

11. cvičení z PSt — 5.5.2023

x	-4	-3	-2	-1	0	1	2	3	4
$\Phi(x)$	0.00003	0.00135	0.02275	0.15866	0.500000	0.84135	0.97725	0.99865	0.99997

- *Markovova nerovnost:* $P(X \geq a) \leq \mathbb{E}(X)/a$ pro $X \geq 0, a > 0$.
- *Čebyševova nerovnost:* $P(|X - \mu| \geq a\sigma) \leq 1/a^2$ pro $\mu = \mathbb{E}(X), \sigma^2 = \text{var}(X)$.
- *Chernoffova nerovnost:* $P(X \geq t) = P(X \leq -t) \leq e^{-t^2/2\sigma^2}$ pro $X = X_1, \dots, X_n$, kde všechna $X_i = \pm 1$ jsou n.n.v. a $\sigma^2 = \text{var}(X)$.
- *Centrální limitní věta:* Pro X_1, X_2, \dots stejně rozdělené n.v. s konečným μ a σ , $Y_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$ platí $Y_n \xrightarrow{d} \Phi$.

Aplikace nerovností a Centrální Limitní Věty

1. Statistik chce odhadnout průměrnou výšku h (v metrech) lidí v nějaké populaci, pomocí n nezávislých vzorků X_1, \dots, X_n , které vybíráme uniformně náhodně ze všech možných lidí. Pro odhad použije výběrový průměr $\bar{X}_n = (X_1 + \dots + X_n)/n$. Odhaduje, že směrodatná odchylka jednoho měření je nejvýše 1 metr.

- (a) Jak velké n má volit, aby směrodatná odchylka \bar{X}_n byla nejvýše 1 cm?
- (b) Pro jaké n zajistí Čebyševova nerovnost, že \bar{X}_n se liší od h nejvýše o 5 cm s pravděpodobností alespoň 99 %?
- (c) Statistik si všimne, že všichni měření lidé mají výšku v intervalu (1.4, 2.1). Jak má upravit odhad směrodatné odchylky? Jak se změní odpovědi na předchozí otázky?

2. Označme $S = \sum_{k=0}^{30} \binom{100}{k}$. Označme dále $X = \sum_{i=1}^{100} X_i$, kde X_i je ± 1 s pravděpodobností 1/2 a veličiny X_1, \dots, X_n jsou nezávislé.

- (a) Spočítejte $P(X = k)$ pro každé k .
- (b) Vyjádřete S pomocí distribuční funkce F_X .
- (c) Použijte CLV na odhad této pravděpodobnosti.
- (d) Případně vyčíslete S vhodným softwarem a srovnejte.

3. Vygenerujeme náhodný bod v jednotkovém čtverci (obě souřadnice budou mít rozdělení $U(0, 1)$). Označme X_i indikátor jevu „ i -tý bod leží ve vepsaném kruhu“. (a) Určete $\mathbb{E}(X_i), \text{var}(X_i)$.

- (b) Položte $S_n = (X_1 + \dots + X_n)/n$. Určete $\mathbb{E}(S_n)$ a $\text{var}(S_n)$.
- (c) Všimněte si, že lze počítat S_n z S_{n-1}, X_n a n (nižší nároky na paměť).
- (d) Pro jaké n čekáte, že dostaneme výsledek správně na jedno desetinné místo? Na dvě, tři, ...?

9. domácí úkol (termín odevzdání je 2. 6. 2023)

1. Odhadněte $\binom{100}{30}$ pomocí CLV. Nápověda: použijte CLV pro odhad $P(29.5 < X < 30.5)$ pro vhodnou n.v. X . Na druhou stranu pro $P(X = 30)$ máme vzorec $\binom{100}{30}/2^{100}$ z binomického rozdělení. Alternativně, můžete použít Moivre–Laplaceho větu. [10 bodů]