Kovariance a korelace

14.11.2018

1. Náhodná veličina X udává počet dcer v náhodně vybrané rodině se třemi dětmi, veličina Y udává počet starších bratrů nejmladšího dítěte v téže rodině.
   (a) Spočítejte kovarianci X a Y.
   (b) Spočtěte korelaci X a Y.

2. Náhodná veličina X udává počet dcer v náhodně vybrané rodině se třemi dětmi, veličina Z udává počet mladších sester nejstaršího dítěte v téže rodině.
   (a) Spočítejte kovarianci X a Z.
   (b) Spočtěte korelaci X a Z.

3. Náhodný vektor $(X, Y)'$ 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,
   (d) Napište varianční matici vektoru $(X, Y)^\top$.
   (e) Spočtěte $E(X + Y)$ a $\mathsf{Var}\,(X + Y)$.

4. 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?

5. Náhodný vektor $(X, Y)'$ má rovnoměrné rozdělení na jednotkovém kruhu, tj. spojité rozdělení s hustotou

$$f(x,y) = \begin{cases} c & \text{pro } x^2 + y^2 \le 1 \\ 0 & \text{jinak .} \end{cases}$$

   (a) Určete konstantu $c > 0$, aby f byla hustota.
   (b) Spočtěte kovarianci a korelaci X a Y.
   (c) Jsou X a Y nezávislé?

6. Uvažujme n nezávislých hodů pravidelnou kostkou a nechť X značí počet pokusů, kdy padla šestka. Dále označme jako náhodnou veličinu Y_i identifikátor toho, zda v i-tém kole padla šestka. Vyjádřete X pomocí veličin $Y_1, \ldots, Y_n$ a pomocí tohoto vyjádření spočtěte $\mathsf{E}X$ a $\mathsf{Var}\,X$.

7. V rybníce plave 100 rybiček, z nichž je a zlatých. Náhodně vylovíme 5 rybiček a náhodná veličina X udává, kolik zlatých rybiček jsme vylovili. Spočtěte očekávanou hodnotu X. *Nápověda: Použijte podobný rozklad na součet 0-1 náhodných veličin jako v příkladu 6.*

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 $-1 \leq \mathsf{Corr}\,(X,Y) \leq 1$. Korelce je mírou lineární závislosti mezi X a Y.

Varianční matici náhodného vektoru $(X,Y)^\top$ značíme jako $\mathsf{Var}\,(X,Y)^\top$ a je tvaru

$$\begin{pmatrix} \mathsf{Var}\,X & \mathsf{Cov}\,(X,Y) \\ \mathsf{Cov}\,(X,Y) & \mathsf{Var}\,Y \end{pmatrix}.$$

Další užitečné vlastnosti: Jestliže $X_1, \dots, X_n$ jsou náhodné veličiny a $a_1, \dots, a_n \in \mathbb{R}$, pak platí (za předpokladu existence daných momentů)

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