

Math 285

Lecture #15

18 May, 2005

Conditional Expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the sample space, $\mathcal{F}$ is the σ -algebra containing all events which can take probability, and P determines the probability of those events. Consider $\mathcal{G}$ to be a sub- σ -algebra of $\mathcal{F}$, (e.g. $\mathcal{G}$ contains the events you have knowledge about, events you can observe). Typically, it is a σ -algebra containing less information on outcomes than what is given by $\mathcal{F}$.) Suppose X is a random variable defined on $(\Omega, \mathcal{F}, P)$ such that $E[|X|] < \infty$. Expectation is under P .

Def: The conditional expectation of X given $\mathcal{G}$ denoted by $E[X | \mathcal{G}]$, is a (unique) random variable Y such that:

(i.) Y is $\mathcal{G}$ -measurable, ($Y : \Omega \rightarrow \mathbb{R}$, for each interval $(a, b]$ in $\mathbb{R}$,

$$Y^{-1}((a, b]) \equiv \{\omega \in \Omega : Y(\omega) \in (a, b]\} \in \mathcal{G}) \text{ (i.e. you only need information}$$

from $\mathcal{G}$ to determine Y . Remember $\mathcal{F}$ is typically bigger than $\mathcal{G}$.)

(ii.) For each set $A \in \mathcal{G}$,

$$E[X \cdot 1_A] = E[Y \cdot 1_A]$$

(i.e. the integral of X on A = the integral of Y on A .)

Existence follows from the Radon-Nikolym Theorem in real analysis. Uniqueness up to almost sure equivalence is proved below.

Lemma: The conditional expectation is unique up to almost sure equivalence.

Proof: If Y and Y' were two representatives for $E[X | \mathcal{G}]$, then let

$$A = \{Y > Y'\} = \{\omega : Y(\omega) > Y'(\omega)\}.$$

Then $A \in \mathcal{G}$ by (i.) In particular,

$$E[Y \cdot 1_A] = E[X \cdot 1_A] = E[Y' \cdot 1_A], \text{ by (ii.)}$$

The above implies

$$E[(Y - Y') \cdot 1_A] = 0$$

Note, on A , $Y - Y' > 0$. Thus the above is stating that an integral of a non-negative function on a set is 0. This can only happen if $(Y - Y') \cdot 1_A = 0$ almost surely on A since $Y > Y'$ on A . This implies that A itself has probability 0. Thus, $P(A) = 0$ and so $P(Y > Y') = 0$. Similarly $P(Y' > Y) = 0$. Hence $P(Y = Y') = 1$. And thus, the conditional expectation is unique up to almost surely equivalence. $\square$

EXAMPLE 1 Discrete Random Variables

If the probability space is finite ($\Omega < \infty$), we can compute these things by normal addition. Consider X_0, X_1 to be random variables taking values in the set $\{0, 1\}$. The set of four possible combinations X_0, X_1 can take on ,

$\{(X_0 = 0, X_1 = 0), (X_0 = 0, X_1 = 1), (X_0 = 1, X_1 = 0), (X_0 = 1, X_1 = 1)\}$, is illustrated through

the following diagram. The set Ω can be partitioned into two sets A_1 and A_2 , where A_1 contains the points at which $X_0 = 0$ and A_2 contains the points of which $X_0 = 1$.

Ω :

A_1 :	$(X_0 = 0, X_1 = 0)$ B_1	$(X_0 = 0, X_1 = 1)$ B_2
A_2 :	$(X_0 = 1, X_1 = 0)$ B_3	$(X_0 = 1, X_1 = 1)$ B_4

Suppose that

1. $P(X_0 = 0, X_1 = 0) = \frac{1}{8} = P(X_0 = 0, X_1 = 1)$
2. $P(X_0 = 1, X_1 = 0) = \frac{1}{4}$
3. $P(X_0 = 1, X_1 = 1) = \frac{1}{2}$

Note, X_0, X_1 are not independent.

Whenever we have a random variable (like X_0), we can associate a σ -algebra to it.

(Note, σ -algebras are collections of sets containing the empty set, and closed under complements and countable unions.)

$\sigma(X_0)$ = σ -algebra generated by X_0 = smallest σ -algebra containing all sets of the form $X_0^{-1}(a, b]: -\infty < a < b < \infty = \{X_0^{-1}(B) : B \text{ is a Borel set in } \mathbb{R}\}$. (Intervals generate the σ -algebra of Borel sets.).

Informally, $\sigma(X_0)$:= collection of events distinguishing the values that X_0 takes.

In our example, $\sigma(X_0) = \{\emptyset, \Omega, A_1, A_2\}$. Note $A_1 \cap A_2 = \emptyset$; $A_1 \cup A_2 = \Omega$.

Suppose $G = \sigma(X_0) = \{\emptyset, \Omega, A_1, A_2\}$ and $F = \sigma\{B_1, B_2, B_3, B_4\}$.

Then, we can compute

$E[X_1 | \sigma(X_0)] = E[X_1 | X_0] = Y$ (equivalent by just different use of notation.). The

random variable Y has to be constant separately on A_1 and A_2 because you can't

refine A_1 or A_2 anymore knowing just X_0 . Thus, to determine Y , we must find the

value it takes on A_1 and the value it takes on A_2 ...

On A_1 : Suppose $Y = y_1$

$$y_1 P(A_1) = E[Y \cdot 1_{A_1}] = E[X_1 \cdot 1_{A_1}]$$

$$= 0 \cdot P(B_1) + 1 \cdot P(B_2) \text{ since only integrate over } A_1.$$

$$= \frac{1}{8}.$$

$$\therefore y_1 = \frac{1/8}{P(A_1)} = \frac{1/8}{1/4} = \frac{1}{2}.$$

On A_2 : Suppose $Y = y_2$:

$$\begin{aligned} y_2 P(A_2) &= E[Y \cdot 1_{A_2}] = E[X_1 \cdot 1_{A_2}] \\ &= 0 \cdot P(B_3) + 1 \cdot P(B_4) \text{ since only integrate over } A_2. \\ &= \frac{1}{2}. \end{aligned}$$

$$\therefore y_2 = \frac{1/2}{P(A_2)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

So now we can specify Y ,

$$Y = \begin{cases} \frac{1}{2} & \text{on } A_1 \\ \frac{2}{3} & \text{on } A_2 \end{cases}.$$

Note, when you know you're on A_1 , in computing the conditional expected value, you renormalize so that the probability mass is all concentrated on A_1 , by dividing by $P(A_1)$.

EXAMPLE 2 Continuous Random Variables

Suppose we have two continuous random variables X_0, X_1 with joint density $f(x_0, x_1)$.

$$\left(\text{Also note that } P((X_0, X_1) \in B) = \iint_B f(x_0, x_1) dx_0 dx_1 \right)$$

Conditional Density

$$f_{X_1|X_0}(x_1 | x_0) = \left\{ \begin{array}{ll} \frac{f(x_0, x_1)}{\int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1}, & \text{if } \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 \neq 0 \\ 0, & \text{if } \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 = 0 \end{array} \right\}, \text{ which is some density}$$

associated with X_1 , given a viable value of X_0 . Note, if $\int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 = 0$, no value of X_1 can occur, if conditioned on X_0 . For clarity, x_1 is a possible value of X_1 , x_0 is a given value of X_0 , and $\bar{x}_1$ is a dummy value of X_1 .

Note, we are not assuming f is continuous.

Consider another measurable function $g : \mathbb{R} \longrightarrow \mathbb{R}$ such that $E[|g(X_1)|] < \infty$.

Find $E[g(X_1) | X_0]$.

Claim:

$$E[g(X_1) | X_0] = \int_{\mathbb{R}} g(\bar{x}_1) f_{X_1|X_0}(\bar{x}_1 | X_0) d\bar{x}_1 = \phi(X_0), \text{ a measurable function of the}$$

random variable X_0 only, since the variable X_1 is integrated out.

Note:

Because X_0 is continuous, the notation below is technically incorrect because it involves conditioning on an event of probability 0. However this notation is commonly used intuitively or informally.

$$E[g(x_1) | X_0 = x_0] = \int_{\mathbb{R}} g(x_1) f_{X_1|X_0}(x_1 | x_0) dx_1 = \phi(x_0).$$

Moving on into our calculation of $E[g(X_1)|X_0] \dots$

$$E[g(X_1)|X_0] = \int_{\mathbb{R}} g(x_1) \cdot \frac{f(X_0, x_1)}{\int_{\mathbb{R}} f(X_0, \bar{x}_1) d\bar{x}_1} dx_1$$

Consider a set $A \in \mathcal{G} = \sigma(X_0) = \{X_0^{-1}(B) : B \text{ Borel set in } \mathbb{R}\}$. There is a Borel set B in $\mathbb{R}$ such that $A = X_0^{-1}(B)$, the pullback of B .

Now, let's show that $E[E[g(X_1)|X_0] \cdot 1_A] = E[g(X_1) \cdot 1_A]$.

Proof:

It is useful to replace A by its pullback set. Thus we want to check if

$$\begin{aligned} E[E[g(X_1)|X_0] \cdot 1_{\{X_0 \in B\}}] &= E[g(X_1) \cdot 1_{\{X_0 \in B\}}] \\ \text{LHS: } \int_{\mathbb{R}} \left[\int_{\mathbb{R}} g(x_1) \cdot \frac{f(x_0, x_1)}{\int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1} dx_1 \right] \cdot 1_B(x_0) f_{X_0}(x_0) dx_0 \\ \text{RHS: } \iint_{\mathbb{R}^2} g(x_1) \cdot 1_B(x_0) f(x_0, x_1) dx_1 dx_0 \end{aligned}$$

Now, we use the definition of marginal density to substitute into the denominator of the left hand side (LHS) above. We substitute: $f_{X_0}(x_0) := \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1$, and thus this marginal density will cancel with the density already present in the denominator of the LHS. We obtain the following:

$$\begin{aligned} \text{LHS: } \iint_{\mathbb{R}^2} g(x_1) f(x_0, x_1) \cdot 1_B(x_0) dx_1 dx_0, \text{ and} \\ \text{RHS: } \iint_{\mathbb{R}^2} g(x_1) f(x_0, x_1) \cdot 1_B(x_0) dx_1 dx_0, \end{aligned}$$

Thus we have shown that $E[E[g(X_1)|X_0] \cdot 1_A] = E[g(X_1) \cdot 1_A]$. $\square$

Various Properties of Conditional Expectation

(For more reading, see Durrett's *Probability Theory and Examples* or the probability book by Billingsley.)

Suppose X is a random variable with finite first moment, $E[|X|] < \infty$ and $\mathcal{G}$ is a sub- σ -algebra of $\mathcal{F}$, where X is defined on $(\Omega, \mathcal{F}, P)$.

1. If X is $\mathcal{G}$ -measurable,

$$E[X | \mathcal{G}] = X.$$

(Knowing $\mathcal{G}$ tells you everything about X .)

2. In contrast to 1., if X is independent of $\mathcal{G}$, i.e.

$$P(\{X \in A\} \cap B) = P(\{X \in A\})P(B), \text{ for any Borel set } A \text{ in } \mathbb{R}, \text{ any set } B \in \mathcal{G},$$

then

$$E[X | \mathcal{G}] = E[X]$$

(Knowing $\mathcal{G}$ tells you nothing about X , since X is independent of $\mathcal{G}$.)

3. If $a \in \mathbb{R}$ and Y is a random variable with finite first moment $E[|Y|] < \infty$, then

$$E[aX + Y | \mathcal{G}] = a \cdot E[X | \mathcal{G}] + E[Y | \mathcal{G}]$$

(Thus, conditional expectation is linear.)

4. If Y is a random variable such that $E[|XY|] < \infty$ and Y is $\mathcal{G}$ -measurable, then

$$E[XY | \mathcal{G}] = Y \cdot E[X | \mathcal{G}]$$

(Knowing $\mathcal{G}$ tells you everything about Y . There is some integrability to check here too hence the condition on XY).

5. If $X \leq Y$ a.s. (almost surely) then

$$E[X | \mathcal{G}] \leq E[Y | \mathcal{G}] \text{ a.s.}$$

6. $|E[X | \mathcal{G}]| \leq E[|X| | \mathcal{G}]$

Note, $|\cdot|$ is a convex function. This property is a particular case of 7.

7. Jensen's Inequality:

If $\phi: \mathbb{R} \rightarrow \mathbb{R}$ is a convex function, $E[|\phi(X)|] < \infty$, then

$$\phi(E[X | \mathcal{G}]) \leq E[\phi(X) | \mathcal{G}]$$

Note for ϕ convex:

$$\lambda \in (0,1), \quad x, y \in \mathbb{R},$$

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda \phi(x) + (1-\lambda)\phi(y)$$

8. Monotonicity

If for each n , X_n is a random variable such that $X_n \geq 0$ and $X_n \uparrow X$ a.s. as $n \rightarrow \infty$

(i.e. the sequence X_n increases to the random variable X), where we have

assumed $E[|X|] < \infty$, then

$$E[X_n | \mathcal{G}] \uparrow E[X | \mathcal{G}] \text{ a.s. as } n \rightarrow \infty$$

9. Dominated Convergence

If $X_n \rightarrow X$ a.s. as $n \rightarrow \infty$, where X is a random variable and $|X_n| < Y$ a.s. for all n

where $E[|Y|] < \infty$. (i.e. the sequence X_n is dominated by the random variable Y),

then

$$E[X_n | \mathcal{G}] \rightarrow E[X | \mathcal{G}] \text{ a.s. as } n \rightarrow \infty .$$

Note, properties **8.** and **9.** are two conditions to pass a limit inside conditional expected value.

Martingales (discrete time, discrete or continuous state)

Applications:

Martingales occur in stochastic models and in statistics.

Examples:

1.) (Stochastic models) Model of a fair game.

2.) (Statistics) Generalization of partial sums of i.i.d. random variables with zero mean. Note, there is a Law of Large Numbers for (backwards) martingales and a Central Limit Theorem for (forwards) martingales.

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Defn: A filtration is a family of sub- σ – algebras of $\mathcal{F}$: $\{\mathcal{F}_n, n = 0, 1, \dots\} = \{\mathcal{F}_n\}$

such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all n . In other words, as time goes on, a filtration keeps track of information up to that time. It does not forget anything. Thus, a filtration is an increasing family of sub- σ – algebras of $\mathcal{F}$.

Defn: A martingale (relative to a filtration $\{\mathcal{F}_n\}$) is a discrete time stochastic process

$M = \{M_n, n = 0, 1, 2, \dots\} = \{M_n, \mathcal{F}_n, n = 0, 1, 2, \dots\}$ such that

(i.) The random variable M_n is an $\mathcal{F}_n$ -measurable random variable. (i.e.

$M_n : \Omega \rightarrow \mathbb{R}$ and $M_n^{-1}(a, b] \in \mathcal{F}_n$ for all $-\infty < a < b < \infty$.) Note, M_n is not

allowed to anticipate the future. We sometimes write $M_n \in \mathcal{F}_n$

(ii.) $E[|M_n|] < \infty$ for all n

(iii.) $E[M_{n+1} | \mathcal{F}_n] = M_n$ for all $n \geq 0$.

e.g. For the example using martingales to model fair games: given $\mathcal{F}_n$,

and wealth M_n at time n , the average value of the wealth at the next time

step $n+1$ is the same as the amount of wealth now (M_n). Thus, it is a *fair*

game.

e.g. For an unfair game, a sub-martingale is defined similarly with the

following in place of (iii):

$$E[M_{n+1} | \mathcal{F}_n] \geq M_n \text{ for all } n \geq 0,$$

where the player unfairly gains money...

A super-martingale is defined when in place of (iii) we have

$$E[M_{n+1} | \mathcal{F}_n] \leq M_n \text{ for all } n \geq 0,$$

where the house unfairly gains money...

Next Lecture: More theory and applications of martingales.

