Lecture 4: Conditional expectation

Mitch Downey[1]

February 5, 2024

[1]IIES, Stockholm University. Slides heavily based on those developed by Markus Jäntti based on the textbook by Bruce Hansen.

Introduction

- Conditional expectations are the foundation of applied econometrics
  - All descriptive and casual work
  - All reduced form work
  - Most structural work
  - All linear models (which mainly rely only on assumptions about conditional means)
  - Most non-linear models (which also require assumptions about other moments)
- The book illustrates some of these properties using US data on hourly wages, conditional on gender, education, and race (see chapter 2)
- This lecture lays the foundation to formalize:
  - OLS,
  - machine learning,
  - causal inference,
  - and the connections between these

Conditional Expectations

- Today we'll talk about random variable y with
  - CDF: $F(\cdot)$
  - pdf/density (which we'll assume exists): $f(\cdot)$
  - mean: $\mu = E(y) = \int_{-\infty}^{\infty} yf(y)dy$
- We want to talk about the mean, even though it is sensitive to changes in the tails
- The conditional expectation function (CEF) is defined as

$$m(X) \equiv \mathrm{E}[y|\boldsymbol{X}] = \int_{-\infty}^{\infty} yf_{y|\boldsymbol{X}}(y)dy \tag{1}$$

  where $f_{y|X}(y|x)$ is defined as

$$f_{y|X}(y|x) \equiv \frac{f_{X,y}(x,y)}{f_X(x)}$$

  which is similar to "probability of both" divided by "probability of x"
- Note that $m(\boldsymbol{X})$ is a function of a random variable $\boldsymbol{X}$, so is a random variable

What is a conditional expectation

$$m(X) = E(y|X)$$

- It is not observable
- Most of applied econometrics is about trying to approximate it
  - True for causal inference, too, as we will discuss
- Next few lectures:
  - How do we approximate it with linear models and what do those really mean?
  - What are non-linear "machine learning" models about?
  - How does this relate to causal inference?
- But first, good opportunity to talk about a few properties
  - Law of iterated expectations and law of total probability
  - The conditional variance function

Law of iterated expectations

- (We'll follow Hansen's definition; different sources use same term for slightly different theorems and different terms for same theorem)
- Law of iterated expectations:

$$E\big(E(y|x_1,x_2)|x_1\big) = E(y|x_1)$$

- **Interpretation**: It doesn't actually matter whether $E(y)$ depends on x_2, if you don't observe x_2 and cannot condition on it, then you might just as well only talk about how $E(y)$ depends on x_1
- Implications:
  - Even if we cannot observe everything in the "data generating process," it is still meaningful to talk about the ways that Y varies with the subset of determinants that we can actually observe
  - "The smaller information set wins"
  - Various implications for applied work, one of which is spelled out in the problem set
  - Common response to applied work:[2] "But your outcome also depends on..."
    - Does this matter? Sometimes. We will discuss this in our next lecture.

[2]Especially from non-economists but also from economists, including applied ones

Law of iterated expectations: Proof

$$
\begin{aligned}
E\big(E(y|x_1,x_2)|x_1\big) &= \int_{\mathbb{R}^{k_2}} E(y|x_1,x_2)f(x_2|x_1)dx_2 \\
&= \int_{\mathbb{R}^{k_2}} \left(\int_{\mathbb{R}} yf(y|x_1,x_2)dy \right) f(x_2|x_1)dx_2 \\
&= \int_{\mathbb{R}^{k_2}} \int_{\mathbb{R}} yf(y|x_1,x_2)f(x_2|x_1)dydx_2
\end{aligned}
$$

$$
\left[f(y|x_1,x_2)f(x_2|x_1) = \frac{f(y,x_1,x_2)}{f(x_1,x_2)} \frac{f(x_1,x_2)}{f(x_1)} = f(y,x_2|x_1) \right]
$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^{k_2}} \int_{\mathbb{R}} yf(y,x_2|x_1)dydx_2 \\
&= \int_{\mathbb{R}} y \int_{\mathbb{R}^{k_2}} f(y,x_2|x_1)dx_2dy \quad \text{(because } E|Y| < \infty\text{)} \\
&= \int_{\mathbb{R}} yf(y|x_1)dy \\
&= E(y|x_1)
\end{aligned}
$$

Practical application of LIE

- You're interested in the effect of community development grants on wages
- Grants are randomly assigned, conditional on a formula for the prob of a grant
- Let $Pr(G_c = 1) = z_c'\beta$ be the probability community c gets the grant
- z_c is known (realistic) but β is not
- Wages of individual i in community c: $w_{i,c} = x_i'\gamma + z_c'\eta + \theta G_c + \varepsilon_{i,c}$
  - Assume x_i and z_c are disjoint sets of variables
  - Assume γ is "big"
- You currently observe community-level characteristics, but collecting individual-level data is costly and time consuming. Should you?

Practical application of LIE

- You're interested in the effect of community development grants on wages
- Grants are randomly assigned, conditional on a formula for the prob of a grant
- Let $Pr(G_c = 1) = z_c'\beta$ be the probability community c gets the grant
- z_c is known (realistic) but β is not
- Wages of individual i in community c: $w_{i,c} = x_i'\gamma + z_c'\eta + \theta G_c + \varepsilon_{i,c}$
  - Assume x_i and z_c are disjoint sets of variables
  - Assume γ is "big"
- You currently observe community-level characteristics, but collecting individual-level data is costly and time consuming. Should you?

$$E\big(E(w_{i,c} \mid z_c, G_c, x_i) \mid z_c, G_c\big) = E(w_{i,c} \mid z_c, G_c)$$

True wage function (not estimable), Estimable conditional expectation function

- Individual-level data could still be useful!
  - Heterogeneity by characteristics from x_i
  - Reduce standard error of $\hat{\theta}$
  - γ might be useful

LIE special case 1

- Hansen: Simple law of iterated expectations
- Most of internet: Law of iterated expectations

$$E\big(E(y|x)\big) = E(y)$$

- Special case where x_1 is a null information set

LIE special case 2

- Suppose that we can partition the sample space into pairwise disjoint sets $B_1, ..., B_n$ whose union is the entire sample space
- Law of total expectations (for random variables, special case of LIE):

$$E(X) = \sum_{i=1}^{n} E(X|B_i)P(B_i)$$

- Law of total probability (for events):

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

  where we treat $P(A|B_i) = P(A_i \cap B_i)/P(B_i) = 0$ when $P(B_n) = 0$ (i.e., $0/0 = 0$)
- Incredibly useful for causal inference: We will return to this

Conditional Expectation Function error

- We define the difference between a variable and its conditional mean as the CEF error: $e \equiv y - m(X)$
- By definition, $E(e|X) = E(y|X) - E(y|X) = 0 \;\; \forall X$
- By the law of iterated expectations: $E(e) = E\big(E(e|X)\big) = E(0) = 0$
- Define the conditional variance function: $\sigma^2(X) \equiv Var(y|X) = E\big[(y - m(x))^2|X\big] = E(e^2|X)$
- If $\sigma^2(X) = \sigma^2 \;\; \forall X$, we say that e is *homoskedastic*
- Otherwise, it is *heteroskedastic*
    - Note: Many undergraduate courses encourage you to test for heteroskedasticity linearly along some intuitive dimensions of X; homoskedasticity assumes $\sigma^2(X)$ is constant non-linearly across all dimensions of X
    - You should basically never assume homoskedasticity
- Hansen Section 4.14 is a good discussion of dealing with unknown forms of heteroskedasticity
- These issues become practically important for hypothesis testing, and you'll return to them with Markus

Summary

- Conditional expectations are largely the point of econometrics
- They are not complicated objects
- But they are not observable, and that's the trick
- They are the foundation for the next several lectures
- **OLS**:
  - What is really happening when we estimate this linear approximation of the conditional expectation?
  - How does this help us interpret the results?
  - Hansen chapters 2.15 - 2.25, 3, 4, 7.1-7.3
- **Machine learning**:
  - What challenges arise from making the conditional mean estimator more flexible?
  - What are the mainstream solutions to those challenges?
- **Causality**:
  - How does the notion of a conditional expectation help us define "causal effects"?
  - How does a conditional mean help us estimate causal effects?