Conditional Expectation

Let $X$ be a random variable and let $Y$ be a random vector.

Conditional Distributions review

If $X$ and $Y$ are both discrete then we can describe the conditional distribution of $X$ given $Y=y$ by the conditional pmf of $X$ given $Y=y$. This conditional pmf gives the conditional probabilities $P(X=x \mid Y=y)$. We denote this conditional pmf by $p_{X \mid Y}(x \mid y)$.

So, we have

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

where $p_{X,Y}(x,y)$ is the joint pmf of $X$ and $Y$, and $p_Y(y)$ is the marginal pmf of $Y$.

In the continuous case we have a similar expression. If $X$ and $Y$ are both continuous, the conditional pdf of $X$ given $Y=y$ is denoted by $f_{X \mid Y}(x \mid y)$ and given by

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

where $f_{X,Y}(x,y)$ is the joint pdf of $X$ and $Y$ and $f_Y(y)$ is the marginal pdf of $Y$. (We are assuming that these joint and marginal pdfs exist).

Conditional Means and Variances

The conditional mean of $X$ given $Y=y$, denoted by $E[X \mid Y=y]$, is the mean of the conditional distribution of $X$ given $Y=y$. That is,
\[
E(X \mid Y = y) = \begin{cases} 
\sum_{x} x \ p_{x \mid y}(x \mid y) & \text{(discrete case)} \\
\int_{-\infty}^{\infty} x \ f_{x \mid y}(x \mid y) \, dx & \text{(continuous case)}
\end{cases}
\]

Similarly, the conditional variance of \(X\) given \(Y = y\), denoted by \(\text{Var}(X \mid Y = y)\), is the variance of the conditional distribution of \(X\) given \(Y = y\):

\[
\text{Var}(X \mid Y = y) = E(X^2 \mid Y = y) - E^2(X \mid Y = y)
\]

\[
= \begin{cases} 
\sum_{x} x^2 \ p_{x \mid y}(x \mid y) - \left( \sum_{x} x \ p_{x \mid y}(x \mid y) \right)^2 & \text{(discrete case)} \\
\int_{-\infty}^{\infty} x^2 f_{x \mid y}(x \mid y) \, dx - \left( \int_{-\infty}^{\infty} x f_{x \mid y}(x \mid y) \, dx \right)^2 & \text{(cont's case)}
\end{cases}
\]

(we are assuming that \(E(X \mid Y = y)\) and \(\text{Var}(X \mid Y = y)\) exist for every \(y\) in the support of \(Y\).

Note that both \(E(X \mid Y = y)\) and \(\text{Var}(X \mid Y = y)\) are functions of \(y\). We can view them as functions of the random vector \(Y\), and write \(E(X \mid Y)\) and \(\text{Var}(X \mid Y)\), where \(E(X \mid Y)\) is the function of \(Y\) whose value when \(Y = y\) is \(E(X \mid Y = y)\), and \(\text{Var}(X \mid Y)\) is the function of \(Y\) whose value when \(Y = y\) is \(\text{Var}(X \mid Y = y)\).

Remark: Although \(E(X \mid Y = y)\) and \(E(X \mid Y)\) look similar notationally, they are quite different mathematical objects; \(E(X \mid Y = y)\) is a number and \(E(X \mid Y)\) is a random variable . Similarly, \(\text{Var}(X \mid Y = y)\) is a number while \(\text{Var}(X \mid Y)\) is a random variable.

Law of Total Expectation:

\[
\text{Theorem: } E[E(X \mid Y)] = E[X]
\]

w.r.t. dist'n of \(Y\) \quad w.r.t. dist'n of \(X\)
Proof (discrete case)

\[
E[E(X \mid Y)] = \sum_y E[X \mid Y=y] \cdot p_Y(y) \\
= \sum_y \sum_x x \cdot p_{X,Y}(x,y) \cdot p_Y(y) \\
= \sum_y \sum_x x \cdot \frac{p_{X,Y}(x,y)}{p_Y(y)} \cdot p_Y(y) \\
= \sum_x \sum_y p_{X,Y}(x,y) \\
= \sum_x p_X(x) \\
= E[X]
\]

Example Let \( X \) have a Geometric(\( \rho \)) distribution.

\( X \) is the first trial where the first success occurs.

Let \( Y = \begin{cases} 1 & \text{if the first trial is a success} \\ 0 & \text{if the first trial is a failure} \end{cases} \)

By the Law of Total Expectation,

\[
E[X] = E[X \mid Y=1] \cdot P(Y=1) + E[X \mid Y=0] \cdot P(Y=0) \\
= (1) \cdot \rho + E[X \mid Y=0] \cdot (1-\rho)
\]

Given that the first trial is a failure, then we can write

\( X = 1 + X_1 \), where \( X_1 \) is the further number of trials required, starting at trial 2, until the first success. But \( X_1 \) has a Geometric(\( \rho \)) distribution, so

\[
E[X \mid Y=0] = E[1 + X_1] = 1 + E[X_1] = 1 + E[X] \\
\]

Then

\[
E[X] = \rho + (1 + E[X]) \cdot (1-\rho)
\]

or

\[
E[X](1 - (1-\rho)) = \rho + 1-\rho \\
\]

or

\[
p \cdot E[X] = 1 \\
\]

or

\[
E[X] = \frac{1}{\rho} 
\]