Conditional Probabilities and Expectations as Random Variables

H. Krieger, Mathematics 157, Harvey Mudd College

Spring, 2005

Let X and Y be random variables defined on the same probability space and suppose that Y has a *discrete* distribution. Let R_Y be the *essential range* of Y, *i.e.* $\{y \in \mathbf{R} : P(Y = y) > 0\}$. Note that R_Y is either a finite or countably infinite subset of $\mathbf{R}$. Then if $a \in \mathbf{R}$ and $y \in R_Y$, we define the *conditional distribution function*

$$F(a|y) = P(X \leq a|Y = y) = P(X \leq a, Y = y)/P(Y = y).$$

Since the probability of an event is the same as the expected value of the indicator random variable for that event, we can also think of conditional probabilities in terms of conditional expectations. Namely, if we let $g_a(X) = I_{(-\infty,a]}(X)$, then we can write

$$E(g_a(X)|Y = y) = P(X \leq a|Y = y),$$

and call this the *conditional expectation* of $g_a(X)$ given $Y = y$ for $y \in R_Y$.

A new point of view: Suppose we fix $a \in \mathbf{R}$ and let y vary over R_Y. Then the values of $F(a|y)$ define a new function $F(a|\cdot)$ on R_Y. In fact we can extend this function to $\mathbf{R}$ by defining its value to be 0 on the complement of R_Y. Consequently, if we compose this new function with the random variable Y, we get a new *random variable* which we can write as

$$F(a|Y) = \sum_{y \in R_Y} F(a|y) I_{\{y\}}(Y).$$

Note that for $0 \leq p \leq 1$ the event

$$\{F(a|Y) = p\} \cap \{Y \in R_Y\} = \{Y \in \bigcup \{y \in R_Y : F(a|y) = p\}\}$$

and, hence, $F(a|Y)$ is a *discrete* random variable taking values in the interval $[0, 1]$. We refer to this random variable as the conditional probability that $X \leq a$ given Y or as the conditional expectation of $g_a(X)$ given Y.

Generalizing these ideas: Suppose $Z = g(X, Y)$ is a random variable, defined on the same sample space as X and Y, such that $E(|Z|) < \infty$. Examples of such random variables are the indicators $g_a(X)$ described above and, more interestingly, $g_Y(X) = I_{X \le Y}$. Then, for $y \in R_Y$, we define the conditional expectation

$$E(Z|y) = E(ZI_{\{y\}}(Y))/P(Y = y).$$

Once again, we can think of this conditional expectation as a function $E(Z|\cdot)$ defined on R_Y and extended to $\mathbf{R}$. So, composing this function with Y, we get a *random variable*

$$E(Z|Y) = \sum_{y \in R_Y} E(Z|y) I_{\{y\}}(Y)$$

which we call the conditional expectation of Z given Y. Note that for $r \in \mathbf{R}$ the event

$$\{E(Z|Y) = r\} \cap \{Y \in R_Y\} = \{Y \in \bigcup \{y \in R_Y : E(Z|y) = r\}\},$$

so that $E(Z|Y)$ is a discrete random variable.

Here are the two most important special cases:

1. Suppose X also has a discrete distribution with essential range R_X. Define the conditional probability mass function $p_{X|Y}(x|y) = P(X = x|Y = y)$ for $x \in R_X$ and $y \in R_Y$. Then

$$E(g(X,Y)|y) = \sum_{x \in R_X} g(x,y) p_{X|Y}(x|y).$$

2. Suppose that for each $y \in R_Y$ there is a conditional probability density function $f_{X|Y}(\cdot|y)$ such that for $a \in \mathbf{R}$

$$P(X \le a|Y = y) = \int_{-\infty}^{a} f_{X|Y}(x|y)\, dx.$$

Then

$$E(g(X,Y)|y) = \int_{\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\, dx.$$

Properties of Conditional Expectations

1. Conditional expectation, given the discrete random variable Y, is a linear operator on the vector space of random variables that are defined on the same sample space as Y and have finite expected values. This operator maps the entire vector space into the subspace consisting of random variables that are functions of Y.

2. The expected value of the conditional expectation random variable is the same as the expected value of the random variable on which the conditioning is made, *i.e.*

$$E((E(Z|Y)) = E(Z).$$

To see this, recall that since the random variable $E(Z|Y)$ is a function of the random variable Y, its expectation can be computed using the distribution of Y. Namely,

$$
\begin{aligned}
E(E(Z|Y)) &= \sum_{y \in R_Y} E(Z|y)P(Y = y) \\
&= \sum_{y \in R_Y} E(ZI_{\{y\}}(Y)) \\
&= E(Z \sum_{y \in R_Y} I_{\{y\}}(Y)) \\
&= E(Z)
\end{aligned}
$$

since $\sum_{y \in R_Y} I_{\{y\}}(Y) = I_{R_Y}(Y) = 1$ with probability one. Note that interchanging the sum and expectation is allowed since $E(|Z|) < \infty$.

3. If h is a bounded function defined on R_Y and is extended to $\mathbf{R}$ as above, then with probability one

$$E(h(Y)Z|Y) = h(Y)E(Z|Y).$$

First, to show that $h(Y)Z$ has finite expectation, observe that since h is bounded on R_Y, there exists a positive number M such that with probability one, $|h(Y)| \leq M$. Therefore, $E(|h(Y)Z|) \leq ME(|Z|) < \infty$. Next, if $y \in R_Y$, then

$$
\begin{aligned}
E(h(Y)Z|y) &= E(h(Y)ZI_{\{y\}}(Y))/P(Y = y) \\
&= h(y)E(ZI_{\{y\}}(Y))/P(Y = y) \\
&= h(y)E(Z|y).
\end{aligned}
$$

Since $P(Y \in R_Y) = 1$, the result follows. We conclude that the linear operator of conditional expectation with respect to Y treats bounded functions of Y as if they were scalars.

4. The previous results show that for every function h which is bounded on R_Y,

$$E(h(Y)E(Z|Y)) = E(h(Y)Z).$$

We claim that $E(Z|Y)$ is, up to a set of probability zero, the *unique* function of the random variable Y with this property. For suppose that $\mathcal{E}(Y)$ is another such function. Then, letting $h = I_{\{y\}}$ for $y \in R_Y$, we see that

$$E(\mathcal{E}(Y)I_{\{y\}}(Y)) = \mathcal{E}(y)P(Y = y) = E(ZI_{\{y\}}(Y)).$$

Therefore, $\mathcal{E}(y) = E(Z|y)$ for every $y \in R_Y$ and, hence, $\mathcal{E}(Y) = E(Z|Y)$ with probability one.

Further generalizations: If the distribution of Y is not discrete, it is still possible to define the conditional expectation operator with respect to Y in such a way that all of the properties listed above will hold. In particular, $E(Z|Y)$ is again a function of the random variable Y, defined uniquely up to a set of probability zero, and has expected value equal to $E(Z)$. Thus, for example, if the distribution of Y is described by a probability density function f_Y, we can compute the expected value of $E(Z|Y)$ and, hence, the expected value of Z by using this density function. The computation goes as follows:

$$E(Z) = E(E(Z|Y)) = \int_{-\infty}^{\infty} E(Z|y) f_Y(y) dy.$$

The most important special case of this situation occurs when $Z = g(X, Y)$ with the joint distribution of X and Y described by a joint probability density function $f_{X,Y}$. Observe that on the one hand we have

$$P(X \leq a) = \int_{-\infty}^{a} f_X(x)\, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{a} f_{X,Y}(x, y)\, dx\, dy$$

and, on the other hand,

$$P(X \leq a) = \int_{-\infty}^{\infty} P(X \leq a|Y = y) f_Y(y)\, dy,$$

Consequently,

$$P(X \leq a|Y = y) = \int_{-\infty}^{a} \frac{f_{X,Y}(x, y)}{f_Y(y)}\, dx$$

so that the conditional distribution of X given $Y = y$ is described by the conditional density function defined, whenever $f_Y(y) > 0$, by

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

Consequently,

$$E(g(X, Y)|y) = \int_{\infty}^{\infty} g(x, y) f_{X|Y}(x|y)\, dx$$

whenever $f_Y(y) > 0$.

The existence of the conditional expectation in general is guaranteed by a famous result from the theory of measure and integration called the Radon-Nikodym Theorem. In the special case that $Z = g_a(X) = I_{(-\infty,a]}(X)$, we continue to call the resulting random variable the conditional *probability* that $X \leq a$ given Y.