

Úvod do pravděpodobnosti

01UP2

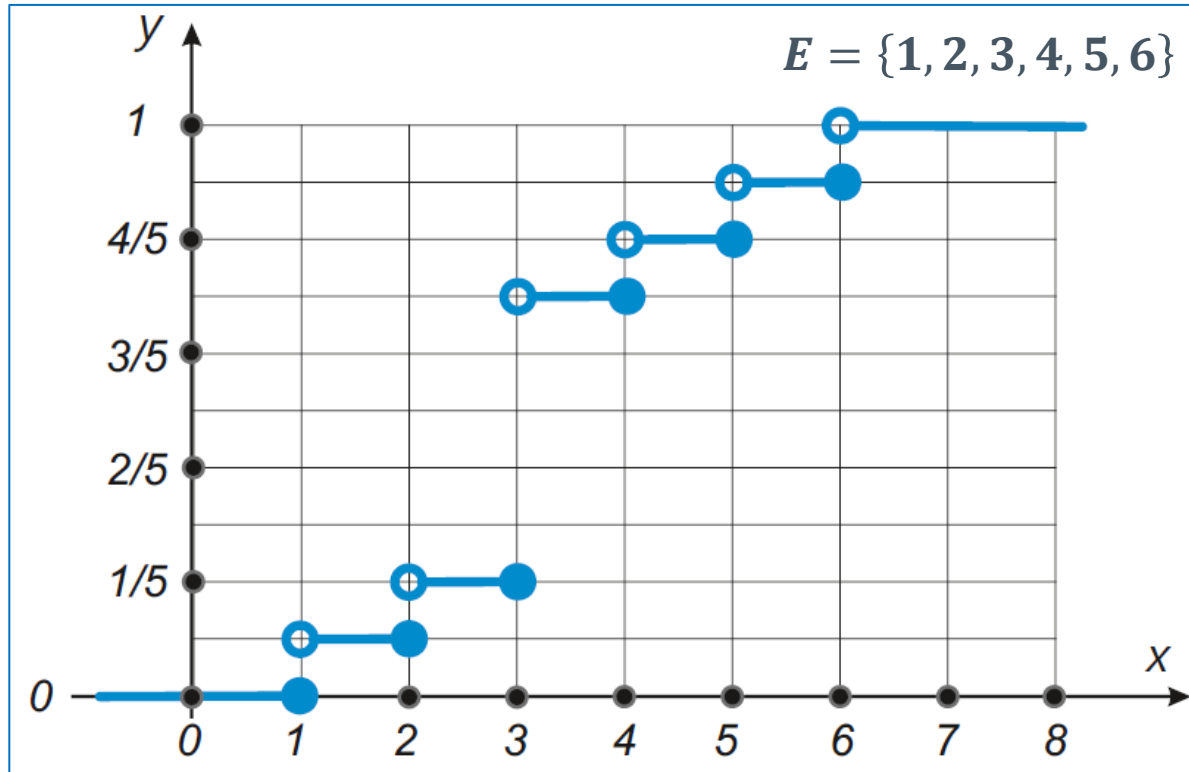
Distribuční funkce

Připomenutí – hod falešnou kostkou

Uvažujme falešnou hrací kostku, při jejímž hodu padá číslo 3 pětkrát častěji než ostatní hodnoty. Všechny ostatní hodnoty jsou stejně pravděpodobné. Výběrovým prostorem je množina $E = \{1, 2, 3, 4, 5, 6\}$. A jejím rozkladem na úplný systém neslučitelných (a zároveň elementárních) jevů je zjevně systém množin (jevů) $B_1 = \{1\}$, $B_2 = \{2\}$, $B_3 = \{3\}$, $B_4 = \{4\}$, $B_5 = \{5\}$ a $B_6 = \{6\}$. S ohledem na výše uvedený fakt, že $\mathcal{P}[B_3] = 5\mathcal{P}[B_k]$ pro všechna $k = 1, 2, 4, 5, 6$, a také že $\sum_{k=1}^6 \mathcal{P}[B_k] = 1$, vidíme, že pravděpodobnostní tabulka zkoumaného jevu vypadá takto:

Elementární jev	Jeho pravděpodobnost
$B_1 = \{1\}$	$\mathcal{P}[B_1] = \frac{1}{10}$
$B_2 = \{2\}$	$\mathcal{P}[B_2] = \frac{1}{10}$
$B_3 = \{3\}$	$\mathcal{P}[B_3] = \frac{5}{10}$
$B_4 = \{4\}$	$\mathcal{P}[B_4] = \frac{1}{10}$
$B_5 = \{5\}$	$\mathcal{P}[B_5] = \frac{1}{10}$
$B_6 = \{6\}$	$\mathcal{P}[B_6] = \frac{1}{10}$

Připomenutí – hod falešnou kostkou



- $F(0,8)=P[X < 0,8] = P[X \in \{\}] = 0$
- $F(2)=P[X < 2] = P[X \in \{1\}] = \frac{1}{10}$
- $F(5,2)=P[X < 5,2] = P[X \in \{1,\dots,5\}] = \frac{9}{10}$

VLASTNOSTI:

- pro diskrétní NV typicky skokový charakter,
- spojitá zleva,
- neklesající,
- obor hodnot z intervalu $\langle 0,1 \rangle$, ...

Příklad – F(x) pro spojitě rovnoměrné rozdělení

- › spojitá rovnoměrně rozdělená náhodná veličina X
- › realizace X nabývají všech hodnot z (a, b) se stejnou pravděpodobností
- › všechny intervaly $\langle a, a + \delta \rangle$ o stejné délce $\delta > 0$, které jsou podmnožinou výběrového prostoru mají stejnou pravděpodobnost
- › distribuční funkce:

$$F(x) = \begin{cases} 0 & x \leq a; \\ \frac{x-a}{b-a} & x \in (a, b); \\ 1 & x \geq b. \end{cases}$$

Jak bude vypadat distribuční funkce pro $a = 1, b = 7$?

VLASTNOSTI:

- všechny výše uvedené
- + kompletně spojitá

Heavisideova funkce

› elementární po částech spojitá funkce $\Theta(x): \mathbb{R} \rightarrow \mathbb{R}$ definovaná:

$$\Theta(x) = \begin{cases} 0 & x \leq 0; \\ 1 & x > 0. \end{cases}$$

- › pro libovolnou funkci $h(x): \mathbb{R} \rightarrow \mathbb{R}$ se definuje součin $\Theta(x)h(x)$ **nulový** pro všechna $x \in I$, $I = (-\infty, 0)$
- › platí i pokud funkce na I není definovaná, nemá v některých bodech I limity nebo jsou limity rovné $\pm\infty$
- › zjednoduší zápis funkcí (skokový průběh); $g(x) = \frac{\Theta(x)}{x} \rightarrow g(0) = 0$
- › **alternativní Heavisideova funkce:** $\Theta(x)^* = 1 - \Theta(-x)$

Základní vlastnosti distribuční funkce

- $\text{Dom}(F) = \mathbf{R}$;
- $\text{Ran}(F) \subset \langle 0, 1 \rangle$;
- $F(x)$ je na $\mathbf{R}$ neklesající;
- $\lim_{x \rightarrow +\infty} F(x) = 1$;
- $\lim_{x \rightarrow -\infty} F(x) = 0$;
- $F(x)$ je na $\mathbf{R}$ spojitá zleva.

› dodatečný předpoklad: **částečná spojitost** $F(x)$

- kompletně spojitá $F(x) \in \mathcal{C}(\mathbf{R})$
- má konečně mnoho bodů nespojitosti, $F(x) \in PC(\mathbf{R})$
- blíže viz definice po částech spojitých funkcí ve skriptech

Důkazy základních vlastností distribuční funkce

$$\text{Ran}(F) \subset \langle 0, 1 \rangle$$

- › definiční obor distribuční funkce jsou všechna $\mathbb{R}$
- › pravděpodobnost $P[X < x]$ má obor hodnot $\langle 0, 1 \rangle$
- › $\text{Ran}(F) \subset \langle 0, 1 \rangle$, pouze ve speciálních případech $\text{Ran}(F) = \langle 0, 1 \rangle$

$$F(x) \text{ je na } \mathbb{R} \text{ neklesající}$$

- › zvolme libovolnou dvojici $x < y$, chceme ukázat $F(x) \leq F(y)$
- › $(-\infty, x) \subset (-\infty, y) \Rightarrow P[(-\infty, x)] \leq P[(-\infty, y)]$...monotonie pravd. míry
- › $F(x) = F(X < x) = \dots?$

$$\lim_{x \rightarrow +\infty} F(x) = 1$$

› princip: $P[X \in (-\infty, +\infty)] = P[\mathbb{R}] = 1$

› posloupnost $(x_k)_{k=1}^{\infty}$, pro kterou platí:

- $x_1 < x_2 < x_3 < \dots$,
- $\lim_{k \rightarrow \infty} x_k = \infty$.

› Pak je posloupnost příslušných distribučních funkčních hodnot $(F_k)_{k=1}^{\infty}$ **neklesající**, protože:

$$F(x_k) = \mathcal{P}[\mathbb{X} < x_k] = \mathcal{P}[\mathbb{X} \in (-\infty, x_k)] \leq \mathcal{P}[\mathbb{X} \in (-\infty, x_{k+1})] = \mathcal{P}[\mathbb{X} < x_{k+1}] = F(x_{k+1})$$

› funkce **omezená** a **neklesající** $\Rightarrow$ musí mít limitu; logicky rovna 1

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

- › analogický postup
- › zavedeme $G(x) = 1 - F(x)$ a aplikujeme předchozí úvahy

$F(x)$ je na $\mathbf{R}$ spojitá zleva

- › předmětem matematické analýzy B4
- › důkaz ve skriptech k 01ANB4

Pravděpodobnost intervalů a bodových množin

Polouzavřený interval

› libovolná $[a < b]$; pravděpodobnostní dotaz $P[X \in \langle a, b \rangle] = ?$

› $\langle a, b \rangle = (-\infty, b) \setminus (-\infty, a)$ a zároveň $(-\infty, a) \subset (-\infty, b)$

› využijeme vlastnosti:

$$A, B \in \mathcal{D} \wedge A \subset B \Rightarrow P[B \setminus A] = P[B] - P[A]$$



$$\mathcal{P}[X \in \langle a, b \rangle] = F(b) - F(a)$$

Otevřený interval

- › libovolná $[a < b]$; pravděpodobnostní dotaz $P[X \in (a, b)] = ?$
- › otevřený interval $(0, 1)$ lze pokrýt disjunktním spočetným sjednocením polouzavřených intervalů:

$$(0, 1) = \bigcup_{n=1}^{\infty} \left(\frac{1}{n+1}, \frac{1}{n} \right)$$

- › ze σ -aditivity víme:
$$P[(a, b)] = \sum_{n=1}^{\infty} P \left[\left(\frac{1}{n+1}, \frac{1}{n} \right) \right]$$
- › pro polouzavřené intervaly:
$$P[(a, b)] = \sum_{n=1}^{\infty} \left(F \left(\frac{1}{n} \right) - F \left(\frac{1}{n+1} \right) \right)$$
- › součet číselné řady = limita posloupnosti částečných součtů

$$P[(0,1)] = \lim_{k \rightarrow \infty} \left(F(1) - F \left(\frac{1}{k+1} \right) \right) = F(1) - \lim_{x \rightarrow 0^+} F(x)$$

- › lze zobecnit pro obecný interval (a, b) , viz skriptá

$$P[(a, b)] = \lim_{k \rightarrow \infty} \left(F(b) - F\left(a + \frac{b-a}{k+1}\right) \right) = F(b) - \lim_{x \rightarrow a^+} F(x)$$

- › obecně tedy: $P[X \in (a, b)] \leq P[X \in \langle a, b \rangle]$
- › totožné pouze když je $F(x)$ spojitá v bodě a

Jednoprvková množina

- › množina $A = \{a\}$, $a \in \mathbb{R}$, tedy $\{a\} \cup (a, b) = \langle a, b \rangle \wedge \{a\} \cap (a, b) = \emptyset$

- › axiom aditivity:

$$P[X = a] = \lim_{x \rightarrow a^+} F(x) - F(a) = [F(x)]_a$$

- › $[F(x)]_a$...délka skoku distribuční funkce
- › $F(x)$ v bodě a spojitá, pak $P[X = a] = 0$

Konečněprvkové množiny

- › množina $K = \{a_1, a_2, \dots, a_m\}$ různých prvků; $P[X \in K] = ?$
- › rozklad množiny na disjunktní sjednocení $A = \{a_1\}$, $B = \{a_2, \dots, a_m\}$
 $\Rightarrow \mathcal{P}[X \in K] = \mathcal{P}[X \in A] + \mathcal{P}[X \in B] = \mathcal{P}[X = a_1] + \mathcal{P}[X \in B]$
- › rozklad B na disjunktní sjednocení $C = \{a_2\}$, $D = \{a_3, \dots, a_m\}$, ...
- › postupné rozklady dalších množin

$$\Rightarrow \mathcal{P}[X \in K] = \sum_{k=1}^m \mathcal{P}[X = a_k] = \sum_{k=1}^m [F(x)]_{a_k}$$

- › součet skoků distribuční funkce v bodech původní množiny

Spočetněprvkové množiny

- › množina $S = \{a_1, a_2, \dots\}$ různých prvků; $P[X \in S] = ?$
- › rozklad množiny na spočetně disjunktní sjednocení jednoprvkových množin
- › aplikace axiomu σ –aditivity pravděpodobnostní míry


$$\mathcal{P}[X \in S] = \sum_{k=1}^{\infty} \mathcal{P}[X = a_k] = \sum_{k=1}^{\infty} [F(x)]_{a_k}$$

ÚLOHY: pravděpodobnost uzavřeného intervalu

- pravděpodobnost uzavřeného intervalu

Na základě předešlých výsledků odvoďte pravděpodobnost $\mathcal{P}[\mathbb{X} \in \langle a, b \rangle]$, že výsledkem experimentu s náhodnou proměnnou $\mathbb{X}$ bude hodnota ležící v uzavřeném intervalu $\langle a, b \rangle$.

- distribuční funkce diskrétní rovnoměrně rozdělené proměnné

Rozmyslete, jakou vlastnost má distribuční funkce diskrétní náhodné proměnné, kdy jsou všechny elementární jevy stejně pravděpodobné. Nalezněte také příklady takového náhodného jevu a příslušnou distribuční funkci vykreslete.

Úvaha: množiny nulové pravděpodobnosti

- **P-nulové množiny** = jejich pravděpodobnostní míra je nulová
- Jde pro $P[X \in A] = 0$ automaticky o jev nemožný?
 - diskrétní NV ano
 - spojitá NV -> četnost jevu je extrémně nízká:

Provedeme-li m opakovaných pokusů s toutéž náhodnou proměnnou $\mathbb{X}$ a zaznameníme-li si do proměnné n počet všech případů, kdy byla výsledkem experimentu hodnota a , pak výraz n/m reprezentuje relativní četnost zkoumaného jevu. Protože číslo n závisí na počtu všech pokusů, píšeme $n = n(m)$. Relativní četnost

$$\kappa(m) = \frac{n(m)}{m}$$

je tedy také funkcí celkového počtu pokusů. A protože je jasné, že se zvyšujícím se počtem pokusů bude relativní četnost $\kappa(m)$ konvergovat k teoretické hodnotě pravděpodobnosti $\mathcal{P}[\mathbb{X} = a]$, vidíme, že

$$\lim_{m \rightarrow +\infty} \kappa(m) = \lim_{m \rightarrow +\infty} \frac{n(m)}{m} = \mathcal{P}[\mathbb{X} = a] = 0.$$