9. cvičení z PST

18. listopadu 2020

9.1 (kovariance)

Nechť veličina X má rovnoměrné rozdělení na intervalu $(-1,1)$ a $Y = X^2$.

(a) Rozhodněte, zda jsou X a Y nezávislé a proč.

(b) Spočtěte $\operatorname{cov}(X,Y)$.

Řešení:

(a) Jak se dá očekávat, pokud jedna veličina závisí svými hodnotami na druhé, nejspíš nezávislé nebudou. Výjimkou je jen jeden případ a celá situaci se dá popsat takto:

Věta: Nechť X a $h(X)$ jsou obě náhodné veličiny, kde $h : \mathbb{R} \to \mathbb{R}$ je borelovsky měřitelná funkce (např. spojitá). Pak X a $h(X)$ jsou nezávislé veličiny právě jen pokud

- $h(X)$ je skoro všude konstantní veličina, tj. ex. $c \in \mathbb{R}$, že $P\big(h(X) = c\big) = 1$.

V našem případě $Y = X^2$ takova není, což snadno ukážeme sporem:

Nechť $c \in \mathbb{R}$ je takové, že $P\big(X^2 = c\big) = 1$. Zřejmě musí být $c \geq 0$. Protože spojitá veličina X má nulovou pravděpodobnost nabytí dané hodnoty, dostáváme, že

$$1 = P\big(X^2 = c\big) = \underbrace{P(X = \sqrt{c})}_{0} + \underbrace{P(X = -\sqrt{c})}_{0} = 0$$

což je spor.

Takže veličiny X a Y jsou *závislé*.

(c) Kovarianci vypočteme ze vztahu

$$\operatorname{cov}(X,Y) = \operatorname{cov}(X,X^2) = E(X \cdot X^2) - E(X) \cdot E(X^2)$$

Hustota f_X je sudá funkce, takže

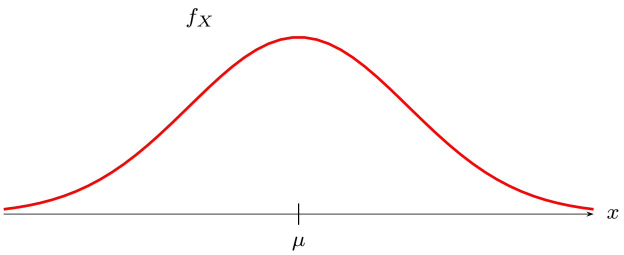
$$E(X^3) = \int_{-1}^{1} \underbrace{x^3 \cdot f_X(x)}_{\text{lichá funkce}} \, dx = 0$$

a podobně je $E(X) = 0$ a dosazením dostaneme, že $\operatorname{cov}(X,Y) = 0$. Máme tedy příklad toho, že veličiny X a Y mohou být závislé a přitom nekorelované.

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

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

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 $t \in \mathbb{R}$.
- 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.)

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

$$0.45 = P(X > 60) = 1 - P(X \leq 60) = 1 - F_X(60) = 1 - \Phi\Big(\frac{60-\mu}{\sigma}\Big)$$

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

a podobně

$$0.2 = P(X > 70) = 1 - P(X \leq 70) = 1 - F_X(70) = 1 - \Phi\Big(\frac{70-\mu}{\sigma}\Big)$$

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

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

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

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

9.3 (normální rozdělení)

Nechť veličina X má normované normální rozdělení $N(0,1)$. Určete $P(X^2 < 3X - 2)$ a najděte takové číslo ε, že $P(|X| < \varepsilon) = 0.95$.

Řešení:
Máme

$$P(X^2 < 3X - 2) = P(X^2 - 3X + 2 < 0) = P\Big((X-1)(X-2) < 0\Big) = P\Big(1 < X < 2\Big) =$$

$$= \Phi(2) - \Phi(1) \doteq 0.97725 - 0.84134 = 0.13591\ .$$

A dále je

$$0.95 = P(|X| < \varepsilon) = P(-\varepsilon < X < \varepsilon) = \Phi(\varepsilon) - \underbrace{\Phi(-\varepsilon)}_{=1-\Phi(\varepsilon)} = 2\Phi(\varepsilon) - 1$$

a tedy

$$\varepsilon = \Phi^{-1}\left(\frac{1+0.95}{2}\right) = \Phi^{-1}(0.975) \doteq 1.96\ .$$

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

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

9.4 (normální rozdělení)

Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě 67 m a 75 m a směrodatné odchylky 6 m a 3 m. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál než Barbora.

Řešení:
Náhodná veličina

$A =$"*délka hodu Anny*"

má rozdělení $N\big(67\ \mathrm{m}, (6\ \mathrm{m})^2\big)$ a veličina

$B =$"*délka hodu Barbory*"

má rozdělení $N\big(75\ \mathrm{m}, (3\ \mathrm{m})^2\big)$.

Zajímá nás $P(A > B) = P(A-B > 0)$. Protože veličiny A a B jsou nezávislé, tak veličina $Z := A-B$ má také normální rozdělení, a sice

$$Z \sim N(67 - 75, 6^2 + 3^2) = N(-8, 45)$$

(jednotky už pro přehlednost nepíšeme).

Takže

$$P(A > B) = P(Z > 0) = P\left(\underbrace{\tfrac{Z-(-8)}{\sqrt{45}}}_{norm(Z)} > \frac{0-(-8)}{\sqrt{45}}\right) = 1 - P\left(norm(Z) \leq \frac{8}{\sqrt{45}}\right) =$$

$$= 1 - \Phi\left(\tfrac{8}{\sqrt{45}}\right) \doteq 1 - \Phi\,(1.1926) \doteq 1 - 0.883 = 0.117\,.$$

POZOR! Zatímco střední hodnota je lineární zobrazení, tak rozptyl se chová jinak! Konkrétně je to takto:

Nechť X a Y jsou veličiny se střední hodnotou a konečným rozptylem. Pak

- $E(X \pm Y) = E(X) \pm E(Y)$
- $D(X \pm Y) = D(X) + D(Y) \pm 2 \cdot \mathrm{cov}(X, Y)$ (≥ 0)

Speciálně, pokud X a Y jsou nezávislé, je $\mathrm{cov}(X, Y) = 0$. Máme tedy:

- X a Y nezávislé $\Rightarrow$ $D(X \pm Y) = D(X) + D(Y)$

Tedy v tomto případě se rozptyly VŽDY sčítají!

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

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

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 (pro všechna $t \in \mathbb{R}$ a $n \in \mathbb{N}$):

$$\left|F_{\mathrm{norm}(\widetilde{X}_n)}(t) - \Phi(t)\right| \leq C_1 \cdot \frac{\varrho}{\sigma^3\sqrt{n}}$$

kde C_1 je nějaké konstanta. Nejlepší současný odhad pro C_1 zatím je, že $C_1 < 0.4748$.

Kromě toho platí ještě odhad (opět pro všechna $t \in \mathbb{R}$ a $n \in \mathbb{N}$):

$$\left|F_{\mathrm{norm}(\widetilde{X}_n)}(t) - \Phi(t)\right| \leq \frac{C_2}{(1+|t|)^3} \cdot \frac{\varrho}{\sigma^3\sqrt{n}}$$

kde C_2 je konstanta. Její současný odhad je $C_2 < 30.84$.

- -

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

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

čímž dostáváme odhady

$$\left|F_{\mathrm{norm}(\widetilde{X}_n)}(t) - \Phi(t)\right| \;\leq\; C_1 \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}} \;<\; 0.4748 \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}$$

a

$$\left|F_{\mathrm{norm}(\widetilde{X}_n)}(t) - \Phi(t)\right| \;\leq\; \frac{C_2}{(1+|t|)^3} \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}} \;<\; \frac{30.84}{(1+|t|)^3} \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}\ .$$

Druhý odhad je tedy lepší než první pro $|t| > 3.0198$.

Pěkně zpracováno je to v bakalářské práci "Rastislav Rehák: Rýchlosť konvergencie v centrálnej limitnej vete". http://is.muni.cz/th/394214/prif_b/BakalarskaPraca.pdf

- -

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

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

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

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

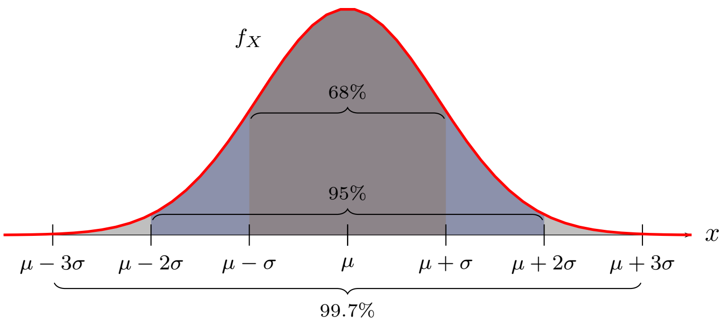
dostaneme postupně

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

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

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

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



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

9.5 (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 $i = 1, \dots, n$ (kde $n = 100$) si zavedeme veličiny

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

Veličiny X_i budou nezávislé s alternativním rozdělením $\mathrm{Alt}(p) = \mathrm{Alt}(0.2)$, 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 $\mathrm{Bi}(n,p) = \mathrm{Bi}(100, 0.2)$. 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:

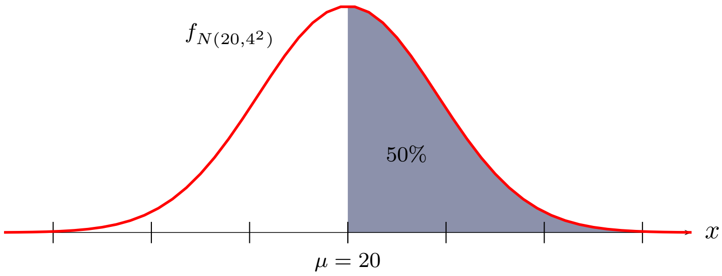
(a)

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

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

(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)

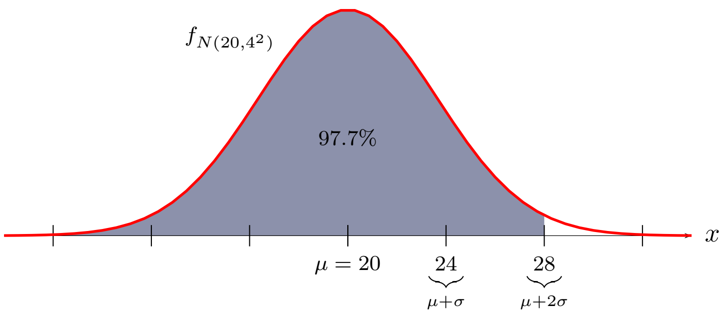
$$P(Z < 28) = P\left(\frac{Z-20}{\sqrt{16}} < \frac{28-20}{\sqrt{16}}\right) = P\Big(norm(Z) < 2\Big) \overset{(CLV)}{\doteq}$$

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

(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)

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

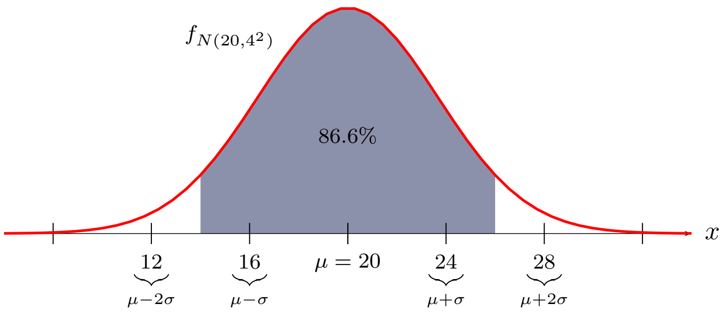
$$= P\Big(norm(Z) \le 1.5\Big) - P\Big(norm(Z) < -1.5\Big) \overset{(CLV)}{\doteq} \Phi(1.5) - \Phi(-1.5) =$$

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

(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).



9.6 (alternativní rozdělení, odhad pravděpodobnosti - CLV a Čebyševova nerovnost)
Házíme 1000 krát symetrickou mincí a počítáme počet rubů. Pro náhodnou veličinu X, která je počtem rubů,

(a) vypočtěte pravděpodobnost, že počet rubů bude mezi 455 a 545, pomocí centrální limitní věty,

(b) odhadněte pravděpodobnost, že počet rubů bude mezi 455 a 545, pomocí Čebyševovy nerovnosti,

(c) určete pravděpodobnost, že rub padne nejvýše 520 krát.

Řešení:
Veličina X počítá počet úspěchů v n pokusech a má tedy binomické rozdělení $\mathrm{Bi}(n,p)$.

(a) Kvůli velkých hodnotám, se kterými pracujeme je ale výhodnější použít CLV.
Pro $i = 1, \dots, n$ (kde $n = 1000$) si označme diskrétní veličiny s alternativním rozdělením

$$X_i = \begin{cases} 1 & , \text{ při } i\text{-tém hodu padne rub,} \\ 0 & , \text{ při } i\text{-tém hodu padne líc} \end{cases}$$

s alternativním rozdělením $\mathrm{Alt}(p)$, $p = \frac{1}{2}$ (tj. $P(X_i = 1) = p$). Veličiny X_i považujeme za nezávislé, se středními hodnotami $E(X_i) = p = \frac{1}{2}$ a rozptyly $D(X_i) = p(1-p) = \frac{1}{4}$. Máme

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

se střední hodnotou

$$E(X) = \sum_{i=1}^{n} E(X_i) = np = 1000 \cdot \frac{1}{2} = 500$$

a rozptylem

$$D(X) = \sum_{i=1}^{n} D(X_i) = np(1-p) = 1000 \cdot \frac{1}{4} = 250 \ .$$

Podle CLV má veličina $\mathrm{norm}(X) = \frac{X-500}{\sqrt{250}}$ přibližně normální rozdělení $N(0,1)$. Máme teď určit

$$P(455 < X < 545) = P\left(\frac{455-500}{\sqrt{250}} < \frac{X-500}{\sqrt{250}} < \frac{545-500}{\sqrt{250}}\right) = P\left(-\frac{9}{\sqrt{10}} < \mathrm{norm}(X) < \frac{9}{\sqrt{10}}\right) =$$

$$\doteq \Phi\Big(\frac{9}{\sqrt{10}}\Big) - \Phi\Big(-\frac{9}{\sqrt{10}}\Big) = 2\cdot\Phi\Big(\frac{9}{\sqrt{10}}\Big) - 1 \doteq$$

$$\doteq 2\cdot\Phi(2.846) - 1 \doteq 2\cdot 0.99779 - 1 = 0.99558\ .$$

Postup se dá ještě trochu zpřesnit protože veličina X má binomické rozdělení a její hodnoty jsou pouze přirozená čísla od 0 do n:

Nechť Y je veličina s normálním rozdělením aproximujícím X, tj. $Y \sim N\Big(E(X), D(X)\Big)$.

Pro $a = 0, 1, \ldots, n-1$ je F_X na intervalu $\langle a, a+1)$ konstantní (s hodnotou $F_X(a)$) a funkce F_Y pro odpovídající normální rozdělení, je zde ostře rostoucí. Z tohoto důvodu je lepší vzít jako aproximaci pro $F_X(a)$ hodnotu $F_Y(a+\frac{1}{2})$.

Tedy pro $a, b \in \{0, 1, \ldots, n-1\}$ se níže uvedená pravděpodobnost aproximuje jako:

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F_X(b) - F_X(a) \doteq$$
$$\doteq F_Y(b+\tfrac{1}{2}) - F_Y(a+\tfrac{1}{2}) = P(a+\tfrac{1}{2} < Y \le b+\tfrac{1}{2})$$

Abychom se tedy při výpočtu dostali k tomuto tvaru je dobré uvědomit si, že (díky diskrétnosti X) přesně platí

$$P(a < X \le b) = P(a+\tfrac{1}{2} < X \le b+\tfrac{1}{2})$$

a toto "posunutí" o $\frac{1}{2}$ se pak používá při přesnější aproximaci (přitom u nerovnosti na pravé straně rovnosti už je jedno, jestli jsou ostré nebo ne):

$$P(455 < X < 545) = P(455 < X \le 544) = P(455.5 < X \le 544.5) = P\Big(\frac{455.5-500}{\sqrt{250}} < \frac{X-500}{\sqrt{250}} \le \frac{544.5-500}{\sqrt{250}}\Big) =$$

$$= P\Big(-\frac{8.9}{\sqrt{10}} < \mathrm{norm}(X) \le \frac{8.9}{\sqrt{10}}\Big) \doteq$$

$$\doteq \Phi\Big(\frac{8.9}{\sqrt{10}}\Big) - \Phi\Big(-\frac{8.9}{\sqrt{10}}\Big) = 2\cdot\Phi\Big(\frac{8.9}{\sqrt{10}}\Big) - 1 \doteq$$

$$\doteq 2\cdot\Phi(2.814) - 1 \doteq 2\cdot 0.99755 - 1 = 0.9951\ .$$

Předchozí postup dál 99.558% a tento postup zase 99.51%, což je rozdíl 0.048%. Větší rozdíly bychom zaznamenali, kdybychom byli v intervalech blíže ke střední hodnotě.

(b) Odhad pomocí Čebyševovy nerovnosti:

$$P(455 < X < 545) = P\Big(\big|X - \underbrace{500}_{E(X)}\big| < 45\Big) \ge 1 - \frac{D(X)}{45^2} = 1 - \frac{250}{45^2} = 1 - \frac{10}{81} \doteq 0.8765$$

Dostáváme tedy spodní odhad 87.65%.

(c) Zde použijeme opět CLV:

$$P(X \le 520) = P\Big(\frac{X-500}{\sqrt{250}} \le \frac{520-500}{\sqrt{250}}\Big) = P\Big(\mathrm{norm}(X) \le \frac{4}{\sqrt{10}}\Big) \doteq \Phi\Big(\frac{4}{\sqrt{10}}\Big) \doteq \Phi(1.265) \doteq 0.8971\ .$$

9.7 (alternativní rozdělení, odhad pravděpodobnosti - CLV)

Tramvaj má intervaly mezi příjezdy 10 minut. Jaká je pravděpodobnost, že během 24 pracovních dnů stráví člověk při cestách do práce a zpět čekáním na tramvaj nejvýše 3 hodiny?

Řešení:
Pro veličinu

$$Z = \text{“}\textit{celková doba čekání během 24 dnů při cestách tam a zpět}\text{”}\ [\text{v hodinách}]$$

nás zajímá $P(Z \leq 3)$.

K řešení opět použijeme centrální limitní větu. Označme si tedy pro $i = 1, 2, ..., n$, kde $n = 24 \cdot 2 = 48$, veličiny

$$X_i = \text{“}\textit{doba strávená čekáním při } i\textit{-tém příchodu na zastávku}\text{”}\ [\text{v hodinách}]$$

které pokládáme za nezávislé. Tramvaj jezdí přesně po 10 minutách, zatímco naše příchody na zastávku budeme pokládat za náhodné s rovnoměrným rozdělením v rámci 10 minutového intervalu. Proto i doba čekání X_i bude mít rovnoměrné rozdělení (v jednotkách hodin) tvaru $\mathrm{Ro}\,(a, b) = \mathrm{Ro}\left(0, \frac{1}{6}\right)$.

Protože opět platí $Z = \sum\limits_{i=1}^{n} X_i$, dostaneme

$$E(X_i) = \frac{a+b}{2} = \frac{0+\frac{1}{6}}{2} = \frac{1}{12} \quad \Rightarrow \quad E(Z) = n \cdot E(X_1) = 48 \cdot \frac{1}{12} = 4$$

$$D(X_i) = \frac{(b-a)^2}{12} = \frac{(\frac{1}{6}-0)^2}{12} = \frac{1}{12 \cdot 36} \quad \Rightarrow \quad D(Z) = n \cdot D(X_1) = 48 \cdot \frac{1}{12 \cdot 36} = \frac{1}{9}$$

$$\Rightarrow \quad \sqrt{D(Z)} = \sqrt{\frac{1}{9}} = \frac{1}{3}\,.$$

Podle CLV bude mít veličina $norm(Z) = \frac{Z-E(Z)}{\sqrt{D(Z)}} = 3(Z-4)$ přibližně rozdělení $N(0,1)$. Můžeme proto psát

$$P(Z \leq 3) = P\Big(\underbrace{3 \cdot (Z-4)}_{norm(Z)} \leq 3 \cdot (3-4)\Big) = P\Big(norm(Z) \leq -3\Big) \overset{(CLV)}{\doteq}$$

$$\overset{(CLV)}{\doteq} \Phi(-3) = 1 - \Phi(3) \doteq 1 - 0.9987 = \mathbf{0.0013}\,.$$

9.8 (normální rozdělení, odhad počtu - CLV a Čebyševova nerovnost)

Napětí v síti je náhodná veličina, která má normální rozdělení $N(\mu, \sigma^2)$, kde $\mu = 235$ V a $\sigma = 3$ V. Kolik měření musíme minimálně provést, aby se průměr naměřených hodnot lišil od 235 V nejvýše o 1 V s pravděpodobností alespoň 0.95? Výpočet proveďte:

(a) pomocí odhadu pravděpodobnosti z Čebyševovy nerovnosti;

(b) z výpočtu pravděpodobnosti z centrální limitní věty.

Řešení:

Pokud budeme předpokládat, že veličiny

$$X_i = \text{“}\textit{hodnota napětí při } i\textit{-tém měření}\text{”}\ [V]$$

(v jednotkách Volt) mají *přesné* normální rozdělení, pak i veličina výběrového průměru

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

bude také mít *přesné* normální rozdělení. Takže CLV, která mluví o *přibližnosti* pomocí normálního rozdělení, ani nebudeme potřebovat. Máme tedy $\overline{X}_n \sim N\Big(E(\overline{X}_n),\ D(\overline{X}_n)\Big)$, kde

$$E(\overline{X}_n) = E(X_i) = \mu = 235$$
$$D(\overline{X}_n) = \frac{D(X_i)}{n} = \frac{\sigma^2}{n} = \frac{9}{n} \quad \Rightarrow \quad \sqrt{D(\overline{X}_n)} = \frac{3}{\sqrt{n}}$$

(a) Hledáme nejmenší $n \in \mathbb{N}$ tak, aby $P(|\overline{X}_n - 235| \le 1) \ge 0.95$

$$0.95 \le P(|\overline{X}_n - 235| \le 1) = P\left(\left|\frac{\overline{X}_n - 235}{\frac{3}{\sqrt{n}}}\right| \le \frac{1}{\frac{3}{\sqrt{n}}}\right) = P\left(\Big|\text{norm}(\overline{X}_n)\Big| \le \frac{\sqrt{n}}{3}\right) =$$
$$= \Phi\Big(\frac{\sqrt{n}}{3}\Big) - \Phi\Big(-\frac{\sqrt{n}}{3}\Big) = 2\cdot\Phi\Big(\frac{\sqrt{n}}{3}\Big) - 1$$

Máme tudíž

$$\Phi\Big(\frac{\sqrt{n}}{3}\Big) = \frac{1+0.95}{2} = 0.975$$
$$n \ge \Big(3\cdot\Phi^{-1}(0.975)\Big)^2 \doteq (3\cdot 1.96)^2 \doteq 34.57$$

Je tedy potřeba udělat alespoň $n = 35$ měření.

(b) Pomocí Čebyševovy nerovnosti odhadujeme nejmenší $n \in \mathbb{N}$ tak, aby $P(|\overline{X}_n - 235| \le 1) \ge 0.95$. Víme, že

$$P(|\overline{X}_n - \underbrace{235}_{E(\overline{X}_n)}| \le 1) \ge 1 - \frac{D(\overline{X}_n)}{1^2} = 1 - \frac{3}{\sqrt{n}}$$

Pokud nyní bude platit, že $1 - \frac{3}{\sqrt{n}} \ge 0.95$, bude určitě odhad pravděpodobnosti splněny (a nic silnějšího nám Čebyševova nerovnosti neumožňuje zjistit). Máme tedy

$$1 - \frac{3}{\sqrt{n}} \ge 0.95 \quad \Leftrightarrow \quad \frac{3}{\sqrt{n}} \le 0.05 \quad \Leftrightarrow \quad n \ge \Big(\frac{3}{0.05}\Big)^2 = 3600$$

Zde tedy máme odhad, že potřebujeme udělat alespoň 3600 měření, což je oproti předchozímu (přesnému) počtu zbytečně mnoho. Čebyševova nerovnost, ale lepší výsledek neumí a to proto, že je splněna pro libovolné rozdělení, které může být velmi vzdálené od normálního.

9.9 (alternativní rozdělení, odhad počtu - CLV a Čebyševova nerovnost)

Házíme pravidelnou hrací kostkou a počítáme výskyt šestek. Určete, kolik musíme provést hodů, aby se relativní výskyt šestek (tj. poměr počtu šestek ku počtu hodů) lišil od $\frac{1}{6}$ nejvýše o $\varepsilon = 0.05$ s pravděpodobností alespoň $p = 0.95$. Použijte

(a) centrální limitní větu.

(b) Čebyševovu nerovnost.

Řešení:
Pro $i \in \mathbb{N}$ si zavedeme veličiny

$$X_i = \text{\textit{„počet šestek, které padnou v i-tém hodu"}} = \begin{cases} 1 & , \text{ \textit{při i-tém hodu padla šestka,}} \\ 0 & , \text{ \textit{při i-tém hodu nepadla šestka.}} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělení s parametrem $p = \frac{1}{6}$ (protože $P(X_i = 1) = p$), střední hodnotou $E(X_i) = p = \frac{1}{6}$ a rozptylem $D(X_i) = p(1-p) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$.

Relativní výskyt šestek při n hodech se pak vyjádří jako veličina

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

se střední hodnotou: $E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = p = \frac{1}{6}$

a rozptylem: $D(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} D(X_i) = \frac{p(1-p)}{n} = \frac{5}{36n}$.

Zajímá nás teď nejmenší n tak, aby

$$P\left(\left| \overline{X}_n - \frac{1}{6} \right| \leq \varepsilon \right) \geq p = 0.95 \ ,$$

kde $\varepsilon = 0.05$.

(a) Podle centrální limitní věty má veličina $\overline{X}_n$ přibližně rozdělení $N(\frac{1}{6}, \frac{5}{36n})$, konkrétně veličina $\overline{X}_n - \frac{1}{6}$ má přibližně rozdělení $N(0, \frac{5}{36n})$. Takže

$$0.95 \leq P\left(\left| \overline{X}_n - \frac{1}{6} \right| \leq 0.05 \right) = P\left(-0.05 \leq \overline{X}_n - \frac{1}{6} \leq 0.05 \right) \doteq$$

$$\doteq \Phi\left(\frac{0.05}{\sqrt{\frac{5}{36n}}} \right) - \Phi\left(-\frac{0.05}{\sqrt{\frac{5}{36n}}} \right) = 2\Phi\Big(\frac{0.3}{\sqrt{5}} \sqrt{n} \Big) - 1$$

tedy

$$\Phi\Big(\frac{0.3}{\sqrt{5}} \sqrt{n} \Big) \geq \frac{1 + 0.95}{2} = 0.975$$

$$n \geq \left(\frac{\sqrt{5}}{0.3} \cdot \Phi^{-1}(0.975) \right)^2 \doteq \left(\frac{\sqrt{5}}{0.3} \cdot 1.96 \right)^2 \doteq 213.42$$

Musíme tedy provést alespoň $n = 214$ hodů.

(b) Použijeme už upravených výrazů. Z Čebyševovy nerovnosti máme:

$$P\left(\left| \overline{X}_n - \frac{1}{6} \right| \leq 0.05 \right) \geq 1 - \frac{D(\overline{X}_n)}{(0.05)^2} = 1 - \frac{\frac{5}{36n}}{(0.05)^2} = 1 - \frac{500}{9n} \ .$$

Pokud bude

$$1 - \frac{500}{9n} \geq 0.95$$

neboli

$$n \geq \frac{500}{9(1 - 0.95)} \doteq 1111.11 \ ,$$

pak máme určitě podmínku $P\left(\left| \overline{X}_n - \frac{1}{6} \right| \leq 0.05 \right) \geq 0.95$ splněnu. Čebyševova nerovnost nám tak dává odhad, že potřebujeme udělat alespoň $n = 1112$ hodů. To je podstatně více než u centrální limitní věty, ale zato to víme přesně.

9.10 (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:

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

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

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

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

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

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).

9.11 (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 $\mathrm{norm}(T) = \frac{T - E(T)}{\sqrt{D(T)}} = \frac{T-3}{\sqrt{0.075}}$ má přibližně normované normální rozdělení $N(0,1)$:

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

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

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

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

(b) Odhad pomocí Čebyševovy nerovnosti:

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

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

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

9.12 (alternativní rozdělení, odhad intervalu - CLV a Čebyševova nerovnost)

Házíme 100-krát pravidelnou mincí a náhodná veličina X je počet rubů. Určete (přibližně) číslo ε tak, aby $P\Big(|X-E(X)| < \varepsilon\Big) = 0.9$:

(a) z výpočtu pravděpodobnosti z centrální limitní věty;

(b) pomocí odhadu pravděpodobnosti z Čebyševovy nerovnosti.

Řešení:
Náhodná veličina X má binomické rozdělení Bi(100, 0.5). Je tedy $E(X) = 100 \cdot 0.5 = 50$ a $D(X) = 100 \cdot 0.5 \cdot (1-0.5) = 25$.

(a) Z centrální limitní věty vyplývá, že můžeme předpokládat pro náhodnou veličinu X přibližně normální rozdělení N(50, 25). Tedy $F_X(t) \doteq \Phi(\frac{t-50}{5})$ pro $t \in \mathbb{R}$.

Číslo ε určíme z rovnice

$$\begin{aligned} 0.9 &= P\Big(|X-E(X)| < \varepsilon\Big) = P\Big(50-\varepsilon < X < 50+\varepsilon\Big) = F_X(50+\varepsilon) - F_X(50-\varepsilon) \doteq \\ &\doteq \Phi\left(\frac{\varepsilon}{5}\right) - \Phi\left(\frac{-\varepsilon}{5}\right) = 2\,\Phi\left(\frac{\varepsilon}{5}\right) - 1 \\ &\Rightarrow \Phi\left(\frac{\varepsilon}{5}\right) = 0.95 \Rightarrow \frac{\varepsilon}{5} = \Phi^{-1}(0.95) \Rightarrow \varepsilon \doteq 5 \cdot 1.645 = 8.225\,. \end{aligned}$$

(b) Odhad ε pomocí Čebyševovy nerovnosti:

$$0.9 = P\Big(|X-E(X)| < \varepsilon\Big) = P\Big(|X-50| < \varepsilon\Big) \geq 1 - \frac{D(X)}{\varepsilon^2} = 1 - \frac{25}{\varepsilon^2}$$

$$\varepsilon \leq \sqrt{\frac{25}{0.1}} \doteq 15.81$$