

Część II

Estymacja

Rozdział 3

Metody estymacji

W statystyce matematycznej zakładamy, że rozkład prawdopodobieństwa opisujący doświadczenie należy do rodziny $\{\mathbb{P}_\theta : \theta \in \Theta\}$, ale nie znamy parametru θ . **Estymatorem** parametru θ nazywamy dowolną statystykę $T = T(X)$ o wartościach w zbiorze Θ . Interpretujemy T jako przybliżenie θ . Zwykle oznaczamy estymator tą samą literką, co wielkość estymowaną, dodając „daszek”: $T = \hat{\theta}$. Zaczniemy od przeglądu prostych metod konstrukcji estymatorów.

3.1 Metody heurystyczne

Podstawianie częstości jest sposobem estymacji, który natychmiast się narzuca i jest zrozumiały dla każdego.

3.1.1 Przykład (Prawdopodobieństwo niewadliowości wyrobu). Rozpatrzmy jeszcze raz model statystycznej kontroli jakości z Przykładów 1.1.7 i 2.1.4. Przypomnijmy, że parametr p oznacza prawdopodobieństwo pojawienia się sztuki prawidłowej. Oczywistym estymatorem p jest

$$\hat{p} = \bar{X} = S/n,$$

gdzie $S = \sum X_i$, czyli frakcja prawidłowych sztuk w próbie losowej. Zastąpiliśmy nieznanne prawdopodobieństwo przez *prawdopodobieństwo empiryczne*. Na tym właśnie polega metoda podstawiania częstości. \diamond

Metoda podstawiania częstości dopuszcza dużą dowolność. Pokazuje to następny przykład.

3.1.2 Przykład (Model Hardy'ego – Weinberga). W populacji mamy osobników o trzech genotypach: 1, 2 i 3. Z rozważań genetycznych wynika, że te trzy typy powinny występować w proporcji $\theta^2 : 2\theta(1 - \theta) : (1 - \theta)^2$. Wybieramy losowo n osobników, wśród których liczby poszczególnych genotypów są, odpowiednio, N_1, N_2 i N_3 . Mamy do czynienia z rozkładem wielomianowym: $(N_1, N_2, N_3) \sim \text{Mult}(n, p_1, p_2, p_3)$:

$$\mathbb{P}_\theta(N_1 = n_1, N_2 = n_2, N_3 = n_3) = \frac{n!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}.$$

Prawdopodobieństwa p_i nie są tu dowolnymi liczbami dodatnimi sumującymi się do jedynki. Wiemy, że

$$p_1 = \theta^2, \quad p_2 = 2\theta(1 - \theta), \quad p_3 = (1 - \theta)^2.$$

Zadanie polega na estymacji θ . Nasuwają się dwa rozwiązania. Z jednej strony mamy $\theta = \sqrt{p_1}$ i za estymator możemy przyjąć

$$\hat{\theta} = \sqrt{\hat{p}_1} = \sqrt{\frac{N_1}{n}}.$$

Z drugiej strony, $\theta = 1 - \sqrt{p_3}$ i równie rozsądny wydaje się estymator

$$\tilde{\theta} = 1 - \sqrt{\hat{p}_3} = 1 - \sqrt{\frac{N_3}{n}}$$

Który estymator wybrać? Do tej kwestii wrócimy później. \diamond

Metoda momentów jest równie prosta. Przyrównujemy momenty rozkładu teoretycznego, zależące od nieznanego parametru, do ich odpowiedników empirycznych. Z powstałych w ten sposób równań wyliczamy parametr. Najczęściej używamy momentów centralnych i równania wyglądają tak:

$$\mu(\theta) = \hat{\mu}; \quad m_k(\theta) = \hat{m}_k,$$

gdzie $\mu(\theta) = \int x f_\theta(x) dx$, $m_k(\theta) = \int (x - \mu(\theta))^k f_\theta(x) dx$, $\hat{\mu} = \bar{X}$ i $\hat{m}_k = (1/n) \sum (X_i - \bar{X})^k$. Układamy tyle równań, ile jest niewiadomych parametrów (współrzędnych wektora θ).

3.1.3 Przykład (Rozkład wykładniczy). Niech X_1, \dots, X_n będzie próbką z rozkładu $\text{Ex}(\theta)$. Wiemy, że $\mu(\theta) = 1/\theta$. Mamy jeden nieznaną parametr, więc wystarczy jedno równanie: $1/\theta = \bar{X}$. Estymator otrzymany metodą momentów jest równy $\hat{\theta} = 1/\bar{X}$. \diamond

3.1.4 Przykład (Rozkład gamma). Niech X_1, \dots, X_n będzie próbką z rozkładu $\text{Gamma}(\alpha, \lambda)$. Mamy dwa nieznaną parametry, więc wykorzystamy dwa momenty: wartość oczekiwaną $\mu(\alpha, \lambda) = \alpha/\lambda$ i wariancję $\sigma^2(\alpha, \lambda) = \alpha/\lambda^2$. Dostajemy układ równań

$$\frac{\alpha}{\lambda} = \bar{X}, \quad \frac{\alpha}{\lambda^2} = \tilde{S}^2,$$

gdzie \tilde{S}^2 jest wariancją z próbki. Rozwiązanie jest następujące:

$$\hat{\lambda} = \frac{\bar{X}}{\tilde{S}^2}, \quad \hat{\alpha} = \frac{\bar{X}^2}{\tilde{S}^2}.$$

\diamond

Przejdziemy do opisu metody, która odgrywa znacznie ważniejszą rolę.

3.2 Wiarygodność

Idea jest taka: wybieramy taki parametr θ , dla którego otrzymane wyniki doświadczenia losowego są „najbardziej prawdopodobne”. Sformułujemy to dokładniej. Niech $f_\theta(x)$ będzie łączną gęstością obserwacji.

3.2.1 DEFINICJA. *Wiarygodność jest to funkcja $\mathcal{L} : \Theta \rightarrow \mathbb{R}$ dana wzorem*

$$\mathcal{L}(\theta) = f_\theta(x).$$

Wiarygodność jest właściwie tym samym, co gęstość prawdopodobieństwa, ale rozważana jako funkcja parametru θ , przy ustalonych wartościach obserwacji $x = X(\omega)$.

3.2.2 DEFINICJA. *Mówimy, że $\hat{\theta} = \hat{\theta}(X)$ jest estymatorem największej wiarygodności parametru θ , jeśli*

$$f_{\hat{\theta}(x)}(x) = \sup_{\theta \in \Theta} f_\theta(x),$$

dla dowolnego x . Symbolicznie będziemy pisać $\hat{\theta} = \text{MLE}(\theta)$.

Oznaczenie MLE pochodzi od angielskiego terminu *Maximum Likelihood Estimator*. Definicję $\text{MLE}(\theta)$ zapiszemy w skrócie

$$\mathcal{L}(\hat{\theta}) = \sup_{\theta \in \Theta} \mathcal{L}(\theta),$$

musimy jednak pamiętać, że zarówno \mathcal{L} , jak i $\hat{\theta}$ zależą od obserwacji. Pominiemy dyskusję na temat istnienia i jednoznaczności MLE.

3.2.3 Przykład (Liczenie ryb w jeziorze). W jeziorze pływa pewna liczba, powiedzmy r , ryb. Żeby oszacować nieznaną liczbę r bez osuszania jeziora, musimy przywołać na pomoc statystykę matematyczną. Najpierw odławiamy m ryb, *znaczymy* je i wpuszczamy z powotem do jeziora. Czekamy, aż znaczone ryby „wymieszają się” z pozostałymi. Następnie wyławiamy n ryb. Stwierdzamy, że jest wśród nich k znaczonych (złapanych powtórnie). W tym doświadczeniu, r jest nieznanym parametrem, m i n są znanymi liczbami, zaś k jest znaną wartością *zmiennej losowej* K , o rozkładzie hipergeometrycznym

$$\mathcal{L}(r) = \mathbb{P}_r(K = k) = \binom{m}{k} \binom{r-m}{n-k} / \binom{r}{n}.$$

Wyznamy $\text{MLE}(r)$. W tym celu trzeba znaleźć maksimum

$$\mathcal{L}(r) = \max_r.$$

Przyjmijmy wygodne oznaczenie $(n)_k = n(n-1)\cdots(n-k+1)$. Ponieważ

$$\frac{\mathcal{L}(r+1)}{\mathcal{L}(r)} = \frac{(r+1-m)_{n-k}}{(r+1)_n} \cdot \frac{(r)_n}{(r-m)_{n-k}} = \frac{r+1-m}{r-m-n+k+1} \cdot \frac{r-n+1}{r+1},$$

więc $\mathcal{L}(r+1) > \mathcal{L}(r)$ wtedy i tylko wtedy gdy $r < mn/k$. Stąd widać, że $\mathcal{L}(r)$ osiąga maksymalną wartość dla najmniejszej liczby całkowitej przekraczającej mn/k . Innymi słowy,

$$\hat{r} = \text{MLE}(r) = \left\lceil \frac{mn}{k} \right\rceil.$$

Wynik jest zgodny ze zdrowym rozsądkiem. \diamond

Prosty chwyt często ułatwia wyznaczenie MLE. Funkcja $\mathcal{L}(\theta)$ osiąga maksimum w tym samym punkcie, co jej *logarytm*. Niech

$$\ell(\theta) = \log \mathcal{L}(\theta).$$

3.2.4 Przykład (Rozkład wykładniczy). Jeśli X_1, \dots, X_n jest próbka z rozkładu $\text{Ex}(\theta)$, to wiarygodność jest dana wzorem

$$\mathcal{L}(\theta) = \prod_{i=1}^n (\theta e^{-\theta x_i}),$$

więc

$$\ell(\theta) = n \log \theta - \theta \sum x_i.$$

Wystarczy teraz przyrównać pochodną do zera:

$$\ell'(\theta) = \frac{n}{\theta} - \sum x_i = 0.$$

Łatwo się przekonać, że $\ell(\theta)$ osiąga maksimum w punkcie

$$\hat{\theta} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}.$$

Zatem $\text{MLE}(\theta) = 1/\bar{X}$. Metoda największej wiarygodności doprowadziła do tego samego estymatora, co metoda momentów. \diamond

3.2.5 Przykład (Rozkład Poissona). Dla próbki z rozkładu $\text{Poiss}(\theta)$ mamy

$$\ell(\theta) = -\theta n + \log(\theta) \sum x_i - \sum \log(x_i!),$$

więc $\ell'(\theta) = -n + \sum x_i/\theta = 0$ dla $\theta = \sum x_i/n$. Otrzymujemy znów dobrze znany estymator: $\text{MLE}(\theta) = \bar{X}$. \diamond

Sposób obliczenia MLE w Przykładach 3.2.4 i 3.2.5 jest bardzo typowy. Jeśli mamy próbkę z rozkładu o gęstości f_θ i zbiór parametrów $\Theta \subseteq \mathbb{R}$ jest przedziałem, to

$$\begin{aligned}\mathcal{L}(\theta) &= f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n), \\ \ell(\theta) &= \log \mathcal{L}(\theta) = \sum_{i=1}^n \log f_\theta(x_i).\end{aligned}$$

MLE wyznaczamy zazwyczaj rozwiązując równanie

$$\ell'(\theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) = 0.$$

Przypadek *wielu* nieznanymi parametrów wymaga tylko oczywistej modyfikacji. Jeśli mamy zbudować estymator wektora $\theta = (\theta_1, \dots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$, to obliczamy pochodne *cząstkowe* logarytmu wiarygodności. Rozwiązujemy układ k równań z k niewiadomymi:

$$\frac{\partial}{\partial \theta_j} \ell(\theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta_j} \log f_\theta(x_i) = 0, \quad (j = 1, \dots, k).$$

3.2.6 Przykład (Rozkład normalny). Rozważmy próbkę X_1, \dots, X_n z rozkładu $N(\mu, \sigma^2)$, z nieznanymi parametrami μ i $\sigma > 0$. Mamy

$$\log f_{\mu, \sigma}(x) = -\frac{1}{2} \log(2\pi) - \log \sigma - \frac{1}{2\sigma^2} (x - \mu)^2.$$

Logarytm wiarygodności dla całej próbki jest równy

$$\ell(\mu, \sigma) = \sum_{i=1}^n \log f_{\mu, \sigma}(x_i) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2,$$

czyli

$$\ell(\mu, \sigma) = \text{const} - n \log \sigma - \frac{1}{2\sigma^2} \left(\sum x_i^2 - 2\mu \sum x_i + n\mu^2 \right).$$

Układ równań przybiera postać

$$\begin{aligned}\frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \left(\sum x_i^2 - 2\mu \sum x_i + n\mu^2 \right) = 0; \\ \frac{\partial \ell}{\partial \mu} &= -\frac{1}{\sigma^2} \sum x_i - \frac{n\mu}{\sigma^2} = 0.\end{aligned}$$

Z drugiego równania łatwo wyliczamy $\mu = \sum x_i / n$. Podstawiając do pierwszego równania otrzymujemy MLE:

$$\begin{aligned}\hat{\mu} &= \bar{x}; \\ \hat{\sigma}^2 &= \frac{1}{n} \left(\sum x_i^2 - \bar{x}^2 \right) = \frac{1}{n} \sum (x_i - \bar{x})^2.\end{aligned}$$

Dobrze znane estymatory \bar{X} i \tilde{S}^2 okazują się być $\text{MLE}(\mu)$ i $\text{MLE}(\sigma^2)$, odpowiednio. \diamond

Dla rodzin wykładniczych metoda największej wiarygodności jest blisko związana (w pewnym sensie nawet identyczna) z metodą momentów. Najłatwiej to zauważyć używając naturalnej parametryzacji (2.4.8). Niech $X_1, \dots, X_n \sim_{\text{iid}} f_\psi$. Ponieważ

$$\ell(\psi) = \sum_{j=1}^k \sum_{i=1}^n T_j(x_i) - nb(\psi),$$

więc MLE $\hat{\psi}$ jest rozwiązaniem układu równań

$$\frac{\partial}{\partial \psi_j} \ell(\psi) = \sum_{i=1}^n T_j(x_i) - n \frac{\partial}{\partial \psi_j} b(\psi) = \mathbb{E}_\psi T_j(X),$$

(porównaj z Zadaniem 2.20 z Rozdziału 2). Otrzymujemy k równań dla momentów statystyk dostatecznych, w skrócie, $\bar{T} = \mathbb{E}_\psi T$.

3.2.7 Przykład (Rozkład Laplace'a). Mówimy, że zmienna losowa X ma rozkład **Laplace'a** z parametrami μ i λ , jeśli ma gęstość daną wzorem

$$f_{\mu,\lambda}(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|}.$$

Inaczej, taki rozkład nazywa się rozkładem *podwójnie wykładniczym*. Symbolicznie, piszemy $X \sim \text{Lapl}(\mu, \lambda)$. Interesujący jest MLE(μ) dla próbki z rozkładu Laplace'a. Mamy

$$\ell(\mu, \lambda) = \sum_{i=1}^n \log f_{\mu,\lambda}(x_i) = -n \log 2 + n \log \lambda - \frac{\lambda}{2} \sum_{i=1}^n |x_i - \mu|.$$

Popatrzmy na $\ell(\mu, \lambda)$ jak na funkcję μ , przy ustalonym λ . Funkcja ℓ jest *kawałkami liniowa* i wklęsła. Jej wykres jest łamaną, której wierzchołki mają współrzędne $\mu = x_i$. Pochodna (poza punktami nieróżniczkowalności) jest równa $\partial \ell / \partial \mu = (\lambda/2) \sum \text{sign}(\mu - x_i)$. Widać, że funkcja jest rosnąca na lewo od *mediany próbkowej* i malejąca na prawo od niej. Zatem

$$\text{MLE}(\mu) = \hat{\text{méd}}.$$

Dla parzystych n mediana próbkowa, a więc i MLE(μ), nie są wyznaczone jednoznacznie. Najlepiej naszkicować wykres $\ell(\mu)$ dla, powiedzmy, $n = 4$ i $n = 5$. Wtedy powyższe rozważania staną się jasne.

Po wyznaczeniu MLE(μ), nietrudno znaleźć MLE(λ). Przyrównujemy do zera $\partial \ell / \partial \lambda$ i natychmiast dostajemy

$$\text{MLE}(\lambda) = \frac{2n}{\sum |\hat{\text{méd}} - x_i|}.$$

◇

3.2.8 Przykład (Rozkład jednostajny). Jeśli X_1, \dots, X_n jest próbką z rozkładu $U(0, \theta)$, z nieznanym parametrem $\theta > 0$, to

$$\mathcal{L}(\theta) = \prod_i \left(\frac{1}{\theta} \mathbb{1}(x_i \geq \theta) \right) = \begin{cases} 1/\theta^n & \text{dla } \theta \geq \max(x_1, \dots, x_n); \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Wystarczy popatrzeć chwilę na ten wzór, żeby stwierdzić, że

$$\text{MLE}(\theta) = \max(X_1, \dots, X_n).$$

◇

Metoda największej wiarygodności jest bardzo giętka i daje się zastosować w różnych „nietypowych” sytuacjach.

3.2.9 Przykład (Dane cenzurowane). Przypuśćmy, że interesuje nas rozkład zmiennych losowych X_1, \dots, X_n ale obserwujemy tylko Y_1, \dots, Y_n , gdzie

$$Y_i = \min(X_i, a),$$

dla pewnej znanej liczby a . Jest to najprostszy model „cenzurowania”. Można sobie wyobrazić, że X_i to „czasy życia” elementów zaś a jest „horyzontem obserwacji”. Dla ustalenia uwagi załóżmy, że X_1, \dots, X_n jest próbką z rozkładu $\text{Ex}(\theta)$. Pojedyncza zmienna losowa Y_i ma rozkład, który nie jest ani absolutnie ciągły ani dyskretny. Niemniej, ten rozkład ma gęstość względem sumy miary Lebesgue’a i miary o masie 1 skupionej w a :

$$f_\theta(y) = \begin{cases} \theta e^{-\theta y} & \text{dla } y < a, \\ e^{-\theta y} & \text{dla } y = a. \end{cases}$$

Istotnie, w punkcie $y < a$ gęstość zmiennej Y_i jest równa gęstości zmiennej X_i . Dla $y = a$ mamy $f_\theta(a) = \mathbb{P}_\theta(Y_i = a) = \mathbb{P}_\theta(X_i \geq a) = e^{-\theta a}$. Wiarygodność jest dana wzorem

$$\mathcal{L}(\theta) = \theta^m \exp \left(-\theta \sum_{i=1}^n y_i \right),$$

gdzie $s = \sum_{i=1}^n y_i$ jest sumą wszystkich obserwacji zaś $m = \sum \mathbb{1}(y_i < a)$ oznacza liczbę obserwacji „nieocenzurowanych”. Stąd otrzymujemy wzór na MLE: $\hat{\theta} = m/s$. ◇

3.3 Zadania

3.1. Samoloty bombowe przedzierają się przez dwie linie obrony przeciwniczej. Każdy samolot, niezależnie od pozostałych, z prawdopodobieństwem θ może zostać strącony przez pierwszą linię obrony. Jeśli pokona pierwszą linię, z prawdopodobieństwem θ może zostać strącony przez drugą linię. Prawdopodobieństwo θ jest nieznanne. Spośród $n = 100$ samolotów, $K_1 = 40$ zostało strąconych przez pierwszą linię, a dalszych $K_2 = 20$ zostało strąconych przez drugą linię.

- (a) Obliczyć wiarygodność dla zaobserwowanych wartości K_1 i K_2 , czyli $\mathcal{L} = \mathbb{P}_\theta(K_1 = 40, K_2 = 20)$.
- (b) Podać *estymator największej wiarygodności* parametru θ .

3.2. W modelu Hardy’ego-Weinberga, Przykład 3.1.2, skonstruować *estymator największej wiarygodności*, $\hat{\theta} = \text{MLE}(\theta)$.

3.3. (Ciąg dalszy). W modelu Hardy’ego-Weinberga, trzy genotypy oznaczymy 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 MLE z poprzedniego zadania jest w istocie estymatorem podstawienia częstości (tutaj: częstości literek A).

3.4. Niech X_1, \dots, 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 wiarygodności* $\hat{\theta}_{\text{MLE}}$ parametru θ w oparciu o próbkę X_1, X_2, \dots, X_n .
- (b) Podać wzór na *estymator metody momentów* $\hat{\theta}_{\text{MME}}$ parametru θ w oparciu o tę samą próbkę.

3.5. 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, \dots, X_n będzie próbką losową z wyżej podanego rozkładu.

- (a) Obliczyć estymator parametru θ *metodą momentów*.
- (b) Obliczyć estymator *największej wiarygodności* parametru θ .

3.6. Niech X_1, \dots, X_n będzie próbką z rozkładu o gęstości

$$f_\theta(x) = \begin{cases} \theta x^{-2} e^{-\theta/x} & \text{dla } x > 0; \\ 0 & \text{w przeciwnym przypadku,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Podać wzór na *estymator największej wiarygodności* parametru θ .

3.7. Niech X_1, \dots, X_n będzie próbką z rozkładu (Weibulla) o gęstości

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

gdzie $\theta > 0$ jest nieznanym parametrem. Podać *estymator największej wiarygodności* parametru θ .

3.8. Niech X_1, \dots, 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 $\hat{\mu}_n = \hat{\mu}(X_1, \dots, X_n)$ i $\hat{a}_n = \hat{a}(X_1, \dots, X_n)$ metodą momentów.

3.9. Niech X_1, X_2, \dots, X_{15} będą niezależnymi zmiennymi losowymi, przy czym X_1, X_2, \dots, X_5 mają rozkład wykładniczy $\text{Ex}(\theta)$, a X_6, X_7, \dots, X_{15} rozkład wykładniczy $\text{Ex}(\theta/3)$. Podać wzór na *estymator największej wiarygodności* parametru θ w oparciu o obserwacje X_1, X_2, \dots, X_{15} .

3.10. Niech X_1, X_2, \dots, X_{15} będą niezależnymi zmiennymi losowymi, przy czym X_1, X_2, \dots, X_5 mają rozkład wykładniczy $\text{Ex}(\theta)$, a X_6, X_7, \dots, X_{15} rozkład gamma $\text{Gamma}(2, \theta)$. Podać wzór na *estymator największej wiarygodności* parametru θ w oparciu o obserwacje X_1, X_2, \dots, X_{15} .

3.11. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi, przy czym X_i ma rozkład normalny $N(\mu, \sigma_i^2)$. Zakładamy, że wariancje $\sigma_1^2, \dots, \sigma_n^2$ są znane (a wartość oczekiwana μ jest jednakowa dla wszystkich zmiennych). Podać wzór na *estymator największej wiarygodności* parametru μ w oparciu o obserwacje X_1, X_2, \dots, X_n .

3.12. Niech X_1, \dots, X_n będzie próbką z rozkładu normalnego $N(0, \sigma^2)$ (o wartości oczekiwanej równej 0 i nieznannej wariancji). Podać wzór na *estymator największej wiarygodności* parametru σ , czyli *odchylenia standardowego*.

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

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

Rozdział 4

Teoria estymacji

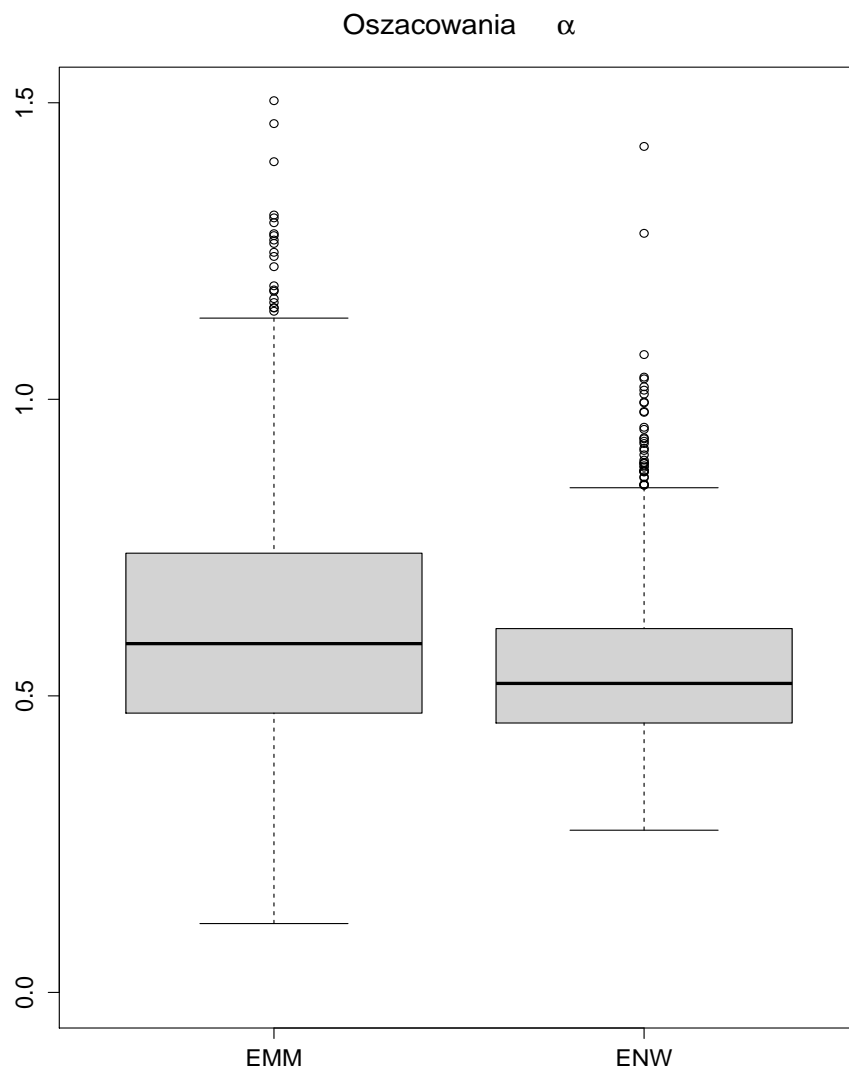
Przykład wstępny

Zasadnicze pytania, na które chcemy odpowiedzieć dotyczą dokładności estymatorów. Jaki błąd popełniamy szacując nieznaną wartość parametru? Co to znaczy, że estymator jest „dobry”? Który z konkurujących ze sobą estymatorów uznać za „lepszy”? Czy można znaleźć estymator „najlepszy”? Zaczniemy od nieformalnej ilustracji w postaci następującego przykładu.

4.0.1 Przykład. Rozważmy próbkę X_1, \dots, X_n z rozkładu $\text{Gamma}(\alpha, \lambda)$. Estymatora największej wiarygodności parametru α nie da się wyrazić prostym wzorem ale można łatwo obliczyć numerycznie w R przy użyciu funkcji `nlm`. Porównajmy MLE z estymatorem otrzymanym metodą momentów w Przykładzie 3.1.4. Próbki rozmiaru $n=25$ były generowane z rozkładu $\text{Gamma}(\alpha, \lambda)$ z $\alpha = 0.5$. Estymowane były oba parametry, ale przedstawimy tylko wyniki dla α . Doświadczenie było powtórzone $m = 1000$ razy. Rysunek 4.1 przedstawia wykresy pudełkowe otrzymanych oszacowań parametru α . Na rysunku 4.2 rozkłady prawdopodobieństwa obu estymatorów są przedstawione w postaci histogramów.

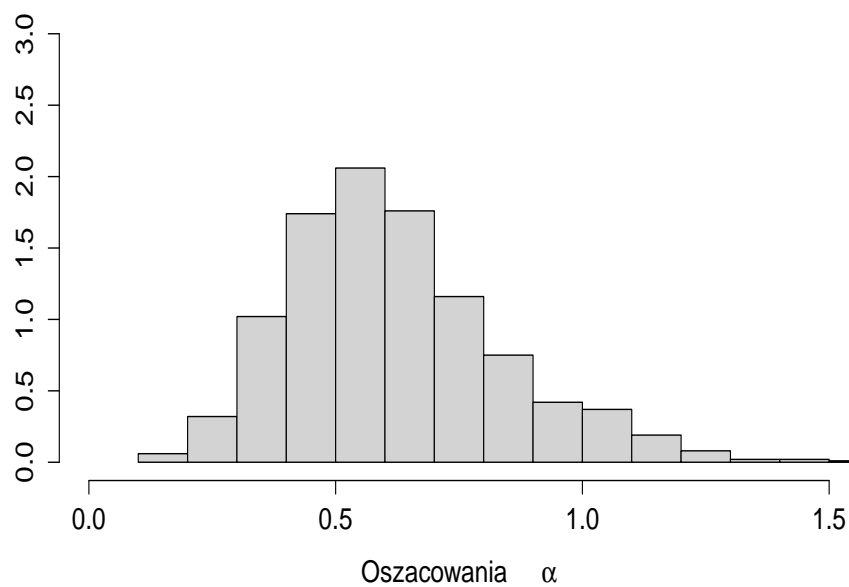
Widać z rysunków, że estymatory MLE są „średnio bliższe” estymowanej wielkości $\alpha = 0.5$ ¹. Pojęcia wprowadzone poniżej w tym rozdziale uściślają i kwantyfikują pojęcie „średniego błędu” estymacji. \diamond

¹W doświadczeniach symulacyjnych „udajemy”, że nie znamy α i obliczamy estymatory, a później porównujemy z prawdziwą wartością. To typowa metodologia.

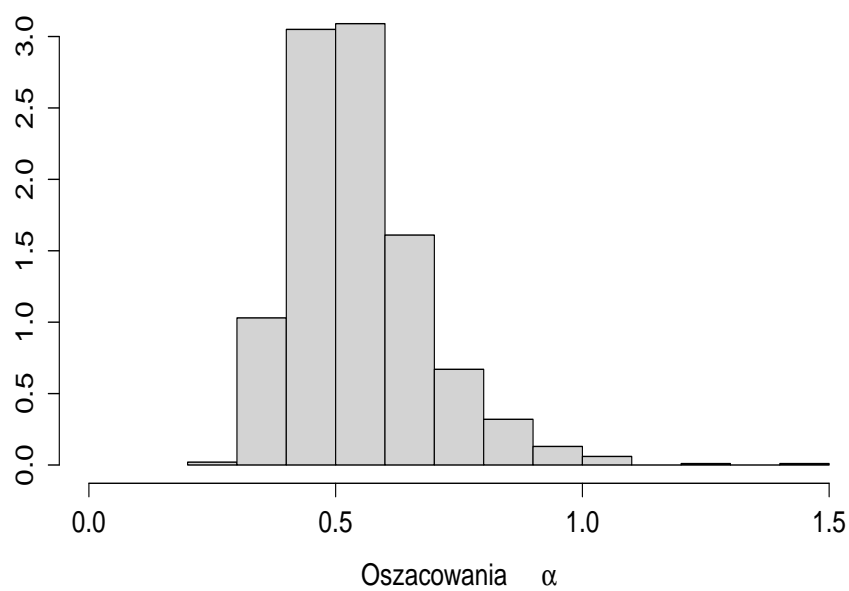


Rysunek 4.1: Rozkłady dwóch estymatorów w Przykładzie 4.0.1: wykresy pudełkowe.

Metoda Momentów



Metoda Największej Wiarygodności



Rysunek 4.2: Rozkłady dwóch estymatorów w Przykładzie 4.0.1: histogramy.

4.1 Błąd średniokwadratowy

Zajmiemy się bardziej systematycznie teorią estymacji. Tak jak zwykle, rozważamy model statystyczny, a więc losową obserwację X i rodzinę $\{\mathbb{P}_\theta; \theta \in \Theta\}$ rozkładów prawdopodobieństwa. Zadanie estymacji sformułujemy dokładniej niż w poprzednim podrozdziale. Przypuścimy, że chcemy estymować parametr θ lub liczbę $g(\theta)$, gdzie $g : \Theta \rightarrow \mathbb{R}$ jest znaną funkcją (ale argument θ jest nieznanymi)². Estymatorem $g(\theta)$ może być dowolna *statystyka* $\hat{g} : \mathcal{X} \rightarrow \mathbb{R}$. Staramy się znaleźć taki estymator, dla którego mały jest „błąd przybliżenia”. Chcemy, żeby różnica

$$\hat{g}(X) - g(\theta),$$

miała możliwie małą wartość bezwzględną dla każdego θ . Są dwie trudności. Po pierwsze, błąd jest zmienną losową. Po drugie, zależy od parametru θ . Jest sposób na ominięcie pierwszej trudności. Skoro wielkość błędu zależy od przypadku, to możemy żądać, żeby błąd był „średnio” możliwie mały. Trudność druga pozostaje: uśredniony błąd zależy od parametru θ , który jest przecież nieznanymi.

4.1.1 DEFINICJA. Niech $\hat{g}(X)$ będzie estymatorem $g(\theta)$. Funkcję

$$R(\theta) = \mathbb{E}_\theta(\hat{g}(X) - g(\theta))^2, \quad (\theta \in \Theta),$$

nazywamy **błędem średniokwadratowym** (MSE) tego estymatora.

Skrót MSE pochodzi od angielskiego terminu *Mean Square Error*. Błąd średniokwadratowy nazywa się także *funkcją ryzyka przy kwadratowej funkcji straty*. Powróćmy na chwilę do dwóch estymatorów w Przykładzie 4.0.1. Obliczony doświadczalnie MSE był równy odpowiednio 0.0658 dla MME i 0.0235 dla MLE.

4.1.2 DEFINICJA. Jeśli statystyka $\hat{g}(X)$ jest estymatorem $g(\theta)$ to

$$b(\theta) = \mathbb{E}_\theta(\hat{g}(X) - g(\theta)), \quad (\theta \in \Theta),$$

nazywamy **obciążeniem** tego estymatora. Mówimy, że estymator jest **nieobciążony**, jeśli $b(\theta) \equiv 0$, to znaczy dla każdego $\theta \in \Theta$,

$$\mathbb{E}_\theta \hat{g}(X) = g(\theta).$$

4.1.3 Stwierdzenie. Jeśli $R(\theta)$ jest błędem średniokwadratowym estymatora \hat{g} i $b(\theta)$ jest jego obciążeniem to

$$R(\theta) = \text{Var}_\theta \hat{g}(X) + b(\theta)^2.$$

²Na przykład zależy nam na dobrym przybliżeniu θ^2 albo e^θ , albo jednej ze współrzędnych w sytuacji, gdy θ jest wektorem.

Dowód. Błąd średniokwadratowy jest, z definicji, równy

$$\begin{aligned}\mathbb{E}_\theta(\hat{g}(X) - g(\theta))^2 &= \mathbb{E}_\theta(\hat{g}(X) - \mathbb{E}_\theta\hat{g}(X) + \mathbb{E}_\theta\hat{g}(X) - g(\theta))^2 \\ &= \mathbb{E}_\theta(\hat{g}(X) - \mathbb{E}_\theta\hat{g}(X))^2 + \mathbb{E}_\theta(\mathbb{E}_\theta\hat{g}(X) - g(\theta))^2 \\ &\quad + 2\mathbb{E}_\theta(\hat{g}(X) - \mathbb{E}_\theta\hat{g}(X))(\mathbb{E}_\theta\hat{g}(X) - g(\theta)) \\ &= \mathbb{E}_\theta(\hat{g}(X) - \mathbb{E}_\theta\hat{g}(X))^2 + \mathbb{E}_\theta(\mathbb{E}_\theta\hat{g}(X) - g(\theta))^2,\end{aligned}$$

bo wartość oczekiwana iloczynu mieszanego jest zerem. \square

Rozkład MSE na dwa składniki: wariancję i kwadrat obciążenia wskazuje, mówiąc nieprecyzyjnie, na dwa odmienne źródła błędu estymacji.

4.1.4 Przykład (Nieporównywalne estymatory). Rozważmy ciąg obserwacji X_1, \dots, X_n będący próbką z rozkładu $N(\mu, 1)$. Wiemy, że wariancja jest równa 1 i jedynym nieznanym parametrem jest μ . Spróbujmy porównać „bardzo rozsądny” estymator ze „zgadywaniem w ciemno”. Niech

$$\hat{\mu}_1 = \bar{X}, \quad \hat{\mu}_2 \equiv 5.$$

Mamy, jak łatwo widzieć, $R_1(\mu) \equiv 1/n$ i $R_2(\mu) = (\mu - 5)^2$. Estymator $\hat{\mu}_1$ jest nieobciążony, zaś $\hat{\mu}_2$ ma zerową wariancję. Nie można twierdzić, że jeden z tych estymatorów ma jednostajnie mniejszy MSE. Jednostajnie, to znaczy dla wszystkich wartości nieznanego parametru μ . Wykresy funkcji R_1 i R_2 się „krzyżują”. \diamond

Przykład jest wyjątkowo jaskrawy ale nieporównywalność estymatorów jest raczej powszechnym zjawiskiem w statystyce.

4.1.5 Przykład (Dwa estymatory wariancji rozkładu normalnego). Niech X_1, \dots, X_n będzie próbką z rozkładu $N(\mu, \sigma^2)$. Rozważmy dwa estymatory wariancji:

$$\begin{aligned}\hat{\sigma}_1^2 &= S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2; \\ \hat{\sigma}_2^2 &= \tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.\end{aligned}$$

Pokazaliśmy już wcześniej, przy okazji twierdzenia Fishera, że $\mathbb{E}_{\mu,\sigma} S^2 = \sigma^2$ i $\text{Var}_{\mu,\sigma} S^2 = (2/(n-1))\sigma^4$. Estymator $\hat{\sigma}_1^2$ ma więc obciążenie $b_1(\theta) = 0$ i ryzyko tego estymatora jest równe

$$R_1 = \frac{2}{n-1}\sigma^4.$$

Ponieważ $\tilde{S}^2 = S^2(n-1)/n$, więc wnioskujemy, że $\mathbb{E}_{\mu,\sigma} \tilde{S}^2 = \sigma^2(n-1)/n$ i $\text{Var}_{\mu,\sigma} \tilde{S}^2 = ((n-1)/n)^2(2/(n-1))\sigma^4 = 2((n-1)/n^2)\sigma^4$. Estymator $\hat{\sigma}_2^2$ ma więc obciążenie $b_2(\theta) = -(1/n)\sigma^2$ i ryzyko

$$R_2 = \frac{2(n-1)}{n^2}\sigma^4 + \frac{1}{n^2}\sigma^4 = \frac{2n-1}{n^2}\sigma^4.$$

Natychmiast można sprawdzić, że zawsze mamy $R_2 < R_1$. Co prawda, \tilde{S}^2 ma *ujemne obciążenie*, czyli systematycznie zaniża wielkość wariancji σ^2 , lecz z drugiej strony ma *mniejszą wariancję* niż S^2 , to znaczy mniejszy „rozrzut losowy”. \diamond

Chociaż estymator \tilde{S}^2 ma jednostajnie mniejszy MSE niż S^2 , przeważnie używa się nieobciążonego estymatora wariancji S^2 . Własność nieobciążoności jest uznawana za istotną zaletę. Rzecz jasna, dla estymatora nieobciążonego, MSE pokrywa się z wariancją:

$$R(\theta) = \text{Var}_\theta \hat{g}(X).$$

Poniższe twierdzenie pokazuje jak można „poprawiać” estymatory nieobciążone.

4.1.6 TWIERDZENIE (Rao-Blackwella). *Jeśli $\hat{g}(X)$ jest nieobciążonym estymatorem $g(\theta)$ i $T = T(X)$ jest statystyką dostateczną, to warunkowa wartość oczekiwana*

$$g^*(T) = \mathbb{E}_\theta(\hat{g}(X)|T)$$

jest nieobciążonym estymatorem $g(\theta)$ i nierówność $\text{Var}_\theta(g^(T)) \leq \text{Var}_\theta(\hat{g}(X))$ zachodzi dla każdego θ .*

Dowód. Najpierw zauważmy, że $g^*(T)$ jest *statystyką*. Wynika to z Definicji 2.3.1: ponieważ rozkład warunkowy X przy danym $T = t$ nie zależy od θ , to wartość oczekiwana definiująca $g^*(T)$ też nie zależy od θ . Nieobciążoność wynika ze wzoru na prawdopodobieństwo całkowite: $\mathbb{E}_\theta \mathbb{E}_\theta(\hat{g}(X)|T) = \mathbb{E}_\theta \hat{g}(X) = g(\theta)$. Ze znanego wzoru na „dekompozycję wariancji” wynika, że $\text{Var}_\theta \hat{g}(X) = \text{Var}_\theta \mathbb{E}_\theta(\hat{g}(X)|T) + \mathbb{E}_\theta \text{Var}_\theta(\hat{g}(X)|T) \geq \text{Var}_\theta \mathbb{E}_\theta(\hat{g}(X)|T)$. \square

4.1.7 Przykład (Rozkład wykładniczy; BUE($e^{-\theta a}$)). Niech X_1, \dots, X_n będzie próbka z rozkładu $\text{Ex}(\theta)$. Interesuje nas estymacja funkcji $g(\theta) = e^{-\theta a} = \mathbb{P}_\theta(X_1 > a)$, dla danego $a > 0$. Estymator $\hat{g}(X_1, \dots, X_n) = \mathbb{1}(X_1 > a)$ jest marny, ale nieobciążony. Poprawimy go korzystając z faktu, że $S = \sum_{i=1}^n X_i$ jest statystyką dostateczną. Zadanie 2.15 w Rozdziale 2 pokazuje, że rozkład warunkowy (X_1, \dots, X_{n-1}) przy danym $S = s$ jest jednostajny na sympleksie $\{(x_1, \dots, x_{n-1}) : x_i \geq 0, \sum_{i=1}^n x_i \leq s\}$. Wobec tego otrzymujemy $g^*(s) = \mathbb{E}_\theta(\hat{g}|S = s) = \mathbb{P}_\theta(X_1 > a|S = s) = (1 - a/s)^{n-1}$ dla $s \geq a$ i zero dla $s < a$. Ostatecznie,

$$g^*(s) = \begin{cases} \left(1 - \frac{a}{S}\right)^{n-1} & \text{dla } s \geq a; \\ 0 & \text{dla } s < a. \end{cases}$$

\diamond

Przykład 4.1.4 dobitnie pokazał, że poszukiwanie najlepszych estymatorów jest zadaniem beznadziejnym. Jeśli ograniczymy się tylko do estymatorów nieobciążonych, sytuacja się zmienia. W wielu ciekawych zadaniach istnieje estymator *najlepszy wśród nieobciążonych*, po angielsku *Best Unbiased Estimator*. Będziemy używać skrótu BUE.

4.1.8 DEFINICJA. Statystyka $g^*(X)$ jest *estymatorem nieobciążonym o minimalnej wariancji* (BUE) wielkości $g(\theta)$, jeśli

- (i) $g^*(X)$ jest estymatorem nieobciążonym $g(\theta)$;
- (ii) dla każdego nieobciążonego estymatora $\hat{g}(X)$ spełniona jest nierówność $\text{Var}_\theta g^*(X) \leq \text{Var}_\theta \hat{g}(X)$ dla każdego $\theta \in \Theta$.

Twierdzenie Rao-Blackwella pozwala konstruować estymatory nieobciążone lepsze od wyjściowych, ale nie daje gwarancji, że są one *najlepsze*, czyli BUE. Estymator skonstruowany w Przykładzie 4.1.7 w rzeczywistości jest BUE. Wynika to z następującego faktu, który przyjmujemy bez dowodu.

4.1.9 Stwierdzenie (BUE w rodzinach wykładniczych). *Założmy, że rodzina rozkładów obserwacji X jest regularną rodziną wykładniczą i $T = (T_1(X), \dots, T_k(X))$ jest naturalną statystyką dostateczną. Jeżeli $g^* = g^*(T)$ i $\mathbb{E}_\theta g^*(T) = g(\theta)$ dla każdego θ , to*

- (i) g^* jest jedynym nieobciążonym estymatorem $g(\theta)$, który jest funkcją T ;
- (ii) g^* jest BUE $g(\theta)$.

Podkreślmy, że w Stwierdzeniu 4.1.9 istotne jest założenie, że rodzina wykładnicza jest *regularna*, to znaczy zbiór „naturalnych parametrów” $\{(\psi_1(\theta), \dots, \psi_k(\theta)) : \theta \in \Theta\}$ ma niepuste wnętrze.

4.2 Informacja Fishera i nierówność Craméra-Rao

4.2.1 DEFINICJA (Informacja Fishera). *Niech X będzie zmienną losową o gęstości f_θ , zależnej od jednowymiarowego parametru $\theta \in \Theta \subseteq \mathbb{R}$. Funkcję*

$$I(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log f_\theta(X) \right)^2$$

nazywamy informacją Fishera zawartą w obserwacji X .

Jak zwykle, gęstość f_θ w powyższej definicji jest rozumiana w szerszym sensie, obejmującym również zmienne dyskretne. Mamy więc

$$(4.2.2) \quad \begin{aligned} I(\theta) &= \int \left(\frac{\partial}{\partial \theta} \log f_\theta(x) \right)^2 f_\theta(x) dx && \text{dla zmiennej ciągłej;} \\ I(\theta) &= \sum_x \left(\frac{\partial}{\partial \theta} \log f_\theta(x) \right)^2 f_\theta(x) && \text{dla zmiennej dyskretnej.} \end{aligned}$$

Jeżeli mówimy o informacji zawartej w wektorze obserwacji $X = (X_1, \dots, X_n)$ to f_θ jest gęstością łączną, zaś całkę należy rozumieć jako całkę wielokrotną.

4.2.3 Przykład (Rozkłady Poissona). Jeśli

$$f_\theta(x) = \mathbb{P}_\theta(X = x) = e^{-\theta} \frac{\theta^x}{x!}, \quad (x = 0, 1, 2, \dots)$$

to $\log f_\theta(x) = -\theta + x \log \theta - \log(x!)$ i $(\partial/\partial\theta) \log f_\theta(x) = -1 + x/\theta$. Mamy więc

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \log f_\theta(x) \right)^2 = \sum_{x=0}^{\infty} \left(\frac{x}{\theta} - 1 \right)^2 e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{\theta^2} \text{Var}_\theta(X) = \frac{1}{\theta}.$$

Ostatecznie, $I(\theta) = 1/\theta$. ◇

4.2.4 Przykład (Rozkłady wykładnicze). Jeśli

$$f_\theta(x) = \theta e^{-\theta x}, \quad (x > 0)$$

to $\log f_\theta(x) = \log \theta - \theta x$ i $(\partial/\partial\theta) \log f_\theta(x) = 1/\theta - x$. Teraz

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \log f_\theta(x) \right)^2 = \int_0^{\infty} \left(\frac{1}{\theta} - x \right)^2 \theta e^{-\theta x} dx = \text{Var}_\theta(X) = \frac{1}{\theta^2}.$$

Zatem $I(\theta) = 1/\theta^2$. ◇

4.2.5 Uwaga (Warunki regularności). Rodzina gęstości musi być „dostatecznie regularna” aby pewne kroki rachunkowe w dalszych rozumowaniach były poprawne. Ścisłe sformułowanie potrzebnych założeń przekracza ramy naszego wykładu. W dużym uproszczeniu zakładamy co następuje.

- (i) Informacja Fishera jest dobrze określona. Zakładamy, że Θ jest przedziałem otwartym, istnieje pochodna $(\partial/\partial\theta) \log f_\theta$, całka/suma we wzorze (4.2.2) jest bezwzględnie zbieżna i $0 < I(\theta) < \infty$
- (ii) Wszystkie gęstości f_θ mają ten sam „nośnik”, to znaczy zbiór $\{x \in \mathcal{X} : f_\theta(x) > 0\}$ nie zależy od θ .
- (iii) Można „przenosić pochodną przed znak całki”, czyli zamienić kolejność operacji różniczkowania $(\partial/\partial\theta)$ i całkowania $\int \cdots dx$. W przypadku zmiennych dyskretnych, całkę należy zastąpić przez sumę.

W wielu sytuacjach wystarczy się powołać na następujący fakt. **Warunki regularności 4.2.5** są spełnione **dla regularnych rodzin wykładniczych** z naturalną parametryzacją. Przypomnijmy sobie Definicję 2.4.1, Uwagę 2.4.10 i mnóstwo przykładów takich rodzin.

4.2.6 Przykład. Rodzina gęstości jednostajnych:

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

gdzie $\theta \in \Theta =]0, \infty[$ nie spełnia warunku (ii). ◇

4.2.7 Stwierdzenie. *Jeśli spełnione są warunki regularności (4.2.5) to*

$$\begin{aligned} (i) \quad & \mathbb{E}_{\theta} \frac{\partial}{\partial \theta} \log f_{\theta}(X) = 0, \\ (ii) \quad & I(\theta) = \text{Var}_{\theta} \left(\frac{\partial}{\partial \theta} \log f_{\theta}(X) \right) \\ (iii) \quad & I(\theta) = -\mathbb{E}_{\theta} \left(\frac{\partial^2}{\partial \theta^2} \log f_{\theta}(X) \right), \end{aligned}$$

Dowód. Posłużymy się uproszczoną notacją, która powinna być zrozumiała. Pominiemy argumenty θ i x , pisząc $f = f_{\theta}(x)$. “Prim” oznacza wszędzie pochodną $(\partial/\partial\theta)$, zaś całkujemy względem x . Teza (i) wynika z następującego rachunku:

$$\begin{aligned} \mathbb{E}_{\theta} \frac{\partial}{\partial \theta} \log f_{\theta}(X) &= \int (\log f)' f \\ &= \int \frac{f'}{f} f = \int (f') = \left(\int f \right)' = 1' = 0. \end{aligned}$$

Pisząc $\int (f') = (f f)'$, zamieniliśmy kolejność operacji $(\partial/\partial\theta)$ i $\int \cdots dx$. Punkt (ii) jest natychmiastową konsekwencją (i). Dowód (iii) jest podobny:

$$\begin{aligned} \mathbb{E}_{\theta} \frac{\partial^2}{\partial \theta^2} \log f_{\theta}(X) &= \int (\log f)'' f \\ &= \int \left(f' \frac{1}{f} \right)' f = \int \frac{f''}{f} f - \int f' \frac{f'}{f^2} f = 0 - \int \left(\frac{f'}{f} \right)^2 f \\ &= - \int \left((\log f)' \right)^2 f = -\mathbb{E}_{\theta} \left(\frac{\partial}{\partial \theta} \log f_{\theta}(X) \right)^2, \end{aligned}$$

Zamiana kolejności operacji $(\partial/\partial\theta)$ i $\int \cdots dx$ jest tu potrzebna aby uzasadnić równość $\int f'' = (f f)'' = 1'' = 0$. □

4.2.8 Stwierdzenie. *Niech $X = (X_1, \dots, X_n)$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie. Jeżeli $I_n(\theta)$ jest informacją Fishera zawartą w n -wymiarowej obserwacji X , zaś $I_1(\theta)$ jest informacją zawartą w pojedynczej współrzędnej X_1 , to*

$$I_n(\theta) = nI_1(\theta).$$

Dowód. Najłatwiej skorzystać ze Stwierdzenia 4.2.7 (ii):

$$\begin{aligned} I_n(\theta) &= \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log(f_\theta(X_1) \cdots f_\theta(X_n)) \right) \\ &= \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \sum_{i=1}^n \log f_\theta(X_i) \right) \\ &= \sum_{i=1}^n \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log f_\theta(X_i) \right) = nI_1(\theta). \end{aligned}$$

Oczywiście, wynika to z niezależności obserwacji i z faktu, że gęstość każdej z nich jest jednakowa. \square

4.2.9 TWIERDZENIE (Nierówność Cramèra-Rao). *Założmy, że model spełnia warunki regularności (4.2.5) i $\hat{g}(X)$ jest nieobciążonym estymatorem $g(\theta)$. Wtedy dla każdego $\theta \in \Theta$,*

$$\text{Var}_\theta \hat{g}(X) \geq \frac{(g'(\theta))^2}{I(\theta)}.$$

W szczególności, jeśli $\hat{\theta}(X)$ jest nieobciążonym estymatorem parametru θ , to

$$\text{Var}_\theta \hat{\theta}(X) \geq 1/I(\theta).$$

Dowód. Zastosujemy nierówność Schwarzera do zmiennych losowych $\hat{g}(X)$ i $(\partial/\partial\theta) \log f_\theta(X)$:

$$\begin{aligned} \text{Var}_\theta \hat{g}(X) I(\theta) &= \text{Var}_\theta \hat{g}(X) \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log f_\theta(X) \right) \\ &\geq \left(\text{Cov}_\theta \left(\hat{g}(X), \frac{\partial}{\partial \theta} \log f_\theta(X) \right) \right)^2 \\ &= \left(\mathbb{E}_\theta \left(\hat{g}(X) \frac{\partial}{\partial \theta} \log f_\theta(X) \right) \right)^2 \\ &= \left(\int \hat{g}(x) \left(\frac{\partial}{\partial \theta} \log f_\theta(x) \right) f_\theta(x) dx \right)^2 = (*) \end{aligned}$$

Skorzystaliśmy kolejno: ze Stwierdzenia 4.2.7 (ii), nierówności Schwarzera i Stwierdzenia 4.2.7 (i).

Dalej,

$$\begin{aligned}
 (*) &= \left(\int \hat{g}(x) \left(\frac{\partial}{\partial \theta} \log f_{\theta}(x) \right) f_{\theta}(x) dx \right)^2 \\
 &= \left(\int \hat{g}(x) \frac{(\partial/\partial \theta) f_{\theta}(x)}{f_{\theta}(x)} f_{\theta}(x) dx \right)^2 \\
 &= \left(\int \frac{\partial}{\partial \theta} \hat{g}(x) f_{\theta}(x) dx \right)^2 = \left(\frac{\partial}{\partial \theta} \int \hat{g}(x) f_{\theta}(x) dx \right)^2 \\
 &= \left(\frac{\partial}{\partial \theta} \mathbb{E}_{\theta} \hat{g}(X) \right)^2 = \left(\frac{\partial}{\partial \theta} g(\theta) \right)^2.
 \end{aligned}$$

Pierwsza równość wynika ze wzoru (4.2.2), druga z „przeniesienia pochodnej przed znak całki” na mocy warunku regularności (4.2.5) (iii). Wreszcie, ostatnia równość wynika z założenia o nieobciążoności estymatora. \square

Nierówność Craméra-Rao stwierdza, że funkcja ryzyka każdego nieobciążonego estymatora nie może nigdzie być mniejsza od pewnej ustalonej funkcji. Istnieje granica jakości nieprzekraczalna dla wszystkich estymatorów *nieobciążonych*. Jeśli ryzyko jakiegoś estymatora nieobciążonego jest *równe* dolnemu oszacowaniu Craméra-Rao, to jest oczywiste, że jest to estymator najlepszy wśród nieobciążonych, a więc BUE.

4.2.10 Wniosek. *Jeśli $\hat{g}(X_1, \dots, X_n)$ jest nieobciążonym estymatorem $g(\theta)$, obliczonym na podstawie niezależnej próbki losowej X_1, \dots, X_n to*

$$\text{Var}_{\theta} \hat{g}(X_1, \dots, X_n) \geq \frac{(g'(\theta))^2}{nI_1(\theta)},$$

pod warunkiem, że spełnione są założenia regularności (4.2.5).

Jakość nieobciążonego estymatora parametru θ można ocenić, porównując wariancję tego estymatora z dolnym ograniczeniem Craméra-Rao. Wielkość $1/I_n(\theta)$ jest po prostu wygodnym punktem odniesienia. Przyjmijmy następującą definicję:

Efektywność nieobciążonego estymatora \hat{g} wielkości $g(\theta)$ określamy jako

$$\text{ef}(\hat{g}) = \frac{(g'(\theta))^2}{\text{Var}_{\theta}(\hat{g})I(\theta)}.$$

Użyteczne jest też pojęcie efektywności względnej dwóch estymatorów. Jeśli \hat{g}_1 i \hat{g}_2 są nieobciążone to

$$\text{ef}(\hat{g}_1, \hat{g}_2) = \frac{\text{Var}_{\theta}(\hat{g}_2)}{\text{Var}_{\theta}(\hat{g}_1)}$$

nazywamy *efektywnością* estymatora \hat{g}_1 względem \hat{g}_2 . Oczywiście, $\text{ef}(\hat{g}_1, \hat{g}_2) = \text{ef}(\hat{g}_1)/\text{ef}(\hat{g}_2)$.

Jeśli spełniona jest nierówność Cramèra-Rao, to $ef(\hat{g}) \leq 1$ dla każdego estymatora nieobciążonego. Jeśli $\hat{g} = \text{BUE}(g)$ to może się zdarzyć, że $ef(\hat{g}) = 1$ ale też może być $ef(\hat{g}) < 1$. Zwróćmy jeszcze uwagę na to, że $ef(\hat{g})$ jest na ogół funkcją parametru θ (a nie liczbą).

4.2.11 Przykład (Rozkład Poissona; $\text{BUE}(\theta)$). Niech X_1, \dots, X_n będzie próbką z rozkładu $\text{Poiss}(\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Średnia z próbki \bar{X} jest $\text{BUE}(\theta)$. Istotnie, wiemy, że $I_n(\theta) = nI(\theta) = n/\theta$. Estymator \bar{X} jest nieobciążony i jego wariancja jest równa dolnemu ograniczeniu Cramèra-Rao, bo

$$\text{Var}_\theta(\bar{X}) = \frac{\theta}{n} = \frac{1}{I_n(\theta)}.$$

Każdy inny estymator nieobciążony musi więc mieć wariancję nie mniejszą, niż wariancja \bar{X} . Udowodniliśmy, że \bar{X} jest BUE bez odwołania się do Stwierdzenia 4.1.9. \diamond

4.2.12 Przykład (Rozkład wykładniczy; $\text{BUE}(1/\theta)$). Niech X_1, \dots, X_n będzie próbką z rozkładu $\text{Ex}(\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Średnia z próbki \bar{X} jest $\text{BUE}(1/\theta)$, bo

$$\text{Var}_\theta(\bar{X}) = \frac{1}{n\theta^2} = \frac{((1/\theta)')^2}{I_n(\theta)}.$$

Stosujemy tu Wniosek 4.2.10 do funkcji $g(\theta) = 1/\theta$. Wiemy, że $I_n(\theta) = nI_1(\theta) = n/\theta^2$. Średnia z próbki jest estymatorem efektywnym, to znaczy ma efektywność 1. \diamond

Twierdzenie 4.2.9 nie daje uniwersalnej metody sprawdzania, że estymator jest BUE . Pokazuje to następujący przykład.

4.2.13 Przykład (Rozkład wykładniczy; $\text{BUE}(\theta)$). Tak jak poprzednio, niech X_1, \dots, X_n będzie próbką z rozkładu $\text{Ex}(\theta)$. Naturalnym estymatorem parametru θ jest $1/\bar{X}$, ale jest to estymator *obciążony*. W istocie, zmienna losowa $Y = \sum X_i$ ma rozkład $\text{Gamma}(n, \theta)$, więc

$$\mathbb{E}_\theta \frac{1}{Y} = \int_0^\infty \frac{1}{y} \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy = \int_0^\infty \frac{\theta^n}{(n-1)!} y^{n-2} e^{-\theta y} dy = \frac{\theta}{n-1}.$$

Widzimy, że $\mathbb{E}_\theta(1/\bar{X}) = n/(n-1)$. Łatwo zmodyfikować nasz wyjściowy estymator tak, żeby „usunąć obciążenie”:

$$\hat{\theta} = \frac{n-1}{Y} = \frac{n-1}{n\bar{X}}$$

jest estymatorem nieobciążonym. Co więcej, $\hat{\theta} = \text{BUE}(\theta)$, co wynika ze Stwierdzenia 4.1.9. Zbadajmy, jaka jest wariancja $\hat{\theta}$. Podobnie jak poprzednio, obliczamy

$$\mathbb{E}_\theta \frac{1}{Y^2} = \int_0^\infty \frac{1}{y^2} \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} dy = \frac{\theta^2}{(n-1)(n-2)}.$$

Stąd $\mathbb{E}_\theta(\hat{\theta}^2) = \theta^2(n-1)/(n-2)$ i $\text{Var}_\theta \hat{\theta} = \theta^2/(n-2)$. Dla rodziny rozkładów wykładniczych znamy informację Fishera, $I_1(\theta) = n/\theta^2$. Możemy teraz porównać wariancję z ograniczeniem wynikającym z nierówności Cramèra-Rao.

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

Nierówność jest *ostra*. BUE nie osiąga dolnego ograniczenia Cramèra-Rao. Mamy $\text{ef}(\hat{\theta}) = (n-2)/n < 1$. \diamond

4.3 Zadania

4.1. Zmienne losowe X_1, \dots, 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 nieznaną 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, $\mathbb{E}_{\mu, \sigma} (\hat{\mu} - \mu)^2$.

4.2. 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 $\tilde{p} = (X + 1)/(n + 2)$.

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

4.3. 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 (jest to ten sam model, co w Zadaniu 2.12 w Rozdziale 3).

- (a) Znaleźć estymator *największej wiarygodności* parametru θ .
- (b) Obliczyć *błąd średniokwadratowy* tego estymatora.

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

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

4.5. Obserwujemy zmienną losową S o rozkładzie $\text{Bin}(n, p)$, czyli liczbę sukcesów w schemacie Bernoulliego z nieznanym prawdopodobieństwem sukcesu p . Interesuje nas estymacja wielkości $g(p) = p^2$. Znaleźć estymator *nieobciążony* $\hat{g}(S)$.

Wskazówka: Można na przykład obliczyć $\mathbb{E}_{\theta}(S^2)$ i odpowiednio „poprawić” estymator $(S/n)^2$. Można też inaczej.

4.6. Niech X_1, \dots, X_n będzie próbką z rozkładu $N(\mu, \sigma^2)$, z nieznanymi parametrami μ i σ . Dobrać stałą c tak, żeby $\hat{\sigma} = cS$ był estymatorem nieobciążonym odchylenia standardowego σ . Jak zwykle, S^2 jest nieobciążonym estymatorem wariancji (Przykład 4.1.5), zaś $S = \sqrt{S^2}$.

4.7. Niech X_1, \dots, X_n będzie próbką losową z rozkładu normalnego $N(\mu, \sigma^2)$ z nieznanymi parametrami μ i σ^2 . Rozważmy następujący estymator parametru σ^2 :

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

gdzie $\mu \in \mathbb{R}$ jest nieznanym parametrem, $c = c_n > 0$ jest ustaloną liczbą.

- Obliczyć obciążenie tego estymatora.
- Obliczyć błąd średniokwadratowy tego estymatora.
- Znaleźć takie $c > 0$, dla którego błąd średniokwadratowy jest najmniejszy (jednostajnie ze względu na parametry).

4.8. Niech X_1, \dots, X_n będzie próbką z rozkładu jednostajnego $U(0, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem.

- Dobrać stałą c tak, żeby $\hat{\theta} = cX_{n:n}$ był estymatorem *nieobciążonym*.
- Obliczyć $\text{Var}_\theta \hat{\theta}$.
- Wyznaczyć estymator *metody momentów*, $\hat{\theta}_{\text{MME}}$ i pokazać jego *nieobciążoność*.
- Obliczyć $\text{Var}_\theta \hat{\theta}_{\text{MME}}$.
- Podać $\text{ef}(\hat{\theta}_{\text{MME}}, \hat{\theta})$.

Wskazówka: Skorzystać z wyników Zadań 1.7 i 1.8 z Rozdziału 1.

4.9. Niech X_1, \dots, X_n będzie próbką z rozkładu o gęstości

$$f_\theta(x) = \begin{cases} \frac{\theta}{(1-x)^2} \cdot e^{-\theta x/(1-x)} & \text{dla } 0 < x < 1 \\ 0 & \text{w przeciwnym przypadku,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem.

- Podać wzór na *estymator największej wiarygodności* $\hat{\theta}$ parametru θ .
- Obliczyć wariancję statystyki $1/\hat{\theta}$.

Wskazówka: Najłatwiej jest skorzystać z następującego faktu: można sprawdzić, że $1/\hat{\theta}$ jest estymatorem *nieobciążonym i efektywnym* wielkości $1/\theta$, a więc ma wariancję równą dolnemu ograniczeniu Craméra-Rao.

4.10. X_1, X_2, \dots, X_{100} jest próbką z rozkładu normalnego $N(\mu, \sigma^2)$. Ponieważ dane, składające się ze 100 liczb, były zapisane w dziesięciu kolumnach, sekretarka policzyła średnie

$$\bar{X}_1 = \frac{1}{10} \sum_{i=1}^{10} X_i, \quad \bar{X}_2 = \frac{1}{10} \sum_{i=11}^{20} X_i, \quad \dots, \quad \bar{X}_{10} = \frac{1}{10} \sum_{i=91}^{100} X_i$$

i zapisała liczby $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_{10}$ na osobnej karteczce. Niestety, oryginalne dane zostały zgubione i pozostała tylko ta karteczka. Należy zbudować estymatory parametrów μ i σ^2 na podstawie dziesięciu średnich. Oczywiście, $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i$.

- (a) Obliczyć $\text{Var} \bar{X}$.
- (b) Podać taką liczbę b , żeby statystyka $b \sum_{i=1}^{10} (\bar{X}_i - \bar{X})^2$ była nieobciążonym estymatorem parametru σ^2 .

4.11. Niech X_1, \dots, X_n będzie próbką z rozkładu normalnego $N(\mu, \sigma^2)$ z nieznanymi parametrami μ i $\sigma > 0$. Aby otrzymać nieobciążony estymator wielkości μ^2 , będący funkcją statystyki dostatecznej (\bar{X}, S^2) , wykonać można poniższe kroki.

- (a) Obliczyć obciążenie „naturalnego” estymatora \bar{X}^2 , czyli $\mathbb{E}_{\mu, \sigma} \bar{X}^2 - \mu^2$.
- (b) Rozważyć estymator postaci $\bar{X}^2 - c \sum (X_i - \bar{X})^2$. Znaleźć stałą c tak, żeby ten estymator był nieobciążony.

Wskazówka: $\text{Var} X = \mathbb{E} X^2 - (\mathbb{E} X)^2$.

4.12. Niech X_1, \dots, X_n będzie próbką z rozkładu o gęstości

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

(wykładniczego przesuniętego), gdzie parametry μ i $\lambda > 0$ nie są znane.

- (a) Korzystając ze statystyki

$$T_1(X_1, \dots, X_n) = \sum_{i=1}^n (X_i - X_{1:n}),$$

skonstruować estymator nieobciążony dla $1/\lambda$.

- (b) Korzystając ze statystyki $T_1(X_1, \dots, X_n)$ i statystyki $T_2(X_1, \dots, X_n) = X_{1:n}$, skonstruować estymator nieobciążony dla parametru μ .

Wskazówka: Skorzystać z wyników Zadania 1.10 z Rozdziału 1.

4.13. Niech X_1, \dots, X_n będzie próbką z rozkładu normalnego $N(\mu, \sigma^2)$.

(a) Dobrać stałą k tak, aby estymator

$$T(X_1, \dots, X_n) = k \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$$

był nieobciążonym estymatorem parametru σ^2 .

(b) Porównać wariancję estymatora $T(X_1, \dots, X_n)$ z wariancją „zwykłego” nieobciążonego estymatora S^2 . Który z estymatorów gwarantuje większą precyzję oszacowania?

4.14. Niech X_1, \dots, X_n będzie próbką z rozkładu Poiss(θ). Chcemy estymować wielkość $g(\theta) = e^{-\theta}$ (jeśli nasz model opisuje liczbę szkód, to interpretujemy tę wielkość jako prawdopodobieństwo braku szkód, bo $g(\theta) = \mathbb{P}_\theta(X_1 = 0)$). Skonstruować BUE($e^{-\theta}$).

Wskazówka: Zastosować metodę z Przykładu 4.1.7, wykorzystując wyniki Zadania 2.14 z Rozdziału 2. Pokazać, że BUE($e^{-\theta}$) jest dany wzorem $\hat{g} = (1 - 1/n)^S$.

4.15. (Ciąg dalszy) Próbka z rozkładu Poiss(θ), estymacja $g(\theta) = e^{-\theta}$

(a) Pokazać przy pomocy bezpośredniego rachunku, że $\mathbb{E}_\theta(1 - 1/n)^S = e^{-\theta}$.

(b) Obliczyć $\text{Var}_\theta(1 - 1/n)^S$.

Wskazówka: Zauważyć, że znany jest wzór na funkcję tworzącą momenty rozkładu Poissona $\mathbb{E}_\theta \exp(tS) = \exp[n\theta(e^t - 1)]$ dla dowolnego t .

4.16. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi takimi, że $\mathbb{E}X_i = \mu$ i $\text{Var}X_i = \sigma_i^2$. Zakładamy, że znamy wariancje $\sigma_1^2, \dots, \sigma_n^2$ a nieznaną wartość oczekiwaną μ jest jednakowa dla wszystkich zmiennych. Rozpatrujemy linowe estymatory wartości oczekiwanej, postaci

$$\hat{\mu} = a + \sum_{i=1}^n b_i X_i.$$

(a) Znaleźć warunki jakie muszą spełniać współczynniki a i b_1, \dots, b_n , aby estymator $\hat{\mu}$ był nieobciążony.

(b) Znaleźć a i b_1, \dots, b_n , takie aby $\hat{\mu}$ miał najmniejszą wariancję pośród nieobciążonych estymatorów liniowych.

Wskazówka: Skorzystać z metody mnożników Lagrange’a.

Uwaga: Wynik jest podobny jak w Zadaniu 3.11 z Rozdziału 3. Teraz jednak nie zakładamy, że X_i mają rozkłady normalne. Otrzymany w tym zadaniu estymator ma piękny angielski skrót: BLUE.

4.17. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi z rozkładu Raileigh'a o gęstości

$$f_\sigma(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} & \text{dla } x > 0; \\ 0 & \text{dla } x \leq 0, \end{cases}$$

gdzie $\sigma > 0$ jest nieznanym parametrem. Wiadomo, że $\mathbb{E}_\sigma X_i = \sigma\sqrt{\pi/2}$ oraz że k -ty moment w rozkładzie Raileigh'a wyraża się wzorem $m_k = \sigma^k 2^{k/2} \Gamma(1 + k/2)$.

- (a) Dobrać stałą c tak, aby statystyka $c\bar{X}$ była estymatorem nieobciążonym parametru σ .
- (b) Obliczyć wariancję tego estymatora nieobciążonego i porównać z ograniczeniem dolnym Craméra-Rao.

Rozdział 5

Asymptotyczne własności estymatorów

Zajmiemy się teraz sytuacją, kiedy rozmiar próbki X_1, \dots, X_n jest duży. Własności asymptotyczne są to, z matematycznego punktu widzenia, twierdzenia graniczne, w których n dąży do nieskończoności. W praktyce, te twierdzenia opisują *w przybliżeniu* zachowanie estymatorów dla „dostatecznie dużych” próbek. Niestety, teoria asymptotyczna (przynajmniej w najprostszej swojej postaci) nie dostarcza informacji o tym, *jak duża* powinna być próbka, żeby przybliżenie było *dostatecznie dobre*.

Wyobraźmy sobie (potencjalnie) nieskończony ciąg obserwacji X_1, \dots, X_n, \dots . Ograniczymy się do rozpatrzenia sytuacji, kiedy obserwacje są *niezależnymi zmiennymi losowymi o jednakowym rozkładzie* (zależnym, jak zwykle, od nieznanego parametru). Przez estymator rozumiemy funkcję zależącą od początkowych n obserwacji X_1, \dots, X_n . Faktycznie, mówiąc o „estymatorze” będziemy mieli na myśli *ciąg* estymatorów $\hat{g}(X_1, \dots, X_n)$, gdzie $n = 1, 2, \dots$. Niekiedy będziemy pisali $\hat{g}_n = \hat{g}(X_1, \dots, X_n)$, żeby uwidocznić rozmiar próbki. Na przykład,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i; \quad \tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2; \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

5.1 Zgodność

5.1.1 DEFINICJA. Estymator $\hat{g}(X_1, \dots, X_n)$ wielkości $g(\theta)$ jest **zgodny**, jeśli dla każdego $\theta \in \Theta$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{g}(X_1, \dots, X_n) - g(\theta)| \leq \varepsilon) = 1,$$

dla każdego $\varepsilon > 0$. Estymator jest **mocno zgodny**, jeśli

$$\mathbb{P}_\theta \left(\lim_{n \rightarrow \infty} \hat{g}(X_1, \dots, X_n) = g(\theta) \right) = 1.$$

Zgodność (mocna zgodność) znaczy tyle, że

$$\hat{g}(X_1, \dots, X_n) \rightarrow g(\theta), \quad (n \rightarrow \infty)$$

według prawdopodobieństwa (prawie na pewno). Innymi słowy, estymator jest zgodny, jeśli zmierza do estymowanej wielkości przy nieograniczonym powiększaniu próbki. W świetle twierdzenia Gliwienki-Cantelliego wydaje się jasne, że od każdego „rozsądnego” estymatora powinniśmy oczekiwać mocnej zgodności. W istocie, niemal wszystkie rozważane przez nas estymatory są mocno zgodne (wyjątkami są specjalnie dobrane „głupie” estymatory konstruowane dla zilustrowania pewnych osobliwości).

5.1.2 Przykład (Momenty z próbki). Średnia z próbki, \bar{X}_n , jest mocno zgodnym estymatorem wartości oczekiwanej $\mu(\theta) = \mathbb{E}_\theta X_1$ dla każdej rodziny rozkładów prawdopodobieństwa takiej, że $\mu(\theta)$ jest dobrze określone. Zbieżność $\bar{X}_n \rightarrow_{\text{p.n.}} \mu(\theta)$ wynika po prostu z Mocnego Prawa Wielkich Liczb (MPWL). Podobnie, \tilde{S}_n^2 i S_n^2 są mocno zgodnymi estymatorami wariancji $\sigma^2(\theta) = \text{Var}_\theta X_1$. Wystarczy zauważyć, że $\tilde{S}_n^2 = \sum_1^n X_i^2/n - \bar{X}_n^2$ i $S_n^2 = \tilde{S}_n^2 n/(n-1)$ i znowu powołać się na MPWL. \diamond

5.1.3 Przykład (Rozkłady Pareto). Niech X_1, X_2, \dots , będzie próbką z rozkładu Pareto(α, λ). Z definicji, jest to rozkład o dystrybuancie $F_{\alpha, \lambda}(x) = 1 - \lambda^\alpha/(x - \lambda)^\alpha$ dla $x > 0$ i gęstości

$$f_{\alpha, \lambda}(x) = \begin{cases} \alpha \lambda^\alpha / (x - \lambda)^{\alpha+1} & \text{dla } x > 0; \\ 0 & \text{wpp.} \end{cases}$$

Jeśli $\alpha > 1$, to wartość oczekiwana rozkładu Pareto jest równa $\lambda/(\alpha - 1)$. Jeśli $\alpha > 2$, to wariancja jest równa $\alpha \lambda^2 / ((\alpha - 1)^2 (\alpha - 2))$. To łatwo można obliczyć metodą całkowania przez części. Dla $\alpha \leq 1$, wartość oczekiwana jest nieskończona. Dla $\alpha \leq 2$, wariancja nie istnieje.

Metoda momentów prowadzi do następujących estymatorów parametrów α i λ :

$$\hat{\alpha}_n = \frac{2\tilde{S}_n^2}{\tilde{S}_n^2 - \bar{X}_n^2}, \quad \hat{\lambda}_n = (\hat{\alpha}_n - 1)\bar{X}_n.$$

Jeśli ograniczymy się do rozkładów Pareto mających skończoną wariancję, czyli rozpatrzymy rodzinę $\{\text{Pareto}(\alpha, \lambda), \alpha > 2, \lambda > 0\}$, to estymatory są mocno zgodne: $\hat{\alpha}_n \rightarrow_{\text{p.n.}} \alpha$ i $\hat{\lambda}_n \rightarrow_{\text{p.n.}} \lambda$. Wynika to z MPWL, podobnie jak w poprzednim przykładzie. Dla $\alpha \leq 2$ sytuacja się zmienia: wariancja rozkładu Pareto nie istnieje i nie należy oczekiwać, że metoda momentów da przyzwoite wyniki. W istocie, dla każdego $\alpha \leq 2$ mamy $\hat{\alpha}_n \rightarrow_{\text{p.n.}} 2$ przy $n \rightarrow \infty$. (łatwo to uzasadnić dla $1 < \alpha \leq 2$, bo wtedy $\tilde{S}_n^2 \rightarrow_{\text{p.n.}} \infty$ i $\bar{X}_n \rightarrow_{\text{p.n.}} \lambda/(\alpha - 1)$). Podsumowując, estymatory otrzymane metodą momentów *nie są zgodne*, jeśli rozpatrujemy rodzinę *wszystkich* rozkładów Pareto $\{\text{Pareto}(\alpha, \lambda), \alpha > 0, \lambda > 0\}$. \diamond

Estymatory największej wiarygodności są *zazwyczaj* mocno zgodne. Dowody (mocnej) zgodności wykorzystują (mocne) PWL, ale szczegóły mogą być żmudne. Nie będziemy się przy tym zatrzymywać, bo zgodność (nawet w mocnym sensie) nie jest specjalnie satysfakcjonującą własnością estymatora. Jest zaledwie minimalnym żądaniem, które powinien spełniać każdy „przyzwoity” estymator.

5.2 Asymptotyczna normalność

5.2.1 DEFINICJA. Mówimy, że estymator $\hat{g}(X_1, \dots, X_n)$ wielkości $g(\theta)$ jest **asymptotycznie normalny**, jeśli dla każdego $\theta \in \Theta$ istnieje funkcja $\sigma^2(\theta)$, zwana asymptotyczną wariancją, taka że

$$\sqrt{n}(\hat{g}(X_1, \dots, X_n) - g(\theta)) \rightarrow_d N(0, \sigma^2(\theta)), \quad (n \rightarrow \infty).$$

Estymator jest asymptotycznie normalny, jeśli

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(\frac{\sqrt{n}}{\sigma(\theta)} (\hat{g}(X_1, \dots, X_n) - g(\theta)) \leq a \right) = \Phi(a).$$

Mówiąc jeszcze inaczej, rozkład prawdopodobieństwa statystyki $\hat{g}(X_1, \dots, X_n)$ jest dla dużych n zbliżony do rozkładu

$$N \left(g(\theta), \frac{\sigma^2(\theta)}{n} \right).$$

Jeśli estymator jest asymptotycznie normalny, to jest zgodny, choć nie musi być *mocno* zgodny. Zazwyczaj dla estymatora asymptotycznie normalnego mamy

$$(5.2.2) \quad \begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}_\theta \hat{g}(X_1, \dots, X_n) &= g(\theta), \\ \lim_{n \rightarrow \infty} n \text{Var}_\theta \hat{g}(X_1, \dots, X_n) &= \sigma^2(\theta), \end{aligned}$$

ale te relacje *nie wynikają* z Definicji 5.2.1. Chodzi o to, że zbieżność według rozkładu nie pociąga za sobą zbieżności wartości oczekiwanych ani wariancji. Kontrprzykłady mają jednak dość “patologiczny” charakter i nie będziemy ich przytaczać.

Dlaczego oczekujemy, że wariancja aproksymującego rozkładu normalnego jest *odwrotnie proporcjonalna* do rozmiaru próbki n ? Najprostszą statystyką jest średnia z próbki. Centralne Twierdzenie Graniczne (CTG) mówi, że

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow_d N(0, \sigma^2),$$

gdzie μ i σ^2 oznaczają wartość oczekiwaną i wariancję pojedynczej obserwacji X_i . Asymptotyczna wariancja σ^2/n pokrywa się w tu po prostu z wariancją statystyki \bar{X} . Dla wielu innych statystyk, asymptotyczna normalność wraz z odpowiednią postacią asymptotycznej wariancji daje się też wywnioskować z CTG. Pomocny jest następujący lemat.

5.2.3 Lemat (Metoda delta). *Jeżeli dla ciągu zmiennych losowych T_n mamy $\sqrt{n}(T_n - \mu) \rightarrow_d N(0, \sigma^2)$ przy $n \rightarrow \infty$ i $h : \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją różniczkowalną w punkcie μ , to*

$$\sqrt{n}(h(T_n) - h(\mu)) \rightarrow_d N(0, \sigma^2(h'(\mu))^2).$$

Dla uproszczenia odstąpiliśmy chwilowo od notacji, jawnie wskazującej na zależność rozkładów prawdopodobieństwa od nieznanego parametru. W zastosowaniach zarówno μ jak i σ^2 będą funkcjami parametru θ .

Dowód. Rozwińmy funkcję h zgodnie ze wzorem Taylora, uwzględniając tylko wyrazy rzędu zerowego i pierwszego. Resztę oznaczmy przez $r(t)$:

$$h(t) = h(\mu) + h'(\mu)(t - \mu) + r(t), \quad \text{gdzie } \frac{r(t)}{t - \mu} \rightarrow 0, \quad (t \rightarrow \mu).$$

Wstawmy teraz zmienną losową T_n w miejsce t i pomnóżmy równanie przez \sqrt{n} . Dostajemy

$$(*) \quad \sqrt{n}(h(T_n) - h(\mu)) = h'(\mu)\sqrt{n}(T_n - \mu) + \sqrt{nr}(T_n).$$

Bezpośrednio z założenia wynika, że

$$(**) \quad h'(\mu)\sqrt{n}(T_n - \mu) \rightarrow_d N(0, \sigma^2(h'(\mu))^2).$$

Pozostaje zająć się wyrazem związanym z resztą we wzorze (*). Pokażemy, że

$$(***) \quad \sqrt{nr}(T_n) \rightarrow_P 0, \quad (n \rightarrow \infty).$$

Istotnie,

$$\sqrt{nr}(T_n) = \frac{r(T_n)}{T_n - \mu} \sqrt{n}(T_n - \mu).$$

Ponieważ $T_n \rightarrow_P \mu$, więc $r(T_n)/(T_n - \mu) \rightarrow_P 0$. Z kolei $\sqrt{n}(T_n - \mu) \rightarrow_d N(0, \sigma^2)$. Na mocy Lematu Śluckiego, iloczyn zmiennych losowych zbieżnych według prawdopodobieństwa do zera i zmiennych zbieżnych według rozkładu dąży według prawdopodobieństwa do zera. Ostatecznie, teza lematu wynika z (*), (**) i (***) \square

5.2.4 Przykład (Model Hardy'ego-Weinberga). Wróćmy do Przykładu 3.1.2. Obserwujemy trójkę zmiennych losowych (N_1, N_2, N_3) o rozkładzie wielomianowym $\text{Mult}(n, p_1, p_2, p_3)$, gdzie $p_1 = \theta^2$, $p_2 = 2\theta(1 - \theta)$ i $p_3 = (1 - \theta)^2$. Niech

$$\hat{\theta} = \sqrt{\frac{N_1}{n}}.$$

Asymptotyczna normalność estymatora $\hat{\theta}$ wynika z zastosowania metody delta do funkcji $h(x) = \sqrt{x}$, gdyż $\sqrt{n}(N_1/n - p_1) \rightarrow_d N(0, p_1(1 - p_1))$ na mocy CTG. Ponieważ $(h'(p_1))^2 = 1/(4p_1)$, więc

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N\left(0, \frac{1}{4}(1 - \theta^2)\right).$$

\diamond

Estymatory największej wiarygodności

5.2.5 TWIERDZENIE (Asymptotyczna normalność MLE). *Rozważmy rodzinę gęstości $\{f_\theta : \theta \in \Theta\}$ spełniającą warunki regularności (4.2.5). Niech X_1, \dots, X_n, \dots będzie próbką z rozkładu o gęstości f_{θ_0} . Zakładamy, że $\Theta \subseteq \mathbb{R}$ i θ_0 jest punktem wewnętrznym zbioru Θ . Niech $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ będzie ciągiem estymatorów takich, że $\sum_1^n (\partial/\partial\theta) \log f_{\hat{\theta}_n}(X_i) = 0$ i $\hat{\theta}_n \rightarrow_P \theta_0$.*

Wtedy

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \rightarrow_d N(0, 1/I_1(\theta_0)), \quad (n \rightarrow \infty).$$

Szkic dowodu. Jak zwykle, „prim” oznaczać będzie różniczkowanie względem parametru θ . Niech $\ell'(\theta) = \sum_1^n (\partial/\partial\theta) \log f_\theta(X_i)$. Napiszmy rozwinięcie tej funkcji wynikające z wzoru Taylora: $\ell'(\theta) = \ell'(\theta_0) + \ell''(\theta_0)(\theta - \theta_0) + r(\theta)$. Biorąc pod uwagę założenie (i), możemy napisać

$$0 = \ell'(\hat{\theta}_n) = \ell'(\theta_0) + \ell''(\theta_0)(\hat{\theta}_n - \theta_0) + r(\hat{\theta}_n).$$

Ponieważ $\hat{\theta}_n \rightarrow \theta_0$, więc reszta $r(\hat{\theta}_n)$ jest mała w porównaniu z pierwszymi wyrazami rozwinięcia. Zaniedbując resztę otrzymujemy *przybliżoną* równość $\hat{\theta}_n - \theta_0 \simeq -\ell'(\theta_0)/\ell''(\theta_0)$. Innymi słowy,

$$(*) \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \simeq -\frac{\ell'(\theta_0)/\sqrt{n}}{\ell''(\theta_0)/n}.$$

Zarówno $\ell'(\theta_0)$, jak i $\ell''(\theta_0)$ są sumami niezależnych zmiennych losowych o jednakowym rozkładzie. Zastosujmy CTG do licznika we wzorze (*):

$$\ell'(\theta_0)/\sqrt{n} = \frac{\sum_1^n (\partial/\partial\theta) \log f_{\theta_0}(X_i)}{\sqrt{n}} \rightarrow_d N(0, I_1(\theta_0)),$$

bo zmienne losowe $(\partial/\partial\theta) \log f_{\theta_0}(X_i)$ są niezależne, mają wartość oczekiwaną zero i wariancję $I_1(\theta_0)$ na mocy Stwierdzenia 4.2.7 (i-ii). Z kolei PWL, zastosowane do mianownika we wzorze (*), daje

$$\ell''(\theta_0)/n = \frac{1}{n} \sum_1^n (\partial^2/\partial\theta^2) \log f_{\theta_0}(X_i) \rightarrow_P -I_1(\theta_0),$$

na mocy Stwierdzenia 4.2.7 (iii). Prawa strona (*) zachowuje się w granicy tak, jak zmienna o rozkładzie $N(0, I_1(\theta_0))$ podzielona przez stałą $I_1(\theta_0)$. Ma więc graniczny rozkład $N(0, I_1(\theta_0)^{-1})$.

Idea dowodu jest prosta. Nicco żmudne jest tylko uzasadnienie tego, że rzeczywiście można zaniedbać resztę we wzorze Taylora. \square

Twierdzenie 5.2.5 dotyczy estymatorów największej wiarygodności takich, że $\ell'(\hat{\theta}_n) = 0$. Nie obejmuje MLE(θ) w Przykładzie 3.2.8 (dla rodziny rozkładów jednostajnych). Zadanie 1.9 w Rozdziale 1 pokazuje, że ten estymator nie jest asymptotycznie normalny. Z drugiej strony, MLE w Przykładzie 3.2.7 (rozkłady Laplace'a) jest asymptotycznie normalny, mimo nieróżniczkowalności funkcji ℓ .

Jeśli rozważamy estymację wielkości $g(\theta)$ to przyjmujemy z *definicji*, że $g(\hat{\theta}) = \text{MLE}(g(\theta))$, gdzie $\hat{\theta} = \text{MLE}(\theta)$. Z Twierdzenia 5.2.5 i Lematu 5.2.3 (delta) otrzymujemy natychmiast asymptotyczną normalność tego estymatora:

5.2.6 Wniosek. *Przy założeniach Twierdzenia 5.2.5, jeśli g jest funkcją różniczkowalną, to*

$$\sqrt{n} \left(g(\hat{\theta}_n) - g(\theta_0) \right) \rightarrow_d N \left(0, \frac{g'(\theta_0)^2}{I_1(\theta_0)} \right), \quad (n \rightarrow \infty).$$

Niech $\hat{g} = \hat{g}_n = \hat{g}(X_1, \dots, X_n)$ będzie estymatorem asymptotycznie normalnym, czyli $\sqrt{n}(\hat{g}_n - g(\theta)) \rightarrow_d N(0, \sigma^2(\theta))$. Jego **asymptotyczną efektywność** określamy jako

$$\text{as.ef}(\hat{g}) = \frac{(g'(\theta))^2}{\sigma^2(\theta)I_1(\theta)}.$$

Jest to oczywista modyfikacja definicji „zwykłej” efektywności: rolę wariancji estymatora nieobciążonego przejęła *asymptotyczna* wariancja estymatora asymptotycznie normalnego. Ponieważ zakładamy, że obserwacje X_1, \dots, X_n tworzą ciąg i.i.d. to zgodnie ze Stwierdzeniem 4.2.8 mamy $I_n(\theta) = nI_1(\theta)$. *Asymptotyczna efektywność względna* dwóch estymatorów jest określona analogicznie. Jeśli $\hat{g}_1 = \hat{g}_n^{(1)}$ i $\hat{g}_2 = \hat{g}_n^{(2)}$ są asymptotycznie normalne i mają asymptotyczne wariancje odpowiednio $\sigma_1^2(\theta)$ i $\sigma_2^2(\theta)$, to

$$\text{as.ef}(\hat{g}_1, \hat{g}_2) = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}.$$

Oczywiście, $\text{as.ef}(\hat{g}_1, \hat{g}_2) = \text{as.ef}(\hat{g}_1)/\text{as.ef}(\hat{g}_2)$. Pojęcie względnej efektywności asymptotycznej ma bardzo intuicyjną interpretację. Wyobraźmy sobie, że $\text{as.ef}(\hat{g}_1, \hat{g}_2) = 3$. Wtedy estymatory $\hat{g}_n^{(1)} = \hat{g}_1(X_1, \dots, X_n)$ i $\hat{g}_{3n}^{(2)} = \hat{g}_2(X_1, \dots, X_{3n})$ mają w przybliżeniu ten sam rozkład $N(g(\theta), \sigma_1^2(\theta)/n) = N(g(\theta), \sigma_2^2(\theta)/(3n))$. To znaczy, że estymator \hat{g}_2 potrzebuje trzykrotnie więcej danych niż \hat{g}_1 żeby osiągnąć podobną dokładność. Krótko mówiąc, jest trzykrotnie mniej efektywny.

5.2.7 Przykład. [Estymacja prawdopodobieństwa braku szkód; rozkład Poissona] Niech X_1, \dots, X_n będzie próbka z rozkładu $\text{Poiss}(\theta)$, gdzie $\theta > 0$. W naukach aktuarialnych, rozkład Poissona często opisuje liczby szkód w poszczególnych latach (dla pojedynczej polisy ubezpieczeniowej lub dla grupy polis). Interesującą wielkością jest $e^{-\theta}$, ponieważ jest to prawdopodobieństwo, że w ciągu roku nie będzie żadnych szkód. Zajmiemy się zagadnieniem estymacji funkcji

$$g(\theta) = e^{-\theta} = \mathbb{P}_\theta(X_1 = 0).$$

Rozpatrzmy dwa estymatory:

$$\hat{g}_1 = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i = 0) = \frac{\text{liczba lat w których nie było szkód}}{\text{liczba lat}},$$

$$\hat{g}_2 = e^{-\bar{X}} = \text{MLE}(e^{-\theta}).$$

Mamy

$$\sqrt{n}(\hat{g}_1 - g(\theta)) \rightarrow_d N(0, \sigma_1^2), \quad \text{gdzie } \sigma_1^2 = \text{Var}_\theta \mathbb{1}(X_1 = 0) = e^{-\theta}(1 - e^{-\theta});$$

$$\sqrt{n}(\hat{g}_2 - g(\theta)) \rightarrow_d N(0, \sigma_2^2), \quad \text{gdzie } \sigma_2^2 = (g'(\theta))^2 \text{Var}_\theta X_1 = e^{-2\theta} \theta.$$

Asymptotyczna normalność \hat{g}_1 wynika wprost z CTG dla schematu Bernoulliego (za sukces uważamy obserwację równą zeru). Do \hat{g}_2 możemy zastosować Lemat 5.2.3. Estymator największej wiarygodności \hat{g}_2 jest asymptotycznie efektywny, $\text{as.ef}(\hat{g}_2) = 1$. Względna efektywność asymptotyczna jest dana wzorem

$$\text{as.ef}(\hat{g}_1, \hat{g}_2) = \frac{\sigma_2^2}{\sigma_1^2} = \frac{\theta e^{-2\theta}}{e^{-\theta}(1 - e^{-\theta})} = \frac{\theta}{e^\theta - 1}$$

i oczywiście $\text{as.ef}(\hat{g}_1) = \text{as.ef}(\hat{g}_1, \hat{g}_2) < 1$. Dla małych θ (czyli dla małej częstości szkód) efektywność obu estymatorów jest zbliżona: $\theta/(e^\theta - 1) \rightarrow 1$ przy $\theta \rightarrow 0$. Dla dużych θ estymator \hat{g}_2 jest *wyraźnie* lepszy niż \hat{g}_1 .

Nie należy pochopnie dyskwalifikować estymatorów mniej efektywnych. Mogą odznaczać się innymi zaletami. W rozpatrywanym modelu estymator \hat{g}_2 jest bardziej efektywny niż \hat{g}_1 . Wyobraźmy sobie jednak, że wysokość roszczeń ma rozkład inny, niż rozkład Poissona. Estymator \hat{g}_1 będzie w dalszym ciągu *zgodnym* estymatorem $\mathbb{P}_\theta(X_1 = 0)$, bo nie wykorzystuje założenia o postaci gęstości. Natomiast \hat{g}_2 faktycznie estymuje wielkość $\exp(-\mathbb{E}_\theta X_1)$, która jest na ogół różna od $\mathbb{P}_\theta(X_1 = 0)$, jeśli rozkład nie jest rozkładem Poissona. \diamond

5.3 Zadania

5.1. Niech X_1, \dots, X_n będzie próbką z rozkładu jednostajnego $U(0, \theta)$. Wykazać bezpośrednio (i szczegółowo) mocną zgodność estymatora największej wiarygodności $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$.

5.2. Niech X_1, \dots, X_n będzie próbką z rozkładu wykładniczego $Ex(\theta)$ o gęstości $f_\theta(x) = \theta e^{-\theta x}$ dla $x > 0$. Niech $\hat{\theta}_n$ będzie estymatorem *największej wiarygodności* 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.

5.3. Obserwujemy zmienną losową S o rozkładzie $\text{Bin}(n, p)$, czyli liczbę sukcesów w schemacie Bernoulliego z nieznanym prawdopodobieństwem sukcesu p . Interesuje nas estymacja wielkości $g(p) = p^2$.

- (a) Zbadać zgodność i asymptotyczną normalność estymatora $\tilde{g}(S) = S^2/n^2$.
- (b) Zbadać zgodność i asymptotyczną normalność nieobciążonego estymatora $\hat{g}(S)$ skonstruowanego w Zadaniu 4.5 w Rozdziale 4.

5.4. Niech X_1, \dots, X_n będzie próbką z rozkładu normalnego $N(\mu, \sigma^2)$. Rozważmy dwa estymatory wielkości μ^2 . Pierwszy estymator jest dany wzorem $\hat{\mu}_1^2 = \bar{X}^2$. Drugi to estymator nieobciążony, skonstruowany w Zadaniu 4.11 z rozdziału 4. Oznaczmy go $\hat{\mu}_2^2$.

- (a) Wykazać asymptotyczną normalność $\hat{\mu}_1^2$. Podać asymptotyczną wariancję.
- (b) Wykazać asymptotyczną normalność $\hat{\mu}_2^2$. Podać asymptotyczną wariancję.
- (c) Obliczyć asymptotyczną efektywność względną $\text{as.ef}(\hat{\mu}_1, \hat{\mu}_2)$.
- (d) Porównać z ograniczeniem dolnym Craméra-Rao. Obliczyć asymptotyczną efektywność $\text{as.ef}(\hat{\mu}_1)$ i $\text{as.ef}(\hat{\mu}_2)$.

5.5. Niech X_1, \dots, X_n będzie próbką z rozkładu wykładniczego $Ex(\theta)$. Interesuje nas estymacja funkcji $g(\theta) = e^{-a\theta} = \mathbb{P}_\theta(X_1 > a)$, gdzie $a > 0$ jest znaną liczbą a $\theta > 0$ jest nieznanym parametrem. Rozpatrzmy dwa estymatory:

$$\hat{g}_1 = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i > a)$$

$$\hat{g}_2 = e^{-a/\bar{X}} = \text{MLE}(e^{-a\theta}).$$

- (a) Wykazać asymptotyczną normalność \hat{g}_1 . Podać asymptotyczną wariancję.

- (b) Wykazać asymptotyczną normalność \hat{g}_2 . Podać asymptotyczną wariancję.
- (c) Obliczyć asymptotyczną efektywność względną $\text{as.ef}(\hat{g}_1, \hat{g}_2)$.
- (d) Porównać z ograniczeniem dolnym Craméra-Rao. Obliczyć asymptotyczną efektywność $\text{as.ef}(\hat{g}_1)$ i $\text{as.ef}(\hat{g}_2)$.

5.6. Kontynuacja Przykładu 5.2.7. Estymujemy $g(\theta) = e^{-\theta}$ na podstawie próbki z rozkładu Poiss(θ). Zbadać asymptotyczną wariancję estymatora

$$\hat{g}_3 = \left(1 - \frac{1}{n}\right)^S,$$

gdzie $S = \sum_{i=1}^n X_i$. Obliczyć asymptotyczną efektywność \hat{g}_3 . Porównać z Zadaniem 4.14 i 4.15 z Rozdziału 4.

5.7. Niech X_1, \dots, X_n będzie próbką losową z rozkładu o gęstości

$$g_\theta(x) = \begin{cases} \theta x^{-2} & \text{dla } \theta \leq x < +\infty; \\ 0 & \text{dla } x < \theta, \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem.

- (a) Podać jednowymiarową statystykę dostateczną dla θ .
- (b) Obliczyć estymator największej wiarygodności $\hat{\theta}$ parametru θ .
- (c) Czy estymator $\hat{\theta}$ jest nieobciążony?
- (d) Jaki jest rozkład asymptotyczny $n(\hat{\theta} - \theta)$? Oczywiście, $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ i $n \rightarrow \infty$.

Rozdział 6

Przedziały ufności

Pojęcie przedziału ufności precyzuje ideę estymacji z określoną dokładnością. Zamiast pojedynczego oszacowania nieznanego parametru, podajemy górną i dolną granicę oszacowania. Nie możemy co prawda zagwarantować, że parametr leży na pewno między tymi granicami. Możemy wymagać, by tak było z odpowiednio dużym prawdopodobieństwem.

6.0.1 DEFINICJA. Niech $g(\theta)$ będzie funkcją nieznanego parametru. Rozważmy dwie statystyki $\underline{g} = \underline{g}(X)$ i $\bar{g} = \bar{g}(X)$. Mówimy, że $[\underline{g}, \bar{g}]$ jest **przedziałem ufności** dla $g(\theta)$ na poziomie ufności $1 - \alpha$, jeśli

$$\mathbb{P}_\theta (\underline{g}(X) \leq g(\theta) \leq \bar{g}(X)) \geq 1 - \alpha$$

dla każdego θ .

Typowo, α jest małą liczbą, na przykład $1 - \alpha = 0.95$ lub $1 - \alpha = 0.99$. Zauważmy, że warunek $\mathbb{P}_\theta(g(\theta) \in [\underline{g}, \bar{g}]) = 1 - \alpha$ należy rozumieć tak: „losowy przedział $[\underline{g}, \bar{g}]$ pokrywa nieznaną liczbę $g(\theta)$ z dużym prawdopodobieństwem”. Jeśli obliczone z próbki wartości statystyk są równe, powiedzmy, $\underline{g} = 5$ i $\bar{g} = 8$, to *nie ma sensu* sformułowanie “wielkość $g(\theta)$ należy do przedziału $[5, 8]$ z prawdopodobieństwem $1 - \alpha$!”. Doświadczenie losowe już zostało wykonane i zakończyło się albo „sukcesem” (to znaczy zaszło zdarzenie losowe $g(\theta) \in [\underline{g}, \bar{g}]$) lub „porażką” ($g(\theta) \notin [\underline{g}, \bar{g}]$). Osobliwość sytuacji polega na tym, że *nie wiemy*, która z tych ewentualności ma miejsce.

6.1 Przykłady

Rozpatrzmy typowe i ważne przykłady konstrukcji przedziałów ufności w modelu normalnym.

6.1.1 Przykład (Przedział ufności dla średniej w modelu normalnym). Niech X_1, \dots, X_n będzie próbką z rozkładu $N(\mu, \sigma^2)$.

(i) *Znana wariancja*. Zakładamy, że znamy σ^2 i tylko μ jest nieznanym parametrem. Ponieważ

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

więc jeśli z jest kwantylem rzędu $1 - \alpha/2$ standardowego rozkładu normalnego $N(0, 1)$, to mamy $\mathbb{P}_\mu(|\sqrt{n}(\bar{X} - \mu)/\sigma| \leq z) = \Phi(z) - \Phi(-z) = 1 - \alpha$. Innymi słowy,

$$\mathbb{P}_\mu \left(\bar{X} - \frac{\sigma z}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\sigma z}{\sqrt{n}} \right) = 1 - \alpha,$$

Przedział ufności dla μ ma zatem postać $[\bar{X} - \sigma z/\sqrt{n}, \bar{X} + \sigma z/\sqrt{n}]$, gdzie $z = z_{1-\alpha/2}$.

(ii) *Nieznana wariancja*. Założymy teraz, że wariancja σ^2 rozkładu normalnego jest nieznana. Interesuje nas przedział ufności dla μ . Niech $S^2 = \sum (X_i - \bar{X})^2 / (n - 1)$ będzie zwykle stosowanym nieobciążonym estymatorem wariancji i $S = \sqrt{S^2}$. Wiemy, że

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1),$$

gdzie $t(n - 1)$ jest rozkładem t-Studenta z $n - 1$ stopniami swobody (zobacz Wniosek 2.2.5 w Rozdziale 2). Jeśli więc t jest kwantylem rzędu $1 - \alpha/2$ to $\mathbb{P}_{\mu, \sigma}(|\sqrt{n}(\bar{X} - \mu)/S| \leq t) = 1 - \alpha$. Innymi słowy,

$$\mathbb{P}_{\mu, \sigma} \left(\bar{X} - \frac{St}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{St}{\sqrt{n}} \right) = 1 - \alpha.$$

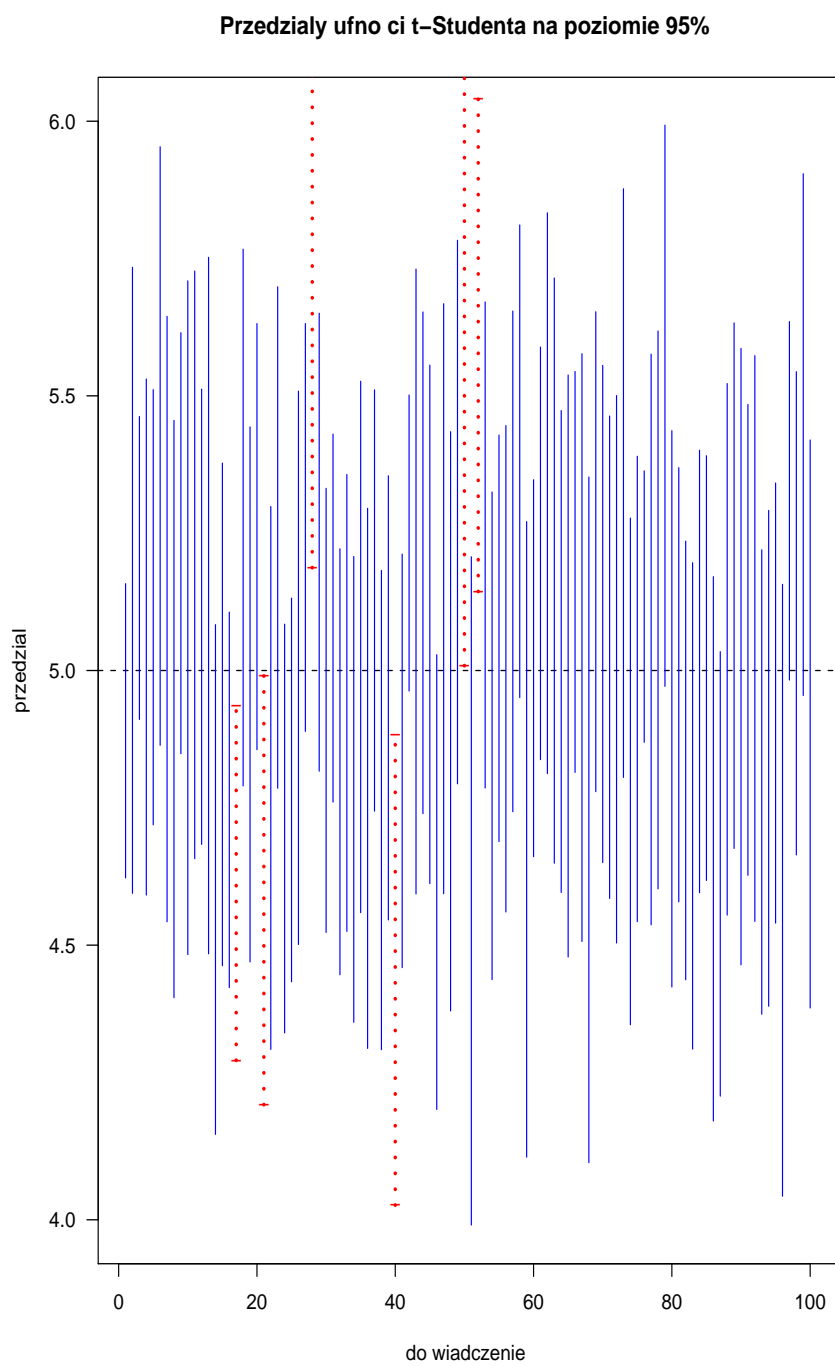
Przedział ufności ma postać $[\bar{X} - St/\sqrt{n}, \bar{X} + St/\sqrt{n}]$, gdzie $t = t_{1-\alpha/2}(n - 1)$.

Przedstawimy graficznie takie przedziały dla 100 symulowanych próbek. Próbki rozmiaru $n = 20$ były generowane z rozkładu $N(\mu, \sigma^2)$. Prawdziwe wartości parametrów użyte w symulacjach to $\mu = 5$ i $\sigma = 1$ ¹ Rysunek 6.1 przedstawia 100 doświadczeń, wśród których zdarzyło się sześć „porażek”. Są one wyróżnione przerywaną linią i kolorem czerwonym. \diamond

6.1.2 Przykład (Przedział ufności dla wariancji). Rozważamy ten sam model, co poprzednio. Mamy próbkę X_1, \dots, X_n z rozkładu $N(\mu, \sigma^2)$ z nieznanymi parametrami μ i σ^2 . Zajmiemy się teraz konstrukcją przedziału ufności dla σ^2 . Wiemy, że

$$\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2(n - 1)$$

¹Jeszcze raz podkreślimy metodologię badań symulacyjnych: generujemy „sztuczne dane” z rozkładu prawdopodobieństwa który znamy, a następnie „udajemy, że go nie znamy” i estymujemy.



Rysunek 6.1: Przedziały ufności dla μ w rozkładzie normalnym. Prawdziwe $\mu = 5$.

(Twierdzenie Fishera, 2.2.3). Niech c_1 i c_2 będą kwantylami rzędu, odpowiednio, $\alpha/2$ i $1-\alpha/2$ rozkładu chi-kwadrat z $n-1$ stopniami swobody. Mamy $\mathbb{P}_{\mu,\sigma}(c_1 \leq (n-1)S^2/\sigma^2 \leq c_2) = 1-\alpha$. Inaczej,

$$\mathbb{P}_{\mu,\sigma} \left(\frac{(n-1)S^2}{c_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{c_1} \right) = 1-\alpha.$$

Przedziałem ufności jest więc $[(n-1)S^2/c_2, (n-1)S^2/c_1]$, gdzie $c_1 = \chi_{\alpha/2}^2(n-1)$ i $c_2 = \chi_{1-\alpha/2}^2(n-1)$. \diamond

6.1.3 Przykład (Przedział ufności dla ilorazu wariancji). Rozważmy dwie niezależne próbki $X_1, \dots, X_k \sim N(\mu_X, \sigma_X^2)$ i $Y_1, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2)$. Nieznanymi parametrami są μ_X , σ_X^2 , μ_Y i σ_Y^2 . Chcemy zbudować przedział ufności dla σ_Y^2/σ_X^2 . Wyobraźmy sobie, że X_i i Y_j są wynikami pomiarów dokonanych przy użyciu dwóch przyrządów. Stosunek wariancji pozwala na porównanie dokładności obu przyrządów (przy tej interpretacji naturalne jest założenie, że $\mu_X = \mu_Y$; nie jest ono jednak potrzebne w dalszych rozważaniach). Wiemy, że

$$\frac{S_X^2 \sigma_Y^2}{S_Y^2 \sigma_X^2} \sim F(k-1, m-1)$$

(zobacz Przykład 2.2.6). W powyższym wzorze, S_X^2 i S_Y^2 oznaczają standardowe nieobciążone estymatory wariancji obliczone, odpowiednio, dla próbki X -ów i Y -ów. Niech f_1 i f_2 będą kwantylami rzędu, odpowiednio, $\alpha/2$ i $1-\alpha/2$ rozkładu F-Snedecora z $k-1$ stopniami swobody licznika i $m-1$ stopniami swobody mianownika. Mamy $\mathbb{P}(f_1 \leq S_X^2 \sigma_Y^2 / (S_Y^2 \sigma_X^2) \leq f_2) = 1-\alpha$. Inaczej,

$$\mathbb{P} \left(f_1 \frac{S_Y^2}{S_X^2} \leq \frac{\sigma_Y^2}{\sigma_X^2} \leq f_2 \frac{S_Y^2}{S_X^2} \right) = 1-\alpha.$$

Rzecz jasna, \mathbb{P} oznacza tu rozkład prawdopodobieństwa zależny od wszystkich parametrów modelu, a więc $\mathbb{P}_{\mu_X, \mu_Y, \sigma_X, \sigma_Y}$. Przedziałem ufności jest $[f_1 S_Y^2 / S_X^2, f_2 S_Y^2 / S_X^2]$, gdzie $f_1 = F_{\alpha/2}(k-1, m-1)$ i $f_2 = F_{1-\alpha/2}(k-1, m-1)$. \diamond

6.2 Asymptotyczne przedziały ufności

Czasami wyznaczenie dokładnego przedziału ufności jest trudne i musimy się zadowolić rozwiązaniami przybliżonymi. Przybliżenia, o których będziemy mówić są uzasadnione w przypadku dużej liczności próbki.

6.2.1 DEFINICJA. *Rozważmy nieskończony ciąg obserwacji X_1, X_2, \dots i dwie statystyki: $\underline{g}_n = \underline{g}(X_1, \dots, X_n)$ i $\bar{g}_n = \bar{g}(X_1, \dots, X_n)$. Powiemy, że $[\underline{g}_n, \bar{g}_n]$ jest **asymptotycznym przedziałem ufności** dla $g(\theta)$ na poziomie ufności $1-\alpha$, jeśli*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(\underline{g}(X_1, \dots, X_n) \leq g(\theta) \leq \bar{g}(X_1, \dots, X_n) \right) \geq 1-\alpha$$

dla każdego θ .

W praktyce, jeśli n jest „odpowiednio duże”, oczekujemy, że warunek $\mathbb{P}_\theta(g(\theta) \in [\underline{g}_n, \bar{g}_n]) \geq 1 - \alpha$ jest w przybliżeniu spełniony.

6.2.2 Przykład (Przedział ufności w schemacie Bernoulliego). Niech $S_n \sim \text{Bin}(n, p)$, nieznanym parametrem jest prawdopodobieństwo sukcesu p . Jeśli liczba prób n jest duża, S_n ma w przybliżeniu rozkład normalny $N(np, np(1-p))$. Innymi słowy,

$$\mathbb{P}_\theta \left(\frac{\sqrt{n} |S_n/n - p|}{\sqrt{p(1-p)}} \leq z \right) \simeq \Phi(z) - \Phi(-z).$$

Weźmy za z kwantyl rzędu $1 - \alpha/2$ standardowego rozkładu normalnego: $\Phi(z) - \Phi(-z) = 1 - \alpha$. Możemy teraz ostatni wzór przepisać w postaci

$$\mathbb{P}_\theta \left(z^2 p(1-p) \geq n (S_n/n - p)^2 \right) \simeq 1 - \alpha.$$

Nierówność w nawiasie jest kwadratowa względem p i możemy ją rozwiązać w standardowy sposób. Wniosujemy, że

$$\left[\frac{S_n + \frac{z^2}{n} - z \sqrt{\frac{S_n(n-S_n)}{n} + \frac{z^2}{4}}}{n + z^2}, \frac{S_n + \frac{z^2}{n} + z \sqrt{\frac{S_n(n-S_n)}{n} + \frac{z^2}{4}}}{n + z^2} \right]$$

jest przybliżonym (asymptotycznym) przedziałem ufności dla p . ◇

Związek z metodą największej wiarygodności

Jeśli $\hat{\theta} = \text{MLE}(\theta)$ i spełnione są założenia Twierdzenia 5.2.5, to zgodnie z Wnioskiem 5.2.6,

$$\sqrt{n} \left(g(\hat{\theta}_n) - g(\theta) \right) \rightarrow_d N \left(0, \frac{(g'(\theta))^2}{I_1(\theta)} \right),$$

dla dowolnej różniczkowalnej funkcji $g(\theta)$. Mamy więc

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(\frac{|g(\hat{\theta}_n) - g(\theta)| \sqrt{n I_1(\theta)}}{|g'(\theta)|} \leq z \right) = \Phi(z) - \Phi(-z).$$

Dla $z = z_{1-\alpha/2}$, prawa strona staje się równa $1 - \alpha$. Z Lematu Slutskiego wynika, że powyższy wzór pozostanie słuszny, gdy zastąpimy po lewej stronie $I_1(\theta)$ i $g'(\theta)$ przez $I_1(\hat{\theta}_n)$

i $g'(\hat{\theta}_n)$. Wynika to ze zgodności MLE : $\hat{\theta} \rightarrow_P \theta$, przy założeniu ciągłości funkcji I_1 i g' . Podsumowując, stwierdzamy, że (przy pewnych założeniach)

$$(6.2.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(g(\hat{\theta}_n) - \frac{z|g'(\hat{\theta}_n)|}{\sqrt{nI_1(\hat{\theta}_n)}} \leq g(\theta) \leq g(\hat{\theta}_n) + \frac{z|g'(\hat{\theta}_n)|}{\sqrt{nI_1(\hat{\theta}_n)}} \right) = 1 - \alpha.$$

W ten sposób otrzymaliśmy bardzo ogólny sposób konstrukcji asymptotycznych przedziałów ufności. Niestety, w praktyce ten sposób nie zawsze jest zadowalający, bo wynikające z asymptotycznych rozważań przybliżenie może okazać się dostatecznie dobre dopiero dla bardzo dużych próbek.

6.2.4 Przykład. Niech X_1, \dots, X_n będzie próbką z rozkładu wykładniczego $\text{Ex}(\theta)$. Skonstruujemy asymptotyczny przedział ufności dla parametru θ . Ponieważ $\hat{\theta}_n = \text{MLE}(\theta) = \bar{X}$ i $I_1(\theta) = 1/\theta^2$, więc otrzymujemy następujący przedział:

$$\left[\frac{1}{\bar{X}} - \frac{z}{\sqrt{n\bar{X}}}, \frac{1}{\bar{X}} + \frac{z}{\sqrt{n\bar{X}}} \right],$$

gdzie $z = z_{1-\alpha/2}$. ◇

6.2.5 Przykład (Jeszcze raz schemat Bernoulliego). Niech $S_n \sim \text{Bin}(n, p)$. Najpopularniejszy (co wcale nie znaczy, że najlepszy) wzór na asymptotyczny przedział ufności dla p wynika wprost z ogólnej metody 6.2.3:

$$\left[\hat{\theta} - \frac{z\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}}, \hat{\theta} + \frac{z\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}} \right],$$

gdzie $\hat{\theta} = S_n/n$ i $z = z_{1-\alpha/2}$. ◇

6.3 Zadania

6.1. 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 $N(\mu, \sigma^2)$ z nieznanymi parametrami μ i σ . Podać przedział ufności dla średniej masy paczki μ na poziomie ufności $1 - \alpha = 0.95$.

6.2. Interesuje nas procent absolwentów warszawskich wyższych uczelni, którzy znajdują pracę nie później niż pół roku po ukończeniu studiów. Zbadano reprezentatywną próbkę $n = 400$ absolwentów i okazało się, że wśród nich $S = 256$ znalazło pracę.

(a) Podać przedział ufności na poziomie $1 - \alpha = 0.95$ dla interesującego nas procentu. Użyć najprostszej metody asymptotycznej, czyli wzoru z Przykładu 6.2.5.

(b) Zrobić to samo, używając wzoru z Przykładu 6.2.2. Porównać.

6.3. 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 ?

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

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

6.6. Wyprowadzić dokładnie wzór podany w Przykładzie 6.2.5 ze wzoru 6.2.3.

6.7. Mamy próbkę X_1, \dots, X_n z rozkładu $N(\mu, \sigma^2)$ z nieznanym parametrem σ^2 i *znanym* μ . Skonstruować przedział ufności dla σ^2 w podobny sposób jak w Przykładzie 6.1.2, wykorzystując jednak znajomość μ . Dobrać stałe c_1 i c_2 tak, aby

$$\mathbb{P}_\sigma \left(\frac{1}{c_2} \sum_{i=1}^n (X_i - \mu)^2 \leq \sigma^2 \leq \frac{1}{c_1} \sum_{i=1}^n (X_i - \mu)^2 \right) = 1 - \alpha.$$

6.8. W sytuacji opisanej w Przykładzie 6.1.2, skonstruować *jednostronny* przedział ufności dla σ^2 postaci $[0, \bar{\sigma}^2]$. Dobrać stałą c tak, aby

$$\mathbb{P}_{\mu, \sigma} \left(\sigma^2 \leq \frac{(n-1)S^2}{c} \right) = 1 - \alpha.$$

6.9. Niech X_1, \dots, X_5 będzie próbką z nieznanego rozkładu o dystrybuancie F . Załóżmy, że m jest jednoznacznie wyznaczoną medianą tego rozkładu, $F(m) = 1/2$. Dla jakiego poziomu ufności $1 - \alpha$ przedział $[X_{2:5}, X_{4:5}]$ jest przedziałem ufności dla m na poziomie $1 - \alpha$? Innymi słowy, oblicz

$$\mathbb{P}_F(X_{2:5} \leq m \leq X_{4:5}).$$

Wskazówka: Porównać z Zadaniem 1.11 w Rozdziale 1.

6.10. Niech X_1, \dots, 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

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

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

6.11. Niech X_1, \dots, X_n będzie próbką z rozkładu Rayleigha o gęstości

$$f_\sigma(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} & \text{dla } x > 0; \\ 0 & \text{dla } x \leq 0, \end{cases}$$

gdzie $\sigma > 0$ jest nieznanym parametrem.

- Wyznaczyć estymator największej wiarygodności dla σ^2 . Oznaczmy ten estymator $\hat{\sigma}^2$.
- Skonstruować przedział ufności dla σ^2 na poziomie ufności $1 - \alpha$ postaci

$$[c_n \hat{\sigma}^2, d_n \hat{\sigma}^2]$$

Wskazówka: Znaleźć rozkład $Y_1 = X_1^2$ i następnie $\text{MLE}(\sigma^2) = \hat{\sigma}^2$.