Domáca úloha 1

Preskúmajte kvalitu nasledovných odhadov variancie:

$$A = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad B = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

kde $x_1, \ldots, x_n$ sú namerané hodnoty a $\bar{x}$ je ich priemer.

1. Zvoľte relatívne nízke n (napr. $n = 10$) a generujte n nezávislých realizácií z rozdelenia $N(\mu, \sigma^2)$ pre zvolené μ a σ^2. Vypočítajte odhady A a B a porovnajte ich so skutočnou σ^2. Toto celé zopakujte N-krát (napr. 1000-krát) a zistite ako ďaleko sú v priemere odhady A a B od σ^2. Teda vypočítajte

$$Err_A = \frac{1}{N}\Big(\sum_{j=1}^{N} A_{(j)}\Big) - \sigma^2, \quad Err_B = \frac{1}{N}\Big(\sum_{j=1}^{N} B_{(j)}\Big) - \sigma^2,$$

   kde $A_{(j)}$ značí odhad A vypočítaný z j-teho opakovania (podobne $B_{(j)}$). Zrejme čím je Err bližšie k nule, tým lepšie.

   (a) Ktorý odhad vyšiel ako lepší?

   (b) Keď výpočet spustíte opakovane, ktorý odhad vychádza častejšie/stále lepší? Zdôvodnite pomocou teoretických poznatkov o S^2.

2. Vyskúšajte meniť počet dát n od veľmi malých (menej ako 10) po relatívne veľké (stovky, až tisíce).

   (a) Čo sa deje s chybovosťami Err_A a Err_B, keď meníme n?

   (b) Aký je vzťah medzi chybovosťou Err_A a Err_B so zvyšujúcim sa n? (Teda napr.: rozdiel v chybách sa zväčšuje/zmenšuje/skáče to hore-dolu okolo nejakej hodnoty bez ohľadu na n.)

   (c) Odpovede v (a) a (b) logicky zdôvodnite.