LECTURE: FINAL EXAM – REVIEW

1. MLE

Example 1:

Let $Y_1, Y_2, \cdots Y_n$ be iid samples from a distribution with pdf

$$f(y) = \begin{cases} \theta y^{\theta-1} & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Here $\theta > -1$ is the parameter of interest

(a) Find the method of moments estimator for θ

STEP 1: First find the mean μ

$$\mu = \int_{-\infty}^{\infty} y f(y) dy = \int_0^1 y \left(\theta y^{\theta-1}\right) dy = \int_0^1 \theta y^{\theta} dy = \theta \left[\frac{y^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

STEP 2: Solve for θ in terms of μ

$$\frac{\theta}{\theta+1} = \mu \Rightarrow \theta = \mu(\theta+1) \Rightarrow \theta = \mu\theta + \mu \Rightarrow \theta(1-\mu) = \mu \Rightarrow \theta = \frac{\mu}{1-\mu}$$

STEP 3: Replace μ with $\bar{Y}$

$$\hat{\theta} = \frac{\bar{Y}}{1-\bar{Y}}$$

(b) Find the MLE for θ

STEP 1: Find $L(Y_1, Y_2, \cdots, Y_n|\theta)$

$$\begin{aligned} L(Y_1, Y_2, \cdots, Y_n|\theta) =& f(Y_1)f(Y_2)\cdots f(Y_n) \\ =& \left(\theta Y_1^{\theta-1}\right)\left(\theta Y_2^{\theta-1}\right)\cdots\left(\theta Y_n^{\theta-1}\right) \\ =& \theta^n \left(Y_1\cdots Y_n\right)^{\theta-1} \end{aligned}$$

STEP 2:

$$\ln\left(L(Y_1, Y_2, \cdots, Y_n|\theta)\right) = \ln\left(\theta^n \left(Y_1\cdots Y_n\right)^{\theta-1}\right) = n\ln(\theta) + (\theta-1)\ln\left(Y_1\cdots Y_n\right)$$

STEP 3: Differentiate with respect to θ

$$\begin{aligned} \frac{d}{d\theta}\ln\left(L(Y_1, Y_2, \cdots, Y_n|\theta)\right) =& \frac{d}{d\theta}\left(n\ln(\theta) + (\theta-1)\ln\left(Y_1\cdots Y_n\right)\right) \\ =& \left(\frac{n}{\theta}\right) + \ln\left(Y_1\cdots Y_n\right) = 0 \end{aligned}$$

$$\frac{n}{\theta} = -\ln\left(Y_1\cdots Y_n\right) \Rightarrow \frac{\theta}{n} = -\frac{1}{\ln\left(Y_1\cdots Y_n\right)} \Rightarrow \theta = \frac{-n}{\ln\left(Y_1\cdots Y_n\right)}$$

STEP 4: Answer

$$\hat{\theta} = \frac{-n}{\ln\left(Y_1\cdots Y_n\right)} = \frac{-n}{\ln(Y_1) + \cdots + \ln(Y_n)} = \frac{-n}{\sum_{i=1}^n \ln\left(Y_i\right)}$$

2. Confidence Intervals

Example 2:

(a) The amount of time it takes you to walk to your classroom is random and has a standard deviation of 6 mins. Suppose you walk to the classroom 81 times. The average amount of time is 18 mins. Give a 95 % confidence interval for the mean walking time, assuming $P(Z \leq -2) = 0.025$

Since the sample size $n = 81$ is large, the confidence interval is

$$\left[\hat{L}, \hat{U}\right] = \left[\overline{Y} - \hat{\sigma}\left(z_{\alpha/2}\right), \overline{Y} + \hat{\sigma}\left(z_{\alpha/2}\right)\right]$$

Here $\overline{Y} = 18$ and since $\sigma = 6$ is known we use

$$\hat{\sigma} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{81}} = \frac{6}{9} = \frac{2}{3}$$

We want $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

From the note above we get $P(Z \leq -2) = 0.025$ and so $z_{\alpha/2} = 2$

Therefore our confidence interval is

$$\left[\hat{L}, \hat{U}\right] = \left[18 - \left(\frac{2}{3}\right)(2), 18 + \left(\frac{2}{3}\right)(2)\right] = \left[\frac{50}{3}, \frac{58}{3}\right] = [16.67, 19.33]$$

(b) Assume the average is still 18 mins and the standard deviation 6, and $P(Z \leq -2) = 0.025$ how many times do you need to walk to the classroom so that the total length of the 95% confidence interval is less than 2 mins?

The length of the interval $\left[\hat{L}, \hat{U}\right] = \left[\overline{Y} - \hat{\sigma}\left(z_{\alpha/2}\right), \overline{Y} + \hat{\sigma}\left(z_{\alpha/2}\right)\right]$ is

$$\overline{Y} + \hat{\sigma}\left(z_{\alpha/2}\right) - \left(\overline{Y} - \hat{\sigma}\left(z_{\alpha/2}\right)\right) = 2\hat{\sigma}\left(z_{\alpha/2}\right)$$

Here $\hat{\sigma} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{n}}$ and since $\alpha = 0.95$ we still have $z_{\alpha/2} = 2$ and so

$$\not{2}\hat{\sigma}\left(z_{\alpha/2}\right) \leq \not{2} \Rightarrow \left(\frac{6}{\sqrt{n}}\right)(2) \leq 1 \Rightarrow \frac{1}{\sqrt{n}} \leq \frac{1}{12} \Rightarrow \sqrt{n} \geq 12 \Rightarrow n \geq 144$$

Hence you need to walk at least 144 times

3. Hypothesis Testing

Example 3:

It is thought that university students do not get enough sleep during the semester.

You claim that students get less than 7 hours of sleep per day

A random sample of 49 students was taken with a sample mean of 6.7 hours and a sample standard deviation 0.7 hours

(a) Set up the 4 components of a hypothesis test

Parameter of interest μ = the average amount of sleep

(1) Alternative hypothesis: $\mu < 7$

(2) Null hypothesis: $\mu = 7$

(3) Test statistic: $\overline{Y} = 6.7$

(4) Rejection Region $\left\{\overline{Y} \leq k\right\}$

(b) Is your claim supported at the 0.05 level?

You could change $\mu \leq 7$ to $-\mu \geq -7$ but here it is easier to do everything from scratch

Since we don't know σ we use S instead

$$\hat{\sigma} = \frac{S}{\sqrt{n}} = \frac{0.7}{\sqrt{49}} = \frac{0.7}{7} = 0.1$$

Since the sample size is large we convert to Z

$$P(\overline{Y} \leq k) = 0.05 \Rightarrow P\left(\frac{\overline{Y}-7}{\hat{\sigma}} \leq \frac{k-7}{\hat{\sigma}}\right) = 0.05 \Rightarrow P\left(Z \leq \frac{k-7}{0.1}\right) = 0.05$$

By the $Z-$table we get $P(Z \leq -1.64) \approx 0.05$ and so

$$\frac{k-7}{0.1} = -1.64 \Rightarrow k = 7 - (1.64)(0.1) = 7 - 0.164 = 6.836$$

Hence our rejection region is $\left\{\overline{Y} \leq 6.836\right\}$

Answer: Since the test statistic $\overline{Y} = 6.7$ is in the rejection region, since $6.7 \leq 6.836$, we reject the null hypothesis, and so our claim is supported with a level of $\alpha = 0.05$

(c) What is the $p-$value of this test?

Here the observed value is $\overline{Y} = 6.7$

$$\begin{aligned} p\text{ -value } &= P(\overline{Y} \leq 6.7 \text{ given Null is true }) \\ &= P(\overline{Y} \leq 6.7 \text{ given } \mu = 7) \\ &= P\left(\frac{\overline{Y}-7}{\hat{\sigma}} \leq \frac{6.7-7}{0.1}\right) \\ &= P(Z \leq -3) \\ &= 0.0013 \end{aligned}$$

(d) Suppose we have reason to suspect the mean number of hours of students sleep is actually 6.

Considering $\mu_a = 6$ what is β for this test?

$$\begin{aligned}\beta =& P(\overline{Y} \text{ is outside RR when Alt is true })\\ =& P(\overline{Y} \geq k) \text{ given } \mu = 6\\ =& P(\overline{Y} \geq 6.836) \text{ given } \mu = 6\\ =& P\left(\frac{\overline{Y} - 6}{\hat{\sigma}} \geq \frac{6.836 - 6}{0.1}\right)\\ =& P(Z \geq 8.36)\\ =& 0.0000001\end{aligned}$$

4. Consistency

Example 4:

There are three boxes, containing $0, \theta, \theta + 1$ jellybeans.

Each of n people opens a box uniformly at random and takes all the jellybeans in the box and the boxes are reset after each person takes their turn

Let $Y_1, Y_2 \cdots, Y_n$ be the number of beans taken by each person

(a) Calculate the Bias of $\hat{\theta} = \overline{Y}$ as an estimator for θ

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$E(\hat{\theta}) = E\left(\overline{Y}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$$

Since each box has a probability of $\frac{1}{3}$ of being picked, we have

$$E(Y_i) = 0\left(\frac{1}{3}\right) + \theta\left(\frac{1}{3}\right) + (\theta+1)\left(\frac{1}{3}\right) = \frac{2\theta+1}{3}$$

$$E(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{2\theta+1}{3}\right) = \frac{2\theta+1}{3}$$

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = \left(\frac{2\theta+1}{3}\right) - \theta = \frac{2\theta+1-3\theta}{3} = \frac{1-\theta}{3}$$

(b) For what value of θ is $\hat{\theta}$ an unbiased estimator?

$$\text{Bias}(\hat{\theta}) = 0 \Rightarrow \frac{1-\theta}{3} = 0 \Rightarrow \theta = 1$$

(c) In that case, show that $\hat{\theta}$ is consistent for θ

Since $\hat{\theta}$ is unbiased, it is enough to show that $\text{Var}(\hat{\theta})$ goes to 0 as $n \to \infty$

Since the Y_i are independent, we have

$$\text{Var}(\hat{\theta}) = \text{Var}(\overline{Y}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(Y_i)$$

By the magic variance formula

$$\text{Var}(Y_i) = E\left(Y_i^2\right) - \left(E(Y_i)\right)^2$$

$$E(Y_i^2) = 0^2\left(\frac{1}{3}\right) + 1^2\left(\frac{1}{3}\right) + 2^2\left(\frac{1}{3}\right) = \frac{5}{3}$$

Since $E(Y_i) = \frac{2(1)+1}{3} = 1$ we get

$$\text{Var}(Y_i) = \frac{5}{3} - (1)^2 = \frac{2}{3}$$

$$\text{Hence }\ \text{Var}(\hat{\theta}) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{2}{3} = \frac{1}{n^2}\left(\frac{2}{3}n\right) = \frac{2}{3}\left(\frac{1}{n}\right) \xrightarrow{n\to\infty} 0$$

Therefore $\hat{\theta}$ is consistent for θ