Statystyka 2 Kartkówka mat. Szkic rozwiąząń

Proszę o pełny i czytelny zapis rozwiązań.

Zad.1 Przyjmując rozkład a priori Π_0 dla parametru $\theta \in \{0.5, 1, 1.5, 2\}$ jako rozkład $\Pi_0(0.5) = \frac{1}{4}$, $\Pi_1(1) = 0$, $\Pi_0(1.5) = \frac{1}{4}$, $\Pi_0(2) = \frac{1}{2}$, wyznaczyć rozkład a posteriori Π_1 parametru θ wiedząc, że cecha x_1 wylosowana z rozkładu jednostajnego na odcinku $[0, \theta]$ ma wartość $x_1 = 0,9$.

Odpowiedź: Niech f_θ- gęstość rozkładu jednostajnego $U[0, \theta]$ Mamy

$$\Pi_1(\theta) = \frac{f_\theta(x_1) \cdot \Pi_0(\theta)}{\sum\limits_{\tilde{\theta}} f_\theta(x_1)\Pi_0(\tilde{\theta})},$$

a zatem:

(a) $\Pi_1(0.5) = 0$

(b) $\Pi_2(1) = 0$

(c) $\Pi_3(1.5) = \frac{\frac{2}{3}\cdot\frac{1}{4}}{\frac{2}{3}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{1}{2}}$

(d) $\Pi_3(2) = 1 - \Pi_3(1.5)$

Zad. 2 Niech $Y = [Y_1, Y_2, Y_3]$ ma wielowymiarowy rozkład normalny o parametrach $\mu = [2, 3, 4]$ oraz $\Sigma = [3, 2, 1; 2, 2, 1; 1, 1, 1]$. Wyznacz rozkład wektora $(X_1, X_2) = (Y_1 - Y_2 + Y_3, Y_1 - Y_2 - Y_3)$. Czy zmienne losowe X_1 i X_2 są niezależne? Czy wektor (X_1, X_2) ma gęstość- jeśli tak, wypisz ją.

Odpowiedź: X jest afinicznym przekształceniem Y, a zatem również ma rozkład normalny. Parametry tego rozkładu to:

$$\mu_X = [1, -1, 1; 1, -1, -1]\mu^T = [3, -5]^T$$

oraz

$$\Sigma_X = A\Sigma A^T = [2, 0; 0, 2], \text{ gdzie } A = [1, -1, 1; 1, -1, -1].$$

Widzimy że $\Sigma_X[2, 1] = 0$, a zatem X_1 i X_2 są nieskorelowane- w naszym przypadku jest to równoważne niezależności. Skoro są niezależne, mamy $f_X(a, b) = f_{X_1}(a) \cdot f_{X_2}(b)$, gdzie f_{X_i} to gęstość X_i, czyli jednowymiarowego rozkładu normalnego z parametrami $\mu_X[i]$ oraz $\sqrt{\Sigma_X[i, i]}$

Zad.3 Niech

$$\mathbb{R}^n \supset V = \{x \in \mathbb{R}^n \colon x_1 = -x_n \text{ oraz } x_i = 0 \text{ dla } i = 2, 3, \ldots, n-1\}$$

będzie przestrzenią liniową oraz niech $Z = V^{\perp}$ będzie przestrzenią liniową prostopadłą do V. Wyznacz macierze rzutowania prostopadłego na przestrzenie V oraz Z.

Odpowiedź: Macierz rzutowania P_V to macierz odzworowania $P(x) = [\frac{x_1 - x_n}{2}, 0, \ldots, 0, \frac{x_n - x_1}{2}]$ a P_Z to macierz odwzorowania $I_n - P$

Zad.4 a) Zdefiniuj histogram. Uzasadnij, że jest to gęstość. b) Rozważmy rozkład jednostajny na odcinku $[0, 2]$ oraz pochodzącą z niego próbkę prostą $x = (0,1; 0,4; 0,5; 0,8; 1,1; 1,5; 1,9)$. Dla histogramu pochodzącego od tej próbki o szerkości pasma $h = 0,3$ wylicz scałkowany błąd średniokwadratowy (MISE).

Odpowiedź: Wyznaczamy histogram ze wzoru z wykładu - dowód, że jest to gęstość był na ćwiczeniach. Druga część zadania jest niestety rachunkowa, mamy do policzenia całkę z funkcji kawałkami stałej $\int\limits_{\mathbb{R}} (f_h(x) - \frac{1}{2} \cdot 1_{[0,2]}(x))^2$.

Zad.5 Na podstawie niezależnych par obserwacji (X,Y):

Tabela 1:

Y	3	3	9
X	1	2	5

wyznacz wartości estymatorów parametrów α oraz β w modelu liniowym:

$$Y = a \cdot X + b + \epsilon,$$

gdzie ϵ ma rozkład normalny $N(0, \sigma)$. Wyznacz odręcznie współczynnik R^2

Odpowiedź: Można skorzystać, ze wzoru z wykładu, lub metodą najmniejszych kwadratów znajdujemy minimum $(3-a-b)^2 + (3-2a-b)^2 + (9-5a-b)^2$ co daje $(a, b) = (\frac{21}{13}, \frac{9}{13})$. Współczynnik R^2 wyliczamy ze znanego nam wzoru

$$R^2 = 1 - \frac{[3 - \frac{21}{13}1 - \frac{9}{13}]^2 + [3 - \frac{21}{13}2 - \frac{9}{13}]^2 + [9 - \frac{21}{13}5 - \frac{9}{13}]^2}{(3-3)^2 + (3-3)^2 + (9-3)^2}$$