

12.1 Náhodné javy a ich pravdepodobnosti

Pravdepodobnostné modely sa zaoberajú popisom procesov, ktoré sú ovplyvňované množstvom činiteľov, pričom tieto činitele poznáme iba čiastočne alebo ich nevieme určiť vôbec. Tieto činitele spôsobujú, že pri rovnakých základných podmienkach výsledok procesu je náhodný. Veličinu X , ktorej hodnoty sú úplne určené výsledkom náhodného procesu, nazývame **náhodnou premennou (veličinou)**. Náhodné premenné delíme na diskrétne a spojité.

Diskrétné náhodné premenné sú často charakterizované pravdepodobnostnou tabuľkou,

x_i	x_1	x_2	\dots	x_n
p_i	p_1	p_2	\dots	p_n

kde $p_i = P(X = x_i) = \sum_{x_i < x} p_i$ a platí rovnosť $\sum_{i=1}^n p_i = 1$. Štatistickým odhadom pravdepodobnosti p_i

je relatívna početnosť $\frac{m}{n}$, kde m je počet tých pokusov z celkového počtu n , pri ktorých nastala hodnota x_i . Jednou z možností popisu rozdelenia je zadanie **distribučnej funkcie**, ktorá každému reálnemu číslu priradí pravdepodobnosť, že náhodná premenná nadobudne hodnoty menšie než toto číslo, t.j.

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i.$$

Spojitou náhodnou premennou nazveme premennú X , pre ktorú existuje taká nezáporná funkcia $f(x)$, že distribučná funkcia

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Funkcia $f(x)$ je **hustota pravdepodobnosti** náhodnej premennej. Je zrejmé, že

$$\int_{-\infty}^{\infty} f(x) dx = F(\infty) = 1$$

a

$$P(x_1 \leq x < x_2) = \int_{x_1}^{x_2} f(t) dt = F(x_2) - F(x_1).$$

Charakteristiky polohy a variability

V praxi často sledujeme priemerné (očakávané) tempo rastu výroby, priemerné ceny, priemerné mzdy a pod. Používame pri tom jednoduchý aritmetický priemer hodnôt sledovaného kvantitatívneho znaku. Podobnú základnú informáciu o rozložení náhodnej premennej X nám poskytuje **stredná (očakávaná) hodnota** $\mu = E(X)$ náhodnej premennej X (tzv. charakteristika polohy), kde

$$E(X) = \sum_i x_i p_i, \quad \text{resp.} \quad E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

pre diskrétnu, resp. pre spojitú náhodnú premennú. Nech postupnosť hodnôt x_1, \dots, x_n predstavuje nezávislé realizácie náhodnej premennej X , potom pre veľké n je

$$E(X) \approx \bar{x} = \frac{1}{n}(x_1 + \dots + x_n).$$

Aritmetický priemer \bar{x} je štatistickým odhadom strednej hodnoty $E(X)$.

Charakteristikou variability náhodnej premennej je **rozptyl** (variancia, disperzia) $D(X)$, pričom

$$D(X) = E([X - E(X)]^2).$$

Druhá odmocnina z rozptylu je **smerodajná odchýlka** $\sigma = \sigma_X = \sqrt{D(X)}$. Ako odhad smerodajnej odchýlky sa používa výberová odchýlka, resp. štatistická smerodajná odchýlka

$$s_x^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{resp. } s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Najdôležitejším rozdelením spojitaj náhodnej premennej je **normálne rozdelenie** $N(\mu, \sigma^2)$, kde hustota pravdepodobnosti je daná vzťahom

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in R,$$

$$\mu = E(X),$$

$$\sigma = \sigma_X = \sqrt{D(X)}.$$

Rozdelenie $N(0,1)$ nazývame **normované normálne rozdelenie**.

Prostriedkom merania „tesnosti“ vzťahu medzi dvoma náhodnými premennými je **kovariancia** a je definovaná vzťahom

$$\text{Cov}_{XY} = E\{[X - E(X)] \cdot [Y - E(Y)]\} = E(XY) - E(X) \cdot E(Y).$$

Ak náhodné premenné X, Y zo systému X, Y sú nezávislé, $\text{Cov}_{XY} = 0$.

Majme systém náhodných premenných (X_1, \dots, X_s) a nech $\text{Cov}_{ij} = \text{Cov}_{X_i Y_j}$ a $\sigma_i = \sigma_{X_i}$. Potom maticu

$$\begin{pmatrix} \sigma_1^2 & \text{Cov}_{12} & \dots & \text{Cov}_{1s} \\ \text{Cov}_{21} & \sigma_2^2 & \dots & \text{Cov}_{2s} \\ \dots & \dots & \dots & \dots \\ \text{Cov}_{s1} & \text{Cov}_{s2} & \dots & \sigma_s^2 \end{pmatrix}$$

nazývame **kovariančnou maticou**. Štatistickým odhadom kovariancie $\text{Cov}_{X_i Y_j}$ je

$$k_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y}.$$

Výberová modifikovaná kovariancia je k_{xy}^* číslo

$$k_{xy}^* = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{n}{n-1} \cdot k_{xy}.$$

Mieru lineárnej závislosti medzi náhodnými premennými X, Y udáva **koeficient korelácie** $Corr_{XY}$ a je daný vzťahom

$$Corr_{XY} = \frac{Cov_{XY}}{\sigma_X \sigma_Y}.$$

Koeficient korelácie nadobúda hodnoty z intervalu $\langle -1, 1 \rangle$. Ak je koeficient korelácie $Corr_{XY}$ nulový, hovoríme, že náhodné premenné X, Y sú nekorelované.

Majme systém náhodných premenných (X_1, \dots, X_s) a nech

$$r_{ij} = \frac{k_{ij}}{s_i s_j} = \frac{k_{ij}^*}{s_i^* s_j^*} \approx Corr_{X_i X_j},$$

potom maticu

$$\begin{pmatrix} 1 & r_{12} & \dots & r_{1s} \\ r_{21} & 1 & \dots & r_{2s} \\ \dots & \dots & \dots & \dots \\ r_{s1} & r_{s2} & \dots & 1 \end{pmatrix},$$

kde $r_{ij} = r_{ji}$ nazývame **korelačnou maticou**.