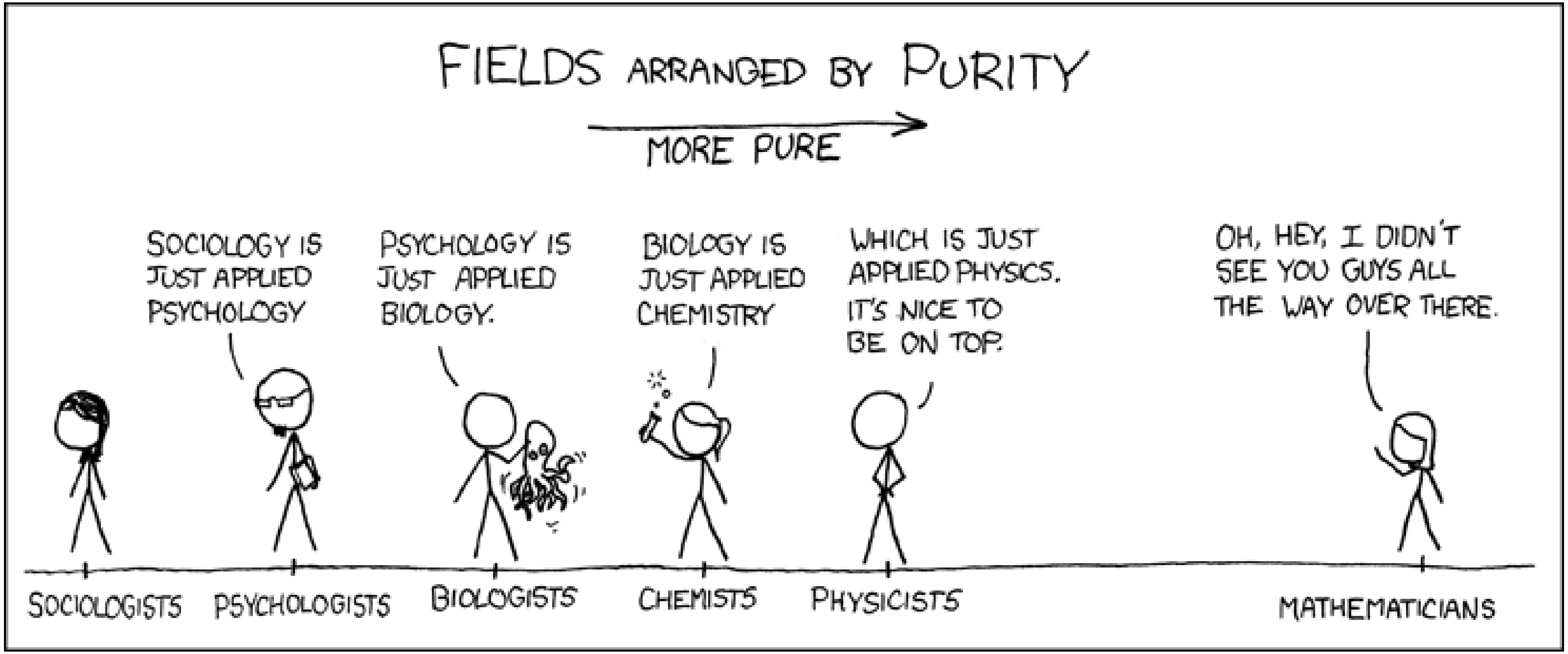
Referred to last time...



https://3d.xkcd.com/435/

Joint Distributions and Marginal Distributions

Lecture 20, CS70 Summer 2025

RVs and Probability Concepts

- **Joint Distributions and Marginal Distributions**
- Independent RVs and Conditional Probability and RVs
- Linearity of Expectation

Additional Info on Important Distributions

- Memoryless Property of Geometric RVs
- The Sum of Two Poisson RVs
- Poisson as the Limit of Binomial RVs

Multiple Random Variables on the Same Sample Space

Consider the sample space of flipping three coins:

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

Sometimes, we define multiple random variables on the same sample space.

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads.

- $X(TTT) = 0$, $\mathrm{Z}(TTT) = 0$
- $X(HTH) = 2$, $\mathrm{Z}(HTH) = 1$

Quick question: Think of events defined by X and Z – are they independent?

We'll get back to this later – checking intuition for now!

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads. Examples:

- $X(TTT) = 0$, $\text{Z}(TTT) = 0$ $X(HTH) = 2$, $\text{Z}(HTH) = 1$

The joint distribution of two discrete random variables X_1 and X_2 is the collection of values $\{((a,b), P(X_1 = a, X_2 = b)) : a \in \text{range}(X_1)\ b \in \text{range}(X_2)\}$.

Example for $X_1 = X$, $X_2 = Z$:

a	b	$P(X = a, Z = b)$	ω
0	0		
0	1		
1	0		
1	1		

a	b	$P(X = a, Z = b)$	ω
2	0		
2	1		
3	0		
3	1		

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads. Examples:

- $X(TTT) = 0$, $\mathrm{Z}(TTT) = 0$ $\quad$ $X(HTH) = 2$, $\mathrm{Z}(HTH) = 1$

The joint distribution of two discrete random variables X_1 and X_2 is the collection of values $\{((a, b), P(X_1 = a, X_2 = b)) : a \in \text{range}(X_1)\ b \in \text{range}(X_2)\}$.

Example for $X_1 = X$, $X_2 = Z$:

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1		
1	0		
1	1		

a	b	$P(X = a, Z = b)$	ω
2	0		
2	1		
3	0		
3	1		

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads. Examples:

- $X(TTT) = 0$, $\mathrm{Z}(TTT) = 0$ $X(HTH) = 2$, $\mathrm{Z}(HTH) = 1$

The joint distribution of two discrete random variables X_1 and X_2 is the collection of values $\{((a, b), P(X_1 = a, X_2 = b)): a \in \text{range}(X_1)\ b \in \text{range}(X_2)\}$.

Example for $X_1 = X$, $X_2 = Z$:

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0		
1	1		

a	b	$P(X = a, Z = b)$	ω
2	0		
2	1		
3	0		
3	1		

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads. Examples:

- $X(TTT) = 0$, $\mathrm{Z}(TTT) = 0$ $X(HTH) = 2$, $\mathrm{Z}(HTH) = 1$

The joint distribution of two discrete random variables X_1 and X_2 is the collection of values $\{((a,b), P(X_1 = a, X_2 = b)) : a \in \mathrm{range}(X_1)\ b \in \mathrm{range}(X_2)\}$.

What other (a, b) has probability 1/4?

Example for $X_1 = X$, $X_2 = Z$:

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1		

a	b	$P(X = a, Z = b)$	ω
2	0		
2	1		
3	0		
3	1		

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads. Examples:

- $X(TTT) = 0$, $\mathrm{Z}(TTT) = 0$ $\qquad$ $X(HTH) = 2$, $\mathrm{Z}(HTH) = 1$

The joint distribution of two discrete random variables X_1 and X_2 is the collection of values $\{((a, b), P(X_1 = a, X_2 = b)) \colon a \in \mathrm{range}(X_1)\ b \in \mathrm{range}(X_2)\}$.

Example for $X_1 = X$, $X_2 = Z$:

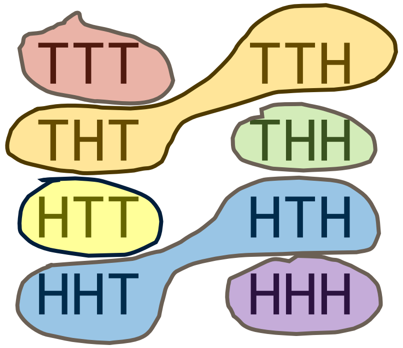
a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Joint Random Variables and Partitions

Joint random variables (also) partition a sample space.

$\Omega =$



a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Marginal Probability

Given a joint distribution $P(X_1 = a, X_2 = b)$, the distribution $P(X_1 = a)$ of X_1 is called the marginal distribution of X_1. We can compute the marginal distribution by summing over all values of X_2.

$$P(X_1 = a) = \sum_{b \in \mathrm{range}(X_2)} P(X_1 = a, X_2 = b)$$

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Example, can compute $P(X = a)$ from $P(X = a, Z = b)$.

a	$P(X = a)$	ω
0		
1		
2		
3		

Marginal Probability

Given a joint distribution $P(X_1 = a, X_2 = b)$, the distribution $P(X_1 = a)$ of X_1 is called the marginal distribution of X_1. We can compute the marginal distribution by summing over all values of X_2.

$$P(X_1 = a) = \sum_{b \in \mathrm{range}(X_2)} P(X_1 = a, X_2 = b)$$

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Example, can compute $P(X = a)$ from $P(X = a, Z = b)$.

a	$P(X = a)$	ω
0		
1		
2		
3		

Marginal Probability

Given a joint distribution $P(X_1 = a, X_2 = b)$, the distribution $P(X_1 = a)$ of X_1 is called the marginal distribution of X_1. We can compute the marginal distribution by summing over all values of X_2.

$$P(X_1 = a) = \sum_{b \in \mathrm{range}(X_2)} P(X_1 = a, X_2 = b)$$

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Example, can compute $P(X = a)$ from $P(X = a, Z = b)$.

a	$P(X = a)$	ω
0	1/8	TTT
1		
2		
3		

Marginal Probability

Given a joint distribution $P(X_1 = a, X_2 = b)$, the distribution $P(X_1 = a)$ of X_1 is called the marginal distribution of X_1. We can compute the marginal distribution by summing over all values of X_2.

$$P(X_1 = a) = \sum_{b \in \mathrm{range}(X_2)} P(X_1 = a, X_2 = b)$$

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Example, can compute $P(X = a)$ from $P(X = a, Z = b)$.

a	$P(X = a)$	ω
0	1/8	TTT
1		
2		
3		

What is $P(X = 1)$?

- Which rows above will be pink?

Marginal Probability

Given a joint distribution $P(X_1 = a, X_2 = b)$, the distribution $P(X_1 = a)$ of X_1 is called the marginal distribution of X_1. We can compute the marginal distribution by summing over all values of X_2.

$$P(X_1 = a) = \sum_{b \in \text{range}(X_2)} P(X_1 = a, X_2 = b)$$

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Example, can compute $P(X = a)$ from $P(X = a, Z = b)$.

a	$P(X = a)$	ω
0	1/8	TTT
1	3/8	TTH, THT, HTT
2		
3		

Marginal Probability

Given a joint distribution $P(X_1 = a, X_2 = b)$, the distribution $P(X_1 = a)$ of X_1 is called the marginal distribution of X_1. We can compute the marginal distribution by summing over all values of X_2.

$$P(X_1 = a) = \sum_{b \in \mathrm{range}(X_2)} P(X_1 = a, X_2 = b)$$

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Example, can compute $P(X = a)$ from $P(X = a, Z = b)$.

a	$P(X = a)$	ω
0	1/8	TTT
1	3/8	TTH, THT, HTT
2	3/8	THH, HTH, HHT
3	1/8	HHH

Independent RVs and Conditional Probability and RVs

Lecture 20, CS70 Summer 2025

RVs and Probability Concepts

- Joint Distributions and Marginal Distributions
- **Independent RVs and Conditional Probability and RVs**
- Linearity of Expectation

Additional Info on Important Distributions

- Memoryless Property of Geometric RVs
- The Sum of Two Poisson RVs
- Poisson as the Limit of Binomial RVs

Independence and Events (review)

Suppose we have two events A_1 and A_2 in the same sample space.

What does it mean for these two events to be independent?

True or false?

- The events share no outcomes, i.e., the Venn diagram has no overlap. False!
- $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$ True!

Independence and Events (review)

Suppose we have two events A_1 and A_2 in the same sample space.

What does it mean for these two events to be independent?

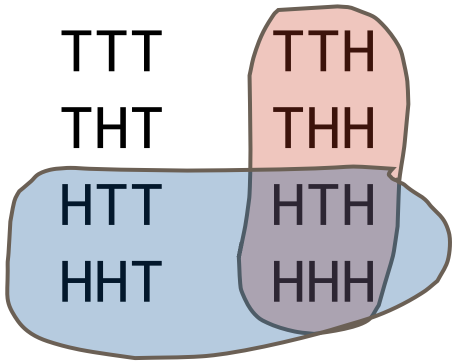
True or false?

- The events share no outcomes, i.e., the Venn diagram has no overlap. False!
- $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$ True!

A_1: First flip is heads.
A_2: Third flip is heads.

$P(A_1 \cap A_2) = P(A_1) \times P(A_2)$

$\Omega =$



Suppose heads comes up 33% of the time:

- $P(A_1 \cap A_2) = 1/9$
- $P(A_1) = 1/3$
- $P(A_2) = 1/3$

Independent Random Variables

Recall that a random variable taking a specific value is just an event (set of outcomes).

Random variables X and Y on the same probability space are said to be independent if the events $X = a$ and $Y = b$ are independent for all values a, b.

- This means: $P(X = a, Y = B) = P(X = a) \cdot P(Y = b), \quad \forall a \in \text{range}(X), b \in \text{range}(Y)$

Example 1: Rolling two dice

Example: Rolling two fair six-sided dice where X is the result of first roll, and Y is the result of the second roll.

These are independent, i.e. we have that:

$$P(X = a, Y = B) = P(X = a) \cdot P(Y = b), \quad \forall a \in \text{range}(X), b \in \text{range}(Y)$$

That is, for $a, b \in \{1, 2, 3, 4, 5, 6\}$, we have that:

- $P(X = a, Y = b) = 1/36$
- $P(X = a) = 1/6$
- $P(Y = b) = 1/6$
- $P(X = a) \cdot P(Y = b) = 1/36$

Example 2: Rolling Two Dice (again)

Example: Suppose we roll two fair six-sided dice where X is the result of first roll, and S is the sum of the two rolls.

Are these independent?

Example 2: Rolling Two Dice (again)

Example: Suppose we roll two fair six-sided dice where X is the result of first roll, and S is the sum of the two rolls.

These random variables are not independent!

- There exist choices a and b such that the probability of event $P(X = a, S = b)$ is not equal to the product of the probabilities of events $P(X = a) \cdot P(S = b)$.

Example: $P(X = 1, S = 8) = 0$, but $P(X = 1) = 1/6$ and $P(S = 8) = 5/36$

- These two events are not independent.

Example 3: Flipping 3 coins – our initial example

Suppose $X(\omega)$ maps each to the number of heads in the sample, $Z(\omega)$ is a binary random variable that is 1 if the first flip is a heads. Examples:

- $X(TTT) = 0$, $Z(TTT) = 0$ $\quad$ $X(HTH) = 2$, $Z(HTH) = 1$

Joint distribution:

a	b	$P(X = a, Z = b)$	ω
0	0	1/8	TTT
0	1	0	∅
1	0	1/4	TTH, THT
1	1	1/8	HTT

a	b	$P(X = a, Z = b)$	ω
2	0	1/8	THH
2	1	1/4	HTH, HHT
3	0	0	∅
3	1	1/8	HHH

Independent? **Justification?** $P(X = 0, Z = 1) = 0$

... but $P(X = 0) = 1/8$ and $P(Z = 1) = 1/2$

Conditional Probability of Events and Random Variables

Suppose we have random variables X_1 through X_5 corresponding to 5 coin flips, where X_i is 1 if the ith flip is heads. Suppose that S is the number of heads, i.e., the random variable S has range $\{0, 1, 2, 3, 4, 5\}$

Naturally, we can write expressions like:

$$P(S = 5 \mid X_1 = 1)$$

There's nothing mathematically new here, this just means the probability of the event $S = 5$ given that $X_1 = 1$.

If coin comes up heads with probability p, then $P(S = 5 \mid X_1 = 1) = p^4$

Random Variables and the Product Rule

Similarly, the product rule applies to events involving random variables.

Again, for our coin example, suppose we have random variables X_1 through X_5 corresponding to 5 coin flips, where X_i is 1 if the ith flip is heads. Suppose that S is the number of heads, i.e., the random variable S has range $\{0, 1, 2, 3, 4, 5\}$

Example: $P(S = 5 \cap X_1 = 1) = P(S = 5 \mid X_1 = 1) \times P(X_1 = 1)$

- $P(S = 5 \cap X_1 = 1) = p^5$
- $P(S = 5 \mid X_1 = 1) = p^4$
- $P(X_1 = 1) = p$

Mutual Independence

We can define mutual independence of three random variables as follows. We saw that random variables X, Y, and Z are mutually independent if:

$$P(X = a, Y = b, Z = c) = P(X = a) \cdot P(Y = b) \cdot P(Z = c)$$

$$\forall a \in \text{range}(X), b \in \text{range}(Y), c \in \text{range}(Z)$$

You can generalize this definition to any number of random variables.

i.i.d. Variables

A common phrase you'll see used in the real world is "independent and identically distributed" random variables, also called i.i.d.

- Here, independent is used to mean mutually independent.

Example: If we roll a die 13 times, the resulting variables X_1 through X_{13} are "independent and identically distributed".

Example: If we take 10 cardboard boxes from an assembly line and test their crush strength in the same way, the resulting observations X_1 through X_{10} are i.i.d.

Linearity of Expectation

Lecture 20, CS70 Summer 2025

RVs and Probability Concepts

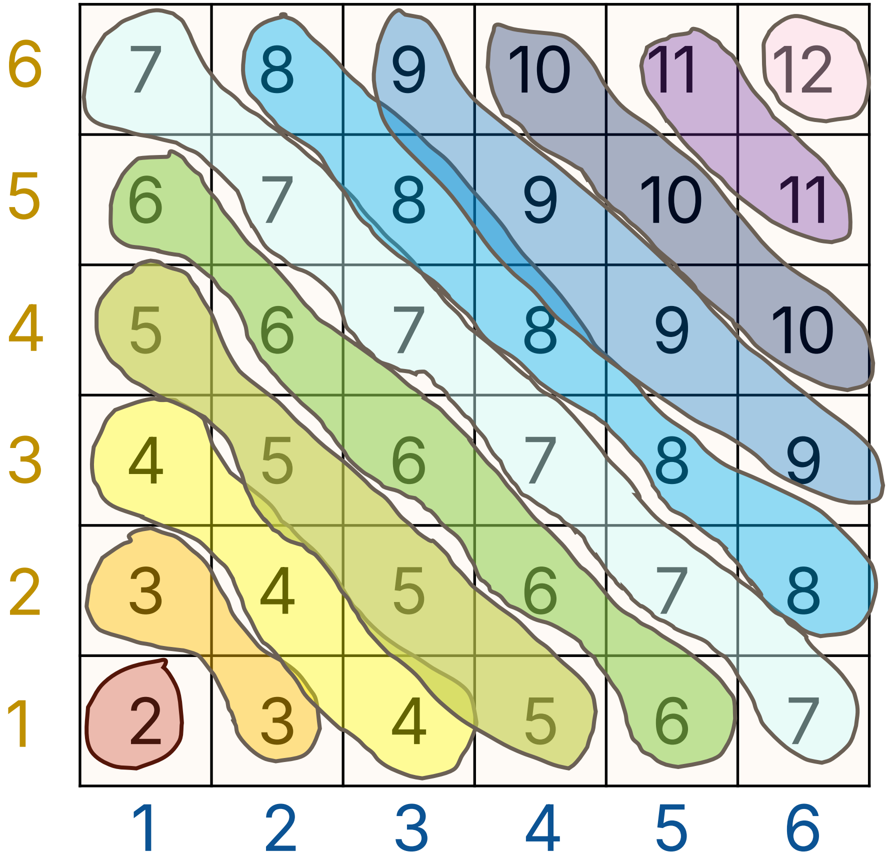
- Joint Distributions and Marginal Distributions
- Independent RVs and Conditional Probability and RVs
- **Linearity of Expectation**

Additional Info on Important Distributions

- Memoryless Property of Geometric RVs
- The Sum of Two Poisson RVs
- Poisson as the Limit of Binomial RVs

Expectation of Two Dice

Let $S = X + Y$ be the result after rolling two six-sided dice. What is $E[S]$?

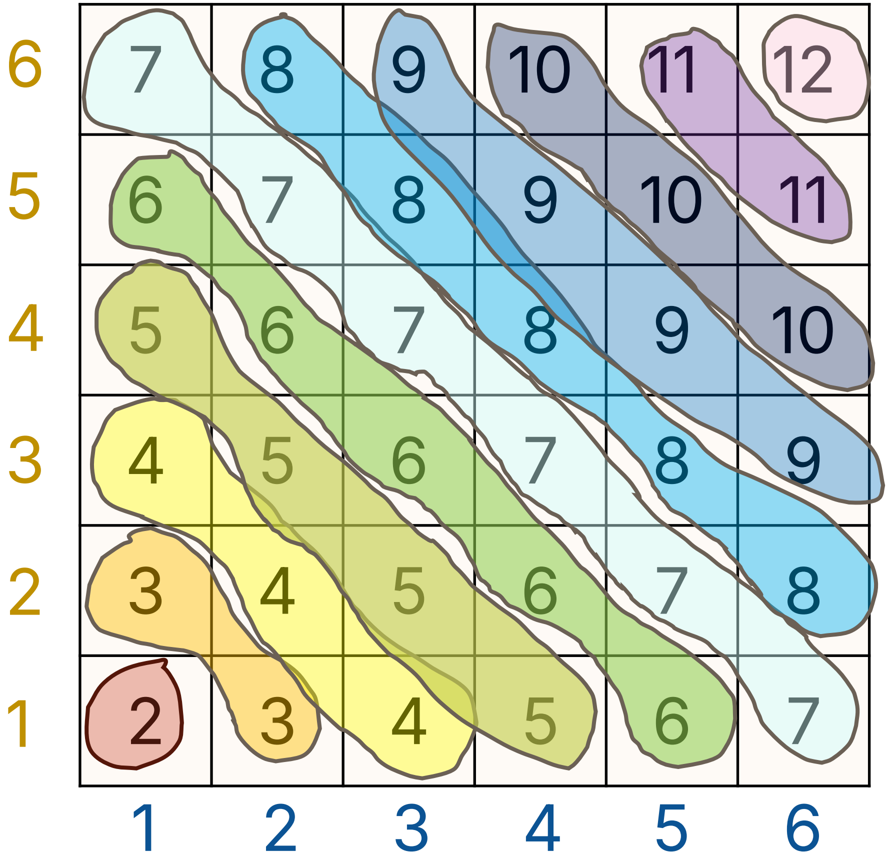


a	$P(S = a)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Expectation of Two Dice

Let $S = X + Y$ be the result after rolling two six dided dice. What is $E[S]$?

The annoying way: $E[S] = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7$



Here we're explicitly summing over the entire range of our random variable S.

a	$P(S = a)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Linearity of Expectation, and Example 1: Two Dice

Note: This is the most important part of the lecture today!

For any two random variables X and Y, $E[X+Y] = E[X] + E[Y]$.

- This is true for any two random variables, **even if not independent**.

Will prove in a moment. First let's consider dice:

- If $S = X + Y$, then $E[S] = E[X] + E[Y]$.
- $E[X] = \frac{1+2+3+4+5+6}{6} = 3.5$
- So $E[S] = 3.5 + 3.5 = 7$ (far easier!)

Prelude: The Mega-Verbose Form for Expectation

Recall that underlying expectation is a sum over all outcomes (even if we don't use this formulation much in practice):

$$E[X] = \sum_{\omega \in \Omega} X(\omega) \times P(\omega)$$

For two dice, it would look something like this, summing explicitly over all 36 outcomes. The verbosity just makes summing more annoying.

$$E[S] = S\big((1,1)\big) \times \frac{1}{36} + S\big((1,2)\big) \times \frac{1}{36} + \cdots + S\big((6,6)\big) \times \frac{1}{36}$$

But for our proof of linearity of expectation, it will be more direct.

Proof of Linearity of Expectation

$$E[X] = \sum_{\omega \in \Omega} X(\omega) \times P(\omega)$$

If $S = X + Y$, we can write $E[S]$ as:

$$E[X + Y] = E[S] = \sum_{\omega \in \Omega} S(\omega) \times P(\omega)$$

$$= \sum_{\omega \in \Omega} \big(X(\omega) + Y(\omega)\big) \times P(\omega)$$

$$= \sum_{\omega \in \Omega} X(\omega) \times P(\omega) \; + \sum_{\omega \in \Omega} Y(\omega) \times P(\omega)$$

$$= E[X] + E[Y]$$

Linearity of Expectation

We can also show that $E[cX] = cE[X]$, where c is some constant. We won't prove this (left as an exercise in the notes).

When we say "Expectation is Linear", we mean that $E[X + Y] = E[X] + E[Y]$, and $E[cX] = cE[X]$.

Note: These two facts can also be combined into one statement, which we also won't prove, but which follows naturally from the other two facts:

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

Example 2: Binomial RV Expectation

Suppose we want to compute the expectation of a random variable $X \sim \text{Binomial}(n, p)$ which has a distribution given by:

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

One way is to use the usual expectation formula:

$$E[X] = \sum_{i=0}^{n} i \times \binom{n}{i} p^i (1-p)^{n-i}$$

This is tedious.

Example 2: Binomial RV Expectation

Suppose we want to compute the expectation of a random variable $X \sim \mathrm{Binomial}(n, p)$ which has a distribution given by:

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

A better approach is to keep in mind that a binomial random variable is just the sum $B_1 + B_2 + \cdots + B_n$ where $\mathrm{B_i} \sim \mathrm{Bernoulli}(p)$.

Since $\mathrm{E}[B_i] = p$, we have that $E[X] = E[B_1] + E[B_2] + \cdots + E[B_n]$

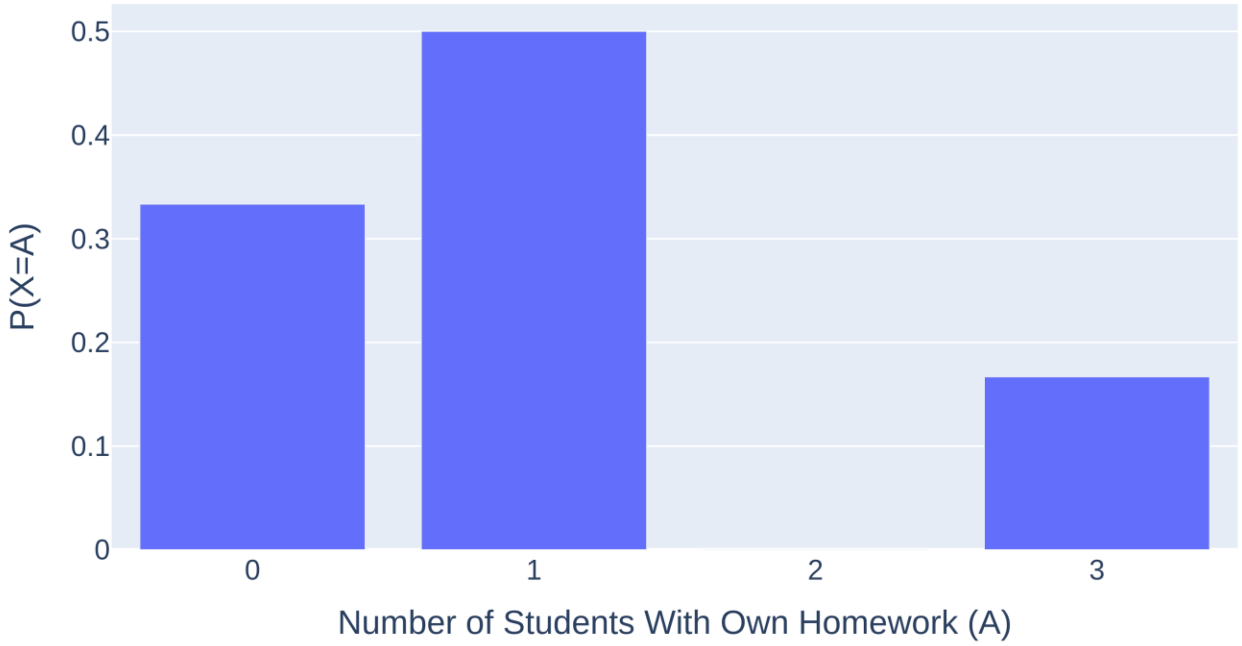
$$= p + p + \cdots + p = np$$

Example 3: Fixed Points

In the previous lecture, we showed the distribution for shuffling 3 students homework and handing it back to them randomly is as shown in this figure.

- Can compute expectation of number of students who get their own homework back as $E[X] = 1 \times \frac{1}{2} + 3 \times \frac{1}{6} = 1$.

Let's try to find the expectation for the case of n students.



Example 3: Fixed Points

Let X_n be the number of students who receive their own homework.

Let I_i be a random variable which is 1 if a student gets their own homework, and 0 otherwise.

- Note, such a random variable is often called an “indicator random variable”. It is 1 when some condition is true, and false otherwise.
- Indicator random variables combined with linearity of expectation is *very powerful*.

Then $X_n = I_1 + I_2 + \cdots + I_n$.

Note: These random variables as NOT independent. Example: If students 1 through n-1 get their own homework, then student n also gets their own homework.

Example 3: Fixed Points

Let X_n be the number of students who receive their own homework.

Let I_i be a random variable which is 1 if a student gets their own homework, and 0 otherwise.

Then $X_n = I_1 + I_2 + \cdots + I_n$.

What is $E[X_n]$?

- Hint: What is $E[I_i]$?

Example 3: Fixed Points

Let X_n be the number of students who receive their own homework.

Let I_i be a random variable which is 1 if a student gets their own homework, and 0 otherwise.

Then $X_n = I_1 + I_2 + \cdots + I_n$.

What is $E[X_n]$?

- Hint: What is $E[I_i]$?
  - The chance a student gets their own homework is $1/n$. That is, each indicator variable is $\mathrm{I_i} \sim \mathrm{Bernoulli}(1/n)$. So $E[I_i] = 1/n$
- Since $E[X_n] = E[I_1] + E[I_2] + \cdots + E[I_n]$, we have $E[X_n] = n \times 1/n = 1$

Are the I_k's independent?

No!

Does it matter?

No!

Memoryless Property of Geometric RVs

Lecture 20, CS70 Summer 2025

RVs and Probability Concepts

- Joint Distributions and Marginal Distributions
- Independent RVs and Conditional Probability and RVs
- Linearity of Expectation

Additional Info on Important Distributions

- **Memoryless Property of Geometric RVs**
- The Sum of Two Poisson RVs
- Poisson as the Limit of Binomial RVs

Geometric Random Variables

Recall that a geometric random variable X has probabilities given by:

$$P(X = i) = (1 - p)^{i-1}p \text{ for } i = 1, 2, \ldots$$

A geometric random variable models the process of waiting for an event to occur for the first time:

- Flipping a coin until you get heads
- Operating a system until it fails
- Betting until you win

  "The lottery is a tax on people who are bad at math."

Geometric Random Variables and the Memoryless Property

The memoryless property

- It basically means: Even if you've waited a long time, the chance of the event occurring soon isn't any better.
- Example: You're flipping a coin and it came up tails 10 times in a row. The chance of the next heads isn't better just because you had 10 tails.
- Example: You're playing poker and have gotten 5 terrible hands in a row. The chance of getting a good hand is the same, and it doesn't matter that you've had 5 terrible hands.

Note: The "Gambler's Fallacy" – and lucky (or unlucky) streaks....

The memoryless property

- It basically means: Even if you've waited a long time, the chance of the event occurring soon isn't any better.

In terms of probability, we can state this as "If we've already tossed a coin m times without getting a head, the probability that we need n additional tosses is the same as if we hadn't tossed the first m coins at all."

Or symbolically we have:

$$P(X > n + m \mid X > m) = P(X > n)$$

Geometric Random Variables and the Memoryless Property

The memoryless property

- It basically means: Even if you've waited a long time, the chance of the event occurring soon isn't any better.

Wait – isn't this supposed to be an intersection?

Yes... and it is... see why?

Proof of $\mathrm{P}(X > n + m \mid X > m) = P(X > n)$

$$\mathrm{P}(X > n + m \mid X > m) = \frac{P(X > n + m)}{P(X > m)}$$

$$= \frac{(1-p)^{n+m}}{(1-p)^{m}}$$

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

$$= P(X > n)$$

The Sum of Two Poisson RVs

Lecture 20, CS70 Summer 2025

RVs and Probability Concepts

- Joint Distributions and Marginal Distributions
- Independent RVs and Conditional Probability and RVs
- Linearity of Expectation

Additional Info on Important Distributions

- Memoryless Property of Geometric RVs
- **The Sum of Two Poisson RVs**
- Poisson as the Limit of Binomial RVs

The Poisson Random Variable

Reminder: If $X \sim \text{Poisson}(\lambda)$, then the distribution is given by:

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \qquad \text{for } i = 0, 1, 2, \ldots$$

The Sum of Independent Poisson RVs

In the previous lecture, we mentioned that if $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$, and $S = X + Y$, then $S \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

- This makes intuitive sense, but let's prove using the idea of joint distributions.

Before we start, recall the Binomial Theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

Example: $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$

Or alternately: $(a+b)^4 = \binom{4}{0}a^4 + \binom{4}{1}a^3b + \binom{4}{2}a^2b^2 + \binom{4}{3}ab^3 + \binom{4}{4}b^4$

The Sum of Independent Poisson RVs

In the previous lecture, we mentioned that if $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, and $S = X + Y$, then $S \sim \text{Poisson}(\lambda + \mu)$. For all $k \in \{0, 1, 2, \ldots\}$, we have:

$$P(S = k) = \sum_{j=0}^{k} P(X = j, Y = k - j)$$

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

$$= \sum_{j=0}^{k} \frac{\lambda^j}{j!} e^{-\lambda} \frac{\mu^{k-j}}{(k-j)!} e^{-\mu} \quad = e^{-(\lambda+\mu)} \sum_{j=0}^{k} \frac{\lambda^j \mu^{k-j}}{j!\,(k-j)!}$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{j!\,(k-j)!} \lambda^j \mu^{k-j}$$

The Sum of Independent Poisson RVs

In the previous lecture, we mentioned that if $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, and $S = X + Y$, then $S \sim \text{Poisson}(\lambda + \mu)$. For all $k \in \{0, 1, 2, \dots\}$, we have:

$$P(S = k) = \sum_{j=0}^{k} P(X = j, Y = k - j)$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{j!\,(k-j)!} \lambda^j \mu^{k-j}$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} (\lambda + \mu)^k$$

$$= \frac{(\lambda + \mu)^k}{k!} e^{-(\lambda+\mu)}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \qquad \text{for } i = 0, 1, 2, \dots$$

Poisson as the Limit of Binomial RVs

Lecture 20, CS70 Summer 2025

RVs and Probability Concepts

- Joint Distributions and Marginal Distributions
- Independent RVs and Conditional Probability and RVs
- Linearity of Expectation

Additional Info on Important Distributions

- Memoryless Property of Geometric RVs
- The Sum of Two Poisson RVs
- **Poisson as the Limit of Binomial RVs**

Poisson Distribution

The distribution of a Poisson random variable is given by:

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \qquad \text{for } i = 0, 1, 2, \dots$$

Unlike our other RVs, we never derived this expression. Let's derive it from scratch.

- Fact from Calculus: $\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e \quad \Longrightarrow \quad \lim_{n \to \infty} \left(1 - \frac{x}{n}\right)^n = e^{-x}$

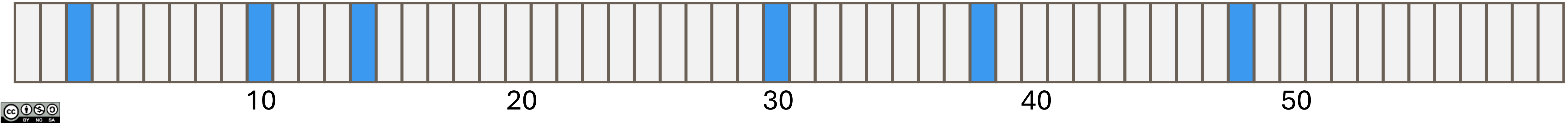
Modeling Calls with a Binomial Distribution

As a thought experiment, let's imagine we want to model the number of phone calls that users of a mobile network will initiate per unit time. We have a vast number of users, any of whom can make a call.

Suppose we pick a unit time of 1 minutes. Let's try modeling the number of calls as a binomial random variable.

- ith coin flip is whether a call is initiated in the ith time interval.
- n is number of intervals, e.g., $n = 60$, we split minute into 60 intervals of 1s each.
- p is the chance of a call in any interval. Assumed independent.

If a call is initiated at times 3.3s, 10.5s, 14.2s, 30.1s, 38.9s, and 48.3s we have:

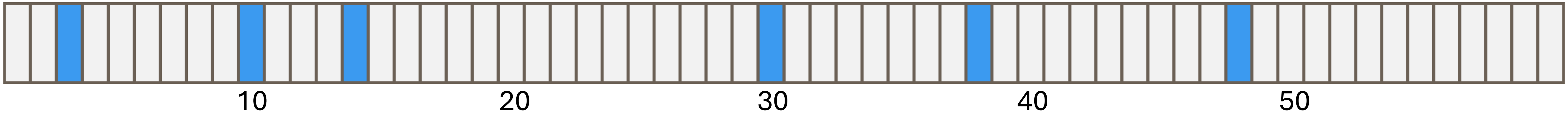


Modeling Calls with a Binomial Distribution

Suppose we pick a unit time of 1 minutes. Let's try modeling the number of calls as a binomial random variable.

- ith coin flip is whether a call is initiated in the ith time interval.
- p is the chance of getting a call in any interval. Assumed independent.
- For example, if $n = 60$, we split a minute into 60 intervals of 1s each.

If a call is initiated at times 3.3s, 10.5s, 14.2s, 30.1s, 38.9s, and 48.3s we have:



Why might a Binomial random variable not work for this modeling process?

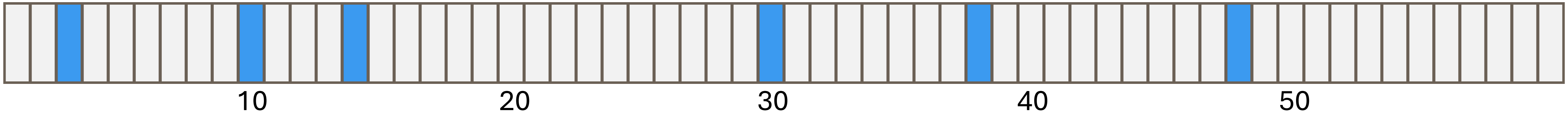
- What could happen in the world that this model cannot handle?

Modeling Calls with a Binomial Distribution

Suppose we pick a unit time of 1 minutes. Let's try modeling the number of calls as a binomial random variable.

- ith coin flip is whether a call is initiated in the ith time interval.
- p is the chance of getting a call in any interval. Assumed independent.
- For example, if $n = 60$, we split a minute into 60 intervals of 1s each.

If a call is initiated at times 3.3s, 10.5s, 14.2s, 30.1s, 38.9s, and 48.3s we have:



Why might a Binomial random variable not work for this modeling process?

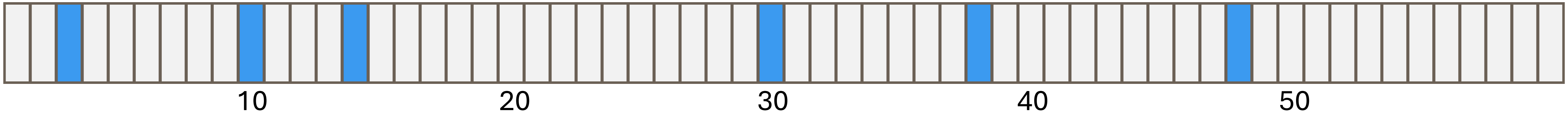
- We might get two calls in the same interval, e.g., a call at 3.3s and 3.8s.
- So what should we do about this?

Modeling Calls with a Binomial Distribution

Suppose we pick a unit time of 1 minutes. Let's try modeling the number of calls as a binomial random variable.

- ith coin flip is whether a call is initiated in the ith time interval.
- p is the chance of getting a call in any interval. Assumed independent.
- For example, if $n = 60$, we split a minute into 60 intervals of 1s each.

If a call is initiated at times 3.3s, 10.5s, 14.2s, 30.1s, 38.9s, and 48.3s we have:



Why might a Binomial random variable not work for this modeling process?

- We might get two calls in the same interval, e.g., a call at 3.3s and 3.8s.
- So what should we do about this? Pick a smaller interval.

Adjusting Time Intervals

Assume that we know that we receive λ calls per minute on average.

If we split the minute into n time intervals (e.g., $n = 60$ means 1 second intervals), and we model the number of calls X as a Binomial random variable, which can we say below?

a) $X \sim \text{Binomial}(n, \lambda)$

b) $X \sim \text{Binomial}\left(n, \frac{1}{\lambda}\right)$

c) $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$

d) $X \sim \text{Binomial}\left(n, \frac{n}{\lambda}\right)$

Adjusting Time Intervals

Assume that we know that we receive λ calls per minute on average.

If we split the minute into n time intervals (e.g., $n = 60$ means 1 second intervals), and we model the number of calls X as a Binomial random variable, which can we say below?

- $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$

Chance of a call coming in during a time interval is going to be number of calls per minutes divided by number of time intervals per minute.

- Example: $\lambda = 100$ calls per minute, and $n = 60{,}000{,}000$, then probability of a call in any microsecond is $p = \frac{100}{60{,}000{,}000}$.
- Note: We want to pick n so large that the probability of two events in one interval is negligible.

Example – a large n

We're modeling the number of calls as $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$.

Example: If $n = 60{,}000{,}000$ time intervals and $\lambda = 120$ calls per minute, then the probability that the number of calls in a given minute is equal to 100 is given by the binomial distribution:

$$P(X = 100) = \binom{60{,}000{,}000}{100}\left(\frac{120}{60{,}000{,}000}\right)^{100}\left(1 - \frac{120}{60{,}000{,}000}\right)^{60{,}000{,}000-100}$$

This is an awkward computation. Luckily:

- As n gets larger, the model is increasingly accurate (less likely that we get two calls being initiated in the same time interval).
- In the limit as $n \to \infty$, this expression is simpler.

The Limit as $n \to \infty$

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X=i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together. And remember out end-goal!

$$= \frac{\lambda^i}{i!}$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right)$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}$$

The limit of blue subexpression is:

$$\left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) = \frac{n \cdot (n-1) \cdots (n-i+1) \cdot (n-i)!}{(n-i)!} \cdot \frac{1}{n^i} = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{(n-i+1)}{n} \to 1$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}$$

The limit of blue subexpression is 1.

What is the limit of the gold?

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

Calculus fact from beginning of analysis...

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}$$

The limit of blue subexpression is 1, and the limit of the gold is $e^{-\lambda}$.

What is the limit of the purple?

$$\left(1 - \frac{\lambda}{n}\right)^{-i} \to (1-0)^{-i} = 1$$

The Large N Limit

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

This rearrangement of terms may seem arbitrary. Rough goal: Get terms that will simplify when n is large all clustered together.

$$= \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}$$

The limit of blue subexpression is 1, the limit of the gold is $e^{-\lambda}$, limit of purple is 1.

Plugging these in we have:

$$\to \frac{\lambda^i}{i!} e^{-\lambda}$$

Poisson Distribution

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$.

In the limit as $n \to \infty$, this becomes $X \sim \text{Poisson}(\lambda)$, i.e., $P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$

Binomial vs Poisson Distribution (Call Center Example)

If $X \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, we have: $P(X = i) = \binom{n}{i} \cdot \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$.

In the limit as $n \to \infty$, this becomes $X \sim \text{Poisson}(\lambda)$, i.e., $P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$

Example: For our call center example, we had $X \sim \text{Binomial}\left(60{,}000{,}000, \frac{120}{60{,}000{,}000}\right)$

$$P(X = 100) = \binom{60{,}000{,}000}{100} \left(\frac{120}{60{,}000{,}000}\right)^{100} \left(1 - \frac{120}{60{,}000{,}000}\right)^{60{,}000{,}000 - 100}$$

If instead we model as $X \sim \text{Poisson}(120)$, we have:

$$P(X = 100) = \frac{120^{100}}{100!} e^{-120}$$

These two values are close if you compute them! We won't.

Summary of the Poisson Distribution

A Poisson random variable tells you the number of events that occur per unit time.

The critical assumptions to be Poisson are:

- Knowing the number of events in one time interval does not provide any information about any other disjoint time interval.
- The number of events that occur per unit time are some average constant λ.

The specific modeling choice that yields the Poisson distribution is to model the number of occurrences over a sequence of very short time intervals as a binomial.

- If n is sufficiently large / p is sufficiently small, the Binomial distribution can be approximated by the Poisson distribution with parameter $\lambda = np$.