12: Independent RVs

Lisa Yan

May 1, 2020

Quick slide reference

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:

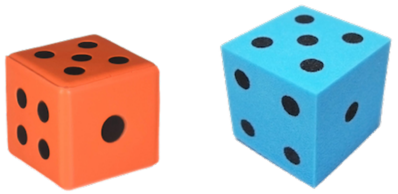
$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

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 6-sided die is an independent trial.
- Random variables D_1 and D_2 are independent.

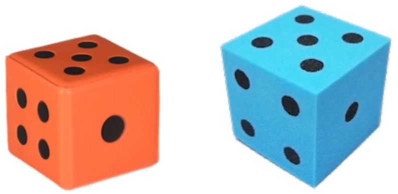
1. Are events $(D_1 = 1)$ and $(S = 7)$ independent?

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?

Event $(S = 7)$: $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$ $|(S=7)| = 6$

$P(D_1=1)\,P(S=7) = 1/6 \cdot 1/6$
$= P(D_1=1, S=7) = 1/36$

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

Event $(S = 5)$: $\{(1,4), (2,3), (3,2), (4,1)\}$ $|(S=5)| = 4$

$P(D_1=1)\,P(S=5) = \frac{1}{6} \cdot \frac{1}{9}$
$\neq P(D_1=1, S=5) = \frac{1}{36}$

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 sums 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

$\Rightarrow$

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

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 also $Z = X + Y$

Holds in general case:

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

$\Rightarrow$

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

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:

known n

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

In particular, for independent discrete random variables X and Y:

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

known PMF @k

the convolution of p_X and p_Y

$Z = X + Y$

$P(Z = k) = \ldots$

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 $\{0, 1, \dots, n, \dots\}$: ☆

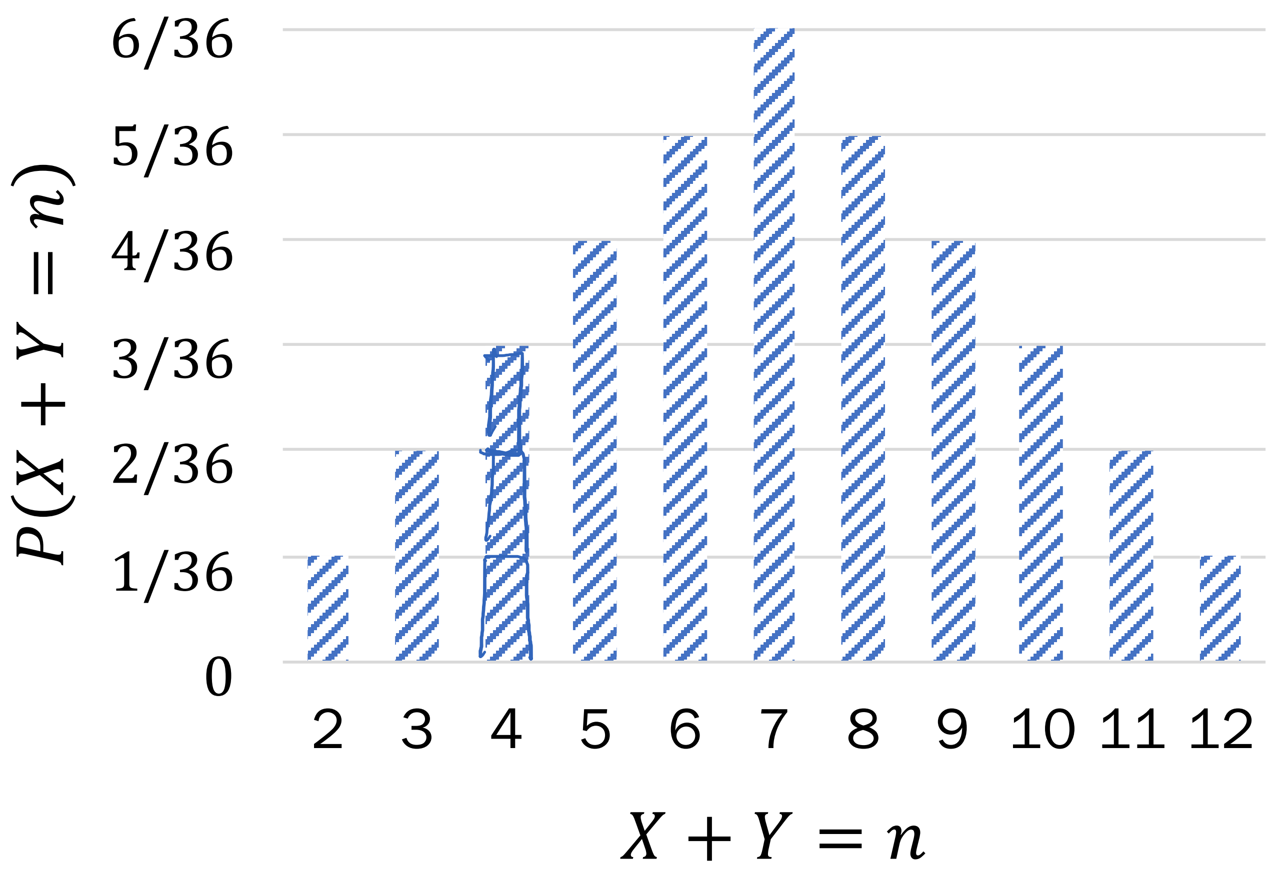
	X						
Y	0	1	2	…	n	$n+1$	…
0					✓		
…				…			
$n-2$			✓				
$n-1$		✓ (1, n−1)					
n	✓ (0, n)						
$n+1$							
…							

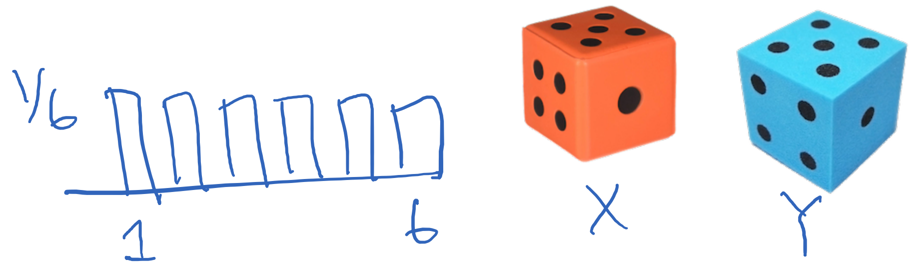
- ✓: event where $X+Y=n$

$$P(X+Y=n)=\sum_{k=0 \leftarrow \min X}^{n \leftarrow \max X,\ \text{b/c } \min Y=0} P(X=k, Y=n-k)$$

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

Sum of 2 dice rolls





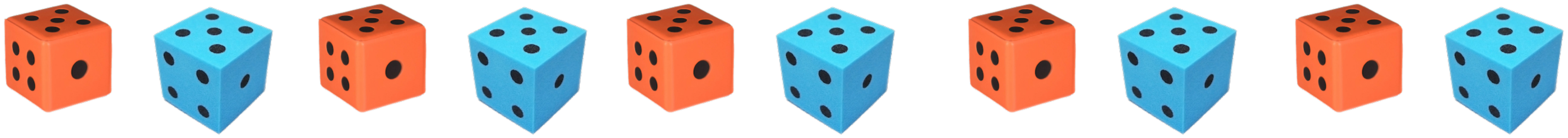
The distribution of a sum of 2 dice rolls is a convolution of 2 PMFs.

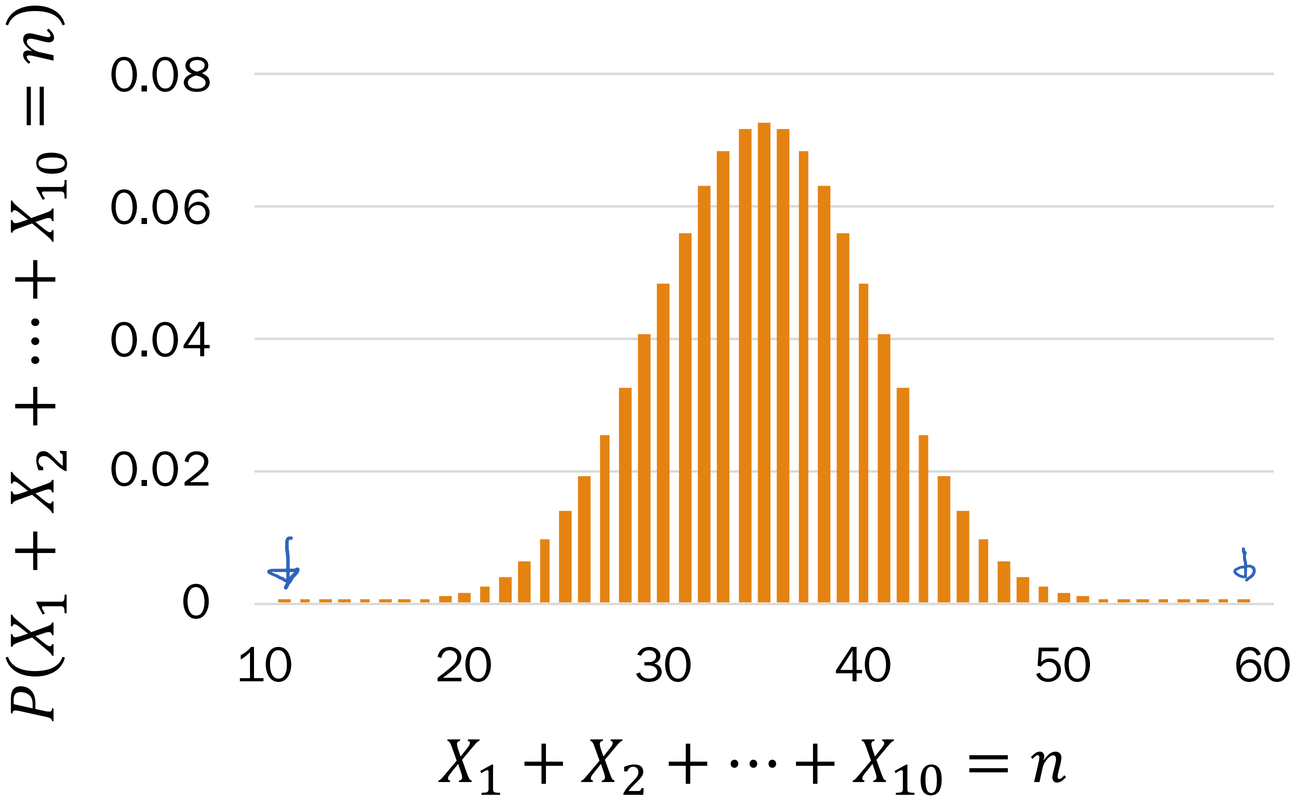
$\{(1,3),(2,2),(3,1)\}$

Example:

$$P(X+Y=4) = P(X=1)P(Y=3) + P(X=2)P(Y=2) + P(X=3)P(Y=1)$$

Sum of 10 dice rolls (fun preview)





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

Looks kinda Normal...???
(more on this in Week 7)

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 independent for $i = 1, \ldots, n$

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

(live)
12: Independent RVs

Lisa Yan

May 1, 2020

Quiz #1: Closing remarks

however...

Quiz #1: Closing remarks

Learning goals:

- This quiz was designed for a range of students to test their knowledge.
- We have kept the rigor the same as regular quarters of CS109.
- 2-hour exam length + typesetting, to be completed in 24 hours

A mid-quarter feedback form will be going out sometime next week

- How the course is going overall
- How you are doing overall
- Quiz 1 feedback (start time, duration), so that we can improve

A word about the Honor Code.
https://communitystandards.stanford.edu/policies-and-guidance/honor-code

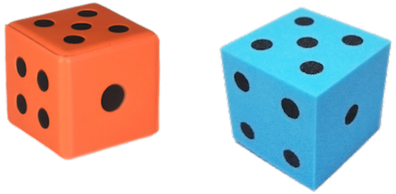
Independent discrete RVs

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.

Think

Slide 22 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/46502

Think by yourself: 2 min

🤔 (by yourself)

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?

(by yourself)

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? ✅

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

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

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

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

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

HHH TTTT HH TTTT

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?

$P(X > 10)$ $\quad X = \sum_{i=1}^{n} X_i \quad X \sim \text{Poi}(\lambda)$

$\lambda = \sum_{i=1}^{n} \lambda_i$

Breakout Rooms

Slide 47 has two questions to go over in groups.

ODD breakout rooms: Try question 1
EVEN breakout rooms: Try question 2

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/46502

Breakout rooms: 5 min. Introduce yourself!

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 (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?

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?

$X \approx A \sim \text{Poi}(\lambda_A = 30(0.01) = 0.3)$ $\quad A + B \sim \text{Poi}(\underbrace{1.3}_{\lambda})$

$Y \approx B \sim \text{Poi}(\lambda_B = 50(0.02) = 1)$

$P(X+Y=2) \approx P(A+B=2) = \frac{\lambda^2}{2!} e^{-\lambda} \approx 0.2302$

- How do we compute $P(X + Y = 2)$ exactly? $\quad X+Y \sim \text{Bin}(80, 0.013?)$?

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

$\{X=0, Y=2;\ X=1, Y=1;\ X=2, Y=0\}$

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

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

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

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

Yes, X and Y are independent!

Interlude for jokes/announcements

Announcements

Quiz #1

Grades/solutions: Next week

Problem Set 3

Due: ~~Monday~~ Friday 5/8 10am

Covers: Up to and including Lecture 11

CS109 Contest

Make up any part(s) of your grade

Details Next 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

CS109 Current Events Spreadsheet

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 $r = 1, \dots, n$:
for all subsets $x_1, x_2, \dots, x_r$:

$$P(X = x_1, X = x_2, \dots, X_r = x_r) = \prod_{i=1}^{r} P(X_i = x_i)$$

Independence is symmetric

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

...duh?

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)

In short: Independence is not always intuitive, but it is symmetric.

Statistics of 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 of sums (you've seen some of this)
- Covariance: a measure of how two RVs vary with *each other*

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

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

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

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
$X_i \sim \text{Ber}(p), E[X_i] = p$

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

Think

Slide 40 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/46502

Think by yourself: 2 min

(by yourself)

Expectations of common RVs: Negative Binomial

$Y \sim \text{NegBin}(r, p) \quad E[Y] = \frac{r}{p}$

# of independent trials with probability of success p until r successes

Recall: $\text{NegBin}(1, p) = \text{Geo}(p)$

$$Y = \sum_{i=1}^{?} Y_i$$

1. How should we define Y_i?
2. How many terms are in our summation?

(by yourself)

Expectations of common RVs: Negative Binomial

$Y \sim \text{NegBin}(r, p) \quad E[Y] = \frac{r}{p}$

# of independent trials with probability of success p until r successes

Recall: $\text{NegBin}(1, p) = \text{Geo}(p)$

$$Y = \sum_{i=1}^{?} Y_i$$

Let Y_i = # trials to get ith success (after $(i-1)$th success)

$Y_i \sim \text{Geo}(p), E[Y_i] = \frac{1}{p}$

$$E[Y] = E\left[\sum_{i=1}^{r} Y_i\right] = \sum_{i=1}^{r} E[Y_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

$P(X=k, Y=j)$

$$\frac{\lambda_1^k e^{-\lambda_1}}{k!} \qquad \frac{\lambda_2^{j+k} e^{-\lambda_2}}{(j+k)!}$$

$P(\text{2 girls} \mid \text{1st girl}) \Rightarrow$ GG, GB

$\hookrightarrow \dfrac{P(\text{2 girls } \& \text{ 1st girl})}{P(\text{1 girl})} = \dfrac{1/4}{2/4}$

$P(\text{both boys} \mid \geq 1 \text{ boy}) = \dfrac{P(\text{both boys } \& \geq 1 \text{ boy})}{P(\geq 1 \text{ boy})} = \dfrac{1/4}{3/4}$

GG GB

BG BB