Náhodná veličina

Jav, ktorý za určitých podmienok v závislosti na náhode môže, ale nemusí nastať, sa nazýva *náhodný jav*. Samotná realizácia podmienok, za ktorých tento jav môže nastať, sa nazýva *náhodný pokus*. Ak je náhodný jav výsledkom náhodného pokusu, ktorý je opakovateľný, nazýva sa *hromadný náhodný jav*. Javy tohto druhu sú predmetom skúmania teórie pravdepodobnosti. Možnosť výskytu hromadného náhodného javu A sa charakterizuje číslom $P(A)$, ktoré nazývame pravdepodobnosťou javu A. Pravdepodobnosť je číslo, nadobúdajúce hodnoty z intervalu $\langle 0, 1\rangle$. Ak je jav *A javom nemožným*, je $P(A) = 0$, ak je jav *A javom istým*, je $P(A) = 1$. Možnosť výskytu javu A je potom tým väčšia, čím je jeho pravdepodobnosť bližšie k 1. Povedali sme, že číslo $P(A)$ udáva pravdepodobnosť, že náhodný jav A za určitých podmienok nastane. K týmto základným podmienkam pripájame niekedy ešte ďalšiu podmienku, že totiž nastal jav B. Zaujíma nás potom pravdepodobnosť javu A za podmienky, že nastal jav B.

Takáto pravdepodobnosť sa nazýva *podmienená pravdepodobnosť* javu A a označuje sa $P(A|B)$. S ohľadom na to sa potom pravdepodobnosť $P(A)$ často nazýva *nepodmienená pravdepodobnosť* javu A. Ak je pravdepodobnosť javu A rôzna podľa toho, či jav B nastal, či nenastal, nazývame javy A, B *závislé*. Pokiaľ sa pravdepodobnosť javu A nemení v závislosti na tom, či jav B nastal či nenastal, nazývajú sa javy *A*, *B nezávislé.*

Výsledok náhodného pokusu môžeme niekedy popísať len slovne. Niekedy však sú možnými výsledkami náhodného pokusu hodnoty nejakej veličiny. Veličinu, ktorá pri náhodnom pokuse môže závisle na náhode nadobúdať rôzne hodnoty, nazývame *náhodná veličina.* Za náhodnú veličinu môžeme považovať napr. „počet bodiek, ktorý padne pri jednom vrhu hracej kocky“. Táto veličina nadobúda závisle na náhode hodnoty 1, 2, 3, 4, 5, 6. *Zákon rozdelenia* náhodnej veličiny X je zákon, ktorý udáva pravdepodobnosti javov, ktoré môžu nastať.

Náhodnou veličinou alebo náhodnou premennou sa nazýva premenná, ktorá vo výsledku pokusu môže nadobudnúť niektorú hodnotu, pričom je vopred neznáme akú konkrétnu.

$$\mathbf{P}\ (\boldsymbol{X} = \boldsymbol{x})$$

Náhodné premenné, ktoré môžu nadobudnúť iba jednotlivé, navzájom rôzne hodnoty, ktoré možno vopred vymenovať, nazývajú sa *nespojitými, alebo diskrétnymi náhodnými premennými.*

Napr.: počet zásahov pri troch výstreloch
počet telefonických hovorov prichádzajúcich na telefónnu linku

Náhodné premenné, ktoré nie sú navzájom oddelené, ale spojite vyplňujú určitý interval, sa nazývajú *spojitými náhodnými premennými*.

Napr.: chyba meracieho prístroja

Náhodné premenné teda delíme na dva základné typy:

Diskrétne: ak náhodná premenná nadobúda konkrétnu hodnotu, pričom počet týchto hodnôt je konečný alebo spočítateľný.
Náhodná premenná X nadobúda hodnoty $x_1, x_2, x_3, \ldots, x_n$. Ak hodnoty x_i a k ním zodpovedajúce pravdepodobnosti p_i usporiadame do tabuľky, tak táto tabuľka s podmienkou

$$\sum_{i=1}^{n} p_i = 1$$

je zákonom rozdelenia pravdepodobnosti diskrétnej náhodnej premennej a voláme ju *pravdepodobnostná tabuľka.*

Spojité: ak náhodná premenná nadobúda všetky hodnoty z uzavretého alebo otvoreného intervalu reálnych čísel.
Pri spojitej náhodnej premennej nemôžme vymenovať všetky hodnoty a k ním určiť všetky pravdepodobnosti (t. j. vytvoriť pravdepodobnostnú tabuľku), lebo ich je na reálnom intervale nespočítateľne veľa, preto tento vzťah popíšeme pomocou funkcie, ktorú nazývame *hustota*.

Aby sme mohli popísať náhodnú premennú musíme poznať:
1. *Množinu všetkých hodnôt*, ktoré náhodná premenná nadobúda – obor hodnôt náhodnej premennej.
2. *Pravdepodobnosť* pri ktorej môže náhodná premenná nadobudnúť svoju hodnotu.

Ak náhodná premenná bude úplne opísaná z pravdepodobnostného hľadiska, a ak zadáme toto rozdelenie, t.j. presne ukážeme, akú pravdepodobnosť má každý z javov, určíme tým tzv. *zákon rozdelenia náhodnej premennej*. Zákon rozdelenia náhodnej premennej je každý predpis, ktorý určuje vzťah možnými hodnotami náhodnej premennej a im zodpovedajúcimi pravdepodobnosťami. Najjednoduchšou formou zadania tohto zákona je tabuľka, v ktorej sú uvedené možné hodnoty náhodnej premennej a im zodpovedajúce pravdepodobnosti. Takúto tabuľku nazývame pravdepodobnostnou tabuľkou.

Príklad:
Vykoná sa jeden pokus, v ktorom sa môže objaviť alebo neobjaviť jav *A*. Pravdepodobnosť javu *A* je 0,3. Vyšetruje sa náhodná premenná X – počet objavení javu *A* v danom pokuse. Premenná X má iba dve hodnoty: 0 a 1. Pravdepodobnostná tabuľka premennej X má tvar:

x_i	0	1
p_i	0,7	0,3

Distribučná funkcia

Pravdepodobnostná tabuľka vyjadruje zákon rozdelenia pravdepodobnosti diskrétnej náhodnej premennej. Pre spojitú náhodnú premennú takúto tabuľku nevieme zostrojiť. Spojitá náhodná premenná nadobúda nespočítateľne veľa hodnôt. Pravdepodobnosť toho, že spojitá náhodná veličina nadobudne určitú konkrétnu číselnú hodnotu je veľmi blízka nule. Budeme uvažovať pravdepodobnosť javu, že náhodná premenná X nadobudne hodnoty menšie ako zadaná hodnota x. Táto pravdepodobnosť $P(X < x)$ je závislá na x. Túto funkciu označujeme $F(x)$ a voláme ju „Distribučnou funkciou“.

Je definovaná vzťahom: $\boldsymbol{F(x) = P(X < x)}$

Distribučná funkcia je všeobecnou formou popisu náhodnej premennej a dá sa použiť ako pre diskrétnu tak aj pre spojitú náhodnú premennú.

Vlastnosti distribučnej funkcie:

1. Distribučná funkcia $F(x)$ je neklesajúcou funkciou svojho argumentu, t.j. pri $x_2 > x_1$ je $F(x_2) \geq F(x_1)$.
2. V bode mínus nekonečno sa hodnota funkcie rovná nule: $F(-\infty) = 0$
3. V bode plus nekonečno sa distribučná hodnota rovná jednej: $F(+\infty) = 1$

Hodnota distribučnej funkcie $F(x)$ sa nazýva aj integrálnou funkciou rozdelenia, alebo integrálnym zákonom rozdelenia. Distribučná funkcia je najuniverzálnejšou charakteristikou náhodnej premennej, ktorá existuje pre všetky náhodné premenné a to diskrétne aj spojité.

Stredná hodnota náhodnej premennej

Strednou hodnotou náhodnej premennej je určité číslo, ktoré ako keby bolo jej reprezentantom a ktoré ju pri približne orientovaných výpočtoch nahradzuje. Parametre polohy charakterizujú „polohu“ náhodnej premennej na číselnej osi a jej „koncentráciu“.

Matematické vyčíslenie strednej hodnoty náhodne premennej:

$$E(X) = \frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \cdots + x_n \cdot p_n}{p_1 + p_2 + \cdots p_n}$$

$$E(X) = \frac{\sum_{i=1}^{n} x_i \cdot p_i}{\sum_{i=1}^{n} p_i}$$

Strednou hodnotou **diskrétnej** náhodnej premennej je súčet súčinov všetkých možných hodnôt náhodnej premennej s pravdepodobnosťami týchto hodnôt.

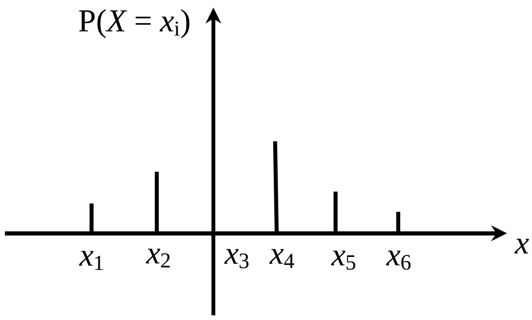
Platí: $\sum_{i=1}^{n} p_i = 1 \qquad E(X) = \sum_{i=1}^{n} x_i \cdot p_i \qquad F(x) = P(X < \mathrm{x}) = \sum_{x_i : x_i < \mathrm{x}} p_i$

p_{i} – pravdepodobnosť pokusu

Stredná hodnota **spojitej** náhodnej premennej je definovaná vzťahom

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, \mathrm{dx}$$

Diskrétne (nespojité) náhodné veličiny sú veličiny, ktoré nadobúdajú iba konečné alebo spočitateľné množstvo rôznych hodnôt, napr. hodnôt $x = 0, 1, 2, \ldots$. Ich zákon rozdelenia je popísaný pravdepodobnosťami jednotlivých hodnôt $p_a = \mathrm{P}(x = x_{\mathrm{a}})$ a zobrazený na obrázku:

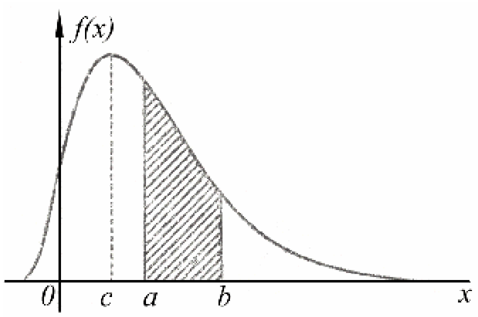


Obrázok: Graf rozdelenia pravdepodobnosti diskrétnej náhodnej premennej nadobúdajúcej hodnoty $x_1, \ldots, x_6$.

Spojité náhodné veličiny sú veličiny, ktoré môžu nadobúdať v rámci určitého intervalu všetky možné hodnoty. Sú to veličiny, ktoré majú tzv. spojitý zákon rozdelenia, ktorý je popísaný hustotou pravdepodobnosti $f(x) \geq 0$, ktorej integráciou na uzavretom intervale $\langle a, b \rangle$ dostaneme pravdepodobnosť, že náhodná veličina bude z intervalu $\langle a, b \rangle$:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)\, \mathrm{dx}\,.$$

Ak je hustota pravdepodobnosti $f(x)$ spojitá, potom $f(x)$ dx vyjadruje pravdepodobnosť toho, že náhodná veličina X bude z intervalu $\langle x, x + \mathrm{d}x \rangle$. Hustota pravdepodobnosti spojitej náhodnej veličiny je zobrazená na obrázku:



Obrázok: Hustota pravdepodobnosti spojitej náhodnej veličiny.
Vyšrafovaná plocha zodpovedá pravdepodobnosti
$P\{a \leq x \leq b\}$.

Distribučná funkcia $F(x_b)$ náhodnej veličiny X je funkcia, ktorá udáva pre každé x_b pravdepodobnosť $P(x < x_b)$.

Distribučnú funkciu diskrétnej náhodnej veličiny vypočítame:

$$F(x) = P(X < x) = \sum_{x_i : x_i < x} p_i$$

$(p_i = P(x = x_i),\ -\infty < x < +\infty)$.

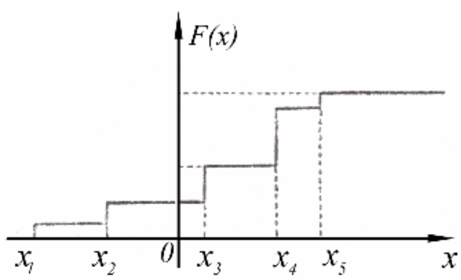
Distribučnú funkciu spojitej náhodnej veličiny vypočítame:

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

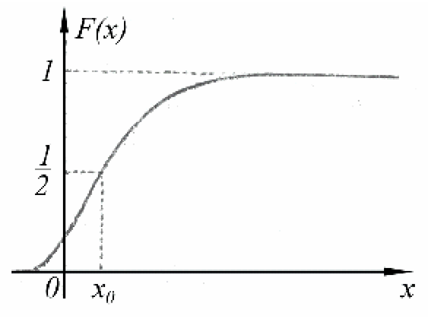
Potom pre spojité náhodné veličiny platí:

$$f(x) = F'(x) = \frac{dF(x)}{dx}\,.$$

Distribučné funkcie náhodných veličín sú znázornené na nasledovných obrázkoch:



Obrázok: Distribučná funkcia diskrétnej náhodnej premennej, ktorá nadobúda hodnoty $x_1, ..., x_5$



Obrázok: Distribučná funkcia spojitej náhodnej premennej

Charakteristikou polohy rozdelenia náhodnej veličiny je jej ***Stredná hodnota E(X)***, ktorá je definovaná pre nespojité (diskrétne) náhodné veličiny nasledovne:

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

Stredná hodnota spojitej náhodnej veličiny x je definovaná:

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

kde $f(x)$ je hustota pravdepodobnosti.

Stredná hodnota vyjadruje priemernú veľkosť (centrálnu tendenciu, stred, ťažisko, polohu) hodnôt príslušnej náhodnej veličiny a značíme ju tiež $M(x)$, alebo μ. Pre všetky typy náhodných veličín možno strednú hodnotu vyjadriť jednotne, Stieltjesovým integrálom:

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

kde $F(x)$ je distribučná funkcia.

Rozptyl (disperzia, variancia) D(x) - predstavuje priemerný štvorec odchýlky od priemeru**.** Značíme ho tiež $D^2(x)$, $V(x)$, alebo σ_x^2 a pre diskrétnu, resp. spojitú náhodnú veličinu je definovaný vzťahmi:

$$D(X) = \sum_{i=1}^{n} (X_i - E(X))^2 \cdot p_i = \sum_{i=1}^{n} X_i^2 \cdot p_i - (E(X))^2\,,$$

kde $E(X)$ je stredná hodnota náhodnej veličiny X, resp. pre spojitú:

$$D(X) = \int_{-\infty}^{\infty} (X_i - E(X))^2 \cdot f(x)\ \mathrm{d}x = \int_{-\infty}^{\infty} X_i^2 \cdot f(x)\ \mathrm{d}x - (E(X))^2.$$

Smerodajná odchýlka **(štandardná, stredná kvadratická odchýlka)** je druhou odmocninou rozptylu, je charakteristikou variability náhodnej veličiny a označujeme ju symbolom σ(*X*). Je odmocninou rozptylu, ktorý označujeme D(*X*) a je definovaná vzťahom:

$$\sigma(X) = \sqrt{D(X)}\,.$$

Variačný koeficient vypočítame zo vzťahu:

$$C(X) = \frac{\sqrt{D(X)}}{E(X)}\,.$$

Šikmosť vypočítame ako:

$$\gamma_3 = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(X_i - E(X))^3}{\sigma^3}\,.$$

Špicatosť **(štatistický exces)** vypočítame ako:

$$\gamma_4 = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(X_i - E(X))^4}{\sigma^4} - 3\,.$$

Medián $\tilde{X}$ je hodnota, pre ktorú platí P(x < $\tilde{X}$) = *0,5*. Je to hodnota prostredného člena z radu členov súboru usporiadaných podľa veľkosti. Ak je počet členov párny, je to priemer dvojice prostredných členov.

P-kvantil $\boldsymbol{x_p}$ je hodnota, pre ktorú P($x < x_p$) = *P*.

Modus diskrétnej, resp. spojitej náhodnej veličiny je hodnota s maximálnou pravdepodobnosťou, resp. hustotou pravdepodobnosti. Modus je tá hodnota nezávislej premennej *X*, ktorá sa v súbore vyskytuje najčastejšie.

Pre lineárnu funkciu $ax + b$ náhodnej veličiny x sú stredná hodnota a rozptyl vyjadrené nasledovnými vzťahmi:

$$E(ax+b) = aE(x) + b\,,$$

$$D(ax+b) = a^2 D(x)\,.$$

Strednú hodnotu a rozptyl všeobecnej funkcie $g(x)$ približne vypočítame podľa vzťahov:

$$E(g(x)) \approx g(E(x)) \qquad D(g(x)) = \left[g'(E(x))\right]^2 D(x).$$

kde $g' = \mathrm{d}g/\mathrm{d}x$, pričom musí platiť podmienka, že $g'(x)$ je v okolí $E(x)$ približne konštantné, povedzme pre $|x - E(x)| \prec 3 \cdot \sqrt{D(x)}$.

Normálne rozdelenie

Normovaným normálnym rozdelením nazývame normálne rozdelenie, ktoré má strednú hodnotu rovnú 0 a rozptyl 1.

Hustota normálneho rozdelenia $\varphi(x)$ a distribučná funkcia $\phi(x)$ sú vyjadrené vzťahmi:

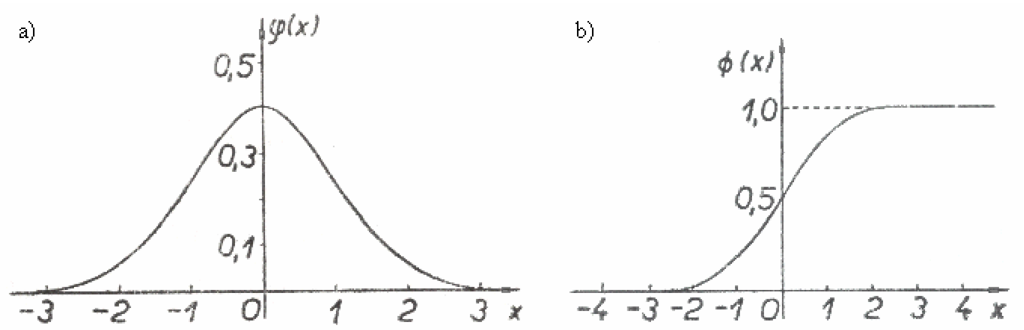
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}},$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, \mathrm{d}t.$$

Pričom platia vzťahy $\varphi(-x) = \varphi(x)$ a $\Phi(-x) = 1 - \Phi(x)$.

Význam normálneho rozdelenia spočíva v tom, že ním možno aproximovať množstvo iných rozdelení a že približne normálne rozdelenie majú mnohé náhodné veličiny, ktorých hodnota je súhrnom účinkov množstva nezávislých činiteľov, ktoré (každý sám o sebe) majú iba nepatrný vplyv.

Normálne rozdelenie sa tiež nazýva *Gaussovým rozdelením* a jeho hustota *Gaussovou krivkou* alebo *Gaussovým zákonom chýb*. Grafy hustoty $\varphi(x)$ a distribučnej funkcie $\Phi(x)$ normovaného normálneho rozdelenia sú znázornené na nasledujúcom obrázku:



Obrázok: Grafické znázornenie a) hustoty pravdepodobnosti, b) distribučnej funkcie normovaného normálneho rozdelenia

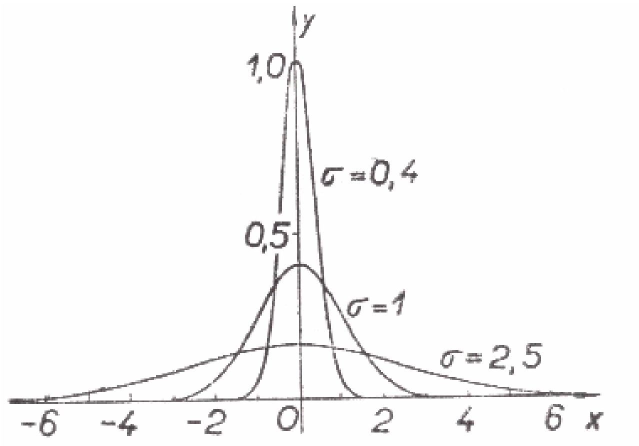
Všeobecným normálnym rozdelením (rozdelením $N(\mu, \sigma^2)$) nazývame rozdelenie so strednou hodnotou μ a rozptylom σ^2, pričom hustota $f(x)$ sa rovná:

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

a distribučná funkcia $F(x)$ je vyjadrená vzťahom:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, \mathrm{d}t,$$

pričom grafické znázornenie hustoty normálneho rozdelenia pre σ = 0,4; 1; 2,5 a μ = 0 je na nasledujúcom obrázku:



Obrázok: Grafické znázornenie normálnych rozdelení pre niektoré prípady σ

Jednotlivé pravdepodobnosti javov určených náhodnou veličinou x s rozdelením $N(\mu, \sigma^2)$ vypočítame podľa vzťahov:

$$P(a \le x \le b) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{a}^{b} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

$$P(|x-\mu| \le a) = 2\Phi\left(\frac{a}{\sigma}\right) - 1, \quad P(|x-\mu| > a) = 2\left[1 - \Phi\left(\frac{a}{\sigma}\right)\right].$$

Lineárna funkcia $ax + b$ náhodnej veličiny x s normálnym rozdelením $N(\mu, \sigma^2)$ má normálne rozdelenie $N(a\mu + b, a^2\sigma^2)$. Normovaná náhodná veličina u, ktorá je vyjadrená vzťahom: $u = \frac{x-\mu}{\sigma}$ má normované normálne rozdelenie $N(0,1)$.

Príklady diskrétnych rozdelení

Alternatívne rozdelenie

Najjednoduchším typom diskrétnych rozdelení je tzv. *alternatívne rozdelenie*. Je to rozdelenie náhodnej veličiny, ktorá môže nadobúdať len 2 hodnoty: hodnoty 1 s pravdepodobnosťou p a hodnoty 0 s pravdepodobnosťou $q=1-p$. Uvedenú náhodnú veličinu nazývame *nula-jedničková*. Ľahko je možné vypočítať jej strednú hodnotu a rozptyl:

$$E(X) = 1\,.\,p + 0\,(1-p) = p \quad a \quad D(X) = (1-p)^2\,.\,p + p^2.(1-p) = p\,(1-p)$$

Nula-jedničkové náhodné veličiny je možné použiť pre popis výskytu určitého náhodného javu. Ak náhodný jav nastal, nadobúda táto veličina hodnotu 1, ak nenastal, tak nadobúda hodnoty 0. (Číslo p je jediným parametrom tohto rozdelenia).

Binomické rozdelenie

Predpokladajme, že určitý pokus opakujeme n-krát za rovnakých podmienok. V každom pokuse môže nastať náhodný jav A s rovnakou pravdepodobnosťou p a nenastať s pravdepodobnosťou $q=1-p$. Uvedená schéma pokusov sa nazýva *Bernoulliho schéma.* Počet realizácií náhodného javu A v Bernoulliho schéme n nezávislých pokusov je zrejme diskrétnou náhodnou veličinou X, ktorá môže nadobúdať hodnoty $0, 1, \dots, n$. Vzhľadom k nezávislosti pokusov pre jej pravdepodobnostnú funkciu platí:

$$P_n(k) = P(X=k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \dots, n$$

Rozdelenie, ktoré je popísané touto pravdepodobnostnou funkciou sa nazýva *binomické rozdelenie* s parametrami n a p a označujeme ho $Bi(n, p)$. (Parametre sú veličiny, ktorých hodnoty musíme poznať, aby sme ľubovoľnému x mohli pomocou $P_n(k)$ priradiť jeho pravdepodobnosť).

Pri výpočte strednej hodnoty a rozptylu náhodnej veličiny s binomickým rozdelením pravdepodobnosti je možné s výhodou použiť rozpis binomickej veličiny X v tvare súčtu nula-jedničkových náhodných veličín. Ak vezmeme do úvahy vlastnosti Bernoulliho schémy, tak platí:

$$X = Y_1 + Y_2 + \dots + Y_n$$

kde $Y_1, Y_2 \dots Y_n$ sú nezávislé náhodné veličiny s týmto alternatívnym rozdelením. Použitím vlastnosti strednej hodnoty a rozptylu náhodných veličín tak dostávame:

$$E(X) = E(Y_1) + E(Y_2) + \dots + E(Y_n) = p + p + \dots + p = np$$

$$D(X) = D(Y_1) + D(Y_2) + \dots + D(Y_n) = p\,(1-p) + \dots + p\,(1-p) = np\,(1-p) = npq$$

Príklad:
Pravdepodobnosť vypestovania zdravej rastliny zo semena je 0,3. Aká je pravdepodobnosť, že vypestujeme aspoň 3 a najviac 5 zdravých rastlín, ak zasadíme 10 náhodne vybraných semien? Aký bude priemerný počet vypestovaných zdravých rastlín?

Riešenie:
Nech X je náhodná veličina, ktorá predstavuje počet vypestovaných zdravých rastlín. Zo zadania úlohy vyplýva, že táto veličina má rozdelenie *Bi*(10, 0,3). Úlohou je stanoviť pravdepodobnosť $P(3 \le X \le 5)$.

$$\mathrm{P}(3 \le X \le 5) = \mathrm{P}(X = 3) + \mathrm{P}(X = 4) + \mathrm{P}(X = 5) =$$

$$= \binom{10}{3} 0{,}3^3 . 0{,}7^7 + \binom{10}{4} 0{,}3^4 . 0{,}7^6 + \binom{10}{5} 0{,}3^5 . 0{,}7^5 = 0{,}57$$

Priemerný počet vypestovaných rastlín určíme zo vzťahu $E(X) = np = 10 \,.\, 0{,}3 = 3$.

Poissonovo rozdelenie

Ak $n \to \infty$ a $p \to 0$ a $np = \lambda$, kde λ je kladná konštanta, tak takýmto spôsobom dostaneme limitný prípad binomického rozdelenia, ktoré nazývame *rozdelenie Poissonovo* s parametrom λ. Budeme ho označovať $Po(\lambda)$. Pravdepodobnostná funkcia tohto rozdelenia má tvar:

$$P_k = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad , \qquad k = 0, 1, 2, \ldots$$

Rozdelenie $Po(\lambda)$ je charakteristické tým, že $E(X) = D(X) = \lambda$. Poissonovo rozdelenie s parametrom $\lambda = np$ býva vhodnou aproximáciou binomického rozdelenia $Bi(n, p)$, ak počet kusov n je veľký (prakticky $n > 30$) a pravdepodobnosť p je malá (prakticky $p < 0{,}1$). Pre Poissonovo rozdelenie sa niekedy používa názov *zákon vzácnych javov* (*zákon riedkych javov*). Popisuje totiž správanie sa tzv. *vzácnych javov*, ktoré majú malú pravdepodobnosť výskytu, takže aj v rozsiahlych súboroch pozorovaní sa vyskytujú zriedka.

Hypergeometrické rozdelenie

Veľa dôležitých úloh môže byť popísaných podľa tejto schémy:

V súbore N prvkov ich má M určitú vlastnosť A. Zo súboru náhodne vyberieme n prvkov, bez toho, aby sme ich po vybratí vracali naspäť do pôvodného súboru (tzv. *výber bez vrátenia*). Počet prvkov s vlastnosťou A, ktoré boli vybrané do uvedeného výberu n prvkov, je zrejme náhodná veličina X, ktorá môže nadobúdať hodnoty $k = \max(0, n + M - N), \ldots, \min(M, n)$ s pravdepodobnosťami

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

Rozdelenie náhodnej veličiny *X*, popísané touto pravdepodobnostnou funkciou sa nazýva *hypergeometrické rozdelenie.*

Je možné ukázať, že stredná hodnota a rozptyl hypergeometrického rozdelenia sú

$$E(X) = np \qquad a \qquad D(X) = npq\frac{N-n}{N-1}.$$

kde $p = \frac{M}{N}$ a $q=1 - p$. Porovnaním s binomickým rozdelením zistíme, že stredné hodnoty sú pri oboch rozdeleniach rovnaké, rozptyl hypergeometrického rozdelenia je menší, alebo tzv. *konečnostný faktor* $\frac{N-n}{N-1}$ je menší než 1.

Ak je rozsah výberu n veľmi malý vzhľadom k rozsahu základného súboru N, je možné hypergeometrické rozdelenie úspešne nahradiť binomickým rozdelením. Výhodná a v praxi často používaná je tiež aproximácia rozdelením Poissonovým s parametrom $\lambda = n\frac{M}{N}$. Náhrada je kvalitná už pri $\frac{M}{N} < 0{,}1$ a $\frac{n}{N} < 0{,}1$.

Príklady spojitých rozdelení

Normálne rozdelenie

Kľúčové postavenie v štatistickej teórii aj aplikáciách má *normálne (Gaussovo) rozdelenie*. Je najfrekventovanejším rozdelením spojitých náhodných veličín. V zásade sa dá povedať, že je adekvátnym pravdepodobnostným modelom takých náhodných veličín, ktoré sú súčtom veľkého počtu nezávislých alebo len slabo závislých veličín (zložiek), pričom príspevky jednotlivých zložiek sú nepatrné.

Význam normálneho rozdelenia je podčiarknutý okolnosťou, že za veľmi širokých podmienok im je možné aproximovať veľa iných rozdelení, dokonca aj rozdelenie diskrétnych náhodných veličín.

Hustota pravdepodobnosti normálneho rozdelenia je daná výrazom

$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{(x-\mu)}{2\sigma}\right)^2}, \quad -\infty < x < +\infty ,$$

kde konštanty $\mu \in (-\infty, \infty)$ a $\sigma^2 > 0$ sú parametre tohto rozdelenia. (Rozdelenie dané touto hustotou býva často skrátene označované $N(\mu, \sigma^2)$. Parameter μ sa rovná strednej hodnote a parameter σ^2 rozptylu normálne rozdelenej náhodnej veličiny X: $E(X) = \mu$ a $D(X) = \sigma^2$.

Grafom hustoty normálneho rozdelenia je tzv. *Gaussova krivka.* Jedná sa o zvonovitú symetrickú krivku, ktorá nadobúda maximá v bode $x = \mu$ a pre $x \to \pm\infty$ sa asymptoticky približuje k osi úsečiek. Pre $x = \mu \pm \sigma$ má Gaussova krivka inflexné body. Zmenou strednej hodnoty μ pri konštantnej veľkosti rozptylu σ^2 sa posúva Gaussova krivka pozdĺž osy x, bez toho, aby sa menil tvar krivky. Ak sa mení parameter σ^2 a μ = konšt., dochádza k zmene tvaru normálnej krivky. Ak $\mu = 0$ a $\sigma^2 = 1$, hovoríme o *normovanom normálnom rozdelení*, ktoré označujeme $N(0, 1)$. Hustota normovaného normálneho rozdelenia, tzv. *normovaná normálna hustota,* má tvar:

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

Tejto hustote pravdepodobnosti zodpovedá distribučná funkcia

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, \mathrm{d}t ,$$

ktorú nazývame *normovaná normálna distribučná funkcia.*

Pre normálne rozdelenie sú charakteristické tieto vlastnosti:

1. Ak náhodná veličina X má rozdelenie $N(\mu,\sigma^2)$, má ľubovoľná lineárna funkcia $Y = aX + b$ normálne rozdelenie s parametrami $a\mu + b$ a $a^2\sigma^2$. Špeciálne to znamená, že normovaná veličina $U = \dfrac{X-\mu}{\sigma}$ má rozdelenie $N(0, 1)$.
2. Ľubovoľná lineárna kombinácia n nezávislých náhodných veličín, ktoré majú všetky normálne rozdelenia, má opäť normálne rozdelenie. Odtiaľ vyplýva, že súčet, rozdiel a priemer nezávislých normálne rozdelených náhodných veličín sú náhodné veličiny, ktoré majú opäť normálne rozdelenie.

Rozdelenie χ^2, t, F

S normálnym rozdelením sú úzko spojené tri dôležité rozdelenia spojitých náhodných veličín:
χ^2 - rozdelenie, Studentovo t-rozdelenie a F-rozdelenie. Majú mimoriadny význam pri analýze štatistických dát, získaných náhodných výberom.

a) χ^2 **- rozdelenie**

Nech $U_1, U_2, \ldots U_v$ sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$.
Rozdelenie náhodnej veličiny

$$\chi^2 = U_1^2 + U_2^2 + \ldots + U_v^2$$

sa nazýva χ^2 - rozdelenie pri k stupňoch voľnosti a označuje sa $\chi^2\,(k)$.

b) **Studentovo t-rozdelenie**

Nech U a V sú nezávislé náhodné veličiny, z ktorých U má rozdelenie $N(0, 1)$ a V má rozdelenie $\chi^2\,(k)$. O náhodnej veličine t, definovanej vzťahom

$$t = \frac{U}{\sqrt{\frac{V}{k}}}$$

budeme hovoriť, že má *Studentovo t-rozdelenie* o k stupňoch voľnosti.

c) **Fisherovo – Snedecorovo rozdelenie (F-rozdelenie)**

Rozdelenie náhodnej veličiny

$$F = \frac{V_1}{n_1} : \frac{V_2}{n_2}$$

kde V_1 resp. V_2 sú nezávislé náhodné veličiny, ktoré majú rozdelenie $\chi^2(n_1)$ resp. $\chi^2(n_2)$, sa nazýva *Fisherovo – Snedecorovo rozdelenie* (alebo len stručne *F*-rozdelenie) o n_1 a n_2 stupňoch voľnosti a označuje sa $F(n_1, n_2)$. Používa sa hlavne v analýze rozptylu.

Momenty

Pre opis základných vlastností súboru sa používa pojem momentu, podobne ako sa v mechanike používa moment inercie na opísanie rozloženia hmoty. V praxi sa najčastejšie používajú momenty dvoch typov: počiatočné a centrálne.

Počiatočným momentom k-tého rádu diskrétnej veličiny X sa nazýva suma typu

$$\alpha_k[X] = \sum x_i . p_i$$

Ako sme uviedli, aritmetický priemer možno vypočítať podľa vzťahu

$$\bar{x} = \sum_{i=1}^{n} x_{i.} . p_i$$

Ak poznáme oba vzťahy a porovnávame ich vidíme, že aritmetický priemer je vlastne počiatočný moment k-tého rádu náhodnej veličiny X a preto sa nazýva aritmetický priemer k-tého stupňa tejto veličiny.

Všeobecný moment k-tého radu

$$\mu_k' = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

Centrálny moment. Najprv si osvetlíme pojem centrovanej náhodnej veličiny. Centrovanou
náhodnou veličinou budeme nazývať odchýlku náhodnej veličiny X, od jej aritmetického priemeru

$$\dot{X} = x - \bar{x}$$

kde $\dot{X}$ je centrovaná náhodná veličina.
Aritmetický priemer centrovanej náhodnej veličiny je rovný nule

$$\bar{X}\left[\dot{X}\right] = \bar{X}\left[x - \bar{x}\right] = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)p_i = \sum_{i=1}^{n} x_i . p_i - \bar{x}\sum_{i=1}^{n} p_i = \bar{x} - \bar{x} = 0$$

Nezabudnime, že $\sum_{i=1}^{n} pi = 1 \quad \sum_{i=1}^{n} x_i . p_i = \bar{x}$

Centrovanie náhodnej veličiny odpovedá prenosu počiatku súradníc do bodu, ktorý má hodnotu aritmetického priemeru. Momenty centrovanej náhodnej veličiny sa nazývajú **centrálne momenty**. Sú analogické pojmu ťažisko v mechanike. Centrálnym momentom k-tého rádu náhodnej veličiny X sa nazýva aritmetický priemer k-tého stupňa tejto centrovanej náhodnej veličiny

$$\mu_k = [X] = \bar{x}\left[\dot{X}^k\right] = X\left[\left(x - \bar{x}\right)^k\right]$$

Pre diskrétnu veličinu sa k-tý centrálny moment vyjadruje sumou

$$\mu_s = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k p_i$$

Ako sme ukázali, prvý centrálny moment je rovný nule

$$\mu = 0$$

Centrálny moment k-tého radu sa vypočíta

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k$$

Vypočítame si druhý centrálny moment

$$\mu_2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 p_i = \sum_{i=1}^{n} x_i^{\,2}.p_i - 2\bar{x}\sum_{i=1}^{n} x_i.p_i + \bar{x}^2\sum_{i=1}^{n} p_i = \alpha_2 - 2\bar{x}^2 + \bar{x}^2 = \alpha_2 - \bar{x}^2$$

Analogicky vypočítame tretí centrálny moment

$$\mu_3 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3 p_i = \sum_{i=1}^{n} x_i^{\,3}.p_i - 3\bar{x}\sum_{i=1}^{n} x_i^{\,2}.p_i + 3\bar{x}^2\sum_{i=1}^{n} x_i.p_i - \bar{x}^3\sum_{i=1}^{n} p_i = \alpha_3 - 3\alpha_2\bar{x} - 2\bar{x}^2$$

Druhý centrálny moment sa nazýva rozptyl (disperzia) náhodnej veličiny

$$\mu_2 = s^2$$

Rozptyl (variabilita, variancia, disperzia) – s^2 je daný vzorcom:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2, kde\ \ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ , \qquad \text{(a)}$$

resp. $s^2 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 f_i, kde\ \ \bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i.f_i$

a pre výberový súbor

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\ \ resp.\ s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 f_i \qquad \text{(b)}$$

Pri výpočte rozptylu používame vzorce (a) vtedy, ak je štatistický súbor malý, a ak poznáme aritmetický priemer. Môžeme použiť aj vzorec odvodený z (1), jeho výhodou je, že nemusíme počítať odchýlky $x_i - \bar{x}$, platí

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - \bar{x}^2 \qquad \text{resp.} \qquad s^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^{\,2}.f_i - \bar{x}^2 ,$$

pre výberový súbor

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right] \text{,resp. } s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2.f_i - n\bar{x}^2\right] \text{,kde } \bar{x}^2 = \left(\frac{1}{n}\sum_{i=1}^{k} x_i.f_i\right)^2$$

Ak sú dané iba hodnoty x_i, prípadne aj ich početnosti f_i, je výhodnejšie použiť vzťahy, ktoré sú odvodené z predchádzajúcich vzorcov

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n} , \qquad \text{resp.} \qquad s^2 = \frac{\sum_{i=1}^{k} x_i^2.f_i - \frac{\left(\sum_{i=1}^{k} x_i.f_i\right)^2}{n}}{n}$$

pre výberový súbor

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i.f_i\right)^2}{n}\right]$$

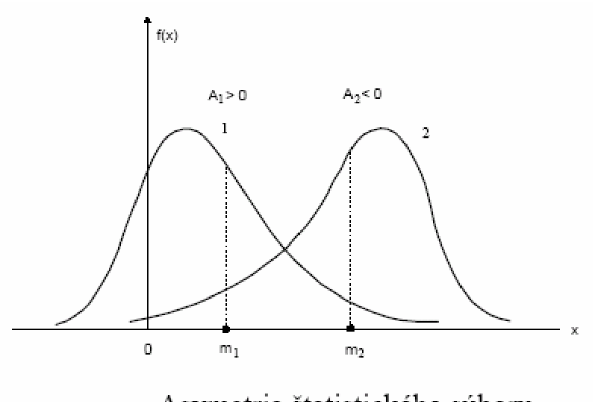
Rozptyl má tú nevýhodu, že nemá rozmer hodnôt znakov. Tenko nedostatok nemá ***smerodajná*** (štandardná) ***odchýlka*** – s

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} \qquad \text{alebo} \qquad s = \sqrt{\frac{1}{n}\sum^{k} x_i^2 - \bar{x}^2}$$

Hodnotu smerodajnej odchýlky dostaneme, ak príslušný tvar rozptylu odmocníme.

Ak použijeme analógiu s mechanikou, tak disperzia je vlastne moment inercie – udáva rozdelenie hmoty okolo ťažiska. **Tretí centrálny moment** sa používa na charakteristiku symetrie resp. asymetrie súboru (vlastne každý nepárny moment sa môže použiť na charakteristiku asymetrie). Aby sme dostali bez rozmernú veličinu, delíme tretí centrálny moment treťou mocninou strednej kvadratickej odchýlky. Získanú hodnotu nazývame asymetria (A) alebo koeficient šikmosti (kš)

$$A = \frac{\mu_3}{s^3} \quad \text{alebo} \quad kš = \frac{3\left(\bar{x} - \hat{x}\right)}{s}$$



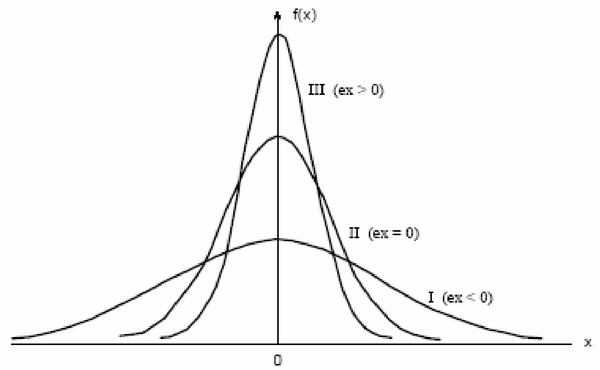
Asymetria štatistického súboru

Ak má hodnotu 0, môžeme konštatovať, že rozdelenie hodnôt je symetrické, ak koeficient šikmosti je kladný, hovoríme o zošikmení doľava (vyskytujú sa častejšie menšie hodnoty), ak je koeficient šikmosti záporný hovoríme o zošikmení doprava (vyskytujú sa častejšie väčšie hodnoty)

Štvrtý centrálny moment sa používa na charakteristiku štíhlosti (špicatosti) súboru. Získanú veličinu nazývame exces:

$$E = \frac{\mu_4}{s^4} - 3$$

Ak ma hodnotu 0, môžeme konštatovať že rozdelenie je rovnako špicaté ako normálne, ak je > 0 rozdelenie je "špicatejšie" ako normálne a ak je < 0 hovoríme o "plochšom" rozdelení ako je normálne.



Exces štatistického súboru

Príklad:

Počet detí do 18 rokov v 40 náhodne vybraných rodinách je uvedený v tabuľke

x_i	0	1	2	3	4	5	6	7
f_i	5	11	8	5	3	3	4	1

Určte!

a) rozpätie
b) priemernú odchýlku
c) rozptyl, smerodajnú odchýlku
d) kvartilové rozpätie
e) viariačný koeficient
f) koeficient šikmosti

Riešenie

a) Rozpätie R = $x_{max} - x_{min} = 7 - 0 = 7$
b) Vytvoríme tabuľku pre ďalšie výpočty

X_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert f$	$x_i^2 f_i$	kf_i
0	5	0	-2,5	12,5	0	5
1	11	11	-1,5	16,5	11	16
2	8	16	-0,5	4	32	24
3	5	15	0,5	2,5	45	29
4	3	12	1,5	4,5	48	32
5	3	15	2,5	7,5	75	35
6	4	24	3,5	14	144	39
7	1	7	4,5	4,5	49	40
	$\sum = 40$	$\sum = 100$		$\sum = 66$	$\sum = 404$	

Priemerná odchýlka $d = \frac{1}{n}\sum_{i=1}^{k}\left|x_i - \bar{x}\right| f_i = \frac{1}{40}\sum_{i=0}^{7}\left|x_i - 2{,}5\right| f_i = \frac{1}{40}.66 = 1{,}5, kde$

$$\bar{x} = \frac{1}{n}\sum_{i=0}^{7} x_i f_i = \frac{1}{40}.100 = 2{,}5$$

c) *Pretože je to výberový súbor, na výpočet variability použijeme vzorec*

$$s^2 = \frac{\sum_{i=1}^{k} x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^{k} x_i \cdot f_i\right)^2}{n}}{n-1} = \frac{\left(404 - \frac{100^2}{40}\right)}{39} = \frac{1}{39}(404 - 250) = 3{,}95$$

Smerodajná odchýlka $s = \sqrt{s^2} = \sqrt{3{,}95} = 1{,}99$

d) *Kvartilové rozpätie*

$$Q = Q_{h-}Q_d$$

$$Q_h = L + \frac{\frac{3n}{4} - \sum_{i=1}^{m-1} f_i}{f_m}.h = 3{,}5 + \frac{30-29}{3}.1 = 3{,}83$$

$$Q_d = L + \frac{\frac{n}{4} - \sum_{i=1}^{m-1} f_i}{f_m}.h = 0{,}5 + \frac{10-5}{11}.1 = 0{,}95$$

$$Q = 3{,}83 - 0{,}95 = 2{,}88$$

e*) Variačný koeficient*

$$v = \frac{s}{\bar{x}} = \frac{0{,}88}{2{,}9} = 0{,}3$$

f) Koeficient šikmosti

$$kš = \frac{3\left(\bar{x} - \hat{x}\right)}{s} = \frac{3.(2{,}5-2)}{2} = 0{,}75, kde$$

$$\hat{x} = L + \frac{\frac{n}{2} - \sum_{i=1}^{m-1} f_i}{f_m}.h = 1{,}5 + \frac{20-16}{8}.1 = 1{,}5 + 0{,}5 = 2$$

Príklad 1 :

Hádžeme dvoma pravidelnými diskami, ktorých strany sú očíslované (1 a 2). Sledujeme podiel čísla na prvom k číslu na druhom disku. Nájdite rozdelenie pravdepodobnosti, strednú hodnotu a disperziu tejto náhodnej premennej. [2]

Riešenie:

x_i	1/2	1/1	2/1
p_i	1/4	2/4	1/4

E(x)=SUM(pi*xi)
D(x)=SUM(pi*xi^2)

F(0)=P(X ≤ 0)=0
F(1/2)=P(X ≤ 1/2)=1/4
F(3/4)=P(X ≤ 3/4)=1/4
F(1)=P(X ≤ 1)=1/4+2/4=3/4
F(2)=P(X ≤ 2)=1/4+1/4+2/4=1

MATLAB:

```
x=[1/2, 1 , 2 ];
p=[1/4, 1/2 , 1/4];
stem(x,p,'r')   -----   'r' - farba
hold
plot(x,p)
```

Príklad 2:

Pri kontrole automatickej prevádzky na výrobu 40% rumu boli zistené tieto výsledky (v percentách, bolo odobrených 20 vzorkou):

40.8   40 .9   40.2   39.8   40.1   40.2   40.3   40.6   40.9   39.1

39.5   39.8   39.9   39.6   40.2   40.7   39.2   40.1   40.2   38.9

Rozhodnite, či je priemerný obsah alkoholu vo vyrábanom rumu väčší alebo menší než deklarovaných 40 % a odhadnite pravdepodobnosť, že náhodne odoberaná vzorka bude mať obsah alkoholu v rozmedzí 39% až 40%.

Riešenie:

Funkcia MAX nájde najväčší a funkcia MIN najmenšiu hodnotu zo zadaných dát.

Funkcia HISTC sa použije v príkaze N=HISTC(X, A), kde X je pole dát a A je vektor krajných bodov triedy. Výstup N je vektor o toľko súradníc, koľko máme tried a jeho súradnice sú počty prvkov v danej triede.

Funkcia BAR vytvorí obrázok histogramu príkazom BAR(X,A, 'HISTC').

Teda: Vložme dáta

X=[40.8; 40.9; 40.2; 39.8; 40.1; 40.2; 40.3; 40.6; 40.9 ;39.1; 39.5; 39.8; 39.9; 39.6; 40.2; 40.7; 39.2; 40.1; 40.2; 38.9];

Vypočítame najväčší a najmenšiu hodnotu

a=max(X) b=min(X)

a = 40.9000   b = 38.9000

Vypočítame dĺžku triedy c=(a-b)/4

c = 0.5000

Teraz vytvoríme krajné body jednotlivých tried. Pretože funkcia HISTC nepracuje úplne tak, ako by sme potrebovali, musíme trochu zmeniť dĺžku triedy - položme c=0.501 a zväčšíme také o trochu maximálnu hodnotu na 41.3.

A=[38.9:0.501:41.3]

A = 38.9000   39.4010   39.9020   40.4030   40.9040

Zavolajme funkciu HISCT   N=histc(X,A)

N =   3   5   7   5   0

Posledná nula znamená počet hodnôt, ktoré sa rovnajú poslednej zložke vektoru A, a to nie je žiadna, práve preto sme trochu modifikovali dĺžku triedy c.

Nakoniec nakreslíme obrázok   BAR(A,N,'histc')

Teraz k druhej časti príkladu: nájdeme výberovú strednú hodnotu (funkcie MEAN) a výberový rozptyl (pomocou funkcie VAR) náhodnej veličiny, je náhodný výber sme urobili.

mean(X)

ans =   40.0500

var(X)

ans =   0.3416

Ak je potom F(x) distribučná funkcia našou náhodnou veličinou, je hľadaná pravdepodobnosť

$$P(0.39 < X < 40) = F(40) – F(39).$$

Hodnoty funkcie F(x) dáva funkcia NORMCDF(X, MU, SIGMA), kde X je bod, v ktorej hľadáme hodnotu, MU je stredná hodnota a SIGMA rozptyl náhodnej veličiny.

Teda: p = normcdf(40,40.05,0.3416) – normcdf(39,40.05,0.3416)

p = 0.4408.