12: Independent RVs

Jerry Cain
April 22, 2022

Table of Contents

Sums of independent Binomial 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:

joint pmf $p_{X,Y}(x, y) = p_X(x)^{\text{marg}}\, p_Y(y)^{\text{marg}}$

- 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**.

Sum of independent Binomials

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

$\Rightarrow$

Sum

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

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…

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.

$Z = 0$

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? ✅

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

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

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

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

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

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

Convolution: Sum of independent Poisson RVs

Convolution: Sum of independent random variables

For any discrete random variables X and Y:

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

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

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

Insight into convolution

For independent discrete random variables X and Y:

$P(X=0)P(Y=n)$

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

the convolution of p_X and p_Y

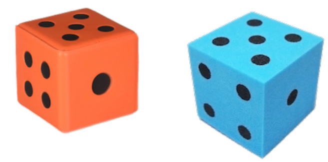
Suppose X and Y are independent, both with support $\{0,1,\ldots,n,\ldots\}$:

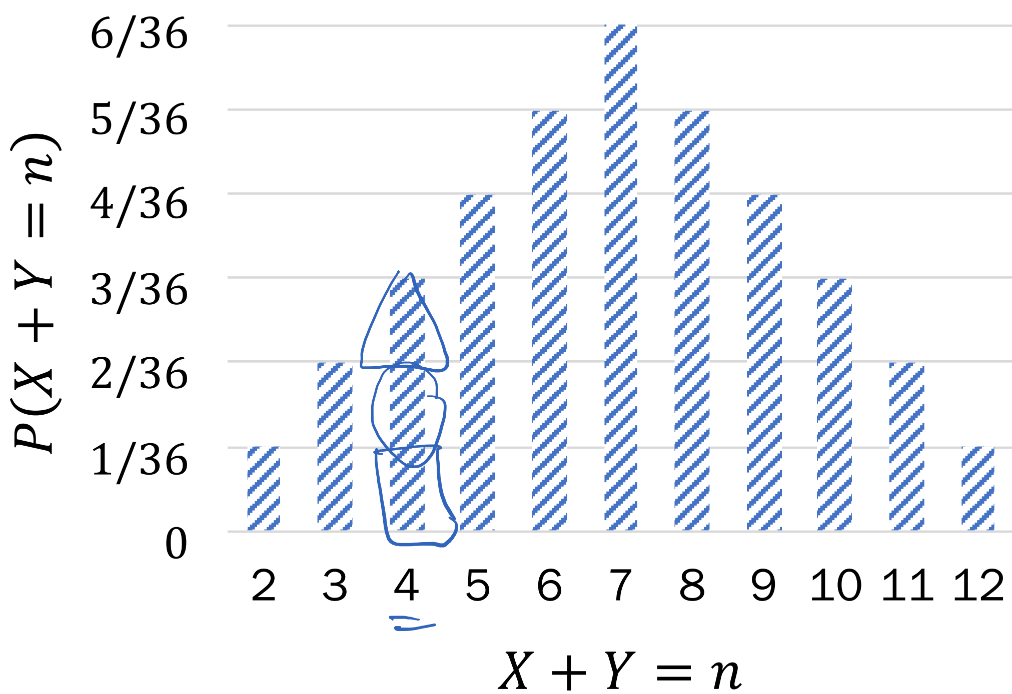
$\sum_{k=0}^{n} P(X=k)P(Y=n-k)$

Y \ X	0	1	2	…	n	$n+1$	…
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 dice rolls



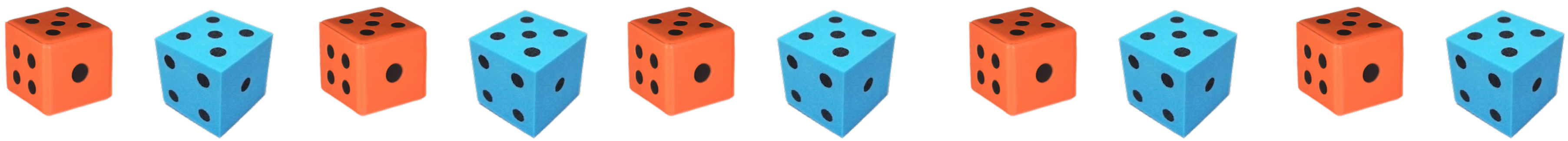


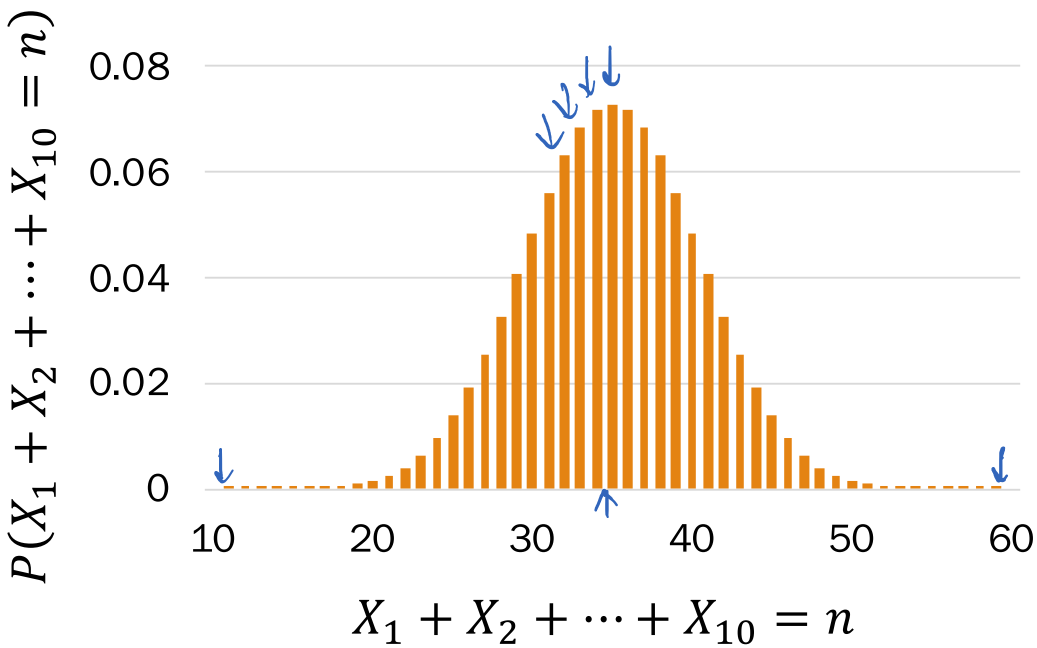
The distribution of a sum of 2 dice rolls is a convolution of 2 PMFs.

Example:

$$\boxed{P(X+Y=4) =}$$
$$P(X=1)P(Y=3)$$
$$+\,P(X=2)P(Y=2)$$
$$+\,P(X=3)P(Y=1)$$

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 $i = 1, \ldots, n$

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

Sum of independent Poissons

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

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

- 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?

$X = \sum_{i=1}^{n} X_i \qquad X \sim \text{Poi}(\lambda), \ \lambda = \sum_{i=1}^{n} \lambda_i$

$P(X > 10) = e^{-\lambda} \sum_{k=11}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda}\left[\frac{\lambda^{11}}{11!} + \frac{\lambda^{12}}{12!} + \frac{\lambda^{13}}{13!} + \cdots\right]$

$= 1 - \sum_{k=0}^{10} \frac{\lambda^k}{k!}$

Exercises

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 $N \sim \text{Poi}(\lambda)$. Ignore for now
   - 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.

$X \approx A \sim \text{Poi}(0.3)$ $\quad$ $Y \approx B \sim \text{Poi}(1.0)$

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

$$P(X+Y=2) \approx P(\underbrace{A+B}_{S \sim \text{Poi}(1.3)}=2) = P(S=2)$$

$$= e^{-\lambda}\frac{\lambda^2}{2!}$$

$$= e^{-1.3}\frac{1.3^2}{2!} = .2302$$

- 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?

$$
\begin{aligned}
P(X=x, Y=y) &= P(X=x, Y=y \mid N=x+y)P(N=x+y) \\
&\quad + P(X=x, Y=y \mid N \neq x+y)P(N \neq x+y) && \text{Law of Total Probability} \\
&= \overbrace{P(X=x \mid N=x+y)}^{\text{Bin}}P(Y=y \mid X=x, N=x+y)P(N=x+y) && \text{Chain Rule} \\
&= \binom{x+y}{x} p^x (1-p)^y \quad \cdot \quad 1 \quad \cdot e^{-\lambda}\frac{\lambda^{x+y}}{(x+y)!} \\
&= \frac{(x+y)!}{x!\,y!} e^{-\lambda} \frac{(\lambda p)^x \big(\lambda(1-p)\big)^y}{(x+y)!} = e^{-\lambda p}\frac{(\lambda p)^x}{x!} \cdot e^{-\lambda(1-p)}\frac{\big(\lambda(1-p)\big)^y}{y!} \\
&= P(X=x)P(Y=y)
\end{aligned}
$$

Given $N = x+y$ indep. trials, $X \mid N = x+y \sim \text{Bin}(x+y, p)$

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

Yes, X and Y are independent!

Independence of multiple random variables

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

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

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

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

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.