CS109: Independent Random Variables

Independent Discrete RVs

Independent discrete RVs

Recall the definition of independent events E and F:

$$P(EF) = P(E)P(F)$$

Two discrete random variables X and Y are **independent** if:

for all x, y:

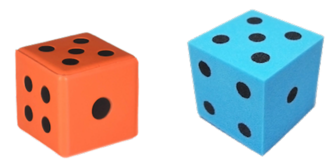
$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Different notation, same idea:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

- Intuitively: knowing value of X tells us nothing about the distribution of Y (and vice versa)
- If two variables are not independent, they are called **dependent**.

Dice (after all this time, still our friends)



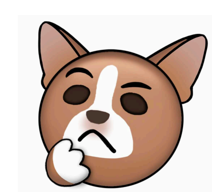
Let: D_1 and D_2 be the outcomes of two rolls
$S = D_1 + D_2$, the sum of two rolls

- Each roll of a fair, 6-sided die is an independent trial.
- Random variables D_1 and D_2 are independent.

1. Are events $(D_1 = 1)$ and $(S = 7)$ independent?

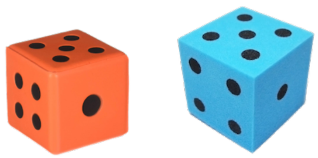
2. Are events $(D_1 = 1)$ and $(S = 5)$ independent?

3. Are random variables D_1 and S independent?



Dice (after all this time, still our friends)

Let: D_1 and D_2 be the outcomes of two rolls
$S = D_1 + D_2$, the sum of two rolls



- Each roll of a 6-sided die is an independent trial.
- Random variables D_1 and D_2 are independent.

1. Are events $(D_1 = 1)$ and $(S = 7)$ independent? ✅
2. Are events $(D_1 = 1)$ and $(S = 5)$ independent? ❌
3. Are random variables D_1 and S independent? ❌

All events $(X = x, Y = y)$ must be independent for X, Y to be independent RVs.

What about continuous random variables?

Continuous random variables can also be independent! We'll see this later.

Today's goal:

How can we model <u>sums</u> of discrete random variables?

Big motivation: Model total successes observed over multiple experiments

Sums of independent Binomial RVs

Sum of independent Binomials

$X \sim \text{Bin}(n_1, p)$
$Y \sim \text{Bin}(n_2, p)$
X, Y independent

$$X + Y \sim \text{Bin}(n_1 + n_2, p)$$

Intuition:

- Each trial in X and Y is independent and has same success probability p
- Define $Z =$# successes in $n_1 + n_2$ independent trials, each with success probability p. $Z \sim \text{Bin}(n_1 + n_2, p)$, and $Z = X + Y$

Holds in general case:

$X_i \sim \text{Bin}(n_i, p)$
X_i independent for $i = 1, \dots, n$

$$\sum_{i=1}^{n} X_i \sim \text{Bin}\left(\sum_{i=1}^{n} n_i, p\right)$$

If only it were always so simple...

Convolution: Sum of independent Poisson RVs

Convolution: Sum of independent random variables

For any discrete random variables X and Y:

$$P(X+Y=n)=\sum_k P(X=k,Y=n-k)$$

More specifically, for independent discrete random variables X and Y:

$$P(X+Y=n)=\sum_k \underbrace{P(X=k)P(Y=n-k)}_{\text{the convolution of } p_X \text{ and } p_Y}$$

Insight into convolution

For independent discrete random variables X and Y:

$$P(X+Y=n) = \sum_k P(X=k)P(Y=n-k)$$

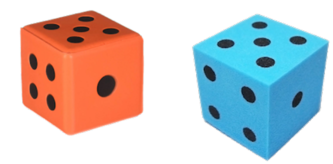
the convolution of p_X and p_Y

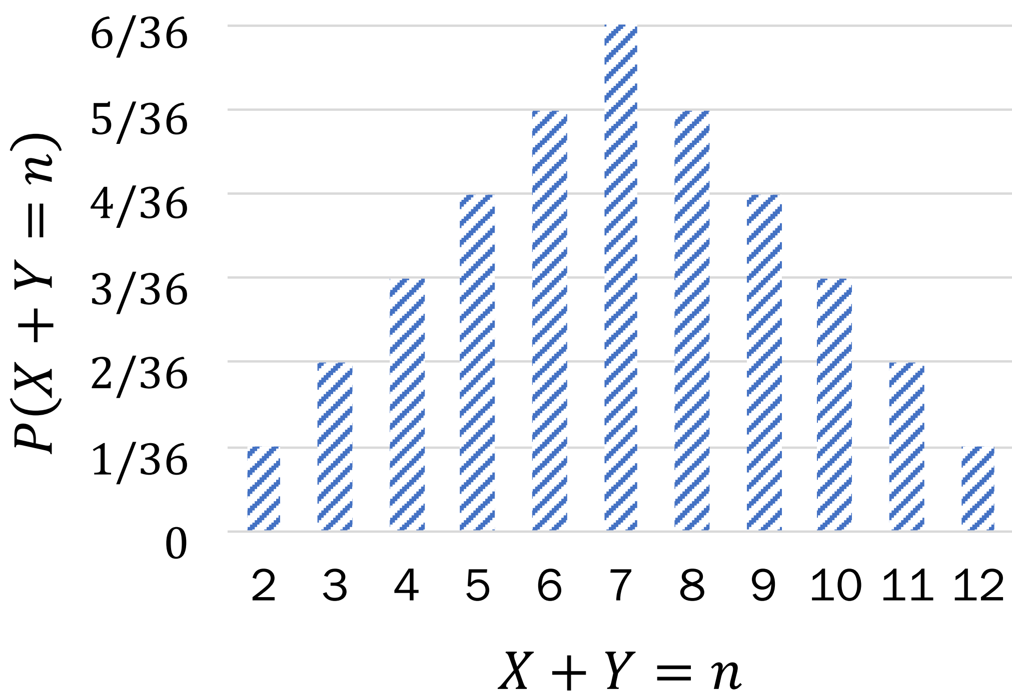
Suppose X and Y are independent, both with support $\{0, 1, \dots, n, \dots\}$:

		X						
		0	1	2	…	n	$n+1$	…
Y	0					✓		
	…				…			
	$n-2$			✓				
	$n-1$		✓					
	n	✓						
	$n+1$							
	…							

- ✓: event where $X+Y=n$
- Each event has probability:
  $P(X=k, Y=n-k)$
  $= P(X=k)P(Y=n-k)$
  (because X, Y are independent)
- $P(X+Y=n)$ = sum of mutually exclusive events

Sum of 2 die rolls





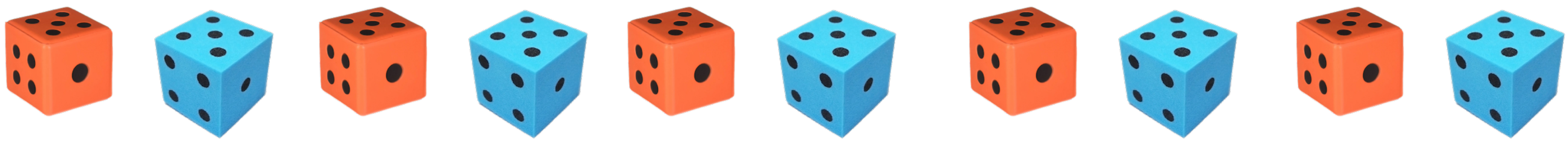
The distribution of a sum of 2 die rolls is a convolution of 2 PMFs.

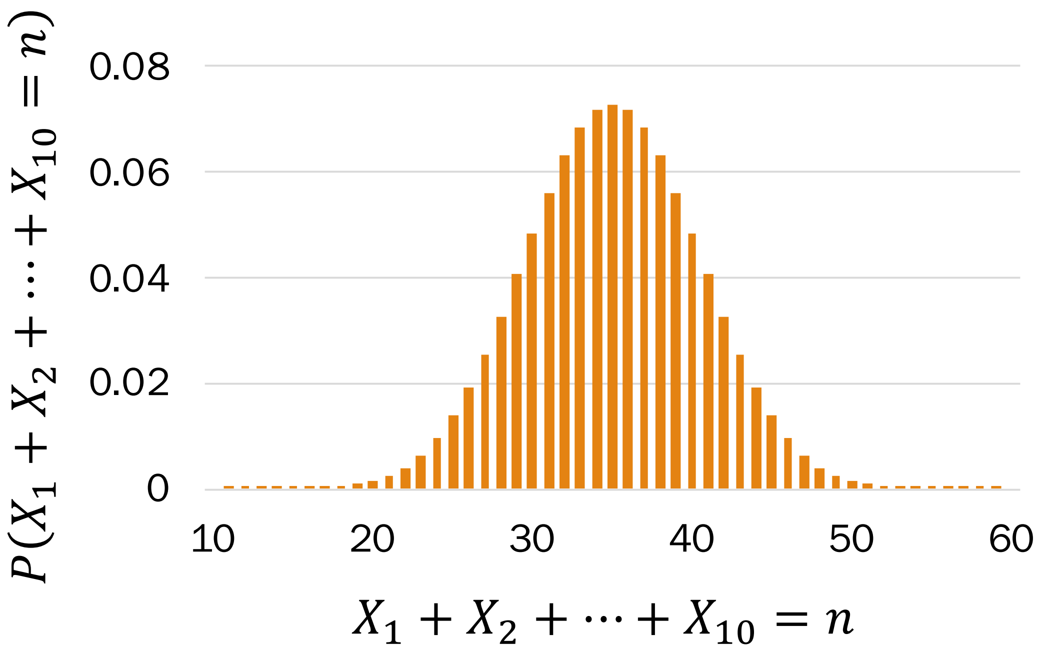
Examples:

$$P(X+Y=2) = P(X=1)P(Y=1)$$

$$\begin{aligned} P(X+Y=4) = P(X=1)P(Y=3) &+ \\ P(X=2)P(Y=2) &+ \\ P(X=3)P(Y=1) & \end{aligned}$$

Sum of 10 die rolls (fun preview)





The distribution of a sum of <u>10</u> die rolls is a convolution <u>10</u> PMFs.

Looks sort of Gaussian, eh?

(more on this in Week 7, but spoiler alert: it is!)

Sum of independent Poissons

$X \sim \text{Poi}(\lambda_1), Y \sim \text{Poi}(\lambda_2)$
X, Y independent

$\Rightarrow$

$X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$

Proof (just for reference):

$$P(X+Y=n) = \sum_k P(X=k)P(Y=n-k)$$

X and Y independent, convolution

$$= \sum_{k=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} = e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k \lambda_2^{n-k}}{k!\,(n-k)!}$$

PMF of Poisson RVs

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \lambda_1^k \lambda_2^{n-k} = \underbrace{\frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1+\lambda_2)^n}_{\text{Poi}(\lambda_1+\lambda_2)}$$

Binomial Theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

General sum of independent Poissons

Holds in general case:

$$X_i \sim \text{Poi}(\lambda_i)$$
$$X_i \text{ independent for } i = 1, \ldots, n$$

$$\Rightarrow \quad \sum_{i=1}^{n} X_i \sim \text{Poi}\left(\sum_{i=1}^{n} \lambda_i\right)$$

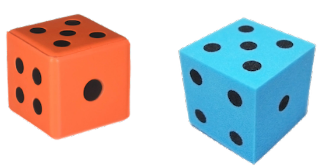
Independent discrete RVs

Two discrete random variables X and Y are **independent** if:

for all x, y:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$



The sum of 2 dice and the outcome of 1st die are **dependent RVs.**

Important: Joint PMF must decompose into product of marginal PMFs for ALL values of X and Y for X, Y to be independent RVs.

Deliberate Binomials

Let’s take a 90-second break to check email.

Slide 18 presents a few questions about sums of Binomial distributions to distract you from your email.

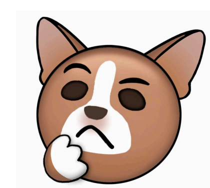
Once you’ve cleared your inbox, we’ll come back to these questions.

Coin flips

Flip a coin with probability p of "heads" a total of $n + m$ times.

Let X = number of heads in first n flips. $X \sim \text{Bin}(n, p)$
Y = number of heads in next m flips. $Y \sim \text{Bin}(m, p)$
Z = total number of heads in $n + m$ flips.

1. Are X and Z independent?
2. Are X and Y independent?



Coin flips

Flip a coin with probability p of "heads" a total of $n + m$ times.

Let X = number of heads in first n flips. $X \sim \text{Bin}(n, p)$
Y = number of heads in next m flips. $Y \sim \text{Bin}(m, p)$
Z = total number of heads in $n + m$ flips.

1. Are X and Z independent? ❌ Counterexample: What if $Z = 0$?
2. Are X and Y independent? ✅

$$P(X = x, Y = y) = P\left(\begin{array}{c}\text{first } n \text{ flips have } x \text{ heads} \\ \text{and next } m \text{ flips have } y \text{ heads}\end{array}\right)$$

# of mutually exclusive outcomes in event: $\binom{n}{x}\binom{m}{y}$

$P(\text{each outcome})$
$= p^x(1-p)^{n-x}p^y(1-p)^{m-y}$

$$= \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}$$

$$= P(X = x)P(Y = y)$$

This probability (found through counting) is the product of the marginal PMFs.

Sum of independent Poissons

$X \sim \text{Poi}(\lambda_1), Y \sim \text{Poi}(\lambda_2)$
X, Y independent

$$X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$$

- n servers with independent number of requests/minute
- Server i's requests each minute can be modeled as $X_i \sim \text{Poi}(\lambda_i)$

What is the probability that the total number of web requests received at all servers in the next minute exceeds 10?

Cogitate Binomial & Poisson

Mondays are hard, so let's take another break to launch our Instagram stories so you can let it run all day to clear them out.

Slide 22 presents two separate questions—both substantial—that combine several key concepts from today and the preceding week.

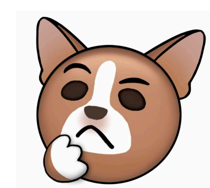
Take 90 more seconds and we'll work through them together afterwards.

Independent questions

1. Let $X \sim \text{Bin}(30, 0.01)$ and $Y \sim \text{Bin}(50, 0.02)$ be independent RVs.
   - How do we compute $P(X + Y = 2)$ using a Poisson approximation?
   - How do we compute $P(X + Y = 2)$ exactly?

2. Let $N = \#$ of requests to a web server per day. Suppose $N \sim \text{Poi}(\lambda)$.
   - Each request independently comes from a human (w.p p) or a bot $(1 - p)$.
   - Let X be # of human requests/day, and Y be # of bot requests/day.

Are X and Y independent? What are their marginal PMFs?



1. Approximating the sum of independent Binomial RVs

Let $X \sim \text{Bin}(30, 0.01)$ and $Y \sim \text{Bin}(50, 0.02)$ be independent RVs.

- How do we compute $P(X + Y = 2)$ using a Poisson approximation?

- How do we compute $P(X + Y = 2)$ exactly?

$$P(X + Y = 2) = \sum_{k=0}^{2} P(X = k)P(Y = 2 - k)$$

$$= \sum_{k=0}^{2} \binom{30}{k} 0.01^k (0.99)^{30-k} \binom{50}{2-k} 0.02^{2-k} 0.98^{50-(2-k)} \approx 0.2327$$

2. Web server requests

Let N = # of requests to a web server per day. Suppose $N \sim \text{Poi}(\lambda)$.

- Each request independently comes from a human (prob. p), or bot $(1-p)$.
- Let X be # of human requests/day, and Y be # of bot requests/day.

Are X and Y independent? What are their marginal PMFs?

$$P(X=n, Y=m) = P(X=n, Y=m \mid N=n+m)P(N=n+m) + P(X=n, Y=m \mid N \neq n+m)P(N \neq n+m)$$

Law of Total Probability

$$= P(X=n \mid N=n+m)P(Y=m \mid X=n, N=n+m)P(N=n+m)$$

Chain Rule

$$= \binom{n+m}{n} p^n (1-p)^m \cdot 1 \cdot e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!}$$

Given $N=n+m$ indep. trials, $X|N=n+m \sim \text{Bin}(n+m, p)$

$$= \frac{(n+m)!}{n!\,m!} e^{-\lambda} \frac{(\lambda p)^n \big(\lambda(1-p)\big)^m}{(n+m)!} = e^{-\lambda p} \frac{(\lambda p)^n}{n!} \cdot e^{-\lambda(1-p)} \frac{\big(\lambda(1-p)\big)^m}{m!}$$

$$= P(X=n)P(Y=m)$$

where $X \sim \text{Poi}(\lambda p)$, $Y \sim \text{Poi}\big(\lambda(1-p)\big)$

Yes, X and Y are independent!

Interlude for Announcements

Announcements

<u>Quiz #1</u>

Grades/solutions today!

<u>Problem Set 3</u>

Due: Friday 2/12 1pm
Covers: Up to and including Lecture 10

<u>CS109 Challenge</u>

Make up any part of your grade
Details later this week!

Interesting probability news

Column: Did Astros beat the Dodgers by cheating? The numbers say no

"...new analyses of the Astros' 2017 season by baseball's corps of unofficial statisticians — "[sabermetricians](#)," to the sport — indicate that the Astros didn't gain anything from their cheating; in fact, it may have hurt them."

https://www.latimes.com/business/story/2020-02-27/astros-cheating-analysis

https://www.theguardian.com/sport/2020/jan/17/houston-astros-sign-stealing-cheating-scandal

Independence of multiple random variables

Recall independence of n events $E_1, E_2, \dots, E_n$:

for $r = 1, \dots, n$:
for every subset $E_1, E_2, \dots, E_r$:
$$P(E_1, E_2, \dots, E_r) = P(E_1)P(E_2) \cdots P(E_r)$$

We have independence of n discrete random variables $X_1, X_2, \dots, X_n$ if
for all $x_1, x_2, \dots, x_n$:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

Independence is symmetric

If X and Y are independent random variables, then
X is independent of Y, and Y is independent of X

Let N be the number of times you roll 2 dice repeatedly until a 4 is rolled (the player wins), or a 7 is rolled (the player loses).

Let X be the value (4 or 7) of the final throw.

- Is N independent of X? $P(N = n | X = 7) = P(N = n)$?
  $P(N = n | X = 4) = P(N = n)$?
- Is X independent of N? $P(X = 4 | N = n) = P(X = 4)$?
  $P(X = 7 | N = n) = P(X = 7)$? (yes, easier to intuit)

Redux: Independence is not always intuitive, but it is **always** symmetric.

LIVE

Statistics on Two RVs

Expectation and Covariance

In real life, we often have many RVs interacting at once.

- We've seen some simpler cases (e.g., sum of independent Poissons).
- Computing joint PMFs in general is hard!
- But often you don't need to model joint RVs completely.

Instead, we'll focus next on reporting ***statistics*** of multiple RVs:

- Expectation: sum of RV expectations == expectation of RV sums
- Covariance: measure of how two RVs vary with each other (coming soon)

Properties of Expectation, extended to two RVs

1. Linearity:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

2. Expectation of a sum = sum of expectation:

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

(we've seen this; we'll prove this next)

3. Unconscious statistician:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

True for both independent and dependent random variables!

Proof of expectation of a sum of RVs

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

$$E[X+Y] = \sum_x \sum_y (x+y) p_{X,Y}(x,y)$$

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

$$= \sum_x \sum_y x p_{X,Y}(x,y) + \sum_x \sum_y y p_{X,Y}(x,y)$$

$$= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y)$$

Linearity of summations
(cont. case: linearity of integrals)

$$= \sum_x x p_X(x) + \sum_y y p_Y(y)$$

Marginal PMFs for X and Y

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

Expectations of common RVs: Binomial

Review

$X \sim \text{Bin}(n, p) \quad E[X] = np$

\# of successes in n independent trials with probability of success p

Recall: $\text{Bin}(1, p) = \text{Ber}(p)$

$$X = \sum_{i=1}^{n} X_i$$

Let $X_i = i$th trial is heads
$X_i \sim \text{Ber}(p), E[X_i] = p$

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$