Lectures 24-25

0.1 Conditional Expectation

When Y is a simple random variable, i.e.

$$Y = \sum_{k=1}^{n} a_k P(A_k)$$

and X is either discrete (with joint pmf f) or continuous with a pdf (i.e absolutely continuous) and is such that $G(x, y) = P(X \le x, Y = y)$ is differentiable in x for each $y = a_k$, then we have defined $f_{Y|X}(a_k|x)$ as

$$P(Y = a_k|X = x) := \begin{cases} \frac{f(x,a_i)}{f_X(x)} & \text{if} \quad X \text{ is discrete} \\ \lim\limits_{h\to\infty} P\big(Y = a_k|X \in (x-h, x+h]\big) & \text{if} \quad X \text{ is absolutely} \\ = \frac{1}{f_X(x)}\frac{dG}{dx}(x, a_k) & \qquad \text{continuous} \end{cases}$$

Recall that we defined EY for simple random variable Y as

$$EY = \sum_{k=1}^{n} a_k P(A_k).$$

Hence we define $E[Y|X = x]$ as follows.

$$E[Y|X = x] \;=\; \sum_{k=1}^{n} a_k P\big(A_k|X = x\big) \;=\; \sum_{k=1}^{n} a_k f_{Y|X}(a_k|x).$$

When Y is non negative, then we define

$$E[Y|X = x] = \lim_{n\to\infty} E[Y_n|X = x]$$

where

$$Y_n \;=\; \sum_{k=1}^{n2^n-1} \frac{k}{2^n} I_{\{k/2^n \le Y < (k+1)/2^n\}}.$$

Now

$$\begin{aligned}
\lim_{n\to\infty} E[Y_n|X=x] &= \lim_{n\to\infty}\sum_{k=1}^{n2^n-1}\frac{k}{2^n} f_{Y_n|X}(k/2^n|x) \\
&= \lim_{n\to\infty}\sum_{k=1}^{n2^n-1}\frac{k}{2^n} P\big(k/2^n \le Y < (k+1)/2^n|X=x\big) \\
&= \lim_{n\to\infty}\sum_{k=1}^{n2^n-1}\frac{k}{2^n}\int_{k/2^n}^{(k+1)/2^n} f_{Y|X}(y|x)dy \\
&= \lim_{n\to\infty}\sum_{k=1}^{n2^n-1}\int_{k/2^n}^{(k+1)/2^n} yf_{Y|X}(y|x)dy \\
& \quad \lim_{n\to\infty}\sum_{k=1}^{n2^n-1}\int_{k/2^n}^{(k+1)/2^n} (k/2^n - y)f_{Y|X}(y|x)dy \\
&= \int_0^\infty yf_{Y|X}(y|x)dy.
\end{aligned}$$

Hence according to our definition, when Y is nonegative such that $f_{Y|X}$ exists,then

$$E[Y|X=x] = \int_0^\infty yf_{Y|X}(y|x)dy.$$

Combining all the above discussion, we get the definitions.

Definition 0.1 *Let X, Y be random variables. Then conditional expectation of Y given $X = x$ is defined as*

$$\begin{aligned}
E[Y\,|\,X = x] &= \sum_{y\in D_Y} yf_{Y\,|\,X}(y\,|\,x)\,, \text{if both } X, Y \text{ discrete} \\
E[Y\,|\,X = x] &= \int_{-\infty}^{\infty} yf_{Y\,|\,X}(y\,|\,x)dy\,, \text{if both } X, Y \text{ continuous with joint pdf } f \\
E[Y\,|\,X = x] &= \int yf_{Y\,|\,X}(y\,|\,x)dy\,, \text{if } X \text{ is discrete, } Y \text{ continuous, } f_{Y|X} \text{ exists} \\
E[Y\,|\,X = x] &= \sum_{y\in D_Y} f_{Y\,|\,X}(y\,|\,x)\,, \text{if } Y \text{ is discrete, } X \text{ continuous, } f_{Y|X} \text{ exists}
\end{aligned}$$

Definition 0.2 *Let X, Y be random variables such that $g(x) = E[Y|X = x]$ is well defined for each x. Then we define $E[Y|X]$ as the random variable*

$$E[Y|X] = g(X).$$

Remark 0.1 *From the above definition, it follows that $E[Y|X]$ is a random variable with respect to $\sigma(X)$.*

Example 0.1 *Let X, Y be independent random variables with geometric distribution of parameter $p > 0$. Calculate $E[Y \mid X+Y]$. Also show that $E\big[E[Y \mid X+Y]\big] = EY$.*

Set

$$Z = X + Y .$$

$$P(Y = y \mid Z = n) = 0 \text{ if } y \geq n+1 .$$

For $y = 0, 1, 2, \cdots, z$

$$P(Y = y|Z = n) = \frac{P\{Y = y, X+Y = n\}}{P\{X+Y = n\}} .$$

Now

$$P\big(X+Y = n\big) = \sum_{x=0}^{n} P\big(X = x,\, Y = n-x\big) = (n+1)p^2(1-p)^n .$$

$$\begin{aligned} P\big(Y = y,\, X+Y = n\big) &= P\big(Y+y,\, X = n-y\big) \\ &= P(Y = y)\, P(X = n-y) \\ &= p(1-p)^y p(1-p)^{n-y} = p^2(1-p)^n . \end{aligned}$$

Therefore

$$P(Y = y|Z = n) = \frac{1}{n+1} .$$

i.e.,

$$f_{Y|Z}(y|z) = \begin{cases} \frac{1}{z+1} & \text{if} \quad y = 0, 1, 2, \cdots, z \\ & = 0 \quad \text{otherwise} \end{cases} .$$

Now

$$E[Y|Z = n] = \sum_y y f_{Y|Z}(y|n) = \sum_{k=1}^{n} k \frac{1}{n+1} = \frac{1}{n+1} \frac{n(n+1)}{2} = \frac{n}{2} .$$

Hence

$$E[Y|X+Y] = \frac{1}{2}(X+Y).$$

Now

$$E\big[E[Y \mid X+Y]\big] = E\Big[\frac{1}{2}(X+Y)\Big] = \frac{1}{2}(EX + EY) = EY,$$

since $EX = EY$.

Example 0.2 *Let X, Y be continuous random variables with joint pdf given by*

$$f(x,y) \;=\; \begin{cases} 6(y-x) & \text{if} \quad 0 \le x \le y \le 1 \\ 0 & \qquad \text{otherwise} \end{cases} .$$

Compute $E[Y|X], EY, E\Big[E[Y|X]\Big]$ and verify that $E\Big[E[Y|X]\Big] = EY$.

Note that

$$f_X(x) \;=\; \int_x^1 6(y-x)dy \;=\; 3(x-1)^2,\; 0 \le x \le 1$$

and $f_X(x) \;=\; 0$ elsewhere. Hence for $0 \le x < 1$,

$$\begin{aligned} E[Y|X=x] \;&=\; \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \\ &=\; \frac{2}{(x-1)^2}\int_x^1 y(y-x)dy \\ &=\; \frac{x^3-3x+2}{3(x-1)^2}. \end{aligned}$$

Also $E[Y|X=x] \;=\; 0$ elsewhere. Therefore

$$E[Y|X] \;=\; \frac{X^3-3X+2}{3(X-1)^2} = \frac{(X-1)^2(X+2)}{3(X-1)^2} = \frac{1}{3}(X+2).$$

$$EX = \frac{1}{4}, \Rightarrow \;\; E[E[Y|X]] = \frac{3}{4}.$$

$$f_Y(y) = 3y^2, 0 < y < 1.$$

Hence

$$EY = \frac{3}{4}.$$

Lemma 0.1 *Let X, Y be random variables which are either discrete or continuous with pdf and such that $f_{Y|X}, f_{X|Y}$ exists. Then*

$$E\big[E[Y \mid X]\big] \;= EY.$$

Proof: Let $g(x) = E[Y|X=x]$. Then $E[Y|X] = g(X)$. Hence

$$Eg(X) \;=\; \begin{cases} \sum_{x \in D_X} g(x)P(X=x) & \text{if} \quad X \text{ is discrete} \\ \int g(x)f_X(x)dx & \text{if} \quad X \text{ is continuous.} \end{cases}$$

When X, Y are discrete, then

$$\begin{aligned}\sum_{x \in D_X} g(x) P(X = x) &= \sum_{x \in D_X} \sum_{y \in D_Y} y f_{Y|X}(y|x) f_X(x) \\ &= \sum_{y \in D_Y} y \sum_{x \in D_X} f(x,y) \\ &= EY.\end{aligned}$$

When X is discrete and Y is continuous,

$$\begin{aligned}\sum_{x \in D_X} g(x) P(X = x) &= \sum_{x \in D_X} \int y f_{Y|X}(y|x) f_X(x) dy \\ &= \sum_{x \in D_X} \int y \frac{dG(x,y)}{dy} dy \\ &= \int y \sum_{x \in D_X} \Big(\frac{dG(x,y)}{dy} \Big) dy \\ &= \int y \frac{d}{dy} \Big(\sum_{x \in D_X} G(x,y) \Big) dy \\ &= \int y \frac{dF_Y(y)}{dy} dy = EY.\end{aligned}$$

Other cases are similar. So left as an exercise. $\square$

Now we state a couple of useful results.

Theorem 0.1 *Let X and Y be random variables and $\phi : \mathbb{R} \to \mathbb{R}$ be continuous, then*

$$E[\varphi(Y)|X = x] = \begin{cases} \sum_{y \in D_Y} \varphi(y) f_{Y|X}(y|x) & \text{if } Y \text{ is discrete} \\ \\ \int_{-\infty}^{\infty} \varphi(y) f_{Y|X}(y|x) dy & \text{if } Y \text{ is continuous} \end{cases}$$

provided the rhs converges absolutely.

Theorem 0.2 *(i) Let X, Y be random variables and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ a continuous function such that such that $E\varphi(Y)$ is finite. Then*
(i)

$$E[\varphi(X,Y)|X = x] = E[\varphi(x,Y)|X = x].$$

(ii)

$$E[\varphi(X,Y)] = \begin{cases} \displaystyle\sum_{x\in D_X} E[\varphi(x,Y)|X=x]f_X(x) & \text{if} \quad X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} E[\varphi(x,Y)|X=x]f_X(x)dx & \text{if} \quad X \text{ is continuous} \end{cases}$$
$$= E[E[\varphi(X,Y)|X]].$$

Example 0.3 *Let* $X \sim$ *uniform* $(0,\ 1)$ *and* $Y \sim$ *uniform* $(0, X)$. *Find* $E(X+Y)^2$.

Note that what is given is $f_X(x) = 1, 0 < x < 1,\ = 0$ *otherwise and the conditional pdf*

$$f_{Y|X}(y|x) = \frac{1}{x},\ 0 < y < x,\ = 0 \text{ otherwise.}$$

Hence

$$\begin{aligned} E(X+Y)^2 &= \int_0^1 E[(x+Y)^2|X=x]dx \\ &= \int_0^1 \int_{-\infty}^{\infty} (x+y)^2 f_{Y|X}(y|x)dydx \\ &= \int_0^1 \int_0^x \frac{(x+y)^2}{x} dydx \\ &= \int_0^1 \frac{1}{x} \int_x^{2x} z^2 dzdx = \frac{7}{9}. \end{aligned}$$

Finally we list some properties of conditional expectation.

Theorem 0.3 *Let* X, Y, Z *be random variables such that* $E[X|Z], E[Y|Z]$ *are defined using* $f_{X|Z}$ *and* $f_{Y|Z}$. *Then*

(1) If $X = g(Z)$, *for some* $g : \mathbb{R} \to \mathbb{R}$, *Borel measurable, then*

$$E[XY|Z] = XE[Y|Z].$$

(2) If X *is independent of* Z, *then*

$$E[X|Z] = EX.$$

Proof:

Using $X = g(Z)$,

$$\begin{aligned} E[g(Z)Y|Z=z] &= E[g(z)Y|Z=z] \\ &= g(z)E[Y|Z=z]. \end{aligned}$$

Hence

$$E[XY|Z] = g(Z)E[Y|Z] = XE[Y|Z].$$

This completes the proof of (1).

When X is independent of Z, then $f_{X|Z}$ when it exists is equal to f_X. Hence

$$E[X|Z=z] \;=\; EX.$$

(Here keep in mind that for different cases, meaning of $f_{X|Z}$ is different.) This proves (2). □

From the above the theorem following is immediate.

Lemma 0.2 *Let X, Y be random variables which falls in one the cases where $E[Y|X]$ is defined through $f_{Y|X}$. Then for each $A \in \sigma(X)$,*

$$E[I_A E[Y|X]] \;=\; E[I_A Y].$$

Proof: For $A \in \sigma(X)$, there exists $B \in \mathcal{B}_{\mathbb{R}}$ such that $A = \{X \in B\}$. Hence using Theorem 0.3, we get

$$E[I_{\{X\in B\}}Y|X] \;=\; I_{\{X\in B\}}E[Y|X].$$

Therefore

$$E[I_A Y|X] \;=\; I_A E[Y|X].$$

Hence taking mean on both sides, we get

$$E[I_A E[Y|X]] \;=\; E[E[I_A Y|X]] = E[I_A Y],$$

where the last equality follows from Theorem 0.2. This completes the proof.

□

Lemma 0.3 *If Z is a random variable which satisfing*

(i) $\sigma(Z) \subseteq \sigma(X)$

(ii)

$$E[I_A Z] = E[I_A X] \text{ for all } A \in \sigma(X).$$

Then $Z = E[Y|X]$.

Proof: The proof follows from the following result. If X, Z is a random variables such that $\sigma(Z) \subseteq \sigma(X)$ such that $E[I_A Z] = 0$ for all $A \in \sigma(X)$ and Z is with finite mean , then $P(Z = 0) = 1$.

I will give a hint for the above. Set $A_n = \{Z \leq -\frac{1}{n}\}$. Then $A_n \in \sigma(X)$ and hence $0 = -E[I_{A_n} Z] \geq E[\frac{1}{n} I_{A_n}] = \frac{1}{n} P(A_n)$. This implies $P(A_n) = 0$ for all n. Hence $P(Z < 0) = 0$.

Then above lemma indicates that $E[Y|X]$ is the only random variable satisfiying the above two properties, hence it becomes a characterizing property for conditional expectation. Hence we have the following definition.

Definition 0.3 *For the random variables X, Y, $E[Y|X]$ is defined as the random variable Z (if exists) satisfying*
(i) $\sigma(Z) \subseteq \sigma(X)$,
(ii) for each $A \in \sigma(X)$,

$$E[I_A Z] \;=\; E[I_A Y].$$

This is where 'advanced theory ' of conditional expectation and conditional probabilty begin and we stop our discussion about conditional stories here.

Chapter 9: Limit Theorems

Key words: a.s convergence, convergence in probability, weak law of large numbers, strong law of large numbers, central limit theorem.

We describe two classes of limit theorems, i.e, "law of large numbers' and "central limit theorems" in this chapter. Limit thorems describes the the asymptotic behavior of random dynamical systems. For example, Law of large numbers describes the asymptotic behavior of the average 'position' S_n at the nth epoch (here we are looking time evolution in discrete time) where as central limit theorem describes the fluctuations of S_n arround its 'average'. Here we describe the situation where 'displacements' X_n's are independent and identically distributed, i.e. when $S_n = X_1 + \cdots + X_n$, where $\{X_n | n \geq 1\}$ is a sequence of random variables.

0.1.1 Types of convergences

To describe the asymptotic behavior, for example in law of large numbers, one should define the meaning of

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} .$$

i.e. one need to talk about convergence of random variables. There are multiple ways one can define convergence of sequence of random variables.

Definition 0.4 *Let $X_n, n \geq 1, X$ be random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Then X_n is said to converges almost surely if*

$$P\big(\lim_{n\to\infty} X_n = X\big) = 1 .$$

If X_n converges to X almost surely, we write $X_n \to X$ a.s.

Definition 0.5 *Let $X_n, n \geq 1, X$ be random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Then X_n is said to converge to X in Probability, if for each $\epsilon > 0$,*

$$\lim_{n\to\infty} P\big(|X_n - X| > \epsilon\big) = 0 .$$

Definition 0.6 *Let $X_n, n \geq 1,\ X$ be random variables with distribution functions $F_n,\ n \geq 1, F$ respectively. We say that X_n converges to X in distribution if*

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ for all } x \notin D,$$

where D is the set of discontinuities of F.

Definition 0.7 *Let $X_n, n \geq 1, X$ be random variables. Then X_n is said to converge to X in mth moment if $X_n, n \geq 1,\ X$ have finite mth moments and*

$$\lim_{n\to\infty} E|X_n - X|^m \ = \ 0.$$

Let $X_n, n \geq 1, X$ be random variables. Then following relations holds.

- $X_n \to X$ a.s. $\Rightarrow X_n \to X$ in Probability .

  Recall that

  $$\Big\{ \lim_{n\to\infty} X_n = X \Big\} \ = \ \bigcap_{\varepsilon > 0} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\}.$$

  Now

  $$\begin{aligned} P\Big(\bigcap_{\varepsilon > 0} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \Big) \ = \ 1 \ &\Rightarrow \ P\Big(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \Big) \ = \ 1 \\ &\quad \text{for all } \varepsilon > 0 \\ &\Rightarrow \ \lim_{n\to\infty} P\Big(\bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \Big) \ = \ 1 \\ &\quad \text{for all } \varepsilon > 0 \\ &\Rightarrow \ \lim_{n\to\infty} P\big(|X_n - X| \leq \varepsilon \big) \ = \ 1 \\ &\quad \text{for all } \varepsilon > 0. \end{aligned}$$

  This implies convergence in probability.

- Converse is not true. i.e. convergence in probability need not imply convergence a.s.

  For example, consider $X_n, n \geq 1,\ X_n \sim$ Bernoulli $(\frac{1}{n})$ and are independent.

  For $\varepsilon > 0$,

  $$P(|X_n| > \varepsilon) = P\big(X_n = 1\big) = \frac{1}{n} \to 0 \text{ as } n \to \infty.$$

  Hence $X_n \to 0$ in probability.

  Now note that

$\sum_{n=1}^{\infty} P\big(X_n = 1\big) \;=\; \infty$. Hence using Borel-Cantelli lemma, it follows that $P\big(X_n = 1 \text{ i.o}\big) = 1$. Hence X_n doen't converge to 0 a.s.

Note that

$$\sum_{n=1}^{\infty} P\big(X_{n^2} = 1\big) \;=\; \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Hence using Borel-Cantelli lemma, we have

$$P\big(X_{n^2} = 1 \text{ i.o.}\, n\big) = 0.$$

i.e.

$$P\Big(\bigcup_n \bigcap_{m \geq n} \{X_{m^2} = 0\}\Big) \;=\; 1.$$

Hence

$$P(\lim_{n\to\infty} X_{n^2} = 0) \;=\; 1.$$

i.e., $X_{n^2} \to 0$ a.s. as $n \to \infty$. So there is a subsequence along which convergence is a.s. This is an illustration for the following result.

- $X_n \to X$ in probability $\Rightarrow$ along a subsequence $X_n \to X$ a.s.