PRAVDEPODOBNOSŤ a ŠTATISTIKA

Náhodná premenná

doc. RNDr. Štefan Peško

Katedra matematických metód a operačnej analýzy, FRI ŽU

11. marca 2019

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor. Reálnu funkciu $\mathbf{X} : \Omega \to \Re$ nazveme náhodná premenná, ak pre každé $x \in \Re$ platí

$$\{\omega \in \Omega : \mathbf{X}(\omega) < x\} \in \mathcal{F}. \tag{1}$$

Hodnoty $\mathbf{X}(\omega)$ nazývame realizáciami náhodnej premenne X.

Dohoda:
Náhodné premenné budeme označovať veľkými tučnými písmenami $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ a ich realizácie malými písmenami x, y, z.

Definujme podmnožinu výberového priestoru Ω

$$[\mathbf{X} = x] = \{\omega \in \Omega : \mathbf{X}(\omega) = x\},$$

ktorá obsahuje všetky elementárne udalosti, ktorým náhodná premenná priraďuje rovnakú hodnotu.

Podobne pre každé $x \in \Re$ a každé $a, b \in \Re$ definujeme množiny

$$[\mathbf{X} < x] = \{\omega \in \Omega : \mathbf{X}(\omega) < x\},$$

$$[\mathbf{X} \geq x] = \{\omega \in \Omega : \mathbf{X}(\omega) \geq x\},$$

$$[a < \mathbf{X} < b] = \{\omega \in \Omega : a < \mathbf{X}(\omega) < b\}.$$

Príklad 3.1

Uvažujme hod dvoma spravodlivými mincami. Analyzujme náhodnú premennú, ktorá každému možnému výsledku hodu priradí počet padnutých znakov (z).

Výberový priestor $\Omega = \{(h,h),(h,z),(z,h),(z,z)\}$ obsahuje 4 rovnako pravdepodobné náhodné udalosti. Náhodná premenná $\mathbf{X}$ môže nadobúdať tri hodnoty z množiny $\{0,1,2\}$

$$
\begin{aligned}
[\mathbf{X} < 0] &= \emptyset, \\
[\mathbf{X} < 1] &= \{(h,h)\}, \\
[\mathbf{X} < 2] &= \{(h,h),(h,z),(z,h)\}, \\
[\mathbf{X} < 3] &= \{(h,h),(h,z),(z,h),(z,z)\}.
\end{aligned}
$$

Na Ω môžeme definovať elementárne pole udalostí $\mathcal{F}_1 = \{\emptyset, \Omega\}$. Ale $\mathbf{X}$ nie je náhodnou premennou v $(\Omega, \mathcal{F}_1, \mathcal{P})$. No jeho postupným dopĺňaním komplemetárnymi udalosťami dostaneme najmenšie neelementárne pole udalosti $\mathcal{F}_2 = 2^\Omega$ také, že $[\mathbf{X} < k] \in \mathcal{F}_2$. Potom už $\mathbf{X}$ je náhodnou premennou v $(\Omega, \mathcal{F}_2, \mathcal{P})$.

Príklad 3.2 (Čakanie na autobus)

Uvažujme náhodnú udalosť, ktorá spočíva v dobe čakania na autobus, ktorý chodí v T-minútových intervaloch. Táto doba $\mathbf{X}$ môže nadobúdat hodnoty z intervalu $\langle 0, T\rangle$. Vytvorme pole udalostí v ktorom bude $\mathbf{X}$ náhodnou premennou.

Nech je x ľubovoľné reálne číslo z intervalu $x \in \langle 0, T\rangle$. Potom je interval $(-\infty, x)$ vhodným vzorom intervalu $\langle 0, x\rangle$. Množina intervalov $\mathcal{B} = \{(-\infty, b) : b \in \Re\}$ je tu poľom udalostí. Z prvkov takýchto otvorených intervalov vytvoríme uzavretý interval

$$\langle 0, x\rangle = (-\infty, x\rangle - (-\infty, 0) = \bigcap_{n=1}^{\infty}\Big(-\infty, x + \frac{1}{n}\Big) - (-\infty, 0),$$

a tak je $[\mathbf{X} < x] \in \mathcal{B}$.

Poznámka:

Možno ukázať, že do poľa udalostí $\mathcal{B}$ patria všetky intervaly tvaru $\langle a, b\rangle, (a, b\rangle, \langle a, b), (a, b), \langle -\infty, b\rangle, \langle -\infty, b), (a, \infty), \langle a, \infty).(-\infty, \infty)$.
Toto pole je známe pod názvom borelovská σ-algebra.

Rozdelenie pravdepodobnosti náhodnej premennej

Nech $(\Omega, \mathcal{B}, \mathcal{P})$ je pravdepodobnostný priestor. Množinovú funkciu $\mathcal{P}_{\mathbf{X}} : \mathcal{B} \to \langle 0, 1\rangle$ definovanú pre všetky $B \in \mathcal{B}$ vzťahom

$$\mathcal{P}_{\mathbf{X}}(B) = \mathcal{P}(\{\omega \in \Omega : \mathbf{X}(\omega) \in B\}). \tag{2}$$

nazývame pravdepodobnostné rozdelenie náhodnej premennej $\mathbf{X}$,

Poznámka:
Pravdepodobnostné chovanie náhodnej veličiny $\mathbf{X} : \Omega \to \Re$ je určené systémom pravdepodobností $\mathcal{P}_{\mathbf{X}}(B)$ v príslušnom poli náhodných udalostí $\mathcal{B}$.

Distribučná funkcia

Nech $\mathbf{X}$ je náhodná premenná definovaná na pp. priestore $(\Omega, \mathcal{B}, \mathcal{P})$. Reálna funkcia

$$F_{\mathbf{X}}(x) = \mathcal{P}(\mathbf{X} < x). \tag{3}$$

sa nazýva ditribučná funkcia náhodnej premennej $\mathbf{X}$.

Dohoda:
Ak nebude hroziť nedorozumenie, budeme dolný index $\mathbf{X}$ v označení distribučnej funkcie vynechávať t.j. budeme písať len $F(x)$. Ďalej budeme používať skrátený zápis $\mathcal{P}(\mathbf{X} < x)$ miesto $\mathcal{P}([\mathbf{X} < x])$.

Príklad 3.3

Nech B je nejaká náhodná udalosť, ktorá nastáva s pp. $\mathcal{P}(B) = p$ a nenastáva s pp. $\mathcal{P}(B^c) = 1 - p$. Definujme náhodnú premennú I_B (indikátor náhodnej udalosti) takto

$$\mathbf{I}_B = \begin{cases} 1 & \text{ak nastala udalosť } B \\ 0 & \text{ak nenastala udalosť } B \end{cases}$$

Distribučná funkcia náhodnej premennej I_B je definovaná takto

$$F_{\mathbf{I}_B}(x) = \mathcal{P}(\mathbf{I}_B < x) = \begin{cases} 0 & \text{ak } x \leq 0, \\ 1 - p & \text{ak } 0 < x \leq 1, \\ 1 & \text{ak } x > 1 \end{cases}$$

Zobrazte graf $F_{\mathbf{I}_B}(x)$.

Tvrdenie 5

Distribučná funkcia $F_{\mathbf{X}}$ je funkcia

a) kde $\forall x \in \Re$ platí
$$0 \leq F_{\mathbf{X}}(x) \leq 1. \tag{4}$$

b) neklesajúca; $\forall x_1, x_2, x_1 < x_2 : F_{\mathbf{X}}(x_1) \leq F_{\mathbf{X}}(x_2)$.

c) zľava spojitá; $\forall a \in \Re : \lim\limits_{x \to a^-} F_{\mathbf{X}}(x) = F_{\mathbf{X}}(a)$.

d) splňuje
$$\lim_{x \to -\infty} F_{\mathbf{X}}(x) = 0, \quad \lim_{x \to \infty} F_{\mathbf{X}}(x) = 1. \tag{5}$$

e) pre ktorú platí
$$\mathcal{P}(\mathbf{X} = a) = F_{\mathbf{X}}(a + 0) - F_{\mathbf{X}}(a), \quad \forall a \in \Re, \tag{6}$$
$$\mathcal{P}(a \leq \mathbf{X} < b) = F_{\mathbf{X}}(b) - F_{\mathbf{X}}(a), \quad \forall a, b \in \Re, a < b. \tag{7}$$

f) má nanajvýš spočítateľne veľa bodov nespojitosti.

Príklad 3.4

Autobusy MHD odchádzajú zo zastávky v 8-minútových intervaloch. Cestujúci môže prísť na zastávku v ľubovoľnom okamihu.

Náhodná premenná $\mathbf{X}$ =„doba čakania cestujúceho na príchod autobusu“ môže nadobúdať hodnoty z intervalu $\langle 0, 8\rangle$. Predpokladajme, že všetky doby z tohoto intervalu sú rovnako pravdepodobné.

Distribučná funkcia náhodnej premennej $\mathbf{X}$ je definovaná takto

$$F_{\mathbf{X}}(x) = \mathcal{P}(\mathbf{X} < x) = \begin{cases} 0 & \text{ak } x \leq 0, \\ \frac{x}{8} & \text{ak } 0 < x \leq 8, \\ 1 & \text{ak } x > 8 \end{cases}$$

Zobrazte graf $F_{\mathbf{X}}(x)$.

Distribučná funkcia diskrétnej náhodnej premennej

Ak náhodná premenná $\mathbf{X}$ nadobúda hodnoty z nanajvýš spočitateľnej množiny H, tak hovoríme o diskrétnej náhodnej premennej, jej distribučná funkcia

$$F_{\mathbf{X}}(x) = \sum_{x_i \in H : x_i < x} \mathcal{P}(\mathbf{X} = x_i), \tag{8}$$

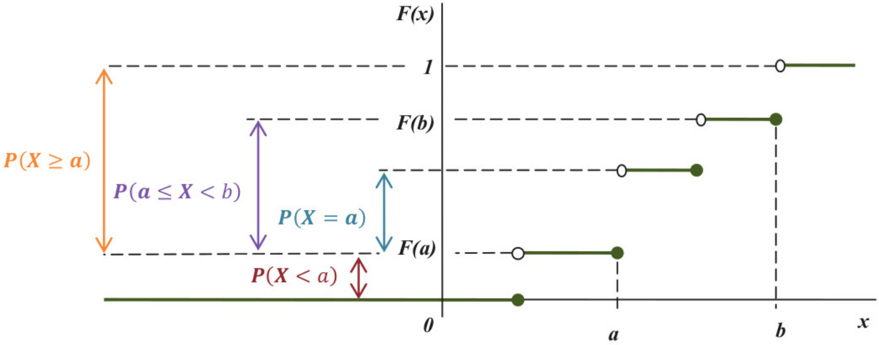
je po častiach konštantná so skokmi v bodoch $x_i \in H$.

Funkciu $p(x_i) = \mathcal{P}(\mathbf{X} = x_i), x_i \in H$ takú, že

$$\sum_{x_i \in H} p(x_i) = 1, \tag{9}$$

nazývame pravdepodobnostnou funkciu náhodnej premennej $\mathbf{X}$.

Interpretácia distribučnej funkcie diskrétnej náhodnej premennej



Obr.: *Graf distribučnej funkcie $F(x)$*

Príklad 3.5

V dielni pracujú nezávisle dva stroje. Prvý stroj má prestoj s pp. 0.2 a druhý s pp. 0.3. Určte pravdepodobnostnú a distribučnú funkciu náhodnej premennej označujúcej počet strojov v prestoji.

Nech $\mathbf{X}$ =„počet strojov v prestoji“, S_1 =„1.stroj v prestoji“ a S_2 =„2.stroj v prestoji“. Zo zadania $\mathcal{P}(S_1) = 0.2, \mathcal{P}(S_2) = 0.3$. Z nezávislosti náhodných udalostí S_1 a S_2 máme pp. funkciu

$$p(x) = \begin{cases} \mathcal{P}(S_1^c) \cdot \mathcal{P}(S_2^c) = 0.56 & \text{ak } x = 0, \\ \mathcal{P}(S_1) \cdot \mathcal{P}(S_2^c) + \mathcal{P}(S_1^c) \cdot \mathcal{P}(S_2) = 0.38 & \text{ak } x = 1, \\ \mathcal{P}(S_1) \cdot \mathcal{P}(S_2) = 0.06 & \text{ak } x = 2. \end{cases}$$

Distribučná funkcia má tvar „schodovitej funkcie“

$$F_{\mathbf{X}}(x) = \mathcal{P}(\mathbf{X} < x) = \begin{cases} 0 & \text{ak } x \leq 0, \\ p(0) & \text{ak } 0 < x \leq 1, \\ p(0) + p(1) & \text{ak } 1 < x \leq 2, \\ 1 & \text{ak } x > 2. \end{cases}$$

Alternatívne rozdelenie

Uvažujme náhodnú premennú $\mathbf{X}$, ktorá nadobúda hodnotu 1 s pp. p, $(0 < p < 1)$ ak pri náhodnom pokuse došlo k výskytu sledovanej udalostia a 0 v opačnom prípade.

Máme $\Omega = \{0, 1\}$ pričom elementárna udalosť $\omega \in \Omega$ udáva počet úspechov v pokuse.

Hovoríme, že náhodná premenná $\mathbf{X}$ má alternatívne rozdelenie s pravdepodobnosťou p, píšeme $\mathbf{X} \sim A(p)$, ak má pp. funkciu

$$p(x) = \mathcal{P}(\mathbf{X} = x) = \begin{cases} p & \text{ak } x = 1, \\ 1 - p & \text{ak } x = 0. \end{cases} \tag{10}$$

Uvažujme n nezávislých pokusov, v každom môže nastať úspech s pp. p a neúspech s pp. $1-p$. Zvolme $\Omega = \{0,1\}^n$. Elementárna udalosť má potom tvar $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, kde ω_i je počet úspechov v i-tom pokuse.

Binomické rozdelenie má náhodná premenná $\mathbf{X}(\omega) = \sum_{i=1}^{n} \omega_i$, udávajúca celkový počet úspechov v n pokusoch. Pre každý z pokusov platí $\mathcal{P}(\omega_i) = p^{\omega_i}(1-p)^{1-\omega_i}$. Z nezávislosti pokusov dostaneme $\mathcal{P}(\omega) = \prod_{i=1}^{n} \mathcal{P}(\omega_i) = p^{\sum_i \omega_i}(1-p)^{1-\sum_i \omega_i}$. Pretože máme $\binom{n}{k}$ elementárnych udalostí, pre ktoré je $\sum_{i=1}^{n} \omega_i = k$, dostávame vzťah

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{11}$$

Tento vzťah udáva rozdelenie náhodnej premennej s binomickým rozdelením s parametrami n, p kde $n \in \mathcal{N}, 0 < p < 1$ čo stručne zapisujeme $\mathbf{X} \sim Bi(n,p)$.

Príklad 3.6

Päť krát hodíme mincou. Pomocou distribučnej funkcie nejakého rozdelenia vyjadrite pravdepodobnosť, že aspoň 2x padol znak.

$\mathbf{X}$ = "počet znakov"$\sim Bi(5, \frac{1}{2})$.

$$p = \mathcal{P}(\mathbf{X} \geq 2) = 1 - \mathcal{P}(\mathbf{X} < 2) = 1 - F(2)$$
$$= 1 - \frac{\binom{5}{0}}{2^5} - \frac{\binom{5}{1}}{2^5} = 1 - \frac{1+5}{32} = \frac{13}{16},$$

kde F je distribučná funkcia binomického rozdelenia $Bi(5, \frac{1}{2})$.

Príklad 3.7 – Newton (1693)

Hazardní hráči A, B, C hádžu spravodlivými kockami v troch krabičkách, každý jeden krát. Pričom hráč

- A má v krabičke 6 kociek a chce hodiť šestku,
- B má v krabičke 12 kociek a chce hodiť dve šestky,
- C má v krabičke 18 kociek a chce hodiť tri šestky,

Majú hráčí A, B a C rovnako ťažkú úlohu pri rovnakom šťastí? Nech $p = \frac{1}{6}$ pp. hodu šestky, ináč $q = \frac{5}{6}$. Pp. úspechu hráčov:

$$\mathcal{P}(A) = \sum_{k=1}^{6} \binom{6}{k} p^k q^{6-k} = 1 - \binom{6}{0} p^0 q^6 = 0.6651,$$

$$\mathcal{P}(B) = \sum_{k=2}^{12} \binom{12}{k} p^k q^{6-k} = 1 - \binom{12}{0} p^0 q^{12} - \binom{12}{1} p^1 q^{11} = 0.6187,$$

$$\mathcal{P}(C) = 1 - \sum_{k=0}^{2} \binom{18}{k} p^k q^{18-k} = 0.5973.$$

Pretože $\mathcal{P}(A) > \mathcal{P}(B) > \mathcal{P}(C)$, úlohy hráčov nie sú rovnako ťažké.

Hypergeometrické rozdelenie

Predpokladajme, že v základnom súbore je N prvkov, z ktorých má M sledovanú vlastnosť a zvyšných $N-M$ prvkov túto vlastnosť nemá. Náhodne vyberieme zo základného súboru n prvkov tak, že žiadny nevraciame späť. Nech náhodná veličina $\mathbf{X}$ udáva počet prvkov s danou vlastnosťou vo výbere n z N prvkov.

Hovoríme, že náhodná premenná $\mathbf{X}$ má hypergeometrické rozdelenie s parametrami N, M, n, píšeme $\mathbf{X} \sim HG(N, M, n)$, ak má pp. funkciu

$$p(k) = \mathcal{P}(\mathbf{X} = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad (12)$$

$$k \in \left\{ \max\{0, M+n-N\}, min\{M, n\} \right\}.$$

Poznámka:
Hypergeometrické rozdelenie je základným rozdelením pri výbere bez vracania.

Príklad 3.8

Medzi 200 vyrobenými súčiastkami je 50 chybných. Aká je pp., že pri nákupe 20 súčiastok bude 8 chybných?

Ide o výber bez vracania pričom sú jednotlivé výbery nezávislé. Definujme náhodnú premennú $\mathbf{X}$ =„počet chybných súčiastok medzi k vybranými".

- $N = 200$ – počet vyrobených súčiastok,
- $M = 50$ – počet vyrobených chybných súčiastok,
- $n = 20$ – počet kúpených súčiastok,
- $k = 8$ – počet chybných kúpených súčiastok.

Teda $\mathbf{X} \sim HG(200, 50, 20)$ a

$$p(8) = \mathcal{P}(\mathbf{X} = 8) = \frac{\binom{50}{8}\binom{200-50}{20-8}}{\binom{200}{20}} = 0.057.$$

Pravdepodobnosť, že medzi 20 súčiastkami bude 8 chybných je 0.057.

Geometrické rozdelenie

Náhodná premenná $\mathbf{X}$ počet nezávislých pokusov do výskytu prvej úspešnej udalosti, ak pravdepodobnosť úspechu je $p, (0 < p < 1)$.

Hovoríme, že náhodná premenná $\mathbf{X}$ má geometrické rozdelenie s parametrom p, píšeme $\mathbf{X} \sim G(p)$, ak má pp. funkciu

$$p(k) = \mathcal{P}(\mathbf{X} = k) = (1 - p)^{k-1} p, \; k = 1, 2, \ldots . \tag{13}$$

Poznámka:
Definícia geometrického rozdelenia nie je ustálená. V niektorých publikáciách a štatistických sofvéroch je $\mathbf{X} =$„počet úspešných udalostí pred prvou neúspešnou udalosťou“.

Príklad 3.9

Aká je pravdepodobnosť pri hodoch spravodlivou kockou, že aby padla šestka musíme hádzať a) $5\times$, b) viac než $3\times$?

Počet hodov potrebných k 1. úspechu (padla „6“) s pp. $p = \frac{1}{6}$ nech je náhodnou premennou $\mathbf{X}$ t.j. $\mathbf{X} \sim G\left(\frac{1}{6}\right)$.

ada) Pravdepodobnosť, že padne „6“ v šiestom hode

$$p(5) = \mathcal{P}(\mathbf{X} = 5) = \left(1 - \frac{1}{6}\right)^{5-1} \cdot \frac{1}{6} \approx 0.08.$$

adb) Pravdepodobnosť, že padne „6“ treba hádzať viac než $3\times$

$$\mathcal{P}(\mathbf{X} > 3) = 1 - \mathcal{P}(\mathbf{X} \leq 3) = 1 - p(1) - p(2) - p(3) \approx 0.578.$$

Negatívne binomické (Pascalovo) rozdelenie

Náhodná premenná $\mathbf{X}$, počet nezávislých pokusov do výskytu k úspešných udalostí, ak pravdepodobnosť úspechu je $p, (0 < p < 1)$.

Hovoríme, že náhodná premenná $\mathbf{X}$ má negatívne binomické rozdelenie s parametrami k a p, píšeme $\mathbf{X} \sim NBi(k, p)$, ak má pp. funkciu

$$p(n) = \mathcal{P}(\mathbf{X} = n) = \binom{n-1}{k-1}(1-p)^{n-k}p^k, \ k \leq n < \infty. \qquad (14)$$

Príklad 3.10

V nemocnici urgentne potrebujú troch darcov krvy skupiny $A+$, ktorá sa vyskytuje u 35% populácie. Ochotných darcov je dosť no nepoznajú spoľahlivo svoju krvnú skupinu.

Aká je pp., že pre nájdenie 3 vhodných darcov treba

a) 10 darcov:
$\mathbf{X}$ =„počet darcov na získanie 3 so skupinou $A+$“
$p(10) = \binom{10-1}{3-1}(1-0.35)^{10-3}(0.35)^3 \approx 0.076.$

b) viac než 9 darcov:
$\mathcal{P}(\mathbf{X} > 9) = 1 - \mathcal{P}(\mathbf{X} \leq 9) =$
$1 - \sum_{n=3}^{9} \binom{n-1}{2}(0.65)^{n-3}(0.35)^3 \approx 0.337.$

c) viac než 7 a menej ako 12 darcov:
$\mathcal{P}(7 < \mathbf{X} < 12) = \sum_{n=8}^{11} \binom{n-1}{2}(0.65)^{n-3}(0.35)^3 \approx 0.332.$

Hovoríme, že náhodná premenná $\mathbf{X}$ má Poissonovo rozdelenie s parametrom $\lambda > 0$, ak má pravdepodobnostnú funkciu

$$p(k) = \mathcal{P}(\mathbf{X} = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots \tag{15}$$

čo značíme $\mathbf{X} \sim Po(\lambda)$.

Presvedčte sa o platnosti vzťahu

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1.$$

Poznámka:
S týmto rozdelením sa stretneme pre sledovaní počtu výskytu náhodnej udalosti behom konkrétneho časového intervalu (počet porúch, katastrof,...) kde λ udáva priemerný počet udalostí, za časovú jednotku.

O aproximácii binomického rozdelenia Poissonovým

Tvrdenie 6 (Poissonova veta)

Nech $\mathbf{X}_n \sim Bi(n, p_n)$, kde $\lim\limits_{n\to\infty} np_n = \lambda \in (0,\infty)$ a $p_n \in (0,1)$, nech $\mathbf{X} \sim Po(\lambda)$. Potom pre $k = 0, 1, 2, \dots$ platí

$$\lim_{n\to\infty} \mathcal{P}(\mathbf{X}_n = k) = \mathcal{P}(\mathbf{X} = k).$$

Dôkaz: Pre $k = 0, 1, 2, \dots$ platí

$$\begin{aligned}\mathcal{P}(\mathbf{X}_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \frac{np_n \cdot (n-1)p_n \cdots (n-k+1)p_n}{(1-p_n)^k} \left(1 - \frac{np_n}{n}\right)^n.\end{aligned}$$

Limita pravej strany je pre $n \to \infty$ rovná $\mathcal{P}(\mathbf{X} = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, pretože $\lim\limits_{n\to\infty} np_n = \lambda$, $\lim\limits_{n\to\infty} p_n = 0$ a $(1 - \frac{\lambda}{n})^n$ je postupnosť funkcií konvergujúcej k $e^{-\lambda}$ rovnomerne na každom obmedzenom intervale.

Príklad 3.11

V aparatúre dochádza k výmene 10-tich súčiastok za rok. Aká je pravdepodobnosť, že v priebehu 1000 hodín dojde k vyradeniu aparatúry v dôsledku poruchy súčiastky.

Počet porúch $\mathbf{X}$ behom t hodín má $Po(\lambda)$, kde $\lambda = \frac{10}{365 \cdot 24} t$.

$$\mathcal{P}(\mathbf{X} \geq 1) = 1 - \mathcal{P}(\mathbf{X} = 0) = 1 - e^{-\lambda} = 1 - e^{-\frac{10000}{365 \cdot 24}} = 0.681$$

Cvičenie – diskrétne rozdelenia

3.1 Náhodne vyberieme jedno z čísel $1, 2, \ldots, 10$. Nech náhodná premenná $\mathbf{X}$ udáva zvyšok po vydelení tohoto čísla 4.
   a) Aká je pp., že $\mathbf{X}$ je nepárne číslo?
   b) Zostrojte pravdepodobnostnú a distribučnú funkciu.

3.2 Zostrojte distribučnú funkciu pre počet bodiek na nespravodlivej kocke, kde padá $2\times$ častejšie 6 než ostatné čísla.

3.3 Pravdepodobnosť, že súčiastka bude vyhovovať všetkým technologickým požiadavkám je 0.9. Popíšte rozdelenie počtu nevyhovujúcich súčiastok medzi 3 súčiastkami:
   a) pomocou pp. funkcie v tvare tabuľky aj v tvare vzorca,
   b) pomocou distribučnej funkcie v tvare vzorca aj grafu.

3.4* V meste boli počas 60-tich dní evidované dopravné nehody v počte dní s 0 až 6 nehodami $(2, 28, 10, 7, 5, 5, 3)$. Formulujte vzorce pre príslušnu pp. a distribučnú funkciu.

3.5* Pravdepodobnosť narodenia chlapca je 0.515. Určte taký počet detí, aby pp., že medzi nimi bude aspoň jedno dievča, bola najmenej $0.9, 0.95, 0.99$.

Distribučná funkcia spojitej náhodnej premennej

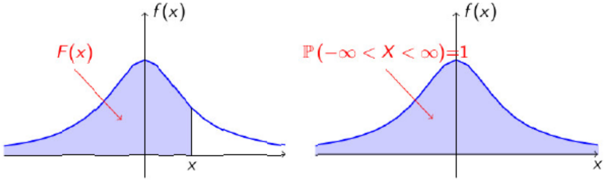
Ak náhodná premenná nadobúda všetky hodnoty z nejakého (ohraničeného aj neohraničeného) intervalu, tak hovoríme o spojitej náhodnej premennej.

Spojitá náhodná premenná **X** má distribučnú funkciu F_X, ak existuje funkcia f_X, ktorú nazývame hustota pravdepodobnosti taká, že

$$F_{\mathbf{X}}(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t. \tag{16}$$

Dohoda:
Ak nebude hroziť nedorozumenie, niekedy budeme dolný index v f_X vynechávať a hovoriť skrátene o hustote.

Geometrická interpretácia



Obr.: *Graf hustoty pravdepodobnosti* $f(x)$ *a distribučnej funkcie* $F(x)$ *spojitej náhodnej premennej* **X**

Tvrdenie 7

Nech $f_{\mathbf{X}}$ je hustota pravdepodobnosti a $F_{\mathbf{X}}$ distribučná funkcia spojitej náhodnej premennej $\mathbf{X}$. Potom platí

- a)
$$\int_{-\infty}^{\infty} f_{\mathbf{X}}(t)\mathrm{d}t = 1, \tag{17}$$
- b) $f_{\mathbf{X}}(x) = F'_{\mathbf{X}}(x)$ pre každé $x \in \Re$ v ktorom je $f_{\mathbf{X}}$ spojitá,
- c) $f_{\mathbf{X}}(x) \geq 0$ pre každé $x \in \Re$,
- d)
$$\mathcal{P}(a \leq \mathbf{X} < b) = \int_{a}^{b} f_{\mathbf{X}}(t)\mathrm{d}t. \tag{18}$$

Dôkaz:

- a)

$$\int_{-\infty}^{\infty} f_{\mathbf{X}}(t)\mathrm{d}t = F_{\mathbf{X}}(\infty) = \mathcal{P}(\Omega) = 1,$$

- b) $f_{\mathbf{X}}(x) = F'_{\mathbf{X}}(x)$ vyplýva z definície $F_{\mathbf{X}}(x) = \int_{-\infty}^{x} f_{\mathbf{X}}(t)\mathrm{d}t$.
- c) $f_{\mathbf{X}}(x) = F'_{\mathbf{X}}(x) = \lim_{\delta \to 0} \frac{F_{\mathbf{X}}(x+\delta) - F_{\mathbf{X}}(x)}{\delta} \geq 0$, $f_{\mathbf{X}}(x)$ je limitou nezápornej funkcie a tak je tiež nezáporná.
- d) Vo vzťahu (7) je $\mathcal{P}(a \leq \mathbf{X} < b) = F_{\mathbf{X}}(b) - F_{\mathbf{X}}(a)$

$$F_{\mathbf{X}}(b) - F_{\mathbf{X}}(a) = \int_{-\infty}^{b} f_{\mathbf{X}}(t)\mathrm{d}t - \int_{-\infty}^{a} f_{\mathbf{X}}(t)\mathrm{d}t = \int_{a}^{b} f_{\mathbf{X}}(t)\mathrm{d}t.$$

Príklad 3.10

Distribučná funkcia spojitej náhodnej premennej $\mathbf{X}$ je určená takto

$$F(x) = \begin{cases} 0 & \text{ak } x \le 0, \\ cx^2 & \text{ak } 0 < x \le 10, \\ 1 & \text{ak } x > 10. \end{cases}$$

Určme parameter c, hustotu pravdepodobnosti a $\mathcal{P}(2 < \mathbf{X} < 6)$.

Distribučná funkcia F musí byť v bode 10 spojitá a tak $F(10) = c \cdot 100 = 1$ odkiaľ $c = \frac{1}{100}$.
Podľa tvrdnie 7b) je hustota pravdepodobnosti $f(x) = F'(x)$

$$f(x) = \begin{cases} 0 & \text{ak } x \le 0, \\ \frac{x}{50} & \text{ak } 0 < x \le 10, \\ 0 & \text{ak } x > 10. \end{cases}$$

$$\mathcal{P}(2 < \mathbf{X} < 6) = \mathcal{P}(2 \le \mathbf{X} < 6) = F(6) - F(2) = \frac{36}{100} - \frac{4}{100} = 0.32$$

Príklad 3.11

Náhodná premenná $\mathbf{X}$ má hustotu pp. $f(x)$ pre $a, b > 0, c \in \Re$

$$f(x) = ae^{-b|x-c|}, x \in \Re$$

Aký je vzťah medzi konštantami a, b, aký ma tvar distribučná funkcia $F(x)$ a čomu je rovná $\mathcal{P}(c - 1 < \mathbf{X} < c + 1)$?

Hustotu pravdepodobnosti môžeme prepísať takto

$$f(x) = \begin{cases} ae^{b(x-c)} & \text{ak } -\infty < x < c, \\ ae^{-b(x-c)} & \text{ak } c \leq x < \infty. \end{cases}$$

Zo vzťahu (17)

$$a\Big(\int\limits_{-\infty}^{c} e^{b(x-c)}\mathrm{d}x + \int\limits_{c}^{\infty} e^{-b(x-c)}\mathrm{d}x\Big) = 1,$$

pomocou substitúcie $y = b(x - c), \mathrm{d}x = \frac{\mathrm{d}y}{b}$ dostávame $a = \frac{1}{2}b$.

Potom distribučná funkcia

$$F(x) = \begin{cases} \int\limits_{-\infty}^{x} \frac{b}{2} e^{b(t-c)} \mathrm{d}t = \frac{1}{2} e^{b(x-c)} & \text{ak } -\infty < x < c, \\ \frac{1}{2} + \int\limits_{c}^{x} \frac{b}{2} e^{-b(t-c)} \mathrm{d}t = 1 - \frac{1}{2} e^{-b(x-c)} & \text{ak } c \leq x < \infty. \end{cases}$$

Pravdepodobnosť, že náhodná premenná $\mathbf{X}$ nadobudne hodnoty z intervalu $(c-1, c+1)$ je

$$\mathcal{P}(c-1 < \mathbf{X} < c+1) = F(c+1) - F(c-1) =$$
$$= 1 - \frac{1}{2} e^{-b} - \frac{1}{2} e^{-b} = 1 - e^{-b}.$$

Rovnomerné rozdelenie

Jedná sa o spojité rozdelenie, ktorého hustota je konštantná na nejakom intervale (a, b).

Hovoríme, že náhodná premenná $\mathbf{X}$ má na intervale (a, b) rovnomerné rozdelenie, píšeme $\mathbf{X} \sim R(a, b)$, ak má

- hustotu pravdepodobnosti:

$$f_{\mathbf{X}}(x) = \begin{cases} \frac{1}{b-a} & \text{ak } x \in (a, b), \\ 0 & \text{ak } x \notin (a, b). \end{cases}$$

- distribučnú funkciu:

$$F_{\mathbf{X}}(x) = \begin{cases} 0 & \text{ak } x \leq a, \\ \frac{x-a}{b-a} & \text{ak } x \in (a, b), \\ 1 & \text{ak } x \geq b. \end{cases}$$

Poznámka:

$R(0, 1)$ zohráva základnú úlohu ako modelovacia technika pri simulácii náhodných udalostí a náhodných procesov.

Príklad 3.11

Majme $\mathbf{X} \sim R(-2,2)$. Formulujte f_X a F_X vzorcom.

Hustota pravdepodobnosti náhodnej premennej $\mathbf{X}$ na intervale $(-2,2)$ s rovnomerným rozdelením

$$f_{\mathbf{X}}(x) = \begin{cases} \frac{1}{4} & \text{ak } x \in (-2,2), \\ 0 & \text{ak } x \notin (-2,2). \end{cases}$$

a jej distribučná funkcia

$$F_{\mathbf{X}}(x) = \begin{cases} 0 & \text{ak } x \leq -2, \\ \frac{x+2}{4} & \text{ak } x \in (-2,2), \\ 1 & \text{ak } x \geq 2. \end{cases}$$

Exponenciálne rozdelenie

Hovoríme, že náhodná premenná $\mathbf{X}$ má exponenciálne rozdelenie s parametrom $\lambda, (\lambda > 0)$, píšeme $\mathbf{X} \sim Exp(\lambda)$, ak má

- hustotu pravdepodobnosti:

$$f_{\mathbf{X}}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{ak } t > 0, \\ 0 & \text{ak } t \leq 0. \end{cases}$$

- distribučnú funkciu:

$$F_{\mathbf{X}}(t) = \begin{cases} 1 - e^{-\lambda t} & \text{ak } t > 0, \\ 0 & \text{ak } t \leq 0. \end{cases}$$

Poznámka:

Exponenciálne rozdelenie hrá dôležitú úlohu v teórii spoľahlivosti (doba do poruchy) a v teórii hromadnej obsluhy (doba čakania v rade). Je totiž vhodným rozdelením pre popis „doby do výskytu prvej udalosti“ resp. „doby medzi udalosťami“.

Bezpamäťová vlastnosť

Príklad 3.12

Ak modelujeme dobu do poruchy $\mathbf{T} \sim Exp(\lambda)$, potom pp., že systém pracoval bez poruchy do doby t_1 a bude ešte pracovať bez poruchy aspoň dobu t_2, sa rovná pp., že systém, ktorý nebol v činnosti, bude pracovať aspoň dobu t_2. Platí totiž

$$\begin{aligned}
\mathcal{P}(\mathbf{T} > t_1 + t_2 | \mathbf{T} > t_1) &= \frac{\mathcal{P}([\mathbf{T} > t_1 + t_2] \cap [\mathbf{T} > t_1])}{\mathcal{P}([\mathbf{T} > t_1])} = \\
&= \frac{\mathcal{P}([\mathbf{T} > t_1 + t_2])}{\mathcal{P}([\mathbf{T} > t_1])} = \frac{e^{-\lambda(t_1+t_2)}}{e^{-\lambda t_1}} = \\
&= e^{-\lambda t_2} = \mathcal{P}(\mathbf{T} > t_2).
\end{aligned}$$

Poznámka:

Exponenciálne rozdelenie dobre popisuje rozdelenie doby života sytémov, v ktorých dochádza k poruche z úplne náhodných príčin, nie v dôsledku opotrebovania.

Hovoríme, že náhodná premenná $\mathbf{X}$ má normálne rozdelenie s parametrami μ, σ^2 $(\mu, \sigma > 0)$, píšeme $\mathbf{X} \sim N(\mu, \sigma^2)$, ak má pre všetky $x \in \Re$

- hustotu pravdepodobnosti:

$$f_{\mathbf{X}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

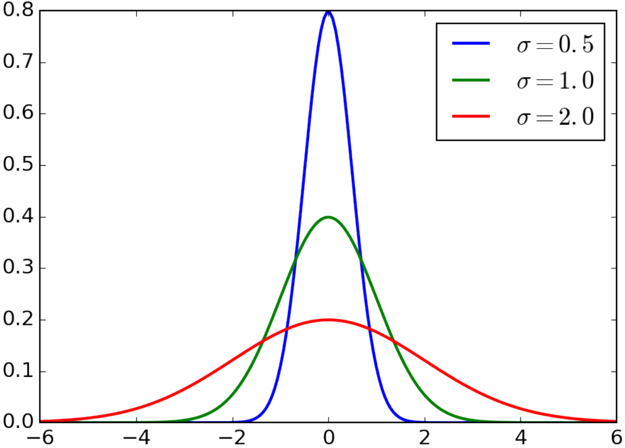
- distribučnú funkciu:

$$F_{\mathbf{X}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t.$$

Poznámka:
$F_{\mathbf{X}}(x)$ nevieme napísať v tvare vzorca ako u predchádzajúcich spojitých rozdelení.

Príklad 3.14

Zostrojte grafy hustôt pravdepodobnosti $f_{\mathbf{X}_1}, f_{\mathbf{X}_2}, f_{\mathbf{X}_3}$ pre $\mathbf{X}_1 \sim N(0, 0.25), \mathbf{X}_2 \sim N(0, 1), \mathbf{X}_3 \sim N(0, 4)$.



Normované normálne rozdelenie

Hovoríme, že náhodná premenná Z má normované (štandartizované) normálne rozdelenie, píšeme $\mathbf{Z} \sim N(0,1)$, ak má pre všetky $z \in \Re$

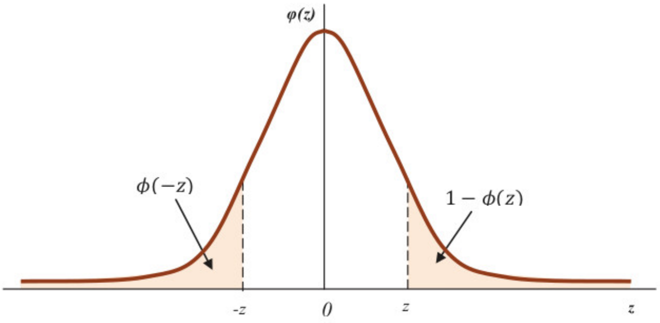
- hustotu pravdepodobnosti:

$$\phi_{\mathbf{Z}}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$

- distribučnú funkciu:

$$\Phi_{\mathbf{Z}}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

Poznámka:
Význam $N(0,1)$ je v tom, že je tabelovaná jeho distribučná funkcia.



Obr.: *Graf pravdepodobnostnej funkcie* $\phi(z)$

Hustota pravdepodobnosti $\phi(z)$ je párna funkcia, takže platí $\phi(-z) = \phi(z)$ pre $-\infty < z < \infty$. Pre distribučnú funkciu $\Phi(z)$ platí $\Phi(z) = 1 - \Phi(-z)$ pre $z \geq 0$.

Vzťah medzi $N(\mu, \sigma^2)$ a $N(0,1)$

Medzi distribučnou funkciou normálnej náhodnej premennej $\mathbf{X} \sim N(\mu, \sigma^2)$ a normovanej náhodnej premennej $\mathbf{Z} \sim N(0,1)$ platí prevodový vzťah

$$F_{\mathbf{X}}(x) = \Phi_{\mathbf{Z}}\left(\frac{x-\mu}{\sigma}\right). \tag{19}$$

Platí

$$F_{\mathbf{X}}(x) = \mathcal{P}(\mathbf{X} < x) = \mathcal{P}(\mathbf{Z}\sigma + \mu < x) = \mathcal{P}\left(\mathbf{Z} < \frac{x-\mu}{\sigma}\right) = \Phi_{\mathbf{Z}}\left(\frac{x-\mu}{\sigma}\right).$$

Príklad 3.15

Bolo zistené, že výsledky bodového hodnotenia testov študentov má rozdelenie $N(83, 64)$. Hoci bolo použité normálne rozdelenie, potrebujeme zistiť pp., s akou sú od 75 do 95 bodov.

Máme $\mathbf{X} \sim N(83, 64)$.

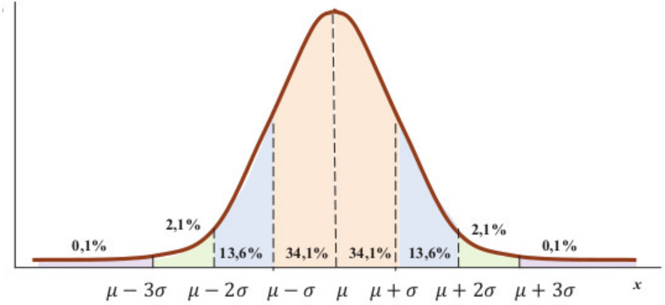
$$
\begin{aligned}
\mathcal{P}(75 \leq \mathbf{X} \leq 95) &= \mathcal{P}(\mathbf{X} \leq 95) - \mathcal{P}(\mathbf{X} \leq 75) \\
&= F_{\mathbf{X}}(95) - F_{\mathbf{X}}(75) = \Phi_{\mathbf{Z}}\left(\frac{95-83}{8}\right) - \Phi_{\mathbf{Z}}\left(\frac{75-83}{8}\right) \\
&= \Phi_{\mathbf{Z}}(1.5) - \Phi_{\mathbf{Z}}(-1) = 0.9332 - 1.1587 = 0.7745
\end{aligned}
$$

Ako sa zmení pp. ak hladáme bodové hodnotene v rozsahu 83 ± 8?

$$
\begin{aligned}
\mathcal{P}(75 \leq \mathbf{X} \leq 91) &= F_{\mathbf{X}}(95) - F_{\mathbf{X}}(91) \\
&= \Phi_{\mathbf{Z}}\left(\frac{91-83}{8}\right) - \Phi_{\mathbf{Z}}\left(\frac{75-83}{8}\right) \\
&= \Phi_{\mathbf{Z}}(1) - \Phi_{\mathbf{Z}}(-1) = 0.8413 - 1.1587 = 0.6826
\end{aligned}
$$

Pravidlo 3σ

Pravidlo 3σ je jedným zo základných princípov, na ktorých je založená kontrola kvality a akosti. Pravidlo hovorí, že ak máme údaje, ktoré sú z rozdelenia $N(\mu, \sigma^2)$, potom takmer všetky (99.8% z nich) ležia v intervale $(\mu - 3\sigma, \mu + 3\sigma)$.



Obr.: *Pravdepodobnosť výskytu realizácie normálnej náhodnej premennej vo vyznačenom intervale*

Príklad 3.16

Určte pravdepodobnosť, že náhodná premenná $\mathbf{X} \sim N(\mu, \sigma^2)$ bude mať hodnotu v intervale $(\mu - k\sigma, \mu + k\sigma)$.

Pre $k > 0$:

$$\begin{aligned}
\mathcal{P}(\mu - k\sigma < \mathbf{X} < \mu + k\sigma) &= F_{\mathbf{X}}(\mu + k\sigma) - F_{\mathbf{X}}(\mu - k\sigma) \\
&= \Phi_{\mathbf{Z}}\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi_{\mathbf{Z}}\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\
&= \Phi_{\mathbf{Z}}(k) - \Phi_{\mathbf{Z}}(-k) = \Phi_{\mathbf{Z}}(k) - (1 - \Phi_{\mathbf{Z}}(k)) \\
&= 2\Phi_{\mathbf{Z}}(k) - 1.
\end{aligned}$$

k	$\mathcal{P}(\mu - k\sigma < \mathbf{X} < \mu + k\sigma)$
1	0.682
2	0.954
3	0.998

3.6 Náhodná premenná $\mathbf{X}$ má distribučnú funkciu

$$F_{\mathbf{X}}(x) = \begin{cases} 0 & \text{ak } x \leq -5, \\ c(x+5) & \text{ak } -5 < x \leq 2, \\ 1 & \text{ak } x > 2. \end{cases}$$

Určte a) konštantu c, b) $\mathcal{P}(-6 < \mathbf{X} < 2)$ a $\mathcal{P}(\mathbf{X} = 2)$, c) zobrazte $f_{\mathbf{X}}$, $F_{\mathbf{X}}$.

3.7 Náhodná premenná $\mathbf{Y}$ má distribučnú funkciu

$$F_{\mathbf{Y}}(y) = \begin{cases} 0 & \text{ak } y \leq 0, \\ a + b\sin(y) & \text{ak } 0 < y \leq \frac{\pi}{2}, \\ 1 & \text{ak } y > \frac{\pi}{2}. \end{cases}$$

Čomu sa rovnajú konštanty a, b a aký tvar má hustota pp.?

3.8* Aká je pravdepodobnosť, že po 200 hodinách prevádzky budú fungovať aspoň 3 výrobky z 5, ak ich životnosť v hodinách má $N(180, 400)$?