12. cvičení z PaSti – 2021-05-26

Z každé kapitolky zkuste aspoň jeden příklad.

Intervalové odhady

1. Máme jedno měření $X \sim N(\mu, 1)$. (Parametrem je tedy $\vartheta = \mu$.)

(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?

2. Tentokrát vybíráme z rozdělení $N(\mu, \sigma^2)$: μ ani σ neznáme, parametr $\vartheta = (\mu, \sigma)$. 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, najděte intervalový odhad pro μ.

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

3. Počet emailů za den modelujeme pomocí Poissonova rozdělení $Pois(\lambda)$. První týden v prosinci jsme dostali postupně 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 %.)

Testování hypotéz

4. (Všimněte si podobnosti a rozdílu oproti prvnímu příkladu.) 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?

5. 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).

Test dobré shody

6. Otestujte náhodnost kostky. Můžete házet ručně, máte-li kostku po ruce, nebo využít vhodný generátor (na webu např. `https://www.random.org/dice/?num=60` – výsledek jde kopírovat a vložit jako posloupnost čísel – nebo softwarový generátor na vašem stroji). Použijte Pearsonův χ^2 test dobré shody.

7. Pojďme zkontrolovat, zda počty emailů jsou opravdu dobře modelovány Poissonovým rozdělením. Můžete použít svoje data, nebo použít data Roberta Šamala, za loňský listopad. (Nejsou to všechny emaily, jenom „ty důležité“ podle klasifikace Gmailu, ale to by na statistických vlastnostech nemělo nic měnit.)

$$0, 6, 14, 8, 8, 9, 3, 3, 12, 12, 15, 7, 15, 2, 5, 13, 5, 17, 15, 11, 9, 2, 16, 8, 9, 11, 6, 2, 2, 9.$$

Rozmyslete, jak použít test dobré shody, zjistěte velikost λ a test proveďte. Zamyslete se nad výsledkem.