Prawdopodobieństwo & Statystyka - rozwiązania.

Ekonometria WNE UW

Sebastian Zalas

s.zalas@uw.edu.pl

Proszę nie rozpowszechniać!

Rozwiązania wybranych zadań.

I. Rachunek prawdopodobieństwa.

Zadanie 1. Pokazać, że $\mathrm{Cov}(x_i, x_j) = \mathbb{E}(x_i, x_j) - \mathbb{E}(x_i)\,\mathbb{E}(x_j)$

Rozwiązanie: skorzystajmy z definicji kowariancji:

$$\mathrm{Cov}(x_i, x_j) = (\mathbb{E}(x_i - \mathbb{E}(x_i)))(\mathbb{E}(x_j - \mathbb{E}(x_j)))$$

$$\mathrm{Cov}(x_i, x_j) = \mathbb{E}(x_i x_j - x_i\,\mathbb{E}(x_j) - x_j\,\mathbb{E}(x_i) + \mathbb{E}(x_j)\,\mathbb{E}(x_i))$$

$$\mathrm{Cov}(x_i, x_j) = \mathbb{E}(x_i x_j) - \mathbb{E}(x_i)\,\mathbb{E}(x_j) - \mathbb{E}(x_j)\,\mathbb{E}(x_i) + \mathbb{E}(x_j)\,\mathbb{E}(x_i)$$

$$\mathrm{Cov}(x_i, x_j) = \mathbb{E}(x_i x_j) - \mathbb{E}(x_i)\,\mathbb{E}(x_j)$$

Zadanie 2. Pokazać, że jeśli $\mathbb{E}(x_i) = 0$ to $\mathbb{V}(x_i) = \mathbb{E}(x_i^2)$

Rozwiązanie: skorzystajmy z definicji wariancji:

$$\mathbb{V}(x_i) = \mathbb{E}((x_i - \mathbb{E}(x_i))^2)$$

$$\mathbb{V}(x_i) = \mathbb{E}(x_i^2 - 2x_i\,\mathbb{E}(x_i) - (\mathbb{E}(x_i))^2)$$

ponieważ $\mathbb{E}(x_i) = 0$

$$\mathbb{V}(x_i) = \mathbb{E}(x_i^2)$$

Zadanie 3. Które z macierzy mogą być macierzami kowariancji?

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

Rozwiązanie: Macierzami kowariancji mogą byc macierze $\boldsymbol{B}$ i $\boldsymbol{C}$, ponieważ są symetryczne.

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

$$\mathbb{E}[\boldsymbol{a} + \boldsymbol{B}\varepsilon] = \boldsymbol{a} + \boldsymbol{B}\,\mathbb{E}[\varepsilon]$$
$$\mathbb{V}[\boldsymbol{a} + \boldsymbol{B}\varepsilon] = \boldsymbol{B}\,\mathbb{V}[\varepsilon]\boldsymbol{B}'$$

Rozwiązanie:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{a} + \boldsymbol{B}\varepsilon] &= \underbrace{\boldsymbol{a}}_{\text{Wektor } \boldsymbol{a} \text{ zawiera tylko stałe, więc możemy opuścić wart. ocz.}} + \mathbb{E}[\boldsymbol{B}\varepsilon] \\
&= \boldsymbol{a} + \underbrace{\boldsymbol{B}}_{\text{Macierz } \boldsymbol{B} \text{ zawiera tylko stałe, więc również możemy opuścić wart. ocz.}} \mathbb{E}[\varepsilon] \\
\mathbb{V}[\boldsymbol{a} + \boldsymbol{B}\varepsilon] &= \underbrace{\mathbb{V}[\boldsymbol{B}\varepsilon]}_{\boldsymbol{a} \text{ - to wektor stałych, możemy go pominąć.}} \\
&= \underbrace{\mathbb{E}[(\boldsymbol{B}\varepsilon - \mathbb{E}[\boldsymbol{B}\varepsilon])(\boldsymbol{B}\varepsilon - \mathbb{E}[\boldsymbol{B}\varepsilon])']}_{\text{Skorzystajmy z definicji wariancji dla } \boldsymbol{B}\varepsilon} \\
&= \mathbb{E}[\boldsymbol{B}(\varepsilon - \mathbb{E}[\varepsilon])(\boldsymbol{B}(\varepsilon - \mathbb{E}[\varepsilon]))'] \\
&= \mathbb{E}[\boldsymbol{B}(\varepsilon - \mathbb{E}[\varepsilon])((\varepsilon - \mathbb{E}[\varepsilon])'\boldsymbol{B}'] \\
&= \boldsymbol{B}\,\underbrace{\mathbb{E}[(\varepsilon - \mathbb{E}[\varepsilon])((\varepsilon - \mathbb{E}[\varepsilon])']}_{\text{definicja waraincji dla } \varepsilon}\boldsymbol{B}' \\
&= \boldsymbol{B}\,\mathbb{V}[\varepsilon]\boldsymbol{B}'
\end{aligned}$$

Zadanie 5. Mamy wektor losowy $\boldsymbol{x}$, przy czym $\mathbb{E}[\boldsymbol{x}] = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, $\mathbb{V}[\boldsymbol{x}] = \left[\begin{smallmatrix}1 & 2\\2 & 5\end{smallmatrix}\right]$. Policz wartość oczekiwaną i wariancję $\boldsymbol{y} = \left[\begin{smallmatrix}x_1+2x_2+5\\x_1+x_2+1\end{smallmatrix}\right]$. Rozwiązanie: Zauważmy najpierw że wektor losowy $\boldsymbol{y}$ to przekształcenie liniowe wektora losowego $\boldsymbol{x}$:

$$\boldsymbol{y} = \begin{bmatrix}1 & 2\\1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}5\\1\end{bmatrix}$$
$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{a}$$

Aby obliczyć wartość oczekiwaną i wariancję $\boldsymbol{y}$, wystarczy zastosować wzory udowodnione w zadaniu 5.

Zadanie 6. Mamy wektor losowy $\boldsymbol{x} = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$, przy czym $\mathbb{E}[\boldsymbol{x}] = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, $\mathbb{V}[\boldsymbol{x}] = \left[\begin{smallmatrix}1 & 2\\2 & 5\end{smallmatrix}\right]$.
Rozwiązanie:

(a) odchylenie standardowe x_1, x_2 - korzystamy z macierzy wariancji-kowariancji:

$$\sigma[x_1] = \sqrt{\mathbb{V}[x_1]} = 1$$
$$\sigma[x_2] = \sqrt{\mathbb{V}[x_5]} = \sqrt{5}$$

(b) współczynnik korelacji między x_1, x_2 - używamy wzoru na wsp. korelacji:

$$\rho_{x_1,x_2} = \frac{\mathrm{Cov}[x_1, x_2]}{\sigma[x_1]\sigma[x_2]} = \frac{2}{2\sqrt{5}}$$

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

$$\begin{aligned}
\mathbb{E}[y] &= \mathbb{E}[5 + x_1 + 2x_2] = 5 + \mathbb{E}[x_1] + 2\,\mathbb{E}[x_2] \\
&= 5 + 1 + 4 = 10 \\
\mathbb{V}[y] &= \mathbb{V}[5 + x_1 + 2x_2] \\
&= \mathbb{V}[x_1 + 2x_2] \\
&= \mathrm{Cov}[x_1, x_1] + \mathrm{Cov}[x_1, 2x_2] + \mathrm{Cov}[2x_2, x_1] + \mathrm{Cov}[2x_2, 2x_2] \\
&= \mathbb{V}[x_1] + 2\,\mathrm{Cov}[x_1, x_2] + 2\,\mathrm{Cov}[x_2, x_1] + 4\,\mathbb{V}[x_2] \\
&= 1 + 4 + 4 + 20 = 29
\end{aligned}$$

Zadanie 7. Udowodnić, że dla dowolnej macierzy losowej $\boldsymbol{A}$: $\mathbb{E}[tr(\boldsymbol{A})] = tr[\mathbb{E}(\boldsymbol{A})]$.

Rozwiązanie: Zacznijmy od definicji śladu:

$$\begin{aligned}
\mathrm{E}\left[\mathrm{tr}(\boldsymbol{A})\right] &= \mathrm{E}\left[\sum_{i=1}^{n} a_{ii}\right] \\
&= \sum_{i=1}^{n} \mathrm{E}\left[a_{ii}\right] \\
&= \mathrm{tr}\left(\begin{bmatrix} \mathrm{E}[a_{11}] & \dots & \mathrm{E}[a_{1n}] \\ \vdots & \ddots & \vdots \\ \mathrm{E}[a_{n1}] & \dots & \mathrm{E}[a_{nn}] \end{bmatrix}\right) \\
&= \mathrm{tr}\left(\mathrm{E}[\boldsymbol{A}]\right) \ .
\end{aligned}$$

Zadanie 8. Załóżmy, że $\mathbb{E}(x) > 0$. Jaka jest relacja między $\mathbb{E}(x)$ i $\mathbb{E}(\frac{1}{x})$?

Rozwiązanie: Nierówność Jensena. Dla wypukłej funkcji $\phi(.)$ zachodzi:

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$$

Zadanie 9. Załóżmy że Y i X są zmiennymi losowymi. Czemu jest równe $\mathbb{E}(\frac{Y}{X}|X)$?

Rozwiązanie:

$$\begin{aligned}
\mathbb{E}(\frac{Y}{X}|X) &= \mathbb{E}(\frac{Y}{x}|X = x) \\
&= \frac{1}{x}\,\mathbb{E}(Y|X = x)
\end{aligned}$$

Zadanie 10. Wiemy, że $\mathbb{E}(x) = 2$ oraz $\mathbb{E}(y|x) = 1 + 2x$. Czemu równe jest $\mathbb{E}(y)$?

Rozwiązanie: Skorzystaj z wzoru na całkowitą wartość oczekiwaną (LIE *law of iterated expecatations*)

$$\mathbb{E}(y) = \mathbb{E}(\mathbb{E}(y|x))$$

II. Własności rozkładu normalnego

Zadanie 2. Pokazać, że dla k-wymiarowego wektora losowego $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, forma kwadratowa $\boldsymbol{\varepsilon}'\mathbf{\Sigma}\boldsymbol{\varepsilon} \sim \chi^2$

Rozwiązanie: Wyznaczmy formę kwadratową, załóżmy że wielowymiarowy rozkład normalny jest standardowy:

$$\begin{aligned}
\boldsymbol{\varepsilon}'\mathbf{\Sigma}\boldsymbol{\varepsilon} &= \begin{bmatrix} \varepsilon_1 & \cdots & \varepsilon_k \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{k \times k} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix} \\
&= \sum_{i=1}^{k} \varepsilon_i^2 \sim \chi_k^2
\end{aligned}$$

Widzimy, że forma kwadratowa $\boldsymbol{\varepsilon}'\mathbf{\Sigma}\boldsymbol{\varepsilon}$ to suma kwadratów k zmiennych losowych o rozkładzie normalnym; taka suma ma rozkład χ^2 o k stopniach swobody.

Zadanie 3. Mamy wektor losowy $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, gdzie $\boldsymbol{\mu} = \left[\begin{smallmatrix} 1 \\ 3 \\ 2 \end{smallmatrix}\right]$, $\mathbb{V}[\boldsymbol{x}] = \left[\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{smallmatrix}\right]$. Jaki rozkład ma zmienna losowa $v = x_1 + 2x_2 + x_3$?

Rozwiązanie: Zmienna losowa v ma rozkład normalny, ponieważ suma zmiennych losowych o rozkładzie normalnym, ma rozkład normalny. Policzmy parametry rozkładu:

$$\begin{aligned}
\mathbb{E}[v] &= \mathbb{E}[x_1 + 2x_2 + x_3] = \mathbb{E}[x_1] + 2\,\mathbb{E}[x_2] + \mathbb{E}[x_3] = 1 + 6 + 2 = 9 \\
\mathbb{V}[v] &= \mathrm{Cov}[x_1 + 2x_2 + x_3, x_1 + 2x_2 + x_3] \\
&= \mathbb{V}[x_1] + 4\,\mathbb{V}[x_2] + \mathbb{V}[x_3] + 2\,\mathrm{Cov}[x_1, 2x_2] + 2\,\mathrm{Cov}[2x_2, x_3] + 2\,\mathrm{Cov}[x_1, x_3] \\
&= 1 + 8 + 2 + 4 + 4 + 1 = 20
\end{aligned}$$

Zatem $v \sim \mathcal{N}(9, 20)$.

Zadanie 4. Mamy wektor losowego $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, gdzie $\boldsymbol{\mu} = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$, $\mathbb{V}[\boldsymbol{x}] = \left[\begin{smallmatrix} 5 & 3 \\ 3 & 2 \end{smallmatrix}\right]$.

(i). Pokazać, że wektor $\boldsymbol{v} = \left[\begin{smallmatrix} x_1 - x_2 + 1 \\ -x_1 + 2x_2 - 4 \end{smallmatrix}\right]$ ma rozkład $\boldsymbol{v} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.

Rozwiązanie: Pokażmy że $\boldsymbol{v}$ jest przekształceniem liniowym $\boldsymbol{x}$:

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$\boldsymbol{v} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}$$

Możemy skorzystać ze wzorów z zadania 4, z pierwszej części zadań (Rachunek Prawdopodobieństwa).

$$\begin{aligned} \mathbb{E}[\boldsymbol{v}] &= \boldsymbol{a} + \boldsymbol{B}\,\mathbb{E}[\boldsymbol{x}] \\ &= \begin{bmatrix} 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \mathbb{V}[\boldsymbol{v}] &= \boldsymbol{B}\,\mathbb{V}[\boldsymbol{x}]\boldsymbol{B}' \\ &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

Istotnie, $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

(ii). Pokazać, że $(x_1 - x_2 + 1)^2 + (-x_1 + 2x_2 - 4)^2 \sim \chi^2_2$

Rozwiązanie: Zauważmy że zmienne losowe $x_1 - x_2 + 1$ oraz $-x_1 + 2x_2 - 4$ są sumami zmiennych losowych o rozkładzie normalnym, a więc one także mają rozkład normalny Z parametrami $\mathcal{N}(0, 1)$ (Parametry proszę sprawdzić samodzielnie!). Zatem zmienna losowa $(x_1 - x_2 - 1)^2 + (-x_1 + 2x_2 + 4)^2$ musi mieć rozkład χ^2 o dwóch stopniach swobody, ponieważ jest sumą dwóch kwadratów zmiennych losowych o standardowym rozkładzie normalnym.

Uwaga! W stosunku do oryginalnej wersji zadania zamieniłem znaki w stałej $\boldsymbol{a}$!.

III. Statystyka

Zadanie 2. Mamy zmienne losowe y_1 i y_2 takie, że $\mathbb{E}(y_1) = \theta$, $\mathbb{E}(y_2) = \frac{1}{2}\theta$, $\mathbb{V}(y_1) = 3\sigma^2$, $\mathbb{V}(y_2) = \sigma^2$, $\mathrm{Cov}(y_1, y_2) = \sigma^2$.

Rozwiązanie:

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

Rozpocznijmy od definicji obciążenia:

$$bias(\hat{\theta}, \theta) = \mathbb{E}(\theta) - \hat{\theta}$$

Aby estymator był nieobciążony, musi spełniać warunek:

$$\mathbb{E}(\hat{\theta}) - \theta = 0$$

w naszym przypadku:

$$\begin{aligned}
\mathbb{E}(a_1 y_1 + a_2 y_2) - \theta &= 0 \\
a_1 \mathbb{E}(y_1) + a_2 \mathbb{E}(y_2) - \theta &= 0 \\
a_1 \theta + a_2 \frac{1}{2} \theta - \theta &= 0 \\
(a_1 + a_2 \frac{1}{2} - 1)\theta &= 0 \\
a_1 + a_2 \frac{1}{2} - 1 &= 0 \qquad (1)
\end{aligned}$$

(ii) Podać jakie powinny być a_1 i a_2, by estymator liniowy $\hat{\theta}$ miał najniższą wariancję i był nieobciążony.

Wariancja estymatora:

$$\begin{aligned}
\mathbb{V}(\hat{\theta}) &= \mathbb{V}(a_1 y_1 + a_2 y_2) \\
\mathbb{V}(\hat{\theta}) &= \mathbb{V}(a_1 y_1) + \mathbb{V}(a_2 y_2) + \mathrm{Cov}(a_1 y_1, a_2 y_2) \\
\mathbb{V}(\hat{\theta}) &= a_1^2 \mathbb{V}(y_1) + a_2^2 \mathbb{V}(y_2) + a_1 a_2 \mathrm{Cov}(y_1, y_2) \\
\mathbb{V}(\hat{\theta}) &= a_1^2 3\sigma^2 + a_2^2 \sigma^2 + a_1 a_2 \sigma^2 \qquad (2)
\end{aligned}$$

Aby znaleźć nieobciążony estymator o najniższej wariancji, należy zminimalizować (2) z ograniczeniem (1).

$$(3)$$

(iii) Dla y_1 i y_2 mających rozkład normalny podać rozkład estymatora $\bar{\theta}$.

Suma dwóch zmiennych losowych o rozkładzie normalnym, ma rozkład normalny:

$$X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y})$$

$$aX + bY \sim N(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{X,Y})$$

Czyli estymator $\hat{\theta}$ ma rozkład normalny z parametrami:

$$\begin{aligned}
\mathbb{E}(\hat{\theta}) &= a_1 \theta + a_2 \frac{1}{2} \theta \\
\mathbb{V}(\hat{\theta}) &= 3a_1^2 \sigma^2 + a_2^2 \sigma^2 + 2a_1 a_2 \sigma^2
\end{aligned}$$

Zadanie 3. Mamy n wymiarowy wektor **x**. Elementy tego wektora mają tę samą wartość oczekiwaną μ i wariancję σ^2 oraz są nieskorelowane.

Rozwiązanie:

(i) Podać postać macierzy wariancji-kowariancji **x**.

Macierz wariancji-kowariancji:

$$\mathbb{V}(\mathbf{x}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

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

Aby $\bar{x}$ była niobciążonym estymatorem μ, musi zachodzić warunek:

$$\begin{aligned} \mathbb{E}(\bar{x}) &= \mu \\ \mathbb{E}(\bar{x}) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(x_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu \\ &= \frac{1}{n} n\mu \\ &= \mu \end{aligned}$$

(iii) Pokazać że wariancja $\bar{x}$ maleje, gdy N rośnie.

Innymi słowy, mamy sprawdzić czy $\bar{x}$ jest zgodnym estymatorem. Wyznaczmy wariancję $\bar{x}$:

$$\begin{aligned} \mathbb{V}(\bar{x}) &= \mathbb{V}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) \\ &= \frac{1}{N^2}\mathbb{V}\left(\sum_{i=1}^{N} x_i\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{V}(x_i) \\ &= \frac{1}{N^2} N\sigma^2 \\ &= \frac{1}{N}\sigma^2 \end{aligned}$$

Przy rosnącym N, wariancja $\bar{x}$ maleje:

$$\lim_{i\to N} \frac{\sigma^2}{N} = 0$$

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

Innymi słowy, mamy sprawdzić czy zachodzi następujący warunek:

$$\mathbb{E}(s^2) = \sigma^2$$

Zapiszmy wariancję z próby jeszcze raz:

$$\begin{aligned}
s^2 &= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\left(\sum_{n=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2)\right) \\
&= \frac{1}{n}\sum_{n=1}^{n}x_i^2 - 2\bar{x}\frac{1}{n}\sum_{n=1}^{n}x_i + \bar{x}^2
\end{aligned}$$