Zadania z Analizy danych

zestaw 5

Zadanie 1

Zmienna losowa podlega rozkładowi dwumiennemu

$$P(k) = \frac{N!}{k!\,(N-k)!} p^k (1-p)^{N-k},$$

gdzie N jest liczbą prób Bernoulliego, k liczbą prób zakończonych sukcesem, a p prawdopodobieństwem sukcesu w pojedynczej próbie. Wykonano n prób wyznaczenia wartości zmiennej k i uzyskano wyniki $k_1, k_2, \dots, k_n$. Oszacuj na tej podstawie wartość parametru p stosując metodą momentów oraz metodą największej wiarygodności.

Rozwiązanie

a) Metoda momentów

$$m_1(0) \equiv M_1 = \bar{k} = Np$$

$$M_1 \approx T_n\big(m_1(0)\big) = \frac{1}{n}\sum_{i=1}^{n} k_i.$$

W powyższym wyrażeniu n jest liczbą prób dla rozkładu dwumiennego (N – liczba prób Bernouliego).

$$Np = \frac{1}{n}\sum_{i=1}^{n} k_i.$$

Stąd

$$p = \frac{1}{nN}\sum_{i=1}^{n} k_i.$$

b) Metoda największej wiarygodności

$$L = \prod_{i=1}^{n} \frac{N!}{k_i!\,(N-k_i)!} p^{k_i}(1-p)^{N-k_i}$$

$$l = \ln L = \sum_{i=1}^{n} \ln\left(\frac{N!}{k_i!\,(N-k_i)!} p^{k_i}(1-p)^{N-k_i}\right)$$

$$= \sum_{i=1}^{n}\left(\ln\frac{N!}{k_i!\,(N-k_i)!} + k_i \ln p + (N-k_i)\ln(1-p)\right)$$

$$\frac{dl}{dp} = \sum_{i=1}^{n}\left(\frac{k_i}{p} - \frac{N-k_i}{1-p}\right) = \sum_{i=1}^{n}\left(\frac{k_i - k_i p - Np + k_i p}{p(1-p)}\right) = \sum_{i=1}^{n}\left(\frac{k_i - Np}{p(1-p)}\right) = 0$$

$$\sum_{i=1}^{n} k_i = nNp$$

$$p = \frac{1}{nN}\sum_{i=1}^{n} k_i.$$

Zadanie 2

Zmienna losowa ma rozkład równomierny na przedziale $[a, b]$. Znajdź metodą momentów estymatory parametrów a, b. Korzystając z tych estymatorów oszacuj wartości tych parametrów na podstawie następującej próby:

3,5; 7,6; 2,9; 4,3; 2,0; 3,8; 6,5; 6,4; 5,0; 2,8; 4,5; 7,8; 3,7; 6,4; 5,0; 4,4; 2,5; 6,7; 2,1; 7,7; 3,2; 6,2; 4,7; 6,3; 7,1.

Rozwiązanie

$$m_1(0) \equiv M_1 = \frac{a+b}{2}$$

$$m_2(0) - \left(m_1(0)\right)^2 = M_2 - M_1^2 = \sigma^2(X) = \frac{(b-a)^2}{12}.$$

$$M_1 \approx T_n\left(m_1(0)\right) = \frac{1}{n}\sum_{i=1}^{n} x_i = 4{,}924, \qquad M_1^2 = 4{,}924^2 = 24{,}245776$$

$$M_2 \approx T_n\left(m_2(0)\right) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 27{,}4884$$

$$\begin{cases} \dfrac{a+b}{2} = 4{,}924 \\ \dfrac{(b-a)^2}{12} = 27{,}4884 - 4{,}924^2 = 3{,}24264 \end{cases}$$

$$\begin{cases} a+b = 9{,}848 \\ (b-a)^2 \approx 38{,}9115 \end{cases}$$

$$\begin{cases} a+b = 9{,}848 \\ b-a \approx 6{,}2379 \end{cases}$$

$$\begin{cases} 2a \approx 3{,}69009 \\ 2b \approx 16{,}0859 \end{cases}$$

$$\begin{cases} a \approx 1{,}81 \\ b \approx 8{,}04 \end{cases}$$

Lub ogólnie

$$\begin{cases} \dfrac{a+b}{2} = M_1 \\ \dfrac{(b-a)^2}{12} = M_2 - M_1^2 \end{cases}$$

$$\begin{cases} a+b = 2M_1 \\ (b-a)^2 = 12M_2 - 12M_1^2 \end{cases}$$

$$\begin{cases} a+b = 2M_1 \\ b-a = 2\cdot\sqrt{3M_2 - 3M_1^2} \end{cases}$$

$$\begin{cases} a = M_1 - \sqrt{3(M_2 - M_1^2)} \\ b = M_1 + \sqrt{3(M_2 - M_1^2)} \end{cases}$$

Zadanie 3

Pewna zmienna losowa $x > 0$ ma rozkład wykładniczy o postaci

$$f(x) = Ce^{-Cx}.$$

W wyniku pomiarów tej zmiennej otrzymano następujące wartości:

2; 3; 3; 5; 6; 7; 7; 8; 8; 9; 10; 12; 13; 14; 15; 23; 24; 27; 29; 35.

Oszacuj metodą momentów oraz metodą największej wiarygodności wartość parametru C.

Rozwiązanie

a) Metoda momentów

$$m_1(0) \equiv M_1 = \frac{1}{C}.$$

$$M_1 \approx T_n(m_1(0)) = \frac{1}{n}\sum_{i=1}^{n} x_i = 13$$

$$\frac{1}{C} = 13$$

$$C = \frac{1}{13} \approx 0{,}0769.$$

b) Metoda największej wiarygodności

$$L = \prod_{i=1}^{n} Ce^{-Cx_i} = C^n e^{-C\sum_{i=1}^{n} x_i}$$

$$l = \ln C^n e^{-V\sum_{i=1}^{20} x_i} = n \ln C - C\sum_{i=1}^{n} x_i$$

$$\frac{dl}{dC} = \frac{n}{C} - \sum_{i=1}^{n} x_i = 0$$

$$C^{-1} = \frac{1}{n}\sum_{i=1}^{n} x_i = 13$$

$$C = \frac{1}{13} \approx 0{,}0769.$$

Zadanie 4

Pobrano czteroelementową próbę zmiennej losowej podlegającej rozkładowi normalnemu. Znajdź przedział ufności wartości oczekiwanej tej zmiennej dla poziomu ufności $\gamma = 0{,}9$. Wyniki pomiarów były następujące: 10,5; 10,2; 10,4; 10,5.

Wskazówka

Skorzystaj z wyrażenia $\bar{x} - t_{n-1,\frac{1}{2}+\frac{\gamma}{2}} \frac{S(X)}{\sqrt{n}} \le \hat{x} \le \bar{x} + t_{n-1,\frac{1}{2}+\frac{\gamma}{2}} \frac{S(X)}{\sqrt{n}}$. Odpowiednie kwantyle można obliczyć w programie MS Excel korzystając z funkcji ROZKŁAD.T.ODWR

Rozwiązanie

$$\bar{x} = 10,4,$$

$$S(X) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}} \approx 0,14142,$$

$$t_{4,\frac{1}{2}+\frac{0,9}{2}} \approx 2,353.$$

$$\bar{x} - t_{4-1,\frac{1}{2}+\frac{0,9}{2}} \frac{S(X)}{\sqrt{4}} \approx 10,23,$$

$$\bar{x} + t_{4-1,\frac{1}{2}+\frac{0,9}{2}} \frac{S(X)}{\sqrt{4}} \approx 10,57,$$

Z prawdopodobieństwem 0,9

$$10,23 \leq \bar{x} \leq 10,57.$$

Zadania 5

Oszacuj przedział ufności odchylenia standardowego na poziomie ufności $\gamma = 0,9$ dla danych z poprzedniego zadania.

Wskazówka

Skorzystaj z wyrażenia $\frac{(n-1)S^2(X)}{(\chi^2_{n-1})_{\frac{1}{2}+\frac{\gamma}{2}}} \leq \sigma^2(X) \leq \frac{(n-1)S^2(X)}{(\chi^2_{n-1})_{\frac{1}{2}-\frac{\gamma}{2}}}$. Odpowiednie kwantyle można obliczyć w programie MS Excel korzystając z funkcji ROZKŁAD.CHI.ODWR.

Rozwiązanie

$$(\chi^2_{4-1})_{\frac{1}{2}+\frac{0,9}{2}} \approx 7,815$$

$$(\chi^2_{4-1})_{\frac{1}{2}-\frac{0,9}{2}} \approx 0,352.$$

$$\frac{(4-1)S^2(X)}{(\chi^2_{4-1})_{\frac{1}{2}+\frac{0,9}{2}}} \approx 0,00768$$

$$\frac{(4-1)S^2(X)}{(\chi^2_{4-1})_{\frac{1}{2}-\frac{0,9}{2}}} \approx 0,171$$

$$0,00768 \leq \sigma^2(X) \leq 0,876$$

$$0,00768 \leq \sigma(X) \leq 0,413$$