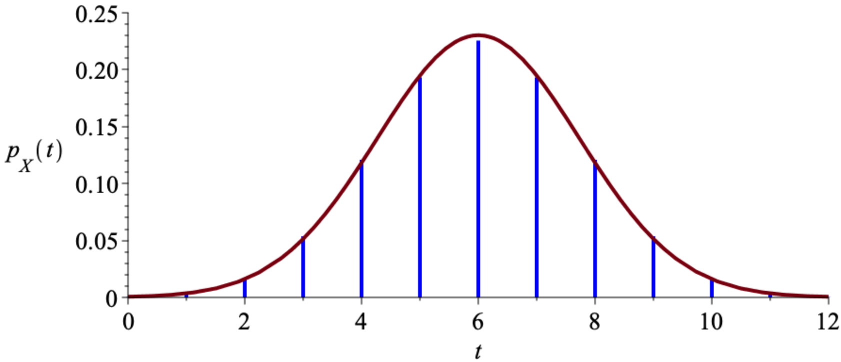
Co (ne)říká centrální limitní věta

Mirko Navara, 29. prosince 2020

Co chceme říci

Když se podíváme na binomické rozdělení z velkého počtu možností n, připomíná tvarem Gaussovu křivku.



Pravděpodobnostní funkce rozdělení $\mathrm{Bi}(12, 1/2)$ *a její proložení Gaussovou křivkou*

Je to důsledek toho, že se jedná o součet mnoha nezávislých náhodných veličin. Jejich rozdělení bylo alternativní, ale podobné výsledky dostaneme i pro mnohá jiná rozdělení.

Předpokládáme náhodné veličiny X_j, $j = 1, 2, \ldots$, které jsou **nezávislé** a mají **stejné rozdělení** (s distribuční funkcí F_X, kvantilovou funkcí q_X, popřípadě pravděpodobnostní funkcí p_X nebo hustotou f_X).
Definujeme posloupnost náhodných veličin U_n, $n = 1, 2, \ldots$, což jsou částečné součty řady,

$$U_n = \sum_{j=1}^{n} X_j \,.$$

Protože nás bude zajímat limitní chování pro $n \to \infty$, potřebujeme nekonečnou posloupnost náhodných veličin X_j, $j \in \mathbb{N}$, i částečných součtů U_n, $n \in \mathbb{N}$.
Centrální limitní věta říká, že pro „velká n“ se rozdělení náhodných veličin U_n „tvarem blíží“ normálnímu (je „asymptoticky normální“). Je však těžké korektně vyjádřit „tvarem se blíží.“

Jak správně formulovat centrální limitní větu

Kdyby X_j měly normální rozdělení $\mathrm{N}(\mathrm{E}X, \mathrm{D}X)$, částečné součty U_n by měly normální rozdělení $\mathrm{N}(\mathrm{E}U_n, \mathrm{D}U_n)$ s parametry

$$\mathrm{E}U_n = n\,\mathrm{E}X \,,$$
$$\mathrm{D}U_n = n\,\mathrm{D}X \,.$$

Tyto parametry má U_n i v obecném případě, pokud existuje rozptyl $\mathrm{D}X$ (a tedy i střední hodnota $\mathrm{E}X$).

„Tvar rozdělení“ je to, co zbyde, když odhlédneme od střední hodnoty a rozptylu, tj. když náhodnou veličinu znormujeme:

$$Y_n = \operatorname{norm} U_n = \frac{U_n - \mathrm{E} U_n}{\sigma_{U_n}} = \frac{U_n - n\,\mathrm{E}X}{\sqrt{n}\,\sigma_X}\,.$$

Stejný výsledek dá i normovaný výběrový průměr,

$$\operatorname{norm} \overline{\boldsymbol{X}}_n = \frac{\overline{\boldsymbol{X}}_n - \mathrm{E}X}{\frac{1}{\sqrt{n}}\,\sigma_X} = \frac{n\,\overline{\boldsymbol{X}}_n - n\,\mathrm{E}X}{\sqrt{n}\,\sigma_X} = Y_n\,,$$

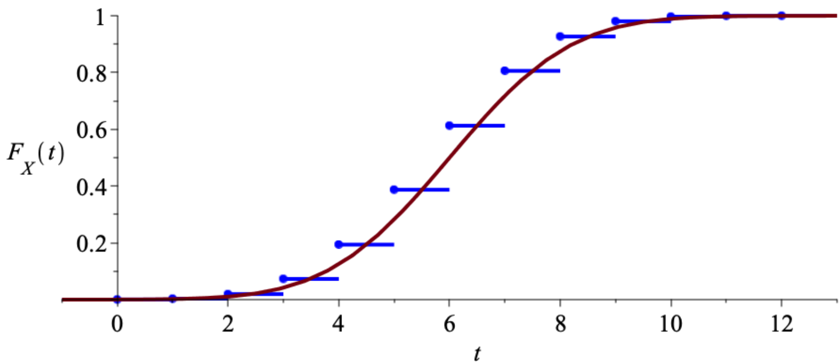
kde $n\,\overline{\boldsymbol{X}}_n = U_n$.
Zpětný převod:

$$U_n = n\,\mathrm{E}X + \sqrt{n}\,\sigma_X\,Y_n\,,$$
$$\overline{\boldsymbol{X}}_n = \mathrm{E}X + \frac{\sigma_X}{\sqrt{n}}\,Y_n\,.$$

Centrální limitní věta říká, že pro $n \to \infty$

$$Y_n \rightsquigarrow \mathrm{N}(0,1)\,, \tag{1}$$

kde $\rightsquigarrow$ značí, že „rozdělení se blíží,“ což bude nutno upřesnit.



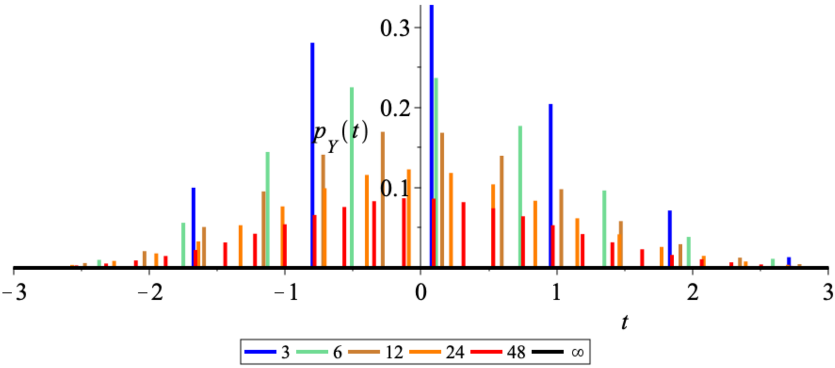
Distribuční funkce binomického rozdělení $\mathrm{Bi}(12, 1/2)$ *a normálního rozdělení se stejnou střední hodnotou a rozptylem*

Kdyby X_j měly normální rozdělení, platila by v (1) rovnost (a centrální limitní větu bychom nepotřebovali).

Centrální limitní věta **neříká**, že **pravděpodobnostní funkce** konvergují (k pravděpodobnostní funkci),

$$\lim_{n\to\infty} p_{Y_n}(t) = p_{\mathrm{N}(0,1)}(t) \qquad \text{pro všechna } t \in \mathbb{R}\,,$$

i když **je to pravda**, ovšem triviální – obě strany jsou nulové.

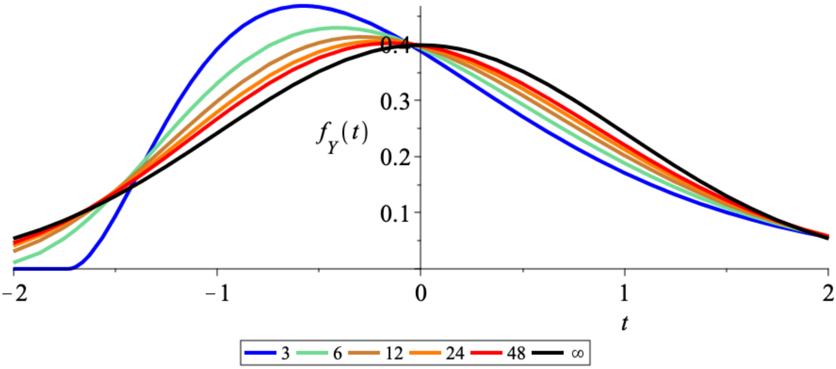


Pravděpodobnostní funkce normovaných součtů různých počtů nezávislých náhodných veličin s alternativním rozdělením Alt$(1/\pi)$

Centrální limitní věta **neříká**, že **hustoty** konvergují (k hustotě),

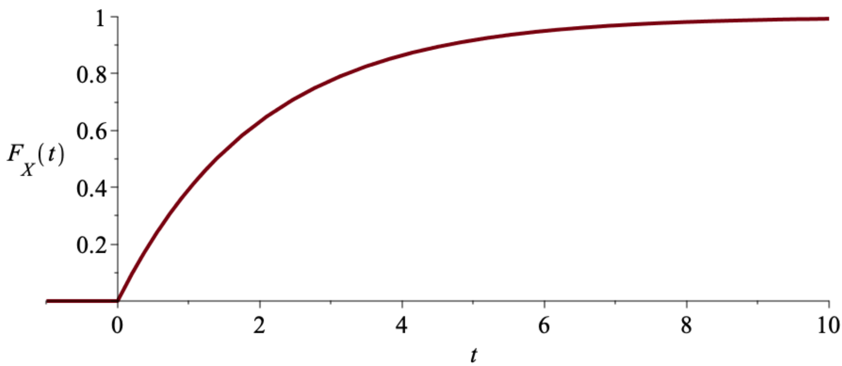
$$\lim_{n\to\infty} f_{Y_n}(t) = f_{\mathrm{N}(0,1)}(t) \qquad \text{pro } \textbf{skoro} \text{ všechna } t \in \mathbb{R}\,;$$

je to pravda jen pro absolutně spojitá rozdělení, ale my se neomezíme jen na ně.

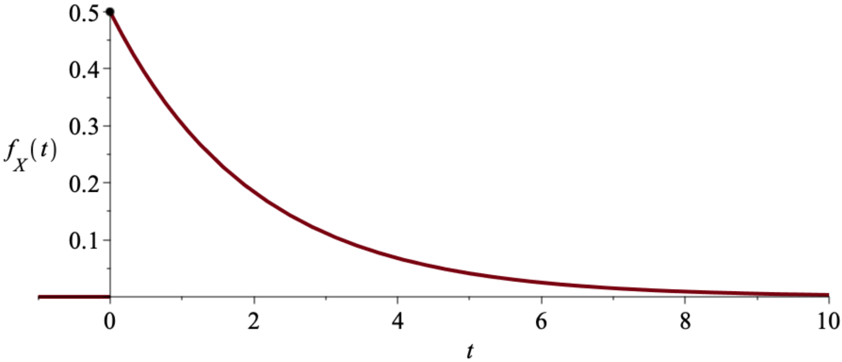


Hustoty normovaných součtů různých počtů nezávislých náhodných veličin

Zde i dále je použit výběr z exponenciálního rozdělení Exp(2).



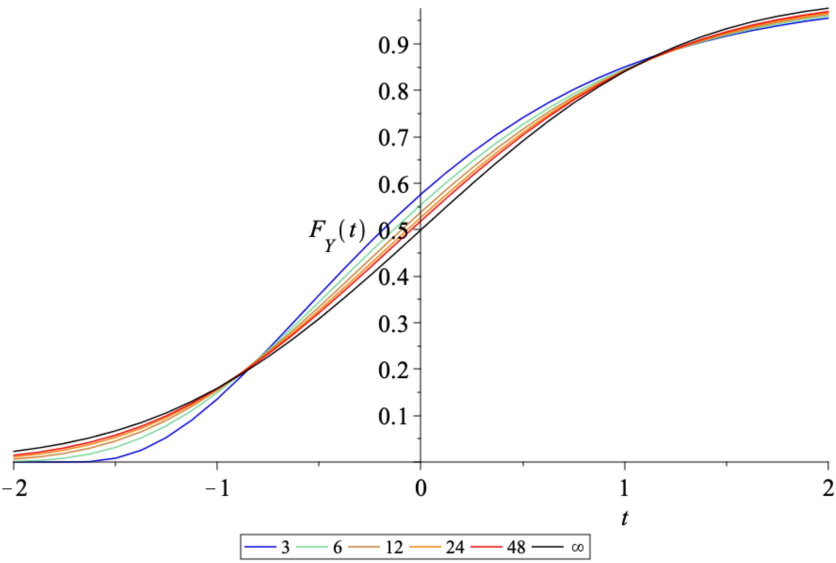
Distribuční funkce exponenciálního rozdělení Exp(2)



Hustota exponenciálního rozdělení Exp(2)

Centrální limitní věta **říká**, že **distribuční funkce** konvergují (k distribuční funkci normovaného normálního rozdělení), a to **bodově**,

$$\lim_{n\to\infty} F_{Y_n}(t) = F_{\mathrm{N}(0,1)}(t) = \Phi(t) \qquad \text{pro všechna } t \in \mathbb{R}\,.$$



Distribuční funkce normovaných součtů různých počtů nezávislých náhodných veličin

Centrální limitní věta **by mohla říkat**, že **distribuční funkce** konvergují (k distribuční funkci) **stejnoměrně**,

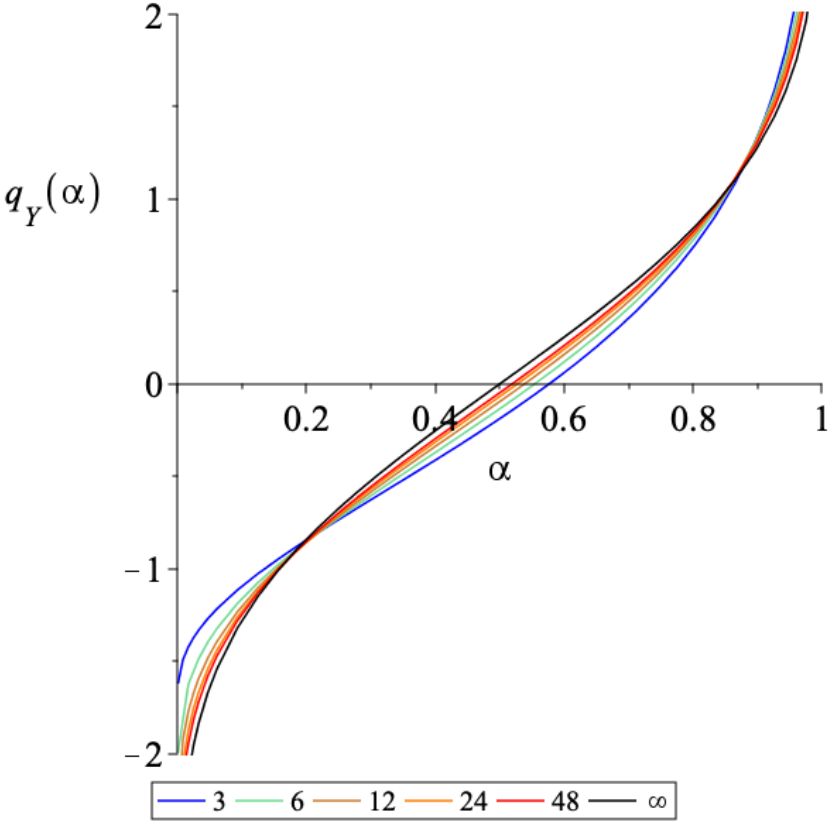
$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F_{Y_n}(t) - \Phi(t)| = 0\,,$$

ale k tomu by byl navíc **nutný předpoklad existence** 3. **centrálního momentu**.

Centrální limitní věta **by mohla říkat**, že **kvantilové funkce** konvergují (ke kvantilové funkci normovaného normálního rozdělení) a to **bodově**,

$$\lim_{n\to\infty} q_{Y_n}(\alpha) = q_{\mathrm{N}(0,1)}(\alpha) = \Phi^{-1}(\alpha) \qquad \text{pro všechna } \alpha \in (0,1)\,.$$

To je důsledek její verze pro distribuční funkce, **ve statistice se hojně používá**.



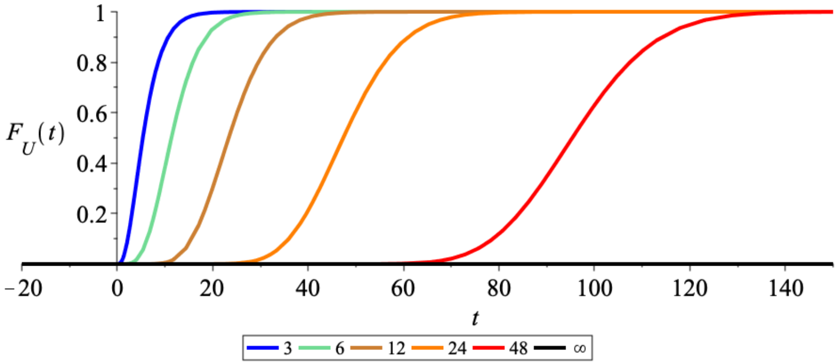
Kvantilové funkce normovaných součtů různých počtů nezávislých náhodných veličin

Co centrální limitní věta neříká

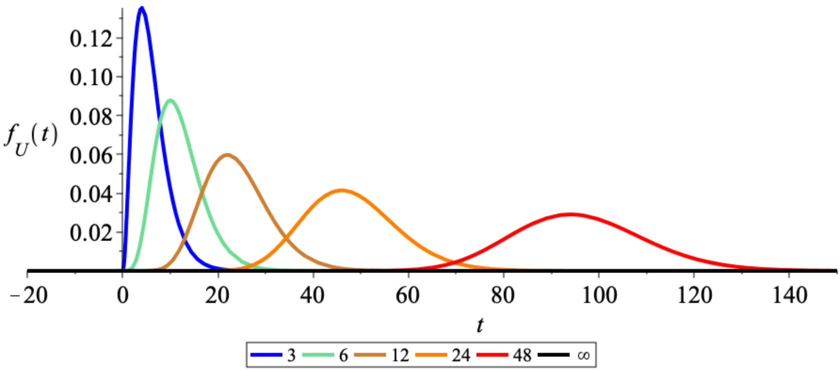
Normování (částečných součtů U_n nebo výběrových průměrů $\overline{\boldsymbol{X}}_n$, se stejným výsledkem) **nelze vynechat**.

Centrální limitní věta **neříká**, že rozdělení **součtů** $U_n = \sum_{j=1}^{n} X_j$ konvergují k normálnímu rozdělení. Např. pro $\mathrm{E}X > 0$:

$$\lim_{n\to\infty} F_{U_n}(t) = 0 \qquad \text{pro všechna } t \in \mathbb{R}\,.$$



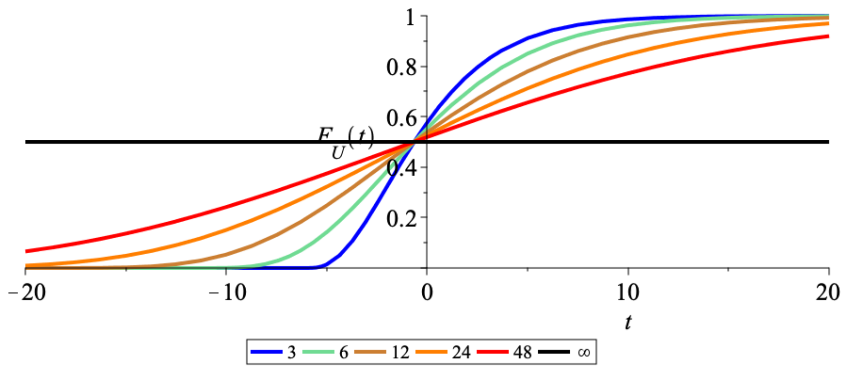
Distribuční funkce součtů různých počtů nezávislých náhodných veličin



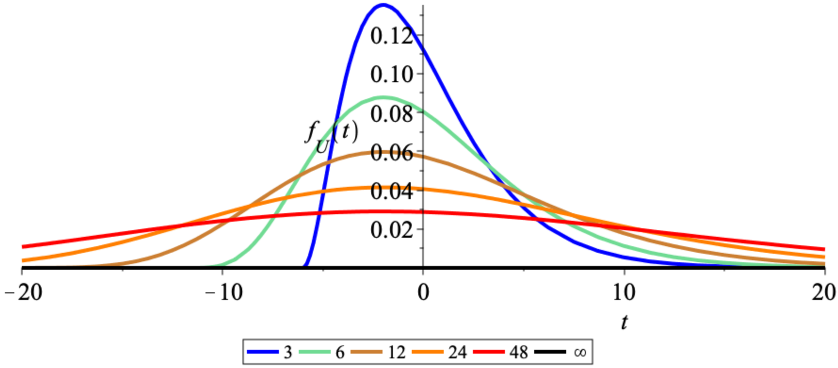
Hustoty součtů různých počtů nezávislých náhodných veličin

Centrální limitní věta **neříká**, že rozdělení **součtů s nulovou střední hodnotou**, $T_n = U_n - \mathrm{E}U_n = \sum_{j=1}^{n}(X_j - \mathrm{E}X) = n\,(\overline{\boldsymbol{X}}_n - \mathrm{E}X)$ konvergují k normálnímu rozdělení:

$$\lim_{n\to\infty} F_{T_n}(t) = \frac{1}{2} \qquad \text{pro všechna } t \in \mathbb{R}\,.$$

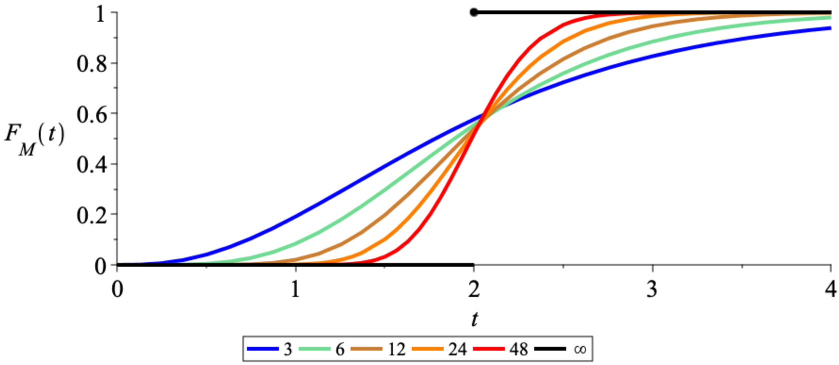


Distribuční funkce součtů různých počtů nezávislých náhodných veličin s nulovou střední hodnotou

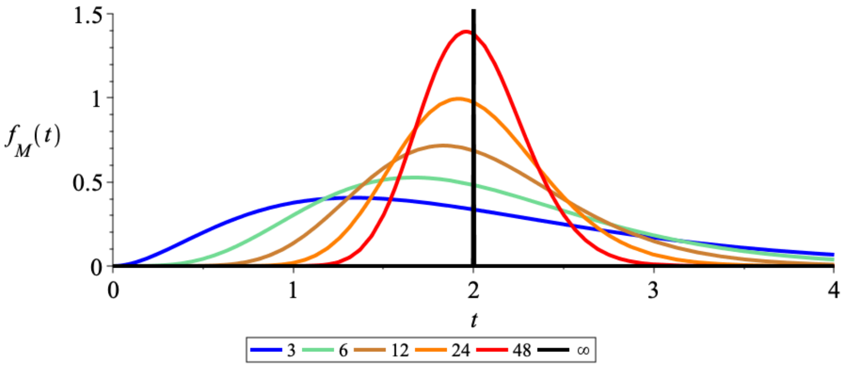


Hustoty součtů různých počtů nezávislých náhodných veličin s nulovou střední hodnotou

Centrální limitní věta **neříká**, že rozdělení **výběrových průměrů**, $\overline{\boldsymbol{X}}_n$ konvergují k normálnímu rozdělení:



Distribuční funkce výběrových průměrů různých počtů nezávislých náhodných veličin



Hustoty výběrových průměrů různých počtů nezávislých náhodných veličin

$$\lim_{n\to\infty} F_{\overline{\mathbf{X}}_n}(t) = \begin{cases} 0\,, & t < \mathrm{E}X\,, \\ 1\,, & t \geq \mathrm{E}X\,. \end{cases} \qquad \text{(Diracovo rozdělení)}$$

Kdy nelze použít centrální limitní větu

1. Pokud původní rozdělení nemá rozptyl, nebo ho má nulový (Diracovo). Takové rozdělení nelze normovat.

2. Pokud máme k dispozici jen „výběr malého rozsahu,“ takže nelze zanedbat rozdíl mezi skutečným a limitním rozdělením. V tom případě centrální limitní věta zůstává v platnosti, ale není nám nic platná, neboť hovoří o situaci příliš vzdálené od reality.