

Econ 103 – Statistics for Economists

Chapter 4 and 5: Discrete Random Variables

Mallick Hossain

University of Pennsylvania

Motivation

Why Should I Care?

- These next few lectures will be a crash course in random variables and probability
- This will help build your toolbox of how to think about real-world problems using statistics.
- Economists are always concerned about the “distribution” of things. This will help you understand why this is such a critical concept.

Random Variables

Definitions

A random variable is neither random nor a variable.

- **Random Variable (RV):** A random variable X is a *fixed* function that assigns a *number* to each basic outcome of a random experiment.
- **Realization:** A realization x of a RV X is a particular numeric value that the RV could take. We write $\{X = x\}$ to denote the event that X took on the value x .
- **Support Set:** The support is the set of all possible realizations of a RV.

Note: RVs are CAPITAL LETTERS while their realizations are lowercase letters.

Definitions

- **Discrete RV:** A random variable X is *discrete* if its support set is discrete (e.g. $\{0, 1, 2\}$).
- **Continuous RV:** A random variable X is *continuous* if its support set is continuous (e.g. $[0, 1], \mathbb{R}$).

What are examples of each kind of random variable?

Example: Coin Flip Random Variable

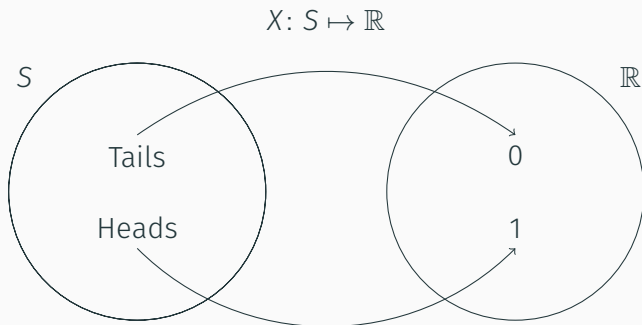


Figure 1: This random variable assigns numeric values to the random experiment of flipping a fair coin once: Heads is assigned 1 and Tails 0.

Which of these is a realization of the Coin Flip RV?

- (a) Tails
- (b) 2
- (c) 0
- (d) Heads
- (e) $1/2$

What is the support set of the Coin Flip RV?

- (a) {Heads, Tails}
- (b) $1/2$
- (c) 0
- (d) $\{0, 1\}$
- (e) 1

Let X denote the Coin Flip RV

What is $P(X = 1)$?

- (a) 0
- (b) 1
- (c) $1/2$
- (d) π
- (e) Not enough information to determine

CDFs and PMFs

Probability Mass Function (PMF)

A function that gives $P(X = x)$ for any x in the support set of a discrete RV X . We use the following notation for the PMF:

$$p(x) = P(X = x)$$

Plug in a realization x , get out a probability $p(x)$.

Probability Mass Function for Coin Flip RV

$$X = \begin{cases} 0, \text{Tails} \\ 1, \text{Heads} \end{cases}$$

$$p(0) = 1/2$$

$$p(1) = 1/2$$

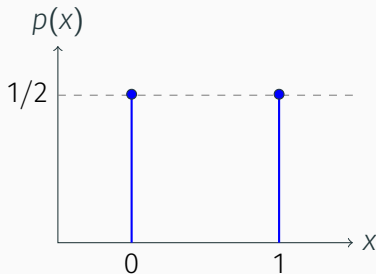


Figure 2: Plot of Coin Flip RV PMF

Important Note about Support Sets

Whenever you write down the PMF of a RV, it is **crucial** to also write down its support (*all possible realizations for a RV*). Outside of the support set, all probabilities are zero. In other words, the PMF is **only defined** on its support.

Properties of Probability Mass Functions

If $p(x)$ is the PMF of a random variable X , then

(i) $0 \leq p(x) \leq 1$ for all x

(ii) $\sum_{\text{all } x} p(x) = 1$

where “all x ” is shorthand for “all x in the support of X .”

Cumulative Distribution Function (CDF)

The CDF gives the probability that a RV X is less than or equal to some threshold x_0 , as a function of x_0

$$F(x_0) = P(X \leq x_0)$$

Important!

The threshold x_0 is allowed to be *any real number*. In particular, it doesn't have to be in the support of X !

Discrete RVs: Sum the PMF to get the CDF

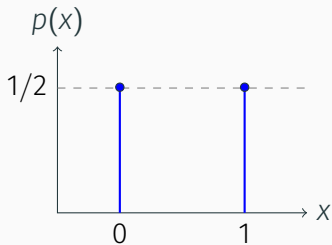
$$F(x_0) = \sum_{x \leq x_0} p(x)$$

Why?

The events $\{X = x\}$ are mutually exclusive, so we sum to get the probability of their union for all $x \leq x_0$:

$$F(x_0) = P(X \leq x_0) = P\left(\bigcup_{x \leq x_0} \{X = x\}\right) = \sum_{x \leq x_0} P(X = x) = \sum_{x \leq x_0} p(x)$$

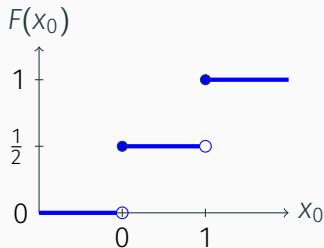
Probability Mass Function



$$p(0) = 1/2$$

$$p(1) = 1/2$$

Cumulative Dist. Function



$$F(x_0) = \begin{cases} 0, & x_0 < 0 \\ 1/2, & 0 \leq x_0 < 1 \\ 1, & x_0 \geq 1 \end{cases}$$

Properties of CDFs

1. $\lim_{x_0 \rightarrow \infty} F(x_0) = 1$
2. $\lim_{x_0 \rightarrow -\infty} F(x_0) = 0$
3. Non-decreasing: $x_0 < x_1 \Rightarrow F(x_0) \leq F(x_1)$
4. Right-continuous (“open” versus “closed” on prev. slide)

Since $F(x_0) = P(X \leq x_0)$, we have $0 \leq F(x_0) \leq 1$ for all x_0

Lotteries

Choose between the following two lotteries:

- **Lottery A:**

- You get \$1 million for sure

- **Lottery B:**

- 10% chance of \$5 million
- 89% chance of \$1 million
- 1% chance of nothing

Choose between the following two lotteries:

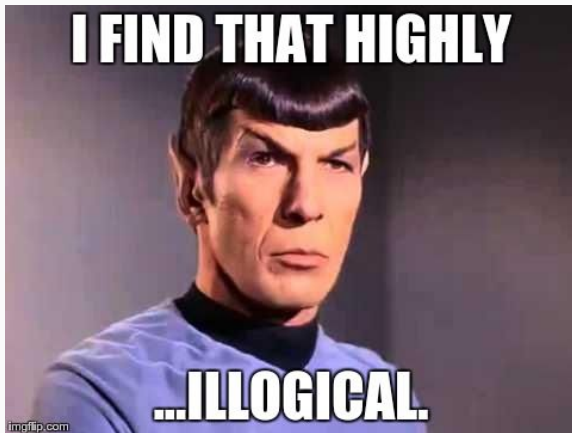
- **Lottery C:**

- 11% chance of \$1 million
- 89% chance of nothing

- **Lottery D:**

- 10% chance of \$5 million
- 90% chance of nothing

Highly Illogical



If you chose $\{A, D\}$ or $\{B, C\}$, you are highly illogical.

Expectation

Expected Value (aka Expectation)

The expected value of a discrete RV X is given by

$$E[X] = \sum_{\text{all } x} x \cdot p(x)$$

In other words, the expected value of a discrete RV is the *probability-weighted average of its realizations*.

Notation

We sometimes write μ as shorthand for $E[X]$.

How Can I Think About Expected Value?

- The long-run average of a random variable
- How much you expect to win from a gamble where the payoffs are the values of the RV

If the realizations of the coin-flip RV were **payoffs**, how much would you expect to win per play *on average* in a long sequence of plays?

$$X = \begin{cases} \$0, \text{Tails} \\ \$1, \text{Heads} \end{cases}$$

Your Turn to Calculate an Expected Value

Let X be a random variable with support set $\{1, 2, 3\}$ where $p(1) = p(2) = 1/3$. Calculate $E[X]$.

Your Turn to Calculate an Expected Value

Let X be a random variable with support set $\{1, 2, 3\}$ where $p(1) = p(2) = 1/3$. Calculate $E[X]$.

$$E[X] = \sum_{\text{all } x} x \cdot p(x) = 1 \times 1/3 + 2 \times 1/3 + 3 \times 1/3 = 2$$

Expectation of a Function of a Discrete RV

Let X be a random variable and g be a function. Then:

$$E[g(X)] = \sum_{\text{all } x} g(x)p(x)$$

Is $E[g(X)] = g(E[X])$? What if $g(x) = x^2$?

$$E[g(X)] \neq g(E[X])$$

(Expected Value of Function \neq Function of Expected Value)

Linearity of Expectation

$$E[a + bX] = a + bE[X]$$

Linearity of Expectation

$$E[a + bX] = a + bE[X]$$

Proof:

$$\begin{aligned} E[a + bX] &= \sum_{\text{all } x} (a + bx)p(x) \\ &= \sum_{\text{all } x} p(x) \cdot a + \sum_{\text{all } x} p(x) \cdot bx \\ &= a \sum_{\text{all } x} p(x) + b \sum_{\text{all } x} x \cdot p(x) \\ &= a + bE[X] \end{aligned}$$

One of the few instances where $E[g(X)] = g(E[X])$

Why Was I Highly Illogical?

This is known as the Allais paradox.

Let $U(x)$ be the utility you get from money (notice, we are not assuming that every dollar is worth the same to you).

If you chose A, then we know that

$$U(1) > 0.1U(5) + 0.89U(1) + 0.01U(0)$$

Why Was I Highly Illogical?

This is known as the Allais paradox.

Let $U(x)$ be the utility you get from money (notice, we are not assuming that every dollar is worth the same to you).

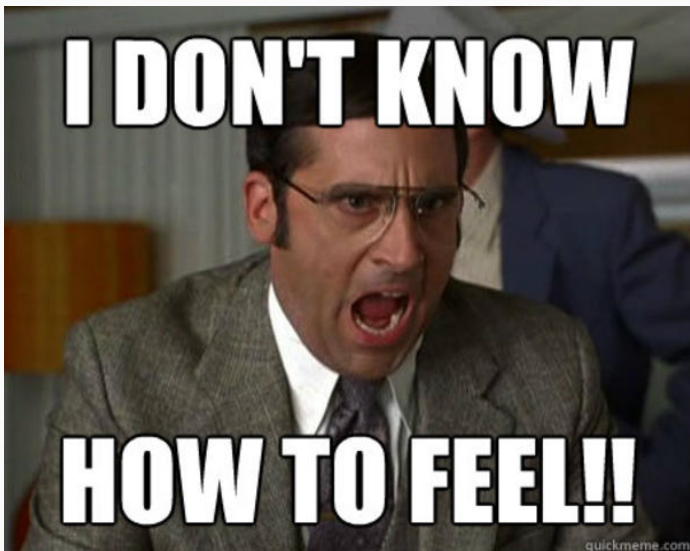
If you chose A, then we know that

$$U(1) > 0.1U(5) + 0.89U(1) + 0.01U(0)$$

If you also chose D, then we know that

$$0.1U(5) + 0.9U(0) > 0.11U(1) + 0.89U(0)$$

$$0.1U(5) + 0.89U(1) + 0.01U(0) > U(1)$$



Variance

Variance and Standard Deviation of a RV

Variance (Var)

$$\sigma^2 = \text{Var}(X) = E[(X - \mu)^2] = E[(X - E[X])^2]$$

Standard Deviation (SD)

$$\sigma = \sqrt{\sigma^2} = \text{SD}(X)$$

These look similar to their definitions from last chapter

Variance and Standard Deviation of a RV

Variance (Var)

$$\sigma^2 = \text{Var}(X) = E[(X - \mu)^2] = E[(X - E[X])^2]$$

Standard Deviation (SD)

$$\sigma = \sqrt{\sigma^2} = \text{SD}(X)$$

These look similar to their definitions from last chapter

Variance and std. dev. are *expectations of functions of a RV*

How To Calculate Variance for Discrete RV?

$$\begin{aligned}\text{Var}(X) &= E[(X - \mu)^2] = \sum_{\text{all } x} (x - \mu)^2 p(x) \\&= \sum_{\text{all } x} (x^2 - 2x\mu + \mu^2) p(x) \\&= \sum_{\text{all } x} x^2 p(x) - 2\mu \sum_{\text{all } x} x p(x) + \mu^2 \sum_{\text{all } x} p(x) \\&= E[X^2] - 2\mu E[X] + \mu^2 \\&= E[X^2] - (E[X])^2\end{aligned}$$

Shortcut Formula For Variance

This is *not* the definition, it's a shortcut for doing calculations:

$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2$$

Variance of a Linear Function

Suppose X is a random variable with $\text{Var}(X) = \sigma^2$ and a, b are constants. What is $\text{Var}(a + bX)$?

- (a) σ^2
- (b) $a + \sigma^2$
- (c) $b\sigma^2$
- (d) $a + b\sigma^2$
- (e) $b^2\sigma^2$

Variance and SD are *NOT* Linear

$$\text{Var}(a + bX) = b^2\sigma^2$$

$$\text{SD}(a + bX) = |b|\sigma$$

These should look familiar from the related results for sample variance and std. dev. that you worked out on an earlier problem set.

Variance of a Linear Transformation

$$\begin{aligned}\text{Var}(a + bX) &= E \left[\{(a + bX) - E(a + bX)\}^2 \right] \\&= E \left[\{(a + bX) - (a + bE[X])\}^2 \right] \\&= E \left[(bX - bE[X])^2 \right] \\&= E[b^2(X - E[X])^2] \\&= b^2 E[(X - E[X])^2] \\&= b^2 \text{Var}(X) = b^2 \sigma^2\end{aligned}$$

The key point here is that variance is defined in terms of expectation and expectation is linear.

Constants Versus Random Variables

This is a crucial distinction that students sometimes miss:

- **Random Variables**

- Suppose X is a RV – the values it takes on are random
- A function $g(X)$ of a RV is itself a RV

- **Constants**

- $E[X]$ and $Var(X)$ are constants
- Realizations x are constants, but *which* realization the RV takes on is random
- Parameters are constants
- Sample size n is a constant

Bernoulli Random Variable

Bernoulli Random Variable – Generalization of Coin Flip

Support Set

$\{0, 1\}$ – 1 traditionally called “success,” 0 “failure”

Probability Mass Function

$$p(0) = 1 - p$$

$$p(1) = p$$

Cumulative Distribution Function

$$F(x_0) = \begin{cases} 0, & x_0 < 0 \\ 1 - p, & 0 \leq x_0 < 1 \\ 1, & x_0 \geq 1 \end{cases}$$

Expected Value of Bernoulli RV

$$X = \begin{cases} 0, \text{Failure: } 1 - p \\ 1, \text{Success: } p \end{cases}$$

$$\sum_{\text{all } x} x \cdot p(x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

Variance of Bernoulli RV – via the Shortcut Formula

Step 1 – $E[X]$

$$\mu = E[X] = \sum_{x \in \{0,1\}} p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p$$

Step 2 – $E[X^2]$

$$E[X^2] = \sum_{x \in \{0,1\}} x^2 p(x) = 0^2(1-p) + 1^2 p = p$$

Step 3 – Combine with Shortcut Formula

$$\sigma^2 = \text{Var}[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$$

Variance of Bernoulli RV – Without Shortcut

You will fill in the missing steps on Problem Set 5.

$$\begin{aligned}\sigma^2 &= \text{Var}(X) = \sum_{x \in \{0,1\}} (x - \mu)^2 p(x) \\ &= \sum_{x \in \{0,1\}} (x - p)^2 p(x) \\ &\vdots \\ &= p(1 - p)\end{aligned}$$

Notation: $X \sim \text{Bernoulli}(p)$

Means X is a Bernoulli RV with $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The tilde is read “distributes as.”

Parameter

Any constant that appears in the definition of a RV, here p .

The St. Petersburg Game

How Much Would You Pay?

How much would you be willing to pay for the right to play the following game?

Imagine a fair coin. The coin is tossed once. If it falls heads, you receive a prize of \$2 and the game stops. If not, it is tossed again. If it falls heads on the second toss, you get \$4 and the game stops. If not, it is tossed again. If it falls heads on the third toss, you get \$8 and the game stops, and so on. The game stops after the first head is thrown. If the first head is thrown on the x^{th} toss, the prize is $\$2^x$

$X = \text{Trial Number of First Head}$

x	2^x	$p(x)$	$2^x \cdot p(x)$
-----	-------	--------	------------------

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

$X = \text{Trial Number of First Head}$

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

$X = \text{Trial Number of First Head}$

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
\vdots	\vdots	\vdots	\vdots
n	2^n	$1/2^n$	1
\vdots	\vdots	\vdots	\vdots

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

$X = \text{Trial Number of First Head}$

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
\vdots	\vdots	\vdots	\vdots
n	2^n	$1/2^n$	1
\vdots	\vdots	\vdots	\vdots

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) = 1 + 1 + 1 + \dots$$

$X = \text{Trial Number of First Head}$

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
\vdots	\vdots	\vdots	\vdots
n	2^n	$1/2^n$	1
\vdots	\vdots	\vdots	\vdots

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) = 1 + 1 + 1 + \dots = \infty$$

Functions of Random Variables

Example: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Support of Y

Example: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Support of Y

$$\{e^0, e^1\} = \{1, e\}$$

Probability Mass Function for Y

Example: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Support of Y

$$\{e^0, e^1\} = \{1, e\}$$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation of $Y = e^X$

Expectation: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation of $Y = e^X$

$$\sum_{y \in \{1, e\}} y \cdot p_Y(y) =$$

Expectation: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation of $Y = e^X$

$$\sum_{y \in \{1, e\}} y \cdot p_Y(y) = (1 - p) \cdot 1 + p \cdot e = 1 + p(e - 1)$$

Expectation: Function of Bernoulli RV

Let $Y = e^X$ where $X \sim \text{Bernoulli}(p)$

Expectation of the Function

$$\sum_{y \in \{1, e\}} y \cdot p_Y(y) = (1 - p) \cdot 1 + p \cdot e = 1 + p(e - 1)$$

Function of the Expectation

$$e^{E[X]} = e^p$$

Binomial Random Variable

Binomial Random Variable

Let X = the sum of n independent Bernoulli trials, each with probability of success p . Then we say that: $X \sim \text{Binomial}(n, p)$

Parameters

p = probability of “success,” n = # of trials

Support

$\{0, 1, 2, \dots, n\}$

Probability Mass Function (PMF)

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

Where does the Binomial PMF come from?

Question

Let's flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Where does the Binomial PMF come from?

Question

Let's flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Answer

Three basic outcomes make up this event: $\{HHT, HTH, THH\}$, each has probability $1/8 = 1/2 \times 1/2 \times 1/2$. Basic outcomes are mutually exclusive, so sum to get $3/8 = 0.375$

Where does the Binomial PMF come from?

Question

Let's flip an *unfair* coin 3 times, where the probability of heads is $1/3$. What is the probability that we get exactly 2 heads?

Answer

All basic outcomes are not equally likely, but those with exactly two heads *still are*

$$P(HHT) = (1/3)^2(1 - 1/3) = 2/27$$

$$P(THH) = 2/27$$

$$P(HTH) = 2/27$$

Summing gives $2/9 \approx 0.22$

Where does the Binomial PMF come from?

Suppose we flip an unfair coin 4 times, where the probability of heads is $1/3$. What is the probability that we get exactly 2 heads?

HHTT	TTHH
HTHT	THTH
HTTH	THHT

Six equally likely, mutually exclusive basic outcomes make up this event:

$$\binom{4}{2} (1/3)^2 (2/3)^2$$

Joint Distributions

Definition

Sometimes we are interested in how two random variables behave together (aka jointly). Because of this, we need a way to characterize their joint behavior.

Joint probability mass function ($p_{XY}(x, y)$): Let X and Y be discrete random variables. The joint probability mass function $p_{XY}(x, y)$ gives the probability of each pair of realizations (x, y) in the support:

$$p_{XY}(x, y) = P(X = x \cap Y = y)$$

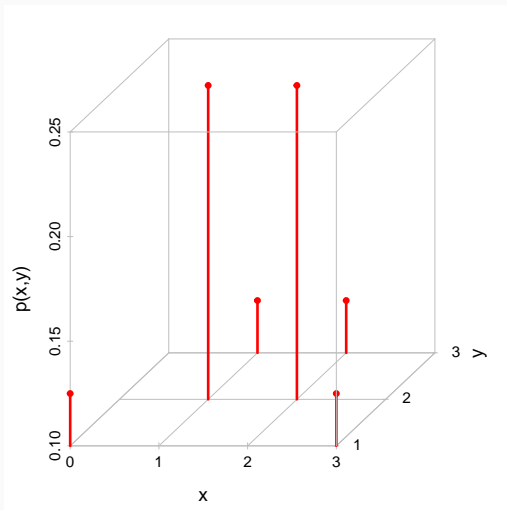
Example: Joint PMF in Tabular Form

For discrete RVs, we often use tables

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

One example might be that X is one's home region (0 is Northeast, 1 is Northwest, 2 is Southeast, 3 is Southwest) and Y is the car they own (1 is sedan, 2 is SUV, 3 is truck).

Plot of Joint PMF



What is $p_{XY}(1, 2)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

What is $p_{XY}(1, 2)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

$$p_{XY}(1, 2) = P(X = 1 \cap Y = 2) = 1/4$$

What is $p_{XY}(2, 1)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

What is $p_{XY}(2, 1)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

$$p_{XY}(2, 1) = P(X = 2 \cap Y = 1) = 0$$

Properties of Joint PMF

1. $0 \leq p_{XY}(x, y) \leq 1$ for any pair (x, y)
2. The sum of $p_{XY}(x, y)$ over all pairs (x, y) in the support is 1:

$$\sum_x \sum_y p(x, y) = 1$$

Does this satisfy the properties of a joint PMF?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

Does this satisfy the properties of a joint PMF?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

1. $p(x,y) \geq 0$ for all pairs (x,y)
2. $\sum_x \sum_y p(x,y) = 1/8 + 1/4 + 1/8 + 1/4 + 1/8 + 1/8 = 1$

Joint versus Marginal PMFs

What if we just care about one RV? Then, use the **marginal PMF**

Marginal PMFs

$$p_X(x) = P(X = x)$$

$$p_Y(y) = P(Y = y)$$

You can't calculate a joint pmf from marginals alone but you *can* calculate marginals from the joint!

Marginals from Joint

$$p_X(x) = \sum_{\text{all } y} p_{XY}(x, y)$$

$$p_Y(y) = \sum_{\text{all } x} p_{XY}(x, y)$$

Why?

$$\begin{aligned} p_Y(y) &= P(Y = y) = P\left(\bigcup_{\text{all } x} \{X = x \cap Y = y\}\right) \\ &= \sum_{\text{all } x} P(X = x \cap Y = y) = \sum_{\text{all } x} p_{XY}(x, y) \end{aligned}$$

To get the marginals sum “into the margins” of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	

To get the marginals sum “into the margins” of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	

$$p_X(0) = 1/8 + 0 + 0 = 1/8$$

To get the marginals sum “into the margins” of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	
	3	1/8	0	0	

$$p_X(0) = 1/8 + 0 + 0 = 1/8$$

$$p_X(1) = 0 + 1/4 + 1/8 = 3/8$$

To get the marginals sum “into the margins” of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	

$$p_X(0) = 1/8 + 0 + 0 = 1/8$$

$$p_X(1) = 0 + 1/4 + 1/8 = 3/8$$

$$p_X(2) = 0 + 1/4 + 1/8 = 3/8$$

To get the marginals sum “into the margins” of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
					1

$$p_X(0) = 1/8 + 0 + 0 = 1/8$$

$$p_X(1) = 0 + 1/4 + 1/8 = 3/8$$

$$p_X(2) = 0 + 1/4 + 1/8 = 3/8$$

$$p_X(3) = 1/8 + 0 + 0 = 1/8$$

Marginal Probabilities

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	

Marginal Probabilities

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4			

$$p_Y(1) = 1/8 + 0 + 0 + 1/8 = 1/4$$

Marginal Probabilities

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2		

$$p_Y(1) = 1/8 + 0 + 0 + 1/8 = 1/4$$

$$p_Y(2) = 0 + 1/4 + 1/4 + 0 = 1/2$$

Marginal Probabilities

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	1

$$p_Y(1) = 1/8 + 0 + 0 + 1/8 = 1/4$$

$$p_Y(2) = 0 + 1/4 + 1/4 + 0 = 1/2$$

$$p_Y(3) = 0 + 1/8 + 1/8 + 0 = 1/4$$

Conditional Distributions

Definition of Conditional PMF

How does the distribution of y change with x ? In other words, if we know x , does that tell us anything about y ?

$$p_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)} = \frac{p_{XY}(x, y)}{p_X(x)}$$

Conditional PMF of Y given $X = 2$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8

$$p_{Y|X}(1|2) = \frac{p_{XY}(2,1)}{p_X(2)} = \frac{0}{3/8} = 0$$

$$p_{Y|X}(2|2) = \frac{p_{XY}(2,2)}{p_X(2)} = \frac{1/4}{3/8} = 2/3$$

$$p_{Y|X}(3|2) = \frac{p_{XY}(2,3)}{p_X(2)} = \frac{1/8}{3/8} = 1/3$$

What is $p_{X|Y}(1|2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	

What is $p_{X|Y}(1|2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	

$$p_{X|Y}(1|2) = \frac{p_{XY}(1,2)}{p_Y(2)} = \frac{1/4}{1/2} = 1/2 \quad (1)$$

What is $p_{X|Y}(1|2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	

$$p_{X|Y}(1|2) = \frac{p_{XY}(1,2)}{p_Y(2)} = \frac{1/4}{1/2} = 1/2 \quad (1)$$

Similarly:

$$p_{X|Y}(0|2) = 0, \quad p_{X|Y}(2|2) = 1/2, \quad p_{X|Y}(3|2) = 0$$

Independent RVs: Joint Equals Product of Marginals

Definition

Two discrete RVs are **independent** if and only if

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

for all pairs (x, y) in the support.

Equivalent Definition

$$p_{Y|X}(y|x) = p_Y(y) \text{ and } p_{X|Y}(x|y) = p_X(x)$$

for all pairs (x, y) in the support.

Are X and Y Independent?

		Y			
		1	2	3	
X	0	$1/8$	0	0	$1/8$
	1	0	$1/4$	$1/8$	$3/8$
	2	0	$1/4$	$1/8$	$3/8$
	3	$1/8$	0	0	$1/8$
		$1/4$	$1/2$	$1/4$	

Are X and Y Independent?

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$p_{XY}(2,1) = 0$$

$$p_X(2) \times p_Y(1) = (3/8) \times (1/4) \neq 0$$

Therefore X and Y are *not* independent.

Conditional Expectation

Conditional Expectation

Intuition

$E[Y|X]$ = “best guess” of realization that Y after observing realization of X .

$E[Y|X]$ is a Random Variable

While $E[Y]$ is a constant, $E[Y|X]$ is a function of X , hence a **Random Variable**.

$E[Y|X = x]$ is a Constant

The constant $E[Y|X = x]$ is the “guess” of Y if we see $X = x$.

Calculating $E[Y|X = x]$

Take the mean of the conditional pmf of Y given $X = x$.

Conditional Expectation: $E[Y|X = 2]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

We showed above that the conditional pmf of $Y|X = 2$ is:

$$p_{Y|X}(1|2) = 0 \quad p_{Y|X}(2|2) = 2/3 \quad p_{Y|X}(3|2) = 1/3$$

Hence

$$E[Y|X = 2] = 2 \times 2/3 + 3 \times 1/3 = 7/3$$

Conditional Expectation: $E[Y|X = 0]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

The conditional pmf of $Y|X = 0$ is

$$p_{Y|X}(1|0) = 1 \quad p_{Y|X}(2|0) = 0 \quad p_{Y|X}(3|0) = 0$$

Hence $E[Y|X = 0] = 1$

Calculate $E[Y|X = 3]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

The conditional pmf of $Y|X = 3$ is

$$p_{Y|X}(1|3) = 1 \quad p_{Y|X}(2|3) = 0 \quad p_{Y|X}(3|3) = 0$$

Hence $E[Y|X = 3] = 1$

Calculate $E[Y|X = 1]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

Calculate $E[Y|X = 1]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

The conditional pmf of $Y|X = 1$ is

$$p_{Y|X}(1|1) = 0 \quad p_{Y|X}(2|1) = 2/3 \quad p_{Y|X}(3|1) = 1/3$$

Hence

$$E[Y|X = 1] = 2 \times 2/3 + 3 \times 1/3 = 7/3$$

$E[Y|X]$ is a Random Variable

For this example:

$$E[Y|X] = \begin{cases} 1 & X = 0 \\ 7/3 & X = 1 \\ 7/3 & X = 2 \\ 1 & X = 3 \end{cases}$$

From above the marginal distribution of X is:

$$P(X = 0) = 1/8 \quad P(X = 1) = 3/8$$

$$P(X = 2) = 3/8 \quad P(X = 3) = 1/8$$

$E[Y|X]$ takes the value 1 with prob. 1/4 and 7/3 with prob. 3/4.

Law of Iterated Expectations

The Law of Iterated Expectations

Remember that $E[Y|X]$ is a random variable, so it makes sense to ask what its expectation is. It turns out that there's a nice solution:

$$E[E[Y|X]] = E[Y]$$

Proof of The Law of Iterated Expectations (Helpful for 104)

$$\begin{aligned}E_X [E_{Y|X} [Y|X]] &= E_X \left[\sum_y y p_{Y|X}(y|X) \right] = \sum_x \left(\sum_y y p_{Y|X}(y|x) \right) p_X(x) \\&= \sum_x \sum_y y p_X(x) p_{Y|X}(y|x) = \sum_x \sum_y y p_{XY}(x, y) \\&= \sum_y y \sum_x p_{XY}(x, y) = \sum_y y p_Y(y) = E[Y]\end{aligned}$$

Law of Iterated Expectations for Our Example

Marginal pmf of Y

$$P(Y = 1) = 1/4$$

$$P(Y = 2) = 1/2$$

$$P(Y = 3) = 1/4$$

$$E[Y|X]$$

$$E[Y|X] = \begin{cases} 1 & \text{w/ prob. } 1/4 \\ 7/3 & \text{w/ prob. } 3/4 \end{cases}$$

$$\begin{aligned} E[E[Y|X]] &= 1 \times 1/4 + 7/3 \times 3/4 \\ &= 2 \end{aligned}$$

$$\begin{aligned} E[Y] &= 1 \times 1/4 + 2 \times 1/2 + 3 \times 1/4 \\ &= 2 \end{aligned}$$

Expectation of Function of Two Discrete RVs

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

Some Extremely Important Examples

Let $\mu_X = E[X]$, $\mu_Y = E[Y]$

Covariance

$$\sigma_{XY} = \text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Correlation

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Shortcut Formula for Covariance

Much easier for calculating:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

I'll mention this again in a few slides...

Calculating $\text{Cov}(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

Calculating $\text{Cov}(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2 + 4) + 1/8 \times (3 + 6 + 3) \\ &= 3 \end{aligned}$$

Calculating $\text{Cov}(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2 + 4) + 1/8 \times (3 + 6 + 3) \\ &= 3 \end{aligned}$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

Calculating $\text{Cov}(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2 + 4) + 1/8 \times (3 + 6 + 3) \\ &= 3 \end{aligned}$$

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= 3 - 3/2 \times 2 = 0 \end{aligned}$$

Calculating $\text{Cov}(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2 + 4) + 1/8 \times (3 + 6 + 3) \\ &= 3 \end{aligned}$$

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= 3 - 3/2 \times 2 = 0 \end{aligned}$$

$$\text{Corr}(X, Y) = \text{Cov}(X, Y) / [SD(X)SD(Y)] = 0$$

Zero Covariance versus Independence

- From this example we learn that zero covariance (correlation) *does not* imply independence.
- However, it turns out that independence *does* imply zero covariance (correlation).

You will prove this on the homework

Linearity of Expectation, Again

In general, $E[g(X, Y)] \neq g(E[X], E[Y])$. The key exception is when g is a linear function:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

You will also show this on the homework

Another Application: Shortcut Formula for Covariance

Similar to Shortcut for Variance: in fact $\text{Var}(X) = \text{Cov}(X, X)$

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\ &\vdots \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

You'll fill in the details for homework...

Revisiting Binomial RVs

Expected Value of Sum = Sum of Expected Values

Repeatedly applying the linearity of expectation,

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n]$$

whether or not they are **dependent or independent**.

Independent and Identically Distributed (iid) RVs

Example

$X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$

Independent

Realization of one of the RVs gives no information about the others.

Identically Distributed

Each X_i is the same kind of RV, with the same values for any parameters. (Hence same pmf, cdf, mean, variance, etc.)

Binomial(n, p) Random Variable

Definition

Sum of n independent Bernoulli RVs, each with probability of “success,” i.e. 1, equal to p

Parameters

p = probability of “success,” n = # of trials

Support

$\{0, 1, 2, \dots, n\}$

Using Our New Notation

Let $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$, $Y = X_1 + X_2 + \dots + X_n$. Then $Y \sim \text{Binomial}(n, p)$.

Which of these is the PMF of a Binomial(n, p) RV?

(a) $p(x) = p^x(1 - p)^{n-x}$

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

(c) $p(x) = \binom{x}{n} p^x$

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

(e) $p(x) = p^n(1 - p)^x$

Which of these is the PMF of a Binomial(n, p) RV?

(a) $p(x) = p^x(1 - p)^{n-x}$

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

(c) $p(x) = \binom{x}{n} p^x$

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

(e) $p(x) = p^n(1 - p)^x$

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

Expected Value of Binomial RV

Use the fact that a Binomial(n, p) RV is defined as the sum of n iid Bernoulli(p) Random Variables and the Linearity of Expectation:

$$\begin{aligned} E[Y] &= E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n] \\ &= p + p + \dots + p \\ &= np \end{aligned}$$

Variance of a Sum \neq Sum of Variances!

$$\text{Var}(aX + bY) = E \left[\{(aX + bY) - E[aX + bY]\}^2 \right]$$

Variance of a Sum \neq Sum of Variances!

$$\begin{aligned}\text{Var}(aX + bY) &= E \left[\{(aX + bY) - E[aX + bY]\}^2 \right] \\ &= E \left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2 \right]\end{aligned}$$

Variance of a Sum \neq Sum of Variances!

$$\begin{aligned}\text{Var}(aX + bY) &= E \left[\{(aX + bY) - E[aX + bY]\}^2 \right] \\ &= E \left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2 \right] \\ &= E \left[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y) \right]\end{aligned}$$

Variance of a Sum \neq Sum of Variances!

$$\begin{aligned}\text{Var}(aX + bY) &= E \left[\{(aX + bY) - E[aX + bY]\}^2 \right] \\&= E \left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2 \right] \\&= E \left[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y) \right] \\&= a^2 E[(X - \mu_X)^2] + b^2 E[(Y - \mu_Y)^2] + 2ab E[(X - \mu_X)(Y - \mu_Y)]\end{aligned}$$

Variance of a Sum \neq Sum of Variances!

$$\begin{aligned}\text{Var}(aX + bY) &= E \left[\{(aX + bY) - E[aX + bY]\}^2 \right] \\&= E \left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2 \right] \\&= E \left[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y) \right] \\&= a^2 E[(X - \mu_X)^2] + b^2 E[(Y - \mu_Y)^2] + 2ab E[(X - \mu_X)(Y - \mu_Y)] \\&= a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)\end{aligned}$$

Since $\sigma_{XY} = \rho\sigma_X\sigma_Y$, this is sometimes written as:

$$\text{Var}(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y$$

$$\text{Independence} \Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

X and Y independent $\implies \text{Cov}(X, Y) = 0$. Hence, independence implies

$$\begin{aligned}\text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ &= \text{Var}(X) + \text{Var}(Y)\end{aligned}$$

Also true for three or more RVs

If X_1, X_2, \dots, X_n are independent, then

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

Crucial Distinction

Expected Value

Always true that

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n]$$

Variance

Not true in general that

$$\text{Var}[X_1 + X_2 + \dots + X_n] = \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n]$$

except in the special case where X_1, \dots, X_n are independent (or at least uncorrelated).

Variance of Binomial Random Variable

Definition from Sequence of Bernoulli Trials

If $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$ then

$$Y = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$$

Using Independence

$$\begin{aligned}\text{Var}[Y] &= \text{Var}[X_1 + X_2 + \dots + X_n] \\ &= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] \\ &= \end{aligned}$$

Variance of Binomial Random Variable

Definition from Sequence of Bernoulli Trials

If $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$ then

$$Y = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$$

Using Independence

$$\begin{aligned}\text{Var}[Y] &= \text{Var}[X_1 + X_2 + \dots + X_n] \\ &= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] \\ &= p(1 - p) + p(1 - p) + \dots + p(1 - p) \\ &= \end{aligned}$$

Variance of Binomial Random Variable

Definition from Sequence of Bernoulli Trials

If $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$ then

$$Y = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$$

Using Independence

$$\begin{aligned}\text{Var}[Y] &= \text{Var}[X_1 + X_2 + \dots + X_n] \\ &= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] \\ &= p(1-p) + p(1-p) + \dots + p(1-p) \\ &= np(1-p)\end{aligned}$$