STATISTIKA I

Martin Řezáč, Marie Budíková
2010

Obsah

1. Popisná statistika

Popisná statistika je disciplína, která popisuje a sumarizuje informace obsažené ve velkém množství dat pomocí tabulek, grafů, funkcionálních a číselných charakteristik. Činí tak pomocí základních matematických operací. Cílem popisné statistiky je zpřehlednit informace „ukryté´" v datových souborech.

Popisná statistika je velmi důležitá minimálně ze dvou důvodů:

- v praxi se často používá (všichni znají takové pojmy, jako je průměr, směrodatná odchylka, tabulka rozložení četností, výsečový graf apod.)
- motivuje pojmy, se kterými pak pracuje počet pravděpodobnosti (např. relativní četnost motivuje pravděpodobnost, hustota četnosti motivuje hustotu pravděpodobnosti, průměr motivuje střední hodnotu apod.)

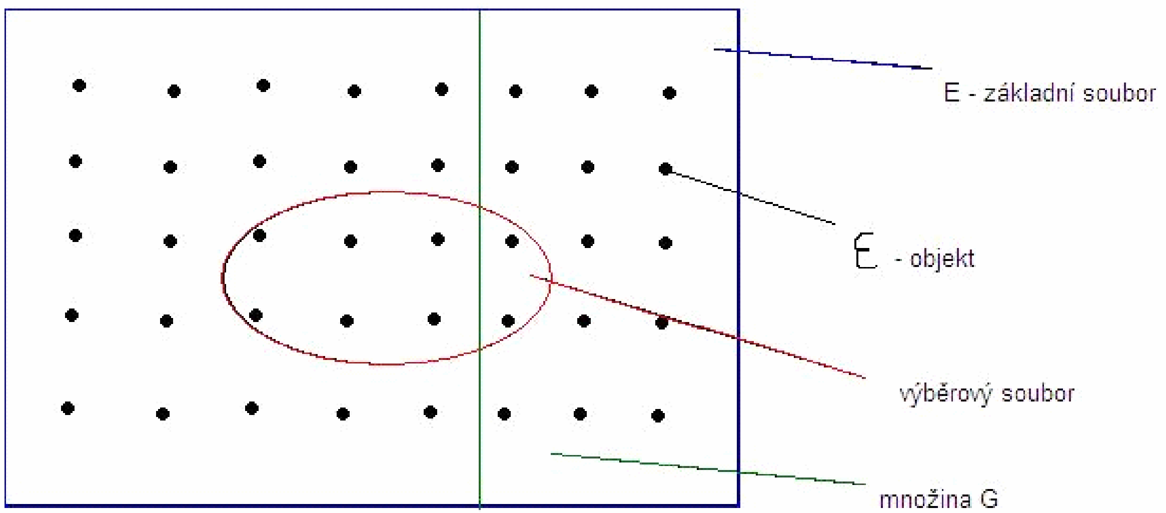
Dobré pochopení pojmů popisné statistiky tedy velmi usnadní studium počtu pravděpodobnosti.

Základní, výběrový a datový soubor

Základním souborem rozumíme libovolnou neprázdnou množinu E. Prvky množiny E značíme ε a nazýváme je objekty. Libovolnou neprázdnou podmnožinu $\{\varepsilon_1, \ldots, \varepsilon_n\}$ základního souboru E nazýváme výběrový soubor rozsahu n. Je-li množina G ⊆ E, pak symbolem N(G) rozumíme absolutní četnost množiny G ve výběrovém souboru, tj. počet těch objektů množiny G, které patří do výběrového souboru. Relativní četnost množiny G ve výběrovém souboru zavedeme vztahem

$$p(G) = \frac{N(G)}{n}.$$

Ilustrace



Příklad

Příklad: Základním souborem E je množina všech ekonomicky zaměřených studentů 1. ročníku českých vysokých škol. Množina G_1 je tvořena těmi studenty, kteří uspěli v prvním zkušebním termínu z matematiky a množina G_2 obsahuje ty studenty, kteří uspěli v prvním zkušebním termínu z angličtiny. Ze základního souboru bylo náhodně vybráno 20 studentů, kteří tvoří výběrový soubor $\{\varepsilon_1, ..., \varepsilon_{20}\}$. Z těchto 20 studentů 12 uspělo v matematice, 15 v angličtině a 11 v obou předmětech. Zapište absolutní a relativní četnosti úspěšných matematiků, angličtinářů a oboustranně úspěšných studentů.

Řešení:

$$N(G_1) = 12, N(G_2) = 15, N(G_1 \cap G_2) = 11, n = 20, p(G_1) = \frac{12}{20} = 0{,}6, p(G_2) = \frac{15}{20} = 0{,}75,$$

$$p(G_1 \cap G_2) = \frac{11}{20} = 0{,}55$$

Vidíme, že úspěšných matematiků je 60%, angličtinářů 75% a oboustranně úspěšných studentů jen 55%.

Relativní četnost

Vlastnosti relativní četnosti: Relativní četnost má následujících 12 vlastností, které jsou obdobné vlastnostem procent.

- $p(\emptyset) = 0$
- $p(G) \geq 0$ (nezápornost)
- $p(G) \leq 1$
- $p(G_1 \cup G_2) + p(G_1 \cap G_2) = p(G_1) + p(G_2)$
- $1 + p(G_1 \cap G_2) \geq p(G_1) + p(G_2)$
- $p(G_1 \cup G_2) + 0 \leq p(G_1) + p(G_2)$ (subaditivita)
- $G_1 \cap G_2 = \emptyset \Rightarrow p(G_1 \cup G_2) = p(G_1) + p(G_2)$ (aditivita)
- $p(G_2 \setminus G_1) = p(G_2) - p(G_1 \cap G_2)$
- $G_1 \subseteq G_2 \Rightarrow p(G_2 \setminus G_1) = p(G_2) - p(G_1)$ (subtraktivita)
- $G_1 \subseteq G_2 \Rightarrow p(G_1) \leq p(G_2)$ (monotonie)
- $p(E) = 1$ (normovanost)
- $p(G) + p(\overline{G}) = 1$ (komplementarita)

Podmíněná relativní četnost

Pokud se v daném základním souboru zajímáme o dvě podmnožiny, můžeme zavést pojem podmíněné relativní četnosti jedné podmnožiny v daném výběrovém souboru za předpokladu, že objekt pochází z druhé množiny.

Nechť E je základní soubor, G_1, G_2 jeho podmnožiny, $\{\varepsilon_1, \ldots, \varepsilon_n\}$ výběrový soubor. Definujeme

podmíněnou relativní četnost množiny G_1 ve výběrovém souboru za předpokladu G_2:

$$p(G_1/G_2) = \frac{N(G_1 \cap G_2)}{N(G_2)} = \frac{p(G_1 \cap G_2)}{p(G_2)} \text{ a}$$

podmíněnou relativní četnost G_2 ve výběrovém souboru za předpokladu G_1:

$$p(G_2/G_1) = \frac{N(G_1 \cap G_2)}{N(G_1)} = \frac{p(G_1 \cap G_2)}{p(G_1)}.$$

Příklad

Příklad: Pro údaje z příkladu o studentech vypočtěte podmíněnou relativní četnost úspěšných matematiků mezi úspěšnými angličtináři a podmíněnou relativní četnost úspěšných angličtinářů mezi úspěšnými matematiky.
(Připomínáme, že z 20 studentů 12 uspělo v matematice, 15 v angličtině a 11 v obou předmětech.)

Řešení:

$N(G_1) = 12, N(G_2) = 15, N(G_1 \cap G_2) = 11, n = 20,$

$p(G_1/G_2) = \frac{N(G_1 \cap G_2)}{N(G_2)} = \frac{11}{15} = 0{,}73$ (tzn., že 73% těch studentů, kteří byli úspěšní v angličtině, uspělo i v matematice)

$p(G_2/G_1) = \frac{N(G_1 \cap G_2)}{N(G_1)} = \frac{11}{12} = 0{,}92$ (tzn., že 92% těch studentů, kteří byli úspěšní v matematice, uspělo i v angličtině)

Četnostní nezávislost

Pojem četnostní nezávislosti dvou množin: O četnostní nezávislosti dvou množin v daném výběrovém souboru hovoříme tehdy, když informace o původu objektu z jedné množiny nijak nemění šance, s nimiž soudíme na jeho původ i z druhé množiny.
V příkladě se studenty by množiny úspěšných matematiků a úspěšných angličtinářů byly četnostně nezávislé, pokud podíl úspěšných matematiků mezi úspěšnými angličtináři by byl stejný jako podíl úspěšných matematiků mezi všemi zkoušenými studenty a stejně tak podíl úspěšných angličtinářů mezi úspěšnými matematiky by byl stejný jako podíl úspěšných angličtinářů mezi všemi zkoušenými studenty, tj.

$$\frac{n(G_1 \cap G_2)}{n(G_2)} = \frac{n(G_1)}{n} \wedge \frac{n(G_1 \cap G_2)}{n(G_1)} = \frac{n(G_2)}{n}.$$

Po snadné úpravě dostaneme multiplikativní vztah

$$\frac{n(G_1 \cap G_2)}{n} = \frac{n(G_1)}{n} \cdot \frac{n(G_2)}{n}, \text{ tj. } p(G_1 \cap G_2) = p(G_1)p(G_2)$$

Řekneme tedy, že množiny G_1, G_2 jsou četnostně nezávislé v daném výběrovém souboru, jestliže $p(G_1 \cap G_2) = p(G_1)p(G_2)$.
(V praxi jen zřídka dojde k tomu, že uvedený vztah platí přesně. Většinou je jen naznačena určitá tendence četnostní nezávislosti.)

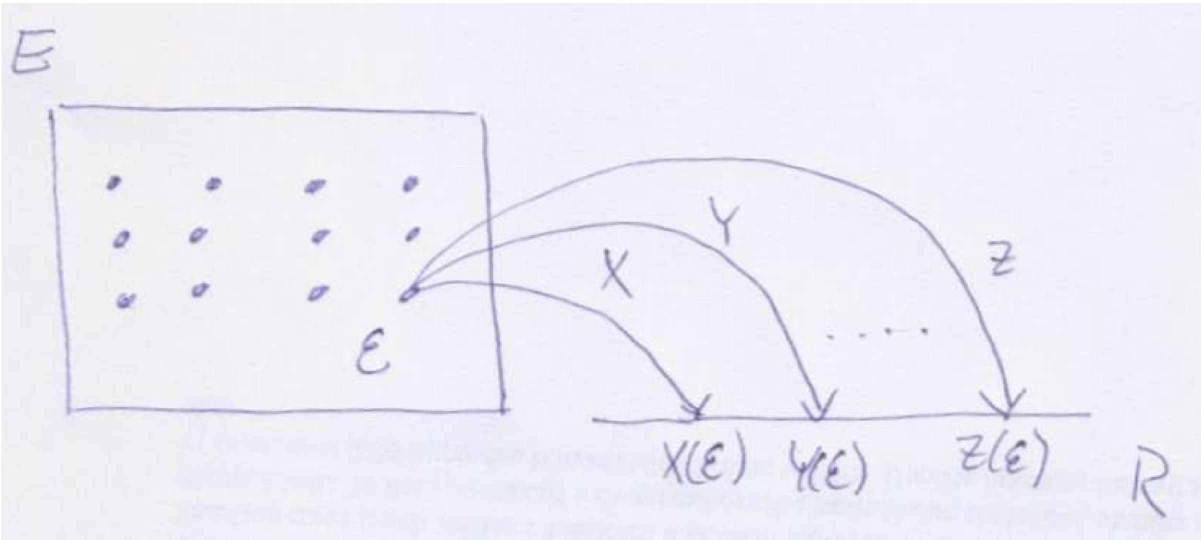
Příklad

Příklad: Pro údaje z příkladu o studentech zjistěte, zda úspěchy v matematice a angličtině jsou v daném výběrovém souboru četnostně nezávislé.
(Připomínáme, že oboustranně úspěšných studentů bylo 55%, úspěšných matematiků 60% a úspěšných angličtinářů 75%.)

Řešení:
$p(G_1 \cap G_2) = 0{,}55$, $p(G_1)p(G_2) = 0{,}6 \times 0{,}75 = 0{,}45$, tedy skutečná relativní četnost oboustranně úspěšných studentů je větší než by odpovídalo četnostní nezávislosti množin G_1, G_2 v daném výběrovém souboru. Znamená to, že úspěch v matematice se zpravidla sdružuje s úspěchem v angličtině a naopak.

Skalární a vektorový znak

Pojem skalárního a vektorového znaku: Vlastnosti objektů vyjadřujeme číselně pomocí znaků.
Nechť E je základní soubor. Funkce
$X: E \rightarrow R, Y: E \rightarrow R, ..., Z: E \rightarrow R$,
které každému objektu přiřazují číslo, se nazývají (skalární) znaky.
Uspořádaná p-tice (X, Y, ..., Z) se nazývá vektorový znak.



Označení: Nechť je dán výběrový soubor $\{\varepsilon_1, ..., \varepsilon_n\} \subseteq E$. Hodnoty znaků X, Y, ..., Z pro i-tý objekt označíme $x_i = X(\varepsilon_i)$, $y_i = Y(\varepsilon_i)$, ..., $z_i = Z(\varepsilon_i)$, $i = 1, ..., n$.

Datový soubor

Pojem datového souboru:

Matice $\begin{pmatrix} x_1 & y_1 & \cdots & z_1 \\ x_2 & y_2 & \cdots & z_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_n & y_n & \cdots & z_n \end{pmatrix}$ typu $n \times p$ se nazývá datový soubor. Její řádky odpovídají jednotlivým objektům, sloupce znakům.

Libovolný sloupec této matice nazýváme jednorozměrným datovým souborem. Jestliže uspořádáme hodnoty některého znaku (např. znaku X) v jednorozměrném datovém souboru vzestupně podle velikosti, dostaneme uspořádaný datový soubor $\begin{pmatrix} x_{(1)} \\ \vdots \\ x_{(n)} \end{pmatrix}$, kde $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Vektor $\begin{pmatrix} x_{[1]} \\ \vdots \\ x_{[r]} \end{pmatrix}$, kde $x_{[1]} < \ldots < x_{[r]}$ jsou navzájem různé hodnoty znaku X, se nazývá vektor variant.

Příklad

Příklad: Pro studenty z výběrového souboru uvedeného výše byly zjišťovány hodnoty znaků X – známka z matematiky v prvním zkušebním termínu, Y – známka z angličtiny v prvním zkušebním termínu, Z – pohlaví studenta (0 ... žena, 1 ... muž). Byl získán datový soubor

$$\begin{pmatrix} 2 & 2 & 0 \\ 1 & 3 & 1 \\ 4 & 3 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 4 & 4 & 1 \\ 3 & 3 & 1 \\ 3 & 4 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \\ 4 & 4 & 0 \\ 2 & 2 & 0 \\ 4 & 3 & 1 \\ 2 & 3 & 1 \\ 4 & 4 & 0 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \\ 4 & 4 & 1 \\ 1 & 3 & 0 \end{pmatrix}$$

Utvořte jednorozměrný uspořádaný i neuspořádaný datový soubor pro známky z matematiky a vektor variant pro známky z matematiky.

Řešení:

$$\begin{pmatrix} 2 \\ 1 \\ 4 \\ 1 \\ 1 \\ 4 \\ 3 \\ 3 \\ 1 \\ 1 \\ 4 \\ 4 \\ 2 \\ 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

Jev

Pojem jevu: Nechť $\{\varepsilon_1, ..., \varepsilon_n\}$ je výběrový soubor, X, Y, ..., Z jsou znaky, B, B_1, …, B_p jsou číselné množiny.
Zápis $\{X \in B\}$ znamená jev „znak X nabyl hodnoty z množiny B“ .
Zápis $\{X \in B_1 \wedge Y \in B_2 \wedge ... \wedge Z \in B_p\}$ znamená jev „znak X nabyl hodnoty z množiny B_1 a současně znak Y nabyl hodnoty z množiny B_2 atd. až znak Z nabyl hodnoty z množiny B_p“.
Symbol $N(X \in B)$ značí absolutní četnost jevu $\{X \in B\}$ ve výběrovém souboru, tj. počet těch objektů ve výběrovém souboru, pro něž $x_i \in B$.
Symbol $p(X \in B)$ znamená relativní četnost jevu $\{X \in B\}$ ve výběrovém souboru, tj. $p(X \in B) = \frac{N(X \in B)}{n}$.
Analogicky $N(X \in B_1 \wedge Y \in B_2 \wedge ... \wedge Z \in B_p)$ resp. $p(X \in B_1 \wedge Y \in B_2 \wedge ... \wedge Z \in B_p)$ znamená absolutní resp. relativní četnost jevu $\{X \in B_1 \wedge Y \in B_2 \wedge ... \wedge Z \in B_p\}$ ve výběrovém souboru.

Příklad

Příklad: Pro datový soubor s údaji o známkách najděte relativní četnost

a) matematických jedničkářů

b) úspěšných matematiků

c) oboustranně neúspěšných studentů.

Datový soubor má tvar:

$$\begin{pmatrix} 2 & 2 & 0 \\ 1 & 3 & 1 \\ 4 & 3 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 4 & 4 & 1 \\ 3 & 3 & 1 \\ 3 & 4 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \\ 4 & 4 & 0 \\ 2 & 2 & 0 \\ 4 & 3 & 1 \\ 2 & 3 & 1 \\ 4 & 4 & 0 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \\ 4 & 4 & 1 \\ 1 & 3 & 0 \end{pmatrix}$$

Řešení:

ad a) $p(X = 1) = \frac{7}{20} = 0{,}35;$

ad b) $p(X \leq 3) = \frac{12}{20} = 0{,}60;$

ad c) $p(X = 4 \wedge Y = 4) = \frac{4}{20} = 0{,}20.$

Zjistili jsme, že jedničku z matematiky mělo 35% studentů, zkoušku z matematiky úspěšně složilo 60% studentů a oboustranně neúspěšných bylo 20% studentů.

Jednorozměrné bodové rozložení četností

Jestliže počet variant znaku X v jednorozměrném datovém souboru není příliš velký, pak přiřazujeme četnosti jednotlivým variantám a hovoříme o bodovém rozložení četností.

Nechť je dán jednorozměrný datový soubor $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, v němž znak X nabývá r variant.

Pro j = 1, ..., r definujeme:

$n_j = N(X = x_{[j]})$ – absolutní četnost varianty $x_{[j]}$ ve výběrovém souboru

$p_j = \frac{n_j}{n}$ - relativní četnost varianty $x_{[j]}$ ve výběrovém souboru

$N_j = N(X \leq x_{[j]}) = n_1 + \dots + n_j$ – absolutní kumulativní četnost prvních j variant ve výběrovém souboru

$F_j = \frac{N_j}{n} = p_1 + \dots + p_j$ – relativní kumulativní četnost prvních j variant ve výb ěrovém souboru

Tabulka typu

$x_{[j]}$	n_j	p_j	N_j	F_j
$x_{[1]}$	n_1	p_1	N_1	F_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{[r]}$	n_r	p_r	N_r	F_r

se nazývá variační řada (nebo též tabulka rozložení četností).

Příklad

Příklad: Máme jednorozměrný datový soubor, který obsahuje údaje o známkách z matematiky (znak X) u 20 studentů.

$$\begin{pmatrix} 2 \\ 1 \\ 4 \\ 1 \\ 1 \\ 4 \\ 3 \\ 3 \\ 1 \\ 1 \\ 4 \\ 4 \\ 2 \\ 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 4 \\ 1 \end{pmatrix}$$

Sestavte tabulku rozložení četností.

Řešení:

$x_{[j]}$	n_j	p_j	N_j	F_j
1	7	7/20=0,35	7	7/20=0,35
2	3	3/20=0,15	10	10/20=0,50
3	2	2/20=0,10	12	12/20=0,60
4	8	8/20=0,40	20	20/20=1,00
$\sum$	20	1,00	-	-

Četnostní funkce, empirická distribuční funkce

Pomocí relativních četností zavedeme četnostní funkci.

Funkce $p(x) = \begin{cases} p_j \text{ pro } x = x_{[j]},\ j = 1, \ldots, r \\ 0 \text{ jinak} \end{cases}$ se nazývá četnostní funkce.

Četnostní funkce je
nezáporná ($\forall x \in R$: $p(x) \geq 0$)
a normovaná ($\sum_{x=-\infty}^{\infty} p(x) = 1$).

Pomocí kumulativních relativních četností zavedeme empirickou distribuční funkci.

Funkce $F(x) = \begin{cases} 0 \text{ pro } x < x_{[1]} \\ F_j \text{ pro } x_{[j]} \leq x < x_{[j+1]},\ j = 1, \ldots, r-1 \\ 1 \text{ pro } x \geq x_{[r]} \end{cases}$ se nazývá empirická distribuční funkce.

Empirická distribuční funkce je
neklesající ($\forall x_1, x_2 \in R$, $x_1 < x_2$: $F(x_1) \leq F(x_2)$),
zprava spojitá ($\forall x_0 \in R$ libovolné, ale pevně dané: $\lim_{x \to x_{0+}} F(x) = F(x_0)$)
a normovaná ($\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$).

Platí $\forall x \in R : F(x) = \sum_{t \leq x} p(t)$.

Příklad

Příklad: Pro známky z matematiky nakreslete graf četnostní funkce a empirické distribuční funkce.

Řešení:

Variační řada

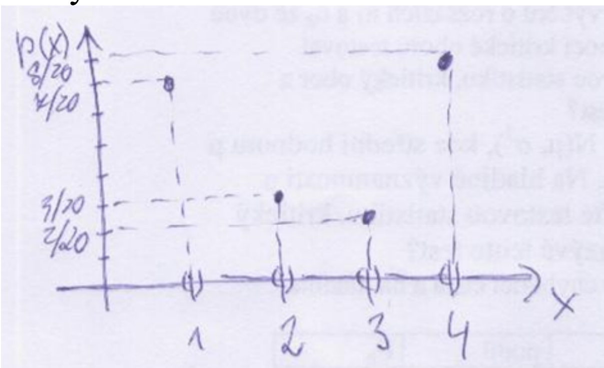
$x_{[j]}$	n_j	p_j	N_j	F_j
1	7	7/20=0,35	7	7/20=0,35
2	3	3/20=0,15	10	10/20=0,50
3	2	2/20=0,10	12	12/20=0,60
4	8	8/20=0,40	20	20/20=1,00
$\sum$	20	1,00	-	-

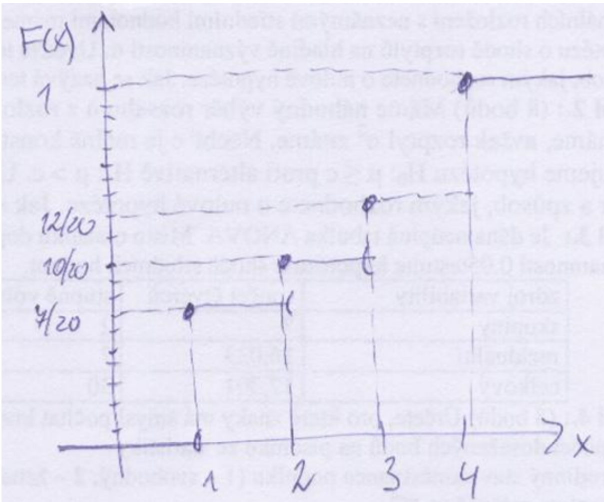
Vzorce

$$p(x) = \begin{cases} p_j \text{ pro } x = x_{[j]},\ j = 1,\ \ldots,\ r \\ 0 \text{ jinak} \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ pro } x < x_{[1]} \\ F_j \text{ pro } x_{[j]} \le x < x_{[j+1]},\ j = 1, \ldots, r-1 \\ 1 \text{ pro } x \ge x_{[r]} \end{cases}$$

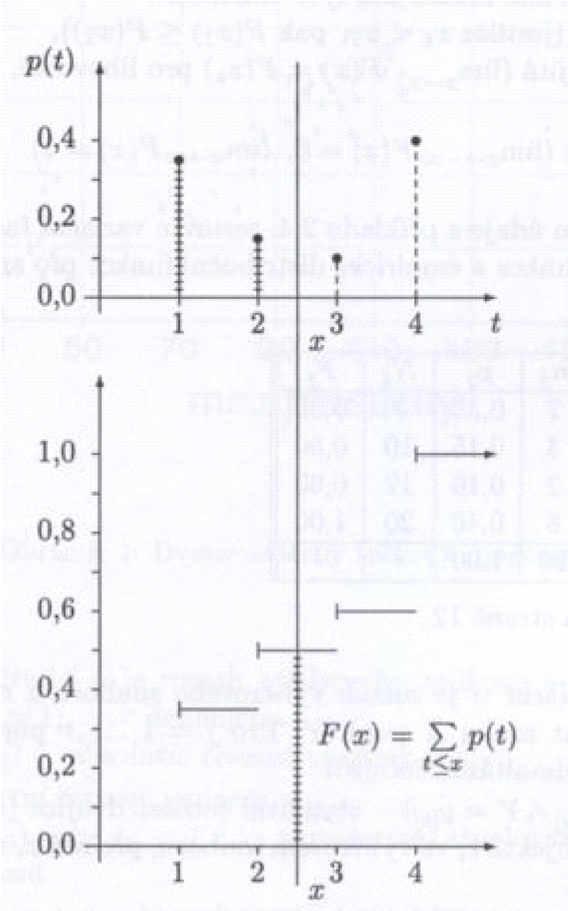
Grafy





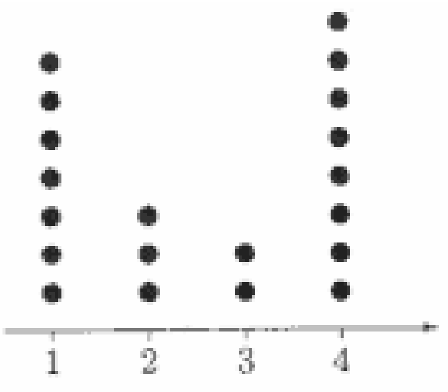
Vztah mezi četnostní funkcí a empirickou distribuční funkcí

$$\forall x \in R : F(x) = \sum_{t \le x} p(t)$$

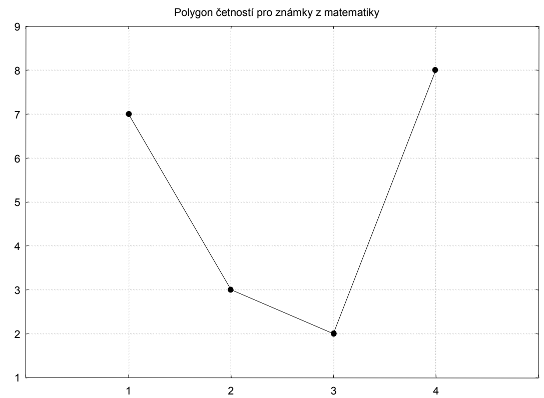


Grafické znázornění jednorozměrného bodového rozdělení četností

Tečkový diagram: na číselné ose vyznačíme jednotlivé varianty znaku X a nad každou variantu nakreslíme tolik teček, jaká je její absolutní četnost.

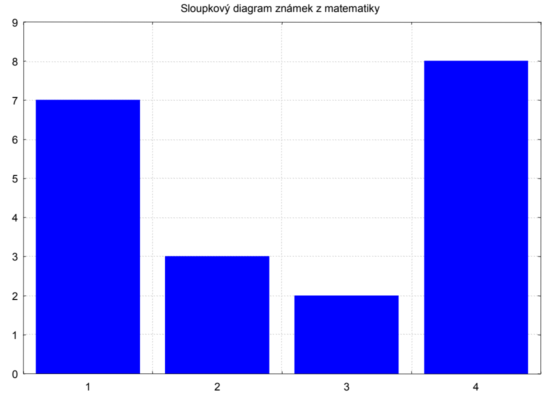


Polygon četnosti: je lomená čára spojující body, jejichž x-ová souřadnice je varianta znaku X a y-ová souřadnice je absolutní či relativní četnost této varianty.

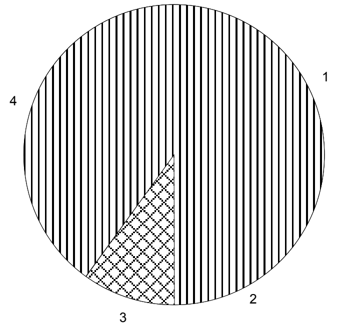


Grafické znázornění jednorozměrného bodového rozdělení četností

Sloupkový diagram: je soustava na sebe nenavazujících obdélníků, kde střed základny je varianta znaku X a výška je absolutní či relativní četnost této varianty.



Výsečový graf: je kruh rozdělený na výseče, jejichž vnější obvod odpovídá absolutním četnostem variant znaku X.



Dvourozměrné bodové rozložení četností

Nechť je dán dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \dots & \dots \\ x_n & y_n \end{pmatrix}$, kde znak X má r variant a znak Y má s variant. Pak definujeme:

$n_{jk} = N(X = x_{[j]} \wedge Y = y_{[k]})$ – simultánní absolutní četnost dvojice $(x_{[j]}, y_{[k]})$ ve výběrovém souboru

$p_{jk} = \frac{n_{jk}}{n}$ – simultánní relativní četnost dvojice $(x_{[j]}, y_{[k]})$ ve výběrovém souboru

$n_{j.} = N(X = x_{[j]}) = n_{j1} + \dots + n_{js}$ – marginální absolutní četnost varianty $x_{[j]}$

$p_{j.} = \frac{n_{j.}}{n} = p_{j1} + \dots + p_{js}$ – marginální relativní četnost varianty $x_{[j]}$

$n_{.k} = N(Y = y_{[k]}) = n_{1k} + \dots + n_{rk}$ – marginální absolutní četnost varianty $y_{[k]}$

$p_{.k} = \frac{n_{.k}}{n} = p_{1k} + \dots + p_{rk}$ – marginální relativní četnost varianty $y_{[k]}$

$N_{jk} = N\left(X \le x_{[j]} \wedge Y \le y_{[k]}\right) = \sum_{u \le j} \sum_{v \le k} n_{uv}$ Absolutní kumulativní četnost dvojice $(x_{[j]}, y_{[k]})$

Simultánní četností zapisujeme do kontinenční tabulky.

Kontingenční tabulka simultánních absolutních četností má tvar:

x \ y, n_{jk}	$y_{[1]}$	...	$y_{[s]}$	$n_{j.}$
$x_{[1]}$	n_{11}	...	n_{1s}	$n_{1.}$
$\vdots$	...	...	...	...
$x_{[r]}$	n_{r1}	...	n_{rs}	$n_{r.}$
$n_{.k}$	$n_{.1}$	...	$n_{.s}$	n

Příklad

Příklad: Máme datový soubor, který obsahuje údaje o známkách z matematiky (znak X), z angličtiny (znak Y) a pohlaví studenta (znak Z, 0 – žena, 1 – muž) u 20 studentů:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X	2	1	4	1	1	4	3	3	1	1	4	4	2	4	2	4	1	4	4	1
Y	2	3	3	1	2	4	3	4	1	1	2	4	2	3	3	4	1	3	4	3
Z	0	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	1	0

Vytvořte kontingenční tabulku simultánních absolutních a relativních četností pro známky z matematiky a angličtiny.

Řešení:

Kontingenční tabulka simultánních absolutních četností

x \ y, n_{jk}	1	2	3	4	$n_{j\cdot}$
1	4	1	2	0	7
2	0	2	1	0	3
3	0	0	1	1	2
4	0	1	3	4	8
$n_{\cdot k}$	4	4	7	5	$n = 20$

Kontingenční tabulka simultánních relativních četností

x \ y, p_{jk}	1	2	3	4	$p_{j\cdot}$
1	0,20	0,05	0,10	0,00	0,35
2	0,00	0,10	0,05	0,00	0,15
3	0,00	0,00	0,05	0,05	0,10
4	0,00	0,05	0,15	0,20	0,40
$p_{\cdot k}$	0,20	0,20	0,35	0,25	1,00

Simultánní a marginální četnostní funkce

Pomocí simultánních relativních četností zavedeme simultánní četnostní funkci:

Funkce

$$p(x,y) = \begin{cases} p_{jk} \text{ pro } x = x_{[j]},\ y = y_{[k]},\ j = 1,\ldots,r,\ k = 1,\ldots,s \\ 0 \text{ jinak} \end{cases}$$

se nazývá simultánní četnostní funkce.

Pomocí marginálních relativních četností zavedeme marginální četnostní funkce pro znaky X a Y. Odlišíme je indexem takto:

$$p_1(x) = \begin{cases} p_{j.} \text{ pro } x = x_{[j]},\ j = 1,\ldots,r \\ 0 \text{ jinak} \end{cases},$$

$$p_2(y) = \begin{cases} p_{.k} \text{ pro } y = y_{[k]},\ k = 1,\ldots,s \\ 0 \text{ jinak} \end{cases}.$$

Mezi simultánní četnostní funkcí a marginálními četnostními funkcemi platí vztahy:

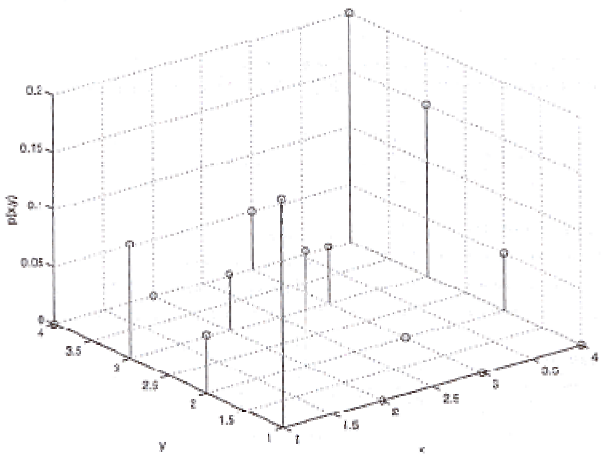
$$p_1(x) = \sum_{y=-\infty}^{\infty} p(x,y),\quad p_2(y) = \sum_{x=-\infty}^{\infty} p(x,y).$$

Příklad

Příklad: Sestrojte graf simultánní četnostní funkce pro známky z matematiky a angličtiny.

Řešení: Vyjdeme z kontingenční tabulky simultánních relativních četností.

	y	1	2	3	4	$p_{j\cdot}$
x	p_{jk}					
1		0,20	0,05	0,10	0,00	0,35
2		0,00	0,10	0,05	0,00	0,15
3		0,00	0,00	0,05	0,05	0,10
4		0,00	0,05	0,15	0,20	0,40
$p_{\cdot k}$		0,20	0,20	0,35	0,25	1,00



Četností nezávislost znaků v daném výběrovém souboru

Řekneme, že znaky X, Y jsou v daném výběrovém souboru četnostně nezávislé, právě když pro všechna j = 1, ..., r a všechna k = 1, ..., s platí multiplikativní vztah:
$p_{jk} = p_{j.}\, p_{.k}$ neboli pro $\forall (x, y) \in R^2$: $p(x, y) = p_1(x)\, p_2(y)$.

Příklad: Ověřte, zda v našem datovém souboru jsou známky z matematiky a angličtiny četnostně nezávislé.

Řešení: Vyjdeme z kontingenční tabulky relativních četností.

x \ y (p_{jk})	1	2	3	4	$p_{j\cdot}$
1	0,20	0,05	0,10	0,00	0,35
2	0,00	0,10	0,05	0,00	0,15
3	0,00	0,00	0,05	0,05	0,10
4	0,00	0,05	0,15	0,20	0,40
$p_{\cdot k}$	0,20	0,20	0,35	0,25	1,00

Známky z matematiky a angličtiny nejsou četnostně nezávislé, protože už pro j = 1, k = 1 je multiplikativní vztah porušen:
$p_{11} = 0{,}20$, $p_{1.} = 0{,}35$, $p_{.1} = 0{,}20$, tudíž $0{,}20 \neq 0{,}35.0{,}20$

Řádkově a sloupcově podmíněné relativní četnosti

Sloupcově podmíněná relativní četnost varianty $x_{[j]}$ za předpokladu $y_{[k]}$

$$p_{j(k)} = \frac{n_{jk}}{n_{\cdot k}}$$

Řádkově podmíněná relativní četnost varianty $y_{[k]}$ za předpokladu $x_{[j]}$

$$p_{(j)k} = \frac{n_{jk}}{n_{j\cdot}}$$

Příklad

Příklad: Pro datový soubor známek z matematiky a angličtiny sestavte kontingenční tabulku sloupcově a poté řádkově podmíněných relativních četností.

Řešení:

Nejprve se budeme zabývat sloupcově podmíněnými relativními četnostmi. Použijeme vzorec $p_{j(k)} = \frac{n_{jk}}{n_{.k}}$.

Vyjdeme z kontingenční tabulky simultánních absolutních četností.

x \ y (n_{jk})	1	2	3	4	$n_{j.}$
1	4	1	2	0	7
2	0	2	1	0	3
3	0	0	1	1	2
4	0	1	3	4	8
$n_{.k}$	4	4	7	5	$n = 20$

x \ y ($p_{j(k)}$)	1	2	3	4
1	1,00	0,25	0,29	0,00
2	0,00	0,50	0,14	0,00
3	0,00	0,00	0,14	0,20
4	0,00	0,25	0,43	0,80
$\sum$	1,00	1,00	1,00	1,00

Interpretujeme např. třetí sloupec: z těch studentů, kteří měli trojku z angličtiny, mělo 2/7 = 29% jedničku z matematiky, 1/7 = 14% dvojku z matematiky, 1/7 = 14% trojku z matematiky a 3/7 = 43% čtyřku z matematiky.

Příklad

Dále se budeme zabývat řádkově podmíněnými relativními četnostmi. Použijeme vzorec $p_{(j)k} = \dfrac{n_{jk}}{n_{j.}}$.

Opět nám poslouží kontingenční tabulka absolutních četností.

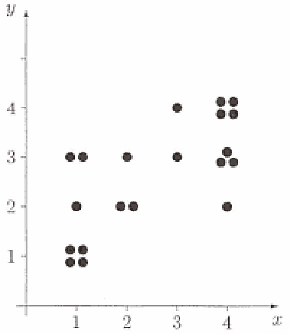
x \ y (n_{jk})	1	2	3	4	$n_{j.}$
1	4	1	2	0	7
2	0	2	1	0	3
3	0	0	1	1	2
4	0	1	3	4	8
$n_{.k}$	4	4	7	5	$n = 20$

x \ y ($p_{(j)k}$)	1	2	3	4	$\sum$
1	0,57	0,14	0,29	0,00	1,00
2	0,00	0,67	0,33	0,00	1,00
3	0,00	0,00	0,50	0,50	1,00
4	0,00	0,12	0,38	0,50	1,00

Interpretujeme např. první řádek: z těch studentů, kteří měli jedničku z matematiky, mělo 4/7 = 57% jedničku z angličtiny, 1/7 = 14% dvojku z angličtiny a 2/7 = 29% trojku z angličtiny.

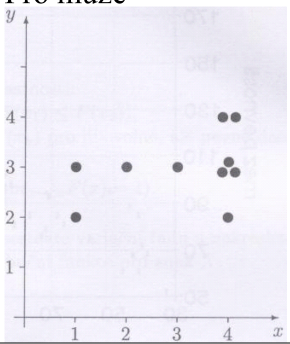
Dvourozměrný tečkový diagram

Dvourozměrné rozložení četností lze znázornit pomocí dvourozměrného tečkového diagramu. Na vodorovnou osu vyneseme varianty znaku X, na svislouvarianty znaku Y a do příslušných průsečíků nakreslíme tolik teček, jaká je absolutní četnost dané dvojice. V našem příkladě se studenty dostaneme tento diagram:

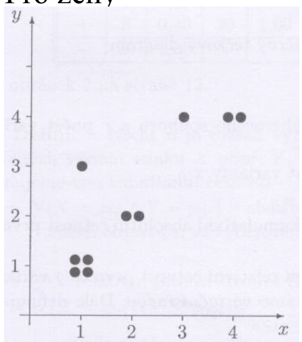


Dvourozměrný tečkový diagram svědčí o nepříliš výrazné tendenci k podobné klasifikaci v obou předmětech.
Zcela odlišný vzhled má diagram pro muže a pro ženy:

Pro muže



Pro ženy



Intervalové rozložení četností

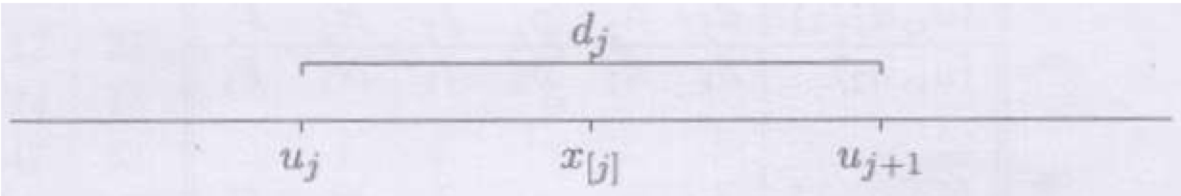
Nechť je dán jednorozměrný datový soubor. Jestliže počet variant znaku X je blízký rozsahu souboru, pak četnosti přiřazujeme nikoliv jednotlivým variantám, ale celým intervalům hodnot. Hovoříme pak o intervalovém rozložení četnosti.

Číselnou osu rozložíme na intervaly typu $(-\infty, u_1\rangle$, $(u_1, u_2\rangle$, ..., $(u_r, u_{r+1}\rangle$, (u_{r+1}, ∞) tak, aby okrajové intervaly neobsahovaly žádnou pozorovanou hodnotu znaku X. Užíváme označení:

$(u_j, u_{j+1}\rangle$ – j-tý třídicí interval znaku X, j = 1, ..., r.

$d_j = u_{j+1} - u_j$ – délka j-tého třídicího intervalu znaku X

$x_{[j]} = \frac{u_j + u_{j+1}}{2}$ – střed j-tého třídicího intervalu znaku X



Intervalové rozložení četností – stanovení počtu tříd

Třídicí intervaly volíme nejčastěji stejně dlouhé. Jejich počet určíme např. pomocí Sturgersova pravidla: r = 1 + 3,3 log n, kde n je rozsah souboru.

- počet tříd (r):
  - do 100 prvků……………6 až 9 tříd
  - do 500 prvků……………10 až 15 tříd
  - nad 500 prvků…………..Sturgesovo pravidlo

$$r \approx 1 + 3{,}3 \log n$$

log…dekadický logaritmus!!!

Sestavení tabulky rozložení četností

Hodnoty znaku X roztřídíme do r třídicích intervalů. Pro j = 1, ..., r definujeme:

$n_j = N(u_j < X \leq u_{j+1})$ – absolutní četnost j-tého třídicího intervalu ve výběrovém souboru

$p_j = \frac{n_j}{n}$ – relativní četnost j-tého třídicího intervalu ve výběrovém souboru

$f_j = \frac{p_j}{d_j}$ – četnostní hustota j-tého třídicího intervalu ve výběrovém souboru

$N_j = N(X \leq u_{j+1}) = n_1 + \ldots + n_j$ – absolutní kumulativní četnost prvních j třídicích intervalů ve výběrovém souboru

$F_j = \frac{N_j}{n} = p_1 + \ldots + p_j$ – relativní kumulativní četnost prvních j třídicích intervalů ve výběrovém souboru.

Tabulka typu

(u_j, u_{j+1})	d_j	n_j	p_j	f_j	N_j	F_j
(u_1, u_2)	d_1	n_1	p_1	f_1	N_1	F_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
(u_r, u_{r+1})	d_r	n_r	p_r	f_r	N_r	F_r
Součet		n	1			

se nazývá tabulka rozložení četností.

Příklad

Příklad: Do laboratoře bylo dodáno 60 vzorků a byly zjištěny a hodnoty znaku X – mez plasticity (v kp/cm^2) a Y – mez pevnosti (v kp/cm^2). Datový soubor má tvar:

154	178	83	98	73	76
133	164	106	111	77	85
58	75	92	104	47	61
145	161	85	103	68	85
94	107	112	118	137	142
113	141	98	102	44	68
86	97	103	108	92	116
121	127	99	119	141	157
119	138	104	128	155	189
112	125	107	118	136	155
85	97	98	140	82	81
41	72	97	115	136	163
96	113	105	101	72	79
45	89	71	93	66	81
99	109	39	69	42	61
51	95	122	147	113	123
101	114	33	52	42	85
160	169	78	117	133	147
87	101	114	137	153	179
88	139	125	149	85	91

a) Pro znak X stanovte optimální počet třídicích intervalů dle Sturgersova pravidla.
b) Sestavte tabulku rozložení četností.

Příklad

Řešení:

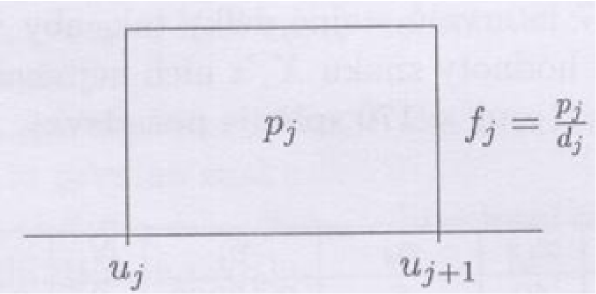
ad a) Rozsah souboru je 60. Podle Sturgersova pravidla je optimální počet třídicích intervalů r = 7. Budeme tedy volit 7 intervalů stejné délky tak, aby v nich byly obsaženy všechny pozorované hodnoty znaku X, z nichž nejmenší je 33, největší 160; volba $u_1 = 30, \ldots, u_8 = 170$ splňuje požadavky.

ad b)

$\langle u_j, u_{j+1} \rangle$	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
$\langle 30, 50 \rangle$	20	40	8	$8/60 = 0{,}1\overline{3}$	8	$8/60 = 0{,}1\overline{3}$	$8/(60 \cdot 20) = 0{,}00\overline{6}$
$\langle 50, 70 \rangle$	20	60	4	$4/60 = 0{,}0\overline{6}$	12	$12/60 = 0{,}2$	$4/(60 \cdot 20) = 0{,}00\overline{3}$
$\langle 70, 90 \rangle$	20	80	13	$13/60 = 0{,}21\overline{6}$	25	$25/60 = 0{,}41\overline{6}$	$13/(60 \cdot 20) = 0{,}0183$
$\langle 90, 110 \rangle$	20	100	15	$15/60 = 0{,}25$	40	$40/60 = 0{,}\overline{6}$	$15/(60 \cdot 20) = 0{,}0125$
$\langle 110, 130 \rangle$	20	120	9	$9/60 = 0{,}15$	49	$49/60 = 0{,}81\overline{6}$	$9/(60 \cdot 20) = 0{,}0075$
$\langle 130, 150 \rangle$	20	140	7	$7/60 = 0{,}11\overline{6}$	56	$56/60 = 0{,}9\overline{3}$	$7/(60 \cdot 20) = 0{,}00583$
$\langle 150, 170 \rangle$	20	160	4	$4/60 = 0{,}0\overline{6}$	60	$60/60 = 1$	$4/(60 \cdot 20) = 0{,}00\overline{3}$
Součty			60	1			

Histogram, hustota četnosti, intervalová empirická distribuční funkce

Intervalové rozložení četností graficky znázorňujeme pomocí histogramu. Je to graf skládající se z r obdélníků, sestrojených nad třídicími intervaly, přičemž obsah j-tého obdélníku je roven relativní četnosti p_j j-tého třídicího intervalu, j = 1, ..., r.



Histogram je shora omezen schodovitou čarou, která je grafem funkce zvané hustota četnosti:

$$f(x) = \begin{cases} f_j \text{ pro } u_j < x \le u_{j+1}, j = 1, \cdots, r \\ 0 \text{ jinak} \end{cases}$$

Pomocí hustoty četnosti zavedeme intervalovou empirickou distribuční funkci:

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

Hustota četnosti je nezáporná ($\forall x \in R : f(x) \ge 0$) a normovaná ($\int_{-\infty}^{\infty} f(x)dx = 1$). Intervalová empirická distribuční funkce je neklesající, spojitá a normovaná ($\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$).

Příklad

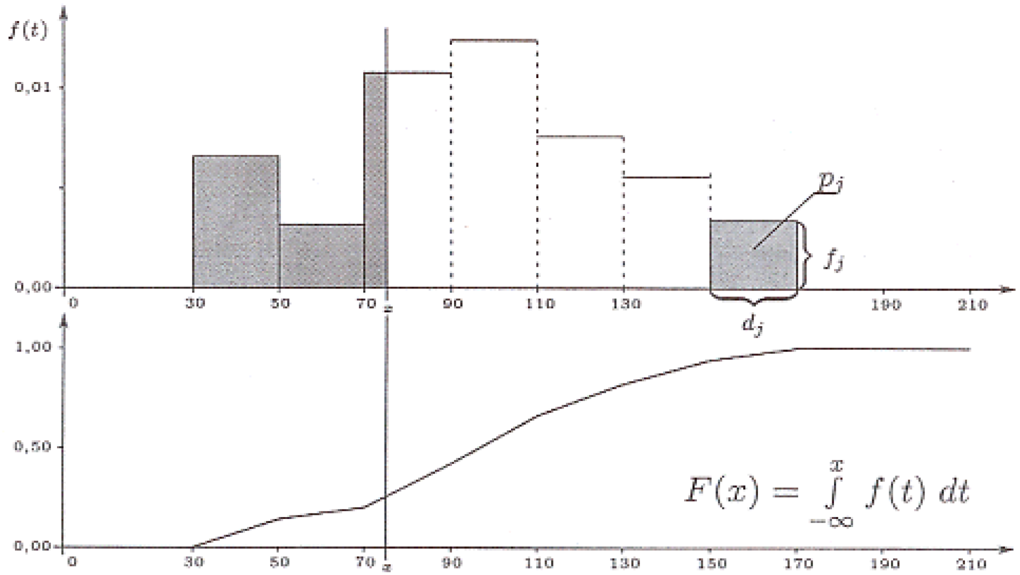
Příklad: Pro mez plasticity oceli nakreslete histogram a pod histogram graf intervalové empirické distribuční funkce.

Řešení: Vyjdeme z tabulky rozložení četností.

$\langle u_j, u_{j+1})$	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
$\langle 30, 50)$	20	40	8	$8/60 = 0{,}1\overline{3}$	8	$8/60 = 0{,}1\overline{3}$	$8/(60 \cdot 20) = 0{,}00\overline{6}$
$\langle 50, 70)$	20	60	4	$4/60 = 0{,}0\overline{6}$	12	$12/60 = 0{,}2$	$4/(60 \cdot 20) = 0{,}00\overline{3}$
$\langle 70, 90)$	20	80	13	$13/60 = 0{,}21\overline{6}$	25	$25/60 = 0{,}41\overline{6}$	$13/(60 \cdot 20) = 0{,}0183$
$\langle 90, 110)$	20	100	15	$15/60 = 0{,}25$	40	$40/60 = 0{,}\overline{6}$	$15/(60 \cdot 20) = 0{,}0125$
$\langle 110, 130)$	20	120	9	$9/60 = 0{,}15$	49	$49/60 = 0{,}81\overline{6}$	$9/(60 \cdot 20) = 0{,}0075$
$\langle 130, 150)$	20	140	7	$7/60 = 0{,}11\overline{6}$	56	$56/60 = 0{,}9\overline{3}$	$7/(60 \cdot 20) = 0{,}00583$
$\langle 150, 170)$	20	160	4	$4/60 = 0{,}0\overline{6}$	60	$60/60 = 1$	$4/(60 \cdot 20) = 0{,}00\overline{3}$
Součty			60	1			

Příklad

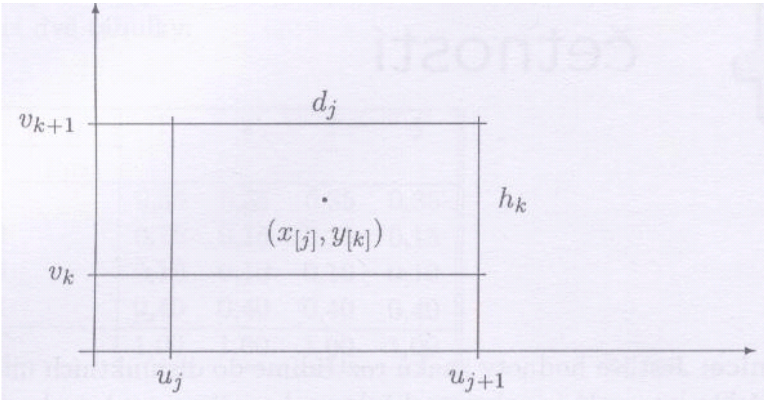
$(u_j, u_{j+1}\rangle$	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
$(30,50\rangle$	20	40	8	$8/60=0{,}1\overline{3}$	8	$8/60=0{,}1\overline{3}$	$8/(60\cdot 20)=0{,}00\overline{6}$
$(50,70\rangle$	20	60	4	$4/60=0{,}0\overline{6}$	12	$12/60=0{,}2$	$4/(60\cdot 20)=0{,}00\overline{3}$
$(70,90\rangle$	20	80	13	$13/60=0{,}21\overline{6}$	25	$25/60=0{,}41\overline{6}$	$13/(60\cdot 20)=0{,}0183$
$(90,110\rangle$	20	100	15	$15/60=0{,}25$	40	$40/60=0{,}\overline{6}$	$15/(60\cdot 20)=0{,}0125$
$(110,130\rangle$	20	120	9	$9/60=0{,}15$	49	$49/60=0{,}81\overline{6}$	$9/(60\cdot 20)=0{,}0075$
$(130,150\rangle$	20	140	7	$7/60=0{,}11\overline{6}$	56	$56/60=0{,}9\overline{3}$	$7/(60\cdot 20)=0{,}00583$
$(150,170\rangle$	20	160	4	$4/60=0{,}0\overline{6}$	60	$60/60=1$	$4/(60\cdot 20)=0{,}00\overline{3}$
Součty			60	1			



Dvourozměrné intervalové rozložení četností

Dále se budeme věnovat dvourozměrnému intervalovému rozložení četností, tj. budeme pracovat s dvourozměrným datovým souborem. Zavedeme podobné pojmy jako u dvourozměrného bodového rozložení četností

Nechť je dán dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \dots & \dots \\ x_n & y_n \end{pmatrix}$, kde hodnoty znaku X roztřídíme do r třídicích intervalů $(u_j, u_{j+1}\rangle$, $j = 1, \dots, r$ s délkami $d_1, \dots, d_r$ a hodnoty znaku Y roztřídíme do s třídicích intervalů $(v_k, v_{k+1}\rangle$, $k = 1, \dots, s$ s délkami $h_1, \dots, h_s$.
Obdélník $(u_j, u_{j+1}\rangle \times (v_k, v_{k+1}\rangle$ se nazývá (j,k) -tý dvourozměrný třídicí interval.



Simultánní a marginální četnosti

$n_{jk} = N(u_j < X \leq u_{j+1} \wedge v_k < Y \leq v_{k+1})$ – simultánní absolutní četnost (j, k)-tého třídicího intervalu.

$p_{jk} = \frac{n_{jk}}{n}$ – simultánní relativní četnost (j, k)-tého třídicího intervalu.

$n_{j.} = n_{j1} + \ldots + n_{js}$ – marginální absolutní četnost j-tého třídicího intervalu pro znak X.

$p_{j.} = \frac{n_{j.}}{n}$ – marginální relativní četnost j-tého třídicího intervalu pro znak X.

$n_{.k} = n_{1k} + \ldots + n_{rk}$ – marginální absolutní četnost k-tého třídicího intervalu pro znak Y.

$p_{.k} = \frac{n_{.k}}{n}$ – marginální relativní četnost k-tého třídicího intervalu pro znak Y.

$f_{jk} = \frac{p_{jk}}{d_j h_k}$ – simultánní četnostní hustota v (j, k)-tém třídicím intervalu.

$f_{j.} = \frac{p_{j.}}{d_j}$ – marginální četnostní hustota v j-tém třídicím intervalu pro znak X.

$f_{.k} = \frac{p_{.k}}{h_k}$ – marginální četnostní hustota v k-tém třídicím intervalu pro znak Y.

Kteroukoliv ze simultánních četností zapisujeme do kontingenční tabulky.

Kontingenční tabulka simultánních absolutních četností:

$(u_j, u_{j+1}\rangle$ \ $(v_k, v_{k+1}\rangle$ / n_{jk}	$(v_1, v_2\rangle$	$\ldots$	$(v_s, v_{s+1}\rangle$	$n_{j.}$
$(u_1, u_2\rangle$	n_{11}	$\ldots$	n_{1s}	$n_{1.}$
$\vdots$				$\vdots$
$(u_r, u_{r+1}\rangle$	n_{r1}	$\ldots$	n_{rs}	$n_{r.}$
$n_{.k}$	$n_{.1}$	$\ldots$	$n_{.s}$	n

Příklad

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti (znak Y) oceli
a) stanovte dle Sturgersova pravidla optimální počet třídicích intervalů pro znak Y
b) sestavte kontingenční tabulku simultánních absolutních četností.

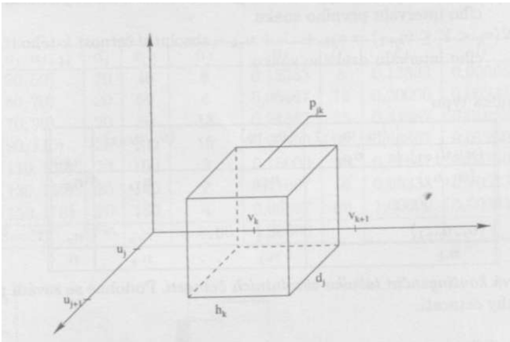
Řešení:
ad a) Rozsah datového souboru je 60. Podle Sturgersova pravidla je tedy optimální počet třídicích intervalů 7. Nejmenší hodnota je 52 a největší 189. Volíme $v_1 = 50$, $v_2 = 70$, ..., $v_8 = 190$.
ad b)

$\langle u_j, u_{j+1})$ \ $\langle v_k, v_{k+1})$ n_{jk}	$\langle 50, 70)$	$\langle 70, 90)$	$\langle 90, 110)$	$\langle 110, 130)$	$\langle 130, 150)$	$\langle 150, 170)$	$\langle 170, 190)$	$n_{j.}$
$\langle 30, 50)$	5	3	0	0	0	0	0	8
$\langle 50, 70)$	0	3	1	0	0	0	0	4
$\langle 70, 90)$	0	4	7	1	1	0	0	13
$\langle 90, 110)$	0	0	6	8	1	0	0	15
$\langle 110, 130)$	0	0	0	4	5	0	0	9
$\langle 130, 150)$	0	0	0	0	2	5	0	7
$\langle 150, 170)$	0	0	0	0	0	1	3	4
$n_{.k}$	5	10	14	13	9	6	3	$n = 60$

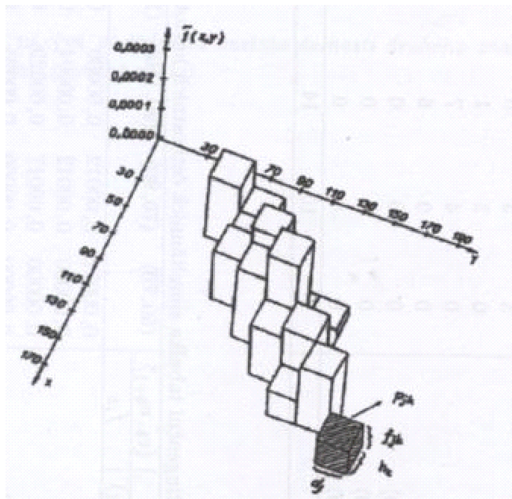
Stereogram

Dvourozměrné intervalové rozložení četností graficky znázorňujeme pomocí stereogramu . Je to graf skládající se z r * s kvádrů, sestrojených nad dvourozměrnými třídicími intervaly, pľičemž objem (j, k)-tého kvádru je roven relativní četnosti p_{jk} (j, k)-tého třídicího intervalu,

j = 1, ..., r, k = 1, …, s. Výška kvádru tedy vyjadřuje simultánní četnostní hustotu.



V našem příkladě s mezí plasticity
a mezí pevnosti oceli bude mít stereogram tvar:



Simultánní a marginální hustota četnosti

Pomocí simultánních četnostních hustot zavedeme simultánní hustotu četnosti:

Funkce $f(x,y)=\begin{cases} f_{jk} \text{ pro } u_j < x \le u_{j+1}, v_k < y \le v_{k+1}, j=1,\cdots,r, k=1,\cdots,s \\ 0 \text{ jinak} \end{cases}$ se nazývá simultánní hustota četnosti. Jejím grafem je schodovitá plocha shora omezující stereogram.

Hustoty četnosti pro znaky X a Y odlišíme indexem takto:

$$f_1(x)=\begin{cases} f_{j.} \text{ pro } u_j < x \le u_{j+1}, j=1,\cdots,r \\ 0 \text{ jinak} \end{cases},$$

$$f_2(y)=\begin{cases} f_{.k} \text{ pro } v_k < y \le v_{k+1}, k=1,\cdots,s \\ 0 \text{ jinak} \end{cases}.$$

Mezi simultánní hustotou četnosti a marginálními hustotami četnosti platí vztahy:

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)dy,\quad f_2(y)=\int_{-\infty}^{\infty} f(x,y)dx\,.$$

Četnostní nezávislost znaků v daném výběrovém souboru při intervalovém rozložení četností

Pomocí simultánních a marginálních četností zavedeme pojem četnostní nezávislosti znaků v daném výběrovém souboru při intervalovém rozložení četností:
Řekneme, že znaky X, Y jsou v daném výběrovém souboru četnostně nezávislé při intervalovém rozložení četností, jestliže pro všechna j = 1, ..., r a všechna k = 1, ...,s platí multiplikativní vztah: $f_{jk} = f_{j.}\, f_{.k}$ neboli pro $\forall (x, y) \in R^2$: $f(x, y) = f_1(x)\, f_2(y)$.

V našem příkladě nejsou mez pevnosti a mez plasticity četnostně nezávislé, protože už pro j = 1, k = 1 je multiplikativní vztah porušen:

$\langle u_j, u_{j+1})$ \ $\langle v_k, v_{k+1})$, n_{jk}	$\langle 50, 70)$	$\langle 70, 90)$	$\langle 90, 110)$	$\langle 110, 130)$	$\langle 130, 150)$	$\langle 150, 170)$	$\langle 170, 190)$	$n_{j.}$
$\langle 30, 50)$	5	3	0	0	0	0	0	8
$\langle 50, 70)$	0	3	1	0	0	0	0	4
$\langle 70, 90)$	0	4	7	1	1	0	0	13
$\langle 90, 110)$	0	0	6	8	1	0	0	15
$\langle 110, 130)$	0	0	0	4	5	0	0	9
$\langle 130, 150)$	0	0	0	0	2	5	0	7
$\langle 150, 170)$	0	0	0	0	0	1	3	4
$n_{.k}$	5	10	14	13	9	6	3	$n = 60$

$f_{11} = \dfrac{5}{60 \cdot 20 \cdot 20} = 0{,}000208$, $f_{1.} = \dfrac{8}{60 \cdot 20} = 0{,}006667$, $f_{.1} = \dfrac{5}{60 \cdot 20} = 0{,}004167$, tudíž
$0{,}000208 \neq 0{,}006667 . 0{,}004167 = 0{,}000028$

2. Číselné charakteristiky znaků

Doposud jsme se zabývali funkcionálními charakteristikami znaků, jako jsou:
- empirická distribuční funkce F(x),
- simultánní četnostní funkce p(x,y),
- marginální četnostní funkce $p_1(x)$, $p_2(y)$,
- simultánní hustots četnosti f(x,y),
- marginální hustoty četnosti $f_1(x)$, $f_2(y)$,

které nesou úplnou informaci o rozložení četností.
Nyní zavedeme číselné charakteristiky, které nás informují o některých rysech tohoto rozložení četností:
- o poloze (úrovni) hodnot znaku,
- o jejich variabilitě (rozptýlení),
- o těsnosti závislosti dvou znaků
- a pod.

Pro různé typy znaků se používají různé číselné charakteristiky, proto se nejdřív seznámíme s jednotlivými typy znaků.

Typy znaků

Nominální znak: připouští obsahovou interpretaci pouze u relace rovnosti =. O dvou variantách nominálního znaku lze pouze konstatovat, že jsou buď stejné nebo různé. Čísla, která přiřadíme jednotlivým variantám znaku, nereprezentují skutečnou hodnotu použitých čísel, ale jsou pouhým označením variant znaku.
Příklady nominálních znaků: lékařská diagnóza, typ profese, barva očí, rodinný stav, národnost, …

Ordinální znak: připouští obsahovou interpretaci nejen u relace rovnosti =, ale též u relace uspořádání <. Můžeme tedy konstatovat, že varianta $x_{[j]}$ je větší (dokonalejší, silnější, vhodnější) než varianta $x_{[k]}$.
Příklad ordinálního znaku: školní klasifikace vyjadřuje menší nebo větší znalosti zkoušených žáků – jedničkář je lepší než dvojkař, ale intervaly mezi známkami nemají obsahovou interpretaci. Nelze tvrdit, že rozdíl ve znalostech mezi jedničkářem a dvojkařem je stejný jako mezi trojkařem a čtyřkařem.
Další příklady: Různá bodování ve sportovních a uměleckých soutěžích, posuzování různých rysů sociálního chování, posuzování stavu pacientů, hodnocení postojů respondentů k různým otázkám, …

Typy znaků

Intervalový znak: kromě relací rovnosti = a uspořádání < umožňuje obsahovou interpretaci také u operace rozdílu -, tj. stejný interval mezi jednou dvojicí hodnot a jinou dvojicí hodnot vyjadřuje i stejný rozdíl v extenzitě zkoumané vlastnosti.
Příklad intervalového znaku: teplota měřená ve stupních Celsia. Např. naměříme-li ve čtyřech po sobě jdoucích dnech polední teploty 0, 2, 4, 6 ºC, znamená to, že každým dnem stouply teploty o 2 ºC. Nelze však říci, že z druhého na třetí den vzrostla teplota dvojnásobně, kdežto ze třetího na čtvrtý den pouze jeden a půl krát.
Další příklady: kalendářní systémy, směr větru, inteligenční kvocient, …
Společný znak intervalových znaků: nula byla stanovena uměle, pouhou konvencí.

Poměrový znak: kromě relací rovnosti = a uspořádání < umožňuje obsahovou interpretaci také u operací rozdílu - a podílu /, tj. stejný poměr mezi jednou dvojicí hodnot a jinou dvojicí hodnot vyjadřuje i stejný podíl v extenzitě zkoumané vlastnosti.
Příklad poměrového znaku: délka předmětu měřená v cm. Má-li jeden předmět délku 8 cm a druhý 16 cm, má smysl prohlásit, že druhý předmět je dvakrát delší než první předmět.
Další příklady: počet dětí v rodině, výška kapesného v Kč, hmotnost osoby, …
Společný znak poměrových znaků: Poměrový znak má přirozený počátek, ke kterému jsou vztahovány všechny další hodnoty znaku.

Mimo uvedenou klasifikaci stojí alternativní znaky, které nabývají jen dvou hodnot, např. 0,1, což znamená absenci a prezenci nějakého jevu. Například 0 bude znamenat neúspěch, 1 úspěch při řešení určité úlohy. Alternativní znaky mohou být ztotožněny s kterýmkoliv z předcházejících typů.

Typy znaků II

- Demografické znaky:
  - **Klienta (věk, pohlaví, rodinný stav, počet dětí, druh bydlení, kraj/okres trvalého bydliště…)**
  - **Prodejního místa (kraj/okres, typ, prodejní plocha,…)**
  - **Prodejce (věk, pohlaví, kraj/okres trvalého bydliště…)**
- Behaviorální znaky:
  - **Klienta („stáří“ klienta, doposud splacená jistina, dlužná jistina, počet dní po splatnosti,…)**
  - **Prodejního místa („stáří“ prodejny, počet uzavřených smluv, objem uzavřených smluv, podíl nesplácených úvěrů,…)**
  - **Prodejce (počet uzavřených smluv, objem uzavřených smluv, podíl nesplácených úvěrů …)**
- Produktové znaky:
  - **Výše úvěru, délka smlouvy, akontace, RPSN,…**

Číselné charakteristiky nominálních znaků

Charakteristika polohy: modus – nejčetnější varianta resp. střed nejčetnějšího třídicího intervalu.

Příklad na stanovení modu
20 náhodně vybraných osob mělo odpovědět na otázku, který z pěti výrobků (označíme je A, B, C, D, E) preferují. Výsledky máme v tabulce:

Výrobek	A	B	C	D	E
Četnost odpovědí	3	5	3	6	3

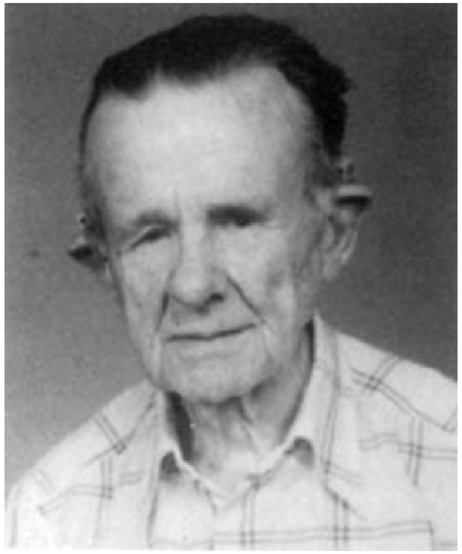
Stanovte modus.

Řešení:
Modus = D

Označení: $\hat{x}$

Cramérův koeficient

Charakteristika těsnosti závislosti dvou nominálních znaků: Cramérův koeficient kontingence.



Carl Harald Cramér (1893 – 1985): Švédský matematik

Cramérův koeficient

Nechť znak X nabývá variant $x_{[1]}, \ldots, x_{[r]}$ a znak Y nabývá variant $y_{[1]}, \ldots, y_{[s]}$. Máme dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \ldots & \ldots \\ x_n & y_n \end{pmatrix}$. Zjistíme absolutní četnosti n_{jk} dvojice variant $(x_{[j]}, y_{[k]})$, j = 1, …, r, k = 1, …, s a uspořádáme je do kontingenční tabulky:

x \ y, n_{jk}	$y_{[1]}$	…	$y_{[s]}$	$n_{j.}$
$x_{[1]}$	n_{11}	…	n_{1s}	$n_{1.}$
⋮	…	…	…	…
$x_{[r]}$	n_{r1}	…	n_{rs}	$n_{r.}$
$n_{.k}$	$n_{.1}$	…	$n_{.s}$	n

Vypočteme tzv. teoretické četnosti $\frac{n_{j.}n_{.k}}{n}$ a s jejich pomocí pak statistiku

$K = \sum_{j=1}^{r}\sum_{k=1}^{s}\frac{\left(n_{jk} - \frac{n_{j.}n_{.k}}{n}\right)^2}{\frac{n_{j.}n_{.k}}{n}}$. Cramérův koeficient: $V = \sqrt{\frac{K}{n(m-1)}}$, kde m = min{r,s}. Tento koeficient nabývá hodnot mezi 0 a 1. Čím blíže je 1, tím je těsnější závislost mezi X a Y, čím blíže je 0, tím je tato závislost volnější.

Cramérův koeficient

Význam hodnot Cramérova koeficientu:

mezi 0 až 0,1 … zanedbatelná závislost,
mezi 0,1 až 0,3 … slabá závislost,
mezi 0,3 až 0,7 … střední závislost,
mezi 0,7 až 1 … silná závislost.

Příklad

Příklad na výpočet Cramérova koeficientu:
686 náhodně vybraných osob bylo dotázáno, zda vlastní auto (znak X, varianty 1 – ano, 2 – ne) a zda jsou ochotny používat MHD (znak Y, varianty 1 – ano, 2 – ne). Výsledky průzkumu jsou uvedeny v kontingenční tabulce

X	Y		$n_{j.}$
	ano	ne	
ano	56	312	368
ne	283	35	318
$n_{.k}$	339	347	686

Vypočtěte a interpretujte Cramérův koeficient.

Řešení: Nejprve vypočteme teoretické četnosti:

$$\frac{n_{1.}n_{.1}}{n} = \frac{368 \cdot 339}{686} = 181{,}8542, \quad \frac{n_{1.}n_{.2}}{n} = \frac{368 \cdot 347}{686} = 186{,}1458,$$

$$\frac{n_{2.}n_{.1}}{n} = \frac{318 \cdot 339}{686} = 157{,}1458, \quad \frac{n_{2.}n_{.2}}{n} = \frac{318 \cdot 347}{686} = 160{,}8542$$

Nyní dosadíme do vzorce pro výpočet statistiky K:

$$K = \frac{(56 - 181{,}8542)^2}{181{,}8542} + \frac{(312 - 186{,}1458)^2}{186{,}1485} + \frac{(283 - 157{,}1458)^2}{157{,}1458} + \frac{(35 - 160{,}8542)^2}{160{,}8542} = 371{,}456$$

Nakonec vypočteme Cramérův koeficient:

$$V = \sqrt{\frac{371{,}456}{686 \cdot 1}} = 0{,}7358$$

Hodnota Cramérova koeficientu svědčí o tom, že mezi znaky X a Y existuje silná závislost.

Číselné charakteristiky ordinálních znaků

Charakteristika polohy: α-kvantil. Je-li $\alpha \in (0;1)$, pak α-kvantil x_α je číslo, které rozděluje uspořádaný datový soubor na dolní úsek, obsahující aspoň podíl α všech dat a na horní úsek obsahující aspoň podíl 1 – α všech dat. Pro výpočet α-kvantilu slouží algoritmus:

$$n\alpha = \begin{cases} \text{celé číslo } c \Rightarrow x_\alpha = \dfrac{x_{(c)} + x_{(c+1)}}{2} \\ \text{necelé číslo} \Rightarrow \text{zaokrouhlíme nahoru na nejbližší celé číslo } c \Rightarrow x_\alpha = x_{(c)} \end{cases}$$

Pro speciálně zvolená α užíváme názvů: $x_{0,50}$ – medián, $x_{0,25}$ – dolní kvartil, $x_{0,75}$ – horní kvartil, $x_{0,1}$, ..., $x_{0,9}$ – decily, $x_{0,01}$, ..., $x_{0,99}$ – percentily.

Charakteristika variability: kvartilová odchylka: $q = x_{0,75} - x_{0,25}$.

Příklad

Příklad na výpočet kvantilů:
U 50 žáků 7. ročníku jedné základní školy byly na pololetním vysvědčení zjištěny známky z matematiky:

známka	1	2	3	4	5
četnost známky	9	15	20	4	2

Určete medián, 1. a 9. decil a kvartilovou odchylku.

Řešení:
Pro snadnější výpočet tabulku doplníme ještě o absolutní kumulativní četnosti:

známka	1	2	3	4	5
n_j	9	15	20	4	2
N_j	9	24	44	48	50

Rozsah souboru n = 50

α	$n\alpha$	c	x_α
0,50	50.0,5=25	25	$\frac{x_{(25)}+x_{(26)}}{2}=\frac{3+3}{2}=3$
0,10	50.0,1 = 5	5	$\frac{x_{(5)}+x_{(6)}}{2}=\frac{1+1}{2}=1$
0,90	50.0,9 = 45	45	$\frac{x_{(45)}+x_{(46)}}{2}=\frac{4+4}{2}=4$
0,25	50.0,25 = 12,5	13	$x_{(13)} = 2$
0,75	50.0,75 = 37,5	38	$x_{(38)} = 3$

Kvartilová odchylka: q = 3 – 2 = 1.

Interpretace např. dolního kvartilu: V souboru žáků je aspoň čtvrtina takových, kteří mají z matematiky jedničku nebo dvojku (neboli v souboru 50 žáků jsou aspoň tři čtvrtiny takových, kteří mají z matematiky dvojku či horší známku).

Příklad

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4	4	5	6	8	8	12	12	13	14	14	14	18	19

$x_{0,25}$ $x_{0,5}$ $x_{0,75}$

	n_j	N_j	p_j	F_j
1	1	1	0,07	0,07
4	2	3	0,13	0,20
5	1	4	0,07	0,27
6	1	5	0,07	0,33
8	2	7	0,13	0,47
12	2	9	0,13	0,60
13	1	10	0,07	0,67
14	3	13	0,20	0,87
18	1	14	0,07	0,93
19	1	15	0,07	1,00
Součet	15	x	1,00	x

$$x_{0,25} = 5$$

$$x_{0,5} = \tilde{x} = 12$$

$$x_{0,75} = 14$$

$$\hat{x} = 14$$

$$q = x_{0,75} - x_{0,25} = 14 - 5 = 9$$

$X_{0,25}$ je tedy hodnota, u které F_j poprvé překročí 0,25.

!!! Pokud ale $F_j=\alpha$ pro nějaké $x_{[j]}$, $x_\alpha = \left(x_{[j]} + x_{[j+1]}\right)/2$

$x_{0,2}=(4+5)/2=4,5$

Modus a kvantily pro intervalově tříděná data

$$\hat{x} = d_m + \frac{n_m - n_{m-1}}{2n_m - n_{m-1} - n_{m+1}} \cdot h$$

d_m	je dolní mez modální třídy,
n_m, n_{m-1}, n_{m+1}	je četnost modální, předcházející a následující třídy,
h	je šířka třídy

$$x_P = d_P + \frac{P - F_{P-1}}{p_P} \cdot h$$

d_P	je dolní mez třídy obsahující příslušný *P*-kvantil,
p_P	je relativní četnost této třídy,
F_{P-1}	je kumulativní relativní četnost předcházející třídy,
h	je šířka třídy

Příklad

Určete modus a medián.

x_i	N_i
méně než 15>	22
(15;20>	34
(20;25>	72
(25;30>	102
(30;35>	127
více než 35	135

Příklad

x_i	N_i
méně než 15>	22
(15;20>	34
(20;25>	72
(25;30>	102
(30;35>	127
více než 35	135

x_i	n_i	p_i	N_i	F_i
12,5	22	0,16	22	0,16
17,5	12	0,09	34	0,25
22,5	38	0,28	72	0,53
27,5	30	0,22	102	0,76
32,5	25	0,19	127	0,94
37,5	8	0,06	135	1,00
Součet	135	1,00	x	x

$$\hat{x} = 20 + \frac{38-12}{2\cdot 38-12-30}\cdot 5$$

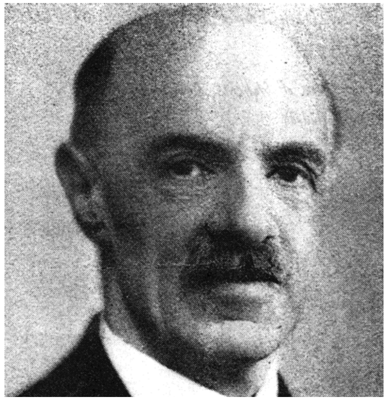
$$= 23,82$$

$$\tilde{x} = 20 + \frac{0,5-0,25}{0,28}\cdot 5$$

$$= 24,46$$

Spearmanův koeficient

Charakteristika těsnosti závislosti dvou ordinálních znaků: Spearmanův koeficient pořadové korelace



Charles Edward Spearman (1863 – 1945): Britský psycholog a statistik

Nejprve je nutné vysvětlit pojem pořadí čísla v posloupnosti čísel.
Nechť $x_1, \ldots, x_n$ je posloupnost reálných čísel.
a) Jsou-li čísla navzájem různá, pak pořadím R_i čísla x_i rozumíme počet těch čísel $x_1, \ldots, x_n$, která jsou menší nebo rovna číslu x_i.
b) Vyskytují-li se mezi danými čísly skupinky stejných čísel, pak každé takové skupince přiřadíme průměrné pořadí.

Příklad

Příklad na stanovení pořadí
a) Jsou dána čísla 9, 4, 5, 7, 3, 1.
b) Jsou dána čísla 6, 7, 7, 9, 6, 10, 8, 6, 6, 9.
Stanovte pořadí těchto čísel.

Řešení
ad a)

usp. čísla	1	3	4	5	7	9
pořadí	1	2	3	4	5	6

ad b)

usp. čísla	6	6	6	6	7	7	8	9	9	10
pořadí	1	2	3	4	5	6	7	8	9	10
prům. pořadí	2,5	2,5	2,5	2,5	5,5	5,5	7	8,5	8,5	10

Spearmanův koeficient

Vzorec pro výpočet Spearmanova koeficientu:

Předpokládejme, že máme dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \cdots & \cdots \\ x_n & y_n \end{pmatrix}$. Označíme R_i pořadí hodnoty x_i a Q_i pořadí hodnoty y_i, $i = 1, \ldots, n$.

Spearmanův koeficient pořadové korelace: $r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - Q_i)^2$.

Vlastnosti Spearmanova koeficientu pořadové korelace:

Koeficient nabývá hodnot mezi –1 a 1. Čím je bližší 1, tím je silnější přímá pořadová závislost mezi znaky X a Y, čím je bližší –1, tím je silnější nepřímá pořadová závislost mezi znaky X a Y.

Je-li $r_S = 1$ resp. $r_S = -1$, pak dvojice (x_i, y_i) leží na nějaké vzestupné resp. klesající funkci.

Hodnoty r_S se nezmění, když provedeme vzestupnou transformaci původních dat.

Hodnoty r_S se vynásobí -1, když provedeme sestupnou transformaci původních dat.

Koeficient je symetrický.

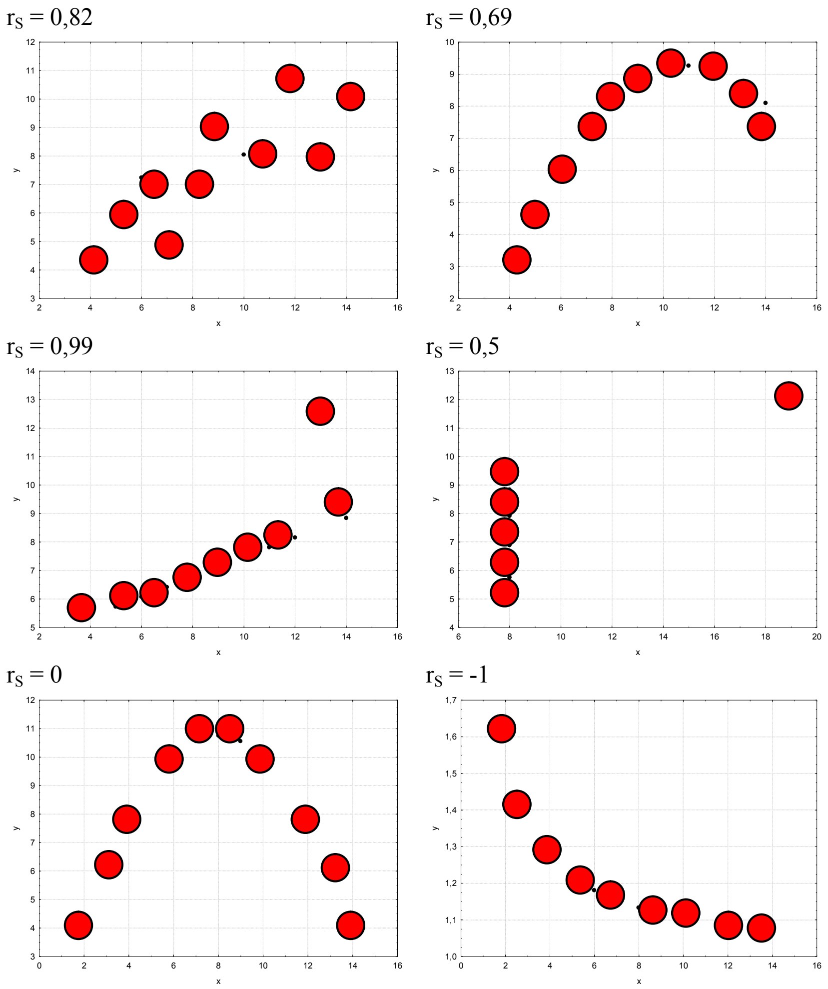
Koeficient je rezistentní vůči odlehlým hodnotám.

Spearmanův koeficient

Význam absolutní hodnoty Spearmanova koeficientu:

mezi 0 až 0,1 … zanedbatelná pořadová závislost,
mezi 0,1 až 0,3 … slabá pořadová závislost,
mezi 0,3 až 0,7 … střední pořadová závislost,
mezi 0,7 až 1 … silná pořadová závislost.

Ilustrace významu Spearmanova koeficientu pořadové korelace



Příklad

Příklad na výpočet Spearmanova koeficientu pořadové korelace:
Je dán dvourozměrný datový soubor

$$\begin{pmatrix} 2{,}5 & 13{,}4 \\ 3{,}4 & 15{,}2 \\ 1{,}3 & 11{,}8 \\ 5{,}8 & 13{,}1 \\ 3{,}6 & 14{,}5 \end{pmatrix}$$

Vypočtěte Spearmanův koeficient pořadové korelace.
Řešení:

x_i	2,5	3,4	1,3	5,8	3,6
y_i	13,4	15,2	11,8	13,1	14,5
R_i	2	3	1	5	4
Q_i	3	5	1	2	4
$(R_i-Q_i)^2$	1	4	0	9	0

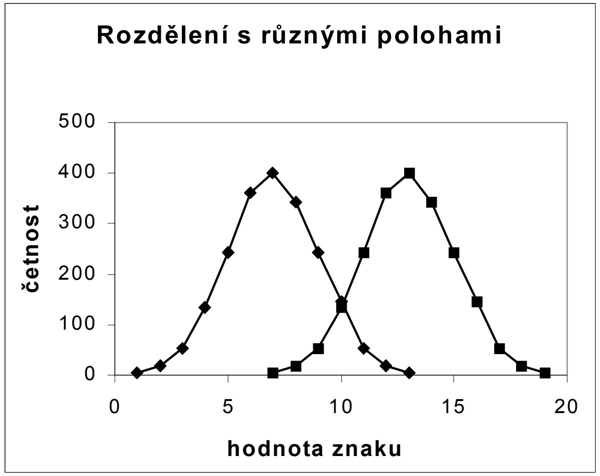
$$r_S = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2 = 1 - \frac{6}{5\cdot 24}(1+4+0+9+0) = 1 - \frac{6\cdot 14}{5\cdot 24} = 0{,}3$$

Znamená to, že mezi znaky X a Y existuje slabá přímá pořadová závislost.

Číselné charakteristiky intervalových znaků

Charakteristika polohy: aritmetický průměr je součet hodnot dělený jejich počtem $m = \frac{1}{n}\sum_{i=1}^{n} x_i$. Pomocí průměru zavedeme i-tou centrovanou hodnotu $x_i - m$ (podle znaménka poznáme, zda i-tá hodnota je podprůměrná či nadprůměrná).

Znázornění rozložení četností dvou datových souborů, které se liší aritmetickým průměrem



Často se aritmetický průměr označuje $\bar{x}$:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

Příklad

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y) vypočtěte aritmetické průměry znaků X, Y.

154	178	83	98	73	76
133	164	106	111	77	85
58	75	92	104	47	61
145	161	85	103	68	85
94	107	112	118	137	142
113	141	98	102	44	68
86	97	103	108	92	116
121	127	99	119	141	157
119	138	104	128	155	189
112	125	107	118	135	155
85	97	98	140	82	81
41	72	97	115	136	163
96	113	105	101	72	79
45	89	71	93	66	81
99	109	39	69	42	61
51	95	122	147	113	123
101	114	33	52	42	85
160	169	78	117	133	147
87	101	114	137	153	179
88	139	125	149	85	91

Řešení:

$$m_1 = \frac{154 + 133 + \ldots + 85}{60} = 95{,}9\,,\quad m_2 = \frac{178 + 164 + \ldots + 91}{60} = 114{,}4$$

Aritmetický průměr

Vlastnosti aritmetického průměru

- Aritmetický průměr si lze představit jako těžiště dat – součet podprůměrných hodnot je stejný jako součet nadprůměrných hodnot – oba součty jsou v rovnováze.

- Průměr centrovaných hodnot je nulový, protože $\frac{1}{n}\sum_{i=1}^{n}(x_i - m) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} m = m - \frac{1}{n} \cdot n \cdot m = 0 = 0.$

- Výraz $\sum_{i=1}^{n}(x_i - a)^2$ (tzv. kvadratická odchylka) nabývá svého minima pro a = m. Uvedený výraz charakterizuje celkovou chybu, které se dopustíme, když datový soubor nahradíme jedinou hodnotou a. Tato chyba je tedy nejmenší, když datový soubor nahradíme aritmetickým průměrem, přičemž za míru chyby považujeme kvadratickou odchylku.

- Aritmetický průměr je silně ovlivněn extrémními hodnotami.

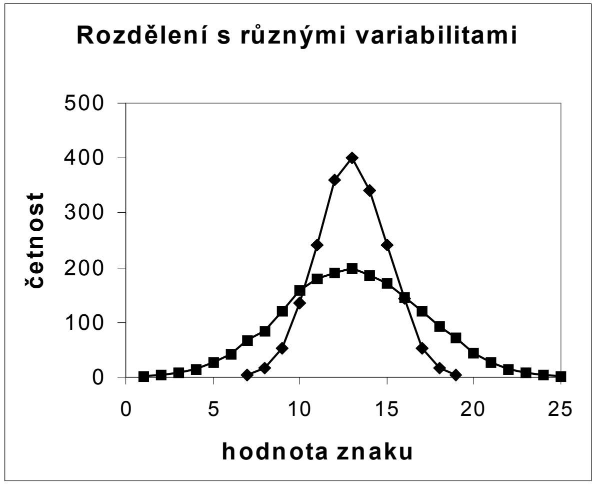
- Aritmetický průměr je vhodné použít, pokud je rozložení dat přibližně symetrické.

Rozptyl, směrodatná odchylka

Charakteristika variability: rozptyl je průměrná kvadratická odchylka hodnot od jejich aritmetického průměru $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$. Kladná odmocnina z rozptylu se nazývá směrodatná odchylka $s = \sqrt{s^2}$. Pomocí směrodatné odchylky zavedeme i-tou standardizovanou hodnotu $\frac{x_i - m}{s}$ (vyjadřuje, o kolik směrodatných odchylek se i-tá hodnota odchýlila od průměru).

Výpočetní tvar vzorce pro rozptyl: $s^2 = \left(\frac{1}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{x_i}^2\right) - m^2$

Znázornění rozložení četností dvou datových souborů, které se liší rozptylem:



Příklad

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y) vypočtěte rozptyly a směrodatné odchylky znaků X, Y. Přitom již víme, že $m_1 = 95{,}5$ a $m_2 = 114{,}4$.

154	178	83	98	73	76
133	164	106	111	77	85
58	75	92	104	47	61
145	161	85	103	68	85
94	107	112	118	137	142
113	141	98	102	44	68
86	97	103	108	92	116
121	127	99	119	141	157
119	138	104	128	155	189
112	125	107	118	136	155
85	97	98	140	82	81
41	72	97	115	136	163
96	113	105	101	72	79
45	89	71	93	66	81
99	109	39	69	42	61
51	95	122	147	113	123
101	114	33	52	42	85
160	169	78	117	133	147
87	101	114	137	153	179
88	139	125	149	85	91

Řešení:

$$s_1^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m_1^2 = \frac{1}{60}\left(154^2 + 133^2 + \ldots + 85^2\right) - 95{,}5^2 = 1052{,}40,\ s_1 = \sqrt{1052{,}40} = 32{,}4$$

$$s_2^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - m_2^2 = \frac{1}{60}\left(178^2 + 164^2 + \ldots + 91^2\right) - 114{,}4^2 = 1057{,}21,\ s_1 = \sqrt{1057{,}21} = 32{,}5$$

Rozptyl, směrodatná odchylka - vlastnosti

Vlastnosti rozptylu a směrodatné odchylky:

- Směrodatná odchylka je nulová pouze tehdy, když jsou všechny hodnoty stejné, jinak je kladná.

- Rozptyl centrovaných hodnot je roven původnímu rozptylu, neboť $\frac{1}{n}\sum_{i=1}^{n}[(x_i - m) - 0]^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = s^2$

- Rozptyl standardizovaných hodnot je 1, protože $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - m}{s} - 0\right)^2 = \frac{1}{s^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{s^2}{s^2} = 1$

- Rozptyl či směrodatná odchylka jsou stejně jako průměr silně ovlivněny extrémními hodnotami.

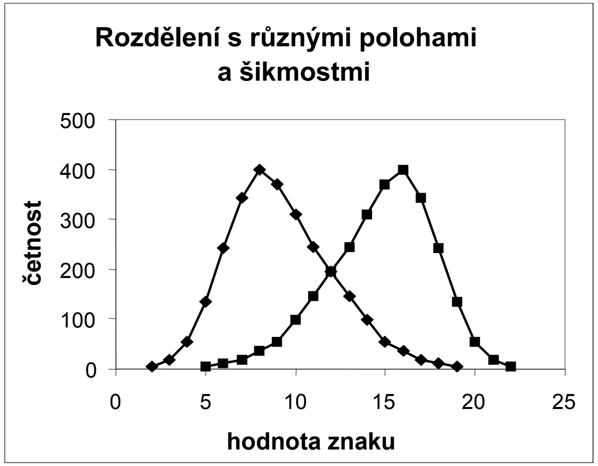
- Rozptyl či směrodatná odchylka se nehodí jako charakteristiky variability, je-li rozložení dat nesymetrické.

Šikmost

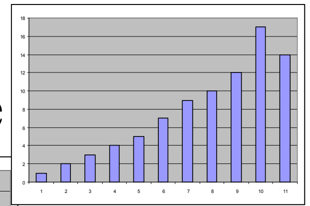
Charakteristika nesymetrie dat: **šikmost** $\alpha_3 = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^3}{s^3}$

Je-li rozložení dat symetrické kolem aritmetického průměru, pak $\alpha_3 = 0$.
Má-li rozložení dat prodloužený pravý konec, jde o kladně zešikmené rozložení $\alpha_3 > 0$.
Má-li rozložení dar prodloužený levý konec, jde o záporně zešikmené rozložení $\alpha_3 < 0$.

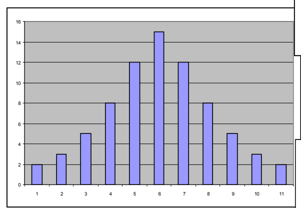
Znázornění rozložení četností dvou datových souborů, které se liší aritmetickým průměrem a šikmostí



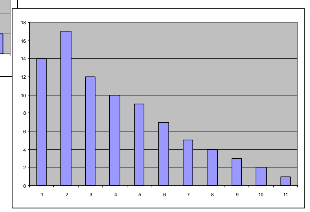
$\alpha_3 < 0$: Pravostranná asymetrie



$\alpha_3 = 0$: Symetrie



$\alpha_3 > 0$: Levostranná asymetrie

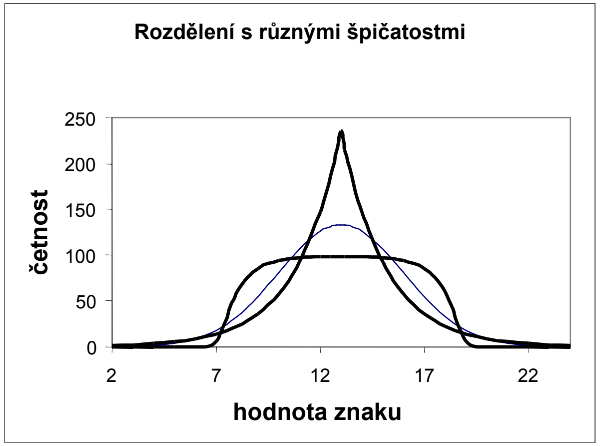


Špičatost

Charakteristika koncentrace dat kolem průměru: špičatost $\alpha_4 = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^4}{s^4} - 3$

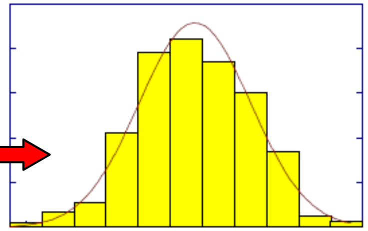
Je-li rozložení dat normální (Gaussovo), pak $\alpha_4 = 0$.
Je-li rozložení dat strmé, pak $\alpha_4 > 0$.
Je-li rozložení dat ploché, pak $\alpha_4 < 0$.

Znázornění rozložení četností dvou datových souborů, které se liší špičatostí

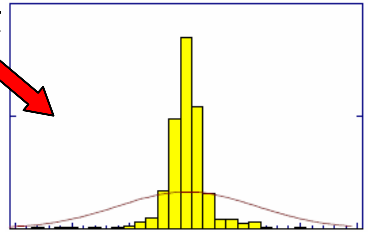


$\alpha_4 < 0$: **Podnormální špičatost**

$\alpha_4 = 0$: **Normální špičatost**



$\alpha_4 > 0$: **Nadnormální špičatost**



Kovariance

Charakteristika společné variability dvou intervalových znaků: kovariance

Předpokládejme, že máme dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \cdots & \cdots \\ x_n & y_n \end{pmatrix}$. Označme m_1, m_2 průměry znaků X, Y a s_1, s_2 směrodatné odchylky znaků X, Y. Zavedeme kovarianci jako charakteristiku společné variability znaků X, Y kolem jejich průměrů

$$s_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_1)(y_i - m_2).$$

Kovariance je průměrem součinů centrovaných hodnot.
Pokud se nadprůměrné (podprůměrné) hodnoty znaku X sdružují s nadprůměrnými (podprůměrnými) hodnotami znaku Y, budou součiny centrovaných hodnot $x_i - m_1$ a $y_i - m_2$ vesměs kladné a jejich průměr (tj. kovariance) rovněž. Znamená to, že mezi znaky X, Y existuje určitý stupeň přímé lineární závislosti. Říkáme, že znaky X, Y jsou kladně korelované.

Pokud se nadprůměrné (podprůměrné) hodnoty znaku X sdružují s podprůměrnými (nadprůměrnými) hodnotami znaku Y, budou součiny centrovaných hodnot vesměs záporné a jejich průměr rovněž. Znamená to, že mezi znaky X a Y existuje určitý stupeň nepřímé lineární závislosti. Říkáme, že znaky X, Y jsou záporně korelované.

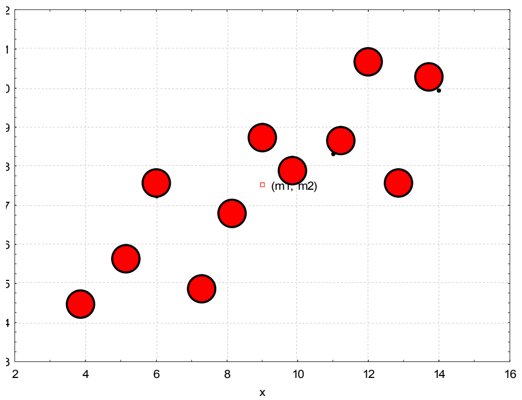
Je-li kovariance nulová, pak řekneme, že znaky X, Y jsou nekorelované a znamená to, že mezi nimi neexistuje žádná lineární závislost.

Pro výpočet kovariance používáme vzorec $s_{12} = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{y}_i\right) - \mathbf{m}_1\mathbf{m}_2$.

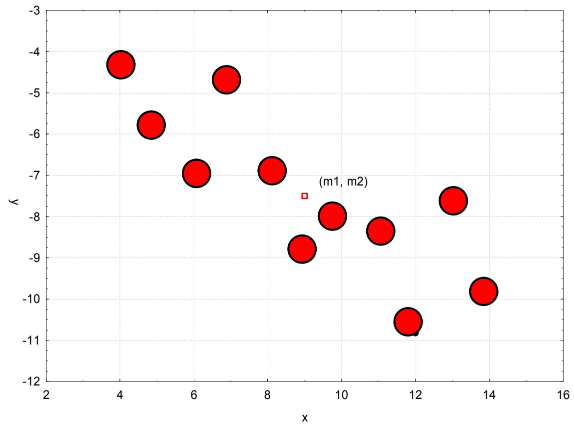
Kovariance

Znázornění významu kovariance

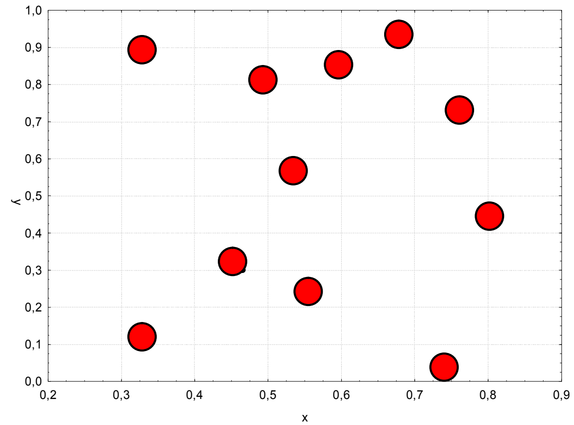
= 5,5



s_{12} = -5,5



s_{12} = 0



Příklad

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y) vypočtěte kovarianci znaků X, Y. Přitom již víme, že $m_1 = 95{,}5$, $m_2 = 114{,}4$, $s_1 = 32{,}4$, $s_2 = 32{,}5$

154	178	83	98	73	76
133	164	106	111	77	85
58	75	92	104	47	61
145	161	85	103	68	85
94	107	112	118	137	142
113	141	98	102	44	68
86	97	103	108	92	116
121	127	99	119	141	157
119	138	104	128	155	189
112	125	107	118	136	155
85	97	98	140	82	81
41	72	97	115	136	163
96	113	105	101	72	79
45	89	71	93	66	81
99	109	39	69	42	61
51	95	122	147	113	123
101	114	33	52	42	85
160	169	78	117	133	147
87	101	114	137	153	179
88	139	125	149	85	91

Řešení:

$$s_{12} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - m_1 m_2 = \frac{1}{60}\left(154\cdot 178 + 133\cdot 164 + \ldots + 85\cdot 91\right) - 95{,}5\cdot 114{,}4 = 985{,}76$$

Pearsonův koeficient korelace

Charakteristika těsnosti závislosti dvou intervalových znaků: Pearsonův koeficient korelace

Jsou-li směrodatné odchylky s_1, s_2 nenulové, pak definujeme Pearsonův koeficient korelace znaků X, Y vzorcem: $r_{12} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - m_1}{s_1}\frac{y_i - m_2}{s_2}$. Je to průměr součinů standardizovaných hodnot. Počítá se podle vzorce $r_{12} = \frac{s_{12}}{s_1 s_2}$.

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y) vypočtěte koeficient korelace znaků X, Y. Přitom již víme, že $m_1 = 95{,}5$, $m_2 = 114{,}4$, $s_1 = 32{,}4$, $s_2 = 32{,}5$, $s_{12} = 985{,}76$.

Řešení:

$$r_{12} = \frac{s_{12}}{s_1 s_2} = \frac{985{,}76}{32{,}4.32{,}5} = 0{,}936$$

Koeficient korelace svědčí o tom, že mezi oběma znaky existuje velmi silná přímá lineární závislost – čím je vyšší mez plasticity, tím je vyšší mez pevnosti a čím je nižší mez plasticity, tím je nižší mez pevnosti.

Vlastnosti Pearsonova koeficientu korelace:

Pro koeficient korelace platí $-1 \le r_{12} \le 1$ a rovnosti je dosaženo právě když mezi hodnotami $x_1, \ldots, x_n$ a $y_1, \ldots, y_n$ existuje úplná lineární závislost, tj. existují konstanty a, b tak, že $y_i = a + bx_i$, $i = 1, \ldots, n$, přičemž znaménko + platí pro $b > 0$, znaménko – pro $b < 0$. (Uvedená nerovnost se nazývá Cauchyova – Schwarzova – Buňakovského nerovnost.)

Tedy čím je r_{12} bližší 1, tím je silnější přímá lineární závislost mezi znaky X a Y, čím je bližší –1, tím je silnější nepřímá lineární závislost mezi X a Y.

Je-li $r_{12} = 1$ resp. $r_{12} = -1$, pak dvojice (x_i, y_i) leží na nějaké rostoucí resp. klesající přímce.

Hodnoty r_{12} se nezmění, když u x-ových a y-ových hodnot současně provedeme vzestupnou resp sestupnou lineární transformaci.

Hodnoty r_{12} se vynásobí -1, když u x-ových hodnot provedeme vzestupnou (resp. sestupnou) a u y-ových hodnot sestupnou (resp. vzestupnou) lineární transformaci.

Koeficient je symetrický, tj. $r_{12} = r_{21}$.

Početní pravidla pro číselné charakteristiky

Početní pravidla pro číselné charakteristiky

Nechť m_1 je aritmetický průměr a s_1^2 rozptyl znaku X. Pak znak **Y = a + bX** má:

aritmetický průměr $\boxed{m_2 = a + bm_1}$, rozptyl $\boxed{s_2^2 = b^2 s_1^2}$

Nechť m_1, m_2 jsou aritmetické průměry, s_1^2, s_2^2 rozptyly a s_{12} kovariance znaků X, Y. Pak znak **U = X + Y** má

aritmetický průměr $\boxed{m_3 = m_1 + m_2}$, rozptyl $\boxed{s_3^2 = s_1^2 + s_2^2 + 2s_{12}}$

Nechť s_{12} je kovariance znaků X, Y a m_1, m_2 jsou aritmetické průměry znaků X, Y. Pak znaky **U = a + bX, V = c + dY** mají kovarianci

$$\boxed{s_{34} = bds_{12}}$$

Příklad

Příklad:
a) Znak X má aritmetický průměr 2 a rozptyl 3. Najděte aritmetický průměr a rozptyl znaku Y = -1 + 3X.

b) Znaky X a Y mají aritmetické průměry 3 a 2, rozptyly 2 a 3, kovarianci 1,5. Vypočtěte aritmetický průměr a rozptyl znaku Z = 5X – 4Y.

c) Součet rozptylů dvou znaků je 120, součin 1000 a rozptyl jejich součtů je 100. Vypočtěte koeficient korelace těchto znaků.

Řešení:
ad a) $m_2 = -1 + 3m_1 = -1 + 3 \times 2 = 5$, $s_2^2 = 3^2 \times s_1^2 = 9 \times 3 = 27$.

ad b) $m_3 = 5m_1 - 4m_2 = 5 \times 3 - 4 \times 2 = 7$, $s_3^2 = 5^2 \times s_1^2 + (-4)^2 \times s_2^2 + 2 \times 5 \times (-4) \times s_{12} = 25 \times 2 + 16 \times 3 - 40 \times 1{,}5 = 38$.

ad c) $s_1^2 + s_2^2 = 120$, $s_1^2 \times s_2^2 = 1000$, $s_{1+2}^2 = 100 = s_1^2 + s_2^2 + 2s_{12} \Rightarrow s_{12} = \frac{1}{2}\left(s_{1+2}^2 - s_1^2 - s_2^2\right) = \frac{1}{2}(100 - 120) = -10$

$$r_{12} = \frac{s_{12}}{s_1 \times s_2} = \frac{-10}{\sqrt{1000}} = -0{,}3162\,.$$

Vážené číselné charakteristiky

Pokud nemáme k dispozici původní datový soubor, ale jenom tabulku rozložení četností (resp. kontingenční tabulku), můžeme vypočítat tzv. vážené číselné charakteristiky.

Vážený aritmetický průměr: $m = \frac{1}{n}\sum_{j=1}^{r} n_j x_{[j]}$

Vážený rozptyl: $s^2 = \frac{1}{n}\sum_{j=1}^{r} n_j \left(x_{[j]} - m\right)^2 = \frac{1}{n}\sum_{j=1}^{r} n_j x_{[j]}{}^2 - m^2$

Vážená kovariance: $s_{12} = \frac{1}{n}\sum_{j=1}^{r}\sum_{k=1}^{s} n_{jk}\left(x_{[j]} - m_1\right)\left(y_{[k]} - m_2\right) = \frac{1}{n}\sum_{j=1}^{r}\sum_{k=1}^{s} n_{jk} x_{[j]} y_{[k]} - m_1 m_2$

Příklad

Příklad na výpočet vážených číselných charakteristik

Z dvourozměrného datového souboru rozsahu 27, v němž znak X má varianty 1, 2, 3 a znak Y má rovněž varianty 1, 2, 3, byly určeny simultánní absolutní četnosti: $n_{11} = 5$, $n_{12} = 1$, $n_{13} = 3$, $n_{21} = 4$, $n_{22} = 3$, $n_{23} = 4$, $n_{31} = 2$, $n_{32} = 3$, $n_{33} = 2$.

a) Vypočtěte průměry a směrodatné odchylky znaků X a Y.

b) Vypočtěte a interpretujte koeficient korelace znaků X a Y.

Řešení:

Kontingenční tabulka simultánních absolutních četností:

x	y			$n_{j.}$
	1	2	3	
1	5	1	3	9
2	4	3	4	11
3	2	3	2	7
$n_{.k}$	11	7	9	27

ad a) $m_1 = \frac{1}{27}(1\cdot 9 + 2\cdot 11 + 3\cdot 7) = \frac{52}{27} = 1{,}926$, $m_2 = \frac{1}{27}(1\cdot 11 + 2\cdot 7 + 3\cdot 9) = \frac{52}{27} = 1{,}926$

$$s_1^2 = \frac{1}{27}(1^2\cdot 9 + 2^2 11 + 3^2\cdot 7) - \left(\frac{52}{27}\right)^2 = \frac{116}{27} - \frac{2704}{729} = \frac{428}{729},\ s_1 = 0{,}766$$

$$s_2^2 = \frac{1}{27}(1^2\cdot 11 + 2^2\cdot 7 + 3^2\cdot 9) - \left(\frac{52}{27}\right)^2 = \frac{120}{27} - \frac{2704}{729} = \frac{536}{729},\ s_2 = 0{,}857$$

ad b)

$$s_{12} = \frac{1}{27}(1\cdot 1\cdot 5 + 1\cdot 2\cdot 1 + 1\cdot 3\cdot 3 + 2\cdot 1\cdot 4 + 2\cdot 2\cdot 3 + 2\cdot 3\cdot 4 + 3\cdot 1\cdot 2 + 3\cdot 2\cdot 3 + 3\cdot 3\cdot 2) - \frac{52}{27}\cdot\frac{52}{27}$$

$$= \frac{102}{27} - \frac{2704}{729} = \frac{2754 - 2704}{729} = \frac{50}{729} = 0{,}0685871$$

$$r_{12} = \frac{\frac{50}{729}}{\sqrt{\frac{428}{729}\cdot\frac{536}{729}}} = 0{,}10439.$$

Mezi znaky X a Y existuje velmi slabá přímá lineární závislost.

Koeficient variace, geometrický průměr

Pro poměrové znaky používáme jako charakteristiku variability koeficient variace $\frac{s}{m}$. Je to bezrozměrné číslo, které se často vyjadřuje v procentech. Umožňuje porovnat variabilitu několika znaků.

Jsou-li všechny hodnoty poměrového znaku kladné, pak jako charakteristiku polohy lze užít geometrický průměr $\sqrt[n]{x_1 \cdot \ldots \cdot x_n}$.

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y) vypočtěte koeficienty variace znaků X, Y. Přitom již víme, že $m_1 = 95{,}5$, $m_2 = 114{,}4$, $s_1 = 32{,}4$, $s_2 = 32{,}5$

Řešení:

$$cv_1 = \frac{s_1}{m_1} = \frac{32{,}4}{95{,}5} = 0{,}339, cv_2 = \frac{s_2}{m_2} = \frac{32{,}5}{114{,}4} = 0{,}284$$

Výpočty zavedením pomocné proměnné

- pomocná proměnná ⇒ $$v_i = \frac{x_i - a}{h}$$

- konstanty:
  - $a \rightarrow$ střed třídy s nejvyšší četností
  - $h \rightarrow$ šířka třídy

Výpočty zavedením pomocné proměnné

$$\overline{v} = \frac{\overline{x} - a}{h} \Rightarrow \overline{x} = \overline{v}h + a$$

$$s_v^2 = \frac{s_x^2}{h^2} \Rightarrow s_x^2 = h^2 s_v^2$$

Příklad

x_i	n_i
<30 – 40)	10
<40 – 50)	31
<50 – 60)	27
<60 – 70)	19
<70 – 80)	13
Celkem	**100**

<u>Vypočítejte:</u>

- aritmetický průměr, rozptyl, směrodatnou odchylku a variační koeficient zavedením pomocné proměnné

Příklad

$a = 45$ $h = 10$

x_i	n_i	v_i
35	10	-1
45	31	0
55	27	1
65	19	2
75	13	3
Součet	**100**	**x**

Příklad

x_i	n_i	v_i	$v_i n_i$
35	10	-1	-10
45	31	0	0
55	27	1	27
65	19	2	38
75	13	3	39
Součet	**100**	**x**	**94**

$$\overline{v} = 0{,}94 \Rightarrow \overline{x} = \overline{v}h + a =$$

$$= 0{,}94 \cdot 10 + 45 = 54{,}4$$

Příklad

x_i	n_i	v_i	$v_i^2 n_i$
35	10	-1	10
45	31	0	0
55	27	1	27
65	19	2	76
75	13	3	117
Součet	**100**	**x**	**230**

$$s_v^2 = \frac{1}{n}\sum_{i=1}^{n}(v_i^2 n_i) - \overline{v}^2 = 2{,}3 - 0{,}8836 = 1{,}4164$$

$$\Rightarrow \quad s_x^2 = h^2 \cdot s_v^2 = 10^2 \cdot 1{,}4164 = 141{,}64$$

Příklad

$$s_x = \sqrt{s_x^2} = \sqrt{141{,}64} = 11{,}9$$

nebo $$s_x = h \cdot s_v^2 = 10 \cdot \sqrt{1{,}4164} = 11{,}9$$

$$cv_x = \frac{s_x}{\bar{x}} = \frac{11{,}9}{55{,}4} = 0{,}2188$$

nebo $$cv_x = \frac{h \cdot s_v}{\bar{v} \cdot h + a} = \frac{10 \cdot 1{,}19}{0{,}94 \cdot 10 + 45} = 0{,}2188$$

Společný rozptyl

$$s^2 = \overline{s}^2 + s_{\overline{x}}^2$$

$\overline{s}^2$ **……vnitroskupinová variabilita** ($s_{\#}^2$)[1]

$s_{\overline{x}}^2$ **……meziskupinová variabilita** (s_{*}^2)

[1]) Značení ze skript „Popisná statistika“

Společný rozptyl

- **vnitroskupinová variabilita**

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k} s_i^2 \cdot n_i$$

- **meziskupinová variabilita**

$$s_{\overline{x}}^2 = \frac{1}{n}\sum_{i=1}^{k} (\overline{x}_i - \overline{x})^2 \cdot n_i$$

Příklad

D1:	**104**	**108**	**79**	**155**
D2:	**93**	**65**	**76**	**111**

<u>Vypočítejte:</u>

- **dílčí průměry,**
- **společný průměr,**
- **dílčí rozptyly,**
- **společný rozptyl.**

Příklad

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i^1 =$$

$$= \frac{1}{4}\cdot(104+108+79+155) = \underline{\underline{111{,}5}}$$

$$\bar{x}_2 = \frac{1}{4}\cdot(93+65+76+111) = \underline{\underline{86{,}25}}$$

$$\bar{x} = \frac{1}{8}\cdot(111{,}5\cdot 4+86{,}25\cdot 4) = \underline{\underline{98{,}875}}$$

Příklad

$$s_{x1}^2 = \frac{1}{n}\sum_{i=}^{n_1}\left(x_i^1 - \bar{x}_1\right)^2 =$$

$$= \frac{7{,}5^2 + 3{,}5^2 + 32{,}5^2 + 43{,}5^2}{4} = \underline{\underline{754{,}25}}$$

$$s_{x2}^2 = \frac{6{,}75^2 + 21{,}25^2 + 10{,}25^2 + 24{,}75^2}{4} = \underline{\underline{303{,}69}}$$

Příklad

$$\bar{s}^2 = \frac{1}{n}\sum_{i=1}^{k} s_i^2 \cdot n_i = \frac{1}{8}\left(s_{x1}^2 \cdot 4 + s_{x2}^2 \cdot 4\right) =$$

$$= \frac{1}{8} \cdot (754{,}25 \cdot 4 + 303{,}69 \cdot 4) = \underline{\underline{528{,}97}}$$

$$s_{\bar{x}}^2 = \frac{1}{n}\sum_{i=1}^{k} \left(\bar{x}_i - \bar{x}\right)^2 \cdot n_i =$$

$$= \frac{(111{,}5 - 98{,}875)^2 \cdot 4) + (86{,}25 - 98{,}875)^2 \cdot 4}{8} = \underline{\underline{159{,}39}}$$

Příklad

$$s^2 = \bar{s}^2 + s_{\bar{x}}^2 =$$

$$= 528{,}97 + 159{,}39 = \underline{\underline{688{,}36}}$$

Pro kontrolu ještě spočteme rozptyl přímo:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{8} \cdot 83717 - 98{,}875^2$$

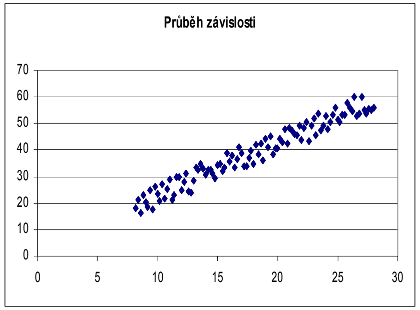
$$= 10464{,}63 - 9776{,}27 = \underline{\underline{688{,}36}}$$

3. Regresní analýza, Počet pravděpodobnosti

Cíl regresní analýzy: vystižení závislosti hodnot znaku Y na hodnotách znaku X. Při tom je nutné vyřešit dva problémy:

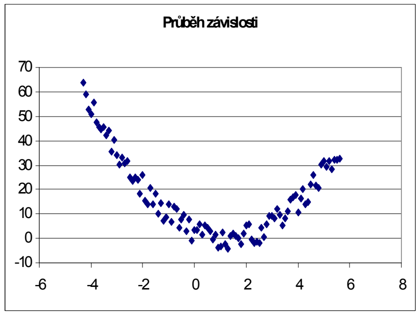
- jaký typ funkce použít k vystižení dané závislosti
- jak stanovit konkrétní parametry zvoleného typu funkce?

Typ funkce určíme buď logickým rozborem zkoumané závislosti nebo se snažíme ho odhadnout pomocí dvourozměrného tečkového diagramu.



přímka

$$y = \beta_0 + \beta_1 x$$



parabola

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Regresní přímka

Zde se omezíme na lineární závislost $y = \beta_0 + \beta_1 x$.

Odhady b_0 a b_1 neznámých regresních parametrů β_0, β_1 získáme na základě datového souboru $\begin{pmatrix} x_1 & y_1 \\ \dots & \dots \\ x_n & y_n \end{pmatrix}$ metodou nejmenších čtverců. Požadujeme, aby výraz $\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ nabýval svého minima vzhledem k β_0 a β_1. Tento výraz je minimální, jsou-li jeho první derivace podle β_0 a β_1 nulové. Stačí tyto derivace spočítat, položit je rovny 0 a řešit systém dvou rovnic o dvou neznámých, tzv. systém normálních rovnic.

Nechť je dán dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \dots & \dots \\ x_n & y_n \end{pmatrix}$ a přímka $y = \beta_0 + \beta_1 x$.

Výraz $q(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ se nazývá rozptyl hodnot znaku Y kolem přímky $y = \beta_0 + \beta_1 x$,

přímka $y = b_0 + b_1 x$, jejíž parametry minimalizují rozptyl $q(\beta_0, \beta_1)$ v celém dvourozměrném prostoru, se nazývá regresní přímka znaku Y na znak X,

$\hat{y}_i = b_0 + b_1 x_i$, $i = 1, \dots, n$ … regresní odhad i-té hodnoty znaku Y,

$r_{12}^2 = ID^2$ … index determinace (Index determinace udává, jakou část variability hodnot znaku Y vystihuje regresní přímka. Nabývá hodnot z intervalu $\langle 0, 1 \rangle$. Čím je bližší 1, tím lépe vystihuje regresní přímka závislost Y na X.)

Odvození odhadů regresních parametrů

Systém normálních rovnic získáme derivováním výrazu

$q(\beta_0,\beta_1)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2$ parciálně podle β_0 a β_1:

$$\frac{\partial q(\beta_0,\beta_1)}{\partial\beta_0}=\frac{2}{n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)(-1)=0$$

$$\frac{\partial q(\beta_0,\beta_1)}{\partial\beta_1}=\frac{2}{n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)(-x_i)=0$$

Systém normálních rovnic:

$$\sum y_i = nb_0 + b_1\sum x_i$$
$$\sum y_i x_i = b_0\sum x_i + b_1\sum x_i^2$$

Řešením tohoto systému získáme odhady

$$b_0=\frac{\sum_{i=1}^{n}x_i^{\,2}\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i y_i}{n\sum_{i=1}^{n}x_i^{\,2}-\left(\sum_{i=1}^{n}x_i\right)^2},\quad b_1=\frac{n\sum_{i=1}^{n}x_i y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^{\,2}-\left(\sum_{i=1}^{n}x_i\right)^2}$$

$$\boxed{y=m_2+\frac{s_{12}}{s_1^2}(x-m_1)}$$

Po jednoduchých úpravách dospějeme ke tvaru $b_1=\frac{s_{12}}{s_1^{\,2}}$, kde s_{12} je kovariance znaků X, Y a $s_1^{\,2}$ je rozptyl znaku X. Dále dostáváme $b_0=m_2-b_1 m_1$, tedy regresní přímku můžeme vyjádřit ve tvaru $y=m_2+\frac{s_{12}}{s_1^{\,2}}(x-m_1)$.

Úsek b_0 regresní přímky udává, jaký je regresní odhad hodnoty znaku Y, nabývá-li znak X hodnoty 0.
Směrnice b_1 udává, o kolik jednotek se změní hodnota znaku Y, změní-li se hodnota znaku X o jednotku. Je-li $b_1 > 0$, dochází s růstem X k růstu Y a hovoříme o přímé závislosti hodnot znaku Y na hodnotách znaku X. Je-li $b_1 < 0$, dochází s růstem X k poklesu Y a hovoříme o nepřímé závislosti hodnot znaku Y na hodnotách znaku X.

Příklad

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y)

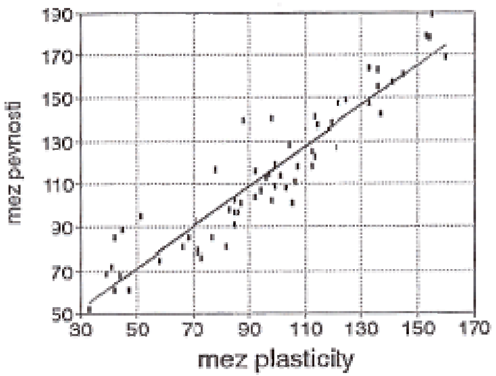
a) Určete regresní přímku meze pevnosti na mez plasticity.
b) Zakreslete regresní přímku do dvourozměrného tečkového diagramu.
c) Jak se změní mez pevnosti, vzroste-li mez plasticity o jednotku?
d) Najděte regresní odhad meze pevnosti pro mez plasticity = 60.
e) Vypočtěte index determinace a interpretujte ho.

Přitom již víme, že $m_1 = 95{,}5$, $m_2 = 114{,}4$, $s_1 = 32{,}4$, $s_2 = 32{,}5$, $s_{12} = 985{,}76$, $r_{12} = 0{,}936$.

Řešení:

ad a) $b_1 = \frac{s_{12}}{s_1^2} = \frac{985{,}76}{1052{,}4} = 0{,}937$, $b_0 = m_2 - b_1 m_1 = 114{,}4 - 0{,}937 \,.\, 95{,}9 = 24{,}5$, $y = 24{,}5 + 0{,}937x$.

ad b)



ad c) Mez pevnosti vzroste o 0,937 kpcm^{-2} – viz parametr b_1 vypočtený v bodě (a)

ad d) $\hat{y} = 24{,}5 + 0{,}937 \times 60 = 80{,}72$.

ad e) $ID^2 = r_{12}^2 = 0{,}936^2 = 0{,}876$. Znamená to, že 87,6% variability hodnot meze pevnosti je vysvětleno regresní přímkou.

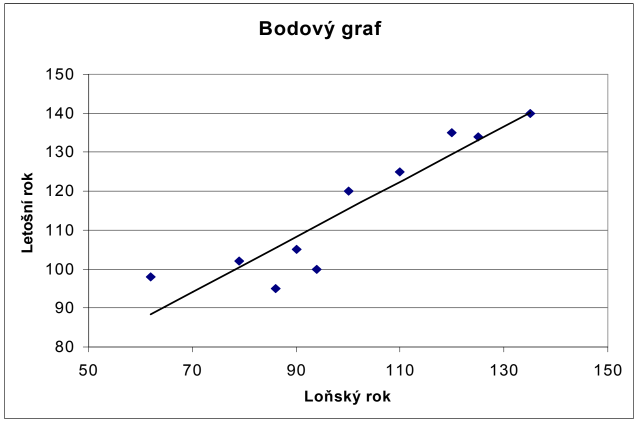
Příklad

x_i	y_i	x_i*y_i	x_i^2
100	120	12 000	10 000
90	105	9 450	8 100
86	95	8 170	7 396
94	100	9 400	8 836
120	135	16 200	14 400
135	140	18 900	18 225
79	102	8 058	6 241
62	98	6 076	3 844
110	125	13 750	12 100
125	134	16 750	15 625
1 001	**1 154**	**118 754**	**104 767**

$$\sum y_i = nb_0 + b_1 \sum x_i$$
$$\sum y_i x_i = b_0 \sum x_i + b_1 \sum x_i^2$$

$$1154 = 10 \cdot b_0 + 1001 \cdot b_1$$
$$118754 = 1001 \cdot b_0 + 104767 \cdot b_1$$

$$y = 44{,}41 + 0{,}709 \cdot x$$



Příklad

Index determinace lze vyjádřit ve tvaru:

$$ID^2 = \frac{\sum \hat{y}_i^2 - \frac{1}{n} \cdot \left(\sum y_i\right)^2}{\sum y_i^2 - \frac{1}{n} \cdot \left(\sum y_i\right)^2}$$

x_i	y_i	$\hat{y}_i$	y_i^2	$\hat{y}_i^2$
100	120	115	14 400	13 301
90	105	108	11 025	11 715
86	95	105	9 025	11 109
94	100	111	10 000	12 337
120	135	130	18 225	16 773
135	140	140	19 600	19 642
79	102	100	10 404	10 088
62	98	88	9 604	7 811
110	125	122	15 625	14 987
125	134	133	17 956	17 704
1 001	**1 154**	**1 154**	**135 864**	**135 468**

$$ID^2 = \frac{135468 - \frac{1}{10} \cdot 1154^2}{135864 - \frac{1}{10} \cdot 1154^2} = 0{,}853$$

Maticové vyjádření MNČ

$$\boxed{b = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot y}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

b	sloupcový vektor 2 neznámých parametrů regresní funkce,
X	matice rozměru $n \times 2$, tvořená konstantou 1 a hodnotami znaku X
y	sloupcový vektor n hodnot znaku Y

Příklad

Nalezněte koeficienty regresní přímky:

$$\mathbf{y} = \begin{bmatrix} 120 \\ 105 \\ 95 \\ 100 \\ 135 \\ 140 \\ 102 \\ 98 \\ 125 \\ 134 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 100 \\ 1 & 90 \\ 1 & 86 \\ 1 & 94 \\ 1 & 120 \\ 1 & 135 \\ 1 & 79 \\ 1 & 62 \\ 1 & 110 \\ 1 & 125 \end{bmatrix}$$

$$\mathbf{X}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 90 & 86 & 94 & 120 & 135 & 79 & 62 & 110 & 125 \end{bmatrix}$$

Příklad

$$g = X^T * y = \begin{bmatrix} 1154 \\ 118754 \end{bmatrix} \qquad A = X^T * X = \begin{bmatrix} 10 & 1001 \\ 1001 & 104767 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 2{,}2941 & -0{,}0219 \\ -0{,}0219 & 0{,}0002 \end{bmatrix}$$

$$b = A^{-1} * g = \begin{bmatrix} 44{,}414 \\ 0{,}709 \end{bmatrix}$$

Sdružené regresní přímky

V některých situacích má smysl zkoumat nejenom závislost znaku Y na znaku X, ale též závislost X na Y. V takovém případě hledáme druhou regresní přímku a souhrnně hovoříme o sdružených regresních přímkách.
Regresní přímkou znaku X na znak Y nazveme tu přímku x = $\bar{b}_0 + \bar{b}_1 y$, jejíž parametry minimalizují rozptyl q($\bar{\beta}_0, \bar{\beta}_1$) = $\sum_{i=1}^{n} (x_i - \bar{\beta}_0 - \bar{\beta}_1 y_i)^2$ v celé rovině. Nazývá se též druhá regresní přímka. Regresní přímka znaku Y na znak X a regresní přímka znaku X na znak Y se nazývají sdružené regresní přímky.

Rovnice regresní přímky znaku X na znak Y má tvar:

$$\boxed{x = m_1 + \frac{s_{12}}{s_2^2}\left(x - m_2\right)}$$

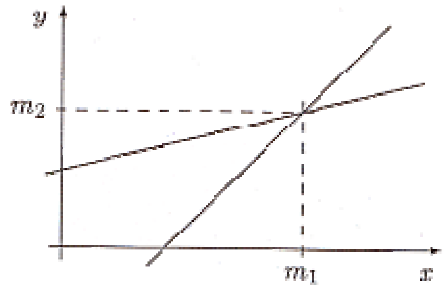
Vlastnosti sdružených regresních přímek

Sdružené regresní přímky se protínají v bodě (m_1,m_2).
Pro regresní parametry $b_1, \bar{b}_1$ platí: $b_1\bar{b}_1 = r_{12}^2$.
Rovnice sdružených regresních přímek můžeme psát ve tvaru

$y = m_2 + r_{12}\frac{s_2}{s_1}(x - m_1)$, $y = m_2 + \frac{1}{r_{12}}\frac{s_2}{s_1}(x - m_1)$ (je-li $r_{12} \neq 0$).

Regresní přímky svírají tím menší úhel, čím méně se od sebe liší r_{12} a $\frac{1}{r_{12}}$.

Regresní přímky splynou, je-li $r_{12}^2 = 1$. K tomu dojde právě tehdy, existuje-li mezi X a Y úplná lineární závislost. Všechny body (x_i, y_i), $i = 1, ..., n$ leží na jedné přímce, tedy ze znalosti x_i můžeme přesně vypočítat y_i, $i = 1, ..., n$.
Jsou-li znaky X, Y nekorelované, pak mají sdružené regresní přímky rovnice $y = m_2$, $x = m_1$ a jsou na sebe kolmé.
Označíme-li α úhel, který svírají sdružené regresní přímky, pak platí:
$\cos \alpha = 0$, právě když mezi X a Y neexistuje žádná lineární závislost,
$\cos \alpha = 1$, právě když mezi X a Y existuje úplná přímá lineární závislost,
$\cos \alpha = -1$, právě když mezi X a Y existuje úplná nepřímá lineární závislost.



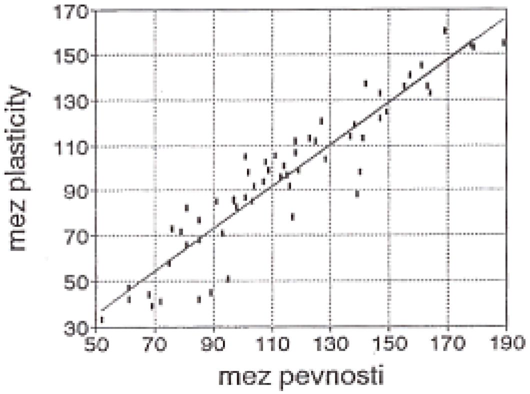
Příklad

Příklad: Pro datový soubor obsahující údaje o mezi plasticity (znak X) a mezi pevnosti oceli (znak Y):
a) Určete regresní přímku meze plasticity na mez pevnosti.
b) Zakreslete tuto druhou regresní přímku do dvourozměrného tečkového diagramu.
Přitom již víme, že $m_1 = 95{,}5$, $m_2 = 114{,}4$, $s_1 = 32{,}4$, $s_2 = 32{,}5$, $s_{12} = 985{,}76$, $r_{12} = 0{,}936$.

Řešení:

ad a) $\overline{b}_1 = \frac{s_{12}}{s_2^{\,2}} = \frac{985{,}76}{1057{,}21} = 0{,}932, \overline{b}_0 = m_1 - \overline{b}_1 m_2 = 95{,}9 - 0{,}932 \times 114{,}4 = -10{,}7$, tedy x = -10,7 + 0,932y.

ad b)



Příklad

Poptávka po vepřovém mase	154	164	123	181	193	105	143	167	158	62
Poptávka po hovězím mase	103	116	98	175	165	90	103	140	113	49

- Sestrojte sdružené regresní přímky.
- Vypočtěte koeficient korelace.

Příklad

x_i	y_i	$x_i{*}y_i$	x_i^2	y_i^2
154	103	15 862	23 716	10 609
164	116	19 024	26 896	13 456
123	98	12 054	15 129	9 604
181	175	31 675	32 761	30 625
193	165	31 845	37 249	27 225
105	90	9 450	11 025	8 100
143	103	14 729	20 449	10 609
167	140	23 380	27 889	19 600
158	113	17 854	24 964	12 769
62	49	3 038	3 844	2 401
1 450	**1 152**	**178 911**	**223 922**	**144 998**

$$m_1 = \frac{1}{10} \cdot 1450 = 145 \qquad m_2 = \frac{1}{10} \cdot 1152 = 115{,}2$$

$$s_1^2 = \frac{1}{10} \cdot 223922 - 145^2 = 1367{,}2$$

$$s_2^2 = \frac{1}{10} \cdot 144998 - 115{,}2^2 = 1228{,}76$$

$$s_{12} = \frac{178911}{10} - 145 \cdot 115{,}2 = 1187{,}1$$

$$b_1 = \frac{1187{,}1}{1367{,}2} = 0{,}868$$

$$\overline{b}_1 = \frac{1187{,}1}{1228{,}76} = 0{,}966$$

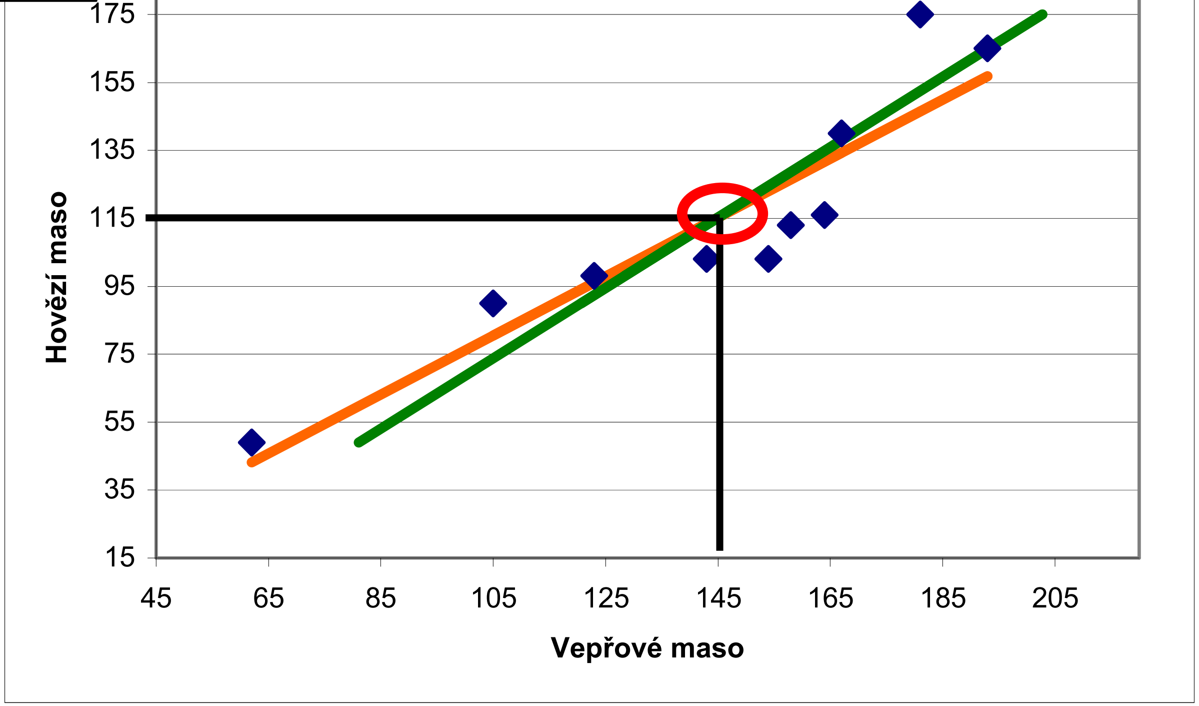
$$b_0 = 115{,}2 - 0{,}868 \cdot 145 = -10{,}66$$

$$\overline{b}_0 = 145 - 0{,}966 \cdot 115{,}2 = 33{,}72$$

Příklad

$$y = -10{,}66 + 0{,}868 \cdot x$$

$$x = 33{,}72 + 0{,}966 \cdot y$$



Příklad

$$r_{12} = \frac{\sum x_i y_i - n m_1 m_2}{\sqrt{\left[\sum x_i^2 - n m_1^2\right] \cdot \left[\sum y_i^2 - n m_2^2\right]}}$$

$$r_{12} = \frac{s_{12}}{s_1 \cdot s_2}$$

$$r_{12} = \mathrm{sgn}(\mathrm{b}_1)\sqrt{b_1 \cdot \overline{b}_1}$$

Příklad

$$r_{12} = \frac{178911 - 10 \cdot 145 \cdot 115{,}2}{\sqrt{\left[223922 - 10 \cdot 145^2\right] \cdot \left[144998 - 10 \cdot 115{,}2^2\right]}} = 0{,}916$$

$$r_{12} = \frac{1187{,}1}{36{,}976 \cdot 35{,}054} = 0{,}916$$

$$r_{12} = \sqrt{0{,}868 \cdot 0{,}966} = 0{,}916$$

Příklad

x_i	y_i
154	103
164	116
123	98
52	175
193	165
105	90
143	103
167	140
158	113
191	49

- Sestrojte sdružené regresní přímky.
- Vypočtěte koeficient korelace.
- Porovnejte výsledky s výsledky předchozího příkladu.

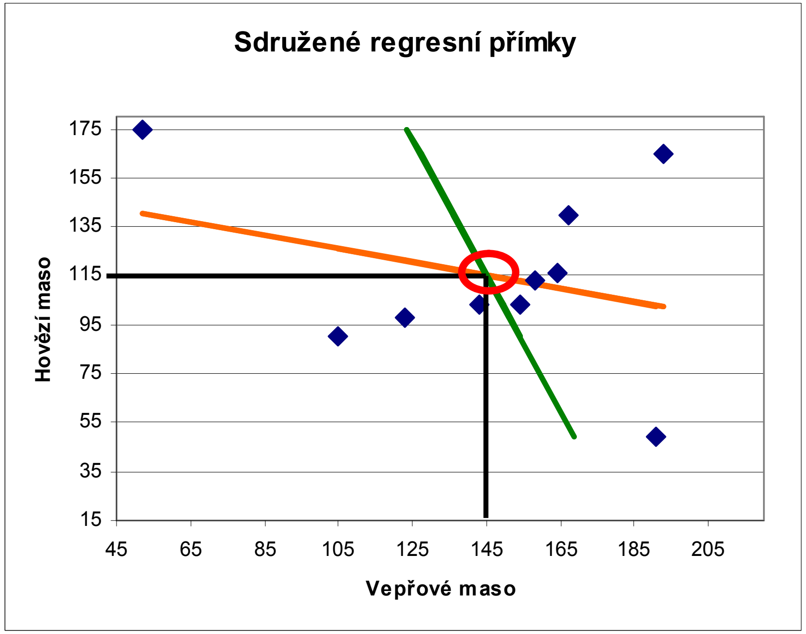
Příklad

x_i	y_i	$x_i{*}y_i$	x_i^2	y_i^2
154	103	15 862	23 716	10 609
164	116	19 024	26 896	13 456
123	98	12 054	15 129	9 604
52	175	9 100	2 704	30 625
193	165	31 845	37 249	27 225
105	90	9 450	11 025	8 100
143	103	14 729	20 449	10 609
167	140	23 380	27 889	19 600
158	113	17 854	24 964	12 769
191	49	9 359	36 481	2 401
1 450	**1 152**	**162 657**	**226 502**	**144 998**

průměry	145,0
	115,2
rozptyly	1625,2
	1228,76
směrodatné odchylky	40,3138
	35,0537

$$y = 154{,}31 - 0{,}270 \cdot x$$

$$x = 186{,}09 - 0{,}357 \cdot y$$



$$r = -\sqrt{0{,}270 \cdot 0{,}357} = -0{,}310$$

Příklad

Rozhodněte zda následující dvojice přímek mohou být sdruženými regresními přímkami:

$$A)\ y = 13 - 2x \qquad x = 2{,}5$$

$$B)\ y = 13 - 2x \qquad x = 0{,}4y$$

$$C)\ y = 13 - 2x \qquad x = 8 - y$$

$$D)\ y = 13 - 2x \qquad x = 6{,}5 - 0{,}5y$$

$$E)\ y = 13 - 2x \qquad x = -2 - 0{,}4y$$

$$F)\ y = 13 - 2x \qquad x = -0{,}5y$$

Příklad

$$A)\ y = 13 - 2x \qquad B)\ y = 13 - 2x \qquad C)\ y = 13 - 2x$$
$$x = 2{,}5 \qquad x = 0{,}4y \qquad x = 8 - y$$
$$D)\ y = 13 - 2x \qquad E)\ y = 13 - 2x \qquad F)\ y = 13 - 2x$$
$$x = 6{,}5 - 0{,}5y \qquad x = -2 - 0{,}4y \qquad x = -0{,}5y$$

1. b_1 i $\bar{b}_1$ mají stejná znaménka
2. je-li jeden roven nule, pak je 0-vý i druhý
3. $r \in [-1, 1]$,tj. $b_1 \cdot \bar{b}_1 \in [0, 1]$
4. pro r=1/r=-1 platí $\bar{b}_0 = -\dfrac{b_0}{b_1}$

A) NE(2)
B) NE(1)
C) NE(3)
D) ANO
E) ANO
F) NE(4)

Počet pravděpodobnosti -úvod

Počet pravděpodobnosti se zabývá studiem zákonitostí v náhodných pokusech. Matematickými prostředky modeluje situace, v nichž hraje roli náhoda. Pod pojmem náhoda rozumíme působení faktorů, které se živelně mění při různých provedeních téhož pokusu a nepodléhají naší kontrole.

Počet pravděpodobnosti jako vědecká disciplína se začal vytvářet **v 17. století** a jeho počátky jsou spjaty se jmény Blaise Pascala, Pierra de Fermata, Christiana Huygense (studovali hazardní hry, zformulovali takové pojmy, jako je pravděpodobnost a střední hodnota, odvodili jejich vlastnosti) a především Jakoba Bernoulliho (dokázal zákon velkých čísel).

V 18. století: Abraham de Moivre a Pierre Simeon Laplace – formulace jedné z forem centrální limitní věty, Georges Buffon odvodil binomickou větu, zavedl diferenciální a integrální počet do teorie pravděpodobnosti, Thomas Bayes odvodil způsob výpočtu aposteriorních pravděpodobností pomocí apriorních pravděpodobností (Bayesův vzorec).

V 19. století: Petrohradská matematická škola – dala teorii pravděpodobnosti pevný logický a matematický základ (Viktor Jakovlevič Buňakovskij, Pafnutij Lvovič Čebyšev, Andrej Andrejevič Markov, Alexandr Michailovič Ljapunov), Karl Fridirich Gauss (mj. vyvinul netodu zpracování experimentálních údajů známou pod názvem metoda nejmenších čtverců), Siméon Denis Poisson (zobecnil Bernoulliho zákon velkých čísel a odvodil speciální zákon rozložení pravděpodobností - Poissonův zákon rozložení).

Ve 20. století: Andrej Nikolajevič Kolmogorov (axiomatická teorie pravděpodobnosti), Norbert Wiener , William Feller (rozvoj reorie stochastických procesů).

Odkaz na zajímavou webovou stránku:

http://www-groups.dcs.st-and.ac.uk/~history

http://turnbull.mcs.st-and.ac.uk/~history/HistTopics/Statistics.html

Základní prostor

Definice (definice pokusu): Pokusem rozumíme jednorázové uskutečnění konstantně vymezeného souboru definičních podmínek. Předpokládáme, že pokus můžeme mnohonásobně nezávisle opakovat za dodržení definičních podmínek (ostatní podmínky se mohou měnit, proto různá opakování pokusu mohou vést k různým výsledkům). Dále předpokládáme, že opakováním pokusu vzniká opět pokus.
Deterministickým pokusem nazýváme takový pokus, jehož každé opakování vede k jedinému možnému výsledku.
Náhodným pokusem nazýváme takový pokus, jehož každé opakování vede k právě jednomu z více možných výsledků, které jsou vzájemně neslučitelné.
Příklad deterministického pokusu: při tlaku 1015 hPa zahříváme vodu na 100 °C. Jediným možným výsledkem je var vody.
Příklady náhodných pokusů: hod hrací kostkou, hod mincí, vylosování čísla z osudí apod.

Definice (definice základního prostoru): Neprázdnou množinu možných výsledků náhodného pokusu značíme Ω a nazýváme ji základní prostor. Možné výsledky značíme ω_t, kde $t \in T$, T je indexová množina.

Příklad

Příklad

a) Náhodný pokus spočívá v hodu kostkou. Možný výsledek ω_i znamená polohu kostky číslem i nahoru, i = 1, …, 6. Základní prostor $\Omega = \{\omega_1, \dots, \omega_6\}$, počet možných výsledků $m(\Omega) = 6$.

b) Náhodný pokus spočívá v hodu dvěma kostkami. Možný výsledek je uspořádaná dvojice $[\omega_i, \omega_j]$, i, j = 1, …, 6. Základní prostor $\Omega = \{[\omega_1, \omega_1], [\omega_1, \omega_2], \dots, [\omega_1, \omega_6], \dots, [\omega_6, \omega_6]\}$, počet možných výsledků $m(\Omega) = 6^2 = 36$.

c) Náhodný pokus spočívá v opakovaném házení mincí tak dlouho, dokud nepadne první líc. Potom základní prostor $\Omega = \{\omega_1, \omega_2, \omega_3 \dots\}$, kde ω_1 znamená, že hned v prvním hodu padl líc, ω_2 znamená, že až ve druhém hodu padl líc, ω_3 znamená, že až ve třetím hodu padl líc atd. Sybmolicky lze zapsat $\omega_1 = [L]$, $\omega_2 = [R, L]$, $\omega_3 = [R, R, L]$, … Tedy základní prostor Ω má nekonečně spočetně mnoho možných výsledků.

Jevové pole

Definice (definice jevového pole): Systém podmnožin $\mathcal{A}$ základního prostoru Ω, který splňuje následující tři axiómy:

J5: $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 - A_2 \in \mathcal{A}$,

J6: $\Omega \in \mathcal{A}$,

J8: $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

se nazývá jevové pole. Jestliže $A \in \mathcal{A}$, pak řekneme, že A je jev. Dvojice $(\Omega, \mathcal{A})$ se nazývá měřitelný prostor.

(Axióm J5 nám říká, že jevové pole obsahuje s každými dvěma množinami i jejich množinový rozdíl. Axióm J6 říká, že jevové pole obsahuje celý základní prostor a konečně axióm J8 říká, že když jevové pole obsahuje každou ze spočetné posloupnosti množin, obsahuje i jejich spočetné sjednocení. Znamená to, že systém $\mathcal{A}$ je uzavřený vzhledem k množinovým operacím.

Protože jevy jsou množiny, pro operace s nimi platí stejné zákony jako pro operace s množinami - komutativní zákon, asociativní zákon, de Morganova pravidla.)

Množinové a pravděpodobnostní pojmy

Poznámka (slovník množinových a pravděpodobnostních pojmů)

Ω se nazývá jistý jev, $\varnothing$ se nazývá nemožný jev

$\omega \in A$ znamená, že možný výsledek ω je příznivý nastoupení jevu A

$A \subseteq B$ znamená, že jev A má za důsledek jev B

$A \cup B$ znamená nastoupení aspoň jednoho z jevů A, B

$A \cap B$ znamená společné nastoupení jevů A, B

$A - B$ znamená nastoupení jevu A za nenastoupení jevu B

$\overline{A} = \Omega - A$ znamená jev opačný k jevu A

$A \cap B = \varnothing$ znamená, že jevy A, B jsou neslučitelné.

Příklad

Příklad: Je dán systém složený ze dvou bloků, který jednorázově použijeme. Nechť jev A_i znamená bezporuchovou funkci i-tého bloku, i = 1, 2. Pomocí jevů A_1, A_2 vyjádřete jevy:

a) bezporuchová funkce aspoň jednoho bloku: $A_1 \cup A_2$

b) bezporuchová funkce obou bloků: $A_1 \cap A_2$

c) porucha aspoň jednoho bloku: $\overline{A_1} \cup \overline{A_2}$

d) porucha obou bloků: $\overline{A_1} \cap \overline{A_2}$

e) porucha právě jednoho bloku: $\left(\overline{A_1} \cap A_2\right) \cup \left(A_1 \cap \overline{A_2}\right)$

Jevové pole - poznámky

Poznámka: Systém axiómů jevového pole je bezesporný (tj. na každém základním prostoru lze sestrojit aspoň jedno jevové pole) a neúplný (tzn., že na každém aspoň dvouprvkovém základním prostoru lze vytvořit jevových polí více).
Neúplnost systému axiómů jevového pole je výhodná, protože umožňuje rozlišovat výsledky náhodného pokusu s různým stupněm podrobnosti.
Např. jevové pole $\mathcal{A}_{min} = \{\Omega, \emptyset\}$ se nazývá minimální jevové pole a charakterizuje krajně „tupozrakého" pozorovatele, který rozliší pouze jev jistý a jev nemožný.
Jevové pole $\mathcal{A}_1 = \{\Omega, \emptyset, A, \overline{A}\}$ již dovolí rozeznat, zda nastal jev A nebo jev opačný $\overline{A}$.
Tak můžeme konstruovat stále bohatší jevová pole, až dostaneme maximální jevové pole $\mathcal{A}_{max} = \{A; A \subseteq \Omega\}$.
To charakterizuje krajně „bystrozrakého" pozorovatele, který rozliší jevy do všech podrobností. Pro libovolné jevové pole $\mathcal{A}$ ovšem platí: $\mathcal{A}_{min} \subseteq \mathcal{A} \subseteq \mathcal{A}_{max}$.

Příklad

Příklad: Sestrojte všechna možná jevová pole na základním prostoru $\Omega = \{\omega_1, \omega_2, \omega_3\}$.

Řešení:

$\mathcal{A}_1 = \{\Omega, \emptyset\} (= \mathcal{A}_{\min})$

$\mathcal{A}_2 = \{\Omega, \emptyset, \{\omega_1\}, \{\omega_2, \omega_3\}\}$

$\mathcal{A}_3 = \{\Omega, \emptyset, \{\omega_2\}, \{\omega_1, \omega_3\}\}$

$\mathcal{A}_4 = \{\Omega, \emptyset, \{\omega_3\}, \{\omega_1, \omega_2\}\}$

$\mathcal{A}_5 = \{\Omega, \emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}\} (= \mathcal{A}_{\max})$

Jevové pole - vlastnosti

Věta (vlastnosti jevového pole): Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. Pak jevové pole $\mathcal{A}$ má následujících 9 vlastností:

J1: $\mathcal{A} \neq \emptyset$,

J2: $\emptyset \in \mathcal{A}$,

J3: $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}$,

J4: $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cap A_2 \in \mathcal{A}$,

J5: $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 - A_2 \in \mathcal{A}$ (axióm),

J6: $\Omega \in \mathcal{A}$ (axióm),

J7: $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$,

J8: $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ (axióm),

J9: $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Důkaz: J1 plyne z J6.

J2 plyne z J5 a J6, protože $\Omega - \Omega = \emptyset$.

J3 plyne z J2 J8 speciální volbou $A_3 = \emptyset, A_4 = \emptyset, \ldots$ Pak $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2$.

J7 plyne J5 a J6, protože $\overline{A} = \Omega - A$.

J9 odvodíme z J7 a J8 užitím de Morganových pravidel $\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}}$:

$A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \overline{A_1}, \overline{A_2}, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} \overline{A_i} \in \mathcal{A} \Rightarrow \overline{\bigcup_{i=1}^{\infty} \overline{A_i}} \in \mathcal{A}$, ovšem $\overline{\bigcup_{i=1}^{\infty} \overline{A_i}} = \bigcap_{i=1}^{\infty} A_i$.

J4 plyne z J9 speciální volbou $A_3 = \Omega, A_4 = \Omega, \ldots$

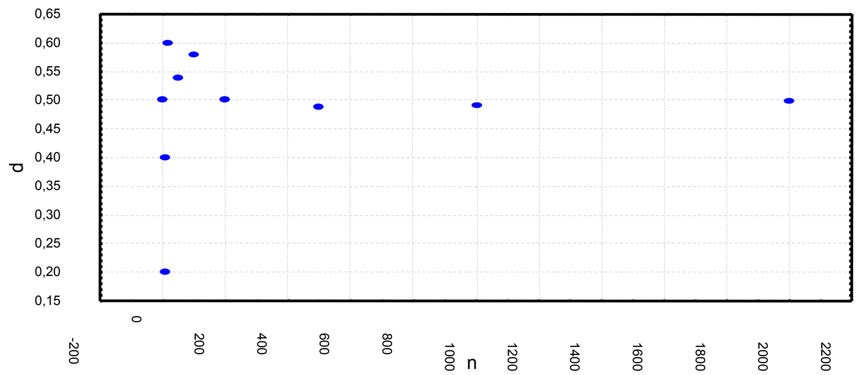
Pravděpodobnostní prostor

Motivace: Provádíme opakovaně nezávisle týž náhodný pokus a v každém pokusu sledujeme nastoupení jevu A, kterému říkáme úspěch. Označme n celkový počet pokusů a N(A) počet těch pokusů, kdy nastal úspěch. S rostoucím n pozorujeme, že relativní četnost úspěchu $\frac{N(A)}{n}$ se blíží číslu P(A), které považujeme za pravděpodobnost úspěchu. (Tento poznatek je znám jako **empirický zákon velkých čísel**).

Ilustrace empirického zákona velkých čísel

Provádíme n nezávislých hodů mincí. Padnutí líce považujeme za úspěch. Budeme sledovat závislost relativní četnosti úspěchu na počtu pokusů. (Počet pokusů volíme 2, 5, 10, 20, 50, 100, 200, 500,1000, 2000.)

n	2	5	10	20	50	100	200	500	1000	2000
p	0,5	0,2	0,4	0,6	0,54	0,58	0,5	0,488	0,49	0,4975



Axiomatická teorie pravděpodobnosti

Vzniká otázka, jak zavést pravděpodobnost, aby byla „zidealizovaným“ protějškem relativní četnosti. Zdálo by se vhodné zavést pravděpodobnost takto:

$$P(A) = \lim_{n\to\infty} \frac{N(A)}{n}.$$

Jde o tzv. **statistickou definici pravděpodobnosti**. Z matematického hlediska tato definice není v pořádku, protože počet pokusů je vždy konečný a nelze se přesvědčit o existenci uvedené limity. Proto ve 30. letech 20. století ruský matematik A. A. Kolmogorov (1903– 1987) vybudoval **axiomatickou teorii pravděpodobnosti.**

Axiomatická teorie pravděpodobnosti zavádí pravděpodobnost jako funkci, která každému jevu přiřazuje číslo mezi 0 a 1 a přitom je zidealizovaným protějškem relativní četnosti. Má tedy všechny vlastnosti relativní četnosti a kromě toho některé další vlastnosti, které vyplývají z vnitřních potřeb matematické teorie.

Pravděpodobnost - definice

Definice: Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. Reálná množinová funkce P: $\mathcal{A} \to R$ se nazývá pravděpodobnost, když splňuje následující 3 axiómy:

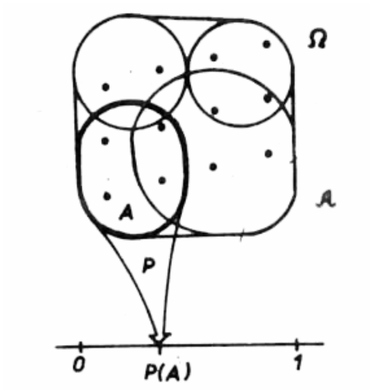
P2: $\forall A \in \mathcal{A}: P(A) \geq 0$ (nezápornost)

P10: $P(\Omega) = 1$ (normovanost)

P15: $A_1, A_2, \ldots \in \mathcal{A}$ jsou neslučitelné $\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ (spočetná aditivita)

Trojice $(\Omega, \mathcal{A}, P)$ se nazývá pravděpodobnostní prostor. (Je to matematický model jednorázového provedení náhodného pokusu.)

Ilustrace pravděpodobnostního prostoru



Poznámka: Systém axiómů pravděpodobnosti je bezesporný (tj. na každém měřitelném prostoru lze sestrojit pravděpodobnost) a neúplný (tj. na každém měřitelném prostoru, jehož jevové pole není minimální, lze sestrojit pravděpodobností více).

Pravděpodobnost - vlastnosti

Věta (vlastnosti pravděpodobnosti): Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $A, A_1, A_2, \ldots \in \mathcal{A}$ libovolné jevy. Pak pravděpodobnost P má následujících 17 vlastností:

P1: $P(\emptyset) = 0$

P2: $P(A) \geq 0$ (nezápornost – axióm)

P3: $P(A_1 \cup A_2) + P(A_1 \cap A_2) = P(A_1) + P(A_2)$

P4: $1 + P(A_1 \cap A_2) \geq P(A_1) + P(A_2)$

P5: $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$ (subaditivita)

P6: $A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$ (aditivita)

P7: $P(A_2 - A_1) = P(A_2) - P(A_1 \cap A_2)$

P8: $A_1 \subseteq A_2 \Rightarrow P(A_2 - A_1) = P(A_2) - P(A_1)$ (subtraktivita)

P9: $A_1 \subseteq A_2 \Rightarrow P(A_2) \leq P(A_1)$ (monotonie)

P10: $P(\Omega) = 1$ (normovanost – axióm)

P11: $P(A) + P(\overline{A}) = 1$ (komplementarita)

P12: $P(A) \leq 1$

P13: $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$ (spočetná subaditivita)

P14: $A_1, A_2, \ldots \in \mathcal{A}$ jsou neslučitelné $\Rightarrow \sum_{i=1}^{\infty} P(A_i) < \infty$ (absolutní konvergence)

P15: $A_1, A_2, \ldots \in \mathcal{A}$ jsou neslučitelné $\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ (spočetná aditivita – axióm)

P16: $A_1 \subseteq A_2 \subseteq \ldots \in \mathcal{A} \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} P(A_i)$ (spojitost pravděpodobnosti zdola)

P17: $A_1 \supseteq A_2 \supseteq \ldots \in \mathcal{A} \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} P(A_i)$ (spojitost pravděpodobnosti shora)

Pravděpodobnost - vlastnosti

P14 Položme $A_0 = \overline{\bigcup_{i=1}^{\infty} A_i}$. Pak jevy $A_0, A_1, A_2, \ldots$ jsou neslučitelné a jejich sjednocením je celý základní prostor, tedy podle axiómu P10 dostáváme: $1 = P(\Omega) = P(\bigcup_{i=0}^{\infty} A_i) = \sum_{i=0}^{\infty} P(A_i)$, přičemž poslední rovnost vyplývá z axiómu P15. $\sum_{i=0}^{\infty} P(A_i)$ tedy absolutně konverguje, tudíž bude konvergovat také $\sum_{i=1}^{\infty} P(A_i)$, kde jsme vynechali první člen.

P1 Položme $A_1 = \emptyset, A_2 = \emptyset, \ldots$ Pak $\bigcup_{i=1}^{\infty} A_i = \emptyset$, tedy podle axiómu P15 $0 = P(\emptyset) = P(\bigcup_{i=1}^{\infty} \emptyset) = \sum_{i=1}^{\infty} P(\emptyset)$, což je možné jen tak, že $P(\emptyset) = 0$.

P6 V axiómu P15 položíme $A_3 = \emptyset, A_4 = \emptyset, \ldots$, tedy $P(\bigcup_{i=1}^{\infty} A_i) = P(A_1 \cup A_2) = \sum_{i=1}^{\infty} P(A_i) = P(A_1) + P(A_2)$.

P11 Plyne z vlastnosti P6 a axiómu P10: $P(A \cup \overline{A}) = P(\Omega) = 1 = P(A) + P(\overline{A})$.

P12 Plyne okamžitě z axiómu P2 a vlastnosti P11.

Pravděpodobnost - vlastnosti

Pro důkaz vlastností P3, P4 a P5 jevy $A_1 \cup A_2$, A_1 a A_2 rozložíme na součet disjunktních sčítanců:
$A_1 \cup A_2 = (A_1 \setminus A_2) \cup (A_1 \cap A_2) \cup (A_2 \setminus A_1)$
$A_1 = (A_1 \setminus A_2) \cup (A_1 \cap A_2)$
$A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2)$

P3 Podle P6 dostáváme: $P(A_1 \cup A_2) + P(A_1 \cap A_2) = P(A_1 \setminus A_2) + P(A_1 \cap A_2) + P(A_2 \setminus A_1) + P(A_1 \cap A_2) = P(A_1) + P(A_2)$.

Protože podle P12 je $P(A_1 \cup A_2) \leq 1$ a podle P2 je $P(A_1 \cap A_2) \geq 0$, dostáváme z P3 okamžitě P4 a P5.

P7 Opět vyjádříme A_2 jako sjednocení neslučitelných jevů: $A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2)$. Podle P3 pak dostaneme: $P(A_2) = P(A_2 \setminus A_1) + P(A_1 \cap A_2)$, tedy $P(A_2 \setminus A_1) = P(A_2) - P(A_1 \cap A_2)$.

P8 Jelikož $A1 \subseteq A_2$, platí $A_1 \cap A_2 = A_1$ a P8 plyne z P7.

P9 Plyne z P8, protože podle P2 je $P(A_2 \setminus A_1) \geq 0$, tudíž $P(A_2) - P(A_1) \geq 0$, tj. $P(A_1) \leq P(A_2)$.

Pravděpodobnost - vlastnosti

P13 Položíme $\bigcup\limits_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \ldots$ Tím jsme dostali sjednocení posloupnosti neslučitelných jevů a aplikujeme axióm P15 a vlastnost P7: $P(\bigcup\limits_{i=1}^{\infty} A_i) = P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus (A_1 \cup A_2)) + \ldots \leq P(A_1) + P(A_2) + P(A_3) + \ldots = \sum\limits_{i=1}^{\infty} P(A_i)$.

P16 Jev $\bigcup\limits_{i=1}^{\infty} A_i$ vyjádříme jako sjednocení neslučitelných jevů. Z předpokladu $A_1 \subseteq A_2 \subseteq \ldots$ plyne $\bigcup\limits_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \ldots \cup (A_i \setminus A_{i-1}) \cup \ldots$, tedy podle axiómu P15 a vlastnosti P8 dostáváme: $P(\bigcup\limits_{i=1}^{\infty} A_i) = P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \ldots + P(A_i \setminus A_{i-1}) + \ldots = P(A_1) + [P(A_2) - P(A_1)] + [P(A_3) + P(A_2)] + \ldots + [P(A_i) + P(A_{i-1})] + \ldots = \lim\limits_{i \to \infty} P(A_i)$.

P17 Podle vlastnosti P16 dostáváme $P(\overline{\bigcup\limits_{i=1}^{\infty} A_i}) = \lim\limits_{i \to \infty} P(\overline{A_i})$. Z de Morganových pravidel plyne $P(\bigcap\limits_{i=1}^{\infty} A_i) = P(\overline{\bigcup\limits_{i=1}^{\infty} \overline{A_i}}) = 1 - P(\bigcup\limits_{i=1}^{\infty} \overline{A_i}) = 1 - \lim\limits_{i \to \infty} P(\overline{A_i}) = 1 - \lim\limits_{i \to \infty} [1 - P(A_i)] = \lim\limits_{i \to \infty} P(A_i)$.

Pravděpodobnost - vlastnosti

Věta (další vlastnosti pravděpodobnosti): Nechť (Ω, $\mathcal{A}$, P) je pravděpodobnostní prostor, $A_1, A_2, \ldots, A_n \in \mathcal{A}$ libovolné jevy. Pak platí:

a) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n)$

(Pro neslučitelné jevy $A_1, \ldots, A_n$ dostáváme $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.) (Věta o sčítání pravděpodobností)

b) $\max_{1 \leq i \leq n} P(A_i) \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$

c) $1 - n + \sum_{i=1}^{n} P(A_i) \leq P\left(\bigcap_{i=1}^{n} A_i\right) \leq \min_{1 \leq i \leq n} P(A_i)$ (nerovnost vlevo se nazývá Bonferroniho nerovnost)

Pravděpodobnost - vlastnosti

Důkaz:

ad a) Vlastnost vyjadřuje princip inkluze a exkluze. Tvrzení o neslučitelných jevech plyne z axiómu P15, kde položíme $A_{n+1} = \varnothing, A_{n+2} = \varnothing, \dots$

ad b) Levá strana: Plyne z monotonie P9. Pro $\forall i \in \{1,\dots,n\}$ je $A_i \subseteq \bigcup_{j=1}^{n} A_j$, tedy pro $\forall i \in \{1,\dots,n\}$ platí $P(A_i) \le P\left(\bigcup_{j=1}^{n} A_j\right)$.

Tvrzení musí platit i pro ten index i, pro který je $P(A_i)$ maximální.

Pravá strana: Plyne ze spočetné subaditivity P13, kde položíme $A_{n+1} = \varnothing, A_{n+2} = \varnothing, \dots$

ad c) Levá strana: $P\left(\bigcap_{i=1}^{n} A_i\right) = P\left(\overline{\bigcup_{i=1}^{n} \overline{A_i}}\right) = 1 - P\left(\bigcup_{i=1}^{n} \overline{A_i}\right) \ge 1 - \sum_{i=1}^{n} P\left(\overline{A_i}\right) = 1 - \sum_{i=1}^{n} [1 - P(A_i)] = 1 - n + \sum_{i=1}^{n} P(A_i)$

Pravá strana: Plyne z monotonie P9. Pro $\forall i \in \{1,\dots,n\}$ je $A_i \supseteq \bigcap_{j=1}^{n} A_j$, tedy pro $\forall i \in \{1,\dots,n\}$ platí

$P(A_i) \ge P\left(\bigcap_{j=1}^{n} A_j\right)$. Tvrzení musí platit i pro ten index i, pro který je $P(A_i)$ minimální.

Příklad

Příklad: Je dán systém složený ze dvou bloků. Jev A_i značí bezporuchovou funkci i-tého bloku, i = 1, 2. Je známo, že $P(A_i) = \vartheta_i$, i = 1, 2.

a) Odhadněte pravděpodobnost správné funkce celého systému, jsou-li bloky zapojeny

α) sériově, β) paralelně.

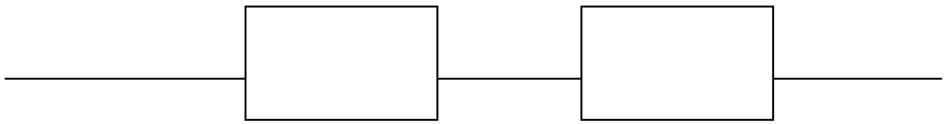
b) Předpokládejme navíc, že $P(A_1 \cap A_2) = \vartheta_{12}$. Vypočtěte nyní pravděpodobnost správné funkce celého systému, jsou-li bloky zapojeny

α) sériově, β) paralelně.

Řešení:

ad a)

Případ sériového zapojení



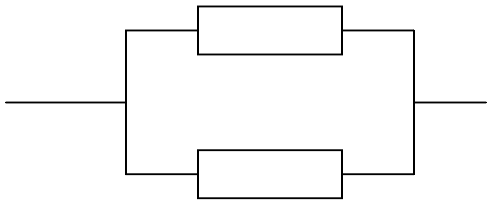
$P(A_1 \cap A_2)$ lze shora i zdola odhadnout pomocí věty 2.5. (c), kde n = 2:

$$1 - 2 + P(A_1) + P(A_2) \le P(A_1 \cap A_2) \le \min\{P(A_1), P(A_2)\}$$

$$\vartheta_1 + \vartheta_2 - 1 \le P(A_1 \cap A_2) \le \min\{\vartheta_1, \vartheta_2\}$$

Příklad

Případ paralelního zapojení



$P(A_1 \cup A_2)$ lze shora i zdola odhadnout pomocí věty 2.5. (b), kde n = 2:

$$\max\{P(A_1), P(A_2)\} \leq P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$
$$\max\{\vartheta_1, \vartheta_2\} \leq P(A_1 \cup A_2) \leq \vartheta_1 + \vartheta_2$$

ad b)
Případ sériového zapojení: $P(A_1 \cap A_2) = \vartheta_{12}$
Případ paralelního zapojení: Podle vlastnosti P3 dostáváme: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \vartheta_1 + \vartheta_2 - \vartheta_{12}$

4. Diskrétní pravděpodobnost. Stochasticky nezávislé jevy.

Motivace: V Kolmogorovově axiomatické definici pravděpodobnosti se nespecifikuje jak na daném měřitelném prostoru (Ω, $\mathbf{A}$) zkonstruovat pravděpodobnost P. Jednou z možností je zavedení tzv. diskrétní pravděpodobnosti. Používá se v situacích, kdy pouze spočetně mnoho možných výsledků náhodného pokusu má reálnou šanci na uskutečnění. Šance těchto výsledků se ohodnotí vahami a pravděpodobnost jevu se pak získá jakou součet těch vah možných výsledků, které jsou příznivé nastoupení daného jevu. Jestliže je možných výsledků pouze konečně mnoho a všechny mají stejnou šanci na uskutečnění, dostáváme klasickou pravděpodobnost jako speciální případ diskrétní pravděpodobnosti.

Konstrukce diskrétní pravděpodobnosti

Označení: Nechť $M \neq \emptyset, B \subseteq M, g: M \to (-\infty, \infty)$ je funkce, která je všude nulová s výjimkou nejvýše spočetné množiny $G \subseteq M$. Pak symbol $\sum_{x \in B} g(x)$ má tento význam:

a) Je-li $B \cap G = \emptyset$, pak $\sum_{x \in B} g(x) = 0$.

b) Je-li $B \cap G$ konečná množina, pak prvky tohoto průniku uspořádáme do konečné posloupnosti $\{x_1, \ldots, x_n\}$ a $\sum_{x \in B} g(x) = \sum_{i=1}^{n} g(x_i)$.

c) Je-li $B \cap G$ spočetná množina, pak prvky tohoto průniku uspořádáme do spočetné posloupnosti $\{x_1, x_2, \ldots\}$ a $\sum_{x \in B} g(x) = \sum_{i=1}^{\infty} g(x_i)$, pokud tato suma absolutně konverguje. Není- li podmínka absolutní konvergence splněna, nemá uvedený symbol smysl.

Je-li $B = (-\infty, \infty)$ nebo $B = (-\infty, x\rangle$, pak píšeme $\sum_{x \in B} g(x) = \sum_{x=-\infty}^{\infty} g(x)$ $\quad \sum_{x \in B} g(x) = \sum_{t \leq x} g(t)$

Konstrukce diskrétní pravděpodobnosti

Definice (definice diskrétní pravděpodobnosti): Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor, $\Gamma \subseteq \Omega$ nejvýše spočetná podmnožina základního prostoru. Funkce $\gamma : \Omega \to R$, která je kladná pouze na Γ a jinak je nulová a splňuje podmínku $\sum_{\omega\in\Omega} \gamma(\omega) = 1$, se nazývá váhová funkce. Reálná množinová funkce P: $\mathcal{A} \to$ R daná vzorcem

$\forall A \in \mathcal{A} : P(A) = \sum_{\omega\in A} \gamma(\omega)$ se nazývá diskrétní pravděpodobnost.

Věta: Diskrétní pravděpodobnost je pravděpodobnost ve smyslu axiomatické definice, tzn., že má vlastnosti P1 – P17.

Důkaz: Stačí ověřit platnost axiómů P2, P10, P15.

P2: P(A) $\geq$ 0 – plyne z definice váhové funkce.

P10: $P(\Omega) = \sum_{\omega\in\Omega} \gamma(\omega) = 1$ - splněno.

P15: $A_1, A_2, \ldots \in \mathcal{A}$ jsou neslučitelné $\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{\omega\in\bigcup_{i=1}^{\infty} A_i} \gamma(\omega) = \sum_{\omega\in A_1} \gamma(\omega) + \sum_{\omega\in A_2} \gamma(\omega) + \ldots = P(A_1) + P(A_2) + \ldots = \sum_{i=1}^{\infty} P(A_i)$ - splněno.

Příklad

Příklad: Házíme kostkou, jejíž těžiště je posunuto z geometrického středu k onomu vrcholu, v němž se stýkají stěny s nejnižším ohodnocením. Vypočtěte pravděpodobnost, že padne sudé číslo.

Řešení: $\Omega = \{\omega_1, \ldots, \omega_6\}$, $\mathcal{A} = \mathcal{A}_{max} = \{A; A \subseteq \Omega\}$. V souladu se zadáním a vzhledem k symetrii úlohy volíme váhy $\gamma(\omega_4) = \gamma(\omega_5) = \gamma(\omega_6) = \frac{\vartheta}{3}$, $\frac{1}{2} < \vartheta < 1$, $\gamma(\omega_1) = \gamma(\omega_2) = \gamma(\omega_3) = \frac{1-\vartheta}{3}$.

$A = \{\omega_2, \omega_4, \omega_6\}, P(A) = \gamma(\omega_2) + \gamma(\omega_4) + \gamma(\omega_6) = \frac{1-\vartheta}{3} + \frac{\vartheta}{3} + \frac{\vartheta}{3} = \frac{1+\vartheta}{3}$.

Např. pro $\vartheta = \frac{2}{3}$ dostáváme $P(A) = \frac{1+\frac{2}{3}}{3} = \frac{5}{9} = 0,\overline{5}$.

Klasická pravděpodobnost

Poznámka: Z matematického hlediska je přípustné zavést ještě možný výsledek ω_0 - poloha kostky na hraně. Tomuto možnému výsledku přiřadíme váhu $\gamma(\omega_0) = 0$. Tím se pravděpodobnostní prostor (Ω, $\mathcal{A}$, P) sice formálně změní, ale výsledky zůstanou stejné. Položíme-li $A_0 = \{\omega_0\}$, pak $P(A_0) = 0$. Jde o jev s pravděpodobností 0, nikoli však nemožný jev.

Definice (definice klasické pravděpodobnosti): Nechť (Ω, $\mathcal{A}$) je měřitelný prostor, Ω konečná množina. Funkce P: $\mathcal{A}$ $\rightarrow$ R daná vzorcem

$$\forall A \in \mathcal{A} : P(A) = \frac{m(A)}{m(\Omega)},$$

kde $m(A)$ je počet možných výsledků příznivých nastoupení jev u A a $m(\Omega)$ je počet všech možných výsledků, se nazývá klasická pravděpodobnost.

Věta: Klasická pravděpodobnost je pravděpodobnost ve smyslu axiomatické definice, tj. má vlastnosti P1 až P17. Kromě toho pro ni platí: $P(A) = 0 \Leftrightarrow A = \emptyset$, $P(A) = 1 \Leftrightarrow A = \Omega$.

Důkaz: Klasická pravděpodobnost je speciálním případem diskrétní pravděpodobnosti s vahami $\gamma(\omega) = \frac{1}{m(\Omega)}$.

Poslední dvě implikace vyplývají z definice klasické pravděpodobnosti.

Příklad

Dřevenou, natřenou krychli o straně 4 cm rozřežeme na jednotkové krychličky. Jaká je pravděpodobnost, že náhodně vybraná krychlička

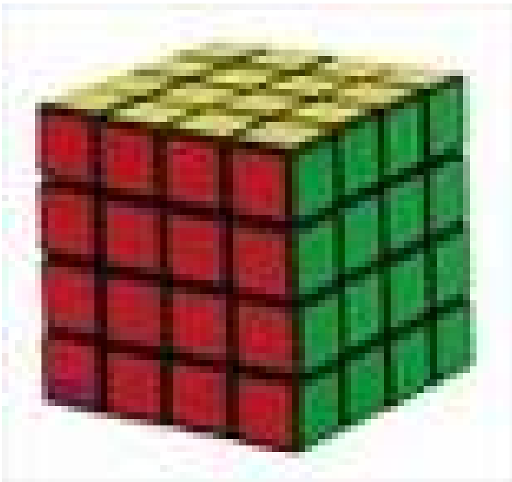
a) má právě 2 natřené stěny,

b) nemá žádnou natřenou stěnu?

Řešení:

$$P(A_1) = \frac{8 \cdot 3}{64} = \frac{3}{8}$$

$$P(A_2) = \frac{2^3}{64} = \frac{1}{8}$$



Příklad

1) Jaká je pravděpodobnost, že slovem náhodně sestaveným z písmen A, A, A, E, I, K, M, M, T, T bude MATEMATIKA?

Řešení: $P(A) = \frac{3! \cdot 2! \cdot 2!}{10!} = \frac{1}{151200}$

2) Jaká je pravděpodobnost, že z n lidí někteří dva slaví narozeniny ve stejný den (n < 365)?

$$P(A) = 1 - P(\overline{A}) = \frac{365 \cdot 364 \cdot \cdots \cdot (365-n+1)}{365^n} = \frac{365!}{(365-n)! \cdot 365^n}$$

Přes pravděpodobnost 1/2 se dostaneme už při n = 23:

n	10	20	22	23	30	40	50	60
p	0,117	0,411	0,476	0,507	0,706	0,891	0,870	0,994

Příklad

Příklad: V dodávce 100 kusů výrobků nemá požadovaný průměr 10 kusů, požadovanou délku 20 kusů a současně nemá požadovaný průměr i délku 5 kusů. Jaká je pravděpodobnost, že náhodně vybraný výrobek z této dodávky má požadovaný průměr i délku?

Řešení:

Jev A spočívá v tom, že výrobek má požadovaný průměr a jev B v tom, že výrobek má požadovanou délku. Počítáme

$$P(A \cap B) = P(\overline{\overline{A} \cup \overline{B}}) = 1 - P(\overline{A} \cup \overline{B}) = 1 - [P(\overline{A}) + P(\overline{B}) - P(\overline{A} \cap \overline{B})] = 1 - \left(\frac{10}{100} + \frac{20}{100} - \frac{5}{100}\right) = 0{,}75$$

Pravděpodobnost, že náhodně vybraný výrobek z dodávky má požadovaný průměr i délku, je 0,75.

Příklad

Jsou dány dvě různé rovnoběžné přímky a, b. Na a je n různých bodu $A_1, \ldots, A_n$, na b m různých bodu $B_1, \ldots, B_m$. Jaká je pravděpodobnost, že 3 náhodně vybrané body tvoří trojúhelník?

Řešení:

$$P(A) = \frac{\binom{n}{2}\binom{m}{1} + \binom{n}{1}\binom{m}{2}}{\binom{n+m}{3}} = \frac{\frac{mn(n-1)}{2} + \frac{nm(m-1)}{2}}{\frac{(m+n)(m+n-1)(m+n-2)}{3\cdot 2\cdot 1}} =$$

$$= \frac{3mn}{(m+n)(m+n-1)}$$

Příklad

Po číselné ose se posuneme o jedničku vpravo, nebo vlevo, podle toho, zda nám na minci padne rub, nebo líc. Začínáme v 0. Jaká je pravděpodobnost, že po 2n krocích budeme v 0?

Řešení: Pravděpodobnostním prostorem mohou být posloupnosti nul a jedniček délky 2n (označující kam v kterém kroku jdeme). Abychom po 2n-tém kroku stáli v nule, musíme jít v průběhu stejněkrát (tj. n krát) doleva a n krát doprava:

$$P(A) = \frac{\binom{2n}{n}}{2^{2n}}$$

Např. pro n=5 (2n=10): $P(A) = \frac{\binom{10}{5}}{2^{10}} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1024} = \frac{63}{256} = 0{,}246$

Příklad

Čísla 1, . . . , n promícháme. Jaká je pravděpodobnost, že aspoň jedno číslo bude na svém místě? Najděte její limitu při n →∞.

Řešení: Nechť jev A_i je „číslo i na svém místě“. Podle vzorce pro sjednocení

$$\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right) = n\frac{1\cdot(n-1)!}{n!} - \binom{n}{2}\frac{1\cdot 1\cdot(n-2)!}{n!} + \cdots + (-1)^{n+1}\binom{n}{n}\frac{1\cdot 1\ldots 1}{n!} = -\sum_{k=1}^{n}(-1)^k\binom{n}{k}\frac{(n-k)!}{n!} = -\sum_{k=1}^{n}\frac{(-1)^k}{k!} \to -\mathrm{e}^{-1}+1.$$

Příklad

Příklad ♣ : Z množiny $\{a_1,\ldots,a_n\}$ zvané základní soubor rozsahu n vybereme k-krát po jednom prvku, který vždy vrátíme zpět. Získáme uspořádanou k-tici $\lfloor a_{i_1},\ldots,a_{i_k}\rfloor$, která se nazývá uspořádaný výběrový soubor rozsahu k s vracením. Předpokládáme, že v základním souboru je právě r prvků označeno, $r \le n$. Zavedeme jevy A_1, B_1, C_1 takto:

A_1 … jev, že každý z prvků základního souboru se ve výběrovém souboru ocitne nejvýše 1x,
B_1 … jev, že předem pevně daný prvek základního souboru se ve výběrovém souboru ocitne aspoň 1x,
C_1 … jev, že ve výběrovém souboru se ocitne právě x označených prvků, $x \le k$.

Řešení:

Ω_1 … množina všech uspořádaných k-tic utvořených z prvků $a_1,\ldots,a_n$, kde se prvky mohou opakovat, $m(\Omega_1) = n^k$.

Jevu A_1 jsou příznivé ty k-tice, kde se prvky neopakují, tedy $m(A_1) = n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)$.

$$P(A_1) = \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right) = \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right)$$

$\overline{B_1}$ … jev, že předem pevně daný prvek základního souboru se do výběrového souboru vůbec nedostane.
Jevu $\overline{B_1}$ jsou příznivé ty k-tice, ve kterých se daný prvek vůbec nevyskytuje. Je zřejmé, že $m(\overline{B_1}) = (n-1)^k$. Pak

$$P(B_1) = 1 - P(\overline{B_1}) = 1 - \frac{(n-1)^k}{n^k} = 1 - \left(1-\frac{1}{n}\right)^k.$$

Jevu C_1 jsou příznivé ty k-tice, které mají na x místech označené prvky a na zbylých k-x místech neoznačené prvky. Pak $m(C_1) = r^x (n-r)^{k-x}\binom{k}{x}$, $P(C_1) = \dfrac{\binom{k}{x} r^x (n-r)^{k-x}}{n^k} = \binom{k}{x}\left(\dfrac{r}{n}\right)^x\left(1-\dfrac{r}{n}\right)^{k-x}$.

Příklad

Příklad ♣♣: Za jinak stejných podmínek jako v př. ♣ nevracíme vybrané prvky zpět do základního souboru (předpokládáme, že $k \le n, x \le r$). Získanou uspořádanou k-tici nazveme uspořádaný výběrový soubor rozsahu k bez vracení. Jevy A_2, B_2, C_2 jsou definovány stejně jako v př. ♣. Vypočtěte jejich pravděpodobnosti.

Řešení:

Ω_2 … množina všech uspořádaných k-tic utvořených z prvků $a_1, \ldots, a_n$, kde se prvky nemohou opakovat,

$$m(\Omega_2) = \frac{n!}{(n-k)!}.$$

$$A_2 = \Omega_2, P(A_2) = 1$$

$$P(B_2) = 1 - P(\overline{B_2}) = 1 - \frac{\frac{(n-1)!}{(n-1-k)!}}{\frac{n!}{(n-k)!}} = 1 - \frac{n-k}{n} = \frac{k}{n}$$

$$P(C_2) = \frac{\frac{r!}{(r-x)!} \cdot \frac{(n-r)!}{(n-r-k+x)!} \cdot \binom{k}{x}}{\frac{n!}{(n-k)!}} = \frac{\frac{k!}{x!(k-x)!} \cdot \frac{r!}{(r-x)!} \cdot \frac{(n-r)!}{(n-r-k+x)!}}{\frac{n!}{(n-k)!}} = \frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}}$$

Příklad

Příklad ♣♣♣: Za jinak stejných podmínek jako v př. ♣♣ . považujeme za výběrový soubor rozsahu k bez vracení nikoliv uspořádanou k-tici, ale podmnožinu $\{a_{i_1},\dots,a_{i_n}\}$. Vypočtěte pravděpodobnosti jevů A_3, B_3, C_3.

Řešení:

Ω_3 … množina všech neuspořádaných k-tic utvořených z prvků $a_1,\dots,a_n$, kde se prvky nemohou opakovat,

$$m(\Omega_2)=\binom{n}{k}.$$

$$A_3=\Omega_3, P(A_3)=1$$

$$P(B_3)=1-P(\overline{B_3})=1-\frac{\binom{n-1}{k}}{\binom{n}{k}}=1-\frac{\frac{(n-1)!}{k!(n-1-k)!}}{\frac{n!}{k!(n-k)!}}=1-\frac{n-k}{n}=\frac{k}{n}$$

$$P(C_3)=\frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}}$$

Dospěli jsme ke stejným výsledkům jako v příkladě ♣♣ , i když jsme použili jiný model. Je to způsobeno tím, že v př ♣♣ . jsme použili variace bez opakování a v př.♣♣♣ kombinace bez opakování. Každé kombinaci $\{a_{i_1},\dots,a_{i_k}\}$ odpovídá právě k! variací.

Příklad

Poznámka: Nemá smysl uvažovat o kombinacích s opakováním. Model by byl nerealistický, protože možné výsledky pokusu by neměly stejnou šanci na uskutečnění. Např. kombinaci $\{a_1,\dots,a_k\}$ odpovídá k! variací (s opakováním), zatímco kombinaci $\{a_1,\dots,a_1\}$ odpovídá pouze jediná variace s opakováním.

Příklad: Nechť r_n je počet předem označených prvků v základním souboru rozsahu n. Nechť rozsah k výběrového souboru je pevný a předpokládejme, že $\lim\limits_{n\to\infty}\frac{r_n}{n}=\vartheta$. Dokažte, že $\lim\limits_{n\to\infty}P(C_2)=P(C_1)$ při označení $\frac{r}{n}=\vartheta$.

Řešení:

$$\lim_{n\to\infty}P(C_2)=\lim_{n\to\infty}\frac{\binom{r_n}{x}\binom{n-r_n}{k-x}}{\binom{n}{k}}=\lim_{n\to\infty}\frac{\frac{r_n!}{x!(r_n-x)!}\cdot\frac{(n-r_n)!}{(k-x)!(n-r_n-k+x)!}}{\frac{n!}{k!(n-k)!}}=$$

$$=\binom{k}{x}\lim_{n\to\infty}\frac{r_n(r_n-1)\cdot\ldots\cdot(r_n-x+1)(n-r_n)(n-r_n-1)\cdot\ldots\cdot(n-r_n-k+x+1)}{n(n-1)\cdot\ldots\cdot(n-k+1)}=$$

$$=\binom{k}{x}\lim_{n\to\infty}\frac{\frac{r_n}{n}\left(\frac{r_n}{n}-\frac{1}{n}\right)\cdot\ldots\cdot\left(\frac{r_n}{n}-\frac{x-1}{n}\right)\left(1-\frac{r_n}{n}\right)\left(1-\frac{r_n+1}{n}\right)\cdot\ldots\cdot\left(1-\frac{r_n}{n}-\frac{k-x-1}{n}\right)}{1\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)}=$$

$$=\binom{k}{x}\vartheta^x(1-\vartheta)^{k-x}=P(C_1)$$

Příklad

Příklad: Předpokládáme, že rozsah k výběrového souboru je pevný. Vypočtěte limity pravděpodobností jevů A_1, B_1, A_2, B_2 z příkladů 3.10. a 3.11. pro $n \to \infty$.

Řešení:

$$\lim_{n\to\infty} P(A_1) = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{n^k} = 1, \quad \lim_{n\to\infty} P(A_2) = \lim_{n\to\infty} 1 = 1$$

$$\lim_{n\to\infty} P(B_1) = \lim_{n\to\infty} \left[1 - \left(1 - \frac{1}{n}\right)^k\right] = 1 - 1 = 0, \quad \lim_{n\to\infty} P(B_2) = \lim_{n\to\infty} \frac{k}{n} = 0$$

Závěr plynoucí z příkladů 3.14. a 3.15.: Pokud je rozsah výběrového souboru konstantní a rozsah základního souboru roste nade všechny meze, stírá se rozdíl mezi výběrovým souborem s vracením a bez vracení.

Stochasticky nezávislé jevy

Motivace: Při provádění pokusu se může stát, že z informace o nastoupení či nenastoupení jednoho jevu jsme schopni odvodit, zda jiný jev nastoupí či nenastoupí, tzn., že platí jedna z inkluzí $A \subseteq B, \overline{A} \subseteq B, A \subseteq \overline{B}, \overline{A} \subseteq \overline{B}$. V takovém případě hovoříme o deterministicky závislých jevech. Jejich protipólem jsou jevy stochasticky nezávislé – informace o nastoupení či nenastoupení jednoho jevu nijak neovlivní šance, s nimiž očekáváme nastoupení jiného jevu.
V popisné statistice jsme zavedli četnostní nezávislost dvou množin G_1, G_2 v daném výběrovém souboru pomocí multiplikativního vztahu: $p(G_1 \cap G_2) = p(G_1)p(G_2)$. V počtu pravděpodobnosti požadujeme pro stochasticky nezávislé jevy A_1, A_2 splnění multiplikativního vztahu: $P(A_1 \cap A_2) = P(A_1)P(A_2)$. Pro tři jevy budeme požadovat, aby i jevy $A_1 \cap A_2$ a A_3 byly stochasticky nezávislé, což vede ke vztahu $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$. Tak můžeme pokračovat pro libovolný počet jevů.

Definice: Nechť (Ω, $\mathcal{A}$, P) je pravděpodobnostní prostor. Řekneme, že jevy A, B $\in \mathcal{A}$ jsou stochasticky nezávislé (vzhledem k P), jestliže P(A $\cap$ B) = P(A) P(B).

Stochasticky nezávislé jevy

Věta: Pro libovolné jevy A, B $\in \mathcal{A}$ platí:

a) $\varnothing$ a A jsou stochasticky nezávislé jevy.

b) Ω a A jsou stochasticky nezávislé jevy.

c) Jsou-li A, B stochasticky nezávislé jevy, pak jsou stochasticky nezávislé též jevy $\overline{A}, B$ a $A, \overline{B}$ a $\overline{A}, \overline{B}$.

Důkaz:

ad a) $P(\varnothing \cap A) = P(\varnothing)P(A) : 0 = 0 \cdot P(A) = 0$

ad b) $P(\Omega \cap A) = P(\Omega)P(A) : P(A) = 1 \cdot P(A) = P(A)$

ad c) $P(\overline{A} \cap B) = P(B - (A \cap B)) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)[1 - P(A)] = P(\overline{A})P(B)$

Tvrzení pro jevy $A, \overline{B}$ se dokáže analogicky.

$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] = 1 - P(A) - P(B) + P(A)P(B) = 1 - P(A) - P(B)[1 - P(A)] =$

$= [1 - P(A)][1 - P(B)] = P(\overline{A})P(\overline{B})$

Stochasticky nezávislé jevy

Definice: Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Jevy $A_1, \dots, A_n \in \mathcal{A}$ jsou stochasticky nezávislé (vzhledem k P), jestliže platí systém multiplikativních vztahů:

$< j \le n:$ $P(A_i \cap A_j) = P(A_i)\,P(A_j)$ (dvojmístný multiplikativní vztah)

$\forall 1 \le i < j < k \le n:$ $P(A_i \cap A_j \cap A_k) = P(A_i)\,P(A_j)\,P(A_k)$ (trojmístný multiplikativní vztah)

$\vdots$

$P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n)$ (n-místný multiplikativní vztah)

Jevy $A_1, A_2, \dots \in \mathcal{A}$ jsou stochasticky nezávislé (vzhledem k P), jestliže pro všechna přirozená n jsou stochasticky nezávislé jevy $A_1, \dots, A_n \in \mathcal{A}$.

Příklad: V osudí jsou 4 lístky s číslicemi 000, 011, 110 a 101. Označme A_i jev, že na náhodně vytaženém lístku je 1 na i-tém místě. Zjistěte, zda jevy A_1, A_2, A_3 jsou stochasticky nezávislé.

Řešení: $P(A_1) = \frac{1}{2}$, $P(A_2) = \frac{1}{2}$, $P(A_3) = \frac{1}{2}$, $P(A_1 \cap A_2) = \frac{1}{4}$, $P(A_1 \cap A_3) = \frac{1}{4}$, $P(A_2 \cap A_3) = \frac{1}{4}$. Vidíme, že dvoumístné multiplikativní vztahy jsou splněny, avšak trojmístný vztah nikoli, neboť $P(A_1 \cap A_2 \cap A_3) = 0$ a $P(A_1)P(A_2)P(A_3) = \frac{1}{8}$. Jevy A_1, A_2, A_3 nejsou stochasticky nezávislé.

Příklad

Ukázka příkladu, kdy jsou jevy po dvou nezávislé, ale jsou celkově závislé. Uvažujme náhodný pokus „hod dvěmi mincemi", kdy sledujeme zda na mincích padl líc (L) nebo (R). Množina všech možných výsledků (elementárních jevů) je tedy $\Omega = \{LL, LR, RL, RR\}$ a všechny elementární jevy jsou stejně pravděpodobné, tj. mají pravděpodobnost $\frac{1}{4}$.

Najděte pravděpodobnost a zjistěte zda jsou nezávislé a po dvou nezávislé jevy

(a) A_1 na první mince padne líc;

(b) A_2 na druhé minci padne líc;

(c) A_3 na obou mincích padne totéž.

Řešení: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\mathcal{P}(\omega_i) = 1/4$ $A_1 = \{\omega_1, \omega_2\}$, $\mathcal{P}(A_1) = 1/2$
$A_2 = \{\omega_1, \omega_3\}$, $\mathcal{P}(A_2) = 1/2$
$A_3 = \{\omega_1, \omega_4\}$, $\mathcal{P}(A_3) = 1/2$
jevy A_1 a A_2 jsou nezávislé, protože $A_1 \cap A_2 = \{\omega_1\}$ a

$$\mathcal{P}(A_1 \cap A_2) = 1/4 = 1/2 \cdot 1/2$$

jevy A_1 a A_3 jsou nezávislé, protože $A_1 \cap A_3 = \{\omega_1\}$ a

$$\mathcal{P}(A_1 \cap A_3) = 1/4 = 1/2 \cdot 1/2$$

jevy A_2 a A_3 jsou nezávislé, protože $A_2 \cap A_3 = \{\omega_1\}$ a

$$\mathcal{P}(A_2 \cap A_3) = 1/4 = 1/2 \cdot 1/2$$

jevy A_1,A_2 a A_3 jsou závislé, protože $A_1 \cap A_2 \cap A_3 = \{\omega_1\}$ a

$$\mathcal{P}(A_1 \cap A_2 \cap A_3) = 1/4 \neq 1/2 \cdot 1/2 \cdot 1/2$$

Příklad

Mohou být neslučitelné (disjunktní) jevy A a B nezávislé?

Řešení: $0 = P(\emptyset) = P(A \cap B) = P(A) \cdot P(B)$

tedy disjunktní jevy mohou být nezávislé, jen když alespoň jeden z nich má nulovou pravděpodobnost.

Stochasticky nezávislé jevy

Příklad: Zjistěte, zda existuje jev, který je stochasticky nezávislý sám se sebou.
Řešení: $P(A \cap A) = P(A)P(A)$, tedy $P(A) = P(A)^2$. To je možné jen tak, že $P(A) = 0$ nebo $P(A) = 1$.

Věta:
a) Jestliže z třídy n stochasticky nezávislých jevů vybereme libovolnou podtřídu r jevů ($2 \le r \le n$), dostaneme opět třídu stochasticky nezávislých jevů.
b) Stochastická nezávislost se neporuší, jestliže některé (nebo i všechny) jevy nahradíme jevy opačnými.
c) Jestliže z třídy n stochasticky nezávislých jevů vybereme r disjunktních podtříd jevů ($2 \le r \le n$) a členy těchto podtříd libovolně sjednotíme nebo pronikneme, pak vzniklá sjednocení a průniky jsou opět stochasticky nezávislé jevy.
d) Neslučitelné jevy nemohou být stochasticky nezávislé (pokud nemají všechny nulovou pravděpodobnost).
e) Nemožný jev je stochasticky nezávislý s každým jevem.
f) Jistý jev je stochasticky nezávislý s každým jevem.

Příklad

Příklad: Firma investovala do tří nezávislých projektů. Pravděpodobnost zisku z těchto projektů je 0,4, 0,5 a 0,7. Jaká je pravděpodobnost, že firma bude mít zisk

a) právě jedenkrát (jev A)
b) alespoň jedenkrát (jev B)
c) právě dvakrát (jev C)
d) aspoň dvakrát (jev D)
e) ze všech tří projektů (jev E)
f) ze žádného projektu? (jev F)

Řešení:

Označme A_i jev, že firma bude mít zisk z i-tého projektu, i = 1, 2, 3.

ad a)

$$P(A) = P(A_1)\cdot P(\overline{A_2})\cdot P(\overline{A_3}) + P(\overline{A_1})\cdot P(A_2)\cdot P(\overline{A_3}) + P(\overline{A_1})\cdot P(\overline{A_2})\cdot P(A_3) =$$

$$= 0{,}4\cdot 0{,}5\cdot 0{,}3 + 0{,}6\cdot 0{,}5\cdot 0{,}3 + 0{,}6\cdot 0{,}5\cdot 0{,}7 = 10^{-3}(60 + 90 + 210) = 0{,}36$$

ad b)

$$P(B) = 1 - P(\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}) = 1 - P(\overline{A_1})\,P(\overline{A_2})\,P(\overline{A_3}) = 1 - 0{,}6\cdot 0{,}5\cdot 0{,}3 = 1 - 0{,}09 = 0{,}91$$

ad c)

$$P(C) = P(A_1)\cdot P(A_2)\cdot P(\overline{A_3}) + P(A_1)\cdot P(\overline{A_2})\cdot P(A_3) + P(\overline{A_1})\cdot P(A_2)\cdot P(A_3) =$$

$$= 0{,}4\cdot 0{,}5\cdot 0{,}3 + 0{,}4\cdot 0{,}5\cdot 0{,}7 + 0{,}6\cdot 0{,}5\cdot 0{,}7 = 10^{-3}(60 + 140 + 210) = 0{,}41$$

ad d)

$$P(D) = P(C) + P(A_1\cap A_2\cap A_3) = P(C) + P(A_1)\cdot P(A_2)\cdot P(A_3) = 0{,}41 + 0{,}4\cdot 0{,}5\cdot 0{,}7 =$$

$$= 0{,}41 + 0{,}14 = 0{,}55$$

ad e)

$$P(E) = P(A_1\cap A_2\cap A_3) = P(A_1)\cdot P(A_2)\cdot P(A_3) = 0{,}4\cdot 0{,}5\cdot 0{,}7 = 0{,}14$$

ad f)

$$P(F) = P(\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}) = P(\overline{A_1})\,P(\overline{A_2})\,P(\overline{A_3}) = 0{,}6\cdot 0{,}5\cdot 0{,}3 = 0{,}09$$

5. Podmíněná pravděpodobnost. Geometrická pravděpodobnost.

Podmíněná pravděpodobnost:

Motivace: Opakovaně nezávisle provádíme týž náhodný pokus a sledujeme nastoupení jevu A v těch pokusech, v nichž nastoupil jev H. Podmíněnou relativní četnost A za podmínky H jsme v popisné statistice zavedli vztahem $p(A/H) = \frac{p(A \cap H)}{p(H)}$. Tato podmíněná relativní četnost se s rostoucím počtem pokusů ustaluje kolem konstanty P(A/H), kterou považujeme za podmíněnou pravděpodobnost jevu A za podmínky H.

Definice: Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $H \in \mathcal{A}$. jev s nenulovou pravděpodobností. Podmíněnou pravděpodobností za podmínky H rozumíme funkci $P(./H)\colon \mathcal{A} \to R$ danou vzorcem: $\forall A \in \mathcal{A}: \boxed{P(A/H) = \frac{P(A \cap H)}{P(H)}.}$

Podmíněná pravděpodobnost

Věta: Podmíněná pravděpodobnost je pravděpodobnost ve smyslu axiomatické definice a kromě toho pro ni platí:

a) $P(A_1 \cap A_2) = P(A_1)\,P(A_2/A_1)$ pro $P(A_1) \neq 0$.

b) $P(A_1 \cap A_2) = P(A_2)\,P(A_1/A_2)$ pro $P(A_2) \neq 0$.

c) Jevy A_1, A_2 jsou stochasticky nezávislé, právě když $P(A_1/A_2) = P(A_1)$ nebo $P(A_2) = 0$ a právě když $P(A_2/A_1) = P(A_2)$ nebo $P(A_1) = 0$.

Důkaz:

Stačí ověřit platnost axiómů P2, P10, P15.

ad a), ad b) Plyne přímo z definičního vzorce.

ad c) Nechť A_1, A_2 jsou stochasticky nezávislé $\Rightarrow P(A_1/A_2) = \dfrac{P(A_1 \cap A_2)}{P(A_2)} = \dfrac{P(A_1)P(A_2)}{P(A_2)} = P(A_1)$.

Nechť naopak $P(A_1/A_2) = P(A_1)$. Z definice: $P(A_1/A_2) = \dfrac{P(A_1 \cap A_2)}{P(A_2)} = P(A_1) \Rightarrow P(A_1 \cap A_2) = P(A_1)P(A_2)$, tedy A_1, A_2 jsou stochasticky nezávislé.

Příklad

Příklad: Jaká je pravděpodobnost, že při hodu kostkou padlo sudé číslo, je-li známo, že padlo číslo menší než 5?

Řešení: $\Omega = \{\omega_1, \dots, \omega_6\}$, A … padlo sudé číslo, $A = \{\omega_2, \omega_4, \omega_6\}$, H … padlo číslo menší než 5, $H = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $A \cap H = \{\omega_2, \omega_4\}$

$$P(A/H) = \frac{P(A \cap H)}{P(H)} = \frac{\frac{2}{6}}{\frac{4}{6}} = \frac{1}{2}$$

Příklad: Dvakrát hodíme kostkou. Jaká je pravděpodobnost, že součet přesáhne 10, víme-li, že padla (aspoň jedna) šestka?

Řešení:

$$P(A|H) = \frac{\frac{|\{[6,5],[5,6],[6,6]\}|}{6 \cdot 6}}{\frac{2 \cdot 5 + 1}{6 \cdot 6}} = \frac{3}{11}$$

Věta o násobení pravděpodobností

Věta: (Věta o násobení pravděpodobností)
Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $A_1, A_2, \ldots, A_n$ takové jevy, že $P(A_1 \cap \ldots \cap A_{n-1}) \neq 0$.
Pak $\boxed{P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cap A_2) \ldots P(A_n/A_1 \cap \ldots \cap A_{n-1})}$.
Důkaz: Matematickou indukcí. Předpokládáme, že vztah platí pro libovolné přirozené $n \geq 2$ a dokážeme jeho platnost pro n+1:

$$P(A_1 \cap \ldots \cap A_n \cap A_{n+1}) = P\left(\bigcap_{i=1}^{n} A_i \cap A_{n+1}\right) = P\left(\bigcap_{i=1}^{n} A_i\right) P\left(A_{n+1} / \bigcap_{i=1}^{n} A_i\right) = P(A_1) P(A_2/A_1) \ldots P(A_{n+1}/A_1 \cap \ldots \cap A_n)$$

Příklad: Ze skupiny 100 výrobků, která obsahuje 10 zmetků, vybereme náhodně bez vracení 3 výrobky. Vypočtěte pravděpodobnost jevu, že první dva výrobky budou kvalitní a třetí bude zmetek.
Řešení:
Jev A_i znamená, že i-tý vybraný výrobek je kvalitní, i = 1, 2, 3.
Počítáme $P(A_1 \cap A_2 \cap \overline{A_3}) = P(A_1)\, P(A_2/A_1)\, P(\overline{A_3}/A_1 \cap A_2) = \frac{90}{100}\frac{89}{99}\frac{10}{98} = 0{,}083$.

Věta o úplné pravděpodobnosti, Bayesův vzorec

Věta (vzorec pro výpočet úplné pravděpodobnosti a Bayesův vzorec)
Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $H_i \in \mathcal{A}$, $i \in I$ (I je nejvýše spočetná indexová množina) takové jevy, že $P(H_i) > 0$, $\bigcup_{i\in I} H_i = \Omega$, $H_i \cap H_j = \emptyset$ pro $i \neq j$ (říkáme, že jevy H_i, $i \in I$ tvoří úplný systém hypotéz).

a) Pro libovolný jev $A \in \mathcal{A}$ platí vzorec úplné pravděpodobnosti: $\boxed{P(A) = \sum_{i\in I} P(H_i)P(A/H_i)}$

b) Pro libovolnou hypotézu H_k, $k \in I$ a jev $A \in \mathcal{A}$ s nenulovou pravděpodobností platí Bayesův vzorec:

$$\boxed{P(H_k/A) = \frac{P(H_k)P(A/H_k)}{P(A)}}.$$

($P(H_k/A)$ se nazývá aposteriorní pravděpodobnost hypotézy H_k, $P(H_k)$ je apriorní pravděpodobnost.)

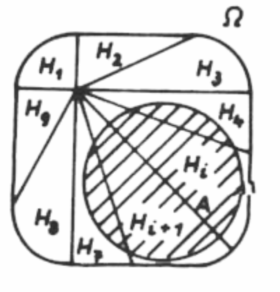
Důkaz:

ad a) Jev A vyjádříme jako sjednocení neslučitelných jevů: $A = \bigcup_{i\in I}(A \cap H_i)$.

Pak $P(A) = P\left(\bigcup_{i\in I}(A\cap H_i)\right) = \sum_{i\in I} P(A\cap H_i) = \sum_{i\in I} P(H_i)P(A/H_i)$

ad b) $P(H_k/A) = \dfrac{P(H_k \cap A)}{P(A)} = \dfrac{P(H_k)P(A/H_k)}{P(A)}$

Ilustrace vzorce pro úplnou pravděpodobnost



Příklad

1) Bez vracení taháme z urny s a černými a b bílými koulemi. Jaká je pravděpodobnost, že ve druhém tahu vytáhneme černou kouli, jestliže v prvním tahu jsme vytáhli kouli bílou?

Řešení:

$$P(A\,|\,H)=\frac{\dfrac{b}{a+b}\cdot\dfrac{a}{a+b-1}}{\dfrac{b}{a+b}\cdot\dfrac{a+b-1}{a+b-1}}=\frac{a}{a+b-1}$$

2) V dostihu zvítězí kůň A (B) s pravděpodobností 0,5 (0,3). Kůň A ztratil na startu příliš a je jisté, že nezvítězí. Jaká je nyní pravděpodobnost, že zvítězí B?

Řešení:

$$P(A\,|\,\overline{H})=\frac{P(A\cap\overline{H})}{P(\overline{H})}=\frac{P(A)}{1-P(H)}=\frac{0{,}3}{0{,}5}=0{,}6$$

Příklad

V první urně je 6 bílých a 2 černé koule, ve druhé jsou 4 bílé a 2 černé koule. Náhodně zvolíme urnu a vytáhneme jednu kouli. Jaká je pravděpodobnost, že bude bílá?

Řešení:

Pravděpodobnost tahu z první (resp. druhé) urny, je 1/2. Označíme-li B =[tah bílé koule], U_i = [tah z urny i], je podle věty o celkové pravděpodobnosti

$$P(B) = P(B \mid U_1) \cdot P(U_1) + P(B \mid U_2) \cdot P(U_2) = \frac{6}{6+2} \cdot \frac{1}{2} + \frac{4}{4+2} \cdot \frac{1}{2} = \frac{17}{24} = 0{,}708$$

Příklad

Automat X vyrobí za směnu dvakrát více výrobku než automat Y. Pravděpodobnost vzniku zmetku je u automatu X 0,02, u Y 0,05. Po skončení směny se výrobky ukládají do jedné bedny. Jaká je pravděpodobnost, že výrobek náhodně vybraný z této bedny není zmetek?

Řešení:

Podle věty o celkové pravděpodobnosti (poměr výrobků v bedně je 2 : 1 ve prospěch automatu X, tj. 2/3 výrobků pochází od X a 1/3 od Y)

$$P(A) = \frac{2}{3} \cdot 0{,}98 + \frac{1}{3} \cdot 0{,}95 = \frac{2{,}91}{3} = 0{,}97$$

Příklad

Mezi 20 střelci jsou 4 výborní, 10 dobrých a 6 průměrných s pravděpodobnostmi zásahu 0,9, 0,7 a 0,5. Jaká je pravděpodobnost, že dva náhodně vybraní střelci oba zasáhnou cíl?

Řešení:

Podle toho, která dvojice bude vybrána

$$P(A) = (0{,}9 \cdot 0{,}9) \cdot \frac{4 \cdot 3}{20 \cdot 19} + (0{,}9 \cdot 0{,}7) \cdot \frac{4 \cdot 10}{20 \cdot 19} + \ldots + (0{,}5 \cdot 0{,}5) \cdot \frac{10 \cdot 9}{20 \cdot 19} = 0{,}46$$

Věta o úplné pravděpodobnosti, Bayesův vzorec

Thomas Bayes (1702 – 1761): Presbytariánský duchovní

Poznámka (Návod na použití vzorce pro výpočet úplné pravděpodobnosti a Bayesova vzorce)
Nejprve podle textu úlohy stanovíme úplný systém hypotéz, tj,. jevy, které se navzájem vylučují a přitom vyčerpávají všechny možnosti.
V úlohách vedoucích na vzorec pro výpočet úplné pravděpodobnosti se zajímáme o pravděpodobnost jevu, který s hypotézami nesouvisí, zatímco v úlohách vedoucích na Bayesův vzorec nás zajímá pravděpodobnost některé hypotézy za podmínky, že nastal jev, který s hypotézami nesouvisí.

Příklad

Příklad: Test obsahuje 100 otázek. Zkoušený si nejprve vylosuje otázku a pak si jeho postup zjednodušeně představíme takto: zná.li správnou odpověď, zatrhne ji. Nezná-li správnou odpověď, zvolí se stejnou pravděpodobností kteroukoliv ze čtyř možných odpovědí. Předpokládejme, že ve skutečnosti zná zkoušený právě k správných odpovědí.

a) S jakou pravděpodobností správně odpoví?

b) S jakou pravděpodobností je při správné odpovědi pravdivé tvrzení, že zkoušený ve skutečnosti jenom hádal?

Řešení: H_1 … zkoušený zná správnou odpověď, H_2 … zkoušený nezná správnou odpověď, A … zkoušený správně odpoví

$$P(H_1)=\frac{k}{100}, P(H_{21})=\frac{100-k}{100}, P(A/H_1)=1, P(A/H_2)=\frac{1}{4}$$

ad a) $P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)=\frac{k}{100}\cdot 1+\frac{100-k}{100}\cdot\frac{1}{4}=\frac{3k+100}{400}$

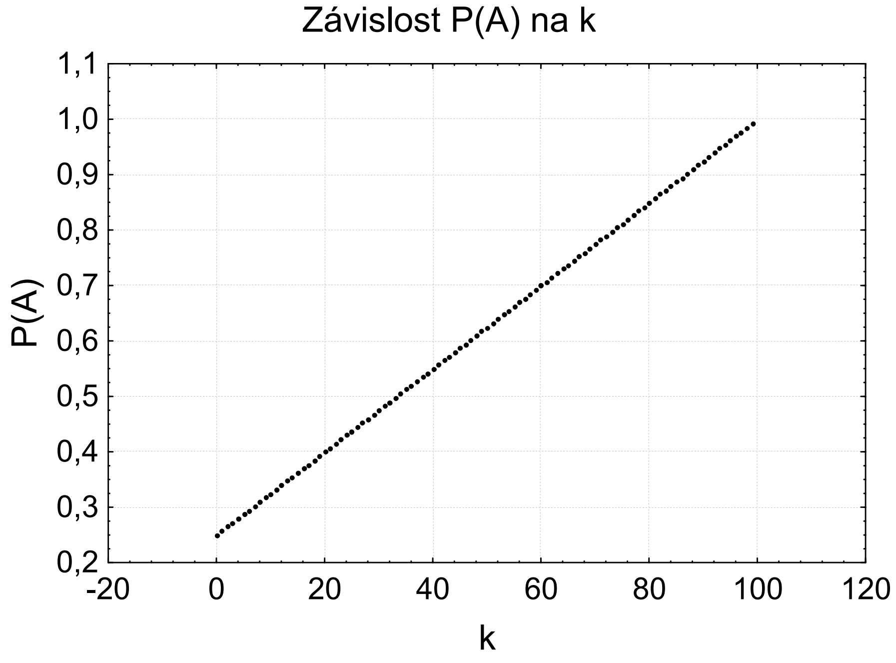
ad b) $P(H_2/A)=\frac{P(H_2)P(A/H_2)}{P(A)}=\frac{\frac{100-k}{100}\cdot\frac{1}{4}}{\frac{3k+100}{400}}=\frac{100-k}{3k+100}$

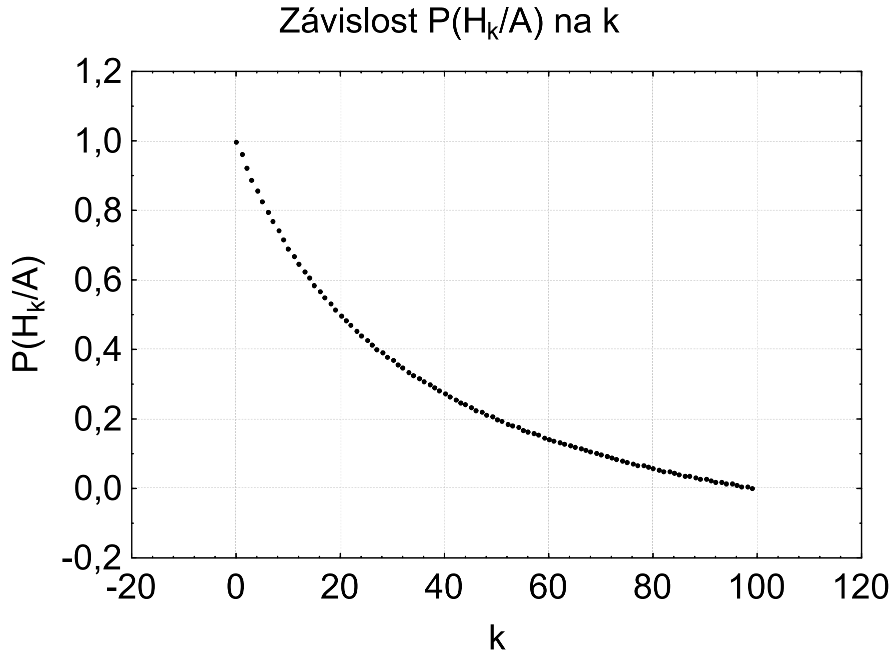
k	0	10	50	90
P(A)	0,25	0,325	0,625	0,925
$P(H_2/A)$	1	0,692	0,2	0,027

Příklad

$$P(A) = \frac{3k + 100}{400}$$

$$P(H_2 / A) = \frac{100 - k}{3k + 100}$$





Příklad

Příklad: K osevu byly vybrány dvě odrůdy pšenice, a to 20% první odrůdy a 80% druhé odrůdy. Pravděpodobnost, že ze zrna vyroste klas, je pro první odrůdu 0,95 a pro druhou odrůdu 0,98. Jaká je pravděpodobnost, že

a) z náhodně vybraného zrna vyroste klas?

b) náhodně vybrané zrno, z něhož vyrostl klas, pocházelo z první odrůdy pšenice?

c) náhodně vybrané zrno, z něhož vyrostl klas, pocházelo z druhé odrůdy pšenice?

d) náhodně vybrané zrno, z něhož nevyrostl klas, pocházelo z první odrůdy pšenice?

e) náhodně vybrané zrno, z něhož nevyrostl klas, pocházelo z druhé odrůdy pšenice?

Řešení:

Jev A … z náhodně vybraného zrna vyroste klas

Jev H_1 … zrno pochází z první odrůdy pšenice

Jev H_2 … zrno pochází z druhé odrůdy pšenice

$P(H_1) = 0{,}2,\ P(A|H_1) = 0{,}95,\ P(H_2) = 0{,}8,\ P(A|H_2) = 0{,}98$

ad a) $P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = 0{,}2 \cdot 0{,}95 + 0{,}8 \cdot 0{,}98 = 0{,}19 + 0{,}784 = 0{,}974$

ad b) $P(H_1/A) = \dfrac{P(H_1)P(A/H_1)}{P(A)} = \dfrac{0{,}2 \cdot 0{,}95}{0{,}974} = 0{,}1951$

ad c) $P(H_2/A) = \dfrac{P(H_2)P(A/H_2)}{P(A)} = \dfrac{0{,}8 \cdot 0{,}98}{0{,}974} = 0{,}8049$

ad d) $P(H_1/\overline{A}) = \dfrac{P(H_1)P(\overline{A}/H_1)}{P(\overline{A})} = \dfrac{0{,}2 \cdot 0{,}05}{1 - 0{,}974} = \dfrac{0{,}01}{0{,}026} = 0{,}3846$

ad e) $P(H_2/\overline{A}) = \dfrac{P(H_2)P(\overline{A}/H_2)}{P(\overline{A})} = \dfrac{0{,}8 \cdot 0{,}02}{1 - 0{,}974} = \dfrac{0{,}016}{0{,}026} = 0{,}6154$

Příklad

1) Jeden ze 3 střelců s pravděpodobnostmi zásahu 0,3, 0,5, 0,8 vystřelil a zasáhl. Jaká je pravděpodobnost, že střílel druhý střelec?

Řešení:

$$P(A) = \frac{0{,}5 \cdot \frac{1}{3}}{0{,}3 \cdot \frac{1}{3} + 0{,}5 \cdot \frac{1}{3} + 0{,}8 \cdot \frac{1}{3}} = \frac{5}{16} = 0{,}3125$$

2) Mezi 20 střelci je 5 výborných, 9 dobrých a 6 průměrných s pravděpodobnostmi zásahu 0,9, 0,8 a 0,7. Náhodně vybraný střelec ze 2 ran trefil jednou. Jaká je pravděpodobnost, že šlo o výborného (dobrého, průměrného) střelce?

Řešení:

$$P(\text{byl to výborný}) = \frac{2 \cdot 0{,}9 \cdot 0{,}1 \cdot \frac{5}{20}}{2 \cdot 0{,}9 \cdot 0{,}1 \cdot \frac{5}{20} + 2 \cdot 0{,}8 \cdot 0{,}2 \cdot \frac{9}{20} + 2 \cdot 0{,}7 \cdot 0{,}3 \cdot \frac{6}{20}} = 0{,}143$$

$$P(\text{byl to dobrý}) = 0{,}457$$

$$P(\text{byl to průměrný}) = 0{,}4$$

Příklad

Víme-li, že pravděpodobnost odhalení AIDS při testu je 0,999, že pravděpodobnost správného otestování zdravého jedince je 0,99 a že AIDS se vyskytuje u 0,006 lidí, jaká je pravděpodobnost, že člověk, u kterého byl test pozitivní, AIDS skutečně má?

Řešení:

Označíme-li A = [má AIDS], T = [test říká AIDS], známe P(T |A) = 0, 999, $P(\overline{T} \mid \overline{A}) = 0{,}99$, P(A) = 0, 006. Bayesova věta nám dá

$$P(A \mid T) = \frac{P(T \mid A)P(A)}{P(T \mid A)P(A) + P(T \mid \overline{A})P(\overline{A})}$$

$$= \frac{0{,}999 \cdot 0{,}006}{0{,}999 \cdot 0{,}006 + (1 - 0{,}99) \cdot (1 - 0{,}006)} = 0{,}376$$

Geometrická pravděpodobnost

Motivace: V některých situacích je vhodné zvolit za základní prostor nikoliv obecnou množinu Ω, ale n-rozměrný prostor $\mathbf{R}^n$ a za možné výsledky reálné vektory $(x_1,\ldots,x_n)$. Za jevové pole však nevezmeme systém všech podmnožin prostoru $\mathbf{R}^n$ (ten totiž obsahuje i tzv. neměřitelné množiny), ale méně podrobné borelovské pole $\mathcal{B}^n$.

Émile Borel (1871 – 1956) – francouzský matematik a politik. Zabýval se teorií míry, teorií pravděpodobnosti A teorií her. Byl poslancem francouzského parlamentu a ministrem námořnictva.

Na borelovském poli pak speciálním způsobem zavedeme geometrickou pravděpodobnost a dostaneme pravděpodobnostní prostor $(\mathbf{R}^n, \mathcal{B}^n, Q)$.

Borelovské pole, Borelovské množiny

Definice

Nechť n je přirozené číslo. Množinu $R^n = (-\infty, \infty) \times \ldots \times (-\infty, \infty) = (-\infty, \infty)^n$ nazýváme n-**rozměrným prostorem**. Minimální jevové pole na R^n obsahující třídu všech polouzavřených intervalů typu $(-\infty, x_1\rangle \times \ldots \times (-\infty, x_n\rangle$ pro $(x_1, \ldots x_n) \in R^n$ nazýváme n-**rozměrným borelovským polem** $\mathcal{B}^n$ a prvky tohoto pole nazýváme **(**n**-rozměrnými) borelovskými množinami.** Dvojice $(R^n, \mathcal{B}^n)$ je tedy měřitelný prostor.

(Není podstatné, že borelovské pole je generováno právě intervaly typu $(-\infty, x_1\rangle \times \ldots \times (-\infty, x_n\rangle$. Mohlo by být generováno i jinými typy intervalů.)

Věta: Borelovské pole je jevové pole, tzn., že splňuje axiómy J2, J6, J8.

Věta: Mezi borelovské množiny náleží zejména prázdná množina, celý základní prostor, všechny jednobodové, konečné a spočetné množiny, intervaly všech typů, všechny uzavřené a otevřené oblasti a všechna konečná a spočetná sjednocení a průniky těchto množin. Rovněž kartézský součin borelovských množin je borelovská množina, ovšem vyšší dimenze.

Borelovsky měřitelná zobrazení, Borelovské funkce

Definice

Nechť $(\Omega, \mathcal{A}), (R^n, \mathcal{B}^n)$ jsou měřitelné prostory. Zobrazení $\mathbf{X} : \Omega \mapsto R^n$ se nazývá **borelovsky měřitelné** (vzhledem k $\mathcal{A}$), právě když úplný vzor každé n-rozměrné borelovské množiny je jev, tj.

$\forall B \in \mathcal{B}^n : \mathbf{X}^{inv}(B) = \{\omega \in \Omega; X(\omega) \in B\} \in \mathcal{A}$.

Ve speciálním případě, kdy $\Omega = R^m$ a $\mathcal{A} = \mathcal{B}^m, \mathbf{X} = \mathbf{g} = (g_1, \ldots, g_n)$, tj.

$\forall B \in \mathcal{B}^n : \mathbf{g}^{inv}(B) =$

$\{(x_1, \ldots, x_m) \in R^m; (g_1(x_1, \ldots, x_m), \ldots, g_n(x_1, \ldots, x_m)) \in B\} \in \mathcal{B}^m$, hovoříme o **borelovské funkci**.

Věta: Mezi borelovské funkce náleží zejména všechny spojité a po částech spojité funkce. Rovněž limita všude konvergentní posloupnosti borelovských funkcí je borelovská funkce.

Definice:

Nechť $(R^n, \mathcal{B}^n)$ je měřitelný prostor a $G \in \mathcal{B}^n$ je borelovská množina. **Objemem** borelovské množiny G rozumíme číslo

$mes(G) = \int_G \ldots \int dx_1 \ldots dx_n$, pokud Riemannův integrál vpravo existuje.

Geometrická pravděpodobnost

Definice:

Nechť objem $mes(G)$ borelovské množiny G je nenulový a konečný. **Geometrickou pravděpodobností** soustředěnou na množině G rozumíme funkci $Q : \mathcal{B}^n \mapsto R$ danou vzorcem

$$\forall B \in \mathcal{B}^n, B \subseteq G : \boxed{Q(B) = \frac{mes(B)}{mes(G)}}, \text{ pokud } mes(B) \text{ existuje.}$$

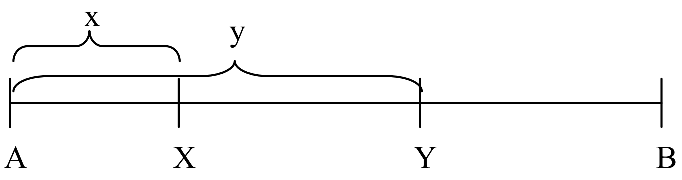
Věta: Geometrická pravděpodobnost je pravděpodobnost ve smyslu axiomatické definice, tj. splňuje axiómy P2, P10, P15. Trojice ($\mathbf{R}^n$, $\mathcal{B}^n$, Q) je tedy pravděpodobnostní prostor.

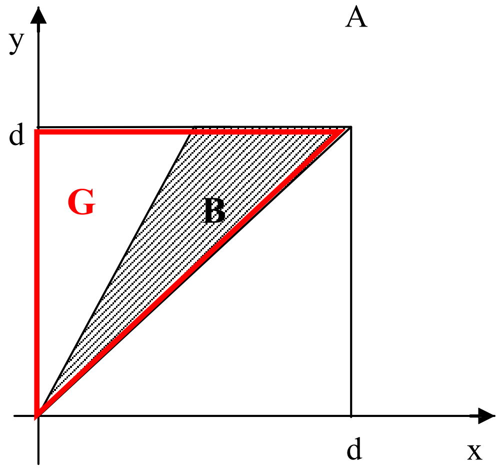
Příklad

Příklad: Na úsečce AB délky d jsou náhodně zvoleny body X a Y, přičemž vzdálenost bodu X od bodu A je menší než vzdálenost bodu Y od bodu A. Jaká je pravděpodobnost, že délka úsečky AX je větší než délka úsečky XY?

Řešení:

$G = \{(x, y) \in R^2; 0 \le x \le d, 0 \le y \le d, x \le y\}$ $\qquad$ $B = \{(x, y) \in G; x > y - x\}$





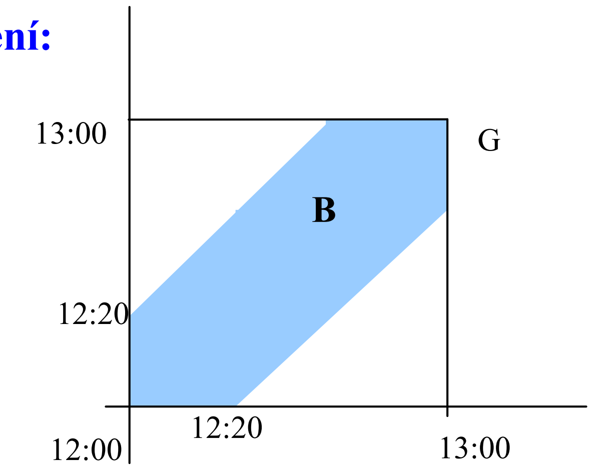
$$\text{mes}(G) = \frac{d^2}{2}, \ \text{mes}(B) = \frac{d^2}{2} - \frac{\frac{d}{2} \cdot d}{2} = \frac{d^2}{4}, \ Q(B) = \frac{\text{mes}(B)}{\text{mes}(G)} = \frac{1}{2}$$

Délka úsečky AX je větší než délka úsečky XY s pravděpodobností 0,5.

Příklad

Dívka a chlapec si smluvili schůzku mezi 12:00 a 13:00. Přijdou náhodně v tomto rozmezí a čekají na sebe 20 minut, nejdéle však do 13:00. Jaká je pravděpodobnost, že se setkají?

Řešení:



$$mes(G) = 1$$

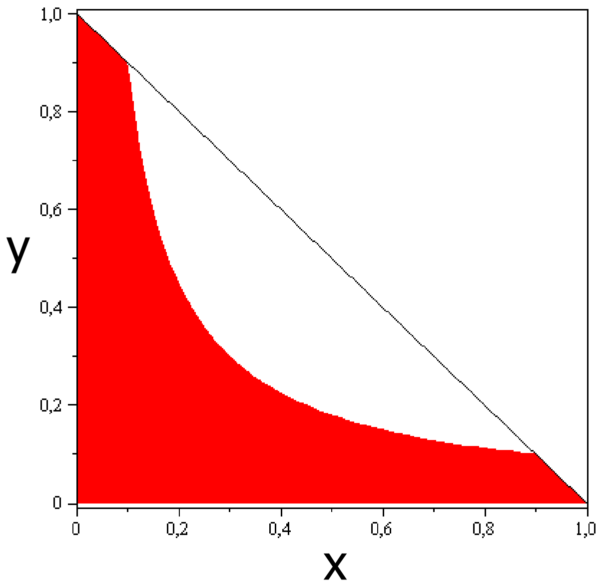
$$mes(B) = \frac{5}{9}$$

$$\Rightarrow \quad Q(B) = \frac{5}{9}$$

Příklad

Volíme náhodně dvě čísla z intervalu (0,1). Jaká je pravděpodobnost, že jejich součet je menší než jedna a současně jejich součin menší než 0,09?

Řešení



$$1 - x = \frac{0{,}09}{x}$$

$$x^2 - x + 0{,}09 = 0$$

$$x_1 = 0{,}1,\ x_2 = 0{,}9$$

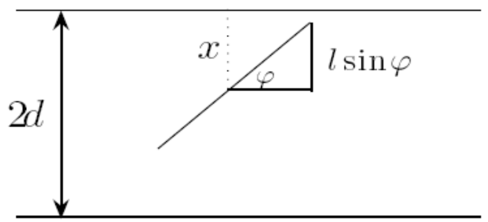
$$Q(B) = \frac{1}{2} - \int_{0{,}1}^{0{,}9} \left(1 - x - \frac{0{,}09}{x}\right) dx = 0{,}29775$$

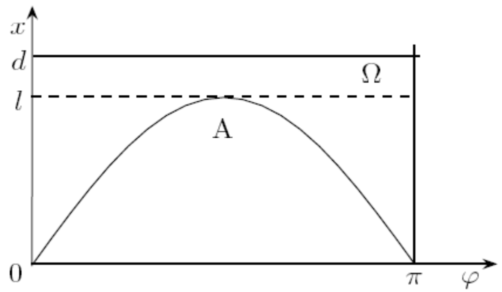
Příklad

Buffonova úloha. V rovině jsou rozmístěny rovnoběžky ve vzdálenosti d > 0. Na rovinu hodíme náhodně jehlu délky 0 < l < d. Jaká je pravděpodobnost, že jehla protne některou rovnoběžku?

Řešení

Předpokládejme, že náhodně znamená, že každá poloha (středu) a každá orientace jehly je stejně pravděpodobná a že tyto dvě nahodile proměnné jsou na sobě nezávislé. Nechť *x* je vzdálenost středu jehly od nejbližší rovnoběžky a φ je úhel, který jehla svírá s rovnoběžkami.

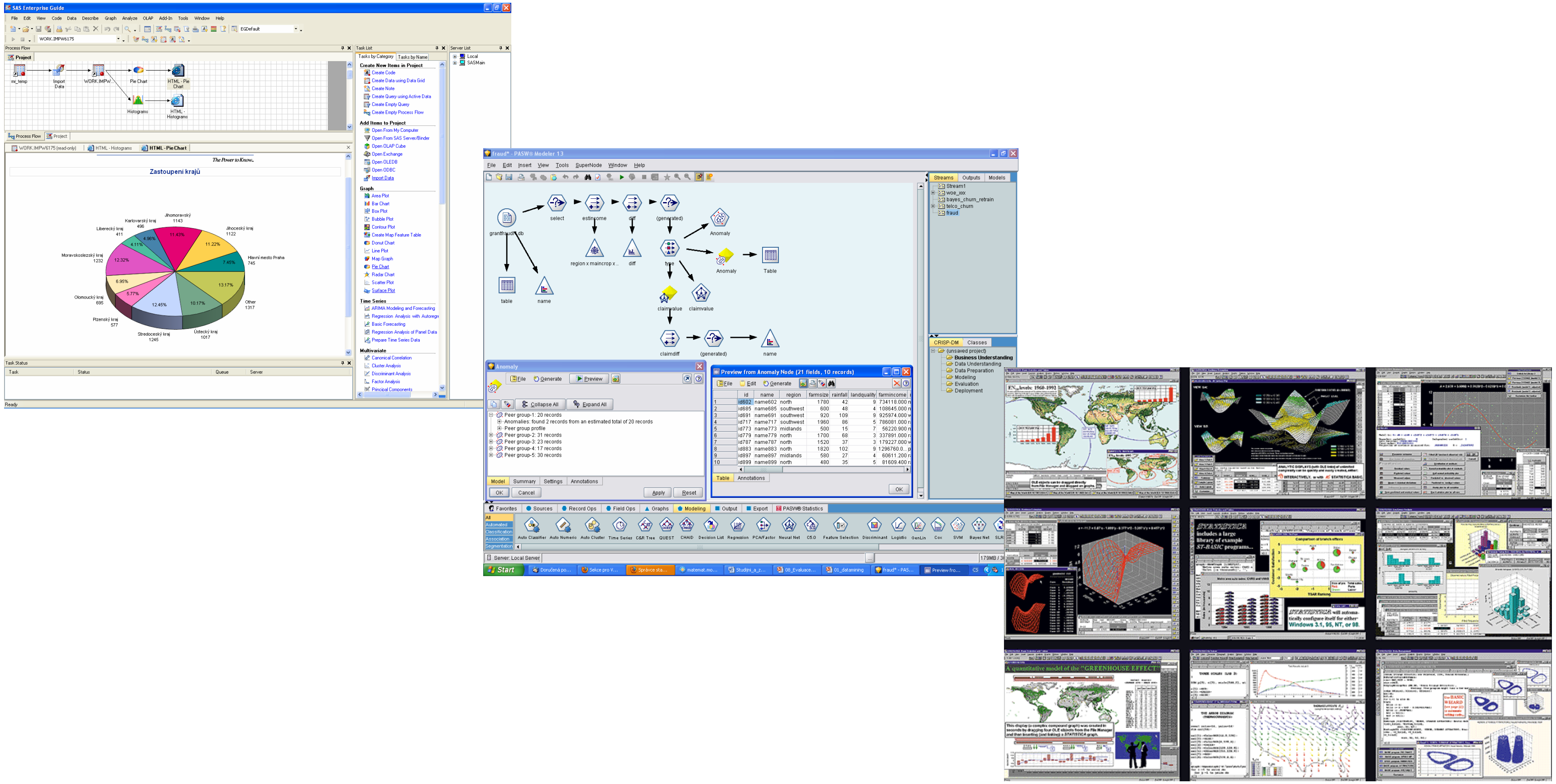




$$\Omega = \{0 \le \varphi < \pi, 0 \le x \le d\}$$

$$A = \{(\varphi, x) \in \Omega : x \le l \sin\varphi\} \qquad Q(A) = \frac{mes(A)}{mes(\Omega)} = \frac{\int_0^\pi l \sin\varphi \, d\varphi}{\pi d} = \frac{2l}{\pi d}$$

6. Statistický software. Vizualizace dat.



Software

AcaStat	GAUSS	MRDCL	RATS	StatsDirect
ADaMSoft	GAUSS	NCSS	RKWard[4]	Statistix
Analyse-it	GenStat	OpenEpi	SalStat	SYSTAT
ASReml	Golden Helix	Origin	SAS	The Unscrambler
Auguri	gretl	Ox programming language	SOCR	UNISTAT
BioStat	JMP	OxMetrics	Stata	VisualStat
BrightStat	MacAnova	Origin	Statgraphics	Winpepi
Dataplot	Mathematica	Partek	STATISTICA	WinSPC
EasyReg	Matlab	Primer	StatIt	XLStat
Epi Info	MedCalc	PSPP	StatPlus	XploRe
EViews	modelQED	R	SPlus	
Excel	Minitab	R Commander[4]	SPSS	

Data miningový software

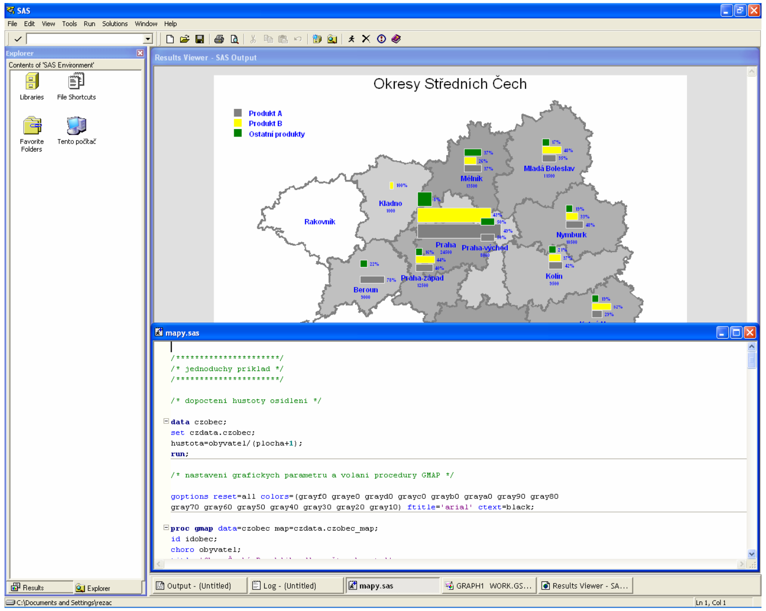
- Cca 20 až 30 dodavatelů
- Hlavní hráči na trhu:
  - Clementine,
  - IBM's Intelligent Miner,
  - SGI's MineSet,
  - SAS's Enterprise Miner.

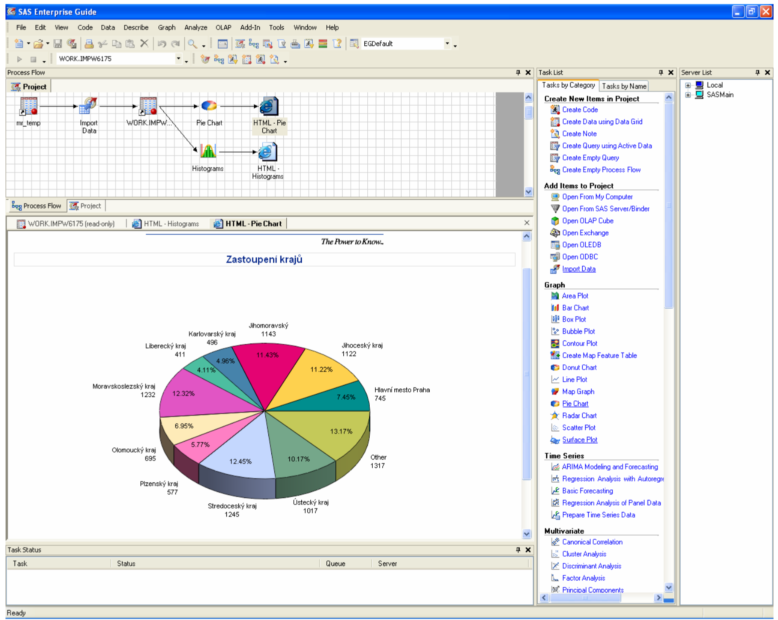
IBM SPSS Modeler (PASW Modeler)

- Řada vestavěných produktů:
  - fraud detection:
  - electronic commerce applications,
  - health care,
  - customer relationship management

Software - SAS

- sas : www.sas.com





SAS

- Společnost SAS Institute
  - Vznik 1976 v univerzitním prostředí
  - Dnes: největší soukromá softwarová společnost na světě (více než 11.000 zaměstnanců)
  - přes 45.000 instalací
  - cca 9 milionů uživatelů ve 118 zemích
  - v USA okolo 1.000 akademických zákazníků (SAS používá většina vyšších a vysokých škol a výzkumných pracovišť)

SAS

Studentské soutěže o hodnotné ceny

- Lokální (ČR) soutěž o nejlepší studentskou práci
- SAS Student Ambassador - celosvětová soutěž o nejlepší práce s využitím SAS
- Možnost účasti a prezentace na SAS konferenci v Seattlu

SAS

Podpora studentů

- Možnost rozšíření licence na domácí instalace pro studenty
- SAS Fellowship Program – software zdarma pro diplomku či dizertaci
- Zadávání a vedení diplomových prací
- Sdílení informací, zkušeností či příkladů v uživatelských skupinách

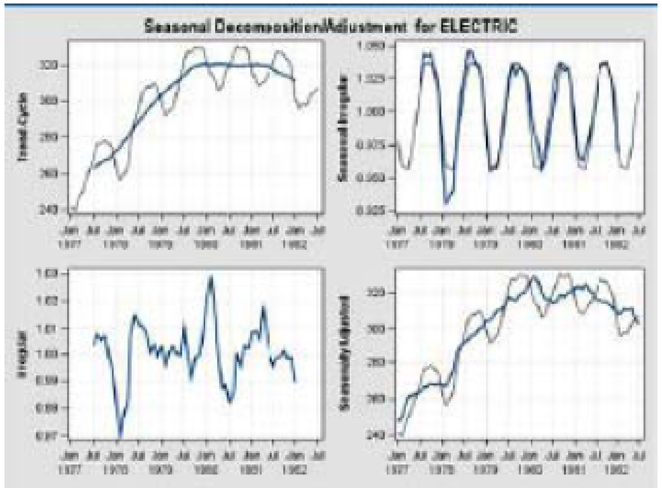
- Interaktivní moduly nebo programovací prostředí
  - Statistická analýza
  - Matice
  - Časové řady
  - Operační výzkum
  - Kontrola kvality

SAS

- Statistická analýza:
  - Popisná statistika
  - Analýza kontingenčních (frekvenčních) tabulek
  - Regresní, korelační, kovarianční analýza
  - Logistická regrese
  - Analýza rozptylu
  - Testování hypotéz
  - Diskriminační analýza
  - Shluková analýza
  - Analýza přežití
  - …

SAS

- Analýza časových řad:
  - Regresní modely
  - Modely se sezónními faktory
  - Autoregresní modely
  - ARIMA
  - Metody exponenciálního vyrovnání
  - …



SAS

- více o SASu: http://www.sas.com/offices/europe/czech/

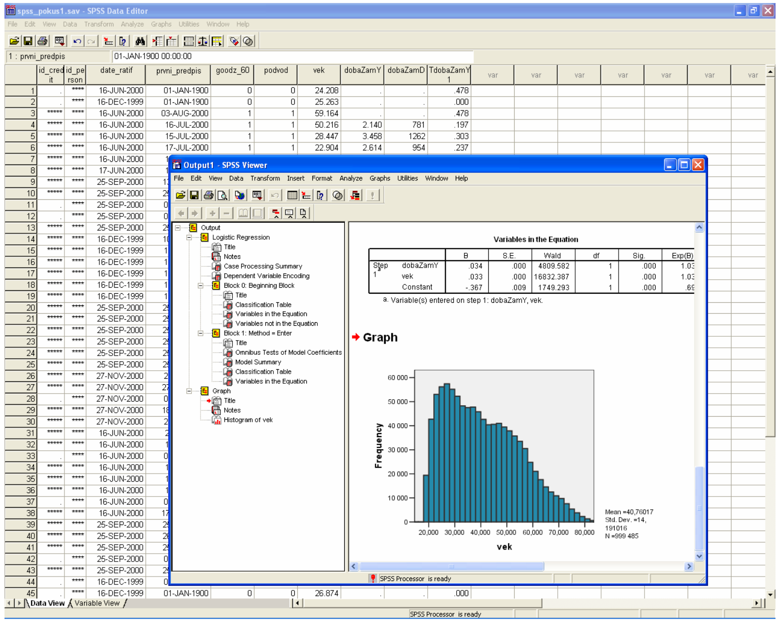
- (neúplný) seznam komerčních společností využívající SAS: http://www.sas.com/offices/europe/czech/reference/list.html

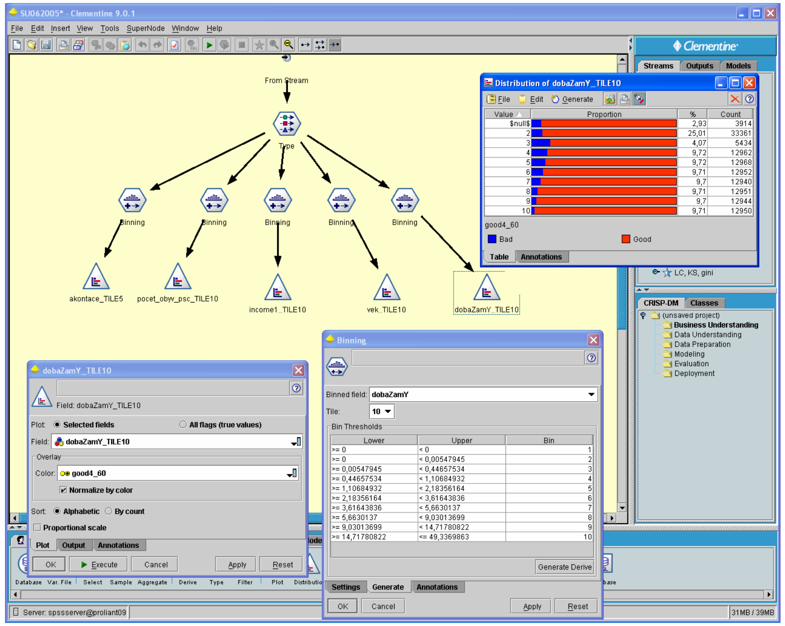
- o akademickém programu: http://www.sas.com/offices/europe/czech/academic/index.html

- o konferenci SAS forum: http://www.sas.com/reg/offer/cz/2010_sas_forum_2010

Software -SPSS

- : www.spss.cz

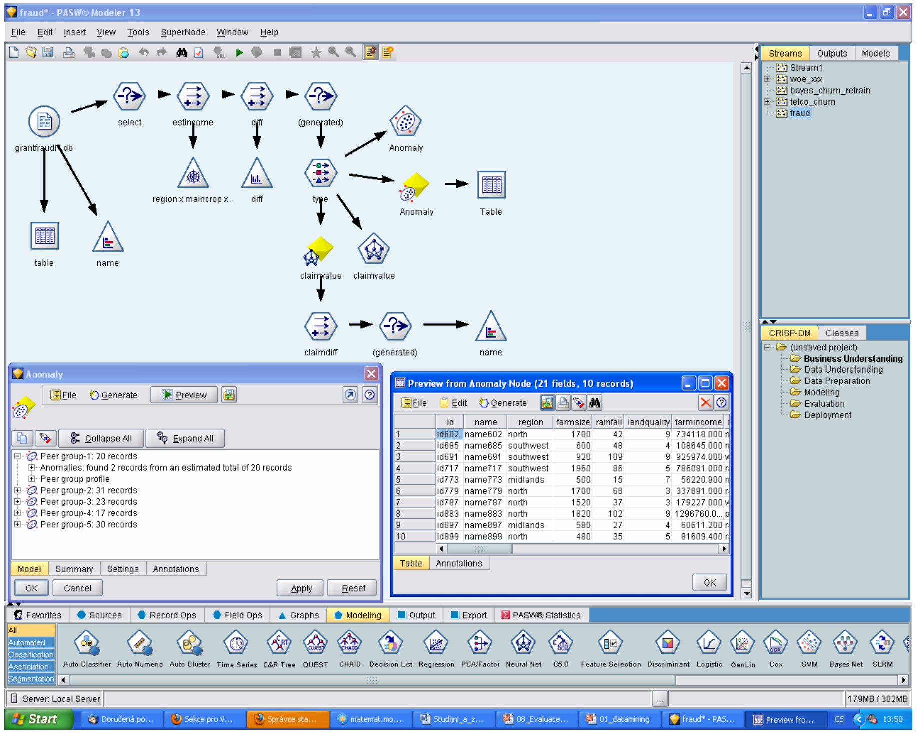


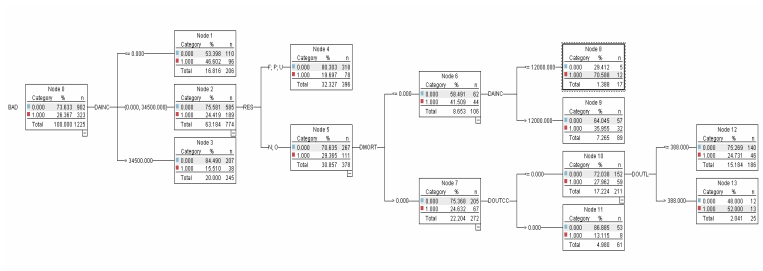


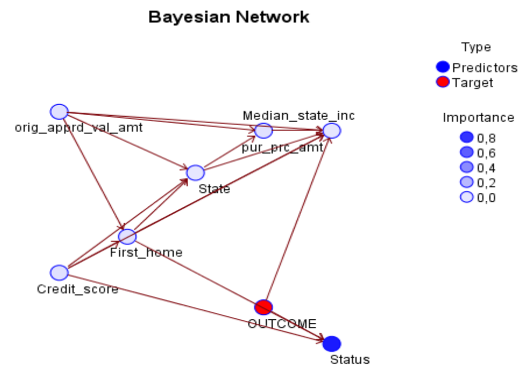
SPSS

- IBM SPSS/ PASW Modeler 13 (dříve Clementine)

  http://www.spss.cz/ibmspss_modeler.htm





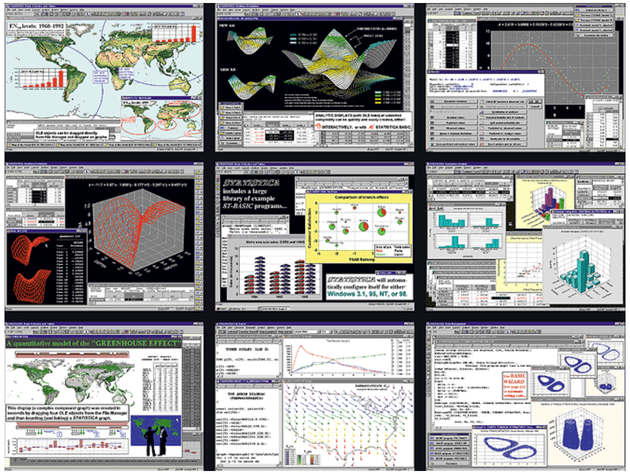


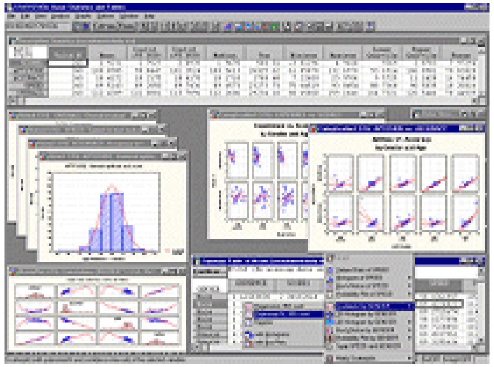
SPSS

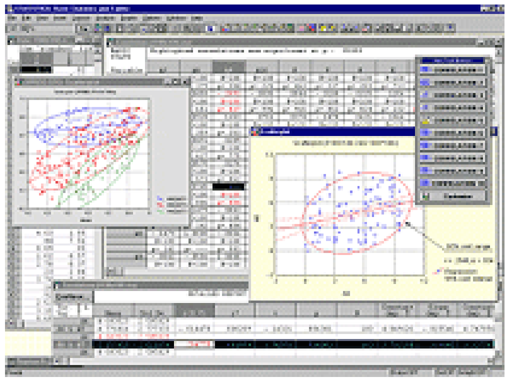
- více o IBM SPSS Modeler 13 (dříve Clementine): http://www.spss.cz/ibmspss_modeler.htm
- (neúplný) seznam zákazníků: http://www.spss.cz/zakaznici.htm
- akademický program: http://www.spss.com/academic/

Software -Statistica

- StatSoft Statistica: www.statistica.cz

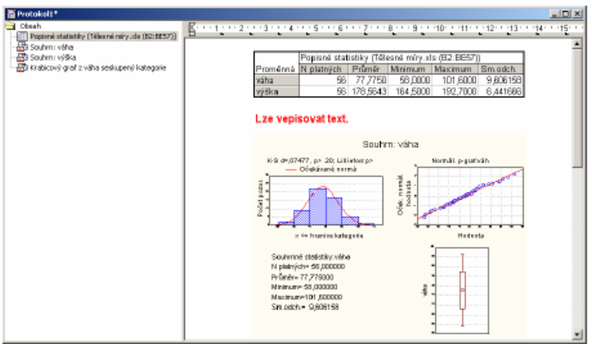






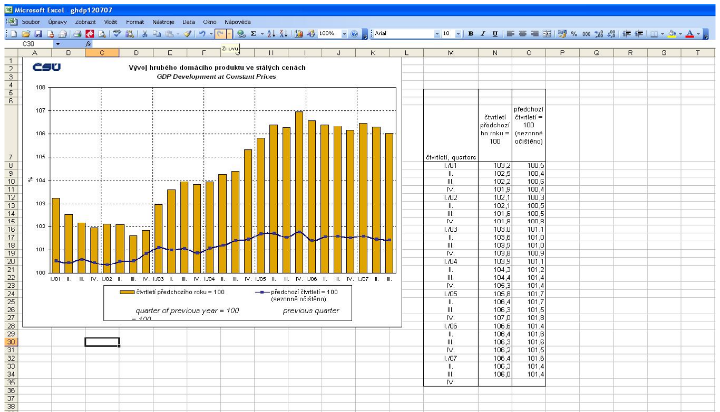
Statistica

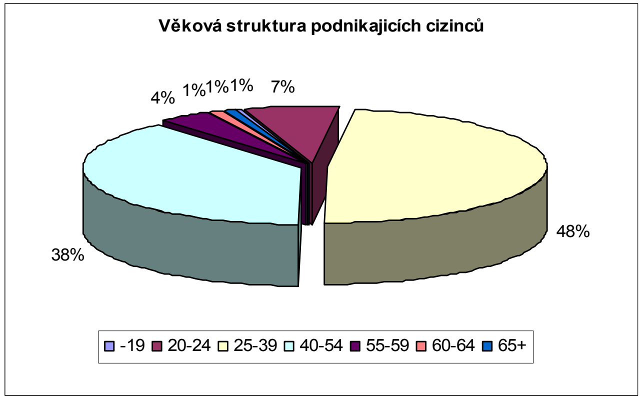
- více o Statistica Data Miner: http://www.statistica.cz/produkty/5-dataminingove-nastroje/21-statistica-data-miner/detail/
- (neúplný) seznam zákazníků: http://www.statsoft.com/customers/
- akademický program: http://www.statsoft.com/academic/
- Petra Beranová – stručný manuál k ovládání programu STATISTICA: http://www.statsoft.cz/download/soubory/STATISTICA_manual.pdf

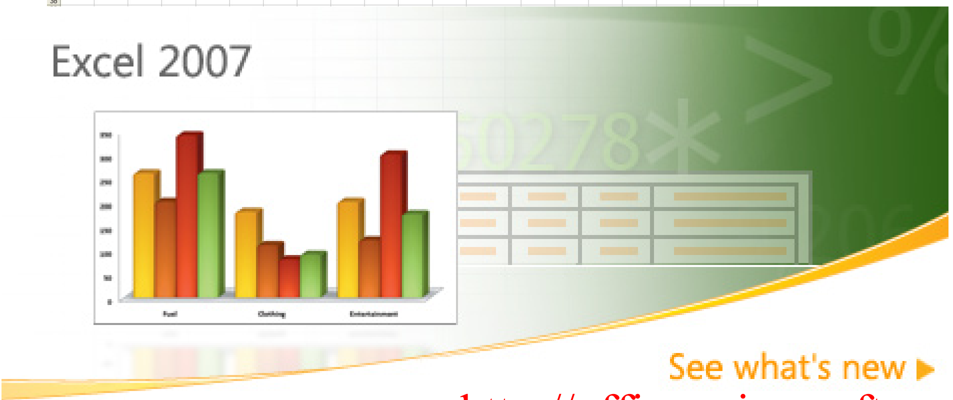


Software

- MS Excel: http://office.microsoft.com/en-us/excel/default.aspx







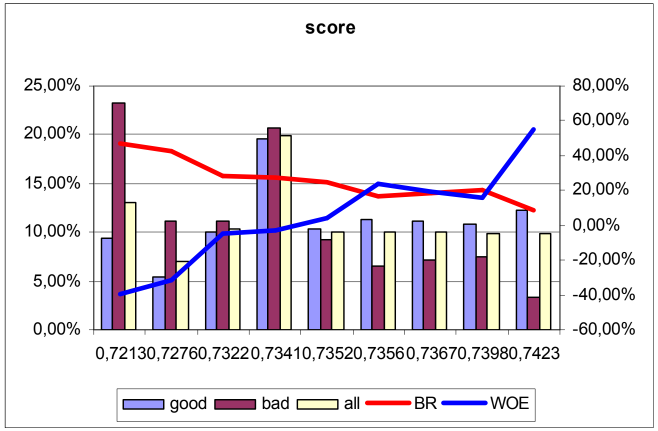
http://office.microsoft.com/en-us/excel/HA100738731033.aspx

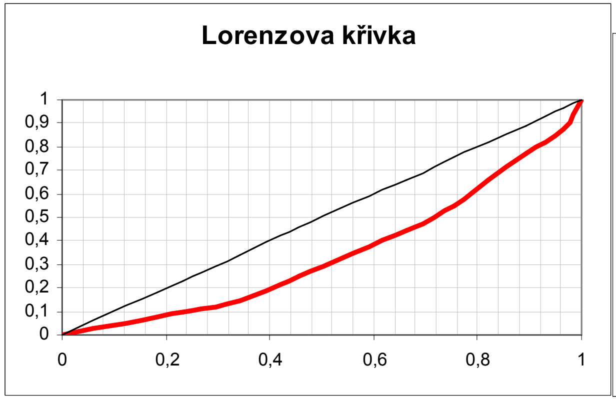
Software

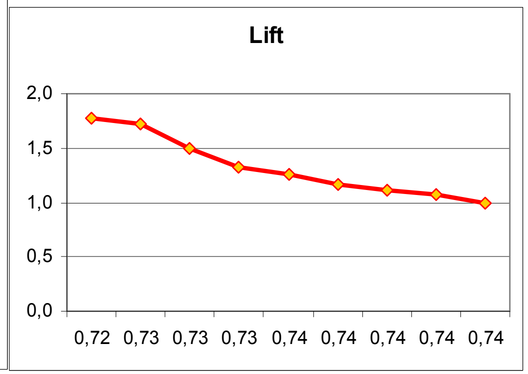
- MS Excel:

Počet z id	BAD		
score k	,00	1,00	Celkový součet
,72	9,42%	23,22%	13,06%
,73	5,43%	11,15%	6,94%
,73	9,98%	11,15%	10,29%
,73	19,51%	20,74%	19,84%
,74	10,31%	9,29%	10,04%
,74	11,31%	6,50%	10,04%
,74	11,09%	7,12%	10,04%
,74	10,75%	7,43%	9,88%
,74	12,20%	3,41%	9,88%
Celkový součet	100,00%	100,00%	100,00%

	good	bad	all	BR	WOE
0,721263	9,42%	23,22%	13,06%	46,88%	-0,392
0,727551	5,43%	11,15%	6,94%	42,35%	-0,312
0,732201	9,98%	11,15%	10,29%	28,57%	-0,048
0,734083	19,51%	20,74%	19,84%	27,57%	-0,027
0,735168	10,31%	9,29%	10,04%	24,39%	0,045
0,735632	11,31%	6,50%	10,04%	17,07%	0,240
0,736706	11,09%	7,12%	10,04%	18,70%	0,192
0,739753	10,75%	7,43%	9,88%	19,83%	0,161
0,742267	12,20%	3,41%	9,88%	9,09%	0,554

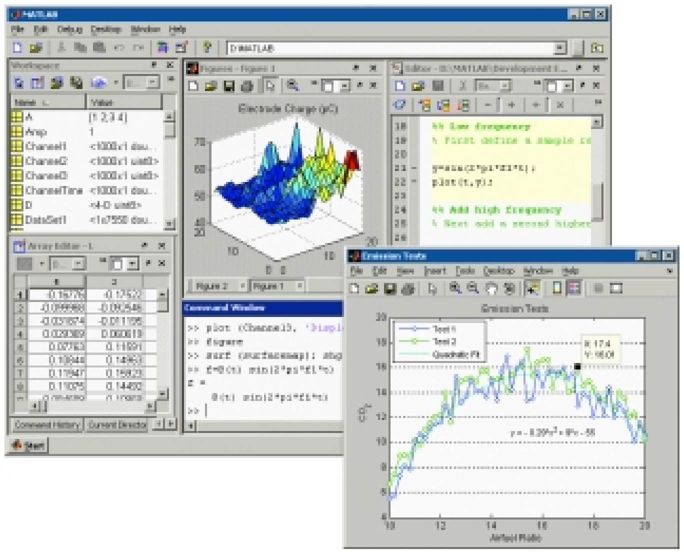


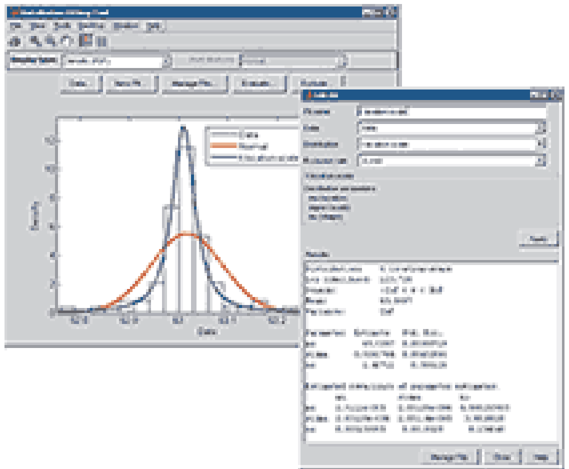




Software

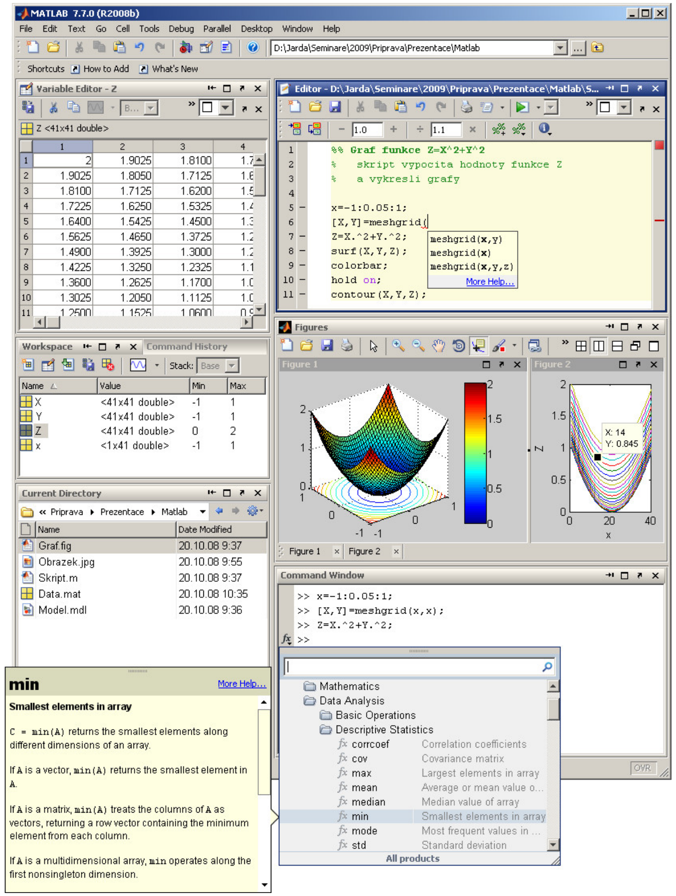
- Matlab : www.mathworks.com,
  www.humusoft.cz



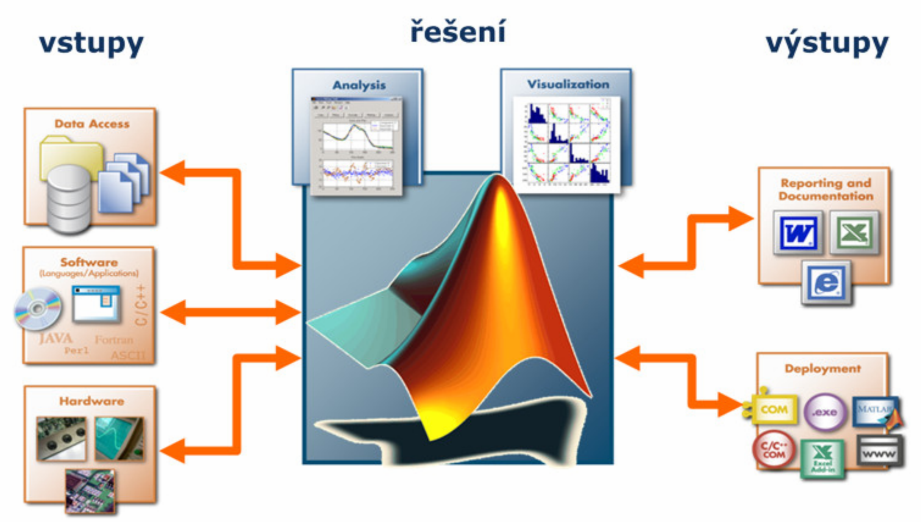


Software

- Matlab :



http://www.humusoft.cz/produkty/matlab/matlab/



Software - MU

- https://inet.muni.cz/app/soft/licence

Nabídka softwaru

Aplikace je určená pro registraci softwaru a následné získání přístupu k instalačním klíčům a dalším informacím (popř. přístup k samotnému softwaru). Přihlášený uživatel si může nechat zobrazit dostupný software podle zvolené kategorie a aktuálnosti. Po zvolení určité kategorie se zobrazí tabulka dostupného softwaru. Po kliknutí na 'Popis' je v některých případech nutné při první návštěvě odsouhlasit licenční ujednání a následně zadat počet licencí (počet počítačů, na kterých bude software provozován). Po potvrzení již budou nabídnuty veškeré dostupné informace ke konkrétnímu softwaru. Zde je možné i nadále měnit počet licencí. Pokud je dostupný soubor s určitou instalační verzí, tak pro jeho stažení na disk stačí jen kliknout odkaz "Stáhnout" a pokračovat dle instrukcí internetového prohlížeče.

Software

Výběr kategorie softwaru: Aplikace

- [x] Pouze aktuální software (platný)
- [x] Pouze volné licence

Název softwaru	Lokalizace	Popis	Platnost od	Platnost do
SPSS CR, spol. s r.o.				
Clementine 13 (PASW Modeler 13)	EN - Anglická verze	Akademická multilicence pro MU 2010	04.01.2010	31.01.2011
SPSS 18 (PASW Statistics 18)	EN - Anglická verze	Akademická multilicence pro MU 2009	09.12.2009	01.02.2011
StatSoft				
Statistica 9.0	CZ - Česká verze	Jednouživatelská verze	09.12.2009	31.12.2010
Statistica 9.0	EN - Anglická verze	Jednouživatelská verze	30.09.2009	31.12.2010
Statistica 9.1	EN - Anglická verze	Jednouživatelská verze	10.03.2010	31.12.2010

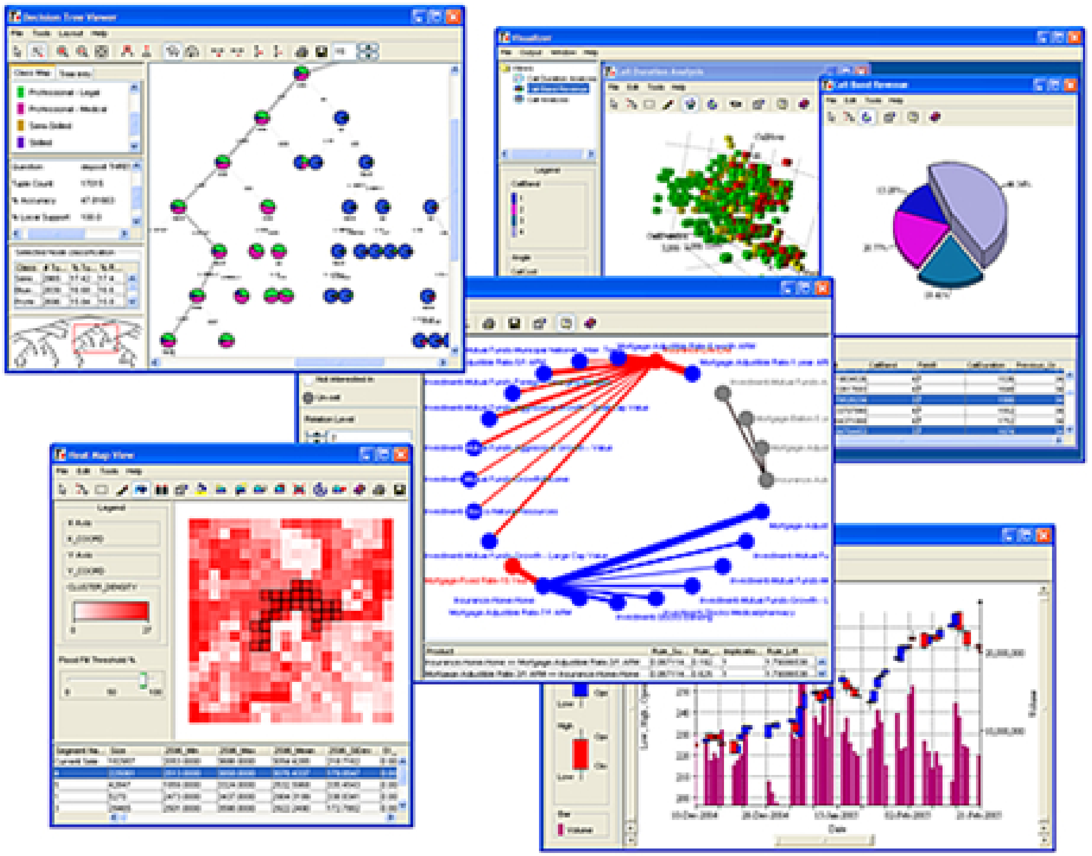
Kontaktní e-mail soft-inet@ics.muni.cz
Informace o serveru inet.muni.cz

- Matlab 2009a: ÚVT MU
  http://www.muni.cz/ics/services/software

GIGO

- Garbage in, Garbage out (smetí dovnitř, smetí ven)
  - sebelepší model/proces/software nevyrobí ze smetí nic jiného než opět smetí.

Vizualizace dat

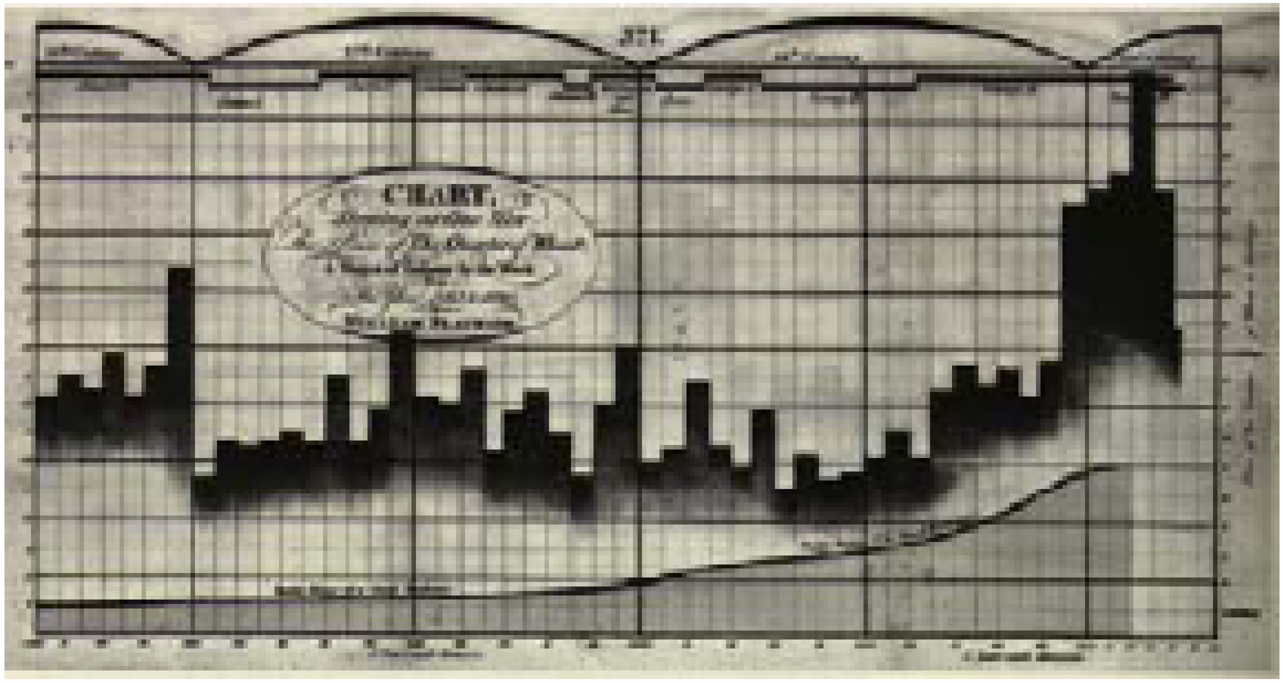


Vizualizace – zdroje

- Na prvním místě se obvykle citují knihy prof. Tufteho, např. Tufte E.R. (1983) The Visual Display of Quantitative Information, Graphic Press, Chesire, Conn.
- Weby o vizualizaci, např.
  - http://www.math.yorku.ca/SCS/Gallery/noframes.html - galerie s poučným výkladem a příklady i nezdařených či lživých grafů
  - http://www.agocg.ac.uk/ - John Lansdown (1992) Aspects of Design in Computer Graphics: Some Notes – http://www.agocg.ac.uk/train/hitch/hitch.htm
- Jiné weby, např. stránky různých vizualizačních programů a organizací
  - http://www.cybergeography.org/atlas/atlas.html nebo http://miner3d.com/products/gallery.html

Vizualizace – historie

- William Playfair, 1786: první publikovaná prezentační grafika

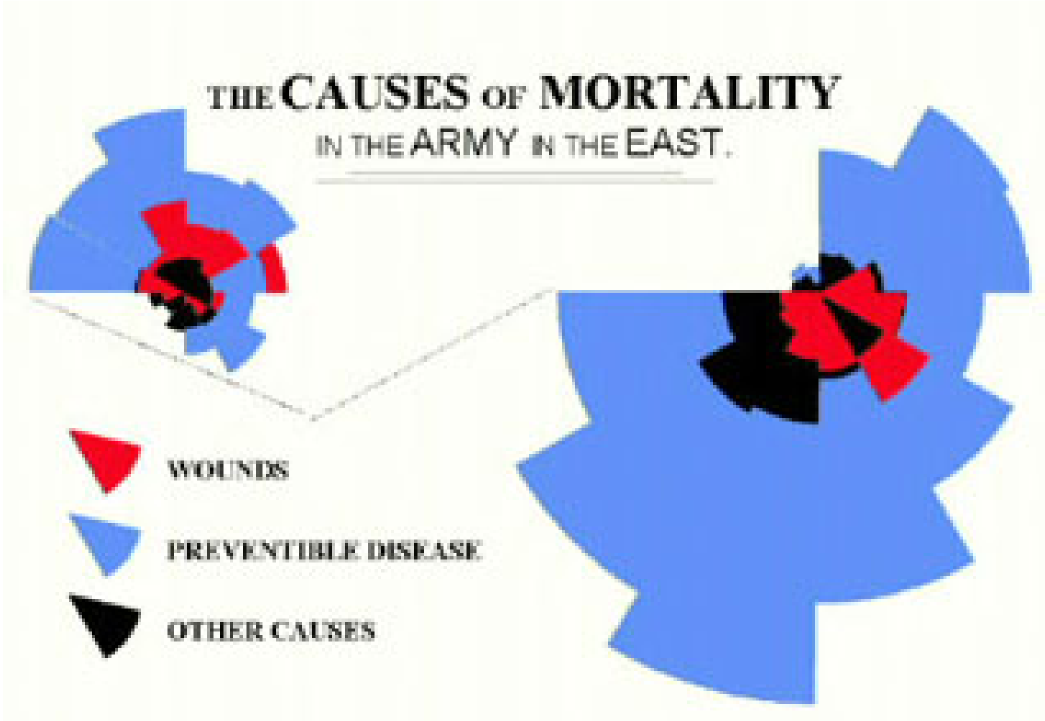


Vizualizace – historie

- Dr. John Snow, 1845: epidemie cholery v Londýně

Vizualizace – historie

- Florence Nightingale, 1858: důvody úmrtí v průběhu Krymské války (1853-1856)



Vizualizace – historie

- Harry Beck, 1931: schéma Londýnského metra

Vizualizace – investigativní analýza

- http://www.i2inc.com/

Law Enforcement

» Counterterrorism
» Narcotics investigations
» Organized crime
» Intelligence analysis
» Fraud
» Missing persons
» Major investigations
» Counterfeiting
» Immigration control
» Major event security
» Money laundering
» Gang investigations

Government

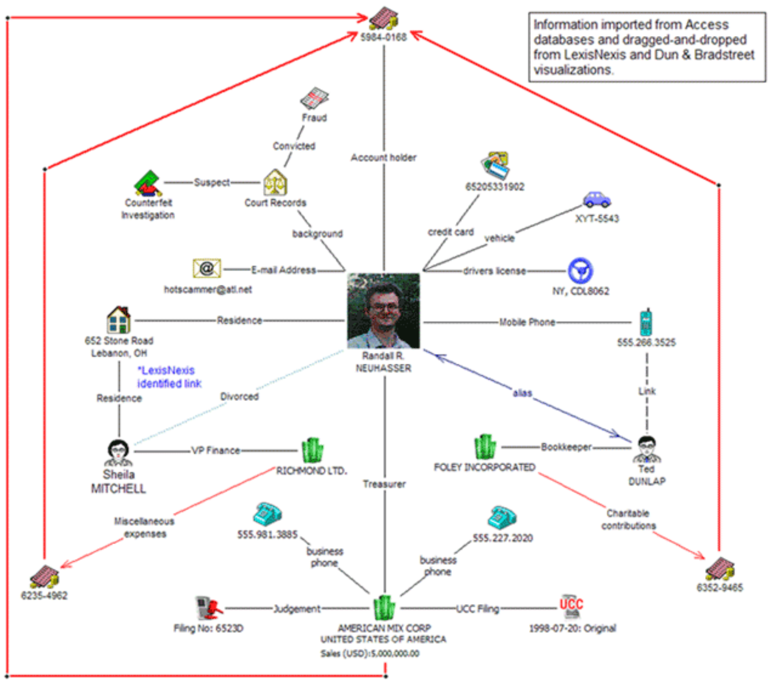
» Criminal prosecutions
» National security
» Military intelligence
» Embassy security
» Postal inspection and fraud
» Prison investigations
» Park and wildlife services
» Antitrust investigations
» Tax fraud investigations
» Customs investigations

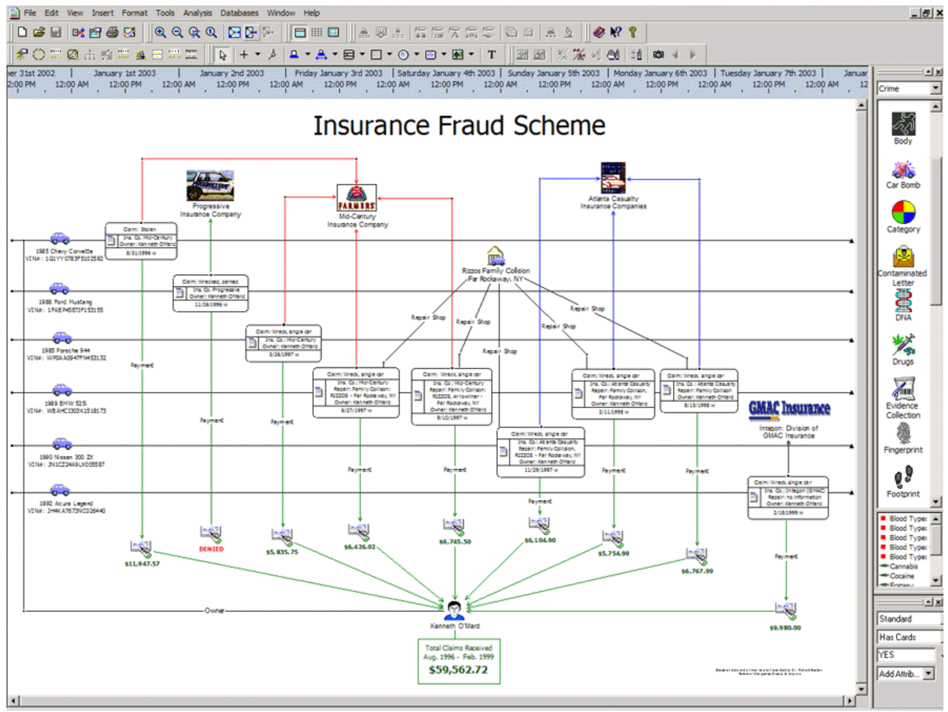
Commercial

» Forensic accounting
» Money laundering
» Insider trading violations
» Corporate security
» Anti-pirating investigations
» Entertainment copyright violations
» Competitive intelligence
» Civil lawsuits
» Fraud:
  » Credit card
  » Insurance
  » Retail
  » Health care
  » Commercial
  » Telephone

Vizualizace – investigativní analýza

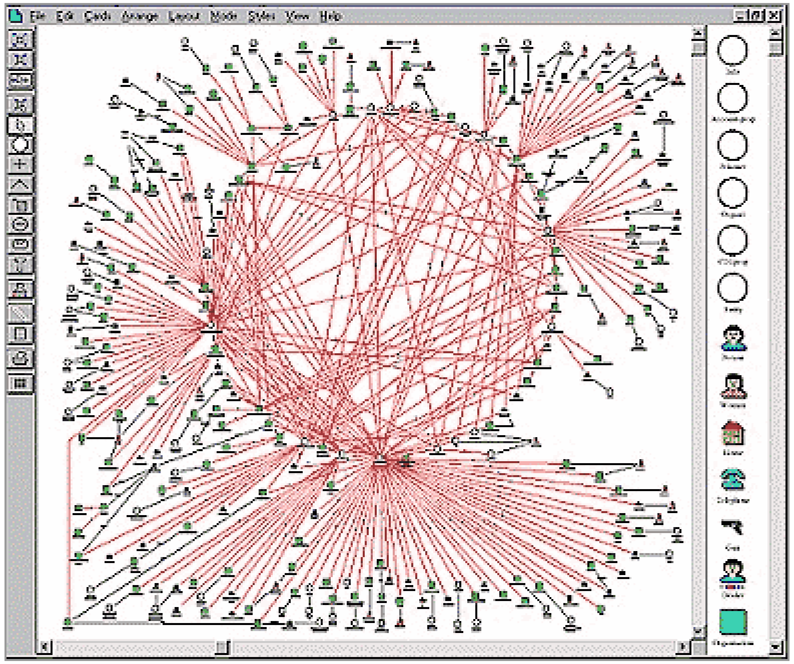
- osobní kontakty, pojistné podvody

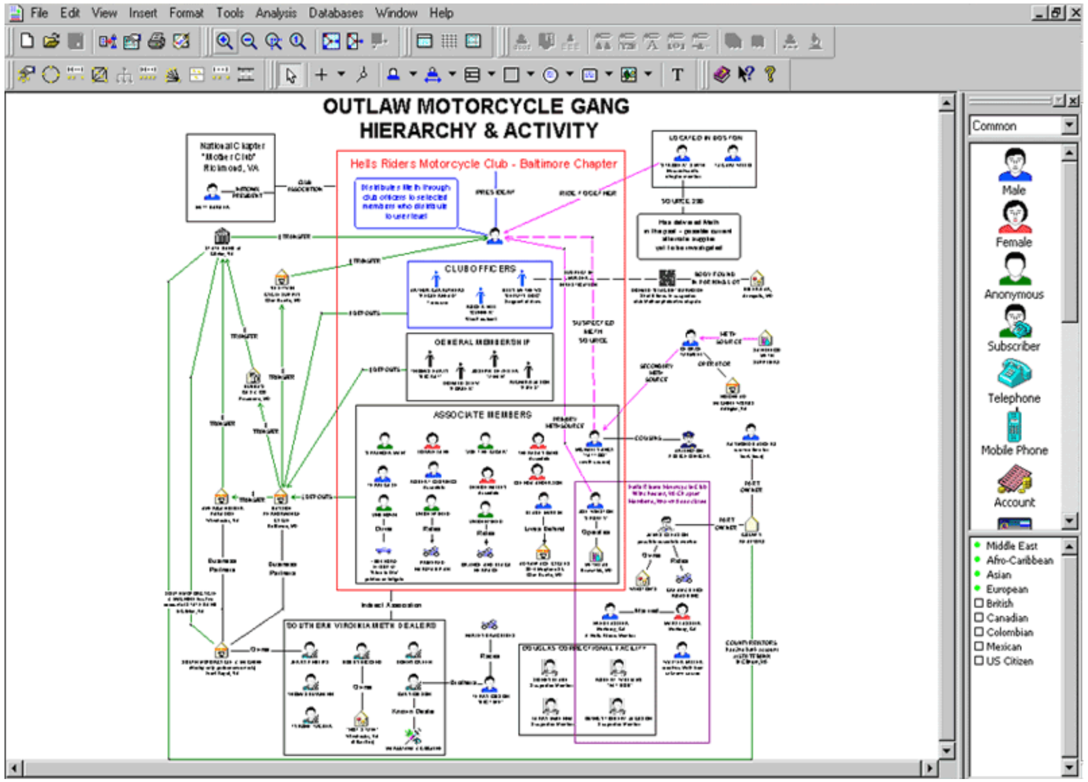




Vizualizace – investigativní analýza

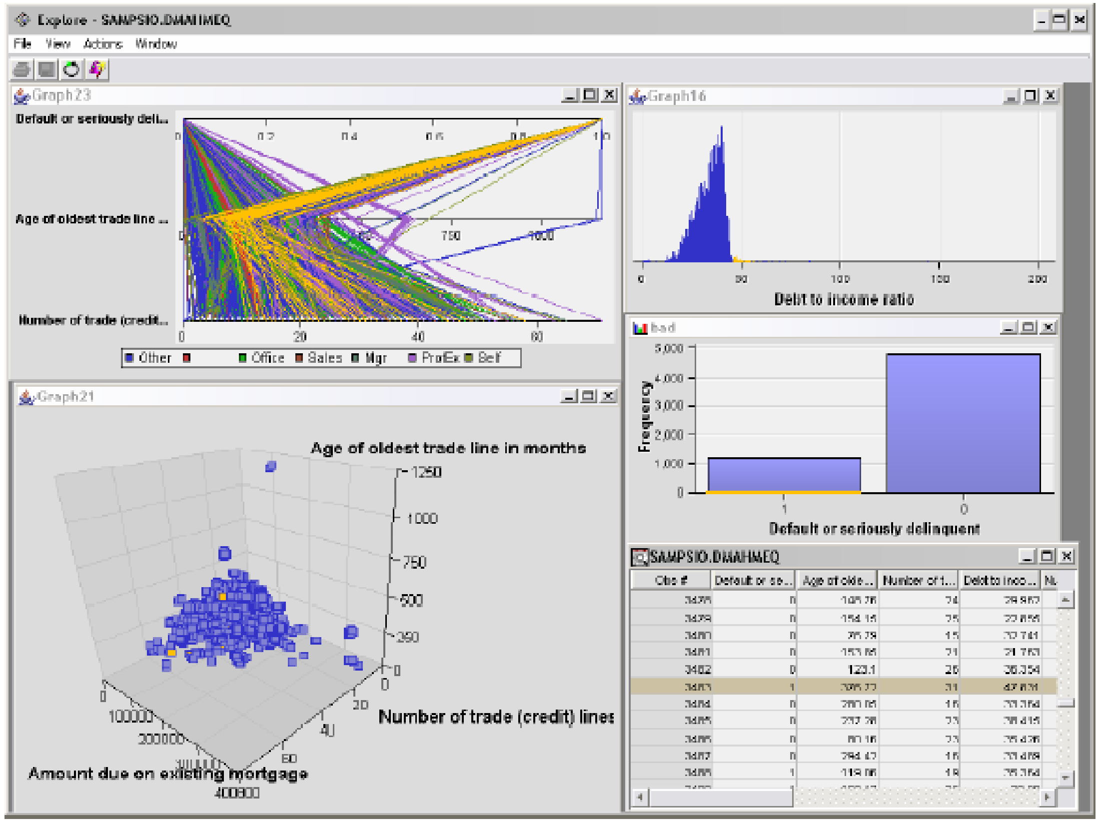
- Praní špinavých peněz, kriminální gangy





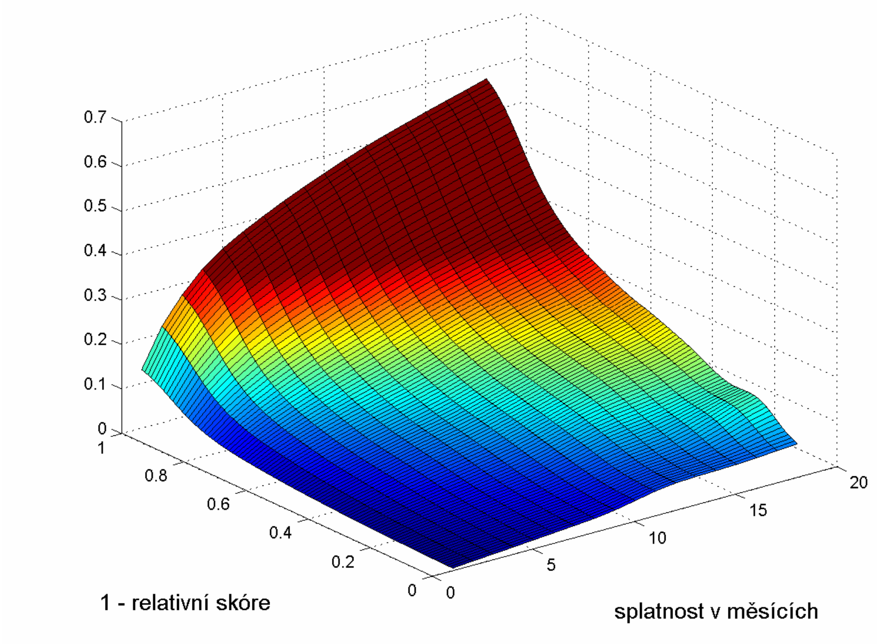
Vizualizace – portfolio management

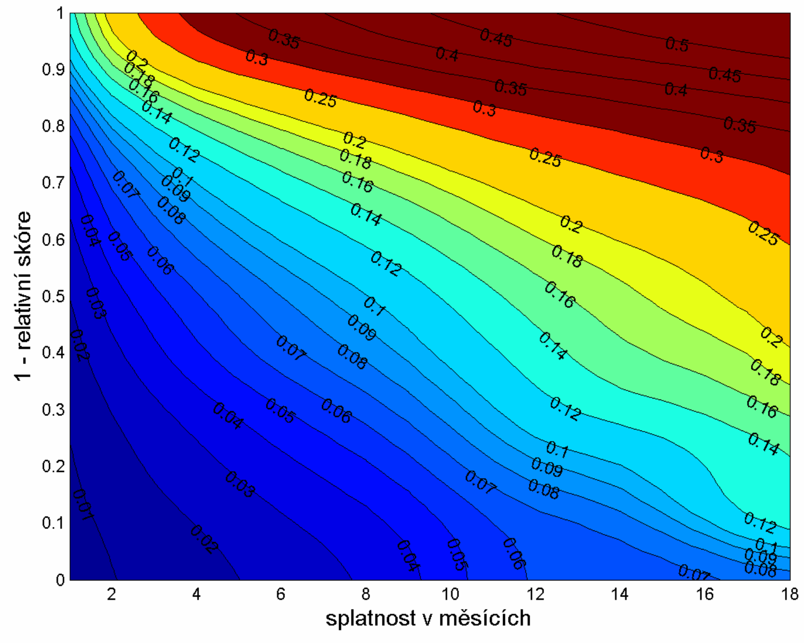
- Hledání závislostí v datech:



Vizualizace – portfolio management

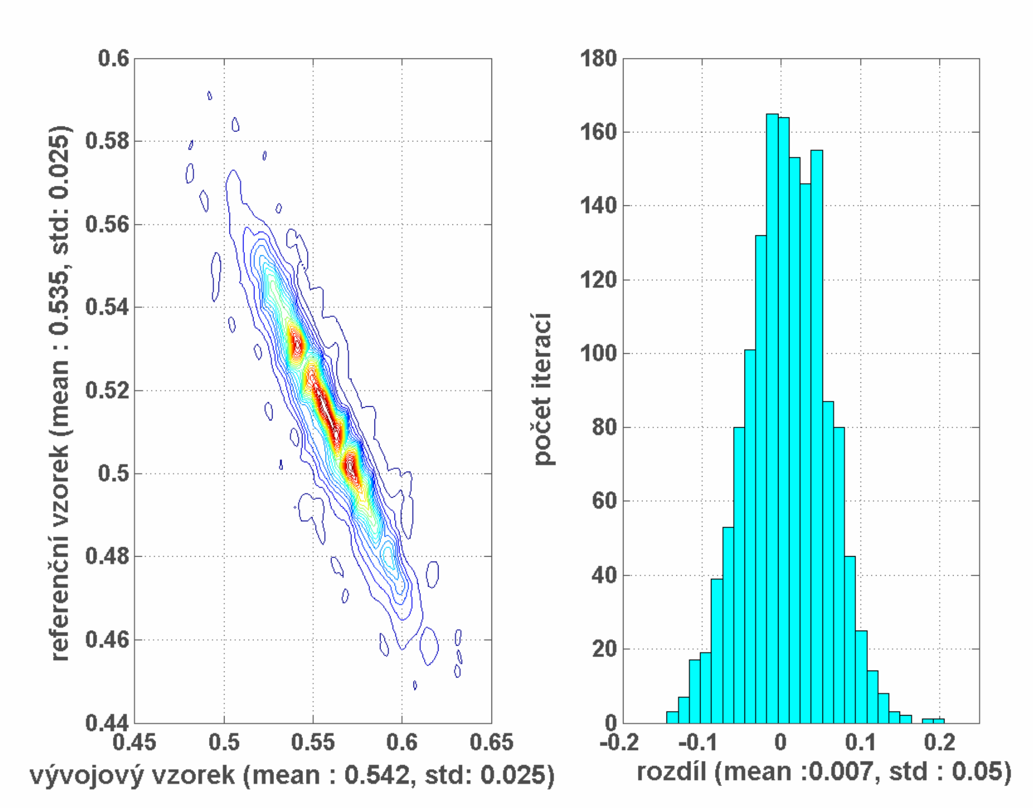
- Distribuční funkce doby do defaultu





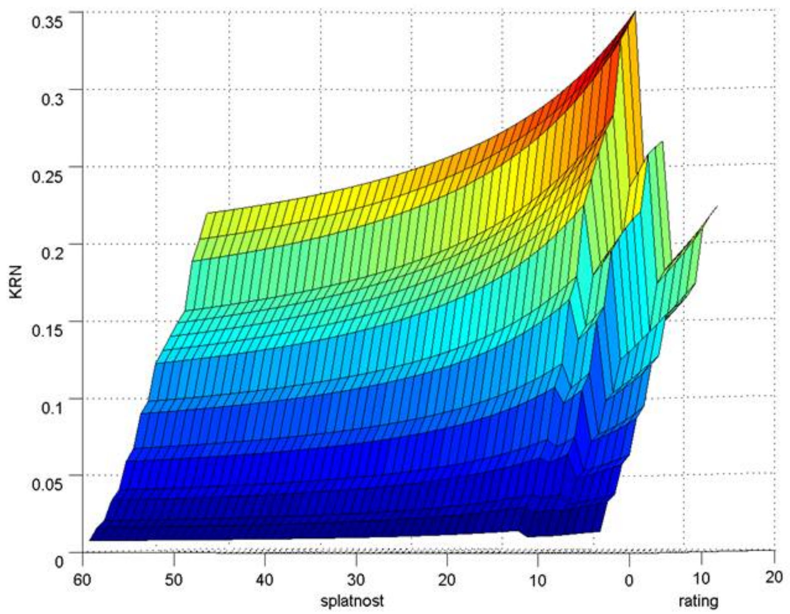
Vizualizace – portfolio management

- Test stability scoringové funkce



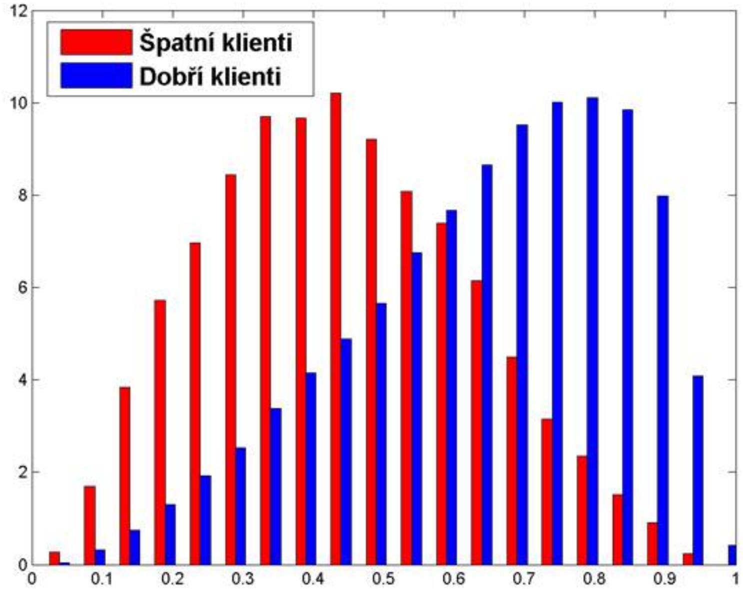
Vizualizace – portfolio management

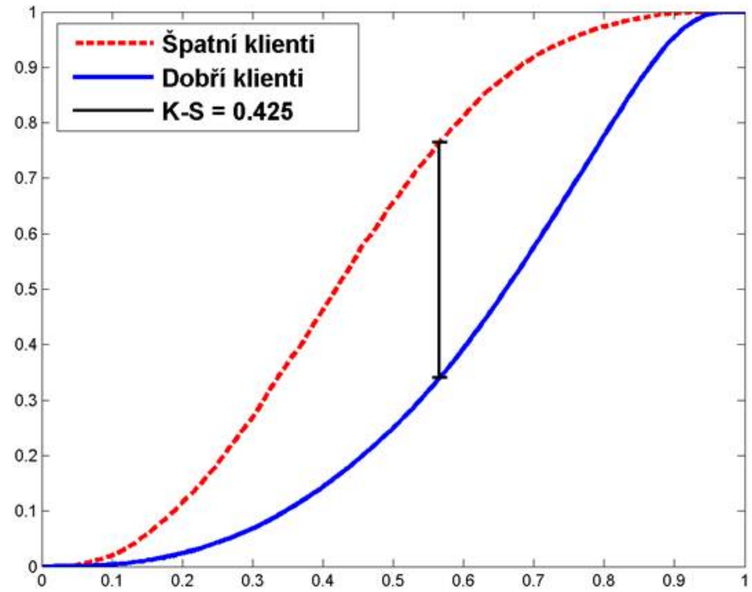
- Vizualizace kreditních rizikových nákladů (KRN)



Vizualizace – portfolio management

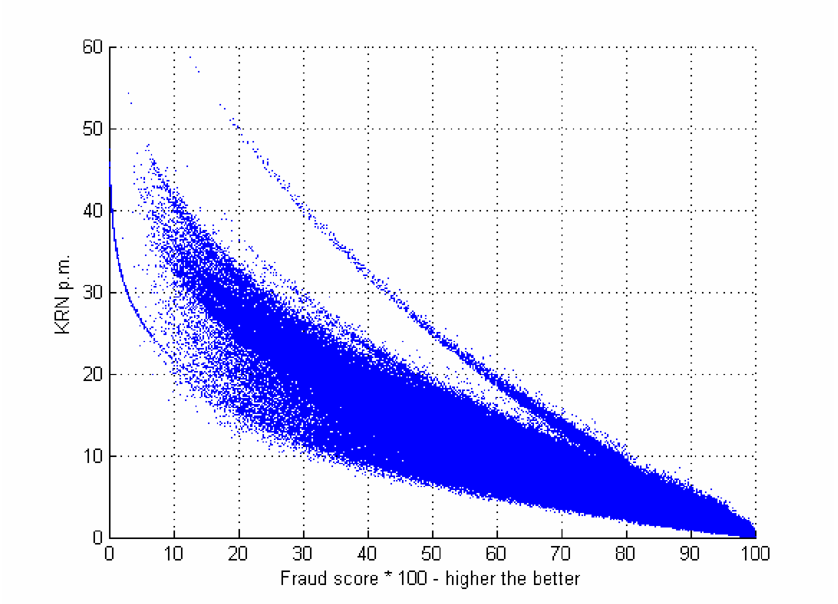
- Histogram, Distribuční funkce

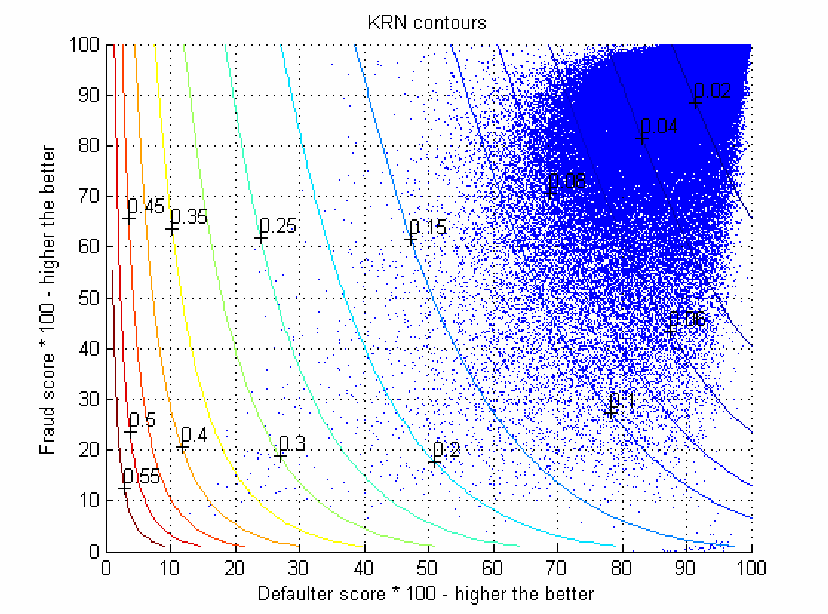




Vizualizace – portfolio management

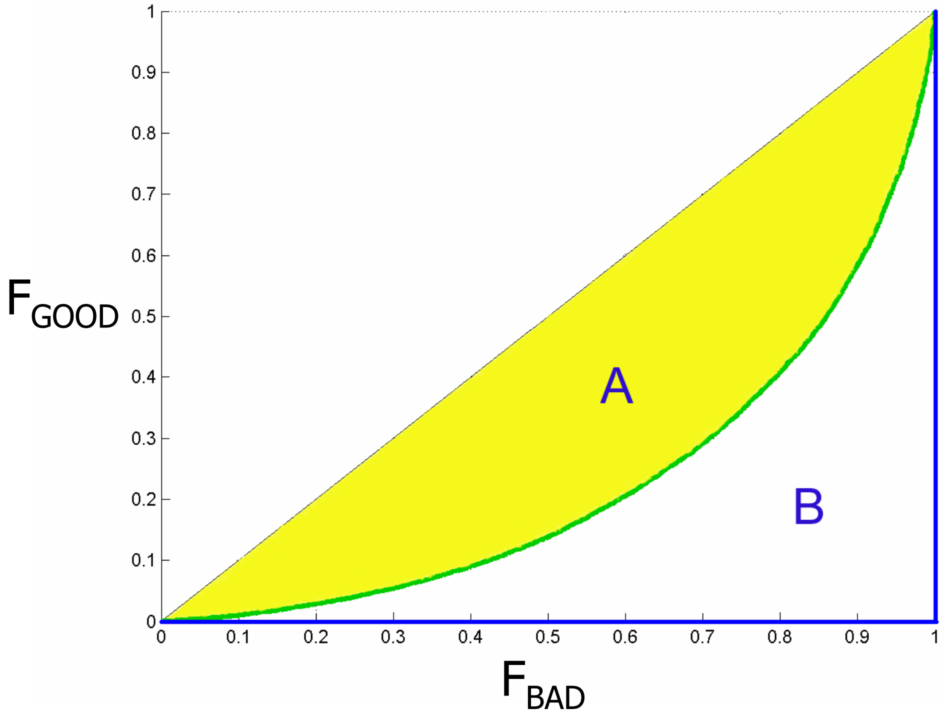
- Bodové grafy





Vizualizace – portfolio management

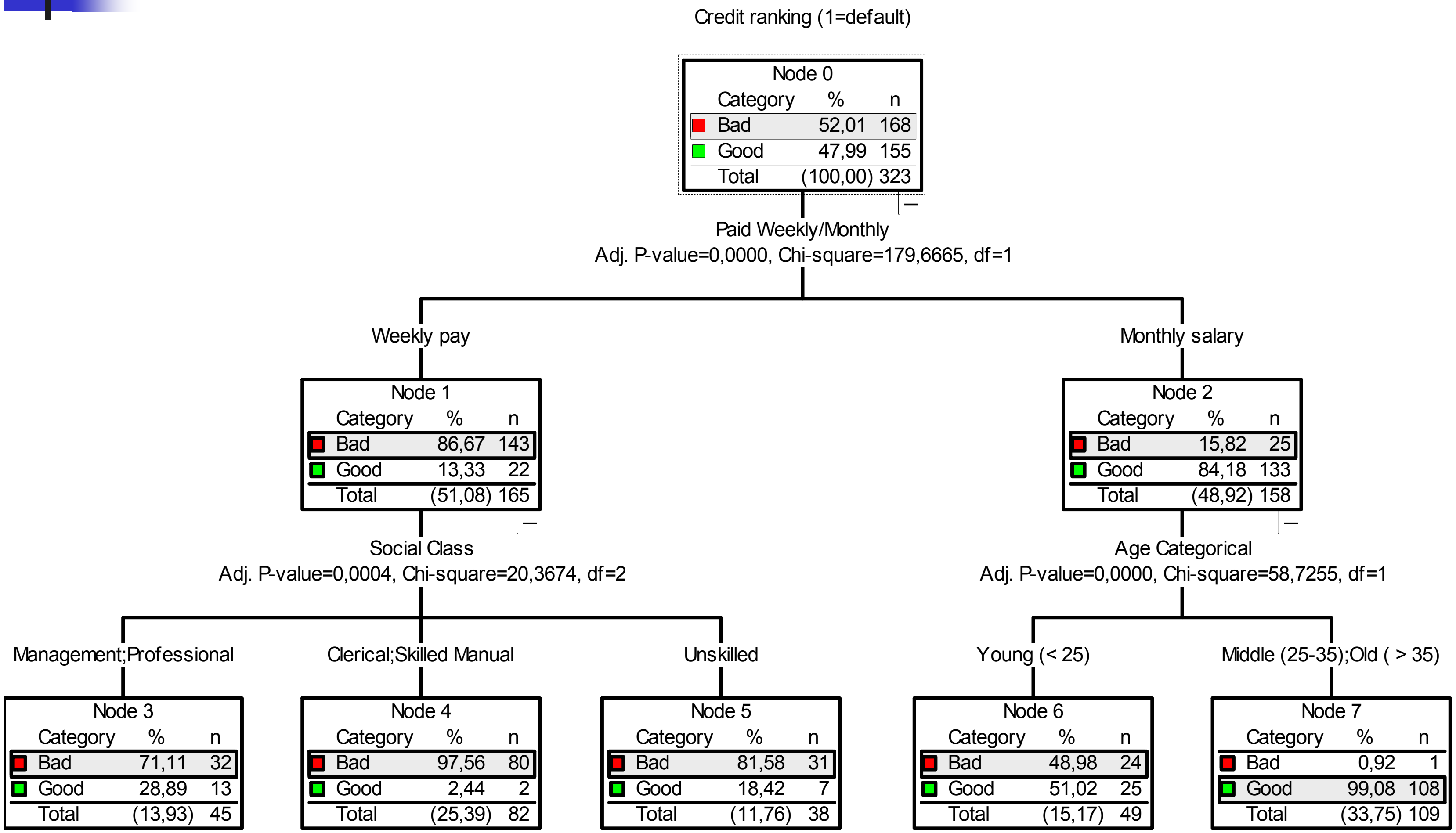
- Lorenzova křivka



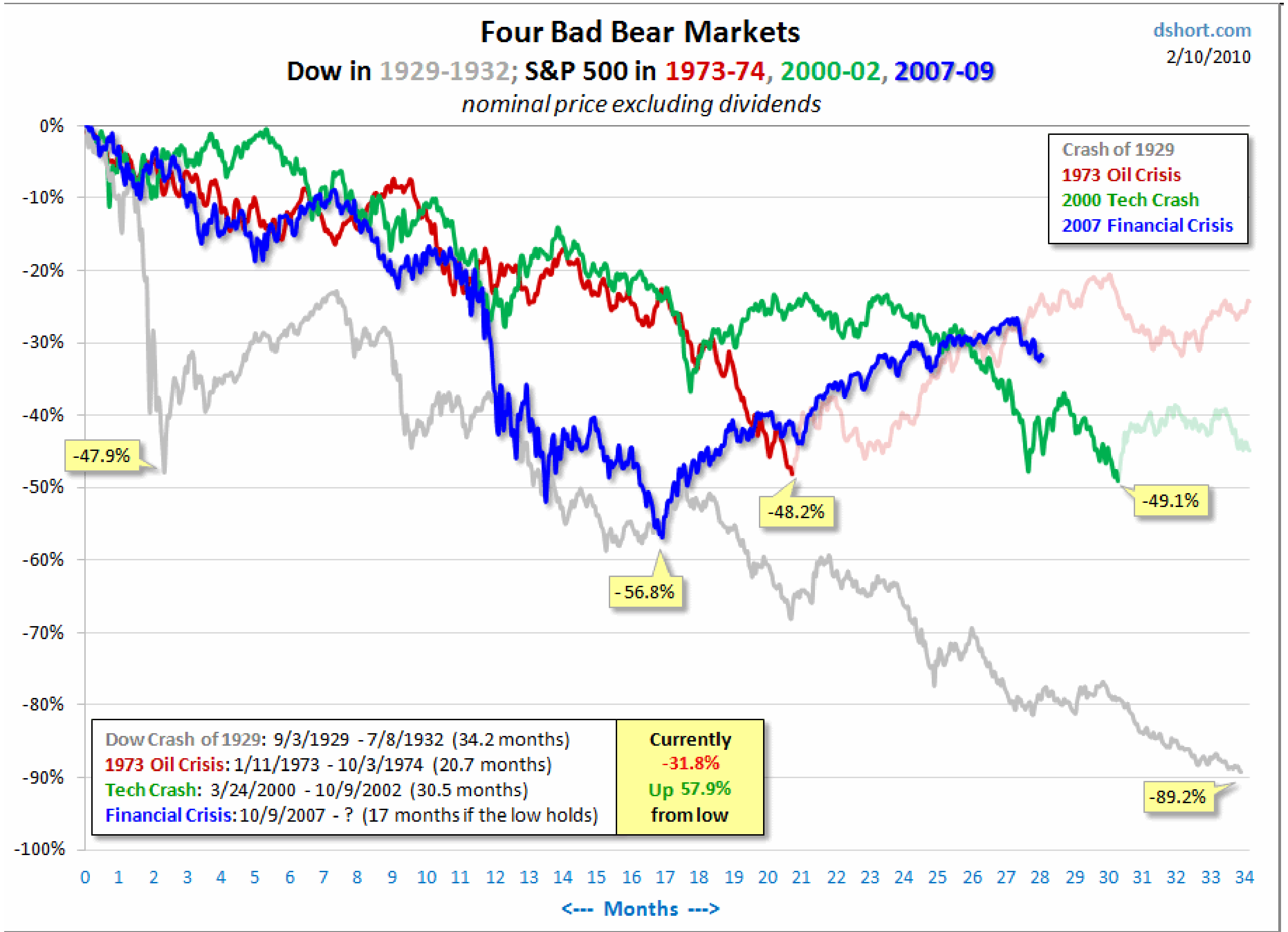
$$Gini = \frac{A}{A+B}$$

$$Gini = 2A$$

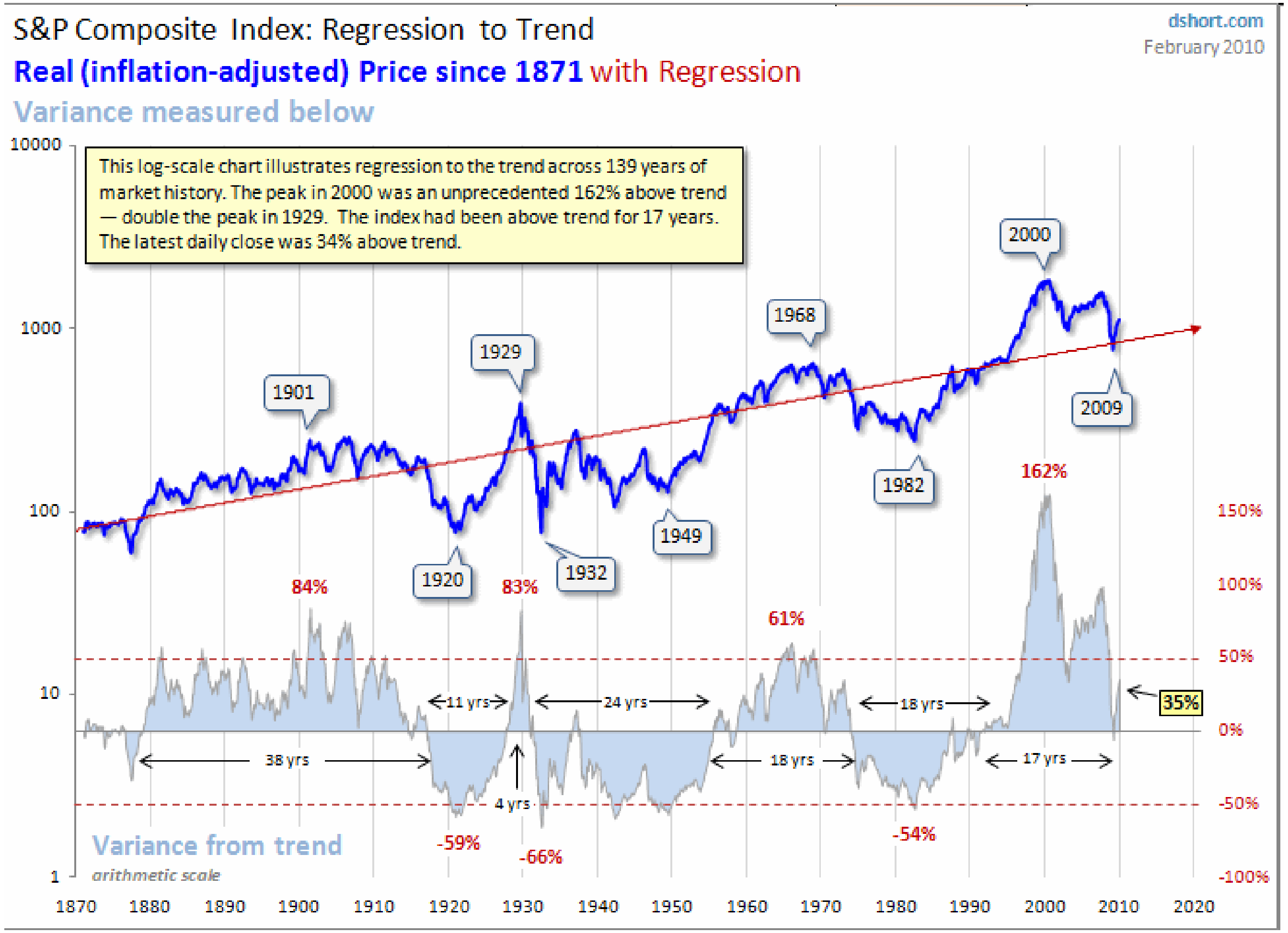
Vizualizace - dendrogram



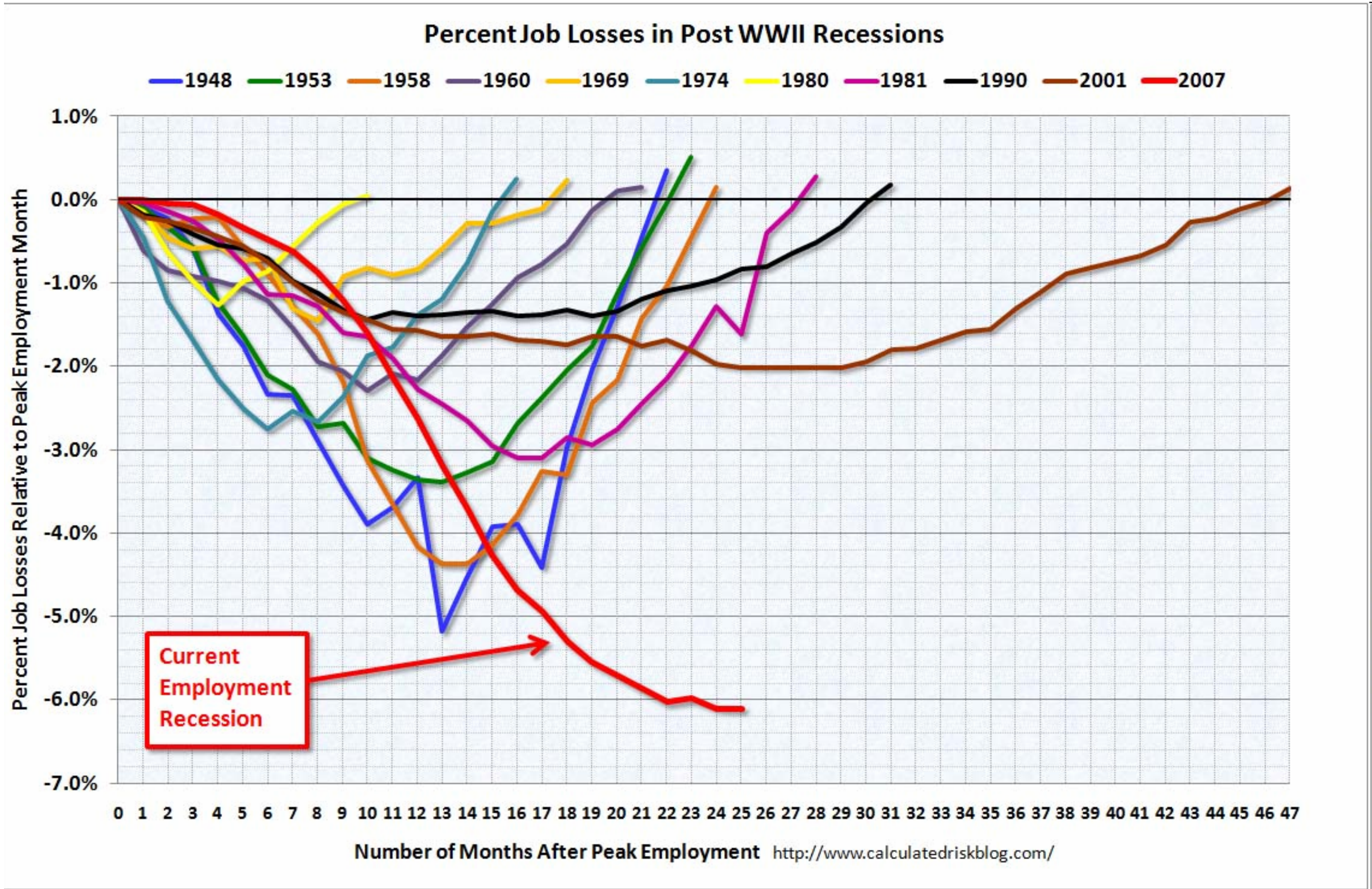
Vizualizace – ekonomie



Vizualizace – ekonomie

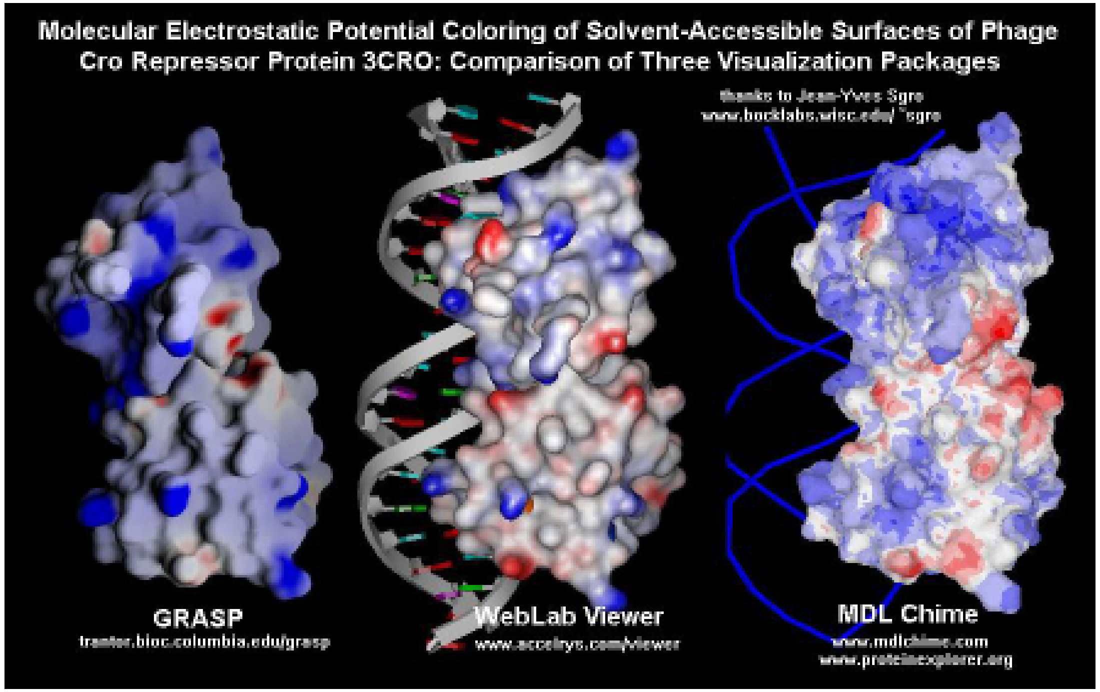


Vizualizace – ekonomie

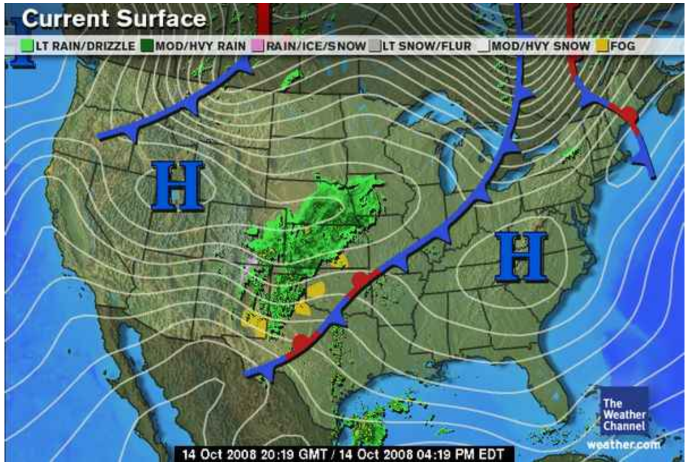


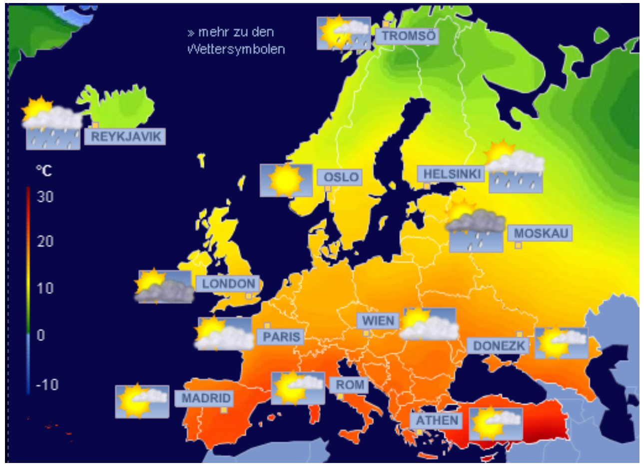
Vizualizace v biologii/chemii

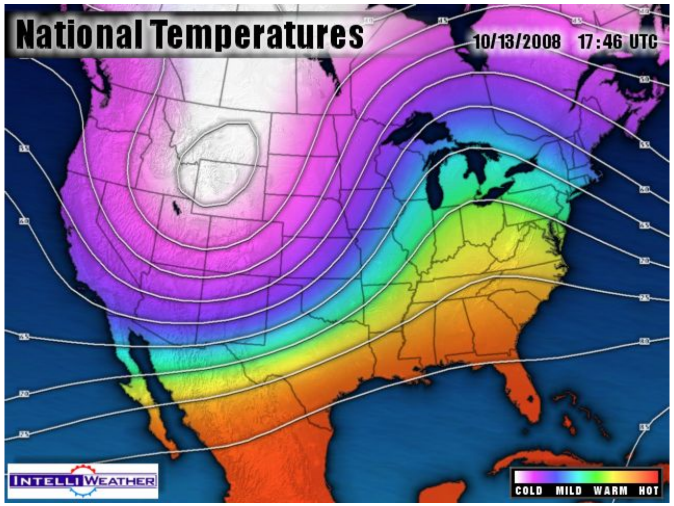
- 3D zobrazení proteinu

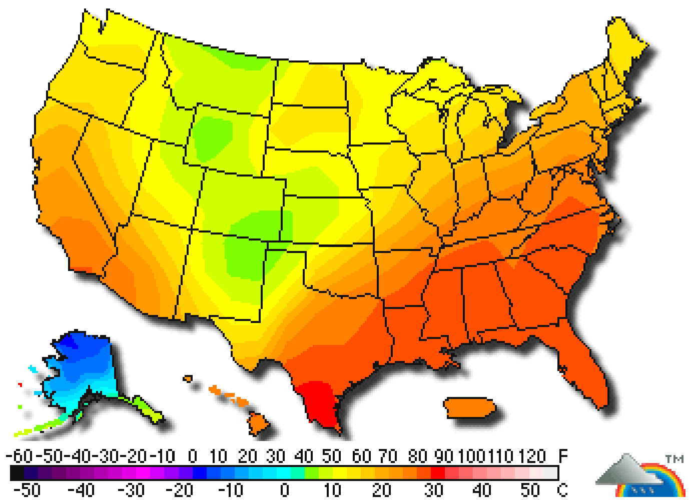


Meteo-vizualizace



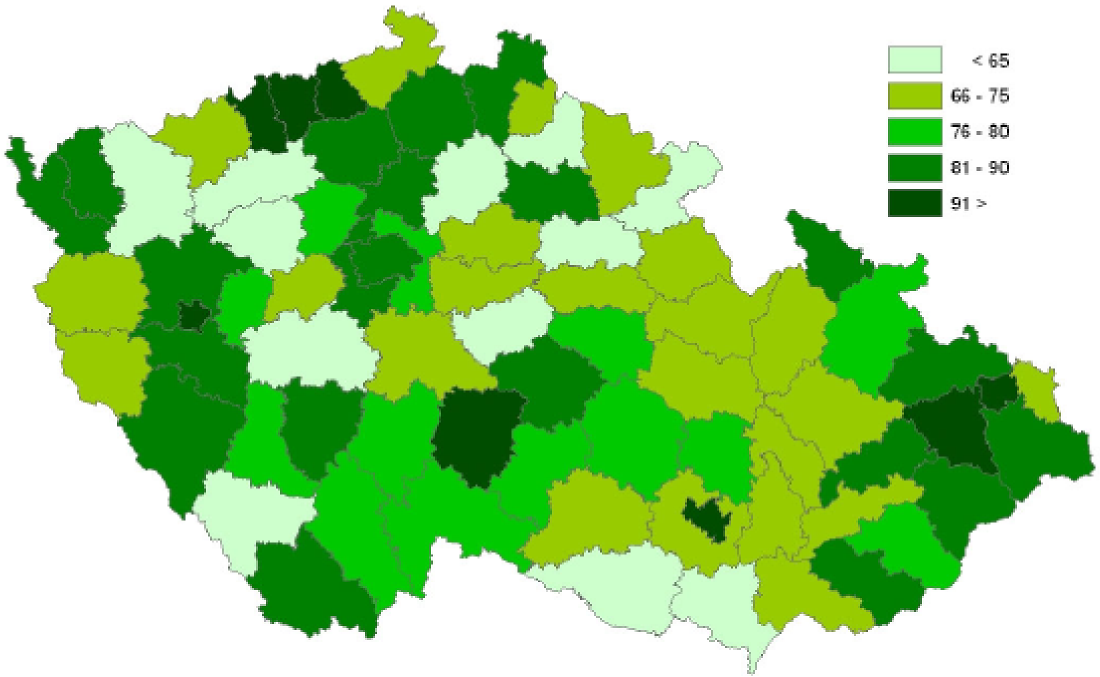




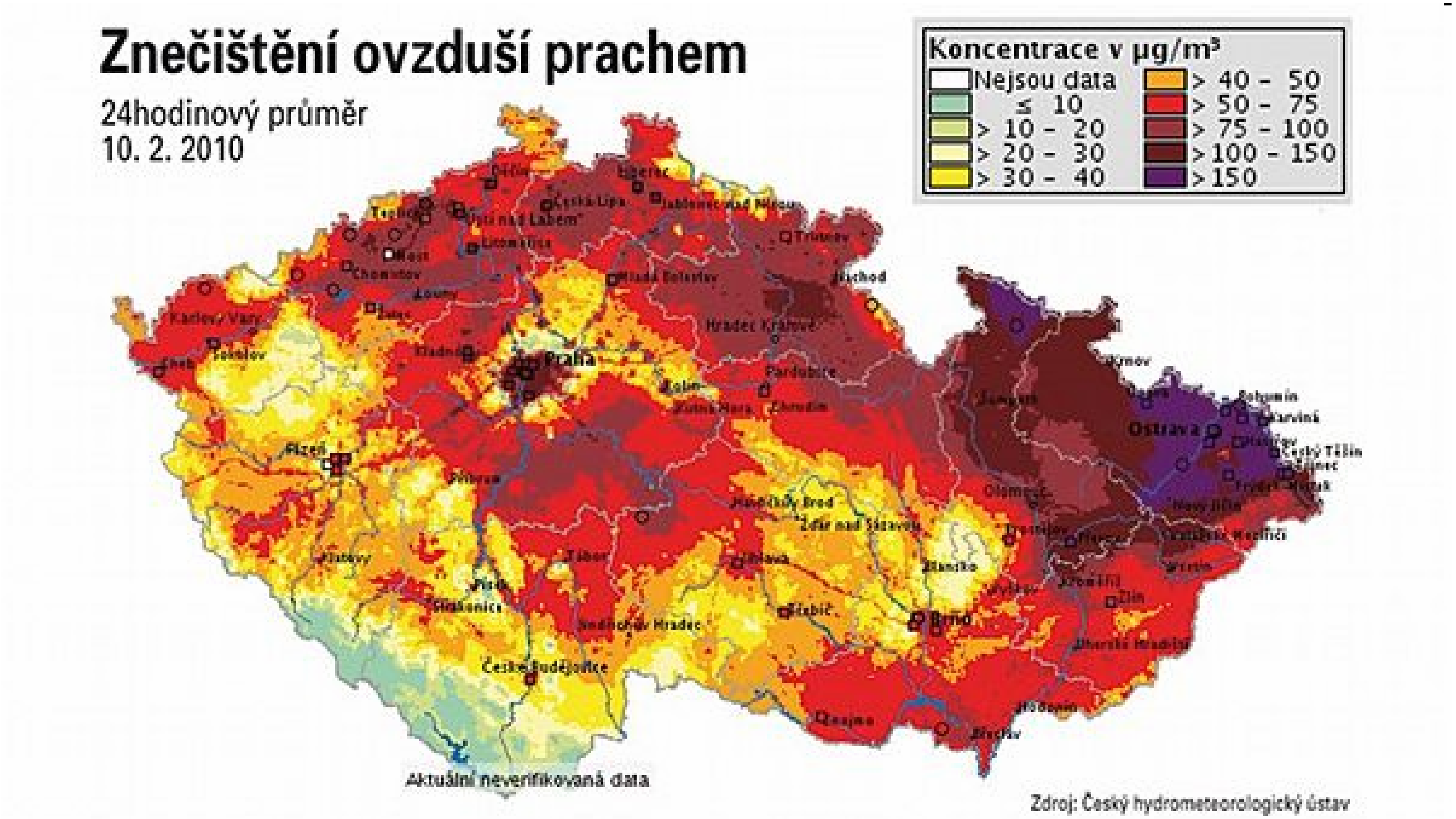


Kartogram

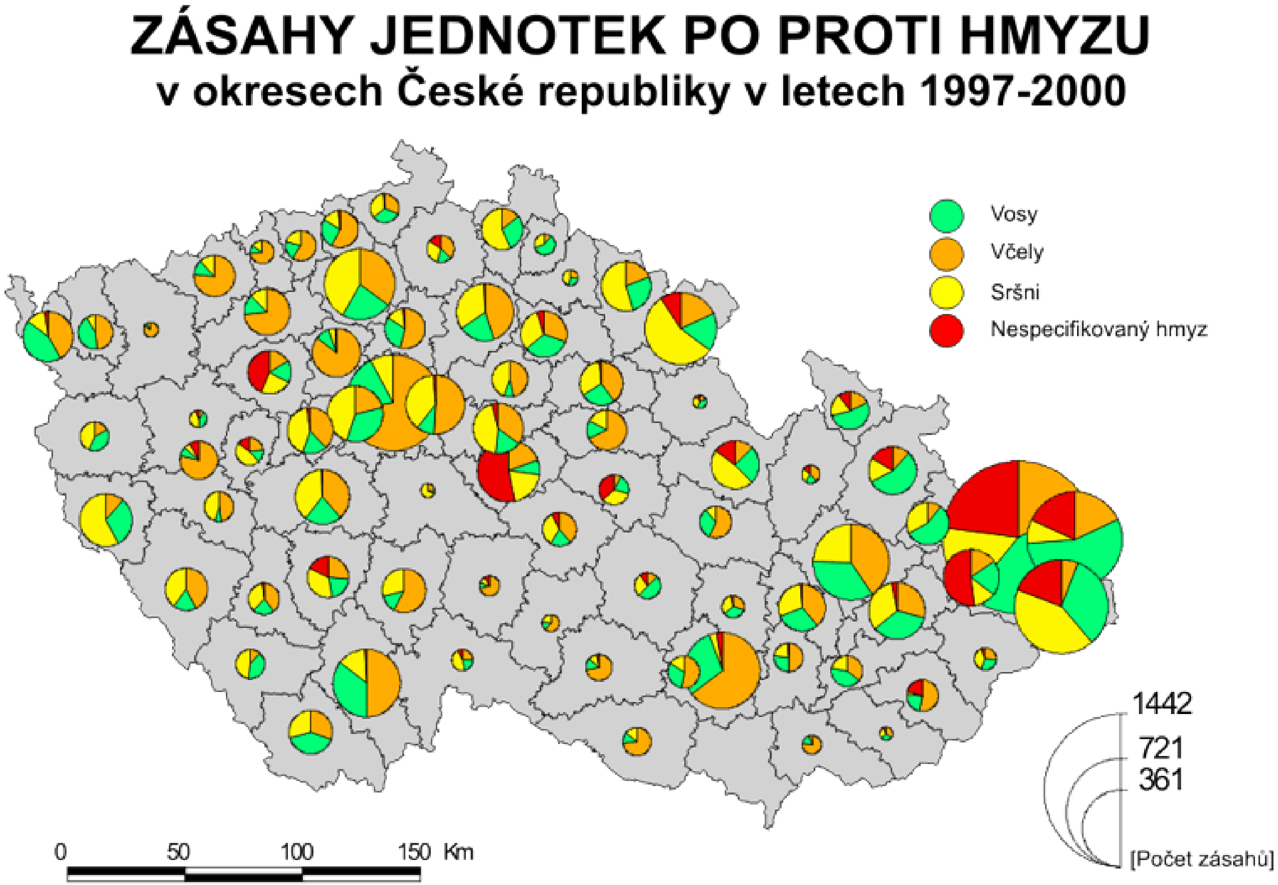
- Obce s počtem 500 a více obyvatel s vysokorychlostním připojením k internetu, podle okresů (%), k 31.12.2006



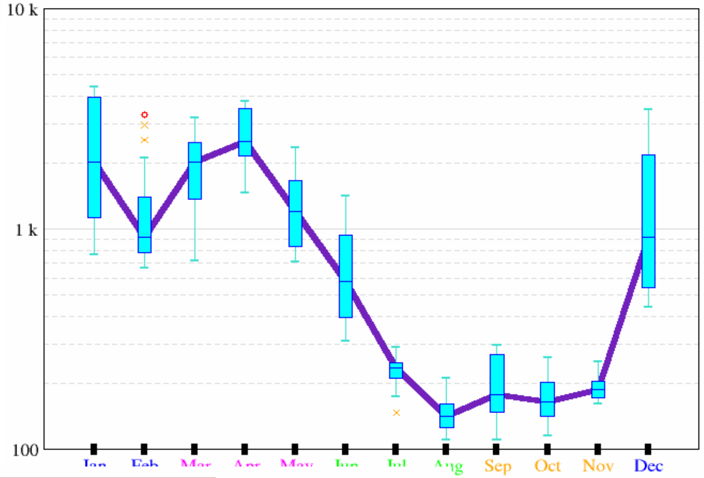
Kartogram

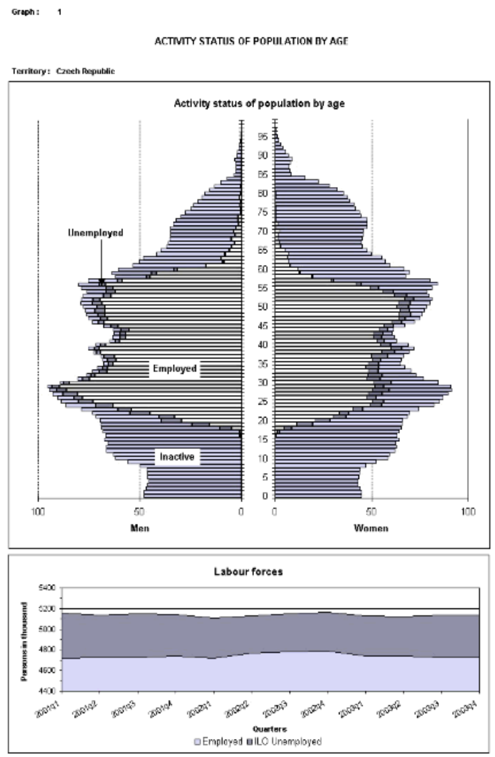

Znečištění ovzduší prachem
24hodinový průměr
10. 2. 2010
Koncentrace v µg/m³
Nejsou data
≤ 10
> 10 - 20
> 20 - 30
> 30 - 40
> 40 - 50
> 50 - 75
> 75 - 100
> 100 - 150
> 150
Praha
Ostrava
Brno
Plzeň
Karlovy Vary
Hradec Králové
Pardubice
České Budějovice
Aktuální neverifikovaná data
Zdroj: Český hydrometeorologický ústav

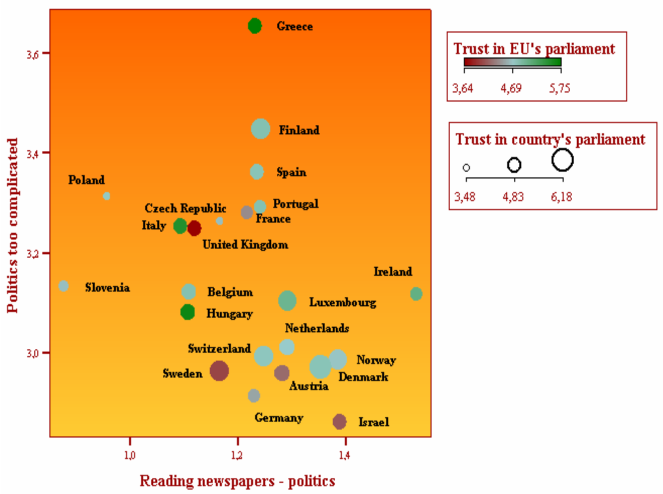
Kartodiagram

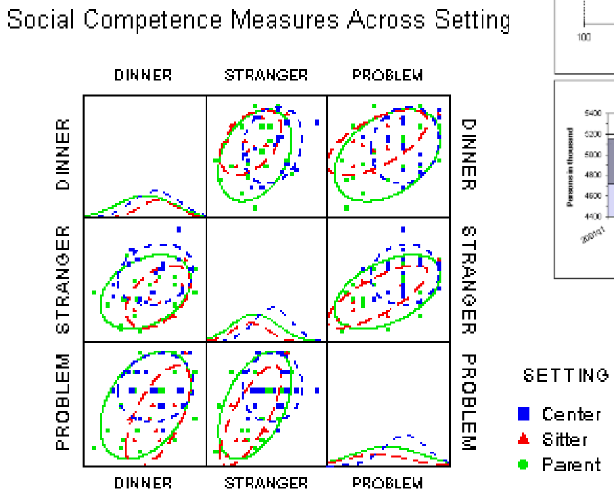


Grafy – další typy

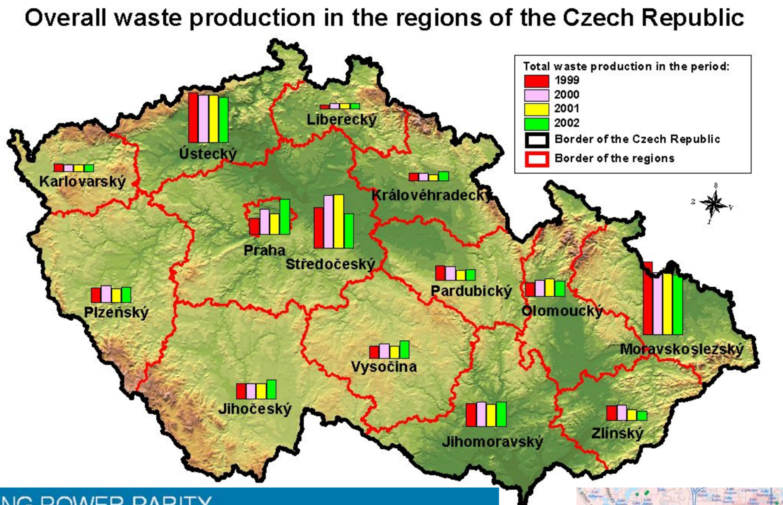


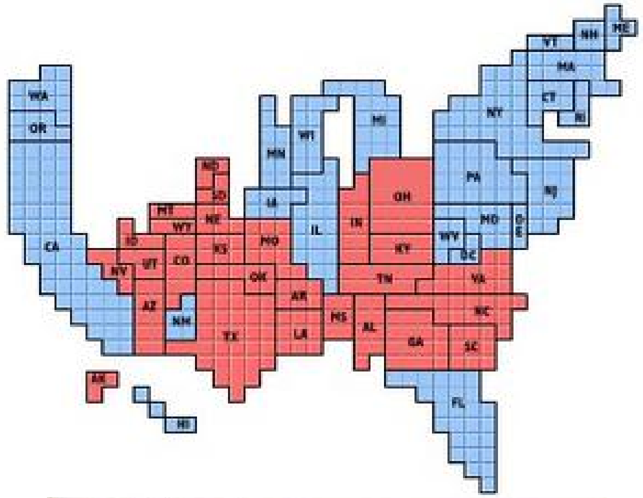


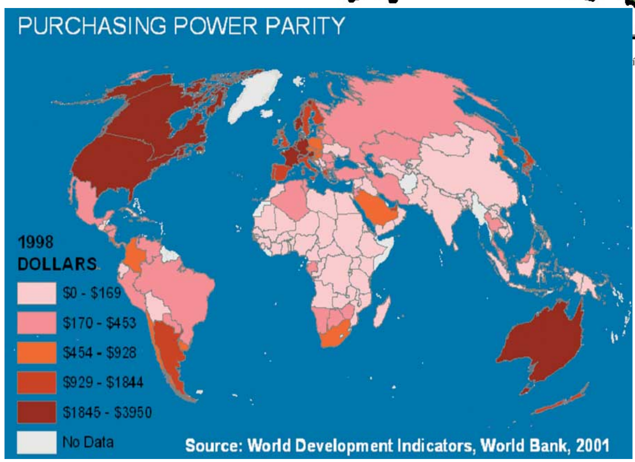


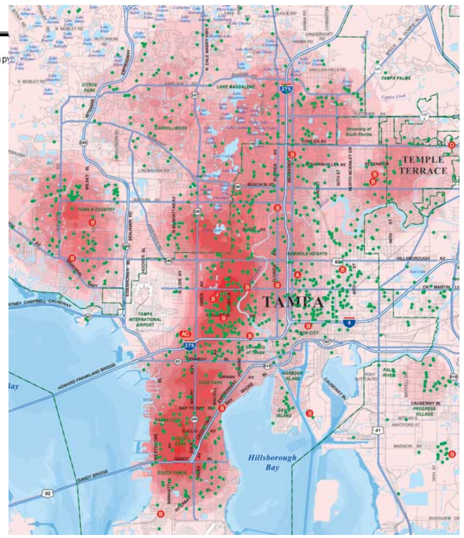


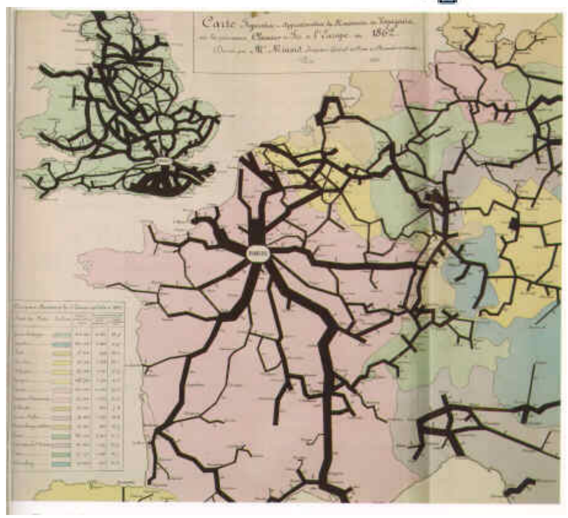
Geografická data











7. Náhodné veličiny

Motivace: Výsledky náhodného pokusu lze popsat reálnými čísly (resp. reálnými vektory) pomocí nějakého zobrazení $X : \Omega \to R$ ($\mathbf{X} = (X_1, \ldots, X_n) : \Omega \to \mathbf{R}^n$). Pokud bude toto zobrazení splňovat určité podmínky, nazveme ho náhodnou veličinou. Příklady náhodných veličin: počet členů náhodně vybrané domácnosti, počet chyb, jichž se dopustí nějaké zařízení za určitou dobu, doba do poruchy nějakého zařízení, hmotnost náhodně vybraného výrobku apod.

Vztah mezi znakem a náhodnou veličinou

Pojem „znak", který jsme zavedli v popisné statistice, je sice blízký pojmu „náhodná veličina", ale není s ním totožný. Znak může být považován za náhodnou veličinu, jestliže jeho hodnoty zjišťujeme na objektech, které byly vybrány ze základního souboru náhodně.

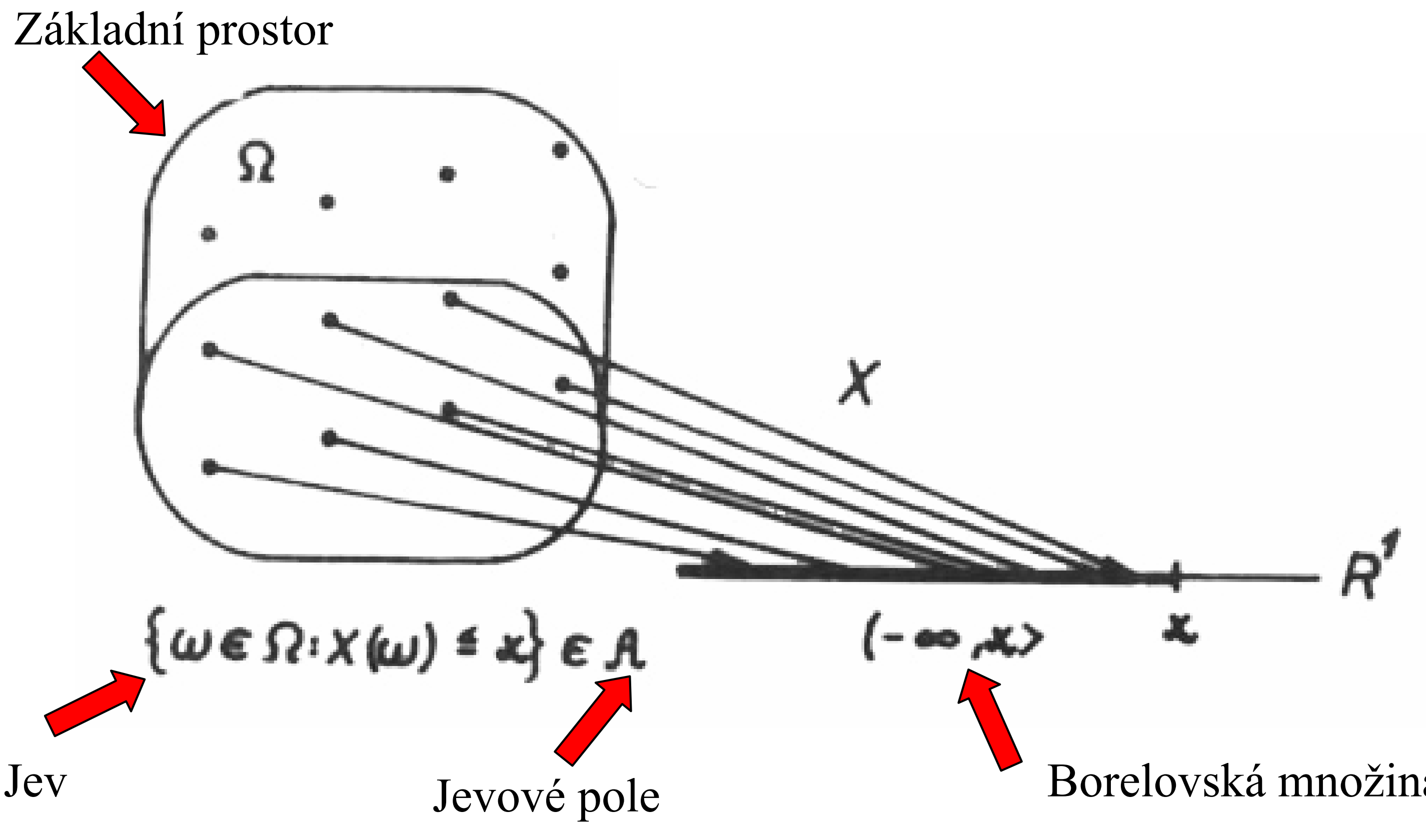
Definice:

Nechť $(\Omega, \mathcal{A}), (R^n, \mathcal{B}^n)$ jsou měřitelné prostory. Zobrazení $\mathbf{X} : \Omega \mapsto R^n$ se nazývá **náhodná veličina** (vzhledem k $\mathcal{A}$), právě když je borelovsky měřitelné (vzhledem k $\mathcal{A}$). Pro $n = 1$ hovoříme o **skalární náhodné veličině**, pro $n \geq 2$ o **náhodném vektoru**. Přitom zobrazení $X_1 : \Omega \to R, \ldots, X_n : \Omega \mapsto R$ se nazývají **složky náhodného vektoru**. Obraz $\mathbf{X}(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ se nazývá **číselná realizace** náhodné veličiny $\mathbf{X}$ příslušná možnému výsledku ω.

Ilustrace náhodné veličiny

Nechť $(\Omega, \mathcal{A}), (R^n, \mathcal{B}^n)$ jsou měřitelné prostory. Zobrazení $\mathbf{X} : \Omega \mapsto R^n$ se nazývá **borelovsky měřitelné** (vzhledem k $\mathcal{A}$), právě když úplný vzor každé n-rozměrné borelovské množiny je jev, tj.

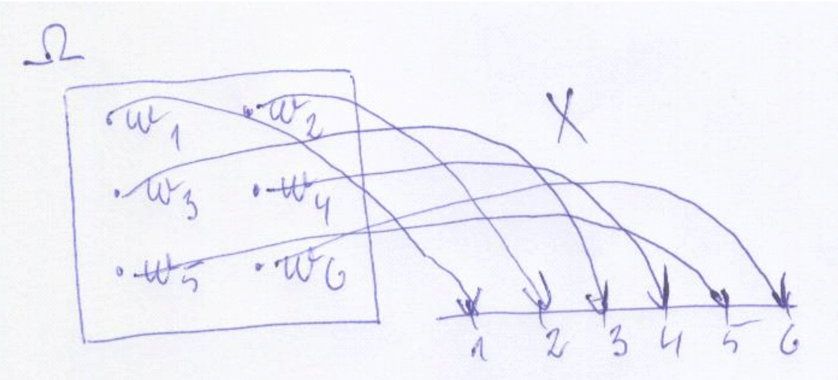
$\forall B \in \mathcal{B}^n : \mathbf{X}^{inv}(B) = \{\omega \in \Omega; X(\omega) \in B\} \in \mathcal{A}$.



Příklad

Příklad: Náhodný pokus spočívá v hodu kostkou. Základní prostor $\Omega = \{\omega_1,\ldots,\omega_6\}$. Uvážíme dvě jevová pole, a to $\mathcal{A}_{max} = \{A; A \subseteq \Omega\}$ a $\mathcal{A} = \{\Omega, \emptyset, \{\omega_1,\omega_3,\omega_5\}, \{\omega_2,\omega_4,\omega_6\}\}$. Zjistěte, zda zobrazení $X:\Omega \to R$, které poloze kostky číslem i nahoru přiřazuje číslo i, i = 1, …, 6, je náhodná veličina vzhledem k $\mathcal{A}_{max}$ a vzhledem k $\mathcal{A}$.

Řešení:



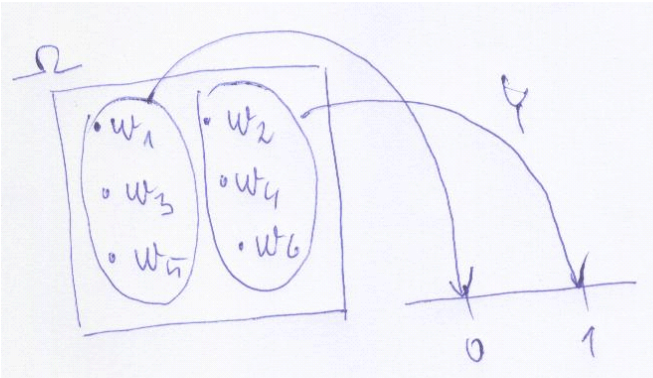
Zobrazení $X:\Omega \to R$ je náhodná veličina vzhledem k $\mathcal{A}_{max}$, protože úplný vzor každé borelovské množiny je jev vzhledem k $\mathcal{A}_{max}$. Vzhledem k $\mathcal{A}$ však X není náhodná veličina:

Úplný vzor množiny $(-\infty, 4\rangle$ je $\{\omega_1,\omega_2,\omega_3,\omega_4\} \notin \mathcal{A}$.

Příklad

Zavedeme zobrazení $\Omega \to R$, které poloze kostky lichým číslem nahoru přiřazuje 0 a sudým 1.

$\mathcal{A} = \{\Omega, \emptyset, \{\omega_1, \omega_3, \omega_5\}, \{\omega_2, \omega_4, \omega_6\}\}$



Toto zobrazení je náhodná veličina vzhledem k $\mathcal{A}$ a nazývá se ukazatel parity.

Náhodná veličina

Označení

a) Jestliže nehrozí nebezpečí nedorozumění, zapisujeme náhodnou veličinu i její číselnou realizaci týmž symbolem $\mathbf{X}$.

b) Množinu $\{\omega \in \Omega; \mathbf{X}(\omega) \in B\}$ zkráceně zapisujeme $\{\mathbf{X} \in B\}$ a čteme: náhodná veličina $\mathbf{X}$ se realizovala v borelovské množině B. Ve speciálním případě, kdy $B = \{\mathbf{x}\}$ resp. $B = (-\infty, \mathbf{x}\rangle$, píšeme $\{\mathbf{X} = \mathbf{x}\}$ resp. $\{\mathbf{X} \leq \mathbf{x}\}$.

c) Zápis pravděpodobnosti zkrátíme takto:

$P(\{\omega \in \Omega; \mathbf{X}(\omega) \in B\}) = P(\mathbf{X} \in B)$

$P(\{\omega \in \Omega; \mathbf{X}(\omega) \in B\}/\{\omega \in \Omega; \mathbf{Y}(\omega) \in C\}) = P(\mathbf{X} \in B/\mathbf{Y} \in C)$.

Transformovaná náhodná veličina

Věta:

Nechť $(\Omega, \mathcal{A}), (R^n, \mathcal{B}^n), (R^m, \mathcal{B}^m)$ jsou měřitelné prostory. Nechť $\mathbf{X} : \Omega \mapsto R^n$ je náhodná veličina a $\mathbf{g} : R^n \mapsto R^m$ je borelovská funkce. Pak složené zobrazení $\mathbf{Y} : \Omega \mapsto R^m$ dané vzorcem $\forall \omega \in \Omega : \mathbf{Y}(\omega) = \mathbf{g}(\mathbf{X}(\omega))$ je náhodná veličina. Nazývá se **transformovaná náhodná veličina**, pro $m = 1$ skalární, pro $m \geq 2$ vektorová.

Důkaz:

$$\begin{tikzcd}(\Omega, \mathcal{A}) \arrow[r, "\mathbb{X}"] \arrow[rr, bend right, "\mathbb{Y}"'] & (R^n, \mathcal{B}^n) \arrow[r, "g"] & (R^m, \mathcal{B}^m)\end{tikzcd}$$

Aby zobrazení $\mathbf{Y}: \Omega \to \mathrm{R}^m$ bylo náhodnou veličinou vzhledem k $\mathcal{A}$, musí platit:

$\forall B \in \mathcal{B}^m$: $\mathbf{Y}^{inv}(B) = \{\omega \in \Omega; \mathbf{Y}(\omega) \in B\} \in \mathcal{A}$. Nechť tedy $B \in \mathcal{B}^m$. Protože $\mathbf{g}$ je borelovská funkce, je $\mathbf{g}^{inv}(B) \in \mathcal{B}^n$. Protože $\mathbf{X}$ je náhodná veličina, je $\mathbf{X}^{inv}(\mathbf{g}^{inv}(B)) \in \mathcal{A}$. Ovšem $\mathbf{X}^{inv}(\mathbf{g}^{inv}(B)) = \mathbf{Y}^{inv}(B)$.

Transformovaná náhodná veličina

Poznámka: (Příklady transformovaných náhodných veličin) Nechť $\mathbf{X} = (X_1, \ldots, X_n)$ je náhodný vektor.
a) Nechť $\{i, \ldots, j\} = \{1, \ldots, n\} - \{k, \ldots, l\}$. Náhodný vektor $(X_i, \ldots, X_j)$ se nazývá vybraný marginální vektor, $(X_k, \ldots, X_l)$ se nazývá zbylý marginální vektor. Původní náhodný vektor $(X_1, \ldots, X_n)$ se v této souvislosti nazývá zbylý marginální vektor.
b) $\sum_{i=1}^{n} X_i, \max\{X_1, \ldots, X_n\}, \sin(X_i), \ldots$ jsou transformované náhodné veličiny.

Definice: Posloupnost $\{X_n\}_{n=1}^{\infty}$ spočetně mnoha náhodných veličin definovaných na témž měřitelném prostoru $(\Omega, \mathcal{A})$ se nazývá náhodná posloupnost.

Distribuční funkce náhodné veličiny

Motivace: Při pozorování realizací náhodné veličiny si povšimneme, že některé její hodnoty se vyskytují s větší pravděpodobností, jiné s menší. Pravděpodobnostní chování náhodné veličiny X budeme popisovat pomocí distribuční funkce, která udává pravděpodobnost jevu, že náhodná veličina X se realizuje hodnotou nejvýše x:

$$\boxed{\forall \mathrm{x} \in \mathrm{R} : \Phi(\mathrm{x}) = P(X \leq x)}$$

Je to zidealizovaný protějšek empirické distribuční funkce zavedené v popisné statistice:

$$\forall \mathrm{x} \in \mathrm{R} : \mathrm{F}(\mathrm{x}) = \frac{\mathrm{N}(\mathrm{X} \leq \mathrm{x})}{\mathrm{n}}$$

Lze očekávat, že s rostoucím rozsahem výběrového souboru se budou hodnoty empirické distribuční funkce F(x) ustalovat kolem hodnot distribuční funkce Φ(x). Vlastnosti empirické distribuční funkce se přenášejí i na distribuční funkci.

Distribuční funkce náhodné veličiny

Definice:

a) Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $X : \Omega \mapsto R$ je skalární náhodná veličina. Funkce $\Phi : R \mapsto R$ daná vzorcem:

$\forall x \in R : \Phi(x) = P(X \leq x)$

se nazývá **distribuční funkce** náhodné veličiny X.

b) Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $\mathbf{X} = (X_1, \ldots, X_n) : \Omega \mapsto R^n$ je náhodný vektor. Funkce $\Phi : R^n \mapsto R$ daná vzorcem:

$\forall (x_1, \ldots, x_n) \in R^n : \Phi(x_1, \ldots, x_n) = P(X_1 \leq x_1 \wedge \ldots \wedge X_n \leq x_n)$

se nazývá **distribuční funkce** náhodného vektoru $\mathbf{X}$.

Příklad

Příklad: Najděte distribuční funkci náhodné veličiny X, která udává, jaké číslo padlo při hodu kostkou a nakreslete graf této distribuční funkce.

Řešení:

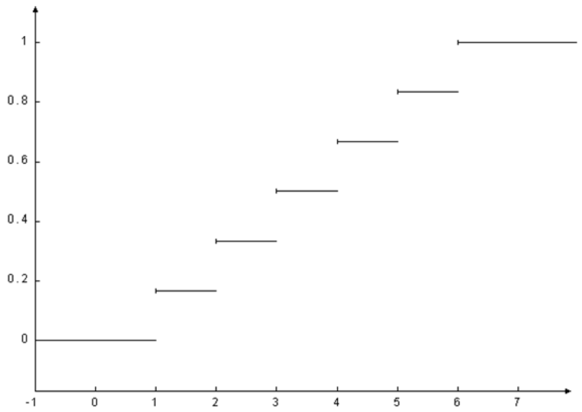
Náhodná veličina X může nabývat hodnot 1, 2, 3, 4, 5, 6. Číselnou osu tedy rozdělíme na 7 intervalů.

$x \in (-\infty, 1)$: $\Phi(x) = P(X \le x) = 0$, $x \in \langle 1, 2)$: $\Phi(x) = P(X \le x) = \frac{1}{6}$

$x \in \langle 2, 3)$: $\Phi(x) = P(X \le x) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$, $x \in \langle 3, 4)$: $\Phi(x) = P(X \le x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$

$x \in \langle 4, 5)$: $\Phi(x) = P(X \le x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6}$, $x \in \langle 5, 6)$: $\Phi(x) = P(X \le x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{5}{6}$

$x \in \langle 6, \infty)$: $\Phi(x) = P(X \le x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$



Vlastnosti distribuční funkce NV

Věta ☼:

Nechť $\Phi(x)$ je distribuční funkce skalární náhodné veličiny X. Pak $\Phi(x)$ má následující vlastnosti:

a) $\Phi(x)$ je neklesající, tj. $\forall x_1 < x_2 : \Phi(x_1) \leq \Phi(x_2)$.

b) $\Phi(x)$ je zprava spojitá, tj. pro libovolné, ale pevně dané $x_0 \in R$ je $\lim\limits_{x \to x_{0+}} \Phi(x) = \Phi(x_0)$.

c) $\Phi(x)$ je normovaná, tj. $\lim\limits_{x \to \infty} \Phi(x) = 1$, $\lim\limits_{x \to -\infty} \Phi(x) = 0$.

d) $\forall a, b \in R, a < b \Rightarrow P(a < X \leq b) = \Phi(b) - \Phi(a)$.

e) Pro libovolné, ale pevně dané $x_0 \in R : P(X = x_0) = \Phi(x_0) - \lim\limits_{x \to x_{0-}} \Phi(x)$.

Důkaz: Jenom náznakem.

ad a) Plyne z monotonie pravděpodobnosti P9.

ad b) Plyne ze spojitosti pravděpodobnosti shora P17.

ad c) $\lim\limits_{x \to -\infty} \Phi(x) = \lim\limits_{x \to -\infty} P(X \leq x) = P(\varnothing) = 0$, $\lim\limits_{x \to \infty} \Phi(x) = \lim\limits_{x \to \infty} P(X \leq x) = P(\Omega) = 1$

ad d) Plyne ze subtraktivity pravděpodobnosti P8.

ad e) Plyne ze spojitosti pravděpodobnosti zdola P16.

P9: $A_1 \subseteq A_2 \Rightarrow P(A_2) \leq P(A_1)$

P17: $A_1 \supseteq A_2 \supseteq \ldots \in \mathcal{A} \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim\limits_{i \to \infty} P(A_i)$

P8: $A_1 \subseteq A_2 \Rightarrow P(A_2 - A_1) = P(A_2) - P(A_1)$

P16: $A_1 \subseteq A_2 \subseteq \ldots \in \mathcal{A} \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim\limits_{i \to \infty} P(A_i)$

Příklad

Příklad: Náhodná veličina X udává denní počet obsazených pokojů v určitém penziónu. Známe její distribuční funkci, tj. pravděpodobnost, že bude obsazeno nejvýše x pokojů:

$$\Phi(x) = \begin{cases} 0 \;\; pro \; x < 7 \\ 0{,}02 \;\; pro \; 7 \le x < 8 \\ 0{,}05 \;\; pro \; 8 \le x < 9 \\ 0{,}12 \;\; pro \; 9 \le x < 10 \\ 1 \;\; pro \; x \ge 10 \end{cases}$$

a) Určete pravděpodobnost, že v náhodně zvolený den bude obsazeno právě 7, 8, 9, 10 pokojů.
b) Jaká je pravděpodobnost, že bude obsazeno nejvýše 10 a nejméně 8 pokojů?

Řešení:

ad a) Využijeme vlastnost (e) z věty☼.

$$P(X = 7) = \Phi(7) - \lim_{x \to 7_-} \Phi(x) = 0{,}02 - 0 = 0{,}02$$

$$P(X = 8) = \Phi(8) - \lim_{x \to 8_-} \Phi(x) = 0{,}05 - 0{,}02 = 0{,}03$$

$$P(X = 9) = \Phi(9) - \lim_{x \to 9_-} \Phi(x) = 0{,}12 - 0{,}05 = 0{,}07$$

$$P(X = 10) = \Phi(10) - \lim_{x \to 10_-} \Phi(x) = 1 - 0{,}12 = 0{,}88$$

ad b) Využijeme vlastnost (d) z věty ☼.

$$P(8 \le X \le 10) = P(7 < X \le 10) = \Phi(10) - \Phi(7) = 1 - 0{,}02 = 0{,}98$$

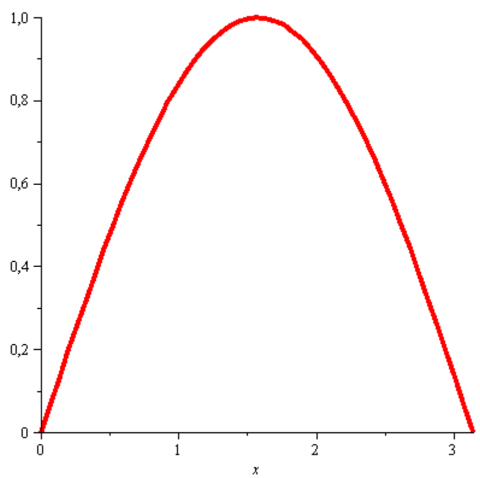
Příklad

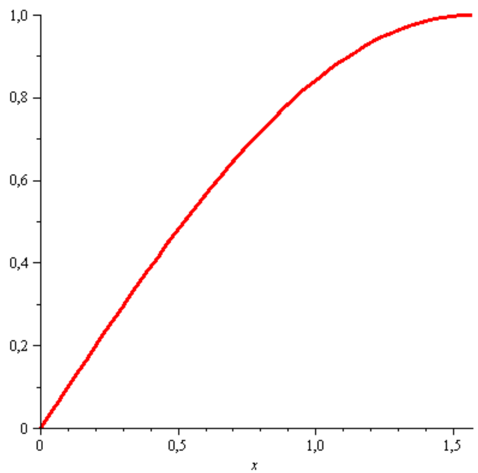
Je funkce $\Phi(x) = \sin x$ distribuční funkcí náhodné veličiny X v intervalu

$a)$ $\langle 0, \pi \rangle$,

$b)$ $\langle 0, \pi/2 \rangle$?

Řešení: a) NE b) ANO





Příklad

Určete

a) konstanty *A, B* tak, aby funkce $\Phi(x) = A + Be^{-x}$ byla distribuční funkcí náhodné veličiny pro $x \in (0, \infty)$,

b) pravděpodobnost $P(1 < X \le 4)$,

Řešení: a) $0 = \lim\limits_{x \to 0} A + Be^{-x} = A + B \lim\limits_{x \to 0} e^{-x} = A + B$

$1 = \lim\limits_{x \to \infty} A + Be^{-x} = A + \lim\limits_{x \to \infty} Be^{-x} = A$

$\Rightarrow$ $A = 1$, $B = -1$

$\Rightarrow$ $\Phi(x) = 1 - e^{-x}$

b) $P(1 < X \le 4) = \Phi(4) - \Phi(1) = \dfrac{e^3 - 1}{e^4} = 0{,}3496$

Vlastnosti distribuční funkce náhodného vektoru

Věta ☼☼: Nechť $\Phi(x_1, \ldots, x_n)$ je distribuční funkce náhodného vektoru $\mathbf{X}$. Pak $\Phi(x_1, \ldots, x_n)$ má následující vlastnosti:

a) $\Phi(x_1, \ldots, x_n)$ je neklesající vzhledem ke každé jednotlivé proměnné.

b) $\Phi(x_1, \ldots, x_n)$ je zprava spojitá vzhledem ke každé jednotlivé proměnné.

c) $\lim\limits_{\substack{x_1 \to \infty \\ \vdots \\ x_n \to \infty}} \Phi(x_1, \ldots, x_n) = 1$

$\forall i \in \{1, \ldots, n\} : \lim\limits_{x_i \to -\infty} \Phi(x_1, \ldots, x_n) = 0$

d) $\forall (x_1, \ldots, x_n) \in R^n, \forall (h_1, \ldots, h_n) \in R_+^n :$

$$P(x_1 < X_1 \leq x_1 + h_1 \wedge \ldots \wedge x_n < X_n \leq x_n + h_n) = \Phi(x_1 + h_1, \ldots, x_n + h_n) - \sum_{i=1}^{n} \Phi(x_1 + h_1, \ldots, x_i, \ldots, x_n + h_n) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Phi(x_1 + h_1, \ldots, x_i, \ldots, x_j, \ldots, x_n + h_n) - \ldots + (-1)^n \Phi(x_1, \ldots, x_n)$$

e) $\forall i \in \{1, \ldots, n\} : \lim\limits_{\substack{x_1 \to \infty \\ \vdots \\ x_{i-1} \to \infty \\ x_{i+1} \to \infty \\ \vdots \\ x_n \to \infty}} \Phi(x_1, \ldots, x_n) = \Phi_i(x_i).$

Vlastnosti distribuční funkce náhodného vektoru

Důkaz: Jenom náznakem.

ad a), ad b) Podobně jako ve skalárním případě.

ad c)

$$\lim_{\substack{x_1\to\infty\\ \vdots\\ x_n\to\infty}} \Phi(x_1,\dots,x_n) = \lim_{\substack{x_1\to\infty\\ \vdots\\ x_n\to\infty}} P(X_1 \le x_1 \wedge \dots \wedge X_n \le x_n) = P(X_1 \in R \wedge \dots \wedge X_n \in R) = P(\Omega) = 1$$

$$\forall i \in \{1,\dots,n\}: \lim_{x_i\to-\infty} \Phi(x_1,\dots,x_n) = \lim_{x_i\to-\infty} P(X_1 \le x_1 \wedge \dots \wedge X_i \le x_i \wedge \dots \wedge X_n \le x_n) = P(X_1 \le x_1 \wedge \dots \wedge \varnothing_i \wedge \dots \wedge X_n \le x_n) = P(\varnothing) = 0$$

ad d) Vlastnost vyjadřuje princip inkluze a exkluze.

ad e)

$$\lim_{\substack{x_1\to\infty\\ \vdots\\ x_{1-1}\to\infty\\ x_{1+1}\to\infty\\ \vdots\\ x_n\to\infty}} \Phi(x_1,\dots,x_n) = \lim_{\substack{x_1\to\infty\\ \vdots\\ x_{1-1}\to\infty\\ x_{1+1}\to\infty\\ \vdots\\ x_n\to\infty}} P(X_1 \le x_1 \wedge \dots \wedge X_n \le x_n) = P(X_1 \in R \wedge \dots \wedge X_i \le x_i \wedge \dots \wedge X_n \in R) = P(\Omega \wedge \dots \wedge X_i \le x_i \wedge \dots \wedge \Omega) =$$

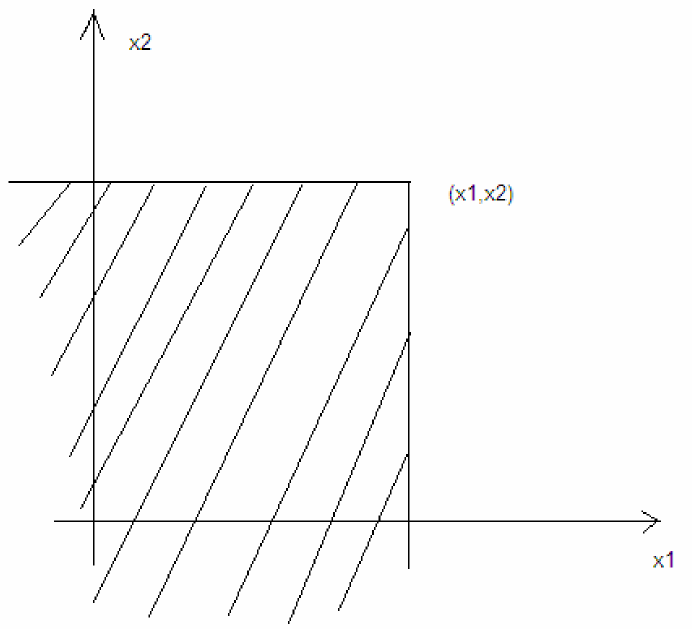
$$= P(X_i \le x_i) = \Phi_i(x_i)$$

Marginální/simultánní distribuční funkce

Funkce $\Phi_i(x_i)$ je distribuční funkce náhodné veličiny X. Nazývá se **marginální distribuční funkce** a $\Phi(x_1, \ldots, x_n)$ se v této souvislosti nazývá **simultánní distribuční funkce**. Analogicky lze zavést marginální distribuční funkce k proměnných, $k \in \{2, 3, \ldots, n-1\}$.

Vlastnosti distribuční funkce náhodného vektoru

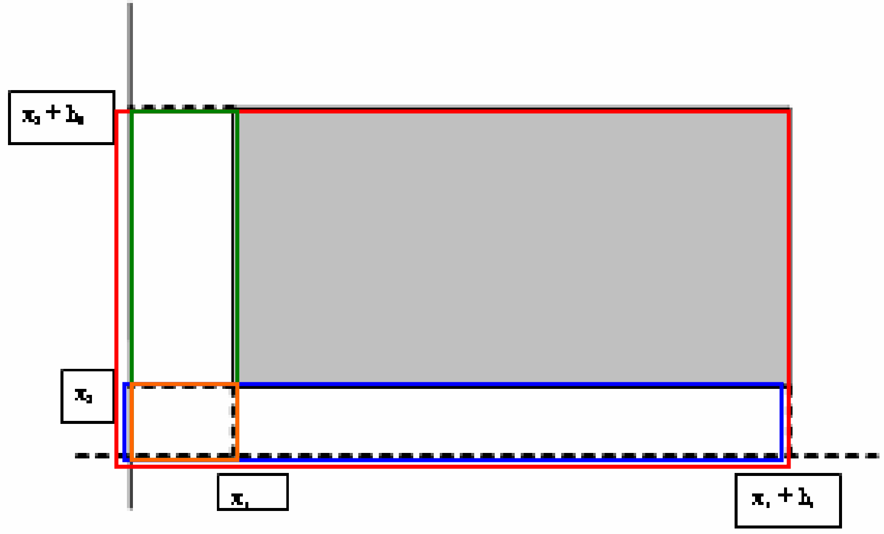
Poznámka: (Ilustrace vlastnosti (d) z věty ☼☼ pro n = 2)
Pro libovolné $(x_1, x_2) \in R^2$ udává $\Phi(x_1, x_2)$ pravděpodobnost, že náhodný vektor (X_1, X_2) se bude realizovat v oblasti $(-\infty, x_1\rangle \times (-\infty, x_2\rangle$:



Vlastnosti distribuční funkce náhodného vektoru

Pro libovolné $h_1 > 0$, $h_2 > 0$ nás zajímá pravděpodobnost, že náhodný vektor (X_1, X_2) se bude realizovat v obdélníku $(x_1, x_1 + h_1\rangle \times (x_2, x_2 + h_2\rangle$:

$$P(x_1 < X_1 \le x_1 + h_1 \wedge x_2 < X_2 \le x_2 + h_2) =$$
$$= \boxed{\Phi(x_1 + h_1, x_2 + h_2)} - \boxed{\Phi(x_1 + h_1, x_2)} - \boxed{\Phi(x_1, x_2 + h_2)} + \boxed{\Phi(x_1, x_2)}$$



Příklad

Příklad: Náhodný vektor (X_1, X_2) má distribuční funkci $\Phi(x_1, x_2) = \frac{1}{\pi^2}(\operatorname{arctg} x_1 + \frac{\pi}{2})(\operatorname{arctg} x_2 + \frac{\pi}{2})$. Vypočtěte pravděpodobnost, že náhodný vektor (X_1, X_2) se bude realizovat v jednotkovém čtverci $(0,1\rangle \times (0,1\rangle$. Najděte obě marginální distribuční funkce $\Phi_1(x_1)$, $\Phi_2(x_2)$.

Řešení:

$P(0 < X_1 \leq 1 \wedge 0 < X_2 \leq 1) = \Phi(1,1) - \Phi(1,0) - \Phi(0,1) + \Phi(0,0) =$

$= \frac{1}{\pi^2}(\frac{\pi}{4}+\frac{\pi}{2})(\frac{\pi}{4}+\frac{\pi}{2}) - \frac{1}{\pi^2}(\frac{\pi}{4}+\frac{\pi}{2})(0 + \frac{\pi}{2}) - \frac{1}{\pi^2}(0 + \frac{\pi}{2})(\frac{\pi}{4}+\frac{\pi}{2}) + \frac{1}{\pi^2}(0 + \frac{\pi}{2})(0 + \frac{\pi}{2}) = \frac{1}{16}$.

$\Phi_1(x_1) = \lim_{x_2\to\infty} \frac{1}{\pi^2}(\operatorname{arctg} x_1 + \frac{\pi}{2})(\operatorname{arctg} x_2 + \frac{\pi}{2}) = \frac{1}{\pi}(\operatorname{arctg} x_1 + \frac{\pi}{2})$

$\Phi_2(x_2) = \lim_{x_1\to\infty} \frac{1}{\pi^2}(\operatorname{arctg} x_1 + \frac{\pi}{2})(\operatorname{arctg} x_2 + \frac{\pi}{2}) = \frac{1}{\pi}(\operatorname{arctg} x_2 + \frac{\pi}{2})$

Existence distribuční funkce

Věta: (existenční věta)

a) Skalární případ: Jestliže funkce $\Phi(x)$ má vlastnosti (a), (b), (c) z věty o vlastnostech distribuční funkce skalární náhodné veličiny, pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a na něm definovaná skalární náhodná veličina X tak, že $\Phi(x)$ je její distribuční funkce.

b) Vektorový případ: Jestliže funkce $\Phi(x_1, \ldots, x_n)$ má vlastnosti (a), (b), (c) z věty o vlastnostech distribuční funkce náhodného vektoru, pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a na něm definovaný náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)$ tak, že $\Phi(x_1, \ldots, x_n)$ je jeho distribuční funkce.

8. Diskrétní a spojité NV

Motivace: Distribuční funkce popisuje pravděpodobnostní chování jakékoliv náhodné veličiny. V praxi však mají význam dva speciální typy náhodných veličin, a to diskrétní a spojité náhodné veličiny.

Diskrétní náhodná veličina nabývá nejvýše spočetně mnoha izolovaných hodnot. Je to např.
počet zásahů do terče při střelbě,
počet chyb, jichž se dopustí nějaké zařízení za určitou dobu,
počet zákazníků ve frontě apod.

Pravděpodobnostní chování diskrétní náhodné veličiny popisujeme pravděpodobnostní funkcí:

$$\forall x \in R : \pi(x) = P(X = x).$$

Je to zidealizovaný protějšek četnostní funkce zavedené v popisné statistice v souvislosti s bodovým rozložením četností:

$$\forall x \in R : p(x) = \frac{N(X = x)}{n}.$$

S rostoucím rozsahem výběrového souboru se budou hodnoty četnostní funkce ustalovat kolem hodnot pravděpodobnostní funkce.

Vlastnosti četnostní funkce se přenášejí i na pravděpodobnostní funkci, tedy pravděpodobnostní funkce

je nezáporná $\forall x \in R : \pi(x) \geq 0$,

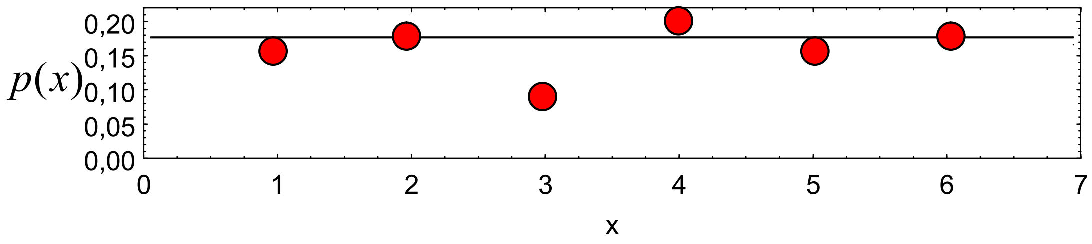
je normovaná $\sum_{x=-\infty}^{\infty} \pi(x) = 1$,

s distribuční funkcí je spjata součtovým vztahem $\forall x \in R : \Phi(x) = \sum_{t \leq x} \pi(t)$

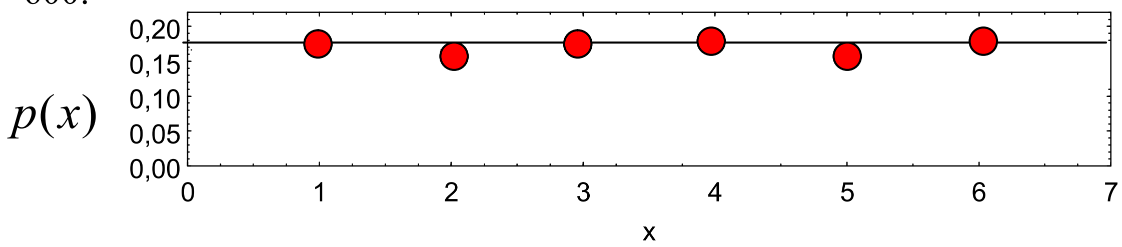
Ilustrace vztahu mezi četnostní funkcí a pravděpodobnostní funkcí

Provedeme n hodů kostkou. Zajímáme se o četnostní funkci počtu ok.

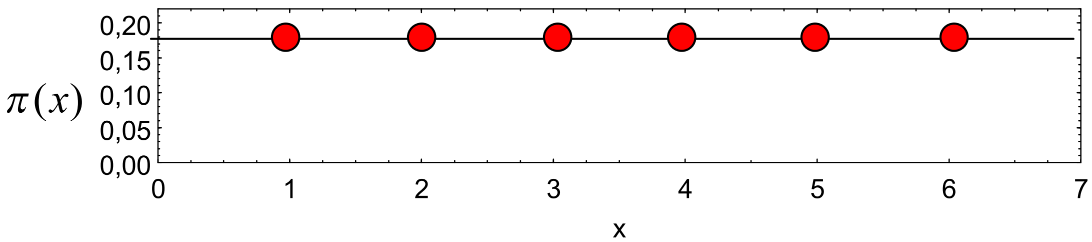
n = 60:



n = 600:



$n \to \infty$:



Spojitá náhodná veličina - motivace

Spojitá náhodná veličina nabývá všech hodnot z nějakého intervalu. Je to např.
výsledek nějakého fyzikálního či chemického měření,
hektarový výnos pšenice,
hmotnost sériově vyráběného výrobku apod.
Pravděpodobnostní chování spojité náhodné veličiny popisujeme hustotou pravděpodobnosti $\varphi(x)$, což je zidealizovaný protějšek hustoty četnosti f(x) zavedené v popisné statistice v souvislosti s intervalovým rozložením četností. S rostoucím rozsahem výběrového souboru a klesajícími šířkami třídicích intervalů se budou hodnoty hustoty četnosti ustalovat kolem hodnot hustoty pravděpodobnosti.
Vlastnosti hustoty četnosti se přenášejí i na hustotu pravděpodobnosti, tedy hustota pravděpodobnosti
je nezáporná $\forall x \in R : \varphi(x) \geq 0$,

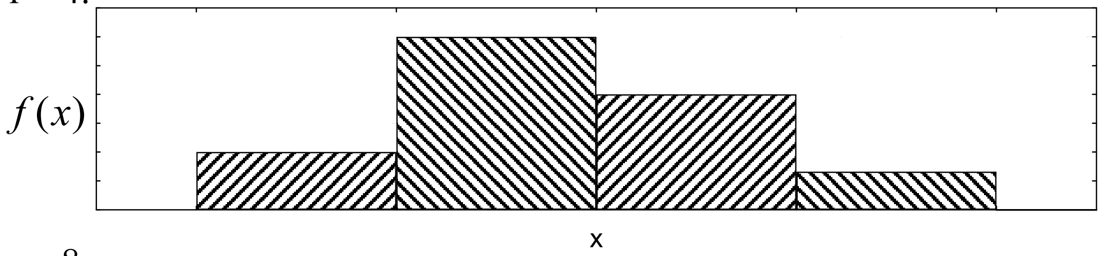
je normovaná $\int_{-\infty}^{\infty} \varphi(x) dx = 1$,

s distribuční funkcí je spjata integrálním vztahem $\forall x \in R : \Phi(x) = \int_{-\infty}^{x} \varphi(t) dt$.

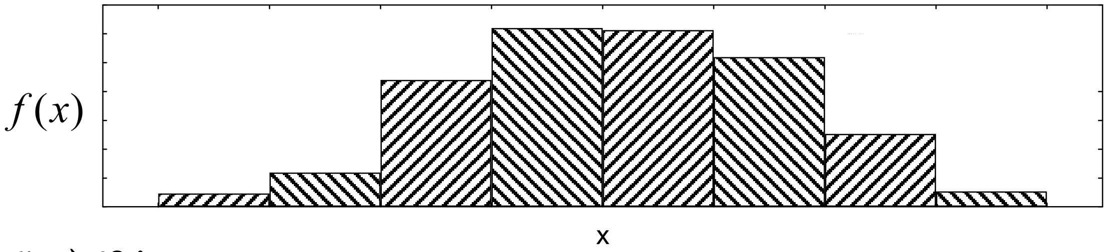
Ilustrace vztahu mezi hustotou četnosti a hustotou pravděpodobnosti

Náhodně vybereme n sériově vyráběných součástek, změříme jejich délku a budeme se zajímat o hustotu četnosti odchylek těchto měření od deklarované délky součástky.

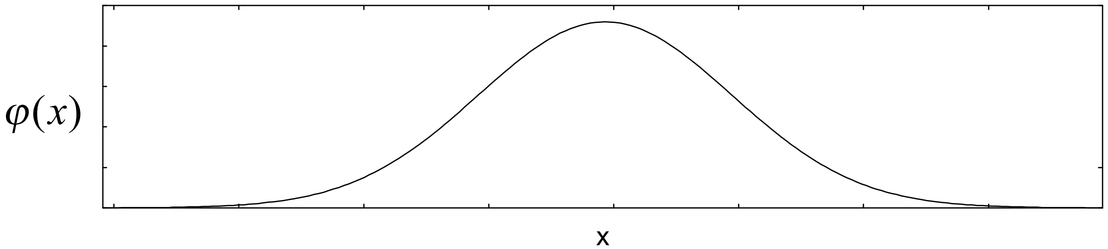
n = 40, r = 4:



n= 400, r = 8:



$n \to \infty, r \to \infty:$



Diskrétní náhodná veličina

Definice:

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, X náhodná veličina definovaná na měřitelném prostoru $(\Omega, \mathcal{A})$, která má distribuční funkci $\Phi(x)$. Řekneme, že náhodná veličina X je **diskrétní** (vzhledem k P), právě když existuje reálná funkce $\pi(x)$, která je nulová v R s výjimkou nejméně jednoho a nejvýše spočetně mnoha bodů, kde je kladná a platí pro ni: $\forall x \in R : \Phi(x) = \sum_{t \leq x} \pi(t)$. Tato funkce se nazývá **pravděpodobnostní funkce** diskrétní náhodné veličiny X.

Vlastnosti pravděpodobnostní funkce

Věta:

Nechť $\pi(x)$ je pravděpodobnostní funkce diskrétní náhodné veličiny X. Pak platí:

a) $\forall x \in R : \pi(x) \geq 0$ (vlastnost D1 - nezápornost)

b) $\sum_{x=-\infty}^{\infty} \pi(x) = 1$ (vlastnost D2 - normovanost)

c) $\forall x \in R : \pi(x) = P(X = x)$

d) $\forall B \in \mathcal{B} : P(X \in B) = \sum_{x \in B} \pi(x)$.

Důkaz:

ad a) Vlastnost D1 je součást definice.

ad b) $\sum_{x=-\infty}^{\infty} \pi(x) = \lim_{t \to \infty} \sum_{x=-\infty}^{t} \pi(x) = \lim_{t \to \infty} \Phi(t) = 1$

ad c) $P(X = x_0) = \Phi(x_0) - \lim_{x \to x_{0-}} \Phi(x) = \sum_{t \leq x_0} \pi(t) - \lim_{x \to x_{0-}} \sum_{t \leq x} \pi(t) = \lim_{x \to x_{0-}} \sum_{x < t \leq x} \pi(t) = \pi(x_0)$

ad d) Označme $G \subseteq R$ tu nejvýše spočetnou množinu, na níž $\pi(x)$ nabývá kladných hodnot. Pak pro libovolnou borelovskou množinu B platí:

$$P(X \in B) = P(X \in B \cap G) + P(X \in B \cap \overline{G}) = P\left(X \in \bigcup_{x \in B \cap G} \{x\}\right) + 0 = \sum_{x \in B \cap G} P(X = x) = \sum_{x \in B} \pi(x)$$

Příklad

Příklad: Náhodná veličina X udává počet líců při hodu dvěma mincemi. Určete její pravděpodobnostní a distribuční funkci a nakreslete jejich grafy.

Řešení:

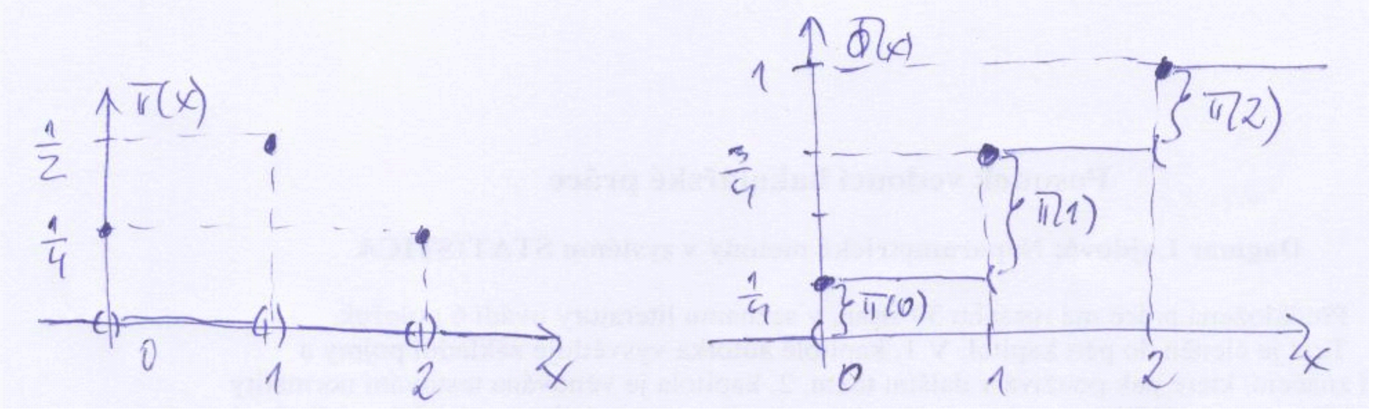
Základní prostor: $\Omega = \{[L,L][L,R][R,L][R,R]\}$, jevové pole: maximální, pravděpodobnost: klasická, náhodná veličina X nabývá hodnot z množiny {0, 1, 2}.

$$\pi(0) = P(X=0) = \frac{1}{4} \qquad x \in (-\infty, 0): \Phi(x) = 0$$

$$\pi(1) = P(X=1) = \frac{2}{4} \qquad x \in \langle 0,1): \Phi(x) = \pi(0) = \frac{1}{4}$$

$$\pi(2) = P(X=2) = \frac{1}{4} \qquad x \in \langle 1,2): \Phi(x) = \pi(0) + \pi(1) = \frac{3}{4}$$

$$\pi(x) = 0 \text{ jinak} \qquad x \in \langle 2,\infty): \Phi(x) = \pi(0) + \pi(1) + \pi(2) = 1$$



Příklad

Dva střelci (s pravděpodobnostmi zásahu p_1 a p_2) se střídají ve střelbě, dokud někdo nezasáhne. Určete pravděpodobnostní funkci počtu výstřelů.

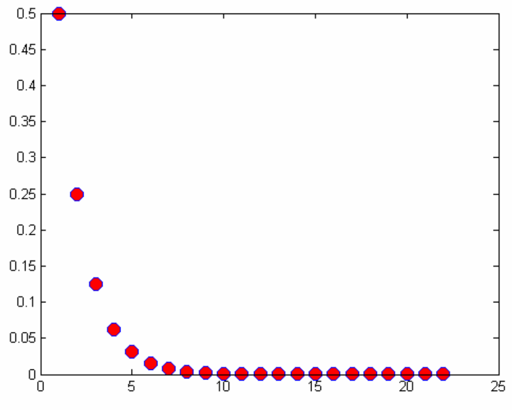
Řešení:

$$\pi(2n+1) = (1-p_1)^n (1-p_2)^n p_1$$

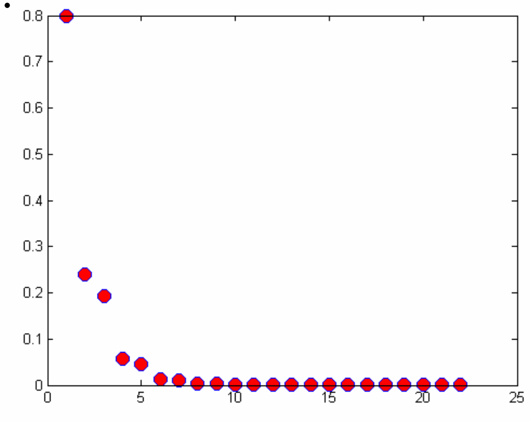
$$\pi(2n+2) = (1-p_1)^{n+1} (1-p_2)^n p_2 \quad n = 0,1,\dots$$

$$\pi(x) = 0 \text{ jinak}$$

Pro $p_1 = p_2 = 0{,}5$:



Pro $p_1 = 0{,}8$ a $p_2 = 0{,}3$:



Příklad

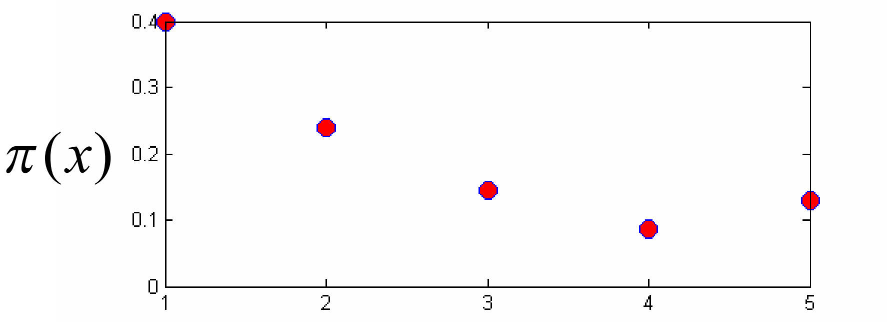
Lovec má 5 patron a pravděpodobnost zásahu 0,4. Střílí, dokud netrefí (a dokud má čím). Určete pravděpodobnostní funkci.

Řešení:

$$\pi(k) = 0{,}6^{k-1} \cdot 0{,}4 \quad k = 1, \ldots, 4$$

$$\pi(5) = 0{,}6^4$$

$$\pi(x) = 0 \text{ jinak}$$



Diskrétní náhodný vektor

Poznámka: Distribuční funkce diskrétní náhodné veličiny má schodovitý průběh. Pravděpodobnostní funkce je distribuční funkcí určena jednoznačně.

Definice:

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $\mathbf{X} = (X_1, \ldots, X_n)$ náhodný vektor definovaný na měřitelném prostoru $(\Omega, \mathcal{A})$. Nechť $\Phi(x_1, \ldots, x_n)$ je jeho distribuční funkce. Řekneme, že náhodný vektor $\mathbf{X}$ je **diskrétní** (vzhledem k P), právě když existuje reálná funkce $\pi(x_1, \ldots, x_n)$, která je nulová v R^n s výjimkou nejméně jednoho a nejvýše spočetně mnoha bodů, kde je kladná a platí pro ni:

$\forall (x_1, \ldots, x_n) \in R^n : \Phi(x_1, \ldots, x_n) = \sum_{t_1 \leq x_1} \cdots \sum_{t_n \leq x_n} \pi(t_1, \ldots, t_n)$. Tato funkce se nazývá **pravděpodobnostní funkce** diskrétního náhodného vektoru $\mathbf{X}$.

Vlastnosti pravděpodobnostní funkce

Věta:
Nechť $\pi(x_1, \ldots, x_n)$ je pravděpodobnostní funkce diskrétního náhodného vektoru $\mathbf{X}$. Pak platí:

a) $\forall (x_1, \ldots, x_n) \in R^n : \pi(x_1, \ldots, x_n) \geq 0$ (vlastnost D1 - nezápornost)

b) $\sum\limits_{x_1=-\infty}^{\infty} \ldots \sum\limits_{x_n=-\infty}^{\infty} \pi(x_1, \ldots, x_n) = 1$ (vlastnost D2 - normovanost)

c) $\forall (x_1, \ldots, x_n) \in R^n : \pi(x_1, \ldots, x_n) = P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$

d) $\forall B \in \mathcal{B}^n : P(\mathbf{X} \in B) = \sum\limits_{(x_1,\ldots,x_n) \in B} \ldots \sum \pi(x_1, \ldots, x_n)$

e) $\forall i \in \{1, \ldots, n\} : \sum\limits_{x_1=-\infty}^{\infty} \ldots \sum\limits_{x_{i-1}=-\infty}^{\infty} \sum\limits_{x_{i+1}=-\infty}^{\infty} \ldots \sum\limits_{x_n=-\infty}^{\infty} \pi(x_1, \ldots, x_n) = \pi_i(x_i)$.

Funkce $\pi_i(x_i)$ je pravděpodobnostní funkce náhodné veličiny X_i. Nazývá se **marginální pravděpodobnostní funkce**. Funkce $\pi(x_1, \ldots, x_n)$ se v této souvislosti nazývá **simultánní pravděpodobnostní funkce**. Podobně lze zavést marginální pravděpodobnostní funkce k proměnných, kde $k \in \{2, 3, \ldots, n-1\}$.

Příklad

Příklad: Je dán systém složený ze dvou bloků. Pravděpodobnost, že i-tý blok správně funguje, je υ_i, $i = 1, 2$ a pravděpodobnost, že správně fungují oba bloky, je υ_{12}. Nechť náhodná veličina X_i je ukazatel fungování i-tého bloku, tj. $X_i = \begin{cases} 1, \text{pokud i - tý blok funguje} \\ 0, \text{pokud i - tý blok nefunguje} \end{cases}$, $i = 1, 2$. Najděte simultánní pravděpodobnostní funkci $\pi(x_1, x_2)$ náhodného vektoru (X_1, X_2) a obě marginální pravděpodobnostní funkce $\pi_1(x_1)$ a $\pi_2(x_2)$.

Řešení:
Hodnoty pravděpodobnostních funkcí zapíšeme do kontingenční tabulky.

x_1	x_2		$\pi_1(x_1)$
	0	1	
0	$1-\vartheta_1-\vartheta_2+\vartheta_{12}$	$\vartheta_2-\vartheta_{12}$	$1-\vartheta_1$
1	$\vartheta_1-\vartheta_{12}$	ϑ_{12}	ϑ_1
$\pi_2(x_2)$	$1-\vartheta_2$	ϑ_2	1

$\pi(0,0) = P(X_1=0 \wedge X_2=0) = 1 - P(X_1=1 \vee X_2=1) = 1 - (\upsilon_1 + \upsilon_2 - \upsilon_{12}) = 1 - \upsilon_1 - \upsilon_2 + \upsilon_{12}$
$\pi(0,1) = P(X_1=0 \wedge X_2=1) = P(X_2=1) - P(X_1=1 \wedge X_2=1) = \upsilon_2 - \upsilon_{12}$
$\pi(1,0) = P(X_1=1 \wedge X_2=0) = P(X_1=1) - P(X_1=1 \wedge X_2=1) = \upsilon_1 - \upsilon_{12}$
$\pi(1,1) = P(X_1=1 \wedge X_2=1) = \upsilon_{12}$
$\pi(x_1,x_2) = 0$ jinak

Existenční věta

Věta (existenční věta)

a) Skalární případ: Jestliže funkce $\pi(x)$ má vlastnosti D1, D2 z věty o vlastnostech pravděpodobnostní funkce skalární náhodné veličiny, pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a na něm definovaná skalární diskrétní náhodná veličina X tak, že $\pi(x)$ je její pravděpodobnostní funkce.

b) Vektorový případ: Jestliže funkce $\pi(x_1, \ldots, x_n)$ má vlastnosti D1, D2 z věty o vlastnostech pravděpodobnostní funkce náhodného vektoru, pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a na něm definovaný diskrétní náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)$ tak, že $\pi(x_1, \ldots, x_n)$ je jeho pravděpodobnostní funkce.

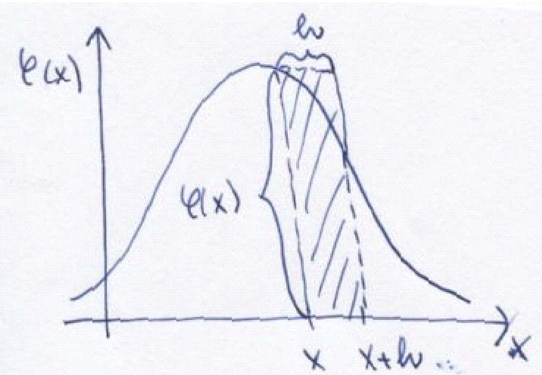
Spojitá náhodná veličina

Definice:

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, X náhodná veličina definovaná na měřitelném prostoru $(\Omega, \mathcal{A})$, která má distribuční funkci $\Phi(x)$. Řekneme, že náhodná veličina X je **spojitá** (vzhledem k P), právě když existuje po částech spojitá nezáporná reálná funkce $\varphi(x)$ tak, že pro $\forall x \in R : \Phi(x) = \int\limits_{-\infty}^{x} \varphi(t)dt$. Tato funkce se nazývá **hustota pravděpodobnosti** spojité náhodné veličiny X.

Spojitá náhodná veličina - poznámka

Poznámka: Na rozdíl od pravděpodobnostní funkce diskrétní náhodné veličiny nemá hustota pravděpodobnosti spojité náhodné veličiny význam pravděpodobnosti. Její význam lze odvodit z integrálního vztahu mezi distribuční funkcí a hustotou pravděpodobnosti.



Pravděpodobnost, že náhodná veličina se bude realizovat v intervalu (x, x+h], je:

$$\boxed{P(x < X \le x + h) = \Phi(x + h) - \Phi(x) = \int_{-\infty}^{x+h} \varphi(t)dt - \int_{-\infty}^{x} \varphi(t)dt = \int_{x}^{x+h} \varphi(t)dt}$$

Bude-li h dostatečně malé číslo, lze plochu pod grafem hustoty nahradit obsahem obdélníka o stranách φ(x) a h, tj.

$$P(x < X \le x + h) \approx \varphi(x) \cdot h$$

Vlastnosti hustoty pravděpodobnosti

Věta:

Nechť $\varphi(x)$ je hustota spojité náhodné veličiny X. Pak platí:

a) $\forall x \in R : \varphi(x) \geq 0$ (vlastnost S1 - nezápornost)

b) $\int\limits_{-\infty}^{\infty} \varphi(x)dx = 1$ (vlastnost S2 - normovanost)

c) $\forall x \in R, \forall h > 0 : P(x < X \leq x + h) = \int\limits_{x}^{x+h} \varphi(t)dt$

d) Pro libovolné, ale pevně dané $x \in R : P(X = x) = 0$.

e) $\varphi(x) = \frac{d\Phi(x)}{dx}$ ve všech bodech spojitosti funkce $\varphi(x)$.

Důkaz:

ad a) Vlastnost S1 je součást definice.

ad b) $\int\limits_{-\infty}^{\infty} \varphi(x)dx = \lim\limits_{x\to\infty} \int\limits_{-\infty}^{x} \varphi(t)dt = \lim\limits_{x\to\infty} \Phi(x) = 1$

ad c) $\int\limits_{x}^{x+h} \varphi(t)dt = \int\limits_{-\infty}^{x+h} \varphi(t)dt - \int\limits_{-\infty}^{x} \varphi(t)dt = \Phi(x+h) - \Phi(x) = P(x < X \leq x + h)$

ad d) $P(X = x) = \int\limits_{x}^{x} \varphi(t)dt = 0$

ad e) $\frac{d\Phi(x)}{dx} = \frac{d}{dx} \int\limits_{-\infty}^{x} \varphi(t)dt = \varphi(x)$ ve všech bodech spojitosti funkce $\varphi(x)$.

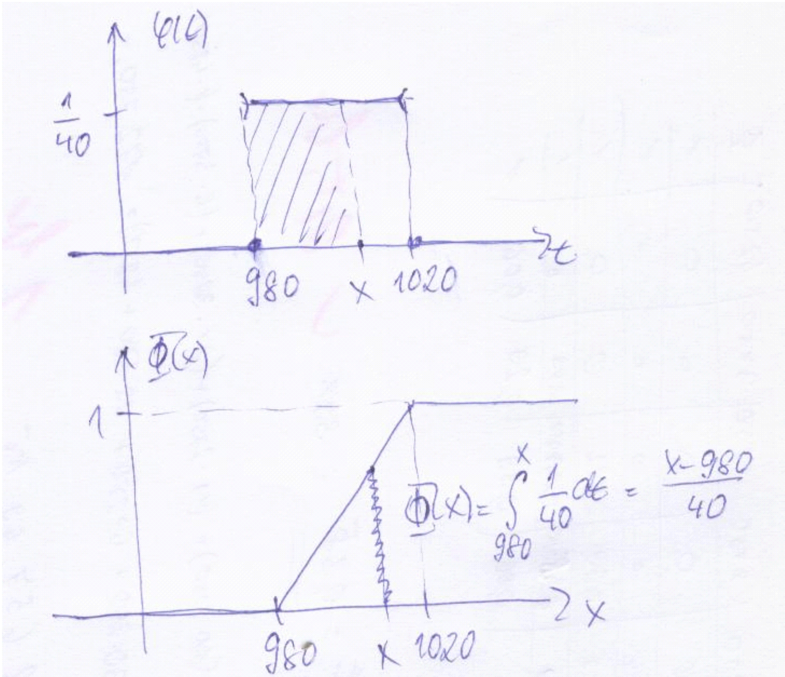
Příklad

Příklad: Na automatické lince se plní láhve mlékem. Každá láhev má obsahovat přesně 1000 ml mléka, ale v důsledku působení náhodných vlivů množství mléka kolísá v intervalu (980 ml, 1020 ml). Každé množství mléka v tomto intervalu považujeme za stejně možné. Náhodná veličina X udává množství mléka v náhodně vybrané lahvi. Najděte její hustotu pravděpodobnosti φ(x) a distribuční funkci Φ(x). Jaká je pravděpodobnost, že v náhodně vybrané láhvi bude aspoň 1000 ml mléka?

Řešení: $\varphi(x) = \begin{cases} k \text{ pro } x \in (980, 1020) \\ 0 \text{ jinak} \end{cases}$

Z normovanosti hustoty plyne: $1 = \int_{980}^{1020} k\,dx = 40\,k$, tedy $k = \frac{1}{40}$.

Pro distribuční funkci platí: $\Phi(x) = \begin{cases} 0 \text{ pro } x \le 980 \\ \int_{980}^{x} \frac{1}{40}\,dt = \frac{x-980}{40} \text{ pro } 980 < x < 1020 \\ 1 \text{ pro } x \ge 1020 \end{cases}$



$$P(X \ge 1000) = \int_{1000}^{1020} \frac{1}{40}\,dx = \frac{1}{40}[x]_{1000}^{1020} = \frac{20}{40} = 0{,}5$$

Příklad

Napište distribuční funkci rozdělení daného hustotou f(x) = x/2 na (0, 1), 1/2 na (1, 2), (3 − x)/2 na (2, 3).

Řešení: Na $(0,1)$:

$$F(x)=\int_0^x f(t)\,dt=\int_0^x \frac{t}{2}\,dt=\frac{1}{2}\left[\frac{t^2}{2}\right]_0^x=\frac{x^2}{4},$$

Na $(1,2)$:

$$F(x)=\frac{1}{4}+\frac{1}{2}(x-1),$$

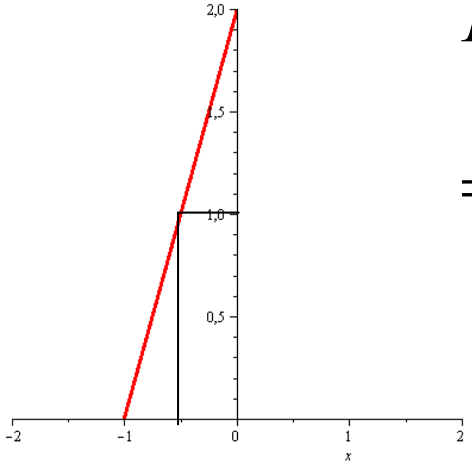
Na $(2,3)$:

$$F(x)=\frac{1}{4}+\frac{1}{2}+\int_2^x \frac{3-t}{2}\,dt=\frac{3}{4}-\frac{1}{2}\left[\frac{(3-t)^2}{2}\right]_2^x=1-\frac{(3-x)^2}{4}.$$

Příklad

Rozdělení náhodné veličiny X je dáno hustotou f(x) = 2x+2, na (−1, 0) a nulovou jinde.
Najděte $P(-2 \leq X \leq -0{,}5)$.

Řešení:



$$P(-2 \leq X \leq -0{,}5) = P(-1 \leq X \leq -0{,}5) =$$

$$= \int_{-1}^{-0{,}5} (2x+2)\,dx = \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4}$$

Příklad

Náhodná veličina X má hustotu $f(x) = \dfrac{a}{1+x^2}$ na R.

Určete a, distribuční funkci, $P\left(X > \sqrt{3}\right)$.

Řešení:

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = a\left(\lim_{x\to\infty} arctg(x) - \lim_{x\to-\infty} arctg(x)\right) =$$

$$= a\left(\frac{\pi}{2} + \frac{\pi}{2}\right) \qquad \Rightarrow a = \frac{1}{\pi}$$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \frac{1}{\pi}\left(arctg(x) + \frac{\pi}{2}\right)$$

$$P\left(X > \sqrt{3}\right) = 1 - F\left(\sqrt{3}\right) = 1 - \frac{1}{\pi}\left(\frac{\pi}{3} + \frac{\pi}{2}\right) = \frac{1}{6}$$

Spojitý náhodný vektor

Definice:

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $\mathbf{X} = (X_1, \ldots, X_n)$ náhodný vektor definovaný na měřitelném prostoru $(\Omega, \mathcal{A})$. Nechť $\Phi(x_1, \ldots, x_n)$ je jeho distribuční funkce. Řekneme, že náhodný vektor $\mathbf{X}$ je **spojitý** (vzhledem k P), právě když existuje po částech spojitá nezáporná reálná funkce $\varphi(x_1, \ldots, x_n)$ tak, že pro

$\forall (x_1, \ldots, x_n) \in R^n : \Phi(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} \varphi(t_1, \ldots, t_n) dt_1 \ldots dt_n$. Tato funkce se nazývá **hustota pravděpodobnosti** spojitého náhodného vektoru $\mathbf{X}$.

Vlastnosti hustoty pravděpodobnosti

Věta: Nechť $\varphi(x_1,\ldots,x_n)$ je hustota pravděpodobnosti spojitého náhodného vektoru $\mathbf{X} = (X_1, \ldots, X_n)$. Pak platí:

a) $\forall (x_1,\ldots,x_n) \in R^n : \varphi(x_1,\ldots,x_n) \geq 0$ (vlastnost S1 - nezápornost)

b) $\int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} \varphi(x_1,\ldots,x_n)dx_1 \ldots dx_n = 1$ (vlastnost S2 - normovanost)

c) $\forall B \in \mathcal{B}^n : P(\mathbf{X} \in B) = \int \ldots \int\limits_B \varphi(x_1,\ldots,x_n)dx_1 \ldots dx_n$

d) $\varphi(x_1,\ldots,x_n) = \frac{\partial^n \Phi(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n}$ ve všech bodech spojitosti funkce $\varphi(x_1,\ldots,x_n)$.

e) $\forall i \in \{1,\ldots,n\} : \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} \varphi(x_1,\ldots,x_n)dx_1 \ldots dx_{i-1}dx_{i+1} \ldots dx_n = \varphi_i(x_i)$.

Funkce $\varphi_i(x_i)$ je hustota náhodné veličiny X_i. Nazývá se **marginální hustota**. Funkce $\varphi(x_1,\ldots,x_n)$ se v této souvislosti nazývá **simultánní hustota**. Podobně lze zavést marginální hustoty k proměnných, kde $k \in \{2, 3, \ldots, n-1\}$.

Existenční věta

Věta: (existenční věta)

a) Skalární případ: Jestliže funkce $\varphi(x)$ má vlastnosti S1, S2 z věty o vlastnostech hustoty skalární náhodné veličiny, pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a na něm definovaná skalární spojitá náhodná veličina X tak, že $\varphi(x)$ je její hustota.

b) Vektorový případ: Jestliže funkce $\varphi(x_1, \dots, x_n)$ má vlastnosti S1, S2 z věty o vlastnostech hustoty náhodného vektoru, pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a na něm definovaný spojitý náhodný vektor $\mathbf{X} = (X_1, \dots, X_n)$ tak, že $\varphi(x_1, \dots, x_n)$ je jeho hustota.

Příklad

Příklad: Spojitý náhodný vektor (X_1, X_2) má simultánní hustotu pravděpodobnosti $\varphi(x_1, x_2) = \dfrac{1}{\pi^2(1+x_1^2)(1+x_2^2)}$.
Najděte obě marginální hustoty $\varphi_1(x_1)$, $\varphi_2(x_2)$.

Řešení:

$$\varphi_1(x_1) = \int_{-\infty}^{\infty} \frac{1}{\pi^2(1+x_1^2)(1+x_2^2)}\,dx_2 = \frac{1}{\pi^2(1+x_1^2)} \int_{-\infty}^{\infty} \frac{1}{1+x_2^2}\,dx_2 =$$

$$= \frac{1}{\pi^2(1+x_1^2)} \left[\operatorname{arctg} x_2\right]_{-\infty}^{\infty} = \frac{1}{\pi^2(1+x_1^2)} \left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = \frac{1}{\pi(1+x_1^2)}.$$

Analogicky dostáváme $\varphi_2(x_2) = \dfrac{1}{\pi(1+x_2^2)}$.

9. Stochasticky nezávislé NV. Vybraná rozložení.

Motivace: Při provedení pokusu se může stát, že se realizace jedné náhodné veličiny Y dají jednoznačně určit ze známé realizace druhé náhodné veličiny X, tedy je mezi nimi funkční vztah Y = g(X). Takové náhodné veličiny se nazývají deterministicky závislé.

Jejich protipólem jsou náhodné veličiny stochasticky nezávislé: informace o realizaci jedné z nich nijak nemění šance, s nimiž při témž pokusu očekáváme realizaci druhé.

Např. náhodný pokus spočívá v hodu dvěma kostkami. Náhodná veličina X udává počet ok, která padla na 1. kostce a náhodná veličina Y udává počet ok, která padla na druhé kostce. Náhodné veličiny X, Y jsou stochasticky nezávislé.

Stochastickou nezávislost náhodných veličin zavádíme na základě analogie s četnostní nezávislostí znaků v daném výběrovém souboru, která se používá v popisné statistice. Musí platit multiplikativní vztah:

$\forall(x,y)\in R^2 : p(x,y) = p_1(x)p_2(y)$ pro bodové rozložení četností,

$\forall(x,y)\in R^2 : f(x,y) = f_1(x)f_2(y)$ pro intervalové rozložení četností.

V počtu pravděpodobnosti nahradíme četnostní funkci pravděpodobnostní funkcí resp. hustotu četnosti nahradíme hustotou pravděpodobnosti. Místo dvou náhodných veličin X, Y můžeme uvažovat n náhodných veličin:

Náhodné veličiny $X_1, \dots, X_n$ jsou stochasticky nezávislé, když platí:

$\forall(x_1,\dots,x_n)\in R^n : \pi(x_1,\dots,x_n) = \pi_1(x_1)\cdot\ldots\cdot\pi_n(x_n)$ v diskrétním případě,

$\forall(x_1,\dots,x_n)\in R^n : \varphi(x_1,\dots,x_n) = \varphi_1(x_1)\cdot\ldots\cdot\varphi_n(x_n)$ ve spojitém případě,

$\forall(x_1,\dots,x_n)\in R^n : \Phi(x_1,\dots,x_n) = \Phi_1(x_1)\cdot\ldots\cdot\Phi_n(x_n)$ v obecném případě.

Stochasticky nezávislé náhodné veličiny

Definice:

a) Obecný případ: Řekneme, že náhodné veličiny $X_1, \dots, X_n$ s marginálními distribučními funkcemi $\Phi_1(x_1), \dots, \Phi_n(x_n)$ a simultánní distribuční funkcí $\Phi(x_1, \dots, x_n)$ jsou **stochasticky nezávislé**, právě když
$\forall (x_1, \dots, x_n) \in R^n : \Phi(x_1, \dots, x_n) = \Phi_1(x_1) \dots \Phi_n(x_n)$.

b) Diskrétní případ: Řekneme, že diskrétní náhodné veličiny $X_1, \dots, X_n$ s marginálními pravděpodobnostními funkcemi $\pi_1(x_1), \dots, \pi_n(x_n)$ a simultánní pravděpodobnostní funkcí $\pi(x_1, \dots, x_n)$ jsou **stochasticky nezávislé**, právě když
$\forall (x_1, \dots, x_n) \in R^n : \pi(x_1, \dots, x_n) = \pi_1(x_1) \dots \pi_n(x_n)$.

c) Spojitý případ: Řekneme, že spojité náhodné veličiny $X_1, \dots, X_n$ s marginálními hustotami $\varphi_1(x_1), \dots, \varphi_n(x_n)$ a simultánní hustotou $\varphi(x_1, \dots, x_n)$ jsou **stochasticky nezávislé**, právě když
$\forall (x_1, \dots, x_n) \in R^n : \varphi(x_1, \dots, x_n) = \varphi_1(x_1) \dots \varphi_n(x_n)$ s případnou výjimkou na množině bodů neovlivňujících integraci.

Definice:

Řekneme, že posloupnost $\{X_n\}_{n=1}^{\infty}$ je **posloupnost stochasticky nezávislých náhodných veličin**, právě když pro všechna přirozená n jsou stochasticky nezávislé náhodné veličiny $X_1, \dots, X_n$.

Příklad

Příklad: Diskrétní náhodný vektor (X_1, X_2) má simultánní pravděpodobnostní funkci $\pi(x_1, x_2)$ danou hodnotami: $\pi(0,0) = \pi(0,2) = \pi(1,1) = \pi(2,0) = \pi(2,2) = 0$, $\pi(0,1) = \pi(1,0) = \pi(1,2) = \pi(2,1) = 0{,}25$. Jsou náhodné veličiny X_1, X_2 stochasticky nezávislé?

Řešení:

Sestavíme kontingenční tabulku, v níž budou hodnoty simultánní pravděpodobnostní funkce a obou marginálních pravděpodobnostních funkcí.

x_1	x_2			$\pi_1(x_1)$
	0	1	2	
0	0	0,25	0	0,25
1	0,25	0	0,25	0,5
2	0	0,25	0	0,25
$\pi_2(x_2)$	0,25	0,5	0,25	1

Ověříme splnění multiplikativního vztahu $\forall (x_1, x_2) \in R^2$: $\pi(x_1, x_2) = \pi_1(x_1)\,\pi_2(x_2)$. Již pro $x_1 = 0$, $x_2 = 0$ vztah splněn není, protože $\pi(0,0) = 0$, avšak $\pi_1(0) = 0{,}25$ a $\pi_2(0) = 0{,}25$. Veličiny X_1, X_2 tedy nejsou stochasticky nezávislé.

Příklad

Příklad:

Nechť spojitý vektor (X_1, X_2) má simultánní hustotu pravděpodobnosti

$\varphi(x_1, x_2) = \begin{cases} 24\, x_1^{\,2} x_2 (1 - x_1) \text{ pro } 0 \le x_1 < 1, 0 \le x_2 < 1 \\ 0 \text{ jinak} \end{cases}$. Dokažte, že náhodné veličiny X_1, X_2 jsou stochasticky nezávislé.

Řešení:

Vypočítáme obě marginální hustoty a ověříme platnost multiplikativního vztahu

$\forall (x_1, ..., x_n) \in R^n$: $\varphi(x_1, ..., x_n) = \varphi_1(x_1) \ldots \varphi_n(x_n)$ s případnou výjimkou na množině bodů neovlivňujících integraci.

$$\varphi_1(x_1) = \int_0^1 24x_1^{\,2} x_2 (1 - x_1)\, dx_2 = 24x_1^2 (1\text{-}x_1) \left[\frac{x_2^{\,2}}{2}\right]_0^1 = 12x_1^2 (1\text{-}x_1) \text{ pro } 0 \le x_1 < 1,$$

$\varphi_1(x_1) = 0$ jinak.

$$\varphi_2(x_2) = \int_0^1 24x_1^{\,2} x_2 (1 - x_1)\, dx_1 = 24x_2 \left[\frac{x_1^{\,3}}{3} - \frac{x_1^{\,4}}{4}\right]_0^1 = 2x_2 \text{ pro } 0 \le x_2 < 1,$$

$\varphi_2(x_2) = 0$ jinak.

Vidíme, že multiplikativní vztah je splněn, tudíž veličiny X_1, X_2 jsou stochasticky nezávislé.

Stochasticky nezávislé náhodné vektory

Definice:

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $\mathbf{X}_1 = (X_{11}, \ldots, X_{p_1 1}), \ldots, \mathbf{X}_n = (X_{1n}, \ldots, X_{p_n n})$ náhodné vektory definované na $(\Omega, \mathcal{A}, P)$. Řekneme, že tyto náhodné vektory jsou **stochasticky nezávislé**, právě když každá složka náhodného vektoru $\mathbf{X}_i$ je stochasticky nezávislá se všemi složkami náhodného vektoru $\mathbf{X}_k$ pro $\forall i \neq k$.

Věta:

Nechť $X_1, \ldots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $g_1, \ldots, g_n$ borelovské funkce. Pak transformované náhodné veličiny $Y_1 = g_1(X_1), \ldots, Y_n = g_n(X_n)$ jsou opět stochasticky nezávislé náhodné veličiny.

(Tvrzení lze zobecnit i pro transformované náhodné vektory.)

Příklad

Příklad:

Nechť $X_1,\dots,X_n$ jsou stochasticky nezávislé náhodné veličiny s distribučními funkcemi $\Phi_1(x_1),\dots,\Phi_n(x_n)$. Zavedeme transformované náhodné veličiny $Y=\max\{X_1,\dots,X_n\}$, $Z=\min\{X_1,\dots,X_n\}$. Odvoďte jejich distribuční funkce $\Phi_{\max}(y),\Phi_{\min}(z)$.

Řešení:

$$\Phi_{\max}(y)=P(Y\le y)=P(\max\{X_1,\dots,X_n\}\le y)=P(X_1\le y\wedge\dots\wedge X_n\le y)=P(X_1\le y)\cdot\ldots\cdot P(X_n\le y)=\Phi_1(y)\cdot\ldots\cdot\Phi_n(y)$$

$$\Phi_{\min}(z)=P(Z\le z)=P(\min\{X_1,\dots,X_n\}\le z)=P(X_1\le z\vee\dots\vee X_n\le z)=1-P(X_1>z\wedge\dots\wedge X_n>z)=1-P(X_1>z)\cdot\ldots\cdot P(X_n>z)=$$
$$=1-[1-P(X_1\le z)]\cdot\ldots\cdot[1-P(X_n\le z)]=1-[1-\Phi_1(z)]\cdot\ldots\cdot[1-\Phi_n(z)]$$

Příklad

Příklad:

Na automatické lince jsou láhve plněny mlékem. Je známo, že množství mléka v láhvích kolísá od od 0,98 l do 1,02 l. V tomto intervalu považujeme každé množství mléka za stejně možné. Za 1 s se naplní 3 láhve. Jaká je pravděpodobnost, že

a) nejméně naplněná láhev obsahuje aspoň 1 l mléka,

b) v nejvíce naplněné láhvi není víc než 1,01 l mléka?

Řešení:

Náhodná veličina X_i udává množství mléka v i-té láhvi, i = 1, 2, 3. Je to spojitá náhodná veličina, její hustota pravděpodobnosti je konstantní na intervalu (0,98; 1,02). Z podmínky normovanosti S2 dostaneme, že hustota

$$\varphi(x) = \begin{cases} \frac{1}{40} \text{ pro } x \in (980, 1020) \\ 0 \text{ jinak} \end{cases}. \quad \text{Distribuční funkce: } \Phi(x) = \begin{cases} 0 \quad \text{pro } x \in (-\infty, 980\rangle \\ \int_{980}^{x} \frac{1}{40} dt = \frac{1}{40}[x]_{980}^{x} = \frac{x-980}{40} \quad \text{pro } x \in (980, 1020) \\ 1 \quad \text{pro } x \in \langle 1020, \infty) \end{cases}$$

ad a) $P(Z \geq 1000) = 1 - P(Z < 1000) = 1 - \Phi_{\min}(1000) = [1 - \Phi(1000)]^3 = \left[1 - \frac{1000-980}{40}\right]^3 = \left(\frac{1}{2}\right)^3 = \frac{1}{8} = 0{,}125$

ad b) $P(Y \leq 1010) = \Phi_{\max}(1010) = [\Phi(1010)]^3 = \left(\frac{3}{4}\right)^3 = \frac{27}{64} = 0{,}42$

Vybraná rozložení diskrétních a spojitých náhodných veličin

Motivace

Nyní se seznámíme s přehledem důležitých pravděpodobnostních funkcí a hustot pravděpodobnosti. Uvedeme nejenom analytické vyjádření těchto funkcí, ale též jejich grafy. Vysvětlíme rovněž, v jakých situacích se lze s uvedenými rozloženími pravděpodobností setkat. Zvláštní pozornost budeme věnovat normálnímu rozložení, které hraje velkou roli v celé řadě praktických aplikací počtu pravděpodobnosti i v matematické statistice.

Označení

Známe-li distribuční funkci $\Phi(x)$ náhodné veličiny X (resp. pravděpodobnostní funkci $\pi(x)$ v diskrétním případě resp. hustotu pravděpodobnosti $\varphi(x)$ ve spojitém případě), pak řekneme, že známe rozložení pravděpodobností (zkráceně rozložení) náhodné veličiny X. Toto rozložení závisí na nějakém parametru ϑ, což je nejčastěji reálné číslo nebo reálný vektor.

Zápis $\boxed{X \sim L(\vartheta)}$ čteme: <u>náhodná veličina X má rozložení L s parametrem ϑ</u>.

<u>Na webu:</u>

<u>http://en.wikipedia.org/wiki/List_of_probability_distributions</u>

Vybraná rozložení diskrétních náhodných veličin

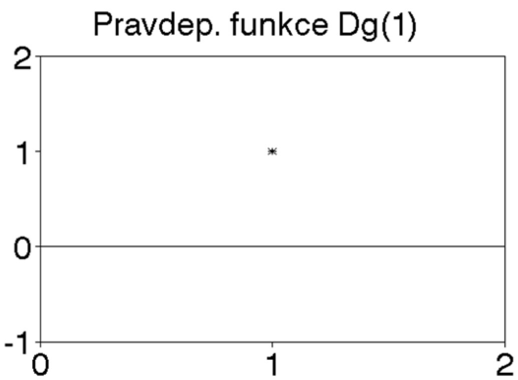
Důležitá diskrétní rozdělení:

- Degenerované rozložení
- Alternativní (Bernoulliho) rozdělení
- Binomické rozdělení
- Multinomické rozdělení
- Poissonovo rozdělení
- Negativně binomické (Pascalovo) rozdělení
- Geometrické rozdělení (zvláštní případ negativně binomického rozdělení)
- Hypergeometrické rozdělení
- Rovnoměrné rozdělení

Degenerované rozložení

Degenerované rozložení: Náhodná veličina X nabývá pouze konstantní hodnoty μ, píšeme $X \sim Dg(\mu)$.

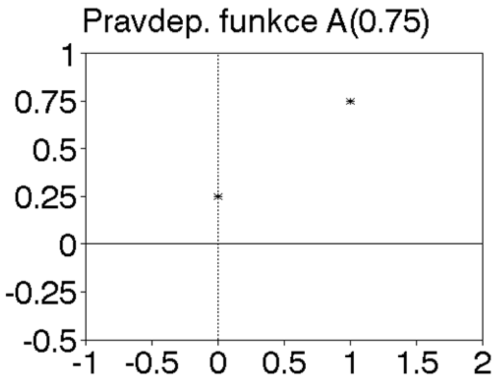
$$\pi(x) = \begin{cases} 1 \text{ pro } x = \mu \\ 0 \text{ jinak} \end{cases}$$



Alternativní rozložení

Alternativní rozložení: Náhodná veličina X udává počet úspěchů v jednom pokusu, přičemž pravděpodobnost úspěchu je ϑ. Píšeme $X \sim A(\vartheta)$.

$$\pi(x) = \begin{cases} 1-\vartheta \text{ pro } x=0 \\ \vartheta \text{ pro } x=1 \\ 0 \text{ jinak} \end{cases} \quad \text{neboli } \pi(x) = \begin{cases} \vartheta^x (1-\vartheta)^{1-x} \text{ pro } x = 0,1 \\ 0 \text{ jinak} \end{cases}$$



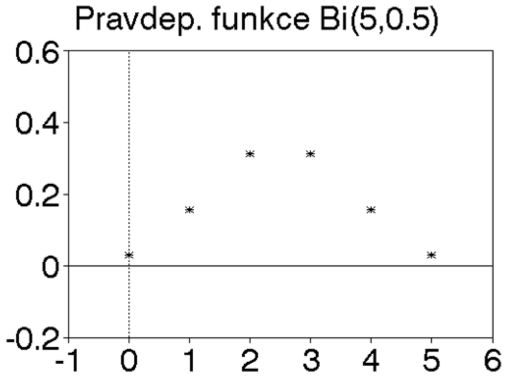
Binomické rozložení

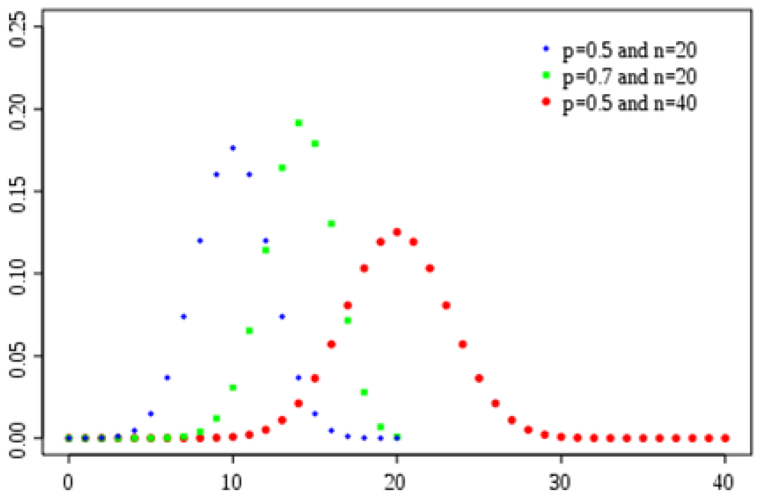
Binomické rozložení: Náhodná veličina X udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ. Píšeme $X \sim Bi(n, \vartheta)$.

$$\pi(x) = \begin{cases} \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} & \text{pro } x = 0, \ldots, n \\ 0 \text{ jinak} \end{cases}$$

(Alternativní rozložení je speciálním případem binomického rozložení pro n = 1.

Jsou-li $X_1, \ldots, X_n$ stochasticky nezávislé náhodné veličiny, $X_i \sim A(\vartheta)$, i = 1, ..., n, pak $X = \sum_{i=1}^{n} X_i \sim Bi(n, \vartheta)$.)





Příklad

Příklad na binomické rozložení pravděpodobností: Firma se účastní čtyř nezávislých výběrových řízení. Pravděpodobnost, že uspěje v kterémkoliv z nich, je pro všechny konkurzy stejná a je rovna 0,7. Jaká je pravděpodobnost, že firma uspěje

a) právě 2x
b) aspoň 2x
c) nejvýše 2x?

Řešení: X … počet úspěšných konkurzů, X ~ Bi(4; 0,7)

ad a) $P(X=2)=\pi(2)=\binom{4}{2}0{,}7^2 0{,}3^2 = 0{,}2646$

ad b) $P(X\geq 2)=\pi(2)+\pi(3)+\pi(4)=\binom{4}{2}0{,}7^2 0{,}3^2+\binom{4}{3}0{,}7^3 0{,}3+\binom{4}{4}0{,}7^4 = 0{,}9163$

ad c) $P(X\leq 2)=\Phi(2)=\pi(0)+\pi(1)+\pi(2)=\binom{4}{0}0{,}3^4+\binom{4}{1}0{,}7\cdot 0{,}3^3+\binom{4}{2}0{,}7^2 0{,}3^2 = 0{,}3483$

Příklad

Pravděpodobnost narození chlapce je 0,515. Určete takový počet dětí, aby pravděpodobnost, že mezi nimi bude aspoň jeden chlapec, byla vetší než 0,99.

Řešení:

Označme jako X veličinu udávající počet chlapců mezi n dětmi, je X~Bi(n,0,515). Hledáme takové n, aby P(X > 0) > 0,99, přitom platí

$$P(X>0)=1-P(X\leq 0)=1-P(X=0)=1-\binom{n}{0}\cdot 0{,}515^{0}\cdot(1-0{,}515)^{n-0}$$

$$\Rightarrow \quad 1-(0{,}485)^{n}>0{,}99$$

$$\Rightarrow \quad n>\frac{\ln 0{,}01}{\ln 0{,}485}\cong 6{,}36$$

$$\Rightarrow \quad n\geq 7$$

Multinomické rozložení

Multinomické rozložení: Zobecnění binomického rozložení. Složky náhodného vektoru $(X_1, \ldots X_k)$ udávají počty úspěchů (nastane jev $A_1, \ldots A_k$) v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnosti úspěchů jsou $\vartheta_1, \ldots, \vartheta_k$. Předpokládáme, že při každém pokusu nastane právě jeden z jevů $A_1, \ldots A_k$, přičemž platí $\vartheta_1 + \ldots + \vartheta_k = 1$. Píšeme $X \sim Mu(n, \vartheta_1, \ldots, \vartheta_k)$

$$\pi(x_1, \ldots, x_k) = \frac{n!}{x_1! \ldots x_k!} \vartheta_1^{x_1} \cdot \ldots \cdot \vartheta_k^{x_k}, \quad x_1, \ldots, x_k \in \{1, \ldots, n\}, \sum_{i=1}^{k} x_i = n$$

$$= 0 \qquad \qquad jinak$$

Platí: $X_j \sim Bi(n, \vartheta_j)$

Multinomické rozložení – příklady využití

- Předvolební průzkum:
  - n – počet tázaných
  - ϑ_j – skutečný podíl voličů j-té strany v populaci
  - X_j – počet (četnost) voličů j-té strany ve výběru

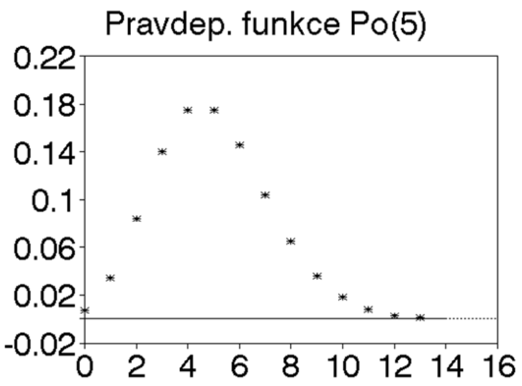
- Hody hrací kostkou:
  - n – počet hodů
  - $\vartheta_1,\dots,\vartheta_6$ – pravděpodobnost jednotlivých stran kostky
  - $X_1,\dots X_6$ – absolutní četnosti jednotlivých stran kostky

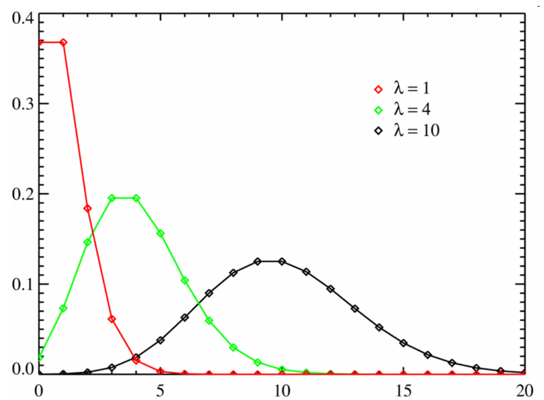
- Krevní skupiny:
  - n=4 (skupiny 0,A,B,AB)
  - $\vartheta_0, \vartheta_A, \vartheta_B, \vartheta_{AB}$ – pravděpodobnosti skupin 0, A, B, AB
  - X_0, X_A, X_B, X_{AB} – počty osob se skupinami 0, A, B, AB

Poissonovo rozložení

Poissonovo rozložení: Náhodná veličina X udává počet událostí, které nastanou v jednotkovém časovém intervalu (resp. jednotkové oblasti), přičemž události nastávají náhodně, jednotlivě a vzájemně nezávisle. Parametr $\lambda > 0$ je střední počet těchto událostí. Píšeme $X \sim Po(\lambda)$.
(Poissonovým rozložením se řídí např. počet výzev, které dojdou na telefonní ústřednu během určitého časového intervalu nebo počet mikroorganizmů v zorném poli mikroskopu. Jde o tzv. řídce se vyskytující jevy.)

$$\pi(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & \text{pro } x = 0, 1, \ldots \\ 0 \text{ jinak} \end{cases}$$





Příklad

Vztah mezi pravděpodobnostní funkcí binomického a Poissonova rozložení:
Nechť náhodná veličina X ~ Po(λ) a náhodná veličina Y ~ Bi(n, ϑ_n). Nechť $\vartheta_n \to 0$ pro $n \to \infty$ a přitom $n\vartheta_n \to \lambda$. Pak pravděpodobnostní funkce náhodné veličiny Y konverguje k pravděpodobnostní funkci náhodné veličiny X, tj.

$$\lim_{n\to\infty}\binom{n}{y}\vartheta_n^{\,y}(1-\vartheta_n)^{n-y} = \frac{\lambda^y}{y!}e^{-\lambda}$$

(Aproximace binomického rozložení pomocí Poissonova rozložení je vyhovující, když $n > 30$ a $\vartheta < 0{,}1$.)

Příklad na Poissonovo rozložení: Dělnice v přádelně obsluhuje 800 vřeten. Pravděpodobnost toho, že se příze přetrhne během časového intervalu délky t, je pro všechna vřetena stejná a je rovna 0,005. Určete pravděpodobnost, že během intervalu délky t dojde k nejvýše 10 přetržením.
Řešení: Y – počet přetržení v časovém intervalu délky t, Y ~ Bi(800;0,005).

Přesný výpočet: $P(Y \le 10) = \sum_{y=0}^{10}\binom{800}{y}0{,}005^y(1-0{,}005)^{800-y} = 0{,}997239$

Aproximativní výpočet: podmínky dobré aproximace jsou splněny, parametr

$\lambda = n\vartheta = 800.0{,}005 = 4$, $P(Y \le 10) = \sum_{y=0}^{10}\frac{4^y}{y!}e^{-4} = 0{,}9971602$

Příklad

1) Průměrný telefonní hovor trvá 1,5 min. Má-li ústředna 10 linek a dochází-li průměrně k 120 hovorům za hodinu, jaká je pravděpodobnost ztráty volání?

Řešení: X udává počet volajících, X~Po(2*1,5). Ke ztrátě volání dojde, pokud chce současně volat více než 10 volajících (tj. není volná linka). Tedy

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{x=0}^{10} \frac{3^x}{x!} e^{-3} \cong 0{,}001.$$

2) Průměrný telefonní hovor trvá 1,5 min. Kolik linek musí ústředna mít, dochází-li průměrně k 240 hovorům za hodinu a pravděpodobnost ztráty volání nemá překročit a) 0,01, b) 0,001?

Řešení: X udává počet volajících, X~Po(240/60*1,5). Hledáme n tak aby $P(X > n) \leq 0{,}01$

tj. $P(X \leq n) \geq 0{,}99 \Rightarrow \sum_{x=0}^{n} \frac{6^x}{x!} e^{-6} \geq 0{,}99 \Rightarrow n = 12.$

Pro případ b) chceme $\sum_{x=0}^{n} \frac{6^x}{x!} e^{-6} \geq 0{,}999 \Rightarrow n = 15.$

Negativní binomické (Pascalovo) rozložení

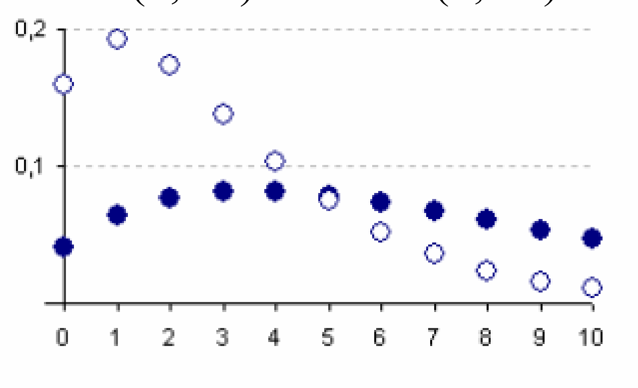
Negativní binomické rozložení (Pascalovo):

Náhodná veličina X udává počet neúspěchů před n-tým úspěchem v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ. Píšeme X ~ NB(n, ϑ).

$$\pi(x) = \binom{n+x-1}{x} \vartheta^n (1-\vartheta)^x, \quad x = 0,1,\ldots, \quad 0 < \vartheta < 1$$

$$= 0 \qquad\qquad jinak$$

$NB(2,0.2) \quad a \quad NB(2,0.4)$

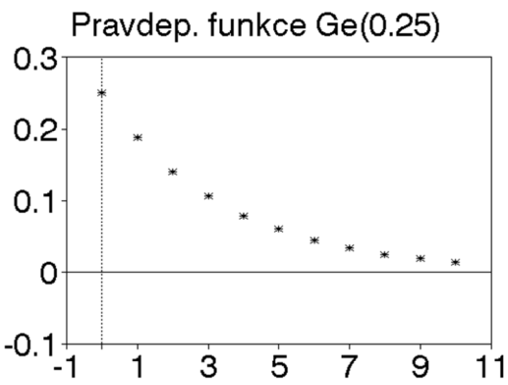


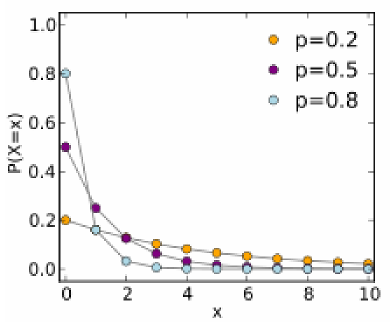
- Negativně binomické rozdělení lze definovat obecněji. Tak jak je zde uvedeno jde o rozdělení Pascalovo.

Geometrické rozložení

Geometrické rozložení: Náhodná veličina X udává počet neúspěchů v posloupnosti opakovaných nezávislých pokusů předcházejících prvnímu úspěchu, přičemž pravděpodobnost úspěchu je v každém pokusu rovna ϑ. Píšeme X ~ Ge(ϑ)

$$\pi(x) = \begin{cases} (1-\vartheta)^x \vartheta \text{ pro } x = 0, 1, \dots \\ 0 \text{ jinak} \end{cases}$$





Příklad

Dva hráči střídavě házejí kostkou. Vyhrává ten, kdo první hodí šestku. Jaká je pravděpodobnost, že vyhraje ten, který začínal?

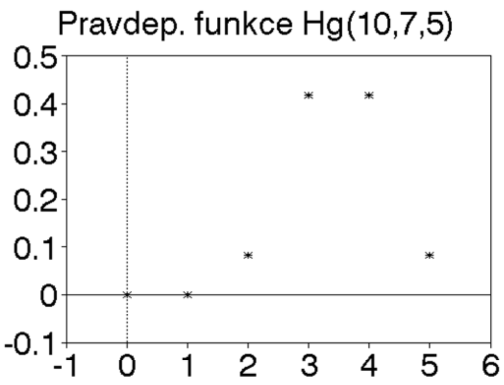
Řešení: X udává počet nehození šestky (neúspěch) před prvním hozením šestky (úspěch), X ~ Ge(1/6). Hledáme tedy pravděpodobnost jevu
A: 1. úspěch po sudém počtu neúspěchů.

$$P(A)=\sum_{k=0}^{\infty}\frac{1}{6}\left(\frac{5}{6}\right)^{2k}=\frac{1}{6}\frac{1}{1-\frac{25}{36}}=\frac{6}{11}=0{,}545$$

Hypergeometrické rozložení

Hypergeometrické rozložení: V souboru N prvků je M prvků označeno. Náhodně vybereme n prvků bez vracení. Náhodná veličina X udává počet vybraných označených prvků. Píšeme X ~ Hg(N, M, n)

$$\pi(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{pro } x = \max\{0, M - N + n\}, \dots, \min\{M, n\} \\ 0 \text{ jinak} \end{cases}$$



Příklad

V klobouku jsou 3 černé a 4 bílé koule. Určete pravděpodobnost, že při vytažení 3 koulí budou aspoň 2 černé.

Řešení: X udává počet vytažených černých koulí, X ~ HG(7,3,3). Hledaná pravděpodobnost je

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - \frac{\binom{3}{0}\binom{4}{3} + \binom{3}{1}\binom{4}{2}}{\binom{7}{3}} =$$

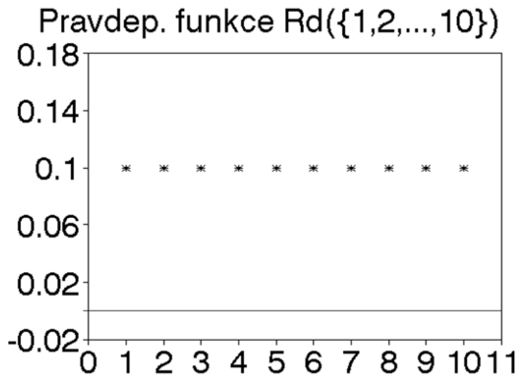
$$1 - \frac{1 \cdot 4 + 3 \cdot 6}{35} = 1 - \frac{22}{35} = \frac{13}{35} = 0{,}371.$$

Rovnoměrné diskrétní rozložení

Rovnoměrné diskrétní rozložení: Nechť G je konečná množina o n prvcích. Náhodná veličina X nabývá se stejnou pravděpodobností každé hodnoty z množiny G. Píšeme X ~ Rd(G)

$$\pi(x) = \begin{cases} \frac{1}{n} \text{ pro } x \in G \\ 0 \text{ jinak} \end{cases}$$

(Typickým příkladem je náhodná veličina udávající počet ok při hodu kostkou.)



Vybraná rozložení spojitých náhodných veličin

Důležitá spojitá rozdělení:

- Rovnoměrné rozdělení
- Normální rozdělení (označované také jako Gaussovo rozdělení)
- Logaritmicko-normální rozdělení (také log-normální rozdělení)
- Studentovo rozdělení
- Fischerovo-Snedecorovo rozdělení
- χ^2 rozdělení (Chí-kvadrát)
- Cauchyho rozdělení
- Exponenciální rozdělení
- Laplaceovo rozdělení (nebo také dvojitě exponenciální rozdělení)
- Weibullovo rozdělení

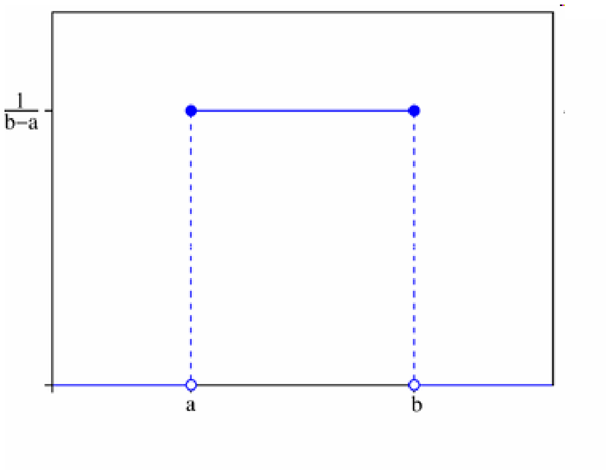
Rovnoměrné spojité rozložení

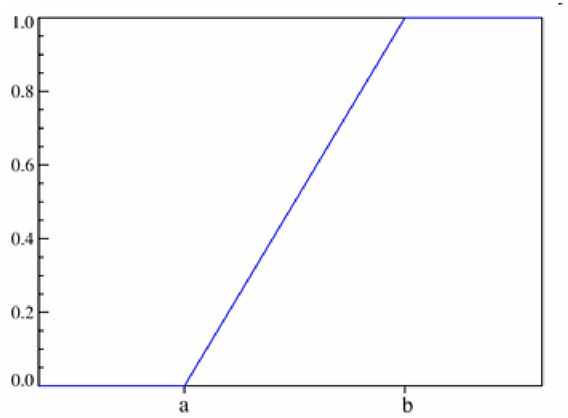
Rovnoměrné spojité rozložení: Předpokládejme, že veličina X
- může nabýt jakékoliv hodnoty mezi čísly a, b
- pravděpodobnost, že nabude hodnoty z jakéhokoliv intervalu v tomto rozmezí je stejná jako pravděpodobnost, že nabude hodnoty z jakéhokoliv jiného intervalu stejné délky.

Jsou-li tyto podmínky splněny, pak X má rovnoměrné spojité rozložení na intervalu (a, b). Hustota pravděpodobnosti náhodné veličiny X je konstantní na intervalu (a, b) a plocha pod křivkou hustoty tvoří obdélník.
Píšeme X ~ Rs(a, b).

$$\varphi(x) = \begin{cases} \dfrac{1}{b-a} & \text{pro } x \in (a,b) \\ 0 & \text{jinak} \end{cases}$$

$$\Phi(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & x \in (a,b) \\ 1 & x \geq b \end{cases}$$

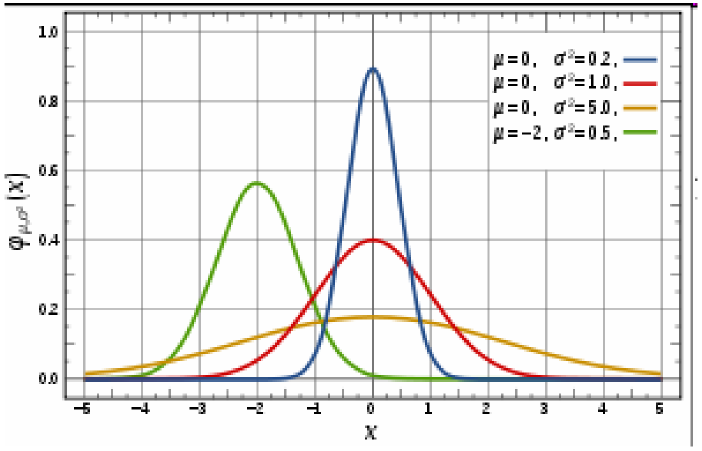


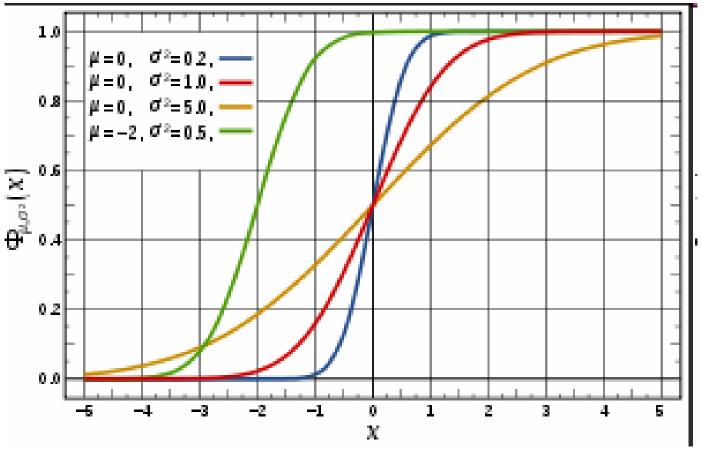


Normální rozložení

Normální rozložení: Tato náhodná veličina vzniká např. tak, že ke konstantě μ se přičítá velké množství nezávislých náhodných vlivů mírně kolísajících kolem nuly. Proměnlivost těchto vlivů je vyjádřena konstantou $\sigma > 0$.

Píšeme $X \sim N(\mu, \sigma^2)$, hustota $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Grafem této hustoty je tzv. Gaussova křivka.

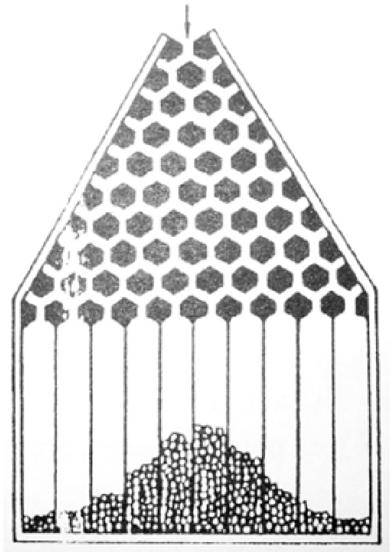




Galtonova deska

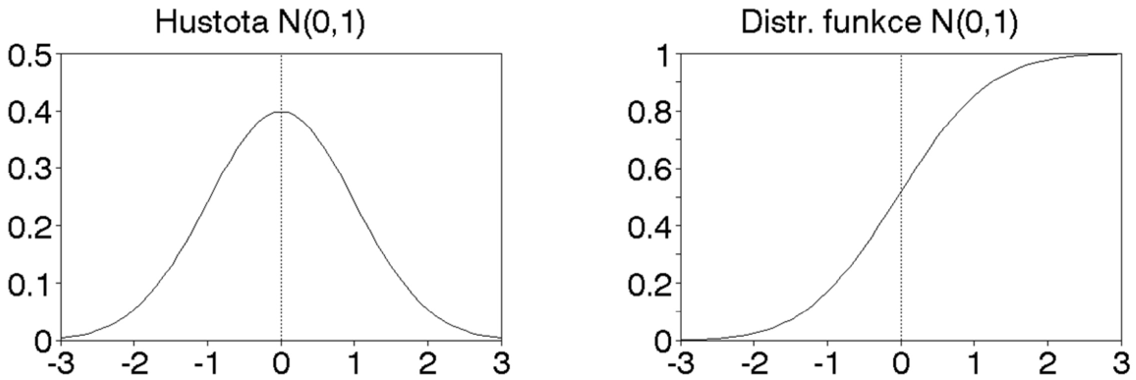
Ilustrace vzniku normálního rozložení pomocí Galtonovy desky:
Deska obsahuje n řad pravidelně uspořádaných klínů, a to tak, že v k-té řadě je právě k klínů. Do otvoru nahoře padají kuličky, které jsou v každé řadě se stejnou pravděpodobností 1/2 vychylovány vlevo nebo vpravo. Pod poslední radou je n - 1 přihrádek, ve kterých se kuličky shromaždují. Nasypeme-li do tohoto systému velké množství kuliček, vytvoří v přihrádkách jakýsi ”kopec”, jehož tvar je velmi podobný tvaru grafu hustoty náhodné veličiny s normálním rozložením.
Náhodné vychylování kuliček jednotlivými řadami překážek je možno chápat jako speciální případ velkého množství chybových faktorů, náhodně působících na nějaký proces, jako působení mnoha blíže nespecifikovatelných vlivů, které ovlivňují zcela náhodně rozložení jeho výsledku.

Obrázek



Standardizované normální rozložení

Standardizované normální rozložení:
Pro $\mu = 0$, $\sigma^2 = 1$ se jedná o standardizované normální rozložení, píšeme $U \sim N(0, 1)$. Hustota pravděpodobnosti má v tomto případě tvar $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$.



$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ je tabelována pro $u \geq 0$, pro $u < 0$ se používá přepočtový vzorec $\Phi(-u) = 1 - \Phi(u)$.

Příklad

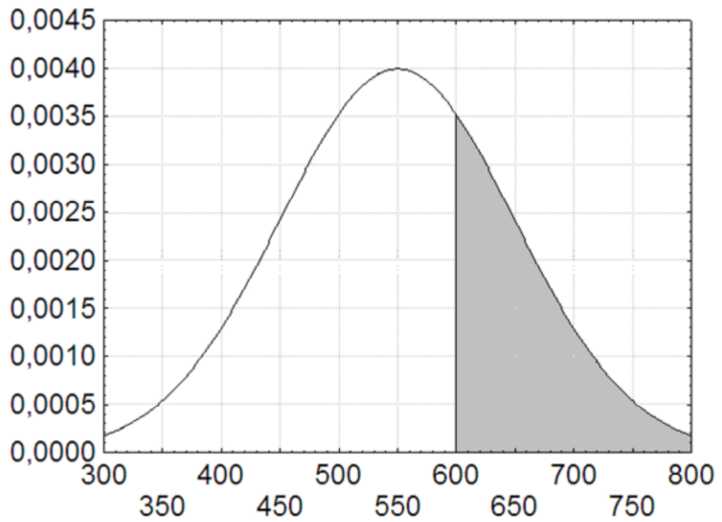
Příklad na normální rozložení: Výsledky u přijímacích zkoušek na jistou VŠ jsou normálně rozloženy s parametry $\mu = 550$ bodů, $\sigma = 100$ bodů. S jakou pravděpodobností bude mít náhodně vybraný uchazeč aspoň 600 bodů?

Řešení:

X – výsledek náhodně vybraného uchazeče, $X \sim N(550, 100^2)$,

$P(X \geq 600) = 1 - P(X \leq 600) + P(X = 600) = 1 - P(X \leq 600) =$

$= 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{600-\mu}{\sigma}\right) = 1 - P\left(U \leq \frac{600-550}{100}\right) = 1 - \Phi(0{,}5) = 1 - 0{,}69146 = 0{,}30854.$



Normální rozložení - vlastnosti

Některé vlastnosti normálního rozložení:

Jestliže $X \sim N(\mu, \sigma^2)$, pak $U = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

Jestliže $X \sim N(\mu, \sigma^2)$, a $Y = a + bX$, pak $Y \sim N(a + b\mu, b^2\sigma^2)$.

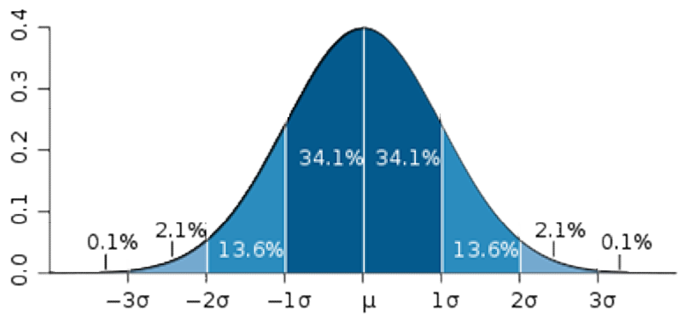
Jestliže $X_1, \ldots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, $Y = \sum_{i=1}^{n} X_i$, pak $Y \sim$ $N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$.

Význam normálního rozložení:

Normální rozložení hraje ústřední roli v počtu pravděpodobnosti i matematické statistice. Jeho význam spočívá jednak v tom, že normálním rozložením se řídí pravděpodobnostní chování mnoha náhodných veličin a jednak v tom, že za určitých podmínek konverguje k normálnímu rozložení součet nezávislých náhodných veličin s týmž rozložením (viz centrální limitní věta).

„koncentrace hodnot“ normální NV:

Přes 68% hodnot „leží“ v intervalu (μ-σ, μ+σ).
Přes 95% hodnot „leží“ v intervalu (μ-2σ, μ+2σ).
Přes 99% hodnot „leží“ v intervalu (μ-3σ, μ+3σ).



Dvojrozměrné normální rozložení

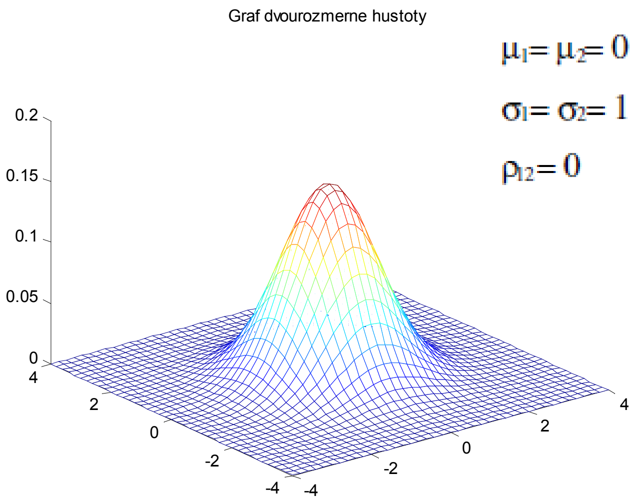
Definice:
O spojitém náhodném vektoru $\mathbf{X} = \binom{X_1}{X_2}$ říkáme, že má dvojrozměrné normální rozložení s parametry $\mu = \binom{\mu_1}{\mu_2}$ a $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, když jeho hustota je dána vzorcem

$$\varphi(\mathbf{x}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\cdot\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}, \quad \mathbf{x} \in \mathbf{R}^2.$$

Zkráceně píšeme $\mathbf{X} = \binom{X_1}{X_2} \sim N_2(\mu, \mathbf{\Sigma})$.
Pro $\mu = \binom{0}{0}$ a $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ mluvíme o standardizovaném dvojrozměrném normálním rozložení.

Poznámka:
Význam parametrů je následující:
$\mu_1 = E(X_1),\ \mu_2 = E(X_2),\ \sigma_1^2 = D(X_1),\ \sigma_2^2 = D(X_2),\ \rho = R(X_1, X_2)$

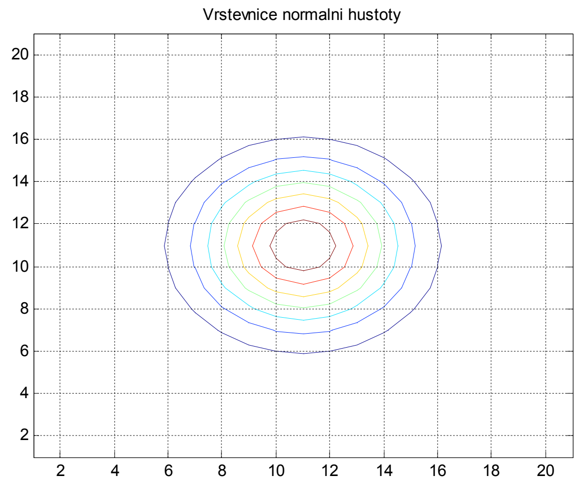
Dvojrozměrné normální rozložení

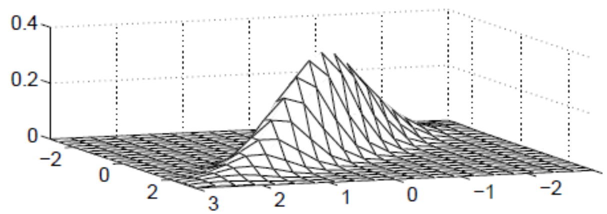


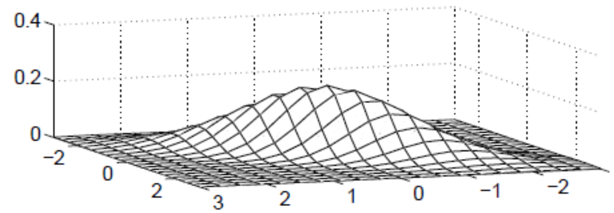
$\mu_1 = \mu_2 = 0$

$\sigma_1 = \sigma_2 = 1$

$\rho_{12} = 0$



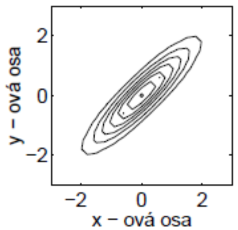




$\mu_1 = \mu_2 = 0$

$\sigma_1 = \sigma_2 = 1$

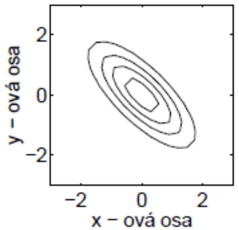
$\rho_{12} = 0.9$



$\mu_1 = \mu_2 = 0$

$\sigma_1 = \sigma_2 = 1$

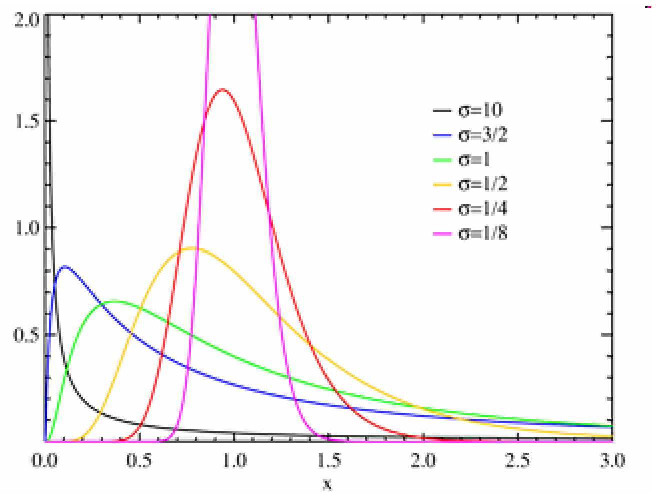
$\rho_{12} = -0.75$

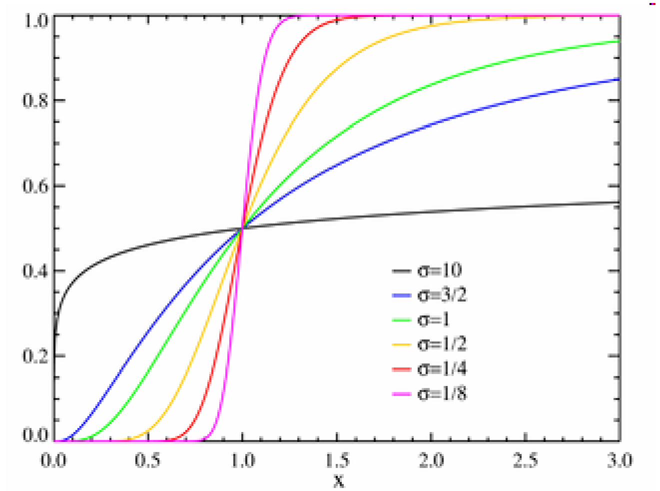


Logaritmicko normální rozložení

Logaritmicko normální rozložení: Náhodná veličina X ~ LN(μ, σ) vzniká v situacích, kdy kladná konstanta logaritmu μ je násobena velkým množstvím nezávislých náhodných veličin, kolísajících mírně kolem jedníčky. Variabilita jejich logaritmů je charakterizována parametrem σ. Logaritmicko normální rozdělení má hustotu

$$\varphi(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \quad x > 0$$

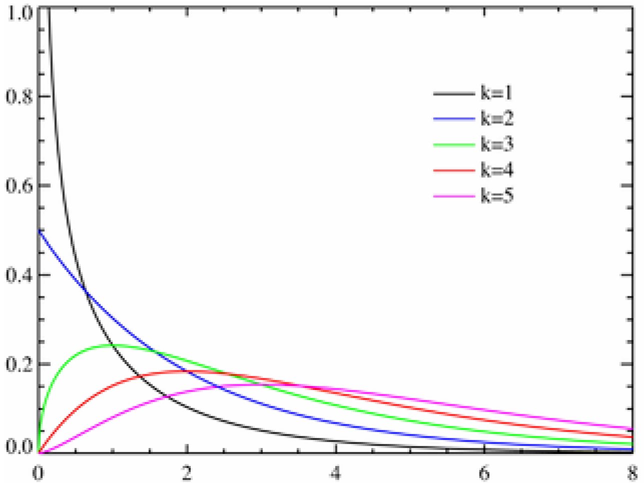


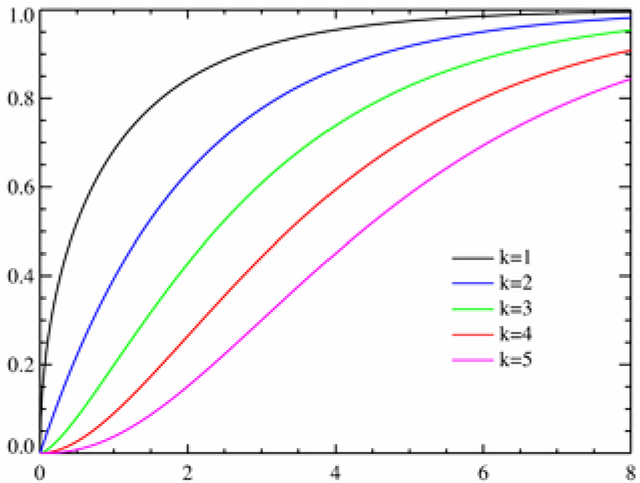


Pearsonovo χ^2 rozložení

Pearsonovo rozložení chí-kvadrát s n stupni volnosti: Nechť $X_1, \ldots, X_k$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1)$, $i = 1, \ldots, k$.
Pak náhodná veličina $X = X_1^2 + \ldots + X_k^2 \sim \chi^2(k)$.

$$\varphi(x,k) = \begin{cases} \dfrac{1}{2^{k/2} \cdot \Gamma(k/2)} \cdot x^{k/2-1} \cdot e^{-x/2} & x > 0 \\ 0 & jinak \end{cases} \qquad \Gamma(s) = \int_0^\infty e^{-t} \cdot t^{s-1}\, dt$$



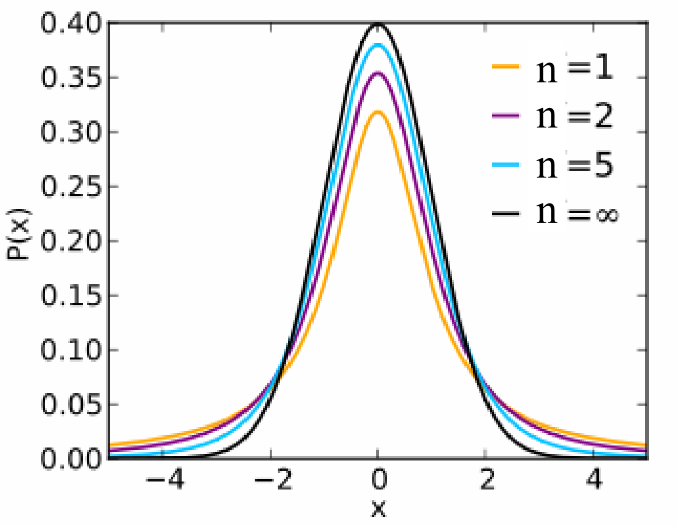


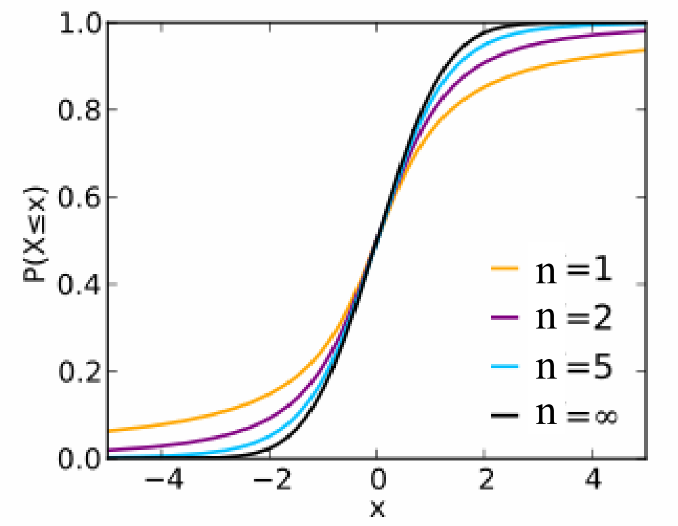
Studentovo rozložení

Studentovo rozložení s n stupni volnosti: Nechť X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_1 \sim N(0, 1)$, $X_2 \sim \chi^2(n)$.

Pak náhodná veličina $X = \dfrac{X_1}{\sqrt{\dfrac{X_2}{n}}} \sim t(n)$.

$$\varphi(x,n) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma(n/2)} \cdot \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad x \in (-\infty, \infty)$$



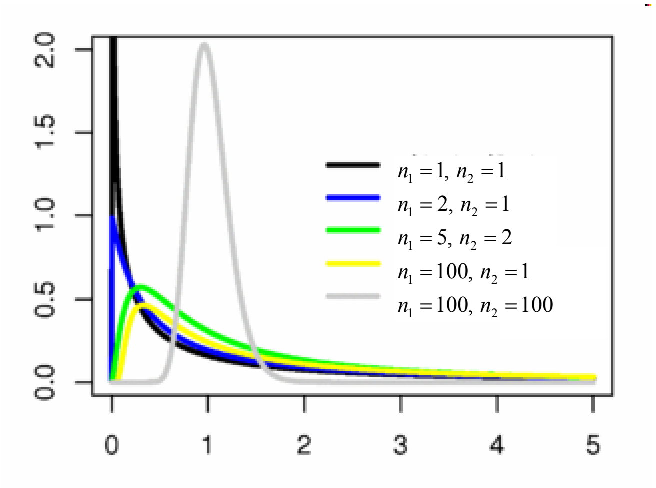


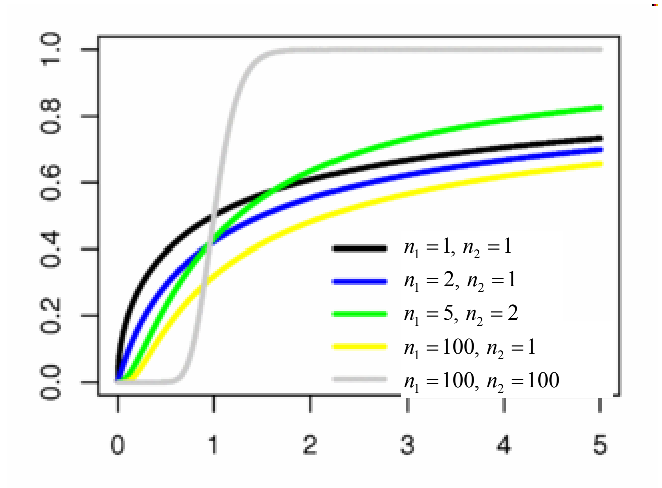
Fisher-Snedecorovo rozložení

Fisherovo-Snedecorovo rozložení s n_1 a n_2 stupni volnosti: Nechť X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \chi^2(n_i)$, i = 1, 2.

Pak náhodná veličina $X = \dfrac{X_1 / n_1}{X_2 / n_2} \sim F(n_1, n_2)$.

$$\varphi(x, n_1, n_2) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right) n_1^{\,n_1/2} n_2^{\,n_2/2}}{\Gamma(n_1/2)\Gamma(n_2/2)} \cdot \left(\frac{x^{(n_1-2)/2}}{(n_2 + n_1 x)^{(n_1+n_2)/2}}\right) \quad \text{pro } x > 0$$





Cauchyho rozložení

Cauchyho rozložení pravděpodobnosti s parametry x_0 a γ, pro $-\infty < x_0 < \infty$ a $\gamma > 0$, je definováno hustotou pravděpodobnosti ve tvaru

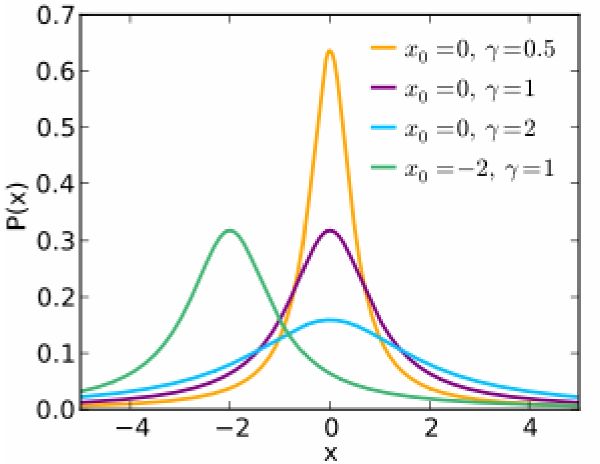
$$\varphi(x;x_0,\gamma) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

$$= \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right]$$

kde x_0 je parametr, určující umístění největší hodnoty rozdělení.
Zvláštní případ, kdy $x_0 = 0$ a $\gamma = 1$ se nazývá **standardní Cauchyho rozdělení** s hustotou pravděpodobnosti vyjádřenou vztahem

$$\varphi(x;0,1) = \frac{1}{\pi(1+x^2)}.$$

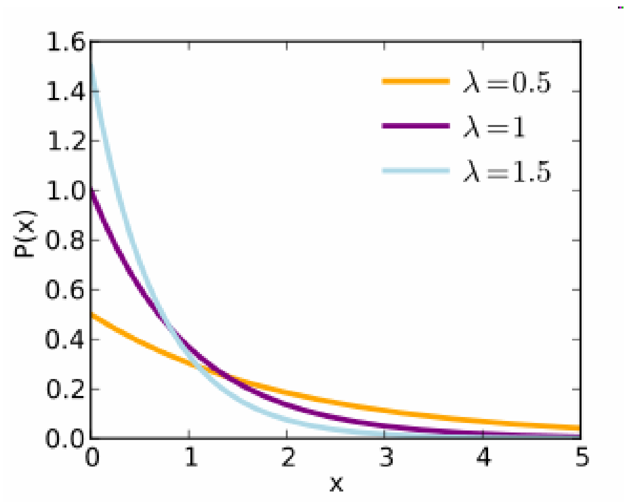
Standardní Cauchyho rozdělení je speciální případ Studentova rozdělení (pro n = 1).

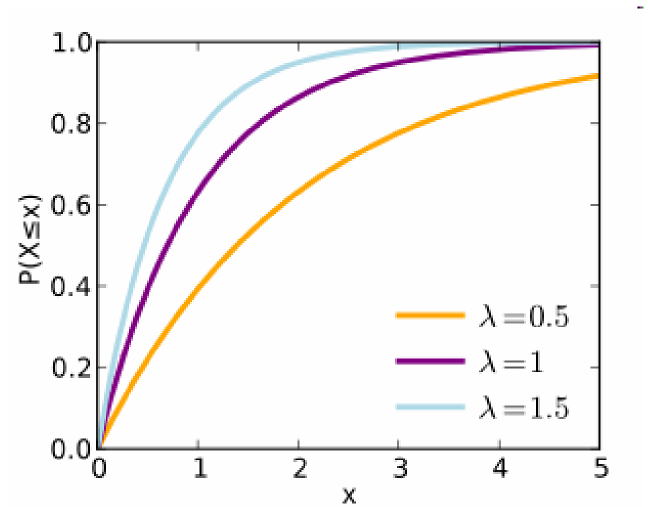


Exponenciální rozložení

Exponenciální rozložení: Náhodná veličina X udává dobu čekání na příchod nějaké události, která se může dostavit každým okamžikem se stejnou šancí bez ohledu na dosud pročekanou dobu. (Jde o tzv. čekání bez paměti.) Přitom $\frac{1}{\lambda}$ vyjadřuje střední dobu čekání. Píšeme $X \sim Ex(\lambda)$

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{jinak} \end{cases}, \quad \Phi(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{jinak} \end{cases}$$





Příklad

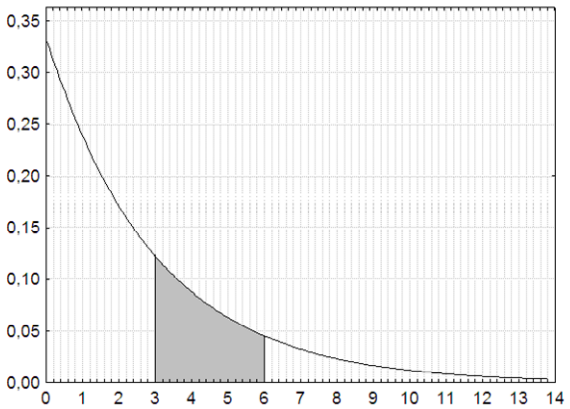
Příklad na exponenciální rozložení: Doba (v minutách) potřebná k obsloužení zákazníka v prodejně potravin je náhodná veličina, která se řídí rozložením $Ex\left(\frac{1}{3}\right)$. Jaká je pravděpodobnost, že doba potřebná k obsloužení náhodně vybraného zákazníka v této prodejně bude v rozmezí od 3 do 6 minut?

Řešení:

X – doba potřebná k obsloužení náhodně vybraného zákazníka, $X \sim Ex\left(\frac{1}{3}\right)$,

$$P(3 \leq X \leq 6) = \int_3^6 \frac{1}{3} e^{-\frac{x}{3}} dx = \frac{1}{3}(-3)\left[e^{-\frac{x}{3}}\right]_3^6 = -e^{-2} + e^{-1} = 0{,}233.$$

S pravděpodobností 0,233 bude zákazník obsloužen v době od 3 do 6 minut.

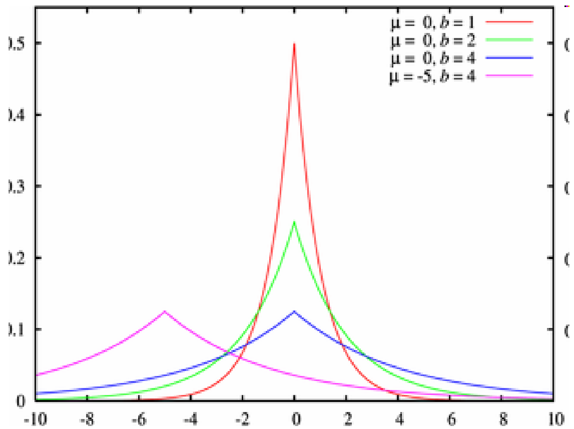


Laplaceovo rozložení

Laplaceovo rozložení: Náhodná veličina, která vznikne rozdílem dvou NV z exponenciálního rozložení, se řídí tímto rozložením. Využití ve fyzice, ekonomii – Brownův pohyb. Hustota je dána vzorcem

$$\varphi(x;\mu,b)=\frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

$$=\frac{1}{2b}\begin{cases}\exp\left(-\frac{\mu-x}{b}\right) & x<\mu\\ \exp\left(-\frac{x-\mu}{b}\right) & x\geq\mu\end{cases}$$



Platí např.:

$$X\sim Laplace(0,b)\Rightarrow|X|\sim Ex\left(\frac{1}{b}\right)$$

$$X_1\sim Ex(\lambda_1),\,X_2\sim Ex(\lambda_2)\Rightarrow\lambda_1X_1-\lambda_2X_2\sim Laplace(0,1)$$

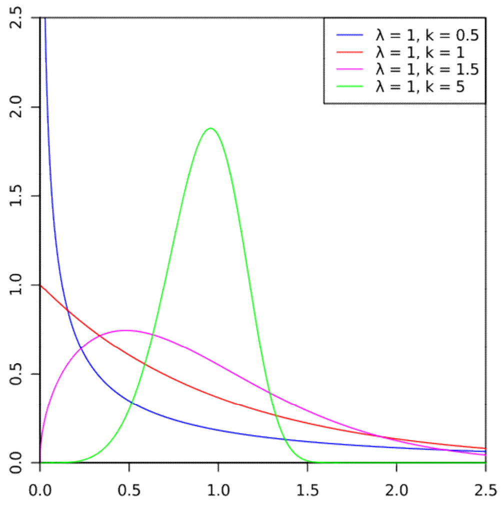
Weibullovo rozložení

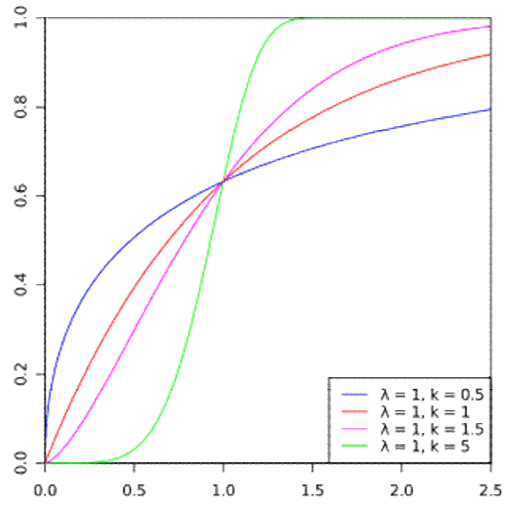
Weibullovo rozdělení: Náhodná veličina X ~ Wb(δ, ε) vyjadřuje dobu čekání na nějakou událost, která se každým okamžikem může dostavit se šancí úměrnou mocninné funkci pročekané doby. Přitom čísla δ > 0 a ε > 0 se nazývají parametry měřítka a formy.

$$\varphi(x;\delta,\varepsilon)=\begin{cases}\varepsilon\cdot\delta\cdot \mathrm{x}^{\varepsilon\text{-}1}e^{-\delta\cdot x^{\varepsilon}} & \text{pro } x>0\\ 0 & \text{pro } x\le 0\end{cases}$$

Jiná forma zápisu:

$$\varphi(x;\lambda,k)=\begin{cases}\dfrac{k}{\lambda}\left(\dfrac{\mathrm{x}}{\lambda}\right)^{k-1}e^{-\left(\frac{x}{\lambda}\right)^{k}} & \text{pro } x>0\\ 0 & \text{pro } x\le 0\end{cases}$$





10. Rozložení transformovaných NV. Číselné charakteristiky NV.

Motivace: Máme náhodnou veličinu X s distribuční funkcí $\Phi(x)$ (resp. pravděpodobnostní funkcí $\pi(x)$ v diskrétním případě resp. hustotou $\varphi(x)$ ve spojitém případě) a borelovskou funkci g: R → R. Zavedeme transformovanou náhodnou veličinu Y = g(X) a hledáme její distribuční funkci $\Phi_*(y)$ (resp. pravděpodobnostní funkcí $\pi_*(y)$ v diskrétním případě resp. hustotu $\varphi_*(y)$ ve spojitém případě).

Věta: Nechť X je diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$ a g je borelovská ryze monotónní funkce, tedy v oblasti $C \subseteq R$ existuje inverzní funkce $g^{-1} = \tau$. Pak pravděpodobnostní funkce $\pi_*(y)$ transformované náhodné veličiny Y = g(X) má tvar:

$$\boxed{\pi_*(y) = \begin{cases} \pi(\tau(y)) \text{pro} y \in C \\ 0 \text{jinak} \end{cases}.}$$

Důkaz: $\pi_*(y) = P(Y = y) = P(g(X) = y) = P\left(X = g^{-1}(y)\right) = P(X = \tau(y)) = \pi(\tau(y))$ pro $y \in C$, $\pi_*(y) = 0$ jinak.

Příklad: $X \sim \pi(x)$, $Y = a + bX$, $\pi_*(y) = ?$

Řešení:

a) $b \neq 0$: $\pi_*(y) = P(Y = y) = P(a + bX = y) = P\left(X = \frac{y-a}{b}\right) = \pi\left(\frac{y-a}{b}\right)$

b) $b = 0$: $Y = a \Rightarrow Y \sim Dg(a)$

Rozložení transformované spojité náhodné veličiny

Věta: Nechť X je spojitá náhodná veličina s hustotou $\varphi(x)$ a g je borelovská ryze monotónní funkce se spojitou a nenulovou derivací v R, tedy v oblasti $C \subseteq R$ existuje inverzní funkce $g^{-1} = \tau$ se spojitou a nenulovou derivací. Pak hustota $\varphi_*(y)$ transformované náhodné veličiny $Y = g(X)$ má tvar:

$$\boxed{\varphi_*(y) = \begin{cases} \varphi(\tau(y))|\tau'(y)| \text{pro} \, y \in C \\ 0 \, \text{jinak} \end{cases}.}$$

Důkaz:

$$\Phi_*(y) = P(Y \le y) = P(g(X) \le y) = \begin{cases} P(X \le \tau(y)) = \Phi(\tau(y)) \, \text{pro g rostoucí} \\ P(X \ge \tau(y)) = 1 - \Phi(\tau(y)) \, \text{pro g klesající} \end{cases}$$

$$\varphi_*(y) = \frac{d\Phi_*(y)}{dy} = \left.\begin{cases} \varphi(\tau(y))\tau'(y) \, \text{pro g rostoucí} \\ -\varphi(\tau(y))\tau'(y) \, \text{pro g klesající} \end{cases}\right\} = \varphi(\tau(y))|\tau'(y)| \text{ pro } y \in C, \; \varphi_*(y) = 0 \text{ jinak}$$

Příklad: $X \sim Rs\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $Y = \mathrm{tg}\, X$, $\varphi_*(y) = ?$

Řešení:

$$\varphi(x) = \begin{cases} \frac{1}{\pi} \, \text{pro} \, x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \\ 0 \, \text{jinak} \end{cases}$$

$$\Phi_*(y) = P(Y \le y) = P(\mathrm{tg}(X) \le y) = P(X \le \mathrm{arctg}(y)) = \Phi(\mathrm{arctg}(y))$$

$$\varphi_*(y) = \frac{d\Phi_*(y)}{dy} = \varphi(\mathrm{arctg}(y))\frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}$$

Říkáme, že Y má Cauchyovo rozložení, píšeme $Y \sim t(1)$.

Nemonotónní transformace

Věta: Není-li transformační funkce g ryze monotónní, pak mezi X a Y neexistuje vzájemně jednoznačný vztah. Distribuční funkce transformované náhodné veličiny Y se vypočte podle vzorce: $\Phi_*(y) = P(X \in \Delta_1) + P(X \in \Delta_2) + \ldots$, kde $\Delta_1, \Delta_2, \ldots$ jsou ty intervaly, pro které $Y \leq y$.

Příklad: $X \sim N(0,1)$, $Y = X^2$, $\varphi_*(y) = ?$

Řešení:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\Phi_*(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \Phi(\sqrt{y}) - [1 - \Phi(\sqrt{y})] =$$

$$= 2\Phi(\sqrt{y}) - 1$$

$$\varphi_*(y) = \frac{d\Phi_*(y)}{dy} = 2\varphi(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \text{ pro } y > 0, \quad \varphi_*(y) = 0 \text{ jinak}$$

Y má χ^2 rozložení s jedním stupněm volnosti, píšeme $Y \sim \chi^2(1)$.

$X \sim \chi^2(k)$:

$$\varphi(x,k) = \begin{cases} \dfrac{1}{2^{k/2} \cdot \Gamma(k/2)} \cdot x^{k/2-1} \cdot e^{-x/2} & x > 0 \\ 0 & jinak \end{cases} \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Rozložení transformovaného náhodného vektoru

Věta (transformace náhodného vektoru **X** = ($X_1, \ldots, X_n$) na skalární náhodnou veličinu $Y = g(X_1, \ldots, X_n)$)

a) Diskrétní případ: **X** = ($X_1, \ldots, X_n$) ~ $\pi(x_1, \ldots, x_n)$, g: $R^n \to R$ je borelovská funkce $\Rightarrow$

$$\boxed{\begin{array}{l} Y = g(x_1, \ldots, x_n) \sim \pi_*(y_1, \ldots, y_n) = \sum\limits_{(x_1,\ldots x_n)\in S(y)} \ldots \sum \pi(x_1,\ldots,x_n), \text{ kde} \\ S(y) = \{(x_1,\ldots,x_n) \in R^n; g(x_1,\ldots,x_n) = y\} \end{array}}$$

b) Spojitý případ: **X** = ($X_1, \ldots, X_n$) ~ $\varphi(x_1, \ldots, x_n)$, g: $R^n \to R$ je borelovská funkce $\Rightarrow$

$$\boxed{\begin{array}{l} Y = g(x_1, \ldots, x_n) \sim \varphi_*(y_1, \ldots, y_n) = \dfrac{d}{dy}\int_{S(y)} \ldots \int \varphi(x_1,\ldots,x_n)dx_1 \ldots dx_n \text{, kde} \\ S(y) = \{(x_1,\ldots,x_n) \in R^n; g(x_1,\ldots,x_n) \le y\} \end{array}}$$

Věta o konvoluci

Věta (věta o konvoluci)

a) Diskrétní případ: X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \pi_i(x_i)$, $i = 1, 2 \Rightarrow$

$$Y = X_1 + X_2 \sim \pi_*(y) = \sum_{x_1=-\infty}^{\infty} \pi_1(x_1)\pi_2(y - x_1) = \sum_{x_2=-\infty}^{\infty} \pi_1(y - x_2)\pi_2(x_2)$$

$\pi_*(y)$ se nazývá konvoluce funkcí $\pi_1(x_1)$, $\pi_2(x_2)$.

b) Spojitý případ: X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \varphi_i(x_i)$, $i = 1, 2 \Rightarrow$

$$Y = X_1 + X_2 \sim \varphi_*(y) = \int_{-\infty}^{\infty} \varphi_1(x_1)\varphi_2(y - x_1)\,dx_1 = \int_{-\infty}^{\infty} \varphi_1(y - x_2)\varphi_2(x_2)\,dx_2$$

$\varphi_*(y)$ se nazývá konvoluce funkcí $\varphi_1(x_1)$, $\varphi_2(x_2)$.

Příklad

Příklad: X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim Po(\lambda_i)$, i = 1, 2, $Y = X_1 + X_2$, $\pi_*(y) = ?$

Řešení:

$$\pi_i(x_i) = \begin{cases} \dfrac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i} \text{ pro } x_i = 0, 1, \dots \\ 0 \text{ jinak} \end{cases}$$

$$\pi_*(y) = \sum_{x_1=-\infty}^{\infty} \pi_1(x_1)\pi_2(y-x_1) = \left| x_1 \geq 0, y - x_1 \geq 0 \Rightarrow 0 \leq x_1 \leq y \right| = \sum_{x_1=0}^{y} \pi_1(x_1)\pi_2(y-x_1) =$$

$$= \sum_{x_1=0}^{y} \frac{\lambda_1^{x_1}}{x_1!} e^{-\lambda_1} \frac{\lambda_2^{y-x_1}}{(y-x_1)!} e^{-\lambda_2} = e^{-(\lambda_1+\lambda_2)} \frac{1}{y!} \sum_{x_1=0}^{y} \binom{y}{x_1} \lambda_1^{x_1} \lambda_2^{y-x_1} = \frac{(\lambda_1+\lambda_2)^y}{y!} e^{-(\lambda_1+\lambda_2)} \text{ pro } y = 0, 1, \dots$$

$\pi_*(y) = 0$ jinak.

Vidíme, že $Y \sim Po(\lambda_1 + \lambda_2)$.

Zobecnění: $X_1, \dots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim Po(\lambda_i)$, $i = 1, 2, \dots, n \Rightarrow$

$Y = X_1 + \dots + X_n \sim Po(\lambda_1 + \dots + \lambda_n)$.

Lineární transformace náhodného vektoru

Věta (lineární transformace n-rozměrného náhodného vektoru)
Nechť $\mathbf{X} = (X_1, \dots, X_n)'$ je náhodný vektor, $\mathbf{a} = (a_1, \dots, a_n)'$ je reálný vektor a $\mathbf{B} = (b_{ij})_{i,j=1,\dots,n}$ je reálná čtvercová pozitivně definitní matice (tj. $\forall \mathbf{x} \in \mathbf{R}^n$ je kvadratická funkce $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$). Pak pro rozložení pravděpodobností transformovaného náhodného vektoru $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$ platí:

a) Diskrétní případ: $\boxed{\pi_*(\mathbf{y}) = \pi(\mathbf{B}^{-1}(\mathbf{y} - \mathbf{a}))}$

b) Spojitý případ: $\boxed{\varphi_*(y) = \det(B)^{-1}\varphi\left(B^{-1}\left(y - a\right)\right)}$

Věta: Nechť náhodný vektor $\mathbf{X} = (X_1, \dots, X_n)'$ má n-rozměrné normální rozložení $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Položme $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$. Pak $\mathbf{Y} \sim N_n(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$.

Číselné charakteristiky náhodných veličin

Motivace: Doposud jsme pracovali s funkcionálními charakteristikami náhodných veličin (např. distribuční funkce, pravděpodobnostní funkce, hustota pravděpodobnosti), které plně popisují pravděpodobnostní chování náhodné veličiny. Číselné charakteristiky vystihují pouze některé rysy tohoto chování, např. popisují polohu realizací náhodné veličiny na číselné ose či jejich proměnlivost (variabilitu). Jsou jednodušší než číselné charakteristiky, ale nesou jen částečnou informaci. Podobně jako v popisné statistice volíme vhodnou číselnou charakteristiku podle toho, jakého typu je daná náhodná veličina - zda je ordinální nebo intervalová či poměrová. Číselné charakteristiky znaků mají své teoretické protějšky v číselných charakteristikách náhodných veličin.

Definice:

Nechť X je náhodná veličina aspoň ordinálního charakteru a $\alpha \in (0,1)$. Číslo $K_\alpha(X)$ se nazývá α-kvantil náhodné veličiny X, jestliže splňuje nerovnosti:

$$\boxed{P(X \le K_\alpha(X)) \ge \alpha \wedge P(X \ge K_\alpha(X)) \ge 1-\alpha}$$

Jiné možné označení kvantilu: x_α

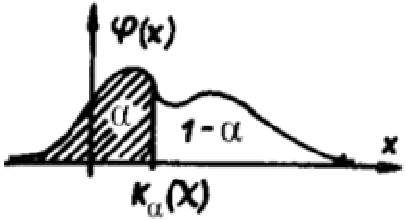
Kvantil $K_{0,50}(X)$ se nazývá medián,
$K_{0,25}(X)$ dolní kvartil,
$K_{0,75}(X)$ horní kvartil,
kvantily $K_{0,10}(X)$, ..., $K_{0,90}(X)$ jsou decily,
$K_{0,01}(X)$, ..., $K_{0,99}(X)$ jsou percentily.
Kterýkoliv α-kvantil je charakteristikou polohy číselných realizací náhodné veličiny na číselné ose.
Jako charakteristika variability slouží kvartilová odchylka $q = K_{0,75}(X) - K_{0,25}(X)$.

Kvantil spojité NV

Důsledek: (pro spojitou náhodnou veličinu)

Je-li X spojitá náhodná veličina, pak $K_\alpha(X)$ je takové číslo, pro které platí: $\alpha = \Phi(K_\alpha(X)) = \int_{-\infty}^{K_\alpha(X)} \varphi(x)dx$.

Ilustrace:



Příklad

Příklad: Nechť X ~ Ex(1). Určete medián a kvartilovou odchylku.

Řešení: $\varphi(x)=\begin{cases} e^{-x} \text{pro} x>0 \\ 0 \text{jinak} \end{cases}$, $\Phi(x)=\begin{cases} 1-e^{-x} \text{pro} x>0 \\ 0 \text{jinak} \end{cases}$

$$\alpha=\Phi(K_\alpha(X))=1-e^{-K_\alpha(X)} \Rightarrow K_\alpha(X)=-\ln(1-\alpha)$$

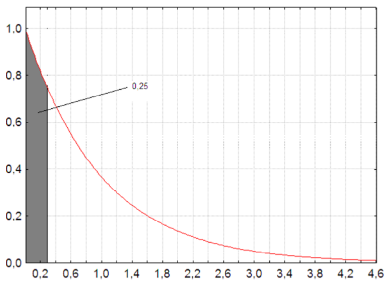
$$K_{0,50}(X)=-\ln(1-0{,}5)=-\ln\frac{1}{2}=\ln 2=0{,}693$$

$$K_{0,25}(X)=-\ln(1-0{,}25)=-\ln\frac{3}{4}=\ln 4-\ln 3=0{,}288$$

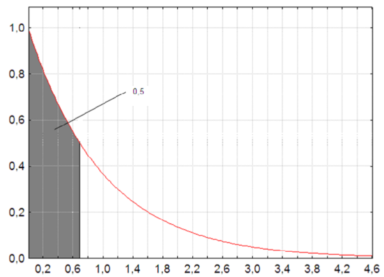
$$K_{0,75}(X)=-\ln(1-0{,}75)=-\ln\frac{1}{4}=\ln 4=1{,}386$$

$$q=K_{0,75}(X)-K_{0,25}(X)=1{,}386-0{,}288=1{,}098$$

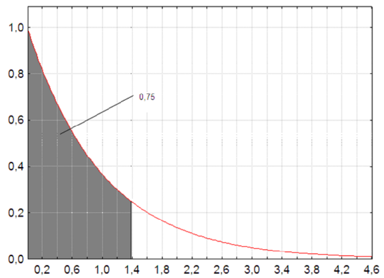
Dolní kvartil



Medián



Horní kvartil



Kvantily vybraných rozložení NV

Označení:

$X \sim N(0, 1) \Rightarrow K_\alpha(X) = u_\alpha$,
$X \sim \chi^2(n) \Rightarrow K_\alpha(X) = \chi^2_\alpha(n)$,
$X \sim t(n) \Rightarrow K_\alpha(X) = t_\alpha(n)$,
$X \sim F(n_1, n_2) \Rightarrow K_\alpha(X) = F_\alpha(n_1, n_2)$.
Tyto kvantily najdeme ve statistických tabulkách. Při jejich hledání používáme vztahy:

$$\boxed{\begin{aligned} &u_\alpha = - u_{1-\alpha}, \\ &t_\alpha(n) = - t_{1-\alpha}(n), \\ &F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}. \end{aligned}}$$

Kvantily lze také vypočítat pomocí statistického software.

Příklad:

a) Nechť $U \sim N(0, 1)$. Najděte medián a horní a dolní kvartil.
b) Určete $\chi^2_{0,025}(25)$.
c) Určete $t_{0,99}(30)$ a $t_{0,05}(14)$.
d) Určete $F_{0,975}(5, 20)$ a $F_{0,05}(2, 10)$.

Řešení:

ad a) $u_{0,50} = 0$, $u_{0,25} = -0,67449$, $u_{0,75} = 0,67449$
ad b) $\chi^2_{0,025}(25) = 13,12$
ad c) $t_{0,99}(30) = 2,4573$, $t_{0,05}(24) = -1,7613$
ad d) $F_{0,975}(5, 20) = 3,2891$, $F_{0,05}(2, 10) = 0,05156$

Kvantily transformované NV

Věta:

Nechť X je spojitá náhodná veličina s distribuční funkcí $\Phi(x)$, $\alpha \in (0,1)$ a $g: R \to R$ ryze monotónní borelovská funkce. Pak pro α-kvantil transformované náhodné veličiny Y = g(X) platí:

a) Je-li g všude rostoucí funkce, pak $\boxed{K_\alpha(Y) = g(K_\alpha(X)).}$

b) Je-li g všude klesající funkce, pak $\boxed{K_\alpha(Y) = g(K_{1-\alpha}(X)).}$

Důkaz:

ad a) $\alpha = \Phi(K_\alpha(X)) = P(X \le K_\alpha(X)) = P(g(X) \le g(K_\alpha(X))) = P(Y \le g(K_\alpha(X))) = \Phi_*(g(K_\alpha(X))) \Rightarrow g(K_\alpha(X)) = K_\alpha(Y)$

ad b) $1-\alpha = \Phi(K_{1-\alpha}(X)) = P(X \le K_{1-\alpha}(X)) = P(g(X) \ge g(K_{1-\alpha}(X))) = 1 - P(Y \le g(K_{1-\alpha}(X))) = 1 - \Phi_*(g(K_{1-\alpha}(X))) \Rightarrow g(K_{1-\alpha}(X)) = K_\alpha(Y)$

Příklad:

Nechť U ~ N(0, 1). Najděte 9. decil transformované náhodné veličiny Y = 3 + 2U.

Řešení:

Funkce y = 3 + 2u je všude rostoucí funkce, tedy $K_{0,90}(Y) = 3 + 2\,u_{0,90} = 3 + 2 \times 1{,}28155 = 5{,}5631$.

Střední hodnota NV

Definice:

Nechť (Ω, A, P) je pravděpodobnostní prostor, X náhodná veličina aspoň intervalového typu definovaná na měřitelném prostoru (Ω, A).

a) Je-li X diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$, pak její střední hodnota (vzhledem k P) je číslo $\boxed{E(X)=\sum_{x=-\infty}^{\infty} x\pi(x)}$, pokud suma vpravo je konečná nebo absolutně konverguje. Jinak řekneme, že střední hodnota neexistuje.

b) Je-li X spojitá náhodná veličina s hustotou pravděpodobnosti $\varphi(x)$, pak její střední hodnota (vzhledem k P) je číslo $\boxed{E(X)=\int_{-\infty}^{\infty} x\varphi(x)dx}$, pokud integrál vpravo je konečný nebo absolutně konverguje. Jinak řekneme, že střední hodnota neexistuje.

(Střední hodnota je číslo, které charakterizuje polohu realizací náhodné veličiny na číselné ose spřihlédnutím k jejich pravděpodobnostem. V diskrétním případě představuje střední hodnota těžiště soustavy hmotných bodů, jejichž hmotnost je popsána pravděpodobnostní funkcí $\pi(x)$ a ve spojitém případě je střední hodnota těžištěm hmotné přímky, na níž je rozprostření hmoty popsáno hustotou pravděpodobnosti $\varphi(x)$. Střední hodnota je teoretickým protějškem váženého aritmetického průměru.)

Příklad

Příklad a):

Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte její střední hodnotu.

Řešení:

$$\pi(x) = \begin{cases} \frac{1}{6} \text{ pro } x = 1, \ldots, 6 \\ 0 \text{ jinak} \end{cases}, \quad E(X) = \sum_{x=1}^{6} x\pi(x) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3{,}5 .$$

Příklad b):

Rozložení náhodné veličiny X je dáno hustotou φ(x) = 2x+2 na (−1, 0) a nulovou jinde. Vypočtěte její střední hodnotu.

Řešení

$$E(X) = \int_R x\varphi(x)\,dx = \int_{-1}^{0} x(2x+2)\,dx = \left[2\frac{x^3}{3} + x^2\right]_{-1}^{0} = \frac{2}{3} - 1 = -\frac{1}{3}$$

Střední hodnota transformované NV

Věta:

a) Diskrétní případ: Nechť X je diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$ (resp. $(X_1, \ldots, X_n)$ je diskrétní náhodný vektor s pravděpodobnostní funkcí $\pi(x_1, \ldots, x_n)$). Nechť $g : R \mapsto R$ je borelovská funkce, $Y = g(X)$ je transformovaná náhodná veličina (resp. $g : R^n \mapsto R$ je borelovská funkce, $Y = g(X_1, \ldots, X_n)$ je transformovaná náhodná veličina). Pak $\boxed{E(Y) = \sum_{x=-\infty}^{\infty} g(x)\pi(x)}$, pokud součet vpravo je konečný nebo absolutně konvergentní (resp. $E(Y) = \sum_{x_1=-\infty}^{\infty} \ldots \sum_{x_n=-\infty}^{\infty} g(x_1, \ldots, x_n)\pi(x_1, \ldots, x_n)$, pokud součet vpravo je konečný nebo absolutně konvergentní).

a) Spojitý případ: Nechť X je spojitá náhodná veličina s hustotou $\varphi(x)$ (resp. $(X_1, \ldots, X_n)$ je spojitý náhodný vektor s hustotou $\varphi(x_1, \ldots, x_n)$). Nechť $g : R \mapsto R$ je borelovská funkce, $Y = g(X)$ je transformovaná náhodná veličina (resp. $g : R^n \mapsto R$ je borelovská funkce, $Y = g(X_1, \ldots, X_n)$ je transformovaná náhodná veličina). Pak $\boxed{E(Y) = \int_{-\infty}^{\infty} g(x)\varphi(x)dx,}$ pokud integrál vpravo je konečný nebo absolutně konvergentní (resp.
$E(Y) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n)\varphi(x_1, \ldots, x_n)dx_1 \ldots dx_n$, pokud integrál vpravo je konečný nebo absolutně konvergentní).

Příklad

Příklad:

Nechť X ~ Ex(λ), $Y = e^{-\gamma X}$, kde $\gamma > 0$ je konstanta. Vypočtěte E(Y).

Řešení:

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{jinak} \end{cases}, \quad E(Y) = \int_0^\infty e^{-\gamma x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda + \gamma}.$$

Rozptyl NV

Definice:

Nechť (Ω, A , P) je pravděpodobnostní prostor, X náhodná veličina aspoň intervalového typu definovaná na měřitelném prostoru (Ω, A), která má střední hodnotu E(X). Rozptylem náhodné veličiny X rozumíme číslo $\boxed{D(X) = E([X-E(X)]^2),}$ pokud střední hodnota vpravo existuje. Číslo $\sqrt{D(X)}$ se nazývá směrodatná odchylka.

(Rozptyl je číslo, které charakterizuje proměnlivost realizací náhodné veličiny kolem její střední hodnoty s přihlédnutím k jejich pravděpodobnostem. Je teoretickým protějškem váženého rozptylu. Je vhodnější počítat rozptyl podle vzorce $D(X) = E(X^2) - [E(X)]^2$, jak bude ukázáno později.)

Důsledek:

V diskrétním případě je rozptyl dán vzorcem $\boxed{D(X) = \sum_{x=-\infty}^{\infty} [x - E(X)]^2 \pi(x) = \sum_{x=-\infty}^{\infty} x^2 \pi(x) - [E(X)]^2}$

a ve spojitém případě vzorcem: $\boxed{D(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 \varphi(x)\, dx = \int_{-\infty}^{\infty} x^2 \varphi(x)\, dx - [E(X)]^2}$

(pokud suma či integrál vpravo absolutně konvergují).

Centrovaná a standardizovaná NV

Definice:

Transformovaná náhodná veličina X– E(X) se nazývá centrovaná náhodná veličina.

Transformovaná náhodná veličina $\frac{X - E(X)}{\sqrt{D(X)}}$ se nazývá standardizovaná náhodná veličina.

Příklad: Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte její rozptyl.

Řešení: $\pi(x) = \begin{cases} \frac{1}{6} \text{ pro } x = 1, \ldots, 6 \\ 0 \text{ jinak} \end{cases}$, E(X) = 3,5 (viz př. 12.10.), $D(X) = \sum_{x=1}^{6} x^2 \frac{1}{6} - 3{,}5^2 = \ldots = \frac{35}{12} = 2{,}92$.

Kovariance a korelace NV

Definice: Kovariancí náhodných veličin X_1, X_2, které mají střední hodnoty $E(X_1)$, $E(X_2)$, rozumíme číslo $\boxed{C(X_1, X_2) = E([X_1 - E(X_1)]\,[X_2 - E(X_2)])}$ (pokud střední hodnoty vpravo existují).

Kovariance je číslo, které charakterizuje proměnlivost realizací náhodných veličin X_1, X_2 kolem jejich středních hodnot s přihlédnutím k jejich pravděpodobnostem. Je-li kovariance kladná (záporná), pak to svědčí o existenci jistého stupně přímé (nepřímé) lineární závislosti mezi realizacemi náhodných veličin X_1, X_2. Je-li kovariance nulová, pak říkáme, že náhodné veličiny X_1, X_2 jsou nekorelované a znamená to, že mezi jejich realizacemi není žádný lineární vztah. Pozor – z nekorelovanosti nevyplývá stochastická nezávislost, zatímco ze stochastické nezávislosti plyne nekorelovanost. Kovariance je teoretickým protějškem vážené kovariance. Je vhodnější počítat kovarianci podle vzorce

$$\boxed{C(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)}.$$

Koeficientem korelace náhodných veličin X_1, X_2 rozumíme číslo

$$R(X_1, X_2) = \begin{cases} E\left(\dfrac{X_1 - E(X_1)}{\sqrt{D(X_1)}} \cdot \dfrac{X_2 - E(X_2)}{\sqrt{D(X_2)}}\right) & \text{pro } \sqrt{D(X_1)}\sqrt{D(X_2)} > 0 \\ 0 & \text{jinak} \end{cases}$$

, pokud střední hodnoty vpravo existují.

Koeficient korelace je číslo, které charakterizuje těsnost lineární závislosti realizací náhodných veličin X_1, X_2. Čím bližší je 1, tím těsnější je přímá lineární závislost, čím bližší je -1, tím těsnější je nepřímá lineární závislost. Je vhodnější počítat koeficient korelace podle vzorce $\boxed{R(X_1, X_2) = \dfrac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}}$

Kovariance NV

Důsledek:

V diskrétním případě je kovariance dána vzorcem

$$C(X_1, X_2) = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} [x_1 - E(X_1)] \cdot [x_2 - E(X_2)]\, \pi(x_1, x_2) = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 x_2 \pi(x_1, x_2) - E(X_1)\, E(X_2)$$

a ve spojitém případě vzorcem

$$C(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - E(X_1)] \cdot [x_2 - E(X_1)]\, \varphi(x_1, x_2)\, dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \varphi(x_1, x_2)\, dx_1 dx_2 - E(X_1)\, E(X_2)$$

Příklad

Příklad:

Náhodná veličina X udává příjem manžela (v tisících dolarů) a náhodná veličina Y příjem manželky (v tisících dolarů. Je známa simultánní pravděpodobnostní funkce $\pi(x,y)$ diskrétního náhodného vektoru (X,Y): $\pi(10,10) = 0,2$, $\pi(10,20) = 0,04$, $\pi(10,30) = 0,01$, $\pi(10,40) = 0$, $\pi(20,10) = 0,1$, $\pi(20,20) = 0,36$, $\pi(20,30) = 0,09$, $\pi(20,40) = 0$, $\pi(30,10) = 0$, $\pi(30,20) = 0,05$, $\pi(30,30) = 0,1$, $\pi(30,40) = 0$, $\pi(40,10) = 0$, $\pi(40,20) = 0$, $\pi(40,30) = 0$, $\pi(40,40) = 0,05$, $\pi(x,y) = 0$ jinak. Vypočtěte koeficient korelace příjmů manžela a manželky.

Řešení: Náhodná veličina X i náhodná veličina Y nabývají hodnot 10, 20, 30, 40. Sestavíme kontingenční tabulku:

X	Y				
	10	20	30	40	$\pi_1(x)$
10	0,20	0,04	0,01	0,00	0,25
20	0,10	0,36	0,09	0,00	0,55
30	0,00	0,05	0,10	0,00	0,15
40	0,00	0,00	0,00	0,05	0,05
$\pi_2(y)$	0,30	0,45	0,20	0,05	1,00

Spočteme

E(X) = 10.0,25+20.0,55+30.0,15+40.0,05 = 20, E(Y) = 10.0,30+20.0,45+30.0,20+40.0,05 = 20,

D(X) = 10^2.0,25+20^2.0,55+30^2.0,15+40^2.0,05 – 20^2 = 60, D(Y) = 10^2.0,30+20^2.0,45+30^2.0,20+40^2.0,05 – 20^2 = 70,

C(X,Y) = 10.10.0,20 + 10.20.0,04 + … 40.40.0,05 – 20.20 = 49,

R(X,Y) = $49/\sqrt{60}\sqrt{70}$ = 0,76.

Střední hodnota a rozptyl vybraných typů rozložení NV

Poznámka: Uvedeme střední hodnoty a rozptyly vybraných typů diskrétních a spojitých rozložení:

a) $X \sim Dg(\mu) \Rightarrow E(X) = \mu,\ D(X) = 0$

b) $X \sim A(\vartheta) \Rightarrow E(X) = \vartheta,\ D(X) = \vartheta\ (1-\vartheta)$

c) $X \sim Bi(n,\ \vartheta) \Rightarrow E(X) = n\vartheta,\ D(X) = n\vartheta\ (1-\vartheta)$

d) $X \sim Ge(\vartheta) \Rightarrow E(X) = \frac{1-\vartheta}{\vartheta},\ D(X) = \frac{1-\vartheta}{\vartheta^2}$

e) $X \sim Hg(N,M,n) \Rightarrow E(X) = \frac{M}{N}n,\ D(X) = \frac{Mn}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$

f) $X \sim Rd(G) \Rightarrow E(X) = \frac{n-1}{2},\ D(X) = \frac{n^2-1}{12}$

g) $X \sim Po(\lambda) \Rightarrow E(X) = \lambda,\ D(X) = \lambda$

h) $X \sim\ Rs(a, b) \Rightarrow E(X) = \frac{a+b}{2},\ D(X) = \frac{(b-a)^2}{12}$

i) $X \sim Ex(\lambda) \Rightarrow E(X) = \frac{1}{\lambda},\ D(X) = \frac{1}{\lambda^2}$

j) $X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu,\ D(X) = \sigma^2$

k) $X \sim \chi^2(n) \Rightarrow E(X) = n,\ D(X) = 2n$

l) $X \sim t(n) \Rightarrow E(X) = 0$ pro $n \geq 2$, pro $n = 1$ $E(X)$ neexistuje, $D(X) = \frac{n}{n-2}$ pro $n \geq 3$, pro $n = 1, 2$ $D(X)$ neexistuje

m) $X \sim F(n_1, n_2) \Rightarrow E(X) = \frac{n_2}{n_2-2}$ pro $n_2 \geq 3$, pro $n_2 = 1, 2$ $E(X)$ neexistuje, $D(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ pro $n_2 \geq 5$, pro $n_2 = 1,$ 2, 3, 4 $D(X)$ neexistuje.

Příklad

Příklad:
V sadě 15 výrobků je 5 zmetků. Náhodně vybereme 4 výrobky. Určete střední hodnotu a rozptyl náhodné vekičiny X, která udává počet zmetků, jestliže výběr provádíme
a) bez vracení,
b) s vracením.

Řešení:
ad a) $X \sim Hg(N, M, n)$, $N = 15$, $M = 5$, $n = 4$

$$E(X) = \frac{M}{N} n = \frac{5}{15} 4 = \frac{4}{3} = 1,\overline{3}, \quad D(X) = \frac{Mn}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1} = \frac{4}{3}\left(1 - \frac{5}{15}\right)\frac{11}{14} = \frac{44}{63} = 0,6984$$

ad b) $X \sim Bi(n, \vartheta)$, $n = 4$, $\vartheta = \frac{5}{15} = \frac{1}{3}$

$$E(X) = n\vartheta = 4\frac{1}{3} = 1,\overline{3}, \quad D(X) = n\vartheta(1-\vartheta) = \frac{4}{3}\left(1 - \frac{1}{3}\right) = \frac{8}{9} = 0,\overline{8}$$

Příklad

Najděte medián rozložení určeného hustotou $\varphi(x) = 1 - x/2$, $0 < x < 2$.

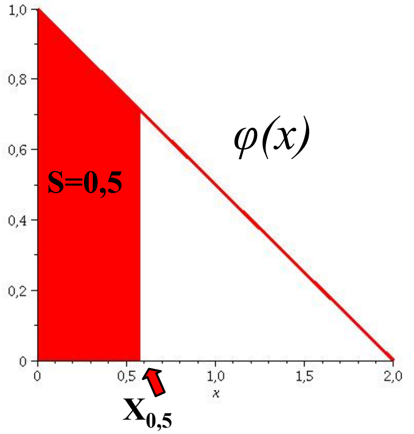
Řešení: Distribuční funkce: $F(x) = 0$ pro $x \leq 0$, $F(x) = 1$ pro $x \geq 2$ a

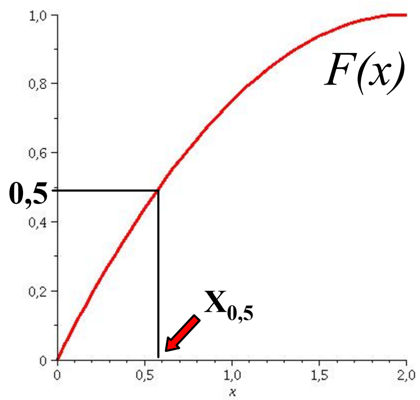
$$F(x) = \int_0^x 1 - \frac{t}{2}\,dt = x - \frac{x^2}{4} \qquad \text{pro } x \in (0,2)$$

Medián $x_{0,5}$ je řešením rovnice $F(x) = 0,5$, tedy

$$x^2 - 4x + 2 = 0 \quad \Rightarrow x_{1,2} = \frac{4 \pm \sqrt{16-8}}{2} = 2 \pm \sqrt{2} = \begin{cases} 3,4142 \\ 0,5857 \end{cases}$$

Protože $3,4142 > 2$, je hledaným řešením $\underline{\underline{x_{0,5} = 0,5857}}$.



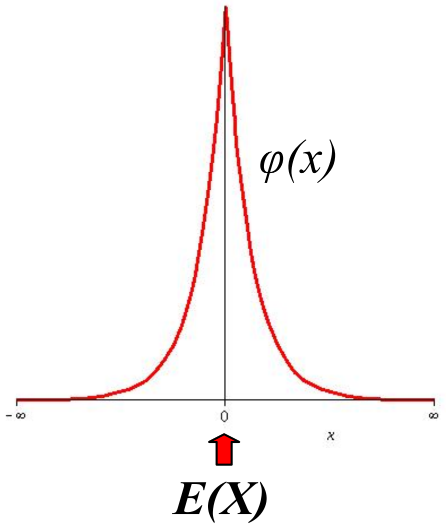


Příklad

Náhodná veličina má hustotu $\varphi(x) = a \cdot e^{-|x|}$ $x \in (-\infty, \infty)$. Určete a, střední hodnotu a rozptyl.

Řešení: $a = 1\Big/\left(2\int_0^\infty e^{-x}\,dx\right) = \dfrac{1}{2}$

$$E(X) = \int_{-\infty}^{\infty} x\varphi(x)\,dx = \int_{-\infty}^{\infty} x\frac{1}{2}e^{-|x|}\,dx = 0 \qquad D(X) = E(X^2) = \int_{-\infty}^{\infty} x^2\frac{1}{2}e^{-|x|}\,dx = 2$$

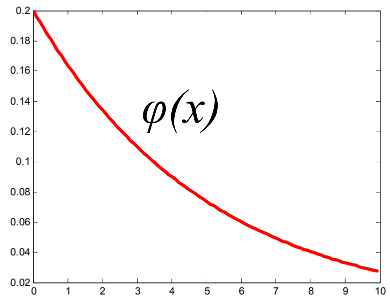


Příklad

Nechť životnost (v letech) výrobků se řídí exponenciálním rozložením s distribuční funkcí $F(x) = 1\text{-}e^{-x/5}$, $x > 0$. Tj. střední doba životnosti je 5 let. Tvar distribuční funkce znamená, že k poruše výrobku dojde s velkou pravděpodobností velmi brzy po jeho prodeji. Jakou záruční dobu stanoví výrobce, nemá-li počet reklamovaných výrobků překročit 10%?

Řešení: Náhodná veličina X udává životnost výrobku. Hledáme takové x, aby platilo

$P(X \leq x) = 0{,}1$ Tedy hledáme 10% kvantil.



$$0{,}1 = 1 - e^{-x/5}$$

$$x = -5\ln(0{,}9) = -5 \cdot (-0{,}10536) = 0{,}5268$$

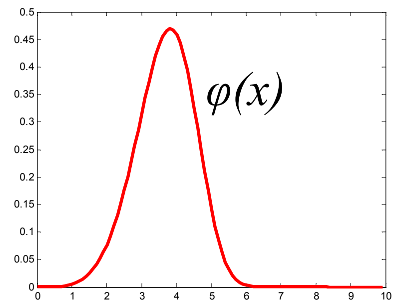
Pro splnění požadované podmínky je třeba stanovit záruční dobu na cca ½ roku.

Příklad

Nechť životnost (v letech) výrobků se řídí Weibullovým rozložením s distribuční funkcí $F(x) = 1\text{-}e^{-(x/4)^\wedge 5}$, $x > 0$. Tj. střední doba životnosti je cca 3.67 let. Tvar distribuční funkce znamená, že k poruše výrobku pravděpodobně nedojde hned po jeho prodeji, ale až po nějaké době. Jakou záruční dobu stanoví výrobce, nemá-li počet reklamovaných výrobků překročit 10%?

Řešení: Náhodná veličina X udává životnost výrobku. Hledáme takové x, aby platilo

$P(X \le x) = 0{,}1$ Tedy hledáme 10% kvantil.



$$0{,}1 = 1 - e^{-(x/4)^5}$$

$$x = 4 \cdot \sqrt[5]{-\ln(0{,}9)} = 2{,}55$$

Pro splnění požadované podmínky je třeba stanovit záruční dobu na cca 2,5 roku.

Momenty, šikmost a špičatost NV

Definice:

Nechť X, X_1, X_2 jsou náhodné veličiny, $k, k_1, k_2 \in R$, $r, s \in N$.

a) Číslo $E([X-k]^r)$ se nazývá r**-tý moment** náhodné veličiny X kolem konstanty k. Je-li $k = 0$, jde o r**-tý počáteční moment**, je-li $k = E(X)$, jedná se o r**-tý centrální moment**.

b) Číslo $E([X_1-k_1]^r[X_2-k_2]^s)$ se nazývá $r \times s$**-tý moment** náhodných veličin X_1, X_2 kolem konstant k_1, k_2. Je-li $k_1 = k_2 = 0$, jde o $r \times s$**-tý počáteční moment**, je-li $k_1 = E(X_1), k_2 = E(X_2)$, jedná se o $r \times s$**-tý centrální moment**.

Číslo $\boxed{A_3(X) = \frac{E\left([X-E(X)]^3\right)}{\left[\sqrt{D(X)}\right]^3}}$ se nazývá šikmost náhodné veličiny X.

Číslo $\boxed{A_4(X) = \frac{E\left([X-E(X)]^4\right)}{\left[\sqrt{D(X)}\right]^4} - 3}$ se nazývá špičatost náhodné veličiny X.

Je-li $A_3(X) = 0$, jde o symetrické rozložení. Je-li $A_3(X) > 0$, jde o kladně sešikmené rozložení a je-li $A_3(X) < 0$, jde o záporně sešikmené rozložení.

Je-li $A_4(X) = 0$, jde o rozložení s normální špičatostí. Je-li $A_4(X) > 0$, jde o špičaté rozložení a je-li $A_4(X) < 0$, jde o ploché rozložení.

Vektor středních hodnot, variační a korelační matice náhodného vektoru

Definice:

Nechť $\mathbf{X} = (X_1, \ldots, X_n)'$ je náhodný vektor. Reálný vektor $E(\mathbf{X}) = (E(X_1), \ldots, E(X_n))'$ se nazývá **vektor středních hodnot**. Reálná čtvercová symetrická matice

$$var(\mathbf{X}) = \begin{pmatrix} D(X_1) & C(X_1, X_2) & \ldots & C(X_1, X_n) \\ \ldots & \ldots & \ldots & \ldots \\ C(X_n, X_1) & C(X_n, X_2) & \ldots & D(X_n) \end{pmatrix}$$

se nazývá **varianční matice** náhodného vektoru $\mathbf{X}$ a reálná čtvercová symetrická matice

$$cor(\mathbf{X}) = \begin{pmatrix} 1 & R(X_1, X_2) & \ldots & R(X_1, X_n) \\ \ldots & \ldots & \ldots & \ldots \\ R(X_n, X_1) & R(X_n, X_2) & \ldots & 1 \end{pmatrix}$$

se nazývá **korelační matice** náhodného vektoru $\mathbf{X}$.

Příklad

Příklad:

Pro náhodný vektor (X, Y) z příkladu 12.19. najděte vektor středních hodnot, varianční a korelační matici.

Řešení:

Bylo spočteno, že

$E(X) = 20,\ \ E(Y) = 20,$

$D(X) = 60,\ \ D(Y) = 70,$

$C(X,Y) = 49,$

$R(X,Y) = 0{,}76.$

$$E(\mathbf{X})=\begin{pmatrix}20\\20\end{pmatrix},\ \mathrm{var}(\mathbf{X})=\begin{pmatrix}60 & 49\\49 & 70\end{pmatrix},\ \mathrm{cor}(\mathbf{X})=\begin{pmatrix}1 & 0{,}76\\0{,}76 & 1\end{pmatrix}$$

11. Vlastnosti číselných charakteristik NV

Věta: Nechť a, a_1, a_2, b, b_1, b_2 jsou reálná čísla, X, X_1, ..., X_n, Y_1, ..., Y_m jsou náhodné veličiny definované na témž pravděpodobnostním prostoru. V následujících vzorcích vždy z existence číselných charakteristik na pravé straně vyplývá existence výrazu na levé straně.

Vlastnosti střední hodnoty

a) $E(a) = a$

b) $E(a + bX) = a + bE(X)$

c) $E(X - E(X)) = 0$

d) $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$

e) Jsou-li náhodné veličiny X_1, ..., X_n stochasticky nezávislé, pak $E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$

Vlastnosti číselných charakteristik NV – kovariance

Vlastnosti kovariance

a) $C(a_1, X_2) = C(X_1, a_2) = C(a_1, a_2) = 0$

b) $C(a_1 + b_1X_1, a_2 + b_2X_2) = b_1b_2C(X_1, X_2)$

c) $C(X, X) = D(X)$

d) $C(X_1, X_2) = C(X_2, X_1)$

e) $C(X_1, X_2) = E(X_1X_2) – E(X_1)E(X_2)$

f) $C\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} C(X_i, Y_j)$

Vlastnosti číselných charakteristik NV – rozptyl

Vlastnosti rozptylu

a) $D(a) = 0$

b) $D(a + bX) = b^2 D(X)$

c) $D(X) = E(X^2) - [E(X)]^2$

d) $D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} C(X_i, X_j)$ (jsou-li náhodné veličiny X_1, ..., X_n nekorelované, pak $D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i)$)

Vlastnosti číselných charakteristik NV – korelace

Vlastnosti koeficientu korelace

a) $R(a_1, X_2) = R(X_1, a_2) = R(a_1, a_2) = 0$

b) $R(a_1 + b_1X_1, a_2 + b_2X_2) = \mathrm{sgn}(b_1b_2)\, R(X_1, X_2)$

c) $R(X, X) = 1$ pro $D(X) \neq 0$, $R(X, X) = 0$ jinak

d) $R(X_1, X_2) = R(X_2, X_1)$

e) $$R(X_1, X_2) = \begin{cases} \dfrac{C(X_1, X_2]}{\sqrt{D(X_1)}\sqrt{D(X_2)}} & \text{pro } \sqrt{D(X_1)}\sqrt{D(X_2)} > 0 \\ 0 \text{ jinak} & \end{cases}$$

Vlastnosti střední hodnoty - důkaz

Důkaz:

Pro vlastnosti střední hodnoty

ad a) X ~ Dg(a), $\pi(x)=\begin{cases}1 \text{ pro } x=a\\ 0 \text{ jinak}\end{cases}$, $E(X)=\sum_{x=-\infty}^{\infty} x\pi(x)=a\pi(a)=a\cdot 1=a$

ad b) Diskrétní případ: $E(a+bX)=\sum_{x=-\infty}^{\infty}(a+bx)\pi(x)=\sum_{x=-\infty}^{\infty}a\pi(x)+\sum_{x=-\infty}^{\infty}bx\pi(x)=a\sum_{x=-\infty}^{\infty}\pi(x)+b\sum_{x=-\infty}^{\infty}x\pi(x)=a+bE(X)$

Spojitý případ: $E(a+bX)=\int_{-\infty}^{\infty}(a+bx)\varphi(x)dx=a\int_{-\infty}^{\infty}\varphi(x)dx+b\int_{-\infty}^{\infty}x\varphi(x)dx=a+bE(X)$

ad c) Plyne z (b), kde a = -E(X), b = 1.

ad d) Spojitý případ:

$$E\left(\sum_{i=1}^{n}X_i\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_1+\ldots+x_n)\varphi(x_1,\ldots,x_n)dx_1\cdots dx_n=$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1\varphi(x_1,\ldots,x_n)dx_1\cdots dx_n+\cdots+\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_n\varphi(x_1,\ldots,x_n)dx_1\cdots dx_n=$$

$$=\int_{-\infty}^{\infty}x_1\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\varphi(x_1,\ldots,x_n)dx_2\cdots dx_n\right]dx_1+\cdots+\int_{-\infty}^{\infty}x_n\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\varphi(x_1,\ldots,x_n)dx_1\cdots dx_{n-1}\right]dx_n=$$

$$=\int_{-\infty}^{\infty}x_1\varphi_1(x_1)dx_1+\cdots+\int_{-\infty}^{\infty}x_n\varphi_n(x_n)dx_n=E(X_1)+\cdots+E(X_n)=\sum_{i=1}^{n}E(X_i)$$

Vlastnosti střední hodnoty - důkaz

ad d) Diskrétní případ: analogicky jako ve spojitém případě.

ad e) Spojitý případ:

$$E\left(\prod_{i=1}^{n} X_i\right) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_1\cdot\ldots\cdot x_n)\varphi(x_1,\ldots,x_n)dx_1\cdots dx_n =$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_1\cdot\ldots\cdot x_n)\varphi(x_1)\cdot\ldots\cdot\varphi(x_n)dx_1\cdots dx_n =$$

$$= \int_{-\infty}^{\infty} x_1\varphi(x_1)dx_1\cdot\ldots\cdot\int_{-\infty}^{\infty} x_n\varphi(x_n)dx_n =$$

$$= \prod_{i=1}^{n} E(X_i)$$

ad e) Diskrétní případ: analogicky jako ve spojitém případě.

Vlastnosti kovariance - důkaz

Pro vlastnosti kovariance:

ad a) $C(a_1, X_2) = E([a_1 - E(a_1)][X_2 - E(X_2)]) = E([a_1 - a_1][X_2 - E(X_2)]) = E(0) = 0$

ad b)

$$C(a_1 + b_1X_1, a_2 + b_2X_2) = E([a_1 + b_1X_1 - (a_1 + b_1E(X_1))][a_2 + b_2X_2 - (a_2 + b_2E(X_2))]) = b_1b_2E([X_1 - E(X_1)][X_2 - E(X_2)]) = b_1b_2C(X_1, X_2)$$

ad c) $C(X, X) = E([X - E(X)][X - E(X)]) = E([X - E(X)]^2) = D(X)$

ad d) $C(X_1, X_2) = E([X_1 - E(X_1)][X_2 - E(X_2)]) = E([X_2 - E(X_2)][X_1 - E(X_1)]) = C(X_2, X_1)$

ad e)

$$C(X_1, X_2) = E([X_1 - E(X_1)][X_2 - E(X_2)]) = E(X_1X_2 - X_2E(X_1) - X_1E(X_2) + E(X_1)E(X_2)) = E(X_1X_2) - E(X_2X_1) - E(X_1)E(X_2) + E(X_1)E(X_2) =$$
$$= E(X_1X_2) - E(X_1)E(X_2)$$

ad f)

$$C\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = E\left(\left[\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)\right]\left[\sum_{j=1}^{m} Y_j - E\left(\sum_{j=1}^{m} Y_j\right)\right]\right) = E\left(\sum_{i=1}^{n}[X_i - E(X_i)]\sum_{j=1}^{m}[Y_j - E(Y_j)]\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}[X_i - E(X_i)][Y_j - E(Y_j)] =$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} C(X_i, Y_j)$$

Vlastnosti rozptylu - důkaz

Pro vlastnosti rozptylu:

ad a) $D(a) = E\left([a - E(a)]^2\right) = E\left([a - a]^2\right) = E(0) = 0$

ad b) $D(a + bX) = E\left([a + bX - E(a + bX)]^2\right) = E\left([a + bX - a - bE(X)]^2\right) =$

$= E\left(b^2 [X - E(X)]^2\right) = b^2 D(X)$

ad c) $D(X) = E\left([X - E(X)]^2\right) = E\left(X^2 - 2XE(X) + [E(X)]^2\right) =$

$= E\left(X^2\right) - 2E(X)E(X) + [E(X)]^2 = E\left(X^2\right) - [E(X)]^2$

ad d) $D\left(\sum_{i=1}^{n} X_i\right) = C\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} C\left(X_i, X_j\right) =$

$= C\left(X_1, X_1\right) + C\left(X_1, X_2\right) + \ldots + C\left(X_1, X_n\right) + \ldots +$

$+ C\left(X_n, X_1\right) + C\left(X_n, X_2\right) + \ldots C\left(X_n, X_n\right) =$

$= \sum_{i=1}^{n} D\left(X_i\right) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} C\left(X_i, X_j\right)$

Vlastnosti korelace - důkaz

Pro vlastnosti koeficientu korelace:

ad a) Plyne přímo z definice, protože $D(a_1) = D(a_2) = 0$,

ad b) $$R(a_1 + b_1X_1, a_2 + b_2X_2) = E\left(\frac{a_1 + b_1X_1 - E(a_1 + b_1X_1)}{\sqrt{D(a_1 + b_1X_1)}} \cdot \frac{a_2 + b_2X_2 - E(a_2 + b_2X_2)}{\sqrt{D(a_2 + b_2X_2)}} \right) =$$

$$= E\left(\frac{a_1 + b_1X_1 - a_1 - b_1E(X_1)}{\sqrt{b_1^{\,2} D(X_1)}} \cdot \frac{a_2 + b_2X_1 - a_2 - b_2E(X_2)}{\sqrt{b_2^{\,2} D(X_2)}} \right) =$$

$$= \frac{b_1}{|b_1|} \cdot \frac{b_2}{|b_2|} R(X_1, X_2) = \operatorname{sgn}(b_1 \cdot b_2) R(X_1, X_2)$$

Vlastnosti korelace - důkaz

ad c) Pro D(X) = 0 plyne přímo z definice, jinak platí

$$R(X,X)=E\left(\frac{X-E(X)}{\sqrt{D(X)}}\cdot\frac{X-E(X)}{\sqrt{D(X)}}\right)=\frac{1}{D(X)}E\left(\left[X-E(X)\right]^2\right)=\frac{1}{D(X)}D(X)=1$$

ad d) Zřejmé.

ad e) $$R(X_1,X_2)=E\left(\frac{X_1-E(X_1)}{\sqrt{D(X_1)}}\cdot\frac{X_2-E(X_2)}{\sqrt{D(X_2)}}\right)=$$

$$=\frac{E\left(\left[X_1-E(X_1)\right]\cdot\left[X_2-E(X_2)\right]\right)}{\sqrt{D(X_1)}\cdot\sqrt{D(X_2)}}=\frac{C(X_1,X_2)}{\sqrt{D(X_1)}\cdot\sqrt{D(X_2)}}$$

Příklad

Příklad:

Vypočtěte střední hodnotu a rozptyl

a) centrované náhodné veličiny Y = X – E(X),

b) standardizované náhodné veličiny $U = \frac{X - E(X)}{\sqrt{D(X)}}$.

Řešení:

ad a) $E(Y) = E(X - \mu) = E(X) - E(\mu) = \mu - \mu = 0$, $D(Y) = D(X - \mu) = D(X) = \sigma^2$,

ad b) $E(U) = E(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma} E(X - \mu) = \frac{1}{\sigma} . 0 = 0$, $D(U) = D(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2} D(X - \mu) = \frac{1}{\sigma^2} . \sigma^2 = 1$.

Příklad:

Náhodné veličiny X, Z jsou náhodné chyby, které vznikají na vstupním zařízení. Mají střední hodnoty E(X) = -2, E(Y) = 4 a rozptyly D(X) = 4, D(Y) = 9. Koeficient korelace těchto chyb je R(X,Y) = -0,5. Chyba na výstupu zařízení souvisí s chybami na vstupu funkční závislostí $Z = 3X^2 - 2XY + Y^2 - 3$. Najděte střední hodnotu chyby na výstupu.

Řešení: $E(Z) = E(3X^2 - 2XY + Y^2 - 3) = 3E(X^2) - 2E(XY) + E(Y^2) - E(3) = 3\{D(X) + [E(X)]^2\} - 2[C(X,Y) + E(X)E(Y)] + D(Y) + [E(Y)]^2 - 3 = 3[D(X) + [E(X)]^2] - 2[R(X,Y)\sqrt{D(X)}\sqrt{D(Y)} + E(X)E(Y)] + D(Y) + [E(Y)]^2 - 3 = 3(4 + 4) -2[-0,5\times2\times3 + (-2) \times4] + 9 + 16 - 3 = 24 + 22 + 25 - 3 = 68$

Příklad

Náhodná veličina X udává počet ok při hodu kostkou. NV Y = 2+ 3X. Vypočtěte:

a) E(X) a D(X),
b) E(Y) a D(Y),
c) C(X,Y),
d) R(X,Y).

Řešení: a) $E(X)=\frac{1}{6}\sum_{x=1}^{6}x=3{,}5$ $\qquad D(X)=\frac{1}{6}\sum_{x=1}^{6}x^2-E(X)^2=\frac{91}{6}-3{,}5^2=2{,}9167$

b) $E(Y)=E(2+3X)=2+3E(X)=2+3\cdot 3{,}5=12{,}5$

$D(Y)=D(2+3X)=3^2D(X)=9\cdot 2{,}9167=26{,}25$

c) $C(X,Y)=C(X,2+3X)=3C(X,X)=3D(X)=3\cdot 2{,}9167=8{,}7501$

d) $R(X,Y)=R(X,2+3X)=\mathrm{sgn}(3)R(X,X)=1\cdot 1=1$

Příklad

Náhodná veličina X udává součet počtu ok při hodu 2 kostkami. Vypočtěte E(X).

Řešení: X_i … počet ok při i-tém hodu, i = 1,…,6

$$E(X_i) = 3{,}5$$

$$E(X) = E\left(\sum_{i=1}^{2} X_i\right) = \sum_{i=1}^{2} E(X_i) = \sum_{i=1}^{2} 3{,}5 = 7$$

Nebo:

součet	počet možností	možnosti
2	1	11
3	2	12 21
4	3	22 13 31
5	4	23 32 41 14
6	5	33 24 42 51 15
7	6	34 43 25 52 16 61
8	5	44 35 53 26 62
9	4	54 45 36 63
10	3	55 64 46
11	2	56 65
12	1	66
Celkem	**36**	

$$E(X) = \sum_{x=2}^{12} x\pi(x) = \frac{1}{36}(2\cdot 1 + 3\cdot 2 + 4\cdot 3 + 5\cdot 4 + 6\cdot 5$$

$$+ 7\cdot 6 + 8\cdot 5 + 9\cdot 4 + 10\cdot 3 + 11\cdot 2 + 12\cdot 1) = \frac{252}{36} = 7$$

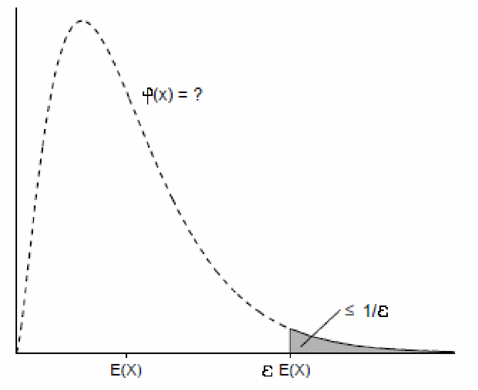
Markovova nerovnost

Věta (Markovova nerovnost):

Nechť pro náhodnou veličinu X se střední hodnotou E(X) platí P(X> 0) = 1. Pak platí Markovova nerovnost:

$\forall \varepsilon > 0: \boxed{P(X > \varepsilon E(X)) \leq \frac{1}{\varepsilon}}.$

Ilustrace pro spojitý případ:



Důkaz: Pro spojitý případ:

$$E(X) = \int_0^{\infty} x\varphi(x)dx \geq \int_{\varepsilon E(X)}^{\infty} x\varphi(x)dx \geq \int_{\varepsilon E(X)}^{\infty} \varepsilon\, E(X)\varphi(x)dx = \varepsilon\, E(X) \int_{\varepsilon E(X)}^{\infty} \varphi(x)dx = \varepsilon\, E(X) P(X > \varepsilon\, E(X))$$

$$\Rightarrow P(X > \varepsilon\, E(X)) \leq \frac{1}{\varepsilon}$$

Příklad

Příklad:

Nechť P(X > 0) = 1 a E(X) = δ, kde δ > 0 je konstanta.

a) Odhadněte $P(X > 3\delta)$.

b) Nechť X ~ $Ex\left(\frac{1}{\delta}\right)$. Vypočtěte $P(X > 3\delta)$.

Řešení:

ad a) $P(X > 3\delta) \le \frac{1}{3} = 0,\overline{3}$

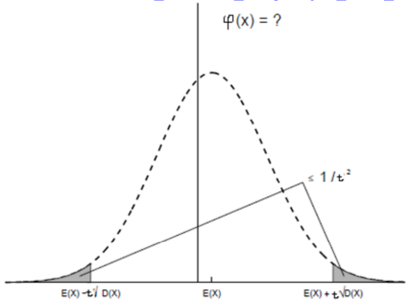
ad b) $X \sim Ex\left(\frac{1}{\delta}\right) \Rightarrow \varphi(x) = \begin{cases} \frac{1}{\delta} e^{-\frac{x}{\delta}} \text{pro } x > 0 \\ 0 \text{ jinak} \end{cases}, \quad E(X) = \delta, P(X > 3\delta) = \int_{3\delta}^{\infty} \frac{1}{\delta} e^{-\frac{x}{\delta}} dx = \left[-e^{-\frac{x}{\delta}}\right]_{3\delta}^{\infty} = e^{-3} = 0,04975$

Čebyševova nerovnost

Věta (Čebyševova nerovnost):
Nechť náhodná veličina X má střední hodnotu E(X) a rozptyl D(X). Pak platí Čebyševova nerovnost:

$$\boxed{\forall t > 0 : P\left(\left|X - E(X)\right| > t\sqrt{D(X)}\right) \le \frac{1}{t^2}.}$$

Ilustrace pro spojitý případ



Důkaz: Pro spojitý případ: Plyne z Markovovy nerovnosti, kde položíme $Y = [X - E(X)]^2$. Pak $P(Y > 0) = 1$ a pro $\forall \varepsilon > 0 : P(Y > \varepsilon E(Y)) \le \frac{1}{\varepsilon}$, tj. pro $\forall \varepsilon > 0 : P\left([X - E(X)]^2 > \varepsilon E\left([X - E(X)]^2\right)\right) \le \frac{1}{\varepsilon}$. Položme $\varepsilon = t^2$. Po odmocnění máme

$$\forall t > 0 : P\left(\left|X - E(X)\right| > t\sqrt{D(X)}\right) \le \frac{1}{t^2}.$$

Příklad

Příklad: Nechť $E(X) = \mu$, $D(X) = \sigma^2$.
a) Odhadněte $P\left(|X-\mu| > 3\sigma\right)$.
b) Jestliže $X \sim N(\mu, \sigma^2)$, vypočtěte $P\left(|X-\mu| > 3\sigma\right)$.

Řešení:

ad a) $P\left(|X-\mu| > 3\sigma\right) \leq \frac{1}{3^2} = \frac{1}{9} = 0{,}\bar{1}$.

(Tomuto výsledku se říká pravidlo 3σ a říká, že nejvýše 11,1% realizací náhodné veličiny leží vně intervalu (μ - 3σ, μ + 3σ).)

ad b) $P\left(|X-\mu| > 3\sigma\right) = 1 - P(-3\sigma \leq X - \mu \leq 3\sigma) = 1 - P(-3 \leq \frac{X-\mu}{\sigma} \leq 3) = 1 - \Phi(3) + \Phi(-3) = 2[1 - \Phi(3)] = 2(1 - 0{,}99865) =$ 0,0027. (Má-li náhodná veličina normální rozložení, pak pouze 0,27% realizací leží vně intervalu (μ – 3σ, μ + 3σ).)

Cauchy – Schwarzova – Buňakovského nerovnost

Věta (Cauchyova – Schwarzova – Buňakovského nerovnost):
Nechť $R(X_1, X_2)$ je koeficient korelace náhodných veličin X_1, X_2. Pak $\boxed{|R(X_1,X_2)| \le 1}$ a rovnost nastane tehdy a jen tehdy, když mezi veličinami X_1, X_2 existuje s pravděpodobností 1 úplná lineární závislost, tj. existují konstanty a, b tak, že $P(X_2 = a + bX_1) = 1$.

Důkaz: Zavedeme standardizované náhodné veličiny $U_i = \dfrac{X_i - E(X_i)}{\sqrt{D(X_i)}}$, $i = 1, 2$.

$0 \le D(U_1 \pm U_2) = D(U_1) \pm 2C(U_1,U_2) + D(U_2) = 2[1 \pm R(X_1,X_2)] \Rightarrow |R(X_1,X_2)| \le 1$.

Předpokládejme nejprve, že $R(X_1, X_2) = 1$. V tomto případě počítáme $D(U_1 - U_2) = 2[1 - R(X_1,X_2)] = 0$. To je možné jen tak, že

$P(U_1 = U_2) = 1$, tj. $P(U_1 - U_2 = 0) = 1$, tj. $1 = P\left(\dfrac{X_1 - E(X_1)}{\sqrt{D(X_1)}} = \dfrac{X_2 - E(X_2)}{\sqrt{D(X_2)}}\right) = P\left(X_2 = E(X_2) - \dfrac{\sqrt{D(X_2)}}{\sqrt{D(X_1)}}E(X_1) + \dfrac{\sqrt{D(X_2)}}{\sqrt{D(X_1)}}X_1\right)$, tudíž

$a = E(X_2) - \dfrac{\sqrt{D(X_2)}}{\sqrt{D(X_1)}}E(X_1)$, $b = \dfrac{\sqrt{D(X_2)}}{\sqrt{D(X_1)}}$.

Předpokládáme-li, že $R(X_1, X_2) = -1$, pak počítáme $D(U_1 + U_2)$.

Nechť naopak $P(X_2 = a + bX_1) = 1$. Pak $R(X_1,X_2) = R(X_1, a + bX_1) = \operatorname{sgn}(b)R(X_1,X_1) = \operatorname{sgn}(b) = \begin{cases} 1 \text{ pro } b > 0 \\ -1 \text{ pro } b < 0 \end{cases}$.

12. Slabý zákon velkých čísel a centrální limitní věta

S rostoucím počtem opakovaných nezávislých pokusů zjišťujeme, že empirické charakteristiky, které popisují výsledky těchto pokusů, se blíží teoretickým charakteristikám. Například relativní četnost úspěchu se blíží pravděpodobnosti úspěchu; průměr měření zatížených náhodnou chybou se blíží hledané neznámé střední hodnotě; empirická distribuční funkce se blíží distribuční funkci. Těmito skutečnostmi se zabývá Slabý zákon velkých čísel, specifikovaný např. Čebyševovou větou, nebo Bernoulliovou větou.

Podstatou centrální limitní věty je tvrzení, že náhodná veličina X, která vznikla jako součet velkého počtu vzájemně nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$ má za velmi obecných podmínek přibližně normální rozdělení. Nejjednodušší specifikací centrální limitní věty je Moivre-Laplaceova věta. Zobecněním Moivre-Laplaceovy věty je věta Lindbergova-Lévyova. Nejobecněji centrální limitní větu formuloval Ljapunov, jeho větu však nebudeme uvádět. V současné době, kdy databáze mají ohromné množství položek, je aplikace CLV nesmírně užitečná.

Při uvedení zmíněných vět se neobejdeme bez pojmu konvergence posloupnosti náhodných veličin. V počtu pravděpodobnosti se nabízí řada způsobů, jak konvergenci posloupnosti náhodných veličin definovat, my si uvedeme následující tři.

Typy konvergence posloupnosti NV

Definice:

Říkáme, že náhodná posloupnost $(X_1, X_2, \ldots, X_n, \ldots)$ konverguje k náhodné veličině X

(i.) *jistě*, právě když všechny realizace náhodné posloupnosti $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega), \ldots)$ konvergují k realizaci náhodné veličiny $X(\omega)$. Tedy platí:

$$\forall \omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)$$

[Jedná se o "obyčejnou" konvergenci číselné posloupnosti]

(ii.) *podle pravděpodobnosti*, právě když pro každé $\varepsilon > 0$ platí:

$$\lim_{n \to \infty} P(|X_n - X| \leq \varepsilon) = 1$$

[Při vzrůstajícím počtu pokusů jsou větší odchylky X_n od X krajně nepravděpodobné]

(iii.) *v distribuci*, právě když pro distribuční funkce $F_1(x_1) \sim X_1, \ldots, F_n(x_n) \sim X_n, \ldots$, popř. $F(x) \sim X$ platí:

$$\lim_{n \to \infty} F_n(x) = F(x) \text{ pro všechna } x, \text{ kde je funkce } F \text{ spojitá}$$

[Jedná se o nejslabší z uvedených typů konvergence, definuje se jen s užitím distribučních funkcí]

Typy konvergence posloupnosti NV

Poznámka:

Náhodná posloupnost $(X_1, X_2, \ldots, X_n, \ldots)$ může konvergovat i ke konstantě, což je v předchozí definici zahrnuto. Stačí uvažovat náhodnou veličinu X degenerovanou.

Věta:

Nechť $(X_1, X_2, \ldots, X_n, \ldots)$ je náhodná posloupnost.

1. Jestliže tato náhodná posloupnost konverguje k náh. vel. X jistě, pak k ní nutně konverguje i podle pravděpodobnosti. Konverguje-li k X podle pravděpodobnosti, pak k ní nutně konverguje i v distribuci. [Obrácené implikace obecně neplatí.]

2. K tomu, aby náhodná posloupnost $(X_1, X_2, \ldots, X_n, \ldots)$ konvergovala podle pravděpodobnosti k číslu μ, stačí splnění podmínek

$$\lim_{n\to\infty} E(X_n) = \mu \quad \wedge \quad \lim_{n\to\infty} D(X_n) = 0$$

Slabý zákon velkých čísel – Čebyševova věta

Věta: Čebyševova (slabý zákon velkých čísel)
Nechť náhodná posloupnost $(X_1, X_2, \ldots, X_n, \ldots)$ je posloupnost stochasticky nezávislých a stejně rozložených náhodných veličin se stejnou střední hodnotou μ a stejným rozptylem σ^2. Potom náhodná posloupnost aritmetických průměrů $(X_1, \frac{1}{2}\sum_{i=1}^{2} X_i, \ldots, \frac{1}{n}\sum_{i=1}^{n} X_i, \ldots)$ konverguje podle pravděpodobnosti ke střední hodnotě μ. Tedy pro každé $\varepsilon > 0$ platí:

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

neboli

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon) = 1$$

[Při velkém počtu nezávislých pokusů můžeme téměr jistě očekávat, že aritmetický průměr jednotlivých pokusů se bude od střední hodnoty μ lišit krajně nepatrně. Proto při dostatečně velkém n lze střední hodnotu μ odhadnout průměrem výsledků jednotlivých pokusů.]

Bernoulliova věta

Věta: Bernoulliova (důsledek Čebyševovy věty)
Nechť náhodná veličina Y_n udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, kdy úspěch nastává v každém pokusu s pravděpodobností ϑ, $0 < \vartheta < 1$. Pak posloupnost relativních četností $(Y_1, \frac{Y_2}{2}, \ldots \frac{Y_n}{n}, \ldots)$ konverguje podle pravděpodobnosti k pravděpodobnosti úspěchu ϑ. Tedy pro každé $\varepsilon > 0$ platí:

$$P(|\frac{Y_n}{n} - \vartheta| < \varepsilon) \geq 1 - \frac{\vartheta(1-\vartheta)}{n\varepsilon^2}$$

neboli

$$\lim_{n\to\infty} P(|\frac{Y_n}{n} - \vartheta| < \varepsilon) = 1$$

Příklad

Příklad:

Pravděpodobnost vyrobení zmetku je $\frac{12}{3000}$. Při výstupní kontrole bylo testováno 3000 výrobků. Jaká je pravděpodobnost, že relativní četnost výskytu zmetků se od pravděpodobnosti výskytu zmetku liší nejvýše o 0,01?

Řešení:

Označme Y_{3000} náhodnou veličinu udávající počet zmetků (úspěchů) v 3000 pokusech.
Potom $Y_{3000} \sim Bi(3000, \frac{12}{3000})$

Relativní četnost úspěchů by se s rostoucím n měla blížit k pravděpodobnosti úspěchu. My chceme určit pravděpodobnost, že pro $n = 3000$ se relativní četnost úspěchů od pravděpodobnosti úspěchu neodchýlí o více, než o 0,01. Tedy v Bernoulliove větě budeme za ε volit 0,01.

Pro každé $\varepsilon > 0$ platí: $P(|\frac{Y_n}{n} - \vartheta| < \varepsilon) \geq 1 - \frac{\vartheta(1-\vartheta)}{n\varepsilon^2}$.
Tedy $P(|\frac{Y_{3000}}{3000} - \frac{12}{3000}| < 0,01) \geq 1 - \frac{\frac{12}{3000}(1-\frac{12}{3000})}{3000 \cdot 0,01^2} \doteq 0,872$

Pokud bychom chtěli využít přímo Čebyševovu větu, pak bychom za X_i volili náhodnou veličinu s alternativním rozložením, kde jednička symbolizuje vyrobení zmetku (úspěch) a nula vyrobení kvalitního výrobku.
Tedy $X_i \sim A(\frac{12}{3000})$, $i = 1, \ldots, 3000$ $\quad E(X_i) = \frac{12}{3000}$ $\quad D(X_i) = \frac{12}{3000}(1 - \frac{12}{3000})$; $X_1, \ldots, X_{3000}$ jsou stoch. nezávislé. Dále stačí za ε volit 0,01 a dosadit do Čebyševovy věty. (Uvědomte si, že binomická náhodná veličina vzniká jako součet nezávislých, stejně rozložených alternativních náhodných veličin.)

Centrální limitní věta

Věta: Lindbergova-Lévyova (centrální limitní věta)
Nechť náhodná posloupnost $(X_1, \ldots, X_n, \ldots)$ je posloupnost stochasticky nezávislých a stejně rozložených náhodných veličin se stejnou střední hodnotou μ a stejným rozptylem σ^2. Uvažme součet $X = \sum_{i=1}^{n} X_i$ a odvoďme střední hodnotu a rozptyl nové náhodné veličiny X.

$E(X) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu$

$D(X) = D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$

Nyní uvažme standardizovaný součet $U_n = \frac{X - E(X)}{\sqrt{D(X)}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ [n můžeme libovolně zvětšovat]

Potom náhodná posloupnost standardizovaných součtů $(U_1, U_2, \ldots, U_n, \ldots)$ konverguje v distribuci k náhodné veličině $U \sim N(0,1)$. Tedy

$$\forall u \in \mathrm{R}: \quad \lim_{n \to \infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

Zkráceně píšeme $U_n \approx N(0,1)$ a říkáme, že U_n se asymptoticky řídí normálním standardizovaným rozložením. [Všimněte si, že $U_n = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}}$. Centrální limitní věta tedy tvrdí, že s rostoucím n se distribuční funkce průměrů náhodných veličin $X_1, \ldots, X_n$ blíží distribuční funkci normálního rozložení se střední hodnotou μ a rozptylem $\frac{\sigma^2}{n}$. Toto nastává bez ohledu na původní rozložení náhodné veličiny X.]

Moivre – Laplaceova věta

Věta: Moivre-Laplaceova (důsledek Lindbergovy-Lévyovy věty)
Nechť $Y_n \sim Bi(n, \vartheta)$, $n = 1, 2, \ldots$. Potom $E(Y_n) = n\vartheta \quad D(Y_n) = n\vartheta(1-\vartheta)$
a $U_n = \frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \approx N(0,1)$

[Moivre-Laplaceova věta říká, že při dostatečně velkém počtu nezávislých pokusů konverguje v distribuci binomické rozdělení k normálnímu.]

Poznámka:

Na základě Moivre-Laplaceova věty se používá přibližný vzorec, který nahrazuje pracný výpočet distribuční funkce binomického rozložení jednoduchým hledáním v tabulkách distribuční funkce normálního standardizovaného rozložení. Porovnejte

Přesný výpočet:

$P(Y_n \leq y) = \sum_{t=0}^{y} \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t}$ … náročná sumace

Aproximace normálním rozložením:

$P(Y_n \leq y) = P\left(\frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \leq \frac{y - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}\right) \approx \Phi\left(\frac{y - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}\right) \sim N(0,1)$,

kde $\Phi(u)$ je tabelovaná distribuční funkce standardizovaného normálního rozložení.

Aproximaci je vhodné použít pokud jsou splněny následující podmínky:
$n\vartheta(1-\vartheta) > 9 \quad \wedge \quad \frac{1}{n+1} < \vartheta < \frac{n}{n+1}$.

Příklad

V určité skupině zaměstnanců je 10% s příjmem, který překračuje celostátní průměr. Kolik zaměstnanců z této skupiny je třeba vybrat, aby s pravděpodobností aspoň 0,95 bylo mezi nimi 8% až 12% zaměstnanců s nadprůměrným příjmem?

Řešení:

X – počet zaměstnanců s nadprůměrným příjmem, X ~ Bi(n, 0,1), E(X) = 0,1n, D(X) = 0,09n,

$$0{,}95 \le P\left(0{,}08 \le \frac{X}{n} \le 0{,}12\right) = P\left(0{,}08n \le X \le 0{,}12n\right) = P\left(\frac{0{,}08n - 0{,}1n}{\sqrt{0{,}09n}} \le \frac{X - 0{,}1n}{\sqrt{0{,}09n}} \le \frac{0{,}12n - 0{,}1n}{\sqrt{0{,}09n}}\right) =$$

$$P\left(\frac{-\sqrt{n}}{15} \le \frac{X - 0{,}1n}{\sqrt{0{,}09n}} \le \frac{\sqrt{n}}{15}\right) \approx \Phi\left(\frac{\sqrt{n}}{15}\right) - \Phi\left(-\frac{\sqrt{n}}{15}\right) = 2\Phi\left(\frac{\sqrt{n}}{15}\right) - 1 \Rightarrow \Phi\left(\frac{\sqrt{n}}{15}\right) \ge 0{,}975,$$

tedy $\dfrac{\sqrt{n}}{15} \ge u_{0{,}975} = 1{,}96 \Rightarrow \sqrt{n} \ge 29{,}4 \Rightarrow n \ge 865.$

Příklad

100-krát nezávisle na sobě házíme kostkou. Jaká je pravděpodobnost, že šestka padne aspoň 20-krát?

Řešení:

Označme Y_{100} náhodnou veličinu, udávající počet padnutých šestek ve 100 hodech, $Y_{100} \sim Bi(100, \frac{1}{6})$.
Nejdříve ověříme podmínky pro použití aproximace normálním rozložením:
$n\vartheta(1-\vartheta) = 100 \cdot \frac{1}{6}(1-\frac{1}{6}) = \frac{500}{36} > 9 \quad \wedge \quad \frac{1}{101} < \frac{1}{6} < \frac{100}{101}$, tedy obě podmínky jsou splněny.
Hledanou pravděpodobnost odhadneme pomocí Moivre-Laplaceovy věty.

$$P(Y_{100} \geq 20) = 1 - P(Y_{100} < 20) = P(Y_{100} \leq 19) = 1 - P\left(\frac{Y_{100} - 100/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}} \leq \frac{19 - 100/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}}\right) =$$
$$1 - P(U_n \leq 0,626) \approx 1 - \Phi(0,626) = 1 - 0,73565 = 0,2635.$$

(Přesný výpočet pomocí softwaru by vyšel 0,2198.)

Aproximace binomického rozložení normálním rozložením nemusí být vždy nejvhodnější. Pro extrémně malé pravděpodobnosti úspěchu ϑ užíváme přibližný vzorec, který vychází z Poissonovy věty.

Poissonova věta

Věta: Poissonova

Nechť $Y_1, Y_2, \ldots$ je posloupnost stochasticky nezávislých náhodných veličin, $Y_n \sim Bi(n, \vartheta_n)$, $n = 1, 2, \ldots$ a platí $\lim\limits_{n\to\infty} n\vartheta_n = \lambda$. Pak posloupnost $Y_1, Y_2, \ldots Y_n, \ldots$ konverguje v distribuci k náhodné veličině $Y \sim Po(\lambda)$, tedy $Y_n \approx Po(\lambda)$.

[Náhodná veličina Y má Poissonovo rozložení s parametrem λ, náhodnou veličinu Y_n s binomickým rozložením lze aproximovat Poissonovýn rozložením.]

Příklad

Poznámka:

Na základě Poissonovy věty se používá přibližný vzorec, který nahrazuje pracný výpočet distribuční (resp. pravděpodobnostní) funkce binomického rozložení jednoduchým hledáním v tabulkách distribuční (resp. pravděpodobnostní) funkce Poissonova rozložení.

• $P(Y_n \leq y) = \sum_{t=0}^{y} \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t} \approx F_{n\vartheta}(y) \sim Po(n\vartheta)$, kde $F_{n\vartheta}(y)$ je distribuční funkce Poissonova rozložení s parametrem $\lambda = n\vartheta$

• $P(Y_n = y) = \binom{n}{y} \vartheta^y (1-\vartheta)^{n-y} \approx \frac{(n\vartheta)^y}{y!} e^{-n\vartheta}$ (Srovnej s pravděpodobnostní funkcí Poissonova rozložení, která je tabelovaná.)

Aproximaci je vhodné použít, pokud jsou splněny následující podmínky:

$n \geq 30 \wedge \vartheta \leq 0,1.$

Příklad

Během zkoušky spolehlivosti se přístroj porouchá s pravděpodobností 0,05. Jaká je pravděpodobnost, že při zkoušení 100 přístrojů se jich porouchá právě 5?

Řešení:

Označme Y_{100} náhodnou veličinu, udávající počet porouchaných přístrojů ve 100 zkouškách, $Y_{100} \sim Bi(100; 0,05)$.
Nejdříve ověříme podmínky pro použití aproximace Poissonovým rozložením:
$100 \geq 30 \quad \wedge \quad 0.05 \leq 0,1.$
Určení hledané pravděpodobnosti aproximací Poissonovým rozložením:

$P(Y_{100} = 5) \approx \frac{(100 \cdot 0,5)^5}{5!} e^{-100 \cdot 0,05}$, což nemusíme počítat, jelikož jde o pravděpodobnostní funkci Poissonova rozložení v bodě 5 s parametrem $\lambda = 100 \cdot 0,05$, která je v tabulkách.
Tedy $p_5(5) = 0,17547$
Určení hledané pravděpodobnosti přesným výpočtem:

$P(Y_{100} = 5) = \binom{100}{5} 0,05^5 (1 - 0,05)^{95} = \ldots = 0,18$

13. Statistické tabulky

Následující tabulky obsahují hodnoty:

- Pravděpodobnostní funkce Binomického rozložení
- Pravděpodobnostní funkce Poissonova rozložení
- Distribuční funkce standardizivaného normálního rozložení
- Kvantilů standardizivaného normálního rozložení
- Kvantilů rozložení $\chi 2$ rozložení
- Kvantilů Studentova rozložení
- Kvantilů Fisherova-Snedecorova rozložení

Pravděpodobnostní funkce binomického rozložení Bi(n,p) 1. část

		p												
n	x	.01	.05	.10	.15	.20	.25	.30	1/3	.35	.40	.45	.49	.50
2	0	.9801	.9025	.8100	.7225	.6400	.5625	.4900	.4444	.4225	.3600	.3025	.2601	.2500
	1	.0198	.0950	.1800	.2550	.3200	.3750	.4200	.4444	.4550	.4800	.4950	.4998	.5000
	2	.0001	.0225	.0100	.0225	.0400	.0625	.0900	.1111	.1225	.1600	.2025	.2401	.2500
3	0	.9703	.8574	.7290	.6141	.5120	.4219	.3430	.2963	.2746	.2160	.1664	.1327	.1250
	1	.0294	.1354	.2430	.3251	.3840	.4219	.4410	.4444	.4436	.4320	.4084	.3823	.3750
	2	.0003	.0071	.0270	.0574	.0960	.1406	.1850	.2222	.2389	.2880	.3341	.3674	.3750
	3	.0000	.0001	.0010	.0034	.0080	.0156	.0270	.0370	.0429	.0640	.0911	.1176	.1250
4	0	.9606	.8145	.6561	.5220	.4096	.3164	.2401	.1975	.1785	.1296	.0915	.0677	.0625
	1	.0388	.1715	.2916	.3685	.4096	.4219	.4116	.3951	.3845	.3456	.2995	.2600	.2500
	2	.0006	.0135	.0486	.0975	.1536	.2109	.2646	.2963	.3105	.3456	.3675	.3747	.3750
	3	.0000	.0005	.0036	.0115	.0256	.0469	.0756	.0988	.1125	.1536	.2005	.2400	.2500
	4	.0000	.0000	.0001	.0005	.0016	.0039	.0081	.0123	.0150	.0256	.0410	.0576	.0625
5	0	.9510	.7738	.5905	.4437	.3277	.2373	.1681	.1317	.1160	.0778	.0503	.0345	.0312
	1	.0480	.2036	.3280	.3915	.4096	.3955	.3602	.3292	.3124	.2592	.2059	.1657	.1562
	2	.0010	.0214	.0729	.1382	.2048	.2637	.3087	.3292	.3364	.3456	.3369	.3185	.3125
	3	.0000	.0011	.0081	.0244	.0512	.0879	.1323	.1646	.1811	.2304	.2757	.3060	.3125
	4	.0000	.0000	.0004	.0022	.0064	.0146	.0284	.0412	.0488	.0768	.1128	.1470	.1562
	5	.0000	.0000	.0000	.0001	.0003	.0010	.0024	.0041	.0053	.0102	.0185	.0283	.0312

Pravděpodobnostní funkce binomického rozložení Bi(n,p) 2. část

		p												
6	0	.9415	.7351	.5314	.3771	.2621	.1780	.1176	.0878	.0754	.0467	.0277	.0176	.0156
	1	.0571	.2321	.3543	.3993	.3932	.3560	.3025	.2634	.2437	.1866	.1359	.1014	.0938
	2	.0014	.0305	.0984	.1762	.2458	.2966	.3241	.3292	.3280	.3110	.2780	.2437	.2344
	3	.0000	.0021	.0146	.0425	.0819	.1318	.1852	.2195	.2355	.2765	.3032	.3121	.3125
	4	.0000	.0001	.0012	.0055	.0154	.0330	.0595	.0823	.0951	.1382	.1861	.2249	.2344
	5	.0000	.0000	.0001	.0004	.0015	.0044	.0102	.0165	.0205	.0369	.0609	.0864	.0938
	6	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0014	.0018	.0041	.0083	.0139	.0156
7	0	.9321	.6983	.4783	.3206	.2097	.1335	.0824	.0585	.0490	.0280	.0152	.0090	.0078
	1	.0659	.2573	.3720	.3960	.3670	.3115	.2471	.2048	.1848	.1306	.0872	.0603	.0547
	2	.0020	.0406	.1240	.2097	.2753	.3115	.3171	.3073	.2985	.2613	.2140	.1740	.1641
	3	.0000	.0036	.0230	.0617	.1147	.1730	.2269	.2561	.2679	.2903	.2918	.2786	.2734
	4	.0000	.0002	.0026	.0109	.0287	.0577	.0972	.1280	.1442	.1935	.2388	.2676	.2734
	5	.0000	.0000	.0002	.0012	.0043	.0115	.0250	.0384	.0466	.0774	.1172	.1543	.1641
	6	.0000	.0000	.0000	.0001	.0004	.0013	.0036	.0064	.0084	.0172	.0320	.0494	.0547
	7	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0006	.0016	.0037	.0068	.0078
8	0	.9227	.6634	.4305	.2725	.1678	.1001	.0576	.0390	.0319	0168	.0084	.0046	.0039
	1	.0746	.2793	.3826	.3847	.3355	.2670	.1977	.1561	.1373	.0896	.0548	.0352	.0312
	2	.0026	.0515	.1488	.2376	.2936	.3115	.2965	.2731	.2587	.2090	.1569	.1183	.1094
	3	.0001	.0054	.0331	.0839	.1468	.2076	.2541	.2731	.2786	.2787	.2568	.2273	.2188
	4	.0000	.0004	.0046	.0185	.0459	.0865	.1361	.1707	.1875	.2322	.2627	.2730	.2734
	5	.0000	.0000	.0004	.0026	.0092	.0231	.0467	.0683	.0808	.1239	.1719	.2098	.2188
	6	.0000	.0000	.0000	.0002	.0011	.0038	.0100	.0171	.0217	.0413	.0703	.1008	.1094
	7	.0000	.0000	.0000	.0000	.0001	.0004	.0012	.0024	.0033	.0079	.0164	.0277	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0002	.0007	.0017	.0033	.0039

Pravděpodobnostní funkce binomického rozložení Bi(n,p) 3. část

		p												
9	0	.9135	.6302	.3874	.2316	.1342	.0751	.0404	.0260	.0207	.0101	.0046	.0023	.0020
	1	.0830	.2985	.3874	.3679	.3020	.2253	.1556	.1171	.1004	.0605	.0339	.0202	.0176
	2	.0034	.0629	.1722	.2597	.3020	.3003	.2668	.2341	.2162	.1612	.1110	.0776	.0703
	3	.0001	.0077	.0446	.1069	.1762	.2336	.2668	.2731	.2716	.2508	.2119	.1739	.1641
	4	.0000	.0006	.0074	.0283	.0661	.1168	.1715	.2048	.2194	.2508	.2600	.2506	.2461
	5	.0000	.0000	.0008	.0050	.0165	.0389	.0735	.1024	.1181	.1672	.2128	.2408	.2461
	6	.0000	.0000	.0001	.0006	.0028	.0087	.0210	.0341	.0424	.0743	.1160	.1542	.1641
	7	.0000	.0000	.0000	.0000	.0003	.0012	.0039	.0073	.0098	.0212	.0407	.0635	.0703
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0009	.0013	.0035	.0083	.0153	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0003	.0008	.0016	.0020
10	0	.9044	.5987	.3487	.1969	.1074	.0563	.0282	.0173	.0135	.0060	.0025	.0012	.0010
	1	.0914	.3151	.3874	.3474	.2684	.1877	.1211	.0867	.0725	.0403	.0207	.0114	.0098
	2	.0042	.0746	.1937	.2759	.3020	.2816	.2335	.1951	.1757	.1209	.0763	.0495	.0439
	3	.0001	.0105	.0574	.1298	.2013	.2503	.2668	.2601	.2522	.2150	.1665	.1267	.1172
	4	.0000	.010	.0112	.0401	.0881	.1460	.2001	.2276	.2377	.2508	.2384	.2130	.2051
	5	.0000	.0001	.0015	.0085	.0264	.0584	.1029	.1366	.1536	.2007	.2340	.2456	.2461
	6	.0000	.0000	.0001	.0012	.0055	.0162	.0368	.0569	.0689	.1115	.1596	.1966	.2051
	7	.0000	.0000	.0000	.0001	.0008	.0031	.0090	.0163	.0212	.0425	.0746	.1080	.1172
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0014	.0030	.0043	.0106	.0229	.0389	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0005	.0016	.0042	.0083	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0010

Pravděpodobnostní funkce Poissonova rozložení Po(λ) 1. část

					λ					
x	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	8187	7408	6703	6065	5488	4966	4493	4066	3679
1	0905	1637	2222	2681	3033	3293	3476	3595	3659	3679
2	0045	0164	0333	0536	0758	0988	1217	1438	1647	1839
3	0002	0011	0033	0072	0126	0198	0284	0383	0494	0613
4	0000	0001	0003	0007	0016	0030	0050	0077	0111	0153
5	0000	0000	0000	0001	0002	0004	0007	0012	0020	0031
6	0000	0000	0000	0000	0000	0000	0001	0002	0003	0005
7	0000	0000	0000	0000	0000	0000	0000	0000	0000	0001

x	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	0,3329	3012	2725	2466	2231	2019	1827	1653	1496	1353
1	3662	3614	3543	3452	3347	3230	3106	2975	2842	2707
2	2014	2169	2303	2417	2510	2584	2640	2678	2700	2707
3	0738	0867	0998	1128	1255	1378	1496	1607	1710	1804
4	0203	0260	0324	0395	0471	0551	0636	0723	0812	0902
5	0045	0062	0084	0111	0141	0176	0216	0260	0309	0361
6	0008	0012	0018	0026	0035	0047	0061	0078	0098	0120
7	0001	0002	0003	0005	0008	0011	0015	0020	0027	0034
8	0000	0000	0001	0001	0001	0002	0003	0005	0006	0009
9	0000	0000	0000	0000	0000	0000	0001	0001	0001	0002

Pravděpodobnostní funkce Poissonova rozložení Po(λ) 2. část

	λ									
x	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0	11,0	12,0
0	0,0498	0183	0067	0025	0009	0003	0001	0000	0000	0000
1	1494	0733	0337	0149	0064	0027	0011	0005	0002	0001
2	2240	1465	0842	0446	0223	0107	0050	0023	0010	0004
3	2240	1954	1404	0892	0521	0286	0150	0076	0037	0018
4	1680	1954	1755	1339	0912	0573	0337	0189	0102	0053
5	1008	1563	1755	1606	1277	0916	0607	0378	0224	0127
6	0504	1042	1462	1606	1490	1221	0911	0631	0411	0255
7	0216	0595	1044	1377	1490	1396	1171	0901	0646	0437
8	0081	0298	0653	1033	1304	1396	1318	1126	0888	0655
9	0027	0132	0363	0688	1014	1241	1318	1251	1085	0874
10	0008	0053	0181	0413	0710	0993	1186	1251	1194	1048
11	0002	0019	0082	0225	0452	0722	0970	1137	1194	1144
12	0002	0006	0034	0113	0264	0481	0728	0948	1094	1144
13		0002	0013	0052	0142	0296	0504	0729	0926	1056
14		0001	0005	0022	0071	0169	0324	0521	0728	0905
15			0002	0009	0033	0090	0194	0347	0534	0724
16				0003	0014	0045	0109	0217	0367	0543
17				0001	0006	0021	0058	0128	0237	0383
18					0002	0009	0029	0071	0145	0256
19					0001	0004	0014	0037	0084	0161
20						0002	0006	0019	0046	0097
21						0001	0003	0009	0024	0055
22							0001	0004	0012	0030
23								0002	0006	0016
24								0001	0003	0008
25									0001	0004
26										0002

Distribuční funkce Φ(u) rozložení N(0,1)

	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)
0,00	0,50000	0,40	0,65542	0,80	0,78814	1,20	0,88493
0,01	0,50399	0,41	0,65910	0,81	0,79103	1,21	0,88686
0,02	0,50798	0,42	0,66276	0,82	0,79389	1,22	0,88877
0,03	0,51197	0,43	0,66640	0,83	0,79673	1,23	0,89065
0,04	0,51595	0,44	0,67003	0,84	0,79955	1,24	0,89251
0,05	0,51994	0,45	0,67364	0,85	0,80234	1,25	0,79435
0,06	0,52392	0,46	0,67724	0,86	0,80511	1,26	0,89617
0,07	0,52790	0,47	0,68082	0,87	0,80785	1,27	0,89796
0,08	0,53188	0,48	0,68439	0,88	0,81057	1,28	0,89973
0,09	0,53586	0,49	0,68793	0,89	0,81327	1,29	0,90147
0,10	0,53983	0,50	0,69146	0,90	0,81594	1,30	0,90320
0,11	0,54380	0,51	0,69497	0,91	0,81859	1,31	0,90490
0,12	0,54776	0,52	0,69847	0,92	0,82121	1,32	0,90658
0,13	0,55172	0,53	0,70194	0,93	0,82381	1,33	0,90824
0,14	0,55567	0,54	0,70540	0,94	0,82639	1,34	0,90988
0,15	0,55962	0,55	0,70884	0,95	0,82894	1,35	0,91149
0,16	0,56356	0,56	0,71226	0,96	0,83147	1,36	0,91309
0,17	0,56749	0,57	0,71655	0,97	0,83398	1,37	0,91466
0,18	0,57142	0,58	0,71904	0,98	0,83646	1,38	0,91621
0,19	0,57535	0,59	0,72240	0,99	0,83891	1,39	0,91774
0,20	0,57926	0,60	0,72575	1,00	0,84134	1,40	0,91924
0,21	0,58317	0,61	0,72907	1,01	0,84375	1,41	0,92073
0,22	0,58706	0,62	0,73237	1,02	0,84614	1,42	0,92220
0,23	0,59095	0,63	0,73565	1,03	0,84850	1,43	0,92364
0,24	0,59483	0,64	0,73891	1,04	0,85083	1,44	0,92507
0,25	0,59871	0,65	0,74215	1,05	0,85314	1,45	0,92647
0,26	0,60257	0,66	0,74537	1,06	0,85543	1,46	0,92786
0,27	0,60642	0,67	0,74857	1,07	0,85769	1,47	0,92922
0,28	0,61026	0,68	0,75175	1,08	0,85993	1,48	0,93056
0,29	0,61409	0,69	0,75490	1,09	0,86214	1,49	0,93189
0,30	0,61791	0,70	0,75804	1,10	0,86433	1,50	0,93319
0,31	0,62172	0,71	0,76115	1,11	0,86650	1,51	0,93448
0,32	0,62552	0,72	0,76424	1,12	0,86864	1,52	0,93574
0,33	0,62930	0,73	0,76730	1,13	0,87076	1,53	0,93699
0,34	0,63307	0,74	0,77035	1,14	0,87286	1,54	0,93822
0,35	0,63683	0,75	0,77337	1,15	0,87493	1,55	0,93943
0,36	0,64058	0,76	0,77637	1,16	0,87698	1,56	0,94062
0,37	0,64431	0,77	0,77935	1,17	0,87900	1,57	0,94179
0,38	0,64803	0,78	0,78230	1,18	0,88100	1,58	0,94295
0,39	0,65173	0,79	0,78524	1,19	0,88298	1,59	0,94408

u	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)
1,60	0,94520	2,00	0,97725	2,40	0,99180	3,10	0,99903
1,61	0,94630	2,01	0,97778	2,41	0,99202	3,12	0,99910
1,62	0,94738	2,02	0,97831	2,42	0,99224	3,14	0,99916
1,63	0,94845	2,03	0,97882	2,43	0,99245	3,16	0,99921
1,64	0,94950	2,04	0,97932	2,44	0,99266	3,18	0,99926
1,65	0,95053	2,05	0,97982	2,45	0,99286	3,20	0,99931
1,66	0,95154	2,06	0,98030	2,46	0,99305	3,22	0,99936
1,67	0,95254	2,07	0,98077	2,47	0,99324	3,24	0,99940
1,68	0,95352	2,08	0,98124	2,48	0,99343	3,26	0,99944
1,69	0,95449	2,09	0,98169	2,49	0,99361	3,28	0,99948
1,70	0,95543	2,10	0,98214	2,50	0,99379	3,30	0,99952
1,71	0,95637	2,11	0,98257	2,52	0,99413	3,32	0,99955
1,72	0,95728	2,12	0,98300	2,54	0,99446	3,34	0,99958
1,73	0,95818	2,13	0,98341	2,56	0,99477	3,36	0,99961
1,74	0,95907	2,14	0,98382	2,58	0,99506	3,38	0,99964
1,75	0,95994	2,15	0,98422	2,60	0,99534	3,40	0,99966
1,76	0,96080	2,16	0,98461	2,62	0,99560	3,42	0,99969
1,77	0,96164	2,17	0,98500	2,64	0,99585	3,44	0,99971
1,78	0,96246	2,18	0,98537	2,66	0,99609	3,46	0,99973
1,79	0,96327	2,19	0,98574	2,68	0,99632	3,48	0,99975
1,80	0,96407	2,20	0,98610	2,70	0,99653	3,50	0,99977
1,81	0,96485	2,21	0,98645	2,72	0,99674	3,55	0,99981
1,82	0,96562	2,22	0,98679	2,74	0,99683	3,60	0,99984
1,83	0,96638	2,23	0,98713	2,76	0,99711	3,65	0,99987
1,84	0,96712	2,24	0,98745	2,78	0,99728	3,70	0,99989
1,85	0,96784	2,25	0,98778	2,80	0,99744	3,72	0,99991
1,86	0,96856	2,26	0,98809	2,82	0,99760	3,80	0,99993
1,87	0,96926	2,27	0,98840	2,84	0,99774	3,85	0,99994
1,88	0,96995	2,28	0,98870	2,86	0,99788	3,90	0,99995
1,89	0,97062	2,29	0,98899	2,88	0,99801	3,95	0,99996
1,90	0,97128	2,30	0,98928	2,90	0,99813	4,00	0,99997
1,91	0,97193	2,31	0,98956	2,92	0,99825	4,05	0,99997
1,92	0,97257	2,32	0,98983	2,94	0,99836	4,10	0,99998
1,93	0,97320	2,33	0,99010	2,96	0,99846	4,15	0,99998
1,94	0,97381	2,34	0,99036	2,98	0,99856	4,20	0,99999
1,95	0,97441	2,35	0,99061	3,00	0,99865	4,25	0,99999
1,96	0,97500	2,36	0,99086	3,02	0,99874	4,30	0,99999
1,97	0,97558	2,37	0,99111	3,04	0,99882	4,35	0,99999
1,98	0,97615	2,38	0,99134	3,06	0,99889	4,40	0,99999
1,99	0,97670	2,39	0,99158	3,08	0,99897	4,45	1,00000

Kvantily standardizovaného normálního rozložení u_P

P	u_P	P	u_P	P	u_P	P	u_P
0,50	0,000	0,75	0,674	0,950	1,645	0,975	1,960
0,51	0,025	0,76	0,706	0,951	1,655	0,976	1,970
0,52	0,050	0,77	0,739	0,952	1,665	0,977	1,995
0,53	0,075	0,78	0,772	0,953	1,675	0,978	2,014
0,54	0,100	0,79	0,806	0,954	1,685	0,979	2,034
0,55	0,126	0,80	0,842	0,955	1,695	0,980	2,054
0,56	0,151	0,81	0,878	0,956	1,706	0,981	2,075
0,57	0,176	0,82	0,915	0,957	1,717	0,982	2,097
0,58	0,202	0,83	0,954	0,958	1,728	0,983	2,120
0,59	0,228	0,84	0,994	0,959	1,739	0,984	2,144
0,60	0,253	0,85	1,036	0,960	1,751	0,985	2,170
0,61	0,279	0,86	1,080	0,961	1,762	0,986	2,197
0,62	0,305	0,87	1,126	0,962	1,774	0,987	2,226
0,63	0,332	0,88	1,175	0,963	1,787	0,988	2,257
0,64	0,358	0,89	1,227	0,964	1,799	0,989	2,290
0,65	0,385	0,90	1,282	0,965	1,812	0,990	2,326
0,66	0,412	0,905	1,311	0,966	1,825	0,991	2,366
0,67	0,440	0,910	1,341	0,967	1,838	0,992	2,409
0,68	0,468	0,915	1,372	0,968	1,852	0,993	2,457
0,69	0,496	0,920	1,405	0,969	1,866	0,994	2,512
0,70	0,524	0,925	1,440	0,970	1,881	0,995	2,576
0,71	0,553	0,930	1,476	0,971	1,896	0,996	2,652
0,72	0,583	0,935	1,514	0,972	1,911	0,997	2,748
0,73	0,613	0,940	1,555	0,973	1,927	0,998	2,878
0,74	0,643	0,945	1,598	0,974	1,943	0,999	3,090

Kvantily Pearsonova rozložení $\chi^2(\nu)$

stupně volnosti	pravděpodobnost				
	0,005	0,01	0,025	0,05	0,1
1	0,0000	0,0002	0,0010	0,0039	0,0158
2	0,0100	0,0201	0,0506	0,1026	0,2107
3	0,0717	0,1148	0,2158	0,3519	0,5844
4	0,2070	0,2971	0,4844	0,7107	1,0636
5	0,4117	0,5543	0,8312	1,1455	1,6103
6	0,6757	0,8721	1,2373	1,6354	2,2041
7	0,9893	1,2390	1,6899	2,1673	2,8331
8	1,3444	1,6465	2,1797	2,7326	3,4895
9	1,7349	2,0879	2,7004	3,3251	4,1682
10	2,1559	2,5582	3,2470	3,9403	4,8652
11	2,6032	3,0535	3,8157	4,5748	5,5778
12	3,0738	3,5706	4,4038	5,2260	6,3038
13	3,5650	4,1069	5,0088	5,8919	7,0415
14	4,0747	4,6604	5,6287	6,5706	7,7895
15	4,6009	5,2293	6,2621	7,2609	8,5468
16	5,1422	5,8122	6,9077	7,9616	9,3122
17	5,6972	6,4078	7,5642	8,6718	10,085
18	6,2648	7,0149	8,2307	9,3905	10,865
19	6,8440	7,6327	8,9065	10,117	11,651
20	7,4338	8,2604	9,5908	10,851	12,443
21	8,0337	8,8972	10,283	11,591	13,240
22	8,6427	9,5425	10,982	12,338	14,041
23	9,2604	10,196	11,689	13,091	14,848
24	9,8862	10,856	12,401	13,848	15,659
25	10,520	11,524	13,120	14,611	16,473
26	11,160	12,198	13,844	15,379	17,292
27	11,808	12,879	14,573	16,151	18,114
28	12,461	13,565	15,308	16,928	18,939
29	13,121	14,256	16,047	17,708	19,768
30	13,787	14,953	16,791	18,493	20,599
40	20,707	22,164	24,433	26,509	29,051
50	27,991	29,707	32,357	34,764	37,689
60	35,534	37,485	40,482	43,188	46,459
70	43,275	45,442	48,758	51,739	55,329
80	51,172	53,540	57,153	60,391	64,278
90	59,196	61,754	65,647	69,126	73,291
100	67,328	70,065	74,222	77,929	82,358
200	152,24	156,43	162,73	168,28	174,84
300	240,66	245,97	253,91	260,88	269,07
500	422,30	429,39	439,94	449,15	459,93

stupně volnosti	pravděpodobnost				
	0,90	0,95	0,975	0,99	0,995
1	2,706	3,841	5,024	6,635	7,879
2	4,605	5,991	7,378	9,210	10,597
3	6,251	7,814	9,348	11,345	12,838
4	7,779	9,488	11,143	13,277	14,860
5	9,236	11,070	12,833	15,086	16,750
6	10,645	12,592	14,449	16,812	18,548
7	12,017	14,067	16,013	18,475	20,278
8	13,362	15,507	17,535	20,090	21,955
9	14,684	16,919	19,023	21,666	23,589
10	15,987	18,307	20,483	23,209	25,188
11	17,275	19,675	21,920	24,725	26,757
12	18,549	21,026	23,337	26,217	28,300
13	19,812	22,362	24,736	27,688	29,819
14	21,064	23,685	26,119	29,141	31,319
15	22,307	27,996	27,488	30,578	32,801
16	23,542	26,296	28,845	32,000	34,267
17	24,769	27,587	30,191	33,409	35,718
18	25,989	28,869	31,526	34,805	37,156
19	27,204	30,144	32,852	36,191	38,582
20	28,412	31,410	34,170	37,566	39,997
21	29,615	32,671	35,479	38,932	41,401
22	30,813	33,924	36,781	40,289	42,796
23	32,007	35,172	38,076	41,638	44,181
24	33,196	36,415	39,364	42,980	45,599
25	34,382	37,652	40,646	44,314	46,928
26	35,563	38,885	41,923	45,642	48,290
27	36,741	40,113	43,195	46,963	49,645
28	37,916	41,337	44,461	48,278	50,993
29	39,087	42,557	45,722	49,588	52,336
30	40,256	43,773	46,979	50,892	53,672
40	51,805	55,758	59,342	63,691	66,766
50	63,167	67,505	71,420	76,154	79,490
60	74,397	79,082	83,298	88,379	91,952
70	85,527	90,531	95,023	100,43	104,21
80	96,578	101,88	106,63	112,33	116,32
90	107,57	113,15	118,14	124,12	128,30
100	118,50	124,34	129,56	135,81	140,17
200	226,02	233,99	241,06	249,45	255,26
300	331,79	341,40	349,87	359,91	366,84
500	540,93	553,13	563,85	576,49	585,21

Kvantily Studentova rozložení t(n)

stupně volnosti	pravděpodobnost				
	0,90	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,303	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,961
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,878
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

Kvantily rozložení $F_{0,95}(\nu_1,\nu_2)$ - 1. část

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385
3	10,128	9,552	9,277	9,117	9,014	8,941	8,887	8,845	8,812
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999
5	6,608	5,786	5,410	5,192	5,050	4,950	4,876	4,818	4,773
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677
8	5,318	4,459	4,066	3,838	3,688	3,581	3,501	3,438	3,388
9	5,117	4,257	3,863	3,633	3,482	2,274	3,293	3,230	3,179
10	4,965	4,103	3,708	3,478	3,326	3,217	3,136	3,072	3,020
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646
15	4,543	3,682	3,287	3,056	2,901	2,791	2,707	2,641	2,588
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393
21	4,325	3,467	3,073	2,840	2,685	2,573	2,488	2,421	2,366
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,266
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211
40	4,085	3,232	2,839	2,606	2,450	2,336	2,249	2,180	2,124
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959
∞	3,842	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880

Kvantily rozložení $F_{0,95}(\nu_1,\nu_2)$ - 2. část

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	241´,88	243,91	245,95	248,01	249,05	250,09	251,14	252,20	253,25	254,32
2	19,396	19,413	19,429	19,446	19,454	19,462	19,471	19,479	19,487	19,496
3	8,786	8,745	8,703	8,660	8,639	8,617	8,594	8,572	8,549	8,527
4	5,964	5,912	5,858	5,803	5,774	5,746	5,717	5,688	5,658	5,628
5	4,735	4,678	4,619	4,558	4,527	4,496	4,464	4,431	4,398	4,365
6	4,060	4,000	3,938	3,874	3,842	3,808	3,774	3,740	3,705	3,669
7	3,637	3,575	3,511	3,445	3,411	3,376	3,340	3,304	3,267	3,230
8	3,347	3,284	3,218	3,150	3,115	3,079	3,043	3,005	2,967	2,928
9	3,137	3,073	3,006	2,937	2,901	2,864	2,826	2,787	2,748	2,707
10	2,978	2,913	2,845	2,774	2,737	2,700	2,661	2,621	2,580	2,538
11	2,854	2,788	2,719	2,646	2,609	2,571	2,531	2,490	2,448	2,405
12	2,753	2,687	2,617	2,544	2,506	2,466	2,426	2,384	2,341	2,296
13	2,671	2,604	2,533	2,459	2,420	2,380	2,339	2,297	2,252	2,206
14	2,602	2,534	2,463	2,388	2,349	2,308	2,266	2,223	2,178	2,131
15	2,544	2,475	2,404	2,328	2,288	2,247	2,204	2,160	2,114	2,066
16	2,494	2,425	2,352	2,276	2,235	2,194	2,151	2,106	2,059	2,010
17	2,450	2,381	2,308	2,230	2,190	2,148	2,104	2,058	2,011	1,960
18	2,412	2,342	2,269	2,191	2,150	2,107	2,063	2,017	1,968	1,917
19	2,378	2,308	2,234	2,156	2,114	2,071	2,026	1,980	1,930	1,878
20	2,348	2,278	2,203	2,124	2,083	2,039	1,994	1,946	1,896	1,843
21	2,321	2,250	2,176	2,096	2,054	2,010	1,965	1,917	1,866	1,812
22	2,297	2,226	2,151	2,071	2,028	1,984	1,938	1,890	1,838	1,783
23	2,275	2,204	2,128	2,048	2,005	1,961	1,914	1,865	1,813	1,757
24	2,255	2,183	2,108	2,027	1,984	1,939	1,892	1,842	1,790	1,733
25	2,237	2,165	2,089	2,008	1,964	1,919	1,872	1,822	1,768	1,711
26	2,220	2,148	2,072	1,990	1,946	1,901	1,853	1,803	1,749	1,691
27	2,204	2,132	2,056	1,974	1,930	1,884	1,836	1,785	1,731	1,672
28	2,190	2,118	2,041	1,959	1,915	1,869	1,820	1,769	1,714	1,654
29	2,177	2,105	2,028	1,945	1,901	1,854	1,806	1,754	1,698	1,638
30	2,165	2,092	2,015	1,932	1,887	1,841	1,792	1,740	1,684	1,622
40	2,077	2,004	1,925	1,839	1,793	1,744	1,693	1,637	1,577	1,509
60	1,993	1,917	1,836	1,748	1,700	1,649	1,594	1,534	1,467	1,389
120	1,911	1,834	1,751	1,659	1,608	1,554	1,495	1,429	1,352	1,254
∞	1,831	1,752	1,666	1,571	1,517	1,459	1,394	1,318	1,221	1,000

Kvantily rozložení $F_{0,975}(\nu_1,\nu_2)$ - 1. část

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28
2	38,506	39,000	39,165	39,248	39,298	39,331	39,355	39,373	39,387
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473
4	12,218	10,649	9,979	9,605	9,365	9,197	9,074	8,980	8,905
5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681
6	8,813	7,260	6,599	6,227	5,988	5,820	5,696	5,600	5,523
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823
8	7,571	6,060	5,416	5,053	4,817	4,652	4,529	4,433	4,357
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880
20	5,872	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837
21	5,827	4,420	3,819	3,475	3,250	3,090	2,969	2,874	2,798
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763
23	5,750	4,349	3,751	3,408	3,184	3,023	2,902	2,808	2,731
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677
26	5,659	4,266	3,670	3,329	3,105	2,945	2,824	2,729	2,653
27	5,633	4,242	3,647	3,307	3,083	2,923	2,802	2,707	2,631
28	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611
29	5,588	4,201	3,607	3,267	3,044	2,884	2,763	2,669	2,592
30	5,568	4,182	3,589	3,250	3,027	2,867	2,746	2,651	2,557
40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452
60	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334
120	5,152	3,805	3,227	2,894	2,674	2,515	2,395	2,299	2,222
∞	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114

Kvantily rozložení $F_{0,975}(\nu_1,\nu_2)$ - 2. část

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	968,93	976,71	984,87	993,10	997,25	1001,4	1005,6	1009,8	1014,0	1018,3
2	39,398	39,415	39,431	39,448	39,456	39,465	39,473	39,481	39,490	39,498
3	14,419	14,337	14,253	14,167	14,124	14,081	14,037	13,992	13,947	13,902
4	8,844	8,751	8,657	8,560	8,511	8,461	8,411	8,360	8,309	8,257
5	6,619	6,525	6,428	6,329	6,278	6,227	6,175	6,125	6,069	6,0115
6	5,461	5,366	5,269	5,168	5,117	5,065	5,013	4,959	4,905	4,849
7	4,761	4,666	4,568	4,467	4,415	4,362	4,309	4,256	4,199	4,142
8	4,295	4,200	4,101	4,000	3,947	3,894	3,840	3,784	3,728	3,670
9	3,964	3,868	3,769	3,667	3,614	3,560	3,506	3,449	3,392	3,333
10	3,717	3,621	3,522	3,419	3,365	3,311	3,255	3,198	3,140	3,080
11	3,526	3,430	3,330	3,226	3,173	3,118	3,061	3,004	2,944	2,883
12	3,374	3,277	3,177	3,073	3,019	2,963	2,906	2,848	2,787	2,725
13	3,250	3,153	3,053	2,948	2,893	2,837	2,7880	2,720	2,659	2,596
14	3,147	3,050	2,949	2,844	2,789	2,732	2,674	2,614	2,552	2,487
15	3,060	2,963	2,862	2,756	2,701	2,644	2,585	2,524	2,461	2,395
16	2,986	2,889	2,788	2,681	2,625	2,568	2,509	2,447	2,383	2,316
17	2,922	2,825	2,723	2,616	2,560	2,502	2,442	2,380	2,315	2,247
18	2,866	2,769	2,667	2,559	2,503	2,445	2,384	2,321	2,256	2,187
19	2,817	2,720	2,617	2,509	2,452	2,394	2,333	2,270	2,203	2,133
20	2,774	2,676	2,573	2,465	2408	2,349	2,287	2,223	2,156	2,085
21	2,735	2,637	2,534	2,425	2,368	2,308	2,247	2,182	2,114	2,042
22	2,700	2,602	2,498	2,389	2,332	2,272	2,210	2,145	2,076	2,003
23	2,668	2,570	2,467	2,357	2,299	2,239	2,176	2,111	2,042	1,968
24	2,640	2,541	2,437	2,327	2,269	2,209	2,146	2,080	2,010	1,935
25	2,614	2,515	2,411	2,301	2,242	2,182	2,118	2,052	1,981	1,906
26	2,590	2,491	2,387	2,276	2,217	2,157	2,093	2,026	1,955	1,878
27	2,568	2,469	2,364	2,253	2,195	2,133	2,069	2,002	1,930	1,853
28	2,547	2,448	2,344	2,232	2,174	2,112	2,048	1,980	1,907	1,829
29	2,529	2,430	2,325	2,213	2,154	2,092	2,028	1,959	1,886	1,807
30	2,511	2,412	2,307	2,195	2,136	2,074	2,009	1,940	1,866	1,787
40	2,388	2,288	2,182	2,068	2,007	1,943	1,875	1,803	1,724	1,637
60	2,270	2,169	2,061	1,945	1,882	1,815	1,744	1,667	1,581	1,482
120	2,157	2,055	1,945	1,825	1,760	1,690	1,614	1,530	1,433	1,310
∞	2,048	1,945	1,833	1,709	1,640	1,556	1,484	1,388	1,268	1,000

Kvantily rozložení $F_{0,99}(\nu_1,\nu_2)$ - 1. část

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	4052,2	4999,5	5403,5	5624,6	5763,7	5859,0	5928,3	5981,6	6022,5
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,639
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158
6	13,745	10,925	9,780	9,148	8,746	8,466	8,260	8,102	7,976
7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719
8	11,259	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911
9	10,561	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351
10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,745	4,632
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,005	3,895
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780
17	8,400	6,112	5,185	4,669	4,336	4,102	3,927	3,791	3,682
18	8,285	6,013	5,092	4,579	4,428	4,015	3,841	3,705	3,597
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457
21	8,017	5,780	4,874	4,369	4,042	3,812	3,640	3,506	3,398
22	7,945	5,719	7,817	4,313	3,988	3,758	3,587	3,453	3,346
23	7,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256
25	7,770	5,568	4,676	4,177	3,855	3,627	3,457	3,324	3,217
26	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,182
27	7,677	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120
29	7,598	5,421	4,538	4,045	3,725	3,500	3,330	3,198	3,092
30	7,563	5,390	4,510	4,018	3,699	3,474	3,305	3,173	3,067
40	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,719
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559
∞	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407

Kvantily rozložení $F_{0,99}(\nu_1,\nu_2)$ - 2. část

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	6055,8	6106,3	6157,3	6208,7	6234,6	6260,7	6286,8	6313,0	6339,4	6366,0
2	99,399	99,416	99,432	99,449	99,458	99,466	99,474	99,483	99,491	99,501
3	27,229	27,052	26,872	26,690	26,598	26,505	26,411	26,316	26,221	26,125
4	14,546	14,374	14,198	14,020	13,929	13,838	13,745	13,652	13,558	13,463
5	10,051	9,888	9,722	9,553	9,467	9,379	9,291	9,202	9,112	9,020
6	7,874	7,718	7,559	7,396	7,313	7,229	7,143	7,057	6,969	6,880
7	6,620	6,469	6,314	6,155	6,074	5,992	5,908	5,824	5,737	5,650
8	5,814	5,667	5,515	5,359	5,279	5,198	5,116	5,032	4,946	4,859
9	5,257	5,111	4,962	4,808	4,729	4,649	4,567	4,483	4,398	4,311
10	4,849	4,706	4,558	4,405	7,327	4,247	4,165	4,082	3,997	3,909
11	4,539	4,397	4,251	4,099	4,021	3,941	3,860	3,776	3,690	3,603
12	4,296	4,155	4,010	3,858	3,781	3,701	3,619	3,536	3,449	3,361
13	4,100	3,960	3,815	3,665	3,587	3,507	3,425	3,341	3,255	3,165
14	3,939	3,800	3,656	3,505	3,427	3,348	3,266	3,181	3,094	3,004
15	3,805	3,666	3,522	3,372	3,294	3,214	3,132	3,047	2,960	2,868
16	3,691	3,553	3,409	3,259	3,181	3,101	3,018	2,933	2,845	2,753
17	3,593	3,455	3,312	3,162	3,084	3,003	2,921	2,835	2,746	2,653
18	3,508	3,371	3,227	3,077	2,999	2,919	2,835	2,749	2,660	2,566
19	3,434	3,297	3,153	3,003	2,925	2,844	2,761	2,674	2,584	2,489
20	3,368	3,231	3,088	2,938	2,859	2,779	2,695	2,608	2,517	2,421
21	3,310	3,173	3,030	2,880	2,801	2,720	2,636	2,548	2,457	2,360
22	3,258	3,121	2,978	2,827	2,749	2,668	2,583	2,495	2,403	2,306
23	3,211	3,074	2,931	2,781	2,702	2,620	2,536	2,447	2,354	2,256
24	3,168	3,032	2,889	2,738	2,659	2,577	2,492	2,404	2,310	2,211
25	3,129	2,993	2,850	2,699	2,620	2,538	2,453	2,364	2,270	2,169
26	3,094	2,958	2,815	2,664	2,585	2,503	2,417	2,327	2,233	2,132
27	3,062	2,926	2,783	2,632	2,552	2,470	2,384	2,294	2,198	2,097
28	3,032	2,896	2,753	2,602	2,522	2,440	2,354	2,263	2,167	2,064
29	3,005	2,869	2,726	2,574	2,495	2,412	2,325	2,234	2,138	2,034
30	2,979	2,843	2,700	2,549	2,469	2,386	2,299	2,208	2,111	2,006
40	2,801	2,665	2,522	2,369	2,288	2,203	2,114	2,019	1,917	1,805
60	2,632	2,496	2,352	2,198	2,115	2,029	1,936	1,836	1,726	1,601
120	2,472	2,336	2,192	2,035	1,950	1,860	1,763	1,656	1,533	1,381
∞	2,321	2,185	2,039	1,878	1,791	1,696	1,592	1,473	1,325	1,000

Kvantily rozložení $F_{0,995}(\nu_1,\nu_2)$ - 1. část

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,37	199,39
3	55,552	49,799	47,467	46,196	45,392	44,838	44,434	44,126	43,882
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	21,352	21,139
5	22,785	18,314	16,530	15,556	14,940	14,513	14,200	13,961	13,772
6	18,635	14,544	12,917	12,028	11,464	11,073	10,786	10,566	10,391
7	16,236	12,404	10,882	10,050	9,522	9,155	8,885	8,678	8,514
8	14,688	11,042	9,597	8,805	8,3302	7,952	7,694	7,496	7,339
9	13,614	10,107	8,717	7,956	7,471	7,134	6,885	6,693	6,541
10	12,826	9,427	8,081	7,343	6,872	6,545	6,303	6,116	5,968
11	12,226	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,537
12	11,754	8,510	7,226	6,521	6,071	5,757	5,525	5,345	5,202
13	11,374	8,187	6,926	6,234	5,791	5,482	5,253	5,076	4,935
14	11,060	7,922	6,680	5,998	5,562	5,257	5,031	4,857	4,717
15	10,798	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,536
16	10,575	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,384
17	10,384	7,354	6,156	5,497	5,075	4,779	4,559	4,389	4,254
18	10,218	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,141
19	10,073	7,094	5,916	5,268	4,853	4,561	4,345	4,177	4,043
20	9,944	6,987	5,818	5,174	4,762	4,472	4,257	4,090	3,956
21	9,830	6,891	5,730	5,091	4,681	4,393	4,179	4,013	3,880
22	9,727	6,806	5,652	5,017	4,609	4,323	4,109	3,944	3,812
23	9,635	6,730	5,582	4,950	4,544	4,259	4,047	3,882	3,750
24	9,551	6,661	5,519	4,890	4,486	4,202	3,991	3,826	3,695
25	9,475	6,598	5,462	4,835	4,433	4,150	3,939	3,776	3,645
26	9,406	6,541	5,409	4,785	4,384	4,103	3,893	3,730	3,599
27	9,342	6,489	5,361	4,740	4,340	4,059	3,850	3,688	3,557
28	9,284	6,440	5,317	4,698	4,300	4,020	3,811	3,649	3,519
29	9,230	6,396	5,276	4,659	4,262	3,983	3,775	3,613	3,483
30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,451
40	8,828	6,066	4,976	4,374	3,986	3,713	3,509	3,350	3,222
60	8,495	5,795	4,729	4,140	3,760	3,492	3,291	3,134	3,008
120	8,179	5,539	4,497	3,921	3,548	3,285	3,087	2,933	2,808
∞	7,879	5,298	4,279	3,715	3,350	3,091	2,897	2,744	2,621

Kvantily rozložení $F_{0,995}(\nu_1,\nu_2)$ - 2. část

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	24,224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	199,40	199,42	199,43	199,45	199,46	199,47	199,47	199,48	199,49	199,51
3	43,686	43,387	43,085	42,778	42,622	42,466	42,308	42,149	41,989	41,829
4	20,967	20,705	20,438	20,167	20,030	19,892	19,752	19,611	19,468	19,325
5	13,618	13,384	13,146	12,903	12,780	12,656	12,530	12,402	12,274	12,144
6	10,250	10,034	9,814	9,589	9,474	9,358	9,241	9,122	9,002	8,879
7	8,380	8,176	7,968	7,754	7,645	7,735	7,423	7,309	7,193	7,076
8	7,211	7,015	6,814	6,608	6,503	6,396	6,288	6,177	6,065	5,951
9	6,417	6,227	6,033	5,832	5,729	5,625	5,519	5,410	5,300	5,188
10	5,847	5,661	5,471	5,274	5,173	5,071	4,966	4,859	4,750	4,639
11	5,418	5,236	5,049	4,855	4,756	4,654	4,551	4,445	4,337	4,226
12	5,086	4,906	4,721	4,530	4,432	4,331	4,228	4,123	4,015	3,904
13	4,820	4,643	4,460	4,270	4,173	4,073	3,970	3,866	3,758	3,647
14	4,603	4,428	4,247	4,059	3,961	3,862	3,760	3,655	3,547	3,436
15	4,424	4,250	4,070	3,883	3,786	3,687	3,585	3,480	3,372	3,260
16	4,272	4,099	3,921	3,734	3,638	3,538	3,437	3,332	3,224	3,112
17	4,142	3,971	3,793	3,607	3,511	3,412	3,311	3,206	3,097	2,984
18	4,031	3,860	3,683	3,498	3,402	3,303	3,201	3,096	2,987	2,873
19	3,933	3,763	3,587	3,402	3,306	3,208	3,106	3,000	2,891	2,776
20	3,847	3,678	3,502	3,318	3,222	3,123	3,022	2,916	2,806	2,690
21	3,771	3,602	3,427	3,243	3,147	3,049	2,947	2,841	2,730	2,614
22	3,703	3,535	3,360	3,176	3,081	2,982	2,880	2,774	2,663	2,546
23	3,642	3,475	3,300	3,117	3,021	2,922	2,820	2,713	2,602	2,484
24	3,587	3,420	3,246	3,062	2,967	2,868	2,765	2,659	2,546	2,428
25	3,537	3,370	3,196	3,013	2,918	2,819	2,716	2,609	2,496	2,377
26	3,492	3,325	3,152	2,969	2,873	2,774	2,671	2,563	2,450	2,330
27	3,450	3,284	3,110	2,928	2,832	2,733	2,630	2,522	2,408	2,287
28	3,412	3,246	3,073	2,890	2,794	2,695	2,592	2,483	2,369	2,247
29	3,377	3,211	3,038	2,855	2,759	2,660	2,557	2,448	2,333	2,210
30	3,344	3,179	3,006	2,823	2,727	2,628	2,524	2,415	2,300	2,176
40	3,117	2,953	2,781	2,598	2,502	2,402	2,296	2,184	3,064	1,932
60	2,904	2,742	2,571	2,387	2,290	2,187	2,079	1,962	1,834	1,688
120	2,705	2,544	2,373	2,188	2,089	1,984	1,871	1,747	1,606	1,431
∞	2,519	2,358	2,187	2,000	1,898	1,789	1,669	1,533	1,364	1,000

Literatura

- Budíková, Marie - Mikoláš, Štěpán - Osecký, Pavel. *Popisná statistika*. 3., doplněné vyd. Brno : Masarykova univerzita, 1998. 52 s. ISBN 80-210-1831-3.

- Budíková, Marie - Mikoláš, Štěpán - Osecký, Pavel. *Teorie pravděpodobnosti a matematická statistika. Sbírka příkladů.* 3. vyd. Brno : Masarykova univerzita, 2004. 127 s. ISBN 80-210-3313-4.

- Michal Friesl - výukové texty (např. Pravděpodobnost a statistika, Posbírané příklady z pravděpodobnosti a statistiky,…): http://home.zcu.cz/~friesl/Archiv/DldTeach.html

- Blanka Šedivá - Pravděpodobnost a statistika: http://home.zcu.cz/~sediva/pse/

- Michal Čihák - výukové texty: http://www.cihak.com/michal/

- Petr Otipka, Vladislav Šmajstrla - Pravděpodobnost a statistika: http://homen.vsb.cz/~oti73/cdpast1/

- Jana Novovičová - Pravděpodobnost a matematická statistika: http://euler.fd.cvut.cz/publikace/files/skripta3.pdf