Pravděpodobnostní rozdělení matematické statistiky 2.

Rozdělení podílu dvou nezávislých kladných náhodných proměnných

Mějme dvě nezávislé náhodné ξ_1, ξ_2 proměnné definované na intervalu $(0, +\infty)$ s hustotami $f_{\xi_1}(x), f_{\xi_2}(x); x \notin \langle 0, +\infty) \Rightarrow f_{\xi_1}(x) = f_{\xi_2}(x) = 0$. Potom sdružená hustota obou náhodných proměnných je $f_{\xi_1 \xi_2}(x, y) = f_{\xi_1}(x) f_{\xi_2}(y)$.

Pro náhodnou proměnnou $\eta = \dfrac{\xi_1}{\xi_2}$, dostaneme její distribuční funkci jako pravděpodobnost:

$\xi_2 > 0$!

$$F_\eta(z) = P(\eta < z) = \boxed{P\left(\frac{\xi_1}{\xi_2} < z\right) = P(\xi_1 < z\xi_2)} = \iint_{x<zy} f_{\xi_1\xi_2}(x, y)\,dxdy = \iint_{x<zy} f_{\xi_1}(x) f_{\xi_2}(y)\,dxdy =$$

$$= \int_0^{+\infty}\int_0^{yz} f_{\xi_1}(x) f_{\xi_2}(y)\,dxdy = \int_0^{+\infty} f_{\xi_2}(y)\left(\int_0^{yz} f_{\xi_1}(x)\,dx\right)dy = \int_0^{+\infty} f_{\xi_2}(y)\left(F_{\xi_1}(yz) - F_{\xi_1}(0)\right)dy =$$

$$= \int_0^{+\infty} f_{\xi_2}(y)\left(F_{\xi_1}(yz) - F_{\xi_1}(0)\right)dy = \int_0^{+\infty} f_{\xi_2}(y) F_{\xi_1}(yz)\,dy - F_{\xi_1}(0)\int_0^{+\infty} f_{\xi_2}(y)\,dy = \int_0^{+\infty} f_{\xi_2}(y) F_{\xi_1}(yz)\,dy - F_{\xi_1}(0).$$

Shrnuto: $\boxed{F_\eta(z) = \int_0^{+\infty} f_{\xi_2}(y) F_{\xi_1}(yz)\,dy}$

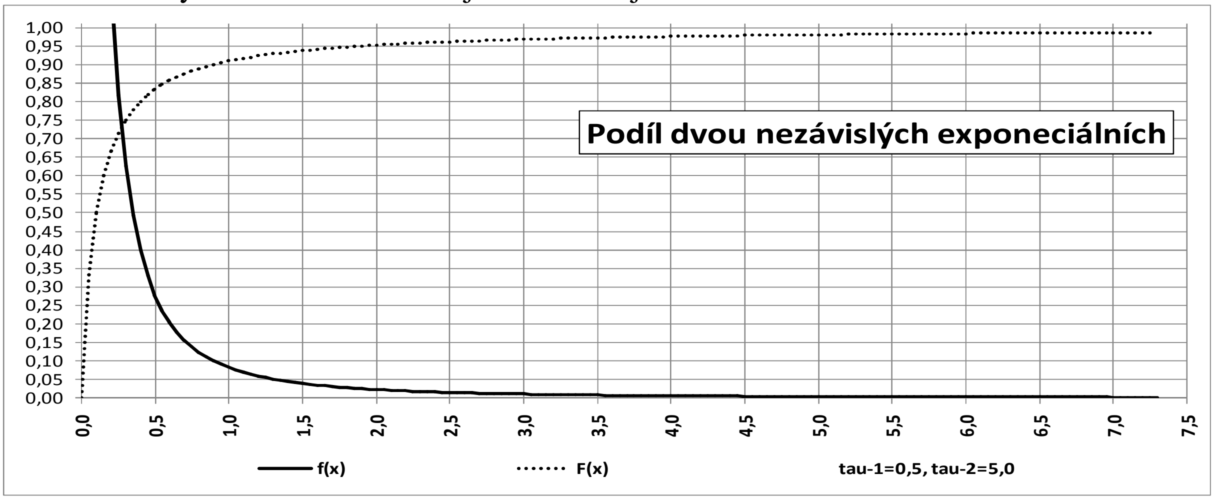
Příklad: Podíl dvou nezávislých exponenciálních náhodných proměnných.

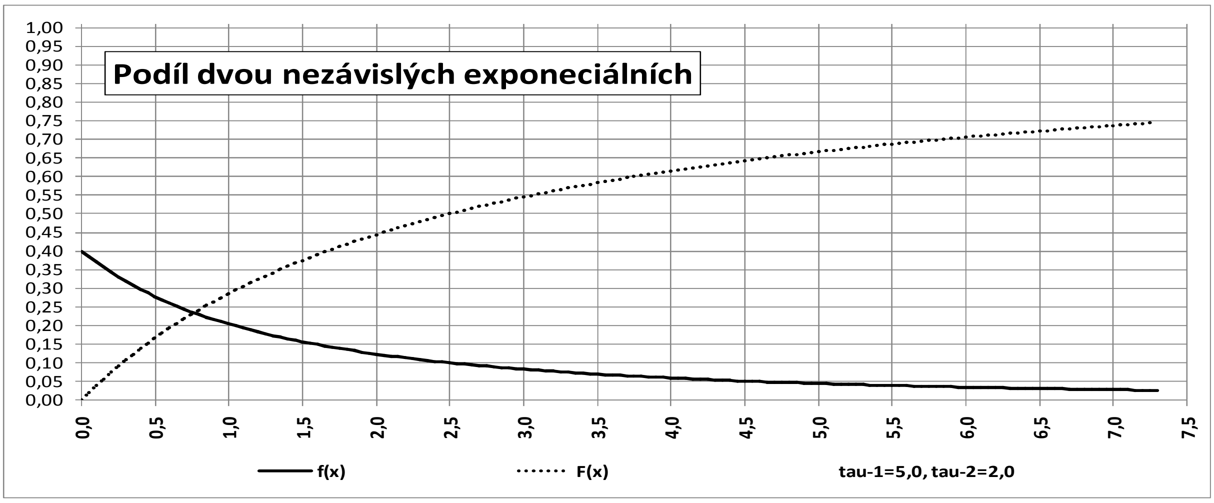
$$f_{\xi_1}(x) = \frac{1}{\tau_1} e^{-\frac{x}{\tau_1}}; f_{\xi_2}(x) = \frac{1}{\tau_2} e^{-\frac{x}{\tau_2}}; x \notin \langle 0, +\infty) \Rightarrow f_{\xi_1}(x) = f_{\xi_2}(x) = 0, \text{ potom:}$$

$$F_\eta(z) = \int_0^{+\infty} f_{\xi_2}(y) F_{\xi_1}(yz)\,dy = \int_0^{+\infty} \frac{1}{\tau_2} e^{-\frac{y}{\tau_2}}\left(1 - e^{-\frac{yz}{\tau_1}}\right)dy = 1 - \frac{1}{\tau_2}\int_0^{+\infty} e^{-y\left(\frac{1}{\tau_2} + \frac{z}{\tau_1}\right)}dy = 1 - \frac{1}{\tau_2}\frac{1}{\frac{1}{\tau_2} + \frac{z}{\tau_1}},$$

shrnuto: $\boxed{F_\eta(x) = 1 - \dfrac{1}{1 + \dfrac{\tau_2}{\tau_1}x} \Rightarrow f_\eta(x) = \dfrac{\tau_2}{\tau_1}\dfrac{1}{\left(1 + \dfrac{\tau_2}{\tau_1}x\right)^2}}$. Medián tohoto rozdělení je: $x_{0,5} = \dfrac{\tau_1}{\tau_2}$.

Průběh hustoty a distribuční funkce je na následujícím obrázku:





Námět: Posuďte souvislost tohoto rozdělení s Paretovým.

Centrální limitní věta[1]

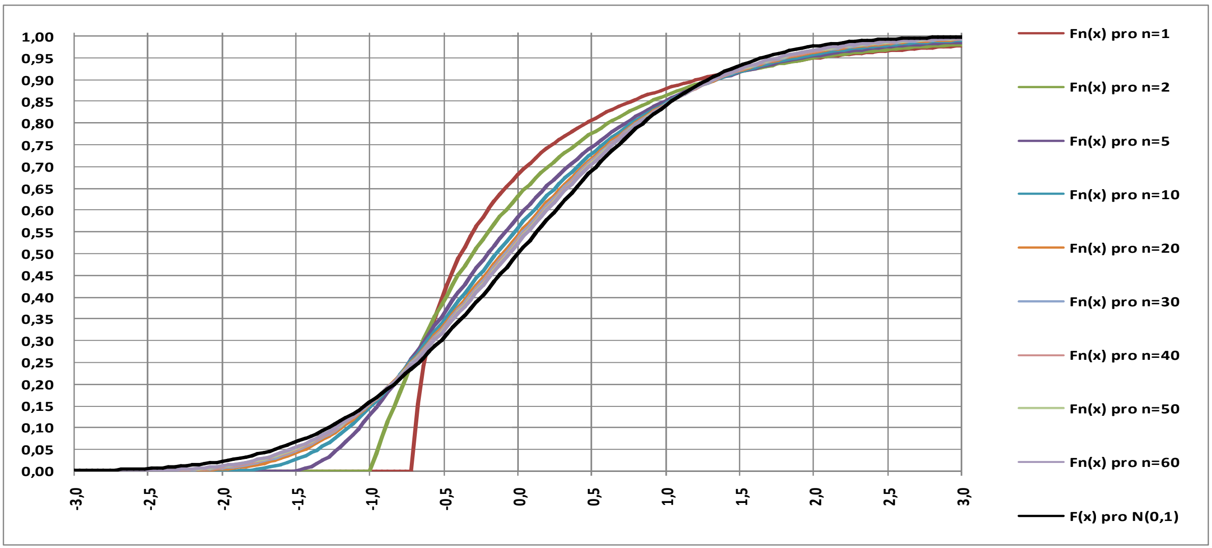
Nechť $\xi_1, \xi_2, \ldots \xi_n, \ldots$ jsou nezávislé náhodné veličiny se stejnou distribuční funkcí a nechť existují $E = E\{\xi_i\}$, $D = \sigma(\xi_i); i = 1,2,\ldots$. Položme $\eta_n = \sum_{i=1}^{n} \xi_i$ a $\eta_n^* = \dfrac{\eta_n - E\{\eta_n\}}{\sigma(\eta_n)}$ a $F_n(x) = F_{\eta_n^*}(x)$ nechť je distribuční funkce náhodné proměnné η_n^*. Potom: $\lim_{n \to +\infty} F_n(x) = \Phi(x)$, pro $-\infty < x < +\infty$. Konvergence nemusí být stejnoměrná. Věta je převzata z Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, str. 374-375, kde lze nalézt i její důkaz.

Volně řečeno: Distribuční funkce centrovaného a normovaného součtu nezávislých stejně rozdělených náhodných proměnných (s existující střední hodnotou rozptylem) konverguje k distribuční funkci centrovaného a normovaného normálního rozdělení (*N(0,1)*).

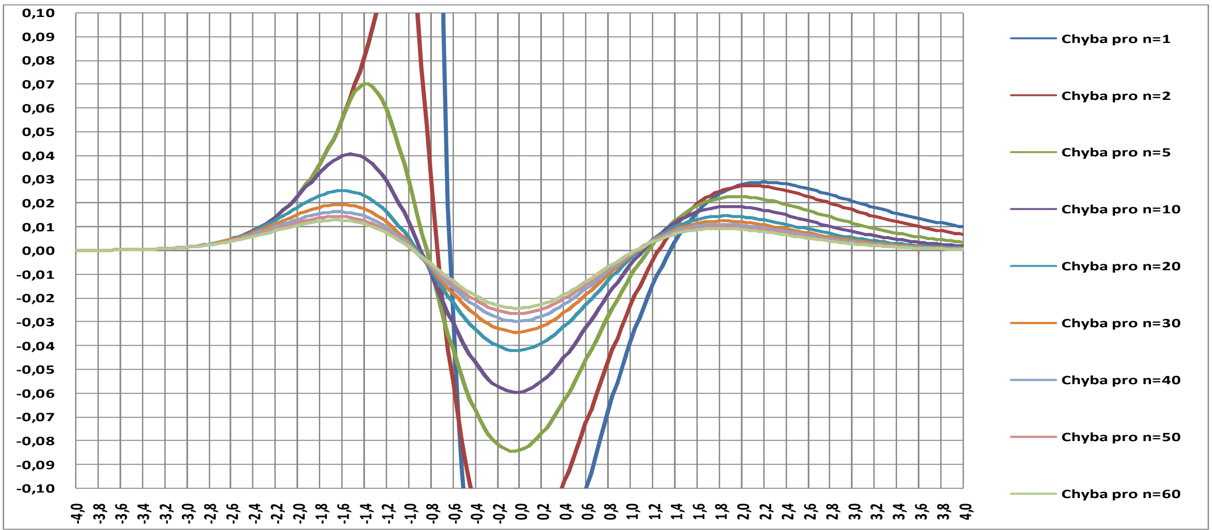
[1] Centrální limitní větu lze dokázat pro mnohem slabší předpoklady. Viz např. Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, str. 374-382.

Demonstrace na příkladu χ_n^2 rozdělení.

χ_n^2 je náhodná proměnná tvořená součtem n-nezávislých náhodných proměnných, z nichž každá je tvořena kvadrátem náhodné proměnné s *N(0,1)* rozdělením. Sčítanci mají střední hodnotu a rozptyl (dokažte). Náhodná proměnná $\eta_n^* = \dfrac{\chi_n^2 - n}{\sqrt{2n}}$ je pak centrovaná a normovaná. Jestliže náhodná proměnná χ_n^2 má distribuční funkci $F_n(x)$, pak náhodná proměnná η_n^* má distribuční funkci $F_n\left(\sqrt{2n}\ x + n\right)$. Následující obrázky demonstrují konvergenci k $\Phi(x)$.



Průběhy distribučních funkcí náhodných proměnných η_n^* "limitní" $\Phi(x)$.



Průběhy chyb „limitní" aproximace, tj. rozdílů $\Phi(x) - F_{\eta_n^*}(x)$.

Námět: Jedná se o stejnoměrnou konvergenci?

Některé souvislosti a vztahy

Použitá označení:

χ_n^2 Chí kvadrát rozdělení s ***n*** stupni volnosti a hustotou: $f_{\chi_n^2}(x) = \dfrac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}; x \geq 0$

$G(n,\tau)$ Gama rozdělení s hustotou: $f_\eta(x) = \dfrac{x^{n-1}e^{-\frac{x}{\tau}}}{\tau^n\Gamma(n)}; x \geq 0$

t_n ***t*** rozdělení s ***n*** stupni volnosti s hustotou: $f_{t_n}(x) = \dfrac{1}{\sqrt{n}}\dfrac{1}{B\left(\frac{1}{2},\frac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$

$Be(p,q)$ beta rozdělení s hustotou: $f_\beta(x) = \dfrac{1}{B(p,q)}x^{p-1}(1-x)^{q-1}; 0 < x < 1; p,q > 0$

$F_{n,m}$ ***F*** rozdělení s ***n,m*** stupni volnosti s hustotou: $f_{F_{n,m}}(x) = \dfrac{1}{B\left(\frac{n}{2},\frac{m}{2}\right)}\left(\dfrac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}\left(1+\dfrac{n}{m}x\right)^{-\frac{n+m}{2}}; x>0$

- $\chi_n^2 \equiv G\left(\frac{n}{2},2\right)$
- Hustota rozdělení χ_n (evidentně jak χ_n, tak i χ_n^2 jsou nezáporné náhodné proměnné). Nechť náhodná proměnná ξ má χ_n^2 rozdělení, zajímá nás rozdělení náhodné veličiny $\eta = \sqrt{\xi} = \sqrt{\chi_n^2}$. Obecně $F_\eta(x) = P(\eta < x) = P(\sqrt{\xi} < x) = P(\xi < x^2) = \ = F_\xi(x^2)$. Odtud

  $$f_\eta(x) = \frac{d}{dx}F_\eta(x) = 2xf_\xi(x^2) = 2x\frac{x^{n-2}e^{-\frac{x^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} = \frac{x^{n-1}e^{-\frac{x^2}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}.$$

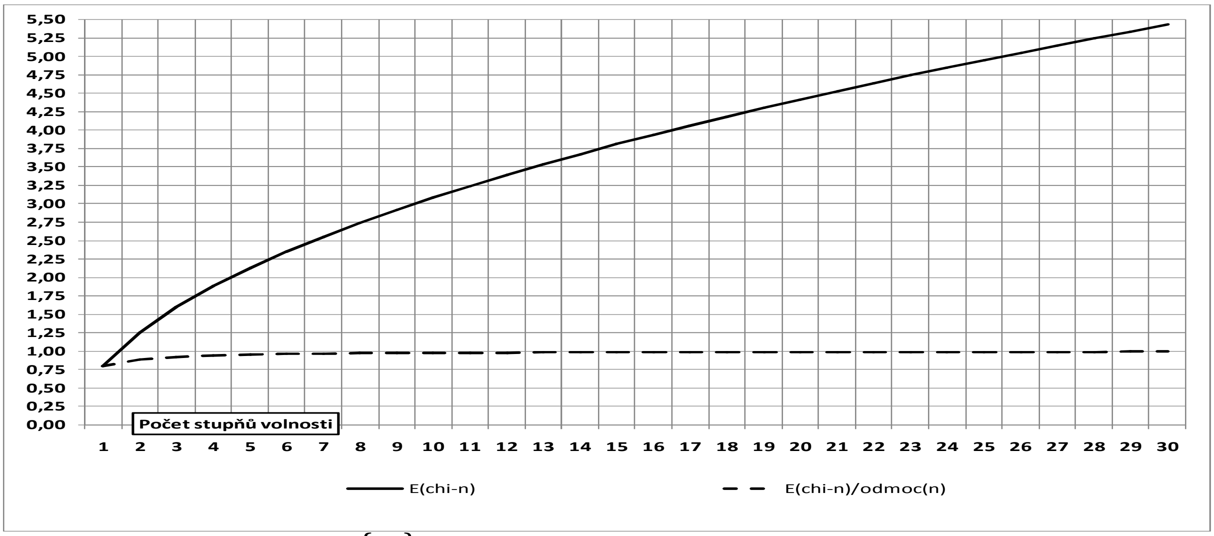
  Shrnuto: $f_\eta(x) = \dfrac{x^{n-1}e^{-\frac{x^2}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}$. $\quad F_{\chi_n}(x) = \displaystyle\int_0^x \frac{z^{n-2}e^{-\frac{z^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}dz = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^x z^{n-2}e^{-\frac{z^2}{2}}dz =$

  $$= \frac{1}{2^{\frac{3}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^{\frac{x^2}{2}} y^{\frac{n-3}{2}}e^{-y}dy = \frac{\Gamma\left(\frac{x^2}{2};\frac{n-1}{2}\right)}{2^{\frac{3}{2}}\Gamma\left(\frac{n}{2}\right)},$$

  kde $\Gamma(x;y) = \int_0^x e^{-t}t^{y-1}dt$ pro $x, y > 0$. Pro náhodnou proměnnou χ_n je užitečná její střední hodnota:

  $$E\{\chi_n\} = \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}\int_0^{+\infty} xx^{n-1}e^{-\frac{x^2}{2}}dx = \frac{2^{\frac{n-1}{2}}\Gamma\left(\frac{n+1}{2}\right)}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} = 2^{\frac{1}{2}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

  Shrnuto: $E\{\chi_n\} = \dfrac{2^{\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$.



Průběh $E\{\chi_n\}$ a $E\{\chi_n\}/\sqrt{n}$ v závislosti na $n =$ počtu stupňů volnosti.

Námět: na základě inspirace předchozím obrázkem dokažte asymptotické chování $\dfrac{E\{\chi_n\}}{\sqrt{n}} \underset{n\to+\infty}{\to} 1$.

- Mějme náhodný výběr $(x_1, x_2, \dots, x_n)$, tj. stejně rozdělené a nezávislé náhodné veličiny z rozdělení $N(\mu, \sigma^2)$, potom transformovaný náhodný výběr $\left(\dfrac{x_1-\mu}{\sigma}, \dfrac{x_2-\mu}{\sigma}, \dots, \dfrac{x_n-\mu}{\sigma}\right)$ je výběrem z $N(0,1)$. Náhodná proměnná $\eta_1 = \sum_{i=1}^{n}\left(\dfrac{x_i-\mu}{\sigma}\right)^2$ má χ_n^2 rozdělení a $\eta_2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(\dfrac{x_i-\mu}{\sigma}\right)^2$ bude mít distribuční funkci $F_{\eta_2}(x) = P(\eta_2 < x) = P\left(\dfrac{\eta_1}{n} < x\right) = P(\eta_1 < nx) = F_{\chi_n^2}(nx)$
- Mějme náhodný výběr $(x_1, x_2, \dots, x_n)$ z rozdělení $N(\mu, \sigma^2)$, nechť $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$. Náhodná proměnná $\eta_3 = \sum_{i=1}^{n}\left(\dfrac{x_i-\bar{x}}{\sigma}\right)^2$ má χ_{n-1}^2 rozdělení neboť platí $0 = \sum_{i=1}^{n}(x_i - \bar{x})$ (dokažte!) a tedy libovolnou z ***n*** proměnných lze vyjádřit pomocí průměru a zbývajících ***n-1*** proměnných (podrobně rozepište a odvoďte). Tento příklad je speciálním případem obecnější situace, která je řešena tzv. Cochranovou větou (Jaroslav Hátle, Jiří Likeš: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, str. 136-137).
- Mějme náhodný výběr $(x_1, x_2, \dots, x_n)$ z rozdělení $N(\mu, \sigma^2)$. Náhodná veličina $\eta_4 = \left(\dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$ má χ_1^2 rozdělení.

- Náhodná veličina $t = \dfrac{\overline{x} - \mu}{\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}$ má Studentovo ***t***-rozdělení s ***n-1*** stupni volnosti.
- Platí $\chi^2_{\sum_{i=1}^{k} k_i} \overset{\text{Prob}}{=} \sum_{i=1}^{k} \chi^2_{k_i}$. Je to důsledek definice χ^2_n rozdělení.
- Další vztahy mezi rozděleními lze nalézt v C. Radhakrishna Rao: Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978, str. 212-214.

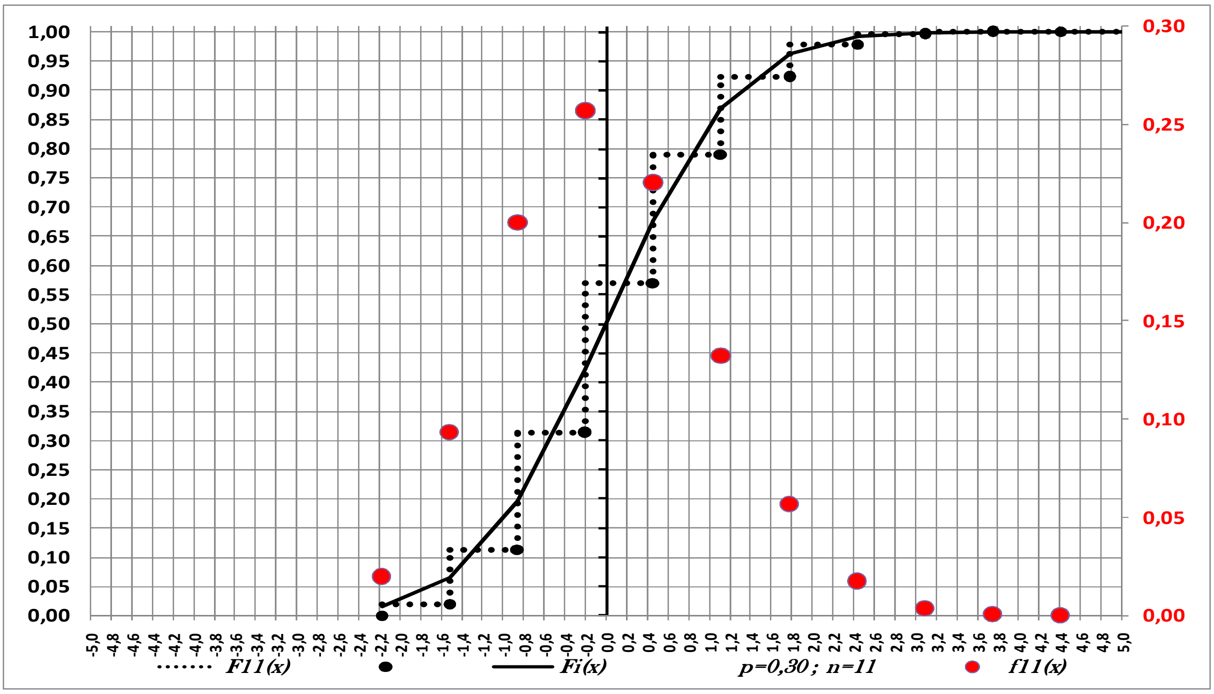
- Vztah mezi Poissonovým rozdělením a χ^2_n rozdělením. Nechť má náhodná veličina ξ Poissonovo rozdělení s parametrem λ, pak platí $P(\xi \le x) = \sum_{i=0}^{x} e^{-\lambda}\dfrac{\lambda^i}{i!} = 1 - F_{\chi^2_{2(x+1)}}(2\lambda)$. Důkaz postupnou integrací per-partes lze nalézt v Jaroslav Hátle, Jiří Likeš: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, str. 141.

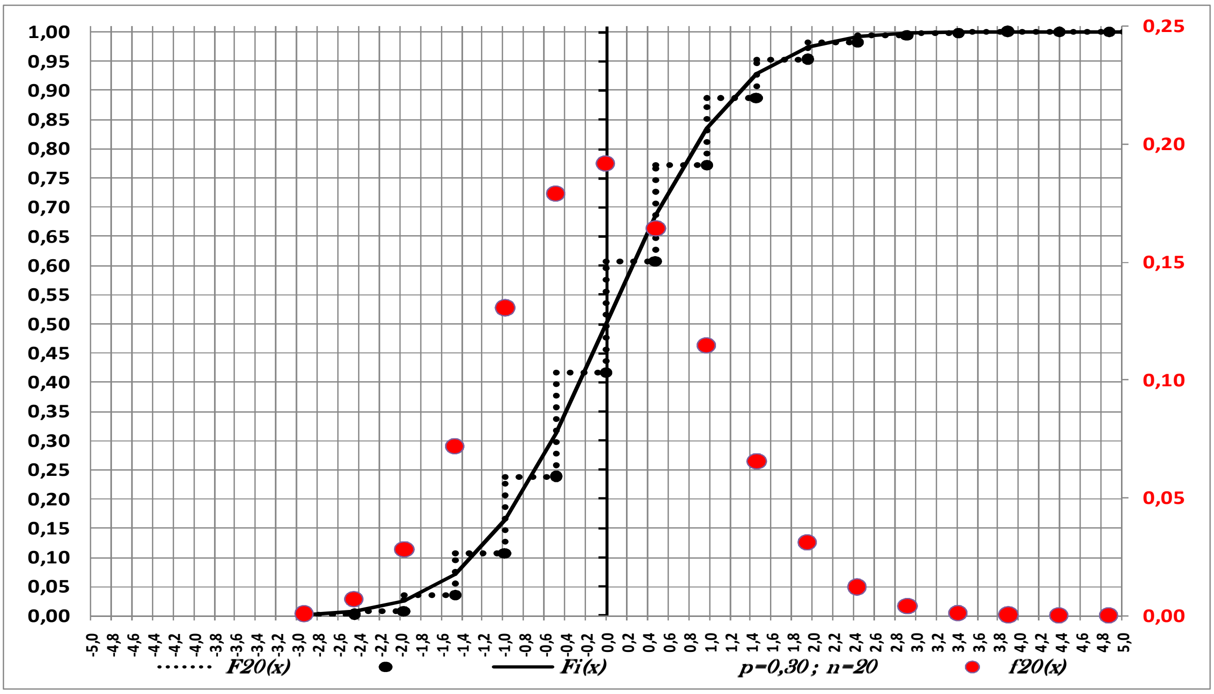
- Vztah mezi binomickým rozdělením a F rozdělením. Nechť má náhodná veličina ξ binomické rozdělení s parametry p, n, pak platí $P(\xi \le x) = \sum_{i=0}^{x}\binom{n}{i} p^i (1-p)^{n-i} = F_{F_{n_1,n_2}}\left(\dfrac{x+1}{n-x}\dfrac{1-p}{p}\right)$, kde $n_1 = 2(n-x)$ a $n_2 = 2(x+1)$. Důkaz opět lze nalézt v Jaroslav Hátle, Jiří Likeš: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, str. 141. Obě předcházející tvrzení platí pro x celé nezáporné. **Námět:** Pro oba předchozí příklady nakreslete průběh distribučních funkcí odpovídajících diskrétních rozdělení a jim přiřazených spojitých rozdělení.

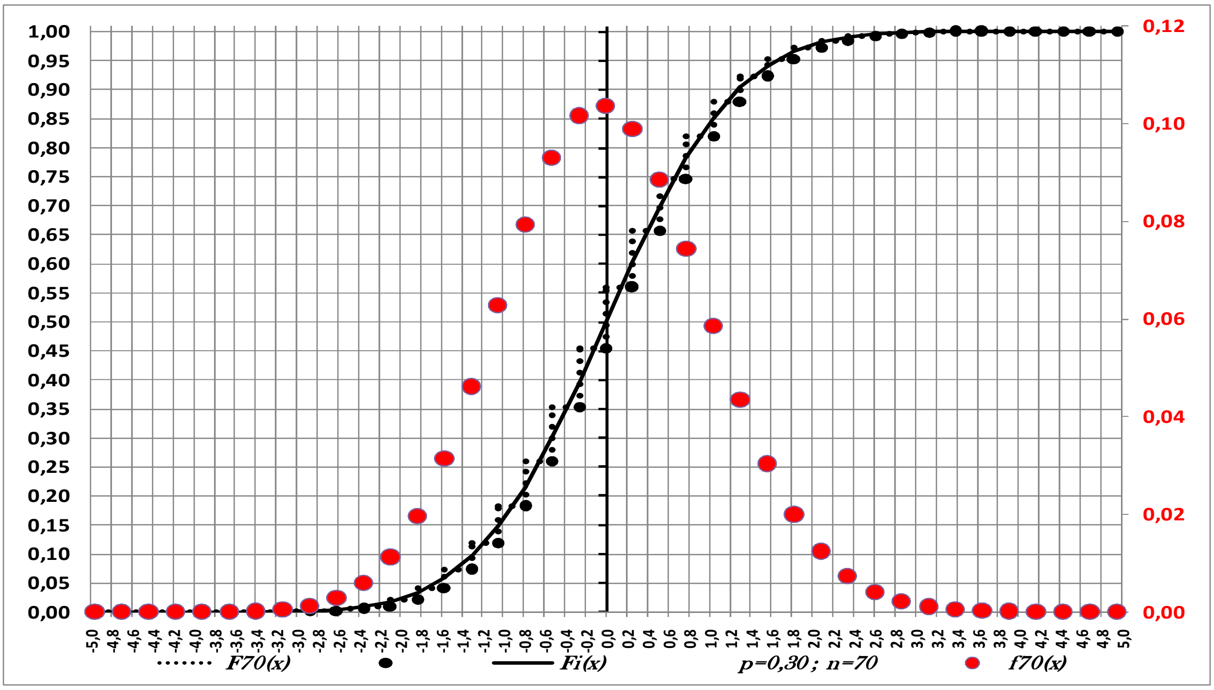
- Vztah mezi binomickým a Poissonovým rozdělením: Vytvořující funkce pravděpodobností pro binomické rozdělení s parametry p, n je $G_{Bi(p,n)}(z) = (pz + (1-p))^n$. Pokud zvolíme $\boldsymbol{p} = \dfrac{\boldsymbol{\lambda}}{\boldsymbol{n}}; \boldsymbol{\lambda} \le \boldsymbol{n}$, dostaneme $G_{Bi\left(\frac{\lambda}{n},n\right)}(z) = \left(\dfrac{\lambda z}{n} + \left(1 - \dfrac{\lambda}{n}\right)\right)^n = \left(1 + \dfrac{\lambda(z-1)}{n}\right)^n$ a odtud $\lim_{n\to+\infty} G_{Bi\left(\frac{\lambda}{n},n\right)}(z) = \lim_{n\to+\infty}\left(1 + \dfrac{\lambda(z-1)}{n}\right)^n = e^{\lambda(z-1)}$. Ale to je vytvořující funkce pravděpodobností Poissonova rozdělení s parametrem λ. Odtud také $\lim_{n\to+\infty}\binom{n}{k}\left(\dfrac{\lambda}{n}\right)^k\left(1 - \dfrac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\dfrac{\lambda^k}{k!}$. Tedy pokud s rostoucím ***n*** klesá pravděpodobnost úspěchu $p = \dfrac{\lambda}{n}$, binomické pravděpodobnosti konvergují k Poissonovým. Detailnější důkaz lze nalézt v Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, str. 133-135. Ve zde uvedeném důkazu chybí důkaz a předpoklady toho, že z konvergence vytvořujících funkcí pravděpodobnosti plyne i konvergence vlastních pravděpodobností.

Poznámka: Pokud považujeme p za pravděpodobnost úspěchu a pozorovaná náhodná proměnná je počet úspěchů, pak se uvedená konvergence uplatní, pokud je pravděpodobnost úspěchu malá a s rostoucím n ještě klesá.

- Vztah mezi binomickým a normálním rozdělením: Binomická náhodná proměnná je tvořena součtem n alternativních náhodných proměnných. Potom podle výše uvedené varianty centrální limitní věty platí, že distribuční funkce normované a centrované binomické náhodné proměnné konverguje s rostoucím n k funkci $\Phi(x)$. Přesněji: $\eta_n = \frac{\sum_{i=1}^{n} \xi_i - np}{\sqrt{np(1-p)}} \Rightarrow \lim_{n \to +\infty} F_{\eta_n}(x) = \Phi(x)$, typické průběhy jsou na následujících obrázcích.







Námět: na základě výše uvedených obrázků odvoďte a spočtěte velikost maximální chyby takové asymptotické aproximace. Je taková maximální chyba určující pro běžné statistické aplikace?

Některá neklasická výpočetní schémata pro

Binomické pravděpodobnosti

$$p_i = \binom{n}{i} p^i (1-p)^{n-i} \quad i=0,\dots,n \qquad p_{i+1} = \binom{n}{i+1} p^{i+1}(1-p)^{n-i-1} = \binom{n}{i+1} \frac{p}{1-p} p^i (1-p)^{n-i} =$$

$$= \frac{n-i}{i+1}\frac{p}{1-p}\binom{n}{i} p^i (1-p)^{n-i} = \frac{n-i}{i+1}\frac{p}{1-p} p_i; \quad p_0 = (1-p)^n \Leftrightarrow \lg p_0 = n \lg(1-p).$$

Shrnuto: $p_{i+1} = \dfrac{n-i}{i+1}\dfrac{p}{1-p} p_i; \quad p_0 = (1-p)^n$

Poissonovy pravděpodobnosti

$$p_i = e^{-\lambda}\frac{\lambda^i}{i!},\ i=0,1,\dots \qquad p_{i+1} = e^{-\lambda}\frac{\lambda^{i+1}}{(i+1)!} = \frac{\lambda}{(i+1)} e^{-\lambda}\frac{\lambda^i}{i!} = \frac{\lambda}{(i+1)} p_i; \quad p_0 = e^{-\lambda}$$

Shrnuto: $p_{i+1} = \dfrac{\lambda}{(i+1)} p_i; \quad p_0 = e^{-\lambda}$

Geometrické pravděpodobnosti

$$p_i = (1-q)q^i,\ i=0,1,\dots \qquad p_{i+1} = (1-q)q^{i+1} = q p_i; \quad p_0 = 1-q$$

Shrnuto: $p_{i+1} = qp_i; \quad p_0 = 1-q$

Námět: Odvoďte obdobná rekurzivní schémata pro negativně binomické a hypergeometrické pravděpodobnosti.

Panjerova rekurze[2]

V uvedeném článku je uveden rekurzivní vztah , $\boldsymbol{p_i = p_{i-1}\left(a + \frac{b}{i}\right); \quad i = 1,2,...,}$ který platí pro třídu diskrétních rozdělení, obsahujících:

1. Poissonovo rozdělení $p_i = e^{-\lambda}\frac{\lambda^i}{i!}$, $i=0,1,\ldots$: $\qquad a = 0; b = \lambda; p_0 = e^{-\lambda}$
2. Binomické rozdělení $p_i = \binom{n}{i} p^i (1-p)^{n-i}$ $i=0,\ldots,n$; $\qquad a = \frac{-p}{1-p}; b = \frac{(n+1)p}{1-p}; p_0 = (1-p)^n$
3. Negativně binomické rozdělení $p_i = \binom{i+r-1}{r-1} p^r (1-p)^i; r \geq 1; i=0,1,\ldots$.

   Námět: určete pro toto rozdělení $a; b; p_0$.
4. Geometrické rozdělení $p_i = (1-q)q^i$; $i=0,1,\ldots$. ***Námět:*** určete pro toto rozdělení $a; b; p_0$.

Námět: Pro všechna uvedená rozdělení detailně prokažte, že do uvedené třídy patří. K důkazu, že do ní nepatří jiná rozdělení, viz: B. SUNDT and W. S. JEWELL: Further Results of Recursive Evaluation of Compound Distributions. ASTIN Volume 12, No. 1, December 1982, Volume 12, No. 1. http://www.casact.org/library/astin/vol12no1/

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://www.casact.org/library/astin/vol12no1/

[2] HARRY H. PANJER (University of Waterloo, Ontario, Canada): RECURSIVE EVALUATION OF A FAMILY OF COMPOUND DISTRIBUTIONS. Astin Bulletin 12 (1981) 22-26. http://www.casact.org/library/astin/vol12no1/