Math-UA.233: Theory of Probability

Lecture 20

Tim Austin

Conditioning with random variables

Next: using conditional probability in connection with RVs.

Our previous study of conditional probability was based on four main formulae:

- the definition
- multiplication rule (= the definition rearranged)
- LOTP
- Bayes.

We will now meet analogs of these involving RVs.

They come in 'discrete' and 'continuous' flavours.

Defining conditional distributions of RVs

Consider a model with two RVs, X and Y. We have various events defined in terms of X, and others defined in terms of Y.

Sometimes we need the conditional probabilities of "X-events" given "Y-events", or vice-versa: that is, something like

$$P(X \in A \mid Y \in B) = \frac{P(X \in A,\ Y \in B)}{P(Y \in B)}$$

for some $A, B \subseteq \mathbb{R}$.

This is not a new concept beyond conditional probability. But for special kinds of RV, some new tools or notation can be useful.

The discrete case (Ross Sec 6.4)

Now suppose X and Y are discrete with joint PMF p.

Then the most common "X-events" and "Y-events" of interest have the form $\{X = x\}$ and $\{Y = y\}$ for some fixed x and y. For the conditional probability, there's a convenient special notation.

Definition

*The **conditional PMF of** X **given** Y is the function*

$$\boxed{\underbrace{p_{X|Y}(x|y)}_{\text{notation}} = P(X = x \mid Y = y) = \frac{p(x, y)}{p_Y(y)}.}$$

CAREFUL: This is defined for any x, and any y for which $p_Y(y) > 0$ (we simply don't use it for other choices of y).

Example (Ross E.g. 6.4a)

Suppose that X and Y have joint PMF given by

$$p(0,0) = 0.4, \quad p(0,1) = 0.2, \quad p(1,0) = 0.1, \quad p(1,1) = 0.3.$$

Calculate the conditional PMF of X given that $Y = 1$.

Example (Ross E.g. 6.4b)

Let X and Y be independent Poi(λ) and Poi(μ) RVs. Calculate the conditional PMF of X given that $X + Y = n$.

MORAL: Calculating cond probs here is exactly the same task as earlier in the course.

What's *new* is that your answer is a function of all possible pairs (x, y), not a single number. (Beware of category errors!)

Conditioning can also involve three or more RVs. This requires care, but still no new ideas.

Example (Special case of Ross E.g. 6.4c)

An experiment has three possible outcomes with respective probabilities p_1, p_2, p_3. After n independent trials are performed, we write X_i for the number of times outcome i occurred, $i = 1, 2, 3$.

Find the conditional distribution (that is, the conditional joint PMF) of (X_1, X_2) given that $X_3 = m$.

WHAT THE QUESTION IS ASKING FOR:

$$p_{X_1,X_2|X_3}(k,\ell|m) = P(X_1 = k, X_2 = \ell \,|\, X_3 = m)$$

for all possible values of k, ℓ and m.

A COMFORTING FACT TO KNOW:

If X and Y are independent, then

$$p_{X|Y}(x|y) = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

So this is just like for events: if X and Y are independent, then knowing the value taken by Y doesn't influence the probability distribution of X.

Defining conditional distributions: continuous case (Ross Sec 6.5)

Now suppose that E is an event and Y is a continuous RV.

We'd like to make sense of the "conditional probability of E given the value taken by Y". But now we need a new idea.

THE PROBLEM:

If Y is a *continuous* RV, then

$$P(Y = y) = 0 \quad \text{for any exact value } y \in \mathbb{R},$$

so we *cannot define the conditional probability* $P(E \mid \{Y = y\})$ for any other event E using the usual formula (would have to divide by zero!).

THE IDEA:

Instead of assuming that Y takes the value y *exactly*, let us condition on Y taking a value in a tiny window around y:

$$\begin{aligned}P(E \mid \{y \leqslant Y \leqslant y + dy\}) &= \frac{P(E \cap \{y \leqslant Y \leqslant y + dy\})}{P\{y \leqslant Y \leqslant y + dy\}} \\ &\approx \frac{P(E \cap \{y \leqslant Y \leqslant y + dy\})}{f_Y(y)\,dy},\end{aligned}$$

where we use the infinitessimal interpretation of $f_Y(y)$. This makes sense provided $f_Y(y) > 0$.

Now let $dy \longrightarrow 0$:

Definition

*The **conditional probability of** E **given that** $Y = y$ is*

$$\boxed{P(E \mid Y = y) = \lim_{dy \longrightarrow 0} \frac{P(E \cap \{y \leqslant Y \leqslant y + dy\})}{f_Y(y)\, dy},}$$

provided the limit exists.

As usual, we won't address here the question of *when* the limit exists — it will in all the cases that matter to us.

(Think of it this way: assuming that this limit exists is a bit like assuming that a function on $\mathbb{R}$ is differentiable at the point y. In a more advanced course on 'analysis', you'll see that these questions are in fact closely related.)

Common setting in which we use the preceding definition:

> *(X, Y) are jointly continuous RVs, and E is an event defined in terms of them, i.e. $E = \{(X, Y) \in A)\}$ for some $A \subseteq \mathbb{R}^2$.*

Example (Ross E.g. 6.5b)

Suppose that X, Y have joint PMF

$$f(x, y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & 0 < x, y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Find $P(X > 1 \mid Y = y)$ for $y > 0$.

Especially important case: computing quantities like

$$P(a < X \leqslant b \mid Y = y)$$

for X and Y jointly continuous. For this task, there's a natural and useful analog of a conditional PMF.

Definition (Ross p250)

*Suppose X and Y are jointly continuous with joint PDF f. The **conditional PDF of** X **given** Y is the function*

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

It's defined for all real values x and all y such that $f_Y(y) > 0$.

CAUTION: This is not a conditional probability, since it's a ratio of *densities*, not of *probability values*.

Usefulness of the conditional PDF:

Proposition (See Ross p251)

Let $a < b$, or $a = \infty$ or $b = -\infty$. Let y be a real value such that $f_Y(y) > 0$. Then

$$\boxed{P(a < X \leqslant b \mid Y = y) = \int_a^b f_{X|Y}(x|y)\,dx.}$$

MORAL: $f_{X|Y}(x|y)$ is a new PDF. It describes the updated probability density of X if we find out that $Y = y$. It has all the other properties that we've already seen for PDFs.

Also, just as in the discrete case, if X and Y are independent then

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

So knowing the value taken by Y (to arbitrary accuracy) doesn't influence the probability distribution of X.

Finding a conditional PDF is just like finding a conditional PMF.

Example (Ross E.g. 6.5a)

The joint PDF of X and Y is

$$f(x,y) = \begin{cases} \frac{12}{5}x(2-x-y) & 0 < x,y < 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Find $f_{X|Y}(x|y)$ for all x and for $0 < y < 1$.

PROCEDURE: Find the marginal f_Y by integrating, then plug into the formula for $f_{X|Y}$.

Let's do this one again, this time using our new tool:

Example (Ross E.g. 6.5b, again)

Suppose that X, Y have joint PMF

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & 0 < x, y < \infty \\ 0 & \textit{otherwise.} \end{cases}$$

Find $P(X > 1 \mid Y = y)$ for $y > 0$.

Here's another important example.

Example (The bivariate standard normal again: Ross E.g. 6.5d)

Let (X, Y) be bivariate standard normal with correlation ρ:

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\Big).$$

Then, conditioned on $Y = y$, X is a $N(\rho y, 1-\rho^2)$ RV, and vice-versa.

See Ross E.g. 6.5d for the case of a general bivariate normal.

The multiplication rule

As before the *multiplication rule* just means re-arranging the definition of conditional probability. We can do the same re-arrangement for conditional PMFs or PDFs:

- Events E and F: $P(E \cap F) = P(E \mid F)P(F)$
- Discrete RVs X and Y: $p(x, y) = p_{X|Y}(x|y)p_Y(y)$.
- Jointly cts RVs X and Y: $f(x, y) = f_{X|Y}(x|y)f_Y(y)$.

Just like for events, we use a multiplication rule when we *assume* a conditional PMF or PDF (e.g., as a modeling choice), and then have to reconstruct a joint PDF.

Example

A stick of unit length is broken at a uniformly random point. Then the left-hand piece is broken again at a uniformly random point. Let Y be the length of the left-most piece after the second break. Find f_Y.

IDEA: Let X be the length of the left-hand piece after the first break.

Step 1: Obtain $f(x, y) = f_X(x) f_{Y|X}(y|x)$ (multiplication rule!).

Step 2: Integrate out x to get $f_Y(y)$.

MAIN DIFFICULTY: Correctly interpreting the question in terms of $f_{Y|X}$.

Example (Ross E.g. 6.5b, yet again)

Suppose that X, Y have joint PMF

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & 0 < x, y < \infty \\ 0 & \textit{otherwise.} \end{cases}$$

How do you suppose Ross came up with this example in the first place?

LOTP with two RVs

So far we've met notation and methods for understanding and computing:

- the conditional PMF $p_{X|Y}$ if X and Y are discrete,
- the conditional PDF $f_{X|Y}$ if X and Y are jointly continuous,
- and conditional probabilities $P((X, Y) \in A) \mid Y = y)$ if X and Y are jointly continuous.

There are also useful versions of LOTP in these settings.

LOTP with two discrete RVs

Once again, there's nothing really new here.

Suppose X and Y are both discrete with possible values x_1, x_2, ... and y_1, y_2, ... and joint PMF p.

Then the events $\{Y = y_1\}$, $\{Y = y_2\}$, ... are mutually exclusive and cover all possible outcomes — they are a *partition*.

Now we can just apply our old version of LOTP to this partition:

$$\begin{aligned} p_X(x_i) &= P(X = x_i) \\ &= \underbrace{\sum_j P(X = x_i \mid Y = y_j) P(Y = y_j)}_{\text{old LOTP}} = \sum_j p_{X|Y}(x_i|y_j) p_Y(y_j). \end{aligned}$$

LOTP with jointly continuous RVs, I

Here we get the obvious analog of the discrete formula. Suppose X and Y are jointly continuous with joint PDF f. Now we get:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{(computing marginal)} \\ &= \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\,dy \quad \text{(multiplication rule)} \end{aligned}$$

LOTP with jointly continuous RVs, II

There's another, more useful version for jointly continuous RVs. Now suppose we care about the event $\{(X, Y) \in A\}$, and want to condition on the value of Y.

Then LOTP looks like this:

$$\boxed{P((X, Y) \in A) = \int_{-\infty}^{\infty} P((X, Y) \in A \mid Y = y) f_Y(y)\, dy.}$$

This one is the main new idea for LOTP in this lecture.

(May sketch proof in class if time permits.)

Example (Version of Ross E.g. 7.5m)

Suppose X is Exp(1), Y is Unif$(0, 1)$, and they are independent. Compute $P(X < Y)$.

(Similar Pishro-Nik E.g. 5.25.)

Example (Ross E.g. 7.5n)

Let X and Y be independent continuous RVs. Use LOTP to re-derive the formula for f_{X+Y}:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx.$$