

Algorytmiczna analiza danych 2026/2027

Lista ćwiczeń

Używam następujących skrótów:

1. ISL = [An Introduction to Statistical Learning](#) autorstwa G. Jamesa i in.
2. ESL = [The Elements of Statistical Learning](#) autorstwa T. Hastie i in.
3. PAC = [Probability and computing](#) autorstwa M. Mitzenmacher i in.
4. ...

Ćwiczenie 1 — Załóżmy, że zmienne losowe X i Y są niezależne. Udowodnij, że

- (a) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
- (b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Czy warunek o niezależności jest niezbędny?

Ćwiczenie 2 — Zmienna X ma rozkład dyskretny:

x	1	2	3	4
$P(X = x)$	0,1	0,3	0,4	0,2

- (a) Narysuj dystrybuantę $F(x)$.
- (b) Oblicz $\mathbb{E}(X)$, $\text{Var}(X)$.
- (c) Znajdź medianę i kwartyle.

Ćwiczenie 3 — Niech $X \sim U(1, 2)$ będzie zmienną losową o rozkładzie jednostajnym na $[1, 2]$.

- (a) Zapisz wzorem i naszkicuj gęstość $f(x)$ i dystrybuantę $F(x)$.
- (b) Oblicz $\mathbb{E}(X)$ i kwartyle.
- (c) Oblicz $\text{Var}(X)$, $\mathbb{SD}(X)$ oraz $IQR = Q_3 - Q_1$.

Ćwiczenie 4 — Niech $X \sim \text{Exp}(\lambda)$ ma rozkład wykładniczy z parametrem $\lambda > 0$.

- (a) Zapisz wzorem i naszkicuj gęstość $f(x)$ i dystrybuantę $F(x)$. Jaki jest wpływ parametru λ na kształt wykresów? Jaka jest kluczowa własność rozkładu wykładniczego?
- (b) Oblicz $\mathbb{E}(X)$ i kwartyle.
- (c) Podaj wzór na $\text{Var}(X)$, $\mathbb{SD}(X)$ oraz IQR .

Ćwiczenie 5 — Podaj przykład zmiennych losowych zależnych X i Y , dla których $\text{Cov}(X, Y) = 0$.

Ćwiczenie 6 — Niech $X_1, \dots, X_n$ będą i.i.d. z rozkładu o średniej $\mu = \mathbb{E}(X_i)$ i wariancji $\sigma^2 = \text{Var}(X_i)$.

- (a) Pokaż, że średnia z próby $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ jest nieobciążonym estymatorem μ , tzn. $\mathbb{E}(\bar{X}) = \mu$.
- (b) Oblicz $\text{Var}(\bar{X})$.
- (c) Pokaż, że $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ jest nieobciążonym estymatorem σ^2 , tzn. $\mathbb{E}(s^2) = \sigma^2$.

Ćwiczenie 7 — Niech $X_1, \dots, X_n$ będą i.i.d. z dystrybuantą F . Definiujemy empiryczną dystrybuantę

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x) .$$

- (a) Zauważ, że $Y_i = \mathbb{I}(X_i \leq x) \sim \text{Bernoulli}(p)$ gdzie $p = P(X_i \leq x) = F(x)$. Jaka jest wartość oczekiwana i wariancja Y_i ?
- (b) Pokaż, że: $\mathbb{E}(\hat{F}_n(x)) = F(x)$.

(c) Oblicz $\text{Var}(\hat{F}_n(x))$. Dla jakiej wartości $F(x)$ wariancja jest największa?

Ćwiczenie 8 — Dana jest próba cen mieszkań (w tys. zł): $x = (280, 320, 350, 410, 450, 520, 600, 750)$.

- (a) Narysuj empiryczną dystrybuantę $\hat{F}_8(t)$.
- (b) Oblicz średnią $\bar{x}$ i odchylenie standardowe s .
- (c) Znajdź medianę, Q_1 , Q_3 i IQR.
- (d) Dodajemy penthouse za 2500 tys. zł (nowa próba ma $n = 9$). Oblicz nowe $\bar{x}$, s , medianę, IQR. Które statystyki zmieniły się najbardziej?

Ćwiczenie 9 — Dla każdej zmiennej podaj jej typ (np. nominalna, porządkowa, dyskretna, ciągła, szereg czasowy): cena mieszkania, dzielnica, piętro, ocena stanu (zły/średni/dobry), rok budowy, średnia cena m² w kolejnych kwartałach. Jaka jest kluczowa różnica między zmiennymi kategorialnymi a liczbowymi?

Ćwiczenie 10 — Wygeneruj próbę 100 liczb z rozkładu $N(700, 100^2)$ reprezentującego ceny mieszkań w tys. zł.

- (a) Oblicz $\bar{x}$, s , medianę, Q_1 , Q_3 , IQR.
- (b) Narysuj histogram (z 10 i 30 przedziałami – porównaj).
- (c) Narysuj empiryczną dystrybuantę.
- (d) Narysuj wykres pudełkowy.
- (e) Dodaj 5 outlierów (np. $x = 1500$) i powtórz obliczenia. Co się zmieniło?

Ćwiczenie 11 — Narysuj histogram licznosci dla danych o czasach obsługi klienta z następującymi przedziałami czasowymi (czas dany w minutach) i licznosciami (łączna liczba obserwacji $n = 120$):

Przedział	[0, 2)	[2, 3)	[3, 4)	[4, 7)
Liczność	40	30	20	30

- (a) Dlaczego taki histogram może wprowadzać w błąd?
- (b) Narysuj histogram gęstości, w którym wysokość j -tego słupka wynosi $w_j = \frac{n_j}{n \cdot s_j}$, gdzie n_j to licznosc, a s_j szerokosc j -tego przedziału. Jak zmienia to interpretację rozkładu?

Ćwiczenie 12 — Niech X, Y będą zmiennymi losowymi o skończonych wariancjach $\text{Var}(X), \text{Var}(Y) > 0$.

- (a) Udowodnij, że współczynnik korelacji Pearsona spełnia $|\rho(X, Y)| \leq 1$.
Wskazówka: Skorzystaj z nierówności Cauchy'ego–Schwarza: $(\mathbb{E}(UV))^2 \leq \mathbb{E}(U^2) \mathbb{E}(V^2)$.
- (b) Korzystając z definicji $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$, udowodnij następujące własności:
 - (i) $\text{Cov}(X, aY + b) = a \text{Cov}(X, Y)$ dla $a, b \in \mathbb{R}$,
 - (ii) $\text{Cov}(X, b) = 0$ dla stałej $b \in \mathbb{R}$,
 - (iii) $\text{Var}(aX + b) = a^2 \text{Var}(X)$ dla $a, b \in \mathbb{R}$.
- (c) Pokaż, że jeśli $Y = aX + b$ dla pewnych $a \neq 0, b \in \mathbb{R}$, to $|\rho(X, Y)| = 1$. Jaki jest związek między znakiem a a znakiem ρ ?

Ćwiczenie 13 — Niech $\rho(X, Y) = \rho$, $U = aX + b$, $V = cY + d$, gdzie $a, c \neq 0, b, d \in \mathbb{R}$.

- (a) Wyraż $\rho(U, V)$ jako funkcję ρ, a, c . Czy b i d mają znaczenie?
- (b) Jakie są konsekwencje praktyczne tej własności? (np. dla standaryzacji danych)
- (c) Czy $\rho(X^2, Y^2) = \rho(X, Y)^2$ dla dowolnych X, Y ? Uzasadnij lub podaj kontrprzykład.

Ćwiczenie 14 — Dane (8 par obserwacji):

x_i	1	2	3	4	5	6	7	100
y_i	2	4	5	4	5	7	8	9

- (a) Oblicz empiryczny współczynnik korelacji Pearsona r_P i Spearmana r_S dla pełnych danych.
- (b) Usuń obserwację (100, 9) i oblicz r_P oraz r_S ponownie.

(c) Porównaj wyniki z (a) i (b). Która miara jest bardziej odporna na wartości odstające i dlaczego?

Ćwiczenie 15 — Niech $X_1, \dots, X_n$ będą i.i.d. z rozkładu $N(\mu, \sigma^2)$, gdzie $\mu \in \mathbb{R}$ i $\sigma^2 > 0$ są nieznanymi. Metodą momentów wyprowadź estymatory $\hat{\mu}$ i $\hat{\sigma}^2$. Czy te estymatory są nieobciążone? Jeśli nie, zaproponuj wersje nieobciążone.

Ćwiczenie 16 — Niech $X_1, \dots, X_n$ będą i.i.d. z rozkładu wykładniczego $\text{Exp}(\lambda)$, $\lambda > 0$.

(a) Metodą momentów wyprowadź estymator $\hat{\lambda}_n$ parametru λ .

(b) Oblicz $\mathbb{E}(\hat{\lambda}_n)$ i uzasadnij, że $\hat{\lambda}_n$ jest asymptotycznie nieobciążony.

Wskazówka: Skorzystaj z faktu, że $S_n = \sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$ i oblicz $\mathbb{E}(1/S_n)$.

(c) Zaproponuj nieobciążony estymator parametru λ .

Ćwiczenie 17 — Dla $n > 2$ oblicz wariancję nieobciążonego estymatora λ wyprowadzonego w poprzednim zadaniu. Wskazówka: Wyznacz $\mathbb{E}(S_n^{-2})$ analogicznie do $\mathbb{E}(S_n^{-1})$.

cdn.

J.L.