Część V

Dodatki

Dodatek A

Ważne rozkłady prawdopodobieństwa

Rozkład DWUMIANOWY $X \sim \mathrm{Bin}(n,p)$

Funkcja prawdopodobieństwa:

$$f(k) = \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad (k=0,1,\ldots,n).$$

Momenty: $\mathbb{E}X = np$, $\mathrm{Var}X = np(1-p)$.

Rozkład POISSONA $X \sim \mathrm{Poiss}(\lambda)$

Funkcja prawdopodobieństwa:

$$f(k) = \mathbb{P}(X=k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \qquad (k=0,1,\ldots).$$

Momenty: $\mathbb{E}X = \lambda$, $\mathrm{Var}X = \lambda$.

Rozkład UJEMNY DWUMIANOWY $X \sim \mathrm{Bin}^-(\alpha,p)$

Funkcja prawdopodobieństwa:

$$f(k) = \mathbb{P}(X=k) = \binom{-\alpha}{k} p^\alpha (p-1)^k = \binom{\alpha+k-1}{k} p^\alpha (1-p)^k, \qquad (k=0,1,\ldots).$$

Momenty: $\mathbb{E}X = \alpha(1-p)/p$, $\mathrm{Var}X = \alpha(1-p)/p^2$.

Rozkład GEOMETRYCZNY $X \sim \mathrm{Geo}(p) = \mathrm{Bin}^-(1,p)$

Funkcja prawdopodobieństwa: $f(k) = \mathbb{P}(X=k) = p(1-p)^k$, $(k=0,1,\ldots)$.

Momenty: $\mathbb{E}X = (1-p)/p$, $\mathrm{Var}X = (1-p)/p^2$.

Rozkład HIPERGEOMETRYCZNY $X \sim \mathrm{H}(n, r, m)$

Funkcja prawdopodobieństwa:

$$f(k) = \mathbb{P}(X = k) = \binom{n}{k}\binom{r-n}{m-k}\Big/\binom{r}{m} = \binom{m}{k}\binom{r-m}{n-k}\Big/\binom{r}{n},$$

$$\Big(\ \max(0, n+m-r) \leqslant k \leqslant \min(n, m)\ \Big).$$

Momenty: $\mathbb{E}X = nm/r$, $\mathrm{Var}X = nm(r-m)(r-n)r^{-2}(r-1)^{-1}$.

Rozkład JEDNOSTAJNY $X \sim \mathrm{U}(a, b)$

Gęstość prawdopodobieństwa i dystrybuanta:

$$f(x) = \frac{1}{b-a}, \qquad F(x) = \frac{x-a}{b-a}, \qquad (a < x < b).$$

Momenty: $\mathbb{E}X = (a+b)/2$, $\mathrm{Var}X = (b-a)^2/12$.

Rozkład NORMALNY $X \sim \mathrm{N}(\mu, \sigma^2)$

Gęstość prawdopodobieństwa i dystrybuanta:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad F(x) = \frac{1}{\sigma}\Phi\left(\frac{x-\mu}{\sigma}\right), \qquad (-\infty < x < \infty).$$

Momenty: $\mathbb{E}X = \mu$, $\mathrm{Var}X = \sigma^2$.

Standardowy rozkład normalny $\mathrm{N}(0,1)$ ma dystrybuantę Φ i gęstość φ:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\mathrm{d}x, \qquad \varphi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-a^2/2}.$$

Rozkład GAMMA $X \sim \mathrm{Gamma}(\alpha, \lambda)$

Gęstość prawdopodobieństwa:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}, \qquad (x > 0).$$

Momenty: $\mathbb{E}X = \alpha/\lambda$, $\mathrm{Var}X = \alpha/\lambda^2$.

Rozkład WYKŁADNICZY $X \sim \mathrm{Ex}(\lambda) = \mathrm{Gamma}(1, \lambda)$

Gęstość prawdopodobieństwa: $f(x) = \lambda\mathrm{e}^{-\lambda x}$; dystrybuanta: $F(x) = 1 - \mathrm{e}^{-\lambda x}$, $(x > 0)$.

Momenty: $\mathbb{E}X = 1/\lambda$, $\mathrm{Var}X = 1/\lambda^2$.

Rozkład CHI-KWADRAT $Y \sim \chi^2(k) = \text{Gamma}(k/2, 1/2)$

Definicja: $Y = Z_1^2 + \cdots + Z_k^2$, gdzie $Z_1, \ldots, Z_k$ są niezależne i $Z_i \sim \text{N}(0,1)$.

Gęstość prawdopodobieństwa:

$$f(y) = \frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2-1} \text{e}^{-y/2}, \quad (y > 0).$$

Rozkład t-STUDENTA $T \sim \text{t}(k)$

Definicja: $T = Z/\sqrt{Y/k}$, gdzie Z i Y są niezależne, $Z \sim \text{N}(0,1)$, $Y \sim \chi^2(k)$.

Gęstość prawdopodobieństwa:

$$f(t) = \frac{\Gamma(k/2+1/2)}{\Gamma(k/2)} \frac{1}{\sqrt{\pi k}} \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}, \quad (-\infty < t < \infty).$$

Rozkład F-SNEDECORA $R \sim \text{F}(k, m)$

Definicja: $T = Z/\sqrt{(Y/k)/(V/m)}$, gdzie Z i V są niezależne, $Y \sim \chi^2(k)$ i $V \sim \chi^2(m)$.

Gęstość prawdopodobieństwa:

$$f(r) = \frac{k^{k/2} m^{m/2} \Gamma(k/2+1/2)}{\Gamma(k/2)\Gamma(m/2)} \frac{r^{k/2-1}}{(kr+m)^{(k+m)/2}}, \quad (r > 0).$$

Rozkład BETA $X \sim \text{Be}(\alpha, \beta)$

Gęstość prawdopodobieństwa:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad (0 < x < 0).$$

Momenty: $\mathbb{E}X = \alpha/(\alpha+\beta)$, $\text{Var}X = \alpha\beta(\alpha+\beta)^{-2}(\alpha+\beta+1)^{-1}$.

Rozkład CAUCHY'ego $X \sim \mathrm{Cauchy}(a, d)$

Gęstość prawdopodobieństwa i dystrybuanta:

$$f(x) = \frac{1}{\pi} \cdot \frac{d}{d^2 + (x-a)^2}, \qquad F(x) = \frac{1}{\pi}\mathrm{arctan}\frac{x-a}{d} + \frac{1}{2}, \qquad (-\infty < x < \infty).$$

Momenty: nie istnieją.

Wielowymiarowy rozkład NORMALNY $X\mathrm{NN}(\mu, \mathbf{V})$ $\mu = (\mu_1, \ldots, \mu_n)^\top$, $\mathbf{V} = (\varrho_{ij}\sigma_i\sigma_j)_{i,j=1,\ldots,n}$

Gęstość prawdopodobieństwa:

$$f(x) = (2\pi)^{-n/2}(\det \mathbf{V})^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)^\top \mathbf{V}^{-1}(x-\mu)\right].$$

Momenty: $\mathbb{E}X = \mu$, $\mathbb{E}X_i = \mu_i$, $\mathrm{VAR}X = \mathbf{V}$, $\mathrm{Var}X_i = \sigma_i^2$, $\mathrm{Cov}(X_i, X_j) = \varrho_{ij}\sigma_i\sigma_j$.

Rozkłady brzegowe: $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$.

Dwuwymiarowy rozkład NORMALNY $(X, Y) \sim \mathrm{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \varrho)$

Gęstość prawdopodobieństwa: $f(x, y) =$

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} \exp\left[-\frac{1}{2(1-\varrho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\varrho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right].$$

Momenty: $\mathbb{E}X = \mu_X$, $\mathbb{E}Y = \mu_Y$, $\mathrm{Var}X = \sigma_X^2$, $\mathrm{Var}Y = \sigma_Y^2$, $\mathrm{Cov}(X, Y) = \varrho\sigma_X\sigma_Y$.

Rozkłady brzegowe: $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$.

Rozkłady warunkowe: jeśli $X = x$ to $Y \sim \mathrm{N}\left(\mu_Y + (x-\mu_X)\varrho\sigma_Y/\sigma_X, \sigma_Y^2(1-\varrho^2)\right)$.

Rozkład WIELOMIANOWY $\mathrm{Mult}(n, p_1, \ldots, p_d)$

Funkcja prawdopodobieństwa:

$$f(n_1, \ldots, n_d) = \mathbb{P}(X_1 = n_1, \ldots, X_k = n_d) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_d^{n_d}, \quad (n_1 + \cdots + n_d = n).$$

Momenty: $\mathbb{E}X_i = np_i$, $\mathrm{Var}X_i = np_i(1-p_i)$, $\mathrm{Cov}(X_i, X_j) = -p_ip_j$ dla $i \neq j$.

Rozkłady brzegowe: $X_i \sim \mathrm{Bin}(n, p_i)$.

Rozkłady warunkowe: jeśli $X_1 = n_1$ to

$$(X_2, \cdots, X_d) \sim \mathrm{Mult}\left(n - n_1, p_2/(1-p_1), \ldots, p_d/(1-p_1)\right).$$

Dodatek B

Rozwiązania wybranych zadań

B.1 Zadania do Rozdziału 1

Rozwiązanie Zadania 1.1.

Obliczyć $\mathbb{E}\hat{F}(x)$, $\mathrm{Var}\hat{F}(x)$.

Używając podstawowych własności wartosci oczekiwanej i definicji $\hat{F}(x)$ wnioskujemy, że $\mathbb{E}\hat{F}(x) = \mathbb{E}n^{-1}\sum_{i=1}^{n}\mathbb{1}(X_i \leqslant x) = \mathbb{E}\mathbb{1}(X_1 \leqslant x) = \mathbb{P}(X_1 \leqslant x) = F(x)$. Podobnie, korzystając z niezależności zmiennych losowych $X_1, \ldots, X_n$, otrzymujemy $\mathrm{Var}\hat{F}(x) = \mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}\mathbb{1}(X_i \leqslant x)\right) = n^{-2}\sum_{i=1}^{n}\mathrm{Var}\,\mathbb{1}(X_i \leqslant x) = n^{-1}\mathrm{Var}\,\mathbb{1}(X_1 \leqslant x) = n^{-1}F(x)(1-F(x))$.

Rozwiązanie Zadania 1.2.

Pokazać, że ciąg zmiennych losowych $\sqrt{n}(\hat{F}_n(x) - F(x))$ jest zbieżny do rozkładu normalnego. Zidentyfikować parametry tego rozkładu.

Wystarczy skorzystać z wyników poprzedniego zadania i z Centralnego Twierdzenia Granicznego (CTG) w najprostszej wersji, dla ciągu zmiennych niezależnych o jednakowym rozkładzie. Otrzymujemy zbieżność według rozkładu $\sqrt{n}(\hat{F}_n(x) - F(x)) \to \mathrm{N}(0, F(x)(1-F(x))$.

Rozwiązanie Zadania 1.3.

Podać granicę $\lim_{n\to\infty}\mathbb{P}(\hat{F}_n(x) \leqslant F(x))$ przy założeniu, że $0 < F(x) < 1$. Dokładnie uzasadnić odpowiedź.

$\lim_{n\to\infty}\mathbb{P}(\hat{F}_n(x) \leqslant F(x)) = 1/2$. Istotnie, dystrybuanta rozkładu normalnego Φ jest funkcją ciągłą. Stąd wynika, że zbieżność według rozkładu $\sqrt{n}(\hat{F}_n(x) - F(x)) \to \mathrm{N}(0, F(x)(1-F(x))$ jest równoważna zbieżności punktowej dystrybuant. Dla każdego z mamy $\mathbb{P}\left(\sqrt{n}(\hat{F}_n(x) - F(x)) \leqslant z\right) \to \Phi(z\sqrt{F(x)(1-F(x))})$. Wystarczy wziąć $z = 0$ aby otrzymać odpowiedź.

Rozwiązanie Zadania 1.4.

Zadanie. Wyprowadzić alternatywny wzór na wariancję próbkową:

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

$\tilde{S}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2 = n^{-1}\left[\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n(\bar{X})^2\right] = n^{-1}\sum_{i=1}^n X_i^2 - 2(\bar{X})^2 + (\bar{X})^2$. Obliczenia są bardzo podobne jak dla wariancji zmiennej losowej.

B.2 Zadania do Rozdziału 2

Rozwiązanie Zadania 2.1.

Zadanie. Rozpatrzmy proces statystycznej kontroli jakości przyjmując te same założenia co w Przykładzie 2.1.4 z tą różnicą, że obserwujemy kolejne wyroby do momentu gdy natrafimy na k wybrakowanych, gdzie k jest ustaloną z góry liczbą. Zbudować model statystyczny.

Za obserwację uznamy wektor $(N, X_1, \ldots, X_N)$, gdzie zmienna losowa N oznacza liczbę sprawdzonych detali, X_i są zmiennymi losowymi o wartościach 0 lub 1. Mamy przy tym $N = \min\{n : \sum_{i=1}^n X_i = k\}$. Przestrzeń obserwacji jest nieco skomplikowana:

$$\mathcal{X} = \bigcup_{n=1}^{\infty} \{n\} \times \mathcal{X}_n, \text{ gdzie } \mathcal{X}_n = \{(x_1, \ldots, x_n) \in \{0,1\}^n : \sum_{i=1}^{n-1} x_i = k-1, x_n = 1\}.$$

Rozkład prawdopodobieństwa jest dany wzorem:

$$\mathbb{P}(N = n, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = 1) = p^k(1-p)^{n-k}.$$

Przestrzenią parametrów jest $\Theta = [0,1]$. Oczywiście, można zredukować model, przechodząc „w pamięci" do przestrzeni wartości statystyki dostatecznej. Na przykład statystyka $T = N - k$ przyjmuje wartości w zbiorze $\mathcal{T} = \{0, 1, \ldots\}$ i ma rozkład ujemny dwumianowy, $\mathrm{Bin}^-(k, p)$.

Rozwiązanie Zadania 2.2.

Zadanie. Uogólnić rozważania z Przykładu 2.1.5 (badanie reprezentacyjne), uwzględniając więcej niż jeden rodzaj jednostek „wyróżnionych". Powiedzmy, że mamy w urnie m_1 kul czerwonych, m_2 zielonych i $r - m_1 - m_2$ białych, gdzie r jest znaną liczbą, a m_1 i m_2 są nieznane i są przedmiotem badania. Opisać dokładnie odpowiedni model statystyczny.

Za obserwację uznamy wektor (X_1, X_2), zawierający liczbę kul czerwonych (X_1) i zielonych (X_2) w próbce. Przestrzenią obserwacji jest podzbiór $\mathcal{X} \subseteq \{0, \ldots, n\}^2$ (nie musimy się bardzo troszczyć o to, żeby wszystkie punkty X miały niezerowe prawdopodobieństwo). Przestrzenią parametrów

jest podzbiór $\Theta \subseteq \{0,\dots,r\}^2$. Rozkład prawdopodobieństwa jest dwuwymiarową wersją rozkładu hipergeometrycznego:

$$\mathbb{P}_{m_1,m_2}(X_1 = x_1, X_2 = X_2) = \binom{m_1}{x_1}\binom{m_2}{x_2}\binom{r-m_1-m_2}{n-x_1-x_2}\Big/\binom{r}{n}.$$

Rozwiązanie Zadania 2.3.

Zadanie. Obliczyć rozkład prawdopodobieństwa zmiennej losowej Z^2, jeśli $Z \sim \mathrm{N}(0,1)$ (obliczyć bezpośrednio dystrybuantę i gęstość rozkładu $\chi^2(1)$).

Najpierw obliczymy dystrybuantę: $F(y) = \mathbb{P}(Z^2 \leqslant y) = \mathbb{P}(|Z| \leqslant \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1$, dla $y > 0$. Gęstość obliczymy różniczkując dystrybuantę: $f(y) = F'(y) = \varphi(\sqrt{y})/\sqrt{y} = (2\pi)^{-1/2} y^{-1/2} \mathrm{e}^{-y/2}$.

Rozwiązanie Zadania 2.4.

Zadanie. Obliczyć rozkład prawdopodobieństwa zmiennej losowej $Z_1^2 + Z_2^2$, jeżeli $Z_i \sim \mathrm{N}(0,1)$ są niezależne dla $i = 1, 2$ (obliczyć bezpośrednio dystrybuantę i gęstość rozkładu $\chi^2(2)$).

Dystrybuanta jest równa

$$F(y) = \mathbb{P}(Z_1^2 + Z_2^2 \leqslant y) = \iint\limits_{z_1^2+z_2^2 \leqslant y} \varphi(z_1)\varphi(z_2)\mathrm{d}z_1\mathrm{d}z_2 = \iint\limits_{z_1^2+z_2^2 \leqslant y} \frac{1}{2\pi} \exp\left[-\frac{z_1^2+z_2^2}{2}\right] \mathrm{d}z_1\mathrm{d}z_2.$$

Tę dwuwymiarową całkę obliczamy przechodząc do współrzędnych biegunowych: $z_1 = r\cos\alpha$, $z_2 = r\cos\alpha$:

$$F(y) = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\sqrt{y}} r\exp\left[-\frac{r^2}{2}\right]\mathrm{d}r\mathrm{d}\alpha = \int_0^{\sqrt{y}} r\mathrm{e}^{-r^2/2}\mathrm{d}r = \left[-\mathrm{e}^{-r^2/2}\right]_{r=0}^{r=\sqrt{y}} = 1 - \mathrm{e}^{-y/2}.$$

Rozwiązanie Zadania 2.5.

Zadanie. Korzystając z Zadania 2.3 oraz z własności rozkładów gamma, udowodnić Uwagę 2.2: gęstość zmiennej losowej $Y \sim \chi^2(k)$ ma postać

$$f(y) = \frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2-1}\mathrm{e}^{-y/2}, \quad (y > 0).$$

Wynik Zadania 2.3 pokazuje, że rozkład $\chi^2(1)$ jest szczególnym przypadkiem rozkładu Gamma, mianowicie Gamma(1/2/1/2). Zatem rozkład $\chi^2(k)$ jest k-krotną potęgą splotową tego rozkładu, czyli jest równy Gamma(k/2/1/2).

Rozwiązanie Zadania 2.6.

Zadanie. Udowodnić zbieżność rozkładów: $\mathrm{t}(k) \to_{\mathrm{d}} \mathrm{N}(0,1)$ dla $k \to \infty$.

Z defincji, $\mathrm{t}(k)$ jest rozkładem zmiennej losowej

$$T_k = \frac{Z_0}{\sqrt{(Z_1^2 + \cdots + Z_k^2)/k}},$$

gdzie $Z_0, Z_1, \ldots, Z_k$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $\mathrm{N}(0,1)$. Zauważmy, że na mocy Prawa Wielkich Liczb, $(Z_1 + \cdots + Z_k)/k \to_{\text{p.n.}} \mathbb{E}Z_1^2 = 1$. Mianownik we wzorze definiującym T_k zmierza do 1 prawie na pewno, a więc tym bardziej według prawdopodobieństwa. Z lematu Słuckiego wynika, że $T_k \to_{\mathrm{d}} Z_0 \sim \mathrm{N}(0,1)$, co należało pokazać.

Rozwiązanie Zadania 2.8.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu (Weibulla) o gęstości

$$f_\theta(x) = \begin{cases} 3\theta x^2 \mathrm{e}^{-\theta x^3} & \text{dla } x > 0; \\ 0 & \text{dla } x \leqslant 0, \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Znaleźć jednowymiarową statystykę dostateczną.

Napiszmy łączną gęstość w postaci:

$$f_\theta(x_1, \ldots, x_n) = \left(3^n \prod_i x_i^2\right) \theta^n \exp\left[-\theta \sum_i x_i^3\right].$$

Skorzystajmy teraz z twierdzenia o faktoryzacji: za czynnik zawierający θ weźmy $\theta^n \exp\left[-\theta \sum_i x_i^3\right]$. Stąd wynika, że $\sum_i X_i^3$ jest statystyką dostateczną.

Rozwiązanie Zadania 2.9.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu $\mathrm{Gamma}(\alpha, \lambda)$. Znaleźć dwuwymiarową statystykę dostateczną, zakładając że $\theta = (\alpha, \lambda)$ jest nieznanym parametrem.

Napiszmy łączną gęstość w postaci

$$f_{\alpha,\lambda}(x_1, \ldots, x_n) = \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} \left(\prod_i x_i\right)^{\alpha-1} \exp\left[-\lambda \sum_i x_i\right].$$

Stąd wynika, że $(\sum_i X_i, \prod_i X_i)$ jest statystyką dostateczną. Równoważnie, możemy powiedzieć, że $(\sum_i X_i, \sum_i \log X_i)$ jest statystyką dostateczną.

Rozwiązanie Zadania 2.10.

Zadanie. Rozważamy rodzinę przesuniętych rozkładów wykładniczych o gęstości

$$f_\mu(x) = \begin{cases} \mathrm{e}^{-(x-\mu)} & \text{dla } x \geqslant \mu; \\ 0 & \text{dla } x < \mu. \end{cases}$$

Niech $X_1, \ldots, X_n$ będzie próbką losową z takiego rozkładu. Znaleźć jednowymiarową statystykę dostateczną dla parametru μ.

Napiszmy łączną gęstość w postaci

$$f_\mu(x_1,\dots,x_n) = \mathrm{e}^{-\sum x_i}\mathrm{e}^{n\mu}\prod_i \mathbb{1}(x_i \geqslant \mu).$$

Zauważmy, że $\prod_i \mathbb{1}(x_i \geqslant \mu) = \mathbb{1}(\min(x_1,\dots,x_n) \geqslant \mu)$. Stąd widać, że $\min(X_1,\dots,X_n)$ jest statystyką dostateczną.

Rozwiązanie Zadania 2.11.

Zadanie Rozważamy rodzinę przesuniętych rozkładów wykładniczych z parametrem skali o gęstości

$$f_{\mu,\lambda}(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda(x-\mu)} & \text{dla } x \geqslant \mu; \\ 0 & \text{dla } x < \mu. \end{cases}$$

Niech $X_1,\dots,X_n$ będzie próbką losową z takiego rozkładu. Znaleźć dwuwymiarową statystykę dostateczną dla parametru (μ,λ).

Mamy

$$f_{\mu,\lambda}(x_1,\dots,x_n) = \lambda^n \mathrm{e}^{n\lambda\mu}\exp\left[-\lambda\sum_i x_i\right]\prod_i \mathbb{1}(x_i \geqslant \mu).$$

Rozumujemy bardzo podobnie, jak w poprzednim zadaniu. Ponieważ mamy $\prod_i \mathbb{1}(x_i \geqslant \mu) = \mathbb{1}(\min(x_1,\dots,x_n) \geqslant \mu)$, więc statystyką dostateczną jest na przykład $(\min(X_1,\dots,X_n), \sum_i X_i)$.

Rozwiązanie Zadania 2.12.

Zadanie. Rozważamy rodzinę rozkładów na przestrzeni $\{0,1,2,\dots\}$:

$$f_\theta(x) = \mathbb{P}_\theta(X = x) = \begin{cases} \theta & \text{dla } x = 0; \\ (1-\theta)/2^x & \text{dla } x \in \{1,2,\dots\}. \end{cases}$$

gdzie $\theta \in]0,1[$ jest nieznanym parametrem. Niech $X_1,\dots,X_n$ będzie próbką losową z wyżej podanego rozkładu. Znaleźć jednowymiarową statystykę dostateczną.

Jeśli przyjmiemy oznaczenie $n_0 = \sum_{i=1}^n \mathbb{1}(x_i = 0)$ (liczba elementów próbki równych 0) to łatwo zauważyć, że

$$f_\theta(x_1,\dots,x_n) = \theta^{n_0}(1-\theta)^{n-n_0}\cdot 2^{-\sum x_i}.$$

Z twierdzenia o faktoryzacji natychmiast wynika, że $N_0 = \sum_{i=1}^n \mathbb{1}(X_i = 0)$ jest statystyką dostateczną.

Rozwiązanie Zadania 2.13.

Zadanie. Niech $X_1,\dots,X_n$ będzie schematem Bernoulliego z prawdopodobieństwem sukcesu θ. Obliczyć warunkowy rozkład prawdopodobieństwa zmiennych losowych $X_1,\dots,X_n$ przy danym $S = s$, gdzie $S = \sum_{i=1}^n X_i$ jest liczbą sukcesów. Zinterpretować fakt, że statystyka S jest dostateczna.

Możemy posłużyć się elementarną definicją prawdopodobieństwa warunkowego. Jeśli $\sum x_i = s$, to

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n | S = s) = \frac{\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)}{\mathbb{P}(S = s)} = \frac{\theta^s (1-\theta)^{n-s}}{\binom{n}{s} \theta^s (1-\theta)^{n-s}} = \frac{1}{\binom{n}{s}}.$$

(jeśli $\sum x_i \neq s$ to, oczywiście, prawdopodobieństwo warunkowe jest równe 0). Otrzymane wyrażenie nie zależy od parametru θ, a więc pokazaliśmy z definicji, że S jest statystyką dostateczną.

Rozwiązanie Zadania 2.14.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu Poiss(θ). Obliczyć warunkowy rozkład prawdopodobieństwa zmiennych losowych $X_1, \ldots, X_n$ przy danym $S = s$, gdzie $S = \sum_{i=1}^n X_i$. Zinterpretować fakt, że statystyka S jest dostateczna.

Podobnie, jak w zadaniu poprzednim, obliczamy elementarne prawdopodobieństwo warunkowe. Załóżmy, że $\sum x_i = s$ i skorzystajmy z tego, że $S \sim \text{Poiss}(n\theta)$. Prawdopodobieństwo warunkowe $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n | S = s)$ jest równe

$$\frac{\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)}{\mathbb{P}(S = s)} = \frac{\mathrm{e}^{-n\theta} \prod \theta^{x_i} / \prod x_i!}{\mathrm{e}^{-n\theta} (n\theta)^s / s!} = \frac{s!}{\prod x_i!} \left(\frac{1}{n}\right)^s.$$

Rozwiązanie Zadania 2.15.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu Ex(θ). Niech $S = \sum_{i=1}^n X_i$. Pokazać, że rozkład warunkowy $(X_1, \ldots, X_{n-1})$ przy danym $S = s$ jest jednostajny na sympleksie $\{(x_1, \ldots, x_{n-1}) : x_i \geqslant 0, \sum_{i=1}^{n-1} x_i \leqslant s\}$. Zinterpretować fakt, że statystyka S jest dostateczna.

Mamy tu do czynienia ze zmiennymi o rozkładach ciągłych. Nie możemy posłużyć się elementarną definicją prawdopodobieństwa warunkowego. Zamiast tego obliczymy gęstość warunkową zmiennych $X_1, \ldots, X_{n-1}$ przy danym $S = s$. Zauważmy, że ta $n-1$-wymiarowa gęstość w pełni wyznacza rozkład warunkowy n-wymiarowego wektora losowego $X_1, \ldots, X_{n-1}, X_n$, ponieważ $X_n = S - X_1, \cdots - X_{n-1}$. Najpierw obliczymy gęstość łączną zmiennych losowych $X_1, \ldots, X_{n-1}, S$. Stosujemy znany wzór na przekształcanie gęstości (inaczej, wzór na całkowanie przez podstawienie): niech J oznacza jakobian przekształcenia $(x_1, \ldots, x_{n-1}, s) \mapsto (x_1, \ldots, x_{n-1}, x_n)$. Łatwo sprawdzić, że $|J| = 1$.

$$\begin{aligned}
f_{X_1,\ldots,X_{n-1},S}(x_1, \ldots, x_{n-1}, s) &= f_{X_1,\ldots,X_{n-1},X_n}(x_1, \ldots, x_{n-1}, s - x_1 - \cdots - x_{n-1}) \cdot |J| \\
&= f_{X_1}(x_1) \cdots f_{X_{n-1}}(x_{n-1}) f_{X_n}(s - x_1 - \cdots - x_{n-1}) \cdot |J| \\
&= \theta^n \exp\left[-\theta(x_1 + \cdots + x_{n-1} + s - x_1 - \cdots - x_{n-1})\right] \\
&= \theta^n \exp\left[-\theta s\right] \\
&= \frac{\theta^n}{(n-1)!} s^{n-1} \mathrm{e}^{-\theta s} \cdot \frac{(n-1)!}{s^{n-1}} \\
&= f_S(s) \cdot f_{X_1,\ldots,X_{n-1}|S}(x_1, \ldots, x_{n-1} | s)
\end{aligned}$$

Stąd widać, że $S \sim \text{Gamma}(n, \theta)$ i $(X_1, \ldots, X_{n-1} | S = s) \sim \text{U}(s\Sigma)$, gdzie $s\Sigma = \{(x_1, \ldots, x_{n-1}) : x_i \geqslant 0, \sum_i x_i \leqslant s\}$. Rozkład warunkowy obserwacji przy danej wartości statystyki $S = s$ nie zależy od parametru θ.

Rozwiązanie Zadania 2.16.

Zadanie. Znaleźć rozkład zmiennej losowej

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

w modelu normalnym. Porównać z twierdzeniem Fishera (Stwierdzenie 2.2.3).

Wystarczy napisać

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

i zauważyć, że $(X_i-\mu)/\sigma\sim \mathrm{N}(0,1)$. Suma kwadratów ma zatem rozkład $\chi^2(n)$.

Rozwiązanie Zadania 2.17.

Zadanie. (Ciąg dalszy). Wyprowadzić tożsamość

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2+\frac{n}{\sigma^2}(\bar{X}-\mu)^2.$$

Jaki jest rozkład prawdopodobieństwa pierwszego i drugiego składnika po prawej stronie?

Tożsamość sprawdza się łatwo, podobnie jak w Zadaniu 1.4. Z Twierdzenia Fishera wynika, że pierwszy składnik ma rozkład $\chi^2(n-1)$. Drugi składnik ma, oczywiście, rozkład $\chi^2(1)$.

Rozwiązanie Zadania 2.18.

Zadanie. Rozważmy jednoparametryczną wykładniczą rodzinę rozkładów z gęstościami danymi wzorem $f_\psi(x)=\exp\big(T(x)\psi-b(\psi)\big)h(x)$. Pokazać, że

$$\mathbb{E}_\psi T(X)=\frac{\partial b(\psi)}{\partial\psi}.$$

Różniczkując pod znakiem całki i korzystając z tego, że $b(\psi)=\log\int_{\mathcal{X}}\big(T(x)\psi\big)h(x)\mathrm{d}x$ otrzymujemy

$$\begin{aligned}\frac{\partial b(\psi)}{\partial\psi}&=\frac{\partial}{\partial\psi}\log\int_{\mathcal{X}}\exp\big(T(x)\psi\big)h(x)\mathrm{d}x\\&=\frac{(\partial/\partial\psi)\int_{\mathcal{X}}\exp\big(T(x)\psi\big)h(x)\mathrm{d}x}{\int_{\mathcal{X}}\exp\big(T(x)\psi\big)h(x)\mathrm{d}x}\\&=\frac{\int_{\mathcal{X}}(\partial/\partial\psi)\exp\big(T(x)\psi\big)h(x)\mathrm{d}x}{\int_{\mathcal{X}}\exp\big(T(x)\psi\big)h(x)\mathrm{d}x}\\&=\frac{\int_{\mathcal{X}}T(x)\exp\big(T(x)\psi\big)h(x)\mathrm{d}x}{\exp\big(b(\psi)\big)}\\&=\int T(x)f_\psi(x)\mathrm{d}x=\mathbb{E}_\psi T(X).\end{aligned}$$

B.3 Zadania do Rozdziału 3

Rozwiązanie Zadania 3.2.

Zadanie. W modelu Hardy'ego-Weinberga, Przykład 3.1.2, skonstruować *estymator największej wiarogodności*, $\hat{\theta} = \mathrm{MLE}(\theta)$.

Obserwujemy trójkę zmiennych losowych (N_1, N_2, N_3) o wielomianowym rozkładzie prawdopodobieństwa $\mathrm{Mult}\,(n, p_1, p_2, p_3) = \mathrm{Mult}\,\big(n, \theta^2, 2\theta(1-\theta), (1-\theta)^2\big)$. Wiarogodność jest dana wzorem

$$f_\theta(n_1, n_2, n_3) = \frac{n!}{n_1! n_2! n_3!} \theta^{2n_1} \big(2\theta(1-\theta)\big)^{n_2} (1-\theta)^{n_3}.$$

Mamy zatem

$$\log f_\theta(n_1, n_2, n_3) = (2n_1 + n_2) \log \theta + (n_2 + 2n_3) \log(1-\theta) + \mathrm{const}.$$

$\mathrm{MLE}(\theta)$ wyznaczamy rozwiązując równanie

$$\frac{\partial}{\partial \theta} \log f_\theta(n_1, n_2, n_3) = \frac{2n_1 + n_2}{\theta} - \frac{n_2 + 2n_3}{1-\theta} = 0.$$

Banalne przekształcenia dają

$$\mathrm{MLE}(\theta) = \frac{2n_1 + n_2}{2n}.$$

Rozwiązanie Zadania 3.3.

Zadanie. (Ciąg dalszy). W modelu Hardy'ego-Weinberga, trzy genotypy oznaczmy symbolami $1 \equiv AA$, $2 \equiv Aa$, $3 \equiv aa$. Wyobraźmy sobie statystykę zliczającą liczbę literek A w próbce. Napisać tę statystykę w terminach N_1, N_2, N_3, obliczyć jej wartość oczekiwaną. Uzasadnić, że ENW z poprzedniego zadania jest w istocie estymatorem podstawienia częstości (tutaj: częstości literek A).

Liczba literek A w próbce jest równa $2n_1 + n_2$. Obliczamy wartość oczekiwaną tej statystyki: $\mathbb{E}_\theta(2N_1 + N_2) = n(2p_1 + p_2) = n(2\theta^2 + 2\theta(1-\theta) = 2n\theta$. Estymator metody momentów (w tym przypadku metoda momentów jest tym samym co metoda podstawienia częstości) otrzymujemy rozwiązując równanie $2n_1 + n_2 = 2n\theta$.

Rozwiązanie Zadania 3.4.

Zadanie. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie (potęgowym) o gęstości

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{dla } 0 < x < 1; \\ 0 & \text{w pozostałych przypadkach,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem.

(a) Podać wzór na *estymator największej wiarogodności* $\hat{\theta}_{\mathrm{MLE}}$ parametru θ w oparciu o próbkę $X_1, X_2, \ldots, X_n$.

(b) Podać wzór na *estymator metody momentów* $\hat{\theta}_{\mathrm{MME}}$ parametru θ w oparciu o tę samą próbkę.

(a) Wiarogodność dla próbki ma postać

$$f_\theta(x_1, \ldots, x_n) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta-1},$$

jeśli $0 < x_i < 1$ dla $i = 1, \ldots, n$. Obliczamy logarytm wiarogodności:

$$\ell(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i.$$

Jest to funkcja wklęsła, która jedyne maksimum przyjmuje w punkcie zerowania się pochodnej. Ponieważ

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i,$$

więc estymatorem największej wiarogodności jest $\hat{\theta}_{\mathrm{MLE}} = -\dfrac{n}{\sum \log x_i}$.

Uwaga: Alternatywna (i dość pouczająca) droga otrzymania tego wyniku jest taka. Można zauważyć, że $-\log X_i$ ma rozkład wykładniczy $\mathrm{Ex}(\theta)$ (proszę sprawdzić). Można skorzystać ze znanej postaci ENW dla próbki z rozkładu wykładniczego.

(b) Obliczenie estymatora metody momentów wymaga tylko wzoru na wartość oczekiwaną:

$$\mathbb{E}X_1 = \int_0^1 x \cdot \theta x^{\theta-1} \mathrm{d}x = \frac{\theta}{\theta+1}.$$

Rozwiązujemy równanie $\dfrac{\theta}{\theta+1} = \bar{x}$ i otrzymujemy $\hat{\theta}_{\mathrm{MME}} = \dfrac{\bar{x}}{1-\bar{x}}$.

Rozwiązanie Zadania 3.5.

Zadanie. Rozważamy rodzinę rozkładów Pareto o gęstości:

$$f_\theta(x) = \begin{cases} \theta/x^{\theta+1} & \text{dla } x > 1; \\ 0 & \text{w przeciwnym przypadku,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Niech $X_1, \ldots, X_n$ będzie próbką losową z wyżej podanego rozkładu.

(a) Obliczyć estymator parametru θ *metodą momentów.*

(b) Obliczyć estymator *największej wiarogodności* parametru θ.

Rozwiązanie jest analogiczne do Zadania 3.4. Należy tylko zwrócić uwagę, że wartość oczekiwana dla rozkładu Pareto może być nieskończona.

Rozwiązanie Zadania 3.8.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu o gęstości

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

gdzie μ, a są nieznanymi parametrami. Wyznaczyć estymatory $\hat{\mu}_n = \hat{\mu}(X_1, \ldots, X_n)$ oraz $\hat{a}_n = \hat{a}(X_1, \ldots, X_n)$ metodą momentów.

Oczywiście, $\mathbb{E}X_1 = \mu$, zatem $\hat{\mu}_n = \bar{x}$. Wariancję obliczymy łatwo, jeśli zauważymy, że rozkład w tym zadaniu jest mieszanką dwóch rozkładów normalnych: $\mathrm{N}(\mu - a, 1)$ i $\mathrm{N}(\mu + a, 1)$. Jeśli $Z \sim \mathrm{N}(0,1)$ i wprowadzimy dodatkowo jest niezależną zmienną losową I o rozkładzie dwupunktowym $\mathbb{P}(I = 1) = \mathbb{P}(I = -1) = 1/2$, to zmienna losowa $aI + Z$ ma rozkład taki zam jak X_1. W konsekwencji $\mathrm{Var}X_1 = a^2 + 1$. Przyrównując to wyrażenie do wariancji próbkowej $\tilde{s}^2 = n^- \sum (x_i - \bar{x})^2$ wnioskujemy, że estymatorem metody momentów jest $\hat{a}_n = \sqrt{\tilde{s}^2 - 1}$.

Rozwiązanie Zadania 3.13.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu jednostajnego $\mathrm{U}(\theta_1, \theta_2)$. Podać wzór na *estymator największej wiarogodności* $\hat{\theta}$ parametru $\theta = (\theta_1, \theta_2)$.

Wiarogodność dla próbki jest dana wzorem

$$\begin{aligned} f_\theta(x_1, \ldots, x_n) &= \frac{1}{\theta_2 - \theta_1} \prod_{i=1}^{n} \mathbb{1}(\theta_1 \leqslant x_i \leqslant \theta_2), \\ &= \frac{1}{\theta_2 - \theta_1} \mathbb{1}(\theta_1 \leqslant \min_{i=1}^{n}(x_i) \leqslant \max_{i=1}^{n}(x_i) \leqslant \theta_2). \end{aligned}$$

Ta funkcja przybiera największą wartość dla $\hat{\theta}_1 = \min_i(x_i)$ i $\hat{\theta}_2 = \max_i(x_i)$.

Rozwiązanie Zadania 3.14.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu jednostajnego $\mathrm{U}(\theta - 1/2, \theta + 1/2)$. Znaleźć *estymator największej wiarogodności* $\hat{\theta}$ parametru (jednowymiarowego) θ. Czy estymator jest wyznaczony jednoznacznie?

Wiarogodność dla próbki jest dana wzorem

$$\begin{aligned} f_\theta(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \mathbb{1}(\theta - 1/2 \leqslant x_i \leqslant \theta + 1/2), \\ &= \mathbb{1}(\theta - 1/2 \leqslant \min_{i=1}^{n}(x_i) \leqslant \max_{i=1}^{n}(x_i) \leqslant \theta + 1/2). \end{aligned}$$

Ta funkcja przybiera największą wartość równą 1 dla dowolnego $\hat{\theta}$ spełniającego warunek $\hat{\theta} - 1/2 \leqslant \min_{i=1}^{n}(x_i) \leqslant \max_{i=1}^{n}(x_i) \leqslant \hat{\theta} + 1/2$. Zbiór takich $\hat{\theta}$ jest przedziałem i każdy z punktów tego przedziału może być uznany za ENW.

B.4 Zadania do Rozdziału 4

Rozwiązanie Zadania 4.1.

Zadanie. Zmienne losowe $X_1, \ldots, X_n$ opisują ceny (w zł.) pewnego artykułu w n różnych sklepach. Zakładamy, że są to zmienne niezależne, o jednakowym rozkładzie normalnym $N(\mu, \sigma^2)$. Interesuje nas estymacja średniej ceny μ. Wyniki wcześniejszych badań sugerują, że nieznana wielkość μ powinna być bliska 200 zł. Wobec tego używamy następującego estymatora:

$$\hat{\mu} = \frac{200 + \bar{X}}{2},$$

gdzie $\bar{X} = \frac{1}{n}\sum X_i$.

(a) Obliczyć *obciążenie* tego estymatora, $\mathbb{E}_{\mu,\sigma}\hat{\mu} - \mu$.

(b) Obliczyć *błąd średniokwadratowy* tego estymatora, $\mathbb{E}_{\mu,\sigma}(\hat{\mu} - \mu)^2$.

Ponieważ $\mathbb{E}_{\mu,\sigma}\hat{\mu} = (200 + \mathbb{E}_{\mu,\sigma}\bar{X})/2 = (200 + \mu)/2$, więc obciążenie naszego estymatora jest równe $(200 - \mu)/2$.

Błąd średniokwadratowy jest równy

$$\underbrace{\left(\frac{200-\mu}{2}\right)^2}_{(\text{obciążenie})^2} + \underbrace{\frac{\sigma^2}{4n}}_{\text{wariancja}}.$$

Uwaga: Estymator $\hat{\theta}$ jest MLE. Estymator $\tilde{\theta}$ jest tak zwanym estymatorem bayesowskim.

Rozwiązanie Zadania 4.2.

Zadanie. Niech $X \sim \text{Bin}(n, p)$ będzie liczbą sukcesów w schemacie Bernoulliego. Obliczyć i porównać *błąd średniokwadratowy* dwóch estymatorów: $\hat{p} = X/n$ oraz $\tilde{p} = (X + 1)/(n + 2)$.

Estymator $\hat{p}$ jest nieobciążony, a wiec jego błąd średniokwadratowy jest równy wariancji, $\text{Var}_p X/n = p(1-p)/n$. Estymator $\tilde{p}$ ma obciążenie $(np + 1)/(n + 2) - p = (1 - 2p)/(n + 2)$ i wariancję $\text{Var}\,\tilde{p} = np(1-p)/(n+2)^2$. Wobec tego, błąd średniokwadratowy estymatora $\tilde{p}$ jest równy

$$\frac{np(1-p) + (1-2p)^2}{(n+2)^2}.$$

Łatwo zauważyć, że estymator $\tilde{p}$ jest lepszy od $\hat{p}$, jeśli nieznany parametr p jest bliski $1/2$ i odwrotnie, estymator $\tilde{p}$ jest gorszyszy od $\hat{p}$, jeśli nieznany parametr p jest bliski 0 lub 1.

Rozwiązanie Zadania 4.4.

Zadanie. W modelu Hardy'ego-Weinberga, Przykład 3.1.2, pokazać, że $\hat{\theta} = \mathrm{MLE}(\theta)$ jest BUE.

Wskazówka: Obliczyć $\mathrm{Var}_\theta \hat{\theta}$ i porównać z dolnym ograniczeniem Craméra-Rao.

Przypomnijmy, że obserwujemy trójkę zmiennych losowych (N_1, N_2, N_3) o rozkładzie wielomianowym $\mathrm{Mult}(n, p_1, p_2, p_3)$, gdzie $p_1 = \theta^2$, $p_2 = 2\theta(1-\theta)$ i $p_3 = (1-\theta)^2$. Pierwszą pochodną logarytmu wiarogodności obliczyliśmy w Zadaniu 3.2. Obliczmy drugą pochodną:

$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(n_1, n_2, n_3) = -\frac{2n_1 + n_2}{\theta^2} - \frac{n_2 + 2n_3}{(1-\theta)^2}.$$

Stąd otrzymamy informację Fishera:

$$I_n(\theta) = -\mathbb{E}_\theta \frac{\partial^2}{\partial\theta^2} \log f_\theta(N_1, N_2, N_3) = \mathbb{E}_\theta \frac{2N_1 + N_2}{\theta^2} + \mathbb{E}_\theta \frac{N_2 + 2N_3}{(1-\theta)^2}$$

$$= 2n\left(\frac{\theta^2 + \theta(1-\theta)}{\theta^2} + \frac{\theta(1-\theta) + (1-\theta)^2}{(1-\theta)^2}\right)$$

$$= 2n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \frac{2n}{\theta(1-\theta)}.$$

Wykorzystaliśmy to, że $\mathbb{E}_\theta N_1 = n\theta^2$, $\mathbb{E}_\theta N_2 = 2n\theta(1-\theta)$ i $\mathbb{E}_\theta N_3 = n(1-\theta)^2$.

Estymator największej wiarogodności parametru θ znaleźliśmy w Zadaniu 3.2:

$$\hat{\theta} = \mathrm{MLE}(\theta) = \frac{2N_1 + N_2}{2n}.$$

Bardzo łatwo sprawdzić, że $\mathbb{E}_\theta \hat{\theta} = \theta$, korzystając z tego, że $\mathbb{E}N_i = np_i$. Wariancję $\hat{\theta}$ można obliczyć korzystając ze wzoru $\mathrm{Var}(2N_1 + N_2) = 4\mathrm{Var}N_1 + \mathrm{Var}N_2 + 4\mathrm{Cov}(N_1, N_2)$. Zastosujemy znane wyrażenia na wariancję i kowariancję dla rozkładu trójmianowego: $\mathrm{Var}N_i = np_i(1-p_i)$, $\mathrm{Cov}(N_i, N_j) = -p_i p_j$. Uwzględniając zależność p_i od θ, po prostych przekształceniach otrzymujemy

$$\mathrm{Var}_\theta \hat{\theta} = \frac{\theta(1-\theta)}{2n} = \frac{1}{I_n(\theta)}.$$

W tym modelu, MLE jest więc BUE .

B.5 Zadania do Rozdziału 5

Rozwiązanie Zadania 5.1.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu jednostajnego $\mathrm{U}(0, \theta)$. Wykazać bezpośredno (i szczegółowo) mocną zgodność estymatora największej wiarogodności $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$.

Wiemy, że $\hat{\theta} = \max(X_1, \ldots, X_n)$. Najpierw pokażemy słabą zgodność, czyli zbieżność według prawdopodobieństwa. Dla ustalonego $\varepsilon > 0$ mamy $\mathbb{P}_\theta(\theta - \varepsilon < \hat{\theta}_n \leqslant \theta) = \mathbb{P}_\theta(\theta - \varepsilon < X_i$ dla $i =$

$1, \ldots, n) = (1 - \varepsilon/\theta)^n \to 0$ przy $n \to \infty$. Aby uzasadnić mocną zgodność, zauważmy, że $\hat{\theta}_n$ jest ciągiem niemalejącym (z prawdopodobieństwem 1), a więc musi mieć granicę w sensie zbieżności prawie na pewno. Ponieważ już wiemy, że $\hat{\theta}_n \to_{\mathrm{P}} \theta$, więc tą granicą musi być θ. Pokazaliśmy, że $\hat{\theta}_n \to_{\text{p.n.}} \theta$.

Rozwiązanie Zadania 5.2.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu wykładniczego $\mathrm{Ex}(\theta)$ o gęstości $f_\theta(x) = \theta \mathrm{e}^{-\theta x}$ dla $x > 0$. Niech $\hat{\theta}_n$ będzie estymatorem *największej wiarogodności* parametru θ.

(a) Wykazać asymptotyczną normalność $\hat{\theta}_n$ korzystając bezpośrednio z CTG i „metody delta" (Lemat 5.2.3). Podać asymptotyczną wariancję.

(b) Otrzymać ten sam rezultat korzystając z Twierdzenia 5.2.5.

(a) Wiemy, że $\hat{\theta} = 1/\bar{X}_n$. Ponieważ $\mathbb{E}_\theta X_1 = 1/\lambda$ i $\mathrm{Var}_\theta X_1 = 1/\lambda^2$, więc CTG implikuje zbieżność $\sqrt{n}(\bar{X}_n - 1/\theta) \to_{\mathrm{d}} \mathrm{N}(0, 1/\theta^2)$. Zastosujemy „metodę delta", przyjmując $h(x) = 1/x$ i $\mu = 1/\theta$. Oczywiście, $h'(\mu) = \theta^2$ i w rezultacie $\sqrt{n}(1/X_n - \theta) \to_{\mathrm{d}} \mathrm{N}(0, \theta^2)$.

(b) W przykładzie 4.2.4 obliczyliśmy, że $I_1(\theta) = 1/\theta^2$, zatem na mocy Stwierdzenia 4.2.8 mamy $I_n(\theta) = n/\theta^2$. Zgodnie z Twierdzeniem 5.2.5 w tym przypadku otrzymujemy $\sqrt{n}(1/\bar{X}_n - \theta) \to_{\mathrm{d}} \mathrm{N}(0, \theta^2)$.

B.6 Zadania do Rozdziału 6

Rozwiązanie Zadania 6.1.

Zadanie. Zważono 10 paczek masła i otrzymano następujące wyniki:

$$245;\ 248;\ 241;\ 251;\ 252;\ 244;\ 246;\ 248;\ 247;\ 248.$$

Zakładamy, że jest to próbka losowa z rozkładu normalnego $\mathrm{N}(\mu, \sigma^2)$ z nieznanymi parametrami μ i σ. Podać przedział ufności dla średniej masy paczki μ na poziomie ufności $1 - \alpha = 0.95$.

Obliczamy $\bar{X} = 247$ i $S = 3.23$. Odpowiedni kwantyl rozkładu t-Studenta jest równy $t = \mathrm{t}_{0.975}(9) = 2.26$. Przedział ufności ma postać $\bar{X} \pm St/\sqrt{n}$. Otrzymujemy wynik $[244.69, 249.31]$.

Rozwiązanie Zadania 6.3.

Zadanie. Dwa laboratoria niezależnie zmierzyły stałą c: prędkość światła w próżni. Każde laboratorium zbudowało przedział ufności dla c na poziomie $1 - \alpha = 0.95$.

(a) Jakie jest prawdopodobieństwo, że *przynajmniej jeden* z dwóch przedziałów zawiera prawdziwą wartość c ?

(b) Jakie jest prawdopodobieństwo, że *oba* przedziały zawierają prawdziwą wartość c ?

Niech A_i oznacza zdarzenie losowe polegające na tym, że przedział obliczony w i-tym laboratorium pokrył prawdziwą wielkość c. Z założenia A_1 i A_2 są niezależne i $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1 - \alpha$. Odpowiedzi: (a) $\mathbb{P}(A_1 \cup A_2) = 1 - \alpha^2$, (b) $\mathbb{P}(A_1 \cap A_2) = (1 - \alpha)^2$.

Rozwiązanie Zadania 6.4.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu $\mathrm{U}(0, \theta)$, z nieznanym parametrem $\theta > 0$, Przykład 3.2.8. Skonstruować przedział ufności dla parametru θ postaci $[M_n, c_n M_n]$, gdzie $M_n = \max(X_1, \ldots, X_n)$. Dobrać stałą c_n tak, aby prawdopodobieństwo pokrycia było równe $1 - \alpha$.

Zawsze zachodzi nierówność $M_n \leqslant \theta$. Należy dobrać c_n tak, żeby $\mathbb{P}_\theta(c_n M_m \geqslant \theta) = 1 - \alpha$. Ale $\mathbb{P}_\theta(c_n M_m \geqslant \theta) = \mathbb{P}_\theta(M_n > \theta / c_n) = 1 - (1/c_n)^n$, skąd wnioskujemy, że $c_n = \alpha^{-1/n}$.

Rozwiązanie Zadania 6.5.

Zadanie. Jaka powinna być minimalna liczebność próbki pochodzącej z rozkładu normalnego $\mathrm{N}(\mu, \sigma^2)$, gdzie $\sigma > 0$ jest znane, aby przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ miał długość nie przekraczającą $2d$, $d > 0$?

Przedział ufności ma postać $\bar{X} \pm \sigma z / \sqrt{n}$, gdzie $z = \mathrm{z}_{1-\alpha/2}$. Wobec tego musimy wybrać takie n, żeby $\sigma z / \sqrt{n} \leqslant d$, czyli $n \geqslant \sigma^2 z^2 / d^2$.

Rozwiązanie Zadania 6.10.

Zadanie. Niech $X_1, ..., X_{400}$ będzie próbką z nieznanego rozkładu o dystrybuancie F. Załóżmy, że m jest jednoznacznie wyznaczoną medianą tego rozkładu, $F(m) = 1/2$. Pokazać, że

$$\mathbb{P}_F(X_{183:400} \leqslant m \leqslant X_{217:400}) \simeq 0.9.$$

Wskazówka: Porównać z poprzednim zadaniem. Użyć przybliżenia rozkładu dwumianowego rozkładem normalnym, czyli CTG.

Rozważmy schemat Bernoulliego, w którym i-te doświadczenie kończy się "sukcesem", jeśli $X_i \leqslant m$ lub "porażką", jeśli $X_i > m$. Oczywiście, prawdopodobieństwo sukcesu jest równe $1/2$. Niech $L_n = \sum_{i=1}^n \mathbb{1}(X_i \leqslant m)$ oznacza liczbę sukcesów.

$$\mathbb{P}_F(X_{183:400} \leqslant m \leqslant X_{217:400}) = \mathbb{P}(183 \leqslant L_{400} < 217).$$

Prawdopodobieństwo w powyższym wzorze nie jest łatwo obliczyć bez pomocy komputera, ale możemy zastosować przybliżenie normalne. Z tablic rozkładu normalnego odczytujemy kwantyl $z = \mathrm{z}_{0.95} = 1.645$. Ponieważ $\mathbb{E}L_{400} = 200$ i $\mathrm{Var}\, L_{400} = 10^2$, więc $\mathbb{P}(183 \leqslant L_{400} < 217) \simeq \mathbb{P}(-1.655 < Z < 1.645) \simeq 0.9$ (Z oznacza tu standardową zmienną normalną, $Z \sim \mathrm{N}(0, 1)$).

B.7 Zadania do Rozdziału 7

Rozwiązanie Zadania 7.1.

Zadanie. W Przykładzie 7.1.7 obliczyć p-wartość testu.

Wskazówka: Rozkład $\chi^2(2)$ jest rozkładem wykładniczym Ex(1/2).

Otrzymana wartość statystyki testowej wynosi $\chi^2 = 2.53$. Przy założeniu prawdziwości hipotezy zerowej, statystyka ma w przybliżeniu rozkład wykładniczy, więc ze wzoru na dystrybuantę rozkładu wykładniczego dostajemy $\mathbb{P}(\chi^2 > 2.53) \simeq \mathrm{e}^{-2.53/2} = 0.282$.

B.8 Zadania do Rozdziału 8

Rozwiązanie Zadania 8.2.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu $\mathrm{N}(\mu, \sigma^2)$ ze *znaną wartością oczekiwaną* μ. Pokazać, że UMPT $\mathrm{H}_0 : \sigma \leqslant \sigma_0$ przeciw $\mathrm{H}_1 : \sigma > \sigma_0$ na poziomie istotności α odrzuca hipotezę zerową gdy $\sum(X_i - \mu)^2 > c$. Jak wybrać próg c?

Rozważmy najpierw zadanie testowania hipotez prostych: $\mathrm{H}'_0 : \sigma = \sigma_0$ przeciw $\mathrm{H}'_1 : \sigma = \sigma_1$, dla ustalonej wartości $\sigma_1 > \sigma_0$. Obliczamy iloraz wiarogodności:

$$\frac{f_{\sigma_1}(x_1, \ldots x_n)}{f_{\sigma_0}(x_0, \ldots x_n)} = \exp\left[\left(\frac{1}{2\sigma_0} - \frac{1}{2\sigma_1}\right)\sum_{i=1}^{n}(x_i - \mu)^2\right].$$

Widać, że iloraz wiarogodności jest rosnącą funkcją statystyki $n\hat{\sigma}^2 = \sum_{i=1}^{n}(x_i - \mu)^2$. Stąd wynika, że najmocniejszy test H'_0 przeciw H'_1 jest postaci $n\hat{\sigma}^2 > c$. Stałą c dobieramy tak, żeby $\mathbb{P}_{\sigma_0}(n\hat{\sigma}^2 > c) = \alpha$. Wiemy (Zadanie 2.16), że przy założeniu prawdziwości H'_0, zmienna losowa $n\hat{\sigma}^2/\sigma_0^2$ ma rozkład chi-kwadrat z n stopniami swobody, a więc $c = \chi^2_{1-\alpha}(n)\sigma_0^2$. Otrzymany test nie zależy od wyboru $\sigma_1 > \sigma_0$. Wynika stąd, że jest to TJNM prostej hipotezy zerowej $\mathrm{H}'_0 : \sigma = \sigma_0$ przeciw złożonej alternatywie $\mathrm{H}_1 : \sigma > \sigma_0$. Pozostawiamy Czytelnikowi sprawdzenie, że tenże test jest TJNM w wyjściowym zadaniu H_0 przeciw H_1.

Rozwiązanie Zadania 8.3.

Zadanie. Obserwujemy pojedynczą zmienną losową X o rozkładzie o gęstości

$$f_\theta(x) = \begin{cases} (\theta+1)x^\theta & \text{jeśli } 0 \leqslant x \leqslant 1; \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Testujemy hipotezę $\mathrm{H}_0 : \theta = 0$ przeciwko $\mathrm{H}_1 : \theta > 0$.

(a) Skonstruować TJNM na poziomie istotności α.

(b) Obliczyć funkcję mocy, $1-\beta(\theta)$, tego testu.

Dla ustalonego $\theta > 0$ obliczamy iloraz wiarogodności:

$$\frac{f_\theta(x)}{f_0(x)} = (\theta+1)x^\theta.$$

Widać, że iloraz wiarogodności jest rosnącą funkcją statystyki X. Najmocniejszy test na poziomie istotności α jest postaci $X > c$ (odrzuca H_0 na rzecz H_1 jeśli $X > c$), dla stałej c dobranej tak, żeby $\mathbb{P}_{\theta_0}(X > c) = \alpha$. Rozkład $\mathbb{P}_{\theta_0}$ jest po prostu rozkładem jednostajnym, a zatem $c = 1-\alpha$. dotychczasowe rozważania dotyczyły ustalonej wartości $\theta > 0$, ale otrzymany test nie zależy od wyboru θ, a więc jest jednostajnie najmocniejszy. Pozostaje obliczyc jego moc:

$$1-\beta(\theta) = \mathbb{P}_{\theta_0}(X > 1-\alpha) = \int_{1-\alpha}^{1} (\theta+1)x^\theta \mathrm{d}x = 1-(1-\alpha)^{\theta+1}.$$

Rozwiązanie Zadania 8.8.

Zadanie. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu o dystrybuancie F. Przeprowadzamy test Kołmogorowa-Smirnowa hipotezy $H_0 : F = F_0$, gdzie F_0 jest dystrybuantą rozkładu wykładniczego Ex(1), przeciwko alternatywie $F \neq F_0$. Statystyką testową jest $D_n = \sup_x |\hat{F}_n(x) - F_0(x)|$. Naprawdę, próbka pochodzi z rozkładu Ex(2).

(a) Zbadać zbieżność ciągu zmiennych losowych D_n prawie na pewno.

(b) Zbadać zbieżność ciągu odpowiednich p-wartości prawie na pewno.

Wskazówka: Wykorzystać Twierdzenie Gliwienki-Cantelliego.

Z twierdzenie Gliwienki-Cantelliego wynika zbieżnośc prawie na pewno do zera ciągu zmiennych losowych $\sup_x |\hat{F}_n(x) - F_1(x)|$, gdzie F_1 jest dystrybuantą rozkładu Ex(2). Stąd natychmiast widać, że $D_n = \sup_x |\hat{F}_n(x) - F_0(x)| \to \sup_x |F_1(x) - F_0(x)|$ w sensie zbieżności prawie na pewno. W naszym konkretnym przykładzie możemy podać jawny wzór na granicę: $\sup_{x>0} |(1-\mathrm{e}^{-2x}) - (1-\mathrm{e}^{-x})| = 1/4$, a więc $D_n \to_{\text{p.n.}} 1/4$.

B.9 Zadania do Rozdziału 9

Rozwiązanie Zadania 9.3.

Zadanie. Wyprowadzić bezpośrednio wzory na $\mathrm{Var}\hat{\beta}_1$ i $\mathrm{Var}\hat{\beta}_0$.

Skorzystamy ze wzorów (9.2.1). Wygodnie będzie nieco przekształcić wzór na $\hat{\beta}_1$. Ponieważ mamy $\sum(x_i - \bar{x}) = 0$, więc

$$\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})Y_i}{\sum(x_i-\bar{x})^2}.$$

Wprowadźmy oznaczenie $\mathrm{SS}_x = \sum(x_i - \bar{x})^2$ (Sum of Squares for x). Zmienne Y_i są niezależne i każda z nich ma wariancję równą σ^2, więc $\mathrm{Var}\sum(x_i - \bar{x})Y_i = \mathrm{SS}_x\sigma^2$. Ostatecznie zatem $\mathrm{Var}\hat{\beta}_1 = \sigma^2/\mathrm{SS}_x$.

Zauważmy teraz, że $\mathrm{Cov}(\bar{Y}, \hat{\beta}_1) = 0$. Istotnie, ponieważ $\mathrm{Cov}(Y_j, Y_i) = \sigma^2\mathbb{1}(j = i)$, więc mamy $\mathrm{Cov}(Y_j, \sum_i(x_i - \bar{x})Y_i) = (x_j - \bar{x})\sigma^2$. Stąd wynika, że $\mathrm{Cov}(\sum_j Y_j, \sum_i(x_i - \bar{x})Y_i) = \sum_j(x_j - \bar{x})\sigma^2 = 0$. Korzystając z faktu nieskorelowania zmiennych losowych $\bar{Y}$ i $\hat{\beta}_1$, możemy teraz obliczyć wariancję $\hat{\beta}_0$. Mamy $\mathrm{Var}\hat{\beta}_0 = \mathrm{Var}\hat{Y} + \bar{x}^2\mathrm{Var}\hat{\beta}_1 = \sigma^2/n + \bar{x}^2\sigma^2/\mathrm{SS}_x$. Po prostych przekształceniach, $\mathrm{Var}\hat{\beta}_0 = \frac{\sigma^2}{n}\frac{\sum x_i^2}{\mathrm{SS}_x}$.

Rozwiązanie Zadania 9.4.

Zadanie. Pokazać, że zmienne losowe $\bar{Y}$ i $\hat{\beta}_1$ są niezależne.

Wskazówka: Skorzystać z poprzednich zadań.

Zmienne losowe $\bar{Y}$ i $\hat{\beta}_1$ mają łączny rozkład normalny (dwuwymiarowy). W rozwiązaniu poprzedniego zadania sprawdziliśmy, że $\mathrm{Cov}(\bar{Y}, \hat{\beta}_1) = 0$. Wiadomo z rachunku prawdopodobieństwa, że dla zmiennych o łącznym rozkładzie normalnym, nieskorelowanie implikuje niezależność.

Rozwiązanie Zadania 9.5.

Zadanie. Wyprowadzić bezpośrednio wzory na $\mathrm{Var}\hat{Y}^*$ i $\mathrm{Var}(Y^* - \hat{Y}^*)$.

Możemy napisać $\hat{Y}_* = \bar{Y} + \hat{\beta}_1(x_* - \bar{x})$, więc

$$\mathrm{Var}\hat{Y}_* = \mathrm{Var}\bar{Y} + (x_* - \bar{x})^2\mathrm{Var}\hat{\beta}_1 = \sigma^2\left[\frac{1}{n} + \frac{(x_* - \bar{x})^2}{\mathrm{SS}_x}\right].$$

Ponieważ zmienna Y_* jest niezależna od $\hat{Y}_*$ (dlaczego?) to

$$\mathrm{Var}(Y_* - \hat{Y}_*) = \mathrm{Var}Y_* + \mathrm{Var}\hat{Y}_* = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{\mathrm{SS}_x}\right].$$

Rozwiązanie Zadania 9.6.

Zadanie. Udowodnić bezpośredno (nie korzystając z geometrycznych rozważań w przestrzeni $\mathbb{R}^n$) podstawową tożsamość analizy wariancji:

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2.$$

Napiszmy lewą stronę w postaci $\sum(\hat{Y}_i - \bar{Y} + Y_i - \hat{Y}_i)^2$. Oczywiście, $(\hat{Y}_i - \bar{Y} + Y_i - \hat{Y}_i)^2 = (\hat{Y}_i - \bar{Y})^2 + (Y_i - \hat{Y}_i)^2 + 2(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$. Zauważmy, że $\hat{Y}_i - \bar{Y} = \hat{\beta}_1(x_i - \bar{x})$ i $Y_i - \hat{Y}_i = Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})$. Stąd $\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \hat{\beta}_1\sum(x_i - \bar{x})(Y_i - \bar{Y}) - \hat{\beta}_1^2\sum(x_i - \bar{x})^2 = 0$. Suma iloczynów mieszanych znika i pozostają tylko dwie sumy kwadratów po prawej stronie tożsamości.

Rozwiązanie Zadania 9.9.

Zadanie. Wyprowadzić wzory na estymatory największej wiarogodności w modelu prostej regresji liniowej *bez wyrazu wolnego*,

$$Y_i = \beta x_i + \varepsilon_i, \qquad (i = 1, \ldots, n),$$

przyjmując Założenie 9.1.2.

Zmienne Y_i są niezależne i $Y_i \sim \mathrm{N}(\beta x_i, \sigma^2)$. Stąd wynika, że wiarogodność jest dana wzorem

$$f_{\beta,\sigma}(y_1, \ldots, y_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right].$$

Estymator największej wiarogodności $\hat{\beta}$ otrzymujemy minimalizując sumę kwadratów

$$\mathrm{RSS} = \sum_{i=1}^{n}(y_i - \beta x_i)^2,$$

czyli rozwiązując równanie $\sum x_i(y_i - \beta x_i) = 0$. W rezultacie $\hat{\beta} = \sum_i x_i y_i / \sum_i x_i^2$. Estymator największej wiarogodności $\hat{\sigma}$ otrzymujemy maksymalizując logarytm wiarogodności. Łatwo się przekonać, że $\hat{\sigma}^2 = \mathrm{RSS}/n$, przy czym RSS obliczamy dla $\beta = \hat{\beta}$.

Rozwiązanie Zadania 9.11.

Zadanie. Pokazać, że statystyka F testu analizy wariancji jest równoważna statystyce ilorazu wiarogodności dla modeli zagnieżdżonych (Punkt 8.3.1 w Podrozdziale 8.3).

Wskazówka: Bardzo podobne rozważania przeprowadziliśmy w Przykładzie 8.3.2. Zbudujemy test ilorazu wiarogodności. Wiarogodność jest dana wzorem

$$\mathcal{L}(\mu, \sigma) = \left(\frac{\mathrm{const}}{\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p}\sum_{i=1}^{n_j}(Y_{ji} - \mu_j)^2\right].$$

Łatwo sprawdzić, że estymatory *bez ograniczeń* (w dużym modelu) są następujące:

$$\hat{\mu}_j = \bar{Y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n_j}(Y_{ji} - \bar{Y}_j)^2.$$

Podmodel jest zupełnie prosty do przeanalizowania. Jeśli hipoteza H_0 jest prawdziwa, to Y jest próbką rozmiaru n z rozkładu $\mathrm{N}(\mu_0, \sigma^2)$. Rolę estymatorów *z ograniczeniami* pełnią doskonale znane, typowe estymatory:

$$\dot{\mu}_0 = \bar{Y} = \frac{1}{n}\sum_{j=1}^{p} n_j \bar{Y}_j = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n_j} Y_{ji}, \qquad \dot{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n_j}(Y_{ji} - \bar{Y})^2.$$

Statystyką ilorazu wiarogodności jest, jak łatwo widzieć, $\mathcal{L}(\hat{\mu}, \hat{\sigma})/\mathcal{L}(\dot{\mu}, \dot{\sigma}) = (\dot{\sigma}/\hat{\sigma})^n$. Test każe odrzucić hipotezę zerową, gdy ta statystyka przekracza odpowiednio ustalony próg. Przekształcimy

ten test do równoważnej, powszechnie używanej postaci. Przypomnijmy oznaczenia na sumy kwadratów (całkowitą – TSS, pomiędzy próbkami – BSS i wewnątrz próbek – WSS). Zauważmy, że $\dot{\sigma}^2 = \mathrm{TSS}/n$ i $\hat{\sigma}^2 = \mathrm{WSS}/n$. Iloraz wiarogodności jest zatem rosnącą funkcją TSS/WSS. Wykorzystując tożsamość analizy wariancji $\mathrm{TSS} = \mathrm{BSS} + \mathrm{WSS}$ możemy napisać $\mathrm{TSS}/\mathrm{WSS} = 1 + \mathrm{BSS}/\mathrm{WSS}$. Stąd już widać, że iloraz wiarogodności jest rosnącą funkcją statystyki Snedecora,

$$F = \frac{\mathrm{BSS}/(p-1)}{\mathrm{WSS}/(n-p)}.$$

Dla ustalenia progu odrzuceń dla LRT nie korzystamy z *asymptotycznego* wyniku wobec znajomości *dokładnego* rozkładu prawdopodobieństwa statystyki F (przy prawdziwości H_0 jest to rozkład F-Snedocora). Hipotezę H_0 odrzucamy, jeśli

$$F > \mathrm{F}_{1-\alpha}(p-1, n-p).$$