Poměrové statistiky – příklad aparátu a motivace

Rozdělení podílu dvou nezávislých kladných náhodných proměnných
Mějme dvě nezávislé náhodné ξ_1, ξ_2 proměnné definované na intervalu $(0, +\infty)$ s hustotami $f_{\xi_1}(x), f_{\xi_2}(x); x \notin \langle 0, +\infty) \Rightarrow f_{\xi_1}(x) = f_{\xi_2}(x) = 0$ a splňující podmínku $F_{\xi_i}(0) = 0; i = 1,2$. 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:

$$\boxed{\xi_2 > 0\ !}$$

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

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

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

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

Obecně ale: $F_\eta(z) = \iint_{x/y<z} f_{\xi_2}(y) f_{\xi_1}(x)\,dxdy$ a $f_\eta(z) = \int_{-\infty}^{+\infty} |y| f_{\xi_2}(y) f_{\xi_1}(yz)dy$

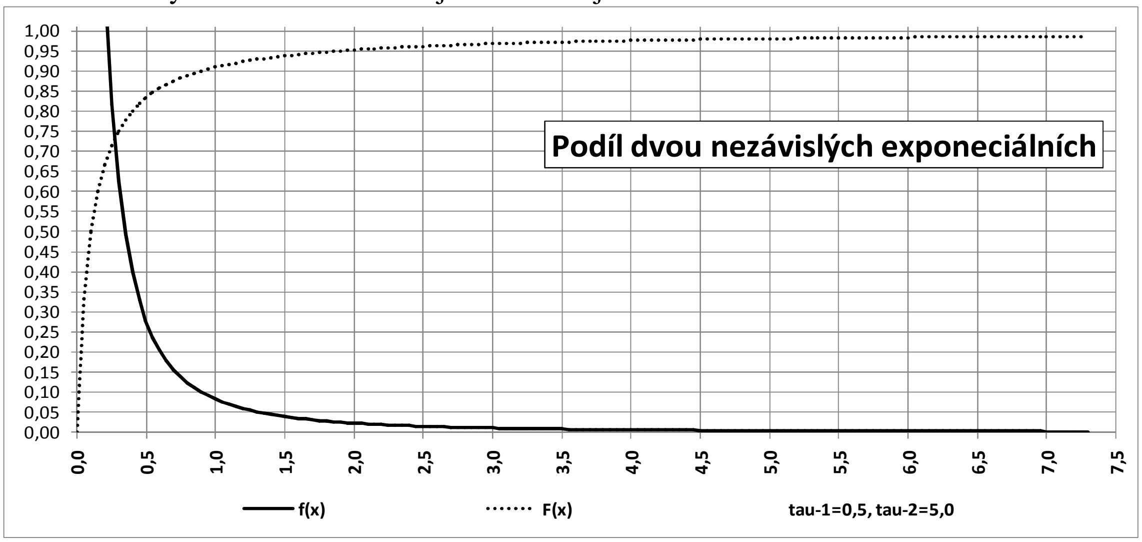
Příklad: Podíl dvou nezávislých exponenciálních náhodných proměnných.

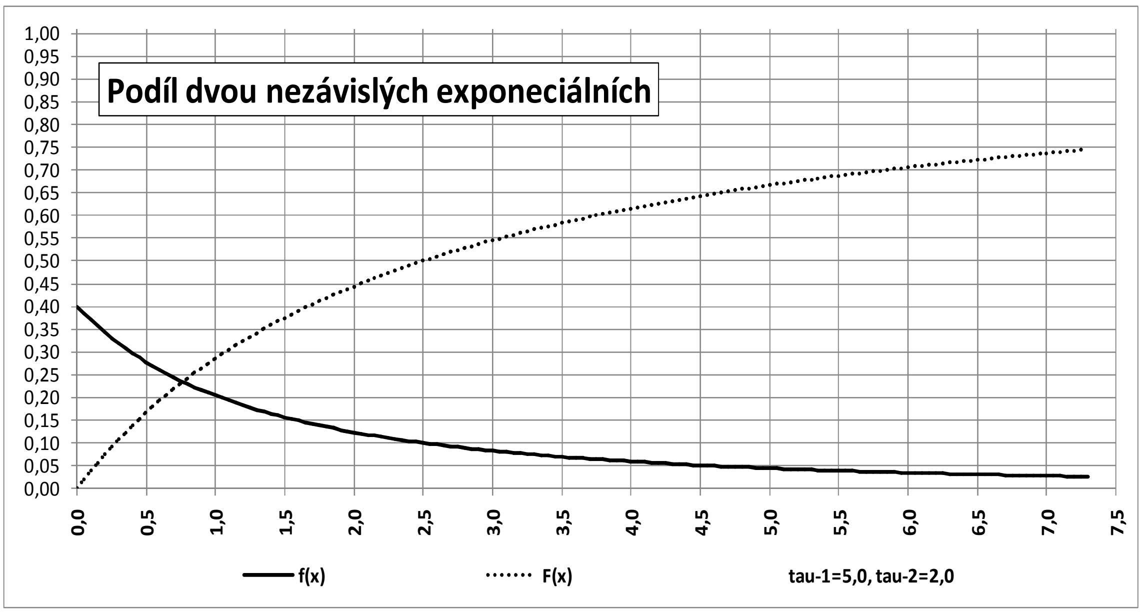
$f_{\xi_1}(x) = \dfrac{1}{\tau_1} e^{-\frac{x}{\tau_1}}; f_{\xi_2}(x) = \dfrac{1}{\tau_2} e^{-\frac{x}{\tau_2}}; x \notin \langle 0, +\infty) \Rightarrow f_{\xi_1}(x) = f_{\xi_2}(x) = 0$, potom:

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

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.

Příklad: Rozdělení náhodné proměnné, která vznikla podílem dvou centrovaných a normovaných, normálně rozdělených náhodných proměnných. Tedy $f_{\xi_1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; f_{\xi_1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; x \in (-\infty, +\infty); \eta = \frac{\xi_1}{\xi_2}$. Pak: $f_\eta(z) = \int_{-\infty}^{+\infty} |y| f_{\xi_2}(y) f_{\xi_1}(yz) dy =$

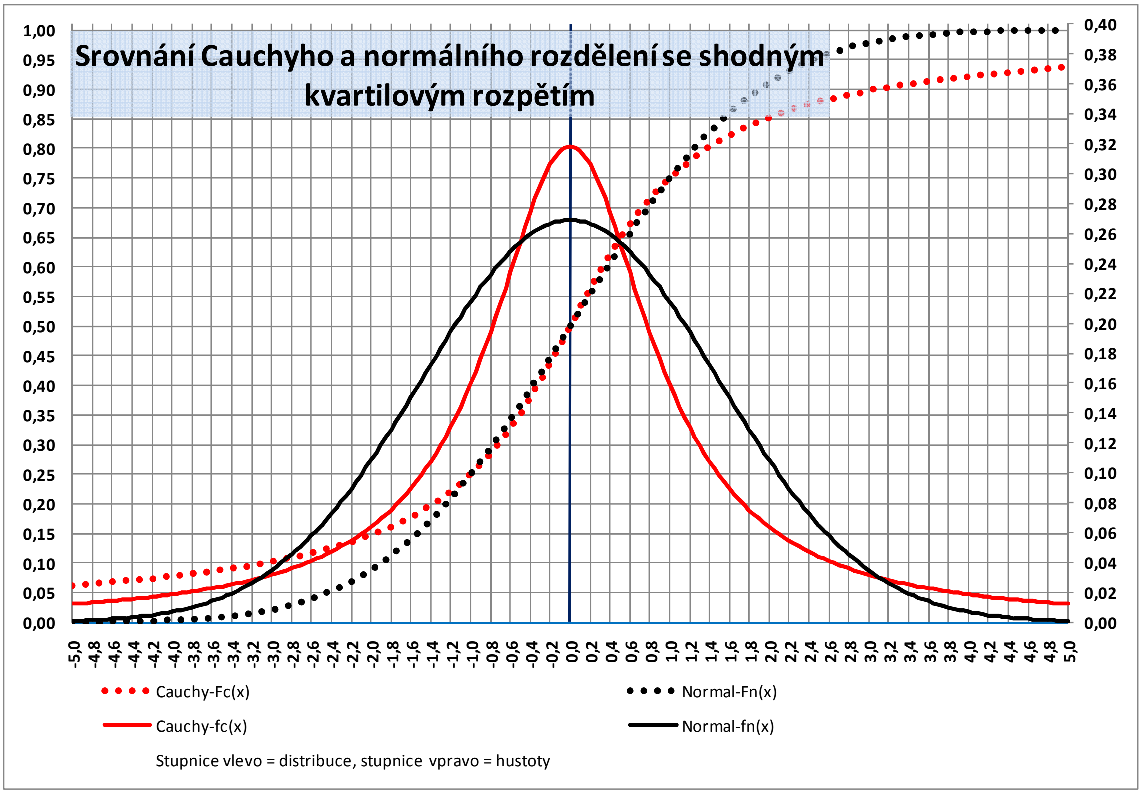
$$= \int_{-\infty}^{+\infty} |y| \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \right] \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(yz)^2}{2}} \right] dy = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |y| e^{-\frac{y^2}{2}(1+z^2)} dy = \frac{1}{\pi} \int_0^{+\infty} y e^{-\frac{y^2}{2}(1+z^2)} dy = (A)$$, po substituci

$x = (1+z^2)\frac{y^2}{2} \;\Rightarrow\; dx = y(1+z^2)dy \;\Rightarrow\; dy = \frac{dx}{y(1+z^2)}$ získáme $(A) = \frac{1}{\pi(1+z^2)} \int_0^{+\infty} e^{-x} dx = \frac{1}{\pi(1+z^2)}$.

Shrnuto: $\boxed{f_\eta(z)=\dfrac{1}{\pi\left(1+z^2\right)} \text{ a } F_\eta(z)=\dfrac{1}{\pi}arctg(x)+\dfrac{1}{2}.}$ To je Cauchyho rozdělení pravděpodobnosti.

Námět: Odvoďte tvar rozdělení, pokud nebudou rozdělení náhodných proměnných vstupujících do podílu centrovaná a normovaná.

Námět: Jak to bude s momenty takového rozdělení, spočtěte medián a případně i modus a kvartilové rozpětí. Odvoďte tvar hustoty a distribuční funkce pro posunuté Cauchyho rozdělení a i se změnou měřítka.



Poměrové statistiky v případě velkých výběrů.

Motivace a důvody

Mějme pozorování $(r_1,s_1),(r_2,s_2),\dots,(r_n,s_n)$, které popisují výsledky nějakých nezávislých experimentů testujících nějakou metodu nebo „měřících" nějaký jev. $r_i,s_i\geq 0;\forall i=1,\dots,n$ a $\sum_{i=1}^{n} r_i>0$. Existují střední hodnoty a rozptyly

$\forall i=1,\dots,n;\ E\{r_i\}=e_{ir};\ \ E\{s_i\}=e_{is};\ \ \sigma^2(r_i)=\sigma^2_{ir};\ \ \sigma^2(s_i)=\sigma^2_{is}$ a

$\forall i=1,\dots,n;\ j=1,\dots,n; E\{(s_i-e_{js})(r_i-e_{jr})\}$. Budeme uvažovat kritérium $\xi_n=\dfrac{\sum_{i=1}^{n} r_i}{\sum_{i=1}^{n} r_i+\sum_{i=1}^{n} s_i}$, pomocí kterého je měřena zvolená kvalita testované metody nebo bude „mírou" nějakého jevu.

Příkladem může být např. měření nezaměstnanosti v celé republice pomocí nezaměstnaností v jednotlivých okresech, kde

r_i je počet registrovaných a dostupných nezaměstnaných v i-tém okrese,

s_i je počet „zaměstnaných" aktivních v i-tém okrese,

n je celkový počet okresů.

Pravděpodobnostní popis

Kritérium $\xi_n = \dfrac{\sum_{i=1}^{n} r_i}{\sum_{i=1}^{n} r_i + \sum_{i=1}^{n} s_i}$ je náhodná proměnná, jejíž hodnota je dána výsledky jednotlivých experimentů (nebo zjišťování). Kritérium $\xi_n = \dfrac{\sum_{i=1}^{n} r_i}{\sum_{i=1}^{n} r_i + \sum_{i=1}^{n} s_i}$ upravíme do tvaru $\xi_n = \dfrac{1}{1 + \dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i}} = \dfrac{1}{1+\eta_n}$. Dále budeme předpokládat, že známe distribuční funkci náhodné proměnné η_n: $F_{\eta_n}(x)$[1].

Potom pro distribuční funkci ξ_n bude platit $F_{\xi_n}(x) = P\{\xi_n < x\} = P\left\{\dfrac{1}{1+\eta_n} < x\right\}$. Vzhledem k uvedeným předpokladům můžeme provést následující úpravy $F_{\xi_n}(x) = P\left\{\dfrac{1}{1+\eta_n} < x\right\} =$
$= P\left\{\dfrac{1}{x} < 1+\eta_n\right\} = P\left\{\dfrac{1}{x} - 1 < \eta_n\right\} = 1 - P\left\{\eta_n < \dfrac{1}{x} - 1\right\} = 1 - F_{\eta_n}(\dfrac{1}{x} - 1)$. Tím je distribuční funkce kritéria ξ_n určena distribuční funkcí náhodné proměnné $\eta_n = \dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i}$.

Shrnuto:

$F_{\xi_n}(x) = 0 \Leftrightarrow x \le 0$

$F_{\xi_n}(x) = 1 - F_{\eta_n}(\dfrac{1}{x} - 1) \Leftrightarrow 0 < x < 1$

$F_{\xi_n}(x) = 1 \Leftrightarrow x \ge 1$

[1] Tato distribuční funkce bude určena dále. Dále nejsou diskutovány „patologické" varianty takového poměru.

Distribuční funkcí náhodné proměnné η_n .

$$F_{\eta_n}(x) = P\{\eta_n < x\} = P\left\{\frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < x\right\} = P\left\{\sum_{i=1}^{n} s_i < x\sum_{i=1}^{n} r_i\right\} = P\left\{\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i < 0\right\} = (A)$$

Použijeme následujících pracovních označení: $e_r = \frac{1}{n}\sum_{i=1}^{n} e_{ir}$; $\quad e_s = \frac{1}{n}\sum_{i=1}^{n} e_{is}$;

Náhodná proměnná $\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i$ má střední hodnotu $E\left\{\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right\} = n(e_s - xe_r)$. Pro její rozptyl lze odvodit:

$$\sigma^2\left\{\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right\} = E\left\{\left[\left(\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right) - n(e_s - xe_r)\right]^2\right\} = E\left\{\sum_{i=1}^{n}[(s_i - e_s) - x(r_i - e_r)]^2\right\} =$$

$$= E\left\{\sum_{i=1}^{n}\left[(s_i - e_s)^2 + x^2(r_i - e_r)^2 - 2x(s_i - e_s)(r_i - e_r)\right]\right\} =$$

$$= \sum_{i=1}^{n}\left[E\{(s_i - e_s)^2\} + x^2E\{(r_i - e_r)^2\} - 2xE\{(s_i - e_s)(r_i - e_r)\}\right] = n\left[\sigma_s^2 + x^2\sigma_s^2 - 2x\rho_{s,r}\sigma_s\sigma_r\right]$$

a tomu odpovídající směrodatná odchylka $\sigma\left\{\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right\} = \sqrt{n\left[\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r\right]}$,

kde:

$$\sigma_r^2 = \frac{1}{n}\sum_{i=1}^{n} E\{(r_i - e_r)^2\};\quad \sigma_s^2 = \frac{1}{n}\sum_{i=1}^{n} E\{(s_i - e_s)^2\};\quad \rho_{s,r}\sigma_s\sigma_r = \frac{1}{n}\sum_{i=1}^{n} E\{(s_i - e_s)(r_i - e_r)\},$$

a x je pevná, daná hodnota ve které počítáme distribuční funkci $F_{\eta_n}(x)$.

Nyní již můžeme pokračovat v odvození distribuční funkce náhodné proměnné η_n:

$$(A) = P\left\{\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i < 0\right\} = P\left\{\left(\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right) - n(e_s - xe_r) < -n(e_s - xe_r)\right\} =$$

$$= P\left\{\frac{\left(\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right) - n(e_s - xe_r)}{\sqrt{n\left[\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r\right]}} < \frac{-n(e_s - xe_r)}{\sqrt{n\left[\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r\right]}}\right\} = (B)$$

Pravou stranu nerovnosti „uvnitř“ pravděpodobnosti lze upravit následovně:

$$\frac{-n(e_s - xe_r)}{\sqrt{n\left[\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r\right]}} = -\sqrt{n}\,\frac{(e_s - xe_r)}{\sqrt{\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r}}.$$

Levá strana této nerovnosti $\frac{\left(\sum_{i=1}^{n} s_i - x\sum_{i=1}^{n} r_i\right) - n(e_s - xe_r)}{\sqrt{n\left[\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r\right]}}$ je centrovaná a normovaná náhodná proměnná, jejíž distribuční funkce s rostoucím n konverguje k distribuční funkci centrovaného a normovaného normálního rozdělení $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}dz$, $\left(\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)$. Důkaz tohoto tvrzení a jeho možné varianty (a další dodatečné předpoklady) lze nalézt např. v Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972 str. 374-382.

Proto pro dostatečně velká n platí $(B) = \Phi\left(-\sqrt{n}\,\frac{(e_s - xe_r)}{\sqrt{\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r}}\right)$.

Shrnuto: $\boxed{F_{\eta_n}(x) = P\{\eta_n < x\} = \Phi\left(-\sqrt{n}\,\frac{(e_s - xe_r)}{\sqrt{\sigma_s^2 + x^2\sigma_r^2 - 2x\rho_{s,r}\sigma_s\sigma_r}}\right)}$

Distribuční funkcí náhodné proměnné ξ_n .

Podle dříve odvozeného vztahu $F_{\xi_n}(x) = 1 - F_{\eta_n}(\frac{1}{x} - 1) \Leftrightarrow 0 < x < 1$ můžeme dostat distribuční funkci kritéria ξ_n

$$F_{\xi_n}(x) = 1 - \Phi\left(-\sqrt{n}\,\frac{\left(e_s - (\frac{1}{x} - 1)e_r\right)}{\sqrt{\sigma_s^2 + \left(\frac{1}{x} - 1\right)^2\sigma_r^2 - 2\left(\frac{1}{x} - 1\right)\rho_{s,r}\sigma_s\sigma_r}}\right) \Leftrightarrow 0 < x < 1.$$

To je poměrně komplikovaný výraz, avšak snadno vyčíslitelný (numerická realizace funkce $\Phi(x)$ je dostupná ve většině programovacích prostředků včetně její inverze).

Reprezentace bodovým odhadem (parametrem)

Pro daný typ rozdělení je nejdostupnějším reprezentativním parametrem medián. Je to řešení rovnice $F_{\xi_n}(x_{med}) = 1/2$. Tedy $e_s = (\frac{1}{x_{med}} - 1)e_r \Rightarrow x_{med} = \frac{e_r}{e_r + e_s}$.

Toleranční meze[2]

Hledáme čísla $x_L < x_U$ taková že platí $P(x_L \le \xi_n \le x_U) = 1 - \alpha$, kde α je hladina významnosti. Vztahem $P(x_L \le \xi_n \le x_U) = 1 - \alpha$ nejsou čísla $x_L < x_U$ určena jednoznačně. Proto zvolíme $\alpha = \alpha_1 + \alpha_2$; $0 < \alpha, \alpha_1, \alpha_2 < 1$ tak, že $P(x_U < \xi_n) = \alpha_1 \Leftrightarrow P(\xi_n < x_U) = 1 - \alpha_1$ a $P(\xi_n < x_L) = \alpha_2$. **Vyjádřeno pomocí aparátu distribučních funkcí:** $F_{\xi_n}(x_U) = 1 - \alpha_1$**,** $F_{\xi_n}(x_L) = \alpha_2$**.**

[2] Pojetí tolerančních mezí námi použité je jednoduché a liší se od [2]. Předpokládáme n dostatečně velké a tím i příslušné bodové odhady s přijatelnou reprezentativní schopností.

x_L pak bude řešením rovnice $1-\Phi\left(-\sqrt{n}\dfrac{\left(e_s-(\frac{1}{x_L}-1)e_r\right)}{\sqrt{\sigma_s^2+\left(\frac{1}{x_L}-1\right)^2\sigma_r^2-2\left(\frac{1}{x_L}-1\right)\rho_{s,r}\sigma_s\sigma_r}}\right)=\alpha_2$,

tedy $1-\alpha_2=\Phi\left(-\sqrt{n}\dfrac{\left(e_s-(\frac{1}{x_L}-1)e_r\right)}{\sqrt{\sigma_s^2+\left(\frac{1}{x_L}-1\right)^2\sigma_r^2-2\left(\frac{1}{x_L}-1\right)\rho_{s,r}\sigma_s\sigma_r}}\right)$, to při využití inverzní funkce $\Phi^{-1}(x)$ k funkci $\Phi(x)$ dá

$\Phi^{-1}(1-\alpha_2)=-\sqrt{n}\dfrac{\left(e_s-(\frac{1}{x_L}-1)e_r\right)}{\sqrt{\sigma_s^2+\left(\frac{1}{x_L}-1\right)^2\sigma_r^2-2\left(\frac{1}{x_L}-1\right)\rho_{s,r}\sigma_s\sigma_r}}$. Použijeme-li pracovní označení

$z=\dfrac{1}{x_L}-1$, budeme hledat kořen rovnice

$\Phi^{-1}(1-\alpha_2)=-\sqrt{n}\dfrac{(e_s-ze_r)}{\sqrt{\sigma_s^2+z^2\sigma_r^2-2z\rho_{s,r}\sigma_s\sigma_r}}\Rightarrow\Phi^{-1}(1-\alpha_2)\sqrt{\sigma_s^2+z^2\sigma_r^2-2z\rho_{s,r}\sigma_s\sigma_r}=-\sqrt{n}(e_s-ze_r)$.

Ten bude obsažen mezi řešeními rovnice

$\left[\Phi^{-1}(1-\alpha_2)\right]^2\left(\sigma_s^2+z^2\sigma_r^2-2z\rho_{s,r}\sigma_s\sigma_r\right)=n(e_s-ze_r)^2$. Pro přehlednost použijeme další pracovní označení $\dfrac{\left[\Phi^{-1}(1-\alpha_2)\right]^2}{n}=a$. Pak hledáme kořen rovnice

$a\left(\sigma_s^2+z^2\sigma_r^2-2z\rho_{s,r}\sigma_s\sigma_r\right)=(e_s-ze_r)^2=e_s^2+z^2e_r^2-2ze_se_r$

a po úpravě: $z^2\left(a\sigma_r^2-e_r^2\right)+2z\left(e_se_r-\rho_{s,r}a\sigma_r\sigma_s\right)+\left(a\sigma_s^2-e_s^2\right)=0$. Tuto rovnici lze, pro přehlednost, přepsat do tvaru: $Az^2+2zB+C=0$,

kde $A=\left(a\sigma_r^2-e_r^2\right)$; $B=\left(e_se_r-\rho_{s,r}a\sigma_r\sigma_s\right)$; $C=\left(a\sigma_s^2-e_s^2\right)$. Nalezneme její kořeny:

$z_{1,2}=\dfrac{-B\pm\sqrt{B^2-AC}}{A}$. Tyto kořeny ověříme pro rovnici:

$\dfrac{\Phi^{-1}(1-\alpha_2)}{\sqrt{n}}\sqrt{\sigma_s^2+z^2\sigma_r^2-2z\rho_{s,r}\sigma_s\sigma_r}+(e_s-ze_r)=0$. Jeden z nich rovnici vyhoví z_*, jeden ne (umocněním jsme v předchozím odvození „přidali“ jeden kořen). Z užité transformace $z_*=\dfrac{1}{x_L}-1$ dostaneme $x_L=\dfrac{1}{1+z_*}$.

Pro výpočet horní meze budeme vycházet ze vztahu:

$$\alpha_1=\Phi\left(-\sqrt{n}\frac{\left(e_s-(\frac{1}{x_U}-1)e_r\right)}{\sqrt{\sigma_s^2+\left(\frac{1}{x_U}-1\right)^2\sigma_r^2-2\left(\frac{1}{x_U}-1\right)\rho_{s,r}\sigma_s\sigma_r}}\right).$$

Postup pro výpočet horní meze bude „shodný“ s postupem pro výpočet meze dolní jen s tím rozdílem, že $\frac{\left[\Phi^{-1}(\alpha_1)\right]^2}{n} = a$ a ověřovací rovnice pro výběr správného kořenu bude $\frac{\Phi^{-1}(\alpha_1)}{\sqrt{n}}\sqrt{\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r} + (e_s - ze_r) = 0$.

Přijatelnost asymptotické reprezentace (praktický pohled)

Předpis pro asymptotickou verzi distribuční funkce $F_{\xi_n}(x)$ obecně nemusí být distribuční funkcí a obvykle jí také není. Nejen to, ξ_n je náhodná proměnná definovaná na intervalu $\langle 0,1 \rangle$. $F_{\xi_n}(x)$ svým předpisem tento interval přesahuje (a to nenulovou nebo ne-jednotkovou hodnotou). Těchto skutečností lze využít k testování přijatelnosti takové asymptotické reprezentace. Pro přijatelnost budou zajímavé dva pohledy.

1. Na intervalu $\langle 0,1 \rangle$ by měla být funkce $F_{\xi_n}(x)$ neklesající.
2. Hodnota $F_{\xi_n}(x)$ „v nule“ by měla být nezáporná a menší než zadané malé nezáporné číslo β_0. Hodnota $F_{\xi_n}(x)$ „v jednotce“ by neměla překročit jednotku, ale měla by být větší než zadané číslo $1-\beta_1$, β_1 nezáporné, dostatečně malé.

První požadavek

Ke splnění prvního požadavku stačí určit $\frac{d}{dx}F_{\xi_n}(x) = f_{\xi_n}(x)$ a testovat nezápornost $f_{\xi_n}(x)$ na intervalu $\langle 0,1 \rangle$. Pro přehlednost opět použijeme pomocnou náhodnou proměnnou $z = \frac{1}{x} - 1$, kde $\frac{dz}{dx} = -\frac{1}{x^2}$ a $F_{\xi_n}(z) = 1 - \Phi\left(-\sqrt{n}\,\frac{(e_s - ze_r)}{\sqrt{\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r}}\right); 0 < z < +\infty$.

Potom: $f_{\xi_n}(x) = \frac{d}{dx}F_{\xi_n}(z) = \frac{d}{dz}F_{\xi_n}(z)\frac{dz}{dx} = \frac{d}{dz}\left[1 - \Phi\left(-\sqrt{n}\,\frac{(e_s - ze_r)}{\sqrt{\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r}}\right)\right]\frac{dz}{dx}$.

Vzhledem ke znaménku $\frac{dz}{dx}$ je pro znaménko $f_{\xi_n}(x)$ rozhodující znaménko

$$\frac{d}{dz}\left[1 - \Phi\left(-\sqrt{n}\,\frac{(e_s - ze_r)}{\sqrt{\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r}}\right)\right].$$

$$\frac{d}{dz}\left[1-\Phi\left(-\sqrt{n}\,\frac{(e_s - ze_r)}{\sqrt{\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r}}\right)\right]=$$

$$=-\sqrt{n}\varphi\left(-\sqrt{n}\,\frac{(e_s - ze_r)}{\sqrt{\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r}}\right)\left(\frac{e_r\left(\sigma_s^2 - z\rho_{s,r}\sigma_s\sigma_r\right)+e_s\left(z\sigma_r^2 - \rho_{s,r}\sigma_s\sigma_r\right)}{\left(\sigma_s^2 + z^2\sigma_r^2 - 2z\rho_{s,r}\sigma_s\sigma_r\right)^{\frac{3}{2}}}\right)$$

Podmínka $f_{\xi_n}(x) \geq 0$ bude proto splněna pro takové $z = \frac{1}{x} - 1$, pro které bude $e_r\left(\sigma_s^2 - z\rho_{s,r}\sigma_s\sigma_r\right)+e_s\left(z\sigma_r^2 - \rho_{s,r}\sigma_s\sigma_r\right) \geq 0$. Pro její testování je rozhodující „zlomový bod“, tj. řešení rovnice $e_r\left(\sigma_s^2 - z\rho_{s,r}\sigma_s\sigma_r\right)+e_s\left(z\sigma_r^2 - \rho_{s,r}\sigma_s\sigma_r\right)=0$ a odtud $z = \frac{\sigma_s}{\sigma_r}\frac{\rho_{s,r}e_s\sigma_r - e_r\sigma_s}{e_s\sigma_r - \rho_{s,r}e_r\sigma_s}$.

Aby takový „zlomový bod x“ byl mimo interval $\langle 0,1\rangle$ stačí $\frac{\rho_{s,r}e_s\sigma_r - e_r\sigma_s}{e_s\sigma_r - \rho_{s,r}e_r\sigma_s} \leq 0$, neboť $x = \frac{1}{1+z}$. Protože v „mediánovém bodě“ $x_{med} = \frac{e_r}{e_r + e_s} \Rightarrow z_{med} = \frac{e_s}{e_r}$ je $e_r\left(\sigma_s^2 - z_{med}\rho_{s,r}\sigma_s\sigma_r\right)+e_s\left(z_{med}\sigma_r^2 - \rho_{s,r}\sigma_s\sigma_r\right)=\frac{(e_r\sigma_s - e_s\sigma_r)^2}{e_r} \geq 0$, je podmínka $\frac{\rho_{s,r}e_s\sigma_r - e_r\sigma_s}{e_s\sigma_r - \rho_{s,r}e_r\sigma_s} \leq 0$ nutnou a postačující podmínkou pro $f_{\xi_n}(x) \geq 0; 0 < x \leq 1$. Pokud bude tato podmínka splněna, řekneme, že asymptotická aproximace je korektní. Korektnost má velký význam pro „malá n“, pro dostatečně velká n“ bude případná „oblast poklesu“ distribuční funkce zanedbatelná z pohledu vzniklé odchylky.

Poznámka: Uvedená podmínka korektnosti nezávisí na $n =$ počtu pozorování.

Druhý požadavek

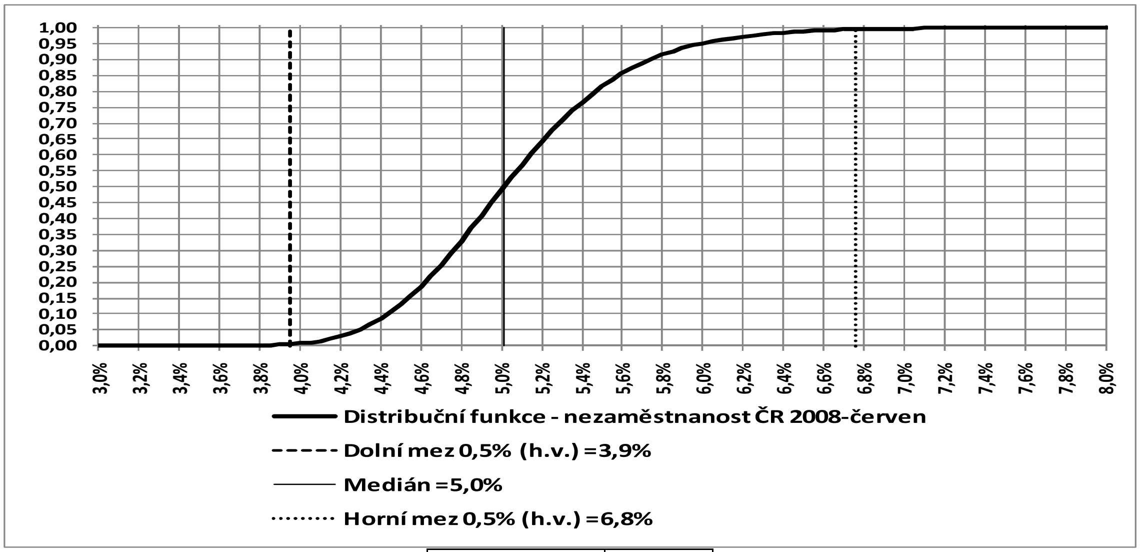
Pokud budou dána čísla $0 < \beta_0, \beta_1, \beta = \beta_0 + \beta_1 < 1$, můžeme o asymptotické aproximaci říci, že je β-přijatelná, když budou splněny nerovnosti $1 \geq F_{\xi_n}(1) \geq 1-\beta_1$ a $0 \leq \lim_{x\to 0} F_{\xi_n}(0) \leq \beta_0$.

$F_{\xi_n}(1) = 1-\Phi\left(-\sqrt{n}\,\frac{e_s}{\sigma_s}\right)$, $\lim_{x\to 0_+} F_{\xi_n}(x) = 1-\Phi\left(\sqrt{n}\,\frac{e_r}{\sigma_r}\right)$ a

$1-\Phi\left(-\sqrt{n}\,\frac{e_s}{\sigma_s}\right) \geq 1-\beta_1 \Rightarrow n \geq \frac{\sigma_s^2}{e_s^2}\left(\Phi^{-1}(\beta_1)\right)^2$, kde $\Phi^{-1}(x)$ je funkce inverzní k funkci $\Phi(x)$, $1-\Phi\left(\sqrt{n}\,\frac{e_r}{\sigma_r}\right) \leq \beta_0 \Rightarrow n \geq \frac{\sigma_r^2}{e_r^2}\left(\Phi^{-1}(1-\beta_0)\right)^2$.

Shrnuto: Na to aby asymptotická aproximace byla $\beta = \beta_0 + \beta_1$ přijatelná, je nutné a stačí $n \geq \max\left\{\frac{\sigma_s^2}{e_s^2}\left(\Phi^{-1}(\beta_1)\right)^2; \frac{\sigma_r^2}{e_r^2}\left(\Phi^{-1}(1-\beta_0)\right)^2\right\}$.

Příklad využití pro toleranční meze u nezaměstnanosti:



Zlomové Z	-239,3
Zlomové X	-0,004
Přijatelné BETA	0,001%
Přijatelný počet	27,25
Skutečný počet	77

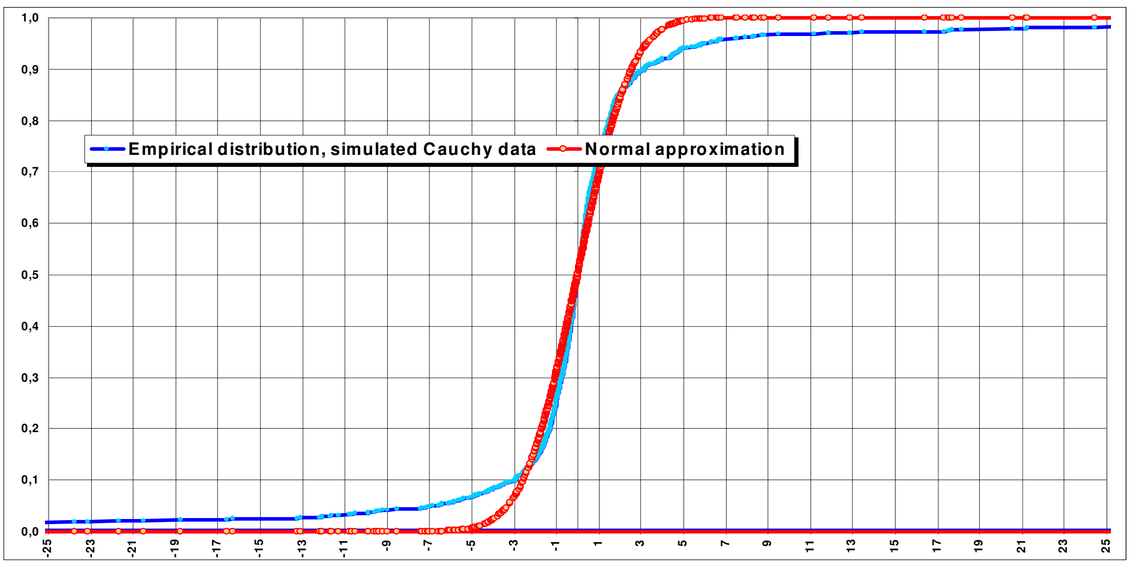
Detaily výpočtu viz soubor: Nezaměstnanost 0806.xls.

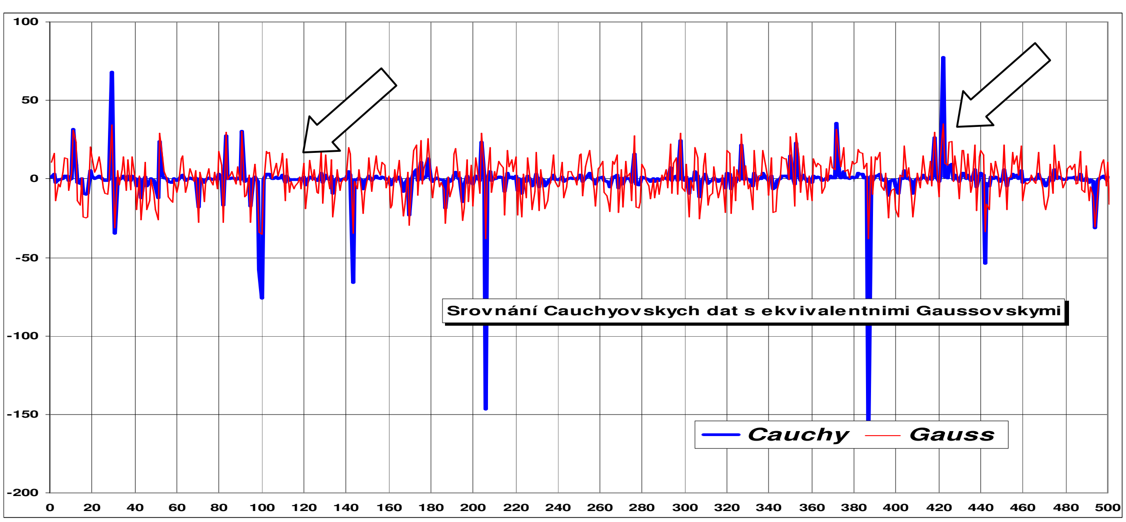
Těžké konce a důsledky pro statistiku.

Rozdělení s těžkými koci jsou taková rozdělení, pro která je pravděpodobnost „extrémních hodnot“ sice malá ale ne nezadatelná jako např. u normálního rozdělení. Příkladem takového rozdělení je výše uvedené Cauchyho rozdělení. Existuje více definic „těžkých konců“ a i více názvů tohoto jevu. Např.:

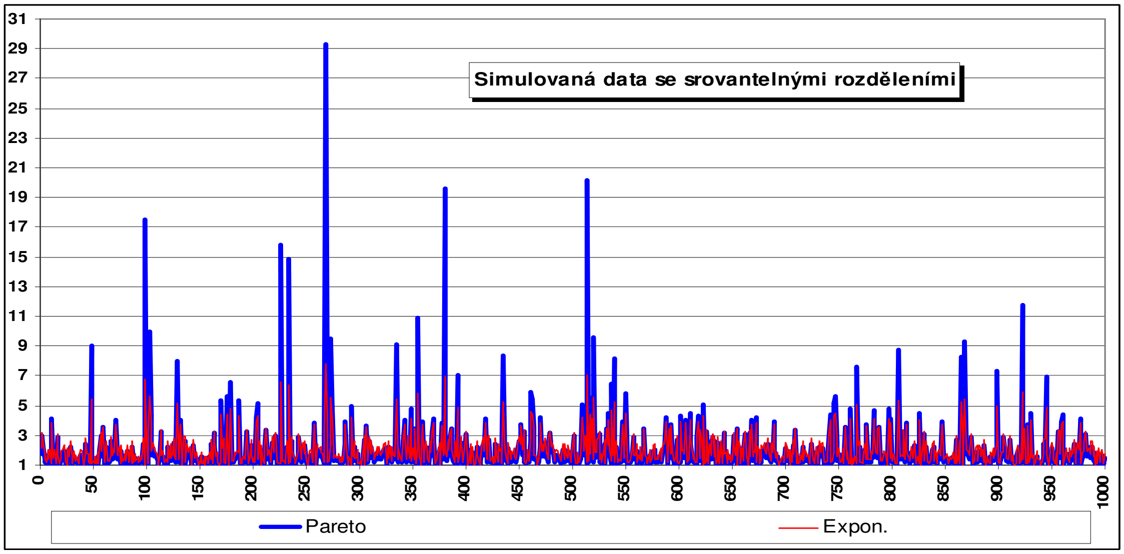
Heavy tailed distributions	***Distribuce s těžkým koncem (konci, zbytky)***
Fat tails	***Tlusté konce (zbytky, zbytkové pravděpodobnosti)***
Long tails	***Dlouhé konce***

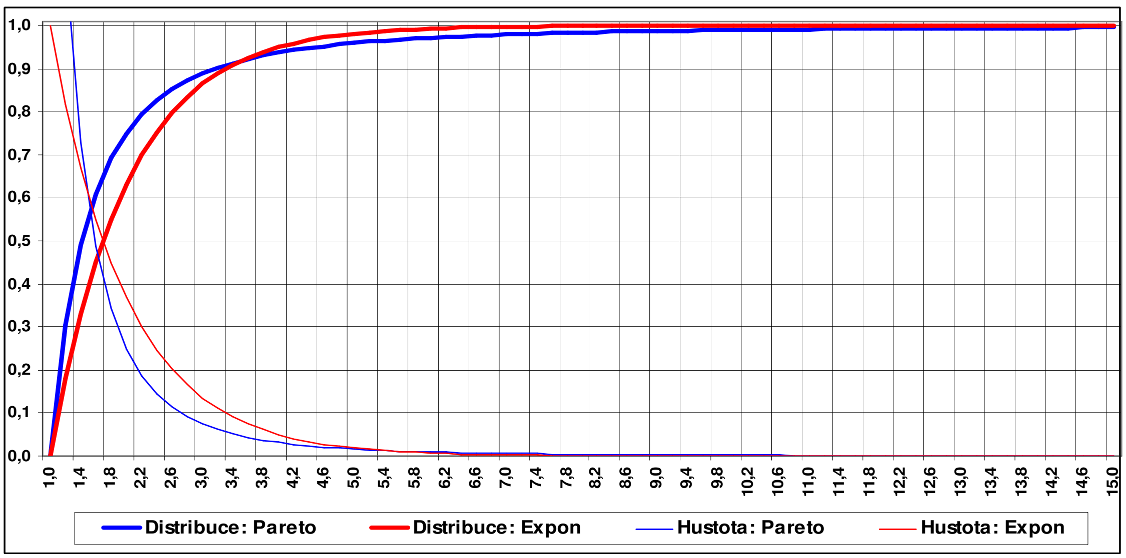
Tento jev reprezentován pomocí empirických distribucí a i simulovaných hodnot vypadá následovně (srovnání cauchyovská a normálně rozdělená data, názorný příklad):





Paretovská data a srovnatelná data z exponenciálním rozdělením (méně názorný příklad).

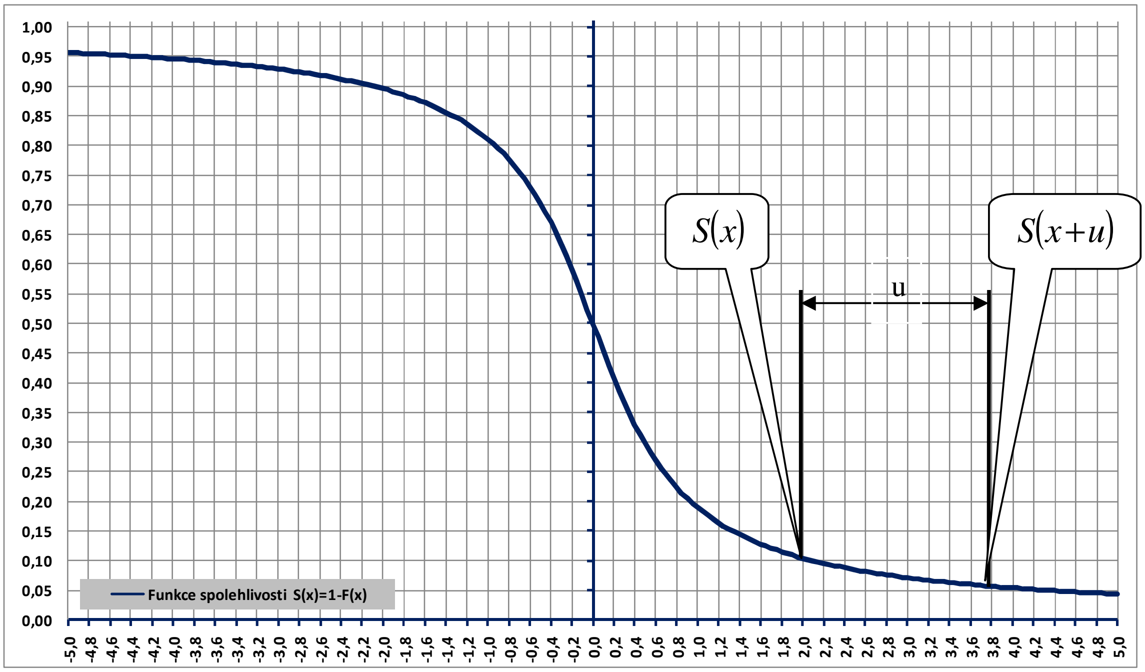




Definice „těžkého konce“ vycházejí z několika pohledů. První z nich používá následující podmíněnou pravděpodobnost pro popis „velkých odchylek“:

$$P\{X \geq x+u / X \geq x\} \text{ pro libovolné } u>0; x \in R_1.$$

Potom: $P\{X \geq x+u / X \geq x\} = \dfrac{P\{X \geq x+u\}}{P\{X \geq x\}} = \dfrac{1-F(x+u)}{1-F(x)} = \dfrac{S(x+u)}{S(x)}$. Rozdělení charakterizující je pak $i(u) = \lim\limits_{x \to +\infty} \dfrac{S(x+u)}{S(x)}$. Takový způsob „měření“ demonstruje následující obrázek:



Demonstrace „indexu konce“.

Primárně limita $i(u)=\lim\limits_{x\to+\infty}\dfrac{S(x+u)}{S(x)}$ vede na neurčitý výraz $\dfrac{0}{0}$, užije se l'Hospitalova pravidla a pak: $i(u)=\lim\limits_{x\to+\infty}\dfrac{f(x+u)}{f(x)}$.

Poznámka: Uvedený formalismus popisuje chování „pravého konce“ (pravého chvostu) rozdělení. Bez problému jej lze přizpůsobit i pro levý konec.

Poznámka: Takový rozbor konců rozdělení má smysl a logiku jen pro rozdělení, která jsou definována nad alespoň, z jedné strany, nekonečným intervalem.

Pro **normální rozdělení s hustotou** $f_X(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, dostaneme:

$$i(u)=\lim_{x\to+\infty}\frac{f(x+u)}{f(x)}=\lim_{x\to+\infty}\frac{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x+u-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}=\lim_{x\to+\infty}\frac{e^{-\frac{(x+u-\mu)^2}{2\sigma^2}}}{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}=\lim_{x\to+\infty}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}-\frac{(x+u-\mu)^2}{2\sigma^2}\right)=$$

$$=\lim_{x\to+\infty}\exp\left(\frac{(x-\mu)^2-(x+u-\mu)^2}{2\sigma^2}\right)=\lim_{x\to+\infty}\exp\left(\frac{-2u(x-\mu)-u^2}{2\sigma}\right)=\lim_{x\to+\infty}\left[\exp\left(\frac{-2u(x-\mu)}{2\sigma}\right)\exp\left(\frac{-u^2}{2\sigma}\right)\right]=0$$

Pro **exponenciální rozdělení s hustotou** $f_X(x)=\lambda e^{-\lambda x}$ dostaneme:

$$i(u)=\lim_{x\to+\infty}\frac{f(x+u)}{f(x)}=\lim_{x\to+\infty}\frac{\lambda e^{-\lambda(x+u)}}{\lambda e^{-\lambda x}}=e^{-\lambda u}$$

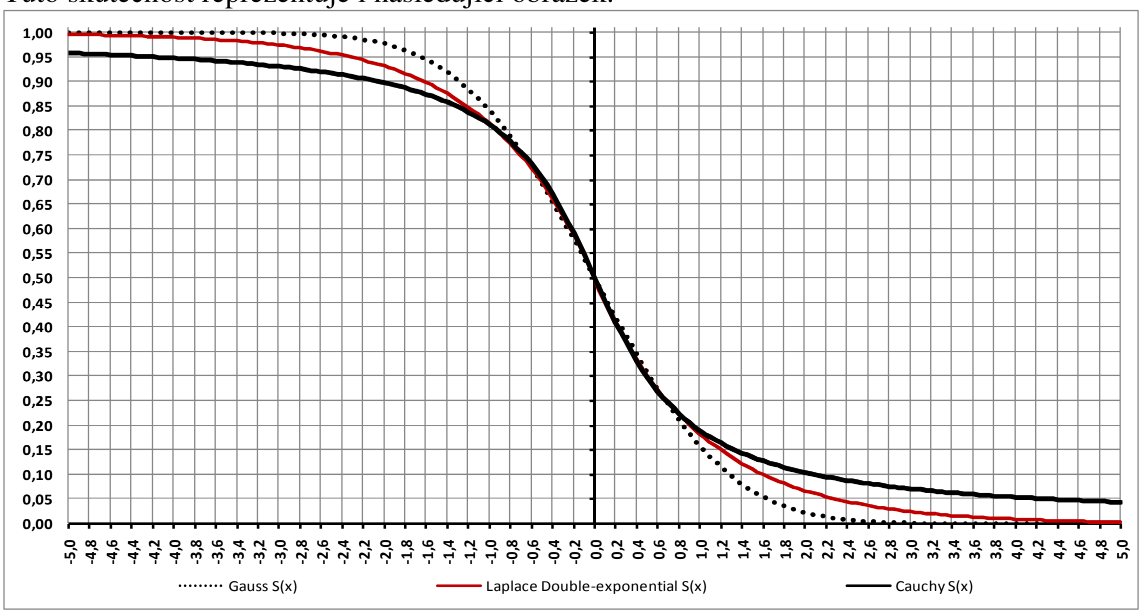
Pro **Paretovo rozdělení s hustotou** $f_X(x) = \dfrac{K}{x^{k+1}}$ dostaneme:

$$i(u) = \lim_{x\to+\infty} \frac{f(x+u)}{f(x)} = \lim_{x\to+\infty} \frac{\dfrac{K}{(x+u)^{k+1}}}{\dfrac{K}{(x)^{k+1}}} = \lim_{x\to+\infty}\left(\frac{x}{x+u}\right)^{k+1} = 1.$$

Uvedené tři příklady „jsou reprezentanty“ tří tříd chování konců u různých rozdělení.

První: Tato rozdělení (zde normální rozdělení) „jdou k nule velice rychle – nad exponenciální rychlostí“ mají tedy zanedbatelné konce.
Druhá: Tato rozdělení (zde exponenciální rozdělení) „jdou k nule exponenciální rychlostí“. Někdy jsou slučována pod jednu třídu exponenciálních rozdělení (s exponenciálními konci).
Třetí: U těchto rozdělení (zde Paretovo rozdělení) je pokles k nule velice pomalý a $P\{X \geq x+u / X \geq u\}$ bude pro dostatečně velká u prakticky jednotková.

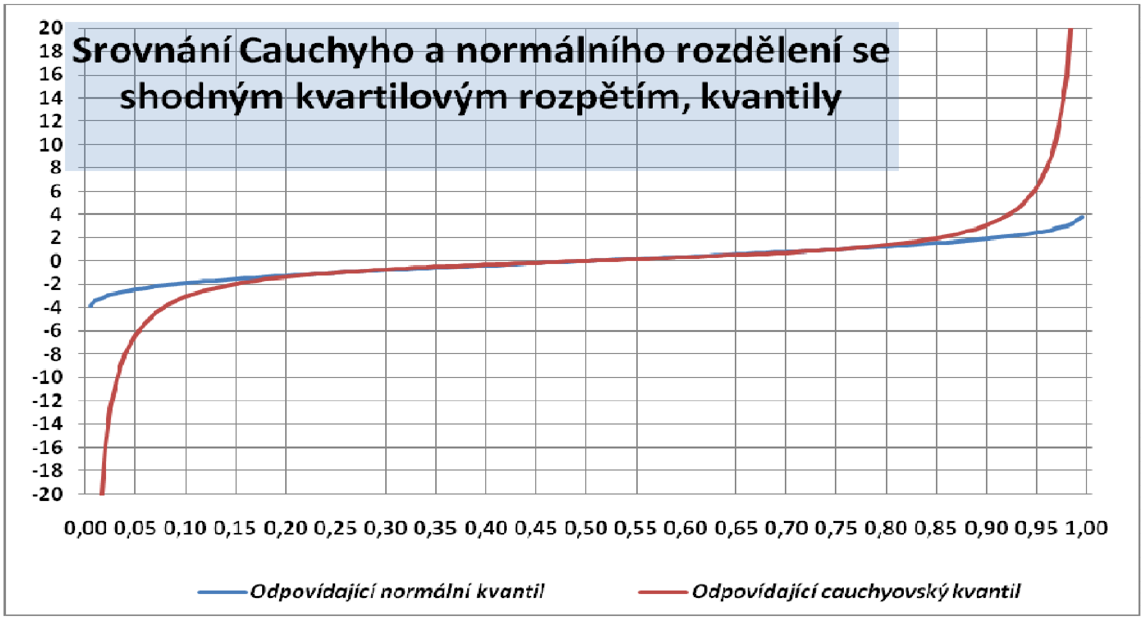
Tuto skutečnost reprezentuje i následující obrázek:



Demonstrace tří skupin „konců“ rozdělení..

Pro statistické potřeby je „jev těžkých konců“ výrazný u výpočtu kvantilů (nejen), specielně jejich ***citlivosti na malou změnu vstupní hodnoty***.

Příklad: průběh hodnoty kvantilů u rozdělení ***N(0,σ)*** a normovaného a centrovaného Cauchyho rozdělení, σ je voleno tak, aby obě rozdělení měla shodné kvartilové rozpětí:



Reálně „měřená data“ obvykle nemívají rozdělení s „těžkými konci“, avšak ty mohou vznikat zpracováním (např. zde demonstrovanými poměry). Ale v ekonometrických (ekonomických) úlohách je poměr (podíl) téměř standardní a neopominutelnou operací, proto je nezbytná velká opatrnost.

Doporučená a zdrojová literatura:

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	
Machek J.	Teorie odhadu, SPN Praha 1974, skripta MFFUK	
Miloš Jílek	Statistické toleranční meze. SNTL Praha 1988.	
Brendan O. Bradley and Murad S. Taqqu	Financial Risk and Heavy Tails.	***Boston University, August 28, 2001,*** To Appear in the in 2002 in the volume ‘‘Heavy-tailed distributions in Finance’’, Svetlozar T. Rachev, editor, North Holland; Copyright Brendan O. Bradley and Murad S. Taqqu.
J. Beirlant (jan.beirlant@wis.kuleuven.ac.be) G. Matthys (gunther.matthys@ucs.kuleuven.ac.be)	Quantile Estimation for Heavy-Tailed Data	*23/03/2000*
Charles M. Goldie and Claudia Kluppelberg.	Subexponential Distributions	Charles M. Goldie School of Mathematical Sciences University of Sussex Brighton BN1 9QH, England email: C.M.Goldie@sussex.ac.uk Claudia Kluppelberg Department of Mathematics Johannes Gutenberg--Universit¨at D--55099 Mainz, Germany email: cklu@mathematik.uni-mainz.de