ARE 210 Final Exam Review Session

Joel Ferguson

2018 Final Exam Q2

Setup: Suppose that the sequence of scalar random variables $\{X_n\}_{n=1}^{\infty}$ converges in distribution to a random variable X. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of constants converging to a. Show that

$$\mathbb{P}(X_n \le a_n) \to \mathbb{P}(X \le a)$$

Solution:
Strategy: The trick to this problem is realizing that we're trying to show that a transformation of X_n converges in distribution to the same tranformation of X. In particular, note that if a sequence of random variables $\{Y_n\}_{n=1}^{\infty} \xrightarrow[n\to\infty]{d} Y$, then by definition $\mathbb{P}(Y_n \le c) \to \mathbb{P}(Y \le c)$. Defining $Y_n = X_n - a_n$, $Y = X - a$, and $c = 0$ yields the solution so long as $X_n - a_n \xrightarrow[n\to\infty]{d} X - a$, which is true by Slutsky's Lemma.

Step 1: Reframe the problem
As outlined in the strategy, it's helpful to think of this problem as showing convergence in distribution of a random variable. Defining $Y_n = X_n - a_n$ abd $Y = X - a$, we can rewrite the problem as trying to show

$$\mathbb{P}(Y_n \le 0) \to \mathbb{P}(Y \le 0)$$

Step 2: Apply Slutsky's Lemma We now know that if we can show $Y_n \xrightarrow[n\to\infty]{d} Y$ we have completed the proof. Since $a_n \to a$ and $X_n \xrightarrow[n\to\infty]{d} X$, we know $X_n - a_n \xrightarrow[n\to\infty]{d} X - a$ by Slutsky's Lemma, and so $Y_n \xrightarrow[n\to\infty]{d} Y$.

2018 Final Exam Q3, parts c-e

Setup: Suppose that $\{X_i\}_{i=1}^{\infty}$ are a sequence of iid random variables and $X_i \in (-1, 1)$ with density

$$f(x, \theta) = \frac{1}{2}(1 + \theta x)\mathbb{I}(x \in (-1, 1))$$

- **c:** Use $\mathbb{E}_\theta[X]$ to propose a method of moments estimator for θ.
- **d:** Use $\mathbb{E}_\theta[X^3]$ to propose a method of moments estimator for θ.
- **e:** Consider now estimating θ by setting up a GMM problem that combines both the moment conditions in parts **c** and **d**. Write down the sample *and* population objective function for the GMM problem for a given positive-definite 2×2 weight matrix W. What would be the optimal choice of the weighting matrix W that minimizes the asymptotic variance of the GMM estimator (you do not need to compute it explicitly).

Solution:
Strategy: Parts **c** and **d** are very similar in approach: we first calculate the relevant expectation, then solve for θ and replace the expectations with sample analogues to obtain our estimators. Mercifully, part **e** only involves writing down the minimization problems, rather than actually solving them,and then stating that the choice of W for the efficient GMM estimator is the inverse of the variance of the moment conditions.

Part c Step 1: Find $\mathbb{E}_\theta[X]$

To find a method of moments estimator for θ based on $\mathbb{E}_\theta[X]$, we first need to calculate $\mathbb{E}_\theta[X]$.

$$\begin{aligned}
\mathbb{E}_\theta[X] &= \int_{-\infty}^{\infty} x\frac{1}{2}(1+\theta x)\mathbb{I}(x \in (-1,1))dx \\
&= \int_{-1}^{1} x\frac{1}{2}(1+\theta x)dx \\
&= \frac{x^2}{4} + \frac{\theta x^3}{6}\Big|_{-1}^{1} \\
&= \frac{3+2\theta}{12} - \frac{3-2\theta}{12} = \frac{\theta}{3}
\end{aligned}$$

Part c Step 2: Solve for θ and apply the analogy principle

Now that we know $\mathbb{E}_\theta[X] = \theta/3$, so our moment condition is $\mathbb{E}_\theta[X] - \theta/3 = 0$. We now need to solve for θ, i.e. $\theta = 3\mathbb{E}_\theta[X]$. To make this an esimator, we apply the analogy principle and swap sample moments for population moments, in this case replacing the expectation with the sample mean

$$\hat{\theta} = \frac{3}{n}\sum_{i=1}^{n} X_i$$

Part d Step 1: Find $\mathbb{E}_\theta[X^3]$

Part **d** proceeds very similarly to part **c**. We first need to calculate $\mathbb{E}_\theta[X^3]$.

$$
\begin{aligned}
\mathbb{E}_\theta[X^3] &= \int_{-\infty}^{\infty} x^3 \frac{1}{2}(1+\theta x)\mathbb{I}(x \in (-1,1))dx \\
&= \int_{-1}^{1} x^3 \frac{1}{2}(1+\theta x)dx \\
&= \frac{x^4}{8} + \frac{\theta x^5}{10}\Big|_{-1}^{1} \\
&= \frac{5+4\theta}{40} - \frac{5-4\theta}{40} = \frac{\theta}{5}
\end{aligned}
$$

Part d Step 2: Solve for θ and apply the analogy principle

Again, we solve $\theta = 5\mathbb{E}_\theta[X^3]$ and apply the analogy principle, where now the relevant relevent moment is the third uncentered moment.

$$\hat{\theta} = \frac{5}{n}\sum_{i=1}^{n} X_i^3$$

Part e Step 1: Write down population minimization problem

Most of this problem just uses the definition of GMM. We first stack our two moment conditions into a vector

$$m(X;\theta) = \begin{pmatrix} X - \frac{\theta}{3} = 0 \\ X^3 - \frac{\theta}{5} = 0 \end{pmatrix}$$

such that

$$\mathbb{E}_\theta[m(X;\theta)] = \begin{pmatrix} \mathbb{E}_\theta[X] - \frac{\theta}{3} = 0 \\ \mathbb{E}_\theta[X^3] - \frac{\theta}{5} = 0 \end{pmatrix}$$

Now we can write the population minimization problem as

$$\theta = argmin_\theta \mathbb{E}_\theta[m(X;\theta)]^T W \mathbb{E}_\theta[m(X;\theta)]$$

Part e Step 2: Apply the analogy principle to get the sample minimization problem

To get the sample minimization problem, we simply apply the analogy principle

$$\hat{\theta}_{GMM} = argmin_\theta \left(\frac{1}{n}\sum_{i=1}^{N} m(X;\theta)\right)^T W \left(\frac{1}{n}\sum_{i=1}^{N} m(X;\theta)\right)$$

Part e Step 3: State the optimal choice of W

Since we don't need to actually calculate it, we only need to state that the optimal choice of W is the inverse of the variance of the moment conditions. More explicitly,

$$W_{Opt} = \mathbb{E}_\theta\left[(m(X;\theta) - \mathbb{E}_\theta[m(X;\theta)])^T(m(X;\theta) - \mathbb{E}_\theta[m(X;\theta)])\right]^{-1}$$

2018 Final Q4

Setup: Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample from a uniform $[0, \theta]$ distribution.

- **a:** Show that the maximal order statistic $X_{(n)}$ is sufficient for θ.
- **b:** Derive the density function for $X_{(n)}$.
- **c:** Show that the statistic $X_{(n)}$ is complete. **Hint:** Use the definition of completeness. You will also find Leibniz's rule helpful

$$\frac{d}{dz}\int_o^{h(z)} q(t)dt = q(h(z))h'(z)$$

- **d:** Find the UMVUE for θ
- **e:** What is the MLE for θ and is its variance larger or smaller than that of the UMVUE?

Solution:
Strategy: As we generally do to show sufficiency, we'll try to use the Factorization Theorem to show that $X_{(n)}$ is sufficient for θ in part **a**. Part **b** will leverage the i.i.d. sample to derive the distribution of $X_{(n)}$, as we've done for other order statistics. For part **c** we'll show that if $\mathbb{E}[g(\theta)X_{(n)}] = 0$ for all θ then $g(\theta) = 0$ for all theta by first writing the expectation in the form of an integral and then taking the derivative of both sides with respect to θ. In part **d**, because there isn't an obvious unbiased estimator we'll take the expectation of $X_{(n)}$ and then try to form an unbiased estimator based on it, which will be the UMVUE by the Lehmann-Scheffe Theorem. Finally, we'll find the MLE in part **e** by noting that the sample likelihood is positive and decreasing in θ, but is 0 for $\theta < x_{(n)}$, meaning that the MLE is $X_{(n)}$. Using the density derived in **b**, we'll find that the variance is smaller than that of the UMVUE, which is possible because the MLE is biased.

Part a Step 1: Write down the sample likelihood

As with almost any sufficiency problem, we'll use the Factorization Theorem and try to write the sample likelihood in the form $h(\mathbf{X})g(X_{(n)}, \theta)$. To do so, we first need the sample likelihood.

$$\begin{aligned}
\mathcal{L}(\mathbf{X}, \theta) &= \prod_{i=1}^n \frac{1}{\theta}\mathbb{I}(X_i \in [0, \theta]) \\
&= \frac{1}{n^\theta}\prod_{i=1}^n \mathbb{I}(X_i \in [0, \theta]) \\
&= \frac{1}{n^\theta}\mathbb{I}(X_{(n)} \leq \theta)
\end{aligned}$$

where the last line follows from the fact that the lower bound of the support does not depend on θ and $X_{(n)} \leq \theta$ implies $X_i \leq \theta$ for all i.

Part a Step 2: Fit into Factorization Theorem

As stated earlier, we'll try to write the likelihood in the form $h(\mathbf{X})g(X_{(n)}, \theta)$. In this case, we can simply set $g(X_{(n)}, \theta) = (X_{(n)}, \theta)$ and $h(mathbfX) = 1$. So $X_{(n)}$ is sufficient for θ.

Part b Step 1: Rewrite definition of CDF

As is typical with problems where we derive the density of an order statistic, we'll start by finding the CDF of $X_{(n)}$ by noting that $\mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x, ..., X_n \leq n)$ and leveraging the i.i.d. sample.

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x, ..., X_n \leq n) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \\ &= \prod_{i=1}^{n} \frac{x}{\theta} \mathbb{I}(x \leq \theta) \\ &= \left(\frac{x}{\theta}\right)^n \mathbb{I}(x \leq \theta) \end{aligned}$$

Part b Step 2: Take derivative

Now that we have $F_{X_{(n)}}(x)$, we simply need to take the derivative with respect to x to get $f_{X_{(n)}}(x)$

$$f_{X_{(n)}}(x) = n\frac{x^{n-1}}{\theta^n}\mathbb{I}(x \leq 0)$$

Part c Step 1: Write the definition of a complete statistic

Since the uniform distribution isn't part of the exponential family, we can't use a trick to quickly find a complete statistic. Per the hint, we'll start by assuming that for all values of θ and a fixed function $g(.)$ we have $\mathbb{E}[g(\theta)X_{(n)}] = 0$ and then try show that $g(\theta) = 0$. To do this, we need to rewrite the expectation as an integral

$$\forall\theta \; \mathbb{E}[g(\theta)X_{(n)}] = \int_0^\infty g(\theta)n\frac{x^{n-1}}{\theta^n}\mathbb{I}(x \leq 0)dx = 0$$

Part c Step 2: Show $g(\theta) = 0$

Per the second part of the hint, we'll probably want to apply Leibniz's rule, meaning we'll want to take a derivative of an integral. To get this integral in the form necessary for Leibniz's rule, we'll need to incorporate the indicator function into the limits of integration: $\forall\theta \; \int_0^\theta g(\theta)n\frac{x^{n-1}}{\theta^n}dx = 0$. Taking the derivative of both sides with respect to θ via Liebniz's rule yields

$$\forall\theta \; g(\theta)\frac{n}{\theta} = 0$$

Since $\theta \in \mathbb{R}_+$, we know $\frac{n}{\theta} > 0$, which implies $g(\theta) = 0$ and thus $X_{(n)}$ is complete.

Part d Step 1: Find $\mathbb{E}[X_{(n)}]$

In this problem we haven't yet found an unbiased estimator and there isn't an obvious candiate since $\mathbb{E}[X] \neq \theta$. We do, however, have a complete and sufficient statistic, which we know by the Lehmann-Scheffe Theorem is the UMVUE for it's mean. We also know that if we can make an estimator that only depends on the data through $X_{(n)}$ and is unbiased for θ we will have found the UMVUE. So a useful approach in this case is to find $\mathbb{E}[X_{(n)}]$ and see if we can multiply it by a scalar to make an unbiased estimator for θ. Using the density we derived in part **b**, we can find

$$\begin{aligned}
\mathbb{E}[X_{(n)}] &= \int_0^\infty xn\frac{x^{n-1}}{\theta^n}\mathbb{I}(x \leq 0)dx \\
&= \int_0^\theta n\left(\frac{x}{\theta}\right)^n dx \\
&= \frac{n}{n+1}\frac{x^{n+1}}{\theta^n}\Big|_0^\theta = \frac{n}{n+1}\theta
\end{aligned}$$

Part d Step 2: Find a constant to make an unbiased estimator

As stated above, if we can find a constant c to multiply $X_{(n)}$ by such that $c\mathbb{E}[X_{(n)}] = \theta$, we'll have found the UMVUE. Since, $\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$, we know $\frac{n+1}{n}\mathbb{E}[X_{(n)}] = \theta$, and thus is the UMVUE.

Part e Step 1: Find the MLE

We now need to find the MLE and its variance. As is typical with distributions where the support depends on a parameter, we can't do unconstrained maximization to find the MLE. Rather, we should note that since $\mathcal{L}(\mathbf{X}, \theta) = \frac{1}{n^\theta}\mathbb{I}(X_{(n)} \leq \theta)$ and $\theta > 0$, we know the likelihood is decreasing and strictly positive in θ, unless $\theta < X_{(n)}$, in which case the likelihood is 0. Thus the likelihood is maximized at $X_{(n)}$, making it the MLE.

Part e Step 2: Find the Variance of the MLE

We don't actually even need to find the variance of the MLE to know that it is less than that of the UMVUE. Since $\hat{\theta}_{UMVUE} = \frac{n+1}{n}\hat{\theta}_{MLE}$, we know $Var(\hat{\theta}_{UMVUE}) = \left(\frac{n+1}{n}\right)^2 Var(\hat{\theta}_{MLE})$, implying $Var(\hat{\theta}_{UMVUE}) > Var(\hat{\theta}_{MLE})$. To actually find the variance of the MLE, we need to find the uncentered second moment of $X_{(n)}$

$$
\begin{aligned}
\mathbb{E}[X_{(n)}^2] &= \int_0^\infty x^2 n \frac{x^{n-1}}{\theta^n}\mathbb{I}(x \leq 0)dx \\
&= \int_0^\theta n\left(\frac{x^{n+1}}{\theta^n}\right)dx \\
&= \frac{n}{n+2}\frac{x^{n+2}}{\theta^n}\Big|_0^\theta = \frac{n}{n+2}\theta^2
\end{aligned}
$$

So $Var(X_{(n)}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2$

2019 Final Exam Q3, Parts b and c

Setup: Suppose we observe an i.i.d. sample $\{Y_i, X_i\}_{i=1}^n$ where Y_i and X_i are both scalars and satisfy

$$Y_i = \Phi(X_i\beta) + \epsilon_i$$

The unobserved variable ϵ_i is independent of X_i and distributed $\mathcal{N}(0,1)$. Assume further that $0 < Var(X_i) < \infty$.

- **b:** Derive the asymptotic distribution of the NLLS estimator and compare it to the asymptotic distribution of the MLE (Note: You do not have to verify the regularity conditions).
- **c:** Compute the Cramer-Rao bound for the parameter β.

Solution:
Strategy: We can actually start part **b** by noting that the NLLS estimator and the MLE objective functions are the same, meaning their asymptotic distributions (and indeed their finite sample distributions) will be the same. To find the asymptotic variance, we'll use the formula for M-estimators. Computing the Cramer-Rao lower bound in part **c** proceeds by noting that since ϵ is a normal random variable, so it Y_i and then applying the formula.

Part b Step 1: Write down the objective function and compare to MLE
To begin, we'll start by writing down the objective function for NLLS. Since $\epsilon_i = Y_i - X_i\beta$, the population minimization problem is

$$min_\beta \mathbb{E}[(Y_i - \Phi(X_i\beta))^2]$$

Applying the analogy principle gives us the sample problem

$$\hat{\beta}_{NLLS} = argmin_\beta \frac{1}{n}\sum_{i=1}^n (Y_i - \Phi(X_i\beta))^2$$

We can make progress on the second part of the question by noting that since ϵ_i is distributed $\mathcal{N}(0,1)$, the MLE problem is

$$\begin{aligned}
\hat{\beta}_{MLE} &= argmax_\beta \prod_{i=1}^{n} \phi(Y_i - \Phi(X_i\beta)) \\
&= argmax_\beta \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(Y_i - \Phi(X_i\beta))^2) \\
&= argmax_\beta \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\sum_{i=1}^{n}(Y_i - \Phi(X_i\beta))^2) \\
&= argmax_\beta - \frac{1}{2}\sum_{i=1}^{n}(Y_i - \Phi(X_i\beta))^2 \\
&= argmin_\beta \frac{1}{n}\sum_{i=1}^{n}(Y_i - \Phi(X_i\beta))^2
\end{aligned}$$

So we know that the NLLS estimator and the MLE are equal, and thus have the same distributions. Now we're free to find the asymptotic variance of either, since we know they will be equal. To align with the provided solution, I'll solve it using the M-Estimator variance formula.

Part b Step 2: Write down the M-estimation variance formula and find its components

From Theorem 8 in Handout 6, we know that the asymptotic distribution of an M-estimator is given by

$$\sqrt{n}(\theta_n - \theta) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, [\mathbb{E}[H(W,\theta)]]^{-1}\,\Sigma\,[\mathbb{E}[H(W,\theta)]]^{-1}\right)$$

We now need to translate this into our problem. If the generic M-estimator is given by $\theta_n = argmax_b n^{-1}\sum_{i=1}^{n} q(W_i, b)$, then our $q(W_i, b)$ is $(Y_i - \Phi(W_i b))^2$. The score, $s(W_i, b) = \partial q(W_i, b)/\partial b$ is $2(Y_i - \Phi(W_i b))\phi(W_i b)W_i$, and the Hessian $H(W_i, b) = \partial^2 q(W_i, b)/\partial b^2$ is given by $2W_i[-\phi(W_i b)^2 W_i + (Y_i - \Phi(W_i b))\phi'(W_i b)W_i]$. Finally, $\Sigma = Var(s(W_i, b))$, which since $(Y_i - \Phi(W_i b))$ is distributed $\mathcal{N}(0,1)$ is given by $(2\phi(W_i b)W_i)^2$. So we have

$$\begin{aligned}
\sqrt{n}(\theta_n - \theta) &\xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, [\mathbb{E}[H(W,\theta)]]^{-1}\,\Sigma\,[\mathbb{E}[H(W,\theta)]]^{-1}\right) \\
&= \mathcal{N}\left(0, \mathbb{E}[2W_i[-\phi(W_i b)^2 W_i + (Y_i - \Phi(W_i b))\phi'(W_i b)W_i]]^{-2}(2\phi(W_i b)W_i)^2\right) \\
&= \mathcal{N}\left(0, \mathbb{E}[-2W_i^2\phi(W_i b)^2]^{-2}(2\phi(W_i b)W_i)^2\right) \\
&= \mathcal{N}\left(0, \frac{1}{4}[W_i\phi(W_i b)]^{-4}4(\phi(W_i b)W_i)^2\right) \\
&= \mathcal{N}\left(0, (\phi(W_i b)W_i)^{-2}\right)
\end{aligned}$$

Where the third line follows from the fact that $\mathbb{E}[Y_i - \Phi(W_i b)] = 0$. Plugging back in X_i for W_i and β for b gives use

$$\sqrt{n}(\hat{\beta}_{NLLS} - \beta) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, (\phi(X_\beta)X_i)^{-2}\right)$$

Part c Step 1: Write down the Cramer-Rao lower bound formula

The Cramer-Rao lower bound of a parameter $\psi(\beta)$ is given by $\frac{d\psi(\beta)}{d\beta}'\mathcal{I}(\beta)^{-1}\frac{d\psi(\beta)}{d\beta}$, where $\mathcal{I}(\beta) = -\mathbb{E}_\beta\left[\frac{\partial^2}{\partial\beta^2}\log p(x,\beta)\right]$. In our case, since we're interested in β itself, $\psi(\beta) = \beta$. The last thing we need to compute is $\mathbb{E}_\beta\left[\frac{\partial^2}{\partial\beta^2}\log p(x,\beta)\right]$. In our case, $p(x,\beta) = \phi(Y_i - \Phi(X_i\beta))$. So we can calculate

$$\begin{aligned}
\frac{\partial}{\partial\beta}\log p(x,\beta) &= \frac{\partial}{\partial\beta}\log\phi(Y_i - \Phi(X_i\beta)) \\
&= -\frac{\partial}{\partial\theta}\frac{1}{2}[(Y_i - \Phi(X_i\beta))^2 + \log(2\pi)] \\
&= (Y_i - \Phi(X_i\beta))\phi(X_i\beta)X_i \\
\frac{\partial^2}{\partial\beta^2}\log p(x,\beta) &= \frac{\partial^2}{\partial\beta^2}(Y_i - \Phi(X_i\beta))\phi(X_i\beta)X_i \\
&= X_i[-\phi(X_i\beta)^2 X_i + (Y_i - \Phi(X_i\beta))\phi'(X_i\beta)X_i] \\
-\mathbb{E}[\frac{\partial^2}{\partial\beta^2}\log p(x,\beta)] &= -\mathbb{E}[X_i[-\phi(X_i\beta)^2 X_i + (Y_i - \Phi(X_i\beta))\phi'(X_i\beta)X_i]] \\
&= (\phi(X_i\beta)X_i)^2
\end{aligned}$$

Again, using the fact that $\mathbb{E}[Y_i - \Phi(W_i b)] = 0$ to get the last line. Putting everything together, we see that the Cramer-Rao lower bound is $(\phi(X_i\beta)X_i)^{-2}$ and the NLLS estimator achieves it.