Math-UA.233: Theory of Probability
Lecture 18

Tim Austin

LOTUS in two dimensions

Let X and Y be two RVs and let $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$. Just like for $E[g(X)]$, there is a simple formula for $E[g(X, Y)]$, with a discrete and a continuous version.

Proposition (Discrete 2D LOTUS; Ross Prop 7.2.1)

If X and Y are discrete, possible values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, and their joint PMF is p, then

$$E[g(X, Y)] = \sum_{i,j} g(x_i, y_j) p(x_i, y_j)$$
$$= \sum_{(x,y) \text{ such that } p(x,y)>0} g(x, y) p(x, y).$$

PROOF: 'Group together' all possible pairs (x_i, y_j) which give the same value for $g(x_i, y_j)$ (just like for the 1D version).

Proposition (Cont$\underline{\text{s}}$ 2D LOTUS; Ross Prop 7.2.1, contd.)

If X, Y are jointly continuous with joint PDF f, then

$$\boxed{E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy.}$$

No proof here (this one is quite tricky). Just remember the analogy with $E[g(X)]$.

LOTUS works just the same for collections of more than two RVs — won't state explicitly here.

Previous example again:

Example (Ross E.g. 6.1d)

I play darts by throwing a dart at a circular dartboard of radius R. Relative to the origin, the dart lands at a random point (X, Y) whose distribution is uniform over the dartboard (I'm not very good at darts).

(b) *Find the marginal density functions of X and Y.*

(c) *Let D be the distance from (X, Y) to the origin. Find $P(D \leqslant r)$ for positive real values r.*

(d) *Find $E[D]$.*

Now let's do part (d) using LOTUS.

Example (Ross E.g. 7.2a)

An accident occurs at a point X that is uniformly distributed along a road of length L. At the time of the accident, an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

NOTE: In this example, "X and Y are independent" implies that the random point (X, Y) is uniformly distributed in the square $\{0 \leqslant x, y \leqslant L\}$.

Important consequence of 2D LOTUS: linearity of expectation for jointly continuous RVs (Ross p282).

Proposition

If X and Y are jointly continuous, then

$$E[X + Y] = E[X] + E[Y].$$

(Of course, we've used linearity of expectation a lot already in the course — the point is just to make the connection with LOTUS in the 2D, jointly continuous case. You should use linearity of expectation whenever it helps you.)

Here's another important application of LOTUS:

Proposition (Ross Prop 7.4.1)

If X and Y are independent, and g and h are any functions $\mathbb{R} \longrightarrow \mathbb{R}$, then

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

PROOF: Will do in class in the jointly continuous case.

BEWARE: *Not true without independence!*

Now recall: the variance of a RV X is

$$\mathrm{Var}(X) = E\big[(X - \mu_X)^2\big] \quad \text{where } \mu_X = E[X].$$

Here's a generalization to two RVs X and Y:

Definition

*Let X and Y be RVs, and let $\mu_X = E[X]$ and $\mu_Y = E[Y]$. The **covariance** of X and Y is*

$$\underbrace{\mathrm{Cov}(X, Y)}_{\text{notation}} = E\big[(X - \mu_X)(Y - \mu_Y)\big]$$

(provided this expectation makes sense.)

First properties:

1. Symmetry: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.
2. Applying with $Y = X$:

$$\boxed{\mathrm{Cov}(X, X) = \mathrm{Var}(X).}$$

3. Like Var, Cov has a useful alternative formula:

$$\boxed{\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y].}$$

4. If X and Y are independent then

$$\boxed{\mathrm{Cov}(X, Y) = E[X - \mu_X] \cdot E[Y - \mu_Y] = 0.}$$

Example (Ross E.g. 7.4d)

Let A and B be events, and let I_A and I_B be their indicator variables. Then

$$\mathrm{Cov}(I_A, I_B) = P(A \cap B) - P(A)P(B).$$

In particular,

$$\mathrm{Var}(I_A) = P(A) - P(A)^2 = P(A)(1 - P(A)).$$

Calculation from previous example again:

$$\mathrm{Cov}(I_A, I_B) = P(A \cap B) - P(A)P(B) = P(B)\big[P(A \,|\, B) - P(A)\big].$$

OBSERVE: this can be *positive or negative*.

- It's positive if $P(A \,|\, B) > P(A)$, i.e. if "knowing B makes A more likely".
- It's negative if $P(A \,|\, B) < P(A)$, i.e. if "knowing B makes A less likely".

So, FIRST REASON WHY $\mathrm{Cov}(X, Y)$ IS IMPORTANT: It gives some indication of how X and Y depend on each other, and how much.

But, WARNING: for *general* RVs,

$$\mathrm{Cov}(X, Y) = 0 \quad \not\Longrightarrow \quad X, Y \text{ independent.}$$

(Examples are plentiful, but messy and not too instructive.)

So remember:

> *Independence implies vanishing covariance, but vanishing covariance is not a <u>guarantee</u> of independence.*

Like variance, we can describe how covariance transforms under sums and products:

Proposition (Ross Prop 7.4.2)

1. *For any RVs X and Y and any $a \in \mathbb{R}$, we have*

$$\boxed{\mathrm{Cov}(aX, Y) = \mathrm{Cov}(X, aY) = a\,\mathrm{Cov}(X, Y)}$$

(so $\mathrm{Var}(aX) = \mathrm{Cov}(aX, aX) = a^2\mathrm{Cov}(X, X) = a^2\mathrm{Var}(X)$ — sanity check).

2. *For any RVs $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, we have*

$$\boxed{\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j).}$$

(so it behaves just like multiplying out a product of sums of numbers.)

In particular, if $m = n$ and $Y_i = X_i$ in part 2 above, we get (Ross eqn (7.4.1))

$$\boxed{\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \underbrace{\sum_{i=1}^{n} \mathrm{Var}(X_i)}_{\text{the 'diagonal terms'}} + 2 \underbrace{\sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j)}_{\text{the 'cross terms'}}}$$

This is the SECOND REASON WHY $\mathrm{Cov}(X, Y)$ IS IMPORTANT: You can use it to find variances of sums.

Special case: if $X_1, \ldots, X_n$ are *independent*, then $\mathrm{Cov}(X_i, X_j) = 0$ whenever $i \neq j$, so we're left with

$$\boxed{\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i).}$$

Here's a recap of what we've learned about sums of RVs so far:

Proposition (Parameters for sums of RVs)

Let $X_1, \ldots, X_n$ be RVs with $E[X_i] = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$ for $i = 1, \ldots, n$.

- *We have*

$$E\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mu_i.$$

- *If $X_1, \ldots, X_n$ are independent, then*

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \sigma_i^2.$$

(If not independent, then terms like $\mathrm{Cov}(X_i, X_j)$ appear on the right.)

Example (Ross E.g. 7.4b)

If X is binom(n, p), then

$$\mathrm{Var}(X) = np(1-p).$$

IDEA: Recall that $X = X_1 + \cdots + X_n$, a sum of independent indicator variables.

Example (Special case of Ross E.g. 7.4f)

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 $\mathrm{Cov}(X_1, X_2)$.

Sometimes, it's more informative to measure covariance *relative to* variance:

Definition

*Let X and Y be RVs with positive, finite variances. Then their **correlation** is*

$$\operatorname{corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Observe that $\operatorname{corr}(aX, bY) = \operatorname{corr}(X, Y)$ for any non-zero constants a, b. For example, if X and Y are two random distance measurements, then their correlation doesn't depend on whether we use inches or centimetres.

Example

Let (X, Y) *be bivariate standard normal with parameter* ρ*.*
(Reminder: their joint PDF is

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

Then X *and* Y *are both* $N(0, 1)$*, and it turns out that*

$$\mathrm{Cov}(X, Y) = \mathrm{corr}(X, Y) = \rho,$$

(justifying our previous name for this parameter ρ*).*

You can prove these facts by direct calculation, but this involves some really nasty integrals. Instead, we'll come back to them later using some other tools.

Distributions of sums of RVs (Ross Sec 6.3)

In 1D, we studied how to find the distribution (i.e., the CDF) of $g(X)$ in terms of that of X. It's sometimes quite hard.

The situation in 2D is even worse. So now we're going to focus on one very special but very important case:

$$g(X, Y) = X + Y.$$

We've already seen

$$E[X + Y] = E[X] + E[Y]$$

and

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y).$$

But sometimes its important to find the actual distribution (i.e. CDF, PMF or PDF) of $X + Y$.

Proposition (Distribution of sum: discrete case)

If X and Y are discrete with possible values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots,$ and joint PMF p, then

$$\boxed{P(X+Y=z) = \sum_{\text{all } i,j \text{ such that } x_i+y_j=z} p(x_i, y_j)}$$

for any $z \in \mathbb{R}$.

PROOF: We have

$$P(X+Y=z) = P\big((X,Y) \in A\big) \text{ for } A = \{(x,y) \in \mathbb{R}^2 \mid x+y=z\}.$$

Now use the general formula for computing probabilities using the joint PMF.

Example (Binomials; Ross E.g. 6.3f)

Let X and Y be independent binomial RVs with params (n, p) and (m, p). Then $X + Y$ is binom$(n + m, p)$.

IDEA: Remember what X and Y are *counting*.

SANITY CHECK: This gives

$E[X+Y] = (n+m)p = np+mp = E[X]+E[Y]$ — as seen before.

Example (Poissons; Ross E.g. 6.3e)

Let X and Y be independent Poi(λ) and Poi(μ), respectively. Then $X + Y$ is Poi$(\lambda + \mu)$.

IDEA: Compute directly, or deduce from Poisson approx to binomial.

The result in the continuous case takes a bit more work.

Proposition (Distribution of sum: continuous case)

If X, Y are jointly continuous with joint PDF f, then $X + Y$ is continuous with PDF

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

INTUITIVE PROOF:

$$f_{X+Y}(z)\,dz \approx P(X+Y \in [z, z+dz])$$

$$= \iint_{\{(x,y)\,|\,x+y\in[z,z+dz]\}} f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \Big[\underbrace{\int_{z-x}^{z+dz-x} f_{X,Y}(x,y)\,dy}_{\approx f_{X,Y}(x,z-x)\,dz} \Big]\,dx$$

(See Ross p239 for a proof by first finding the CDF of $X+Y$.)