

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 is the probability of those events. Consider $\mathcal{G}$ to be a ^{sub} σ -algebra of $\mathcal{F}$. ^{eg.} $\mathcal{G}$ contains the events you have knowledge about, events you ^{can} observe, ~~or do not~~. ^{Typically it} It is a small σ -algebra containing less information on outcomes of probability space than what is possible.) 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, i.e. $Y \in \mathcal{G}$, $(Y: \Omega \longrightarrow \mathbb{R})$, for each interval $(a, b]$ in $\mathbb{R}$,

$Y^{-1}((a, b]) = \{\omega \in \Omega: Y(\omega) \in (a, b]\} \in \mathcal{G}$ (i.e. you only need information on $\mathcal{G}$. Remember $\mathcal{F}$ is 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 ~~over~~ ^{of} X on A = the integral ~~over~~ ^{of} Y on A .)

Existence follows from the Radon-Nikolym Theorem in real analysis. Uniqueness is proved below.

The conditional expectation is unique up to almost sure equivalence.

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

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

This implies $A \in \mathcal{G}$ 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. $\square$

Examples

EXAMPLE 1 Discrete Random Variables

If probability space is finite ($\Omega < \infty$), can compute these things by normal addition.

Consider X_0, X_1 are 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 table. The Ω set can be partitioned into two sets A_1 and A_2 , where A_1 contains the ~~pairs~~ ^{points at} which $X_0 = 0$ and A_2 contains the ~~pairs~~ ^{points at} 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 (sets containing empty set, and closed under complements and countable unions) by it.

$\sigma(X_0)$ = smallest σ -algebra, generated by X_0 , containing all sets

$\{X_0^{-1}(a, b]: -\infty < a < b < \infty\} = \{X_0^{-1}(\beta): \beta \text{ is a Borel set in } \mathbb{R}\}$. (Intervals generate the σ -algebra of

Borel sets). We have $\sigma(X_0) :=$ collection of events discriminating what values X_0

can take. $\sigma(X_0) = \{\emptyset, \Omega, A_1, A_2\}$. Note $A_1 \cap A_2 = \emptyset$; $A_1 \cup A_2 = \Omega$.

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

Then, we can compute

$$E[X_1 | \sigma(X_0)] = E[X_1 | X_0] = Y \text{ (by different use of notation).}$$

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. 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 $Y = y_1$

$$y_1 P(A_1) = E[X_i 1_{A_1}] = \underset{\substack{\uparrow \\ \text{by def.} \\ \text{of cond.} \\ \text{Exp. Value}}}{0 \cdot P(B_1) + 1 \cdot P(B_2)} = 1/8$$

(only integrate over A_1)

$$\begin{aligned} \Rightarrow y_1 &= \frac{1/8}{P(A_1)} \\ &= \frac{1/8}{1/4} \\ &= \underline{\underline{1/2}} \end{aligned}$$

On A_2 : Suppose $Y = y_2$:

$$y_2 \cdot P(A_2) = E[Y 1_{A_2}] = E[X_i \cdot 1_{A_2}]$$

$$= 0 + 1 \cdot P(B_4) = 1 \cdot 1/2 = 1/2$$

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

So now we can specify y !

$$y = \begin{cases} 1/2 & \text{on } A_1 \\ 2/3 & \text{on } A_2 \end{cases}$$

EXAMPLE 2 CONTINUOUS RANDOM VARIABLES

Suppose we have 2 continuous random variables X_0, X_1 , with joint density function $f(X_0, X_1)$

non-negative function on plane

$$P((X_0, X_1) \in B) = \iint_B f(x_0, x_1) dx_0 dx_1$$

"any set on plane"

CONDITIONAL DENSITY:

$$f_{X_1|X_0}(x_1|x_0) = \begin{cases} \frac{f(x_0, x_1)}{\int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1} & \text{if } \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 \neq 0 \\ 0 & \text{if } \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 = 0 \end{cases}$$

renormalize to be a probability density.

this is some density associated with X_1 , given a value of X_0 .

x_1 - possible value of X_1

x_0 - given value of X_0

X_1 - dummy variable

0

$$\text{if } \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 = 0$$

$$\text{if } \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1 = 0$$

$\Rightarrow$ no values of X_1 would occur if conditioned on X_0 .

(note: we're not assuming f is continuous)

Consider another function

$$g: \mathbb{R} \rightarrow \mathbb{R}$$

such that

note, if g is bounded, this is automatic

$$E[|g(X_1)|] < \infty. \text{ Find}$$

$E[g(X_1) | X_0]$

$[X_0]$

(6)

CLAIM: $E[g(X_1) | \bar{X}_0] = \int_{\mathbb{R}} g(\bar{x}_1) f_{\bar{X}_1 | \bar{X}_0}(\bar{x}_1 | \bar{X}_0) d\bar{x}_1$

$$= \Phi(\bar{X}_0)$$

↑ function of $\bar{X}_0$ because we're integrating out $\bar{X}_1$.

Side note, the below notation is really incorrect, because it is conditioning on an event that takes on probability 0.

$$E[g(X_1) | \bar{X}_0 = x_0] = \int_{\mathbb{R}} g(x_1) f_{\bar{X}_1 | \bar{X}_0}(x_1 | x_0) dx_1$$

$$E[g(X_1) | \bar{X}_0] = \int_{\mathbb{R}} g(x_1) \frac{f(\bar{X}_0, x_1)}{\int_{\mathbb{R}} f(\bar{X}_0, \bar{x}_1) d\bar{x}_1} dx_1$$

$$= \Phi(\bar{X}_0)$$

↑ $\bar{X}_0$ measurable random variable
 ↑ measurable function of $\bar{X}_0$.
 ↑ Φ is a measurable function since f is a density.

Consider a set $A \in \mathcal{G} = \sigma(\bar{X}_0) = \{ \bar{X}_0^{-1}(B) : B \text{ is a Borel set on } \mathbb{R} \}$

There is a B , Borel set in $\mathbb{R}$, such that $A = \bar{X}_0^{-1}(B)$. (A is a pullback of B .)

check

$$\underbrace{E[E[g(X_1) | \bar{X}_0] 1_A]}_{\text{LHS}} \stackrel{?}{=} \underbrace{E[g(X_1) 1_A]}_{\text{RHS}}$$

* it is useful to replace A by its pullback set $\rightarrow$

$$= E \left[\underbrace{E[g(x_1) | \mathcal{F}_0]}_{\substack{\uparrow \text{function of } \mathcal{F}_0, \phi(\mathcal{F}_0) \\ \text{since } \mathcal{G} \text{ is generated} \\ \text{by } \mathcal{F}_0}} \mathbb{1}_{\{x_0 \in B\}} \right] = E \left[g(x_1) \mathbb{1}_{\{x_0 \in B\}} \right]$$

$\uparrow$
function of $\mathcal{F}_0, \mathcal{F}_1$

$$= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} g(x_1) \frac{f(x_0, x_1)}{\int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1} \mathbb{1}_B(x_0) f_{x_0}(x_0) dx_0 \right] dx_1$$

LHS

$$\stackrel{?}{=} \underbrace{\int_{\mathbb{R}^2} g(x_1) \mathbb{1}_B(x_0) f(x_0, x_1) dx_0 dx_1}_{\text{RHS}}$$

Note: $f_{x_0}(x_0) = \int_{\mathbb{R}} f(x_0, \bar{x}_1) d\bar{x}_1$ // marginal density
so substitute this in LHS, and the
marginal densities $f_{x_0}(x_0)$ cancel.

$$\Rightarrow \text{LHS} = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_1) f(x_0, x_1) \mathbb{1}_B(x_0) dx_1 dx_0$$

$$\Rightarrow \text{RHS} = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_1) f(x_0, x_1) \mathbb{1}_B(x_0) dx_0 dx_1$$

$\swarrow$ change order of integration, since f is a density, by Fubini's theorem.

Various Properties: (To read more,

- Durrett, Probability Theory & Examples
- Billingsley,

Suppose X is a random variable, $E[|X|] < \infty$
and $\mathcal{G}$ is a sub- σ -algebra of $\mathcal{F}$.

① If X is $\mathcal{G}$ -measurable, $E[X|\mathcal{G}] = X$

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

② If X is independent of $\mathcal{G}$, i.e. $P(\{X \in A\} \cap B) = P(\{X \in A\}) P(B)$
then $E[X|\mathcal{G}] = E[X]$

(Knowing $\mathcal{G}$ tells you nothing about X .)

for any $B \in \mathcal{G}$
" " A Borel in $\mathbb{R}$.

③ If $a \in \mathbb{R}$, Y random variable, $E[|Y|] < \infty$
— then

$$E[aX + Y | \mathcal{G}]$$

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

(Conditional expected value is Linear)

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

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

(Given $\mathcal{G}$, we know Y .)

↳ e.g. can use ④ to prove ①

⑤ If $X \leq Y$ almost surely
then $E[X|G] \leq E[Y|G]$ almost surely.

⑥ $|E[X|G]| \leq E[|X|G]$ $\leftarrow$ ex of ⑦ with $\phi(t) = |t|$

⑦ Jensen's Inequality:

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

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

note. ϕ convex : $\lambda \in (0,1)$, $x, y \in \mathbb{R}$

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

⑧ Monotonicity

If X_n is a random var. such that $X_n \geq 0$
and $X_n \uparrow X$ almost surely as $n \rightarrow \infty$,
where $E[X] < \infty$

$X_n \uparrow X$ = "Sequence X_n increases to random var. X "

then $E[X_n|G] \uparrow E[X|G]$ almost surely
as $n \rightarrow \infty$

⑨ Dominated Convergence

If $X_n \rightarrow X$ a.s. as $n \rightarrow \infty$ X , r.v.

$|X_n| \leq Y$ for all n where $E[Y] < \infty$

(sequence is dominated by r.v. Y)

then $E[X_n|G] \rightarrow E[X|G]$ a.s. as $n \rightarrow \infty$

⑧, ⑨ are the 2 conditions to pass limit inside expected value ⑩

MARTINGALES (discrete time, ^{discrete or} continuous state)

- intuition:

↑
Occur in stochastic models
e.g. model of a fair game

↑
Occur in statistics
e.g. generalization of partial sums of
i.i.d. random variables with 0-mean.

- there is a Law of Large Numbers ^(backwards) for Martingales
" " Central Limit Theorem ^(forwards)

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

DEFN: A filtration is a family of sub- σ -algebras
of $\mathcal{F}$: $\{\mathcal{F}_n, n=0,1,\dots\} = \{\mathcal{F}_n\}$ (since discrete time)
such that

$$\mathcal{F}_n \subset \mathcal{F}_{n+1}.$$

In other words, as time goes on, a filtration
keeps track of information up to that time.

It never forgets anything. (It 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 $M = \{M_n, n=0,1,2,\dots\}$
 $= \{M_n, \mathcal{F}_n, n=0,1,\dots\}$

Such that

(i) R.V. M_n is an $\mathcal{F}_n$ -measurable r.v.

(i.e. $M_n: \Omega \rightarrow \mathbb{R}$)

$M_n^{-1}(a,b] \in \mathcal{F}_n$ for all $-\infty < a < b < \infty$)

↑ not allowed to anticipate the future.

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

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

e.g. Given I_n and fortune M_n at $t=n$, the amount of fortune in time $t+1$ is same as " " " now (M_n).

Unfair game " $=$ " becomes " $\leq$ " or " $\geq$ "

sub / sup Martingale is defined similarly but with $\geq / \leq$ instead of $=$ in (iii.)

Next lecture: examples,
theory
and applications