MTH 102: Probability and Statistics

Quiz 6

22/05/2020

Sanjit K. Kaul

Question 1. [30 marks] A random variable X is uniformly distributed over the interval $(0, b)$, where b is unknown. You can perform an experiment that returns a value governed by the distribution of X. You can perform the experiment $n \geq 1$ times.

Answer the following questions.

1) Propose an unbiased estimator $\hat{B}_n$ of b. Justify your choice using the definition of an unbiased estimator. Provide the corresponding detailed math in the upload. Only mention the unbiased estimator in the provided text box.
2) Propose a biased estimator of b that is asymptotically unbiased. Justify your choice using the definition of an asymptotically unbiased estimator. Provide the corresponding detailed math in the upload. Only mention the asymptotically unbiased estimator in the provided text box.

Question 2. [70 marks] A classroom consists of 500 seats. You are interested in the expected occupancy of the classroom during a lecture on probability and statistics. Occupancy of the classroom is defined as the total number of seats occupied by students. The expected occupancy of the room is unknown and you must estimate the same by measuring the occupancy of the room on n different days. Answer the following.

1) (20 marks) Propose an unbiased estimator of the expected occupancy. Justify your choice using the definition of an unbiased estimator. Provide the corresponding detailed math in the upload. Only mention the unbiased estimator in the provided text box.
2) (50 marks) Derive the minimum number of days for which you must plan to measure the occupancy so that 90% of the times the resulting estimate is within ± 5 of the true expected occupancy? Provide the corresponding detailed math in the upload. Only mention the (name of the) inequality you used and (if you have calculated it) the minimum number of days in the provided text box.

Question 1. [30 marks] A random variable X is uniformly distributed over the interval $(0, b)$, where b is unknown. You can perform an experiment that returns a value governed by the distribution of X. You can perform the experiment $n \geq 1$ times.

Answer the following questions.

1) Propose an unbiased estimator $\hat{B}_n$ of b. Justify your choice using the definition of an unbiased estimator. Provide the corresponding detailed math in the upload. Only mention the unbiased estimator in the provided text box.
2) Propose a biased estimator of b that is asymptotically unbiased. Justify your choice using the definition of an asymptotically unbiased estimator. Provide the corresponding detailed math in the upload. Only mention the asymptotically unbiased estimator in the provided text box.

(1) Note that $b = 2\mu_X$.

Consider $\hat{B}_n = 2M_n(X)$, where $M_n(X)$ is the sample mean.

$$E[\hat{B}_n] = E[2M_n(X)] = 2E[M_n(X)] = 2\mu_X = b.$$

$\therefore$ $\hat{B}_n = 2M_n(X)$ is an unbiased estimator of b.

15

(2) A biased estimator, which is asymptotically unbiased. That is the bias must go to zero as $n \to \infty$.

An example is $\hat{B}_n = 2M_n(X) + \frac{1}{n}$

$E[\hat{B}_n] = 2b + \frac{1}{n}$. Clearly, biased.

$\lim_{n\to\infty} E[\hat{B}_n] = 2b$. Thus, asymptotically biased.

15

Of course, there are other possibilities!

Question 2. [70 marks] A classroom consists of 500 seats. You are interested in the expected occupancy of the classroom during a lecture on probability and statistics. Occupancy of the classroom is defined as the total number of seats occupied by students. The expected occupancy of the room is unknown and you must estimate the same by measuring the occupancy of the room on n different days. Answer the following.

1) (20 marks) Propose an unbiased estimator of the expected occupancy. Justify your choice using the definition of an unbiased estimator. Provide the corresponding detailed math in the upload. Only mention the unbiased estimator in the provided text box.
2) (50 marks) Derive the minimum number of days for which you must plan to measure the occupancy so that 90% of the times the resulting estimate is within ± 5 of the true expected occupancy? Provide the corresponding detailed math in the upload. Only mention the (name of the) inequality you used and (if you have calculated it) the minimum number of days in the provided text box.

(1) Let Y be the RV that governs the occupancy of the room.
The unbiased estimator is $M_n(Y)$. We know that $E[M_n(Y)] = E[Y]$.

20/20

(2) We will use the Chebyshev Inequality.

$$P\left[|M_n(Y) - E[M_n(Y)]| > c\right] \le \frac{Var[M_n(Y)]}{c^2}$$

20/50

Since $M_n(Y) = E[Y]$,

$$P\left[|M_n(Y) - E[Y]| > c\right] \le \frac{Var[M_n(Y)]}{c^2} = \frac{Var[Y]}{nc^2}$$

20/50

Since Y is a Binomial RV,

$$Var[Y] = 500p(1-p),$$

where p is the prob that a seat is occupied. Of course, p is unknown.

$$P\left[|M_n(Y) - E[Y]| > c\right] \le \frac{Var[Y]}{nc^2}$$

Since we want the sample mean to be within ± 5, 90% of the times, for any p, we will assume the worst case $Var[Y] = (0.5)^2 500$

We require $\frac{Var[Y]}{nc^2} = 0.1$, where $c = 5$

10/50

$\therefore$ The smallest value of n is

$$\frac{(0.5)^2 500}{(0.1)(5)^2} = \frac{500}{(100)(0.1)} = 50$$