Poznámky k předmětu Aplikovaná statistika, 2. téma

1. Náhodné veličiny

Motivace: Ohlédneme-li se za příkladem 1 z minulého tématu, můžeme hod spravedlivou šestistěnnou kostkou modelovat funkcí X, jejímž definičním oborem je množina elementárních jevů $\Omega = \{\omega_1, \ldots, \omega_6\}$ a oborem hodnot množina $\{1, \ldots, 6\}$ počtů ok vyskytujících se na kostce. Funkce X je tedy definovaná předpisem

$$X(\omega_i) = i, \ i = 1, \ldots, 6,$$

(např. elementárnímu jevu „na kostce padla čtyřka“ přiřadí X číslo 4).

Poznámka 1. *Anebo ještě jinak. V deskové hře „Marrakech“, ve které se na několik desítek minut můžete stát prodejcem orientálních koberců, se používá zvláštní šestistěnná kostka K. Na dvou stěnách má číslo dva, na dalších dvou číslo tři a na zbylých dvou stěnách jsou čísla jedna a čtyři. Ovšem tuto kostku se vám povedlo ztratit. Nicméně „Marrakech“ je možné hrát i s klasickou kostkou, modelujeme-li výsledek hodu kostkou K (pomocí klasické kostky) funkcí*

$$Y : \{\omega_1, \ldots, \omega_6\} \to \{1, 2, 3, 4\}$$

definovanou následovně

$$\begin{aligned} Y(\omega_1) &= 1, \\ Y(\omega_4) &= 4, \\ Y(\omega_2) &= Y(\omega_5) = 2, \\ Y(\omega_3) &= Y(\omega_6) = 3. \end{aligned}$$

Náhodná veličina $X : \Omega \to \mathbb{R}$ je funkce, která každému elementárnímu jevu $\omega \in \Omega$ přiřadí reálné číslo $X(\omega) \in \mathbb{R}$.
Náhodné veličiny obvykle značíme velkými písmeny z konce abecedy (X, Y, Z).

Z povahy výsledků pokusů lze rozdělit náhodné veličiny na

- **diskrétní** – mohou nabývat nejvýše spočetně mnoha hodnot,
  - konečně (zmíněný hod kostkou),
  - spočetně (pořadí hodu kostkou, v němž poprvé padne šestka; což nás velmi zajímá, hrajeme-li „Člověče, nezlob se!“),
- **spojité** – nabývají nespočetně mnoha hodnot (výška člověka, koncentrace kyseliny).

Házíme-li kostkou opakovaně, jistě nás zajímá, zdali jsou relativní četnosti výskytu elementárních jevů $\omega_1, \ldots, \omega_6$ stejné, aneb zdali je naše kostka „spravedlivá“. Intuitivně je jasné, že pravděpodobnost, že náhodná veličina X je rovna např. čtyřem, by měla být 1/6, stručněji zapsáno

$$\mathbb{P}\big(\{\omega \in \Omega; X(\omega) = 4\}\big) = \frac{1}{6}.$$

To nás přivádí k tzv. **rozdělení pravděpodobnosti náhodné veličiny** X. Zhruba lze říci, že je to rozdělení pravděpodobnosti mezi podmnožiny $\mathbb{R}$, tj.

- musíme znát pravděpodobnosti hodnot, kterých může náhodná veličina nabývat (v případě diskrétní náhodné veličiny),
- či pravděpodobnosti jejích hodnot ležících v intervalech (v případě spojité náhodné veličiny).

Popišme si blíže rozdělení diskrétní a spojité náhodné veličiny:

Diskrétní náhodná veličina X nabývá nejvýše spočetně mnoha hodnot z množiny $M = \{x_1, x_2, \ldots\}$ s kladnými pravděpodobnostmi $p(x_i) = \mathbb{P}(X = x_i)$. Funkce p se nazývá **pravděpodobnostní funkce** a na celém $\mathbb{R}$ se definuje předpisem

$$p(a) = \begin{cases} \mathbb{P}(X = a) & , \quad a \in M, \\ 0 & , \quad a \notin M. \end{cases}$$

Určuje rozdělení pravděpodobnosti náhodné veličiny X.

Poznámka 2. *Je třeba vysvětlit značení $\mathbb{P}(X = a)$. Je to stručnější a přehlednější zápis pro*

$$\mathbb{P}\big(\{\omega \in \Omega;\ X(\omega) = a\}\big).$$

V dalším textu ho budeme používat.

Příklad 1. *Dítě při návštěvě kina vyhraje $1, 2, 3$ nebo 4 bonbóny s pravděpodobnostmi $0,5$; $0,3$; $0,1$; $0,1$.*

Vlastnosti pravděpodobnostní funkce

- $p(x_i) > 0,\ x_i \in M,$
- $\sum_{x_i \in M} p(x_i) = 1,$
- $\mathbb{P}(X \leq x) = \sum_{x_i \leq x, x_i \in M} p(x_i),$
- analogicky $\mathbb{P}(a < X \leq b)$, $\mathbb{P}(a < X)$, $\mathbb{P}(a \leq X \leq b)$.

Spojitá náhodná veličina X nabývá obecně libovolných reálných hodnot, většinou z nějakého intervalu $I \subseteq \mathbb{R}$. Její rozdělení pravděpodobnosti lze určit **hustotou** f **náhodné veličiny** X (je to analogie pravděpodobnostní funkce u diskrétní náhodné veličiny), což je taková **nezáporná** funkce, že

$$\mathbb{P}(a < X \leq b) = \int_a^b f(t)\mathrm{d}t,\ a, b \in \mathbb{R}, a \leq b.$$

Příklad 2. *Spojitá náhodná veličina je zadána hustotou*

$$f(x) = \begin{cases} 2(x+1) & , \quad x \in [-1, 0], \\ 0 & , \quad \text{jinde.} \end{cases}$$

Vlastnosti spojité náhodné veličiny

- $\int_{-\infty}^{+\infty} f(t)\mathrm{d}t = 1,$
- $\mathbb{P}(X = a) = 0,\ a \in \mathbb{R},$

- důsledkem předchozí rovnosti jsou další vztahy

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b).$$

Poznámka 3. *Geometricky lze pravděpodobnost $\mathbb{P}(a < X \leq b)$ interpretovat jako obsah plochy ohraničené grafem hustoty f na intervalu $(a, b]$ a osou x.*

Rozdělení pravděpodobnosti libovolné náhodné veličiny X (nejen diskrétní, či spojité!) lze jednoznačně zadat **distribuční funkcí** F

$$F(x) = \mathbb{P}\big(\{\omega \in \Omega; X \leq x\}\big),\ x \in \mathbb{R},$$

stručněji zapsáno

$$F(x) = \mathbb{P}\big(X \leq x) \equiv \mathbb{P}\big(X \in (\infty, x\,]\big),\ x \in \mathbb{R}.$$

Distribuční funkce je vlastně „kumulovaná pravděpodobnost“. V bodě x určuje celkovou pravděpodobnost jevu, že náhodná veličina nepřekročí hodnotu x.

Vlastnosti distribuční funkce

- F je neklesající,
- F je zprava spojitá s nejvýše spočetně mnoha body nespojitosti,
- $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$,
- pro libovolné $a < b \in \mathbb{R}$ platí

$$\mathbb{P}(a < X \leq b) \equiv P\big(X \in (a, b]\big) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

Vztah distribuční funkce a

(D) pravděpodobnostní funkce

- $F(x) = \sum_{x_i \leq x, x_i \in M} p(x_i)$,
- $p(x) = F(x) - \lim_{y\to x-} F(y),\ x \in \mathbb{R}$, což znamená, že pro x
  - bod spojitosti F je $p(x) = 0$,
  - bod nespojitosti F je hodnota p rovna „velikosti skoku“ v tomto bodě.

(S) hustoty

- $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t,\ x \in \mathbb{R}$,
- $F'(x) = f(x)$ pro všechna $x \in \mathbb{R}$, pro která $F'(x)$ existuje.

Příklad 3. *Dítě při návštěvě kina vyhraje $1, 2, 3$ nebo 4 bonbóny s pravděpodobnostmi $0,5$; $0,3$; $0,1$; $0,1$. Nakreslete graf distribuční a pravděpodobnostní funkce a určete $F(-1)$, $F(0,8)$, $F(1)$, $F(1,8)$, $F(2)$.*

Cvičení 1. Pro pochopení pravidel hry „Osadníci z Katanu“ se nám hodí vyřešit následující úlohu. Náhodná veličina X udává součet čísel při hodu dvěma klasickými kostkami. Určete

(a) pravděpodobnostní funkci náhodné veličiny X (tabulkou i grafem),

(b) distribuční funkci (tabulkou i grafem),

(c) pravděpodobnosti $\mathbb{P}(X \leq 3)$, $\mathbb{P}(X > 9)$, $\mathbb{P}(8 \leq X \leq 11)$, $\mathbb{P}(X = 2)$.

Příklad 4. *Modelujme výšku květin v záhonku spojitou náhodnou veličinou X s hustotou*

$$f(x) = \begin{cases} c \sin x & , \quad x \in [0, \pi], \\ 0 & , \quad \text{jinde.} \end{cases}$$

Nalezněte

(a) *konstantu c tak, aby f byla skutečně hustota, a nakreslete graf f,*

(b) $\mathbb{P}(\frac{\pi}{3} \leq X \leq \frac{4\pi}{3})$, $\mathbb{P}(X = \frac{\pi}{6})$,

(c) *distribuční funkci F náhodné veličiny X včetně jejího grafu, a pak s využitím F určete $\mathbb{P}\big(X \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]\big)$. Určete tuto pravděpodobnost i graficky.*

Cvičení 2. Náhodná veličina X je dána pravděpodobnostní funkcí

$$p(k) = \begin{cases} c\,(k-2)^2 & , \quad k \in \{-1, 0, 1, 3\}, \\ 0 & , \quad \text{jinak.} \end{cases}$$

Určete

(a) c tak, aby p byla pravděpodobnostní funkce,

(b) distribuční funkci náhodné veličiny X a její graf,

(c) $\mathbb{P}(X \geq 1)$, $\mathbb{P}\big(X \in \{0, 3\}\big)$.

Domácí cvičení 1. Náhodné veličina X je dána distribuční funkcí

$$F(x) = \begin{cases} \frac{1}{16}x^4 & , \quad x \in [0, 2], \\ 0 & , \quad x < 0, \\ 1 & , \quad x > 2. \end{cases}$$

Nakreslete graf F, spočítejte hustotu náhodné veličiny X a nakrestelte její graf, dále určete pravděpodobnosti $\mathbb{P}\big(X \in (1/2, 1)\big)$, $\mathbb{P}(X > 1)$, $\mathbb{P}(X > 3)$.