Wstęp do statystyki matematycznej
Lista 7

1. Uzasadnić średniokwadratową różniczkowalność oraz wyznaczyć wektor wynikowy i macierz informacji Fishera dla rodziny rozkładów:

   (a) ujemnie dwumianowych $\mathcal{NB}(r_0, p)$, z parametrem $p \in (0,1)$, r_0 ustalone i znane;

   (b) gamma $\mathcal{G}(\alpha, \beta)$;

   (c) beta $\mathcal{B}e(\alpha, \beta)$;

   (d) normalnych $\mathcal{N}(\sigma, \sigma^2)$, $\sigma > 0$;

   (e) mieszanek rozkładów normalnych $p_{\sigma\theta}(x) = \theta\varphi(x) + (1-\theta)\frac{1}{\sigma}\varphi(\frac{x}{\sigma})$, $\sigma > 0, \theta \in (0,1)$.

2. Wykazać, że rodzina rozkładów Pareto $p_\theta(x) = \alpha\theta^\alpha x^{-\alpha-1}\mathbf{1}_{(\theta,\infty)}(x)$, $\theta > 0$, α ustalone i znane, nie jest średniokwadratowo różniczkowalna.

3. Niech $p_0(x)$ będzie gęstością dodatnią na prostej o ciągłej pochodnej. Podać warunki wystarczające na to, aby rodzina rozkładów z parametrami położenia i skali wyznaczona przez p_0 była średniokwadratowo różniczkowalna . Wyznaczyć wektor wynikowy i macierz informacji Fishera. Jako przykład rozważyć rodzinę rozkładów Cauchy'ego.

4. Niech $X = (X_1, ..., X_n)$ będzie próbą z rozkładu Bernoulliego $\mathcal{B}(1, p)$, $p \in (0,1)$. Uzasadnić, że $T(X) = \frac{n}{n-1}\overline{X}(1 - \overline{X})$ jest estymatorem UMVU funkcji parametrycznej $g(p) = p(1-p)$. Sprawdzić, czy $T(X)$ jest efektywny lub asymptotycznie efektywny.

5. Niech $X_1, ..., X_n$ będzie próbą z rozkładu

   (a) normalnego $\mathcal{N}(\mu, 1)$. Pokazać, że estymator $\overline{X}$ parametru μ jest efektywny. Dla funkcji parametrycznej $g(\mu) = e^{t_0\mu}$, t_0 ustalone i znane, wyznaczyć estymator UMVU i wyznaczyć jego efektywność asymptotyczną.

   (b) normalnego $\mathcal{N}(0, \sigma)$. Pokazać, że estymator $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ jest nieobciążony i efektywny.

6. Niech $X = (X_1, ..., X_n)$ będzie próbą z rozkładu normalnego $\mathcal{N}(\mu, \sigma)$, μ ustalone i znane. Wykazać, że estymator $T(X) = \frac{\sqrt{\pi}}{n\sqrt{2}}\sum_{i=1}^{n}|X_i - \mu|$ parametru σ jest nieobciążony i ma efektywność $\frac{1}{\pi - 2}$.

7. Rozważmy rodzinę rozkładów dyskretnych na zbiorze $\{1, 2, 3\}$ o gęstości $p_\theta(1) = \theta^2$, $p_\theta(2) = 2\theta(1-\theta)$, $p_\theta(3) = (1-\theta)^2$, $\theta \in (0,1)$. W oparciu o próbę $X_1, ..., X_n$ wyznaczyć estymator parametru θ metodą momentów i metodą podstawienia.

8. Niech $X_1, ..., X_n$ będzie próba z rozkładu o gęstości $p_\alpha(x) = (2\alpha x + 1 - \alpha)\mathbf{1}_{(0,1)}(x)$, $\alpha \in (-1, 1)$. Wyznaczyć dwa estymatory zgodne parametru α metodą momentów.

9. Niech $X = (X_1, ..., X_n)$ będzie próbą z rozkładu gamma $\mathcal{G}(\alpha, \beta)$.

   (a) Wyznaczyć estymatory parametrów metodą momentów;

   (b) Powtórzyć to samo dla rodziny jednoparametrowej, gdy $\alpha = \alpha_0$ jest ustalone i znane. Obliczyć efektywność tak otrzymanego estymatora (po usunięciu obciążenia).

Wstęp do statystyki matematycznej
Lista 8

1. Niech $X_1, ..., X_n$ będzie próba z rozkładu Weibulla $\mathcal{W}(\alpha, \beta)$, α ustalone i znane. Znaleźć estymator MLE funkcji parametrycznej β^α. Sprawdzić, czy jest nieobciążony. Wykazać, żę jest zgodny i asymptotycznie normalny.

2. Niech $X_1, ..., X_n$ będzie próbą z rozkładu
   (a) gamma $\mathcal{G}(\alpha, \beta)$, α ustalone i znane (por. zad. 9 poprzedniej listy);
   (b) geometrycznego $\mathcal{G}e(p)$, $p \in (0, 1)$;
   (c) jednostajnego na odcinku $[m, m + 1]$, $m \in R$;
   (d) normalnego $\mathcal{N}(\sigma, \sigma^2)$.

   Znaleźć estymatory MLE parametrów rozkładu. Zbadać ich nieobciążoność i zgodność.

3. Niech $X_1, ..., X_n$ będzie próbą z rozkładu Pareto $\mathcal{P}a(1, \alpha)$, $\alpha > 1$. Znaleźć estymatory parametru α metodą momentów i metodą największej wiarogodności. Wykazać, że są estymatorami zgodnymi i dla $\alpha > 2$ asymptotycznie normalnymi. Porównać wariancje asymptotyczne tych estymatorów.

4. W jeziorze jest nieznana liczba N ryb. W celu oszacowania N złowiono m ryb, oznakowano i wpuszczono do jeziora. Po dłuższym czasie złowiono ponownie m ryb i okazało się, że k z nich jest oznakowanych. Znaleźć estymator MLE parametru N.

5. Niech $X_1, ..., X_n$ będzie próbą z rozkładu o gęstości $\theta\, p(\theta x)$, $\theta > 0$, gdzie p jest różniczkowalną gęstością na $(0, \infty)$ lub symetryczną na R. Pokazać, że równanie wiarogodnosci ma jedno rozwiązanie, gdy $xp'(x)/p(x)$ jest malejąca na $(0, \infty)$. Sprawdzić, że warunek jest spełniony dla rozkładu Cauchy'ego i znaleźć rozkład asymptotyczny estymatora otrzymanego w ten sposób.

6. Niech $(X_1, Y_1), ..., (X_n, Y_n)$ będzie próbą z dwuwymiarowego rozkładu normalnego o średniej 0, wariancjach 1 i kowariancji $\theta \in (-1, 1)$. Pokazać, że równanie wiarogodności jest równaniem trzeciego stopnia względem θ i dla $n \to \infty$ prawdopodobieństwo, że ma ono jedno rozwiązanie dąży do 1. Pokazać, że to rozwiązanie jest estymatorem asymptotycznie normalnym $\mathcal{N}(\theta, (1 - \theta^2)^2/n(1 + \theta^2))$.

7. Rozważmy model regresji liniowej jednokrotnej, w której błędy mają wariancje różne dla każdej obserwacji, wynoszą $\sigma_1^2, ..., \sigma_n^2$ i są znane. Metoda ważonych najmniejszych kwadratów minimalizuje wyrażenie $\sum_{i=1}^{n}(Y_i - \vartheta_1 - \vartheta_2 x_i)^2/\sigma_i^2$. Znaleźć estymatory parametrów a i b, sprowadzając ten problem do zagadnienia regresji liniowej wielokrotnej.

8. Rozważmy model regresji wielomianowej $Y_i = \vartheta_1 + \vartheta_2 x_i + \vartheta_3 x_i^2 + \varepsilon_i$, $i = 1, ..., n$. Wyznaczyć estymator LSE parametru wektorowego ϑ. Tak jak w poprzednim zadaniu przeformułować problem na zagadnienie regresji liniowej wielokrotnej. Podać warunek istnienia estymatora.

9. Rozważmy model regresji wykładniczej $Y_i = \alpha + \beta e^{\gamma x_i} + \varepsilon_i$, $i = 1, ..., n$. Wyznaczyć oszacowania LSE parametrów, jeśli dla wartości x wynoszących 0.1, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20 zaobserwowano wartości y odpowiednio 3.04, 3.46, 3.55, 3.90, 4.00, 4.36, 4.39, 4.69, 4.83, 4.71, 4.87, 4.99, 4.90, 4.99. Oszacować nieznaną wariancję błędu.