12: Independent RVs

Lisa Yan
May 1, 2020
Quick slide reference

3 Independent discrete RVs 12a_independent_rvs
8 Sums of Independent Binomial RVs 12b_sum_binomial
10 Convolution: Sum of independent Poisson RVs 12c_discrete_conv
17 Exercises LIVE
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
\[ S = D_1 + D_2, \text{ the sum of two rolls} \]

- 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

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 $Y$:

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

$Z = X + Y$

$P(Z = k)$ is the convolution of $p_X$ and $p_Y$. 

Lisa Yan, CS109, 2020
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>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\checkmark$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$n - 2$</td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n - 1$</td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>$\checkmark$</td>
<td>$(1, n-1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n + 1$</td>
<td></td>
<td></td>
<td>$\checkmark$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ldots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\ldots$</td>
<td></td>
</tr>
</tbody>
</table>

- $\checkmark$: event where $X + Y = n$

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

$$= \sum_{k=0}^{n} P(X = k)P(Y = n-k)$$
Sum of 2 dice rolls

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

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):

\[
P(X + Y = n) = \sum_{k=0}^{n} P(X = k)P(Y = n - k)
\]

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

\[
= e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \frac{\lambda_1^k \lambda_2^{n-k}}{n!}
= e^{-(\lambda_1 + \lambda_2)} \frac{1}{n!} (\lambda_1 + \lambda_2)^n
\]

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

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

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

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 1\textsuperscript{st} die are dependent RVs.

\textbf{Important:} Joint PMF must decompose into product of marginal PMFs for ALL values of $X$ and $Y$ for $X, Y$ to be independent RVs.
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
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 $X = \text{number of heads in first } n \text{ flips. } X \sim \text{Bin}(n, p)$

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

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

1. Are $X$ and $Z$ independent?  
   Counterexample: What if $Z = 0$?

2. Are $X$ and $Y$ independent? ✓

\[
P(X = x, Y = y) = P\left(\begin{array}{c}
\text{first } n \text{ flips have } x \text{ heads}
\vspace{1em}
\text{and next } m \text{ flips have } y \text{ heads}
\end{array}\right)
\]

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

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

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

  $X \sim A \sim \text{Poi}(\lambda_A = 30(0.01) = 0.3)$
  $Y \sim B \sim \text{Poi}(\lambda_B = 50(0.02) = 1)$

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

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

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

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

\[
P(X = n, Y = m) = \frac{(n + m)!}{n! m!} \cdot e^{-\lambda} \cdot \frac{(\lambda p)^n (\lambda (1 - p))^m}{(n + m)!} \cdot 1 \\
= P(X = n)P(Y = m)
\]

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

Law of Total Probability

Chain Rule

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

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

Quiz #1
Grades/solutions: Next week

Problem Set 3
Due: Monday 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."


CS109 Current Events Spreadsheet
Independence of multiple random variables

Recall independence of $n$ events $E_1, E_2, ..., E_n$:

for $r = 1, ..., n$:

for every subset $E_1, E_2, ..., E_r$:

$$P(E_1, E_2, ..., E_r) = P(E_1)P(E_2) \cdots P(E_r)$$

We have independence of $n$ discrete random variables $X_1, X_2, ..., X_n$ if for $r = 1, ..., n$:

for all subsets $x_1, x_2, ..., x_r$:

$$P(X = x_1, X = x_2, ..., X_r = x_r) = \prod_{i=1}^{r} P(X_i = x_i)$$
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 $X$ 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:
   \[ E[X + Y] = E[X] + E[Y] \]

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

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

\[ = \sum_x \sum_y xp_{X,Y}(x, y) + \sum_x \sum_y yp_{X,Y}(x, y) \]

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

\[ = \sum_x xp_X(x) + \sum_y yp_Y(y) \]

\[ = E[X] + E[Y] \]

LOTUS,
\[ g(X, Y) = X + Y \]

Linearity of summations
(\text{cont. case: linearity of integrals})

Marginal PMFs for \(X\) and \(Y\)
Expectations of common RVs: Binomial

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

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

\[
E[X] = E\left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np
\]
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
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?
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 \( Y_i = \# \text{ trials to get } i\text{th success (after } (i - 1)\text{th success)} \)

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

\[ E[Y] = E \left[ \sum_{i=1}^{r} Y_i \right] = \sum_{i=1}^{r} E[Y_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p} \]
\[ P(X = k, Y = n) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \cdot \frac{\lambda^n \cdot e^{-\lambda}}{n!} = \frac{\lambda^{k+n} \cdot e^{-2\lambda}}{(k+n)!} \]
\[
P(\text{2 girls} \mid \text{1st girl}) = \frac{6}{6} \\
\geq \frac{P(\text{2 girls} \mid \text{1st girl})}{P(\text{1 girl})} = \frac{1}{4} \\
P(\text{both boys} \mid \geq 1 \text{ boy}) = \frac{P(\text{both boys} \mid \geq 1 \text{ boy})}{P(\geq 1 \text{ boy})} = \frac{1}{4} \\
\]