3. prednáška

Rozdelenia spojitej náhodnej premennej

V praxi, či už v prírode alebo v ľudskej činnosti, sa často vyskytujú niektoré náhodné hodnoty, na ktorých sa dajú sledovať spoločné zákonistosti ich výskytu a popísať rozdelením ich pravdepodobnosti. S rozdeleniami diskrétnych náhodných premenných ste sa stretli už v predmete "Diskrétna pravdepodobnosť". Teraz si povieme o niektorých často sa vyskytujúcich rozdeleniach spojitých náhodných premenných.

Rovnomerné rozdelenie (Uniform Distribution) - ozn. $R(a,b)$

Náhodná premenná sa riadi rovnomerným rozdelením pravdepodobnosti, ak sa hodnoty náhodnej premennej vyskytujú s rovnakou pravdepodobnosťou na nejakom intervale (a,b). Hustota rozdelenia pravdepodobnosti náhodnej premennej X s rovnomerným rozdelením je daná predpisom:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases}$$

a, b sú parametre rozdelenia, ktoré je možné interpretovať ako dolné a horné ohraničenie hodnôt, ktoré môže náhodná premenná nadobúdať.
Rovnomerným rozdelením sa riadia napr. chyby pri zaokruhľovaní v numerických výpočtoch. Ak zaokruhľujeme na k–desatinných miest, tak chyba zaokruhľovania je náhodná premenná X, ktorá ma rovnomerné rozdelnie $Ro(-5 \cdot 10^{-k-1}; 5 \cdot 10^{-k-1})$.
Nájdime distribučnú funkciu rovnomerného rozdelenia:

Distribučnú funkciu zapíšeme v tvare:

$$F(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases}$$

Spočítajme strednú hodnotu a disperziu náhodnej premennej s rovnomerným rozdelením:

Zhrnieme: $E(X) = \dfrac{a+b}{2}, \qquad D(X) = \dfrac{(a-b)^2}{12}.$

Exponenciálne rozdelenie (Exponential distribution) - ozn. $Exp(\lambda)$

Náhodná premenná X, ktorá nadobúda nezáporné hodnoty, sa riadi exponenciálnym rozdelením pravdepodobnosti, ak hustota rozdelenia pravdepodobnosti má tvar:

$$f(x) = \begin{cases} \lambda \cdot \mathrm{e}^{-\lambda x} & x > 0,\ \lambda > 0,\ \lambda \in R \\ 0 & x \leq 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x} & x > 0,\ \lambda > 0,\ \lambda \in R \\ 0 & x \leq 0 \end{cases}$$

Nájdime charakteristiky náhodnej premennej s exponenciálnym rozdelením:

Zhrnieme: $\qquad E(X) = \dfrac{1}{\lambda}, \qquad D(X) = \dfrac{1}{\lambda^2}.$

Exponenciálne rozdelenie pravdepodobnosti má náhodná premenná, ktorá predstavuje dĺžku doby životnosti zariadenia, u ktorého dochádza k poruchám v dôsledku náhody (nie v dôsledku opotrebovania).
Rovnako sa exponenciálnym rozdelením riadi dĺžka intervalu medzi výskytmi dvoch udalostí v prípade, že počet výskytov udalostí počas časového intervalu sa riadi Poissonovým rozdelením pravdepodobnosti s parametrom λ. Poissonovo rozdelenie poznáte z predmetu "Diskrétna pravdepodobnosť" – je to rozdelenie "zriedkavých javov". Zriedkavé javy majú príležitosť, aby nastali, ale napriek tomu nastávajú zriedkavo, napr. počet leteckých katastrof v porovnaní s počtom letov za rok, alebo úmrtie ako dôsledok kopnutia koňom v porovnaní s ostatnými príčinami úmrtia za rok.
Exponenciálne rozdelenie je tzv. rozdelenie "bez pamäti". Ak má doba do výskytu udalosti (napr. poruchy zariadenia) exponenciálne rozdelenie, tak informácia o tom, že udalosť nenastala po dobu t_1, nemení pravdepodobnosť výskytu udalosti v nasledujúcom období dĺžky t_2. Ako keby zariadenie "zabudlo" na skôr odpracovanú dobu. Napr. ak doba do nastatia poruchy monitoru má exponenciálne rozdelenie, tak pravdepodobnosť, že monitor sa pokazí za viac ako 200 hodín od tejto chvíle, nijako nezávisí od toho, aký starý je monitor.

Normálne rozdelenie (Normal Distribution) - ozn. $N(m, \sigma^2)$

Normálne rozdelenie, ktoré má v štatistike kľúčový význam tak z teoretického, ako aj z praktického hľadiska, sa niekedy nazýva aj **Gaussovo**. Ako prvý ho popísal Abraham de Moivre v roku 1733. Jeho pôvodným cieľom bolo nájsť vhodnú spojitú aproximáciu diskrétneho binomického rozdelenia v súvislosti s hazardnými hrami.
Náhodné premenné s týmto rozdelením sa často vyskytujú ako výsledok rôznych meraní a pozorovaní. To, že náhodná premenná X nadobúdajúca reálne hodnoty sa riadi normálnym rozdelením $N(m, \sigma^2)$, píšeme $X \sim N(m, \sigma^2)$ a hustota rozdelenia pravdepodobnosti má tvar:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \,\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}} \qquad x \in (-\infty, \infty),$$

kde $m \in R$ a $\sigma > 0$ sú dané parametre.
Distribučná funkcia má tvar:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{(t-m)^2}{2\sigma^2}} \,\mathrm{dt} \qquad \mathrm{x} \in (-\infty, \infty)$$

a tento integrál nie je možné vyčísliť explicitne. A keďže, hlavne v štatistike, je potrebné poznať hodnoty distribučnej funkcie, boli vyvinuté metódy pre jej výpočet. V súčastnosti túto funkciu obsahuje každý štatistický softvér, vrátane Excelu.

Tvrdenie, že normálne rozdelenie hrá kľúčovú úlohu nielen medzi spojitými rozdeleniami, potvrdzuje prakticky fakt, že každý empirický súbor dát býva ako prvý testovaný na normálne rozdelenie. Navyše, ako si ukážeme nesôr, náhodná premenná, ktorá je súčtom veľkého počtu "malých" náhodných premenných, konverguje k normálnemu rozdeleniu. Normálne rozdelenie sa používa:

- ako model náhodnej premennej, ktorá je súčtom veľkého počtu drobných náhodných vplyvov;
- aproximácia iných rozdelení.

Ak položíme $m = 0$ $\sigma = 1$, dostaneme **normované normálne rozdelenie**, ozn. $N(0,1)$. V tomto prípade normovanej náhodnej premennej Z prislúcha hustota rozdelenia pravdepodobnosti:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{z^2}{2}} \qquad z \in (-\infty, \infty).$$

Ak si uvedomíme, že hodnoty $m = 0$ $\sigma = 1$ sú skôr výnimočné, musí existovať spôsob, ako z normálneho rozdelenia s ľubovoľnými parametrami m a σ^2 prejsť k normovanému normálnemu rozdeleniu. Postup, ktorý tento prechod umožňuje, sa nazýva normovanie a poznáme ho už z predchádzajúcich úvah. Ak $X \sim N(m, \sigma^2)$, tak náhodná premenná $Z = \dfrac{X - m}{\sigma} \sim N(0,1)$. Distribučná funkcia náhodnej premennej Z má tvar:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t \qquad z \in (-\infty, \infty)$$

Zo symetrie hustoty rozdelenia náhodnej premennej Z dostávame:

$$\Phi(-z) = 1 - \Phi(z)$$

a

$$\varphi(z) = \varphi(-z)$$

Hodnoty distribučnej funkcie $\Phi(z)$, resp. hustoty rozdelenia pravdepodobnosti $\varphi(z)$ sú tabelované vďaka hore uvedeným vlastnostiam len pre $z \geq 0$.
Pravdepodobnosť, že náhodná premenná $X \sim N(m, \sigma^2)$ nadobudne hodnoty z intervalu $< a, b)$ vyrátame podľa vzťahu:

Normálne rozdelenie má dôležité vlastnosti:

1. $E(X) = m_e = m_o$;
2. rozdelenie je symetrické okolo osi y, ktorá prechádza bodom m;

3. v intervale $m \pm \sigma$ leží 68% hodnôt, v intervale $m \pm 2\sigma$ leží 95% hodnôt a v intervale $m \pm 3\sigma$ leží 99,7% hodnôt.

Posledné pravidlo sa nazýva **pravidlo troch sigma** a často sa interpretuje tak, že pri jednom pokuse je takmer nemožné dostať náhodnú premennú X líšiacu sa od m o viac ako 3σ. Odkiaľ sa vzali tie percentá?

Trojuholníkové rozdelenie (Triangular Distribution) - ozn. $Tri(a, b, c)$

Náhodná premenná X má trojuholníkové rozdelenie na intervale $< a, b >$, $a, b \in R$, $a < b$, ak hustota rozdelenia pravdepodobnosti má tvar:

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & c \leq x \leq b \\ 0 & \text{inde} \end{cases}$$

pričom $a < c < b$. Zapisujeme $X \sim Tri(a, b, c)$, a, b, c sú parametre rozdelenia, kde a je minimálna, b je maximálna a c je najpravdepodobnejšia hodnota, ktorú môže náhodná premenná X nadobudnúť.

Distribučná funkcia má tvar:

$$F(x) = \begin{cases} 0 & x \leq a \\ \dfrac{(x-a)^2}{(b-a)(c-a)} & a < x \leq c \\ 1 - \dfrac{(b-x)^2}{(b-a)(b-c)} & c < x < b \\ 1 & x \geq b \end{cases}$$

$$E(X) = \frac{a+b+c}{3} \qquad D(X) = \frac{a^2+b^2+c^2-ab-ac-bc}{18}$$

Trojuholníkové rozdelenie pravdepodobnosti má bohaté využitie v aplikáciách projektového manažmentu, manažmentu rizika, pri analýze metód rozhodovania, ale tiež opisuje čas, ktorý uplynie medzi dvomi náhodne prichádzajúcimi signálmi. Hodí sa na modelovanie situácií, keď vieme povedať najhorší, najlepší a najpravdepodobnejší scenár. V praxi sa často používa trojuholníkové rozdelenie $Tri(a,b,c)$ s parametrom $c = \frac{a+b}{2}$. Stredná hodnota je potom $E(X) = \frac{a+b}{2}$ a disperzia $D(X) = \frac{(b-a)^2}{24}$.

<u>Poznámka:</u>

Ak X, Y sú nezávislé náhodné premenné rovnomerne rozdelené na intervale $\left\langle \frac{a}{2}, \frac{b}{2} \right\rangle$, potom náhodná premenná $X+Y$ má trojuholníkové rozdelenie na intervale $< a, b >$.

Gamma rozdelenie (Gamma Distribution) - ozn. $\Gamma(a,b)$

Hustota rozdelenia pravdepodobnosti náhodnej premennej X s Gamma rozdelením má tvar:

$$f(x) = \begin{cases} \dfrac{b^a}{\Gamma(a)} \cdot x^{a-1} \cdot \mathrm{e}^{-bx} & x > 0,\ a, b > 0, \\ 0 & x \leq 0 \end{cases},$$

kde Γ je gamma funkcia definovaná vzťahom $\Gamma(a) = \int_0^\infty x^{a-1}\,\mathrm{e}^{-x}\,\mathrm{d}x$.

Stredná hodnota $E(X) = \dfrac{a}{b}$ a rozptyl $D(X) = \dfrac{a}{b^2}$.

Pre celočíselnú hodnotu parametra a Gamma rozdelenie vzniká ako súčet a nezávislých náhodných premenných s exponenciálnym rozdelením pravdepodobnosti s parametrom b.

Chí-kvadrát rozdelenie (Chi-Square Distribution) - ozn. $\chi^2(n)$

Hustota rozdelenia pravdepodobnosti náhodnej premennej X s Chí-kvadrát rozdelením má tvar:

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n-2}{2}} \mathrm{e}^{-\frac{n}{2}} \qquad x > 0.$$

Zapisujeme $X \sim \chi^2(n)$. Toto rozdelenie má jeden parameter a tým je n, ktorý voláme počet stupňov voľnosti. Stredná hodnota náhodnej premenne s Chí-kvadrát rozdelením je $E(X) = n$ a rozptyl $D(X) = 2n$. Toto rozdelenie vzniká ako súčet druhých mocnín nezávislých náhodných premenných s normálnym normovaným rozdelením: $X_i \sim N(0,1)$, $i = 1, 2, \ldots, n$.
Teda ak $Y = X_1^2 + X_2^2 + \cdots + X_n^2$, tak $Y \sim \chi^2(n)$.

Studentovo rozdelenie (Student Distribution) - ozn. $t(n)$

Je významným rozdelením – stretávať sa s ním budeme hlavne v štatistike. Jeho odvodenie sa pripisuje Williamovi Gossetovi, ktorý bol v oblasti štatistiky samoukom a bol preto známy pod pseudonymom "Student". Týnto pseudonymom podpisoval na začiatku 20. storočia svoje štatistické práce chemik pivovaru Guiness. Gosset je jedným zo zakladateľov aplikácií induktívnej štatistiky, a to v oblasti bezpochyby významnej – v zabezpečení kvality piva ☺. Náhodnú premennú so Studentovým rozdelením pravdepodobnosti dostaneme ako podiel dvoch nezávislých náhodných premenných X a Y.
Nech $X \sim N(0,1)$ a $Y \sim \chi^2(n)$. Potom náhodná premenná

$$t = \frac{X}{\sqrt{\frac{Y}{n}}} = \frac{X}{\sqrt{Y}} \cdot \sqrt{n}$$

má Studentovo rozdelenie o n– stupňoch voľnosti. Zapisujeme $t \sim t(n)$. Pre $n > 2$ platí, že stredná hodnota $E(t) = 0$ a rozptyl je $D(t) = \dfrac{n}{n-2}$.
S rastúcim n sa t– rozdelenie blíži k normálnemu normovanému rozdeleniu. Pre $n > 30$ je tvar oboch rozdelení prakticky zhodný.

Fisherovo-Snedecoroco F–rozdelenie (Fisher-Snedecor Distribution) - ozn. $F(n_1, n_2)$

Nech X a Y sú nezávislé náhodné premenné, kde $X \sim \chi^2(n_1)$ a $Y \sim \chi^2(n_2)$. Potom náhodná premenná

$$F = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}} = \frac{X}{Y}\frac{n_2}{n_1}$$

má F–rozdelenie s n_1, n_2 stupňami voľnosti, čo označujeme $F \sim F(n_1, n_2)$. Hodnoty n_1 a n_2 sú parametre rozdelenia, na poradí parametrov tvar rozdelenia pochopiteľne závisí. Pre $n_2 > 4$ platí:

$$E(F) = \frac{n_2}{n_2 - 2} \quad \text{a} \quad D(F) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$$