Random Variables II (Variance and Covariance)

CS70: Discrete Mathematics and Probability Theory

UC Berkeley – Summer 2026

Lecture 20

Ref: Note 17

Today

Conditional expectation
- Basic properties
- Law of Total Expectation

Functions of a random variable
- Sums and products
- Products of independent random variables

Variance
- Definitions and meaning
- Variance of common distributions
- Properties of variance

Covariance
- Definition and meaning
- Positively and negatively correlated random variables

Correlation: Scaled covariance

Conditional Expectation

One more basic property of RVs that didn't fit in the last lecture – recall:

$$\mathbb{P}[X = x \mid Y = y] = \frac{\mathbb{P}[X = x, Y = y]}{\mathbb{P}[Y = y]}$$

Definition: For RVs X and Y, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \sum_x x \cdot \mathbb{P}[X = x \mid Y = y]$$

Interpretation: Expected value of X restricted to $Y = y$ as a sample space.

Some obvious (or at least easy to prove) properties:

- ⇒ If X and Y are independent, and $\mathbb{P}[Y = y] \neq 0$, then $\mathbb{E}[X|Y = y] = \mathbb{E}[X]$
- ⇒ Linearity of expectation – for RVs X, Y, and Z, and constants a and b:

$$\mathbb{E}[aX + bY \mid Z = z] = a\mathbb{E}[X \mid Z = z] + b\mathbb{E}[Y \mid Z = z]$$

And a less obvious property on the next slide...

Conditional Expectation: Total Expectation

$\mathbb{E}[X \mid Y = y]$ is a single number

⇒ "$Y = y$" removes randomness from Y, expectation removes from X

Does $\mathbb{E}[X \mid Y]$ make sense?

⇒ Yes, but randomness not removed from Y, so this is a *random variable*

⇒ ... so we can consider $\mathbb{E}[\mathbb{E}[X \mid Y]]$

Theorem (Law of Total Expectation):

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

Proof:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \sum_y \mathbb{E}[X \mid Y = y]\mathbb{P}[Y = y] \qquad (1) \\
&= \sum_y \sum_x x \cdot \mathbb{P}[X = x \mid Y = y]\mathbb{P}[Y = y] \\
&= \sum_x x \sum_y \mathbb{P}[X = x \mid Y = y]\mathbb{P}[Y = y] \\
&= \sum_x x\mathbb{P}[X = x] = \mathbb{E}[X] \qquad \text{... by Total Probability}
\end{aligned}$$

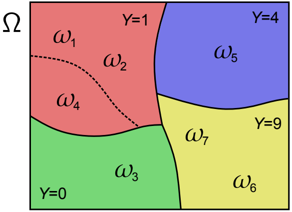
Useful applications of conditional expectation:

⇒ Calculating $\mathbb{E}[X]$ by cases, using (1)

⇒ In *prediction* problems

Functions of a Random Variable

For function $f(x)$, writing $Y = f(X)$ for r.v. X defines Y per outcome: $Y(\omega) = f(X(\omega))$
Example below with $Y = X^2$



$$\begin{array}{lll} X(\omega_1)=1 & f(X(\omega_1))=1 & Y(\omega_1)=1 \\ X(\omega_2)=1 & f(X(\omega_2))=1 & Y(\omega_2)=1 \\ X(\omega_3)=0 & f(X(\omega_3))=0 & Y(\omega_3)=0 \\ X(\omega_4)=-1 & f(X(\omega_4))=1 & Y(\omega_4)=1 \\ X(\omega_5)=-2 & f(X(\omega_5))=4 & Y(\omega_5)=4 \\ X(\omega_6)=3 & f(X(\omega_6))=9 & Y(\omega_6)=9 \\ X(\omega_7)=-3 & f(X(\omega_7))=9 & Y(\omega_7)=9 \end{array}$$

$Y = y$ partitions Ω and for each y the preimage values $f^{-1}(y) = \{x \mid f(x) = y\}$ partition $Y = y$

- $\Rightarrow$ Events $Y = 0$, $Y = 1$, $Y = 4$, and $Y = 9$ partition Ω
- $\Rightarrow$ Preimage $f^{-1}(1) = \{-1, 1\} \implies Y = 1$ partitioned into $X = 1$ and $X = -1$

So what is $\mathbb{P}[Y = 1]$?

$\Rightarrow \mathbb{P}[Y = 1] = \mathbb{P}[X = 1] + \mathbb{P}[X = -1] = \sum_{x \in f^{-1}(1)} \mathbb{P}[X = x]$ (partitions $Y = 1$, disjoint $\rightarrow$ add)

In general

$$\mathbb{P}[Y = y] = \sum_{x \in f^{-1}(y)} \mathbb{P}[X = x]$$

Functions of Random Variable(s): Expectation

For $Y = f(X)$, what is $\mathbb{E}[Y]$?

$$\mathbb{E}[Y] = \sum_y y \cdot \mathbb{P}[Y = y] = \sum_y y \sum_{x \in f^{-1}(y)} \mathbb{P}[X = x] = \sum_y \sum_{x \in f^{-1}(y)} f(x) \cdot \mathbb{P}[X = x]$$

Since the y values partition Ω and x's partition each y, this is just

$$\mathbb{E}[Y] = \sum_x f(x) \cdot \mathbb{P}[X = x]$$

We sometimes write this as $\mathbb{E}[f(x)]$.

This generalizes to functions of more than one r.v., using joint distribution.

For random variables X and Y and function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, if $Z = f(X, Y)$ then

$$\mathbb{E}[Z] = \mathbb{E}[f(X, Y)] = \sum_{x,y} f(x, y) \cdot \mathbb{P}[X = x, Y = y]$$

Simple Combinations of Two Random Variables

Random variables are measurements on an outcome, and we can use them in a calculation/function.

Example 1 (Sum): Experiment is to roll two 6-sided dice

Random variable $X =$ value of first roll

Random variable $Y =$ value of second roll

Calculation $X + Y$ – really $X(\omega) + Y(\omega)$ for same outcome ω

What we know:

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ – *always* (linearity of expectation)

Doesn't matter if X and Y are independent, dependent, or even the same

Example 2 (Product): Experiment is to flip a coin and roll a 6-sided die

Random variable M (the "multiplier") is 1 if coin is `T`, 2 if coin is `H`

Random variable $R =$ value of die roll

Calculation $M \cdot R$ – really $M(\omega) \cdot R(\omega)$ for same outcome ω

What we know:

$\mathbb{E}[M \cdot R] =$??? – let's explore this more...

Products of Random Variables

Experiment: Flip a fair coin $\quad \Omega = \{\mathtt{H}, \mathtt{T}\}$

$$X_1(\omega) = \begin{cases} 1, & \text{if coin is heads} \\ 0, & \text{otherwise} \end{cases} \qquad X_2(\omega) = \begin{cases} 1, & \text{if coin is tails} \\ 0, & \text{otherwise} \end{cases}$$

ω	$\mathbb{P}[\omega]$	X_1	X_2	$X_1 \cdot X_2$
H	1/2	1	0	0
T	1/2	0	1	0

Expected val:

$\mathbb{E}[X_1] \cdot \mathbb{E}[X_2] =$ $\qquad$ but $E[X_1 \cdot X_2] =$

Conclusion: It is *not* generally true that $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Anything interesting here? Are X_1 and X_2 independent?

Products of Independent Random Variables

Theorem: If X and Y are independent random variables, then

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

Proof: Using the joint distribution of X and Y and the definition of independent random variables,

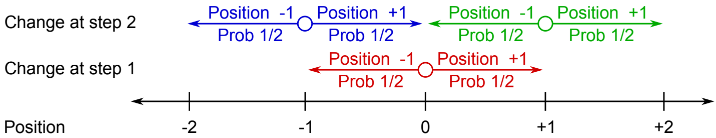
$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_a \sum_b (a \cdot b) \cdot \mathbb{P}[X = a, Y = b] \\
&= \sum_a \sum_b a \cdot b \cdot \mathbb{P}[X = a] \cdot \mathbb{P}[Y = b] \\
&= \sum_a \left(a \cdot \mathbb{P}[X = a] \cdot \sum_b b \cdot \mathbb{P}[Y = b] \right) \\
&= \left(\sum_a a \cdot \mathbb{P}[X = a] \right) \cdot \left(\sum_b b \cdot \mathbb{P}[Y = b] \right) \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}$$

$\square$

Where did we use independence?

New Problem: Random Walk

Random walk: Start at position 0, and move left or right with probability 1/2



In picture above: make red choice at step one, and blue or green at step 2

Random variable $X_n =$ position after step n

Movement: random variable S_i is change in position at step i

$\Rightarrow \quad X_n = S_1 + S_2 + \cdots + S_n$

Expected movement at step i: $\mathbb{E}[S_i] =$

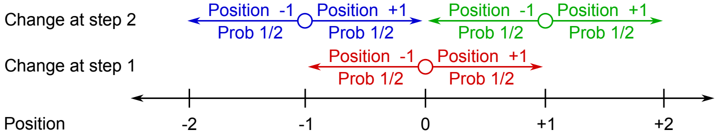
Expected *position* after step i:

$$\mathbb{E}[X_n] = \mathbb{E}[S_1 + S_2 + \cdots + S_n] = \mathbb{E}[S_1] + \mathbb{E}[S_2] + \cdots + \mathbb{E}[S_n] = 0 + 0 + \cdots + 0 = 0.$$

"Expected movement" is.... none?

Random Walk: Expected Distance from Origin

Random walk: Start at position 0, and move left or right with probability 1/2



Expected position after step i: $\mathbb{E}[X_n] = 0$

Issue: Symmetric about origin, so left positions cancel out right positions

$\Rightarrow$ No information about how far we stray from the origin

Idea 1: Take absolute value of position (no more cancellation)

So try $\mathbb{E}[\,|X_n|\,]$

Decent idea, but... absolute value makes this hard to calculate

Idea 2: Square the distance — also turns negatives into positives

Now want $\mathbb{E}[(X_n)^2]$

Is this the same as $(\mathbb{E}[X_n])^2$?

Fortunately no, because this would be zero too!

This would be the same if X_n and X_n were... independent? Nonsense.

Variance

Trying $\mathbb{E}[(X_n)^2]$ as a measure of expected distance from origin.

$$\begin{aligned}
\mathbb{E}[(X_n)^2] &= \mathbb{E}[(S_1 + S_2 + \cdots + S_n)^2] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} (S_i)^2 + \sum_{i \neq j} S_i \cdot S_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[(S_i)^2\right] + \sum_{i \neq j} \mathbb{E}\left[S_i \cdot S_j\right]
\end{aligned}$$

For first term: $\mathbb{E}[(S_i)^2] =$

S_i and S_j are independent for all $i \neq j$, so $\mathbb{E}[S_i \cdot S_j] =$

$$\implies \mathbb{E}[(X_n)^2] = \sum_{i=1}^{n} \mathbb{E}[(S_i)^2] = n$$

So "expected squared distance from start" is n.

- $\Rightarrow$ Does this mean "expected distance from start" is $\sqrt{n}$?
- $\Rightarrow$ Function of a r.v.: if $Z = X_n^2$, does $\mathbb{E}[\sqrt{Z}] = \sqrt{\mathbb{E}[Z]}$? Unfortunately, no...

Variance

Let $\mu_X = \mathbb{E}[X]$ be the mean of random variable X

$\Rightarrow$ Then $\mathbb{E}[(X-\mu_X)^2]$ is the expected squared distance from the mean

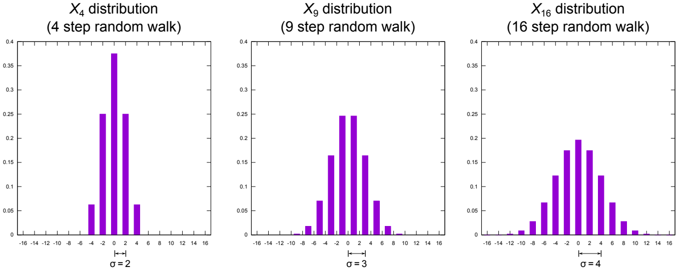
$\Rightarrow$ Note how μ_X notation helped – otherwise this is $\mathbb{E}[(X-\mathbb{E}[X])^2]$

Definition: The **Variance** of random variable X is $\text{Var}(X) = \mathbb{E}[(X-\mu_X)^2]$.

Units of variance are "distance squared" so we often look at:

$\sigma_X = \sqrt{\text{Var}(X)}$ – the **standard deviation** of X (same units as μ_x now)

Warning: Standard deviation in statistics is subtle (population vs sample stddev)



All have mean 0 (i.e., $\mathbb{E}[X_n] = 0$) – σ reflects "spread"

Variance

Variance is often easier to compute using the following theorem.

Theorem: $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Proof: Write μ_x for $\mathbb{E}[X]$ to highlight that it's just a constant value.

(*And to avoid ugly, confusing nested* $\mathbb{E}$*'s*)

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[(X-\mu_X)^2] \\
&= \mathbb{E}[X^2 - 2X\cdot\mu_X + (\mu_X)^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\cdot\mu_X + (\mu_X)^2 \\
&= \mathbb{E}[X^2] - 2(\mu_X)^2 + (\mu_X)^2 \\
&= \mathbb{E}[X^2] - (\mu_X)^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \qquad \square
\end{aligned}$$

Common Distributions: Variance of Bernoulli

Let $X \sim \text{Bernoulli}(p)$, so X is a random variable X with two values: $\{0,1\}$

$\mathbb{P}[X=1]=p$
$\mathbb{P}[X=0]=1-p$

$\mathbb{E}[X]=p$
$\text{Var}(X)=???$

Using the last theorem and the observation that $X^2=X$:

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

Back to:

Does square root of the mean distance squared equal the mean distance?

For $X \sim \text{Bernoulli}(1/4)$, $\mu=\mathbb{E}[X]=1/4$

$\mathbb{P}[X=0]=3/4$ and when $X=0$ we have $|X-\mu|=1/4$
$\mathbb{P}[X=1]=1/4$ and when $X=1$ we have $|X-\mu|=3/4$
$\Rightarrow\ \ \mathbb{E}[|X-\mu|]=\frac{3}{4}\cdot\frac{1}{4}+\frac{1}{4}\cdot\frac{3}{4}=\frac{3}{8}$ $\longleftarrow$ mean distance

$\text{Var}(X)=\frac{1}{4}\cdot\frac{3}{4}=\frac{3}{16}$, so $\sigma=\sqrt{\text{Var}(X)}=\frac{\sqrt{3}}{4}$ $\left(=\frac{\sqrt{12}}{8}\right)$ Not the same!

Common Distributions: Uniform

X is a random variable with $\mathbb{P}[X=i]=\frac{1}{n}$ for $i \in \{1,\ldots,n\}$. (roll an n-sided die)

$$\mathbb{E}[X]=\sum_{i=1}^{n} i\cdot\mathbb{P}[X=i]=\frac{1}{n}\sum_{i=1}^{n} i=\frac{1}{n}\frac{n(n+1)}{2}=\frac{n+1}{2}.$$

Also,

$$\begin{aligned}\mathbb{E}[X^2]&=\sum_{i=1}^{n} i^2\cdot\mathbb{P}[X=i]=\sum_{i=1}^{n} i^2\cdot\frac{1}{n}=\frac{1}{n}\sum_{i=1}^{n} i^2\\ &=\frac{1}{n}\cdot\frac{(n)(n+1)(2n+1)}{6} \qquad \text{“trust me” (or look it up)}\\ &=\frac{2n^2+3n+1}{6}\end{aligned}$$

This gives

$$\mathrm{Var}(X)=\frac{2n^2+3n+1}{6}-\left(\frac{n+1}{2}\right)^2=\frac{2\cdot(2n^2+3n+1)-3\cdot(n^2+2n+1)}{12}=\frac{n^2-1}{12}$$

Application: Fixed Points

Experiment: Random permutation of $1, 2, \ldots, n$ (or, hand back n assignments)

$X =$ number of fixed points (or # students that get their own assignment back)

Event $A_m =$ student m gets their own assignment back
$\Rightarrow$ $I_m =$ indicator variable for event A_m $(\mathbb{P}[I_m = 1] = \frac{1}{n})$

So $X = I_1 + \cdots + I_n$. Linearity of expectation gave $\mathbb{E}[X] = n \cdot \frac{1}{n} = 1$.

What about Var(X)? $\mathbb{E}[X^2] = \sum_{i=0}^{n} \mathbb{E}[(I_i)^2] + 2 \sum_{i<j} \mathbb{E}[I_i I_j]$

$\mathbb{E}[I_i^2] = 1 \cdot Pr[I_i = 1] + 0 \cdot Pr[I_i = 0] = \frac{1}{n}$

$\mathbb{E}[I_i I_j] = 1 \cdot \mathbb{P}[I_i = 1 \cap I_j = 1] + 0 \cdot \mathbb{P}[\text{“anything else”}] = 1 \times \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

Combining: $\mathbb{E}[X^2] = n \cdot \mathbb{E}[I_1^2] + 2 \cdot \binom{n}{2} \cdot \mathbb{E}[I_1 I_2] = n \cdot \frac{1}{n} + 2 \cdot \binom{n}{2} \cdot \frac{1}{n(n-1)} = 1 + 1 = 2$

Conclusion: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 2 - 1 = 1$

Informal conclusion: Not much “spread” in distribution (and doesn’t depend on n).

Fixed Points: Poll

True or False?

1. I_i and I_j are independent.
2. $\mathbb{E}[I_i I_j] = \mathbb{P}[I_i I_j = 1]$
3. $\mathbb{P}[I_i I_j] = \frac{(n-2)!}{n!}$
4. $I_i^2 = I_i$

Properties of Variance

Shifting or scaling a random variable has predictable effects on the variance.

Theorem: For any random variable X and constant c,

$$\mathrm{Var}(X+c) = \mathrm{Var}(X) \qquad \text{(shifting X)}$$

and

$$\mathrm{Var}(c \cdot X) = c^2 \cdot \mathrm{Var}(X) \qquad \text{(scaling X)}$$

Proof:

$$\begin{aligned}
\mathrm{Var}(X+c) &= \mathbb{E}[((X+c) - \mathbb{E}[X+c])^2] \\
&= \mathbb{E}[(X+c-\mathbb{E}[X]-c)^2] \\
&= \mathbb{E}[(X-\mathbb{E}(X))^2] = \mathrm{Var}(X) \\
\mathrm{Var}(cX) &= \mathbb{E}[(cX)^2] - (\mathbb{E}[cX])^2 \\
&= c^2 \cdot \mathbb{E}[X^2] - (c \cdot \mathbb{E}[X])^2 \\
&= c^2(\mathbb{E}[X^2] - (\mathbb{E}[X])^2) \\
&= c^2 \cdot \mathrm{Var}(X)
\end{aligned}$$

$\square$

Variance: Sum of Two Independent Random Variables

Theorem: If X and Y are independent random variables, then

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

Proof: We can evaluate the variance of $X+Y$ as follows:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 \\
&= \mathbb{E}[X^2+2XY+Y^2] - (\mathbb{E}[X]+\mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2 \\
&= \left(\mathbb{E}[X^2] - (\mathbb{E}[X])^2\right) + \left(\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2\right) + 2\,(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])
\end{aligned}$$

The last part of this formula, in red, has a special meaning, but for now...

X and Y are independent, so $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

$\implies \quad \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$

Therefore, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. $\square$

Variance: Sum of Independent Random Variables

Theorem: If $X_1, X_2, \dots, X_n$ are pairwise independent, then

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n).$$

Proof: For all $i = 1, \dots, n$, define $\widehat{X}_i = X_i - \mathbb{E}[X_i]$, so $\mathbb{E}[\widehat{X}_i] = 0$.

$\Rightarrow$ Sum of zero-mean r.v.s has zero mean: $\mathbb{E}[\widehat{X_1} + \widehat{X_2} + \cdots + \widehat{X_n}] = 0$

$\Rightarrow$ Variance is unaffected by shifts, so $\mathrm{Var}(X_i) = \mathrm{Var}(\widehat{X}_i) = \mathbb{E}[\widehat{X}_i^{\,2}]$.

$$\text{So:} \quad \mathrm{Var}(\widehat{X_1} + \widehat{X_2} + \cdots + \widehat{X_n}) = \mathbb{E}\left[\left(\widehat{X_1} + \widehat{X_2} + \cdots + \widehat{X_n}\right)^2\right]$$

$$= \mathbb{E}\left[\sum_i \widehat{X}_i^{\,2} + \sum_{i \neq j} \widehat{X}_i \widehat{X}_j\right]$$

$$= \sum_i \mathbb{E}[\widehat{X}_i^{\,2}] + \sum_{i \neq j} \mathbb{E}[\widehat{X}_i \widehat{X}_j]$$

By independence, $\mathbb{E}[\widehat{X}_i \widehat{X}_j] = \mathbb{E}[\widehat{X}_i] E[\widehat{X}_j] = 0$, and since $\mathbb{E}[\widehat{X}_i^{\,2}] = \mathrm{Var}(\widehat{X}_i)$ this simplifies to $\mathrm{Var}(\widehat{X_1}) + \mathrm{Var}(\widehat{X_2}) + \mathrm{Var}(\widehat{X_n})$.

Removing the shift between the $\widehat{X}_i$'s and the X_i's gives the theorem. $\square$

Sums of iid Random Variables and Binomial

Random variables that are independent and have the same distribution are called **independent, identically distributed** (or **iid**) random variables.

If $X_1, X_2, \ldots, X_n$ are iid random variables, and $X = X_1 + X_2 + \cdots + X_n$, then:

$\mathbb{E}[X] = n \cdot \mathbb{E}[X_1]$

$\text{Var}(X) = n \cdot \text{Var}(X_1)$

Distribution of X might be very different from X_i – this only gives mean and variance.

$X \sim \text{Binomial}(n, p) \quad \Longleftrightarrow \quad X = X_1 + \cdots + X_n$ where each $X_i \sim \text{Bernoulli}(p)$

We have already computed expected value and variance for Bernoulli, so:

Bernoulli		Binomial
$\mathbb{E}[X_1] = p$ $\text{Var}(X_1) = p(1-p)$	$\Longrightarrow$	$\mathbb{E}[X] = np$ $\text{Var}(X) = np(1-p)$

Can also be used for rolling n dice (sum of n iid uniform random variables)

Covariance

Definition: For random variables X and Y with $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$, the **covariance** of X and Y is

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

Notice that

$$\begin{aligned}
\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] &= \mathbb{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\
&= \mathbb{E}[XY] - \mu_Y\mathbb{E}[X] - \mu_X\mathbb{E}[Y] + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_Y\mu_X \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

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

Saw this on a previous slide where X and Y were independent, so was zero.

- $\Rightarrow$ How much different from zero? Think of as "how dependent are they"
- $\Rightarrow$ However, $\text{cov}(X, Y) = 0$ does *not* mean that X and Y are independent!

Also: $\text{cov}(X, X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{Var}(X)$

Covariance – Example 1

Consider the following joint distribution on variables X and Y:

X \ Y	-1	0	1
-1	0	1/5	0
0	1/5	1/5	1/5
1	0	1/5	0

Are X and Y independent?

Remember our earlier tip: Look at zero entries...

$\mathbb{E}[X] = \frac{1}{5} \cdot (-1) + \frac{3}{5} \cdot 0 + \frac{1}{5} \cdot 1 = 0 \quad \Rightarrow$ Similarly, $\mathbb{E}[Y] = 0$

$\mathbb{E}[XY]$? One of X or Y is always zero $\Rightarrow$ product is always zero

So: $\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 \cdot 0 = 0$

Covariance being zero does not mean variables are independent!

Covariance – Example 2

What can we tell if covariance is *not* zero? Consider:

$X \backslash Y$	-1	0	1
-1	0	1/7	1/7
0	1/7	1/7	1/7
1	1/7	1/7	0

$\mathbb{E}[X] = \mathbb{E}[Y] = 0$ – write it out, or note symmetry

$\mathbb{E}[XY]$? Only two cells with non-zero probability and non-zero product:

$\Rightarrow \quad \mathbb{E}[XY] = \frac{1}{7}\cdot(-1)\cdot 1 + \frac{1}{7}\cdot 1\cdot(-1) = -\frac{1}{7} - \frac{1}{7} = -\frac{2}{7}$

$\Rightarrow \quad \mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = -\frac{2}{7} - 0\cdot 0 = -\frac{2}{7}$

First observation: non-zero $\implies$ variables are not independent

Negative covariance means variables are "negatively correlated"

As X gets larger, Y tends to get smaller

If positively correlated: When X gets larger, Y tends get larger

Example: Consider $Z = X + Y$ and $\mathrm{cov}(X,Z)$

Covariance – Example 2 – continued

Sign of covariance give positive/negative correlation – what about magnitude?

X \ Y	-1	0	1
-1	0	1/7	1/7
0	1/7	1/7	1/7
1	1/7	1/7	0

W \ Z	-100	0	100
-100	0	1/7	1/7
0	1/7	1/7	1/7
100	1/7	1/7	0

$\text{cov}(X, Y) = -\frac{2}{7} \approx -0.285714..$

$\text{cov}(W, Z) = -\frac{20000}{7} \approx -2{,}857.14..$

Are W and Z any more correlated than X and Y?
Just a matter of scale or units....

Idea: For "how correlated are these variables" *normalize* units

Correlation

Definition: The correlation of X, Y, $\text{Corr}(X, Y)$ is

$$\text{Corr}(X,Y) : \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Theorem: $-1 \leq \text{Corr}(X,Y) \leq 1$. (proved in notes)

$X \backslash Y$	-1	0	1
-1	0	1/7	1/7
0	1/7	1/7	1/7
1	1/7	1/7	0

$W \backslash Z$	-100	0	100
-100	0	1/7	1/7
0	1/7	1/7	1/7
100	1/7	1/7	0

$\text{cov}(X,Y) = -\frac{2}{7}$
$\text{Var}(X) = \text{Var}(Y) = \mathbb{E}[X^2] = \frac{4}{7}$
$\text{Corr}(X,Y) = -\frac{1}{2}$

$\text{cov}(W,Z) = -\frac{20000}{7}$
$\text{Var}(W) = \text{Var}(Z) = \mathbb{E}[W^2] = \frac{40000}{7}$
$\text{Corr}(X,Z) = -\frac{1}{2}$

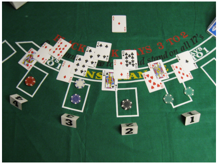
So equally (negatively) correlated — just values on a different scale. Nice!

Covariance – Example 3

Recall card counting example - are first and second cards correlated?

Good example because:
  Real world and interesting

Bad example because:
  Have to use decimal approximations
  (Yuck)



Define random variables:

V_1 ="Value of first card" (1,2,3,...,10,10,10,10,1,2,3...)
V_2 ="Value of second card"

$\mathbb{E}[V_1] = \mathbb{E}[V_2] \approx 6.53846$ (is exactly 85/13) $\Rightarrow$ $\mathbb{E}[V_1]\mathbb{E}[V_2] \approx 42.75148$

Expected product: $\mathbb{E}[V_1 V_2] \approx 42.55656$
$\Rightarrow$ $\text{cov}(V_1, V_2) = \mathbb{E}[V_1 V_2] - \mathbb{E}[V_1]\mathbb{E}[V_2] \approx -0.19492$

Negatively correlated! Larger first card tends to lead to smaller second card
  We're gonna be zillionaires! ...but wait... what about scale?

$\text{Var}(V_1) = \text{Var}(V_2) \approx 9.94083$ $\Rightarrow$ $\text{Corr}(V_1, V_2) \approx -0.019608$
Actually pretty weak... maybe we have to do something productive to be zillionaires...

Summary

Conditional expectation

Basic properties: $\mathbb{E}[X \mid Y = y] = \sum_x x \cdot \mathbb{P}[X = x \mid Y = y]$

Law of Total Expectation: $\mathbb{E}[X] = \sum_y \mathbb{E}[X \mid Y = y] \cdot \mathbb{P}[Y = y]$

Functions of a random variable

Sums and products

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$; *when independent,* $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Variance

Definitions and meaning

$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = (\mathbb{E}[X])^2 - \mathbb{E}[X^2]$

Variance of common distributions: Bernoulli, Binomial, ...

Properties of variance

$\mathrm{Var}(X + c) = \mathrm{Var}(x)$; $\mathrm{Var}(c \cdot X) = c^2 \cdot \mathrm{Var}(X)$; *Ind*: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Covariance

Definition and meaning: $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Positively and negatively correlated random variables

Correlation: Scaled covariance: $\mathrm{Corr}(X, Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$