5.3: Conditional Distributions

Taylor

University of Virginia

Motivation

"The distribution of Y can depend strongly on the value of another variable X. For example, if X is height and Y is weight..."

We want something like $P(A|B)$, but for random variables instead of events.

Sometimes we "know" one and not the other, so we can't really use the joint density.

Definition

Let X and Y be two discrete random variables with joint pmf $p(x, y)$ and marginal $p_X(x)$. Then, for any x such that $p_X(x) > 0$, the **conditional probability mass function** of Y given $X = x$ is

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

Let X and Y be two cts rvs with joint pdf $f(x, y)$ and marginal (for X) $f_X(x)$. Then, for any x such that $f_X(x) > 0$, the **conditional probability density function** of Y given $X = x$ is

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

Definition

We also have the conditional expectation

$$E[Y|X = x] = \int_{-\infty}^{\infty} y \; f_{Y|X=x}(y|x)dy$$

or

$$E[Y|X = x] = \sum_{y} y \; p_{Y|X=x}(y|x)$$

Definition

This is more general. LOTUS for conditional distributions. For any function $h(\cdot)$

$$E[h(Y)|X = x] = \int_{-\infty}^{\infty} h(y)\, f_{Y|X=x}(y|x)dy$$

or

$$E[h(Y)|X = x] = \sum_{y} h(y)\, p_{Y|X=x}(y|x)$$

Definition

One more definition before an example:

The **conditional variance** of Y given $X = x$ is

$$V(Y|X = x) = E\left\{[Y - E(Y|X = x)]^2|X = x\right\}$$

Definition

One more definition before an example:

The **conditional variance** of Y given $X = x$ is

$$V(Y|X = x) = E\left\{[Y - E(Y|X = x)]^2 | X = x\right\}$$

There are two conditional expectations here. First, we have $[Y - E(Y|X = x)]^2$

Then we take the conditional expectation of that, too.

Proposition

We have this result again, too

$$V(Y|X = x) = E[Y^2|X = x] - (E[Y|x = x])^2$$

Example

Check out the table on page 253. This is example 5.18. Find the conditional distribution of Y given $X = 250$

1. $p(Y = 0|X = 250) = \frac{p_{X,Y}(250{,}0)}{p_X(250)} = \frac{.05}{(.05+.15+.3)}$
2. $p(Y = 100|X = 250) = \frac{p_{X,Y}(250{,}100)}{p_X(250)} = \frac{.15}{(.05+.15+.3)}$
3. $p(Y = 200|X = 250) = \frac{p_{X,Y}(250{,}200)}{p_X(250)} = \frac{.30}{(.05+.15+.3)}$

Example

Check out the table on page 253. This is example 5.18. Find the conditional distribution of Y given $X = 250$

1. $p(Y = 0|X = 250) = \frac{p_{X,Y}(250,0)}{p_X(250)} = \frac{.05}{(.05+.15+.3)}$
2. $p(Y = 100|X = 250) = \frac{p_{X,Y}(250,100)}{p_X(250)} = \frac{.15}{(.05+.15+.3)}$
3. $p(Y = 200|X = 250) = \frac{p_{X,Y}(250,200)}{p_X(250)} = \frac{.30}{(.05+.15+.3)}$

Now find the conditional expectation of Y given $X = 250$

$$E[Y|X = x] = \sum_y y\ p_{Y|X=250}$$

$$= 0\frac{.05}{.05 + .15 + .3} + 100\frac{.15}{.05 + .15 + .3} + 200\frac{.30}{.05 + .15 + .3}$$

Independence

Remember how we said X and Y were independent if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$? If we divide both sides $f_X(x)$, then we get something in terms of a conditional density: $\frac{f_{X,Y}(x,y)}{f_X(x)} = f_Y(y)$, or

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

In words, that means Y does not depend on X.

Proposition

If we fix $X = x$, then $E[Y|X = x]$ is a number. However, this can also be an rv (it's a transformation of X, right?).
Depending on what I mean, I will either write $E[Y|X = x]$ (a function of a specific non-random number x) or $E[Y|X]$ (a transformation of X and still random).

Proposition

“The Law of Total Expectation” and “The Law of Total Variance.”

$$E[Y] = E[E(Y|X)]$$

$$V(Y) = V[E(Y|X)] + E[V(Y|X)]$$

Motivation

BE CAREFUL WHAT WEIGHTS YOU'RE USING. NOT ALL Es ARE THE SAME!

proof for the first one in the cts case:

$$\begin{aligned}
E[E(Y|X)] &= E\left[\int y f_{Y|X=x}(y|x)dy\right] \\
&= \int \left[\int y f_{Y|X=x}(y|x)dy\right] f_X(x)dx \\
&= \int \int y f_{Y|X=x}(y|x) f_X(x) dydx \\
&= \int \int y \frac{f_{X,Y}(x,y)}{f_X(x)} f_X(x) dydx \\
&= \int \int y f_{X,Y}(x,y) dydx \\
&= \int y \left[\int f_{X,Y}(x,y)dx\right] dy
\end{aligned}$$

A Special Property

You can "pull out" stuff if it depends what you're conditioning on. For any g,

$$E[g(X)Y|X] = g(X)E[Y|X].$$

This is used often with LTE to break down a weird expectation into known parts:
$E[g(X)Y] = E[E(g(X)Y|X)] = E[g(X)E[Y|X]]$

Example

Show that $E[Y|X]$ and $Y - E[Y|X]$ are uncorrelated.

Example

Show that $E[Y|X]$ and $Y - E[Y|X]$ are uncorrelated.

$$
\begin{aligned}
\mathrm{Cov}(E[Y|X], Y - E[Y|X]) &= \mathrm{Cov}(E[Y|X], Y) - \mathrm{Cov}(E[Y|X], E[Y|X]) \\
&= \mathrm{Cov}(E[Y|X], Y) - \mathrm{Var}(E[Y|X]) \\
&= E(E[Y|X]Y) - E[E(Y|X)]E[Y] - \mathrm{Var}(E[Y|X \\
&= E(E[Y|X]Y) - E[Y]^2 - \mathrm{Var}(E[Y|X]) \\
&= E(E[Y|X]Y) - E[(E[Y|X])^2] \\
&= E[E(E[Y|X]Y|X)] - E[(E[Y|X])^2] \\
&= E[E(Y|X)^2] - E[(E[Y|X])^2] = 0
\end{aligned}
$$

This helps us understand LTV as a Pythagorean theorem for random variables.

Example

Yet another entropy: "Conditional Entropy."

$$H(X|Y) = E[-\log f(X|Y)] = \iint -\log f_{X|Y}(x|y) f(x,y)\,dxdy$$

or

$$H(X|Y) = E[-\log p(X|Y)] = \sum_{x} \sum_{y} -\log p_{X|Y}(x|y) p(x,y)$$

Example

Another meaningful expectation: "Mutual Information." It's more general than correlation.

$$MI(X, Y) = E\left[-\log\left(\frac{f_X(X)f_Y(Y)}{f_{X,Y}(X, Y)}\right)\right]$$

The expectation is taken with respect to the joint distribution.