Math-UA.233: Theory of Probability
Lecture 21

Tim Austin

From last time... 1

If X and Y are discrete, the **conditional PMF of X given Y** is

$$p_{X|Y}(x|y) = P(X = x \mid Y = y) = \frac{p(x,y)}{p_Y(y)} \quad \text{provided } p_Y(y) > 0.$$

If Y is continuous and E is an event, the "conditional probability of E given $Y = y$" is

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

provided $f_Y(y) > 0$.

If X, Y are jointly cts, the **conditional PDF of X given Y** is

$$f_{X|Y}(x|y) = f(x,y)/f_Y(y) \quad \text{provided } f_Y(y) > 0.$$

From last time... 2

Extensions of LOTP:

Discrete:

$$p_X(x_i) = \sum_j p_{X\,|\,Y}(x_i \,|\, y_j) p_Y(y_j).$$

Continuous, version I:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X\,|\,Y}(x|y) f_Y(y)\, dy.$$

Continuous, version II:

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

Bayes' formula

Bayes for X and Y discrete:

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

This is literally just the old Bayes' formula written out for the events $\{X = x\}$ and $\{Y = y\}$.

Bayes for (X, Y) jointly continuous:

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

(Exactly like the discrete formula, but with f in place of p.)

MEANING in jointly cts case: apply old Bayes' formula to $\{x \leqslant X \leqslant x + dx\}$ and $\{y \leqslant Y \leqslant y + dy\}$, and let $dx, dy \longrightarrow 0$.

Example (contd.)

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

$$X = \textit{length of left-hand piece after first break, and}$$

$$Y = \textit{length of left-most piece after second break.}$$

Find $P(X > \frac{1}{2} \mid Y = \frac{1}{3})$.

IDEA: Step 1: $f_{X|Y}(x|y) = f_{Y|X}(y|x) f_X(x) / f_Y(y)$ by Bayes, and last time we showed $f_Y(y) = \log \frac{1}{y}$ for $0 < y < 1$.

Step 2: Integrate to find

$$P(X > 1/2 \mid Y = 1/3) = \int_{1/2}^{1} f_{X|Y}(x|1/3)\, dx.$$

Mixtures of discrete and continuous

Some situations are best modeled using both some continuous RVs and some discrete RVs. This is sometimes called a model of 'mixed type'.

Consider a continuous RV X and a discrete RV N (on the same sample space, as always).

Very often, the model is constructed so that:

> *The value of one of the RVs determines the probabilities of events for the other.*

This can go two ways around: N determines the probabilities of X-events, or vice-versa.

Examples:

1. *Continuous influencing discrete*. Electronic components can be more or less durable, and this is measured by a parameter between 0 and 1. The parameter of a new component is a continuous RV X lying in that range. Then it survives each use independently with probability X. The number of uses before failure is a discrete RV N whose distribution depends on X.
2. *Discrete influencing continuous*. The number of people who enter an elevator is a Poi(3) RV N, and then each of their weights is an independent continuous RV. The total weight X in the elevator is a continuous RV whose distribution depends on N.

Mathematically, in these models, we're being *given* conditional probabilities, and will have to find unconditioned probabilities using the multiplication rule or LOTP.

In these models, one also often wants to turn the conditioning around — leading to uses of Bayes' formula.

A lot of the subtlety here is in recognizing correctly what kind of conditional probability information is being *given* to you.

In the setting of a discrete and a continuous RV on the same sample space, the rigorous theory gets even trickier. But the formulae that result are simple analogs of those we've already seen.

So here let's just assume we are given conditional probability information in such a setting, and practice using those formulae.

Conditional probabilities:

1. *Cts influencing discrete*. X is continuous, N is discrete, and we are given the PDF f_X and a conditional PMF

$$p_{N|X}(n|x) \text{ for each } x \in \mathbb{R}.$$

2. *Discrete influencing cts*. N is discrete, X is continuous, and we are given the PMF p_N and a conditional PDF

$$f_{X|N}(x|n) \text{ for each possible value } n \text{ of } N.$$

Undoing the conditioning: LOTP

1. *Cts influencing discrete*. For any possible n of N, we get

$$\boxed{p_N(n) = \int_{-\infty}^{\infty} p_{N|X}(n|x) f_X(x)\,dx.}$$

2. *Discrete influencing cts*. For X itself, not conditioned on knowing the value of N, we find it is continuous with PDF

$$\boxed{f_X(x) = \sum_{n:\ p_N(n)>0} f_{X|N}(x|n) p_N(n).}$$

Example (Ross E.g. 7.5l)

A coin is manufactured in an unreliable way. As a result, its bias is a RV X which is Unif$(0,1)$. Suppose we flip it n times. Let N be the number of heads obtained. Find the PMF of N.

DATA GIVEN:

$$f_X(x) = 1 \quad (0 < x < 1)$$

and

$$p_{N|X}(k|x) = \binom{n}{k} x^k (1-x)^{n-k} \quad (k = 0, 1, \ldots, n).$$

Now use LOTP.

(A full answer requires the fact

$$\int_0^1 x^k (1-x)^{n-k}\, dx = \frac{k!(n-k)!}{(n+1)!}$$

— will prove this it time permits.)

Another common task: switching from the description from "cts influencing discrete" to "discrete influencing cts". This is a job for Bayes' formula.

Bayes' formula for N discrete and X continuous (see Ross p255):

$$\boxed{f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)}}$$

(can re-arrange if your task is to get $p_{N|X}(n|x)$ in terms of the others.)

Example (Ross E.g. 6.5e)

Consider again our unreliable coin. Suppose we flip it n times. What is the conditional PDF of its bias X given that we obtain k heads?

(The answer is a so-called *beta distribution* — see Ross for more.)

Conditional expectation (Ross Secs 7.5 and 7.6)

Suppose X and Y are discrete RVs and that y is a possible value of Y. Fix a possible value y of Y, and consider $p_{X|Y}(x|y)$ as a function of x.

Then this function obeys all the usual rules of a PMF — because it *is* a PMF. It describes the RV X in the new probability model we get by conditioning all our probabilities on $\{Y = y\}$.

SIMPLE IDEA: For a fixed choice of y, consider the expectation of X in this new model:

$$\boxed{E[X \mid Y = y] = \sum_{x \text{ such that } p_{X|Y}(x|y) > 0} x \cdot p_{X|Y}(x|y).}$$

This is the **conditional expectation of X given that $Y = y$**.

Continuous analog:

We all know the rules by now. If X, Y are jointly continuous with joint PDF f, then we define

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

(There are also analogs in mixed situations, which behave similarly, but let's leave those aside.)

Example (Ross e.g. 7.5a)

Suppose X and Y are two independent binom(n, p) RVs. Calculate $E[X \mid X + Y = m]$.

ANS boils down to: hypergeometric$(m, 2n, n)$

Example (Ross e.g. 7.5b)

Suppose that the joint PDF of X, Y is

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

Compute $E[X \mid Y = y]$.

Conditional expectation is extremely useful as a calculational tool. The key is the right point of view.

DEEPER IDEA: The value $E[X|Y = y]$ depends on the choice of y. This y is some value that Y can take. So let's define a *new random variable*, call it $E[X \mid Y]$, as follows:

$$E[X \mid Y] \text{ takes the value } E[X \mid Y = y] \text{ whenever } Y = y.$$

This new RV is the **conditional expectation of X given Y.**

This new RV has many useful properties. Here I'll focus on just one.

Theorem (The Law of Total Expectation, 'LOTE'; Ross Prop 7.5.1)

$$\boxed{E[E[X|Y]] = E[X].}$$

(This holds if X and Y are both discrete, jointly continuous, mixed, or in any other situation, provided both sides are defined.)

A connection with something we've already met:

Example

Let E be an event, let Y be a discrete RV, and let X be the indicator variable of E. Then

$$E[X \mid Y = y] = P(E \mid Y = y),$$

and therefore

$$P(E) = E[X] = \underbrace{E[E[X \mid Y]]}_{\text{LOTE}} = \underbrace{\sum_{y:\, p_Y(y)>0} P(E \mid Y = y) p_Y(y)}_{\text{LOTP}}.$$

In this way, LOTE is a generalization of LOTP. (You can do this example with Y being continuous, too.)

LOTE can be very useful when handling models that are given in terms of some conditional information.

Example (Pishro-Nik E.g. 5.26)

Suppose X is Unif$(1, 2)$, and that, given $X = x$, Y is Exp(x). Find (a) $E[Y]$ and (b) $\mathrm{Var}(Y)$.

KEY REALIZATION: We are told that

$$f_{Y|X}(y|x) = x\mathrm{e}^{-xy} \quad (1 < x < 2, y > 0),$$

and therefore

$$E[Y \mid X = x] = 1/x \quad \text{and} \quad E[Y^2 \mid X = x] = 2/x^2,$$

by previous calculations for Exp RVs.

Another nice application:

Example (Expectation of a sum of a random number of RVs; Ross E.g. 7.5d)

Suppose that the number of people entering a store on a given day is a RV N with mean 50. Suppose further that the amounts of money they spend are independent RVs all having mean \$8, and that these amounts are all independent of N. What is the expected amount of money spent in the store on that day?

General formula that we get in this situation:

$$E\Big[\sum_{i=1}^{N} X_i\Big] = E[X_1]E[N].$$