Základy biostatistiky

(MD710P09)
ak. rok 2008/2009

Karel Zvára

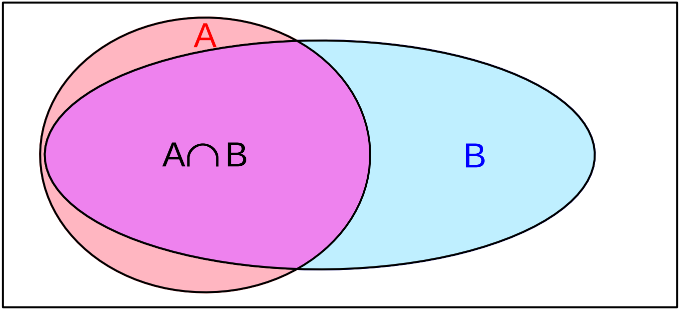
karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara
katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 9. března 2009)

podmíněná pravděpodobnost

když víme, že nastalo A (je to jisté, pst A za podmínky A je rovna 1), pak **podmíněná pst** jevu B za podmínky A bude rovna relativní velikosti $B \cap A$ vzhledem k velikosti A

$$\boxed{\mathrm{P}(B|A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)}}$$



příklad rodina

B – jediná dívka, *D* nejmladší je dívka

ω_i	D	B	$D \cap B$
(m, m, m)			
(f, m, m)	$+$	$+$	$+$
(m, f, m)		$+$	
(f, f, m)	$+$		
(f, f, f)	$+$		
(m, f, f)			
(f, m, f)	$+$		
(m, m, f)		$+$	

- jaká je pst jevu B (v rodině je jediná dívka) **když víme**, že platí D (nejmladší je dívka)?

$$P(B|D) = \frac{m_{B \cap D}}{m_D} = \frac{1/8}{4/8} = \frac{1}{4} = \frac{2}{8}$$

- jaká je pst jevu B (v rodině je jediná dívka) **když víme**, že platí $\overline{D}$ (nejml. **není** dívka)?

$$P(B|\overline{D}) = \frac{m_{B \cap \overline{D}}}{m_{\overline{D}}} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{4}{8}$$

$$P(B|D) = \frac{2}{8} < P(B) = \frac{3}{8} < P(B|\overline{D}) = \frac{4}{8}$$

jev B tedy **závisí** na jevu D!

příklad rodina

B – jediná dívka, *D* nejmladší je dívka

ω_i	D	B	$D \cap B$
(m, m, m)			
(f, m, m)	$+$	$+$	$+$
(m, f, m)		$+$	
(f, f, m)	$+$		
(f, f, f)	$+$		
(m, f, f)			
(f, m, f)	$+$		
(m, m, f)		$+$	

- jaká je pst jevu B (v rodině je jediná dívka) **když víme**, že platí D (nejmladší je dívka)?

$$\mathrm{P}(B|D) = \frac{m_{B\cap D}}{m_D} = \frac{1/8}{4/8} = \frac{1}{4} = \frac{2}{8}$$

- jaká je pst jevu B (v rodině je jediná dívka) **když víme**, že platí $\overline{D}$ (nejml. **není** dívka)?

$$\mathrm{P}(B|\overline{D}) = \frac{m_{B\cap \overline{D}}}{m_{\overline{D}}} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{4}{8}$$

$$\mathrm{P}(B|D) = \frac{2}{8} < \mathrm{P}(B) = \frac{3}{8} < \mathrm{P}(B|\overline{D}) = \frac{4}{8}$$

jev B tedy **závisí** na jevu D!

nezávislost náhodných jevů

ω_i	D	C
(m, m, m)		$+$
(f, m, m)	$+$	$+$
(m, f, m)		$+$
(f, f, m)	$+$	$+$
(f, f, f)	$+$	
(m, f, f)		
(f, m, f)	$+$	
(m, m, f)		

- když víme, že nejstarší je hoch (C), jaká je pak pst, že nejmladší je dívka (D)?
- dva ze čtyř elementárních jevů, tedy

$$\boxed{\frac{m_{D\cap C}}{m_C} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{4}{8} = \frac{m_D}{m}}$$

- zde pst jevu D **nezávisí** na tom, zda platí C
- **nezávislost**: pst jevu D nezávisí na tom, zda C nastal či nenastal

můžeme upravit na **definici nezávislosti náhodných jevů**

$$\frac{m_{D\cap C}}{m} = \frac{m_D}{m}\frac{m_C}{m}$$

$$\boxed{\mathrm{P}(D \cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

nezávislost náhodných jevů

ω_i	D	C
(m, m, m)		$+$
(f, m, m)	$+$	$+$
(m, f, m)		$+$
(f, f, m)	$+$	$+$
(f, f, f)	$+$	
(m, f, f)		
(f, m, f)	$+$	
(m, m, f)		

- když víme, že nejstarší je hoch (C), jaká je pak pst, že nejmladší je dívka (D)?
- dva ze čtyř elementárních jevů, tedy

$$\boxed{\frac{m_{D\cap C}}{m_C} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{4}{8} = \frac{m_D}{m}}$$

- zde pst jevu D **nezávisí** na tom, zda platí C
- **nezávislost**: pst jevu D nezávisí na tom, zda C nastal či nenastal

můžeme upravit na **definici nezávislosti náhodných jevů**

$$\frac{m_{D\cap C}}{m} = \frac{m_D}{m}\frac{m_C}{m}$$

$$\boxed{\mathrm{P}(D \cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

nezávislost náhodných jevů

ω_i	D	C
(m,m,m)		$+$
(f,m,m)	$+$	$+$
(m,f,m)		$+$
(f,f,m)	$+$	$+$
(f,f,f)	$+$	
(m,f,f)		
(f,m,f)	$+$	
(m,m,f)		

- když víme, že nejstarší je hoch (C), jaká je pak pst, že nejmladší je dívka (D)?
- dva ze čtyř elementárních jevů, tedy

$$\boxed{\frac{m_{D\cap C}}{m_C} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{4}{8} = \frac{m_D}{m}}$$

- zde pst jevu D **nezávisí** na tom, zda platí C
- **nezávislost**: pst jevu D nezávisí na tom, zda C nastal či nenastal

můžeme upravit na **definici nezávislosti náhodných jevů**

$$\frac{m_{D\cap C}}{m} = \frac{m_D}{m}\frac{m_C}{m}$$

$$\boxed{\mathrm{P}(D\cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

nezávislost náhodných jevů

ω_i	D	C
(m,m,m)		$+$
(f,m,m)	$+$	$+$
(m,f,m)		$+$
(f,f,m)	$+$	$+$
(f,f,f)	$+$	
(m,f,f)		
(f,m,f)	$+$	
(m,m,f)		

- když víme, že nejstarší je hoch (C), jaká je pak pst, že nejmladší je dívka (D)?
- dva ze čtyř elementárních jevů, tedy

$$\boxed{\frac{m_{D\cap C}}{m_C} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{4}{8} = \frac{m_D}{m}}$$

- zde pst jevu D **nezávisí** na tom, zda platí C
- **nezávislost**: pst jevu D nezávisí na tom, zda C nastal či nenastal

můžeme upravit na **definici nezávislosti náhodných jevů**

$$\frac{m_{D\cap C}}{m} = \frac{m_D}{m}\frac{m_C}{m}$$

$$\boxed{\mathrm{P}(D\cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

vlastnosti podmíněné pravděpodobnost a nezávislosti

- pravděpodobnost jevu D za podmínky jevu C

$$\boxed{\mathrm{P}(D|C) = \frac{m_{D\cap C}}{m_C} = \frac{m_{D\cap C}/m}{m_C/m} = \frac{\mathrm{P}(D\cap C)}{\mathrm{P}(C)}}$$

- pravděpodobnost průniku jevů D, C obecně

$$\mathrm{P}(D\cap C) = \mathrm{P}(D|C)\mathrm{P}(C)$$
$$\mathrm{P}(C\cap D) = \mathrm{P}(C|D)\mathrm{P}(D)$$

(ale $\mathrm{P}(C\cap D) = \mathrm{P}(D\cap C)$, neboť $D\cap C = D\cap C$)

- pro **nezávislé jevy** platí (**násobení pstí**)

$$\boxed{\mathrm{P}(D\cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

vlastnosti podmíněné pravděpodobnost a nezávislosti

- pravděpodobnost jevu D za podmínky jevu C

$$\boxed{\mathrm{P}(D|C) = \frac{m_{D\cap C}}{m_C} = \frac{m_{D\cap C}/m}{m_C/m} = \frac{\mathrm{P}(D\cap C)}{\mathrm{P}(C)}}$$

- pravděpodobnost průniku jevů D, C obecně

$$\mathrm{P}(D\cap C) = \mathrm{P}(D|C)\mathrm{P}(C)$$
$$\mathrm{P}(C\cap D) = \mathrm{P}(C|D)\mathrm{P}(D)$$

(ale $\mathrm{P}(C\cap D) = \mathrm{P}(D\cap C)$, neboť $D\cap C = D\cap C$)

- pro **nezávislé jevy** platí (**násobení pstí**)

$$\boxed{\mathrm{P}(D\cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

vlastnosti podmíněné pravděpodobnost a nezávislosti

- pravděpodobnost jevu D za podmínky jevu C

$$\boxed{\mathrm{P}(D|C) = \frac{m_{D\cap C}}{m_C} = \frac{m_{D\cap C}/m}{m_C/m} = \frac{\mathrm{P}(D\cap C)}{\mathrm{P}(C)}}$$

- pravděpodobnost průniku jevů D, C obecně

$$\mathrm{P}(D\cap C) = \mathrm{P}(D|C)\mathrm{P}(C)$$
$$\mathrm{P}(C\cap D) = \mathrm{P}(C|D)\mathrm{P}(D)$$

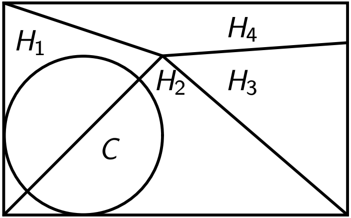
(ale $\mathrm{P}(C\cap D) = \mathrm{P}(D\cap C)$, neboť $D\cap C = D\cap C$)

- pro **nezávislé jevy** platí (**násobení pstí**)

$$\boxed{\mathrm{P}(D\cap C) = \mathrm{P}(D)\mathrm{P}(C)}$$

vzorec pro úplnou pst, Bayesův vzorec

počítáme $\mathrm{P}(H_1|C)$, např. C – správná odpověď, H_j – správná známka j



$$
\begin{aligned}
\mathrm{P}(H_1) &= 0{,}231\\
\mathrm{P}(H_2) &= 0{,}375\\
\mathrm{P}(H_3) &= 0{,}219\\
\mathrm{P}(H_4) &= 0{,}175\\
\mathrm{P}(C|H_1) &= 0{,}589\\
\mathrm{P}(C|H_2) &= 0{,}362 \quad \left(\text{proč je } \mathrm{P}(C|H_2) < \mathrm{P}(C|H_1)?\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{P}(C) &= \mathrm{P}(C \cap H_1) + \mathrm{P}(C \cap H_2)\\
\mathrm{P}(C \cap H_1) &= \mathrm{P}(C|H_1)\mathrm{P}(H_1), \quad \mathrm{P}(C \cap H_2) = \mathrm{P}(C|H_2)\mathrm{P}(H_2)\\
\mathrm{P}(H_1 \cap C) &= \mathrm{P}(H_1|C)\mathrm{P}(C)\\
\mathrm{P}(H_1|C) &= \frac{\mathrm{P}(H_1 \cap C)}{\mathrm{P}(C)} = \frac{\mathrm{P}(C|H_1)\mathrm{P}(H_1)}{\mathrm{P}(C|H_1)\mathrm{P}(H_1) + \mathrm{P}(C|H_2)\mathrm{P}(H_2)} = \frac{1}{2}
\end{aligned}
$$

obecný vzorec pro úplnou pravděpodobnost

(totéž, ale obecně))

- $H_1, \ldots, H_k$ neslučitelné (tj. $H_i \cap H_j = \emptyset$ pro $i \neq j$)
- sjednocení $H_1, \ldots, H_k$ dá jev jistý (tj. $H_1 \cup \ldots \cup H_k = \Omega$)

z definice podmíněné psti plyne $\mathrm{P}(C \cap H_j) = \mathrm{P}(C|H_j) \cdot \mathrm{P}(H_j)$

$$\begin{aligned}
\mathrm{P}(C) &= \mathrm{P}(C \cap \Omega) = \mathrm{P}(C \cap (H_1 \cup H_2 \cup \ldots \cup H_k)) \\
&= \mathrm{P}((C \cap H_1) \cup (C \cap H_2) \cup \ldots \cup (C \cap H_k)) \text{ (neslučitelné jevy)} \\
&= \mathrm{P}(C \cap H_1) + \mathrm{P}(C \cap H_2) + \ldots + \mathrm{P}(C \cap H_k) \\
&= \mathrm{P}(C|H_1)\mathrm{P}(H_1) + \mathrm{P}(C|H_2)\mathrm{P}(H_2) + \ldots + \mathrm{P}(C|H_k)\mathrm{P}(H_k)
\end{aligned}$$

tedy obecně $$\boxed{\mathrm{P}(C) = \sum_{j=1}^{k} \mathrm{P}(C|H_j)\mathrm{P}(H_j)}$$

$\mathrm{P}(C)$ je váženým průměrem podmíněných pstí $\mathrm{P}(C|H_j)$

Bayesův vzorec [Bayes formula]

stejné předpoklady: H_j neslučitelné, sjednocení všech jistý jev

$$\mathrm{P}(H_i|C) = \frac{\mathrm{P}(H_i \cap C)}{\mathrm{P}(C)}, \qquad \mathrm{P}(C|H_i) = \frac{\mathrm{P}(C \cap H_i)}{\mathrm{P}(H_i)}$$

odtud je pro libovolně zvolené i

$$\mathrm{P}(H_i \cap C) = \mathrm{P}(C \cap H_i) = \mathrm{P}(C|H_i)\mathrm{P}(H_i)$$

proto pro každé i, $i = 1, \ldots, k$ platí

$$\boxed{\mathrm{P}(H_i|C) = \frac{\mathrm{P}(H_i \cap C)}{\mathrm{P}(C)} = \frac{\mathrm{P}(C|H_i)\mathrm{P}(H_i)}{\mathrm{P}(C)} = \frac{\mathrm{P}(C|H_i)\mathrm{P}(H_i)}{\sum_{j=1}^{k} \mathrm{P}(C|H_j)\mathrm{P}(H_j)}}$$

$H_1, \ldots, H_k$ – **hypotézy**, $\mathrm{P}(H_1|C), \ldots, \mathrm{P}(H_k|C)$ – **aposteriorní** psti
$\mathrm{P}(H_1), \ldots, \mathrm{P}(H_k)$ – **apriorní** psti (nutně $\mathrm{P}(H_1) + \ldots + \mathrm{P}(H_k) = 1$)

příklad: zkoušení

H_j – student si zaslouží známku j, učitel studenta (tedy j) nezná
C – student správně odpoví na položenou otázku
$\mathrm{P}(H_j)$ – apriorní představa učitele o neznámém studentovi
$\mathrm{P}(C|H_j)$ – obtížnost otázky, volí učitel

H_j	$\mathrm{P}(H_j)$	$\mathrm{P}(C\|H_j)$	$\mathrm{P}(C\|H_j)\mathrm{P}(H_j)$	$\mathrm{P}(H_j\|C)$	$\mathrm{P}(H_j\|C_2)$	$\mathrm{P}(H_j\|C_3)$
1	0,20	1,00	0,2000	0,2694	0,3451	0,4230
2	0,35	0,80	0,2800	0,3771	0,3865	0,3790
3	0,25	0,65	0,1625	0,2189	0,1822	0,1452
4	0,20	0,50	0,1000	0,1347	0,0863	0,0529
Σ	1,00		0,7425	1,0000	1,0000	1,0000

$\mathrm{P}(C) = 0{,}7425$
podobně C_2, C_3 správné odpovědi na další stejně obtížné otázky, když použijeme předchozí aposteriorní psti jako apriorní

senzitivita, specificita, prevalence

- D – subjekt je nemocen, **prevalence** – podíl nemocných v populaci $\mathrm{P}(D)$, zvolme $\mathrm{P}(D) = 0{,}001$
- nemoc je skrytá, vyhledáváme ji pomocí testu s vlastnostmi:
  - $\mathrm{P}(P|D)$ – pravděpodobnost pozitivního výsledku u nemocného (**senzitivita**, pokud možno velká, zvolme $\mathrm{P}(P|D) = 0{,}98$ pozitivně reaguje 98 % nemocných)
  - $\mathrm{P}(\overline{P}|\overline{D})$ – pravděpodobnost negativního výsledku u zdravého (**specificita**, pokud možno velká, zvolme $\mathrm{P}(\overline{P}|\overline{D}) = 0{,}99$ pozitivně reaguje jen 1 % zdravých)

senzitivita, specificita, prevalence

- D – subjekt je nemocen, **prevalence** – podíl nemocných v populaci $\mathrm{P}(D)$, zvolme $\mathrm{P}(D) = 0{,}001$
- nemoc je skrytá, vyhledáváme ji pomocí testu s vlastnostmi:
  - $\mathrm{P}(P|D)$ – pravděpodobnost pozitivního výsledku u nemocného (**senzitivita**, pokud možno velká, zvolme $\mathrm{P}(P|D) = 0{,}98$ pozitivně reaguje 98 % nemocných)
  - $\mathrm{P}(\overline{P}|\overline{D})$ – pravděpodobnost negativního výsledku u zdravého (**specificita**, pokud možno velká, zvolme $\mathrm{P}(\overline{P}|\overline{D}) = 0{,}99$ pozitivně reaguje jen 1 % zdravých)

senzitivita, specificita, prevalence

- D – subjekt je nemocen, **prevalence** – podíl nemocných v populaci $\mathrm{P}(D)$, zvolme $\mathrm{P}(D) = 0{,}001$
- nemoc je skrytá, vyhledáváme ji pomocí testu s vlastnostmi:
  - $\mathrm{P}(P|D)$ – pravděpodobnost pozitivního výsledku u nemocného (**senzitivita**, pokud možno velká, zvolme $\mathrm{P}(P|D) = 0{,}98$ pozitivně reaguje 98 % nemocných)
  - $\mathrm{P}(\overline{P}|\overline{D})$ – pravděpodobnost negativního výsledku u zdravého (**specificita**, pokud možno velká, zvolme $\mathrm{P}(\overline{P}|\overline{D}) = 0{,}99$ pozitivně reaguje jen 1 % zdravých)

senzitivita, specificita, prevalence

- D – subjekt je nemocen, **prevalence** – podíl nemocných v populaci $\mathrm{P}(D)$, zvolme $\mathrm{P}(D) = 0{,}001$
- nemoc je skrytá, vyhledáváme ji pomocí testu s vlastnostmi:
  - $\mathrm{P}(P|D)$ – pravděpodobnost pozitivního výsledku u nemocného (**senzitivita**, pokud možno velká, zvolme $\mathrm{P}(P|D) = 0{,}98$ pozitivně reaguje 98 % nemocných)
  - $\mathrm{P}(\overline{P}|\overline{D})$ – pravděpodobnost negativního výsledku u zdravého (**specificita**, pokud možno velká, zvolme $\mathrm{P}(\overline{P}|\overline{D}) = 0{,}99$ pozitivně reaguje jen 1 % zdravých)

senzitivita, specificita, prevalence

- jaká je pst, že pozitivně reagující je opravdu nemocný?

$$\mathrm{P}(D|P) = \frac{\mathrm{P}(P|D)\mathrm{P}(D)}{\mathrm{P}(P|D)\mathrm{P}(D) + \mathrm{P}(P|\overline{D})\mathrm{P}(\overline{D})} = \frac{0{,}98 \cdot 0{,}001}{0{,}98 \cdot 0{,}001 + 0{,}01 \cdot 0{,}999} \doteq 0{,}089$$

- jaká je pst, že jde o zdravého člověka v případě, že test byl negativní?

$$\mathrm{P}(\overline{D}|\overline{P}) = \frac{\mathrm{P}(\overline{P}|\overline{D})\mathrm{P}(\overline{D})}{\mathrm{P}(\overline{P}|\overline{D})\mathrm{P}(\overline{D}) + \mathrm{P}(\overline{P}|D)\mathrm{P}(D)} = \frac{0{,}99 \cdot 0{,}999}{0{,}99 \cdot 0{,}999 + 0{,}02 \cdot 0{,}001} = 0{,}99998$$

- porovnej s apriorními pstmi: 0,001 resp. 0,999

senzitivita, specificita, prevalence

- jaká je pst, že pozitivně reagující je opravdu nemocný?

$$\mathrm{P}(D|P) = \frac{\mathrm{P}(P|D)\mathrm{P}(D)}{\mathrm{P}(P|D)\mathrm{P}(D) + \mathrm{P}(P|\overline{D})\mathrm{P}(\overline{D})} = \frac{0{,}98 \cdot 0{,}001}{0{,}98 \cdot 0{,}001 + 0{,}01 \cdot 0{,}999} \doteq 0{,}089$$

- jaká je pst, že jde o zdravého člověka v případě, že test byl negativní?

$$\mathrm{P}(\overline{D}|\overline{P}) = \frac{\mathrm{P}(\overline{P}|\overline{D})\mathrm{P}(\overline{D})}{\mathrm{P}(\overline{P}|\overline{D})\mathrm{P}(\overline{D}) + \mathrm{P}(\overline{P}|D)\mathrm{P}(D)} = \frac{0{,}99 \cdot 0{,}999}{0{,}99 \cdot 0{,}999 + 0{,}02 \cdot 0{,}001} = 0{,}99998$$

- porovnej s apriorními pstmi: 0,001 resp. 0,999

senzitivita, specificita, prevalence

- jaká je pst, že pozitivně reagující je opravdu nemocný?

$$\begin{aligned}\mathrm{P}(D|P) &= \frac{\mathrm{P}(P|D)\mathrm{P}(D)}{\mathrm{P}(P|D)\mathrm{P}(D) + \mathrm{P}(P|\overline{D})\mathrm{P}(\overline{D})} \\ &= \frac{0{,}98 \cdot 0{,}001}{0{,}98 \cdot 0{,}001 + 0{,}01 \cdot 0{,}999} \doteq 0{,}089\end{aligned}$$

- jaká je pst, že jde o zdravého člověka v případě, že test byl negativní?

$$\begin{aligned}\mathrm{P}(\overline{D}|\overline{P}) &= \frac{\mathrm{P}(\overline{P}|\overline{D})\mathrm{P}(\overline{D})}{\mathrm{P}(\overline{P}|\overline{D})\mathrm{P}(\overline{D}) + \mathrm{P}(\overline{P}|D)\mathrm{P}(D)} \\ &= \frac{0{,}99 \cdot 0{,}999}{0{,}99 \cdot 0{,}999 + 0{,}02 \cdot 0{,}001} = 0{,}99998\end{aligned}$$

- porovnej s apriorními pstmi: 0,001 resp. 0,999

náhodná veličina

[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $P(X = x_1^*), P(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $P(X = x_1^*), P(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace ...)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \dots$
  - psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace …)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina

[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \dots$
  - psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace ...)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $P(X = x_1^*), P(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace ...)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina

[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina

[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \dots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \dots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina

[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \dots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \dots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina

[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \dots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \dots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace ...)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \ldots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \ldots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace . . .)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

náhodná veličina
[random variable]

- číselně vyjádřený výsledek náhodného pokusu
- předem nevíme, který výsledek vyjde, známe jen
  - možné hodnoty
  - jejich pravděpodobnosti
- každému elementárnímu jevu přiřadíme reálné číslo
- **diskrétní rozdělení** náhodné veličiny X
  - model pro počty případů (četnosti)
  - možné hodnoty $x_1^*, x_2^*, \dots$
  - psti hodnot $\mathrm{P}(X = x_1^*), \mathrm{P}(X = x_2^*), \dots$ (pstní funkce)
- **spojité rozdělení** náhodné veličiny X
  - model pro spojitou veličiny (délka, váha, koncentrace $\dots$)
  - obor (množina) možných hodnot X
  - hustota $f(x)$

příklad: rodina

náhodná veličina X – počet děvčat

rozdělení X dáno hodnotami x_j^* a pstmi těchto hodnot $\mathrm{P}(X = x_j^*)$

ω_i	x_i	x_j^*
(m, m, m)	0	0
(m, m, f)	1	
(m, f, m)	1	1
(f, m, m)	1	
(f, f, m)	2	
(f, m, f)	2	2
(m, f, f)	2	
(f, f, f)	3	3

j	x_j^*	m_j	$\mathrm{P}(X = x_j^*)$
1	**0**	1	**1/8**
2	**1**	3	**3/8**
3	**2**	3	**3/8**
4	**3**	1	**1/8**
součet		8	8/8

$$m = \sum_{j=1}^{4} m_j = 8$$

distribuční funkce

protějšek empirické distribuční funkce (str. 16), [(cumulative) distribution function]

- pst, že X nepřekročí x $\boxed{F_X(x) = \mathrm{P}(X \le x)}$
- diskrétní rozdělení: $\boxed{F(x) = \sum_{t \le x} \mathrm{P}(X = t)}$
- spojité rozdělení: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$, kde $f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$
- vlastnosti distribuční funkce

$$0 \le F(x) \le 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$

$$\boxed{\mathrm{P}(x_1 < X \le x_2) = F(x_2) - F(x_1)}$$

$$\mathrm{P}(X \le x_2) = \mathrm{P}(X \le x_1) + \mathrm{P}(x_1 < X \le x_2)$$
$$F(x_2) = F(x_1) + \mathrm{P}(x_1 < X \le x_2)$$

distribuční funkce

protějšek empirické distribuční funkce (str. 16), [(cumulative) distribution function]

- pst, že X nepřekročí x $\boxed{F_X(x) = \mathrm{P}(X \le x)}$
- diskrétní rozdělení: $\boxed{F(x) = \sum_{t \le x} \mathrm{P}(X = t)}$
- spojité rozdělení: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$, kde $f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$
- vlastnosti distribuční funkce

$$0 \le F(x) \le 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$

$$\boxed{\mathrm{P}(x_1 < X \le x_2) = F(x_2) - F(x_1)}$$

$$\mathrm{P}(X \le x_2) = \mathrm{P}(X \le x_1) + \mathrm{P}(x_1 < X \le x_2)$$
$$F(x_2) = F(x_1) + \mathrm{P}(x_1 < X \le x_2)$$

distribuční funkce

protějšek empirické distribuční funkce (str. 16), [(cumulative) distribution function]

- pst, že X nepřekročí x $\boxed{F_X(x) = \mathrm{P}(X \leq x)}$
- diskrétní rozdělení: $\boxed{F(x) = \sum_{t \leq x} \mathrm{P}(X = t)}$
- spojité rozdělení: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$, kde $f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$
- vlastnosti distribuční funkce

$$0 \leq F(x) \leq 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$

$$\boxed{\mathrm{P}(x_1 < X \leq x_2) = F(x_2) - F(x_1)}$$

$$\mathrm{P}(X \leq x_2) = \mathrm{P}(X \leq x_1) + \mathrm{P}(x_1 < X \leq x_2)$$
$$F(x_2) = F(x_1) + \mathrm{P}(x_1 < X \leq x_2)$$

distribuční funkce

protějšek empirické distribuční funkce (str. 16), [(cumulative) distribution function]

- pst, že X nepřekročí x $\boxed{F_X(x) = \mathrm{P}(X \leq x)}$
- diskrétní rozdělení: $\boxed{F(x) = \sum_{t \leq x} \mathrm{P}(X = t)}$
- spojité rozdělení: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$, kde $f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$
- vlastnosti distribuční funkce

$$0 \leq F(x) \leq 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$

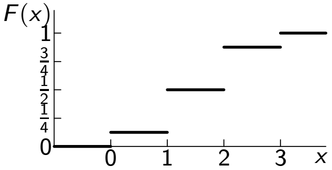
$$\boxed{\mathrm{P}(x_1 < X \leq x_2) = F(x_2) - F(x_1)}$$

$$\mathrm{P}(X \leq x_2) = \mathrm{P}(X \leq x_1) + \mathrm{P}(x_1 < X \leq x_2)$$
$$F(x_2) = F(x_1) + \mathrm{P}(x_1 < X \leq x_2)$$

příklad diskrétního rozdělení

rozdělení počtu děvčat X

j	x_j^*	$\mathrm{P}(X = x_j^*)$	$F_X(x_j^*)$
1	0	$1/8$	$1/8$
2	1	$3/8$	$4/8$
3	2	$3/8$	$7/8$
4	3	$1/8$	$8/8$
součet		$8/8$	

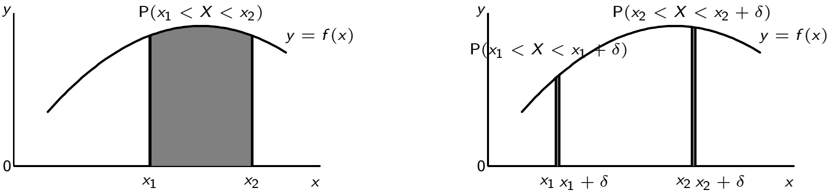


hustota spojitého rozdělení
[density function]

- plocha pod celou hustotou je rovna jedné

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

- plocha pod hustotou nad intervalem x_1, x_2 je rovna pravděpodobnosti, že X je mezi x_1, x_2



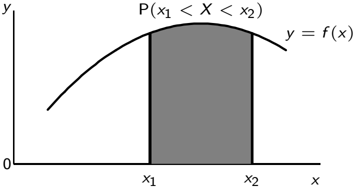
hustota spojitého rozdělení

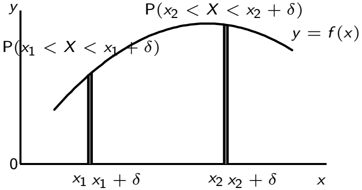
[density function]

- plocha pod celou hustotou je rovna jedné

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

- plocha pod hustotou nad intervalem x_1, x_2 je rovna pravděpodobnosti, že X je mezi x_1, x_2



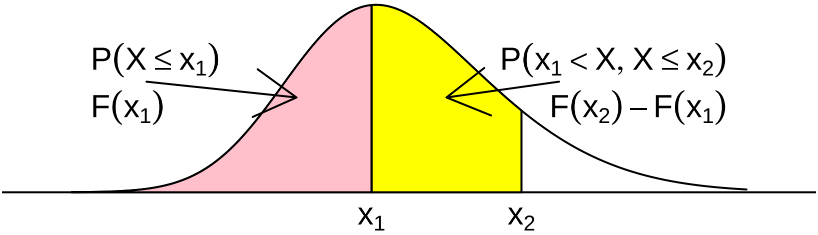


geometrický význam hustoty

$P(x_1 < X, X <= x_2) = P(x_1 < X <= x_2)$, vpravo stručnější, používaný zápis

$$F(x_2) = P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$
$$= F(x_1) + P(x_1 < X \leq x_2)$$

odtud $\boxed{P(x_1 < X \leq x_2) = F(x_2) - F(x_1)}$

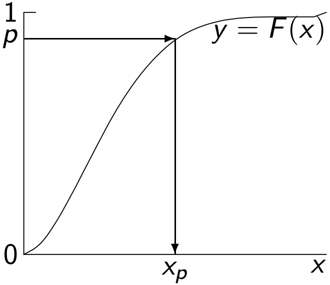


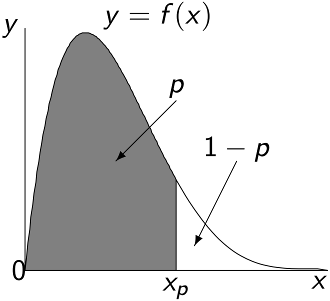
p-kvantil x_p

x_p je hodnota, pod kterou je $100p$ procent pravděpodobnosti

$$\boxed{\mathrm{P}(X \leq x_p) = p}$$

např. [qnorm(0.975)] dá $1{,}959964 \doteq 1{,}96$



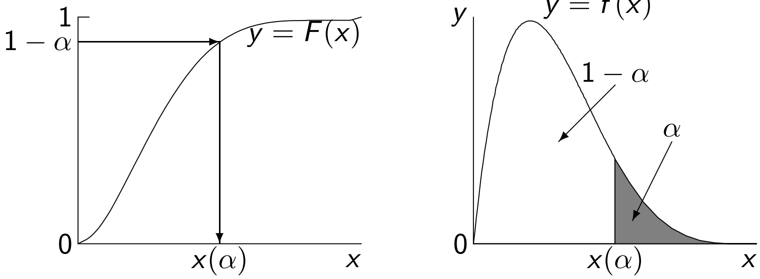


kritická hodnota $x(\alpha)$

kritická hodnota $x(\alpha)$ je překročena s pstí α

$$\boxed{\mathrm{P}(X \geq x(\alpha)) = \alpha}$$

např. [qnorm(1-0.025)] dá $1{,}959964 \doteq 1{,}96$



střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr**
  [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x) \mathrm{d}x$$

střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr** [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x) \mathrm{d}x$$

střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr** [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x) \mathrm{d}x$$

střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr** [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x) \mathrm{d}x$$

střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr** [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x) \mathrm{d}x$$

střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr** [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\, f_X(x)\mathrm{d}x$$

střední hodnota

pokračujeme v idealizovaných představách

- míra polohy, **populační průměr**
  [expected value, mean value]
- metoda výpočtu se značí $\mathrm{E}\,X$
- vypočtená hodnota se značí μ nebo úplněji μ_X
- **vážený průměr možných hodnot**
- ideální protějšek výběrového průměru
- diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\boxed{\mu_X = \mathrm{E}\,X = \sum_j x_j^* \mathrm{P}(X = x_j^*)}$$

- spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = \mathrm{E}\,X = \int_{-\infty}^{\infty} x\ f_X(x)\mathrm{d}x$$

příklad rodina

X – počet děvčat mezi třemi dětmi

j	x_j^*	$\mathrm{P}(X = x_j^*)$	$x_j^* \cdot \mathrm{P}(X = x_j^*)$
1	0	0,125	0,000
2	1	0,375	0,375
3	2	0,375	0,750
4	3	0,125	0,375
součet		1,000	1,500

$$\begin{aligned}\mu_X &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} \\ &= 0 \cdot 0{,}125 + 1 \cdot 0{,}375 + 2 \cdot 0{,}375 + 3 \cdot 0{,}125 \\ &= 1{,}5\end{aligned}$$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí $\operatorname{var} X$
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\quad \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí $\operatorname{var} X$
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j (x_j^* - \mu_X)^2\, \mathrm{P}\,(X = x_j^*)}$$

- spojité rozdělení $\quad \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ
[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí var X
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\quad \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí $\operatorname{var} X$
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\qquad \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí var X
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\quad \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí var X
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\qquad \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí $\operatorname{var} X$
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X-\mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\mathrm{d}x$

rozptyl σ^2, směrodatná odchylka σ

[variance, standard deviation]

- míra variability, **populační rozptyl, popul. směr. odchylka**
- udává velikost kolísání (variabilitu) kolem střední hodnoty
- metoda výpočtu se značí $\operatorname{var} X$
- vypočtená hodnota σ^2, úplněji σ_X^2
- lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \operatorname{var} X = \mathrm{E}\,(X - \mu_X)^2 = \mathrm{E}\,(X^2) - (\mu_X)^2$$

- ideální protějšky výběrového rozptylu, směr. odchylky
- diskrétní rozdělení

$$\boxed{\sigma_X^2 = \operatorname{var} X = \sum_j \left(x_j^* - \mu_X\right)^2 \mathrm{P}\left(X = x_j^*\right)}$$

- spojité rozdělení $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$

příklad rodina

X – počet děvčat mezi třemi dětmi, $\mu_X = 1{,}5$

j	x_j^*	p_j	$x_j^* - \mu_X$	$\left(x_j^* - \mu_X\right)^2$	$\left(x_j^* - \mu_X\right)^2 p_j$
1	0	0,125	-1,5	2,25	0,28125
2	1	0,375	-0,5	0,25	0,09375
3	2	0,375	0,5	0,25	0,09375
4	3	0,125	1,5	2,25	0,28125
$\sum$		1,000			0,75000

$$\begin{aligned}
\sigma_X^2 &= \sum_j (x_j^* - \mu_X)^2 p_j \\
&= (0 - 1{,}5)^2 \cdot 0{,}125 + (1 - 1{,}5)^2 \cdot 0{,}375 \\
&\quad + (2 - 1{,}5)^2 \cdot 0{,}375 + (3 - 1{,}5)^2 \cdot 0{,}125 = 0{,}75 \\
\sigma_X &= \sqrt{0{,}75} \quad = \quad 0{,}866025
\end{aligned}$$

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,. . .) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,. . .) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
- (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené, marginální a podmíněné rozdělení

sdružené rozdělení – popisuje **společné chování** X, Y

$$\boxed{\mathrm{P}(X = x_i^*, Y = y_j^*)} \text{ resp. } \boxed{f_{X,Y}(x, y)}$$

marginální rozdělení: chování jedné bez ohledu na hodnotu druhé

$$\boxed{\mathrm{P}(X = x_i^*) = \sum_j \mathrm{P}(X = x_i^*, Y = y_j^*) \qquad \forall x_i^*}$$

$$\boxed{\mathrm{P}(Y = y_j^*) = \sum_i \mathrm{P}(X = x_i^*, Y = y_j^*) \qquad \forall y_j^*}$$

podmíněné rozdělení: chování Y při **dané** hodnotě X

$$\mathrm{P}(Y = y_j^* | X = x_i^*) = \frac{\mathrm{P}(X = x_i^*, Y = y_j^*)}{\mathrm{P}(X = x_i^*)}$$

příklad rodina

X počet děvčat, Y počet děvčat mezi dvěma staršími dětmi

sdružené, marginální a podmíněné rozdělení

ω_i	x_i	y_i
(m, m, m)	0	0
(m, m, f)	1	1
(m, f, m)	1	1
(f, m, m)	1	0
(f, f, m)	2	1
(f, m, f)	2	1
(m, f, f)	2	2
(f, f, f)	3	2

x_i^*	y_j^* 0	1	2	celkem
0	1/8	0	0	1/8
1	1/8	2/8	0	3/8
2	0	2/8	1/8	3/8
3	0	0	1/8	1/8
	2/8	4/8	2/8	1

x_i^*	y_j^* 0	1	2	celkem
0	1	0	0	1
1	1/3	2/3	0	1
2	0	2/3	1/3	1
3	0	0	1	1

kovariance

protějšek s_{xy}, [covariance]

kovariance vyjadřuje vzájemnou závislost náhodných veličin:

$$\boxed{\sigma_{X,Y} = \mathrm{E}\,(X - \mu_X)(Y - \mu_Y)}$$

$$\boxed{\sigma_{X,Y} = \sum_i \sum_j (x_i^* - \mu_X)(y_j^* - \mu_Y)\mathrm{P}(X = x_i^*, Y = y_j^*)}$$

označení metody výpočtu: $\mathrm{cov}(X, Y)$

zřejmě platí $\boxed{\mathrm{cov}(X, X) = \mathrm{var}\, X}$ tj. $\boxed{\sigma_{X,X} = \sigma_X^2}$

pro **nezávislé** náhodné veličiny platí

(ze znalosti hodnoty jedné nic nevíme o druhé)

$$\boxed{\mathrm{P}(X = x_i^*, Y = y_j^*) = \mathrm{P}(X = x_i^*) \cdot \mathrm{P}(Y = y_j^*), \qquad \forall (x_i^*, y_j^*)}$$

jsou-li X, Y – nezávislé $\Rightarrow$ $\boxed{\sigma_{X,Y} = 0}$ (nikoliv obrácená implikace)

příklad rodina

x_i^*	y_j^* 0	1	2	celkem
0	0,125	0	0	0,125
1	0,125	0,250	0	0,375
2	0	0,250	0,125	0,375
3	0	0	0,125	0,125
celkem	0,250	0,500	0,250	1,000

$$\mu_X = 0 \cdot 0{,}125 + 1 \cdot 0{,}375 + 2 \cdot 0{,}375 + 3 \cdot 0{,}125 = 1{,}5$$
$$\mu_Y = 0 \cdot 0{,}250 + 1 \cdot 0{,}500 + 2 \cdot 0{,}250 = 1$$
$$\sigma_X^2 = (0 - 1{,}5)^2 \cdot 0{,}125 + \ldots + (3 - 1{,}5)^2 \cdot 0{,}125 = 0{,}75$$
$$\sigma_Y^2 = (0 - 1)^2 \cdot 0{,}25 + (1 - 1)^2 \cdot 0{,}5 + (2 - 1)^2 \cdot 0{,}25 = 0{,}5$$
$$\sigma_{XY} = (0 - 1{,}5) \cdot (0 - 1) \cdot 0{,}125 + \ldots = 0{,}5$$

X, Y jsou závislé, neboť např. $0{,}25 \cdot 0{,}125 \neq 0{,}125$