12. Náhodný výběr

Při sledování a studiu vlastností náhodných výsledků poznáme charakter rozdělení z toho, že opakovaný náhodný pokus nám dává za stejných podmínek různé výsledky. Ty odpovídají hodnotám jednotlivých realizací náhodné veličiny, která popisuje příslušný náhodný proces. Základním pojmem statistiky se tak stává pojem náhodného výběru, který je modelem popsané situace.

12.1. Definice: Náhodný výběr je uspořádaná n-tice (náhodný vektor) $(X_1, X_2, \ldots, X_n)$ náhodných veličin $X_i,\ 1 \leq i \leq n$, které jsou nezávislé a mají stejné rozdělení.

Poznámka: Je-li F distribuční funkce popisující rozdělení náhodných veličin X_i, pak sdružená distribuční funkce náhodného výběru je rovna $F(x_1).F(x_2)\ldots F(x_n)$. Obdobně pro sdruženou hustotu či pravděpodobnostní funkci dostaneme vyjádření ve tvaru
$f(x_1).f(x_2)\ldots f(x_n)$, resp. $p(x_1).p(x_2)\ldots p(x_n)$ je-li f hustota resp. p pravděpodobnostní funkce náhodné veličiny X_i.

Studium vlastností rozdělení obvykle provádíme pomocí vhodně zvolené funkce náhodného výběru (statistiky). Uvedeme ty, které nejčastěji používáme.

12.2. Definice: Výběrový průměr, výběrový rozptyl. Jsou-li $(X_1, X_2, \ldots, X_n)$ a $(Y_1, Y_2, \ldots, Y_n)$ náhodné výběry, pak označujeme a nazýváme statistiku:

$\widetilde{X} = \sum\limits_{i=1}^{n} X_i$ - *výběrovým úhrnem;*

$\overline{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ - *výběrovým průměrem;*

$S_X^2 = S^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \overline{X})^2$ - *výběrovým rozptylem;*

$s^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \overline{X})^2$; - *střední kvadratickou odchylkou;*

$S_{XY} = \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$ - *výběrovým koeficientem kovariance;*

$s_{XY} = \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$

$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{s_{XY}}{s_X s_Y}$ - *výběrovým koeficientem korelace.*

Poznámka: Pro vyčíslení výběrových charakteristik používáme někdy jiného vyjádření. Je totiž
$(n-1)S^2 = ns^2 = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} \left[X_i^2 - 2X_i\overline{X} + \left(\overline{X}\right)^2\right] =$
$= \sum_{i=1}^{n} X_i^2 - n\left(\overline{X}\right)^2.$

Je tedy

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\left(\overline{X}\right)^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right]$$

a

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2.$$

Je také

$$(n-1)S_{XY} = \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_iY_i) - n(\overline{XY}) = ns_{XY},$$

$$r_{XY} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_iY_i - (\overline{XY})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - (\overline{Y})^2\right)}}$$

Pro uvedené statistiky platí několik tvrzení, která si postupně uvedeme.

12.3.Věta: Vlastnosti výběrového průměru a úhrnu. Je-li $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení, kde $E(X_i) = \mu$ a $D(X_i) = \sigma^2$, pak:
$E(\overline{X}) = \mu,\ E(\widetilde{X}) = n\mu$ a $D(\overline{X}) = \frac{\sigma^2}{n},\ D(\widetilde{X}) = n\sigma^2.$
Důkaz: $E(\widetilde{X}) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = n\mu,$
$E(\overline{X}) = E(\frac{1}{n}\widetilde{X}) = \frac{1}{n}E(\widetilde{X}) = \mu.$

Pro nezávislé náhodné veličiny ve výběru je dostaneme:
$D(\widetilde{X}) = D(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} D(X_i) = n\sigma^2$,
$D(\overline{X}) = D(\frac{1}{n}\widetilde{X}) = \frac{1}{n^2}D(\widetilde{X}) = \frac{\sigma^2}{n}$.

12.4. Věta: Vlastnosti výběrového rozptylu. Je-li $(X_1, X_2, \ldots, X_n)$ náhodný výběr z rozdělení, kde $E(X_i) = \mu$ a $D(X_i) = \sigma^2$, pak platí:
$E(S^2) = \sigma^2,\ E(s^2) = \dfrac{n}{n-1}\sigma^2$ a $D(S^2) = \dfrac{1}{n}E(X_i^4) - \dfrac{n-3}{n(n-1)}\sigma^4, n \geq 3$.

Důkaz: Úpravou postupně dostaneme
$\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \sum\limits_{i=1}^{n} (X_i - \mu)^2 - n\left(\overline{X} - \mu\right)^2$.

Odtud dostaneme, že
$E\left(\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2\right) = \sum\limits_{i=1}^{n} E\left((X_i - \mu)^2\right) - nE\left(\left(\overline{X} - \mu\right)^2\right) =$
$= (n-1)\sigma^2$,

když jsme použili postupně skutečnosti
$E(X_i) = E(\overline{X}) = \mu,\ E\left((X_i - \mu)^2\right) = D(X_i) = \sigma^2$ a
$E\left((\overline{X} - \mu)^2\right) = D(\overline{X}) = \frac{1}{n}\sigma^2$.

Je tedy $E(S^2) = \sigma^2$, a tedy $E(s^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2$.

Odvození vzorce pro rozptyl výběrového průměru nebudeme uvádět.

12.5. Náhodný výběr z normálního rozdělení. Jestliže mají nezávislé náhodné veličiny $X_i,\ 1 \leq i \leq n$ normální rozdělení $N(\mu, \sigma^2)$ (náhodný výběr z normálního rozdělení), má pak

výběrový úhrn $\widetilde{X} = \sum\limits_{i=1}^{n} X_i$ rozdělení $N(n\mu, n\sigma^2)$

a **výběrový průměr** $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ rozdělení $N(\mu, \frac{\sigma^2}{n})$.

12.6. Poznámka: Vlastnosti rozdělení výběrového průměru a úhrnu vyplývají z centrální limitní věty. Za předpokladů, že se jedná o výběr z rozdělení s konečnou střední hodnotou μ a konečným rozptylem σ^2 má výběrový úhrn $\widetilde{X}$ v limitě normální rozdělení $N(n\mu, n\sigma^2)$ a výběrový průměr $\overline{X}$ má v limitě normální rozdělení $N(\mu, \frac{\sigma^2}{n})$. Tyto skutečnosti můžeme zapsat vztahy pro distribuční funkce. Je

$$\lim_{n\to\infty} P\left(\frac{\widetilde{X}-n\mu}{\sigma\sqrt{n}} \le x\right) = \lim_{n\to\infty} P\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \le x\right) =$$

$$= \Phi(x),\ x \in \boldsymbol{R},$$

kde Φ je distribuční funkce normovaného rozdělení $N(0,1)$.

Některá rozdělení funkcí náhodného výběru (statistik).

12.13. Rozdělení chí kvadrát $\chi^2(n)$ o n stupních volnosti je rozdělení, které má náhodná veličina $X = \sum\limits_{i=1}^{n} U_i^2$, kde U_i, $1 \le i \le n$ jsou nezávislé náhodné veličiny s normovaným normálním rozdělením $N(0,1)$. Pro toto rozdělení je $E(X) = n$ a $D(X) = 2n$. Hustota f tohoto rozdělení je dána předpisem

$$f(x) = \left\langle \begin{array}{ll} 0, & x \le 0, \\ \frac{1}{\sqrt{2^n}\Gamma(\frac{n}{2})}\, x^{\frac{n}{2}-1}\, \mathrm{e}^{-\frac{x}{2}}, & x > 0. \end{array} \right.$$

Rozdělení je výrazně asymetrické, kvantily jsou kladné a jsou tabelovány. Až pro výrazně veliké hodnoty parametru n je možné toto rozdělení nahradit rozdělením normálním $N(n;2n)$. Pro velké hodnoty n, $n > 100$, má náhodná veličina

$$U = \frac{X-n}{\sqrt{2n}} \Leftrightarrow X = n + \sqrt{2n}U$$

přibližně normované normální rozdělení $N(0,1)$. Tvrzení vyplývá ze skutečnosti, že se jedná o rzdělení součtu náhodných veličin s konečnou střední hodnotou a rozptylem. Pro kvantily x_p náhodné veličiny X pak platí přibližný vzorec

$$x_p \doteq n + u_p\sqrt{2n},$$

kde u_p jsou kvantily rozdělení $N(0;1)$. Průběh hustoty rozdělení $\chi^2(n)$, pro dva stupně volnosti je znázorněn na obrázku Obr. 12.4.

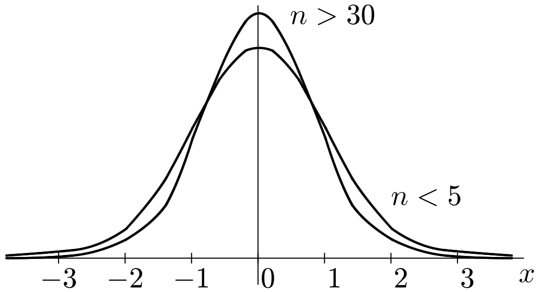
12.14. Studentovo rozdělení (t- rozdělení) $t(n)$ o n stupních volnosti má náhodná veličina

$$T = \frac{U\sqrt{n}}{Z},$$

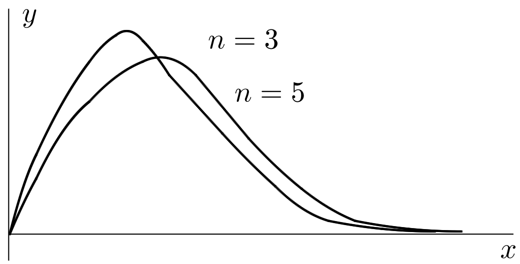
kde náhodná veličina U má normované normální rozdělení $N(0,1)$ a náhodná veličina Z má rozdělení $\chi^2(n)$. Rozdělení je symetrické vzhledem k počátku, je $E(T) = 0$, $D(T) = \frac{n}{n-2}$,
$n > 2$ a pro hodnoty $n > 30$ jej nahrazujeme normovanýn normálním rozdělením $N(0,1)$. Pro kvantily platí $t_p = -t_{1-p}$. Hustota f Studentova rozdělení je dána vzorcem

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \boldsymbol{R}.$$

Průběh hustoty $t-$rozdělení pro dvě hodnoty stupňů volnosti je znázorněn na obrázku Obr. 12.3.



Obr.12.3.



Obr. 12.4.

12.15. Fischerovo-Snedecorovo rozdělení ($F-$rozdělení) $F_{m,n}$ o m a n stupních volnosti má náhodná veličina

$$F = \frac{Xn}{Ym},$$

kde náhodná veličina X má rozdělení $\chi^2(m)$ a náhodná veličina Y má rozdělení $\chi^2(n)$. Náhodná veličina F nabývá pouze kladných hodnot a je $E(F) = \frac{n}{n-2}$, $n > 2$ a
$D(F) = \frac{2n^2(n+m-2)}{m(n-2)^2(n-4)}$, $n > 4$. Hustota f náhodné veličiny F je dána vzorcem

$$f(x) = \frac{1}{B(\frac{m}{2}, \frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, \; x > 0.$$

Pro kvantily $F_p(m,n)$ rozdělení platí

$$F_p(m,n) = \frac{1}{F_{1-p}(n,m)}, \; 0 < p < 1.$$

Je totiž kvantil F_p náhodné veličiny $F(m,n)$ určen podmínkou

$$P\left(F(m,n) \leq F_p\right) = p \Leftrightarrow P\left(\frac{Xn}{Ym} \leq F_p\right) = p.$$

Odtud plyne, že

$$P\left(\frac{1}{F_p} \leq \frac{Ym}{Xn}\right) = 1 - P\left(\frac{Ym}{Xn} \leq \frac{1}{F_p}\right) = p \Rightarrow$$

$$P\left(\frac{Ym}{Xn} \leq \frac{1}{F_p}\right) = P\left(F(n,m) \leq \frac{1}{F_p}\right) = 1 - p,$$

což je podmínka pro kvantil náhodné veličiny $F(n,m)$. Je tedy
$F_{1-p}(n,m) = \dfrac{1}{F_p(m,n)}$.

Eulerovy funkce

Definice: 1. Eulerovy funkce 1. a 2. druhu definujeme:
Funkci *Gama* (Eulerova funkce 2. druhu) vztahem

$$\Gamma(z) = \int_0^\infty \mathrm{e}^{-x} x^{z-1}\,\mathrm{d}x, \quad z > 0.$$

Funkci *Beta* (Eulerova funkce 1. druhu) vztahem

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,\mathrm{d}x, \quad p > 0,\ q > 0.$$

Poznámka: Funkce voláme v
MAPLE: *GAMMA(z), BETA(p,q);*
OpenOffice: nemá zavedeny.

Věta: 2. Pro funkci Γ platí:
1. $\Gamma(z+1) = z\Gamma(z),\ z > 0$;
Formuli odvodíme integrací per partes. Je

$$\Gamma(z+1) = \int_0^\infty \mathrm{e}^{-x} x^z\,\mathrm{d}x = \left| \begin{array}{ll} u' = \mathrm{e}^{-x}, & v = x^z \\ u = -\mathrm{e}^{-x}, & v' = zx^{z-1} \end{array} \right| =$$

$$= \left[-\mathrm{e}^{-x}x^z\right]_0^\infty + +z \int_0^\infty \mathrm{e}^{-x} x^{z-1}\,\mathrm{d}x = z\Gamma(z),$$

jestliže použijeme skutečnosti, že $\lim\limits_{x\to\infty} (\mathrm{e}^{-x}x^z) = 0$.
2. $\Gamma(n+1) = n!,\ n \in \boldsymbol{N}; \quad \Gamma(1) = \Gamma(2) = 1;$
Pro $z = 1$ dostaneme

$$\Gamma(1) = \int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x = \left[-\mathrm{e}^{-x}\right]_0^\infty = 1.$$

Dále je

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \ldots 1\Gamma(1) = n!.$$

3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}, \quad \Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}.$
Pro $z = \frac{1}{2}$ dostaneme

$$\Gamma(\frac{1}{2})) = \int_0^\infty \mathrm{e}^{-x} x^{-\frac{1}{2}}\,\mathrm{d}x = \left| \begin{array}{ll} x = t^2, & x^{-\frac{1}{2}} = t^{-1} \\ \mathrm{d}x = 2t\mathrm{d}t, & \end{array} \right| =$$

$$= \int_0^\infty e^{-t^2} t^{-1} 2t \mathrm{d}t = 2\int_0^\infty e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}.$$

Dále je $\Gamma(\frac{3}{2}) = \Gamma(\frac{1}{2}+1) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$.

4. $\lim\limits_{z\to 0+} \Gamma(z) = \lim\limits_{z\to\infty} \Gamma(z) = \infty$;

Ze vztahu 2 dostaneme

$$\Gamma(z) = \frac{\Gamma(z+1)}{z} \Rightarrow \lim_{z\to 0+} \frac{\Gamma(z+1)}{z} = \left|\frac{1}{0+}\right| = \infty.$$

Obdobně dostaneme pro $z \to \infty$

$$\lim_{z\to\infty} \Gamma(z) = \lim_{n\to\infty} \Gamma(n+1) = \lim_{n\to\infty} n! = \infty.$$

5. *Stirlingova formule*

$\Gamma(z+1) \approx z^z.e^{-z}\sqrt{2\pi z}, \quad n! \approx n^n.e^{-n}\sqrt{2\pi n},\ z, n \to \infty.$

Odvození formule se vymyká z rámce kurzu.

Věta: 3. Pro Eulerovu funkci B platí:

1. $B(p,q) = B(q,p)$;

Vztah plyne ze vzorce pro provedení jednoduché substituce. Je

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,\mathrm{d}x = \left| \begin{array}{cc} 1-x=t,\ x=1-t & 1\to 0 \\ -\mathrm{d}x = \mathrm{d}t, & 0\to 1 \end{array} \right| =$$

$$= -\int_1^0 (1-t)^{p-1} t^{q-1}\,\mathrm{d}t = B(q,p).$$

2. $B(p,1) = \frac{1}{p}, \quad B(1,q) = \frac{1}{q}, \quad B(1,1) = 1$;

Po dosazení dostaneme

$$B(p,1) = \int_0^1 x^{p-1}\,\mathrm{d}x = \left[\frac{x^p}{p}\right]_0^1 = \frac{1}{p};$$

$$B(1,q) = B(q,1) = \frac{1}{q}; \quad \text{pro } p=1\ B(1,1) = 1.$$

3. $B(p,q+1) = \frac{q}{p}B(p+1,q)$;

Integrací per partes dostaneme

$$B(p,q+1) = \int_0^1 x^{p-1}(1-x)^q\,\mathrm{d}x = \left| \begin{array}{ll} u' = x^{p-1}, & v = (1-x)^q \\ u = \frac{x^p}{p}, & v' = -qx^{q-1} \end{array} \right| =$$

$$= \left[\frac{x^p}{p}(1-x)^q\right]_0^\infty + \frac{q}{p}\int_0^1 x^p(1-x)^{q-1}\,\mathrm{d}x = \frac{q}{p}B(p+1,q).$$

4. $B(m,n) = \dfrac{(m-1)!(n-1)!}{(m+n-1)!}$;

5. $B(p,q) = \dfrac{\Gamma(p).\Gamma(q)}{\Gamma(p+q)}$.