Zadanie:

1. Simulačne odhadnite pravdepodobnosť chyby prvého druhu pre t-test, keď dáta generujeme z normálneho rozdelenia. Mení sa táto pravdepodobnosť s meniacim sa počtom dát (nie počtom simulácií)? Vysvetlite.

2. Zopakujte 1. pre dáta generované z exponenciálneho rozdelenia s vhodnou voľbou λ (teda takou, aby bola ozaj meraná chyba prvého druhu, a nie nejaká iná pravdepodobnosť).

   (a) Ako sa mení táto pravdepodobnosť s meniacim sa počtom dát?

   (b) Jav v časti (a) má teoretické vysvetlenie. Aké?

   (c) Ako sa menia výsledky, keď meníme počet simulácií? Vysvetlite.

Poznámka: toto je len príklad, ako sprievodný text môže vyzerať. Vôbec sa nemusíte držať takéhoto štýlu alebo formátu. Výstup môže byť spísaný napr. aj viac "slohovo" alebo viac formálne (prípadne miestami ešte stručnejšie).

Domáca úloha 0, Samuel Rosa

1. Normálne rozdelenie

- Generoval som $n = 20$ realizácií z normálneho rozdelenia $N(170, 15^2)$ a testoval som $H_0 : E(X_i) = 170$. To som opakoval $N = 1000$-krát.

- Odhad pravdepodobnosti vyšiel približne 5 percent, konkrétne 0,048. Došlo teda k 48 zamietnutiam H_0 spomedzi 1 000.

- Pri zmene n (skúšal som medzi 10 a 100) sa odhad pravdepodobnosti držal okolo 5 percent.

- Vysvetlenie: tieto dáta spĺňajú predpoklady t-testu, a ten má z teórie 5-percentnú pravdepodobnosť chyby prvého druhu. Preto odhady tejto chyby kolísali okolo 5 percent, bez ohľadu na počet meraní.

2. Exponenciálne rozdelenie

- Tentokrát som generoval z $Exp(\lambda)$ (rovnaké n, N a H_0). Keďže $E(X) = 1/\lambda$ pre $X \sim Exp(\lambda)$, tak λ som zvolil ako 1/170, aby dáta spĺňali H_0.

- Odhad pravdepodobnosti chyby prvého druhu vyšiel 7,3%.

- (a) Pri rastúcom počte meraní klesala pravdepodobnosť k 5 percentám, viď obrázok v Rku *(alebo sa sem dá vložiť)*.

- (b) Vysvetľuje ho centrálna limitná veta: priemer nezávislých rovnako rozdelených náhodných premenných za istých predpokladov s rastúcim počtom meraní konverguje k normálnemu rozdeleniu – a teda príslušná t-štatistika konverguje k t rozdeleniu. Pre malý počet meraní je výsledné rozdelenie vzdialené od t-rozdelenia, čiže nie sú splnené predpoklady testu a vidíme vyššiu chybovosť. S rastúcim počtom meraním sa rozdelenie testovej štatistiky približuje k t-rozdeleniu a teda pravdepodobnosť chyby prvého druhu sa približuje k príslušným 5 percentám.

- (c) S rastúcim N správanie ostáva podobné, iba odhadované pravdepodobnosti sú menej "rozkolísané" a sú bližšie k "peknej" klesajúcej krivke (viď obrázok v Rku). To je vysvetlené tým, že s väčším N je môj získaný odhad pravdepodobnosti bližší k skutočnej pravdepodobnosti chyby prvého druhu, ktorá k 5 percentám v skúmanom prípade zrejme klesá po krivke, okolo ktorej kolísajú získané odhady.

Spolupráca: úlohu som vypracoval samostatne. *(Prípadne: s vysvetlením pomocou centrálnej limitnej vety mi pomohol William Gosset.)*