Lecture 16 Worsheet

Chrysafis Vogiatzis

We will solve this in-class activity before breaking out into groups (as usual) and working on our worksheets. Just as a reminder: the Lecture 15 worksheet is due **tomorrow** by noon on gradescope (instead of the original deadline).

Activity 1: Uniform distribution

Last time, we worked with estimating the upper bound of a discrete uniformly distributed population X; today, we begin with the continuous version! Assume we are trying to estimate the upper bound of a continuous uniformly distributed population X. In English: if we have a random number generator producing (real) numbers between 0 and α, how can we estimate what α is after collecting some data from the generator?

Problem 1: Intuition

For example, say we have obtained $X_1 = 7.7, X_2 = 15.21, X_3 = 9.1$, then what is a "good" estimate for the true value of α? Is saying that all numbers that are produced by this generator are between 0 and 10 true? How about between 0 and 20?

Answer to Problem 1.

A good estimate we can come up with for the upper bound of the generator is that it is ...

Problem 2: Bias for $n = 3$

Assume that we use the following estimator:

$$\hat{\Theta} = \max\left\{X_1, X_2, \ldots, X_n\right\},$$

where n is the number of data points we have collected. For example, if $X_1 = 7.7, X_2 = 15.21, X_3 = 9.1$ are the data obtained then $n = 3$, and $\hat{\theta}$ (the estimate obtained) is 15.21. **What is the bias of the estimator?** Assume that we have already obtained [1] that

$$E\left[\hat{\Theta}\right] = \frac{n}{n+1} \cdot \alpha.$$

[1] Ask me why, if you are interested!

Answer to Problem 2.

$bias\left[\hat{\Theta}\right] =$

Problem 3: Bias for $n \to \infty$

What happens to the bias as we increase the number of observations from the generator? Specifically, what is $\lim_{n\to\infty} bias\left[\hat{\Theta}\right]$?

Answer to Problem 3.

We see that the bias goes to 0 as we obtain more and more observations. This is great news! We define **a consistent estimator** (sometimes also called an *asymptotically* consistent estimator) as an estimator that has bias converging to 0 as the number of data points used in its calculation increases. In practice, this means that the estimates obtained get closer and closer to the true value of the parameter we are estimating as we use more and more data.

Problem 4: Variance for $n = 3$

Let's go back to our estimator $\hat{\Theta} = \max\{X_1, X_2, \ldots, X_n\}$ for $n = 3$ (i.e., only 3 data points from the generator). What is the variance of the estimator? Assume we have already calculated that [2]

$$E\left[\hat{\Theta}^2\right] = \frac{n}{n+2} \cdot \alpha^2.$$

[2] Again: ask me for the derivation if you are interested!

Answer to Problem 4.

$Var\left[\hat{\Theta}\right] =$

Problem 5: The mean square error

What is the MSE for the estimator $\hat{\Theta}$ for $n = 3$? How about for $n = 10$? When $n \to \infty$?

Answer to Problem 5.

For $n = 3$: $MSE = bias^2\left[\hat{\Theta}\right] + Var\left[\hat{\Theta}\right] =$

For $n = 10$:

$$MSE =$$

As $n \to \infty$:

$$\lim_{n\to\infty} MSE =$$

Activity 2: Comparing point estimators

Assume that a population is distributed with pdf $f(x) = \theta\left(x - \frac{1}{2}\right) + 1$, $0 \leq x \leq 1$, where θ is an unknown parameter. We have also been able to collect a sample of n observations and have calculated their sample average as $\overline{X}$. We have been trying estimators for θ and we want to compare three of them and pick the best:

1. $\hat{\Theta}_1 = 2\overline{X} - 1$
2. $\hat{\Theta}_2 = 12\overline{X} - 6$

Problem 6: MSE

Which one of the two has the smallest mean square error? [3]

[3] As a reminder, for an estimator $\hat{\Theta}$, the mean square error is:

$$MSE = bias\left[\Theta\right]^2 + Var\left[\hat{\Theta}\right].$$

Answer to Problem 6.

Problem 7: Application

For the previous population from Problem 6, we have collected a sample of 6 items and found them equal to $X_1 = 0.8, X_2 = 0.83, X_3 = 0.95, X_4 = 0.72, X_5 = 0.85, X_6 = 0.65$. What is a good estimate for θ? Use the point estimator that provided the best MSE from Problem 8.

Answer to Problem 7.

Activity 3: Point estimators for the exponential distribution

Assume that a population X is exponentially distributed; however, we have no idea what λ is. You recalled one thing though from IE 300: **the expected time between events is equal to $\frac{1}{\lambda}$**. This gives you an idea. You will wait to observe the times between some events and use them to estimate λ. More specifically, you will estimate λ as 1 over the average time between events.

For example, if the sampled times between two consecutive events are equal to 3 minutes, 2 minutes, 5 minutes, 4 minutes, 2 minutes, 2 minutes, you will estimate λ as 1 over 3 minutes!

Problem 8: Biased or not?

Is this estimate for λ biased or not? In mathematical terms, is $\hat{\lambda} = \frac{1}{\overline{X}}$, where $\overline{X}$ is the average of n observations, biased? [4]

[4] Can you make the claim that $E\left[\frac{1}{X}\right] = \frac{1}{E[X]}$? Yes or no?

Answer to Problem 8.

Problem 9: MSE

What is the mean square error for the estimator $\hat{\lambda} = \frac{1}{\overline{X}}$ for the unknown rate of an exponential distribution after observing n data points?

Answer to Problem 9.