12: Independent RVs

Jerry Cain
April 23, 2021
## Quick slide reference

<table>
<thead>
<tr>
<th>3</th>
<th>Independent discrete RVs</th>
<th>12a_independent_rvs</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Sums of Independent Binomial RVs</td>
<td>12b_sum_binomial</td>
</tr>
<tr>
<td>10</td>
<td>Convolution: Sum of independent Poisson RVs</td>
<td>12c_discrete_conv</td>
</tr>
<tr>
<td>17</td>
<td>Exercises</td>
<td>LIVE</td>
</tr>
</tbody>
</table>
Independent Discrete RVs
Independent discrete RVs

Recall the definition of independent events $E$ and $F$:

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

\[
P(X = x, Y = y) = P(X = x)P(Y = y)
\]

or

\[
p_{X,Y}(x, y) = p_X(x)p_Y(y)
\]

for all $x, y$:

- 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?  ✔

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 **sums** of discrete random variables?

Big motivation: Model total successes observed over multiple experiments
Sums of independent Binomial RVs
Sum of independent Binomials

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

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

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 also \( Z = X + Y \)

Holds in general case:
- \( X_i \sim \text{Bin}(n_i, p) \)
  - \( X_i \) independent for \( i = 1, \ldots, n \)

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

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

the convolution of $p_X$ 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 $\{0, 1, \ldots, n, \ldots\}$:

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$0$</th>
<th>$1$</th>
<th>$2$</th>
<th>$\ldots$</th>
<th>$n$</th>
<th>$n + 1$</th>
<th>$\ldots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\checkmark$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $\checkmark$: 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)
- $P(X + Y = n) = \text{sum of mutually exclusive events}$
The distribution of a sum of 2 dice rolls is a convolution of 2 PMFs.

Example:
\[ P(X + Y = 4) = P(X = 1)P(Y = 3) + P(X = 2)P(Y = 2) + P(X = 3)P(Y = 1) \]
The distribution of a sum of 10 dice rolls is a convolution of 10 PMFs.

Looks kinda Normal...???
(more on this in Week 7)
Sum of independent Poissons

\[ X \sim \text{Poi}(\lambda_1), Y \sim \text{Poi}(\lambda_2) \]
\[ X, Y \text{ independent} \]
\[ X + Y \sim \text{Poi}(\lambda_1 + \lambda_2) \]

Proof (just for reference):

\[
P(X + Y = n) = \sum_{k=0}^{n} P(X = k)P(Y = n - k)
\]
\[
= \sum_{k=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} = e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k \lambda_2^{n-k}}{k! (n-k)!}
\]
\[
= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k! (n-k)!} \frac{\lambda_1^k \lambda_2^{n-k}}{k! (n-k)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n
\]

\[ \text{PMF of Poisson RVs} \]

\[ X \text{ and } Y \text{ independent, convolution} \]

\[ \text{Binomial Theorem:} \]
\[ (a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \]
General sum of independent Poissons

Holds in general case:

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

\[ \sum_{i=1}^{n} X_i \sim \text{Poi} \left( \sum_{i=1}^{n} \lambda_i \right) \]
12: Independent RVs (live)

Jerry Cain
April 23, 2021
Quiz #1 is D-O-N-E done!
Independent discrete RVs

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

for all $x, y$:

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

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

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

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

Slide 21 has a question to go over by yourself.

Post any clarifications here!

https://edstem.org/us/courses/5090/discussion/377748

Think by yourself: 2 min
Coin flips

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

Let $X = \text{number of heads in first } n \text{ flips. } X \sim \text{Bin}(n, p)$

$Y = \text{number of heads in next } m \text{ flips. } Y \sim \text{Bin}(m, p)$

$Z = \text{total number of heads in } n + m \text{ flips.}$

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? ✗
   
   Countercase: What if $Z = 0$?

2. Are $X$ and $Y$ independent? ✓

\[
P(X = x, Y = y) = P \begin{pmatrix} 
\text{first } n \text{ flips have } x \text{ heads} \\
\text{and next } m \text{ flips have } y \text{ heads}
\end{pmatrix}
\]

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

\[
= P(X = x)P(Y = y)
\]

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

\[ X \sim \text{Poi}(\lambda_1), \ Y \sim \text{Poi}(\lambda_2) \]

\[ X, Y \text{ independent} \]

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

- \( n \) 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?
Breakout Rooms

Slide 25 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://edstem.org/us/courses/5090/discussion/377748](https://edstem.org/us/courses/5090/discussion/377748)

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

$$P(X = x, Y = y) = P(X = x, Y = y \mid N = x + y)P(N = x + y)$$

$$= P(X = x \mid N = x + y)P(Y = y \mid X = x, N = x + y)P(N = x + y)$$

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

$$= \frac{(x + y)!}{x! y!} e^{-\lambda p} \frac{\lambda^x (1 - p)^y}{(x + y)!} = e^{-\lambda p} \frac{(\lambda p)^x}{x!} \cdot e^{-\lambda (1 - p)} \frac{(\lambda (1 - p))^y}{y!}$$

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

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

Yes, $X$ and $Y$ are independent!
Interlude for announcements
Announcements

Quiz #1
Grades/solutions: Next week

Problem Set 3
Due: Friday 4/30 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.”


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 $X$ 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.