LECTURE: CONSISTENCY

1. Consistency

We've seen 2 ways of measuring how good our estimator $\hat{\theta}$ is: Bias and Mean-Square Error (MSE)

Recall:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$
$$\text{MSE}(\hat{\theta}) = [\text{Bias}(\hat{\theta})]^2 + \text{Var}(\hat{\theta})$$

Main Problem: Although the MSE is a good measure of the accuracy of an estimator, but it does not take into account the sample size n

Ideally, we would like to define an estimator as "good" if the probability of "missing" the true parameter θ goes to 0 as the $n \to \infty$. This is called **consistency**.

Setting: Let $Y_1, \dots, Y_n$ be iid samples from a population, and let $\hat{\theta}_n$ be an estimator for a parameter θ

Here we use the subscript n to emphasize that the estimator depends on the sample size n

Example: To estimate $\theta = \mu$, we use the estimator

$$\hat{\theta}_n = \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Definition:

$\hat{\theta}_n$ is **consistent** for θ if for all $\epsilon > 0$

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) = 0$$

In other words, $\hat{\theta}_n$ is consistent if the probability that $\hat{\theta}_n$ misses the true parameter θ (the "error") goes to 0 as $n \to \infty$

Note: This type of conv is called **convergence in probability**

In the case of unbiased estimators, we get a much easier formula:

Magic Consistency Formula:

Let $\hat{\theta}_n$ be an **unbiased** estimator for θ

Then $\hat{\theta}_n$ is consistent for θ if $\lim_{n\to\infty} \text{Var}(\hat{\theta}_n) = 0$

Why? Let $\epsilon > 0$. Then by Chebyshev we get

$$P(|\hat{\theta}_n - E(\hat{\theta}_n)| \geq \epsilon) \leq \frac{\text{Var}(\hat{\theta}_n)}{\epsilon^2}$$

Since $\hat{\theta}_n$ is unbiased $E(\hat{\theta}_n) = \theta$, hence we get

$$P(|\hat{\theta}_n - \theta| \geq \epsilon) \leq \frac{\text{Var}(\hat{\theta}_n)}{\epsilon^2}$$

Taking the limit of both sides,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) \leq \frac{1}{\epsilon^2} \lim_{n\to\infty} \text{Var}(\hat{\theta}_n) = 0 \text{ (by assumption)}$$

Hence $\hat{\theta}_n$ is consistent for θ

Example 1:

Show that the sample mean $\overline{Y}_n$ is a consistent estimator for μ

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

We have shown that $E\left(\bar{Y}_n\right) = \mu$ for $\bar{Y}_n$ is unbiased for μ

We also know that $\mathrm{Var}(\bar{Y}_n) = \frac{\sigma^2}{n} \overset{n\to\infty}{\longrightarrow} 0$

Hence $\bar{Y}_n$ is consistent for μ.

This implies one of the most fundamental results in probability:

Weak Law of Large Number:

Let $Y_1, \cdots, Y_n$ be iid with $E(Y_i) = \mu$ then

$$\overline{Y}_n = \frac{Y_1 + \cdots + Y_n}{n} \longrightarrow \mu \text{ as } n \to \infty$$

Here the convergence is convergence in probability

Why important? Intuitively, we can think of expected value as taking the average of a large number of experiments, which is what $\overline{Y}_n$ is. As n gets large this says that $\bar{Y}_n$ approaches the "true" average μ

Note: We can similarly show that our unbiased estimator for variance S^2 is a consistent estimator for the population variance.

2. Method of Moments

So far we have only discussed three estimators: $\bar{Y}$, S^2, and $\hat{p}$

In practice, there are many more parameters we might like to estimate: How can we construct estimators for them?

In this section we look at two methods of constructing estimators: the method of moments and the maximum likelihood estimator (MLE)

Method of Moments: This method is based on the fact that the sample mean $\bar{Y}$ is close to the population mean μ.

The method of moments works as follows:

(1) Write the parameter of interest in terms of μ

(2) Replace μ by $\bar{Y}$ to get our estimator

Example 2:

Suppose $X \sim$ Unif $[0, b]$

(a) Find the method of moments estimator for b

$$\mu = E(X) = \frac{0+b}{2} = \frac{b}{2} \Rightarrow b = 2\mu$$

Replacing μ with $\bar{Y}$ we then get $\hat{b} = 2\bar{Y}$

(b) Is $\hat{b}$ consistent for b?

$$E(\hat{b}) = E\left(2\bar{Y}\right) = 2\mu = b$$

Hence $\hat{b}$ is unbiased for b

Moreover, by the Magic Consistency Formula

$$\mathrm{Var}(\hat{b}) = \mathrm{Var}(2\bar{Y}) = 4\,\mathrm{Var}(\bar{Y}) = 4\left(\frac{\sigma^2}{n}\right) = 4\left(\frac{b^2}{12n}\right) = \frac{b^2}{3n} \overset{n\to\infty}{\longrightarrow} 0$$

Here we used that if $X \sim$ Unif $[a,b]$ then $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$

Hence $\hat{b}$ is consistent for b

Example 3:

Suppose $X \sim$ Geom (p)

Find the method of moments estimator for p

$$\mu = E(X) = \frac{1}{p} \Rightarrow p = \frac{1}{\mu} \Rightarrow \hat{p} = \frac{1}{\bar{\bar{Y}}}$$

Example 4:

Suppose $X \sim$ Poi (λ)

Find the method of moments estimator for λ

$$\mu = E(X) = \lambda \Rightarrow \lambda = \mu \Rightarrow \hat{\lambda} = \bar{Y}$$

Note: Sometimes we cannot construct an method of moments estimator in this way, i.e. we cannot solve for the parameter in terms of the population mean μ.

Example 5:

Suppose $X \sim$ Unif $[a, b]$

Find the method of moments estimators for a and b

Sketch: Here $\mu = E(X) = \frac{a+b}{2}$ but we cannot solve for either a or b in terms of μ.

In that case we *also* use $\sigma^2 = \frac{(b-a)^2}{12}$ and then we get two equations for the two unknowns a and b.

You solve for a and b in terms of μ and σ^2 and replace μ by $\overline{Y}$ and σ^2 by S^2 to get your estimators $\hat{a}$ and $\hat{b}$

Since the algebra gets messy really quickly, we will not be pursuing this any further.

3. MLE: Motivation

The method of moments is very intuitive but often does not lead to the best estimators. A second way of constructing estimators is called the maximum likelihood estimator (MLE)

Example 6:

You have a bag which contains three balls, which are either red or white

Suppose we draw out two balls from the bag without replacement, and they are both red

What is a good estimate of the number of red balls in the bag?

Since we drew two red balls, the bag either contains two red balls or three red balls.

Scenario 1: The bag contains two red balls (and one white)

Then the probability of "Drawing two red balls" (= our draw) is $\frac{1}{3}$ (draw a tree diagram)

Scenario 2: The bag contains three red balls

Then the probability of our draw is 1

Therefore, scenario 2 is more likely, so a reasonable estimate for the number of red balls in the bag is 3, since that maximizes the probability of our draw.

The method illustrated in the example above is known as the method of **maximum likelihood**, since we choose the parameter which maximizes the probability of having our sample.