3. The sample mean

Problem 3. The sample mean

5/5 points (graded)

Let X be a continuous random variable. We know that it takes values between 0 and 6, but we do not know its distribution or its mean and variance, although we know that its variance is at most 4. We are interested in estimating the mean of X, which we denote by h. To estimate h, we take n i.i.d. samples $X_1, \ldots, X_n$, which all have the same distribution as X, and compute the sample mean

$$H = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. Express your answers for this part in terms of h and n using standard notation.

$\mathbf{E}[H] =$ h

h

Answer: h

Given the available information, the smallest upper bound for $\mathsf{Var}(H)$ that we can assert/guarantee is:

$\mathsf{Var}(H) \leq$ 4/n

$\frac{4}{n}$

Answer: 4/n

2. Calculate the smallest possible value of n such that the standard deviation of H is guaranteed to be at most 0.01.

This minimum value of n is: 40000

Answer: 40000

3. We would like to be at least 96% sure that our estimate is within 0.02 of the true mean h. Using the Chebyshev inequality, calculate the minimum value of n that will achieve this.

This minimum value of n is: 250000

Answer: 250000

4. Suppose now that X is uniformly distributed on $[h-3, h+3]$, for some unknown h. Using the Central Limit Theorem, identify the most appropriate expression for a 95% confidence interval for h. You may want to refer to the normal table.

Normal Table

Show

- $\left[H - \frac{\sqrt{1.96 \cdot 3}}{\sqrt{n}}, H + \frac{\sqrt{1.96 \cdot 3}}{\sqrt{n}}\right]$
- $\left[H - \frac{1.96}{\sqrt{3n}}, H + \frac{1.96}{\sqrt{3n}}\right]$
- $\left[H - \frac{1.96 \cdot \sqrt{3}}{\sqrt{n}}, H + \frac{1.96 \cdot \sqrt{3}}{\sqrt{n}}\right]$ (selected)

- $\left[H - \frac{1.96 \cdot 3}{\sqrt{n}}, H + \frac{1.96 \cdot 3}{\sqrt{n}}\right]$

STANDARD NOTATION

Solution:

1. We have

$$
\begin{aligned}
H &= \frac{X_1 + X_2 + \cdots + X_n}{n}, \\
\mathbf{E}[H] &= \frac{\mathbf{E}[X_1 + \cdots + X_n]}{n} = \frac{n \cdot \mathbf{E}[X]}{n} = h, \\
\sigma_H^2 &= \mathsf{Var}(H) = \frac{n \cdot \mathsf{Var}(X)}{n^2} \le \frac{4}{n}.
\end{aligned}
$$

2. From the previous part, we know that $\sigma_H \le 2/\sqrt{n}$. In order to guarantee that it is at most 0.01, we solve, $2/\sqrt{n} \le 0.001$ for n to obtain $n \ge 40000$.

3. We apply the Chebyshev inequality to H, with $\mathbf{E}[H]$ and $\mathsf{Var}(H)$ from part (1):

$$\mathbf{P}(|H - h| \ge 0.02) \le \frac{\sigma_H^2}{0.02^2} \quad \text{or} \quad \mathbf{P}(|H - h| \le 0.02) \ge 1 - \frac{\sigma_H^2}{0.02^2}.$$

   Substituting in our upper bound on σ_H^2, we obtain

$$1 - \frac{\sigma_H^2}{0.02^2} \ge 1 - \frac{2^2}{n \cdot 0.02^2}.$$

   Hence, to guarantee that our estimate is within 0.02 of the true mean h with probability of at least 99%, it suffices to have,

$$1 - \frac{2^2}{n \cdot 0.02^2} \ge 0.96.$$

   Solving this for n, we have that n must satisfy,

$$n \ge 250000.$$

4. Since X is uniform in the interval $[h-3, h+3]$, we know that the expected value of X is h and its variance, denoted by σ_H^2, is 3. Using the standard normal table, and the Central Limit Theorem, we know that for sufficiently large n,

$$\mathbf{P}\left(\left|\frac{H - h}{\sigma_H/\sqrt{n}}\right| \le 1.96\right) \approx 0.95.$$

   Hence,

$$\mathbf{P}\left(H - \frac{1.96 \cdot \sqrt{3}}{\sqrt{n}} \le h \le H + \frac{1.96 \cdot \sqrt{3}}{\sqrt{n}}\right) \approx 0.95.$$

   Therefore, the 95% confidence interval for h is $\left[H - \frac{1.96 \cdot \sqrt{3}}{\sqrt{n}}, H + \frac{1.96 \cdot \sqrt{3}}{\sqrt{n}}\right]$.