SECTION 5
ZOO OF RANDOM VARIABLES
— Welcome back, everyone! —
AGENDA

01 ANNOUNCEMENTS
02 VARIANCE
03 INDEPENDENT RV
04 ZOO OF RANDOM VARIABLES
ANNOUNCEMENTS
Schedule Reminders

Pset3 Grades Were Released
(regrade requests open and close after a week)

Pset4 Due Yesterday

Pset5 Was Released
Coding: task 7 due Feb 9th
Other tasks due Feb 7th
LOE REMINDER
When working with linearity of expectation, remember to

\textbf{first} define the RVs and the summation relationships
\emph{don’t worry how the individual RVs are distributed}

\textbf{then} apply linearity of expectation and find each value
Variance is another property of RVs (like expectation) that measures how much the values in the RV "vary"
VARIANCE - how “different” are values from the expectation “on average”

\[ \text{Var}(X) = E[(X - E(X))^2] = \sum_x (P(X=x) \cdot (x - E(X))^2) \]

- expected value of the squared distance between each RV outcome and the expected value of RV
- add up all the squared distances weighted by their probabilities

Properties

\[ \text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X) \]
\[ \text{Var}(X) = E[X^2] - (E[X])^2 \]
What does independence mean for random variables?
Random variables $X$ and $Y$ are **independent** if –

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

*Knowing the value of $X$ doesn’t help “guess” what $Y$ is*

**it’s a useful property! if $X$ and $Y$ are independent random variables then —**

$$E(X \cdot Y) = E[X] \cdot E[Y]$$

$$\text{Var}(X + Y) = \text{Var}[X] + \text{Var}[Y]$$
Random variables $X$ and $Y$ are **independent** if –

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

Knowing the value of $X$ doesn’t help “guess” what $Y$ is

Additionally, there’s **independent and identically distributed (aka, “i.i.d.”)** random variables

Identically distributed means the random variables **have the same pmf** –

$$P(X=k) = P(Y=k) \quad \text{for any value } k$$

For example, rolling a die twice, where $X$ is the first roll number and $Y$ is the second roll number
ZOO OF RV’S

zoo of discrete random variables!
ZOO OF DISCRETE RANDOM VARIABLES

Random variables allow us to represent different random experiments/situations

We’ve seen how tedious computing pmfs, expectations, and variances can be.

There are some common situations that call for a random variable that is too complex to analyze, so we derive pmfs, expectations, and variance for this “zoo” of RVs.
**UNIFORM**

**MODELS SITUATIONS WHERE EACH OUTCOME IS EQUALLY LIKELY**

\[ X \sim \text{Uniform}(a, b) \text{ if } X \text{ is equally likely to take on any value between } a \text{ and } b \]

\[
\begin{align*}
  p_X(k) &= \frac{1}{b - a + 1} \\
  \mathbb{E}[X] &= \frac{a + b}{2} \\
  \text{Var}(X) &= \frac{(b - a)(b - a + 2)}{12}
\end{align*}
\]

A random variable X representing the outcome of rolling a fair 6 sided dice

\[ X \sim \text{Uniform}(1, 6) \]

choosing a random value between 1 and 6 with each outcome equally likely
Bernoulli (indicator) models situations where the RV can take on 0 or 1 (whether success or not)

\[ X \sim \text{Bernoulli}(p) \text{ if } X \text{ is 1 with probability of } p \]

\[
p_X(k) = \begin{cases} 
p, & k = 1 \\ 
1 - p, & k = 0 
\end{cases}
\]

\[ \mathbb{E}[X] = p \quad \text{Var}(X) = p(1 - p) \]

X represents whether outcome of rolling a fair 6 sided dice is even (1) or not (0)

**X - Bernoulli(3/6)**

probability of 3/6 for “success”
Binomial models situations when we count the # times an event occurs in n tries.

\[ X \sim \text{Binomial}(n, p) \] means \( X \) represents the number of times an event with probability \( p \) happens after \( n \) trials.

\[
p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}
\]

\[ \mathbb{E}[X] = np \quad Var(X) = np(1-p) \]

\( X \) represents the number of times the dice rolled to a 6 during 9 dice rolls

\( X \sim \text{Binomial}\left(\frac{1}{6}, 9\right) \)

probability of success (rolling a 6) on a single dice roll is \( \frac{1}{6} \), and 9 trials (rolls)
Geometric

Models situations when we count the # trials until some event occurs

$X \sim \text{Geometric}(p)$ means $X$ represents the number of trials before success (an event with probability $p$ happens)

$$p_X(k) = (1 - p)^{k-1} p,$$
$$\mathbb{E}[X] = \frac{1}{p},$$
$$\text{Var}(X) = \frac{1-p}{p^2}$$

$X$ represents the number of times we roll a 6 sided die, before it rolls a 6

$X \sim \text{Geometric}(\frac{1}{6})$

on a single dice roll, there’s a probability of $\frac{1}{6}$ for success (that it rolls a 6)
**Negative Binomial**

(related to geometric)

Models situations where we count number of trials to get some number of successes.

\[ X \sim \text{NegBin}(r,p) \] means \( X \) represents the number of trials to get \( r \) successes (probability of success on a single trial is \( p \)).

\[
p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}
\]

\[
\mathbb{E}[X] = \frac{r}{p} \\
\text{Var}(X) = \frac{r(1-p)}{p^2}
\]

\( X \) represents number of dice rolls before we get 4 rolls with a 6

\( X \sim \text{NegBin}(4, 1/6) \)

because we want to have 4 trials (rolls) with success, and a success (rolling 6) has probability 1/6.
Poisson models situations with time - how many successes in a unit of time

\[ X \sim \text{Poisson}(\lambda) \text{ means } X \text{ represents the number of success in a unit of time, where } \lambda \text{ is average rate of successes per unit of time} \]

\[ p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \]
\[ E[X] = \lambda \]
\[ Var(X) = \lambda \]

\[ X \text{ represents number of people born during a particular minute} \]

\[ X \sim \text{Poisson}(\lambda) \]

where \( \lambda \) represents the average birth rate per minute
HYPERGEOMETRIC

MODELS SITUATIONS WITH CHOOSING - HOW MANY "SUCCESSES" DO YOU GET WHEN CHOOSING WITHOUT REPLACEMENT

\[ X \sim \text{HypGeo}(N, K, n) \]

X represents the number of successes out of n draws from N items with K successes.

\[ p_X(k) = \binom{K}{k} \binom{N-K}{n-k} \binom{N}{n} \]

\[ E[X] = n \frac{K}{N} \]

\[ Var(X) = n \cdot \frac{K(N-K)(N-n)}{N^2(2N-1)} \]

X represents the number of Kit-Kats we will get when drawing 30 candies from a bowl of 100 candies that contain 10 Kit-Kats:

\[ X \sim \text{HypGeo}(100, 10, 30) \]

because we draw 30 from 100 items with 10 successes (Kit-Kats).
LET'S TRY IT!

Let’s identify these distributions in some real examples!
THANKS!