Economics 696F: Lecture Note 12

Partial Identification and Bounds

This is based on the Manski articles; the book *Identification Problems in the Social Sciences* is also an excellent source for material on bounds.

Censored Data

Suppose that Y and X are variables of interest, but that Y is subject to *censoring*: we do not always observe Y. Formally, assume there is a population joint distribution for (Y, X, T), where T is a binary variable. We obtain a random sample from this population, where we always observe X and T, but we only observe Y if $T = 1$. If $T = 0$, we do not observe Y for that individual.

We will focus on identification. In large samples we can learn $P(X, T)$ and $P(Y|X, T = 1)$. Here, the notation $P(\cdot|\cdot)$ indicates the relevant condition or joint distribution, or conditional probability. The specific meaning will be clear from the arguments of the function.

Does this sampling process identify $P(Y|X)$ (and by extension $P(Y, X)$)? No: we can write

$$P(Y|X) = P(Y|X, T = 1)P(T = 1|X) + P(Y|X, T = 0)P(T = 0|X).$$

We cannot learn $P(Y|X, T = 0)$, so even with an arbitrarily large sample, we cannot identify $P(Y|X)$ without further assumptions.

Missing at Random: suppose we assume

$$P(Y|X, T = 1) = P(Y|X, T = 0).$$

Another way to put this is

$$Y \perp T|X.$$

Then

$$P(Y|X) = P(Y|X, T = 1)P(T = 1|X) + P(Y|X, T = 1)P(T = 0|X).$$

Every piece on the right hand side is identified, hence $P(Y|X)$ is identified.

On the other extreme, suppose we know *nothing* about $P(Y|X, T = 0)$. Suppose we want to estimate the conditional mean of Y given X:

$$E[Y|X] = E[Y|X, T = 1]P(T = 1|X) + E[Y|X, T = 0]P(T = 0|X).$$

If Y has unbounded range, then $E[Y|X, T = 0]$ could in principle take on any value between $-\infty$ and $+\infty$. Thus, $E[Y|X]$ could take on any value.

However, we can actually say a bit more about other quantities. For example, suppose we want to know the conditional probability that Y is in some set B:

$$P(Y \in B|X) = P(Y \in B|X, T = 1)P(T = 1|X) + P(Y \in B|X, T = 0)P(T = 0|X).$$

We don't know $P(Y \in B|X, T = 0)$, but since it is a probability, it must be between 0 and 1. By plugging in the extreme possibilities, we can say that

$$P(Y \in B|X, T = 1)P(T = 1|X) \leq P(Y \in B|X),$$

and

$$P(Y \in B|X) \leq P(Y \in B|X, T = 1)P(T = 1|X) + P(T = 0|X).$$

So $P(Y \in B|X)$ must lie in the interval

$$\Big[P(Y \in B|X, T = 1)P(T = 1|X), \quad P(Y \in B|X, T = 1)P(T = 1|X) + P(T = 0|X)\Big].$$

The width of this interval is $P(T = 0|X)$. If there is a lot of missing data, so $P(T = 0|X)$ is large, then the interval will be wide.

The quantity $P(Y \in B|X)$ cannot be determined by the distribution of the observable data, so it is not "point-identified." However, we can determine that it lies within a certain set of values. So we say that $P(Y \in B|X)$ is "partially identified." It turns out that the expression above provides the tightest bounds (the bounds are "sharp") given the available information.

There are various additional assumptions that may be reasonable to assume, which could be used to further narrow these bounds. One possibility is the following:

Exclusion restriction: suppose that X can be subdivided into two components $X = (W, V)$, and that $P(Y|W, V)$ does not depend on V. So $P(Y|W, V) = P(Y|W)$. Thus V is "excluded" from determination of Y. It is essentially an instrumental variable.

For simplicity, suppose that V can take on two values, v_1 and v_2. Then we must have

$$P(Y \in B|W, V = v_1, T = 1)P(T = 1|W, V = v_1) \leq P(Y \in B|W),$$

and

$$P(Y \in B|W, V = v_2, T = 1)P(T = 1|W, V = v_2) \leq P(Y \in B|W),$$

Likewise

$$P(Y \in B|W) \leq P(Y \in B|W, V = v_1, T = 1)P(T = 1|W = v_1) + P(T = 0|W, V = v_1).$$

$$P(Y \in B|W) \leq P(Y \in B|W, V = v_2, T = 1)P(T = 1|W = v_2) + P(T = 0|W, V = v_2).$$

So the lower bound on $P(Y \in B|W)$ is

$$\max_{j=1,2} P(Y \in B|W, V = v_j, T = 1)P(T = 1|W, V = v_j)$$

The upper bound is

$$\min_{j=1,2} P(Y \in B|W, V = v_j, T = 1)P(T = 1|W = v_j) + P(T = 0|W, V = v_j).$$

This narrows the bounds on $P(Y \in B|W)$.

Treatment Effects

The problem of estimating treatment effects is closely related to censored sampling problems. For individuals with $T = 1$ we observe $Y(1)$, but $Y(0)$ is censored. For individuals with $T = 0$, we observe $Y(0)$ but $Y(1)$ is censored.

Formally, there is a population distribution for $(Y(0), Y(1), X, T)$, where T is binary. We observe a random sample of (X, T, Y), where $Y = T \cdot Y(1) + (1 - T) \cdot Y(0)$.

In large samples, we can learn $P(X, T)$ (and therefore $P(T|X)$) along with

$$P(Y(1)|X, T = 1), \quad P(Y(0)|X, T = 0).$$

However, we do not learn $P(Y(1)|X, T = 0)$ or $P(Y(0)|X, T = 1)$.

Let

$$ATE(X) := E[Y(1)|X] - E[Y(0)|X].$$

We can write

$$E[Y(1)|X] = E[Y(1)|X, T = 1]P(T = 1|X) + E[Y(1)|X, T = 0]P(T = 0|X).$$

If $E[Y(1)|X, T = 0]$ can take values in the range $-\infty$ to $+\infty$, then we cannot say anything about $E[Y(1)|X]$ without further assumptions.

Similarly if $E[Y(0)|X, T = 1]$ has unbounded range, then we cannot say anything about $E[Y(0)|X]$. Thus, without further assumptions, $ATE(X)$ is not identified, and the bounds are infinitely wide.

Suppose instead that $Y(1)$ and $Y(0)$ are binary outcomes. Then it must be the case that

$$0 \le E[Y(1)|X, T = 0] \le 1$$

and

$$0 \le E[Y(0)|X, T = 1] \le 1.$$

Therefore, an upper bound on $ATE(X)$ is:

$$E[Y(1)|X, T = 1]P(T = 1|X) + P(T = 0|X) - E[Y(0)|X, T = 0]P(T = 0|X).$$

A lower bound on $ATE(X)$ is:

$$E[Y(1)|X, T = 1]P(T = 1|X) - E[Y(0)|X, T = 0]P(T = 0|X) - P(T = 1|X).$$

The resulting interval has width $P(T = 0|X) + P(T = 1|X) = 1$.

For Y with general range, if we are interested instead in $P(Y(1) \in B|X) - P(Y(0) \in B|X)$, essentially the same argument can be used. Again, the interval has width 1.

There are various assumptions that, if correct, can narrow the bounds on $ATE(X)$ and related quantities. In some cases, we can narrow the interval to a single point (so that $ATE(X)$ is "point-identified").

<u>Unconfounded Treatment Assignment</u>: if we assume that

$$T \perp Y(1), Y(0)|X,$$

Then

$$E[Y(1)|X, T = 1] = E[Y(1)|X, T = 0]$$

and

$$E[Y(0)|X, T = 1] = E[Y(0)|X, T = 0].$$

Then, $ATE(X)$ is point-identified, even if Y is unbounded.

<u>Other possible assumptions</u>:

1. Ordered outcomes: $Y(1) \geq Y(0)$ for all individuals.
2. Roy Model: individuals choose T leading to highest outcome:

$$Y(1) > Y(0) \Rightarrow T = 1,$$

$$Y(1) \leq Y(0) \Rightarrow T = 0.$$

The Mixing Problem

In this section, we'll drop the conditioning on X for simplicity. The same results hold if we condition throughout on X.

Recall that a randomized experiment identifies $P(Y(0))$ and $P(Y(1))$. That is, in large samples we can learn the <u>marginal distributions</u> of $Y(0)$ and $Y(1)$. So we know the outcome distributions under the policy of assigning everyone to $T = 0$, and the outcome distribution under the policy of assigning everyone to $T = 1$.

Suppose instead that we are interested in the outcome distribution under some other treatment assignment mechanism.

Example: we make the treatment available to everyone, but voluntary.

Example: we make the treatment available to some specific subgroup, and then case-workers (for a job training program) or doctors (for a medical intervention) decide whether to give the treatment.

Manski calls this the *mixing problem*. Formally, let m denote a treatment assignment mechanism, and let T^m denote the treatment chosen under m. So the outcome will be

$$Y^m = T^m Y(1) + (1 - T^m) Y(0).$$

If we actually apply treatment assignment mechanism m to a random sample from the population, then clearly we can learn the distribution of Y^m.

What if we only have the results from a randomized experiment? Assume we know $P(Y(0))$ and $P(Y(1))$. What does this tell us about $P(Y^m)$?

Let's focus on $P(Y^m \in B)$, the probablity that Y^m falls in some set B. It will turn out that we cannot point-identify this quantity, but we can obtain bounds on it.

First, notice that if both $Y(0)$ and $Y(1)$ are in B, then so is Y^m, regardless of T^m. Likewise, if neither $Y(0)$ nor $Y(1)$ is in B, then Y^m also is not in B. We can write this as:

$$Y(1) \in B \quad \cap \quad Y(0) \in B \quad \Rightarrow \quad Y^m \in B,$$
$$Y(1) \notin B \quad \cap Y(0) \notin B \quad \Rightarrow \quad Y^m \notin B.$$

Now suppose $Y(1) \notin B$ and $Y(0) \in B$. Then $P(Y^m \in B)$ is minimized by a rule that sets $T^m = 1$.

Likewise, if $Y(1) \in B$ and $Y(0) \notin B$, then $P(Y^m \in B)$ is minimized by a rule that sets $T^m = 0$.

Thus, the smallest possible value of $P(Y^m \in B)$ is

$$P\left[\{Y(0) \in B\} \cap \{Y(1) \in B\}\right].$$

By similar reasoning, the largest possible value of $P(Y^m \in B)$ is

$$P\left[\{Y(0) \in B\} \cup \{Y(1) \in B\}\right].$$

However, these bounds involve the joint probability distribution of $Y(0), Y(1)$, whereas we only know the marginals.

So we need to get a lower bound on $P\left[\{Y(0) \in B\} \cap \{Y(1) \in B\}\right]$ that involves only the marginal distributions, and then an "upper bound on the upper bound."

Frechet Bounds: a classic result in probability theory gives bounds on joint probabilities in terms of marginal probabilities.

$$\max\{0, P(Y(1) \in B) + P(Y(0) \in B) - 1\} \leq P\left[\{Y(0) \in B\} \cap \{Y(1) \in B\}\right].$$

Therefore,

$$\max\{0, P(Y(1) \in B) + P(Y(0) \in B) - 1\} \leq P(Y^m \in B)$$

and this is our lower bound on $P(Y^m \in B)$ based on the marginals. It can be shown to be sharp.

To get an upper bound, note that

$$P[\{Y(1) \in B\} \cup \{Y(0) \in B\}] = P(Y(1) \in B) + P(Y(0) \in B) - P[\{Y(1) \in B\} \cap \{Y(0) \in B\}].$$

By the Frechet bound for the probability of the intersection, this is less than or equal to

$$P(Y(1) \in B) + P(Y(0) \in B) - \max\{0, P(Y(1) \in B) + P(Y(0) \in B) - 1\}$$

which equals

$$\min\{P(Y(1) \in B) + P(Y(0) \in B), 1\}.$$

Therefore the bounds are

$$\max\{0, P(Y(1) \in B) + P(Y(0) \in B) - 1\} \leq P(Y^m \in B) \leq \min\{P(Y(1) \in B) + P(Y(0) \in B), 1\}.$$