1 Rachunek

1. Pojęcie wektora losowego
2. Wartość oczekiwana wektora losowego
3. Warunkowa wartość oczekiwana wektora losowego
4. Macierz wariancji kowariancji wektora losowego
5. Własności rozkładu normalnego, rozkładu χ^2, rozkładu t i F.

1.1 Zadania: własności wartości oczekiwanej i wariancji

1. Pokazać, że $\mathrm{Cov}\left(x_i, x_j\right) = \mathrm{E}\left(x_i x_j\right) - \mathrm{E}\left(x_i\right)\mathrm{E}\left(x_j\right)$
2. Pokazać, że jeśli $\mathrm{E}\left(x_i\right) = 0$ to $\mathrm{Var}\left(x_i\right) = \mathrm{E}\left(x_i^2\right)$
3. Które z macierzy

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 2 & 1 \\ 4 & 2 & 3 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 2 & 2 \\ 2 & 2 & 3 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

   mogą być macierzami kowariancji?

4. Pokazać, że dla dowolnego wektora losowego $\boldsymbol{\varepsilon}$, wektora nielosowego $\mathbf{a}$ i macierzy nielosowej $\mathbf{B}$

$$\begin{aligned} \mathrm{E}\left(\mathbf{a} + \mathbf{B}\boldsymbol{\varepsilon}\right) &= \mathbf{a} + \mathbf{B}\,\mathrm{E}\left(\boldsymbol{\varepsilon}\right) \\ \mathrm{Var}\left(\mathbf{a} + \mathbf{B}\boldsymbol{\varepsilon}\right) &= \mathbf{B}\,\mathrm{Var}\left(\varepsilon\right)\mathbf{B}' \end{aligned}$$

5. Mamy wektor losowy $\mathbf{x}$, przy czym $\mathrm{E}\left(\mathbf{x}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathrm{Var}\left(\mathbf{x}\right) = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$. Policz wartość oczekiwaną i wariancję $\mathbf{y} = \begin{bmatrix} x_1 + 2x_2 + 5 \\ x_1 + x_2 + 1 \end{bmatrix}$.

6. Mamy wektor losowy $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, przy czym $\mathrm{E}\left(\mathbf{x}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathrm{Var}\left(\mathbf{x}\right) = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$. Policzyć:

   (a) odchylenie standardowe x_1, x_2

   (b) współczynnik korelacji między x_1, x_2

(c) wartość oczekiwaną i wariancję dla $y = y = 5 + x_1 + 2x_2$

7. Udowodnić, że dla dowolnej macierzy losowej $\mathbf{A}$: $\mathrm{E}\left[\mathrm{tr}\left(\mathbf{A}\right)\right] = \mathrm{tr}\left[\mathrm{E}\left(\mathbf{A}\right)\right]$
8. Załóżmy, że $\mathrm{E}\left(x\right) > 0$. Jaka jest relacja między $\mathrm{E}\left(x\right)$ i $\mathrm{E}\left(\frac{1}{x}\right)$?

   Podpowiedź: wykorzystaj twierdzenie Jensena.
9. Załóżmy, że y i x są zmiennymi losowymi, czemu równe jest $\mathrm{E}\left(\left.\frac{y}{x}\right| x\right)$?
10. Wiemy, że $\mathrm{E}\left(x\right) = 2$ oraz $\mathrm{E}\left(y\right| x\left.\right) = 1 + 2x$. Czemu równe jest $\mathrm{E}\left(y\right)$?

2 Własności rozkładu normalnego

1. Jaki rozkład ma $\mathbf{v} = \mathbf{a} + \mathbf{B}\boldsymbol{\varepsilon}$, jeśli $\boldsymbol{\varepsilon} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$?
2. Pokazać, że dla k-wymiarowego wektora losowego $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ forma kwadratowa $\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \sim \boldsymbol{\chi}_k^2$
3. Mamy wektor losowy $\mathbf{x} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, gdzie $\mu = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Jaki rozkład ma zmienna losowa $v = x_1 + 2x_2 + x_3$?
4. Mamy wektor losowego $\mathbf{x} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, gdzie $\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$. Pokazać, że wektor $v = \begin{bmatrix} x_1 - x_2 - 1 \\ -x_1 + 2x_2 + 4 \end{bmatrix}$ ma rozkład $v \sim N\left(0, \mathbf{I}\right)$. Udowodnić, że $\left(x_1 - x_2 - 1\right)^2 + \left(-x_1 + 2x_2 + 4\right)^2 \sim \chi_2^2$.

   Pokazać ten sam wynik przy użyciu faktu, że $\boldsymbol{\Sigma}^{-\frac{1}{2}} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$
5. (*)

   (a) Pokazać, dla macierzy idempotentej $\mathbf{M}$ wartości własne są równe 0 lub 1. Jaka formę będzie miała macierz $\boldsymbol{\Lambda}$ w dekompozycji spektralnej $\mathbf{M} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$

   (b) Jaki rozkład ma $\mathbf{C}\boldsymbol{\varepsilon}$ jeśli $\mathbf{C}\mathbf{C}' = \mathbf{I}$ a $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\sigma}^2 I\right)$

   (c) Pokazać, że dla macierzy idempotentnej rząd macierzy jest równy jej śladowi

   (d) Pokazać, że dla n-wymiarowego wektora losowego $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\sigma}^2 I\right)$ i dowolnej macierzy idempotentnej $\mathbf{M}$ rzędu k forma kwadratowa $\frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{\sigma^2} \sim \boldsymbol{\chi}_k^2$
   Podpowiedź: skorzystaj z punktu 5a.

6. (*)

   (a) Pokazać, że $\mathbf{x} - \overline{\mathbf{x}} = \mathbf{x}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)$ i macierz $\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$ jest macierzą idempotentną rzędu $n-1$

   (b) Pokazać, że dla n-wymiarowego wektora losowego $\mathbf{x} \sim N\left(0, \boldsymbol{\sigma}^2\mathbf{I}\right)$ suma $\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \sim \chi^2_{n-1}$
   Podpowiedź: wykorzystaj wynik z punktu (5d).

   (c) (*) Pokazać, że $\overline{\mathbf{x}}$ i $\mathbf{x} - \overline{\mathbf{x}}$ są nieskorelowane. Załóżmy, że $\mathbf{x} \sim N\left(0, \boldsymbol{\sigma}^2\mathbf{I}\right)$ ma rozkład normalny. Pokazać, że w tym przypadku $\overline{x}$ i $\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \left(\mathbf{x} - \overline{\mathbf{x}}\right)'\left(\mathbf{x} - \overline{\mathbf{x}}\right)$ są niezależne.

   (d) Pokaż, że że dla n-wymiarowego wektora losowego $\mathbf{x} \sim N\left(0, \boldsymbol{\sigma}^2\mathbf{I}\right)$ staystyka $\frac{\overline{x}}{\sqrt{s_x^2}} \sim t_{n-1}$ a $\frac{\overline{x}^2}{s_x^2} \sim F\left(1, n-1\right)$, gdzie $s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$
   Podpowiedź:Wykorzystaj wynik z punktów (6c i 6b).

3 Statystyka

1. Pojęcie estymatora
2. Nieobciążoność estymatora
3. Wariancja estymatora i efektywność
4. Przedziały ufności
5. Testowanie hipotez statystycznych, wartości krytyczne i wartości p

3.1 Zadania

1. Pokazać, że jeśli mamy dwa estymatory $\widehat{\boldsymbol{\theta}}$ i $\widetilde{\boldsymbol{\theta}}$ wektora parametrów $\boldsymbol{\theta}$ o wariancjach $\widetilde{\boldsymbol{\Sigma}}$ i $\widehat{\boldsymbol{\Sigma}}$ i różnica $\widehat{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{\Sigma}}$ jest dodatnio określona, to dla każdego $\boldsymbol{\delta} \neq \mathbf{0}$

$$\operatorname{Var}\left(\boldsymbol{\delta}'\widehat{\boldsymbol{\theta}}\right) > \operatorname{Var}\left(\boldsymbol{\delta}'\widetilde{\boldsymbol{\theta}}\right)$$

2. Mamy zmienne losowe y_1 i y_2 takie, że $\mathrm{E}\left(y_1\right) = \theta$, $\mathrm{E}\left(y_2\right) = \frac{1}{2}\theta$, $\operatorname{Var}\left(y_1\right) = 3\sigma^2$, $\operatorname{Var}\left(y_2\right) = \sigma^2$, $\operatorname{Cov}\left(y_1, y_2\right) = \sigma^2$.

   (a) podać warunek jaki muszą spełniać a_1 i a_2, by estymator liniowy $\widehat{\theta} = a_1 y_1 + a_2 y_2$ był nieobciążony

(b) podać jakie powinny być a_1 i a_2, by estymator liniowy $\widehat{\theta}$ miał najniższą wariancję i był nieobciążony

(c) dla y_1 i y_2 mających rozkład normalny podać rozkład estymatora $\widehat{\theta}$

3. Mamy n wymiarowy wektor $\mathbf{x}$. Elementy tego wektora mają tę samą wartość oczekiwaną μ i wariancję σ^2 oraz są nieskorelowane.

   (a) Podać postać macierzy wariancji kowariancji $\mathbf{x}$

   (b) Udowodnić, że $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ jest nieobciążonym estymatorem μ

   (c) Pokazać, że wariancja $\overline{x}$ maleje, gdy N rośnie

   (d) (*) Pokazać, że estymator σ^2 postaci $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ jest nieobciążony

   Podpowiedź: możesz wykorzystać fakt, że $\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$, $\mathrm{E}(x_i x_j) = \mathrm{Cov}(x_i x_j) + \mathrm{E}(x_i)\mathrm{E}(x_j)$, $\left(\sum_{i=1}^{n} a_i\right)^2 = \sum_{i=1}^{n} a_i^2 + (n-1)\sum_{i \neq j} a_i a_j$

4. Mamy estymator $\widehat{\theta}$ parametru θ i oszacowanie jego błędu standardowego $se\left(\widehat{\theta}\right)$. Wiemy, że $\frac{\widehat{\theta}-\theta}{se\left(\widehat{\theta}\right)} \sim t_s$ gdzie s jest liczbą obserwacji. Dla $\widehat{\theta} = 1$, $se\left(\widehat{\theta}\right) = 0.5$, $s = 10$

   (a) zbudować 95% przedział ufności dla $\widehat{\theta}$

   (b) co się stanie z przedziałem ufności jeśli zamiast przedziału 95% policzymy przedział 90%?

   (c) co sie najprawdopodobniej stanie z przedziałem ufoności jeśli zwiększy się liczba obserwacji?

   (d) zweryfikować hipotezę, $H_0 = 0$ dla $\alpha = 0.05$

   Podpowiedź: wartość dystrybuanty $t_{10}(2) = 0.07$