9. cvičení z PSt — 7. a 10.12.2020

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 (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 vám budou hodit $dbinom$, $dgeom$, $dpois$, dt, $dunif$, $dexp$

Testování hypotéz a intervalové odhady

1. Máme jedno měření $X \sim N(\mu, 1)$. Chceme ověřit hypotézu H_0: $\mu = 5$ s hladinou významnosti $\alpha = 5\%$.

(a) Jaký zvolíme kritický obor – množinu měření, ve které hypotézu zamítneme?

(b) Místo jednoho měření jich provedeme n (pochopitelně nezávislých). Jaký bude kritický obor pro $\bar{X}_n$?

(c) Pokud je ve skutečnosti $\mu = 4$ a máme $n = 10$ měření, jaká je pravděpodobnost, že hypotézu nezamítneme?

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

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

(a) Spočtěte vyběrový průměr a výběrový rozptyl.

(b) Kdybychom věřili, že spočtený výběrový rozptyl je skutečná hodnota σ^2, posuďte na hladině významnosti 5% hypotézu $\mu = 9$.

(c) Zopakujte předchozí část použitím Studentova t-rozdělení.

3. Podle slibu výrobce bude jeho stroj dělat chyby nejvýše ve 3% případů. Z 600 pokusů došlo k chybě v 28 případech. Posuďte slib výrobce (coby nulovou hypotézu) na hladině významnosti 5%.

(a) Počet chyb modelujte přesně, tj. pomocí binomického rozdělení.

(b) Počet chyb modelujte přibližně pomocí normálního rozdělení (s vhodným μ, σ^2).

4. Počet emailů za den modelujeme pomocí Poissonova rozdělení $Pois(\lambda)$. Dlouhodobým měřením jsme určili $\lambda = 35$. První týden v prosinci jsme dostali 34,35,29,31,30 emailů. Chceme otestovat nulovou hypotézu (i v prosinci je $\lambda = 35$) na hladině významnosti 5%.

Připomeňme, že střední hodnota i rozptyl Poissonova rozdělení je rovna λ. A také to, že pokud $X_1 \sim Pois(\lambda_1)$, $X_2 \sim Pois(\lambda_2)$, tak $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$.

(a) Použijte CLV, tj. průměrný počet emailů modelujte přibližně pomocí normálního rozdělení (s vhodným μ, σ^2).

(b) Použijte to, že součet Poissonů je Poisson, tj. průměrný počet emailů modelujte přesně pomocí vhodného Poissonova rozdělení.

5. Bezpečnostní kamera pozoruje hlídanou oblast a naměří signál $X = W$, pokud není přítomen lupič (hypotéza H_0) nebo $X = W + \vartheta$, pokud lupič přítomen je (hypotéza H_1). Víme, že $\vartheta > 0$, ale přesnou hodnotu neznáme (závisí na velikosti lupiče).

Předpokládáme, že $W \sim N(0, 0.5)$.

(a) Naměříme jednu hodnotu $X = 0.96$. Máme nulovou hypotézu zamítnout na hladině 5%?

(b) Naměříme pět nezávislých hodnot $0.96, -0.34, 0.85, 0.51, -0.24$. Máme nulovou hypotézu zamítnout na hladině 5%?

(c) Opakujte část (b) pomocí t-distribuce (bez předpokladu znalosti rozptylu).