ECON 5213: ADVANCED ECONOMETRICS

PROF. LE WANG

Problem Set #6

Part I: Analytical Questions

Question 2. [Conditional Distribution and Truncated Distribution]. Show the following theorems are indeed true:

(1) (**Theorem 19.1**) If a continuous variable Y has pdf $f(y)$ and L is a constant, then

$$f(y|Y > L) = \frac{f(y)}{1 - F(L)}$$

Hint: You can write $f(y \mid Y > L) = \frac{d}{dy}F(y \mid Y > L) = \frac{d}{dy}\Pr[Y \leq y \mid Y > L]$. Then, apply the definition of conditional distribution to show the results.

(2) (**Theorem 19.**2) Using the result in **Theorem 19.1** and the result on the density for truncated normal variables (in your slides), show the first two moments are indeed of the following form ONLY for the case of $Y > L$. If $Y \sim \mathcal{N}(\mu, \sigma^2)$ and L is a constant, then

$$\mathbb{E}[Y|\text{truncated sample}] = \mu + \sigma\lambda(\alpha)$$
$$Var[Y|\text{truncated sample}] = \sigma^2(1 - \sigma(\alpha))$$

where $\alpha = \frac{L-\mu}{\sigma}$, $\phi(\cdot)$ ($\Phi(\cdot)$) is the **standard** normal density (distribution) function and

$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \text{if truncated sample is defined as } Y > L$$
$$\lambda(\alpha) = -\frac{\phi(\alpha)}{\Phi(\alpha)} \quad \text{if truncated sample is defined as } Y < L$$

and $\sigma(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$. **Hint: It might also be helpful to look at the example in Greene on truncated uniform distribution.** And Intuitively discuss whether or not your results indeed make sense. For example, as suggested Kasra, this depends on the shape of the distribution.

Question 4. [Skorohod Representation]. In class, we showed that conditioning on X, for a continuously distributed random variable Y, there always exists $U \sim \mathcal{U}(0,1)$ (the standard uniform distribution) such that

$$Y = F_{U|X}^{-1}(u|x) = m(x,u)$$

holds almost surely. Now, show that

$$U \perp X$$

Note that you just need to show whether or not

$$\Pr[U \leq \tau \mid X] = \Pr[U \leq \tau]$$

Question 4. [Conditional and Unconditional Independence]. Show that if $Y(1) \perp D \mid Y(0)$ and $Y(0) \perp D$, then

$$Y(1), Y(0) \perp D$$

Use a different approach than the one discussed in class.

Question 5. [CEF and Independence].

(1) Show that $Y \perp X$ implies $\mathbb{E}[Y \mid X] = \mathbb{E}[Y]$
(2) Show that $Y \perp D \mid X$ implies $\mathbb{E}[Y \mid D, X] = \mathbb{E}[Y \mid X]$. You can make the assumption that everything is continuous if that makes it easier to prove.

Question 6. [More General Law of Iterated Expectations]. Show that if $\mathbb{E}[|y|] < \infty$, then for any random vectors x_1, x_2

$$\mathbb{E}\left[\mathbb{E}\left[y|x_1, x_2\right] \mid x_1\right] = \mathbb{E}\left[y \mid x_1\right]$$

If you have time, please also try the even more general case in the slides.

Question 7. [Decomposition Property]. As discussed in class, any random variable y can be decomposed as follows

$$y = \mathbb{E}[y|x] + \epsilon$$

where ϵ is a random variable satisfying the following conditions.

(1) $\mathbb{E}\left[\epsilon \mid x\right] = 0$, which implies $\mathbb{E}\left[\epsilon\right] = 0$
(2) $\mathbb{E}[h(x)\epsilon] = 0$,where $h(x)$ can be any arbitrary function of x.

Show that the second condition holds. Also, show the following is true.

$$Cov(x, y) = Cov(x, \mathbb{E}[y|x])$$

Part II: Applied Questions (Optional)

Do the following exercises in Stata. Turn in your answer (along with your do file, log file, and graphs). Set the seed at 123456 at the start of each problem.

Question 1. [Simulation and Theoretical Results of Truncated Distribution]. It is important to verify our theoretical results using Monte Carlo Simulations, at least to see whether they hold for some of the experimental designs. To guide you to perform such exercises, follow the following steps

(1) Simulate a sample of 10000 (if your computer cannot handle this large dataset, you can reduce the sample size to 1000) observations of a random variable normally distributed with mean $\mu = 1$ and variance $\sigma^2 = 4$.
(2) Set up the cut off point to be $L = 0$. And generate an indicator equal to 1 if $Y > 0$, zero otherwise.
(3) Use Stata's Statistical functions to calculate the theoretical values of $\phi(\alpha)$ (density function for standard normal) and $\Phi(\alpha)$ (CDF for standard normal), and then construct $\lambda(\alpha)$. I am not telling you what functions to use because I would like to see if you could find out such functions on your own.
(4) Obtain the theoretical value of $\mu + \sigma\lambda(\alpha)$, the derived formual for truncated mean.
(5) Calculate the truncated mean simply using Stata's `-sum()-` command.
(6) Are they similar or not? What if you increase the sample size to 1000000?
(7) Repeat this process for the truncated variance as well for the case $Y > 0$.