Econ. 517, Econometric Theory I — Fall 2001 — Chris Sims

Notes on Conditional Expectation*

1. Abstract definitions

1.1. $E[X \mid \mathcal{G}]$**,** $\mathcal{G}$ **a** σ**-field.** Suppose we have a space $\mathcal{S}$, a σ-field $\mathcal{F}$ defined on it, and a probability P defined on $\mathcal{F}$. (This is sometimes called a probability triple, and written $(\mathcal{S}, \mathcal{F}, P)$). There will be an associated expectation operator E that maps bounded $\mathcal{F}$-measurable functions on S into $\mathbb{R}$. Now suppose that we have a second σ-field $\mathcal{G}$ and that $\mathcal{G} \subset \mathcal{F}$. The expectation of any random variable X on S "conditional on $\mathcal{G}$" is defined as a random variable $E[X \mid \mathcal{G}]$ that has two properties:

- $E[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.
- For any bounded $\mathcal{G}$-measurable random variable Z,

$$E\big[Z \cdot E[X \mid \mathcal{G}]\big] = E[Z \cdot X]\,.$$

The second of these is stronger than necessary. It is enough to require that it holds for all Z's of the form $\mathbf{1}_A$ with $A \subset \mathcal{G}$. This then implies the stronger condition.

1.2. **Uniqueness.** Conditional expectations are only "almost" unique. Suppose a random variable Y is equal to $E[X \mid \mathcal{G}]$ everywhere on $\mathcal{S}$ except for a subset $A \subset \mathcal{S}$ with $P[A] = 0$. Clearly these differences on a set of probability zero don't matter in taking expectations, so $E[E[X \mid \mathcal{G}]Z] = E[YZ] = E[XZ]$ for all bounded $\mathcal{G}$-measurable Z, and therefore Y has the defining properties of a conditional expectation of X given $\mathcal{G}$. In other words, there can be different versions of a conditional expectation that differ only on sets of probability zero.

1.3. $E[X \mid Y]$**,** Y **a random variable.** In applied work, the most common way to generate a sub-σ-field like $\mathcal{G}$ is to consider the σ-field generated by some random variable, say Y. This is the σ-field $\mathcal{F}_Y$ generated by (i.e., the smallest σ-field containing) all sets of the form $\{x \in \mathcal{S} \mid Y(x) \leq a\}$, where a is some real number. Any random variable measurable with respect to $\mathcal{F}_Y$ has the form $f(Y(x))$, i.e. its value is a function of Y. (This is obvious for finite discrete $\mathcal{S}$ and true, if not quite obvious, also for more general S). We then write $E[X \mid Y]$ as a replacement for $E[X \mid \mathcal{F}_Y]$.

1.4. $E[X \mid A]$**,** $A \in \mathcal{F}$**.** There is a related notion, the conditional expectation given a single set $A \in \mathcal{F}$, written $E[X \mid A]$. It is the same thing as $E[X \mid \mathbf{1}_A]$, evaluated at points in A (for all of which, of course, it has the same value).

1.5. **Conditional Probablility.** We can derive the conditional probability of a set from conditional expectation using the usual relation between probability of a set and expectation of its indicator function: $P[A \mid \mathcal{G}] = E[\mathbf{1}_A \mid \mathcal{G}]$. In cases we will be concerned with, $P[A \mid Y]$ for any fixed value of the random variable Y forms a well-behaved probability on $\mathcal{F}$ except for a set of Y values of probability 0. There are pathological situations, though, where this fails to hold.

2. Calculations with discrete probability and density functions on $\mathbb{R}^k$

- When $P[B] > 0$,
$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$
- When probability is defined by a density function over $\mathbb{R}^k$, we often consider the random variables of the form X_i, where for $x \in \mathbb{R}^k$, $X_i(x) = x_i$. Then the conditional probability $P[\cdot \mid X_1, \ldots, X_j]$ is defined by the density function
$$p(x_{j+1}, \ldots, x_k \mid x_1, \ldots, x_j) = \frac{p(\vec{x})}{\int p(x_1, \ldots, x_k)\, dx_{j+1}, \ldots, dx_k}\,.$$
- The most common way to form $E[X \mid Y]$ is to find a joint pdf for X and Y on $\mathbb{R}^2$, construct the conditional pdf as above, and then calculate $E[X \mid Y = y] = \int x p(x \mid y)\, dx$.
- When we have a density and want to find $p(X \mid A)$ for a set A, we can do so by setting $p(x \mid A)$ to zero for $x \notin A$ and $p(x \mid A) = \kappa p(x)$ for $x \in A$, where κ is one over the integral over A of $p(x)$.

3. Examples

- $\mathcal{S} = \{1, 2, 3\}$. $\mathcal{F}$ is the σ-field generated by the three individual points in $\mathcal{S}$ (i.e. the class of all $2^3 = 8$ subsets of $\mathcal{S}$). $P[\{1\}] = .2$, $P[\{2\}] = .4$, $P[\{3\}] = .6$. Then
$$P[\{2,3\} \mid \{1,2\}] = \frac{.4}{.6} = \frac{2}{3}$$
- $\mathcal{S}$ consists of all sequences x_1, x_2, x_3 where each x_i is either 0 or 1. You can think of these points as the vertices of a unit cube in $\mathbb{R}^3$. All 8 points have the same probability .125. The random variables Y and X_1 are defined by $Y(\vec{x}) = x_3 \cdot (x_2 + x_1)$, $X_1(\vec{x}) = x_1$. Then $E[Y \mid X_1]$ is a random variable depending only on x_1. We can find it by finding the conditional probabilities, then using them to form expectations. The conditional probabilities are, for $x_1 = 0$, .25 on each of (0,1,1), (0,0,1), (0,1,0), (0,0,0) and, for $x_1 = 1$, .25 each on the remaining four points. So the conditional expectation is, for $x_1 = 0$, $(1 + 0 + 0 + 0)/4 = .25$, and for $x_1 = 1$, $(2 + 1 + 0 + 0)/4 = .75$.

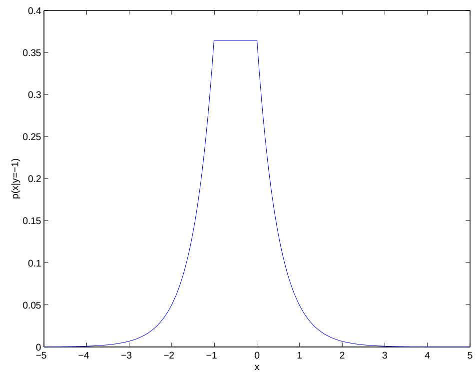

FIGURE 1. Conditional pdf of x given y=-1

- X is a random variable on $[0, \infty)$ with pdf e^{-x}. Suppose we would like to know $E[X \mid X > a]$. The conditional pdf is
$$p(x \mid x > a) = \begin{cases} e^{-(x-a)} & \text{for } x \geq a \\ 0 & \text{for } x < a \end{cases} .$$
The conditional expectation is therefore (recalling how to do integration by parts)
$$\int_a^\infty x e^{-x+a} = a + 1$$
- Suppose the two random variables X and Y have as pdf
$$p(x, y) = .25 e^{-|x|-|y-x|} . \tag{1}$$
To find $E[Y \mid X]$ we form the conditional pdf of $Y \mid X$, which is easily seen here to be $.5e^{-|y-x|}$. Since this is symmetric about x as a function of y, it implies $E[Y \mid X] = X$. A more interesting question is to find $E[X \mid Y]$. The conditional pdf is of course proportional to (1). The exponent in that expression is $2x - y$ for $x < \min\{y, 0\}$, $e^{-|y|}$ for x between y and 0, and $y - 2x$ for $y > \max\{y, 0\}$. The joint pdf is plotted as a function of x, for a particular $y < 0$ and $y > 0$, in Figures 1 and 2. It is easy to see that these sections of the joint pdf are, for fixed y, symmetric in x about $y/2$, so we conclude $E[X \mid Y] = Y/2$. As a further aid to intuition, you can look at the contour plot of $p(x, y)$ in Figure 3.
- **Food for thought:** Suppose you are introduced to someone you don't know, and she says she has two children. As you are talking, a girl walks by. "There's one of my children," says your conversation partner, pointing to the girl walking by. Your conversation breaks off. Later, you decide you need to know the probability that *both* of your conversation partner's children are girls. This would seem to be a conditional probability. There are four possible patterns for first and second child's sex: BB, BG, GB, GG. We

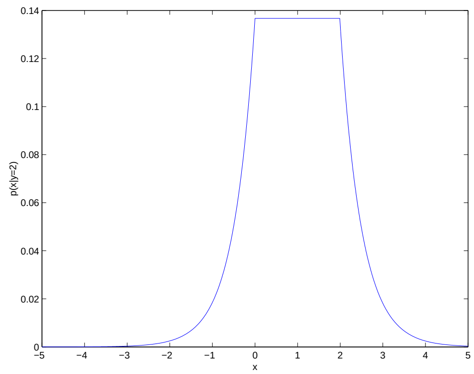

FIGURE 2. Conditional pdf of x given y=2

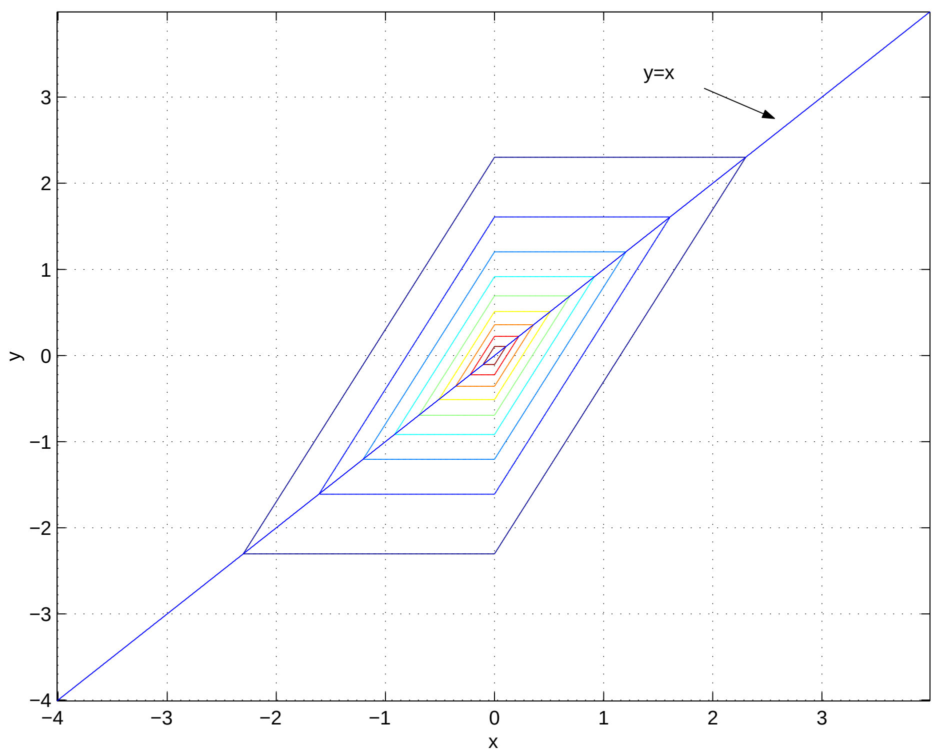

FIGURE 3. Contours of joint pdf of x,y

assume these each has equal probability before we have any other information. So we are interested in the probability of GG given what we now know. What is it? There is some tendency for people to give the wrong answer to this question.