EE5110: Probability Foundations for Electrical Engineers **July-November 2015**

Lecture 23: Conditional Expectation

Lecturer: Dr. Krishna Jagannathan *Scribe: Sudharsan Parthasarathy*

Let X and Y be discrete random variables with joint probability mass function $p_{X,Y}(x,y)$, then the conditional probability mass function was defined in previous lectures as

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

assuming $p_Y(y) > 0$. Let us define

$$\mathbb{E}[X|Y=y] = \sum_x x p_{X|Y}(x|y).$$

$\psi(y) = \mathbb{E}[X|Y=y]$ changes with y. The random variable $\psi(Y)$ is the conditional expectation of X given Y and denoted as $\mathbb{E}[X|Y]$.

Let X and Y be continuous random variables with joint probability density function $f_{X,Y}(x,y)$. Recall the conditional probability density function

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

when $f_Y(y) > 0$. Define

$$\mathbb{E}[X|Y=y] = \int_x x f_{X|Y}(x|y)\mathrm{d}x.$$

The random variable $\psi(Y)$ is the conditional expectation of X given Y and denoted as $\mathbb{E}[X|Y]$.

Example 1: Find $\mathbb{E}[Y|X]$ if the joint probability density function is $f_{X,Y}(x,y)=\frac{1}{x}$; $0 < y \leq x \leq 1$.

Solution: $f_X(x)=\int_0^x \frac{1}{x}dy=1$, $0 \leq x \leq 1$

$f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{x}$, $0 < y \leq x$

$\mathbb{E}[Y|X=x] = \int_0^x y f_{Y|X}(y|x)\mathrm{d}y = \int_0^x \frac{y}{x}\mathrm{d}y = \frac{x}{2}$

The conditional expectation $\mathbb{E}[Y|X] = \frac{X}{2}$.

Theorem 23.1 *Law of Iterated Expectation:*

$$\mathbb{E}[Y] = \mathbb{E}_X[\mathbb{E}[Y|X]].$$

Proof: We prove the result for discrete random variables. We have

$$\begin{aligned}
\mathbb{E}_X[\mathbb{E}[Y|X]] &= \sum_x p_X(x)\mathbb{E}[Y|X=x] \\
&= \sum_x p_X(x) \sum_y y p_{Y|X}(y|x) \\
&= \sum_x p_X(x) \sum_y y \frac{p_{X,Y}(x,y)}{p_X(x)} \\
&= \sum_{x,y} y p_{X,Y}(x,y) \\
&= \sum_y y \sum_x p_{X,Y}(x,y) \\
&= \sum_y y p_Y(y) \\
&= \mathbb{E}[Y].
\end{aligned}$$

■

Similarly law of iterated expectation for jointly continuous random variables can also be proved.

Application of the law of iterated expectation:

$S_N = \sum_{i=1}^{N} X_i$, where $\{X_1, ...X_N\}$ are independent and identically distributed random variables. N is a non-negative random variable independent of X_i $\forall$ $i \in \{1, ..N\}$. From the law of iterative expectation, $\mathbb{E}[S_N] = \mathbb{E}_N[\mathbb{E}[S_N|N]]$. Consider

$$\mathbb{E}[S_N|N=n] = \mathbb{E}\left[\sum_{i=1}^{N} X_i|N=n\right] \tag{23.1}$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} X_i|N=n\right]. \tag{23.2}$$

As N is independent of X_i, $\mathbb{E}\left[\sum_{i=1}^{n} X_i|N=n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}[X]$.

Thus $\mathbb{E}[S_N|N] = N\mathbb{E}[X]$, $\mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[X]$.

Theorem 23.2 *Generalized form of Law of Iterated Expectation:*

For any measurable function g with $\mathbb{E}[|g(X)|] < \infty$,

$$\mathbb{E}[Yg(X)] = \mathbb{E}[\mathbb{E}[Y|X]g(X)].$$

Proof: We prove the result for discrete random variables. We have

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]g(X)] &= \sum_x p_X(x)\mathbb{E}[Y|X=x]g(x) \\ &= \sum_x p_X(x)g(x)\sum_y y p_{Y|X}(y|x) \\ &= \sum_x p_X(x)g(x)\sum_y y\frac{p_{X,Y}(x,y)}{p_X(x)} \\ &= \sum_{x,y} yg(x)p_{X,Y}(x,y) \\ &= \mathbb{E}[Yg(X)].\end{aligned}$$

■

Exercise: Prove $\mathbb{E}[Yg(X)] = \mathbb{E}[\mathbb{E}[Y|X]g(X)]$ if X and Y are jointly continuous random variables.

This theorem implies that

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])g(X)] = 0. \tag{23.3}$$

The conditional expectation $\mathbb{E}[Y|X]$ can be viewed as an estimator of Y given X. $Y - \mathbb{E}(Y|X)$ is then the *estimation error* for this estimator. The above theorem implies that the estimation error is uncorrelated with every function of X.

Observe that in this lecture, we have not dealt with conditional expectations in a general framework. Instead, we have separately defined it for discrete and jointly continuous random variables. In a more general development of the topic, (23.3) is in fact taken as the defining property of the conditional expectation. Specifically, for any $g(X)$, one can prove the existence and uniqueness (up to measure zero) of a $\sigma(X)$-measurable random variable $\psi(X)$, that satisfies $\mathbb{E}[(\psi(X) - Y)g(X)] = 0$. Such a $\psi(X)$ is then defined as the conditional expectation $\mathbb{E}[Y|X]$. For a more detailed discussion, refer Chapter 9 in [1].

Minimum Mean Square Error Estimator:

We have seen that $\mathbb{E}[Y|X]$ is an estimator of Y given X. In the next theorem we will prove that this is indeed an optimal estimate of Y given X, in the sense that the conditional expectation minimizes the mean-squared error.

Theorem 23.3 *If $\mathbb{E}(Y^2) < \infty$, then for any measurable function g,*

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])^2] \leq \mathbb{E}[(Y - g(X))^2].$$

Proof:

$$\begin{aligned}\mathbb{E}[(Y - g(X))^2] &= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + \mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2] + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))] \\ &\geq \mathbb{E}[(Y - \mathbb{E}[Y|X])^2].\end{aligned}$$

This is because $\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))]$ =0 (by (23.3)), and $\mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2] \geq 0$.

$\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))]$ =0 as from (23.3) we know that $\mathbb{E}[(\mathbb{E}[Y|X] - Y)\psi(X)] = 0$. Here $\psi(X) = (\mathbb{E}[Y|X] - g(X))$. ■

From (23.3) we observe that the estimation error $Y - (\mathbb{E}[Y|X)]$ is orthogonal to any measurable function of X. In the Hilbert Space of square integrable random variables, $\mathbb{E}[Y|X]$ can be viewed as the projection of Y onto the subspace $\mathcal{L}_2(\sigma(X))$ of $\sigma(X)$ measurable random variables. As depicted in Figure 23.1, it is quite intuitive that the conditional expectation (which is the projection of Y onto the subspace) minimizes the mean-squared error among all random variables from the subspace $\mathcal{L}_2(\sigma(X))$.

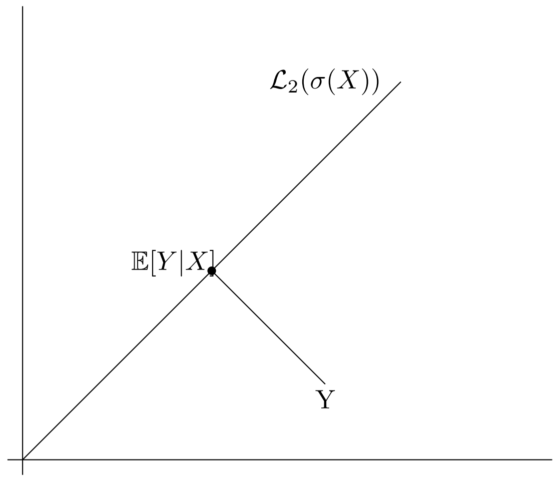

Figure 23.1: Geometric interpretation of MMSE

23.1 Exercises

1. Prove the law of iterated expectation for jointly continuous random variables.
2. (i) Given is the table for Joint PMF of random variables X and Y.

	X=0	X =1
Y=0	$\frac{1}{5}$	$\frac{2}{5}$
Y=1	$\frac{2}{5}$	0

Let $Z = \mathbb{E}[X|Y]$ and $V = Var(X|Y)$. Find the PMF of Z and V, and compute $\mathbb{E}[Z]$ and $\mathbb{E}[V]$.

(ii) Consider a sequence of i.i.d. random variables $\{Z_i\}$ where $\mathbb{P}(Z_i = 0) = \mathbb{P}(Z_i = 1) = \frac{1}{2}$. Using this sequence, define a new sequence of random variables $\{X_n\}$ as follows:
$X_0 = 0$,
$X_1 = 2Z_1 - 1$, and
$X_n = X_{n-1} + (1 + Z_1 + ... + Z_{n-1})(2Z_n$ - 1) for $n \geq 2$.
Show that $\mathbb{E}[X_{n+1}|X_0, X_1, ..., X_n] = X_n$ a.s. for all n.

3. (a) *[MIT OCW problem set]* The number of people that enter a pizzeria in a period of 15 minutes is a (nonnegative integer) random variable K with known moment generating function $M_K(s)$. Each person who comes in buys a pizza. There are n types of pizzas, and each person is equally likely to choose any type of pizza, independently of what anyone else chooses. Give a formula, in terms of $M_K(.)$, for the expected number of different types of pizzas ordered.

   (b) John takes a taxi to home everyday after work. Every evening, he waits by the road to get a taxi but every taxi that comes by is occupied with a probability 0.8 independent of each other . He counts the number of taxis he missed till he gets an unoccupied taxi. Once he gets inside the taxi, he throws a fair six faced die for a number of times equal to the number of taxis he missed. He counts the output of the die throws and gives a tip to the driver equal to that. Find the expected amount of tip that John gives everyday.

References

[1] D. Williams, "Probability with Martingales",*Cambridge University Press*, Fourteenth Printing, 2011.