11. cvičení z PSt — 18.5.2021

Ve většině příkladů je potřeba vyčíslit distribuční funkce (nebo jejich inverzní funkce) v konkrétních bodech. Pokud k tomu nemáte technické prostředky, můžete výsledek nechat nevyčíslený. Zábavnější ale bude skutečně dosadit. Nejlépe pomocí R (viz níže), ale můžete použít i např. `https://www.wolframalpha.com` nebo vhodné tabulky online, např. `https://en.wikipedia.org/wiki/Student%27s_t-distribution#Table_of_selected_values`.

Vhodné funkce R

- $pnorm(q, mean, sd)$ distribuční funkce normálního rozdělení v bodě q. Pokud se parametry neuvedou, je $mean = 0$, $sd = 1$, tj. standardní normální rozdělení. Pozor, uvádí se směrodatná odchylka, ne rozptyl!
- $qnorm(p, mean, sd)$ inverzní distribuční funkce (kvantilová funkce, quantile function) normálního rozdělení v bodě p: pro jaké q je distribuční funkce rovna p?
- $rnorm(n, mean, sd)$ vygeneruje n čísel z $N(mean, sd^2)$
- $dnorm(x, mean, sd)$ hustota (density) v bodě x
- pro kontrolu $qnorm(1) - qnorm(-1) \doteq 0.68$, $qnorm(0.5) = 0$, $pnorm(0) = 0.5$
- pro ostatní distribuce jsou také k dispozici funkce s prvním písmenkem p, q, r, d
- zejména se nám budou hodit $dbinom$, $dgeom$, $dpois$, dt, $dunif$, $dexp$

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 možnych lidí. Pro odhad použije výběrový průměr $\bar{X}_n = (X_1 + \cdots + X_n)/n$. 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 pravděpodobnost, 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. Je tedy $X \sim Bin(100, 1/2)$.

(a) Vyjádřete S pomocí distribuční funkce F_X.

(b) Použijte CLV na odhad této pravděpodobnosti.

(c) Případně vyčíslete S vhodným softwarem a srovnejte.

3. Odhadněte $\binom{100}{30}$ pomocí CLV. Nápověda: použijte CLV pro odhad $P(29.5 < X < 30.5)$. Na druhou stranu pro $P(X = 30)$ máme vzorec $\binom{100}{30}/2^{100}$ z binomického rozdělení.

4. Chceme odhadnout, zda naše mince (a způsob jak s ní házíme) je spravedlivá. Pokud ze sta hodů padne orel více než 55-krát, řekneme, že spravedlivá není. Jaká je pravděpodobnost, že se zmýlíme?

Intervalové odhady

5. Máme jedno měření $X \sim N(\mu, 1)$. (Tj. parametr $\vartheta = \mu$.)

(a) Najděte intervalový odhad pro μ se spolehlivostí 95 %.

(b) Místo jednoho měření jich provedeme n (pochopitelně nezávislých). Jaký bude teď intervalový odhad pro μ?

(c) Nechť X má stále střední hodnotu μ a rozptyl 1, ale není už nutně normální. Co se změní?

6. Tentokrát vybíráme z rozdělení $N(\mu, \sigma^2)$: μ ani σ neznáme, parametr $\vartheta = (\mu, \vartheta)$. Naměřili jsme hodnoty 8.47, 10.91, 10.87, 9.46, 10.40.

(a) Spočtěte vyberový průměr a výběrový rozptyl.

(b) Kdybychom věřili, že spočtený výběrový rozptyl je skutečná hodnota σ^2, najděte intervalový odhad pro μ.

(c) Najděte intervalový odhad pro μ použitím Studentova t-rozdělení.

7. Počet emailů za den modelujeme pomocí Poissonova rozdělení $Pois(\lambda)$. První týden v prosinci jsme dostali 34,35,29,31,30 emailů. Najděte pro λ intervalový odhad se spolehlivostí 95 %.

Použijte k tomu poslední metodu z přednášky – tu využívající Studentova rozdělení. (Poissonovo rozdělení sice není normální, ale pro dostatečně vysokou hodnotu λ je normálnímu dost podobné, metoda bude mít spolehlivost blízkou 95 %.)