8 DALŠÍ SPOJITÁ ROZDĚLENÍ PRAVDĚPODOBNOSTI

Čas ke studiu kapitoly: 60 minut

Cíl: Po prostudování tohoto odstavce budete umět:

- charakterizovat další typy spojitých rozdělení: χ^2 , Studentovo, Fischer-Snedocorovo

Výklad:

Kapitolu o spojitých rozděleních věnujeme třem důležitým rozdělením nacházejícím uplatnění zejména při odhadech a testování statistických hypotéz.

8.1 χ^2 – rozdělení

Mějme nezávislé náhodné veličiny $Z_1; Z_2; \ldots ; Z_n$, z nichž každá má normované normální rozdělení. Pak součet čtverců těchto náhodných veličin má rozdělení χ_n^2 (chí-kvadrát) s n stupni volnosti (degree of freedom, DF).

$$X = \sum_{i=1}^{n} Z_i^2$$

To, že má náhodná veličina X rozdělení χ_n^2 označujeme:

$$X \to \chi_n^2$$

Počet stupňů volnosti tedy označuje počet sčítaných nezávislých náhodných veličin a je jediným parametrem tohoto rozdělení. Zároveň je definice rozdělení χ_n^2 zřejmé, že náhodná veličina s tímto rozdělením může nabývat pouze kladných hodnot.

Hustotu pravděpodobnosti v obecném tvaru (pro n stupňů volnosti) nebudeme pro značnou komplikovanost vztahu uvádět.

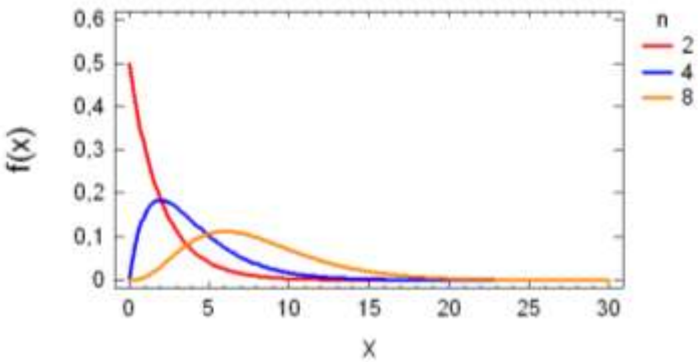
Střední hodnota: $E(\chi_n^2) = n$

Rozptyl: $D(\chi_n^2) = 2n$

100p%-ní kvantily:

Pro některá významná p (0,01; 0,05; 0,10; ...) a pro některá n jsou kvantily χ_n^2 rozdělení tabelovány (viz. příloha – Tabulka 3). Běžně lze také kvantily tohoto rozdělení stanovit pomocí statistického software.

Grafické znázornění hustoty pravděpodobnosti pro různé stupně volnosti



8.1.1 Vlastnosti rozdělení χ^2

1. Pro nezávislé náhodné veličiny s χ^2 rozdělením se dá dokázat, že jejich součet má opět χ^2 rozdělení a počet stupňů volnosti je roven součtu stupňů volnosti jednotlivých veličin v součtu.

$$\text{Nechť } X_i \to \chi^2_{n_i},\ X = \sum_{(i)} \chi^2_{n_i}\text{, pak } X \to \chi^2_{\sum_{(i)} n_i}.$$

2. Pokud náhodné veličiny $X_1, X_2, \ldots, X_n$ mají normální rozdělení $N(\mu,\sigma^2)$ a jsou navzájem nezávislé, pak výběrový rozptyl s^2 vynásobený (n-1) a vydělený σ^2 má rozdělení χ^2_{n-1}. (Plyne to bezprostředně z toho, že tento výraz se dá převést na součet čtverců (n-1) náhodných veličin s rozdělením N(0,1))

Důkaz: $$s^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(X_i-\bar{x})^2 \Rightarrow \frac{s^2}{\sigma^2}\cdot(n-1) = \sum_{i=1}^{n}\frac{(X_i-\bar{x})^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{x}}{\sigma}\right)^2,$$

Pomocí dalších úprav (zdlouhavé), které vedou na nahrazení průměru střední hodnotou, bychom zjistili, že:

$$\frac{s^2}{\sigma^2}\cdot(n-1) = \sum_{i=1}^{n-1}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n-1} Z_i^2$$

(Nahrazení průměru střední hodnotou vede na ztrátu jednoho stupně volnosti.)

Tuto skutečnost můžeme stručně zapsat takto:

$$\frac{(n-1)\,s^2}{\sigma^2} \to \chi^2_{n-1}$$

8.1.2 Použití rozdělení χ^2

1. Vlastnosti, že:

$$\frac{(n-1)\,s^2}{\sigma^2} \to \chi^2_{n-1}$$

(výběrový rozptyl s^2 vynásobený (n-1) a vydělený σ^2 má rozdělení χ^2_{n-1}) se využívá při testování toho, zda rozptyl základního souboru s normálním rozdělením je roven σ^2. (viz. Testování hypotéz)

2. χ^2_n rozdělení se používá pro ověření nezávislosti kategoriálních proměnných (test nezávislosti v kombinační tabulce)

3. Pokud testujeme, zda náhodné veličiny (naměřená data) pocházejí z určitého rozdělení, můžeme také s úspěchem použít chí-kvadrát rozdělení. Tento test je znám pod názvem "test dobré shody".

Řešený příklad:

Odvoďte distribuční funkci a hustotu pravděpodobnosti náhodné veličiny X, která má rozdělení χ_n^2 s jedním stupněm volnosti.

Řešení:

$$X = Z^2$$

$$Z \to N(0;1) \quad \Rightarrow \quad X \to \chi_1^2$$

Náhodná veličina X je funkcí náhodné veličiny Z a proto budeme při hledání její distribuční funkce dále postupovat již známým způsobem (pouze vezmeme v úvahu, že náhodná veličina s rozdělením χ_n^2 nabývá pouze kladných hodnot):

$pro\ x > 0:$

$$F(x) = P(X < x) = P(Z^2 < x) = P(-\sqrt{x} < Z < \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = \Phi(\sqrt{x}) - [1 - \Phi(\sqrt{x})] =$$

$$= 2\Phi(\sqrt{x}) - 1 = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{t}{2}} dt - 1 = \sqrt{\frac{2}{\pi}} \cdot \int_0^{\sqrt{x}} e^{-\frac{t}{2}} dt - 1$$

$pro\ x \le 0:$

$F(x) = 0$

Hustotu pravděpodobnosti pak určíme jednoduše jako derivaci distribuční funkce:

$$f(x) = \frac{dF(x)}{dx} = \begin{matrix} 2 \cdot \frac{1}{2\sqrt{x}} \cdot \varphi(\sqrt{x}) = \frac{1}{\sqrt{x}} \cdot \varphi(\sqrt{x}) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \le 0 \end{matrix}$$

Výklad:

8.2 Studentovo t rozdělení

Uvažujme dvě nezávislé náhodné veličiny Z a χ_n^2. Náhodná veličina Z má normované normální rozdělení, náhodná veličina χ_n^2 má rozdělení χ^2 s n stupni volnosti. Potom náhodná veličina t;

$$T_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}}$$

má Studentovo t rozdělení s n stupni volnosti. Počet stupňů volnosti je jediný parametr tohoto rozdělení.

Pro $n \to \infty$ (vysoký počet stupňů volnosti, v praxi pro n>30) se Studentovo t rozdělení blíží normovanému normálnímu rozdělení.

Hustotu pravděpodobnosti nebudeme ani v tomto případě pro její složitost uvádět.

Střední hodnota: $E(T_n) = 0$

Rozptyl: $D(T_n) = \frac{n}{n-2}$

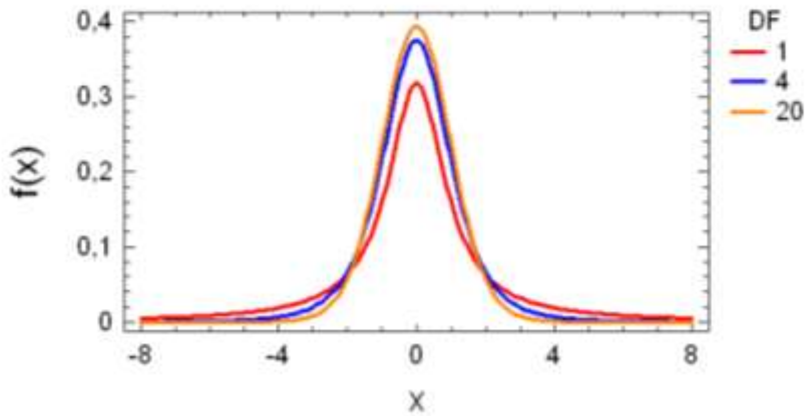
100p%-ní kvantily - t_p:

Pro vybraná p a pro vybrané stupně volnosti n jsou 100p%-ní kvantily tabelovány (viz. příloha – Tabulka 2). Většinou je tato tabelace provedena pouze pro p<0,5. Kvantily t_p pro p>0,5 získáme pomocí vztahu:

$$t_p = -t_{1-p}$$

Běžně se při určování kvantilů využívá rovněž statistický software.

Grafické znázornění hustoty pravděpodobnosti pro různé stupně volnosti



8.2.1 Vlastnosti Studentova t rozdělení

Náhodná veličina definována jako:

$$\frac{\overline{X}-\mu}{s}\cdot\sqrt{n}$$

má Studentovo t rozdělení s (n-1) stupni volnosti.

$$\frac{\overline{X}-\mu}{s}\cdot\sqrt{n}\to t_{n-1}$$

Odvození bude provedeno v následujícím Průvodci studiem.

8.2.2 Použití Studentova t rozdělení

Studentovo t rozdělení má široké uplatnění. Uvedeme alespoň některé možnosti použití.

1. Užívá se k testování hypotéz o střední hodnotě náhodného výběru, pokud je rozptyl neznámý. Mělo by platit, že tento náhodný výběr pochází z normálního rozdělení, ale tento předpoklad je většinou alespoň přibližně splněn, jak dále poznáme.

2. Užívá se k testování hypotéz o shodě středních hodnot dvou náhodných výběrů, se stejnými předpoklady jako v předcházejícím případě - navíc musí být tyto výběry nezávislé.

3. Rozdělení je vhodným prostředkem pro analýzu výsledků regresní analýzy.

Průvodce studiem:

Původ názvu Studentovo má zajímavou historii. Irský statistik W. S. Gosset poprvé publikoval toto rozdělení anonymně pod pseudonymem "Student", protože jeho zaměstnavatel, pivovar Guiness v Dublinu, zakázal svým zaměstnancům publikovat pod svým vlastním jménem z obavy, že konkurence by odhalila tajemství jejich excelentního piva. Ve svém původním článku, Gosset použil označení "t" pro svoji statistiku. Odtud Studentovo t rozdělení pravděpodobnosti. Na práci Gosseta navázalo množství dalších statistiků; jmenujme alespoň R. A. Fishera, jehož jméno můžeme najít téměř ve všech směrech dalšího vývoje statistiky.

Následující odvození je opět určeno zájemcům o matematické pozadí používaných vztahů.

- ***Odvození vlastností, že:*** $\frac{\overline{X}-\mu}{s}\cdot\sqrt{n}\to t_{n-1}$

Pokud náhodné veličiny $X_1,X_2,\dots,X_n$ mají normální rozdělení $N(\mu,\sigma^2)$ a jsou navzájem nezávislé, pak lze snadno ukázat (viz. Centrální limitní věta), že

$$\overline{X}\to N(\mu,\frac{\sigma^2}{n})$$

a dále po standardizaci (transformaci normální na normovanou normální náhodnou veličinu) platí:

$$\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \rightarrow N(0,1)$$

Dále víme, že:

$$\frac{(n-1)\,s^2}{\sigma^2} \quad \rightarrow \chi^2_{n-1}$$

Protože Studentova náhodná veličina je definována jako:

$$T_n = \frac{Z}{\sqrt{\frac{\chi^2_n}{n}}}$$

Proto:

$$T_{n-1} = \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$$

Jako náhodné veličiny Z a χ^2_{n-1} *tedy mějme:*

$$Z = \frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}; \qquad \chi^2_{n-1} = \frac{(n-1)\,s^2}{\sigma^2}$$

Pak:

$$\frac{\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)\,s^2}{\sigma^2}}{n-1}}} \rightarrow t_{n-1}$$

Po úpravě:

$$\frac{\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)\,s^2}{\sigma^2}}{n-1}}} = \frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \cdot \frac{\sigma}{s} = \frac{\overline{X}-\mu}{s} \cdot \sqrt{n} \rightarrow t_{n-1}$$

Výklad:

8.3 Fisherovo-Snedecorovo rozdělení - F rozdělení

Posledním spojitým rozdělením, kterým se budeme zabývat, je F rozdělení. Mějme dvě nezávislé náhodné veličiny χ^2_m a χ^2_n. Obě mají rozdělení chí-kvadrát. První z nich má počet stupňů volnosti m, druhá má počet stupňů volnosti n (obecně mají různý počet stupňů volnosti). Pak náhodná veličina $F_{m,n}$;

$$F_{m,n} = \frac{\frac{\chi^2_m}{m}}{\frac{\chi^2_n}{n}}$$

má Fisherovo-Snedecorovo rozdělení o m a n stupních volnosti. Toto rozdělení má tedy dva parametry- n a m.

Ani v tomto případě nebudeme uvádět vztah pro hustotu pravděpodobnosti (značně složitý).

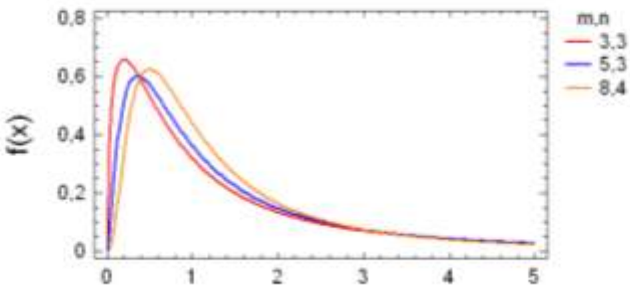
Střední hodnota: $E(F_{m,n}) = \frac{n}{n-2}$

Rozptyl: $D(F_{m,n}) = \frac{2n^2\left(1+\frac{n-2}{m}\right)}{(n-2)^2\,(n-4)}$

100p%-ní kvantily - F_p:

Pro praktické aplikace jsou pro vybrané pravděpodobnosti (p>0,5) a vybrané stupně volnosti n a m tabelovány kvantily F_p (viz. příloha – Tabulka 4). Pro p>0,5 se kvantily F_p určí ze vztahu:

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



Grafické zobrazení hustoty pravděpodobnosti pro různé hodnoty m a n

8.3.1 Vlastnosti Fischerova-Snedecorova rozdělení

Je zřejmé použití tohoto rozdělení – jako rozdělení výběrových rozptylů dvou nezávislých náhodných vektorů, se stejnou směrodatnou odchylkou σ.

Nechť máme dva náhodné vektory: $\overline{X_1}; \overline{X_2}$

$$X_{1j} \to N(\mu_1, \sigma); \quad \mathrm{j} = 1, \mathrm{n}_1$$
$$X_{2j} \to N(\mu_2, \sigma); \quad \mathrm{j} = 1, \mathrm{n}_2$$

S_1; S_2 jsou náhodné veličiny definované jako:

$$S_1^2 = \frac{\sum_{j=1}^{n_i} \left(X_{1j} - \overline{X_1}\right)^2}{n_1 - 1}; \qquad S_2^2 = \frac{\sum_{j=1}^{n_i} \left(X_{2j} - \overline{X_2}\right)^2}{n_2 - 1}$$

Pak:

$$\frac{S_1^2}{S_2^2} \to F_{n_1, n_2}$$

8.3.2 Použití Fischerova-Snedecorova rozdělení

Toto rozdělení má opět široké uplatnění, především při hodnocení výsledků statistických analýz. Používá se především:

1. k testu o shodnosti rozptylů dvou náhodných výběrů
2. k testům o shodě středních hodnot pro více náhodných výběrů, v analýze rozptylu.
3. k testům v regresní analýze.

Shrnutí:

Spojitá rozdělení, s kterými jsme se v této kapitole seznámili nacházejí uplatnění zejména při odhadech a testování statistických hypotéz. Jejich 100p%-ní kvantily jsou pro vybrané p a vybrané parametry tabelovány

Název rozdělení	**Definice NV s daným rozdělením**	**Parametry**	**Vlastnosti**
χ_n^2	$\chi_n^2 = \sum_{i=1}^{n} Z_i^2$	n	$\frac{(n-1)\,s^2}{\sigma^2} \to \chi_{n-1}^2$
Studentovo	$T_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}}$	n	$\frac{\overline{X}-\mu}{s} \cdot \sqrt{n} \to t_{n-1}$
Fischerovo-Snedecorovo	$F_{n,m} = \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}}$	m, n	$\frac{S_1^2}{S_2^2} \to F_{n_1,n_2}$

Otázky

1. Definujte náhodné veličiny: χ_n^2, Studentovu a Fischer-Snedecorovu (prokažte orientaci v tabulkách kvantilů těchto rozdělení)