ARE 210: Nov. 10, 2021 Section Notes

Joel Ferguson

Final 2018: Q1 (various sub-parts)

Setup: Suppose $\{X_i\}_{i=1}^{\infty}$ is an iid sample with the density of X_i given by

$$p(x, \theta, \nu) = \frac{\theta \nu^{\theta}}{x^{\theta+1}} \mathbb{I}(x \geq \nu)$$

where $\nu > 0$ and $\theta > 0$.

- **b)** Find a two-dimensional sufficient statistic for (ν, θ)
- **c)** Assume θ is known. Calculate $\mathbb{E}[X]$ and use this to propose an unbiased estimator for ν, called $\hat{\nu}_1$. How does your answer change if $\theta = 1$?
- **d)** Assume θ is known. Propose another unbiased estimator for ν that has variance no larger than that of $\hat{\nu}_1$.
- **g)** Compute the MLE for ν and denote it by $\hat{\nu}_{MLE}$. Does the value of the MLE depend on whether θ is known?
- **h)** Derive the exact finite sample distribution of $\hat{\nu}_{MLE}$.
- **i)** Compute the expectation of $\hat{\nu}_{MLE}$. Is it unbiased for ν?
- **n)** Compute the MLE for θ.

Solution:
Strategy:
Questions that ask us to derive sufficient statistics are generally easiest to solve via the factorization theorem, so for part **b** we'll write down the sample likelihood and try to decompose it into a function of just data and a function that combines statistics of the sample and the parameters.

Part **c** is essentially a method of moments problem, we'll first calculate the expected value of a given X_i, construct a moment condition that identifies ν, and then apply the analogy principle. Questions regarding the construction of GMM estimators generally follow this flavor.

Part **d** essentially just asks us to apply the Rao-Blackwell theorem. Identifying this requires thinking about what we have so far, sufficient statistics and an unbiased estimator, and results we know that combine those two things.

Part **g** is an MLE problem similar to one we did in last section, it has an ill-behaved likelihood function and finding the MLE requires noting this. One way to realize it is to note that because the support depends on ν, the likelihood function will discontinuously jump down to 0 at the minimum observed value of x. It turns out that the minimum order statistic is indeed the MLE.

Part **h** is a bit similar to problems from the first half of the semester, we need to use the properties of an iid sample to find the distribution of the minimum order statistic by finding the probability that all observed values are greater than a given constant.

Part **i** is also similar to some problems from the first half in that it simply asks us to calculate an expectation given a density.

Part **n** is more like a straightforward MLE problem with a well-defined interior solution that we can find via standard optimization results.

Part b) Step 1: Write down the sample likelihood

As mentioned in the strategy, we're going to use the factorization theorem to find a set of sufficient statistics. As such, we're looking to factor

$$\mathcal{L}(x_1, ..., x_n; \theta, \nu) = g(T_1(x_1, ..., x_n), T_2(x_1, ..., x_n); \theta, \nu) h(x_1, ..., x_n)$$

Which requires first finding $\mathcal{L}(x_1, ..., x_n; \theta, \nu)$. Since our sample is iid, we can write

$$\begin{aligned}\mathcal{L}(x_1, ..., x_n; \theta, \nu) &= \prod_{i=1}^{n} \frac{\theta \nu^\theta}{x_i^{\theta+1}} \mathbb{I}(x_i \geq \nu) \\ &= \frac{(\theta \nu^\theta)^n}{(\prod\limits_{i=1}^{n} x_i)^{\theta+1}} \mathbb{I}(x_{(1)} \geq \nu)\end{aligned}$$

Where the indicator function is that $x_{(1)} \geq \nu$ because the product of the indicator functions in the density is only equal to 1 if all observations are greater than ν.

Part b) Step 2: Factor the sample likelihood

We now need to factor the likelihood function according to the factorization theorem. There are two sample statistics in our expression of the likelihood, $\prod\limits_{i=1}^{n} x_i$ and x_1. As such, we can write

$$\begin{aligned}T_1(x_1, ..., x_n) &= \prod_{i=1}^{n} x_i \\ T_2(x_1, ..., x_n) &= x_{(1)} \\ g(T_1(x_1, ..., x_n), T_2(x_1, ..., x_n); \theta, \nu) &= \frac{(\theta \nu^\theta)^n}{T_1(x_1, ..., x_n)^{\theta+1}} T_2(x_1, ..., x_n)\end{aligned}$$

Finally, setting $h(x_1, ..., x_n) = 1$, we have $\mathcal{L}(x_1, ..., x_n; \theta, \nu) = g(T_1(x_1, ..., x_n), T_2(x_1, ..., x_n); \theta, \nu)h(x_1, ..., x_n)$
Thus, our two dimensional sufficient statistic is $(\prod_{i=1}^{n} x_i, x_1)$.

Part c) Step 1: Calculate the expected value

The first step is to simply calculate the expected value of X_i using the density.

$$
\begin{aligned}
\mathbb{E}[X_i] &= \int_{-\infty}^{\infty} x \frac{\theta\nu^\theta}{x^{\theta+1}} \mathbb{I}(x \geq \nu) dx \\
&= \int_{\nu}^{\infty} \frac{\theta\nu^\theta}{x^\theta} dx = -\frac{\theta\nu^\theta}{\theta-1} x^{-(\theta-1)} \Big|_{\nu}^{\infty} \\
&= 0 - [-\frac{\theta\nu^{\theta-(\theta-1)}}{\theta-1}] = \frac{\theta\nu}{\theta-1}
\end{aligned}
$$

Part c) Step 2: Form a moment condition

Now that we know the expected value of X_i, we want to form a moment condition. Essentially, this involves finding an expression of the expected value that is equal to ν and subtracting ν so that we have something equal to 0. We know from the previous step that $\nu = \frac{\theta-1}{\theta}\mathbb{E}[X_i]$, so our moment condition is

$$\frac{\theta-1}{\theta}\mathbb{E}[X_i] - \nu = 0$$

Part c) Step 3: Apply the analogy principal

Since we don't know the population expectation, we need to apply the analogy principal to derive our estimator

$$\hat{\nu}_1 = \frac{\theta-1}{n\theta}\sum_{i=1}^{n} x_i$$

Part c) Step 4: What if $\theta = 1$?

If $\theta = 1$, then $\mathbb{E}[X_i]$ doesn't exist since

$$\int_{\nu}^{\infty} \frac{\nu}{x} = \nu \log(x) \Big|_{\nu}^{\infty}$$

Part d): Apply Rao-Blackwell This question is all about realizing that we now have a sufficient statistic and an unbiased estimator for ν, so we can apply the Rao-Blackwell theorem to say that

$$\hat{\nu}_2 = \mathbb{E}[\hat{\nu}_1 | X_{(1)} = x_{(1)}]$$

has weakly lower variance. Explicitly writing this estimator as a function of the data is difficult because it requires finding the distribution of $\hat{\nu}_1$ conditional on the sample minimum.

Part g): Consider the sample likelihood again

This problem is tricky because it requires carefully thinking about the sample likelihood. Note that if there were no indicator, the likelihood would be an increasing function of ν, suggesting we would want to set $\nu = \infty$. However, we know that the likelihood is 0 if any observation is greater than ν. As such, the value of ν that maximizes the likelihood of the sample is the sample minimum, ie $\hat{\nu}_{MLE} = x_{(1)}$.

Part h) Step 1: Find $F_X(x)$ We know how to find the distribution of order statistics with an iid sample, but this requires the distribution of the underlying random variables.

$$\mathbb{P}(X_i \leq x) = \int_{\nu}^{x} \frac{\theta \nu^\theta}{z^{\theta+1}} dz = -\nu^\theta z^{-\theta}\Big|_{\nu}^{x} = 1 - \left(\frac{\nu}{x}\right)^\theta$$

Part h) Step 2: Write down $\hat{\nu}_{MLE}$'s distribution

Since $\{X_i\}_{i=1}^n$ are iid, we know

$$\begin{aligned} F_{X_{(1)}}(x) = \mathbb{P}(X_{(1)} \leq x) &= 1 - \mathbb{P}(X_{(1)} > x) \\ &= 1 - \prod_{i=1}^{n} \mathbb{P}(X_i > x) \\ &= 1 - \prod_{i=1}^{n} \left(\frac{\nu}{x}\right)^\theta = 1 - \left(\frac{\nu}{x}\right)^{n\theta} \end{aligned}$$

Part i) Step 1: Derive the density of the MLE

From the previous question we know the distribution of the MLE, now we need to find its density.

$$\begin{aligned} f_{X_{(1)}}(x) &= \frac{\partial}{\partial x} 1 - \left(\frac{\nu}{x}\right)^{n\theta} \\ &= n\theta \nu^{n\theta} x^{-(n\theta+1)} \end{aligned}$$

Part i) Step 2: Calculate the expectation

Now that we have the density, we can calculate the expectation. Note that the domain of the minimum starts at ν

$$\begin{aligned}
\mathbb{E}[\hat{\nu}_{MLE}] &= \int_{\nu}^{\infty} xn\theta\nu^{n\theta}x^{-(n\theta+1)}dx \\
&= \int_{\nu}^{\infty} n\theta\nu^{n\theta}x^{-n\theta}dx \\
&= -\frac{n\theta}{n\theta-1}\nu^{n\theta}x^{-(n\theta-1)}\Big|_{\nu}^{\infty} \\
&= \frac{n\theta}{n\theta-1}\nu
\end{aligned}$$

So the MLE is biased.

Part n): Take the derivative of the log-likelihood function

We've already derived the likelihood function for part *b*. Since there's no funny business with the support for θ, we can take the FOC of the log-likelihood

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log(\mathcal{L}(x_1,...,x_n);\theta,\nu) &= \frac{\partial}{\partial\theta}[n\log(\theta) + n\theta\log(\nu) - (\theta+1)\sum_{i=1}^{n} log(x_i) + log(x_{(1)} \geq \nu)] \\
&= \frac{n}{\theta} + n\log(\nu) - \sum_{i=1}^{n}\log(x_i) = 0
\end{aligned}$$

We can now solve this for θ to get $\theta = \frac{n}{\sum_{i=1}^{n}\log(x_i) - n\log(\nu)}$. However, we can't use this because it depends on the unknown parameter ν. We need to think of this problem in terms of finding the MLE of (θ, ν). In part **g** we were able to find that the MLE for ν is $x_{(1)}$, no matter the value of θ. Plugging this in to the FOC gives us

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n}\log(\frac{x_i}{x_{(1)}})}$$