Stat 201B: P-Set 4

Graham Northrup

October 6, 2019

1)

First the MLE

$$L(\theta|X) = \prod_{i=1}^{n} \frac{1}{\theta}$$
$$log(L) = -\sum_{i=1}^{n} log(\theta)$$
$$\frac{dlog(L)}{d\theta} = -\frac{n}{\theta}$$

Now intuitively the likelihood of any θ is 0 if $\theta < x_{(n)}$, and for any value greater than $x_{(n)}$ the likelihood is decreasing as seen above. So the log likelihood is maximized for $\hat{\theta} = x_{(n)}$. Now we found the MSE of this estimator in homework 2 question 1b to be as follows

MSE of MLE The bias of $\hat{\theta_n}$ is

$$\begin{aligned}
\mathbb{E}[\hat{\theta_n}] - \theta &= \int_0^\theta n(\frac{x}{\theta})^n dx - \theta \\
&= \frac{n}{n+1}\theta - \theta \\
bias &= \frac{-\theta}{n+1}
\end{aligned}$$

The variance

$$\begin{aligned}
Var(\hat{\theta_n}) &= \mathbb{E}[\hat{\theta_n}^2] - (\frac{n-1}{n}\theta)^2 \\
&= \int_0^\theta n\frac{x^{n+1}}{\theta^n}dx - (\frac{n-1}{n}\theta)^2 \\
&= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 \\
&= \frac{n\theta^2}{(n+2)(n+1)^2}
\end{aligned}$$

The MSE is simply the variance plus the bias squared so we get

$$\begin{aligned}
MSE(\hat{\theta_n}) &= \frac{n\theta^2}{(n+2)(n+1)^2} + (\frac{\theta}{n+1})^2 \\
&= \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{(n+2)\theta^2}{(n+1)^2} \\
&= \frac{2(n+1)\theta^2}{(n+2)(n+1)^2}
\end{aligned}$$

Next we find the MOM estimator

$$\mathbb{E}_\theta[X] = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\frac{\hat{\theta}}{2} = \bar{X_n}$$
$$\hat{\theta} = 2\bar{X_n}$$

Which has MSE calcualted like we did in HW 2 question 1c

MSE of MOM: First the bias

$$\begin{aligned} bias &= \mathbb{E}[2\bar{X_n}] - \theta \\ &= 2\mathbb{E}[X_1] - \theta \\ &= \theta - \theta = 0 \end{aligned}$$

Now the Var

$$\begin{aligned} Var(2\bar{X_n}) &= 4Var(\bar{X_n}) \\ &= 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n} \end{aligned}$$

And the MSE

$$MSE = \frac{\theta^2}{3n} + 0^2 = \frac{\theta^2}{3n}$$

2)

To get the total likelihood we multiply together the likelihood for each Y_i

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(Y_i - (\beta_0 + \beta_1 X_i))^2}{2\sigma^2})$$

Now we can take the log of this to get a sum, and break it up into parts resulting in

$$\begin{aligned} log(L(\beta_0, \beta_1, \sigma^2)) &= \sum_{i=1}^{n} log(\frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(Y_i - (\beta_0 + \beta_1 X_i))^2}{2\sigma^2})) \\ &= -\frac{n}{2}log(2\pi) - nlog(\sigma) - \sum_{i=1}^{n} \frac{(Y_i - (\beta_0 + \beta_1 X_i))^2}{2\sigma^2} \\ &= -\frac{n}{2}log(2\pi) - nlog(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_i))^2 \end{aligned}$$

Now to get the maximum likelihood estimate for each of the parameters, we take the derivative of each and set equal to 0. First for β_0

$$0 = \frac{1}{\sigma^2}\sum_{i=1}^{n} Y_i - (\beta_0 + \beta_1 X_i)$$
$$0 = \sum_{i=1}^{n} Y_i - (\beta_0 + \beta_1 X_1)$$
$$0 = \frac{1}{n}\sum_{i=1}^{n} Y_i - (\beta_0 + \beta_1 X_1)$$
$$\hat{\beta_0} = \bar{Y} + \hat{\beta_1}\bar{X}$$

Now for β_1

$$\begin{aligned}
0 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_i))X_i \\
0 &= \frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \hat{\beta}_0 X_1 - \beta_1 X_i^2 \\
0 &= \bar{XY} - \hat{\beta}_0\bar{X} - \hat{\beta}_1\bar{X^2} \\
0 &= \bar{XY} - \bar{X}\bar{Y} + \hat{\beta}_1\bar{X}^2 - \hat{\beta}_1\bar{X^2}
\end{aligned}$$

$$\hat{\beta}_1 = \frac{\bar{XY} - \bar{X}\bar{Y}}{\bar{X^2} - \bar{X}^2}$$

Which we recognize as the sample covariance of X and Y over the sample covariance of X. And lastly we find this for σ^2

$$\begin{aligned}
0 &= \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_i))^2 \\
n\sigma^2 &= \sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_i))^2 \\
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_i))^2
\end{aligned}$$

Which we recognize as the MSE.

3)

As we showed above, the MLE is simply $\hat{\theta} = x_{(n)}$. Now to show it is consistent we need to show as n gets large $X_{(n)}$ converges in probability to θ.

$$\begin{aligned}
\mathcal{P}(|x_{(n)} - \theta| > \epsilon) &= \mathcal{P}(x_{(n)} > \theta + \epsilon) + \mathcal{P}(x_{(n)} < \theta - epsilon) \\
&= (\frac{\theta - \epsilon}{\theta})^n
\end{aligned}$$

Which is true because each x_i is indepedent and for the max to be less than a number, each individual x_i must also be less than that number. Then we get a probability that goes to 0 and n gets large, hence the MLE is consistent.

4)

We know the Fisher information is $I(\mu) = -\mathbb{E}[logL''(x;\mu)]$ so we start with the log likelihood

$$\begin{aligned}
logL(X;\mu) &= -\frac{n}{2}log(2\pi) - \sum_{i=1}^{n}\frac{(x_i - \mu)^2}{2} \\
\frac{dlogL(X;\mu)}{d\mu} &= \sum_{i=1}^{n} x_i - \mu \\
\frac{d^2logL(X;\mu)}{d\mu^2} &= -n \\
I(\mu) &= -\mathbb{E}[\frac{d^2logL(X;\mu)}{d\mu^2}] = n
\end{aligned}$$

The important thing to note here is that the first information does not depend on the value of μ, meaning we get the same amount of information regardless of the location of the distribution. The thing that gives us more information is more observations from the distribution. As n increases the Fisher information increases as well. This makes sense as μ is simply a location parameter, and does not change the scale of the normal distribution.

5)

Note: Debdeep Pati's notes on MLE estimators was consulted for part a of this problem

a)

The likelihood function is

$$L(\beta) = \prod_{i=1}^{n} \frac{x_i^{\alpha-1} e^{x_i \beta}}{\beta^\alpha \Gamma(\alpha)}$$

so we get

$$\begin{aligned}
log(L) &= \sum_{i=1}^{n} log(\frac{x_i^{\alpha-1} e^{x_i \beta}}{\beta^\alpha \Gamma(\alpha)}) \\
&= (\alpha - 1) \sum log(x_i) - \frac{\sum x_i}{\beta} - n\alpha log(\beta) - nlog(\Gamma(\alpha)) \\
\frac{dlog(L)}{d\beta} &= \frac{\sum x_i}{\beta^2} - \frac{n\alpha}{\beta} = 0 \\
\sum x_i - n\alpha\beta &= 0 \\
\hat{\beta} &= \frac{\sum x_i}{n\alpha} \\
\hat{\beta} &= \frac{\bar{X_n}}{\alpha}
\end{aligned}$$

b)

First I take the second derivative

$$\begin{aligned}
\frac{d^2 log(L)}{d\beta^2} &= -2\frac{\sum x_i}{\beta^3} + \frac{n\alpha}{\beta^2} \\
I(\beta) = \mathbb{E}[\frac{d^2 log(L)}{d\beta^2}] &= -\frac{2n\alpha}{\beta^4} + \frac{n\alpha}{\beta^2} \\
&= \frac{n\alpha}{\beta^2}(1 - \frac{2}{\beta^2})
\end{aligned}$$

Which gives us the confidence interval

$$\frac{\bar{X_n}}{\alpha} \pm \frac{1.96}{\frac{n\alpha}{\beta^2}(1 - \frac{2}{\beta^2})}$$

c)

When we augment the sample code to work for our data and a gamma distribution we get the MLE

$$\hat{\alpha} = 6.204123823$$
$$\hat{\beta} = 0.004650913$$

And we get the following confidence intervals

$$CI(\alpha) = [4.799790126, 7.608457520]$$
$$CI(\beta) = [0.003608113, 0.005693713]$$

To understand whether or not we have reached a global extrema rather than a local one, we can examine the Hessian and see that the determinants of the principal sub matrices are all positive. Meaning the Hessian is positive definite and we have a convex function, guaranteeing we have reached the global maximum.