Feb 3, 2016

- Last
  - Expectation
- Today
  - Conditional expectation
  - MMSE estimation

**Conditional expectation**

Let \((X, Y) \sim f_{X,Y}(x,y)\)

Recall that:

\[f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{(if } f_Y(y) > 0)\]

Consider the expectation \(g(X, Y)\) conditioned on \(Y = y\) (w.r.t. \(f_{X|Y}(x|y)\))

\[
E[g(X,Y)|Y=y] = \int g(x,y) f_{X|Y}(x|y) \, dx
\]

\[= \phi(y) \quad \text{(some function of } y)\]

**Examples**

1. \(E[X|Y=y] = \int x f_{X|Y}(x|y) \, dx\)
2. \(E[Y|Y=y] = \int y f_{X|Y}(x|y) \, dx = y \int f_{X|Y}(x|y) \, dx = y\)
3. \(E[XY|Y=y] = y E[X|Y=y] = y \int x f_{X|Y}(x|y) \, dx\)

Let \(\phi(y) = E[g(X,Y)|Y=y]\)

Then the conditional expectation of \(g(X, Y)\) given \(Y\) is defined as:

\[E[g(X,Y)|Y] = \phi(Y)\]

**Properties**

1. \(E[g(X,Y)|Y]\) is a function of \(Y\) and hence is a r.v. itself.
2. Law of iterated expectation:

\[
E[E[g(X,Y)|Y]] = \int \phi(y) f_Y(y) \, dy
\]

\[= \int \int g(x,y) f_{X|Y}(x|y) \, dx f_Y(y) \, dy
\]

\[= \int \int g(x,y) f_{X,Y}(x,y) \, dx \, dy = E[g(X,Y)]\]
Examples

(i) Let \( f_{X,Y}(x,y) = \begin{cases} 2, & x, y > 0, \, x+y \leq 1 \\ 0, & \text{otherwise} \end{cases} \)

(a) Find \( E(X|Y) \)
First, recall that \( X|\{Y=y\} \sim U(0,1-y) \)
Therefore: \( E(X|Y=y) = \frac{1-y}{2} \), \( 0 \leq y \leq 1 \)
Hence: \( E(X|Y) = \frac{1-y}{2} \)

(b) Find the pdf of \( E(X|Y) \). Let \( Z = E(X|Y) = \frac{1-y}{2} \)
Recall that: \( f_y(y) = \)
\[
\begin{array}{c}
\text{2} \\
\text{0} \\
\text{1} \\
y
\end{array}
\]
Hence: \( f_z(z) = \)
\[
\begin{array}{c}
\text{4} \\
\text{0} \\
\text{1/2} \\
z
\end{array}
\]
\( F_z(z) = P(\{Z \leq z\}) = P(\{\frac{1-y}{2} \leq z\}) = P(\{Y \geq 1-2z\}) \)
\( = 1 - F_y(1-2z) \Rightarrow f_z(z) = 2(1-2z) \)

(c) Compare \( E(X) \) and \( E(Z) \)
\( E(X) = \int_0^1 x(2(1-x)) \, dx = \int_0^1 2x \, dx = \frac{2x^3}{3} \bigg|_0^1 = \frac{1}{3} \)
\( E(Z) = \int_0^{1/2} z(\frac{1}{z}) \, dz = \frac{8}{3} z^3 \bigg|_0^{1/2} = \frac{1}{3} \)
\( E(Z) = E(E(X|Y)) = E(X) \)

(d) Find \( E(XY) \)
\( E(XY|Y=y) = y E(X|Y=y) = \frac{y(1-y)}{2} \)
Hence: \( E(XY|Y) = \frac{Y(1-Y)}{2} \)
By law of iterated expectation:
\[
E(xy) = E[E(xy|y)] = E\left(\frac{y(1-y)}{2}\right)
\]
\[
= \int_0^1 y(1-y) \times 2(1-y) dy = \int_0^1 y(1-y)^2 dy = \frac{1}{12} \quad \text{(Check!)}
\]

(2) Let \(P \sim f_p(p) = \begin{cases} 2(1-p) & p \in [0,1] \\ 0 & \text{otherwise} \end{cases} \)

Given \(\{P=p\} \), \(N \sim \text{Binom}(n,p)\).

(Experiment: Pick a bias of a coin at random according to \(f_p(p)\). Then flip the coin \(n\) times and count the number of heads as \(N\).)

Find \(E[N]\).

Observe: \(E[N|P=p] = np\)

Therefore: \(E[N] = E[E(N|P)] = E(np) = nE[P] = \frac{n}{3}\).

(3) Let \(Y \sim U[0,1]\) and \(E(X|Y) = y^2\).

Then: \(E(X) = E(Y^2) = \int_0^1 y^2 dy = \frac{1}{3}\).

**Conditional Variance**

The (conditional) variance of \(X\) w.r.t. \(f_{x|y}(x,y)\) is:

\[
\text{var}(X|Y=y) = E\left((X - E(X|Y=y))^2|Y=y\right)
\]

\[
= E[X^2|Y=y] - (E[X|Y=y])^2
\]

check!

Similarly:

The conditional variance of \(X\) given \(Y\) is:

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

**Properties**

(1) \(\text{var}(X|Y)\) is a function of \(Y\) and is a r.v. itself.

(2) \(\text{var}(X) = E[\text{var}(X|Y)] + \text{var}(E(X|Y))\)
MMSE (Minimum Mean Squared Error) estimation

\[ X \xrightarrow{\text{noisy channel}} Y \xrightarrow{\text{estimate}} \hat{X} = g(Y) \]

Goal: find \( g^*(y) \) such that: \( E[(X-\hat{X})^2] = E[(X-g(Y))^2] \) is minimized.

Let \( X \sim f_X(x) \).

Problem:
- Find \( a^* \) such that \( E[(X-a)^2] \) is minimized.

Answer: \( a^* = E[X] \).

Proof: \( E[(X-a)^2] = E[(X-EX+EX-a)^2] = E[(X-EX)^2] + (EX-a)^2 \)

\[ = \text{var}(X) + (EX-a)^2 \geq \text{var}(X) \] (MSE)

Theorem: \( g^*(y) = E[X|Y=y] \) and the resulting mean squared error \( \hat{g}(x) \) is:

\[ E[	ext{var}(X|Y)] \]

Proof: Consider any \( g(y) \) then:

\[ E[(X-g(y))^2|Y=y] \geq E[(X-E(X|Y=y))^2|Y=y] \]

Taking expectation w.r.t. \( Y \) on both sides and using the law of iterated expectation:

\[ E[(X-g(Y))^2] \geq E[(X-E(X|Y))^2] = E[	ext{var}(X|Y)] \]

Properties (\( \hat{X} = E(X|Y) \))

1. \( E(\hat{X}) = E[X] \) (Check!)
   In other words, it is unbiased.
2. \( E[\hat{X}-X|Y=y] = 0 \) for all \( y \).
3. If \( X \) and \( Y \) are independent, then:
   \( \hat{X} = E(X) \)
4. Orthogonality: For any function \( g(y) \):
   \( X-\hat{X} \) is orthogonal to \( g(Y) \)
   \[ i.e., \quad E[(X-\hat{X})g(Y)] = 0 \]
Proof: Consider $E[(X-\hat{X})g(Y) | Y=y] = g(y)E[X-\hat{X} | Y=y] = 0$
Taking expectation w.r.t. $Y$ to conclude the proof.

(5) $E[(X-\hat{X})^2] = E[(X-\hat{X})X] - E[(X-\hat{X})\hat{X}]$
    $= E[X^2] - E[X\hat{X}]$
    $\frac{E[XE(X|Y)]}{E[X]} = E[E(XE(X|Y) | Y)]$
    $= E[X^2] - E[E(X|Y)^2] = E[E(X|Y))^2]