EKT-816: Midterm - Suggested Solutions

1. Consider the joint distribution

$$f_{XY}(x,y) \quad = \quad k \cdot x^y \tag{1}$$

over the domain $0 < x < 1$, $0 < y < 1$, for some $k > 0$.

(a) What value should k have for f to be a proper density?

Answer: We integrate and set the answer equal to one, then solve for k. It's easier to integrate with respect to x first:

$$\int_0^1 kx^y dx = k\left[\frac{x^{y+1}}{y+1}\right]_{x=0}^{x=1} = \frac{k}{y+1}$$

Then, we integrate with respect to y:

$$\int_0^1 \frac{k}{y+1} dy = [k\log(y+1)]_{y=0}^{y=1} = k\log(2)$$

Thus, $k = 1/\log 2 \approx 1.44$.

(b) Find the marginal densities of X and Y.

Hint: $x^y = \exp[y \cdot \log(x)]$**.**

To find the marginal density of X, we have to integrate out Y. Integrating with respect to y, we get

$$f_X(x) = \int_0^1 kx^y dy = \int_0^1 k\exp(y\log x)dy = k\left[\frac{1}{\log x}e^{y\log x}\right]_{y=0}^{y=1} = \frac{k}{\log x}(x-1)$$

We already have the ingredients to find the marginal density of Y:

$$f_Y(y) = \int_0^1 kx^y dx = \frac{k}{y+1}$$

(c) Find the mean of Y.

$$E[Y] = \int_0^1 yf_y(y)dy = \int_0^1 \frac{ky}{y+1}dy = k\int_0^1 [1-(1+y)^-1]dy = k\left[y - \log(y+1)\right]_0^1 = k[1-\log(2)] = k-1 \approx 0.443$$

(d) Find the conditional mean of Y, given X.

Answer: The conditional density of Y, given X, is

$$f_{Y|X}(y|X=x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{kx^y}{k(x-1)/\log x} = \frac{x^y\log(x)}{x-1}$$

So, integrating by parts, the conditional mean is

$$
\begin{aligned}
E[Y|X=x] &= \frac{\log(x)}{x-1}\int_0^1 yx^y dy \\
&= \frac{\log(x)}{x-1}\left\{\left[y\cdot\frac{x^y}{\log(x)}\right]_{y=0}^{y=1} - \int_0^1 \frac{x^y}{\log(x)}dy\right\} \\
&= \frac{\log(x)}{x-1}\left\{\frac{x}{\log(x)} - \frac{1}{\log(x)}\left[\frac{x^y}{\log(x)}\right]_{y=0}^{y=1}\right\} \\
&= \frac{1}{x-1}\left\{x - \left[\frac{x}{\log(x)} - \frac{1}{\log(x)}\right]\right\} \\
&= \frac{x}{x-1} - \frac{1}{\log(x)}
\end{aligned}
$$

[$4 \times 5 = 20$ points]

2. We will consider some simple aspects of the demand for education in this question, namely that education has financial payoffs but can be unpleasant to acquire. So, think of a population of students who differ along two dimensions: ability A, and their subjective disutility for studying (say, V). Measure these variables in logarithmic units (i.e. a student of ability $a > 0$ is approximately $a\%$ better than average, and similarly for the preference parameter v).

If a student of ability a completes a university degree, she earns a salary of $w_1 e^a$, where $w_1 > 0$ is the wage in the white-collar labor market. Suppose also that disutility of studying acts in such a way as to decrease the value of consumption by a factor of e^{-v}, which is smaller than one for $v > 0$. Of course, some people may enjoy studying, if $v < 0$.

On the other hand, if a student does not go to university, they earn a salary of w_0, which is the wage in the blue-collar sector (and which is the same for everyone). Assume the joint distribution of (A, V) is bivariate normal, with mean zero. Use the notation σ_{AV} for the covariance between A and V, σ_A^2 for the variance of A, and similarly σ_V^2 for the variance of V.

(a) What is a reasonable decision rule for a (prospective) student facing this choice? You may assume she knows her own ability and preferences. Let $D = 1$ if a given student decides to go, and $D = 0$ if not.

Hint: you may want to define $\Delta = a - v$, and consider its distribution.

Answer: If someone of ability A and with preferences V chooses $D = 1$, they get a payoff of $w_1 e^{a-v}$. If they choose $D = 0$, they get w_0. So the decision rule should be: choose $D = 1$ if and only if $\Delta > \log(w_0/w_1)$.

(b) Given the decision rule you wrote down in (a) above, what fraction of the population will choose to go to university?

Answer: Given the joint distribution of (A, V), we can see the distribution of Δ is normal, with mean zero and variance

$$\sigma_\Delta^2 = \sigma_A^2 + \sigma_V^2 - 2\sigma_{AV}$$

(c) What is the joint distribution of (A, Δ)?

Answer: We know that linear combinations of normals are normal, and we have already computed the variance of Δ. So we just need to find the covariance between A and Δ, which is $\text{cov}(A, \Delta) = \sigma_A^2 - \sigma_{AV}$. The means of both A and Δ are zero.

(d) Express a as a linear function of Δ and an independent noise term U. What is the distribution of U?

Hint: look up the properties of the bivariate normal distribution, or use the hint from Problem Set 1, question 3.

Answer: Directly applying the hint, we have

$$A = \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta^2}\Delta + U$$

where we have already expressed σ_Δ^2 in terms of known quantities and U is a normal random variable with mean zero which is independent of Δ. So, to describe the distribution of U we just need to compute its variance. Take variances on both sides and solve to get that

$$\sigma_U^2 = \sigma_A^2 - \frac{(\sigma_A^2 - \sigma_{AV})^2}{\sigma_\Delta^2}$$

(e) Let $Y_1(A, V)$ be the potential (log) earnings of a graduate with ability A and preferences V, and similarly for Y_0. If you wanted to determine the causal effect of university education on earnings, would a comparison of the average earnings of graduates with non-graduates deliver a consistent estimate of either the ATE or the ATT? If not, would this naive comparison be biased upwards or downwards?

Hint: Do graduates have higher ability, on average, than non-graduates?

Another hint: look up the properties of the truncated normal distribution.

Answer: First, let's define the objects we want to compute. The ATT is

$$ATT = E[Y_1 - Y_0|D = 1]$$

while the ATE is $ATE = E[Y_1 - Y_0]$. What do we get out of the naive comparison of graduates with non-graduates? This is

$$E[Y_1|D = 1] - E[Y_0|D = 0].$$

From part (d), we have

$$A = \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta^2}\Delta + U$$

where U is independent of Δ. Now, consider the average (log) earnings of someone who chooses to go to university:

$$\begin{aligned}
E[Y_1|D = 1] &= E[\log w_1 + A|A - V > \log(w_0/w_1)] \\
&= \log w_1 + E\left[\frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta^2}\Delta + U|\Delta > \log(w_0/w_1)\right] \\
&= \log w_1 + \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta^2}E[\Delta|\Delta > \log(w_0/w_1)] + 0 \\
&= \log w_1 + \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta}E\left[\frac{\Delta}{\sigma_\Delta}|\frac{\Delta}{\sigma_\Delta} > \frac{\log(w_0/w_1)}{\sigma_\Delta}\right] \\
&= \log w_1 + \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta}\lambda\left(\frac{\log(w_0/w_1)}{\sigma_\Delta}\right)
\end{aligned}$$

where $\lambda(c) = E[Z|Z > c]$ is the mean of a truncated standard normal. You can verify (by integrating by parts) that

$$\lambda(c) = \int_c^\infty z\phi(c)dz = \frac{\phi(c)}{1 - \Phi(c)}$$

where $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ is the standard normal density and Φ is the CDF of a standard normal. Importantly, $\lambda(c) > 0$ for all c.

Now, what would be the average log earnings in the white-collar sector of a randomly selected person? Equivalently, what would be average earnings in the white-collar sector if people selected into it randomly (i.e. without knowledge or concern for their earnings)? This is $E[Y_1] = \log w_1$. So, we have

$$\begin{aligned}
ATT &= E[Y_1 - Y_0|D = 1] \\
&= \log w_1 + \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta}\lambda\left(\frac{\log(w_0/w_1)}{\sigma_\Delta}\right) - \log w_0 \\
&= \log w_1 - \log w_0 + \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta}\lambda\left(\frac{\log(w_0/w_1)}{\sigma_\Delta}\right)
\end{aligned}$$

and

$$ATE = E[Y_1 - Y_0] = \log w_1 - \log w_0$$

Now, what we would estimate by just comparing the earnings of graduates with non-graduates is

$$E[Y_1|D = 1] - E[Y_0|D = 0] = \log w_1 + \frac{(\sigma_A^2 - \sigma_{AV})}{\sigma_\Delta}\lambda\left(\frac{\log(w_0/w_1)}{\sigma_\Delta}\right) - \log w_0$$

which is not in general equal to the ATE.[1]

In fact, we can see that the sign of the bias term depends on the sign of $\text{cov}(A, \Delta) = \sigma_A^2 - \sigma_{AV}$. We are used to thinking about the "positive selection" case, where $\sigma_A^2 - \sigma_{AV} > 0$, so graduates have higher-than-average ability. But could this selection term be negative? Yes, it could: all we know is that $\sigma_A^2\sigma_V^2 \geq \sigma_{AV}^2$, from the Cauchy-Schwartz inequality (or, just the requirement that the variance-covariance matrix of (A, V) be positive definite).

Here is an interpretation of the negative case, where graduates actually have lower ability than a randomly selected person. We know people choose $D = 1$ if they have a high value of $\Delta = A - V$. This could mean they have high ability (and thus high returns from going to university) but it could also be that they have a low disutility, i.e. they like studying (or dislike it less than others).

Now, suppose the variance of ability is low, so attendance decisions are mostly driven by people's preferences. Then, it's possible that $\sigma_A^2 < \sigma_{AV}$ and $\sigma_V^2 > \sigma_{AV}^2/\sigma_A^2$, so the variance in preferences is relatively large. In this situation, when we observe someone choosing $D = 1$, this is mostly informative that they have a low enough V, and in fact this could mean they have a lower-than-average A.

(f) Now, consider a "merit scholarship" (or, equivalently, a wage subsidy), that depends on students' academic ability as follows: a student of ability a will have her wage augmented by a factor of e^{ka} for some $k > 0$ if she graduates, but if she does not, she still receives the blue-collar wage w_0.

Let Z be an indicator for the receipt of the scholarship, and define $D_1(A, V)$ to be an indicator for whether a student with ability A and preferences V would graduate when subject to the

[1] In this case, the naive difference in earnings *does* consistently estimate the ATT, although this is a bit of a coincidence due to the fact that everyone gets the same outcome in the blue-collar sector. If we had another skill that was exclusively useful in the $D = 0$ sector, we would not be able to estimate the ATT either.

scholarship. Similarly, let $D_0(A,V)$ be the outcome without the scholarship. Describe the joint distribution of (D_1, D_0).

Hint: you may find it helpful to draw a picture of the (A,V) plane and illustrate the different subsets of the population on it. Who would graduate whether the scholarship was given or not? Who would never graduate? There are four possibilities in total.

Answer: Such a policy would cause both inflows *and* outflows from the white-collar sector. It's helpful to draw a picture of the (V,A) plane here and illustrate *who* selects into college under each policy. For simplicity, suppose $\log(w_0/w_1) = 0$, so we are comparing the set $\mathcal{C}(k=0) = \{(v,a) : a - v > 0\}$ with the set $\mathcal{C}(k>0) = \{(v,a) : (1+k)a - v > 0\}$.

There are four possibilities: some people will go to college under both the $k=0$ and the $k>0$ policy; call them the "always-takers". These would be people who have high ability and a low disutility for studying (they may even enjoy it). Some people would never go to college, under either policy; these would be people with low ability a and a high disutility v. Call them the "never-takers".

Then there are some people with intermediate ability and disutility who can be induced by the subsidy to go to college. These are people for whom $a - v < 0 < (1+k)a - v$. Notice that these inequalities imply $a > 0$ and $v > 0$. We might think of these people as the "smart but lazy" ones. Call them the "compliers".

Finally, there are the people for whom $a - v > 0 > (1+k)a - v$. Who are they? These are people who are not very good at studying, but they enjoy it (notice that this group necessarily has $v < 0$ and $a < 0$). The imposition of the subsidy actually *lowers* their wage in the white-collar sector, because it exaggerates their low productivity levels. Since their outside option in the blue-collar sector remains unchanged, some of these students will choose to drop out. (We might think of these people as the "jocks".) In the language of instrumental variables, this set of people would be considered "defiers" with respect to the instrument defined by the subsidy, although when we study IV later we will see that such a subsidy is not a valid instrument because it changes the potential outcomes.

(g) If the scholarship was given to all students leaving high school (or to a random subset of them), how would this policy affect the number of graduates?

Answer: Without the scholarship, the fraction of the population that graduates is $Pr(D_0 = 1) = Pr(A - V > \log(w_0/w_1))$. With the policy, the graduation rate is $Pr(D_1 = 1) = Pr((1+k)A - V > \log(w_0/w_1))$.

[$7 \times 5 = 35$ points]

3. Let $X_1, X_2, \ldots X_n$ be a mutually independent sequence of exponentially distributed random variables, with $\lambda = 1$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ be the sample mean.

(a) What is $\text{plim}_{n\to\infty}\overline{X}_n$?

Answer: The law of large numbers tells us that $\text{plim}_{n\to\infty}\overline{X}_n = E[X]$. In this case, that's $E[X] = \lambda^{-1} = 1$.

(b) Use the delta method to find the asymptotic distribution of $Y_n = 1/\overline{X}_n$.

Answer: The central limit theorem tells us that

$$\frac{\overline{X}_n - E[X]}{\sqrt{V[X]/n}} \longrightarrow N(0,1)$$

[$5 + 10 = 15$ points]

4. Read Neal and Johnson (1996) and answer the following questions:

(a) What type of data (cross-sectional, longitudinal, synthetic panel, etc) do they use to establish their "basic result"?

(b) Explain their omitted variables argument. What evidence is there in their Table 1 that AFQT is an omitted variable in the basic wage regressions?

(c) Why do the authors *not* control for e.g. post-secondary schooling? Is this an instance of the "bad control" problem (see Ch. 3.2.3 of (**???**))? Which assumptions would justify including these controls, and which would not?

[$3 \times 10 = 30$ points]

References

Neal, Derek, and William Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (5): 869–95. https://www.jstor.org/stable/2138945.