Lesson 9

Conditional Expectation

"Conditional Probability Revisited"

Remember that the conditional pmf of Y given X is

$$P_{Y|X}(y|x) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(X=x)} = \mathbb{P}(Y=y|X=x)$$

$$= \frac{P_{X,Y}(x,y)}{P_X(x)}$$

More generally, the conditional pmf of the discrete r.v. Y on $(\Omega, \mathcal{F}, \mathbb{P})$ given $A \in \mathcal{F}$

is defined as:

$$p_{Y|A}(y) =$$

provided that $\mathbb{P}(A) \boxed{\quad}$

Def: Assume that Y is a discrete r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ and $A \in \mathcal{F}$ and $\mathbb{P}(A) > 0$. The conditional expectation of Y given A is defined as:

$$\mathbb{E}[Y|A] =$$

In particular, if A is the event $\{X=x\}=\{\omega \mid X(\omega)=x\}$

then

$$\mathbb{E}[Y \mid A] = \mathbb{E}[Y \mid X=x]$$

$$=$$

The Law of The Unconscious Statistician (~~L~~otus) for Conditional Expectation

If X and Y are discrete r.v.'s, and $g: \mathbb{R} \to \mathbb{R}$

is a function, then

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

$=$

Example (p.58, Lesson 7)

The joint pmf of X, Y is given:

$$p_{X,Y}(x,y) = \begin{cases} \frac{2}{n(n+1)}\left(\frac{x}{x+1}\right)^y & x=0,\dots,n-1 \\ & y=0,1,2,\dots \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathbb{E}[Y \mid X = x]$.

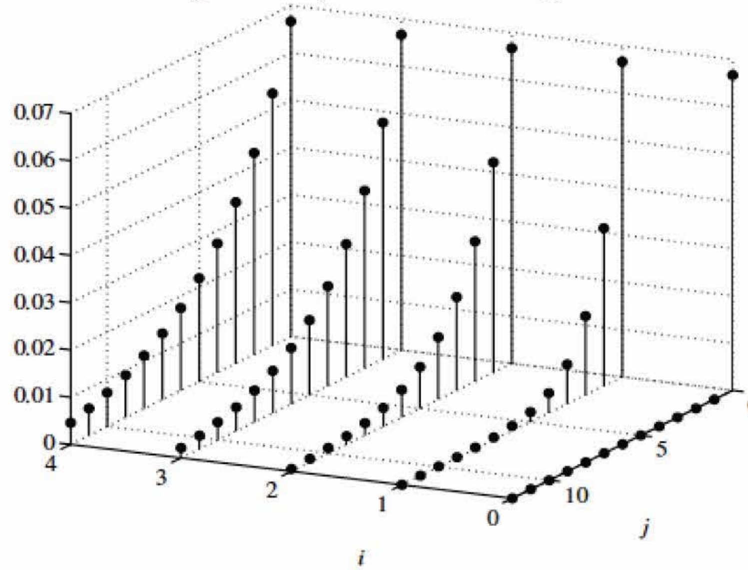


***Figure* 3.3.** Sketch of bivariate probability mass function $p_{XY}(i,j)$ of Example 3.10 with $n=5$. For fixed i, $p_{XY}(i,j)$ as a function of j is proportional to $p_{Y|X}(j|i)$, which is $\text{geometric}_0(i/(i+1))$. The special case $i=0$ results in $p_{Y|X}(j|0) \sim \text{geometric}_0(0)$, which corresponds to a constant random variable that takes the value $j=0$ with probability one.

We saw that $Y \mid X = x \sim \text{Geo}_0\left(\frac{1}{x+1}\right)$

i.e. $P_{Y|X}(y|x) = \left(\frac{1}{x+1}\right)\left(\frac{x}{x+1}\right)^y$

Then $\mathbb{E}[Y \mid X = x] =$

Substitution Law for

Conditional Expectation

Recall that

$$\mathbb{P}(g(X,Y)=z \mid X=x)$$

$$=\mathbb{P}(g(x,Y)=z \mid X=x)$$

The conditional law of the unconscious statistician states that

$$\mathbb{E}[g(X,Y) \mid X=x]$$

$$=$$

Therefore

$$\mathbb{E}[\,g(X,Y) \mid X=x\,] = \mathbb{E}[\,g(x,Y) \mid X=x\,]$$

which is the substitution law for

Conditional expectation.

Exercise: If $g(x,y) = f(x)h(y)$,

what is $\mathbb{E}[g(X,Y) \mid X=x]$?

The Law of Total Expectation

We saw that $\mathbb{E}[g(X,Y) \mid X=x]$

is a function of x, $e(x)$.

We can use LOTUS to

calculate $\mathbb{E}[e(X)]$:

$$\mathbb{E}[e(X)] = \sum_x e(x) P_X(x)$$

but $e(x) = \mathbb{E}[g(X,Y) \mid X = x]$

$$= \sum$$

therefore,

$$\sum_x e(x) P_X(x) =$$

That means

$$\mathbb{E}[g(X,Y)] = \sum_{x} \mathbb{E}[g(X,Y) \mid X=x] P_X(x)$$

$$=$$

$$\mathbb{E}[e(X)]$$

where

$$e(x) = \mathbb{E}[g(X,Y) \mid X=x]$$

If g is only a function of Y:

$$\mathbb{E}[g(Y)] = \sum_{x} \mathbb{E}[g(Y) \mid X=x] P_X(x)$$

Corollary:

$$\sum_{x} \mathbb{E}[Y \mid X = x]\, P_X(x) = \mathbb{E}[Y]$$

Proof: $g(Y) = Y$ in the law of total expectation.

Example: Light of intensity λ is directed at a photomultiplier that generates X primaries, where $X \sim \text{Pois}(\lambda)$. The photomultiplier generates Y

secondaries, where given $X=x$, Y is $\mathrm{Bin}(x+1, 1/2)$. Find the expected number of secondaries and the correlation between X and Y, defined as $E[XY]$.

Total Expectation for Partitions of Sample Space

Assume that $\{B_i \mid i \in I\}$ is a partition of the sample space Ω and I is countable.

Also, assume that $\forall i,\ B_i \in \mathcal{F}$.

Define
$\omega \in B_i \Leftrightarrow X(\omega) = i$

Then X is a r.v. (why?)

$$P_X(i) = \mathbb{P}(X = i) = \mathbb{P}\left(\{\omega \mid X(\omega) = i\}\right)$$

$=$

X is a discrete r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$.

$\mathbb{E}[\cancel{X} \mid X = i] =$

Applying the law of total expectation:

$\mathbb{E}[Y] = \sum_i$

Example: Two manufacturer of light bulbs compete in Los Angeles. M_1's light bulbs work for an average of 6,000 hours and M_2's light bulbs work for

an average of 3,000 hours. M_1 supplies 30% of the light bulbs in L.A. and M_2 supplies 70% of the light bulbs in L.A. What is the average life of a light bulb in L.A.?

L = Life Time of A light Bulb

$\mathbb{E}[L] =$

The Conditional Expectation as A Random Variable

Def:
Let X and Y be two discrete r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ and define

$$e : \mathbb{R} \to \mathbb{R}$$
$$e(x) = \mathbb{E}[Y \mid X = x\}$$

$e(X)$ is a r.v.; which is called the Conditional Expectation of Y given X

$$\mathbb{E}[Y \mid X].$$

Example: Roll a die until a 6 is observed. Let X be the total number of the rolls, and let Y be the number of 1's that are observed.

Find $\mathbb{E}[Y|X=x]$ and $\mathbb{E}[Y|X]$ and try to interprete $\mathbb{E}[Y|X]$

Conditioning on

Multiple Random Variables

Let X, Y, Z be discrete r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$. Conditioning on the event $A = \{\omega \mid X(\omega) = x\} \cap \{\omega \mid Y(\omega) = y\}$

results in

$$\mathbb{E}[Z \mid A] = \mathbb{E}[Z \mid Y = y, X = x]$$

Denote $\mathbb{E}[Z \mid Y = y, X = x] = e(x, y)$.

Then

$e(X,Y)$ is a r.v. ~~...~~ that is called the conditional expectation of Z given X and Y and is denoted as $\mathbb{E}[Z|X,Y]$.

Remark: Because usually more than one r.v. is involved in calculating expectations and conditional expectations, sometimes a subscript is used to show the distribution with respect to which the expectation is calculated:

$$\mathbb{E}_{Y|X}[g(X)h(Y)|X] = \sum_{y} g(x)h(y)P_{Y|X}(y|x)$$

$$\mathbb{E}_{Y}[g(X)h(Y)] = \sum_{y} g(x)h(y)P_{Y}(y)$$

$$\mathbb{E}_{X,Y}[g(X)h(Y)] = \sum_{x,y} g(x)h(y)P_{X,Y}(x,y)$$

Exercise: Calculate $\mathbb{E}[Z \mid X=x, Y=y]$ using conditional pmfs.

Properties of Conditional Expectation

Theorem:
Assume that X, Y, and Z are ~~d.~~ discrete r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ and $a, b \in \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$.

a) $\mathbb{E}[aX|X] =$

b) $\mathbb{E}[aY+bZ|X] =$

c) If $Y \geq 0$ a.s. then

$\mathbb{E}[Y|X]$ $\boxed{}$ a.s.

d) If X and Y are independent

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

(e) (The Tower Property) The Law of Iterated Expectations

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

or in general

f) (Total Expectation)

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

Remark: Using the subscripts,

$$\mathbb{E}_{\square}[\mathbb{E}_{\square}[Y|X]] = \mathbb{E}_{\square}[Y]$$

and

$$\mathbb{E}_{\square}[\mathbb{E}_{\square}[g(Y)|X]] = \mathbb{E}_{\square}[g(Y)]$$

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

Remark: (g) means that the operator $\mathbb{E}[\cdot|X]$ treats functions of X as constants.

h) $\mathbb{E}[g(X)|X] = g(X)$

(i) $\mathbb{E}[Y|X, g(X)] = \mathbb{E}[Y|X]$

Proof:

Exercise: Show the following property of conditional expectation

$$\mathbb{E}_{Z|Y}\left[\ \mathbb{E}_{X|Y,Z}[X|Y,Z]\ \middle|\ Y\right] = \mathbb{E}[X|Y]$$

Note: It may be useful to calculate the RHS for $Y=y$.

Note: You may want to check that, intuitively, both RHS and LHS are functions of Y.

Existence of Conditional Expectation

Lemma: If X is a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ and X is integrable or non-negative,

$\mathbb{E}[|Y| \mid X=x]$ is finite, for all x for which $P_X(x) \neq 0$.

Proof:

Remark: The converse is not necessarily true, i.e. there are cases for which $\mathbb{E}[|Y| \mid X=x]$ is finite for all x, but $\mathbb{E}[|Y|]=\infty$.

Exercise: Find an example for the above Remark

Conditional Probability as Conditional Expectation

Recall that the probability $\mathbb{P}(X \in B)$ can be written as $\mathbb{E}[I_B(X)]$.

Similarly:

$$\mathbb{P}(Y \in B \mid x) = \mathbb{E}_{\boxed{}}$$

Proof: Left as an exercise for X, Y discrete.

Remark: In fact, the concept of conditional expectation provides a rigorous foundation for defining probability. $\mathbb{E}[\cdot|X]$ is more fundamental than $\mathbb{P}(\cdot|X=x)$.
(c.f. Probability with Martingales by Williams)

Exercise (The Law of Iterated Expectation/ Total Expectation)

Assume that $X \sim dU(0,5)$

and $Y|X=x \sim Pois(x+1)$.

Calculate $E[Y^2 X]$.

Wald's Equality:

Let N be a random variable that takes non-negative integer values. Let $X_1, X_2, \ldots$ be a sequence of i.i.d discrete random variables

that have finite expectation and are independent from N. Wald Equality states that

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] =$$

Higher Order Conditional Moments

Def: For all $r \in \mathbb{N}$,

$\mathbb{E}[Y^r | X]$ is called the

r^{th} moment of Y given X.
(conditional)

Def: For all $r \in \mathbb{N}$,

$$\mathbb{E}\left[\left(Y - \mathbb{E}[Y|X]\right)^r \middle| X\right]$$

is called the r^{th} conditional central moment of Y given X.

Remark: In particular

$$\text{Var}(Y|X) = E\left[\left(Y - E[Y|X]\right)^2 \middle| X\right]$$

is called the conditional variance of Y given X

and $\sigma_{Y|X} = \sqrt{\text{Var}(Y|X)}$ is called the conditional standard deviation of Y given X.

Remark: Conditional Variance inherits the properties of variance. For example:

Lemma) $Var(Y|X) = E[$

Proof:

Example: Roll a die until a 6 is observed. Let X be the total number of the rolls, and let Y be the number of 1's that are observed. Find $Var(Y|X)$.

Exercise: Let X and Y be independent random variables and $X \sim dU(0,3)$ and $Y \sim dU(0,3)$. Let $Z = X+Y$. Find $\mathbb{E}[Z|X]$, $\mathbb{E}[X|Z]$, $\mathbb{E}[XZ|X]$, $\mathbb{E}[XZ|Z]$.

Projections, The Principle of Orthogonality, and The Projection Theorem

Def: Let C be a subset of the inner product vector space $\mathcal{V}$.

For some $X \in \mathcal{V}$, we call $\hat{X} \in C$ a projection of X onto C if

$$\|X-\hat{X}\| \leq \|X-Y\| \quad \forall Y \in C$$

Note: $\hat{X}$ can be viewed as an estimate of X, and $\|X-\hat{X}\|$

can be seen as ~~a~~ the estimation error.

The Principle of Orthogonality

Let S be a subspace of V.
(i.e. S is itself a vector space and is a subset of V).

The "estimate" $\hat{X} \in S$ of $X \in V$ is ~~the~~ a minimum error estimate of X, i.e.

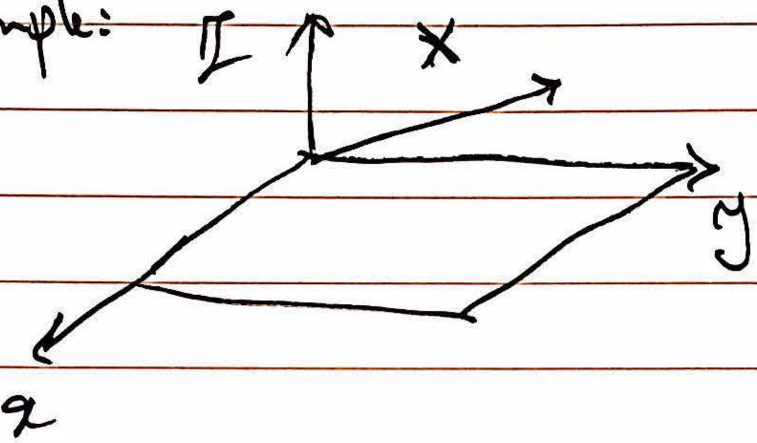
$$\|X - \hat{X}\| \leq \|X - Y\| \quad \forall Y \in S$$

if and only if

$$\langle X - \hat{X}, Y \rangle = 0 \quad \forall Y \in S$$

Remark: The principle of Orthogonality means for $\hat{X}$ an estimate of X to be the "minimum error" estimate, the estimation error has to be orthogonal to the subspace in which the estimate resides.

Example:



$S = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \middle| \, z = 0 \right\}$

S is a subspace

What is the "minimum error" estimate of X, $\hat{X} \in S$?

Remark: The Principle of Orthogonality does NOT guarantee the existence of $\hat{X}$. The Projection Theorem guarantees the existence of $\hat{X}$.

Projection Theorem: If S is a "closed" subspace of $\mathcal{V}$, and $X \in \mathcal{V}$ then there exists a unique $\hat{X}$ such that

$$\| X - \hat{X} \| \leq \| X - Y \| \quad \forall Y \in S$$

Remark:

Def: S is called a closed set if any sequence X_n in S converges, it converges to a member of S, i.e. if

$\lim_{n \to \infty} \| X_n - X \| = 0$ then

$X \in S$.

Remark: One can see that we used the notion of "norm" to define convergence of sequences more generally.

Conditional Expectation as An Estimator

Theorem: Let $g: \mathbb{R} \to \mathbb{R}$ be a function and $X, Y, g(X)$ are random variables

Then: $\mathbb{E}[\mathbb{E}[Yg(X)|X]] =$

$$\mathbb{E}[\mathbb{E}[Y|X]\, g(X)]$$

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

Remark: when $g(X)=1$, we have the tower property.

Proof

Remark: So far, we have proved the results involving conditional expectation ~~and~~ and expectation for discrete r.v's. Later on, we see that most of them are valid for other types of random variables.

Interpretation:

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

$$\Rightarrow \mathbb{E}\left[\left(\mathbb{E}[Y|X] - Y\right) g(X)\right] = 0$$

One can view the set of all $g(X)$'s as a subspace of the space of random variables. Because $\mathbb{E}[Y|X]$ is a function of X, it is a member of this subspace.

Let us think of $\mathbb{E}[Y|X]$

as an estimate of Y, so

$\hat{Y} = \mathbb{E}[Y|X]$, then:

$$\mathbb{E}[(\hat{Y} - Y)g(X)] = 0$$

We learned that $\mathbb{E}[Y_1 Y_2]$ is

the inner product between Y_1 and

Y_2, therefore:

$$\mathbb{E}[(\hat{Y} - Y)g(X)] = 0$$

$\Rightarrow$

$$\langle \hat{Y} - Y, g(X) \rangle = 0$$

In other words, the estimation error $\hat{Y} - Y$ is orthogonal to all possible $g(X)$'s, which consist a subspace.

By the projection theorem

$\mathbb{E}[Y|X]$ is the minimum error estimate of Y in the subspace of $g(X)$'s, i.e.

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

We can prove ~~this~~ it for this particular case $\hat{Y} = \mathbb{E}[Y|X]$.

Theorem: Suppose that $\mathbb{E}[X^2]<\infty$, where X is a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. ~~and~~ For any $g:\mathbb{R}\to\mathbb{R}$ for which $g(X)$ is a random variable

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

Remark: The right hand side of the above equation is called the Mean-Squared Error (MSE) of the estimator $\mathbb{E}[Y|X]$.

Consequently, $\mathbb{E}[Y|X]$ is called the Minimum Mean-Squared Error Estimate (MMSE) of Y.

Proof:

A More Rigorous Definition of Conditional Expectation

Definition: Assume that $f: \Omega_1 \to \Omega_2$ is a function and $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are measurable spaces (i.e. $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-fields). We say f is $\mathcal{F}_1, \mathcal{F}_2$ measurable if the pre-image of any $B \in \mathcal{F}_2$

under f is a measurable set in $\mathcal{F}_1$, i.e.

$$\forall B \in \mathcal{F}_2, \quad \overleftarrow{f}(B_2) \in \mathcal{F}_1$$

Remark: A random variable

$$X: \Omega \to \mathbb{R}$$

is a $\mathcal{B}, \mathcal{F}$ measurable function, because $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $(\Omega, \mathcal{F})$ are measurable spaces and

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \overleftarrow{X}(B) \in \mathcal{F}.$$

Def.(Conditional Expectation of Y given X):

Assume that X, Y are r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$. A random variable of the form $e(X)$,

where $e: \mathbb{R} \to \mathbb{R}$ is a measurable function is called a conditional expectation of Y given X iff for all measurable functions $g: \mathbb{R} \to \mathbb{R}$:

Note:

(a) The existence of $e(X)$ is guaranteed if Y is integrable. (We proved it for discrete r.v.'s)

(b) Any two functions e_1 and e_2 that satisfy the above definition are equal with probability one:

$$\mathbb{P}(\{\omega \mid e_1(X(\omega)) = e_2(X(\omega))\}) = 1$$

Therefore, when using "=" sign for conditional expectation, we must remember that we imply equality with probability one. In other words, Conditional expectation of Y given X is actually a class of r.v.'s that are equal with probability 1.

For Further Discussion, see:
Probability with Martingales, by Williams.