Centrální limitní věta
12.12.2017

1. Předpokládáme-li, že je v běžné populaci 10% leváků, jaká je pravděpodobnost, že mezi 100 náhodně vybranými lidmi bude více než 15 leváků?
   (a) Vyjádřete tuto pravděpodobnost přesně.
   (b) Spočtěte přibližnou hodnotu pomocí centrální limitní věty.

2. Délky telefonních hovorů paní úřednice (v min) jsou stejně rozdělené navzájem nezávislé náhodné veličiny se střední hodnotou 5 a rozptylem 25. Tento měsíc paní úřednice provedla 50 hovorů. S jakou pravděpodobností volala dohromady nejvýše 280 minut, což je objem volných minut na jejím tarifu?
   (Lze předpokládat, že se zaznamenávají přesné délky hovorů, které se na konci měsíce sečtou.)

3. Na server má přístup 100 uživatelů. Z dřívějších zkušeností víme, že uživatel má na serveru průměrně 1200MB dat, směrodatná odchylka (tj. odmocnina z rozptylu) množství dat je 400 MB. Jak velký diskový prostor potřebujeme, aby s pravděpodobností 99% nedošlo k jeho zaplnění?

4. Pojišťovna má pojištěno 1 000 osob stejného věku. Pravděpodobnost úmrtí v daném roce je u každého pojištěného 0,01. Pojištěnci platí roční pojistné 1 200 Kč a v případě úmrtí je oprávněné osobě vyplaceno 80 000 Kč.
   (a) Jaký je v daném roce očekávaný zisk pojišťovny?
   (b) Jaká je pravděpodobnost, že pojišťovna utrpí v daném roce ztrátu?

5. Na jednom z minulých cvičeních jsme odhadovali pravděpodobnost p, s jakou je náhodně vybraný student MFF levák. Zjistěte, kolik studentů musíme pro náš průzkum oslovit, aby s pravděpodobností vyšší než 0.95 nebyl náš odhad od skutečné hodnoty p dále než o 0.01? Použijte CLV a skutečnost, že pro $p \in (0, 1)$ je $p(1-p) \leq 1/4$.

Tabulka distribuční funkce a kvantilové funkce $\mathsf{N}(0,1)$

x	0.000	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
$\Phi(x)$	0.500	0.540	0.579	0.618	0.655	0.691	0.726	0.758	0.788	0.816
x	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000
$\Phi(x)$	0.540	0.579	0.618	0.655	0.691	0.726	0.758	0.788	0.816	0.841
x	1.100	1.200	1.300	1.400	1.500	1.600	1.700	1.800	1.900	2.000
$\Phi(x)$	0.864	0.885	0.903	0.919	0.933	0.945	0.955	0.964	0.971	0.977
x	2.100	2.200	2.300	2.400	2.500	2.600	2.700	2.800	2.900	3.000
$\Phi(x)$	0.982	0.986	0.989	0.992	0.994	0.995	0.997	0.997	0.998	0.999

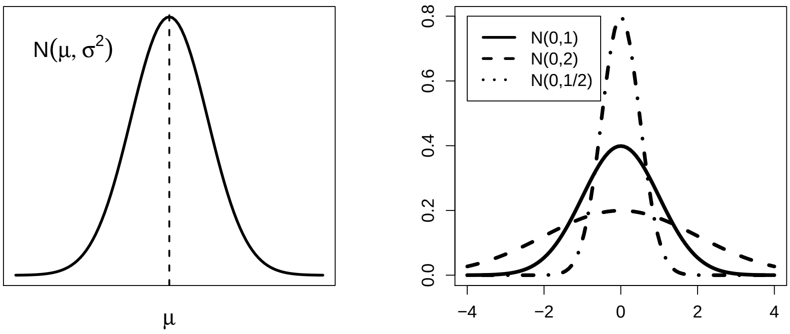
α	0.5	0.8	0.9	0.95	0.975	0.99	0.995
$\Phi^{-1}(\alpha) = q_\alpha$	0	0.842	1.282	1.645	1.960	2.326	2.576

Opakování z přednášky

Normální rozdělení. Normální rozdělení $\mathsf{N}(\mu, \sigma^2)$ má hustotu

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad x \in \mathbb{R},$$

kde $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ jsou parametry. Je-li $\mu = 0$ a $\sigma^2 = 1$, tj. $f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}$, pak se toto rozdělení nazývá **normované** normální rozdělení a značí se $\mathsf{N}(0,1)$.



- Je-li $X \sim \mathsf{N}(\mu, \sigma^2)$, pak $\mathsf{E}X = \mu$ a $\mathsf{Var}\, X = \sigma^2$. Dále pak $\dfrac{X-\mu}{\sigma} \sim \mathsf{N}(0,1)$, tj. obecné normální rozdělení můžeme snadno převést na normované $\mathsf{N}(0,1)$.
- **Distribuční funkce** rozdělení $\mathsf{N}(0,1)$ se značí jako Φ, tj. $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\} dt$. Tento určitý integrál je možné spočítat jen **numericky**, a proto hodnoty funkce Φ nalezneme **v tabulkách** (nebo dostaneme pomocí vhodného softwaru).
- Ze symetrie platí
$$\Phi(x) + \Phi(-x) = 1 \qquad \text{pro všechna } x \in \mathbb{R}.$$
- Hodnoty $\Phi^{-1}(\alpha) = q_\alpha$ (tzv. kvantily $\mathsf{N}(0,1)$) jsou také uvedeny v tabulkách a platí $q_\alpha = -q_{1-\alpha}$.
- Normální rozdělení má v pravděpodobnosti zcela zásadní význam, jak ilustruje následující **centrální limitní věta**.

Centrální limitní věta (CLV): Nechť $\{X_n\}$ je posloupnost nezávislých stejně rozdělených náhodných veličin s $0 < \mathsf{Var}\, X_1 < \infty$. Pak

$$\mathsf{P}\left(\frac{\sum_{k=1}^{n} X_k - n\mathsf{E}X_1}{\sqrt{n\,\mathsf{Var}\, X_1}} \le x \right) \xrightarrow{n\to\infty} \Phi(x), \quad x \in \mathbb{R},$$

neboli ekvivalentně

$$\mathsf{P}\left(\sqrt{n} \cdot \frac{\overline{X}_n - \mathsf{E}X_1}{\sqrt{\mathsf{Var}\, X_1}} \le x \right) \xrightarrow{n\to\infty} \Phi(x), \quad x \in \mathbb{R},$$

kde Φ je distribuční funkci normálního rozdělení $\mathsf{N}(0,1)$.
Zkráceně píšeme

$$Z_n := \frac{\sum_{k=1}^{n} X_k - n\mathsf{E}X_1}{\sqrt{n\mathsf{Var}\, X_1}} = \sqrt{n} \cdot \frac{\overline{X}_n - \mathsf{E}X_1}{\sqrt{\mathsf{Var}\, X_1}} \overset{asympt.}{\sim} \mathsf{N}(0,1)$$

a říkáme, že Z_n má asymptoticky normální rozdělení $\mathsf{N}(0,1)$.

CLV nám tedy říká, že distribuční funkce F_n veličiny Z_n se při $n \to \infty$ blíží k Φ. Pro n dost velké tedy lze uvažovat

$$\mathsf{P}(Z_n \le x) \doteq \Phi(x).$$