Stat 135 Spring 2018 Quiz 2 Solutions

Name: Solutions
SID: Solutions
Section Number: Solutions

Problem 1:

Suppose we observe *fixed* (x_i, y_i) for $i = 1, \dots, n$ and suppose that we know that

$$Y_i = ax_i + b + \epsilon_i$$

Where a and b are unknown parameters to be estimated, ϵ_i are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables (also independent of x and y), and y_i is some observation of Y_i.

Recall that if $Z \sim \mathcal{N}(\mu, \sigma^2)$, then $f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$.

(a) Find the distribution of Y_i and write out the likelihood and log likelihood functions of the y_i's. **(2 points)**
Solution: From the form of Y, we have

$$Y_i \sim \mathcal{N}(ax_i + b, \sigma^2)$$

The likelihood function is

$$\mathcal{L}(a, b \,|\, y_1, \dots, y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(y_i - (ax_i + b))^2}{2\sigma^2}}$$

The log likelihood follows

$$\ell(a, b \,|\, y_1, \dots, y_n) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(y_i - (ax_i + b))^2}{2\sigma^2}$$

(b) Solve for the maximum likelihood estimates of a and b using your solution to (a). **(3 points)**
Solution:

$$0 = \frac{\partial \ell}{\partial a} = \sum_{i=1}^{n} \frac{1}{\sigma^2} x_i (y_i - (ax_i + b)) \Rightarrow 0 = \sum_{i=1}^{n} \frac{1}{n} x_i (y_i - ax_i - b)$$

$$0 = \frac{\partial \ell}{\partial b} = \sum_{i=1}^{n} \frac{1}{\sigma^2} (y_i - ax_i - b) \Rightarrow 0 = \sum_{i=1}^{n} \frac{1}{n} (y_i - ax_i - b)$$

We can solve for these

$$\hat{a}_{MLE} = \frac{\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)\left(\frac{1}{n} \sum_{i=1}^{n} y_i\right)}{\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{b}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{a}_{MLE} \frac{1}{n} \sum_{i=1}^{n} x_i$$

(c) Using your solution to (b), replace y_i with Y_i and compute the expectation of $\hat{a}_{MLE}$ assuming b is known and likewise compute the expectation of $\hat{b}_{MLE}$ assuming a is known. **(3 points)**
Solution: Both $\hat{a}$ and $\hat{b}$ are unbiased:

$$\mathbb{E}[\hat{b}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - a\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b + \mathbb{E}[\epsilon_i]) - a\frac{1}{n}\sum_{i=1}^{n} x_i = b$$

First, rewrite $\hat{a}$:

$$\hat{a} = \frac{-b + \frac{1}{n}\sum_{i=1}^{n} Y_i}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

Then its expectation follows:

$$\mathbb{E}[\hat{a}] = \mathbb{E}\left[\frac{-b + \frac{1}{n}\sum_{i=1}^{n} Y_i}{\frac{1}{n}\sum_{i=1}^{n} x_i}\right] = \frac{-b + \frac{1}{n}\sum_{i=1}^{n}(ax_i + b + \mathbb{E}[\epsilon_i])}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{-b + a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + b + 0}{\frac{1}{n}\sum_{i=1}^{n} x_i} = a\frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} x_i} = a$$

(d) Using your solution to (a), compute the Fisher Information for $\hat{a}_{MLE}$ assuming b is known and likewise compute the Fisher Information of $\hat{b}_{MLE}$ assuming a is known. Using this and your solution to part (c), construct an approximate 95% confidence interval for a and b. For half credit, you may compute directly the variance of $\hat{a}$ and $\hat{b}$ with your solution to (b). **(4 points)**
Solution: The second derivatives are

$$\frac{\partial^2 \ell}{\partial a^2} = \sum_{i=1}^{n}\frac{-x_i^2}{\sigma^2} \Rightarrow I_n(a) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial a^2}\right] = \sum_{i=1}^{n}\frac{x_i^2}{\sigma^2}, \quad \frac{\partial^2 \ell}{\partial b^2} = \sum_{i=1}^{n}\frac{-1}{\sigma^2} = \frac{-n}{\sigma^2} \Rightarrow I_n(b) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial b^2}\right] = \frac{n}{\sigma^2}$$

Now the confidence intervals follow easily from the asymptotic normality of MLE's

$$\left(\hat{a} - 1.96 \times \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}}, \hat{a} + 1.96 \times \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}}\right), \quad \left(\hat{b} - 1.96 \times \sqrt{\frac{\sigma^2}{n}}, \hat{b} + 1.96 \times \sqrt{\frac{\sigma^2}{n}}\right)$$

Alternatively, the variances follow relatively trivially from $\hat{b}$ in (b):

$$\mathrm{Var}(\hat{b}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \hat{a}\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i) = \frac{\sigma^2}{n}$$

$$\mathrm{Var}(\hat{a}) = \mathrm{Var}\left(\frac{\frac{1}{n}\sum_{i=1}^{n} y_i - b}{\frac{1}{n}\sum_{i=1}^{n} x_i}\right) = \frac{n\sigma^2}{\left(\sum_{i=1}^{n} x_i\right)^2}$$

Three notes: (1) the variances above are not equivalent to the true OLS variances because we allowed for the assumption of the opposing parameter being known; (2) if you compute $\mathrm{Var}(\hat{a})$ using the $\hat{a}$ from part (b) rather than rearranging as done above, then you should get a variance that matches the true OLS variance, though this is more computational effort; (3) this problem was meant to introduce you to the concept of linear regression via the method of maximum likelihood, which requires only distributional assumptions are on the errors ϵ_i. Technically, you have seen how to do linear regression in Stat 134 with bivariate normals, but in that context you put distributional assumptions on X_i and Y_i as well. When doing so, you can essentially deduce the same $\hat{a}$ and $\hat{b}$ purely from conditional expectations and covariances. Later at the end of this course, we will remove *all* distributional assumptions and find that we can deduce the same coefficients and approximately the same standard errors as above via the method of least squares.

Problem 2:

Suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(\theta, 1)$ with $0 < \theta$ and θ unknown.

(a) Write out the likelihood function and find the sufficient statistic, $T(X)$. **(2 points)**
Solution:

$$\mathcal{L}(\theta \mid x_1, \dots, x_n) = \frac{1}{(1-\theta)^n} \mathbb{1}\{\theta \le X_{(1)}\}$$

By the Factorization Theorem, $h(X) = 1$ and $g_\theta(T(X)) = \frac{1}{(1-\theta)^n} \mathbb{1}\{\theta \le X_{(1)}\}$ so $T(X) = X_{(1)}$.

(b) Find the method of moments estimator for θ and compute its mean square error. **(2 points)**
Solution: The MoM estimator is as follows

$$\mathbb{E}[X] = \frac{\theta + 1}{2} = \overline{X} \Rightarrow \hat{\theta}_{MoM} = 2\overline{X} - 1$$

Its MSE is as follows

$$MSE(\hat{\theta}_{MoM}) = (\mathbb{E}[2\overline{X} - 1] - \theta)^2 + \text{Var}(2\overline{X} - 1) = (\theta - \theta)^2 + \frac{4}{n}\text{Var}(X_i) = \frac{4}{n}\frac{(1-\theta)^2}{12} = \frac{(1-\theta)^2}{3n}$$

(c) Using your answer to part (a) and (b), compute the mean square error of the following estimator

$$\tilde{\theta} = \mathbb{E}\left[\hat{\theta}_{MoM} \middle| T(X)\right]$$

Where $\hat{\theta}_{MoM}$ is your method of moments estimator for θ. Comment briefly on the mean square errors of $\hat{\theta}_{MoM}$ and $\tilde{\theta}$ for large n. For half credit, you may compute the MSE for $\tilde{\theta}$ for the case when $X_i \sim \text{Uniform}(0, \theta)$ and $\theta < \infty$ as from lecture and homework. **(4 points)**
You may find the following facts helpful:

- $\frac{\text{Uniform}(a,b) - a}{b-a} \overset{d}{=} \text{Uniform}(0,1)$
- If $X_i \sim \text{Uniform}(a, b)$, then $[X_i \mid X_{(1)} = c, X_i \ne X_{(1)}] \sim \text{Uniform}(c, b)$
- If $X_i \sim \text{Uniform}(0, 1)$, then $\min_{1 \le i \le n} X_i \sim \text{Beta}(1, n)$
- For $X \sim \text{Beta}(\alpha, \beta)$,

$$f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$$
$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta}$$
$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Solution: First, note that by the Tower Property,

$$\mathbb{E}[\tilde{\theta}] = \mathbb{E}\left[\mathbb{E}\left[\hat{\theta}_{MoM} \middle| T(X)\right]\right] = \mathbb{E}[\hat{\theta}_{MoM}] = \theta$$

Therefore $\tilde{\theta}$ is unbiased. Next, we can rewrite $\tilde{\theta}$ directly as a function of $T(X) = X_{(1)}$ as follows

$$\begin{aligned}
\tilde{\theta} &= \mathbb{E}\left[\hat{\theta}_{MoM}\middle| T(X)\right] \\
&= \mathbb{E}[2\overline{X} - 1 \,|\, X_{(1)}] \\
&= 2\mathbb{E}[X_1 \,|\, X_{(1)}] - 1 \\
&= 2\left(\mathbb{E}[X_1 \,|\, X_{(1)}, X_1 = X_{(1)}]\mathbb{P}(X_1 = X_{(1)}) + \mathbb{E}[X_1 \,|\, X_{(1)}, X_1 \neq X_{(1)}]\mathbb{P}(X_1 \neq X_{(1)})\right) - 1 \\
&= 2\left(X_{(1)}\frac{1}{n} + \frac{X_{(1)} + 1}{2}\frac{n-1}{n}\right) - 1 \\
&= \frac{n+1}{n}X_{(1)} - \frac{1}{n}
\end{aligned}$$

Note you could use this to compute $\mathbb{E}[\tilde{\theta}]$ as above, but there is no need. Hence, we have

$$\begin{aligned}
MSE(\tilde{\theta}) &= bias(\tilde{\theta})^2 + \mathrm{Var}(\tilde{\theta}) \\
&= (0)^2 + \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(X_{(1)}) \\
&= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(\mathrm{Beta}(1,n)(1-\theta) + \theta) \\
&= \left(\frac{n+1}{n}\right)^2 (1-\theta)^2 \mathrm{Var}(\mathrm{Beta}(1,n)) \\
&= \left(\frac{n+1}{n}\right)^2 (1-\theta)^2 \frac{n}{(n+1)^2(n+2)} \\
&= \frac{(1-\theta)^2}{n(n+2)}
\end{aligned}$$

Notice that the MSE for the MoM estimator decays on the order of n^{-1} while the MSE $\tilde{\theta}$ decay on the order of n^{-2}. Note further that

$$MSE(\hat{\theta}_{MLE}) = (1-\theta)^2\left(\frac{2}{(n+1)(n+2)}\right)$$

Hence, we have $MSE(\hat{\theta}_{MLE}) = O(n^{-2})$ as well. This shows us that via a simple Rao-Blackwellization (you didn't need to know this for this problem) of our MoM estimator, we have gained a order of convergence, giving efficiency on par to MLE, while maintaining finite sample unbiasedness. One could argue that for these reasons, $\tilde{\theta}$ is "better" than $\hat{\theta}_{MLE}$.