

# Joint Distributions

CSE 312 Summer 25  
Lecture 18

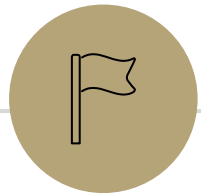
# Announcements

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

$$E[U \mid V=4]$$

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
$V=4$	$2/16$	$2/16$	$2/16$	$1/16$

$$\Rightarrow \frac{7}{16} \quad \frac{5}{16} \quad \frac{3}{16} \quad \frac{1}{16}$$

# Joint Expectation

$$\underline{X+Y}$$

$$\underline{X \cdot Y}$$

$$X^2 + 7Y$$

## 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?

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

# 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

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

# Law of Total Expectation

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

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

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

$$\frac{1}{\lambda}$$

$\{0, 1, 2\}$

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  $\mathbb{E}[Y]$ ?

$$Y \sim \text{Exp}(0+1) \quad \frac{1}{\lambda} = 1$$

$\mathbb{E}[Y]$

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

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

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

$$\underbrace{\quad} \quad \underbrace{\quad} \quad \underbrace{\quad}$$

# 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

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

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

$$p_{U|V}(4|1) = 0$$

$\frac{2}{7}$   
 $\frac{2}{3}$



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

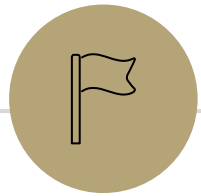
$$p_{V|U}(2|1) = \frac{p_{V,U}(2,1)}{p_U(1)} = \frac{2/16}{7/16} = \frac{2}{7}$$

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

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

$$\Omega_U = \{1, 2, 3, 4\}$$
$$\Omega_V = \{1, 2, 3, 4\}$$

$p_{U,V}$	$U=1$	$U=2$	$U=3$	$U=4$
$V=1$	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  $x \cdot (\text{cond})\text{density}(x)$

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

# A note on independence

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

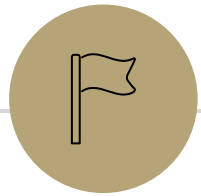
$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \text{ or } f_{X,Y}(x, y) = f_X(x)f_Y(y) \text{ (as appropriate)}$$

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

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

## Covariance

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

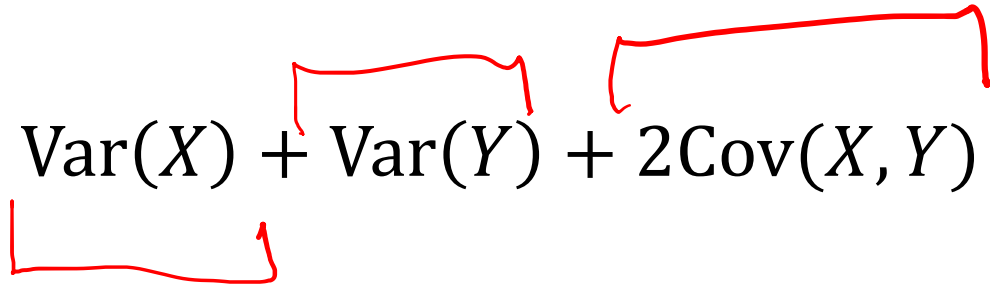
If  $X, Y$  go in the same direction

$x > \mathbb{E}[x] \quad + \quad \rightarrow \quad +$   
 $y > \mathbb{E}[y] \quad +$   
 $x < \mathbb{E}[x] \quad - \quad \rightarrow \quad +$   
 $y < \mathbb{E}[y] \quad -$


If  $X, Y$  go in the opposite directions

$x > \mathbb{E}[x] \quad + \quad \rightarrow \quad -$   
 $y < \mathbb{E}[y] \quad -$   
 $x < \mathbb{E}[x] \quad - \quad \rightarrow \quad -$   
 $y > \mathbb{E}[y] \quad +$

# Covariance

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$
The equation is annotated with three hand-drawn red brackets. One bracket is under the  $\text{Var}(X)$  term. A second bracket is above the  $\text{Var}(Y)$  term. A third, longer bracket is above the  $2\text{Cov}(X, Y)$  term.

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

A hand-drawn red underline is positioned under the expression  $\mathbb{E}[X]\mathbb{E}[Y]$  in the text above.

# Covariance

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

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

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

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

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

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

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

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

# 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  $X$

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

$$= (\underbrace{1 \cdot 1}_{p} + \underbrace{0 \cdot 0}_{[1-p]}) - p \cdot p$$

$$= p - p^2 = \underbrace{p(1-p)}$$

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

$$\begin{aligned} & \text{Cov}(aX, Y) \\ &= a \text{Cov}(X, Y) \end{aligned}$$

# 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  $X$

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

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

$$= (p^2) - p^2 = \underline{0}$$

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

# 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.