Joint Distributions

CSE 312 Summer 25
Lecture 18

Announcements

Homework 4 solutions available outside Allen 206

Homework 5 due tonight

Homework 6 will release later today

Quiz 6 on Friday

Multiple Random Variables

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?

$\sum_{u\in\Omega_U} \sum_{v\in\Omega_V} uv \cdot p_{U,V}(u,v)$

$= 1 \cdot 1 \cdot \frac{1}{16} + 1 \cdot 2 \cdot \frac{2}{16} + 1 \cdot 3 \cdot \frac{2}{16} + 2 \cdot 2 \cdot \frac{1}{16} + 2 \cdot 3 \cdot \frac{2}{16} + 2 \cdot 4 \cdot \frac{2}{16} +$

$3 \cdot 3 \cdot \frac{1}{16} + 3 \cdot 4 \cdot \frac{2}{16} + 4 \cdot 4 \cdot \frac{1}{16}$

$= \frac{92}{16} = \frac{23}{4} = 5.75.$

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 $A_1, A_2, \ldots, A_k$ 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 $\mathrm{Exp}(X+1)$.

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 $\mathrm{Exp}(X+1)$.

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

$\mathbb{E}[Y]$

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

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

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

Different dice

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

What is $\mathbb{P}(U = 2|V = 3)$?

$$\frac{\mathbb{P}(U=2 \cap V=3)}{\mathbb{P}(V=3)} = \frac{2/16}{5/16} = \frac{2}{5}$$

$$p_{U|V}(2|3) = \frac{2}{5}$$

$p_{U,V}$	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

Different dice

Find these values

$p_{V|U}(2|1) =$

$p_{U|V}(1|2) =$

$p_{U|V}(4|1) =$

$p_{U,V}$	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

Different dice

Find these values

$$p_{V|U}(2|1) = \frac{p_{V,U}(2,1)}{p_U(1)} = \frac{2/16}{7/16} = \frac{2}{7}$$

$$p_{U|V}(1|2) = \frac{p_{U,V}(1,2)}{p_V(2)} = \frac{2/16}{3/16} = \frac{2}{3}$$

$$p_{U|V}(4|1) = \frac{p_{U,V}(4,1)}{p_V(1)} = \frac{0}{1/16} = 0$$

$p_{U,V}$	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

Continuous Joint Distributions

What about the continuous versions?

In the continuous case, everything is still a density function, not a mass function.

Joint density

Marginal density

Conditional density

Expectations, conditional expectations integrate $x \cdot$ (cond)density(x)

You aren't getting a probability, you're getting a density; have to integrate to get a value.

Analogues 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	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$
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	$p_{X\mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_{x} x p_{X\mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X\mid Y}(x \mid y)\,dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

A note on independence

The definition of independence says X, Y independent if and only if

$p_{X,Y}(x, y) = p_X(x)p_Y(y)$ or $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (as appropriate)

There's often a nice shortcut. If X, Y are independent then joint support of X, Y (denoted $\Omega_{X,Y}$) must be $\Omega_X \times \Omega_Y$.

Joint support is $\{(x, y): p_{X,Y}(x, y) > 0\}$.

Often easier to verify dependence when those are different (especially in the continuous case).

But note this is a single implication not an if-and-only-if.

Covariance

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

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

Covariance

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

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, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).

Covariance

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

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 $\mathrm{Var}(X+Y)$? What is $\mathrm{Cov}(X,Y)$?

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 <u>is</u> X, not an iid copy, literally the same experiment)

Let $Z = -X$

Let W be an independent Bernoulli, indentically 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, indentically 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, indentically 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, indentically distributed to X

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

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

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

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

A Few Notes

Covariance is an un-normalized number.

It measures both how intertwined X, Y are and in some sense how much X, Y vary in the first place (if you multiply both X, Y by 2, the strength of the relationship intuitively is the same, but covariance increases).

If you want just the strength of the relationship, you probably want the "correlation coefficient": $\frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$ always between -1 and 1.

Covariance directly measures only "linear" relationships; if Y depends on X^2, the covariance might not be as high as you expect.

If dealing with real data, look at a plot to see if you should be looking for a linear relationship in the first place.