9. cvičení z PST

20. listopadu 2019

9.1 (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

$$0.95 = P(|X| < \varepsilon) = P(-\varepsilon < X < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi(\varepsilon) - 1$$

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

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

9.2 (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 $N\big(67 \text{ m}, (6 \text{ m})^2\big)$ a veličina

$B =$“*délka hodu Barbory*”

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

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

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

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

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

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

Zde $\operatorname{cov}(X, Y)$ je tzv. kovariance (viz poznámky níže). Speciálně, pokud X a Y jsou nezávislé, je $\operatorname{cov}(X, Y) = 0$. 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 $X_i : \Omega \to \mathbb{R}$ pro $i \in \mathbb{N}$ je posloupnost nezávislých náhodných veličin, které mají stejné rozdělení se střední hodnotou μ a (konečným) rozptylem σ^2. Položme

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

a

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

Pak

$$\operatorname{norm}(\widetilde{X}_n) = \frac{\widetilde{X}_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} = \operatorname{norm}(\overline{X}_n)$$

a

$$\lim_{n\to\infty} F_{\operatorname{norm}(\widetilde{X}_n)}(t) = \lim_{n\to\infty} P\Big(\operatorname{norm}(\widetilde{X}_n) \leq t\Big) = \Phi(t)$$

pro každé $t \in \mathbb{R}$ (kde Φ je distribuční funkce normovaného normálního rozdělení, tj. pro $N(0, 1)$).

- -

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 $n \in \mathbb{N}$):

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

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 $n \in \mathbb{N}$):

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

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

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

a

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

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 $N(\mu, \frac{\sigma^2}{n})$.

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.

9.3 (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 $i = 1, \ldots, n$ (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 $a = 0, 1, \dots, n-1$ 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 $a, b \in \{0, 1, \dots, n-1\}$ 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:

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

9.4 (normální rozdělení, odhad pravděpodobnosti - CLV a Čebyševova nerovnost)

Výška mužů (určitého věku) je náhodná veličina (s normálním rozdělením) o střední hodnotě 180 cm a směrodatnou odchylkou 12 cm. Určete pravděpodobnost, že průměrná výška $n = 150$ mužů bude

(a) v intervalu 177.5 cm a 182.5 cm pomocí centrální limitní věty,

(b) v intervalu 177.5 cm a 182.5 cm odhadem pomocí Čebyševovy nerovnosti,

(c) alespoň 178 cm.

Řešení:
Výšku i-tého muže (pro $i = 1, \ldots, n$) si označíme jako X_i. Veličiny X_i se spojitým rozdělením považujeme za nezávislé, se střední hodnotou $E(X_i) = 180$ cm a směrodatnou odchylkou $\sigma_i = \sqrt{D(X_i)} = 10$ cm.

Průměrná výška se pak vyjádří jako

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

Zajímá nás $P(177.5 \leq \overline{X}_n \leq 182.5)$. Budeme potřebovat:

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = 180$$

$$D(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = \frac{12^2}{n} = \frac{144}{150} = 0.96$$

(a) Odhad pomocí centrální limitní věty - veličina $\overline{X}_n$ má přibližně rozdělení $N(180, 0.96)$:

$$\begin{aligned} P(177.5 \leq \overline{X}_n \leq 182.5) &= F_{\overline{X}_n}(182.5) - F_{\overline{X}_n}(177.5) \doteq \Phi\Big(\frac{182.5-180}{\sqrt{0.96}}\Big) - \Phi\Big(\frac{177.5-180}{\sqrt{0.96}}\Big) = \\ &= \Phi\Big(\frac{2.5}{\sqrt{0.96}}\Big) - \Phi\Big(-\frac{2.5}{\sqrt{0.96}}\Big) = 2\cdot\Phi\Big(\frac{2.5}{\sqrt{0.96}}\Big) - 1 \doteq \\ &\doteq 2\cdot\Phi\Big(2.5516\Big) - 1 \doteq 2\cdot 0.99464 - 1 = 0.98928 \ . \end{aligned}$$

Tedy asi 98.93%.

(b) Odhad pomocí Čebyševovy nerovnosti:

$$P(177.5 \leq \overline{X}_n \leq 182.5) = P\left(\left|\overline{X}_n - 180\right| \leq 2.5\right) \geq 1 - \frac{D(\overline{X}_n)}{(2.5)^2} = 1 - \frac{0.96}{(2.5)^2} = \frac{5.29}{6.25} = 0.8464$$

Dostáváme tedy spodní odhad 84.64%.

(c) Odhad pomocí centrální limitní věty - veličina $\overline{X}_n$ má přibližně rozdělení $N(180, 0.96)$:

$$\begin{aligned} P(178 \leq \overline{X}_n) &= 1 - P(\overline{X}_n < 178) = 1 - F_{\overline{X}_n}(178) \doteq 1 - \Phi\Big(\frac{178-180}{\sqrt{0.96}}\Big) = \\ &= 1 - \Phi\Big(-\frac{2}{\sqrt{0.96}}\Big) = \Phi\Big(\frac{2}{\sqrt{0.96}}\Big) \doteq \Phi\Big(2.0412\Big) \doteq 0.979 \ . \end{aligned}$$

Tedy asi 97.9%.

9.5 (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

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

(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 $P(|\overline{X}_n - 235| \leq 1) \geq 0.95$

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

Máme tudíž

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

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

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

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

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

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

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.

9.6 (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

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

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

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

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

Musíme tedy provést alespoň $n = 214$ hodů.

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

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

Pokud bude

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

neboli

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

pak máme určitě podmínku $P\left(\left|\overline{X}_n - \frac{1}{6}\right| \le 0.05\right) \ge 0.95$ 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.

9.7 (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 $N(0,1)$:

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

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

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:

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

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

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

9.8 (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:

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

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

9.9 (Poissonovo rozdělení - Čebyševova nerovnost)

Pro náhodnou veličinu X s Poissonovým rozdělením Poiss(λ), kde $\lambda = 5$, odhadněte pomocí Čebyševovy nerovnosti pravděpodobnost $P\big(X < 12\big)$.

Řešení:
Pro Čebyševovu nerovnost potřebujeme tvar $P\big(|X - E(X)| < \varepsilon\big) \geq 1 - \frac{D(X)}{\varepsilon^2}$. Musíme si tedy tento tvar vyrobit. Máme $E(X) = D(X) = \lambda = 5$. Protože hodnoty X jsou $0, 1, 2, \ldots$, tak můžeme psát

$$0 \leq X < 12 \quad \Leftrightarrow \quad -5 \leq X - 5 < 7 \quad \Leftrightarrow \quad |X - 5| < 7$$

Tudíž dostáváme

$$P(X < 12) = P\big(|X - 5| < 7\big) \geq 1 - \frac{D(X)}{7^2} = 1 - \frac{5}{7^2} = \frac{44}{49} \doteq 0.898$$

Skutečná pravděpodobnost je pak

$$P(X < 12) = e^{-5} \sum_{i=0}^{11} \frac{5^i}{i!} = e^{-5} \frac{58519565}{399168} \doteq 0.9878 \ .$$