1 True/False

Omitted.

2 Conditional Expectations

Consider order statistics: $X_{(1)}, ..., X_{(n)}$.

Problem 2.1. What is $\mathbb{E}\left[X_1|X_{(1)}, ..., X_{(n)}\right]$?

Solution. Note that

$$\sum_{i=1}^{n} \mathbb{E}\left[X_i|X_{(1)}, ..., X_{(n)}\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i|X_{(1)}, ..., X_{(n)}\right] = \sum_{i=1}^{n} X_i$$

Therefore:

$$\mathbb{E}\left[X_1|X_{(1)}, ..., X_{(n)}\right] = \frac{1}{n}\sum_{i=1}^{n} X_{(i)}$$

■

Problem 2.2. What is $\mathbb{E}\left[I\left(X_1 \leq x\right)|X_{(1)}, ..., X_{(n)}\right]$?

Solution. We conjecture that

$$\mathbb{E}\left[I\left(X_1 \leq x\right)|X_{(1)}, ..., X_{(n)}\right] = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right)$$

To show that this is indeed the conditional expectation, recall the definition of a conditional expectation:

$$\mathbb{E}\left[\left(I\left(X_1 \leq x\right) - f\left(X_{(1)}, ..., X_{(n)}\right)\right) I\left\{X_{(1)}, ..., X_{(n)} \in \mathcal{B}\right\}\right]$$

for some Borel set $\mathcal{B}$.

- ▷ We can rewrite the above as:

$$\mathbb{E}_{X_{(i)}}\left[\mathbb{E}_{X_1}\left[\left(I\left(X_1 \leq x\right) - f\left(X_{(1)}, ..., X_{(n)}\right)\right)\right] I\left\{X_{(1)}, ..., X_{(n)} \in \mathcal{B}\right\}\right]$$

  which implies that we need to find f such that

$$\mathbb{E}_{X_1}\left[\left(I\left(X_1 \leq x\right) - f\left(X_{(1)}, ..., X_{(n)}\right)\right)\right] = 0$$

- ▷ The conjecture satisfies the above.

■

3 Conditional Expectations

Problem 3.1. Provide interpretation of a linear regression when $\mathbb{E}\left[Xu\right] = 0$.

Solution. One is

$$\arg\min \mathbb{E}\left[(Y - b_0 - b_1 X)^2\right]$$

and the other is

$$\arg\min \mathbb{E}\left[(\mathbb{E}\left[Y|X\right] - b_0 - b_1 X)^2\right]$$

■

4 Maximum Likelihood Estimation

Let X_i be an iid sequence of random variables with common pdf:

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

Here $c > 0$ is unknown and $\theta > 0$ is the unknown parameter of interest.

Problem 4.1. Show that MLE $\hat{\theta}_n$ is given by

$$\hat{\theta}_n = \frac{n}{\sum_i \log(X_i/c)}$$

Solution. The likelihood is constructed as:

$$\prod_{i=1}^n f(\theta|x_i) = \prod_{i=1}^n \frac{\theta c^\theta}{x^{\theta+1}}$$

so the loglikelihood is:

$$\begin{aligned} \ell_n(\theta) &= \sum_{i=1}^n \log\left(\theta c^\theta x^{-(\theta+1)}\right) \\ &= n\log\theta + \theta\sum_{i=1}^n \log c - (\theta+1)\sum_{i=1}^n \log x \end{aligned}$$

Taking the derivative with respect to θ:

$$[\theta] : \frac{n}{\theta} + \sum_{i=1}^n \log c = \sum_{i=1}^n \log x$$

Rearranging:

$$\hat{\theta}_n = \frac{n}{\sum_i \log(X_i/c)}$$

■

Problem 4.2. Show that MLE $\hat{\theta}_n$ is a consistent estimator of θ.

Solution. Since X_is are i.i.d., we have that

$$\frac{1}{n}\sum_i \log\left(\frac{X_i}{c}\right) \xrightarrow{p} \mathbb{E}[\log X_i] - \log c = \frac{1}{\theta}$$

where

$$\mathbb{E}[\log X_i/c] = \frac{1}{\theta}$$

Therefore, by CMT, we have that $\hat{\theta}_n$ is a consistent estimator of θ. ■

Problem 4.3. Describe a score test for $\theta = 1$.

Solution. You compute the derivative of the log likelihood and plug in $\tilde{\theta} = 1$:

$$\begin{aligned}\sqrt{n} D_\theta L_n\left(\tilde{\theta}\right) &= \sqrt{n}\left(1 - \frac{1}{n}\sum_{i=1}^{n}\log\frac{X_i}{c}\right)\\ &= \sqrt{n}\left(1 - \frac{1}{\hat{\theta}_n}\right)\end{aligned}$$

Using this, we have:

$$\sqrt{n}\left(1 - \frac{1}{\hat{\theta}_n}\right) V \sqrt{n}\left(1 - \frac{1}{\hat{\theta}_n}\right)' \xrightarrow{d} \chi_1^2$$

To find V, take the second derivative of the log density

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\log\theta + \theta\log c - (\theta+1)\log c\right] &= \frac{1}{\theta}\\ \frac{\partial^2}{\partial^2\theta}\left[\log\theta + \theta\log c - (\theta+1)\log c\right] &= -\frac{1}{\theta^2}\end{aligned}$$

which means

$$V = [-B]^{-1} = \theta^2|_{\tilde{\theta}=1} = 1$$

so we have

$$n\left(1 - \frac{1}{\hat{\theta}_n}\right)^2 \sim \chi_1^2$$

Therefore, construct the test function as:

$$\phi_n\left(X_1, ..., X_n\right) = I\left\{T_n \geq c_n\right\}, T_n = n\left(1 - \frac{1}{\hat{\theta}_n}\right)^2, c_n = \chi_{1-\alpha}^2(1)$$

■

Problem 4.4. Use Delta Method to derive the limiting distribution of $\tau_n\left(\hat{\theta}_n - \theta\right)$ for an appropriate choice of τ_n.

Solution. From CLT, we had:

$$\sqrt{n}\left(\frac{1}{n}\sum_i \log\left(\frac{X_i}{c}\right) - \frac{1}{\theta}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)$$

Applying the Delta Method for $f(x) = 1/x$, we have

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\right)$$

■

Problem 4.5. Show that the information matrix equality holds.

Solution. It is obvious from here:

$$\frac{\partial}{\partial\theta}\left[\log\theta + \theta\log c - (\theta+1)\log c\right] = \frac{1}{\theta}$$
$$\frac{\partial^2}{\partial^2\theta}\left[\log\theta + \theta\log c - (\theta+1)\log c\right] = -\frac{1}{\theta^2}$$

■

Problem 4.6. Describe a Wald test for $H_0 : \theta = 1$.

Solution. Recall that we had:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\right)$$

from the Delta Method. Therefore:

$$\frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\hat{\theta}_n^2}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

Consider a test

$$\phi_n\left(X_1, .., X_n\right) = I\left\{T_n \geq c_n\right\}, \quad T_n = \left|\frac{\sqrt{n}\left(\hat{\theta}_n - 1\right)}{\hat{\theta}_n}\right|, c_n = z_{1-\alpha/2}$$

■

Problem 4.7. Does the information matrix equality still hold if $\log X_i/c$ is normally distributed with mean μ_f and variance σ_f^2 ?

Solution. Now we have

$$\mathbb{E}\left[\log\frac{X_i}{c}\right] = \mu_f, \quad \mathrm{Var}\left[\log\frac{X_i}{c}\right] = \sigma_f^2$$

In this case, the limiting distribution is

$$\sqrt{n}\left(\hat{\theta}_n - \frac{1}{\mu_f}\right) \xrightarrow{d} \mathcal{N}\left(0, \mu_f^4\sigma_f^2\right)$$

Then we can explicitly show that the derivatives are not equal and conclude that the matrix equality does not hold. ■

5 Switching Regressions

You have a sample (Y_i, Z_i, D_i) in the usual switching regressions setup. $Y_{1,i}$ and $Y_{0,i}$ denote the potential outcomes.

For the first part, suppose that $Y_{1,i} - Y_{0,i}$ equals a constant c.

Problem 5.1. Is the slope estimator from OLS regression of Y_i on a constant and D_i yield a consistent estimator of c?

Solution. We consider the following regression specification:

$$Y_i = \alpha + \beta D_i + u$$

in which case the estimate would be

$$\beta = \frac{\text{Cov}\left(Y_i, D_i\right)}{\text{Var}\left[D_i\right]} = \frac{\mathbb{E}\left[Y_i D_i\right] - \mathbb{E}\left[Y_i\right]\mathbb{E}\left[D_i\right]}{\mathbb{E}\left[D_i^2\right] - \mathbb{E}\left[D_i\right]^2}$$

▷ Focusing on the numerator:

$$\begin{aligned}
\mathbb{E}\left[Y_i D_i\right] &= \mathbb{E}\left[Y_{1,i} | D_i = 1\right]\mathbb{E}\left[D_i\right] \\
\mathbb{E}\left[Y_i\right]\mathbb{E}\left[D_i\right] &= \left(\mathbb{E}\left[Y_{1,i} | D_i = 1\right] P\left(D_i = 1\right) + \mathbb{E}\left[Y_{0,i} | D_i = 0\right] P\left(D_i = 0\right)\right)\mathbb{E}\left[D_i\right] \\
&= \mathbb{E}\left[Y_{1,i} | D_i = 1\right]\mathbb{E}\left[D_i\right]^2 + \mathbb{E}\left[Y_{0,i} | D_i = 0\right]\mathbb{E}\left[D_i\right]\left(1 - \mathbb{E}\left[D_i\right]\right) \\
&= \mathbb{E}\left[D_i\right]\left(1 - \mathbb{E}\left[D_i\right]\right)\left[\mathbb{E}\left[Y_{1,i} | D_i = 1\right] - \mathbb{E}\left[Y_{0,i} | D_i = 0\right]\right]
\end{aligned}$$

▷ Focusing on the denominator:

$$\mathbb{E}\left[D_i^2\right] - \mathbb{E}\left[D_i\right]^2 = \mathbb{E}\left[D_i\right]\left(1 - \mathbb{E}\left[D_i\right]\right)$$

▷ Therefore:

$$\beta = \frac{\text{Cov}\left(Y_i, D_i\right)}{\text{Var}\left[D_i\right]} = \mathbb{E}\left[Y_{1,i} | D_i = 1\right] - \mathbb{E}\left[Y_{0,i} | D_i = 0\right] = c$$

■

Problem 5.2. Is the slope estimator from a TSLS regression of Y_i on a constant and D_i with Z_i as an instrument for D_i a consistent estimator of c?

Solution. First, we claim that Z is a valid instrument. (Argument omitted). Since Z is a valid instrument, IV produces a consistent estimate of β. ■

Now suppose that $Y_{1,i} - Y_{0,i}$ is not necessarily constant.

Problem 5.3. Express the limit in probability of the slope estimator from TSLS regression of Y_i on a constant and D_i with Z_i as an instrument for D_i in terms of a "local average treatment effect."

Solution. The Wald estimand is

$$\beta = \frac{\text{Cov}\left(Y_i, Z_i\right)}{\text{Cov}\left(D_i, Z_i\right)} = \frac{\mathbb{E}\left[Y_i | Z_i = 1\right] - \mathbb{E}\left[Y_i | Z_i = 0\right]}{\mathbb{E}\left[D_i | Z_i = 1\right] - \mathbb{E}\left[D_i | Z_i = 0\right]}$$

▷ Focusing on the numerator:

$$\begin{aligned}
\mathbb{E}\left[Y_i | Z_i = 1\right] &= \mathbb{E}\left[Y_{i,1} D_i + Y_{i,0}\left(1 - D_i\right) | Z_i = 1\right] \\
&= \mathbb{E}\left[Y_{i,1} D_{i,1} + Y_{i,0}\left(1 - D_{i,1}\right)\right] \\
\mathbb{E}\left[Y_i | Z_i = 0\right] &= \mathbb{E}\left[Y_{i,1} D_i + Y_{i,0}\left(1 - D_i\right) | Z_i = 0\right] \\
&= \mathbb{E}\left[Y_{i,1} D_{i,0} + Y_{i,0}\left(1 - D_{i,0}\right)\right]
\end{aligned}$$

so combining the expression yields:

$$\mathbb{E}\left[Y_i | Z_i = 1\right] - \mathbb{E}\left[Y_i | Z_i = 0\right] = \mathbb{E}\left[\left(Y_{i,1} - Y_{i,0}\right)\left(D_{i,1} - D_{i,0}\right)\right]$$

Using monotonicity:

$$= \mathbb{E}\left[\left(Y_{i,1} - Y_{i,0}\right) | D_{i,1} > D_{i,0}\right] P\left(D_{i,1} > D_{i,0}\right)$$

▷ Focusing on the denominator:

$$\begin{aligned}
\mathbb{E}\left[D_i | Z_i = 1\right] &= \mathbb{E}\left[D_{i,1} | Z_i = 1\right] = \mathbb{E}\left[D_{i,1}\right] \\
\mathbb{E}\left[D_i | Z_i = 0\right] &= \mathbb{E}\left[D_{i,0} | Z_i = 0\right] = \mathbb{E}\left[D_{i,0}\right]
\end{aligned}$$

Using monotonicity:

$$\mathbb{E}\left[D_i | Z_i = 1\right] - \mathbb{E}\left[D_i | Z_i = 0\right] = \mathbb{E}\left[D_{i,1} - D_{i,0} | D_{i,1} > D_{0,i}\right] P\left(D_{i,1} > D_{i,0}\right) = P\left(D_{i,1} > D_{i,0}\right)$$

▷ Combining the expressions:

$$\beta = \frac{\text{Cov}\left(Y_i, Z_i\right)}{\text{Cov}\left(D_i, Z_i\right)} = \mathbb{E}\left[\left(Y_{i,1} - Y_{i,0}\right) | D_{i,1} > D_{i,0}\right]$$

which is the LATE.

■

Problem 5.4. How does the quantity above relate to "average treatment effect on the treated"?

Solution. Note that

$$\begin{aligned}
\mathbb{E}\left[Y_{i,1} - Y_{i,0} | D_{i,1} > D_{i,0}\right] &= \mathbb{E}\left[Y_{i,1} - Y_{i,0} | D_{i,1} = 1\right] \\
&= \mathbb{E}\left[Y_{i,1} - Y_{i,0} | D_{i,1} = 1, Z = 1\right] \\
&= \mathbb{E}\left[Y_{i,1} - Y_{i,0} | D_i = 1\right]
\end{aligned}$$

■