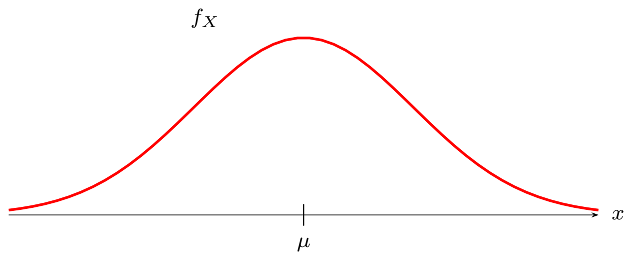
9. cvičení z PST

15. listopadu 2021

Poznámky k normálnímu rozdělení:
Veličina X má normální rozdělení $N(\mu, \sigma^2)$ (kde $\mu \in \mathbb{R}$ a $\sigma > 0$), jestliže má hustotu

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{pro} \quad x \in \mathbb{R} \; .$$



Je to tedy spojité rozdělení, $E(X) = \mu$, $D(X) = \sigma^2$ a oborem hodnot veličiny X je celá reálná osa. Všimněme si ještě, že hustota f_X je symetrická vzhledem ke středu μ a proto platí $F_X(\mu) = \frac{1}{2}$.

Toto rozdělení je limitním rozdělením, které aproximuje součty nezávislých stejně (nebo podobně) rozdělených veličin. Typicky se tedy objevuje u veličin, jejichž hodnoty jsou ovlivněny mnoha drobnými odchylkami (např. u chyb měření, výšky člověka apod.)

U zmíněné výšky člověka (která může být samozřejmě jen kladná) nebo u veličin s hodnotami omezenými na nějaký interval, je přesto použití normálního rozdělení (které může nabývat libovolných hodnot) přiměřené. Je to tím, že u dané veličiny Y předpokládáme aproximaci pomocí normálního rozdělení obvykle jen ve vhodném okolí kolem střední hodnoty $\mu := E(Y)$. Je to podobná situace, jako když aproximujeme funkci pomocí jejího Taylorova polynomu v okolí daného bodu.

Přesněji to vystihuje toto tvrzení:

Věta: Nechť Y je veličina s hustotou f_Y, střední hodnotou μ a rozptylem $\sigma^2 \neq 0$. Nechť $X \sim N(\mu, \sigma^2)$. Jestliže se hustoty f_X a f_Y rovnají na nějakém intervalu $(a, b) \subseteq \mathbb{R}$ takovém, že $\mu \in (a, b)$ a pokud $F_Y(\mu) = \frac{1}{2}$, pak

$$F_Y(t) = F_X(t) \quad \text{pro všechna} \quad t \in (a, b) \; .$$

Důležité vlastnosti normálního rozdělení:

- $X \sim N(\mu, \sigma^2) \;\Rightarrow\; Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$ (je to tzv. normované normální rozdělení s hodnotami v tabulkách) dist. funkce pro $N(0, 1)$ se značí Φ.

  V tomto případě pak máme $F_X(t) = P(X \leq t) = P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{=Y} \leq \frac{t-\mu}{\sigma}\Big) = \Phi\Big(\frac{t-\mu}{\sigma}\Big)$ pro všechna $t \in \mathbb{R}$.
- hustota $f_{N(0,1)}$ je symetrická $\;\Rightarrow\;$ $\Phi(t) + \Phi(-t) = 1$ pro všechna $t \in \mathbb{R}$ a dále $\Phi^{-1}(\alpha) = -\Phi^{-1}(1 - \alpha)$ pro $\alpha \in (0, 1)$.
- Nechť $X_i \sim N(\mu_i, \sigma_i^2)$, pro $i = 1, 2$, jsou nezávislé. Pak $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (tj. speciálně součet nezávislých normálních rozdělení je zase normální.)

Pro lepší představu o tom, jakou roli pro veličinu s normálním rozdělením $X \sim N(\mu, \sigma^2)$ hraje směrodatná odchylka σ se používá tzv.

pravidlo tří-sigma (`https://cs.wikipedia.org/wiki/Pravidlo_t\%C5\%99\%C3\%AD\_sigma`)

které je ovšem čistě jen technickou pomůckou:

Jestliže si budeme počítat pravděpodobnosti

$$P(|X - \mu| \leq k \cdot \sigma) = P\Big(\Big|\underbrace{\frac{X-\mu}{\sigma}}_{\sim N(0,1)}\Big| \leq k\Big) = \Phi(k) - \Phi(-k) = 2 \cdot \Phi(k) - 1 \quad \text{pro} \quad k = 1, 2, 3, \ldots$$

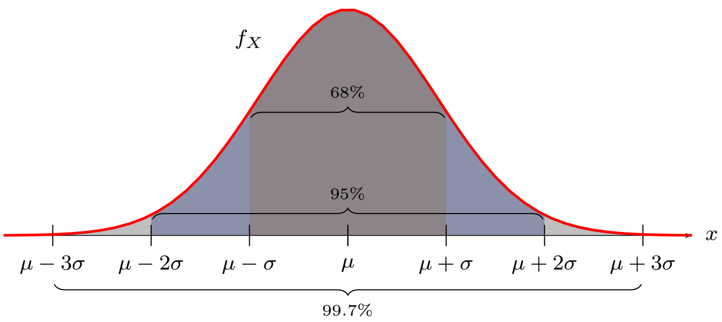
dostaneme postupně

$P(|X - \mu| \leq \sigma) = 2 \cdot \Phi(1) - 1 \doteq 2 \cdot 0.8413 - 1 = 0.6826 \doteq 68\%$

$P(|X - \mu| \leq 2 \cdot \sigma) = 2 \cdot \Phi(2) - 1 \doteq 2 \cdot 0.9772 - 1 = 0.9544 \doteq 95\%$

$P(|X - \mu| \leq 3 \cdot \sigma) = 2 \cdot \Phi(3) - 1 \doteq 2 \cdot 0.99865 - 1 = 0.9973 \doteq 99.7\%$

Pro vyšší hodnoty, tj. $k \geq 4$ už jsou pravděpodobnosti vpodstatě rovny 1, takže se v praxi příliš nepoužívají (záleží samozřejmě na zvolené přesnosti).



9.1 (normální rozdělení)

Bylo zjištěno, že zpoždění odjezdu autobusu ze zastávky se přibližně řídí normálním rozdělení se střední hodnotou 10 min a rozptylem 25 min^2. Spočtěte:

(a) pravděpodobnost, že autobus bude mít zpoždění více než 20 min a pravděpodobnost, že autobus odjede dříve než má.

(b) pravděpodobnost, že autobus bude mít zpoždění více než 20 min, jestliže už má zpoždění 15 min.

(c) čas, kdy bychom měli být na zastávce, aby nám autobus neujel alespoň na 90%.

Řešení:
Označme si $X =$ *"čas zpoždění odjezdu autobusu ze zastávky"*. Podle zadání je $X \sim N(\mu, \sigma^2)$, kde $\mu = 10$ min a $\sigma^2 = 25$ min^2.

(a) $P(X > 20\,\text{min}) = 1 - P(X \leq 20) = 1 - F_X(20) = 1 - \Phi\left(\frac{20-10}{\sqrt{25}}\right) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$.
$P(X < 0\,\text{min}) = F_X(0) = \Phi\left(\frac{0-10}{\sqrt{25}}\right) = \Phi(-2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$.

(b) $P\left(X > 20\,\text{min} | X > 15\,\text{min}\right) = \frac{P(X>20\ \&\ X>15)}{P(X>15)} = \frac{P(X>20)}{P(X>15)} = \frac{0.0228}{1-\Phi\left(\frac{15-10}{\sqrt{25}}\right)} = \frac{0.0228}{1-\Phi(1)} =$

$= \frac{0.0228}{1-0.8413} = \frac{0.0228}{0.1587} = 0.1437$.

(c) Jestliže t_0 bude čas odjezdu autobusu podle jízdního řádu, pak veličina $Y = X + t_0$ představuje čas skutečného odjezdu autobusu. Označme si $t =$ *"čas, kdy přicházíme na zastávku"*. Chceme, aby $P(Y \geq t) = 0.9$. Dostaneme tedy

$$0.9 = P(Y \geq t) = P(X + t_0 \geq t) = P(X \geq \underbrace{t - t_0}_{\Delta t}) = 1 - F_X(\Delta t) = 1 - \Phi\left(\frac{\Delta t - 10}{\sqrt{25}}\right)$$

$$\Phi\left(\frac{\Delta t - 10}{\sqrt{25}}\right) = 0.1$$

$$\frac{\Delta t - 10}{\sqrt{25}} = \Phi^{-1}(0.1) = -\underbrace{\Phi^{-1}(0.9)}_{1.282} \quad \Longrightarrow \quad \Delta t = 10 - 5 \cdot 1.282 = 3.59\ \text{min}\ .$$

Tedy $t = t_0 + 3.59$, neboli stačí přijít na zastávku nejpozději 3.59 min po čase, kdy má autobus pravidelně odjíždět podle jízdního řádu.

9.2 (normální rozdělení)

Rychlost aut v úseku, kde je omezení na maximální povolenou rychlost 50 km/hod, je náhodná veličina X, která má normální rozdělení $N(\mu, \sigma^2)$. Z naměřených hodnot vyplývá, že

$$P(X > 60) = 0.45 \quad \text{a} \quad P(X > 70) = 0.2 \; .$$

Určete

(a) parametry μ a σ^2;

(b) pravděpodobnost $P(X < 50)$, tj. že v daném úseku je dodržována maximální povolená rychlost.

Řešení:

(a) Máme

$$\begin{aligned} 0.45 = P(X > 60) = 1 - P(X \le 60) = 1 - F_X(60) &= 1 - \Phi\Big(\frac{60-\mu}{\sigma}\Big) \\ \Phi\Big(\frac{60-\mu}{\sigma}\Big) &= 1 - 0.45 = 0.55 \\ \frac{60-\mu}{\sigma} &= \Phi^{-1}(0.55) \\ 60 &= \Phi^{-1}(0.55)\cdot\sigma + \mu \end{aligned}$$

a podobně

$$\begin{aligned} 0.2 = P(X > 70) = 1 - P(X \le 70) = 1 - F_X(70) &= 1 - \Phi\Big(\frac{70-\mu}{\sigma}\Big) \\ \Phi\Big(\frac{70-\mu}{\sigma}\Big) &= 1 - 0.2 = 0.8 \\ \frac{70-\mu}{\sigma} &= \Phi^{-1}(0.8) \\ 70 &= \Phi^{-1}(0.8)\cdot\sigma + \mu \end{aligned}$$

Hodnoty funkce Φ^{-1} (tj. kvantilové funkce) zjistíme pomocí tabulek (nebo softwaru). Zde máme $\Phi^{-1}(0.55) \doteq 0.126$ a $\Phi^{-1}(0.8) \doteq 0.842$. Máme tedy (lineární) soustavu rovnic

$$\begin{aligned} 60 &= 0.126\cdot\sigma + \mu \\ 70 &= 0.842\cdot\sigma + \mu \end{aligned}$$

Odsud odečtením získáme $70 - 60 = (0.842 - 0.126)\cdot\sigma$ tedy

$$\begin{aligned} \sigma &= \frac{10}{0.716} \doteq 13.97 \\ \mu &= 70 - 0.842\cdot\sigma \doteq 70 - 0.842\cdot 13.97 \doteq 58.24 \end{aligned}$$

(b) Z vypočítaných μ a σ dostaneme

$$P(X < 50) = F_X(50) = \Phi\Big(\frac{50-\mu}{\sigma}\Big) = \Phi\Big(\frac{50-58.24}{13.97}\Big) \doteq$$

$$\doteq \Phi\Big(-0.59\Big) = 1 - \Phi(0.59) \doteq 1 - 0.722 = 0.278 \; .$$

Připomenutí: Pro náhodnou veličinu X s konečným rozptylem, položme

$$\mathrm{norm}(X) := \frac{X - E(X)}{\sqrt{D(X)}} \ .$$

Speciálně tedy vidíme, že $E\big(\mathrm{norm}(X)\big) = 0$ a $D\big(\mathrm{norm}(X)\big) = 1$.

Platí: Pro takovouto veličinu X a konstanty $a > 0$ a $b \in \mathbb{R}$ je

$$\mathrm{norm}(aX + b) = \mathrm{norm}(X) \ .$$

Centrální limitní věta (CLV):

Nechť X_i, pro $i = 1, 2, \ldots$ je posloupnost nezávislých náhodných veličin, které mají stejná rozdělení se střední hodnotou μ a (konečným) rozptylem σ^2. Pak pro veličiny

$$Z_n = \sum_{i=1}^{n} X_i$$

platí, že

$$\lim_{n\to\infty} P\Big(\mathrm{norm}(Z_n) \leq t\Big) = \Phi(t) \quad \text{pro každé} \quad t \in \mathbb{R} \ .$$

Neboli: pro velká n má veličina $\mathrm{norm}(Z_n)$ přibližně normální rozdělení $N(0, 1)$.

Centrální limitní větu můžeme formulovat (namísto pro Z_n) také pro tzv. výběrový průměr, tj. veličiny

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \cdot Z_n \ .$$

protože pro ně platí $\mathrm{norm}(\overline{X}_n) = \mathrm{norm}(Z_n)$.

- -

Rychlost konvergence v CLV: Pokud pro veličiny X_i v CLV navíc ještě je $\varrho := E(|X_i - \mu|^3) < \infty$, pak platí Berry–Esseenův odhad chyby (pro všechna $t \in \mathbb{R}$ a $n \in \mathbb{N}$):

$$\Big|F_{\mathrm{norm}(Z_n)}(t) - \Phi(t)\Big| < 0.4748 \cdot \frac{\varrho}{\sigma^3 \sqrt{n}}$$

- -

Odhad chyby v CLV pro alternativní rozdělení: Pokud mají veličiny X_i alternativní rozdělení s parametrem p, tj. $P(X_i = 1) = p$, pak

$$\mu = E(X_i) = p, \qquad \sigma = \sqrt{D(X_i)} = \sqrt{p(1-p)}, \qquad \varrho = E\left(|X_i - \mu|^3\right) = p(1-p)\big(p^2 + (1-p)^2\big) = \sigma^2\big(p^2 + (1-p)^2\big)$$

čímž pro binomické rozdělení $Z_n = \sum_{i=1}^{n} X_i \sim \mathrm{Bi}(n, p)$ dostáváme odhad

$$\Big|F_{\mathrm{norm}(Z_n)}(t) - \Phi(t)\Big| \quad < \quad 0.4748 \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}} = 0.4748 \cdot \frac{p^2 + (1-p)^2}{\sqrt{D(Z_n)}} \ .$$

Aproximace CLV se obvykle používá pro $D(Z_n) \geq 9$. Pak je odhad chyby nejvýše: $\Big|F_{\mathrm{norm}(Z_n)}(t) - \Phi(t)\Big| \ < \ \frac{0.4748}{\sqrt{9}} = 0.159$.

Odhad chyby v CLV pro rovnoměrné rozdělení: Pokud mají veličiny X_i rovnoměrné rozdělení na intervalu $\langle a, b\rangle$, pak

$$\mu = E(X_i) = \tfrac{a+b}{2}, \qquad \sigma = \sqrt{D(X_i)} = \sqrt{\tfrac{(b-a)^2}{12}} = \tfrac{b-a}{2\sqrt{3}}, \qquad \varrho = E\left(|X_i - \mu|^3\right) = \tfrac{(b-a)^3}{32}$$

čímž pro $Z_n = \sum_{i=1}^{n} X_i$ dostáváme odhad

$$\Big|F_{\mathrm{norm}(Z_n)}(t) - \Phi(t)\Big| \quad < \quad 0.4748 \cdot \frac{3\sqrt{3}}{4\sqrt{n}} \quad < \quad \frac{0.62}{\sqrt{n}} \ .$$

Obvyklý způsob použití CLV: Veličina norm(Z_n) = norm$(\overline{X}_n)$ má přibližně rozdělení $N(0,1)$. Pro výpočty se tedy užívá, že

- veličina Z_n se střední hodnotou $E(Z_n) = n\mu$ a rozptylem $D(Z_n) = n\sigma^2$ má přibližně rozdělení $N(n\mu, n\sigma^2)$,
- veličina $\overline{X}_n$ se střední hodnotou $E(\overline{X}_n) = \mu$ a rozptylem $D(\overline{X}_n) = \frac{\sigma^2}{n}$ má přibližně rozdělení $N(\mu, \frac{\sigma^2}{n})$.

Důležitá poznámka: Veličina X, která počítá počet úspěchů v n pokusech s pravděpodobností úspěchu p, má binomické rozdělení Bi(n,p) a dá se zapsat jako $X = \sum_{i=1}^{n} X_i$, kde X_i jsou nezávislé veličiny a alternativním rozdělením Alt(p) popisujícími úspěch v i-tém pokusu.

9.3 (alternativní rozdělení, odhad pravděpodobnosti - CLV)

Pravděpodobnost toho, že se za dobu T porouchá přístroj je $p = 0.2$. S jakou pravděpodobností se za dobu T ze 100 (nezávisle pracujících) přístrojů porouchá

(a) alespoň 20,

(b) méně než 28,

(c) 14 až 26 přístrojů?

Řešení:

Pro $i = 1, \ldots, n$ (kde $n = 100$) si zavedeme veličiny

$$X_i = \begin{cases} 1 & , \text{ } i\text{-tý přístroj se porouchá,} \\ 0 & , \text{ } i\text{-tý přístroj bude v pořádku.} \end{cases}$$

Veličiny X_i budou nezávislé s alternativním rozdělením Alt(p) = Alt(0.2), protože $P(X_i = 1) = 0.2$. Počet porouchaných přístrojů je tedy veličina

$$Z = \sum_{i=1}^{100} X_i$$

která má tudíž binomické rozdělení Bi(n,p) = Bi$(100, 0.2)$. To sice umíme přesně popsat, ale vyčíslování součtu mnoha velmi malých členu by vedlo ke značným numerickým chybám (a bez softwaru by ani nebylo možné). Proto použijeme CLV, která velmi dobře aproximuje hledané pravděpodobnosti.

Pro $Z = \sum_{i=1}^{n} X_i$ tedy máme

$$\begin{aligned} E(Z) &= n \cdot E(X_1) = n \cdot p = 100 \cdot 0.2 = 20 \\ D(Z) &= n \cdot D(X_1) = n \cdot p \cdot (1-p) = 100 \cdot 0.2 \cdot 0.8 = 16 \\ &\Rightarrow \quad \sqrt{D(Z)} = \sqrt{16} = 4 \end{aligned}$$

Podle CLV můžeme předpokládat, že veličina $norm(Z) = \frac{Z - E(Z)}{\sqrt{D(Z)}} = \frac{Z-20}{\sqrt{16}}$ má přibližně normované normální rozdělení $N(0,1)$.

To také můžeme chápat tak, že veličina Z má *přibližně* normální rozdělení

$$N\Big(E(Z), D(Z)\Big) = N(20, 4^2)$$

tedy že

$$F_Z(t) \doteq \Phi\left(\frac{t-20}{\sqrt{16}}\right) \quad \text{pro} \quad t \in \mathbb{R}\ .$$

.

Pak tedy máme:

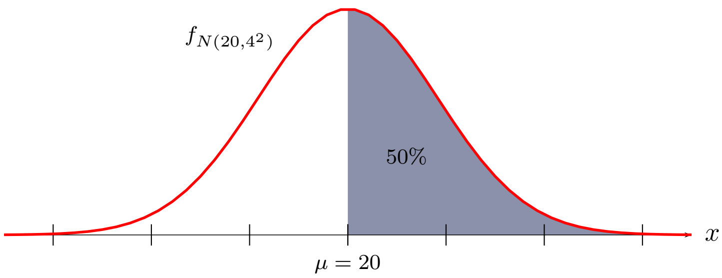
(a)

$$P(Z \geq 20) = P\left(\frac{Z-20}{\sqrt{16}} \geq \frac{20-20}{\sqrt{16}}\right) = P\big(norm(Z) \geq 0\big) =$$

$$= 1 - P\big(norm(Z) < 0\big) \overset{(CLV)}{\doteq} 1 - \Phi\left(0\right) = 1 - 0.5 = \mathbf{0.5}\ .$$

(Pro srovnání: skutečná hodnota pro binomické rozdělení je **0.5398**.)

Když budeme uvažovat (spojité) rozdělení $N(20, 4^2)$, které ALE POUZE aproximuje *původní diskrétní* binomické rozdělení veličiny Z, můžeme si představit hledanou pravděpodobnost (z HLEDISKA VÝPOČTU) takto:



(b)

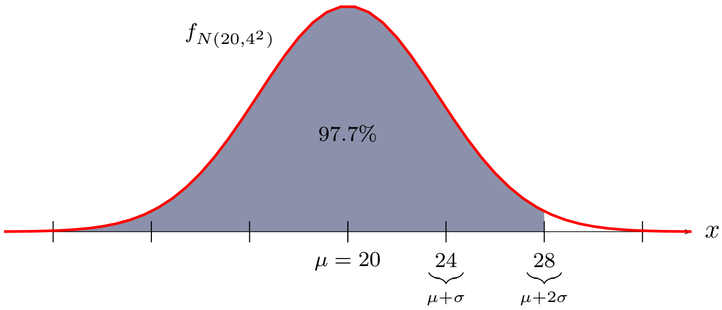
$$P(Z < 28) = P\left(\frac{Z-20}{\sqrt{16}} < \frac{28-20}{\sqrt{16}}\right) = P\Big(norm(Z) < 2\Big) \overset{(CLV)}{\doteq}$$

$$\overset{(CLV)}{\doteq} \Phi(2) \doteq \mathbf{0.977}\ .$$

(Pro srovnání: skutečná hodnota pro binomické rozdělení je **0.9658**.)

Když opět budeme uvažovat (spojité) rozdělení $N(20, 4^2)$, které POUZE aproximuje *původní diskrétní* binomické rozdělení veličiny Z a vezmeme do úvahy pravidlo tří sigma, můžeme uvažovat (z HLEDISKA VÝPOČTU) takto:

Protože $28 = \mu + 2\sigma$, tak hodnota $P(Z < 28) \doteq P\Big(N(20, 4^2) < 28\Big)$ nám musí vyjít větší než 95% (viz náčrt).



(c)

$$P(14 \le Z \le 26) = P\left(\frac{14-20}{\sqrt{16}} \le \frac{Z-20}{\sqrt{16}} \le \frac{26-20}{\sqrt{16}}\right) = P\Big(-1.5 \le norm(Z) \le 1.5\Big) =$$

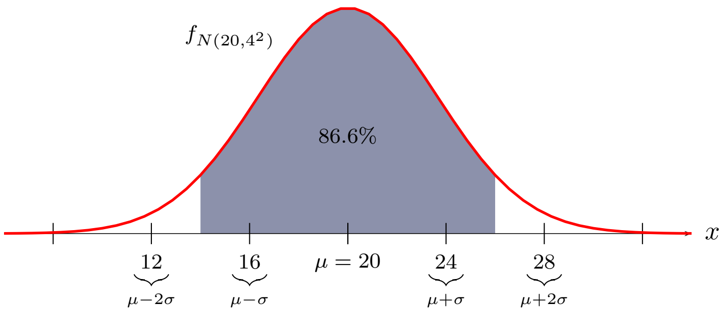
$$= P\Big(norm(Z) \le 1.5\Big) - P\Big(norm(Z) < -1.5\Big) \overset{(CLV)}{\doteq} \Phi(1.5) - \Phi(-1.5) =$$

$$= 2 \cdot \Phi(1.5) - 1 \doteq 2 \cdot 0.933 - 1 = \mathbf{0.866} \ .$$

(Pro srovnání: skutečná hodnota pro binomické rozdělení je **0.8973**.)

Když opět budeme uvažovat (spojité) rozdělení $N(20, 4^2)$, které POUZE aproximuje *původní diskrétní* binomické rozdělení veličiny Z a vezmeme do úvahy pravidlo tří sigma, můžeme uvažovat (z HLEDISKA VÝPOČTU) takto:

Protože $\mu - 2\sigma = 12 < 14 < 16 = \mu - \sigma$ a $\mu + \sigma = 24 < 16 < 28 = \mu + \sigma$, tak hodnota $P(14 \le Z \le 26) \doteq P\Big(14 \le N(20, 4^2) \le 26\Big)$ nám musí vyjít mezi 68% a 95% (viz náčrt).



Odhad chyby binomického rozdělení $Z \sim \mathrm{Bi}(100, 0.2)$ pomocí aproximace CLV je

$$\Big|F_{\mathrm{norm}(Z)}(t) - \Phi(t)\Big| < 0.4748 \cdot \frac{0.2^2 + 0.8^2}{\sqrt{16}} = 0.081 \ .$$

9.4 (spojité rozdělení, odhad pravděpodobnosti a intervalu - CLV)

Doba výdrže baterie má střední hodnotu 5.7 hodin a směrodatnou odchylku 2 hodiny. Nakoupili jsme sto baterií, používáme vždy jednu a po vybití vyměníme.

(a) Jaká je pravděpodobnost, že nám sto baterií vydrží dohromady alespoň 550 hodin?

(b) Určete dobu, u které se můžeme na 90 % spolehnout, že nám daných sto baterií vydrží.

Řešení:
Budeme postupovat podobně jako v příkladu **9.3**. K řešení opět použijeme centrální limitní větu. Označme si tedy pro $i = 1, 2, ..., n$, kde $n = 100$, veličiny

$$X_i = \text{“}\textit{doba výdrže i-té baterie}\text{”} \ \ [\mathrm{h}]$$

které budou nezávislé. Máme $\mu = E(X_i) = 5.7$ h, $\sigma = \sqrt{D(X_i)} = 2$ h.

Zajímá nás veličina $Z =$ *“doba výdrže série 100 baterií”* $= \sum_{i=1}^{n} X_i$. Pro ni máme

$$E(Z) = n \cdot E(X_1) = 100 \cdot 5.7 = 570$$
$$D(Z) = n \cdot D(X_1) = 100 \cdot 2^2 = 400$$

O rozdělení X_i nic nevíme. Jediným nástrojem nám tak zůstává CLV. Z podstaty zadání můžeme ale předpokládat, že X_i mají přibližně normální rozdělení a $n = 100$ je již dostatečně vysoké, abychom mohli odhad pomocí CLV použít. (V případě přesně normálního rozdělení bychom ani CLV nepotřebovali.)

(a)

$$P(Z \geq 550) = P\left(\frac{Z - E(Z)}{\sqrt{D(Z)}} \geq \frac{550 - 570}{\sqrt{400}} \right) = P\Big(norm(Z) \geq -1\Big) = 1 - P\Big(norm(Z) < -1\Big) \overset{(CLV)}{\doteq}$$
$$\overset{(CLV)}{\doteq} \ 1 - \Phi(-1) = \Phi(1) = \mathbf{0.8413} \ .$$

(b) Hledáme t tak, aby $P(Z \geq t) = 0.9$. Máme tedy podobně jako výše

$$0.9 = P(Z \geq t) = P\left(\frac{Z - E(Z)}{\sqrt{D(Z)}} \geq \frac{t - 570}{\sqrt{400}} \right) = P\Big(norm(Z) \geq \tfrac{t-570}{20}\Big) = 1 - P\Big(norm(Z) < \tfrac{t-570}{20}\Big) \overset{(CLV)}{\doteq}$$
$$\overset{(CLV)}{\doteq} \ 1 - \Phi(\tfrac{t-570}{20})$$

Máme tudíž

$$\Phi\Big(\tfrac{t-570}{20}\Big) \doteq 1 - 0.9 = 0.1$$
$$\tfrac{t-570}{20} \doteq \Phi^{-1}(0.1) = -\Phi^{-1}(0.9) \doteq -1.282$$
$$t \doteq 570 - 20 \cdot 1.282 \doteq \mathbf{544.36} \text{ h.}$$

9.5 (spojité rovnoměrné rozdělení, odhad intervalu - CLV)

Zaokrouhlovací chyba na celé jednotky má rovnoměrné rozložení na intervalu $(-0.5, 0.5)$. Uvažujme součet 100 (nezávislých) zaokrouhlovacích chyb,

(a) Spočtěte pravděpodobnost, že tento součet bude v absolutní hodnotě menší než 4.

(b) Určete $\varepsilon > 0$ tak, aby 90% interval spolehlivosti pro S byl $(-\varepsilon, \varepsilon)$.

Řešení:

K řešení opět použijeme centrální limitní větu. Označme si tedy pro $i = 1, 2, ..., n$, kde $n = 100$, veličiny

$$X_i = \text{“}\textit{hodnota } i\textit{-té zaokrouhlovací chyby}\text{”}$$

které budou nezávislé a mají rovnoměrné rozdělení $\mathrm{Ro}\,(a, b) = \mathrm{Ro}(-0.5, 0.5)$.

Zajímá nás veličina $Z = \sum_{i=1}^{n} X_i$. Pro ni máme

$$E(X_i) = \frac{a+b}{2} = \frac{-0.5+0.5}{2} = 0 \quad \Rightarrow \quad E(Z) = n \cdot E(X_1) = 100 \cdot 0 = 0$$

$$D(X_i) = \frac{(b-a)^2}{12} = \frac{\big(0.5-(-0.5)\big)^2}{12} = \frac{1}{12} \quad \Rightarrow \quad D(Z) = n \cdot D(X_1) = 100 \cdot \frac{1}{12} = \frac{25}{3}$$

$$\Rightarrow \quad \sqrt{D(Z)} = \sqrt{\frac{25}{3}} = \frac{5}{\sqrt{3}}\,.$$

Podle CLV bude mít veličina $norm(Z) = \frac{Z-E(Z)}{\sqrt{D(Z)}} = \frac{\sqrt{3}}{5}Z$ přibližně rozdělení $N(0, 1)$.

(a)

$$P(|Z| \le 4) = P\Big(\Big| \underbrace{\tfrac{\sqrt{3}}{5}Z}_{norm(Z)} \Big| \le \frac{\sqrt{3}}{5} \cdot 4 \Big) = P\Big(|norm(Z)| \le \underbrace{\tfrac{\sqrt{3}}{5} \cdot 4}_{\doteq 1.386} \Big) \overset{(CLV)}{\doteq}$$

$$\overset{(CLV)}{\doteq} \Phi(1.386) - \Phi(-1.386) = 2 \cdot \Phi(1.386) - 1 \doteq 2 \cdot 0.917 - 1 = \mathbf{0.834}\,.$$

(b) Hledáme $\varepsilon > 0$ tak, aby $P(|Z| < \varepsilon) = 0.9$. Máme tedy podobně jako výše

$$0.9 = P(|Z| < \varepsilon) = P\Big(\Big| \underbrace{\tfrac{\sqrt{3}}{5}Z}_{norm(Z)} \Big| < \frac{\sqrt{3}}{5} \cdot \varepsilon \Big) = P\Big(|norm(Z)| < \frac{\sqrt{3}}{5} \cdot \varepsilon\Big) \overset{(CLV)}{\doteq}$$

$$\overset{(CLV)}{\doteq} \Phi(\tfrac{\sqrt{3}}{5} \cdot \varepsilon) - \Phi(-\tfrac{\sqrt{3}}{5} \cdot \varepsilon) = 2 \cdot \Phi(\tfrac{\sqrt{3}}{5} \cdot \varepsilon) - 1$$

Máme tudíž

$$\Phi\Big(\tfrac{\sqrt{3}}{5} \cdot \varepsilon\Big) \doteq \frac{1+0.9}{2} = 0.95$$

$$\varepsilon \doteq \tfrac{5}{\sqrt{3}} \cdot \Phi^{-1}(0.95) \doteq \tfrac{5}{\sqrt{3}} \cdot 1.645 \doteq \mathbf{4.75}\,.$$

Odhad chyby rozdělení veličiny Z pomocí aproximace CLV je

$$\Big|F_{\mathrm{norm}(Z)}(t) - \Phi(t)\Big| < \frac{0.62}{\sqrt{n}} = \frac{0.62}{\sqrt{100}} = 0.062\,.$$

Pro nezávislé veličiny $U_i \sim \mathrm{Ro}(0, 1)$ má veličina $Y = \sum_{i=1}^{n} U_i$ tzv. Irwin-Hallovo rozdělení Ir-Ha(n).

V našem případě máme $X_i + 0.5 \sim \mathrm{Ro}(0, 1)$, takže $Z + \frac{n}{2} = \sum_{i=1}^{n} (X_i + 0.5) \sim$ Ir-Ha(n).

9.6 (spojité rovnoměrné rozdělení, odhad intervalu - CLV a Čebyševova nerovnost)

Chodíme náhodně k tramvaji, která jezdí po 6 minutách. Jezdíme dvakrát denně, 20 dní v měsíci. Jestliže označíme X_i náhodnou veličinu dobu čekání (při i-tém příchodu) a T průměrnou dobu čekání za měsíc, pak

(a) vypočtěte pomocí centrální limitní věty a

(b) odhadněte pomocí Čebyševovy nerovnosti

číslo $\varepsilon > 0$ tak, aby $P\Big(|T-3| < \varepsilon\Big) = 0.9$.

Řešení:

Předpokládáme, že tramvaje jezdí vždy přesně po 6 minutách, ale zato naše příchody jsou náhodné a rovnoměrně rozdělené v nějakém časovém intervalu (jehož délka je celočíselným násobkem 6 minut). Tedy doba čekání X_i má rovnoměrné rozdělení v intervalu $\langle 0, 6\rangle$ (v minutách).

Jednotlivé veličiny X_i považujeme za nezávislé, se střední hodnotou (která je uprostřed daného intervalu)

$$E(X_i) = \frac{0+6}{2} = 3 \ \text{ min}$$

a rozptylem

$$D(X_i) = E\Big((X_i - E(X_i))^2\Big) = E\Big((X_i-3)^2\Big) =$$

$$= \int_{-\infty}^{\infty} (t-3)^2 f_X(t)\ \mathrm{d}t = \int_0^6 (t-3)^2 \cdot \tfrac{1}{6}\ \mathrm{d}t = \left[\tfrac{(t-3)^3}{18}\right]_0^6 = 3 \ \text{ min}^2 \ .$$

.

Celkový počet příchodů na zastávku během měsíce je $n = 2 \cdot 20 = 40$. Průměrná doba čekání za měsíc pak je

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i$$

se střední hodnotou: $E(T) = E(X_i) = 3 \ \text{ min}$

a rozptylem: $D(T) = \frac{D(X_i)}{n} = \frac{3}{40} = 0.075 \ \text{ min}^2$.

(a) Odhad ε pomocí centrální limitní věty - veličina $\text{norm}(T) = \frac{T-E(T)}{\sqrt{D(T)}} = \frac{T-3}{\sqrt{0.075}}$ má přibližně normované normální rozdělení $N(0,1)$:

$$0.9 = P\Big(|T-3|<\varepsilon\Big) = P\Big(\Big|\underbrace{\tfrac{T-3}{\sqrt{0.075}}}_{\text{norm}(T)}\Big| < \frac{\varepsilon}{\sqrt{0.075}}\Big) = P\Big(-\frac{\varepsilon}{\sqrt{0.075}} < \text{norm}(T) < \frac{\varepsilon}{\sqrt{0.075}}\Big) \doteq$$

$$\doteq \Phi\Big(\frac{\varepsilon}{\sqrt{0.075}}\Big) - \Phi\Big(-\frac{\varepsilon}{\sqrt{0.075}}\Big) = 2\cdot\Phi\Big(\frac{\varepsilon}{\sqrt{0.075}}\Big) - 1 \ .$$

Dostaneme tedy (přibližnou) rovnost

$$0.9 = 2\cdot\Phi\Big(\frac{\varepsilon}{\sqrt{0.075}}\Big) - 1$$

$$\Phi\Big(\frac{\varepsilon}{\sqrt{0.075}}\Big) = \frac{1+0.9}{2} = 0.95$$

a konečně

$$\varepsilon = \sqrt{0.075}\cdot\Phi^{-1}(0.95) \doteq 0.2739 \cdot 1.645 \doteq 0.4505 \ .$$

Odhad chyby rozdělení veličiny T pomocí aproximace CLV je

$$\left|F_{\mathrm{norm}(T)}(t) - \Phi(t)\right| < \frac{0.62}{\sqrt{n}} = \frac{0.62}{\sqrt{40}} \doteq 0.1 \ .$$

(b) Odhad pomocí Čebyševovy nerovnosti:

$$0.9 = P\left(\left|T-3\right| < \varepsilon\right) \geq 1 - \frac{D(T)}{\varepsilon^2} = 1 - \frac{0.075}{\varepsilon^2}$$

$$\frac{0.075}{\varepsilon^2} \geq 0.1$$

$$\varepsilon \leq \sqrt{0.75} \doteq 0.866 \ .$$

9.7 (CLV - odhad počtu)

Na oboru má studovat 600 studentů. Přibližně jen 2/3 z přijatých studentů se pak nakonec zapíše na studium. Kolik studentů by se mělo přijmout, aby počet zapsaných byl co největší, ale aby překročil 600 s pravděpodobností nejvýše 5%? Jaký pak bude průměrný počet zapsaných studentů?

Řešení:
Nechť $n \in \mathbb{N}$ je počet přijatých studentů. Pro $i = 1, \dots, n$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & , \ i\text{-tý přijatý student se zapíše na studium,} \\ \\ 0 & , \ i\text{-tý přijatý student se nezapíše na studium.} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělením, kde $P(X_i = 1) = \frac{2}{3}$.
Počet zapsaných studentů bude veličina

$$Z_n = \sum_{i=1}^{n} X_i \ ,$$

která má binomické rozdělení $\mathrm{Bi}\Big(n, \frac{2}{3}\Big)$. Zajímá nás teď největší n takové, že

$$P\Big(Z_n > 600\Big) \leq 0.05 \ .$$

Opět použijeme centrální limitní větu na normovanou veličinu

$$\mathrm{norm}(Z_n) = \frac{Z_n - E(Z_n)}{\sqrt{D(Z_n)}} = \frac{Z_n - \frac{2}{3}n}{\sqrt{\frac{2}{3} \cdot \frac{1}{3}n}} = \frac{3Z_n - 2n}{\sqrt{2n}} \ .$$

Pomocí úprav nerovností teď můžeme psát :

$$0.05 \geq P\Big(Z_n > 600\Big) = P\left(\frac{3Z_n - 2n}{\sqrt{2n}} > \frac{3 \cdot 600 - 2n}{\sqrt{2n}}\right) =$$

$$= P\left(\mathrm{norm}(Z_n) > \frac{1800 - 2n}{\sqrt{2n}}\right) = 1 - P\left(\mathrm{norm}(Z_n) \leq \frac{1800 - 2n}{\sqrt{2n}}\right) =$$

$$= 1 - F_{\mathrm{norm}(Z_n)}\left(\frac{1800 - 2n}{\sqrt{2n}}\right) \doteq 1 - \Phi\left(\frac{1800 - 2n}{\sqrt{2n}}\right)$$

tedy dostáváme přibližnou nerovnost (kterou budeme řešit, jako přesnou nerovnost)

$$\Phi\left(\frac{1800-2n}{\sqrt{2n}}\right) \geq 0.95$$

a

$$\left(\sqrt{2n}\right)^2 + \underbrace{\Phi^{-1}(0.95)}_{1.645} \cdot \sqrt{2n} - 1800 \leq 0 \ .$$

Dostali jsme kvadratickou nerovnost a hledáme největší $n \in \mathbb{N}$, které ji splňuje. Z kořenů kvadratické rovnice si tedy vezmeme ten, co je více vpravo. Dostaneme tak

$$\sqrt{2n} \leq \frac{-1.645 + \sqrt{1.645^2 + 4 \cdot 1800}}{2}$$

a

$$n \leq \frac{\left(\sqrt{1.645^2 + 7200} - 1.645\right)^2}{8} \doteq 865.774$$

Mělo by se tedy přijmout $n = 865$ studentů. Průměrný počet těch, co se zapíší tak bude

$$E(Z_n) = \frac{2}{3} \cdot n = \frac{2}{3} \cdot 865 \doteq 576,67 \ .$$