October 6		**October 2018**
	Tutorial 10: Random Processes	
Lecturer: Utpal Mukherji/ Parimal Parag		*Prepared by: Karthik*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Contents

10.1 Conditional Expectation

We begin this lecture by recalling the definition of conditional probability of an event, given another event. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $A, B \in \mathcal{F}$ be two events, with $P(B) > 0$. Then, the conditional probability of event A, given event B, is denoted by $P(A|B)$ and is defined as

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

10.1.1 The Case of Jointly Discrete Random Variables

Suppose now that X and Y are two $\mathcal{F}$-measurable jointly discrete random variables with joint pmf $p_{X,Y}(x,y)$, $x, y \in \mathbb{R}$. Further, suppose that $E[Y] < \infty$. Then, for any two real numbers $x, y \in \mathbb{R}$, making the substitutions

$$A = \{\omega \in \Omega : Y(\omega) = y\}, \quad B = \{\omega \in \Omega : X(\omega) = x\},$$

we get the conditional probability of the event $\{Y = y\}$ given the event $\{X = x\}$ as

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p_{X,Y}(x,y)}{p_X(x)},$$

where in writing the above equation, we assume that $p_X(x) > 0$. We immediately recognize that the last term of the above equation is $p_{Y|X}(y|x)$, the conditional pmf of Y given X.

With this, the conditional expectation of Y given the event $\{X = x\}$ is denoted by $E[Y|X = x]$ and is defined as

$$\begin{aligned} E[Y|X = x] &:= \sum_{\text{all feasible } y} y \cdot P(Y = y|X = x) \\ &= \sum_{\text{all feasible } y} y \cdot p_{Y|X}(y|x). \end{aligned} \tag{10.1}$$

Remark 1. In the above definition, the summation is over the set of all values of y for which $p_{Y|X}(y|x)$ is well-defined, which may in turn depend on x.

We now present two examples.

1. *Example 1:*
   Suppose $Y = X$ in definition (10.1). Then,

$$\begin{aligned} E[Y|X = x] &= E[X|X = x] \\ &= \sum_{\text{all feasible } y} y \cdot P(X = y|X = x) \\ &= \sum_{\text{all feasible } y} y \cdot 1_{\{y = x\}} \\ &= x, \end{aligned}$$

   where the second line in the above set of equations follows by noting that $P(X = y|X = x)$ is equal to 1 if and only if $y = x$, and is zero otherwise.

2. *Example 2:*
Suppose that $x \in \{-1, 0, 1\}$. Further, let the conditional pmf of Y given X be specified as

$$p_{Y|X}(y|x) = \frac{1}{2}\, 1_{\{|y-x|=1\}}, \quad x \in \{-1, 0, 1\}.$$

Then, we have

$$\begin{aligned} E[Y|X=x] &= \sum_{\text{all feasible } y} y \cdot p_{Y|X}(y|x) \\ &= \sum_{\text{all feasible } y} y \cdot \frac{1}{2} \cdot 1_{\{|y-x|=1\}} \\ &\overset{(a)}{=} \sum_{\text{all feasible } y} y \cdot \frac{1}{2} \cdot \left(1_{\{y-x=1\}} + 1_{\{y-x=-1\}}\right) \\ &= \sum_{\text{all feasible } y} y \cdot \frac{1}{2} \cdot \left(1_{\{y=x+1\}} + 1_{\{y=x-1\}}\right) \\ &= \frac{1}{2} \cdot (x-1) + \frac{1}{2} \cdot (x+1) \\ &= x, \end{aligned}$$

where (a) above follows by noting that

(a) $|y-x| = 1$ implies that <u>either</u> $y - x = 1$ <u>or</u> $y - x = -1$ (recall that "either or" translates to union in set theory); also, these are disjoint events, and

(b) $1_{A \cup B} = 1_A + 1_B$ for any two disjoint events A and B.

3. *Example 3:*
Let X and Y be two <u>independent</u> Poisson random variables with means λ_1 and λ_2 respectively. Compute the conditional expectation of X, given that $X + Y = n$, where $n \in \{0, 1, 2, \ldots\}$.
Solution:
We first compute the conditional pmf of X, given that $X + Y = n$. Towards this, for any $k \in \{0, 1, 2, \ldots, n\}$, we have

$$\begin{aligned} P(X = k | X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\ &= \frac{P(X = k, Y = n - k)}{P(X + Y = n)} \\ &= \frac{P(X = k) \cdot P(Y = n - k)}{P(X + Y = n)}, \end{aligned}$$

where the last line above follows from independence of X and Y.
Exercise: Show that if X and Y are independent and Poisson distributed with means λ_1 and λ_2 respectively, then $X + Y$ is Poisson distributed with mean $\lambda_1 + \lambda_2$.
Using the result from the above exercise, we get

$$P(X = k | X + Y = n) = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.$$

That is, the conditional distribution of X given that $X + Y = n$ is binomial with parameters n and $\frac{\lambda_1}{\lambda_1 + \lambda_2}$. Thus, it follows that

$$E[X | X + Y = n] = n \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Exercise: What is $E[X|X+Y]$ in the above question?

In summary, we note the following:

1. $E[Y|X=x]$ is a well-defined quantity only when $P(X=x)>0$, and
2. $E[Y|X=x]$ is a function of x.

Going further, we define a new function $f:\mathbb{R}\to\mathbb{R}$ as

$$f(x) := \begin{cases} E[Y|X=x], & \text{if } P(X=x)>0, \\ 0, & \text{otherwise.} \end{cases} \tag{10.2}$$

Remark 2. It can be shown that $f(\cdot)$ as defined in (10.2) is a Borel-measurable function. However, a proof of this is beyond the scope of this course.

We then define the conditional expectation of Y given X is denoted by $E[Y|X]$ and is defined as

$$E[Y|X] := f(X).$$

Thus, we note that $E[Y|X]$, being a Borel-measurable function of X, is also an $\mathcal{F}$-measurable random variable.

Exercise: In examples 1 and 2 presented before, what is $E[Y|X]$?

10.1.2 Computing $E[Y|X]$ When Y is Continuous and X is Discrete

We now turn attention to the case when Y is a continuous random variable and X is a discrete random variable. We present the modifications necessary to define the conditional expectation $E[Y|X]$ in this case. Towards this, let $(\Omega,\mathcal{F},P)$ be a probability space, and X and Y be two $\mathcal{F}$-measurable random variables, with $E[Y]<\infty$. Further, let $F_{X,Y}(x,y)$, $x,y\in\mathbb{R}$, denote the joint CDF of X and Y. Also, let $f_Y(\cdot)$ denote the pdf of Y and let $p_X(\cdot)$ denote the pmf of X.

Remark 3. Irrespective of whether X and Y are discrete or continuous, their joint CDF is always well-defined.

In what follows, we define the conditional pdf of Y given X. Towards this, we first consider the conditional CDF of Y given $\{X=x\}$ as a function g defined below:

$$g(y|x) := \begin{cases} P(Y\le y|X=x), & \text{if } P(X=x)>0, \\ 0, & \text{otherwise.} \end{cases}$$

Assuming without loss of generality that for each $x\in\mathbb{R}$, $g(\cdot|x)$ is differentiable, let

$$h(y|x) := \frac{d}{dy}g(y|x).$$

Then, the conditional expectation of Y given the event $\{X=x\}$ is defined as

$$E[Y|X=x] := \int_{\text{all feasible } y} y\cdot h(y|x)\,dy.$$

As before, we define a new function $u : \mathbb{R} \to \mathbb{R}$ as

$$u(x) := \begin{cases} E[Y|X = x], & \text{if } P(X = x) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and note that $u(\cdot)$ is a Borel-measurable function. Then, the conditional expectation of Y given X is denoted by $E[Y|X]$ and is defined as

$$E[Y|X] := u(X).$$

Since $u(\cdot)$ is a Borel-measurable function, it follows that $E[Y|X]$ is also an $\mathcal{F}$-measurable random variable.

10.1.3 Computing $E[Y|X]$ When X and Y are Jointly Continuous

We now present the definition of conditional expectation for the case of jointly continuous random variables X and Y. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two jointly continuous random variables with joint pdf $f_{X,Y}(x, y)$, $x, y \in \mathbb{R}$, such that $E[Y] < \infty$. Further, let

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

denote the conditional pdf of Y given X, which is well-defined for all $x \in \mathbb{R}$ such that $f_X(x) > 0$. Then, for any $x \in \mathbb{R}$ satisfying $f_X(x) > 0$, the conditional expectation of Y given the event $\{X = x\}$ is denoted by $E[Y|X = x]$ and is defined as

$$E[Y|X = x] := \int_{\text{all feasible } y} y \cdot f_{Y|X}(y|x)\, dy.$$

Defining $v : \mathbb{R} \to \mathbb{R}$ as

$$v(x) := \begin{cases} E[Y|X = x], & \text{if } f_X(x) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and noting that $v(\cdot)$ is a Borel-measurable function, the conditional expectation of Y given X is denoted by $E[Y|X]$ and is defined as

$$E[Y|X] := v(X).$$

Since $v(\cdot)$ is a Borel-measurable function, it follows that $E[Y|X]$ is also an $\mathcal{F}$-measurable random variable.

Example:
Let X and Y have the joint pdf given by

$$f_{X,Y}(x, y) = \begin{cases} cx(y - x)e^{-y}, & 0 \le x \le y < \infty, \\ 0, & \text{otherwise,} \end{cases}$$

where $c \in \mathbb{R}$ is a normalizing constant to make the above a joint pdf. It can be easily verified that

$$f_{Y|X}(y|x) = (y - x)e^{x-y}, \quad 0 \le x \le y < \infty.$$

Therefore, we get

$$\begin{aligned} E[Y|X=x] &= \int_{\text{all feasible } y} y \cdot f_{Y|X}(y|x)\,dy \\ &= \int_x^\infty y\cdot (y-x)e^{x-y}\,dy \\ &= e^x \left(\int_x^\infty y^2\, e^{-y}\,dy - x\int_x^\infty y\, e^{-y}\,dy \right) \\ &= x+2, \end{aligned}$$

whence it follows that $E[Y|X] = X + 2$.

Exercise: For this question, evaluate $E[X|Y]$.

10.1.4 The Case When Y is Discrete and X is Continuous

In this subsection, we present the formula for $E[Y|X = x]$ for the case when Y is a discrete random variable and X is a continuous random variable. Our approach will be to arrive at a function similar to $h(y|x)$ as presented in Section 10.1.2, and thereafter use it to compute $E[Y|X = x]$.

Towards this, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable functions with joint CDF $F_{X,Y}(x, y)$, $x, y \in \mathbb{R}$, such that $E[Y] < \infty$. For any $x, y \in \mathbb{R}$, let

$$i(x,y) := \begin{cases} P(X \le x, Y = y), & \text{if } P(Y=y) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

denote the joint probability of the events $\{X \le x\}$ and $\{Y = y\}$. Assuming without loss of generality that for each $y \in \mathbb{R}$, $i(\cdot, y)$ is differentiable for all feasible values of x, we define a new function $q(y|x)$ as

$$q(y|x) := \frac{1}{f_X(x)} \frac{d}{dx} i(x,y)$$

whenever $f_X(x) > 0$. Then, the expectation of Y given the event $\{X = x\}$ is denoted by $E[Y|X = x]$ and is defined as

$$E[Y|X=x] := \sum_{\text{all feasible } y} y \cdot q(y|x)$$

for x such that $f_X(x) > 0$. We may then define $w : \mathbb{R} \to \mathbb{R}$ as

$$w(x) = \begin{cases} E[Y|X=x], & \text{if } f_X(x) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Noting that $w(\cdot)$ is a Borel-measurable function, the conditional expectation of Y given X is defined as

$$E[Y|X] := w(X).$$

Exercise:

Let X be Gaussian distributed with mean 0 and variance 1. Suppose that

$$q(y|x) = \frac{1}{2}\, 1_{\{|y-\text{sgn}(x)|=1\}}, \quad x \in \mathbb{R},$$

where $\text{sgn}(x)$ denotes the sign of x, defined as

$$\text{sgn}(x) := \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

What is $E[Y|X]$?

Remark 4. Each of the above definitions for $E[Y|X = x]$ can be extended for $E[b(Y)|X = x]$, where $b : \mathbb{R} \to \mathbb{R}$ is any Borel measurable function. These extensions are as follows:

1. For the case when X and Y are jointly discrete, we have

$$E[b(Y)|X = x] = \sum_{\text{all feasible } y} b(y)\, p_{Y|X}(y|x).$$

2. For the case when Y is continuous and X is discrete, we have

$$E[b(Y)|X = x] = \int_{\text{all feasible } y} b(y)\, h(y|x)\, dy,$$

   where $h(y|x)$ is as defined in Section 10.1.2.

3. For the case when X and Y are jointly continuous, we have

$$E[b(Y)|X = x] = \int_{\text{all feasible } y} b(y)\, f_{Y|X}(y|x)\, dy.$$

4. For the case when X is continuous and Y is discrete, we have

$$E[b(Y)|X = x] = \sum_{\text{all feasible } y} b(y)\, q(y|x),$$

   where $q(y|x)$ is as defined in Section 10.1.4.

Remark 5. In each of the four cases (for X being discrete or continuous, Y being discrete or continuous) presented above, if X and Y are independent, then $E[Y|X] = E[Y]$. (Check this!)

Remark 6. In each of the four cases presented above, if Y is a Borel-measurable function of X, say $Y = t(X)$, then $E[Y|X] = Y = t(X)$. (Check this!) Thus, for example, $E[X^2|X] = X^2$, $E[e^{3X}|X] = e^{3X}$ and so on.

10.2 The Law of Iterated Expectations

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables. The previous sections dealt with the definitions for $E[Y|X]$ for the individual cases when X is discrete (or continuous) and Y is discrete (or continuous). Below, we present an important result in probability theory known as the law of iterated expectations (or the law of total expectations) that states that expectations may be computed in an iterative manner using conditional expectations. The formal statement is as follows:

Theorem 10.2.1. *(Law of iterated expectations) Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables. If $E[Y] < \infty$, then*

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

Remark 7. Notice that $E[Y|X]$ is a function of X. Therefore, the outer expectation in the above theorem is with respect to the distribution of X.

Remark 8. The specific form of $E[E[Y|X]]$ for the cases when X and Y are either discrete or continuous is as follows:

1. For the case when X and Y are jointly discrete, or when X is discrete and Y is continuous, we have

$$E[E[Y|X]] = \sum_{x:p_X(x)>0} E[Y|X=x] \cdot p_X(x).$$

2. For the case when X and Y are jointly continuous, or when X is continuous and Y is discrete, we have

$$E[E[Y|X]] = \int_{x:f_X(x)>0} E[Y|X=x] \cdot f_X(x)\, dx.$$

The proof of the above theorem is left as exercise.

Example:
Let $(\Omega, \mathcal{F}, P)$ be a probability space and let Y be geometric with parameter $p \in (0,1)$. Evaluate $E[Y]$ using iterated expectations.

Solution:

Recall that

$$P(Y > k) = (1-p)^k, \quad k \in \{1, 2, \ldots\}.$$

Define a set A and a random variable X as follows:

$$A = \{\omega \in \Omega : Y(\omega) = 1\}, \quad X = 1_A.$$

Then, by the law of iterated expectations, we have

$$\begin{aligned} E[Y] &= E[E[Y|X]] \\ &= \sum_{x=0}^{1} E[Y|X=x] \cdot p_X(x) \\ &= E[Y|X=0] \cdot p_X(0) + E[Y|X=1] \cdot p_X(1) \end{aligned}$$

We now note the following:

1. $p_X(1) = P(X=1) = P(A) = P(Y=1) = p = 1 - p_X(0)$.
2. $X = 1$ if and only if $1_A = 1$, which is true if and only if event A has occurred, which in turn is true if and only if $Y = 1$. Therefore, $\{X = 1\} = \{Y = 1\}$, and it follows that

$$E[Y|X=1] = E[Y|Y=1] = 1.$$

3. $X = 0$ if and only if $1_A = 0$, which is true if and only if event A has not occurred, which in turn is true if and only if $Y > 1$. Therefore, $\{X = 0\} = \{Y > 1\}$, and it follows that

$$E[Y|X = 0] = E[Y|Y > 1] = 1 + E[Y - 1|Y > 1].$$

We now claim that $E[Y - 1|Y > 1] = E[Y]$. Indeed, by the memoryless property of geometric distribution, for any $k \in \{1, 2, \ldots\}$, we have

$$P(Y - 1 > k|Y > 1) = P(Y > k + 1|Y > 1) = (1 - p)^k,$$

from which it follows that conditioned on the event $\{Y > 1\}$, the random variable $Y - 1$ is distributed geometrically with parameter p, hence proving the claim. Using the above facts, we get

$$E[Y] = p + (1 - p)(1 + E[Y]) = 1 + (1 - p)E[Y],$$

from which we get $E[Y] = 1/p$.

10.3 Miscellaneous Examples

1. Let X be exponentially distributed with parameter $\lambda = 1$. Compute $E[X|X > 1]$.

   Solution:

   The idea is to first compute the conditional pdf of X given that $X > 1$ (which we shall denote by $f_{X|X>1}(x)$). Using this conditional pdf, we may then compute the desired conditional expectation.

   Towards computing the conditional pdf of X given $X > 1$, we note that

$$\begin{aligned} P(X \le x, X > 1) &= \begin{cases} 0, & x \le 1, \\ P(1 < X \le x), & x > 1 \end{cases} \\ &= \begin{cases} 0, & x \le 1, \\ e^{-1} - e^{-x}, & x > 1. \end{cases} \end{aligned}$$

   Thus, we notice that

$$P(X \le x|X > 1) = \frac{P(X \le x, X > 1)}{P(X > 1)} = \begin{cases} 0, & x \le 1, \\ 1 - e^{1-x}, & x > 1. \end{cases}$$

   Differentiating the above expression, we get the conditional pdf of X given the event $\{X > 1\}$ as

$$f_{X|X>1}(x) = \begin{cases} 0, & x \le 1, \\ e^{1-x}, & x > 1. \end{cases}$$

   With this, we get

$$\begin{aligned} E[X|X > 1] &= \int_{-\infty}^{\infty} x\, f_{X|X>1}(x)\, dx \\ &= e \int_{1}^{\infty} x\, e^{-x}\, dx \\ &= e \cdot \frac{2}{e} \\ &= 2. \end{aligned}$$

Exercise: Show that when X is exponentially distributed with parameter $\lambda > 0$,

$$E[X|X>1] = \frac{1}{\lambda} + 1.$$

Exercise: Let U be uniformly distributed over $[-1, 1]$. Evaluate $E[U|U>0]$.

2. Let X and Y be jointly uniformly distributed over a right-angled triangle with vertices at $(0,0)$, $(1,0)$ and $(0,2)$. Evaluate $E[X|Y>1]$.

   Solution:
   The idea is to first compute the conditional pdf of X given that $Y > 1$ (which we shall denote by $f_{X|Y>1}(x)$). Using this conditional pdf, we may then compute the desired conditional expectation.

   We note that the joint pdf of X and Y may be expressed mathematically as

$$f_{X,Y}(x,y) = \begin{cases} 1, & 0 \le x \le 1, \quad 0 \le y \le 2-2x, \\ 0, & \text{otherwise.} \end{cases}$$

   We now observe that if $Y > 1$, the X takes values only in $[0, 0.5)$ with nonzero probability. Thus, we get

$$P(X \le x, Y > 1) = \begin{cases} 0, & x \le 0, \\ \int\limits_0^x \int\limits_1^{2-2t} f_{X,Y}(t,y)\,dy\,dt, & 0 < x < \dfrac{1}{2}, \\ \int\limits_0^{0.5} \int\limits_1^{2-2t} f_{X,Y}(t,y)\,dy\,dt, & x \ge \dfrac{1}{2}. \end{cases}$$

$$= \begin{cases} 0, & x \le 0, \\ x - x^2, & 0 < x < \dfrac{1}{2}, \\ \frac{1}{4}, & x \ge \dfrac{1}{2}, \end{cases}$$

   from which we get

$$P(X \le x|Y>1) = \frac{P(X \le x, Y > 1)}{P(Y>1)} = \begin{cases} 0, & x \le 0, \\ 4x(1-x), & 0 < x < \dfrac{1}{2}, \\ 1, & x \ge \dfrac{1}{2}. \end{cases}$$

   We arrive at the conditional pdf of X given the event $\{Y > 1\}$ as

$$f_{X|Y>1}(x) = \frac{d}{dx} P(X \le x|Y>1) = \begin{cases} 4(1-2x), & 0 < x < \dfrac{1}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

   Using this, we have

$$E[X|Y>1] = \int\limits_{-\infty}^{\infty} x \cdot f_{X|Y>1}(x)\,dx = \int\limits_0^{0.5} 4x(1-2x)\,dx = \frac{1}{6}.$$

3. Suppose $X_1, X_2, \ldots$ are iid random variables with $E[X_1] < \infty$. Suppose N is another random variable independent of X_n for all $n \in \mathbb{N}$, taking values in $\{1, 2, \ldots\}$, such that $E[N] < \infty$. Then, show that

$$E\left[\sum_{n=1}^{N} X_n\right] = E[N] \cdot E[X_1].$$

Remark 9. Note that the summation on the left hand side of the above equation is a "random" sum, i.e., the upper limit N of summation is a random quantity. Thus, we cannot blindly pass the expectation inside this random sum. If we do so, then we get

$$E\left[\sum_{k=1}^{N} X_k\right] = \sum_{k=1}^{N} E[X_k] = N \cdot E[X_1].$$

Now, the left hand side of the above equation is the expectation of some quantity which is a <u>deterministic</u> constant, whereas the right hand side is $N \cdot E[X_1]$ which is a <u>random</u> variable!

Solution:
Let

$$S_N := \sum_{k=1}^{N} X_k.$$

We compute $E[S_N]$ using iterated expectations as

$$E\left[S_N\right] = E\left[E\left[S_N | N\right]\right].$$

For any n such that $P(N = n) > 0$, we have

$$\begin{aligned}
E[S_N|N = n] &= E\left[\sum_{k=1}^{N} X_k \middle| N = n\right] \\
&= E\left[\sum_{k=1}^{n} X_k \middle| N = n\right] \\
&\overset{(a)}{=} \sum_{k=1}^{n} E\left[X_k | N = n\right] \\
&\overset{(b)}{=} \sum_{k=1}^{n} E\left[X_k\right] \\
&= n \cdot E[X_1],
\end{aligned}$$

where (a) above follows by noting that the upper limit n of the summation is a deterministic number, and (b) follows from the independence of X_n's and N. Thus, we get

$$\begin{aligned}
E[S_N] &= E[E[S_N|N]] \\
&= \sum_{n:P(N=n)>0} E[S_N|N = n] \cdot P(N = n) \\
&= \sum_{n:P(N=n)>0} n \cdot E[X_1] \cdot P(N = n) \\
&= E[X_1] \cdot \sum_{n:P(N=n)>0} n \cdot P(N = n) \\
&= E[N] \cdot E[X_1].
\end{aligned}$$

10.4 Geometric Interpretation for Conditional Expectation

We now give a geometric interpretation for the quantity $E[Y|X]$. For the purpose of this section, we assume that all random variables we work with have finite variance (or equivalently, finite second moment). Specifically, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables such that $E[X^2] < \infty$ and $E[Y^2] < \infty$.

We now think of a scenario where Y is unknown and the goal is to estimate Y by observing X. For example, Y could be temperature at a place, and X could be humidity, and the goal is to estimate the temperature with the knowledge of humidity.

Suppose that we know the conditional distribution of Y given X. For example, say for a given value of humidity $x > 0$, the temperature Y is exponentially distributed with parameter x. Note that this suffices to compute $E[Y|X]$. We prove in this section the following key statement:

"Among all estimates of a random variable Y that may be obtained by observing another random variable X, the "best" estimate is $E[Y|X]$, where the notion of "best" is in terms of minimizing the mean-squared error in estimation."

Towards proving this, we first state and prove the following theorem.

Theorem 10.4.1. *Let (Ω, F, P) be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables with $E[X^2] < \infty$ and $E[Y^2] < \infty$. Then, for any Borel measurable function $\psi : \mathbb{R} \to \mathbb{R}$ such that $E[(\psi(X))^2] < \infty$, we have*

$$E[(Y - E[Y|X])\psi(X)] = 0,$$

or equivalently,

$$E[Y\psi(X)] = E[E[Y|X]\psi(X)].$$

Interpretation:
Consider the problem of estimating the unknown Y by observing X. The best we can hope to do to estimate Y from X is to construct a function of X, say $\psi(X)$ (assume that $\psi(X)$ has finite variance), and use this as an estimate for Y. We note that $\psi(X)$ and $E[Y|X]$ are both functions of the observation X.

The above theorem then states that among all possible estimates that we may construct for Y by observing X, the estimate $E[Y|X]$ is such that the $Y - E[Y|X]$ is orthogonal to any other estimate $\psi(X)$, where the notion of orthogonality is that U and V are orthogonal if $E[UV] = 0$.

Proof. We note that by Cauchy-Schwartz inequality,

$$E[Y\psi(X)] \leq \sqrt{E[Y^2]}\sqrt{E[(\psi(x))^2]} < \infty,$$

and thus $E[Y\psi(X)|X]$ is well-defined. The proof of the theorem then follows from the following steps:

$$\begin{aligned} E[Y\psi(X)] &= E[E[Y\psi(X)|X]] \\ &= E[\psi(X)E[Y|X]], \end{aligned}$$

where the last line follows by noting that given X, $\psi(X)$ may be treated like a constant, and therefore

$$E[Y\psi(X)|X] = \psi(X)E[Y|X].$$

□

An immediate consequence of the above theorem is the following result which formally states that $E[Y|X]$ is the minimum mean-squared error estimate (abbreviated as MMSE estimate) of Y among all possible finite variance estimates $\psi(X)$ that may be obtained from observing X.

Theorem 10.4.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let Y be an $\mathcal{F}$-measurable random variable with $E[Y^2] < \infty$. Then,*

$$E[(Y - E[Y|X])^2] \leq E[(Y - \psi(X))^2]$$

for any Borel measurable function ψ such that $E[(\psi(X))^2] < \infty$.

Remark 10. The quantity $E[(Y-\psi(X))^2]$ for any function $\psi(X)$ denotes the mean-squared error in estimating Y by using $\psi(X)$ as an estimate for Y based on X.

Proof. The proof follows from the following steps:

$$\begin{aligned} E[(Y - \psi(X))^2] &= E[((Y - E[Y|X]) + (E[Y|X] - \psi(X)))^2] \\ &= E[(Y - E[Y|X])^2] + \underbrace{E[(E[Y|X] - \psi(X))^2]}_{\geq 0} + 2E[(Y - E[Y|X])(X - \psi(X))] \\ &\geq E[(Y - E[Y|X])^2] + 2E[(Y - E[Y|X])(X - \psi(X))]. \end{aligned}$$

We now note that

$$E[(X - \psi(X))^2] = E[X^2] + E[(\psi(X))^2] - 2E[X\psi(X)],$$

and by Cauchy-Schwartz inequality, the last term on the right hand side of the above equation is finite, hence implying that $X - \psi(X)$ is a finite variance function of X. Thus, applying theorem 10.4.1 with $\psi(X)$ replaced by $X - \psi(X)$, we get

$$E[(Y - E[Y|X])(X - \psi(X))] = 0,$$

therefore implying that

$$E[(Y - \psi(X))^2] \geq E[(Y - E[Y|X])^2].$$

□

10.5 Exercises

1. Let X and Y be jointly continuous with joint pdf given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}\, y e^{-xy}, & 0 < x < \infty, \quad 0 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

   Compute $E[e^{X/2}|Y = 1]$.

2. Let X and Y have joint pdf given by

$$f(x,y) = \begin{cases} 3y, & -1 \leq x \leq 1, \ 0 \leq y \leq |x|, \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Compute $P(X \geq Y + 0.5)$.
   (b) Evaluate $E[Y|X \geq Y + 0.5]$.
   (c) Evaluate $E[Y|X]$ and verify the relation $E[Y] = E[E[Y|X]]$ for this problem.