8. cvičení z PST

22. listopadu 2017

8.1 (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 $P(X > 60) = 0.45$ a $P(X > 70) = 0.2$. Určete

(a) parametry μ a σ^2;

(b) pravděpodobnosti $P(X < 50)$ a $P(X > 90)$.

> **Řešení:**
> Příklad 17: `https://math.feld.cvut.cz/prucha/psi/wcv6.pdf`

8.2 (normální rozdělení)

Náhodná veličina X má normální rozdělení $N(\mu, \sigma^2)$ a $P(X < 85) = 0.9$, $P(X < 95) = 0.95$. Určete parametry rozdělení a pravděpodobnost $P(X > 60)$.

> **Řešení:**
> Příklad 8: `https://math.feld.cvut.cz/prucha/psi/wcv6.pdf`

8.3 (normální rozdělení)

Náhodná veličina U má normované normální rozdělení $N(0, 1)$. Pro hodnoty koeficientu spolehlivosti $1 - \alpha = 0.95$ stanovte jednostranné a oboustranné intervaly spolehlivosti.

> **Řešení:**
> Příklad 10: `https://math.feld.cvut.cz/prucha/psi/wcv6.pdf`

8.4 (normální rozdělení)

Chyby při měření mají normální rozdělení $N(0, 0.01)$. Kolik měření musíme provést, aby byla chyba průměru měření menší než $\varepsilon = 0.04$ s pravděpodobností $P = 0.95$.

> **Řešení:**
> Příklad 12: `https://math.feld.cvut.cz/prucha/psi/wcv6.pdf`

8.5 (normální rozdělení - 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 0.95? Výpočet proveďte:

(a) pomocí odhadu pravděpodobnosti z Čebyševovy nerovnosti;

(b) z výpočtu pravděpodobnosti z centrální limitní věty.

Řešení:
Příklad 4: `https://math.feld.cvut.cz/prucha/psi/wcv7.pdf`

8.6 (Poissonovo rozdělení - Čebyševova nerovnost)
Náhodná veličina X, která má Poissonovo rozdělení s pravděpodobnostní funkcí $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k \in \{0, 1, 2, \dots\}$. Pro hodnotu parametru $\lambda = 5$ odhadněte pomocí Čebyševovy nerovnosti pravděpodobnost $P\big(X < 18\big)$.

Řešení:
Příklad 6: `https://math.feld.cvut.cz/prucha/psi/wcv7.pdf`

8.7 (spojité rovnoměrné rozdělení - Čebyševova nerovnost)
Náhodná veličina X má rovnoměrné rozdělení v intervalu $(\mu - h, \mu + h)$. Určete 90% interval spolehlivosti se středem ve střední hodnotě $E(X)$:

(a) pomocí odhadu pravděpodobnosti z Čebyševovy nerovnosti:

(b) z výpočtu pravděpodobnosti.

Řešení:
Příklad 7: `https://math.feld.cvut.cz/prucha/psi/wcv7.pdf`

8.8 (geometrické rozdělení - Čebyševova nerovnost)
Náhodná veličina X má diskrétní rozdělení s pravděpodobnostní funkcí $p_X(k) = p(1-p)^{k-1}$, $k \in \mathbb{N}$. Pro hodnotu $p = 0.01$

(a) odhadněte pomocí Čebyševovy nerovnosti,

(b) vypočtěte pravděpodobnost $P\big(|X - E(X)| < 200\big)$.

Řešení:
Příklad 8: `https://math.feld.cvut.cz/prucha/psi/wcv7.pdf`

8.9 (alternativní rozdělení - CLV a Čebyševova nerovnost)
Házíme 1000 krát mincí a počítáme počet rubů. Pro náhodnou veličinu X, která je počtem rubů určete:

(a) odhad P o pravděpodobnosti $P\big(|X - E(X)| < 50\big)$ pomocí Čebyševovy nerovnosti;

(b) vypočtěte pravděpodobnost $P\big(|X - E(X)| < 50\big)$ pomocí centrální limitní věty.

Řešení:
Příklad 2: `https://math.feld.cvut.cz/prucha/psi/wcv7.pdf`

8.10 (alternativní rozdělení - 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:
Připomeneme si, co říká centrální limitní věta. Pro náhodnou veličinu X s konečným rozptylem, položme

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

Speciálně tedy vidíme, že $\|\operatorname{norm}(X)\| = 1$.

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

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

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 $\rho := E(|X_i - \mu|^3) < \infty$, pak platí Berry–Esseenův odhad (pro všechna $t \in \mathbb{R}$ a $n \in \mathbb{N}$):

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

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

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

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

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

čímž dostáváme odhady

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

a

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

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

Teď tedy vyřešíme zadaný příklad. Pro $i \in \mathbb{N}$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & \text{, při } i\text{-tém hodu padla šestka,} \\ \\ 0 & \text{, při } i\text{-tém hodu nepadla šestka.} \end{cases}$$

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

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

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

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

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

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:

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

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.

Detaily toho, jak velká byla nepřesnost při použití CLV, nechávám čtenáři ...

8.11 (obecné rozdělení - CLV a Čebyševova nerovnost)

Výška mužů (určitého věku) je náhodná veličina o střední hodnotě 180 cm a směrodatnou odchylkou 12 cm. Určete pravděpodobnost, že průměrná výška $n = 150$ mužů bude v intervalu 177.5 cm a 182.5 cm

(a) pomocí centrální limitní věty,

(b) pomocí Čebyševovy nerovnosti.

Řešení:

Výšku i-tého muže (pro $i = 1, \ldots, n$) si označíme jako X_i. Veličiny X_i 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:

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

a

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

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

$$P(177.5 \leq \overline{X}_n \leq 182.5) \doteq \Phi\Big(\frac{177.5-180}{\sqrt{0.96}}\Big) - \Phi\Big(\frac{182.5-180}{\sqrt{0.96}}\Big) = \Phi\Big(\frac{2.5}{\sqrt{0.96}}\Big) - \Phi\Big(-\frac{2.5}{\sqrt{0.96}}\Big) =$$

$$= 2\cdot\Phi\Big(\frac{2.5}{\sqrt{0.96}}\Big) - 1 \doteq 2\cdot\Phi\Big(2.5516\Big) - 1 \doteq 2\cdot 0.99464 - 1 = 0.98928\ .$$

Tedy asi 98.93%.

(b) Odhad pomocí Čebyševovy nerovnosti - použijeme tvar, který už máme:

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

Dostáváme tedy spodní odhad 84.64%.