

2. Niech \mathbf{X} będzie macierzą wymiaru $n \times k$. Oczywiście $\text{Im}\mathbf{X}^T\mathbf{X} \subseteq \text{Im}\mathbf{X}^T$. Należy pokazać, że $\text{Im}\mathbf{X}^T \subseteq \text{Im}\mathbf{X}^T\mathbf{X}$. Niech $y \in \mathbf{R}^n$. Wtedy (korzystając z wskazówki) istnieją wektory y_1 i y_2 takie, że $y = y_1 + y_2$ i $y_1 \perp y_2$, $y_1 \in \text{Im}\mathbf{X}$, $y_2 \in (\text{Ker}\mathbf{X}^T)^\perp$. Zatem istnieje $z \in \mathbf{R}^k$ taki, że $y_1 = \mathbf{X}z$, a stąd

$$\mathbf{X}^T\mathbf{X}z = \mathbf{X}^T y_1 = \mathbf{X}^T(y_1 + y_2) = \mathbf{X}^T y.$$

3.

$$E_\beta \hat{\beta} = E_\beta(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E_\beta Y = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta = \beta.$$

$$\begin{aligned} \text{Var}_\beta \hat{\beta} &= \text{Var}_\beta(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \text{Var}_\beta Y [(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]^T \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}. \end{aligned}$$

4. $EMNK[\mu] = \bar{Y}$

5. $EMNK[\mu_1 - \mu_2] = \sum_{i=1}^{n_1} Y_{i,1}/n_1 - \sum_{i=1}^{n_2} Y_{i,2}/n_2$, gdzie $Y_{i,1}$ są wynikami pomiarów przy pierwszej, a $Y_{i,2}$ - przy drugiej technologii.

6. Macierz

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ a_1 & a_2 & \dots & a_n \end{bmatrix}^T$$

ma rząd 2 wtedy i tylko wtedy, gdy nie wszystkie a_i są równe. Wtedy estymatory $\hat{\beta}_0$ i $\hat{\beta}_1$ otrzymujemy wyznaczając $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$. Estymatory są liniowymi funkcjami zmiennych Y_1, Y_2, \dots, Y_n o rozkładzie normalnym, stąd są zmiennymi losowymi o rozkładzie normalnym. Macierz kowariancji wektora $[\hat{\beta}_0, \hat{\beta}_1]^T$ jest równa

$$\sigma^2(\mathbf{X}^T\mathbf{X})^{-1} = \frac{\sigma^2}{\sum_{i=1}^n (a_i - \bar{a})^2} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n a_i^2 & \bar{a} \\ \bar{a} & 1 \end{bmatrix}.$$

8. Niech

$$Y = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 - 2\pi \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}.$$

Wtedy

$$EMNK[\theta] = \hat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y,$$

zatem $\hat{\theta}_i = x_i - \bar{x} + \pi/2$, $i = 1, 2, 3, 4$. Korzystając z postaci estymatora nieobciążonego wariancji σ^2 (patrz str. 90) otrzymujemy

$$\hat{\sigma}^2 = 4\left(\bar{x} - \frac{\pi}{2}\right)^2.$$

Przy dodatkowych warunkach $\theta_1 = \theta_3$ i $\theta_2 = \theta_4$ wielkości Y , \mathbf{X} i θ mają postać

$$Y = \begin{bmatrix} x_1 \\ x_2 - \pi \\ x_3 \\ x_4 - \pi \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad \theta = \theta_1.$$

Zatem

$$\hat{\theta}_1 = \hat{\theta}_3 = \frac{x_1 + x_3}{2} - \bar{x} + \frac{\pi}{2}, \quad \hat{\theta}_2 = \hat{\theta}_4 = \frac{x_2 + x_4}{2} - \bar{x} + \frac{\pi}{2}$$

oraz

$$\hat{\sigma}^2 = \frac{(x_1 - x_3 + 2\eta)^2 + (x_2 - x_4 + 2\eta)^2 + (x_3 - x_1 + 2\eta)^2 + (x_4 - x_2 + 2\eta)^2}{12},$$

gdzie $\eta = \bar{x} - \pi/2$. Powyższe estymatory kątów można również otrzymać minimalizując funkcję

$$f(\theta_1, \theta_2, \theta_3, \theta_4) = \sum_{i=1}^4 (x_i - \theta_i)^2$$

przy odpowiednich warunkach na θ_i , $i = 1, 2, 3, 4$.

Wykład VII. Teoria decyzji statystycznych

2. Niech $m = \text{med}(P)$ i $t > m$. Pokażemy, że $E_P|X - t| \geq E_P|X - m|$ (analogicznie pokazuje się, że powyższa nierówność jest prawdziwa dla $t < m$). Mamy

$$\begin{aligned} E_P|X - t| &= \int_{\{x \leq t\}} (t - x)P(dx) + \int_{\{x > t\}} (x - t)P(dx) \\ &= \int_{\{x \leq t\}} (m - x)P(dx) + \int_{\{x > t\}} (x - m)P(dx) + (t - m)[P\{X \leq t\} - P\{X > t\}] \\ &= E_P|X - m| + (t - m)[P\{X \leq m\} - P\{X > m\}] + 2P\{m < X \leq t\} \\ &\quad + 2 \int_{\{m < x \leq t\}} (m - x)P(dx) \geq E_P|X - m|. \end{aligned}$$

3. Niech $n = 2k + 1$, wtedy $\text{med}(X) = X_{k+1:n}$. Korzystając ze wskazówki otrzymujemy, że dla każdego $j = 1, 2, \dots, k$

$$E(X_{k+1:n} - X_{j:n} | \bar{X}) = E(X_{n-j+1:n} - X_{k+1:n} | \bar{X}),$$

czyli

$$2E(X_{k+1:n}|\bar{X}) = E(X_{n-j+1:n} + X_{j:n}|\bar{X}).$$

Sumując stronami dla $j = 1, 2, \dots, k$ i dodając do obu stron $E(X_{k+1:n}|\bar{X})$ otrzymujemy

$$E(X_{k+1:n}|\bar{X}) = E(\bar{X}|\bar{X}),$$

czyli $E(\text{med}(X)|\bar{X}) = \bar{X}$.

Gdy $n = 2k$ definiujemy $\text{med}(X) = \frac{1}{2}(X_{k:n} + X_{k+1:n})$ i postępujemy jak wyżej.

4. Rozkładem *a posteriori* jest rozkład normalny

$$N\left(\frac{\frac{\nu}{\tau^2} + \frac{x}{\sigma^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{1}{\sigma^2}}\right).$$

Estymator bayesowski parametru μ jest równy wartości oczekiwanej, a jego ryzyko wariancji rozkładu *a posteriori*.

5. Niech $\mathbf{X} = (X_1, X_2, \dots, X_k)$ będzie średnią z n -elementowej próby losowej z k -wymiarowego rozkładu normalnego $N(\boldsymbol{\mu}, \mathbf{I})$. Wtedy $\mathbf{X} \sim N(\boldsymbol{\mu}, \frac{1}{n}\mathbf{I})$. Niech $S^2 = n \sum_{i=1}^k X_i^2$ oraz $\hat{\boldsymbol{\mu}} = (1 - \frac{k-2}{S^2})\mathbf{X}$. Różnica

$$\begin{aligned} r = R(\boldsymbol{\mu}, \mathbf{X}) - R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) &= \int \dots \int \left(\|\mathbf{x} - \boldsymbol{\mu}\|^2 - \left\| \mathbf{x} \left(1 - \frac{k-2}{S^2}\right) - \boldsymbol{\mu} \right\|^2 \right) P(d\mathbf{x}) \\ &= \int \dots \int \left(2\frac{k-2}{S^2} \sum_{i=1}^k x_i(x_i - \mu_i) - \frac{(k-2)^2}{nS^2} \right) P(d\mathbf{x}). \end{aligned}$$

Wykorzystując całkowanie przez części otrzymujemy

$$\begin{aligned} &\int x_i(x_i - \mu_i) \exp\left(-\frac{(x_i - \mu_i)^2 n}{2}\right) dx_i \\ &= \int \frac{1}{n} \left(\frac{1}{S^2} - \frac{2nx_i^2}{S^4} \right) \exp\left(-\frac{(x_i - \mu_i)^2 n}{2}\right) dx_i. \end{aligned}$$

Zatem

$$\begin{aligned} r &= \int \dots \int \left[\frac{2(k-2)}{n} \sum_{i=1}^k \left(\frac{1}{S^2} - \frac{2nx_i^2}{S^4} \right) - \frac{(k-2)^2}{nS^2} \right] P(d\mathbf{x}) \\ &= \int \dots \int \frac{(k-2)^2}{nS^2} P(d\mathbf{x}) = E_P \frac{(k-2)^2}{nS^2} > 0. \end{aligned}$$

6. Niech Θ^* będzie zbiorem rozkładów *a priori* parametru θ . Z założenia otrzymujemy

$$\inf_{d \in \mathcal{D}} \sup_{\pi \in \Theta^*} r(\pi, d) \leq \sup_{\theta \in \Theta} R(\theta, \delta) \leq r(\tau, \delta) = \inf_{d \in \mathcal{D}} r(\tau, d) \leq \sup_{\pi \in \Theta^*} \inf_{d \in \mathcal{D}} r(\pi, d),$$

dotatkowo zachodzi

$$\sup_{\pi \in \Theta^*} \inf_{d \in \mathcal{D}} r(\pi, d) \leq \inf_{d \in \mathcal{D}} \sup_{\pi \in \Theta^*} r(\pi, d)$$

oraz

$$\sup_{\pi \in \Theta^*} r(\pi, d) = \sup_{\theta \in \Theta} R(\theta, d).$$

Zatem

$$\inf_{d \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d) = \sup_{\theta \in \Theta} R(\theta, \delta),$$

a to oznacza, że δ jest regułą minimaxową.

7. Dla każdego $\varepsilon > 0$ dla dostatecznie dużych n zachodzi

$$\sup_{\theta \in \Theta} R(\theta, \delta) - \varepsilon < r(\tau_n, \delta_n) = \inf_{d \in \mathcal{D}} r(\tau_n, d) \leq \sup_{\pi \in \Theta^*} \inf_{d \in \mathcal{D}} r(\pi, d).$$

Zatem

$$\inf_{d \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d) \leq \sup_{\theta \in \Theta} R(\theta, \delta) \leq \sup_{\pi \in \Theta^*} \inf_{d \in \mathcal{D}} r(\pi, d) \leq \inf_{d \in \mathcal{D}} \sup_{\pi \in \Theta^*} r(\pi, d).$$

8. Ryzyko estymatora $\bar{\mathbf{X}}$ jest równe $R(\boldsymbol{\mu}, \bar{\mathbf{X}}) = k/n$, gdzie n jest liczebnością próby losowej, i nie zależy od parametru $\boldsymbol{\mu}$. Rozważmy jako rozkład *a priori* rozkład $N(0, v^2 \mathbf{I})$. Wtedy rozkład *a posteriori* jest rozkładem normalnym

$$N\left(\frac{n\bar{\mathbf{X}}}{n+v^{-2}}, \frac{1}{n+v^{-2}}\mathbf{I}\right)$$

i estymator bayesowski jest równy $\delta_v = \frac{\bar{\mathbf{X}}}{n+v^{-2}}$, a jego ryzyko bayesowskie, równe $\frac{k}{n+v^{-2}}$, dąży do liczby k/n , gdy v dąży do $+\infty$. Zatem na mocy twierdzenia 4 str. 99 estymator $\bar{\mathbf{X}}$ jest minimaxowy.

9. Niech Π^b oznacza rozkład *a priori* postaci

$$\Pi^b\{\mu\} = b, \quad \Pi^b\{\nu\} = 1 - b.$$

Rozkład *a posteriori* jest równy

$$\Pi_x^b\{\mu\} = \frac{bp_\mu(x)}{bp_\mu(x) + (1-b)p_\nu(x)}, \quad \Pi_x^b\{\nu\} = \frac{(1-b)p_\nu(x)}{bp_\mu(x) + (1-b)p_\nu(x)},$$

gdzie $p_\mu(x)$ i $p_\nu(x)$ oznaczają gęstości rozkładów normalnych $N(\mu, \sigma^2)$ i $N(\nu, \sigma^2)$. Reguła bayesowska ma postać

$$\delta(x) = \begin{cases} \mu, & \text{gdy } x \leq \frac{\sigma^2}{\nu - \mu} \ln \frac{b}{1-b} + \frac{\mu + \nu}{2}, \\ \nu, & \text{w przeciwnym przypadku.} \end{cases}$$

Reguła $\delta = \delta_m$, jeśli

$$b = \frac{\exp \left[\left(m - \frac{\mu + \nu}{2} \right) \frac{\nu - \mu}{\sigma^2} \right]}{1 + \exp \left[\left(m - \frac{\mu + \nu}{2} \right) \frac{\nu - \mu}{\sigma^2} \right]}.$$

Reguła δ_m jest regułą minimaxową, jeśli

$$P_\mu \{x > m\} = P_\nu \{x \leq m\} \iff \Phi \left(\frac{m - \nu}{\sigma} \right) + \Phi \left(\frac{m - \mu}{\sigma} \right) = 1,$$

stąd $m = (\mu + \nu)/2$.

10. a) Niech Π^b oznacza rozkład *a priori* postaci

$$\Pi^b \{N(\mu, \sigma^2)\} = b, \quad \Pi^b \{N(\nu, \tau^2)\} = 1 - b.$$

Reguła bayesowska wybiera rozkład $N(\mu, \sigma^2)$, jeśli

$$\tau^2(x - \mu)^2 - \sigma^2(x - \nu)^2 < 2\sigma^2\tau^2 \ln \left(\frac{b\tau}{(1-b)\sigma} \right);$$

w przeciwnym przypadku wybiera rozkład $N(\nu, \tau^2)$.

b) Przeprowadzając rozumowanie analogiczne jak w zadaniu 9 otrzymujemy następującą postać reguły bayesowskiej:

$$\delta(x_1, x_2, \dots, x_k) = \begin{cases} \mu, & \text{gdy } \sum_{i=1}^k (\nu_i - \mu_i)(2x_i - \mu_i - \nu_i) < 2 \ln \frac{b}{1-b}, \\ \nu, & \text{w przeciwnym przypadku.} \end{cases}$$

TABLICA ROZKŁADÓW

Podstawowe rozkłady prawdopodobieństwa

Gęstość *	Nazwa rozkładu	Oznaczenie
$(\sigma\sqrt{2\pi})^{-1} \exp[-(x - \mu)^2 / 2\sigma^2]$	<i>normalny</i>	$N(\mu, \sigma^2)$
$(2\lambda)^{-1} \exp[- x - \theta /\lambda]$	<i>Laplace'a</i>	$DE(\theta, \lambda)$
$\frac{\lambda}{\pi} \frac{1}{\lambda^2 + (x - \theta)^2}$	<i>Cauchy'ego</i>	$C(\theta, \lambda)$
$\frac{1}{\lambda} \frac{\exp[-(x - \theta)/\lambda]}{(1 + \exp[-(x - \theta)/\lambda])^2}$	<i>logistyczny</i>	$L(\theta, \lambda)$
$(\lambda)^{-1} \exp[-(x - \theta)/\lambda]$	<i>wykładniczy</i>	$E(\theta, \lambda)$
$(\lambda)^{-1}$	<i>jednostajny</i>	$U(\theta - \frac{\lambda}{2}, \theta + \frac{\lambda}{2})$
$(\Gamma(\alpha)\lambda^\alpha)^{-1} x^{\alpha-1} e^{-x/\lambda}$	<i>gamma</i>	$\Gamma(\alpha, \lambda)$
$(\Gamma(\frac{n}{2})2^{n/2})^{-1} x^{n/2-1} e^{-x/2}$	<i>chi-kwadrat</i>	χ_n^2
$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1}$	<i>beta</i>	$B(\alpha, \beta)$
$\binom{n}{x} p^x (1 - p)^{n-x}$	<i>dwumianowy</i>	$b(p, n)$
$(x!)^{-1} \lambda^x e^{-\lambda}$	<i>Poissona</i>	$P(\lambda)$
$\binom{m+x-1}{m-1} p^m (1-p)^x$	<i>ujemny dwumianowy</i>	$Nb(p, m)$
$\frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$	<i>hipergeometryczny</i>	$H(N, M, n)$

* Gęstość względem miary Lebesgue'a dla pierwszych dziewięciu rozkładów i względem miary liczącej dla czterech pozostałych.

LITERATURA

J.R.Barra, *Matematyczne podstawy statystyki*, PWN, Warszawa 1982.

J.Bartoszewicz, *Wykłady ze statystyki matematycznej*, PWN, Warszawa 1989, 1996

E.L.Lehmann, *Theory of point estimation*, Wiley, New York 1983.

E.L.Lehmann, *Testing statistical hypothesis*, Second edition, Wiley, New York 1986 (polski przekład pierwszego wydania: *Testowanie hipotez statystycznych*, PWN, Warszawa 1968).

S.D.Silvey, *Wnioskowanie statystyczne*, PWN, Warszawa 1978.

SKOROWIDZ

- akcja 93
- analiza wariancji 56,59

- błąd pierwszego rodzaju 60
 - drugiego rodzaju 61
 - średniokwadratowy 33

- decyzja 93
- dyskryminacja 100
- dystrybuanta empiryczna 11

- EMNK* 84,85
- ENMW* 34
- ENW* 72
- estymacja 10
 - przedziałowa 10
 - punktowa 10
 - z zadaną precyzją 11
- estymator liniowy 34
 - największej wiarygodności (*ENW*) 72
 - nieobciążony 34
 - o minimalnej wariancji (*ENMW*) 34
 - optymalny 10
 - uzyskany metodą najmniejszych kwadratów (*EMNK*) 84,85

- funkcja mocy 61
 - ryzyka 33,93
 - straty 33,93

- hipoteza alternatywna 61
 - konkurencyjna 61
 - prosta 46
 - statystyczna 11,46
 - złożona 46

- JNM* 61

- lemat Neymana–Pearsona 61

- metoda najmniejszych kwadratów (*MNK*) 83,84
- minimalna statystyka dostateczna 23
- MNK* 83,84
- moc testu 61
- model liniowy 85
 - nieparametryczny 52
 - statystyczny 9
- monotoniczny iloraz wiarygodności 66
- MWW* 52,53

- obserwacja 9
- obszar krytyczny 60

- planowanie eksperymentów statystycznych 83
- podstawowy lemat Neymana–Pearsona 61
- populacja 8
- poziom istotności 46
 - ufności 10
- próba 8
 - losowa 8
 - z populacji 9
 - z rozkładu 9
- przedział ufności 10
- przestrzeń próby 9
 - statystyczna 9

- randomizacja 55
- randomizowana reguła decyzyjna 93
- randomizowany test 55,60
- rangi 53

- regresja 83,86
- reguła bayesowska 96
 - decyzyjna 93
- reguła minimaksowa 99
- resztowa suma kwadratów 89
- rodzina wykładnicza 28
- rozkład chi–kwadrat 14
 - F 15
 - F Snedecora 15
 - logarytmnormalny 82
 - geometryczny 75
 - t Studenta 15
- rozmiar testu 46
- równoważne rodziny rozkładów 24
 - zmienne losowe 17
- ryzyko 33,93
 - bayesowskie 96

- statystyka 9
 - dostateczna 19
 - dostateczna minimalna 23
 - Kołmogorowa 47
 - swobodna 26
 - zupełna 27

- test istotności 46
 - jednostajnie najmocniejszy (*JNM*) 61
 - Kołmogorowa 46
 - kombinatoryczny 53
 - Manna–Whitneya 53
- test Manna–Whitneya–Wilcoxona (*MWW*) 52,53
 - najmocniejszy 61
 - nieparametryczny 52
 - nierandomizowany 60
 - oparty na ilorazie wiarygodności 77
 - permutacyjny 53
 - randomizowany 55,60
 - Wilcoxona 53
 - zgodności Kołmogorowa 46
- twierdzenie Basu 27
 - Cochran–Fishera 56
 - Gaussa–Markowa 88
 - Gliwienki–Cantelliego 13
 - Rao–Blackwella 34,95

- wartość krytyczna 46
- weryfikacja hipotez statystycznych 11
- wiarygodność 71
- wynik eksperymentu 9
 - obserwacji 9
 - pomiaru 9

- zasada największej wiarygodności 72,77
- zbiór krytyczny 60
 - zerowy 24
- zmienne losowe równoważne 17
- zupełna rodzina rozkładów 27

SPIS TREŚCI

Przedmowa	5
Wykład I. Model statystyczny	7
1. Przykłady wprowadzające	7
2. Model statystyczny	9
3. Podstawowe problemy statystyki matematycznej	10
4. Podstawowe twierdzenie statystyki matematycznej	11
5. Zadania	13
Wykład II. Statystyki dostateczne	16
1. Preliminaria	16
2. Przykład wprowadzający	18
3. Definicja statystyki dostatecznej. Przykłady	19
4. Kryterium faktoryzacji	21
5. Minimalne statystyki dostateczne	23
6. Statystyki swobodne. Statystyki zupełne. Twierdzenie Basu	26
7. Rodziny wykładnicze rozkładów	28
8. Zadania	31
Wykład III. Estymatory nieobciążone o minimalnej wariancji	33
1. Sformułowanie problemu	33
2. Konstrukcja	34
3. <i>ENMW</i> w jednopróbkowym modelu gaussowskim	37
3.1. Statystyki	37
3.2. Estymacja μ , gdy σ jest znane	38
3.3. Estymacja σ^2 , gdy μ jest znane	38
3.4. Przypadek, gdy μ oraz σ nie są znane	38
3.5. Estymacja kwantyla rozkładu $N(\mu, \sigma^2)$	39
3.6. Estymacja prawdopodobieństwa $P_{\mu, \sigma}\{X \leq u\}$	39
4. Kłopoty z <i>ENMW</i>	42
5. Zadania	43

Wykład IV. Testowanie hipotez statystycznych	45
1. Wprowadzenie	45
2. Test zgodności Kołmogorowa	46
2.1. Oznaczenia	46
2.2. Hipoteza prosta	47
2.3. Hipoteza złożona	48
2.3.1. Uwagi ogólne	48
2.3.2. Hipoteza $H^-: G \leq F$	49
2.3.3. Hipoteza o normalności rozkładu	50
3. Porównywanie średnich dwóch rozkładów normalnych	51
3.1. Sformułowanie zagadnienia	51
3.2. Przypadek rozkładów normalnych o jednakowej wariancji	51
3.3. Przypadek dowolnych rozkładów normalnych	52
4. Hipoteza o parametrze położenia	52
5. Porównanie k średnich (analiza wariancji)	56
6. Porównywanie testów. Teoria Neymana–Pearsona	59
6.1. Wprowadzenie	59
6.2. Podstawowy lemat Neymana–Pearsona	61
6.3. Testy JNM w modelach z monotonicznym ilorazem wiarygodności	66
6.4. Przykład, gdy test JNM nie istnieje	69
7. Zadania	69
Wykład V. Wiarygodność	71
1. Koncepcja	71
2. Estymatory największej wiarygodności	72
2.1. Konstrukcja	72
2.2. Błąd średniokwadratowy ENW	72
2.3. ENW w złożonych doświadczeniach	74
2.4. Kłopoty z ENW	75
3. Testy oparte na ilorazie wiarygodności	77
3.1. Konstrukcja	77
3.2. Przykłady	78
4. Zadania	81
Wykład VI. Metoda najmniejszych kwadratów. Modele liniowe	83
1. Przykłady wprowadzające	83
2. Idea MNK	84
3. $EMNK$ w modelach liniowych	85
3.1. Ogólna postać modelu liniowego	85
3.2. $EMNK$ w modelu liniowym. Twierdzenie Gaussa–Markowa	86
3.3. $EMNK$ w gaussowskim modelu liniowym	90
4. Zadania	91

Wykład VII. Teoria decyzji statystycznych	93
1. Sformułowanie problemu	93
2. Optymalne reguły decyzyjne	94
2.1. Wprowadzenie	94
2.2. Redukcja przez dostateczność	95
2.3. Bayesowskie reguły decyzyjne	96
2.4. Minimaksowe reguły decyzyjne	99
3. Zastosowania w dyskryminacji	100
4. Zadania	103
Wskazówki	105
Rozwiązania	109
Tablica rozkładów prawdopodobieństwa	126
Literatura	127
Skorowidz	128