MITx: 6.041x Introduction to Probability - The Science of Uncertainty

Bookmarks

- Unit 0: Overview
- Entrance Survey
- Unit 1: Probability models and axioms
- Unit 2: Conditioning and independence
- Unit 3: Counting
- Unit 4: Discrete random variables
- Exam 1
- Unit 5: Continuous random variables

Unit 8: Limit theorems and classical statistics > Problem Set 8 > Problem 3 Vertical: The sample mean

Bookmark

Problem 3: The sample mean

(5/5 points)

Let X be a continuous random variable. We know that it takes values between 0 and 3, but we do not know its distribution or its mean and variance. We are interested in estimating the mean of X, which we denote by h. We will use 1.5 as a conservative value (upper bound) for the standard deviation of X. To estimate h, we take n i.i.d. samples $X_1, X_2, \ldots, X_n$, 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 ✓ **Answer:** h

Given the available information, the smallest upper bound for $\textbf{var}(H)$ is:

- Unit 6: Further topics on random variables
- Unit 7: Bayesian inference
- Exam 2
- **Unit 8: Limit theorems and classical statistics**

Unit overview

Lec. 18: Inequalities, convergence, and the Weak Law of Large Numbers
Exercises 18 due Apr 27, 2016 at 23:59 UTC

Lec. 19: The Central Limit Theorem (CLT)
Exercises 19 due Apr 27, 2016 at 23:59 UTC

Lec. 20: An introduction to classical statistics
Exercises 20 due Apr 27, 2016 at 23:59 UTC

2.25/n ✓ **Answer:** 2.25/n

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

This minimum value of n is: 22500 ✓ **Answer:** 22500

3. We would like to be at least 99% sure that our estimate is within 0.05 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: 90000 ✓ **Answer:** 90000

4. Assume that X is uniformly distributed on $[0,3]$. Using the Central Limit Theorem, identify the most appropriate expression for a 95% confidence interval for h.

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

○ $\left[H-\frac{\sqrt{1.96\cdot 3}}{\sqrt{4n}}, H+\frac{\sqrt{1.96\cdot 3}}{\sqrt{4n}}\right]$

Solved problems

Additional theoretical material

Problem Set 8
Problem Set 8 due Apr 27, 2016 at 23:59 UTC

Unit summary

- Unit 9: Bernoulli and Poisson processes
- Unit 10: Markov chains
- Exit Survey
- Final Exam

● $\left[H - \frac{1.96\sqrt{3}}{\sqrt{4n}}, H + \frac{1.96\sqrt{3}}{\sqrt{4n}}\right]$ ✓

○ $\left[H - \frac{1.96 \cdot 3}{\sqrt{4n}}, H + \frac{1.96 \cdot 3}{\sqrt{4n}}\right]$

Answer:

1.

$$\begin{aligned} H &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\ \mathbf{E}[H] &= \frac{\mathbf{E}[X_1 + X_2 + \cdots + X_n]}{n} = \frac{n \cdot \mathbf{E}[X]}{n} = h \\ \sigma_H^2 &= \mathrm{var}(H) = \frac{n \cdot \mathrm{var}(X)}{n^2} \leq \frac{1.5^2}{n} \end{aligned}$$

2. From part (a), we know that $\sigma_H \leq 1.5/\sqrt{n}$. Therefore, we solve $1.5/\sqrt{n} \leq 0.01$ for n to obtain $n \geq 22500$.

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

$$\mathbf{P}(|H-h|\geq 0.05)\leq \frac{\sigma_H^2}{0.05^2},\quad \text{or}\quad \mathbf{P}(|H-h|\leq 0.05)\geq 1-\frac{\sigma_H^2}{0.05^2}.$$

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

$$1-\frac{\sigma_H^2}{0.05^2}\geq 1-\frac{1.5^2}{n\cdot 0.05^2}.$$

Hence, to guarantee that our estimate is within 0.05 of the true mean h with probability at least 99%, it suffices to have

$$1-\frac{1.5^2}{n\cdot 0.05^2}\geq 0.99.$$

Solving for n, we have that n must satisfy

$$n\geq \left(\frac{1.5}{0.05}\right)^2\frac{1}{0.01}=90000.$$

4.

Since X is uniform in the interval $[0,3]$, we know that the expected value of X, denoted by h, is 1.5 and the variance of X, denoted by σ_H^2, is $3/4$. Using the standard normal table and the Central Limit Theorem, we know that for sufficiently large n,

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

Hence,

$$\mathbf{P}\left(H - \frac{1.96\cdot\sqrt{3/4}}{\sqrt{n}} \leq h \leq H + \frac{1.96\cdot\sqrt{3/4}}{\sqrt{n}}\right) \approx 0.95.$$

Therefore, the 95% confidence interval for h is $\left[H - \frac{1.96\sqrt{3}}{\sqrt{4n}}, H + \frac{1.96\sqrt{3}}{\sqrt{4n}}\right]$.

You have used 1 of 2 submissions

DISCUSSION

Click “Show Discussion" below to see discussions on this problem.