Political Science 236

ATE, ATT and potential outcomes

Rocio Titiunik

Fall 2007

1 Review of Law of Iterated Expectations

Theorem 1.1 *Law of Iterated Expectations (LIE). If X and Y are any two random variables, then*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$$

Proof. (Continuous case). Let $f_{X,Y}(x,y)$ be the joint PDF of (X,Y), $f(x \mid y)$ be the conditional PDF of given $Y = y$ and $f_Y(y)$ be the marginal PDF of Y. By definition of unconditional expectation we have

$$\mathbb{E}[X] = \int x f_X(x)\, dx$$

Since the marginal distribution of X can be recovered from the joint distribution of (X,Y), i.e. $f_X(x) = \int f_{X,Y}(x,y)\, dy$ we can replace $f_X(x)$ to obtain

$$\begin{aligned}\mathbb{E}[X] &= \int x f_X(x)\, dx \\ &= \int \int x f_{X,Y}(x,y)\, dx dy\end{aligned}$$

Since $f_{X,Y}(x,y) = f(x \mid y) f_Y(y)$ we can replace $f_{X,Y}(x,y)$ to obtain

$$\begin{aligned}\mathbb{E}[X] &= \int\int x f(x \mid y) f_Y(y)\, dx dy \\ &= \int \left[\int x f(x \mid y)\, dx\right] f_Y(y)\, dy\end{aligned}$$

But the expression in brackets is the conditional expectation $\mathbb{E}[X \mid y]$ and therefore we have

$$\begin{aligned}\mathbb{E}[X] &= \int \mathbb{E}[X \mid y] f_Y(y)\, dy \\ &= \mathbb{E}[\mathbb{E}[X \mid Y]]\end{aligned}$$

which was what we wanted. (Can you prove it for the discrete case?). ■

Note that there is a subtlety in the last equation, since we are using the same term ”$\mathbb{E}$” to denote different expectations. The expectation in the left-hand side is taken with respect to the marginal distribution of X, the first expectation in the right-hand side is taken with respect to the marginal distribution of Y, and the second expectation in the right-hand side is taken with respect to the conditional distribution of X given Y. Nonetheless, there is no ambiguity because in every case the expectation allows only one interpretation. (For example, $\mathbb{E}[X \mid Y]$ can only be computed over the conditional distribution of $X \mid Y$. Computing $\mathbb{E}[X \mid Y]$ over, say, the marginal distribution of X just doesn't make sense).

Note that in order for the law of iterated expectations to hold we need *not* assume independence. The LIE holds for *any* two random variables.

2 Potential Outcomes

I will now clarify some things about the potential outcomes framework. As usual, we are interested in estimating the average effect of a binary treatment on some outcome of interest. Let i index the number of units, $i = 1, 2, ...N$. Let $Y_i(0)$ and $Y_i(1)$ denote the two potential outcomes of unit i under the control treatment and the active treatment, respectively. The variable $T_i \in \{0, 1\}$

indicates the type of treatment received. Remember that this is a missing data problem: for unit i we only observe T_i, the outcome of this treatment

$$Y_i = \begin{array}{ll} Y_i\left(1\right), & \text{if } T_i = 1 \\ Y_i\left(0\right), & \text{if } T_i = 0 \end{array}$$

and a vector of pretreatment covariates, denoted by X_i. We can rewrite the observed outcome as:

$$\begin{aligned} Y_i &= \left(1 - T_i\right) Y_i\left(0\right) + T_i Y_i\left(1\right) \\ &= Y_i\left(0\right) - T_i Y_i\left(0\right) + T_i Y_i\left(1\right) \\ &= Y_i\left(0\right) + T_i\left(Y_i\left(1\right) - Y_i\left(0\right)\right) \end{aligned}$$

The unit level treatment effect is

$$\tau_i = Y_i\left(1\right) - Y_i\left(0\right)$$

We are interested in $Y\left(1\right) - Y\left(0\right)$, the difference of the outcome with and without treatment. The difference $Y\left(1\right) - Y\left(0\right)$ is a random variable, and we must decide what aspect of its distribution we are interested in estimating. Some of the most common moments that we are interested in estimating are the average treatment effect (ATE), τ, and the average treatment effect on the treated (ATT),

$$ATE = \mathbb{E}\left[Y\left(1\right) - Y\left(0\right)\right]$$

$$ATT = \mathbb{E}\left[Y\left(1\right) - Y\left(0\right) \mid T = 1\right]$$

In general, we will assume some version of the selection on observables assumption, and therefore we will be interested in the conditional versions of the ATE and the ATT:

$$ATE\left(X\right) = \mathbb{E}\left[Y\left(1\right) - Y\left(0\right) \mid X\right]$$

$$ATT\left(X\right) = \mathbb{E}\left[Y\left(1\right) - Y\left(0\right) \mid T = 1, X\right]$$

The question is how to estimate ATE and ATT with a random sample on Y (observed outcome) and T (treatment assignment).

2.1 Random Assignment: $T \perp (Y(1), Y(0))$

Suppose that treatment assignment is independent of $(Y(1), Y(0))$, which would occur if the treatment is randomized across agents. In this case, ATE=ATT. To see why, just note that

$$ATT = \mathbb{E}[Y(1) - Y(0) \mid T = 1] = \mathbb{E}[Y(1) - Y(0)] = ATE$$

since conditioning on T is irrelevant due to independence.

In this case, estimation of ATE straightforward. To see why, note that the expectation of the observed outcome conditional on $T = 1$ and conditional on $T = 0$ can be written as:

$$\begin{aligned}
\mathbb{E}[Y \mid T = 1] &= \mathbb{E}[Y(0) + T(Y(1) - Y(0)) \mid T = 1] \\
&= \mathbb{E}[Y(1) \mid T = 1] \\
&= \mathbb{E}[Y(1)]
\end{aligned}$$

where the last step follows from the assumption of independence. Similarly,

$$\begin{aligned}
\mathbb{E}[Y \mid T = 0] &= \mathbb{E}[Y(0) + T(Y(1) - Y(0)) \mid T = 0] \\
&= \mathbb{E}[Y(0) \mid T = 0] \\
&= \mathbb{E}[Y(0)]
\end{aligned}$$

It follows that

$$\begin{aligned}
ATE &= \mathbb{E}[Y(1) - Y(0)] \\
&= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\
&= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]
\end{aligned}$$

And since we showed that $ATE = \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1) - Y(0) \mid T = 1] = ATT$ (by independence), we have

$$\begin{aligned}
\mathbb{E}[Y(1) - Y(0) \mid T = 1] &= \mathbb{E}[Y(1) - Y(0)] \\
&= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\
&= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]
\end{aligned}$$

Therefore

$$ATE = ATT = \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]$$

Under these assumptions, we can estimate ATE and ATT very easily:

$$\mathbb{E}[Y \mid T = \widehat{1] - \mathbb{E}}[Y \mid T = 0] = \frac{1}{N_1}\sum_{i:T_i=1}^{N} Y_i - \frac{1}{N_0}\sum_{i:T_i=0}^{N} Y_i$$

And this is the reason why we just compute a simple difference in means between treatment outcome and control outcome when we have random assignment. A randomized treatment assignment guarantees that the difference-in-means estimator is unbiased, consistent and asymptotically normal.

Note: this result also holds if we replace the assumption of independence by the assumption of mean independence. This is, $\mathbb{E}[Y(1) \mid T] = \mathbb{E}[Y(1)]$ and $\mathbb{E}[Y(0) \mid T] = \mathbb{E}[Y(0)]$.

2.2 Non-random assignment

In general, we don't have random assignment. Rather, people usually determine whether to receive treatment and this decision is usually related to the benefits of treatment, i.e. to $Y(1) - Y(0)$. When the T and $(Y(1), Y(0))$ are correlated, we need some assumptions in order to identify treatment effects. These assumptions will involve X, a vector of pretreatment covariates. Rosenbaum and Rubin (1983) introduced the assumption of unconfoundness. Let $\mathbb{X}$ denote the support of of X. Formally, unconfoundness requires:

A1. [*Unconfoundness*] For almost every $x \in \mathbb{X}$, T is independent of $(Y(1), Y(0))$ conditional on $X = x$.

The idea behind this assumption is that even though $(Y(1), Y(0))$ are correlated with T, if we observe enough information (contained in X) that determines treatment assignment, then $(Y(1), Y(0))$ will be independent of T, conditional on X. This is usually referred to as the

selection on observables assumption. Note that assumption A1 implies that $\mathbb{E}[Y(1) \mid T, X] = \mathbb{E}[Y(1) \mid X]$ and $\mathbb{E}[Y(0) \mid T, X] = \mathbb{E}[Y(0) \mid X]$

It is interesting to note that under assumption A1 the ATE conditional on X and the ATT conditional on X are equal:

$$\begin{aligned} ATT(X) &= \mathbb{E}[Y(1) - Y(0) \mid X, T = 1] \\ &= \mathbb{E}[Y(1) \mid X, T = 1] - \mathbb{E}[Y(0) \mid X, T = 1] \\ &= \mathbb{E}[Y(1) \mid X] - \mathbb{E}[Y(0) \mid X] \\ &= \mathbb{E}[Y(1) - Y(0) \mid X] \\ &= ATE(X) \end{aligned}$$

So at $X = x$ the average effect of treatment is:

$$ATE(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x]$$

To see how we can estimate this, notice that

$$\begin{aligned} \mathbb{E}[Y \mid X, T = 1] &= \mathbb{E}[Y(0) + T(Y(1) - Y(0)) \mid X, T = 1] \\ &= \mathbb{E}[Y(1) \mid X, T = 1] \\ &= \mathbb{E}[Y(1) \mid X] \end{aligned}$$

where the last step follows from the unconfoundness assumption (A1). Similarly,

$$\begin{aligned} \mathbb{E}[Y \mid X, T = 0] &= \mathbb{E}[Y(0) + T(Y(1) - Y(0)) \mid X, T = 0] \\ &= \mathbb{E}[Y(0) \mid X, T = 0] \\ &= \mathbb{E}[Y(0) \mid X] \end{aligned}$$

It follows that

$$\begin{aligned} ATE(X) &= \mathbb{E}[Y(1) - Y(0) \mid X] \\ &= \mathbb{E}[Y(1) \mid X] - \mathbb{E}[Y(0) \mid X] \\ &= \mathbb{E}[Y \mid X, T = 1] - \mathbb{E}[Y \mid X, T = 0] \end{aligned}$$

so we can estimate $ATE(X)$ at $X = x$ by subtracting the sample mean of Y for control units whose value of X is equal to x from the sample mean of Y for treatment units whose value of X is also equal to x. Note that $ATE(X)$ is a function of X that tells us the average effect of treatment for every value x of X. Thus, $ATE(x)$ tells us the average treatment effect for the subpopulation with $X = x$ and $ATT(x)$ tells us the average treatment effect for the subpopulation with $X = x$ and $T = 1$. By now, the parallelism with Rubin's response surfaces should be clear:

$$R_1(x) \equiv \mathbb{E}[Y \mid X, T = 1]$$

$$R_2(x) \equiv \mathbb{E}[Y \mid X, T = 0]$$

So the response surfaces are just the conditional expectation of Y given X in both the treated and the control populations. Rubin defined the effect of the treatment variable at $X = x$ as

$$R_1(x) - R_2(x)$$

which under this notation is just

$$ATE(x) = ATT(x) = \mathbb{E}[Y \mid X = x, T = 1] - \mathbb{E}[Y \mid X = x, T = 0]$$

If $ATE(x) \equiv R_1(x) - R_2(x)$ does not depend on x (i.e. is constant) the response surfaces are parallel and the objective of the study will be to study this constant difference. If $ATE(x) \equiv R_1(x) - R_2(x)$ does depend on x the response surfaces are non-parallel and *there is no single parameter that completely summarizes the effect of the treatment variable.* When this happens we will be interested in estimating some average effect of the treatment variable. In particular, we are usually interested in estimating the average difference between non-parallel response surfaces over P_1 (the treated population). This will be called the (unconditional) average treatment effect on the treated (ATT).

How can we recover the average effect of treatment from $ATE(x) \equiv \mathbb{E}[Y \mid X = x, T = 1] - \mathbb{E}[Y \mid X = x, T = 0]$? The answer is: by averaging $ATE(x)$ over the distribution of X (we can only do this under some sort of identification assumption, which we will assume holds). And when we average $ATE\left(x\right)$ over the distribution of X we are actually using the law of iterated expectations. If we average over the entire distribution of X we get the average treatment effect (ATE):

$$\begin{aligned}
\mathbb{E}\left[ATE\left(x\right)\right] &= \mathbb{E}\left\{\mathbb{E}\left[Y \mid X=x, T=1\right]-\mathbb{E}\left[Y \mid X=x, T=0\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[Y\left(1\right) \mid X\right]-\mathbb{E}\left[Y\left(0\right) \mid X\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[Y\left(1\right) \mid X\right]\right\}-\mathbb{E}\left\{\mathbb{E}\left[Y\left(0\right) \mid X\right]\right\} \\
&= \mathbb{E}\left[Y\left(1\right)\right]-\mathbb{E}\left[Y\left(0\right)\right] \\
&= \mathbb{E}\left[Y\left(1\right)-Y\left(0\right)\right] \\
&\equiv ATE
\end{aligned}$$

where we have used the law of iterated expectations to move from the third to the fourth line.

Remember that $ATE\left(x\right)=ATT\left(x\right)$. The average effect of treatment on the treated can be recovered by averaging $ATT\left(x\right)\equiv\mathbb{E}\left[Y \mid X=x, T=1\right]-\mathbb{E}\left[Y \mid X=x, T=0\right]$ over the distribution of X conditional on $T=1$:

$$\begin{aligned}
\mathbb{E}\left[ATT\left(x\right) \mid T=1\right] &= \\
&= \mathbb{E}\left\{\mathbb{E}\left[Y \mid X=x, T=1\right]-\mathbb{E}\left[Y \mid X=x, T=0\right] \mid T=1\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[Y\left(1\right) \mid X\right]-\mathbb{E}\left[Y\left(0\right) \mid X\right] \mid T=1\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[Y\left(1\right) \mid X\right] \mid T=1\right\}-\mathbb{E}\left\{\mathbb{E}\left[Y\left(0\right) \mid X\right] \mid T=1\right\} \\
&= \mathbb{E}\left[Y\left(1\right) \mid T=1\right]-\mathbb{E}\left[Y\left(0\right) \mid T=1\right] \\
&= \mathbb{E}\left[Y\left(1\right)-Y\left(0\right) \mid T=1\right] \\
&\equiv ATT
\end{aligned}$$

where the move between the fourth and fifth line is valid because in $\mathbb{E}\left[ATT\left(x\right) \mid T=1\right]$ the outer expectation is taken with respect to the distribution of X in the treatment group.

In Rubin's notation the ATT is

$$ATT=\mathbb{E}_1\left[R_1\left(x\right)-R_2\left(x\right)\right]$$

There is one reason why Rubin's notation is actually more clear. For Rubin, $\mathbb{E}_1\{\cdot\}$ is the expectation over the distribution of X in the treated population (which he calls P_1). In the other notation, $\mathbb{E}\{\cdot \mid T = 1\}$ means exactly the same thing, but it is less clear because when we condition on $T = 1$ it is not obvious that we are actually conditioning on the distribution of X in the treated population. If you are not careful, you may think that you are conditioning on the fact that treatment was received, which is *not* what we are doing. I think this is the reason why ATT seems so confusing some times.