Kovariance a korelace

13.11.2019

1. V šuplíku je 6 ponožek: 2 bílé, 2 černé a 2 oranžové. Potmě náhodně vytáhneme z šuplíku 3 ponožky. Označme jako X počet vytažených bílých ponožek a Y je počet vytažených černých ponožek.
   (a) Napište tabulku rozdělení náhodného vektoru $(X, Y)^\top$.
   (b) Jsou X a Y nezávislé?
   (c) Spočtěte kovarianci a korelaci X a Y.
   (d) Spočtěte rozdělení, střední hodnotu a rozptyl náhodné veličiny $U = X - Y$.
   (e) Nechť Z značí počet černých ponožek, které zůstaly v šuplíku. Určete pravděpodobnostní rozdělení $(X, Z)^\top$ a kovarianci X a Z.

2. Náhodný vektor $(X, Y)^\top$ má spojité rozdělení charakterizované sdruženou hustotu
$$f(x,y) = \begin{cases} cxy & \text{pro } 0 < x < 1,\ 0 < y < 1, \\ 0 & \text{jinak.} \end{cases}$$
   (a) Určete $c > 0$ tak, aby f byla hustota.
   (b) Spočtěte kovarianci $\mathsf{Cov}\,(X, Y)$.
   (c) Rozhodněte, zda jsou X a Y nezávislé,

3. Náhodná veličina X má rovnoměrné rozdělení na intervalu $(-1, 1)$. Označme $Y = X^2$. Spočtěte kovarianci veličin X a Y a jejich korelační koeficient ρ_{XY}. Jsou X a Y nezávislé?

4. Náhodný vektor $(X, Y)^\top$ má rovnoměrné rozdělení na množině $M = \{(x, y) : 0 \le y \le 1 - x, \quad 0 \le x \le 1\}$, tj. spojité rozdělení s hustotou
$$f(x,y) = \begin{cases} c & \text{pro } 0 \le y \le 1 - x, \quad 0 \le x \le 1, \\ 0 & \text{jinak ,} \end{cases}$$
   (a) Určete konstantu $c > 0$, aby f byla hustota.
   (b) Rozhodněte, zda jsou X a Y nezávislé.
   (c) Spočtěte kovarianci a korelaci X a Y.

5. Řešte předchozí příklad pro množinu $M = \{(x, y) : x - 1 \le y \le 1 - x, \quad 0 \le x \le 1\}$.

6. Uvažujme posloupnost n znaků, ve které se každý znak zkreslí s pravděpodobností $p \in (0, 1)$, a to nezávisle na ostatních znacích. Nechť X značí počet takto zkreslených znaků. Dále označme jako Y_i identifikátor toho, zda se i-tý znak zkreslí.Vyjádřete X pomocí veličin $Y_1, \dots, Y_n$ a pomocí tohoto vyjádření spočtěte $\mathsf{E}X$ a $\mathsf{Var}\,X$.

Opakování

Kovariance a korelace: Kovariance $\mathsf{Cov}\,(X,Y)$ náhodných veličin X,Y je definována jako

$$\mathsf{Cov}\,(X,Y) = \mathsf{E}(X-\mathsf{E}X)(Y-\mathsf{E}Y) = \mathsf{E}(XY) - (\mathsf{E}X)(\mathsf{E}Y),$$

je-li $\mathsf{E}X^2 < \infty$, $\mathsf{E}Y^2 < \infty$.
Korelace $\mathsf{Corr}\,(X,Y)$ je definována jako

$$\rho_{X,Y} = \mathsf{Corr}\,(X,Y) = \frac{\mathsf{Cov}\,(X,Y)}{\sqrt{\mathsf{Var}\,X}\sqrt{\mathsf{Var}\,Y}},$$

je-li $\mathsf{Var}\,X, \mathsf{Var}\,Y > 0$. Platí vždy $-1 \leq \mathsf{Corr}\,(X,Y) \leq 1$. Korelace je mírou lineární závislosti mezi X a Y.

Užitečné vlastnosti: Nechť X,Y jsou náhodné veličiny a $a,b,c,d \in \mathbb{R}$. Pak (za předpokladu existence daných momentů)

- $\mathsf{Cov}\,(aX+c, bY+d) = ab\mathsf{Cov}\,(X,Y)$,
- $\mathsf{E}(aX+bY) = a\mathsf{E}X + b\mathsf{E}Y$,
- $\mathsf{Var}\,(aX+bY) = a^2\mathsf{Var}\,X + b^2\mathsf{Var}\,Y + 2ab\mathsf{Cov}\,(X,Y)$.

Obecně, jsou-li $X_1,\ldots,X_n$ jsou náhodné veličiny a $a_1,\ldots,a_n \in \mathbb{R}$, pak platí (za předpokladu existence daných momentů)

- $\mathsf{E}(\sum_{i=1}^n a_iX_i) = \sum_{i=1}^n a_i\mathsf{E}X_i$
- $\mathsf{Var}\,(\sum_{i=1}^n a_iX_i) = \sum_{i=1}^n a_i^2\mathsf{Var}\,X_i + \sum\sum_{i\neq j} a_ia_j\mathsf{Cov}\,(X_i,X_j)$