# PRAVDEPODOBNOSŤ a ŠTATISTIKA

## Charakteristiky náhodnej premennej

doc. RNDr. Štefan Peško, CSc.

Katedra matematických metód a operačnej analýzy, FRI ŽU

29. marca 2019

Číselnými charakteristikami rozdelenia rozumieme čísla, pomocou ktorých získavame informácie o náhodnej premennej. Delíme ich na

- charakteristiky polohy – popisujú istý druh „stredu“ rozdelenia, okolo ktorého kolíšu hodnoty náhodnej premennej – stredná hodnota, medián, modus, kvantil,
- charakteristiky variability – popisujú rozptýlenosť náhodnej premennej okolo strednej hodnoty – disperzia, smerodajná odchylka,
- doplňujúce charakteristiky – dávajú ďalšie údaje o rozptýlení hodnôt okolo strednej hodnoty – koeficient šikmosti, koeficient špicatosti.

# Stredná hodnota

Nech $\mathbf{X}$ je diskrétna náhodná premenná nadobúdajúca hodnotu $x$ s pravdepodobnosťou $p_X(x), x \in H$. Potom strednou hodnotou náhodnej premennej $\mathbf{X}$ rozumieme číslo

$$E(\mathbf{X}) = \sum_{x \in H} x p_{\mathbf{X}}(x). \tag{1}$$

Nech $X$ je spojitá náhodná premenná a $f_X$ hustota pp.. Potom strednou hodnotou náhodnej premennej $\mathbf{X}$ rozumieme číslo

$$E(\mathbf{X}) = \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x) \mathrm{d}x. \tag{2}$$

ak nevlastný integrál (2) konverguje absolútne.

Poznámka:

$E(\mathbf{X})$ popisuje hodnotu, okolo ktorej náhodne kolíšu hodnoty náhodnej premennej $\mathbf{X}$.

## Príklad 5.1

Hádžeme spravodlivou kockou. Ak padne párne číslo, vyhráme toľko bodov, koľko padlo na kocke a v prípade nepárneho čísla zodpovedajúcu sumu strácame. Je pre nás táto hra výhodná?

Náhodná premenná $X$ nadobúda hodnoty z možných výhier $x \in H = \{-1, 2, -3, 4, -5, 6\}$ s pp. $p(x) = \frac{1}{6}$. Strednú hodnotu vypočítame podľa (1)

$$E(\mathbf{X}) = -1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} - 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} - 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{1}{2}.$$

Pri dlhodobom hraní získame z každej hry 50 centov a tak sa nám túto hru oplati hrať.

## Príklad 5.2

Trolejbusy MHD odchádzajú zo zastávky v 10-minútových intervaloch. Cestujúci môžu prísť na zastávku v ľubovoľnom okamihu. Aká je stredná doba čakania na autobus?

Doba čakania $\mathbf{T} \sim R(0, 10)$ a tak má hustotu pp.

$$f_{\mathbf{T}}(t) = \begin{cases} \frac{1}{10} & \text{ak } t \in (0, 10), \\ 0 & \text{ak } t \notin (0, 10). \end{cases}$$

Podľa vzťahu (2) pre stredn[ hodnotu dostávame

$$E(\mathbf{T}) = \int_{-\infty}^{\infty} t f_{\mathbf{T}}(t)\mathrm{d}t = \int_{0}^{10} \frac{t}{10}\mathrm{d}t = \frac{1}{10}\left[\frac{t^2}{2}\right]_0^{10} = 5.$$

Cestujúci budú v priemere čakať 5 minút, čo nie je prekvapivý ale ;-) očakávaný výsledok. Prečo?

### Tvrdenie 7

Nech $\mathbf{X}$ je náhodná premenná a $a, b \in \Re$. Potom platí

a) $E(b) = b$,

b) $E(a\mathbf{X} + b) = aE(\mathbf{X}) + b$.

Dôkaz:

Add a) Konštantu $b$ možno považovať za istú hodnotu diskrétnej náhodnej premennej, takže $E(b) = b \cdot 1 = b$.

Add b) Ak je $\mathbf{X}$ diskrétna resp. spojitá náhodná premenná a $\mathbf{Y} = a\mathbf{X} + b$, potom

$$E(\mathbf{Y}) = \sum_{x \in H} (ax + b) p_{\mathbf{X}}(x) = a \sum_{x \in H} x p_{\mathbf{X}}(x) + b \sum_{x \in H} p_{\mathbf{X}}(x) = aE(\mathbf{X}) + b,$$

*resp.*

$$E(\mathbf{Y}) = \int_{-\infty}^{\infty} (ax + b) f_{\mathbf{X}}(x) \mathrm{d}x = a \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x) \mathrm{d}x + b \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) \mathrm{d}x = aE(\mathbf{X}) + b.$$

# Začiatočné momenty

Nech $\mathbf{X}$ je diskrétna náhodná premenná nadobúdajúca hodnotu $x$ s pravdepodobnosťou $p_{\mathbf{X}}(x), x \in H$. Potom začiatočným momentom $k$-teho rádu náhodnej premennej $\mathbf{X}$ rozumieme číslo

$$\nu_k(\mathbf{X}) = \sum_{x \in H} x^k p_{\mathbf{X}}(x), k \in \mathbb{N}, \qquad (3)$$

ak rad (3) konverguje absolútne.

Nech $\mathbf{X}$ je spojitá náhodná premenná a $f_{\mathbf{X}}$ hustota pp.. Potom začiatočným momentom $k$-teho rádu náhodnej premennej $\mathbf{X}$ rozumieme číslo

$$\nu_k(\mathbf{X}) = \int_{-\infty}^{\infty} x^k f_{\mathbf{X}}(x)\mathrm{d}x, k \in \mathbb{N} \qquad (4)$$

ak nevlastný integrál (4) konverguje absolútne.

Poznámka:

Zrejme $E(\mathbf{X}) = \nu_1(\mathbf{X})$ .

## Príklad 5.3

Premenná $\mathbf{X}$ má konštantnú hustotu pp. v intervale $(0, a)$, určte

1) $E(2\mathbf{X} + 3)$,
2) $E(3\mathbf{X}^2 - 2\mathbf{X} + 1)$.

$\mathbf{X} \sim R(0, a)$ s hustotou pravdepodobnosti

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

má začiatočné momenty

$$\nu_1(\mathbf{X}) = E(\mathbf{X}) = \int_0^a \frac{x}{a} \mathrm{d}x = \frac{a}{2},$$

$$\nu_2(\mathbf{X}) = E(\mathbf{X}^2) = \int_0^a \frac{x^2}{a} \mathrm{d}x = \frac{a^2}{3}.$$

add1) $E(2\mathbf{X} + 3) = 2E(X) + 3 = a + 3$,

add2) $E(3\mathbf{X}^2 - 2\mathbf{X} + 1) = 3\nu_2(\mathbf{X}) - 2E(\mathbf{X}) + 1 = a^2 - a + 1.$

# Centrálne momenty

Nech $\mathbf{X}$ je diskrétna náhodná premenná nadobúdajúca hodnotu $x$ s pravdepodobnosťou $p_{\mathbf{X}}(x), x \in H$. Potom centrálnym momentom $k$-teho rádu náhodnej premennej $\mathbf{X}$ rozumieme číslo

$$\mu_k(\mathbf{X}) = \sum_{x \in H} (x - E(\mathbf{X}))^k p_{\mathbf{X}}(x), k \in \mathbb{N}, \tag{5}$$

ak rad (5) konverguje absolútne.

Nech $\mathbf{X}$ je spojitá náhodná premenná a $f_X$ hustota pp.. Potom centrálnym momentom $k$-teho rádu náhodnej premennej $\mathbf{X}$ rozumieme číslo

$$\mu_k(\mathbf{X}) = \int_{-\infty}^{\infty} (x - E(\mathbf{X}))^k f_X(x) \mathrm{d}x, k \in \mathbb{N}, \tag{6}$$

ak nevlastný integrál (6) konverguje absolútne.

# Rozptyl

Nech $\mathbf{X}$ je diskrétna náhodná premenná nadobúdajúca hodnotu $x$ s pravdepodobnosťou $p_{\mathbf{X}}(x), x \in H$. Potom centrálny moment druhého rádu $\mu_2(\mathbf{X})$ nazývame disperzia alebo rozptyl náhodnej premennej $\mathbf{X}$, označujeme ho $D(\mathbf{X})$ t.j.

$$D(\mathbf{X}) = \sum_{x \in H} (x - E(\mathbf{X}))^2 p_{\mathbf{X}}(x). \tag{7}$$

Nech $\mathbf{X}$ je spojitá náhodná premenná a $f_{\mathbf{X}}$ hustota pp.. Potom centrálny moment druhého rádu $\mu_2(\mathbf{X})$ nazývame disperzia alebo rozptyl náhodnej premennej $\mathbf{X}$ , označujeme ho $D(\mathbf{X})$ t.j.

$$D(\mathbf{X}) = \int_{-\infty}^{\infty} (x - E(\mathbf{X}))^2 f_{\mathbf{X}}(x) \mathrm{d}x. \tag{8}$$

Poznámka:

Rozptyl je parameter rozdelenia, ktorý vyjadruje variabilitu „rozptýlenosť“ náhodnej premennej okolo strednej hodnoty.

## Tvrdenie 8

Ak existujú začiatočné momenty prvého a druhého rádu náhodnej premennej $\mathbf{X}$ potom platí vzťah

$$D(\mathbf{X}) = E(\mathbf{X}^2) - E(\mathbf{X})^2. \tag{9}$$

Dôkaz: Nech $\mathbf{X}$ je spojitá náhodná premenná. Potom

$$\begin{aligned} D(\mathbf{X}) &= \int_{-\infty}^{\infty} (x - E(\mathbf{X}))^2 f_{\mathbf{X}}(x)\mathrm{d}x = \int_{-\infty}^{\infty} (x^2 - 2xE(\mathbf{X}) + E(\mathbf{X})^2) f_{\mathbf{X}}(x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} x^2 f_{\mathbf{X}}(x)\mathrm{d}x - 2E(\mathbf{X}) \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x)\mathrm{d}x + E(\mathbf{X})^2 \int_{-\infty}^{\infty} f_{\mathbf{X}}(x)\mathrm{d}x \\ &= E(\mathbf{X}^2) - 2E(\mathbf{X})E(\mathbf{X}) + E(\mathbf{X})^2 = E(\mathbf{X}^2) - E(\mathbf{X})^2. \end{aligned}$$

Analogicky pre diskrétnu náhodnú premennú $\mathbf{X}$.

Poznámka:

Vzťah (9) sa v praxi používa častejšie než (7) a (8).

Druhú odmocninu z disperzie náhodnej premennej $\mathbf{X}$ nazývame smerodajná odchylka tejto premennej a označujeme ju

$$\sigma(\mathbf{X}) = \sqrt{D(\mathbf{X})}. \tag{10}$$

Poznámka:
V prípade, že náhodná premenná $\mathbf{X}$ sa meria v nejakých jednotkách (napr. v metroch), tak rozptyl $D(\mathbf{X})$ má rozmer vo štvorcoch týchto jednotiek (napr. v metroch štvorcových). Preto sa v praxi používa smerodajná odchýlka $\sigma(\mathbf{X})$ , ktorej rozmer je v rovnakých jednotkách ako pôvodná náhodná premenná.

### Tvrdenie 9

Nech $\mathbf{X}$ je náhodná premenná, $a, b \in \Re$. Potom rozptyl $D(\mathbf{X})$ má tieto vlastnosti

a) $D(a) = 0$,
b) $D(a\mathbf{X}) = a^2 D(\mathbf{X})$,
c) $D(\mathbf{X} + b) = D(\mathbf{X})$,
d) $D(a\mathbf{X} + b) = a^2 D(\mathbf{X})$.

Dôkaz: Add d) Z vlastnosti strednej hodnoty a vzťahu (9)

$$
\begin{aligned}
D(a\mathbf{X} + b) &= E((a\mathbf{X} + b)^2) - E^2(a\mathbf{X} + b) \\
&= E(a^2\mathbf{X}^2 + 2ab\mathbf{X} + b^2) - \big(E(a\mathbf{X} + b)\big)^2 \\
&= a^2 E(\mathbf{X}^2) + 2abE(\mathbf{X}) + b^2 - \big(aE(\mathbf{X}) + b\big)^2 \\
&= a^2 E(\mathbf{X}^2) + 2abE(\mathbf{X}) + b^2 - a^2 E(\mathbf{X})^2 - 2abE(\mathbf{X}) - b^2 \\
&= a^2 (E(\mathbf{X}^2) - E(\mathbf{X})^2) = a^2 D(\mathbf{X}).
\end{aligned}
$$

## Príklad 5.4

Nech je náhodná premenná $\mathbf{X}$ určená hustotou pravdepodobnosti

$$f_{\mathbf{X}}(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{ak } x \in (-1,1), \\ 0 & \text{ak } x \notin (-1,1). \end{cases}$$

Určte strednú hodnotu a disperziu náhodnej veličiny $\mathbf{Y} = 5\mathbf{X} + 6$.

$$E(\mathbf{X}) = \int_{-1}^{1} x\frac{3}{4}(1-x^2)\mathrm{d}x = \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{-1}^{1} = 0,$$

$$E(\mathbf{X}^2) = \int_{-1}^{1} x^2\frac{3}{4}(1-x^2)\mathrm{d}x = \frac{3}{4}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1} = 0.2,$$

$$E(\mathbf{Y}) = E(5\mathbf{X}+6) = 5E(\mathbf{X}) + 6 = 6,$$

$$D(\mathbf{Y}) = D(5\mathbf{X}+6) = 25D(\mathbf{X}) = 25\big(E(\mathbf{X}^2) - E(\mathbf{X})^2\big) = 5.$$

# Štandartizácia (normovanie)

Hovoríme, že náhodná premenná $\mathbf{U}$ je normovaná ak $E(\mathbf{U}) = 0$ a $D(\mathbf{U}) = 1$.

## Tvrdenie 12

Nech $\mathbf{X}$ je náhodná premenná, pre ktorú existuje stredná hodnota $E(\mathbf{X})$ a disperzia $D(\mathbf{X}) > 0$. Potom je náhodná premenná

$$\mathbf{U} = \frac{\mathbf{X} - E(\mathbf{X})}{\sigma(\mathbf{X})} \tag{11}$$

normovaná náhodná premenná.

Dôkaz: Platí

$$E(\mathbf{U}) = E\left(\frac{\mathbf{X} - E(\mathbf{X})}{\sigma(\mathbf{X})}\right) = \frac{1}{\sigma(\mathbf{X})}\left(E(\mathbf{X}) - E(E(\mathbf{X}))\right) = 0,$$

$$D(\mathbf{U}) = \frac{1}{\sigma(\mathbf{X})^2} D(\mathbf{X} - E(\mathbf{X})) = \frac{D(\mathbf{X})}{D(\mathbf{X})} = 1.$$

## Príklad 5.5

Nech $\mathbf{X} \sim N(\mu, \sigma^2)$. Netriviálnym výpočtom možno odvodiť, že $E(\mathbf{X}) = \mu$ a $D(\mathbf{X}) = \sigma^2$. Overte, že $\mathbf{Z} = \frac{\mathbf{X}-E(\mathbf{X})}{\sigma(\mathbf{X})} \sim N(0,1)$.

Platí

$$E(\mathbf{Z}) = E\left(\frac{\mathbf{X} - E(\mathbf{X})}{\sigma(\mathbf{X})}\right) = \frac{1}{\sigma}(E(\mathbf{X}) - \mu) = 0,$$

$$D(\mathbf{Z}) = D\left(\frac{\mathbf{X} - E(\mathbf{X})}{\sigma(\mathbf{X})}\right) = \frac{1}{\sigma^2} D(\mathbf{X}) = 1.$$

Poznamenajme, že hustoty pravdepodobnosti sú pre náhodné premenné $\mathbf{X}$ a $\mathbf{Z}$ definovaná takto

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

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

Podiel centrálneho momentu tretieho rádu a tretej mocniny smerodajnej odchýlky náhodnej premennej **X** sa nazýva koeficient šikmosti (skewness). Označujeme ho $\alpha_1(\mathbf{X})$

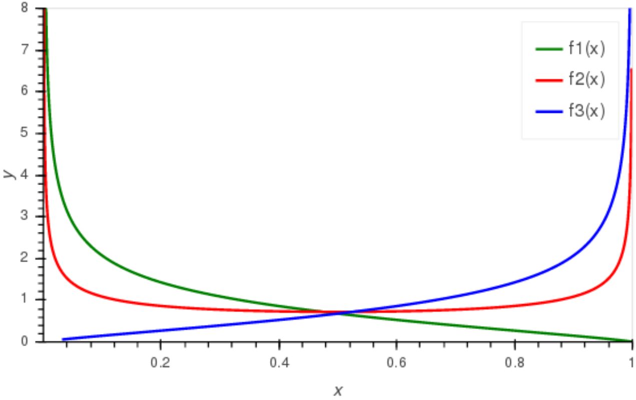
$$\alpha_1(\mathbf{X}) = \frac{\mu_3(\mathbf{X})}{\sigma(\mathbf{X})^3}. \tag{12}$$

Ak $\alpha_1 = 0$ pre symetrické rozdelenie.
Ak $\alpha_1 > 0$, pri rozdelení pretiahnutejšom doprava, hodnoty sa hromadia ľavej časti.

Ak $\alpha_1 < 0$, pri rozdelení pretiahnutejšom doľava, hodnoty sa hromadia pravej časti.

Koeficient špicatosti (kurtosis), ktorý označujeme $\alpha_2(\mathbf{X})$ je určený vzťahom

$$\alpha_2(\mathbf{X}) = \frac{\mu_4(\mathbf{X})}{\sigma(\mathbf{X})^4} - 3. \tag{13}$$




$\alpha_1(\mathbf{X}_1) = 0.968$, $\alpha_1(\mathbf{X}_2) = 0$, $\alpha_1(\mathbf{X}_3) = -0.968$
$\alpha_2(\mathbf{X}_1) = 0.034$, $\alpha_2(\mathbf{X}_2) = -1.429$, $\alpha_2(\mathbf{X}_3) = 0.034$

Poznámka:
Symetrické rozdelenie $\mathbf{X}$ má $\alpha_1(\mathbf{X}) = 0$. Z $\alpha_1(\mathbf{X}) = 0$ ale neplynie, že rozdelenie je symetrické.

# Kvantily

Číslo $x_\alpha, \alpha \in (0,1)$ nazývame $\alpha$- kvantil náhodnej premennej $\mathbf{X}$ (tiež $\alpha$%- kvantil) ak platí

$$\mathcal{P}(\mathbf{X} \le x_\alpha) = \alpha, \ \mathcal{P}(\mathbf{X} \ge x_\alpha) = 1 - \alpha. \tag{14}$$

- median: $x_{0.5}$,
- dolný kvantil: $x_{0.25}$,
- horný kvantil: $x_{0.75}$,
- percentil: $x_{0.01}, x_{0.02}, \ldots, x_{0.99}$.

Vybrané kvantily dôležitých rozdelení v štatistke sú tabelované pod názvom kritické hodnoty.

Poznámka:
Pre spojitú rastúcu distribučnú funkciu $F_{\mathbf{X}}(x)$ platí $F_{\mathbf{X}}(x_\alpha) = \alpha$ odkiaľ $x_\alpha = F_{\mathbf{X}}^{-1}(\alpha)$. $\alpha$%- kvantil je hodnota, ktorá delí plochu pod hustotou pravdepodobnosti v pomere $\alpha : (1-\alpha)$.

## Príklad 5.6

Majme náhodnú premennú $\mathbf{X}$ s hustotou $f_{\mathbf{X}}$ pre ktorú je

$$f_{\mathbf{X}}(x) = \begin{cases} 1 - |x-1| & \text{ak } 0 \leq x \leq 2, \\ 0 & \text{ináč.} \end{cases}$$

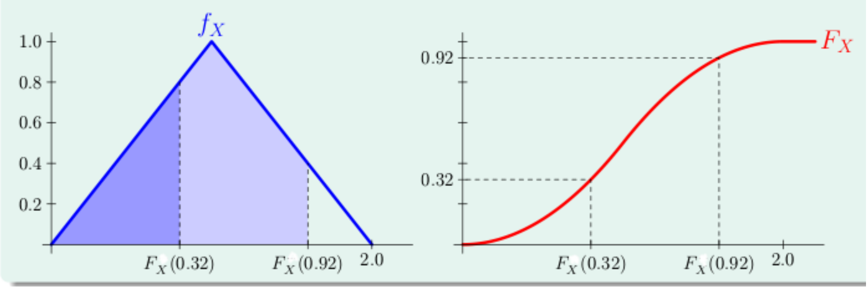
Vypopočítajte kvantily $x_{0.32}$ a $x_{0.92}$

Distribučná funkcia je v tvare

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

Kvantil $x_\alpha = F_{\mathbf{X}}^{-1}(\alpha)$ a tak pre $x_{0.32}$ máme $0.32 = \alpha < \frac{1}{2}$ rovnicu $\alpha = \frac{x_\alpha^2}{2} \Rightarrow x_\alpha = \sqrt{2\alpha}$, teda $x_{0.32} = \sqrt{2 \cdot 0.32} = 0.8$. Podobne pre $0.92 = \alpha \geq \frac{1}{2}$ máme $\alpha = 1 - \frac{(2-x_\alpha)^2}{2} \Rightarrow x_\alpha = 1 - \sqrt{2(1-\alpha)}$ a tak dostaneme $x_{0.92} = 1.6$.

# Príklad 5.7 – grafický význam



## Príklad 5.8

Vypočítajte medián a strednú hodnotu náhodnej premennej $\mathbf{X} \sim Exp(5)$.

$F_{\mathbf{X}}(x_{0.5}) = 1 - e^{-5x_{0.5}} \Rightarrow e^{-5x_{0.5}} = \frac{1}{2}$ a $x_{0.5} = \frac{ln(2)}{2} \approx 0.139.$

$E(\mathbf{X}) = \int_0^{\infty} x \cdot 5e^{-5x}\mathrm{d}x = \cdots = \frac{1}{5} = 0.2$

# Modus

Nech **X** je diskrétna náhodná premenná. Modus $\hat{x}$ je jej najpravdepodobnejšia hodnota, teda pre každé $x \in H$ platí

$$p(\hat{x}) \geq p(x). \tag{15}$$

Nech **X** je spojitá náhodná premenná. Modus $\hat{x}$ je hodnota, kde hustota pravdepodobnosti nadobúda svoje maximum, teda pre každé $x \in \Re$ platí

$$f(\hat{x}) \geq f(x). \tag{16}$$

## Príklad 5.9

Náhodná premenná $X$ má hustotu pravdepodobnosti

$$f_{\mathbf{X}}(x) = \begin{cases} \frac{3}{4}(x-2)(4-x) & \text{ak } 2 < x < 4, \\ 0 & \text{ináč.} \end{cases}$$

Hustota je nenulová funkcia na intervale $(2,4)$. Modus $\hat{x}$ hľadáme ako je stacionárny bod t.j. $f'_{\mathbf{X}}(x) = \frac{3}{4}(6-2x) = 0$ s riešením $x = 3$.

5.1 Čas čakania na autobus je modelovaný náhodnou premennou $\mathbf{X}$:

$$F_{\mathbf{X}}(x) = \begin{cases} 0 & \text{ak } x \leq 0, \\ \frac{x}{3} & \text{ak } 0 < x \leq 2, \\ \frac{x+2}{6}(2x-1) & \text{ak } 2 < x \leq 4, \\ 1 & \text{ináč.} \end{cases}$$

Vypočítajte je charakteristiky a interpretujte ich.

5.2* Vypočítajte postupne počiatočné momenty $\nu_k(\mathbf{X})$, $k = 1, 2, 3, 4$ ak má náhodná premenná $\mathbf{X}$ hustotu pp. $f_{\mathbf{X}}(x) = Ae^{-|x|}$ pre $x \in \Re$.

5.3* Vypočítajte a) parameter $\alpha$, b) $E(\mathbf{X}), D(\mathbf{X})$, c) $x_{0.5}, \hat{x}$ ak náhodná premenná $\mathbf{X}$ má $F_{\mathbf{X}}(x) = \alpha x^3$ na intervale $(0, 2)$.