1 Rozptyl a kovariance

Nechť X je náhodná veličina s konečnou střední hodnotou EX. Potom *rozptyl* náhodné veličiny X definujeme jako:

$$DX = E(X - EX)^2 \,,$$

pokud střední hodnota na pravé straně existuje. Podobně jako střední hodnota náhodné veličiny v jistém smyslu popisovala její rozdělení pravděpodobnosti (např. v případě diskrétní náhodné veličiny to byl vlastně vážený průměr všech hodnot, kterých náhodná veličina nabývala), vypovídá i rozptyl cosi o tvaru rozdělení pravděpodobnosti náhodné veličiny. Jak je z definice rozptylu patrno, rozptyl udává, jak moc náhodná veličina X kolísá kolem střední hodnoty EX, protože počítá střední hodnotu z kvadratických odchylek X od EX. Odmocnina z rozptylu $\sqrt{DX}$ se nazývá *směrodatná odchylka*. Někdy se rozptyl značí jako $\mathrm{var}X$.

Věta 1.1 *Pro náhodnou veličinu X s konečným rozptylem a libovolná reálná čísla a, b platí:*

$$\begin{aligned} DX &= E(X^2) - (EX)^2 \,, & (1) \\ D(a + bX) &= b^2 DX \,. & (2) \end{aligned}$$

Vztah (1) je pro výpočet rozptylu obvykle výhodnější než definice.

Příklad 1.2 Nechť X je spojitá náhodná veličina s hustotou pravděpodobnosti:

$$f(x) = \begin{cases} 3x^2 \,, & x \in (0,1) \,, \\ 0 \,, & \text{jinak} \,. \end{cases}$$

Spočtěte její střední hodnotu EX a rozptyl DX.

Z definice střední hodnoty dostáváme:

$$EX = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \int_0^1 x 3x^2 \mathrm{d}x = 3 \int_0^1 x^3 \mathrm{d}x = 3 \left[\frac{x^4}{4} \right]_0^1 = \frac{3}{4} \,.$$

Z věty o střední hodnotě transformované náhodné veličiny máme:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x = \int_0^1 x^2 3x^2 \mathrm{d}x = 3 \int_0^1 x^4 \mathrm{d}x = 3 \left[\frac{x^5}{5} \right]_0^1 = \frac{3}{5} \,.$$

A nakonec z věty 1.1 dostaneme:

$$DX = E(X^2) - (EX)^2 = \frac{3}{5} - \left(\frac{3}{4} \right)^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80} \,.$$

O závislosti dvou náhodných veličin do jisté míry vypovídá pojem *kovariance*, který se definuje následovně:

$$\mathrm{cov}(X,Y) = E\left[(X - EX)(Y - EY)\right] .$$

Zřejmě platí $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ a $\mathrm{cov}(X,X) = DX$. Podobně jako u rozptylu je možné ukázat, že $\mathrm{cov}(X,Y) = E(XY) - (EX)(EY)$.

Věta 1.3 *Pro náhodné veličiny X, Y, jejichž rozptyly existují, platí:*

$$D(X+Y) = DX + DY + 2\mathrm{cov}(X,Y) .$$

Pokud jsou navíc X a Y nezávislé, platí:

$$D(X+Y) = DX + DY .$$

Všimeme si, že z předchozí věty plyne, že pokud jsou X a Y nezávislé, pak $\mathrm{cov}(X,Y) = 0$. Upozorňujeme však, že obrácené tvrzení neplatí (tzn. nulovost $\mathrm{cov}(X,Y)$ nezaručuje nezávislost X a Y). Nicméně pokud $\mathrm{cov}(X,Y) \neq 0$, pak již X, Y musí být závislé.

Příklad 1.4 Určete střední hodnotu a rozptyl dikrétní náhodné veličiny X s binomickým rozdělením pravděpodobnosti s parametry n, p.

Připomeňme, že binomické rozdělení udává pravděpodobnosti počtů výskytů náhodného jevu A takového, že $P(A) = p$, pokud jsme sledovali jeho výskyt v n nezávislých pokusech. Binomické rozdělení je dáno vztahem:

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} , \quad k = 0, 1, \ldots, n .$$

Zaveďme nové náhodné veličiny Y_i udávající počet výskytů jevu A v i-tém pokusu. Náhodná veličina Y_i tedy může nabývat pouze dvou hodnot a to 0 pokud jev A v i-tém pokusu nenastal nebo 1 pokud nastal. Náhodná veličina X se potom dá vyjádřit jako:

$$X = \sum_{i=1}^{n} Y_i ,$$

protože udává celkový počet výskytů A během n pokusů.

Spočítejme střední hodnotu a rozptyl Y_i. Je zřejmé, že všechny náhodné veličiny $Y_1, Y_2, \ldots, Y_n$ budou mít stejné rozdělení pravděpodobnosti. Konkrétně:

$$P[Y_i = 0] = 1 - p , \quad P[Y_i = 1] = p .$$

Střední hodnota je tedy:

$$EY_i = 0(1-p) + 1p = p .$$

Dále

$$E(Y_i^2) = 0^2(1-p) + 1^2 p = p .$$

Rozptyl tedy je:

$$DY_i = E(Y_i^2) - (EY_i)^2 = p - p^2 = p(1-p)\,.$$

Protože platí $E(X+Y) = EX + EY$, dostáváme pro střední hodnotu X:

$$EX = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} EY_i = \sum_{i=1}^{n} p = np\,.$$

Protože jsou jednotlivé pokusy nezávislé, jsou nezávislé i Y_i. Z věty 1.3 pak dostaneme, že $D(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} DY_i$, a proto platí:

$$DX = D\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} DY_i = \sum_{i=1}^{n} p(1-p) = np(1-p)\,.$$

Zobecněním pojmu rozptylu pro náhodné vektory je tzv. kovarianční matice. Pro náhodný vektor $(X,Y)^T$ je definována následovně:

$$\mathbb{C} = \begin{pmatrix} \operatorname{cov}(X,X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & \operatorname{cov}(Y,Y) \end{pmatrix} = \begin{pmatrix} DX & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & DY \end{pmatrix}.$$

Příklad 1.5 Máme symetrickou minci a hrací kostku. Nejprve hodíme mincí. Pokud padne rub hodíme kostkou dvakrát a padne-li líc hodíme jednou. Náhodná veličina X nabývá hodnoty 0 padne-li rub a 1 padne-li líc. Náhodná veličina Y udává počet padlých šestek na kostce (t.j. může být buď 0, 1 nebo 2). Nalezněte sdružené rozdělení pravděpodobnosti náhodného vektoru $(X,Y)^T$ a kovarianční matici.

Sdružené rozdělení pravděpodobnosti je dáno:

	$Y=0$	$Y=1$	$Y=2$
$X=0$	$\frac{1}{2}\cdot\frac{5}{6}\cdot\frac{5}{6}$	$2\cdot\frac{1}{2}\cdot\frac{5}{6}\cdot\frac{1}{6}$	$\frac{1}{2}\cdot\frac{1}{6}\cdot\frac{1}{6}$
$X=1$	$\frac{1}{2}\cdot\frac{5}{6}$	$\frac{1}{2}\cdot\frac{1}{6}$	0

Spočítejme hodnoty uvnitř tabulky a najděme marginální rozdělení pravděpodobnosti P_X a P_Y.

	$Y=0$	$Y=1$	$Y=2$	P_X
$X=0$	$\frac{25}{72}$	$\frac{10}{72}$	$\frac{1}{72}$	$\frac{1}{2}$
$X=1$	$\frac{30}{72}$	$\frac{6}{72}$	0	$\frac{1}{2}$
P_Y	$\frac{55}{72}$	$\frac{16}{72}$	$\frac{1}{72}$	

Pro střední hodnoty a rozptyly tedy dostáváme:

$$EX = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}\,,$$

$$EX^2 = 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = \frac{1}{2}\,,$$

$$DX = EX^2 - (EX)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}\,,$$

$$EY = 0 \cdot \frac{55}{72} + 1 \cdot \frac{16}{72} + 2 \cdot \frac{1}{72} = \frac{1}{4}\,,$$

$$EY^2 = 0^2 \cdot \frac{55}{72} + 1^2 \cdot \frac{16}{72} + 2^2 \cdot \frac{1}{72} = \frac{5}{18}\,,$$

$$DY = EY^2 - (EY)^2 = \frac{5}{18} - \frac{1}{16} = \frac{31}{144}\,.$$

Pro hodnotu kovariance je potřeba určit $E(XY)$.

$$E(XY) = 0 \cdot 0 \cdot \frac{25}{72} + 0 \cdot 1 \cdot \frac{10}{72} + \cdots + 1 \cdot 1 \cdot \frac{6}{72} + 1 \cdot 2 \cdot 0 = \frac{6}{72} = \frac{1}{12}\,.$$

Máme tedy:

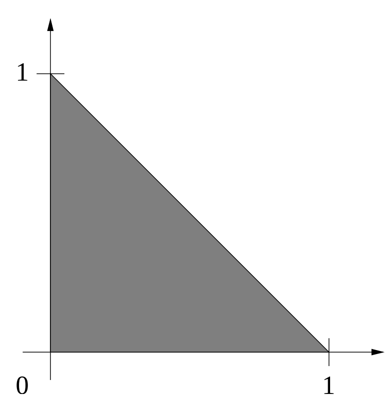
$$\operatorname{cov}(X,Y) = E(XY) - (EX)(EY) = \frac{1}{12} - \frac{1}{2} \cdot \frac{1}{4} = -\frac{1}{24}\,.$$

A kovarianční matice tedy je:

$$\mathbb{C} = \begin{pmatrix} 1/4 & -1/24 \\ -1/24 & 31/144 \end{pmatrix}.$$

Příklad 1.6 Nechť $T = \{(x,y) \mid x \geq 0, y \geq 0, x + y \leq 1\}$. Spojitý náhodný vektor $(X,Y)^T$ má rovnoměrné rozdělení na T. Najděte sdruženou hustotu pravděpodobnosti $f(x,y)$ a vyplňte prvky kovarianční matice $\mathbb{C}$.

Množina T je pravoúhlý trojúhleník ležící v prvním kvadrantu a vymezený přímkou $y = 1 - x$ viz obrázek:



Protože plocha trojúhelníka je $\frac{1}{2}$, dostaneme:

$$f(x,y) = \begin{cases} 2 & (x,y) \in T\,, \\ 0 & (x,y) \notin T\,. \end{cases}$$

Abychom mohli určit prvky kovarianční matice potřebujeme nejprve zjistit hustotu rozdělení X a Y, t.j. marginální hustotu.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y = \begin{cases} 0, & x \notin \langle 0,1\rangle\,, \\ \int_0^{1-x} 2\mathrm{d}y = 2(1-x), & x \in \langle 0,1\rangle\,. \end{cases}$$

Marginální hustota Y je ze symetrie trojúhelníku T stejná. Nyní můžeme určit střední hodnoty a rozptyly.

$$EX = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int_0^1 x2(1-x)\mathrm{d}x = 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{3}\,.$$

Ze symetrie dostáváme, že také $EY = \frac{1}{3}$. Dále

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\mathrm{d}x = \int_0^1 x^2 2(1-x)\mathrm{d}x = 2\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{1}{6}\,.$$

Rozptyl tedy je:

$$DX = EX^2 - (EX)^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}\,.$$

Ze symetrie opět dostaneme, že rovněž $DY = \frac{1}{18}$.

Pro hodnotu $\mathrm{cov}(X,Y)$ potřebujeme ještě určit $E(XY)$.

$$\begin{aligned} E(XY) &= \iint_T xy f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_T xy2\,\mathrm{d}x\mathrm{d}y \\ &= \int_0^1 \mathrm{d}x \int_0^{1-x} 2xy\,\mathrm{d}y = \int_0^1 x(1-x)^2\mathrm{d}x \\ &= \int_0^1 (x - 2x^2 + x^3)\mathrm{d}x = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{1}{12}\,. \end{aligned}$$

Máme tedy:

$$\mathrm{cov}(X,Y) = E(XY) - (EX)(EY) = \frac{1}{12} - \frac{1}{3}\cdot\frac{1}{3} = -\frac{1}{36}\,.$$

Veličiny X a Y jsou tedy závislé. Pro kovarianční matici dostáváme:

$$\mathbb{C} = \begin{pmatrix} 1/18 & -1/36 \\ -1/36 & 1/18 \end{pmatrix}.$$