EML Assignment 3 - Problem 2

Muhammed Saeed and Ali Salaheldin Ali Ahmed

December 2022

We'll refer to the original sample as S, where s_i is the ith element from S. Similarly, B refers to the bootstrap sample, and b_i is the ith observation from the bootstrap sample.

Question 1

The sample is uniformly sampled from S. Thus, $P(b_1 \neq s_j)$ is the probability of *picking any of the other $n-1$ observation from S*, where $P(b_1 = s_i) = {}^1\!/_n, \forall s_i \in S$.

$$P(b_1 \neq s_j) = \sum_{i \in \{1 \ldots n\} \setminus \{j\}} P(b_1 = s_i) = \sum_{1 \leq i \leq n-1} \frac{1}{n} = \frac{n-1}{n} = 1 - \frac{1}{n}$$

Question 2

When constructing the bootstrap sample, the same process from *Question 1* is independently repeated n times (i.e. sampling n samples with replacement from S). Hence, for sample to not be included in the bootstrap sample, it must have not been selected as the b_i, $\forall i, 1 \leq i \leq n$.

$$P(s_j \notin B) = P(\bigwedge_{1 \leq i \leq n} (b_i \neq s_j)) = \prod_{1 \leq i \leq n} P(b_i \neq s_j) = \prod_{1 \leq i \leq n} (1 - \frac{1}{n}) = (1 - \frac{1}{n})^n$$

Question 3

The probability from *Question 1* increases with n, and approaches a probability of 1 for *very large values* of n ($n = 0 \rightarrow P \approx 0.99$).

On the other hand, while the probability from *Question 2* is also proportional to n, it more rapidly plateaus at a probability of $e^{-1} \approx 0.368$ for *sufficiently large values* of n ($n = 30 \rightarrow P \approx 0.362$). This means that on average, a little over the third of the original sample is not included in the bootstrap sample.