8 Střední hodnota a rozptyl

Břetislav Fajmon, UMAT FEKT, VUT Brno

Této přednášce odpovídá kapitola 10 ze skript [1]. Také je k dispozici sbírka úloh [2], kde si můžete procvičit příklady z kapitol 2, 3 a 4.

K samostatnému procvičení této přednášky: po přečtení kapitoly 10 ve skriptech [1] můžete absolvovat str. 165-168 (otázky i příklady).

Nejdůležitějšími pojmy této přednášky jsou

- **průměr, modus a medián souboru naměřených hodnot ; rozptyl souboru naměřených hodnot (= empirický rozptyl); směrodatná odchylka souboru naměřených hodnot (= empirická směrodatná odchylka) (JEDNODUCHÉ, nebudeme probírat, prosím přečtěte si ve skriptech [1] strany 154-159)**

- **distribuční funkce náhodné veličiny X – vzorce pro diskrétní i pro spojitou veličinu;**

- **střední hodnota a rozptyl náhodné veličiny X – vzorce pro diskrétní i pro spojitou veličinu.**

Uvedené pojmy budou vysvětleny na příkladech 2 a 5 z minulé přednášky a na dalších čtyřech příkladech v této přednášce.

Vraťme se k příkladu 2 z minulé přednášky (X = počet šestek v pěti hodech kostkou; situace popsána tzv. <u>pravděpodobnostní funkcí</u>, veličina X je *diskrétní veličina*).

A vraťme se také k příkladu 5 z minulé přednášky (X = životnost žárovky do temné komory; situace popsána tzv. <u>hustotou</u>, veličina X je *spojitá veličina*).

Existuje nějaký matematický pojem, který by dokázal spojit jak popis diskrétní veličiny (pravděpodobnosti počítáme pomocí sumy), tak popis spojité veličiny (pravděpodobnosti počítáme pomocí integrálu)?

Existuje, a jmenuje se <u>distribuční funkce</u>.

Definice: Distribuční funkce $F(x_0)$ náhodné veličiny X se definuje jako pravděpodobnost, že náhodná veličina X nabude hodnotu menší než x_0.

$$F(x_0) = P(X < x_0).$$

Způsob výpočtu:

$$F(x_0) = P(X \in (-\infty, x_0)) = \begin{cases} \sum_{k<x_0} p(k) & \textbf{pro diskrétní } X; \\ \int_{-\infty}^{x_0} f(x)\, dx & \textbf{pro spojitou } X. \end{cases} \quad (1)$$

(POZOR, ROZLIŠUJTE f hustotu od F distribuční fce! – case-sensitivity) Je jasné, že pak hodnoty distribuční funkce vypočteme buď diskrétním, nebo spojitým přístupem, ale samotný pojem <u>distribuční funkce</u> popisuje jak veličiny spojité, tak veličiny diskrétní.

Vypočtěme distribuční funkci v příkladech 2 a 5 z minulé přednášky:

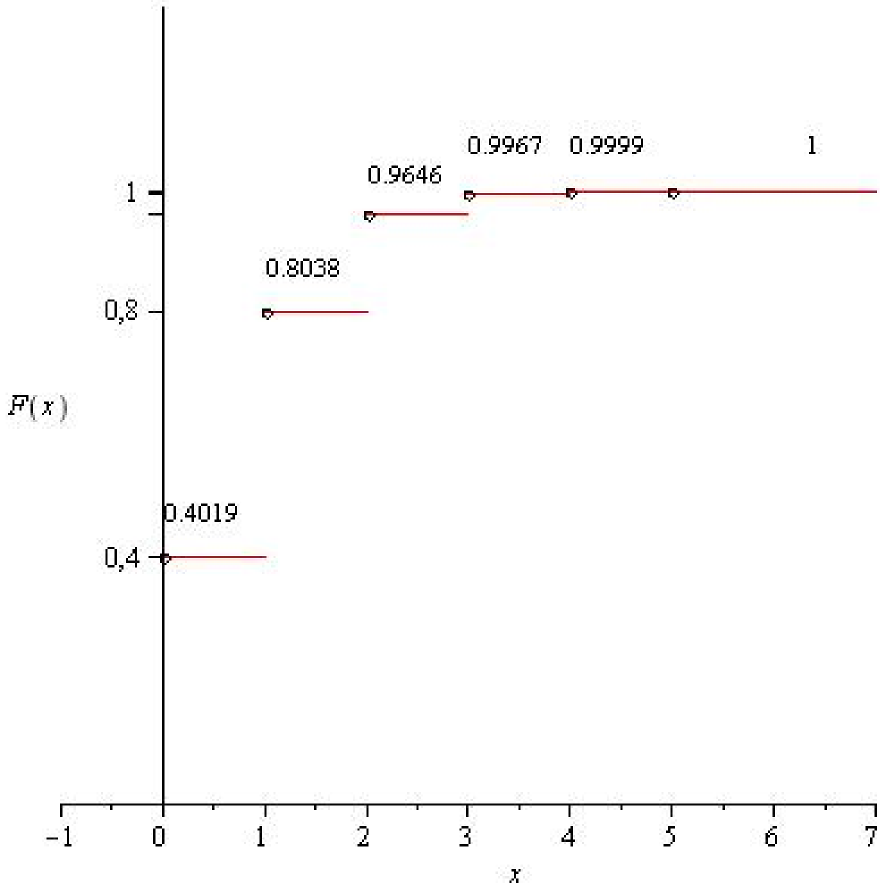
Ad příklad 2: v tabulce jsou dány hodnoty pravděpodobnostní funkce $p(k)$:

k	0	1	2	3	4	5
$p(k)$	0,4019	0,4019	0,1608	0,0321	0,0032	0,0001

.

Sestrojte příslušnou distribuční funkci $F(x)$ této veličiny.

Řešení: užijme vzorec 1 pro každé reálné x – dostaneme funkci $F(x)$ na obrázku:

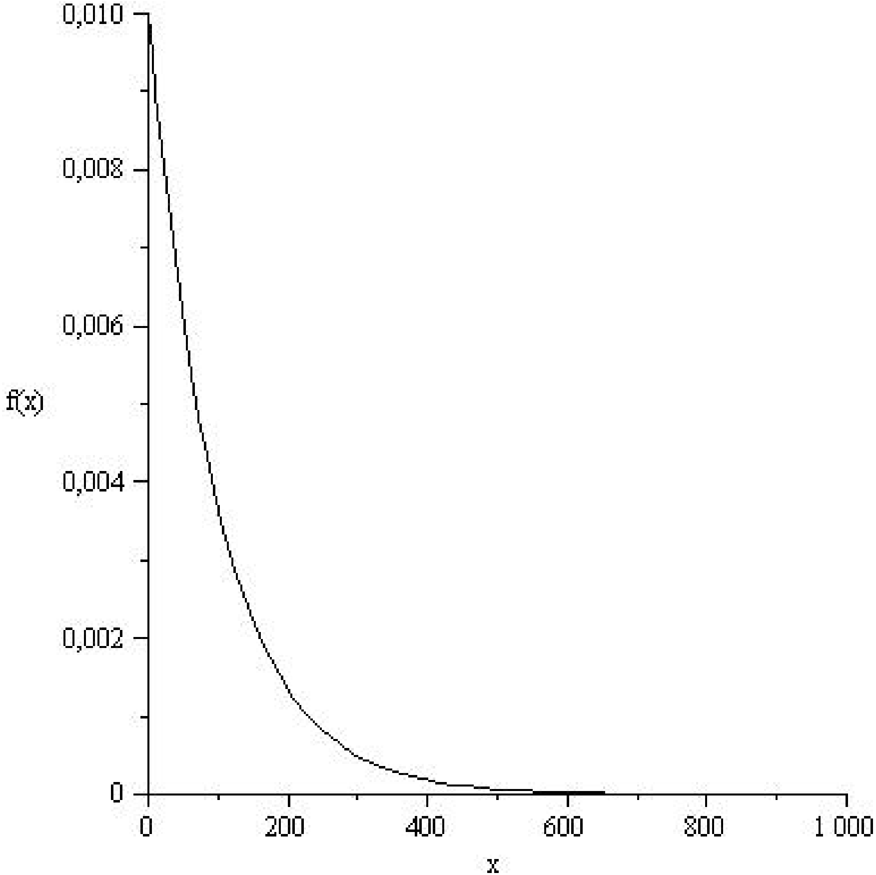

0.9967
0.9999
1
0.9646
1
0.8038
0,8
F(x)
0.4019
0,4
-1
0
1
2
3
4
5
6
7
x

Ad příklad 5: hustota životnosti X žárovky je dána vztahem

$$f(x) = \begin{cases} 0 & \textbf{pro } x < 0; \\ \frac{1}{100} \cdot e^{-\frac{1}{100}x} & \textbf{pro } x \geq 0. \end{cases}$$

Vypočtěte distribuční funkci $F(x)$ této veličiny (opět POZOR – rozlišujte f od F).

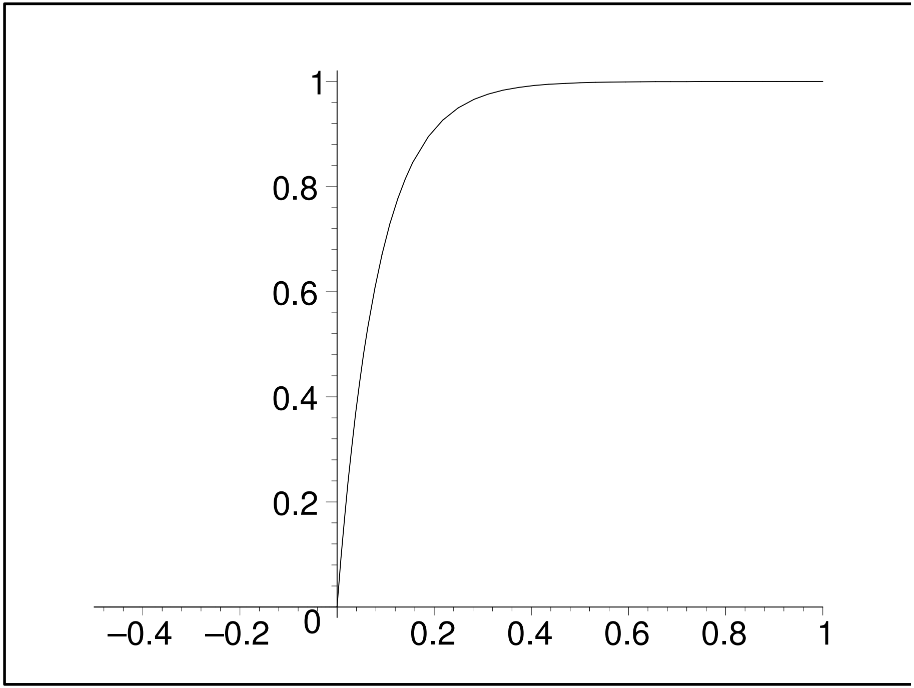
Už v minulé přednášce jsme viděli obrázek hustoty $f(x)$:


0,010
0,008
0,006
0,004
0,002
0
f(x)
0
200
400
600
800
1 000
x

Nyní se podíváme na obrázek distribuční funkce $F(x)$:

Tato byla získána integrací hustoty $f(x)$ podle vztahu 1:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \textbf{pro } x \leq 0; \\ 1 - e^{-\frac{1}{100}x} & \textbf{pro } x > 0. \end{cases}$$



Co mají společného grafy na stranách 7 a 11?

Jedná se o distribuční funkce v případě diskrétním (ad př. 2 z minulé přesnášky) i v případě spojitém (ad příklad 5 z minulé přednášky). Přestože spojitý případ se od diskrétního liší, existuje několik vlastností distribuční funkce, které platí pro všechny situace. Tyto vlastnosti lze odvodit z definice distribuční funkce (z toho, že distribuční funkce je jistá pravděpodobnost, tj. hodnota v intervalu $\langle 0; 1\rangle$).

Například pro $x_1 < x_2$ máme z definice

$$F(x_1) = P(X \in (-\infty, x_1)), \quad F(x_2) = P(X \in (-\infty, x_2)),$$

a protože druhý z intervalů je delší než ten první, musí platit $F(x_1) \leq F(x_2)$. vlastnosti přehledně:

1. $\lim_{x\to-\infty} F(x) = 0; \quad \lim_{x\to\infty} F(x) = 1.$

2. F **je neklesající funkce (jedná se o jakousi kumulativní pravděpodobnost, která pro rostoucí hodnoty proměnné** x **může jen růst, nikoli klesat).**

3. **Funkce** F **je zleva spojitá v každém bodě, tj. limita zleva se rovná funkční hodnotě v každém bodě:**
$$\lim_{x \to x_{0-}} F(x) = F(x_0)$$
pro libovolné reálné x_0**.**

4. **U obou typů veličin (diskrétní i spojité) platí**
$$P(X \in \langle a; b)) = F(b) - F(a).$$

Klíčová je vlastnost číslo 4: pro diskrétní i spojitou veličinu X platí vztah $P(X \in \langle a; b)) = F(b) - F(a)$. (tento vztah pro spojitou veličinu není nic než jiného než Newton-Leibnizova formule pro výpočet určitého integrálu, ale po zavedení pojmu distribuční funkce je jeho platnost rozšířena i pro diskrétní veličiny).

Pro výpočet pravděpodobností platí tedy tento klíčový vztah:

$$P(X \in \langle a; b)) = F(b) - F(a) = \begin{cases} \sum_{k \in \langle a; b)} p(k) & \textbf{pro diskrétní } X; \\ \int_a^b f(x)\, dx & \textbf{pro spojitou } X. \end{cases}$$

Podívejme se nyní na další příklady, ve kterých se naučíme pracovat s distribuční funkcí.

Příklad 8.1. Náhodná veličina X je popsána distribuční funkcí

$$F(x) = \begin{cases} 0{,}5 & \textbf{pro } 0 < x \leq 1; \\ 0{,}7 & \textbf{pro } 1 < x \leq 3; \\ ?? & \textbf{jinak} \ . \end{cases}$$

Určete následující pravděpodobnosti, pokud je to možné: a) $P(X > 1)$; b) $P(X < 1)$; c) $P(X \geq 1)$; d) $P(X \leq 1)$.

Řešení: Distribuční funkci známe jen na části reálné osy, ale jsme schopni pomocí ní určit všechny uvedené pravděpodobnosti:

Nejednodušší je b), protože to je přesně definice funkce $F(x)$ v bodě $x = 1$:

$$P(X < 1) = F(1) = 0{,}5.$$

Ze zadání jsme také schopni určit hodnotu pravděpodobnostní funkce p v bodě $x = 1$, protože ta je rovna „výšce schodu“ distribuční funkce v bodě $x = 1$, A TEDY $p(1) = 0{,}2$. Odtud lze určit b):

$$P(X \geq 1) = P(X < 1) + P(X = 1) = F(1) + p(1) = 0{,}5 + 0{,}2 = 0{,}7.$$

a) a c) dopočítáme z toho faktu, že od hodnoty 1

odečteme pravděpodobnost opačného jevu:

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - 0{,}7 = 0{,}3; \\ P(X \geq 1) &= 1 - P(X < 1) = 1 - 0{,}5 = 0{,}5. \end{aligned}$$

Tyto patálie se mohou objevit pouze u diskrétní veličiny. U spojité veličiny $P(X = 1) = 0$ vždy, a proto vždy platí pro spojitou veličinu $P(X < 1) = P(X \leq 1)$.

Podívejme se tedy také na distribuční funkci spojité veličiny: Vezměme opět příklad 5 z minulé přednášky:

Příklad 8.2. Distribuční funkce životnosti žárovky X je dána vztahem

$$F(x) = \begin{cases} 0 & \textbf{pro } x \leq 0; \\ 1 - e^{-\frac{1}{100}x} & \textbf{pro } x > 0. \end{cases}$$

Určete pravděpodobnosti a) $P(X < 1) = P(X \leq 1)$**;**
b) $P(X \geq 30) = P(X > 30)$**;**
c) $P(X \in \langle 80; 120\rangle)$**.**

Řešení: u spojité veličiny nás nezajímá, zda je v intervalech kulatá či ostrá závorka, protože to při integraci nehraje roli. Také je ideální, že je zadána

distribuční funkce. Ta totiž v sobě uchovává výsledek integrace – a pro výpočet integrálů stačí už jen dosadit meze!!

$$\begin{aligned} P(X < 1) &= F(1) = 1 - e^{\frac{-1}{100}} \doteq 0{,}00995; \\ P(X > 30) &= 1 - P(X \leq 30) = 1 - F(30) = e^{\frac{-30}{100}} \doteq 0{,}74082; \\ P(80 < X < 120) &= F(120) - F(80) = e^{\frac{-80}{100}} - e^{\frac{-120}{100}} = 0{,}14813. \end{aligned}$$

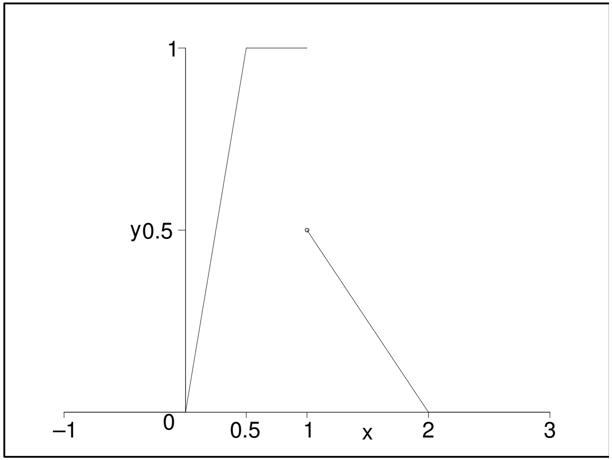
Viz prostřední řádek výpočtu: s využitím opačného jevu se snažíme nerovnosti typu $X > x_0$ převést na výraz využívající pravděpodobnost $X < x_0$, a pak dosadit distribuční funkci v daném x_0.

Důležitým cvičením je v každém z uvedených tří případů nakreslit obrázek plochy, jejíž obsah počítáme (viz minulá přednáška).

Příklad 8.3. Domácí úkol na příští přednášku: s využitím vztahu

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

vypočtěte $F(x)$, jeli zadána hustota $f(x)$ na obrázku:



Důležitými parametry pro popis náhodné veličiny jsou střední hodnota a rozptyl.

Střední hodnota náhodné veličiny X je, zhruba řečeno, průměrná hodnota, kterou naměříme. Přesněji řečeno, jedná se o „teoretický průměr“ = průměr hodnot měření veličiny X, který bychom většinou spočetli, pokud by se měřená veličina chovala přesně podle teoretického popisu zadaného distribuční funkcí $F(x)$. Studenti musí opět pečlivě vážit, zda při výpočtu užít „diskrétní“ či „spojitou“ variantu vzorce:

$$EX = \begin{cases} \sum_{k\in\Omega} k \cdot p(k) & \textbf{pro diskrétní } X; \\ \int_{-\infty}^{+\infty} x \cdot f(x)\, dx & \textbf{pro spojitou } X. \end{cases} \quad (3)$$

Dalším důležitým parametrem je <u>rozptyl</u>: ten udává, jak moc by se odchylovaly měřené hodnoty X od své střední hodnoty EX, pokud by se veličina X chovala přesně podle teoretického popisu zadaného distribuční funkcí $F(x)$. Vzorec pro výpočet:

$$DX = \begin{cases} \left(\sum_{k\in\Omega} k^2 \cdot p(k)\right) - (EX)^2 & \textbf{pro diskrétní } X; \\ \int_{-\infty}^{+\infty} x^2 \cdot f(x)\, dx - (EX)^2 & \textbf{pro spojitou } X. \end{cases} \quad (4)$$

Protože rozměr veličiny DX je čtverec rozměru veličiny X, pro praktické účely potřebujeme znát tzv. směrodatnou odchylku, definovanou jako

$$\sigma(X) = \sqrt{DX}. \tag{5}$$

Směrodatná odchylka je tedy veličina, která nám reprezentuje míru odchylování veličiny X od její střední hodnoty EX – a to ve stejných jednotkách jako jsou jednotky veličiny X.

Pro názornost: zhruba řečeno, u mnoha veličin platí tzv. pravidlo šesti σ:

„asi $97{,}5$ procent měření veličiny X (tedy naprostá většina měření veličiny) leží v intervalu $\langle EX - 3\sigma; EX + 3\sigma\rangle$“.

Tedy rozptýlenost měření veličiny X lze popsat intervalem šířky 6σ a středem v bodě EX. Více o tomto pravidle bude řečeno u normálního rozdělení pravděpodobnosti.

Poznámka: pokud máme k dispozici měření $x_1, x_2, \dots, x_n$ veličiny X, tak odhadem její střední hodnoty EX je průměr těchto měření

$$\overline{x} = \frac{1}{n}\sum_{1}^{n} x_i,$$

odhadem jejího rozptylu DX je empirický rozptyl naměřených hodnot

$$s^2 = \frac{1}{n}\sum_{1}^{n}(x - \overline{x})^2 = \ldots = \left(\frac{1}{n}\sum_{1}^{n} x_i^2\right) - (\overline{x})^2$$

(pomůcka pro zapamatování: průměr čtverců minus čtverec průměru).

Příklad 8.4. **Diskrétní náhodná veličina udává počet gramatických chyb v krátkém textu, platí pro ni**

$$F(x) = \begin{cases} 0 & \textbf{pro } x \leq 0 \\ 0{,}5 & \textbf{pro } x \in (0;1\rangle; \\ 0{,}8 & \textbf{pro } x \in (1;2\rangle; \\ 0{,}9 & \textbf{pro } x \in (2;3\rangle; \\ 0{,}95 & \textbf{pro } x \in (3;4\rangle; \\ 1 & \textbf{pro } x \in (4;\infty). \end{cases}$$

Určete pravděpodobnost toho, že v textu budou a) méně než dvě chyby; b) právě tři chyby; c) vypočtěte EX**,** DX**.**

Řešení: Ad a) Nejjednodušší úkol: pouze dosadíme zadání:

$$P(X < 2) = F(2) = 0{,}8.$$

Ad b,c) Veličina X **může nabývat hodnot** $0, 1, 2, 3, 4$ – **to plyne z charakteru funkce** $F(x)$**. Hodnoty pravděpodobnostní funkce jsou rovny výšce schodů v jednotlivých bodech skoku. Pak** $P(X = 3) = 0{,}05$**; pro střední hodnotu platí (vzorec 3)**

$$\begin{aligned}
EX &= 0 \cdot (0{,}5 - 0) + 1 \cdot (0{,}8 - 0{,}5) + 2 \cdot (0{,}9 - 0{,}8) + 3 \cdot (0{,}95 - 0{,}9) \\
&\quad + 4 \cdot (1 - 0{,}95) = 0{,}85. \\
DX &= 1 \cdot 0{,}3 + 2^2 \cdot 0{,}1 + 3^2 \cdot 0{,}05 \\
&\quad + 4^2 \cdot 0{,}05 - 0{,}85^2 = 1{,}2275. \\
\sigma &= \sqrt{DX} \doteq 1{,}08.
\end{aligned}$$

Příklad 8.5. Spojitá náhodná veličina je popsána distribuční funkcí

$$F(x) = \begin{cases} 0 & \textbf{pro } x \leq 0 \\ \frac{x}{4} & \textbf{pro } x \in (0; 2\rangle; \\ \frac{x^2}{2} - 2x + \frac{5}{2} & \textbf{pro } x \in (2; 3\rangle; \\ 1 & \textbf{pro } x \in (3; \infty). \end{cases}$$

Určete a) $P(X \in (1{,}5; 2{,}5))$**;** $P(X > 2,5)$ **b)** EX**,** DX**;**
c) x_0 **tak, aby veličina** X **nabývala hodnoty menší než** x_0 **s pravděpodobností** $0{,}8$**.**

Řešení: ad a) získáme pomocí vztahu 2:

$$P(X \in (1{,}5; 2{,}5)) = F(2{,}5) - F(1,5) = 0{,}25;$$
$$P(X > 2{,}5) = 1 - F(2{,}5) = 1 - \frac{5}{8} = 0{,}375.$$

Ad c) vlastně tentýž příklad jako a), jen máme zadaný výsledek pravděpodobnosti a máme určit x_0**. Jen musíme dát pozor, který předpis pro** $F(x)$ **platí pro dané** x_0**: z toho, že**

$$P(X < x_0) = 0{,}8 \Rightarrow F(x_0) = 0{,}8$$

plyne, že x_0 **nemůže být z intervalu** $(0; 2\rangle$**, protože rovnice** $\frac{x}{4} = 0{,}8$ **nemá řešení na intervalu** $(0; 2\rangle$**. Tedy**

$x \in (2;3\rangle$ **a lze do** $F(x_0) = 0{,}8$ **dosadit předpis**

$$\frac{x_0^2}{2} - 2x_0 + \frac{5}{2} = 0{,}8;$$

A nyní víme, že při řešení této kvadratické rovnice musíme hledat řešení v intervalu $x \in (2;3\rangle$**, tedy ze dvou řešení kvadratické rovnice nyní odpovídá situaci pouze** $x_0 = 2{,}7746$**.**

Ad b) pozor, case-sensitive vzorec pro EX **vyžaduje, abychom z** $F(x)$ **určili hustotu** $f(x)$**. Díky vzorci 1 víme, že** $F(x)$ **je integrálem z** $f(x)$ **– naopak** $f(x)$

tedy najdeme jako derivaci distribuční funkce:

$$f(x) = F'(x) = \begin{cases} 0 & \textbf{pro } x \leq 0 \\ \frac{1}{4} & \textbf{pro } x \in (0;2\rangle; \\ x-2 & \textbf{pro } x \in (2;3\rangle; \\ 0 & \textbf{pro } x \in (3;\infty). \end{cases}$$

Nyní pomocí vzorce 3 můžeme počítat

$$\begin{aligned} EX &= \int_0^2 x \cdot \frac{1}{4}\, dx + \int_2^3 x \cdot (x-2)\, dx = \\ &= \left[\frac{x^2}{8}\right]_0^2 + \left[\frac{x^3}{3} - x^2\right]_2^3 = 1{,}8333. \end{aligned}$$

a podle 4

$$DX = \int_0^2 x^2 \cdot \frac{1}{2}\,dx + \int_2^3 x^2 \cdot (x-2)\,dx - 1{,}8333^2 =$$
$$= \left[\frac{x^3}{6}\right]_0^2 + \left[\frac{x^4}{4} - \frac{2x^3}{3}\right]_2^3 - 1{,}8333^2 = 0{,}97222.$$

Zakončíme dalšími shrnujícími otázkami k opakování, jejichž odpovědi musí každý student znát:

otázka č. 6**. Jak se definuje distribuční funkce** $F(x)$**?**

Odpověď:

$$F(x) = P(X < x).$$

otázka č. 7**. Jak se čte zápis z otázky číslo** 6**?**

Odpověď: Hodnota funkce F **v bodě** x **se rovná pravděpodobnosti, že veličina** X **nabude hodnoty menší než** x**, tj. hodnoty z intervalu** $(-\infty; x)$**.**

otázka č. 8**. Jaké jsou vlastnosti distribuční funkce** $F(x)$ **u diskrétní i spojité veličiny** X**?**

Odpověď:

1. $\lim_{x\to-\infty} F(x) = 0; \quad \lim_{x\to\infty} F(x) = 1.$
2. F **je neklesající funkce (jedná se o jakousi kumulativní pravděpodobnost, která pro rostoucí hodnoty proměnné x může jen růst, nikoli klesat).**
3. **Funkce F je zleva spojitá v každém bodě, tj. limita zleva se rovná funkční hodnotě v každém bodě:**

$$\lim_{x\to x_{0-}} F(x) = F(x_0)$$

   pro libovolné reálné x_0.
4. **U obou typů veličin (diskrétní i spojité) platí**

$$P(X \in \langle a; b)) = F(b) - F(a).$$

otázka č. 9. Je nějaký rozdíl mezi hodnotami pravděpodobnosti $P(X \in (a;b))$**,** $P(X \in \langle a;b))$**,** $P(X \in (a;b\rangle)$**,** $P(X \in \langle a;b\rangle)$**?**

Odpověď: U spojité veličiny NE, u diskrétní veličiny ANO. přesněji u diskrétní veličiny uvedené čtyři pravděpodobnosti mohou být navzájem různé:

$$
\begin{aligned}
P(X \in \langle a;b)) &= F(b) - F(a);\\
P(X \in \langle a;b\rangle) &= F(b) - F(a) + P(X = b);\\
P(X \in (a;b)) &= F(b) - F(a) - P(X = a);\\
P(X \in (a;b\rangle) &= F(b) - F(a) + P(X = b) - P(X = a).
\end{aligned}
$$

Jednoduše řečeno, pokud by se uvedené vzorce někomu zdály kostrbaté, vždy lze z funkce $F(x)$

vyjádřit pravděpodobnostní funkci $p(x)$ a vzorec

$$P(X \in I) = \sum_{k \in I} p(k)$$

platí pro jakýkoli tvar intervalu I.

otázka č. 10. Jak se vypočte střední hodnota náhodné veličiny X?
Odpověď: Pro diskrétní veličinu (s pravděpodobnostní funkcí $p(k)$)

$$EX = \sum_{k \in \Omega} k \cdot p(k);$$

(s hustotou $f(x)$ – pozor, case-sensitive!)

$$EX = \int_{-\infty}^{\infty} x \cdot f(x)dx.$$

otázka č. 11. **Jak se vypočte rozptyl náhodné veličiny X?**
odpověď: Pro diskrétní veličinu (s pravděpodobnostní funkcí $p(k)$)

$$DX = \left(\sum_{k \in \Omega} k^2 \cdot p(k)\right) - (EX)^2;$$

(s hustotou $f(x)$ – pozor, case-sensitive!)

$$DX = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - (EX)^2.$$

otázka č. 12. **Sice platí $F'(x) = f(x)$, ale $F(x) \neq \int f(x)$. Jak je to možné?**

Odpověď: $\int f(x)$ je označení celé třídy funkcí, které se liší o konstantu. Jen jedna z nich má graf ležící celý v pásu mezi přímkami $y = 0$ a $y = 1$ a je dost malá pravděpodobnost, že je to zrovna ta, pro niž je konstanta rovna nule. Správný vztah je

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Dosazením mezí je výpočet správné konstanty zaručen.

Literatura

[1] Fajmon, B., Růžičková, I.: Matematika 3. Skriptum FEKT 2003 (identifikační číslo v informačním systému VUT: MAT103). http://www.rozhovor.cz/souvislosti/matematika3.pdf.

[2] Hlavičková, I., Hliněná, D.: Sbírka úloh z pravděpodobnosti. Skriptum FEKT 2008.