Pravděpodobnostní rozdělení matematické statistiky 1.

χ^2 rozdělení

Mějme náhodnou proměnnou ξ s hustotou pravděpodobnosti $f_\xi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, tj. normální náhodnou centrovanou a normovanou proměnnou $N(0,1)$. Otázkou je rozdělení kvadrátu této náhodné proměnné $\eta=\xi^2$.

Nejprve obecně:

$$\eta=\xi^2 \Rightarrow F_\eta(x)=P(\eta<x)=P\left(\xi^2<x\right)=P\left(-\sqrt{x}<\xi<\sqrt{x}\right)=F_\xi\left(\sqrt{x}\right)-F_\xi\left(-\sqrt{x}\right)$$

$$f_\eta(x)=\frac{d}{dx}\left(F_\xi\left(\sqrt{x}\right)-F_\xi\left(-\sqrt{x}\right)\right)=\frac{1}{2\sqrt{x}}\left(f_\xi\left(\sqrt{x}\right)+f_\xi\left(-\sqrt{x}\right)\right)$$

Pro náš konkrétní případ $f_\eta(x)=\frac{1}{2\sqrt{x}}\left(f_\xi\left(\sqrt{x}\right)+f_\xi\left(-\sqrt{x}\right)\right)=\begin{matrix}\frac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}} & \Leftrightarrow x>0 \\ 0 & \Leftrightarrow x\le 0\end{matrix}$.

Charakteristická funkce bude $C_\eta(j\omega)=E\left\{e^{j\omega\eta}\right\}=\int_0^{+\infty}e^{j\omega s}f_\eta(s)ds=\int_0^{+\infty}e^{j\omega s}\frac{1}{\sqrt{2\pi s}}e^{-\frac{s}{2}}ds=$

$$=\frac{1}{\sqrt{2\pi}}\int_0^{+\infty}s^{-\frac{1}{2}}e^{j\omega s-\frac{s}{2}}ds=\frac{1}{\sqrt{1-2j\omega}}=\boxed{(1-2j\omega)^{-\frac{1}{2}}}.$$

Odtud dostaneme pro střední hodnotu (obecně): $\frac{d^k}{d\omega^k}C_\eta(j\omega)\Big|_{\omega=0}=(j)^k m_k(\eta)$, kde $m_k(\eta)$ je k-tý obecný moment náhodné veličiny η.

$$\frac{d}{d\omega}C_\eta(j\omega)\Big|_{\omega=0}=j(1-2j\omega)^{-\frac{3}{2}}\Big|_{\omega=0}=j=(j)^1 m_1(\eta)\Rightarrow E\left\{\xi^2\right\}=1$$

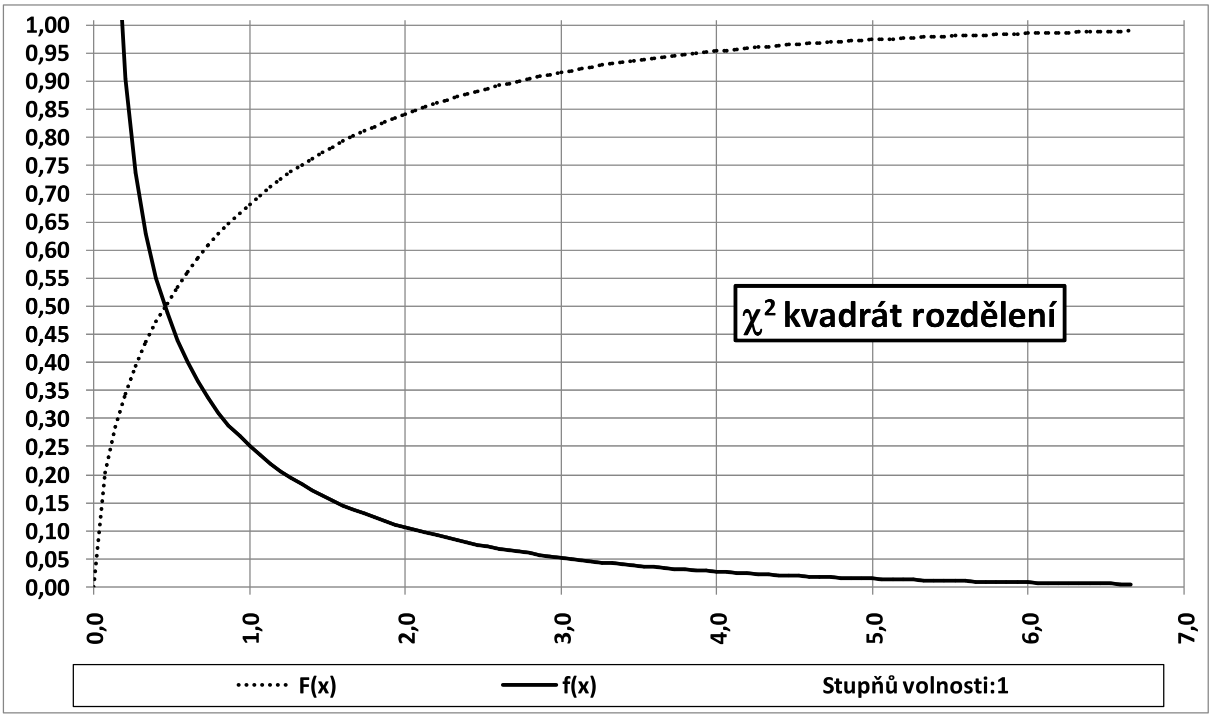
$$\frac{d^2}{d\omega^2}C_\eta(j\omega)\Big|_{\omega=0}=-3(1-2j\omega)^{-\frac{5}{2}}\Big|_{\omega=0}=-3=(j)^2 m_2(\eta)\Rightarrow E\left\{\left(\xi^2\right)^2\right\}=3$$ a odtud

$\sigma^2\left(\xi^2\right)=E\left\{\left(\xi^2\right)^2\right\}-\left(E\left\{\xi^2\right\}\right)^2=3-1=2$.

Shrnuto: $E\left\{\xi^2\right\}=1$ a $\sigma^2\left(\xi^2\right)=2$

Náhodná proměnná $\eta=\xi^2$ se za uvedených předpokladů nazývá χ^2 **náhodná proměnná** s ***jedním*** stupněm volnosti.

Průběh její hustoty $f_{\xi^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \Leftrightarrow x > 0$ a distribuční funkce $F_{\xi^2}(x) = \int_0^x \frac{1}{\sqrt{2\pi z}} e^{-\frac{z}{2}} dz$ je následujícím obrázku:



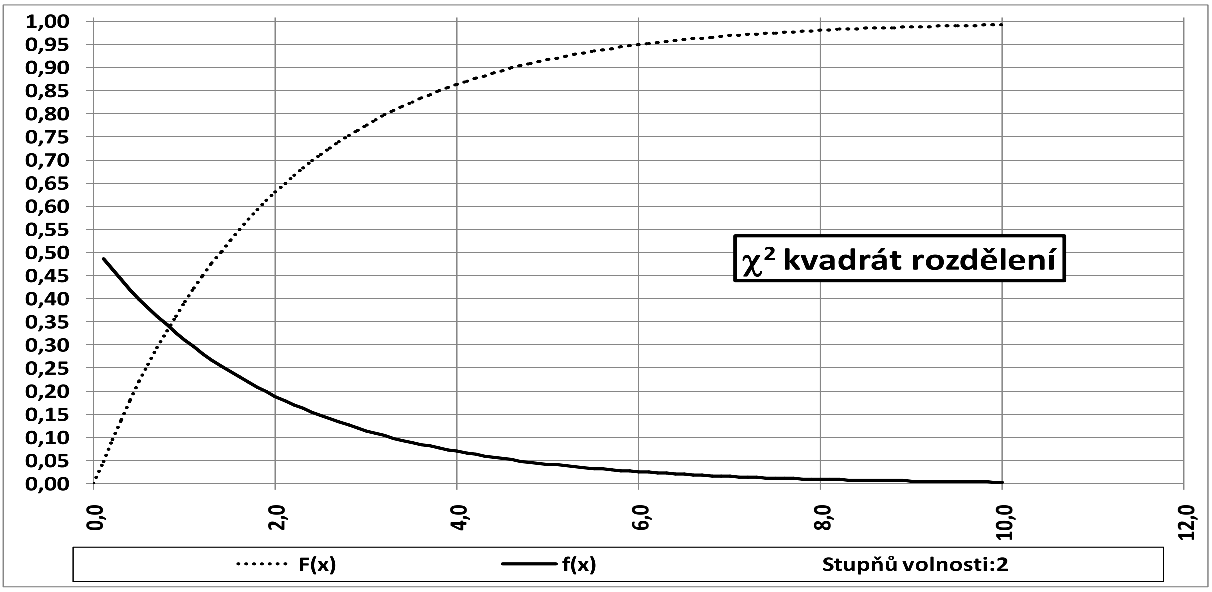
Mějme dále posloupnost nezávislých a stejně rozdělených náhodných proměnných $\xi_1, \xi_2, \dots, \xi_n$ s rozdělením $N(0,1)$. Zajímá nás rozdělení náhodné proměnné $\boxed{\chi_n^2 = \sum_{i=1}^{n} \xi_i^2}$, taková náhodná proměnná se nazývá $\boldsymbol{\chi^2}$ náhodná proměnná s ***n*** stupni volnosti.

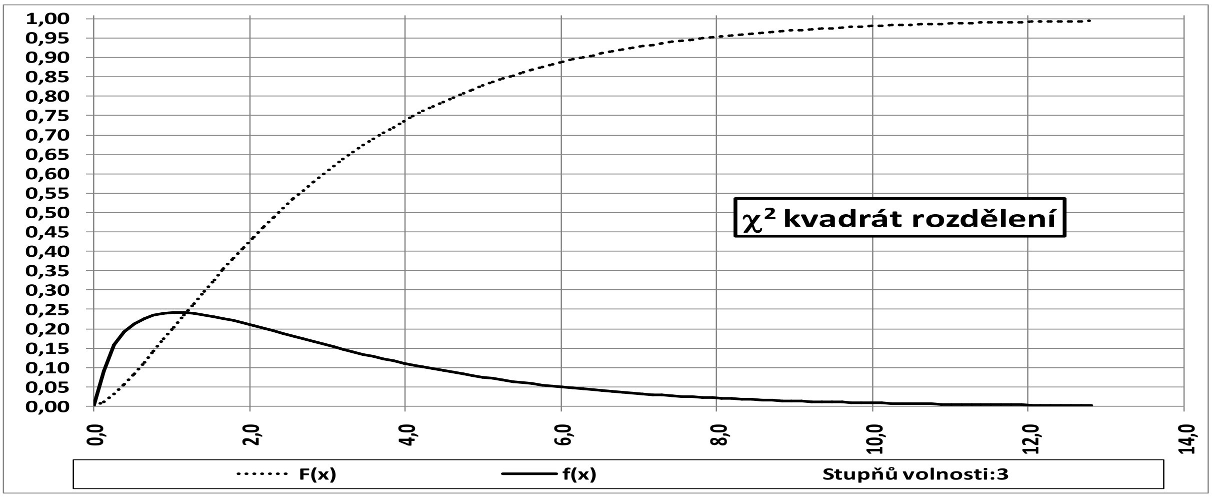
Její charakteristická funkce bude $C_{\chi_n^2}(j\omega) = (1 - 2j\omega)^{-\frac{n}{2}}$.

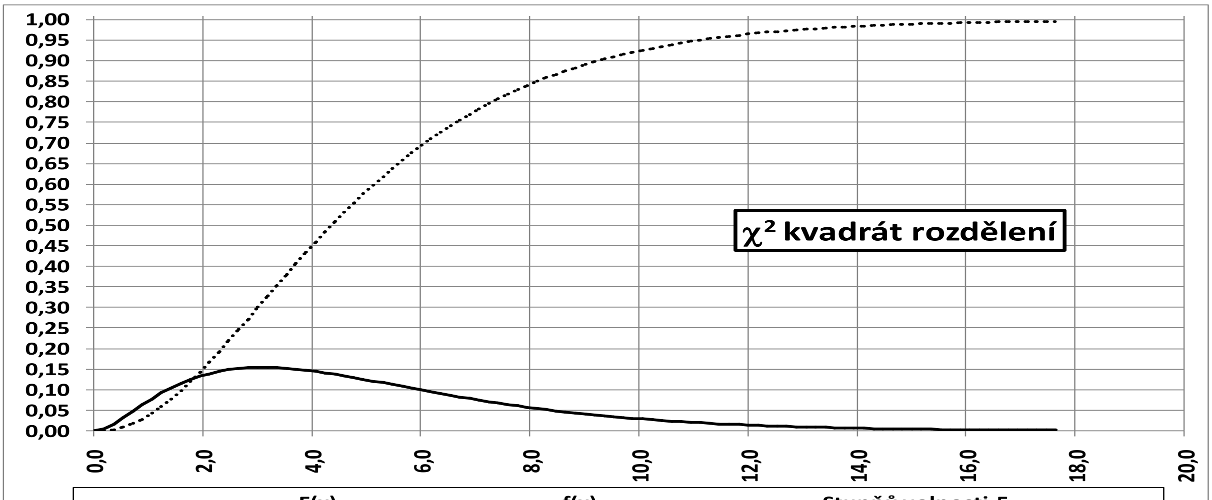
Pro hustotu této náhodné proměnné platí $\boxed{f_{\chi_n^2}(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}}$. Lze dokázat buď tím, že se najde vzor k charakteristické funkci (charakteristická funkce je až na násobící konstantu Fourierovou transformací hustoty) nebo indukcí podle stupňů volnosti: $\chi_{n+1}^2 = \chi_n^2 + \xi^2$ a pro hustoty pak platí $f_{\chi_{n+1}^2}(x) = \int_0^{+\infty} f_{\chi_n^2}(z) f_{\chi_1^2}(x-z) dz$, kde $f_{\chi_1^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} = \frac{x^{\frac{1}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)}$ a $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ (viz gama funkce v příloze tohoto textu).

Distribuční funkce pak bude $F_{\chi_n^2}(x) = \int_0^x \frac{z^{\frac{n}{2}-1} e^{-\frac{z}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} dz$. Tu lze snadno počítat pomocí neúplné gama funkce.

Příklady průběhů hustoty a distribučních funkcí χ^2 rozdělení jsou na následujících obrázcích:

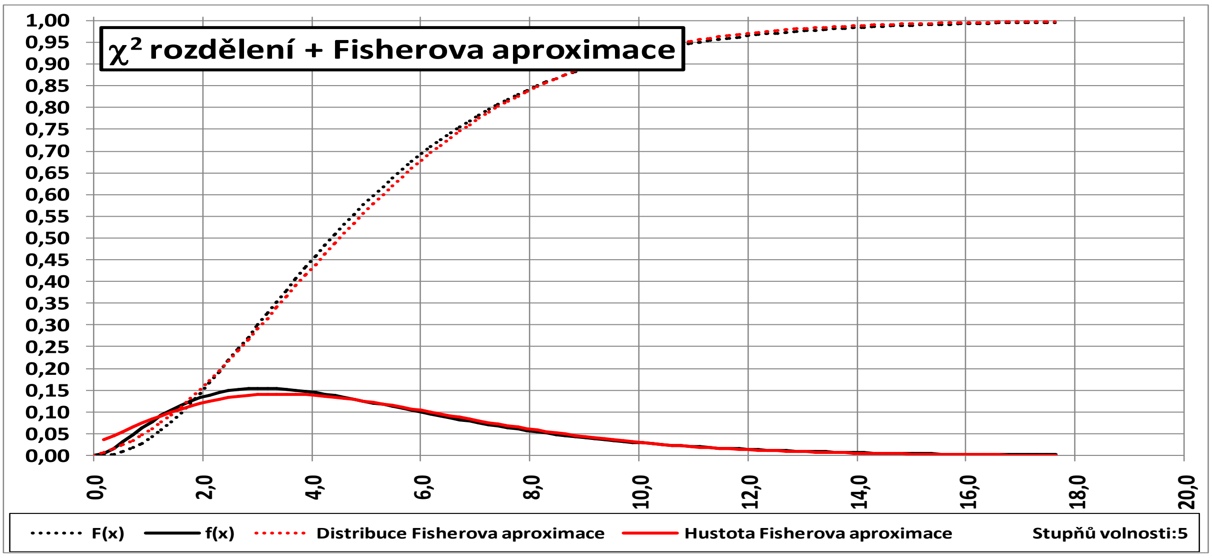


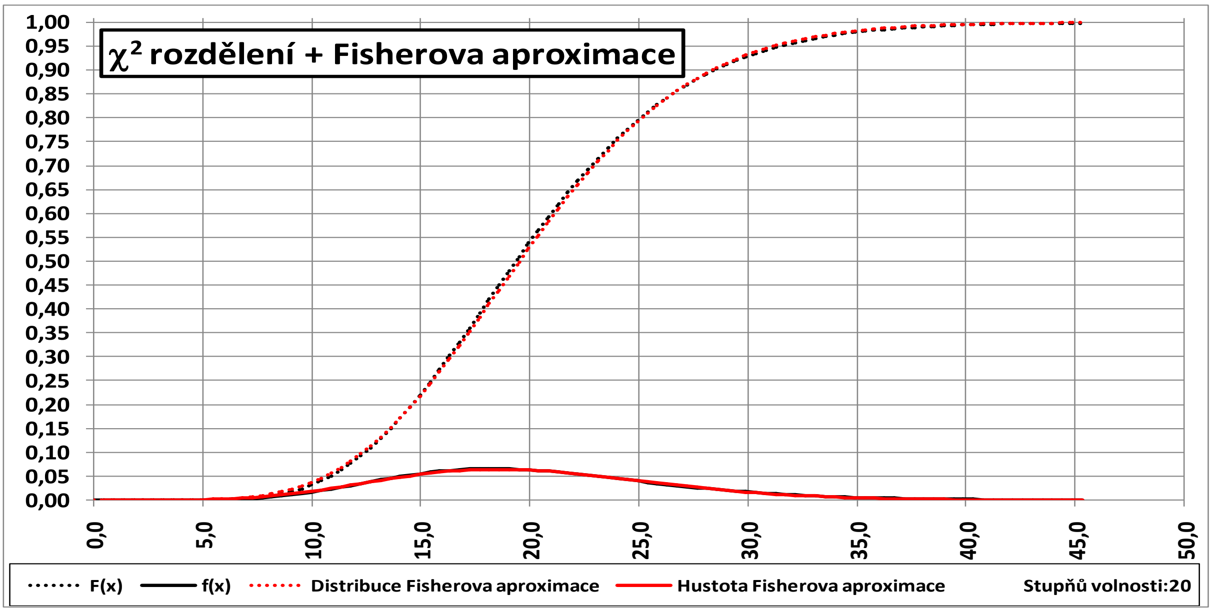




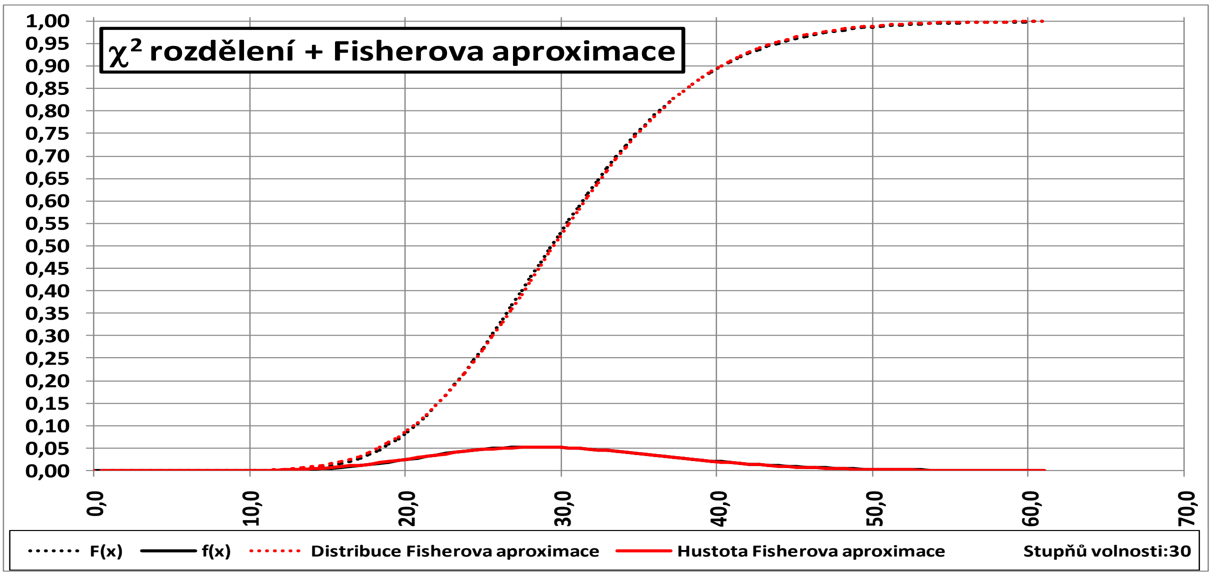
Pro velká $n > 30$ se používá Fisherova[1] aproximace hustoty a distribuce χ^2 rozdělení: $f_{\chi_n^2}(x) \approx \frac{\varphi\left(\sqrt{2x} - \sqrt{2n-1}\right)}{\sqrt{2x}}$ a $F_{\chi_n^2}(x) \approx \Phi\left(\sqrt{2x} - \sqrt{2n-1}\right)$. Viz Jaroslav Hátle, Jiří Likeš: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, str. 134-135.

Vztah aproximace ke skutečnému průběhu je demonstrován na následujících obrázcích:





[1] Sir Ronald Aylmer Fisher, FRS (17 February 1890 – 29 July 1962) was an English statistician, evolutionary biologist, eugenecist and geneticist. Zdroj http://en.wikipedia.org/wiki/ , http://www.genealogy.math.ndsu.nodak.edu/index.php .



Námět: Analyzujte rozdíl mezi „přesným“ vyjádřením distribuční funkce a její Fisherovou aproximací. Vysvětlete, zda a jak případně budou oblasti maximální odchylky vadit pro klasické statistické účely!

Námět: Odvoďte hustotu náhodné proměnné $\chi_n = \sqrt{\sum_{i=1}^{n} \xi_i^2}$.

Gama rozdělení

Mějme náhodnou proměnnou ξ s hustotou pravděpodobnosti $f_\xi(x) = \frac{1}{\tau} e^{-\frac{x}{\tau}}; x \geq 0$, tj. neposunuté exponenciální rozdělení se střední hodnotou τ. Otázkou je rozdělení součtu n takto shodně rozdělených nezávislých náhodných proměnných. Charakteristická funkce takového neposunutého exponenciálního rozdělení bude:

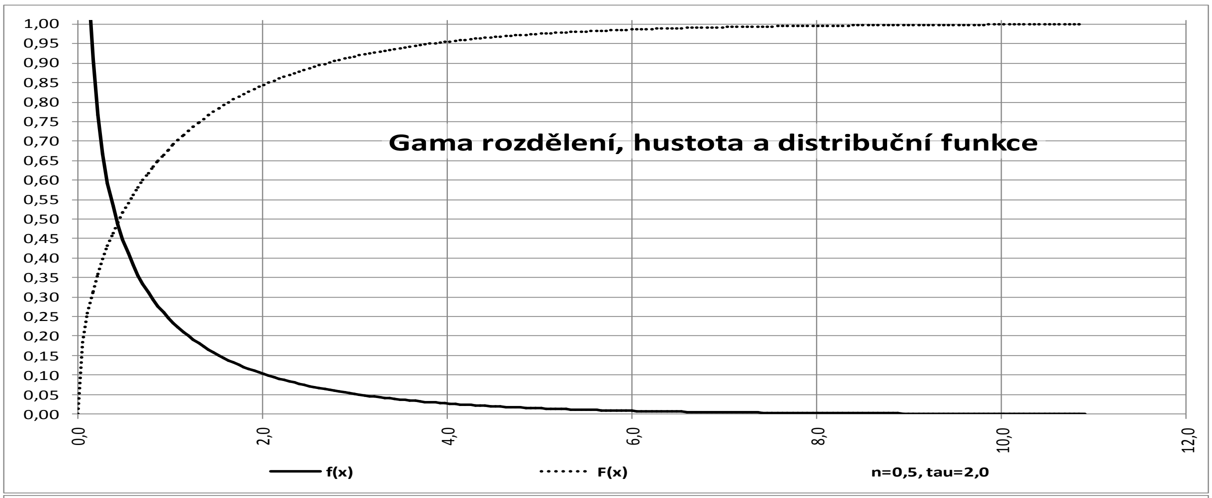
$$C_\xi(j\omega) = E\left\{e^{j\omega\xi}\right\} = \int_0^{+\infty} e^{j\omega s} f_\xi(s)ds = \int_0^{+\infty} e^{j\omega s} \frac{1}{\tau} e^{-\frac{s}{\tau}} ds = \frac{1}{\tau} \int_0^{+\infty} e^{-s\left(\frac{1}{\tau} - j\omega\right)} ds = \frac{1}{1 - j\omega\tau} = (1 - j\omega\tau)^{-1}.$$

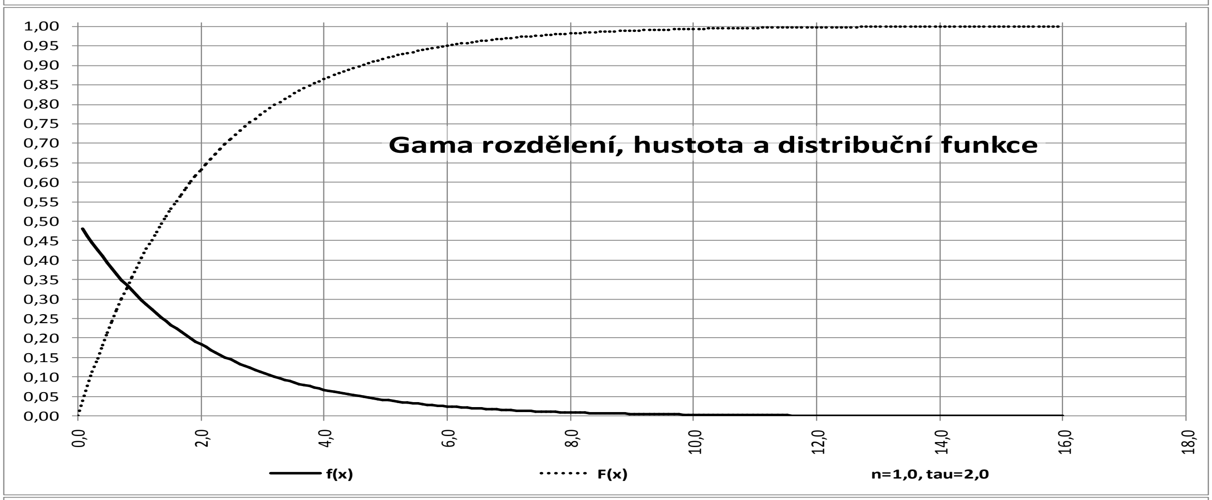
Charakteristická funkce náhodné proměnné $\eta = \sum_{i=1}^{n} \xi_i$ bude $C_\eta(j\omega) = (1 - j\omega\tau)^{-n}$. Zpětnou Fourierovou transformací dostaneme: $\boxed{f_\eta(x) = \frac{x^{n-1} e^{-\frac{x}{\tau}}}{\tau^n \Gamma(n)}; x \geq 0,}$ $\Gamma(n)$ je opět „úplnou“ gama funkcí. Takové rozdělení budeme označovat $G(n,\tau)$(neslučovat s neúplnou gama funkcí, pozor, může být i jiná konvence označování vycházející z jiné parametrizace výchozího exponenciálního rozdělení $f_\xi(x) = \lambda e^{-\lambda x}; x \geq 0$). Obdobně jako u χ^2 rozdělení lze hustotu gama rozdělení odvodit rekurentně posloupností konvolucí. Pro takové postupy viz Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, str. 180-187. Není problémem ověřit, že taková hustota je hustotou i pro reálné $n > 0$, nejen celočíselné.

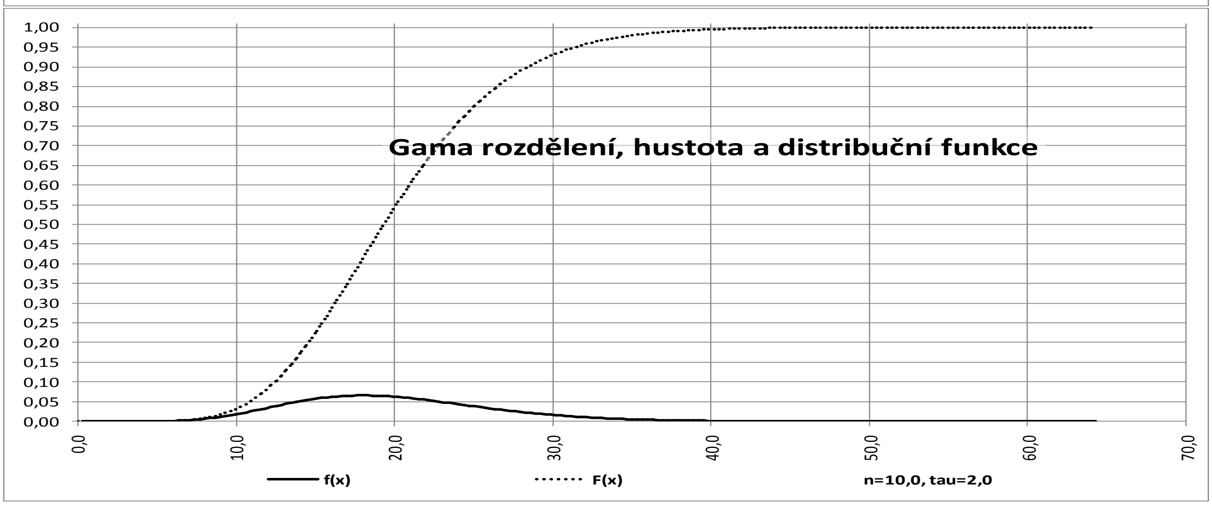
Distribuční funkce gama rozdělení bude

$$F_\eta(x)=\int_0^x \frac{s^{n-1}e^{-\frac{s}{\tau}}}{\tau^n\Gamma(n)}\,ds=\frac{1}{\tau^n\Gamma(n)}\int_0^x s^{n-1}e^{-\frac{s}{\tau}}\,ds=\frac{1}{\Gamma(n)}\int_0^{\frac{x}{\tau}} z^{n-1}e^{-z}\,dz=\boxed{\frac{\Gamma\left(\frac{x}{\tau};n\right)}{\Gamma(n)},}$$

kde $\Gamma\left(\frac{x}{\tau};n\right)$ je neúplná gama funkce argumentů $\frac{x}{\tau};n$. Pro střední hodnotu a rozptyl tohoto rozdělení dostaneme: $E\{\eta\}=n\tau$ a $\sigma^2(\eta)=n\tau^2$. Průběhy hustot a distribučních funkcí jsou demonstrovány na následujících obrázcích:







Gama rozdělení je velice užitečné, protože pomocí něj lze reprezentovat i jiná rozdělení, např. χ^2 rozdělení s $\boldsymbol{n}$ stupni volnosti je vlastně $\Gamma\left(\frac{n}{2},2\right)$. Numerický výpočet distribuční funkce může však někdy činit problémy (naznačte, úplná gama funkce je zobecněním faktoriálu, …).

Námět: Vyjádřete rozdělení $\chi_n = \sqrt{\sum_{i=1}^{n} \xi_i^2}$ ve tvaru gama rozdělení!

Námět: Pro $n \geq 40$ lze pro rozdělení $\Gamma(n,\tau)$ užít jeho asymptotickou normální aproximaci $N(n\tau, n\tau^2)$, analyzujte chybu a její vliv na používání v klasických statistických úlohách.

Studentovo[2] *t* rozdělení (Gossetovo)

Mějme náhodnou veličinu υ s rozdělením $N(0,1)$ a náhodnou veličinu $\chi_n = \sqrt{\sum_{i=1}^{n} \xi_i^2}$, rozdělení podílu $\boxed{t_n = \frac{\upsilon}{\sqrt{\frac{\sum_{i=1}^{n} \xi_i^2}{n}}}}$, kde $\sum_{i=1}^{n} \xi_i^2$ je náhodná proměnná χ_n^2 s n stupni volnosti.

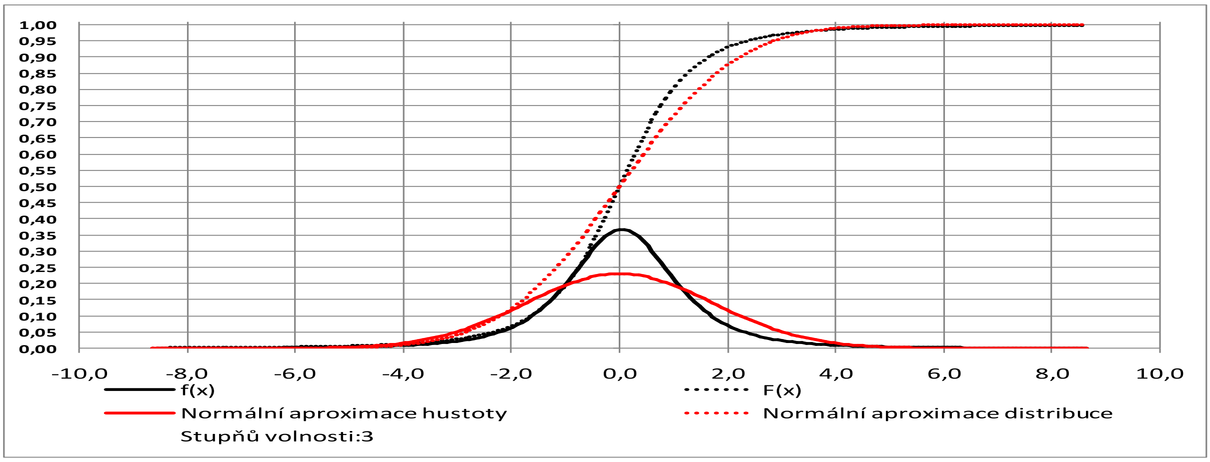
Náhodné proměnné υ a $\sum_{i=1}^{n} \xi_i^2$ jsou nezávislé. Hustota náhodné proměnné t_n je: $\boxed{f_{t_n}(x) = \frac{1}{\sqrt{n}} \frac{1}{B\left(\frac{1}{2},\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}}$, kde $B\left(\frac{1}{2},\frac{n}{2}\right)$ je „úplná“ beta funkce argumentů $\frac{1}{2},\frac{n}{2}$.

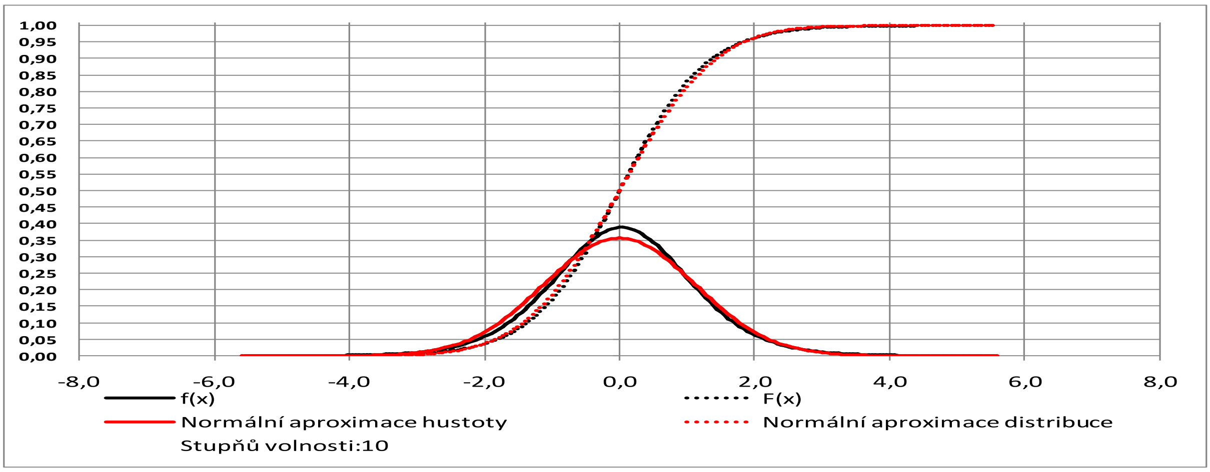
Odvození lze nalézt např. v Jaroslav Hátle, Jiří Likeš: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, str. 137-138. Distribuční funkce bude:

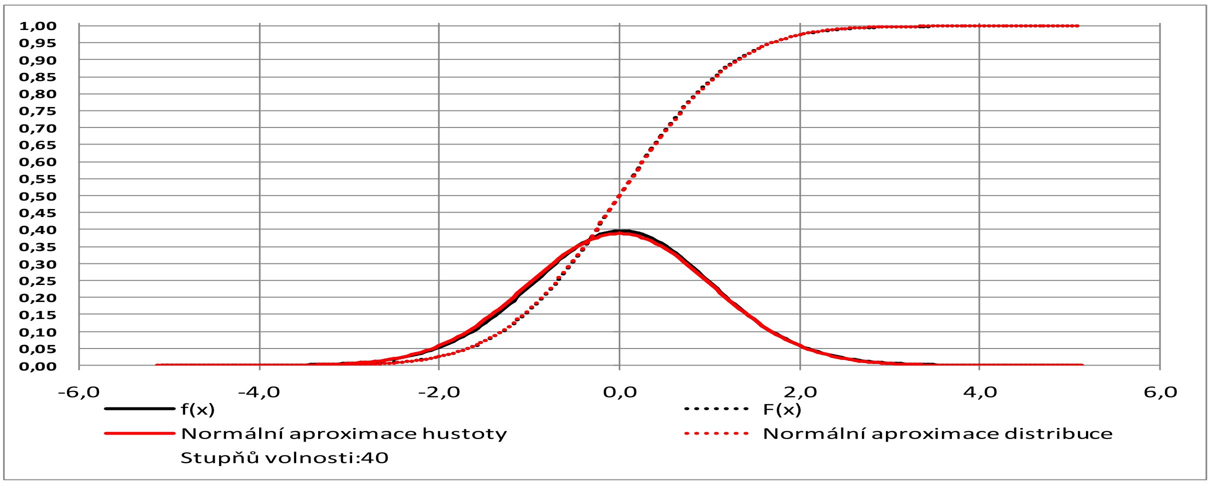
$$F_{t_n}(x) = \frac{1}{\sqrt{n}} \frac{1}{B\left(\frac{1}{2},\frac{n}{2}\right)} \int_{-\infty}^{x} \left(1+\frac{s^2}{n}\right)^{-\frac{n+1}{2}} ds .$$

Střední hodnota tohoto rozdělení je $0; n>1$, jinak neexistuje; rozptyl $\frac{n}{n-2}; n>2$, jinak neexistuje. Průběhy hustoty a distribuce jsou na následujících obrázcích.

[2] The derivation of the t-distribution was first published in 1908 by William Sealy Gosset, while he worked at a Guinness Brewery in Dublin. Due to proprietary issues, the paper was written under the pseudonym Student. The t-test and the associated theory became well-known through the work of R.A. Fisher, who called the distribution "Student's distribution". Zdroj http://en.wikipedia.org/wiki/, http://www.genealogy.math.ndsu.nodak.edu/index.php .







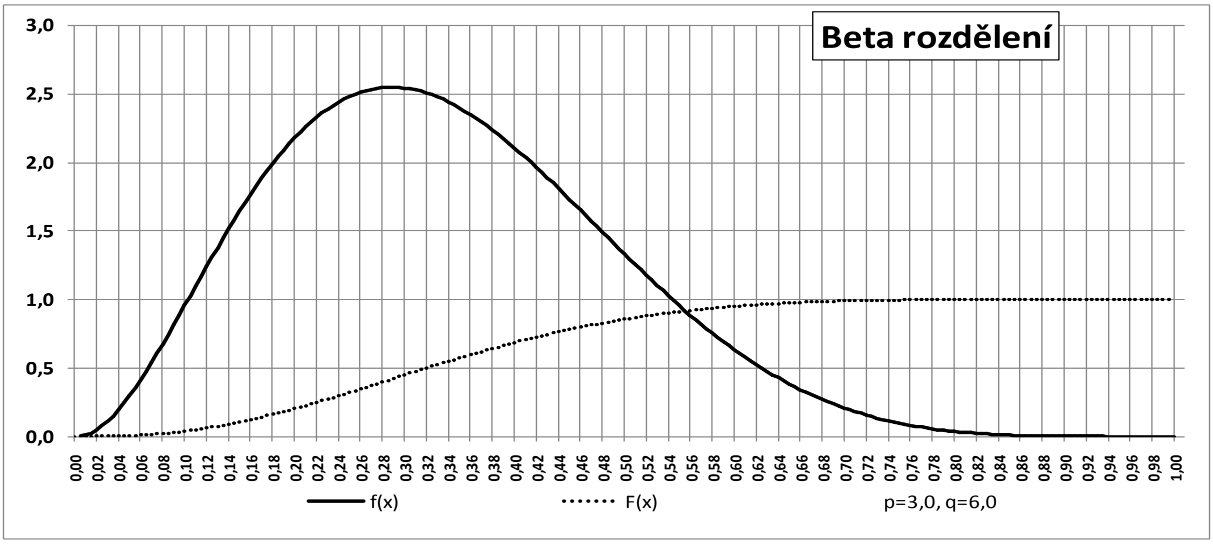
Námět: Zkoumejte oblast užití normální aproximace a její chybovost s ohledem na běžná statistická užití tohoto rozdělení.

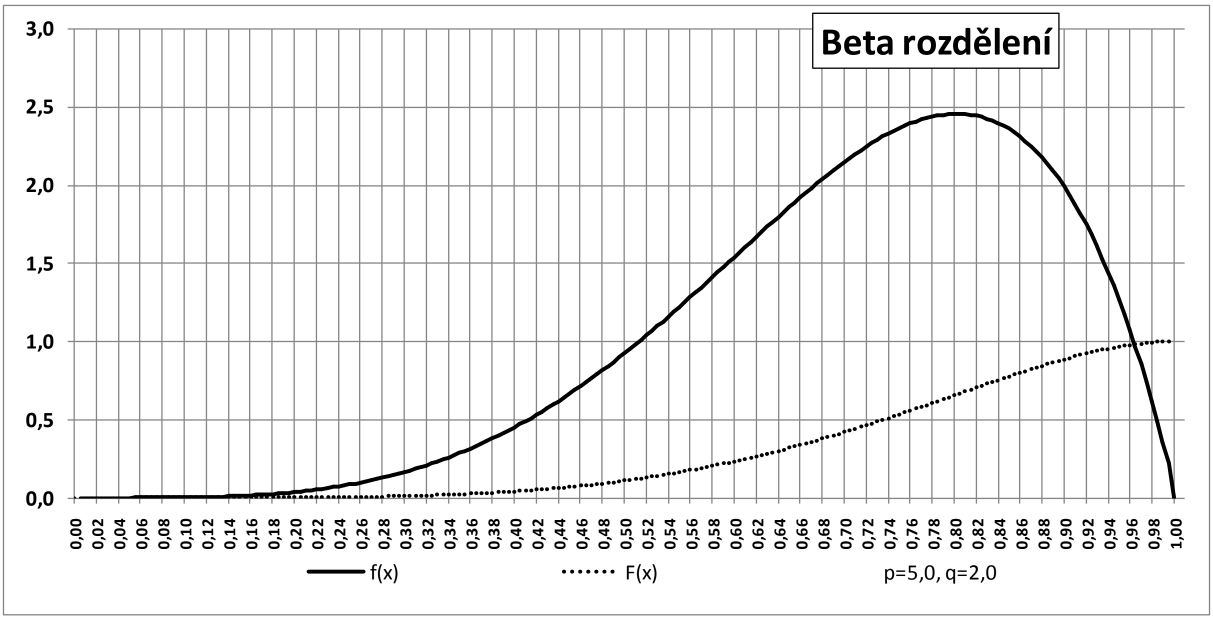
Beta rozdělení

Rozdělení náhodné proměnné s hustotou $\boxed{f_\beta(x)=\frac{1}{B(p,q)}x^{p-1}(1-x)^{q-1};0<x<1;\, p,q>0}$.

Střední hodnota a rozptyl takové náhodné proměnné jsou: $E\{\beta\}=\frac{p}{p+q};\sigma^2(\beta)=\frac{pq}{(p+q)^2(p+q+1)}$.

Distribuční funkce $\boxed{F_\beta(x)=\frac{1}{B(p,q)}\int_0^x s^{p-1}(1-s)^{q-1}ds;0<x<1;\, p,q>0}$, kde $B(p,q)=\int_0^1 s^{p-1}(1-s)^{q-1}ds$ je „úplná“ beta funkce s parametry p,q a $B(x;p,q)=\int_0^x s^{p-1}(1-s)^{q-1}ds$ je její „neúplná“ varianta. Opět následují průběhy hustot a distribučních funkcí.

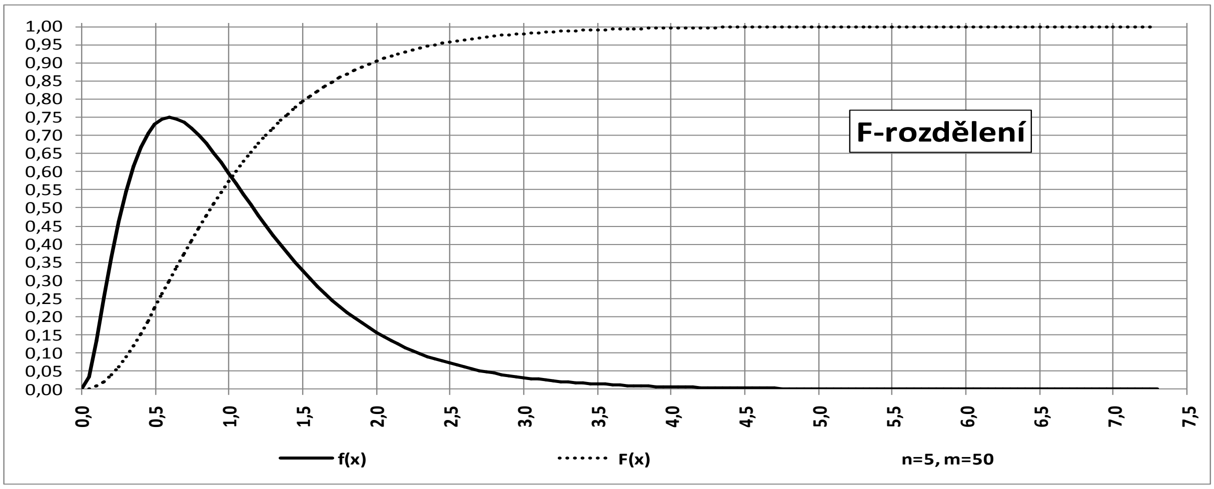


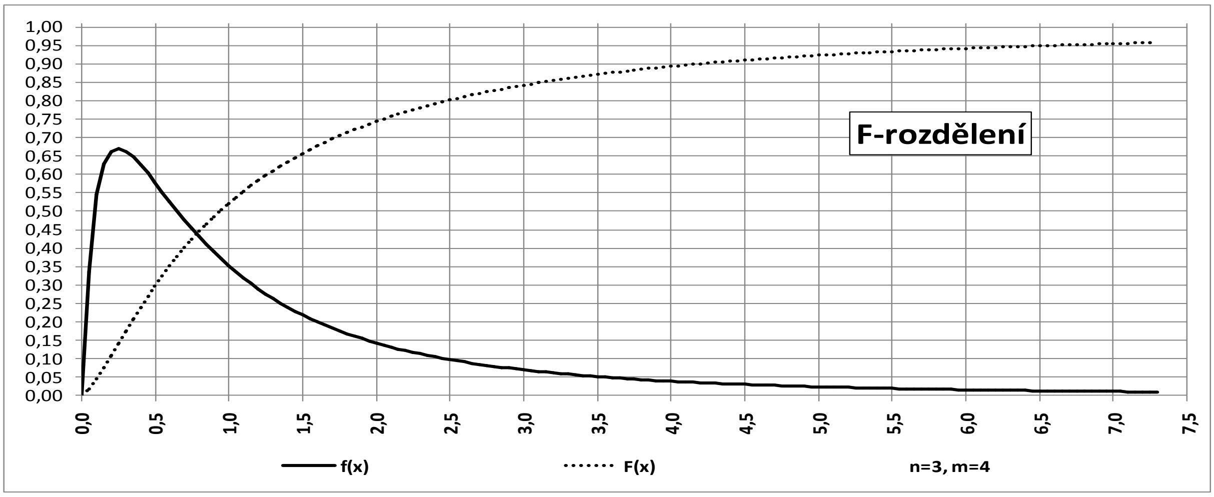


Beta rozdělení je užíváno pro popis náhodných proměnných s hodnotami v intervalu $(0,1)$.

Fisher Snedecorovo[3] *F* rozdělení

Mějme dvě nezávislé náhodné proměnné $\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$ a $\chi_m^2 = \sum_{i=1}^{m} \xi_i^2$. Rozdělení náhodné proměnné $\boxed{F_{n,m} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}}$ je nazýváno *F*-rozdělením s *n,m* stupni volnosti (evidentně se jedná o nezápornou náhodnou proměnnou). Její hustota je: $\boxed{f_{F_{n,m}}(x) = \frac{1}{B\left(\frac{n}{2},\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} \left(1+\frac{n}{m}x\right)^{-\frac{n+m}{2}} ; x>0}$ a distribuční funkce $\boxed{F_{F_{n,m}}(x) = \frac{1}{B\left(\frac{n}{2},\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \int_0^x s^{\frac{n}{2}-1} \left(1+\frac{n}{m}s\right)^{-\frac{n+m}{2}} ds;\ x>0}$. Jejich průběhy následují.





Námět: Spočtěte střední hodnotu a rozptyl tohoto rozdělení.

---

[3] George Waddel Snedecor (October 20, 1881 – February 15, 1974) was an American mathematician and statistician. He contributed to the foundations of analysis of variance, data analysis, experimental design, and statistical methodology. Snedecor's F distribution and the George W. Snedecor Award of the American Statistical Association are named after him. Zdroj http://en.wikipedia.org/wiki/, http://www.genealogy.math.ndsu.nodak.edu/index.php .

Doporučená a zdrojová literatura:

Jiří Reif	Metody matematické statistiky, ZČU v Plzni 2004
Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
Alfréd Rényi	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
	http://en.wikipedia.org/wiki/ http://www.genealogy.math.ndsu.nodak.edu/index.php

Některé užitečné pojmy a vztahy:

Beta funkce: $B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx \quad a,b>0$

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)};\quad B(a,b)=B(b,a);\quad B(a,b)=\frac{a-1}{a+b-1}B(a-1,b);\ B(a,b)=\frac{b-1}{a+b-1}B(a,b-1);$$

$$\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}=\frac{n!}{(k-1)!(n-k)!}\int_0^p x^{k-1}(1-x)^{n-k}dx = k\binom{n}{k}\int_0^p x^{k-1}(1-x)^{n-k}dx=$$

$$=k\binom{n}{k}\boldsymbol{B(p;k,n-k+1)},\quad \text{kde}\quad \boldsymbol{B(x;a,b)}=\int_0^x z^{a-1}(1-z)^{b-1}dz \quad a,b>0;\ 0\le x\le 1,$$

pro $a,b\ge 1$, celé(!):

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}=\frac{(a-1)!(b-1)!}{(a+b-1)!}=\frac{(a-1)!(b-1)!}{(a+b-1)(a+b-2)!}=$$

$$=\frac{1}{(a+b-1)\binom{a+b-2}{a-1}}=\frac{1}{(a+b-1)\binom{a+b-2}{b-1}}\Rightarrow$$

$$B(k+1,n-k+1)=\frac{1}{(n+1)\binom{n}{k}}$$

$$B(n-k+1,k+1)=\frac{1}{(n+1)\binom{n}{k}}$$

Neúplná beta funkce: $B(x;a,b)=\int_0^x z^{a-1}(1-z)^{b-1}dz \quad a,b>0;\ 0\le x\le 1$

Gama funkce: $\Gamma(x)=\int_0^\infty e^{-t}t^{x-1}dt$ pro $x>0$,

$$\Gamma(x+1)=x\Gamma(x);\Gamma(n)=(n-1)!;\Gamma(1)=\int_0^\infty e^{-t}dt=1;\Gamma(x)\Gamma(1-x)=\frac{\pi}{\sin(\pi x)};$$

$$\Gamma(x)\Gamma(x+\tfrac{1}{2})=\frac{\sqrt{\pi}}{2^{2x-1}}\Gamma(2x)\qquad \Rightarrow \Gamma(\tfrac{1}{2})\Gamma(1)=\sqrt{\pi}\Gamma(1)\Rightarrow\Gamma(\tfrac{1}{2})=\sqrt{\pi}$$

$$\Gamma(n)=(n-1)!$$

Neúplná gama funkce: $\Gamma(x;y)=\int_0^x e^{-t}t^{y-1}dt$ pro $x,y>0$,

Dirichletův integrál: $$\int\limits_{\forall x_i>0;\sum_{i=1}^{k} x_i=1} \prod_{i=1}^{k} x_i^{\alpha_i-1} dx_1 \ldots dx_k = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}; \forall \alpha_i > 0; i = 1,\ldots,k$$

Gaussův (Laplaceův) integrál: $$\int_0^\infty e^{-a^2t^2} dt = \frac{\sqrt{\pi}}{2a} \quad \text{pro } a > 0\text{,}$$

$$\int_0^\infty t^n e^{-at^2} dt = \frac{\Gamma\left(\frac{n+1}{2}\right)}{2a^{\frac{n+1}{2}}} \quad \text{pro } a > 0,\ n > -1$$

$$\int_0^\infty t^n e^{-\frac{t^2}{2}} dt = 2^{\frac{n-1}{2}} \Gamma\left(\frac{n+1}{2}\right) \quad \text{pro } n > -1$$