

Joint Distributions

CSE 312 Winter 25
Lecture 18

Announcements

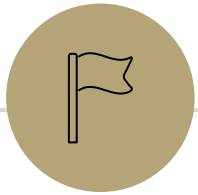
HW6 is out (back to the Wed->Wed schedule)

More gradescope questions than usual

Since we're just passed the midterm, we're hoping saving you the extra writing will make this week more manageable.

Also a coding question this week

You'll want to read the motivation in the chapter to understand why you're looking at these random variables.



Multiple Random Variables

This lecture and next lecture

Somewhat out-of-place content.

When we introduced multiple random variables, we've always had them be independent.

Because it's hard to deal with non-independent random variables.

Today and Monday are a crash-course in the toolkit for when you have multiple random variables and they aren't independent.

Going to focus on discrete RVs, we'll talk about continuous at the end.

Why

Independent random variables are easier to interact with.

But sometimes you **want** to interact with the dependence

ML/Data science takes advantage of dependence: Netflix knows you like movie A; people liking movie A is dependent on people liking movie B, and so recommends you movie B

Random variables might be indicators for specific individual people liking movies, or “if we select a person at random, will they like this movie”

Our examples today and Monday are artificial/simple; we’re just hoping to get the tools down today and Monday.

Joint PMF, support

For two (discrete) random variables X, Y their joint pmf

$$p_{X,Y}(x, y) = \mathbb{P}(X = x \cap Y = y)$$

When X, Y are independent then $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.

Examples

Roll a blue die and a red die. Each die is 4-sided. Let X be the blue die's result and Y be the red die's result.

Each die (individually) is fair. But not all results are equally likely when looking at them both together.

$$p_{X,Y}(1,2) = 3/16.$$

$p_{X,Y}$	$X=1$	$X=2$	$X=3$	$X=4$
$Y=1$	1/16	1/16	1/16	1/16
$Y=2$	3/16	0	0	1/16
$Y=3$	0	2/16	0	2/16
$Y=4$	0	1/16	3/16	0

Marginals

What if I just want to talk about X ?

Well, use the law of total probability:

$$\mathbb{P}(X = k) = \sum_{\text{partition } \{E_i\}} \mathbb{P}(X = k | E_i) \mathbb{P}(E_i)$$

and use E_i to be possible outcomes for Y For the dice example

$$\mathbb{P}(X = k) = \sum_{\ell=1}^4 \mathbb{P}(X = k | Y = \ell) \mathbb{P}(Y = \ell)$$

$$= \sum_{\ell=1}^4 \mathbb{P}(X = k \cap Y = \ell)$$

$$p_X(k) = \sum_{\ell=1}^4 p_{X,Y}(k, \ell)$$

$p_X(k)$ is called the “marginal” distribution for X (we “marginalized out” Y) it’s the same pmf we’ve always used; the name comes from being in the margin of the paper when people printed these on paper.

Marginals

$$p_X(k) = \sum_{\ell=1}^4 p_{X,Y}(k, \ell)$$

So

$$p_X(2) = \frac{1}{16} + 0 + \frac{2}{16} + \frac{1}{16} = \frac{4}{16}$$

$p_{X,Y}$	$X=1$	$X=2$	$X=3$	$X=4$
$Y=1$	1/16	1/16	1/16	1/16
$Y=2$	3/16	0	0	1/16
$Y=3$	0	2/16	0	2/16
$Y=4$	0	1/16	3/16	0

Different dice

Roll two fair dice independently.
Let U be the minimum of the two rolls and V be the maximum

Are U and V independent?

Write the joint distribution in the table

What's $p_U(z)$? (the marginal for U)

$p_{U,V}$	$U=1$	$U=2$	$U=3$	$U=4$
$V=1$				
$V=2$				
$V=3$				
$V=4$				

Different dice

Roll two fair dice independently.
Let U be the minimum of the two rolls and V be the maximum

$$p_U(z) = \begin{cases} \frac{7}{16} & \text{if } z = 1 \\ \frac{5}{16} & \text{if } z = 2 \\ \frac{3}{16} & \text{if } z = 3 \\ \frac{1}{16} & \text{if } z = 4 \\ 0 & \text{otherwise} \end{cases}$$

$p_{U,V}$	$U=1$	$U=2$	$U=3$	$U=4$
$V=1$	1/16	0	0	0
$V=2$	2/16	1/16	0	0
$V=3$	2/16	2/16	1/16	0
$V=4$	2/16	2/16	2/16	1/16

Joint Expectation

Expectations of joint functions

For a function $g(X, Y)$, the expectation can be written in terms of the joint pmf.

$$\mathbb{E}[g(X, Y)] = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g(x, y) \cdot p_{X,Y}(x, y)$$

This definition hopefully isn't surprising at this point (it's the value of g times the probability g takes on that value), but it's good to see.

Expectation of a function of two RVs

What's $\mathbb{E}[UV]$ for U, V from the last slide?

Expectation of a function of two RVs

What's $\mathbb{E}[UV]$ for U, V from the last slide?

$$\begin{aligned} & \sum_{u \in \Omega_U} \sum_{v \in \Omega_V} uv \cdot p_{U,V}(u, v) \\ &= 1 \cdot 1 \cdot \frac{1}{16} + 1 \cdot 2 \cdot \frac{2}{16} + 1 \cdot 3 \cdot \frac{2}{16} + 2 \cdot 2 \cdot \frac{1}{16} + 2 \cdot 3 \cdot \frac{2}{16} + 2 \cdot 4 \cdot \frac{2}{16} + \\ & \quad 3 \cdot 3 \cdot \frac{1}{16} + 3 \cdot 4 \cdot \frac{2}{16} + 4 \cdot 4 \cdot \frac{1}{16} \\ &= \frac{92}{16} = \frac{23}{4} = 5.75. \end{aligned}$$

Conditional Expectation

Waaaaaay back when, we said conditioning on an event creates a new probability space, with all the laws holding.

So we can define things like “conditional expectations” which is the expectation of a random variable in that new probability space.

$$\mathbb{E}[X|E] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|E)$$

$$\mathbb{E}[X|Y = y] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|Y = y)$$

Conditional Expectations

All your favorite theorems are still true.

For example, linearity of expectation still holds

$$\mathbb{E}[(aX + bY + c) | E] = a\mathbb{E}[X|E] + b\mathbb{E}[Y|E] + c$$

Law of Total Expectation

Let $A_1, A_2, \dots, A_k$ be a partition of the sample space, then

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X|A_i] \mathbb{P}(A_i)$$

Let X, Y be discrete random variables, then

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y = y] \mathbb{P}(Y = y)$$

Similar in form to law of total probability, and the proof goes that way as well.

LTE

You will flip 2 (independent, fair coins). Call the number of heads X . Then (independently of the coin flips) draw an exponential random variable Y from the distribution $\text{Exp}(X + 1)$.

What is $\mathbb{E}[Y]$?

LTE

You will flip 2 (independent, fair coins). Call the number of heads X . Then (independently of the coin flips) draw an exponential random variable Y from the distribution $\text{Exp}(X + 1)$.

What is $\mathbb{E}[Y]$?

$$\mathbb{E}[Y]$$

$$= \mathbb{E}[Y|X = 0]\mathbb{P}(X = 0) + \mathbb{E}[Y|X = 1]\mathbb{P}(X = 1) + \mathbb{E}[Y|X = 2]\mathbb{P}(X = 2)$$

$$= \mathbb{E}[Y|X = 0] \cdot \frac{1}{4} + \mathbb{E}[Y|X = 1] \cdot \frac{1}{2} + \mathbb{E}[Y|X = 2] \cdot \frac{1}{4}$$

$$= \frac{1}{0+1} \cdot \frac{1}{4} + \frac{1}{1+1} \cdot \frac{1}{2} + \frac{1}{2+1} \cdot \frac{1}{4} = \frac{7}{12}.$$

Analogue for continuous

Everything we saw today has a continuous version.

There are “no surprises”– replace pmf with pdf and sums with integrals.

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x, y) = P(X = x, Y = y)$	$f_{X,Y}(x, y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x, y) = \sum_{t \leq x} \sum_{s \leq y} p_{X,Y}(t, s)$	$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(t, s) ds dt$
Normalization	$\sum_x \sum_y p_{X,Y}(x, y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x, y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$
Expectation	$E[g(X, Y)] = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$	$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$
Conditional PMF/PDF	$p_{X Y}(x y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$	$f_{X Y}(x y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$
Conditional Expectation	$E[X Y = y] = \sum_x x p_{X Y}(x y)$	$E[X Y = y] = \int_{-\infty}^{\infty} x f_{X Y}(x y) dx$
Independence	$\forall x, y, p_{X,Y}(x, y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x, y) = f_X(x) f_Y(y)$

Covariance

We sometimes want to measure how “intertwined” X and Y are – how much knowing about one of them will affect the other.

If X turns out “big” how likely is it that Y will be “big” how much do they “vary together”

Covariance

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Covariance

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

If X, Y go in the same direction

If X, Y go in the opposite directions

Covariance

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

That's consistent with our previous knowledge for independent variables. (for X, Y independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\text{Var}(X + Y)$?

Before you calculate, make a prediction. What should it be?

Covariance

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\text{Var}(X + Y)$?

$$\text{Var}(X) = \text{Var}(Y) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1 - 0^2 = 1$$

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$\mathbb{E}[XY] = \frac{1}{2} \cdot (-1 \cdot 1) + \frac{1}{2} (1 \cdot -1) = -1$$

$$\text{Cov}(X, Y) = -1 - 0 \cdot 0 = -1.$$

$$\text{Var}(X + Y) = 1 + 1 + 2 \cdot -1 = 0$$

Covariance, Another example

Let X be a Bernoulli RV with probability p of success.

Let $Y = X$ (Y is X , not an iid copy, literally the same experiment)

Let $Z = -X$

Let W be an independent Bernoulli, identically distributed to X

Find

$\text{Cov}(X, Y)$, $\text{Cov}(X, Z)$, $\text{Cov}(X, W)$

Covariance, Another example

Let X be a Bernoulli RV with probability p of success.

Let $Y = X$ (Y is X , not an iid copy, literally the same experiment)

Let $Z = -X$

Let W be an independent Bernoulli, identically distributed to X

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$= (1 \cdot 1 \cdot p + 0 \cdot 0 \cdot [1 - p]) - p \cdot p$$

$$= p - p^2 = p(1 - p)$$

Hey, that's the variance of X . This is a pattern: $\text{Cov}(X, X) = \text{Var}(X)$

Covariance, Another example

Let X be a Bernoulli RV with probability p of success.

Let $Y = X$ (Y is X , not an iid copy, literally the same experiment)

Let $Z = -X$

Let W be an independent Bernoulli, identically distributed to X

$$\text{Cov}(X, Z) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$= (1 \cdot -1 \cdot p + 0 \cdot -0 \cdot [1 - p]) - (p \cdot [-p])$$

$$= -p - [-p^2] = -p(1 - p)$$

General pattern: $\text{Cov}(X, -Y) = -\text{Cov}(X, Y)$

Covariance, Another example

Let X be a Bernoulli RV with probability p of success.

Let $Y = X$ (Y is X , not an iid copy, literally the same experiment)

Let $Z = -X$

Let W be an independent Bernoulli, identically distributed to X

$$\text{Cov}(X, W) = \mathbb{E}[XW] - \mathbb{E}[X]\mathbb{E}[W]$$

$$= (1 \cdot 1 \cdot p^2 + 1 \cdot 0 \cdot p[1 - p] + 0 \cdot 1 \cdot [1 - p]p + 0 \cdot 0 \cdot [1 - p]^2) - (p \cdot [p])$$

$$= (p^2) - p^2 = 0$$

General pattern: if X, Y independent $\text{Cov}(X, Y) = 0$