

Quantitative Methods in Economics

Conditional Expectations

Maximilian Kasy

Harvard University, fall 2016

Roadmap, Part I

1. Linear predictors and least squares regression
2. **Conditional expectations**
3. Some functional forms for linear regression
4. Regression with controls and residual regression
5. Panel data and generalized least squares

Takeaways for these slides

- ▶ Probability review
- ▶ Conditional probability \leftrightarrow share within subpopulation
- ▶ Conditional expectations minimize average squared prediction error (population concept)
- ▶ Like best linear predictor, but dropping linearity restriction

Probability review

- ▶ Assume for simplicity finite state space.
- ▶ Set of States of Nature
 $S = \{s_1, \dots, s_M\}.$
- ▶ Probability Distribution (Measure):
 $P(s_j) \geq 0, \quad \sum_{j=1}^M P(s_j) = 1.$
- ▶ An *event* A is a subset of S and has probability
 $P(A) = \sum_{s \in A} P(s).$

- ▶ A *random variable* Y is a mapping from S to \mathcal{R} . Its *expectation* is

$$E(Y) = \sum_{j=1}^M Y(s_j)P(s_j).$$

- ▶ Distribution of Y : $F_Y(y) = \sum_{s: Y(s)=y} P(s)$.

Questions for you

Express $E(Y)$ in terms of F_Y

Solution:

$$E(Y) = \sum_{j=1}^M Y(s_j)P(s_j) = \sum_{l=1}^L y_l \left(\sum_{j: Y(s_j)=y_l} P(s_j) \right) = \sum_{l=1}^L y_l F(y_l).$$

Conditional Probability

- ▶ If $P(B) > 0$ and $s_j \in B$,

$$P(s_j | B) = \frac{P(s_j)}{\sum_{s \in B} P(s)} = P(s_j)/P(B); \quad P(s_j | B) = 0 \quad \text{if } s_j \notin B.$$

- ▶ Same properties as unconditional probability:

$$P(s_j | B) \geq 0, \quad \sum_{j=1}^M P(s_j | B) = 1.$$

- ▶ Conditional probability for events:

$$P(A | B) = \sum_{s \in A} P(s | B) = \frac{\sum_{s \in A \cap B} P(s)}{P(B)} = P(A \cap B)/P(B).$$

Properties of conditional probability

- ▶ Partition Formula: If the collection of disjoint subsets $\{B_i\}$ forms a partition of S , then

$$P(A) = \sum_j P(A|B_j)P(B_j).$$

- ▶ Bayes' Rule:

$$\begin{aligned} P(B_i | A) &= \frac{P(A|B_i)P(B_i)}{P(A)} \\ &= \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}. \end{aligned}$$

Questions for you

Prove these formulas.

Conditional Expectation

- ▶ Conditional expectation given event B :

$$E(Y|B) = \sum_{j=1}^M Y(s_j)P(s_j|B).$$

- ▶ Let $X: S \rightarrow \{x_1, \dots, x_K\}$ be another random variable.
- ▶ The regression function $r: \{x_1, \dots, x_K\} \rightarrow \mathcal{R}$ has

$$r(x) = E(Y|X=x) = E(Y|\{s \in S : X(s)=x\}).$$

- ▶ \Rightarrow Conditional expectation given random variable X

- ▶ Joint Distribution. The random vector (X, Y) maps S into

$$\mathcal{X} \times \mathcal{Y} = \{(x, y) : x \in \{x_1, \dots, x_K\}, y \in \{y_1, \dots, y_L\}\},$$

and induces

$$\begin{aligned} F_{XY}(x, y) &= P(X = x, Y = y) \\ &= \sum_{s \in S: X(s)=x, Y(s)=y} P(s) \quad \text{for } (x, y) \in \mathcal{X} \times \mathcal{Y}. \end{aligned}$$

- ▶ Marginal Distribution:

$$F_X: \mathcal{X} \rightarrow [0, 1], \quad F_X(x) = \sum_{y \in \mathcal{Y}} F_{XY}(x, y),$$

$$F_Y: \mathcal{Y} \rightarrow [0, 1], \quad F_Y(y) = \sum_{x \in \mathcal{X}} F_{XY}(x, y).$$

- ▶ Conditional Distribution. If $P(X = x) \neq 0$,

$$F_{Y|X}(y | x) = F_{YX}(x, y) / F_X(x).$$

Optimal Prediction

- ▶ Recall the definition of the best linear predictor:

$$\hat{Y} = \beta_0 + \beta_1 X, \quad \min_{\beta_0, \beta_1} E(Y - \hat{Y})^2$$

- ▶ Now consider the same problem, without linearity:

$$\min_g E[Y - g(X)]^2.$$

Questions for you

- ▶ Rewrite $E[Y - g(X)]^2$ in terms of $F_{XY}(x, y)$.
- ▶ Now rewrite it using $F_{Y|X}(y|x)$ and $F_X(x)$.
- ▶ Solve for the optimal g .

Solution:

- ▶ Rewriting:

$$\begin{aligned} E[Y - g(X)]^2 &= \sum_{x,y} [y - g(x)]^2 F_{XY}(x,y) \\ &= \sum_x \left(\sum_y [y - g(x)]^2 F_{Y|X}(y|x) \right) F_X(x). \end{aligned}$$

- ▶ Thus:

$$\arg \min_c \sum_y (y - c)^2 F(y|x) = E(Y|X=x).$$

- ▶ So the optimal choice of the function g is the regression function r .

General state spaces and continuous distributions

- ▶ So far, we considered discrete state spaces.
- ▶ For continuous distributions, we can derive most of the same results.
- ▶ Let's start with the uniform distribution on $[a, b]$:

$$P([c, d]) = \frac{1}{b-a} \int_c^d dx = \frac{d-c}{b-a} \quad (a \leq c < d \leq b).$$

- ▶ If $A \subset [a, b]$,

$$P(A) = \frac{1}{b-a} \int_A dx.$$

Expectation

- ▶ General notation:

$$\begin{aligned} E(Y) &= \int_S Y(s) dP(s) \\ &= \int y dF_Y(y), \quad \text{with} \quad F_Y(B) = P\{s : Y(s) \in B\}. \end{aligned}$$

- ▶ If the induced distribution for Y is uniform on $[a, b]$, then

$$E(Y) = \frac{1}{b-a} \int_a^b y dy.$$

Generalizing our previous definitions

- ▶ Joint Distribution:

$$F_{XY}(B \times A) = P\{s : X(s) \in B, Y(s) \in A\}.$$

- ▶ Marginal Distribution:

$$F_X(B) = F_{XY}(B \times \mathcal{R}), \quad F_Y(A) = F_{XY}(\mathcal{R} \times A).$$

- ▶ Conditional Distribution:

$$F_{Y|X}(A|x) = P(Y \in A | X = x).$$

- ▶ Partition Formula:

$$F_{XY}(B \times A) = \int_B F_{Y|X}(A|x) dF_X(x).$$

- ▶ Bayes' Rule:

$$P(X \in B | Y \in A) = \frac{\int_B P(Y \in A | X = x) dF_X(x)}{\int P(Y \in A | X = x) dF_X(x)}.$$

Densities

- ▶ Joint Density f_{XY} (with respect to Lebesgue measure on \mathcal{R}^2):

$$\begin{aligned} F_{XY}(B \times A) &= \int_{B \times A} f_{XY}(x, y) dx dy \\ &= \int_A \left(\int_B f_{XY}(x, y) dx \right) dy = \int_B \left(\int_A f_{XY}(x, y) dy \right) dx. \end{aligned}$$

- ▶ Marginal Density:

$$f_X(x) = \int f_{XY}(x, y) dy, \quad f_Y(y) = \int f_{XY}(x, y) dx.$$

- ▶ Conditional Density:

$$f_{Y|X}(y|x) = f_{XY}(x, y)/f_X(x).$$

$$F_{Y|X}(A|x) = \int_A f_{Y|X}(y|x) dy.$$

- ▶ The joint density factors into the product of the conditional density and the marginal density:

$$f_{XY}(x, y) = f_{Y|X}(y|x)f_X(x).$$

- ▶ Partition Formula:

$$f_Y(y) = \int f_{XY}(x, y) dx = \int f_{Y|X}(y|x)f_X(x) dx.$$

- ▶ Bayes' Rule:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x) dx}.$$

Law of iterated expectations

- For Random Variables X and Y , $E[E[Y|X]] = E[Y]$

Proof for the continuous case:

$$\begin{aligned}E[E[Y|X]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy f_X(x) dx \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x) dy dx \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x,y) dy dx \\&= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy \\&= \int_{-\infty}^{\infty} y f_Y(y) dy = E[Y]\end{aligned}$$