APMA 1650 – Homework 7

Due Monday, August 1, 2016

Homework is due during class or by 3:45 pm in the homework drop box in 182 George St.
Show all of your work used in deriving your solutions.

1. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with probability density function parameterized by θ given by

$$f_\theta(y) = \begin{cases} \theta y^{\theta-1} & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ is the parameter of interest.

(a) Show that the sample mean $\bar{Y}$ is an unbiased estimator for $\frac{\theta}{\theta+1}$.

Since the expected value of $\bar{Y}$ is the population mean μ, first we find μ.

$$\mu = \int_0^1 y\theta y^{\theta-1} dy$$
$$= \int_0^1 \theta y^\theta dy$$
$$= \theta \frac{y^{\theta+1}}{\theta+1}\Big|_1^0$$
$$= \frac{\theta}{\theta+1}$$

(b) Show that the sample mean $\bar{Y}$ is a consistent estimator for $\frac{\theta}{\theta+1}$.

Since the estimator is unbiased, for consistency all we have to do is show that its variance goes to 0 as n goes to infinity. We use the Magic Variance Formula to find the population variance. Let Y be a sample from the population. Then

$$\mathbb{E}(Y^2) = \int_0^1 y^2\theta y^{\theta-1} dy$$
$$= \int_0^1 \theta y^{\theta+1} dy$$
$$= \theta \frac{y^{\theta+2}}{\theta+2}\Big|_0^1$$
$$= \frac{\theta}{\theta+2}$$

By the Magic Variance Formula, for the population variance we have

$$\sigma^2 = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2$$
$$= \frac{\theta}{\theta+2} - \left[\frac{\theta}{\theta+1}\right]^2$$

To get the variance of the sample mean, we divide this by n. This gives us:

$$Var(\bar{Y}) = \frac{1}{n}\left(\frac{\theta}{\theta+2} - \left[\frac{\theta}{\theta+1}\right]^2\right)$$

Since everything inside the parentheses is constant, this goes to 0 as n goes to infinity. It is enough to note that since the population variance is finite, the sample variance must go to 0 as n goes to infinity.

2. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with probability density function parameterized by θ given by

$$f_\theta(y) = \begin{cases} (\theta+1)y^\theta & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > -1$ is the parameter of interest. Find an estimator for θ using the method of moments.

In the method of moments, we set the population mean $\mu = \bar{Y}$ and solve for θ.

$$\mu = \int_0^1 y(\theta+1)y^\theta dy$$
$$= \int_0^1 (\theta+1)y^{\theta+1} dy$$
$$= (\theta+1)\frac{y^{\theta+2}}{\theta+2}\Big|_0^1$$
$$= \frac{\theta+1}{\theta+2}$$

Now we set $\mu = \bar{Y}$ and solve for θ.

$$\frac{\theta+1}{\theta+2} = \bar{Y}$$
$$\theta+1 = \bar{Y}(\theta+2) = \bar{Y}\theta + 2\bar{Y}$$
$$\theta(1-\bar{Y}) = 2\bar{Y} - 1$$
$$\theta = \frac{2\bar{Y}-1}{1-\bar{Y}}$$

Thus the method of moments estimator is:

$$\hat{\theta} = \frac{2\bar{Y}-1}{1-\bar{Y}}$$

3. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with probability density function parameterized by θ given by

$$f_\theta(y) = \begin{cases} (\theta+1)y^\theta & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > -1$ is the parameter of interest. Find the maximum likelihood estimator (MLE) for θ. Compare this to your answer from the previous problem.

First we find the likelihood function. For this density function, we have:

$$L(Y_1, \ldots, Y_n | \theta) = (\theta + 1)Y_1^\theta \cdots (\theta + 1)Y_n^\theta$$
$$= (\theta + 1)^n (Y_1 \cdots Y_n)^\theta$$

We want to find the value of θ which maximizes this. To do so, we can take the derivative with respect to θ and set it equal to 0. Since we have a product of things involving θ, it is easier to take the log to turn the product into a sum and to then maximize the log likelihood function. This is the same thing we did for the geometric distribution. First we find the log likelihood function.

$$\log L(Y_1, \ldots, Y_n | \theta) = \log \left[(\theta + 1)^n (Y_1 \cdots Y_n)^\theta \right]$$
$$= n \log(\theta + 1) + \theta \log(Y_1 \cdots Y_n)$$
$$= n \log(\theta + 1) + \theta \sum_{i=1}^{n} \log(Y_i)$$

To find the value of θ which maximizes this, we take the derivative with respect to θ and set it equal to 0.

$$\frac{d}{d\theta} \log L(Y_1, \ldots, Y_n | \theta) = \frac{d}{d\theta} \left[n \log(\theta + 1) + \theta \sum_{i=1}^{n} \log(Y_i) \right]$$
$$= \frac{n}{\theta + 1} + \sum_{i=1}^{n} \log(Y_i)$$

Setting this equal to 0:

$$\frac{n}{\theta + 1} = -\sum_{i=1}^{n} \log(Y_i)$$
$$\theta + 1 = -\frac{n}{\sum_{i=1}^{n} \log(Y_i)}$$
$$\theta = -1 - \frac{n}{\sum_{i=1}^{n} \log(Y_i)}$$

4. Once again, you are the quality control manager for the Acme Widget Company. Your line of MiniWidgets has been so successful that the MiniWidget machines are running nonstop to satisfy the high customer demand. For a properly functioning MiniWidget machine, the probability of producing a defective MiniWidget is 1% (or less). As part of the quality control process, you will take a sample of 100 MiniWidgets from a machine to determine whether it needs repair. To make a statistically-sound decision, you design a hypothesis test to aid you in this process. You desire a level of $\alpha = 0.05$ for your hypothesis test.

(a) State the alternative hypothesis, null hypothesis, and test statistic for your hypothesis test.

We are interested in the proportion of defective widgets produced by the machine. Alternative hypothesis: $p > 0.01$. Null hypothesis: $p = 0.01$. Test statistic $\hat{p} = Y/n$, where $n = 100$ and Y is the number of defective widgets in your sample of 100.

(b) You sample 100 MiniWidgets from one of your machines and find that 3 of them are defective. At the level of $\alpha = 0.05$, does this machine need to be repaired?

To compute the rejection region, we need to find the standard deviation of our test statistic $\hat{p}$. Since we don't know the true population proportion p, we will estimate it with the sample proportion $\hat{p}$ since the sample size is large. In this case $\hat{p} = 3/100 = 0.03$

$$
\begin{aligned}
\sigma_{\hat{p}} &= \sqrt{\frac{p(1-p)}{n}} \\
&\approx \sqrt{\frac{\hat{p}(1-h\hat{p})}{n}} \\
&= \sqrt{\frac{(0.03)(0.97)}{100}} \\
&= 0.017
\end{aligned}
$$

The rejection region is given by:

$$
\begin{aligned}
\hat{p} &= p_0 + z_\alpha \sigma_{\hat{p}} \\
&= 0.01 + 1.65(0.017) \\
&= 0.038
\end{aligned}
$$

Since $\hat{p}$ does not fall inside the rejection region, we do not reject the null hypothesis, thus we conclude with level $\alpha = 0.05$ that the machine does not need to be repaired.

(c) What is the p-value for this hypothesis test?

The p-value is the smallest value of α for which we will reject the null hypothesis.

$$
\begin{aligned}
p - \text{value} &= \mathbb{P}(\hat{p} \geq 0.03 \text{ given } p = p_0 = 0.01) \\
&= \mathbb{P}\left(\frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \geq \frac{0.03 - p_0}{\sigma_{\hat{p}}} \right) \\
&= \mathbb{P}\left(Z \geq \frac{0.03 - 0.01}{0.017} \right) \\
&= \mathbb{P}(Z \geq 1.18) \\
&= 0.1190
\end{aligned}
$$

In this case, since we do not reject the null hypothesis at a level $\alpha = 0.05$, the p-values is greater than this value of α.

5. A random sample of 500 measurements on the length of stay in hospitals had a sample mean of 5.4 days and a sample standard deviation of 3.1 days. A federal regulatory agency hypothesizes that the average length of stay is greater than 5 days.

   (a) Do the data support this hypothesis with a level of $\alpha = 0.05$?

   The null hypothesis is $\mu = 5$, alternative hypothesis is $\mu > 5$, and test statistic is $\bar{Y}$. For the rejection region, we will approximate the population standard deviation σ with the sample standard deviation S since the sample size is large. The rejection region is given by:

   $$\bar{Y} \geq \mu_0 + z_\alpha \frac{S}{\sqrt{n}}$$
   $$= 5 + 1.65 \frac{3.1}{\sqrt{500}}$$
   $$= 5.23$$

   Since the rejection region is $\bar{Y} \geq 5.23$, our value of $\bar{Y}$ falls within the rejection region, thus we reject the null hypothesis, and so the claim is supported by the data at a level of $\alpha = 0.05$.

   (b) What is the p-value for this hypothesis test?

   For the p-value:

   $$p - \text{value} = \mathbb{P}(\bar{Y} \geq 5.4 \text{ given } \mu = \mu_0 = 5)$$
   $$= \mathbb{P}\left(\frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \geq \frac{5.4 - \mu_0}{S/\sqrt{n}}\right)$$
   $$= \mathbb{P}\left(Z \geq \frac{5.4 - 5.0}{3.1/\sqrt{500}}\right)$$
   $$= \mathbb{P}(Z \geq 2.89)$$
   $$= 0.0019$$

   (c) Using the rejection region found in the previous part, calculate β for the specific value of the alternative hypothesis $\mu_a = 5.5$.

$$\beta = \mathbb{P}\left(Z \leq \frac{k - \mu_a}{S/\sqrt{n}}\right)$$
$$= \mathbb{P}\left(Z \leq \frac{5.23 - 5.5}{3.1/\sqrt{500}}\right)$$
$$= \mathbb{P}(Z \leq -0.41)$$
$$= \mathbb{P}(Z \leq -1.95)$$
$$= 0.0256$$

(d) How large should the sample size be if we require that $\alpha = 0.01$ and $\beta = 0.05$, where we use the specific value of the alternative hypothesis $\mu_a = 5.5$.

Using the formula from the notes, approximating σ^2 with the sample variance S^2, we get

$$n = \frac{(z_\alpha + z_\beta)^2 S^2}{(\mu_a - \mu_0)^2}$$
$$= \frac{(2.33 + 1.65)^2 3.1^2}{(5.5 - 5.0)^2}$$
$$= 608.905$$

Rounding up, we should have a sample size of 609.