Statistika

1 Popisná statistika

Relativní četnost

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

$0 \leq p(G) \leq 1$

Podmíněná relativní četnost Množiny G_1 za předpokladu G_2: $p(G_1/G_2) = \frac{n(G_1 \cap G_2)}{n(G_2)} = \frac{p(G_1 \cap G_2)}{p(G_2)}$.

Četnostní nezávislost dvou množin $\frac{n(G_1 \cap G_2)}{n(G_2)} = \frac{n(G_1)}{n} \wedge \frac{n(G_1 \cap G_2)}{n(G_1)} = \frac{n(G_2)}{n}$. Po úpravě dostaneme vztah $\frac{n(G_1 \cap G_2)}{n} = \frac{n(G_1)}{n} \cdot \frac{n(G_2)}{n}$, tj. $p(G_1 \cap G_2) = p(G_1) \cdot p(G_2)$.

Absolutní a relativní kumulativní četnost prvních j variant $N_j = n(X \leq x_{[j]}) = n_1 + \cdots + n_j$. $F_j = \frac{N_j}{n} = p_1 + \cdots + p_j$.

Variační řada

$x_{[j]}$	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

Četnostní funkce

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

Četnostní funkce je nezáporná a normovaná ($\sum_{x=-\infty}^{\infty} p(x) = 1$)

Emprická distribuční funkce

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

Empricická distribuční funkce je neklesající, zprava spojitá a normovaná ($\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$)

Simultánní absolutní a relativní četnost dvojice $(X_{[j]}, Y_{[k]})$

$$n_{jk} = n(X = x_{[j]} \wedge Y = y_{[k]})$$

$$p_{jk} = \frac{n_{jk}}{n}$$

Marginální absolutní četnost varianty $x_{[j]}$

$$n_{j\cdot} = n(X = x_{[j]}) = n_{j1} + \cdots + n_{js}$$

Absolutní kumulativní četnost dvojice $(X_{[j]}, Y_{[k]})$

$$N_{jk} = n(X \leq x_{[j]} \wedge Y \leq y_{[k]}) = \sum_{u \leq j} \sum_{v \leq k} n_{uv}$$

Simultánní četnostní funkce

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

Marginální četnostní funkce pro znaky X a Y

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

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

Mezi simultánní a marginální četnostní funkcí platí vztahy:

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

Četnostně nezávislé znaky Pro všechna $j = 1, \ldots, r$ a $k = 1, \ldots, s$ platí $p_{jk} = p_{j\cdot} \cdot p_{\cdot k}$, neboli $\forall (x, y) \in \mathbb{R}^2 : p(x, y) = p_1(x) \cdot p_2(y)$.

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

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

Intervalové rozložení četností Používáno, jestliže počet variant je blízký rozsahu souboru.

j-tý třídící interval znaku X $(u_j, u_{j+1}\rangle$

délka j-tého třídícího intervalu $d_j = u_{j+1} - u_j$

střed j-tého třídíciho intervalu $x_{[j]} = \frac{u_j + u_{j+1}}{2}$

Sturgesovo pravidlo Ke stanovení počtu tříd.

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

Četnostní hustota j-tého třídícího intervalu

$$f_j = \frac{p_j}{d_j}$$

Histogram Graf znázorňující rozložení četností skládající se z r obdélníků sestrojených nad třídícími intervaly. Obsah j-tého obdélníku je roven relativní četnosti p_j j-tého třídícího intervalu.

Hustota četnosti Schodovitá čára, shora omezuje histogram.

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

Je nezáporná a normovaná (intervalová empirická funkce se rovná jedné).

Intervalová empirická distribuční funkce

$$F(x) = \int_{-\infty}^{\infty} f(x)\, \mathrm{d}x$$

Je neklesající, spojitá a normovaná ($\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} = 1$).

Simultánní četnostní hustota v (j,k)-tém třídícím intervalu

$$f_{jk} = \frac{p_{jk}}{d_j \cdot d_k}$$

Marginální četnostní hustota v j-tém třídícím intervalu pro znak X

$$f_{j\cdot} = \frac{p_{j\cdot}}{d_j}$$

Simultánní hustota četnosti

$$f(x,y) = \begin{cases} f_{jk} & \text{pro } u_j < x \le u_{j+1}, v_k < y \le v_{k+1}, \\ & j = 1, \dots, r,\ k = 1, \dots, s \\ 0 & \text{jinak} \end{cases}$$

Hustota četnosti

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

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\, \mathrm{d}y$$

2 Číselné charakteristiky znaků

Nominální znaky Dají se porovnávat jen na rovnost (např. profese, barva očí, …).

Ordinální znaky Mají definovanou i relaci uspořádání (např. školní klasifikace).

Intervalové znaky Umožňují i operaci rozdílu, tj. stejný interval mezi jednou dvojicí a druhou dvojicí hodnot vyjadřuje i stejný rozdíl v extenzitě vlastnosti (např. teplota, IQ).

Poměrové znaky Umožňují i operaci podílu, tj. poměr mezi jednou dvojicí a druhou dvojicí hodnot vyjadřuje i stejný podíl v extenzionalitě vlastnosti (např. délka v cm).

Alternativní znaky Nabývají jen dvou hodnot (ano/ne, …).

Modus Nejčetnější varianta, respektive střed nejčetnějnějšího třídícího intervalu. Značí se $\hat{x}$.

Cramérův koeficient kontingence

$$K = \sum_{j=1}^{r} \sum_{k=1}^{s} \frac{\left(n_{jk} - \frac{n_{j\cdot} \cdot n_{\cdot k}}{n}\right)^2}{\frac{n_{j\cdot} \cdot n_{\cdot k}}{n}}$$

Cramérův koeficient $V = \sqrt{\frac{K}{n \cdot (m-1)}}$, kde $m = \min\{r, s\}$.

Význam hodnot Cramérova koeficientu:

$\langle 0; 0{,}1)$	zanedbatelná závislost
$\langle 0{,}1; 0{,}3)$	slabá závislost
$\langle 0{,}3; 0{,}7)$	střední závislost
$\langle 0{,}7; 1\rangle$	silná závislost

α-kvantil Je charakteristikou polohy. Je-li $\alpha \in (0; 1)$, pak α-kvantil x_α je číslo rozdělující uspořádaný datový soubor na dolní úsek, obsahující alespoň podíl α všech dat, a horní podíl obsahující $1 - \alpha$ všech dat.

$$n\alpha = \begin{cases} x_\alpha = \frac{x_{(c)} + x_{(c+1)}}{2} & \text{pro celé c} \\ x_\alpha = x_{(\lceil c \rceil)} & \text{pro necelé c} \end{cases}$$

Pro speciálně zvolená α užíváme názvů medián ($x_{0,5}$), dolní kvartil ($x_{0,25}$), horní kvartil ($x_{0,75}$), decily ($x_{0,1}, \ldots, x_{0,9}$) a percentily ($x_{0,01}, \ldots, x_{0,99}$)

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

Modus pro intervalově tříděná data

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

kde d_m je dolní mez modální třídy, n_m je četnost modální třídy a h je šířka třídy.

Kvantily pro intervalově tříděná data

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

kde d_P je dolní mez třídy obsahující příslušný P-kvantil, p_P je relativní četnost této třídy, F_{P-1} je kumulativní četnost předchozí třídy a h je šířka třídy.

Pořadí čísla v posloupnosti čísel Nechť $x_1, \ldots, x_n$ je posloupnost reálných čísel. Jsou-li čísla navzíjem různá, pak pořadím R_i čísla x_i rozumíme počet těch čísel x_l, ktera jsou menší nebo rovna x_i. Vyskytují-li se mezi v dané posloupnosti skupinky stejných čísel, pak každé takové skupince přiřadíme průměrné pořadí.

Spearmanův koeficient pořadové korelace

$$r_S = 1 - \frac{6}{n \cdot (n^2 - 1)} \cdot \sum_{i=1}^{n} (R_i - Q_i)^2,$$

kde R_i je pořadí hodnoty x_i a Q_i pořadí hodnoty y_i.

Koeficient nabývá hodnot $(-1, 1)$. Čím bližší je 1, tím silnější je přímá pořadová závislost mezi znaky X a Y, čím je bližší -1, tím silnější je nepřímá závislost. Koeficient je symetrický a rezistentní proti odlehlým hodnotám.

Aritmetický průměr $m_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

i-tá centrovaná hodnota $x_i - \bar{x}$, podle znaménka poznáme, je-li hodnota nad- či podprůměrná.

Průměr představuje „težiště“ dat – součet podprůměrných a nadprůměrných hodnot je v rovnováze. Je silně ovlivněn extrémními hodnotami a je vhodné ho použít při přibližně symetrickém rozložení dat.

Kvadratická odchylka

$$\sum_{i=1}^{n} (x_i - a)^2$$

Charakterizuje celkovou chybu, které se dopustíme, nahradíme-li celý soubor jedinou hodnotou a. Je minimální pro $a = \bar{x}$.

Rozptyl Průměrná kvadratická odchylka hodnot od jejich aritmetického průměru.

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = (\frac{1}{n} \sum_{i=1}^{n} x_i^2) - \bar{x}^2$$

. Je silně ovlivněn extrémními hodnotami a nehodí se pro nesymetrické rozložení dat. Rozptyl standardizovaných hodnot je roven 1.

Směrodatná odchylka Kladná odmocnina z rozptylu $s = \sqrt{s^2}$. Je nulová, pokud jsou všechny hodnoty stejné, jinak je kladná. Je silně ovlivněna extrémními hodnotami a nehodí se pro nesymetrické rozložení dat.

i-tá standardizovaná hodnota

$$\frac{x_i - \bar{x}}{s}$$

vyjadřuje, o kolik směrodatných odchylek se i-tá hodnota odchýlila od průměru.

Šikmost Charakteristika nesymetrie

$$\alpha_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}.$$

Je-li rozložení dat symetrické kolem aritmetického průměru, pak $\alpha_3 = 0$, má-li rozložení prodloužený pravý konec, pak $\alpha_3 > 0$ a jde o kladně zešikmené rozložení (levostranou symetrii) a naopak.

Špičatost Charaktreristika koncentrace dat kolem průměru

$$\alpha_4 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3.$$

Pokud $\alpha_4 = 0$ jde o normální špičatost, $\alpha_4 < 0$ – podnormální špičatost, $\alpha_4 > 0$ – nadnormální špičatost.

Kovariance Charakteristika společné variability dvou intervalových znaků.

$$s_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = (\frac{1}{n}\sum_{i=1}^{n} x_i y_y) - \bar{x}\bar{y}$$

Pokud se nadprůměrné (podprůměrné) hodnoty znaku X sdružují s nadprůměrnými (podprůměrnými) hodnotami znaku Y, budou kovariance vesměs kladné. Znamená to, že je mezi nimi určitý stupeň přímé lineární závislosti – jsou kladně korelované.

Pokud se nadprůměrné (podprůměrné) hodnoty znaku X sdružují s podprůměrnými (nadprůměrnými) hodnotami znaku Y, budou kovariance vesměs zápornéé. Znamená to, že je mezi nimi určitý stupeň nepřímé lineární závislosti – jsou záporně korelované.

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

Pearsonův koeficient korelace Charakteristika těsnosti závislosti dvou intervalových znaků.

$$r_{12} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s1} \cdot \frac{y_i - \bar{y}}{s_2} = \frac{s_{12}}{s_1 \cdot s_2}$$

$-1 \leq r_{12} \leq 1$, je-li $r_{12} = \pm 1$, pak dvojice (x_i, y_i) leží na rostoucí, resp. klesající přímce. Koeficient je symetrický, tj. $r_{12} = r_{21}$.

Početní pravidla pro číselné charakteristiky Je-li znak $Y = a + bX$, pak $\bar{y} = a + b\bar{x}$ a $s_2^2 = b^2 s_1^2$.

Je-li $U = X + Y$, pak $\bar{u} = \bar{x} + \bar{y}$ a $s_3^2 = s_1^2 + s_2^2 + 2s_{12}$.

Pro znaky $U = a + bX$ a $V = c + dY$ je jejich kovariance $s_{34} = b \cdot c \cdot s_{12}$.

Vážený aritmetický průměr

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{r} n_j \cdot x_{[j]}$$

Vážený rozptyl

$$s^2 = \frac{1}{n}\sum_{j=1}^{r} n_j \cdot (x_{[j]} - \bar{x})^2 = \frac{1}{n}\sum_{j=1}^{r} n_j \cdot x_{[j]}^2 - \bar{x}^2$$

Vážená kovariance

$$\begin{aligned} s_{12} &= \frac{1}{n}\sum_{j=1}^{r}\sum_{k=1}^{s} n_{jk} \cdot (x_{[j]} - \bar{x})^2 (y_{[k]} - \bar{y})^2 \\ &= \frac{1}{n}\sum_{j=1}^{r}\sum_{k=1}^{s} n_{jk} \cdot x_{[j]} \cdot y_{[k]} - \bar{x}\bar{y} \end{aligned}$$

Koeficient variace $\frac{s}{m}$ Užívaný pro poměrové znaky, jde o bezrozměrné číslo, často vyjádřené v procentech.

Geometrický průměr $\sqrt[n]{x_1 \cdots x_n}$. Užívaný, pokud jsou všechny hodnoty poměrového znaku kladné.

Vypočty zavedením pomocné proměnné $v_i = \frac{x_i - a}{h}$, kde a je střed třídy s nejvyšší četností a h je šířka třídy.

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

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

Vnitroskupinová variabilita

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

Někdy také značeno jako $s_{\#}^2$.

Meziskupinová variabilita

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

Někdy také značeno jako s_*^2.

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

3 Regresní analýza

Regresní přímka Lineární závislost $y = \beta_0 + \beta_1 x$ Požadujeme, aby výraz $q(\beta_0, \beta_1)$ (rozptyl) $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ nabýval minima vzhledem k β_0 a β_1, tj. právě když jsou jeho první derivace nulové.

Regresní přímka znaku Y na X $y = b_0 + b_1 x$, jejíž parametry minimalizují rozptyl $q(\beta_0, \beta_1)$.

Regresní odhad i-té hodnoty znaku Y

$$\hat{y} = b_0 + b_1 x_i,\ i = 1, \dots, n$$

Index determinace $\mathrm{ID}^2 = r_{12}^2 \in \langle 0, 1\rangle$ Čím bližší 1, tím lépe vystihuje přímka vystihuje závislost Y na X

Systém nominálních rovnic

$$\sum y_i = nb_0 + b_1 \sum x_i$$

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

Regresní přímka

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

Maticové vyjádření MNČ

$$b = (x^T \cdot X)^{-1} \cdot X^T \cdot y$$

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

Regresní přímka znaku X na Y $x = \overline{b_0} + \overline{b_1} y$, někdy též nazývána druhá regresní přímka.

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

Obě regresní přímky dohromady se nazývají **sdružené regresní přímky**.

Vlastnosti sdružených regresních přímek Protínají se v bodě (m_1, m_2), pro regresní parametry $b_1, \overline{b_1}$ platí:

- $b_1 \cdot \overline{b_1} = r_{12}^2$,
- b_1 i $\overline{b_1}$ mají stejná znaménka,
- je-li jeden nulový, pak je i druhý nulový
- $r \in \langle -1, 1\rangle \Rightarrow b_1 \cdot \overline{b_1} \in \langle 0, 1\rangle$,
- pro $r = \pm 1$ platí $b_0 = -\frac{b_0}{b_1}$

4 Počet pravděpodobnosti

Pokus 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.

Deterministický pokus Pokus, jehož opakování vede k jedinému možnému výsledku.

Náhodný pokus Pokus, jehož opakování vede právě k jednomu z možných výsledků.

Základní prostor Neprázdná množina možných výsledků náhodného pokusu. $\Omega = \{\omega_t | t \in T\}$, kde T je indexová množina.

Jevové pole Systém podmnožin $\mathcal{A}$ základního jevového prostoru Ω splňující následující axiomy:

- $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 - A_2 \in \mathcal{A}$
- $\Omega \in \mathcal{A}$
- $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Jev $A \in \mathcal{A}$

Měřitelný prostor Dvojice $(\Omega, \mathcal{A})$

Jistý jev Ω

Nemožný jev $\emptyset$

Jev A ma za následek jev B $A \subseteq B$

Minimální jevové pole $\mathcal{A}_{min} = \{\emptyset, \Omega\}$ rozliší pouze jev jistý a nemožný.

Jevové pole $\mathcal{A}_1 = \{\emptyset, \Omega, A, \bar{A}\}$ umožňuje rozeznat, zda nastal jev A, nebo opečný jev $\bar{A}$.

Maximální jevové pole $\mathcal{A}_{max} = \{A | A \subseteq \Omega\}$ rozliší jevy do všech podrobností.

Statistická definice pravděpodobnosti

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

Pravděpodobnost je reálná množinová funkce $P : \mathcal{A} \to \mathbb{R}$ splňující následující axiomy:

- $\forall A \in \mathcal{A} : P(A) \geq 0$ (nezápornost)
- $P(\Omega) = 1$ (normovanost)
- $A_1, A_2, \ldots \in \mathcal{A}$ jsou neslučitelné $\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ (společná aditivita)

Pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ Matematický model jednorázového provedení náhodného pokusu.

4.1 Diskrétní pravděpodobnost

Váhová funkce $\gamma : \Omega \to \mathbb{R}$ je kladná pouze na $\Gamma \subseteq \Omega$, jinak je nulová a splňuje podmínku $\sum_{\omega \subseteq \Omega} \gamma(\omega) = 1$

Diskrétní pravděpodobnost reálná množinová funkce $P : \mathcal{A} \to \mathbb{R}$, daná vzorcem $\forall A \in \mathcal{A} : P(A) = \sum_{\omega \subseteq \mathcal{A}} \gamma(\omega)$

4.2 Klasická pravděpodobnost

Klasická pravděpodobnost je funkce $P : \mathcal{A} \to \mathbb{R}$, daná vzorcem $\forall A \in \mathcal{A} : P(A) = \frac{m(A)}{m(\Omega)}$, kde $m(A)$ je počet příznivých výsledků a $m(\Omega)$ je počet všech možných výsledků (Ω je konečná).

Platí $P(A) = 0 \Leftrightarrow A = \emptyset$ a $P(A) = 1 \Leftrightarrow A = \Omega$.

4.3 Stochasticky nezávislé jevy

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

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

- $\emptyset$ a A jsou stochasticky nezávislé jevy,
- Ω a A jsou stochasticky nezávislé jevy,
- Jsou-li A, B stochasticky nezávislé jevy, pak i jevy $\bar{A}, B$, $A, \bar{B}$ a $\bar{A}, \bar{B}$ jsou stochasticky nezávislé.

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

Pokud jev v každém z pokusů nastane s pravděpodobností $\theta \in (0, 1)$, pak pravděpodobnost, že jev nastane:

- právě x-krát: $P_n(x) = \binom{n}{x} \cdot \vartheta^x \cdot (1 - \vartheta)^{n-x}$
- aspoň x_0-krát: $\sum_{x=x_0}^{n} P_n(x)$
- nejvýše x_1-krát: $\sum_{x=0}^{x_1} P_n(x)$
- aspoň x_0-krát a nejvýše x_1-krát: $\sum_{x=x_0}^{x_1} P_n(x)$

4.4 Podmíněná pravděpodobnost

Podmíněná pravděpodobnost za podmínky H funkce $P(./H) : \mathcal{A} \to \mathbb{R}$ daná vzorcem $\forall A \in \mathcal{A} : P(A/H) = \frac{P(A \cap H)}{P(H)}$, kde $H \in \mathcal{A}$ je jev s nenulovou pravděpodobností.

Pro podmíněnou pravděpodobnost platí:

- $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2/A_1) = P(A_2) \cdot P(A_1/A_2)$, pro $P(A_1) \neq 0 \neq P(A_2)$
- 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 $P(A_1) = 0$

Věta o násobení pravděpodobností Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $A_1, A_2, \ldots, A_n$ takové jevy, že $P(A_1 \cap \cdots \cap A_n) \neq 0$. Pak $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdot \cdots \cdot P(A_n/A_1 \cap \cdots \cap A_{n-1})$.

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $i \in I$ (I je nejvýše spočetná indexová množina), $H_i \in \mathcal{A}$ 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 tvoří úplný systém hypotéz)

Věta o úplné pravděpodobnosti $P(A) = \sum_{i \in I} P(H_i) \cdot P(A/H_i)$ pro libovolný jev $A \in \mathcal{A}$.

Bayesův vzorec $P(H_k/A) = \frac{P(H_k) \cdot P(A/H_k)}{P(A)}$, pro libovolnou hypotézu $H_k, k \in I$ a jev $A \in \mathcal{A}$ s nenulovou pravděpodobností.

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

4.5 Geometrická pravděpodobnost

n-rozměrný prostor Množina $R^n = (-\infty, \infty) \times \cdots \times (-\infty, \infty) = (-\infty, \infty)^n$, kde $n \in \mathbb{N}$.

n-rozměrné Borelovské pole $\mathcal{B}^n$ je minimální jevové pole R^n obsahující třídu polouzavřených intervalů $(-\infty, x_1\rangle \times \cdots \times (-\infty, x_n\rangle$. Prvky tohoto pole nazýváme **(n-rozměrnými) borelovskými množinami**

Borelovsky měřitelné zobrazení Zobrazení $X : \Omega \to R^n$ právě když úplný vzor každé borelovské množiny je jev.

Objem borelovské množiny $mes(G) = \int_G \dots \int \mathrm{d}x_1 \dots \mathrm{d}x_n$ pokud Riemanův integrál vpravo existuje. $(R^n, \mathcal{B}^n)$ je měřitelný prostor a $G \in \mathcal{B}^n$ borelovská množina.

Geometrická pravděpodobnost $\forall B \in \mathcal{B}^n, B \subseteq G : Q(B) = \frac{mes(B)}{mes(G)}$ pokud $mes(B)$ existuje a objem $mes(G)$ borelovské množiny G je nenulový a konečný.

5 Náhodné veličiny

Náhodná veličina borelovsky měřitelné zobrazení $X : \Omega \to R^n$. Pro $n = 1$ jde o **skalární NV**, jinak mluvíme o **náhodném vektoru**.

Transformovaná NV Je-li $X : \Omega \to R^n$ NV a $g : R^n \to R^m$ borelovsky měřitelná funkce, pak složené zobrazení $Y : \Omega \to R^m$, dané vzorcem $Y(\omega) = g(X(\omega))$ je transformovaná NV.

Distribuční funkce NV Udává pravděpodobnost jevu, že NV X se realizuje hodnotou nejvýše x. $\forall x \in R : \Phi(x) = P(X \leq x)$.

Vlastnosti:

- neklesající
- zprava spojitá, tj. pro libovolné, pevně dané, $x_0 \in R$ je $\lim_{x \to x_0^+} \Phi(x) = \Phi(x_0)$
- normovaná ($\lim_{x \to \infty} \Phi(x) = 1$, $\lim_{x \to -\infty} \Phi(x) = 0$)
- $\forall a, b \in R, a < b, \implies P(a < X \leq b) = \Phi(b) - \Phi(a)$
- pro libovolné $x_0 \in R : P(X = x_0) = \Phi(x_0) - \lim_{x \to x_0^-} \Phi(x)$

Marginální distribuční funkce Distribuční funkce $\Phi_i(x_i)$.

Simultánní distribuční funkce Distribuční funkce $\Phi(x_1, \dots, x_n)$.

$P(x_1 < X_1 \leq x_1 + h_1 \wedge x_2 < X_2 \leq x_2 + h_2) = \Phi(x_1+h_1, x_2+h_2) - \Phi(x_1+h_1, x_2) - \Phi(x_1, x_2+h_2) + \Phi(x_1, x_2)$

5.1 Diskrétní NV

Nabývá nejvýše spočetně mnoha izolovaných hodnot.

Pravděpodobnostní funkce $\pi(x) = P(X = x)$

$\Phi(x) = \sum_{t \leq x} \pi(t)$

5.2 Spojitá NV

Nabývá všech hodnot z určitého intervalu.

Hustota pravděpodobnosti $\Phi(x) = \int_{-\infty}^{x} \varphi(t)$

Nemá význam pravděpodobnosti

5.3 Stochasticky nezávislé NV

Náhodné veličiny jsou stochasticky nezávislé, když platí $\forall (X_1, \dots, X_n) \in R^n : \Phi(x_1, \dots, x_n) = \Phi_1(X_1) \cdot \dots \cdot \Phi_n(X_n)$.

Posloupnost stochasticky nezávislých NV $\{X_i\}_{i=1}^{\infty}$, právě, když pro všechna přirozená i jsou stochasticky nezávislé NV $X_1, \dots, X_n$.

5.4 Rozložení diskrétních NV

$X \sim L(\vartheta)$ – NV X má rozložení L s parametrem ϑ.

Degenerované rozložení NV X nabývá konstantní hodnoty μ, píšeme $X \sim Dg(\mu)$.

Alternativní rozložení NV X udává počet úspěchů v jednom pokusu, přičemž pravděpodobnost úspěchu je ϑ. Píšeme $X \sim A(\vartheta)$.

$$\pi(x) = \begin{cases} 1-\vartheta & \text{pro } x = 0 \\ \vartheta & \text{pro } x = 1 \\ 0 & \text{jinak} \end{cases}$$

$$\text{neboli } \pi(x) = \begin{cases} \vartheta^x (1-\vartheta)^{1-x} & \text{pro } x = 0, 1 \\ 0 & \text{jinak} \end{cases}$$

Binomické rozložení Náhodná veličina udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ. Píšeme $X \sim Bi(n, \vartheta)$.

$$\pi(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} \vartheta^x (1-\vartheta)^{n-x} & \text{pro } x = 0, \dots, n \\ 0 & \text{jinak} \end{cases}$$

Multinomické rozložení Zobecnění binomického rozložení. Složky náhodného vektoru $(X_1, \dots, X_k)$ udávají počty úspěchů (nastane jev $A_1, \dots, A_k$) v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnosti úspěchů jsou $\vartheta_1, \dots, \vartheta_k$. Předpokládáme, že při každém pokusu nastane právě jeden z jevů $A_, \dots, A_k$, přičemž platí $\vartheta_1 + \dots + \vartheta_k = 1$. Píšeme $X \sim Mu(n, \vartheta_1, \dots, \vartheta_k)$.

$$\pi(x_1, \dots, x_k) = \begin{cases} \frac{n!}{x_1! \cdot \dots \cdot x_k!} & \cdot \vartheta_1^{x_1} \cdot \dots \cdot \vartheta_k^{x_k} \\ & \text{pro } x_1, \dots, x_k \\ & \in \{1, \dots, n\}, \sum_{i=1}^k x_i = n \\ 0 & \text{jinak} \end{cases}$$

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 $X \sim Po(\lambda)$.

$$\pi(x) = \begin{cases} \frac{\lambda^x}{x!} \cdot \mathrm{e}^{-\lambda} & \text{pro } x = 0, 1, \dots \\ 0 & \text{jinak} \end{cases}$$

Negativní binomické (Pascalovo) rozložení 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 $X \sim NB(n, \vartheta)$.

$$\pi(x) = \begin{cases} \begin{pmatrix} n+x+1 \\ x \end{pmatrix} & \cdot \vartheta^n \cdot (1-\vartheta)^x \\ & x = 0, 1, \dots \\ 0 & \text{jinak} \end{cases}$$

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 $X \sim Ge(\vartheta)$.

$$\pi(X) = \begin{cases} (1-\vartheta)^x \cdot \vartheta & \text{pro } x = 0, 1 \\ 0 & \text{jinak} \end{cases}$$

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 $X \sim Hg(N, M, n)$.

$$\pi(x) = \begin{cases} \dfrac{\begin{pmatrix} M \\ x \end{pmatrix} \cdot \begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} & \text{pro } x = \max\{0, M-N+n\}, \\ & \dots, \min\{M, n\} \\ 0 & \text{jinak} \end{cases}$$

Rovnoměrné diskrétní rozložení Nechť G je konečná množina o n prvcích. Náhodná veličina X nabývá se stejnou pravděpodobností každé hodnoty z množiny G. Píšeme $X \sim Rd(G)$.

5.5 Rozložení spojitých NV

Rovnoměrné spojité rozložení Veličina X může nabýt jakéhokoliv hodnoty z intervalu (a,b) a to všude se stejnou pravděpodobností. Píšeme $X \sim Rs(a,b)$.

$$\varphi(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a,b) \\ 0 & \text{jinak} \end{cases}$$

$$\Phi(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & x \in (a,b) \\ 1 & x \ge b \end{cases}$$

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 $X \sim N(\mu, \sigma^2)$.

$\varphi(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$.

Standardizované normální rozložení Píšeme $U \sim N(0,1)$. $\Phi(-u) = 1 - \Phi(u)$.

Vlastnosti:

- $X \sim N(\mu, \sigma^2)$, pak $U = \frac{X-\mu}{\sigma} \sim N(0,1)$
- $X \sim N(\mu, \sigma^2)$ a $Y = a + bX$, pak $Y \sim (a + b\mu, b^2\sigma^2)$
- $X_1, \ldots, X_n$ jsou stochasticky nezávislé, $X_i \sim (\mu_i, \sigma_i^2)$, $Y = \sum_{i=1}^n X_i$, pak $Y \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
- 68 % hodnot leží v intervalu $(\mu-\sigma, \mu+\sigma)$, 95 % hodnot leží v intervalu $(\mu - 2\sigma, \mu + 2\sigma)$, 99 % hodnot leží v intervalu $(\mu - 3\sigma, \mu + 3\sigma)$

Dvojrozměrné normální rozložení Spojitý náhodný vektor $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ s parametry $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ a $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Píšeme $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\mu, \Sigma)$.

$\varphi(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \mathrm{e}^{\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\cdot\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)}$.

Logaritmicko normální rozložení Náhodná veličina $X \sim LN(\mu, \sigma)$ vzniká v situacích, kdy kladná konstanta logaritmu μ je násobena velkým množstvím nezávislých náhodných veličin, kolísajících mírně kolem jedničky.

$$\varphi(x) = \frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

Pearsonovo χ^2 rozložení $X = X_1 + \cdots + X_k \sim \chi^2(k)$

$$\varphi(x,k) = \begin{cases} \frac{1}{2^{k/2} \cdot \Gamma(k/2)} \cdot x^{k/2-1} \cdot \mathrm{e}^{-x/2} & x > 0 \\ 0 & \text{jinak} \end{cases}$$

$$\Gamma(s) = \int_0^\infty \mathrm{e}^{-t} \cdot t^{s-1} \mathrm{d}t$$

Studentovo rozložení Nechť X_1, X_2 jsou stochasticky nezávislé NV, $X_1 \sim N(0,1)$, $X_2 \sim \chi^2(n)$. Pak $X = \frac{X_1}{\sqrt{\frac{X_2}{n}}} \sim t(n)$.

$\varphi(x,n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n \cdot \pi} \cdot \Gamma(\frac{n}{2})} \cdot (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$

Fisher-Snedecorovo rozložení X_1, X_2 jsou stochasticky nezávislé NV, $X_i \sim \chi^2(n_i)$. Pak $X = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2)$.

$\varphi(x, n_1, n_2) = \frac{\Gamma(\frac{n_1+n_2}{2} n_1^{n_1/2} n_2^{n_2/2}}{\Gamma(n_1/2)\Gamma(n_2/2)} \cdot (\frac{x^{(n_1-2)/2}}{(n_2+n_1x)^{(n_1+n_2)/2}})$.

Cauchyho rozložení s parametry x_0 a $\gamma > 0$, kde x_0 je parametr určující umístění největší hodnoty rozdělení.

$\varphi(x, x_0, \gamma) = \frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} = \frac{1}{\pi}\left(\frac{\gamma}{(x-x_0)^2+\gamma^2}\right)$

Pro $x_0 = 0$ a $\gamma = 1$ mluvíme o **standardním Cauchyho rozdělení**. $\varphi(x,0,1) = \frac{1}{\pi \cdot (1+x^2)}$.

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 $X \sim Ex(\lambda)$.

$$\varphi(x) = \begin{cases} \lambda\mathrm{e}^{-\lambda x} & x > 0 \\ 0 & \text{jinak} \end{cases}$$

$$\Phi(x) = \begin{cases} 1-\mathrm{e}^{-\lambda x} & x>0 \\ 0 & \text{jinak} \end{cases}$$

$E(X) = \frac{1}{\lambda}$, $D(X) = \frac{1}{\lambda^2}$

Laplaceovo rozložení Náhodná veličina, která vznikne rozdílem dvou NV z exponenciálního rozložení, se řídí tímto rozložením. $\varphi(x,\mu,b) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b}) = \frac{1}{2b}\begin{cases}\exp(-\frac{\mu-x}{b}) & x<\mu \\ \exp(-\frac{x-\mu}{b}) & x\geq\mu\end{cases}$

Weibullovo rozložení Náhodná veličina $X \sim Wb(\delta,\epsilon)$ vyjadřuje dobu čekání na nějakou událost, která se každým okamžikem může dostavit se šancí úměrnou mocninné funkci pročekané doby. Přitom čísla $\delta > 0$ a $\epsilon > 0$ se nazývají parametry měřítka a formy.

$$\varphi(x,\delta\epsilon) = \begin{cases} \epsilon\cdot\delta\cdot x^{\epsilon-1}\cdot \mathrm{e}^{-\delta\cdot x^{\epsilon}} & x>0 \\ 0 & x\leq 0\end{cases}$$

5.6 Rozložení transformovaných NV

Nechť X je diskrétní náhodná veličina s pravděpodobnostní funkcí $\pi(x)$ a g je borelovská ryze monotónní funkce, tedy v oblasti $C \in R$ existuje inverzní funkce $g^{-1} = \tau$. Pak pravděpodobnostní funkce $\pi_*(y)$ transformované náhodné veličiny $Y = g(X)$ má tvar:

$$\pi_*(y) = \begin{cases}\pi(\tau(y)) & \text{pro } y\in C \\ 0 & \text{jinak.}\end{cases}$$

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 $C \in R$ existuje inverzní funkce $g^{-1} = \tau$ se spojitou a nenulovou derivací. Pak hustota $\varphi_*(y)$ transformované náhodné veličiny $Y = g(X)$ má tvar:

$$\varphi_*(y) = \begin{cases}\varphi(\tau(y))|\tau'(y)| & \text{pro } y\in C \\ 0 & \text{jinak.}\end{cases}$$

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: $\Phi_*(y) = P(X\in\Delta_1) + P(X\in\Delta_2) + \dots$, kde $\Delta_1,\Delta_2,\dots$ jsou ty intervaly, pro které $Y \leq y$.

Věta o konvoluci X_1, X_2 jsou stochasticky nezávislé diskrétní náhodné veličiny, $X_i \sim \pi_i(x_i), i\in\{1,2\} \implies Y = X_1+X_2 \sim \pi_*(y) = \sum_{x_1=-\infty}^{\infty}\pi_1(x_1)\pi_2(y-x_1) = \sum_{x_2=-\infty}^{\infty}\pi_1(y-x_2)\pi_2(x_2)$

$\pi_*(y)$ se nazývá konvoluce funkcí $\pi_1(x_1)$, $\pi_2(x_2)$.

Analogicky platí i pro spojité NV.

5.7 Číselné charakteristiky NV

Kvantil X je NV aspoň ordinálního charakteru a $\alpha \in (0,1)$. Číslo $K_\alpha(X)$ se nazývá α-kvantil NV X, jestliže splňuje nerovnosti $P(X \leq K_\alpha(X))) \geq \alpha \wedge P(X \geq K_\alpha(X)) \geq 1-\alpha$.

Jiné označení kvantilu: x_α.

Je-li X spojitá náhodná veličina, pak $K_\alpha(X)$ je takové číslo, pro které platí: $\alpha = \Phi(K_\alpha(X)) = \int_{-\infty}^{K_\alpha(X)}\varphi(x)\mathrm{d}x$.

- $X \sim N(0,1) \implies K_\alpha(X) \equiv u_\alpha$. $u_\alpha = -u_{1-\alpha}$.
- $X \sim t(n) \implies K_\alpha(X) \equiv t_\alpha(n)$. $t_\alpha(n) = -t_{1-\alpha}(n)$.
- $X \sim F(n_1,n_2) \implies K_\alpha(X) \equiv F_\alpha(n_1,n_2)$. $F_\alpha(n_1,n_2) = \frac{1}{F_{1-\alpha}(n_1,n_2)}$.

Nechť X je spojitá náhodná veličina s distribuční funkcí $F(x)$, $\alpha \in (0,1)$ a $g : R \to R$ ryze monotónní borelovská funkce. Pak pro α-kvantil transformované náhodné veličiny $Y = g(X)$ platí:

- Je-li g všude rostoucí funkce, pak $K_\alpha(Y) = g(K_alpha(X))$.
- Je-li g všude klesající funkce, pak $K_\alpha(Y) = g(K_{1-\alpha}(X))$.

Střední hodnota $E(X) = \sum_{x=-\infty}^{\infty} x\pi(X)$ pro diskrétní NV, pokud suma je konečná nebo absolutně konverguje. $E(X) = \int_{x=-\infty}^{\infty} x\varphi(x)\mathrm{d}x$ pro spojitou NV, pokud je integrál konečný nebo absolutně konverguje. Jinak řekneme, že střední hodnota neexistuje.

Pro transformované diskrétní NV platí $E(Y) = \sum_{x=-\infty}^{\infty} g(x)\pi(x)$. Analogicky i pro spojité NV.

Rozptyl $D(X) = E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2$.

V diskrétním případě je rozptyl dán vzorcem $D(X) = \sum_{x=-\infty}^{\infty}(x - E(X))^2\pi(x) = \sum_{x=-\infty}^{\infty} x^2\pi(x) - (E(X))^2$.

Ve spojitém případě je rozptyl dán vzorcem $D(X) = \int_{x=-\infty}^{\infty}(x - E(X))^2\varphi(x)\mathrm{d}x = \int_{x=-\infty}^{\infty} x^2\varphi(x)\mathrm{d}x - (E(X))^2$.

Směrodatná odchylka $\sqrt{D(X)}$

Centrovaná NV $X - E(X)$

Standardizovaná NV $\frac{X-E(X)}{\sqrt{D(X)}}$

Kovariance $C(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2)))$

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. Pro výpočet je vhodné použít vzorec $C(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$.

Koeficient kovariance Koeficient korelace je číslo, které charakterizuje těsnost lineární závislosti realizací náhodných veličin X_1, X_2. $R(X_1, X_2) = \frac{C(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$.

Důsledkem je $C(X_1, X_2) = \sum_{x_1=-\infty}^{\infty}\sum_{x_2=-\infty}^{\infty} x_1x_2\pi(x_1, x_2) - E(X_1)E(X_2)$ pro diskrétní NV. Analogicky platí i pro spojité NV.

r**-tý moment** $E((X - k)^2)$, pro $k = 0$ jde o r-tý **počáteční moment**, $k = E(X)$ jde o r-tý **centrální moment**

$r \times s$**-tý moment** $E((X - k_1)^r(X - k_2)^s)$, pro $k_1 = k_2 = 0$ jde o $r \times s$-tý **počáteční moment**, $k_1 = E(X_1)$, $k_2 = (X_2)$ jde o $r \times s$-tý **centrální moment**

Šikmost $A_3(X) = \frac{E((X-E(X))^3)}{(\sqrt{D(X)})^3}$

Špičatost $A_4(X) = \frac{E((X-E(X))^4)}{(\sqrt{D(X)})^4} - 3$

5.8 Vlastnosti číselných charakteristik NV

Střední hodnota

- $E(a) = a$
- $E(a + bX) = a + bE(X)$
- $E(X - E(X)) = 0$
- $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$
- pro stochasticky nezávislé NV $X_1, \ldots, X_n$: $E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i)$

Kovariance

- $C(a_1, X_2) = C(X_1, a_2) = C(a_1, a_2) = 0$
- $C(a_1 + b_1X_1, a_2 + b_2X_2) = b_1b_2C(X_1, X_2)$
- $C(X, X) = D(X)$
- $C(X_1, X_2) = C(X_2, X_1)$
- $C(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
- $C(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} C(X_i, Y_j)$

Rozptyl

- $D(a) = 0$
- $D(a + bX) = b^2D(X)$
- $D(X) = E(X^2) - (E(X))^2$
- $D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} C(X_i, X_j)$, pro nekorelované $D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i)$

Koeficient korelace

- $R(a_1, X_2) = R(X_1, a_2) = R(a_1, a_2) = 0$
- $R(a_1 + b_1X_1, a_2 + b_2X_2) = \mathrm{sgn}(b_1b_2)R(X_1, X_2)$

- $R(X, X) = 1$ pro $D(X) \neq 0$, $R(X, X) = 0$ jinak
- $R(X_1, X_2) = R(X_2, X_1)$
- $R(X_1, X_2) = \frac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$ pro $\sqrt{D(X_1)}\sqrt{D(X_2)} > 0$, $R(X_1, X_2) = 0$ jinak

Markovova nerovnost Nechť NV X má střední hodnotu $E(X)$ a $P(X > 0) = 1$. Pak platí $\forall \epsilon > 0 : P(X > \epsilon E(X)) \leq \frac{1}{\epsilon}$.

Čebyševova nerovnost Nechť NV X má střední hodnotu $E(X)$ a rozptyl $D(X)$. Pak platí $\forall t > 0 : P\left(|X - E(X)| > t\sqrt{D(X)}\right) \leq \frac{1}{t^2}$.

Cauchy-Schwarzova-Buňakovského nerovnost Nechť $R(X_1, X_2)$ je koeficient korelace náhodných veličin X_1, X_2. Pak $|R(X_1, X_2)| \leq 1$ 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 $P(X_2 = a + bX_1) = 1$.

5.9 Slabý zákon velkých čísel a centrální limitní věta

Čebyševova věta Nechť náhodná posloupnost $(X_1, X_2, \ldots, X_n, \ldots)$ je posloupnost stochasticky nezávislých a stejně rozložených NV se stejnou střední hodnotou μ a stejným rozptylem σ^2. Potom náhodná posloupnost aritmetických průměrů $(X_1, {}^1/_2 \sum_{i=1}^{2} X_i, \ldots, {}^1/_n \sum_{i=1}^{n} X_i, \ldots)$ konverguje podle pravděpodobnosti ke střední hodnotě μ. Tedy pro každé $\epsilon > 0$ platí:

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

neboli:

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \epsilon) = 1$$

Při velkém počtu nezávislých pokusů můžeme téměř 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 Nechť NV 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 $\epsilon > 0$ platí:

$$P(|\frac{Y_n}{n} - \vartheta| < \epsilon) \geq 1 - \frac{\vartheta(1-\vartheta)}{n\epsilon^2}$$

neboli:

$$\lim_{n\to\infty} P(|\frac{Y_n}{n} - \vartheta| < \epsilon) = 1$$

Lindbergova-Lévyova věta (centrální limitní věta)

Nechť náhodná posloupnost $(X_1, \ldots, X_n, \ldots)$ je posloupnost stochasticky nezávislých a stejně rozložených NV se stejnou střední hodnotou μ a stejným rozptylem σ^2. Uvažme součet $X = \sum_{i=1}^{n} X_i$ a odvoďmě střední hodnotu a rozptyl nové náhodné veličiny X: $E(X) = E(\sum_{i=1}^{n} X_i) = \cdots = n\mu$, $D(X) = D(\sum_{i=1}^{n} X_i) = \cdots = n\sigma^2$.

Nyní uvažme standardizovaný součet $U_n = \frac{X - E(X)}{\sqrt{D(X)}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$.

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 $\forall u \in R : \lim_{n\to\infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$

Moivre-Laplaceova věta Nechť $Y_n \sim Bi(n, \vartheta), n = 1, 2, \ldots$. Potom $E(Y_n) = n\vartheta$, $D(Y_n) = n\vartheta$ a $U_n = \frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \approx N(0, 1)$.

Poissonova věta 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 $\lim_{n\to\infty} n\vartheta_n = \lambda$. Pak posloupnost $Y_1, Y_2, \ldots, Y_n, \ldots$ konverguje v distribuci k náhodné veličině $Y \sim Po(\lambda)$, tedy $Y_n \approx Po(\lambda)$.