Poisson Random Variables

Lecture 20, CS70 Fall 2026

Poisson Random Variables

- Derivation of Poisson RV

Jointly Distributed RVs

- Joint Distributions and Marginal Distributions
- Independent RVs, Conditional Probability and RVs
- The Sum of Two Poisson RVs

Linearity of Expectation

CC BY NC SA

Poisson Distribution

Suppose we want to know how many alpha particles will be emitted by a radiation source. Let X be the number of emissions per unit time.

Assumptions:

- The average number of emissions is λ alpha particles per unit time.
- Emissions in disjoint time intervals are independent, e.g. if 2 particles were emitted in the interval [0, 2], this tells us absolutely nothing about how many were emitted in the interval [2.1, 4].

If assumptions hold, $X \sim \text{Poisson}(\lambda)$.

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

CC BY NC SA

Poisson Distribution Example

Suppose we want to know how many alpha particles will be emitted by a radiation source. Let X be the number of emissions per unit time.

Suppose we model this radiation source as $X \sim \text{Poisson}(\lambda)$, where $\lambda = 2$ emissions per minute.

- What is the probability of seeing 3 emissions in any given minute?

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

$$P(X = 3) = \frac{2^3}{3!} e^{-2} = \frac{8}{6} e^{-2} = 18.04\%$$

CC BY NC SA

Useful Poisson Distribution Facts

The Poisson distribution comes from the large n limit for Binomial RVs.

- That is, can think of $\frac{\lambda^i}{i!}e^{-\lambda}$ as arising from an infinite number of Bernoulli trials.
- We'll derive this shortly. For now it's just **magic**.

The expected value of a Poisson random variable is λ.

- This is by definition! λ was defined as the average number of events per unit time.

If we add independent $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, and $Z = X + Y$, then $Z \sim \text{Poisson}(\lambda + \mu)$.

- We're not equipped to prove this yet. We'll prove it later today.

CC BY NC SA

Poisson Distribution Example 2

Suppose we have two radiation sources. We model one as $X_1 \sim \text{Poisson}(\lambda_1)$, and the other as $X_2 \sim \text{Poisson}(\lambda_2)$.

- If $\lambda_1 = 2$ emissions per minute, and $\lambda_2 = 5$ emissions per minute, what is the chance that the total number of emissions in a given minute is 4 emissions?

I claimed that the sum of two independent Poisson RVs is also Poisson. So $S = X_1 + X_2$ is a random variable where $S \sim \text{Poisson}(7)$.

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

$$P(S = 4) = \frac{7^4}{4!} e^{-7} \quad = 9.1\%$$

Poisson Distribution Example 3

Suppose we want to know how many alpha particles will be emitted by a radiation source. Let X be the number of emissions per unit time.

Suppose we model this radiation source as $X \sim \text{Poisson}(\lambda)$, where $\lambda = 2$ emissions per minute.

- What is the probability that we get 100 emissions in an hour?

CC BY NC SA

Poisson Distribution Example 3

Suppose we want to know how many alpha particles will be emitted by a radiation source. Let X be the number of emissions per unit time.

Suppose we model this radiation source as $X \sim \text{Poisson}(\lambda)$, where $\lambda = 2$ emissions per minute.

- What is the probability that we get 100 emissions in an hour?
- Need to convert our units. 2 emissions per minute is 120 emissions per hour.
- Let Y be X, but in units of emissions per hour. Then $Y \sim \text{Poisson}(120)$.

$$P(Y = i) = \frac{\lambda^i}{i!} e^{-\lambda} \qquad P(Y = 100) = \frac{120^{100}}{100!} e^{-120} = 0.68\%$$

CC BY NC SA

Poisson Distribution Example 3 (Alternate Solution)

Suppose we want to know how many alpha particles will be emitted by a radiation source. Let X be the number of emissions per unit time.

Suppose we model this radiation source as $X \sim \text{Poisson}(\lambda)$, where $\lambda = 2$ emissions per minute.

- What is the probability that we get 100 emissions in an hour?
- The number of emissions per hour can be thought of as the sum of 60 independent RVs that are $\sim \text{Poisson}(\lambda)$.
- Let S be this sum. Then $S \sim \text{Poisson}(120)$.

$$P(S = i) = \frac{\lambda^i}{i!} e^{-\lambda} \qquad P(S = 100) = \frac{120^{100}}{100!} e^{-120} = 0.68\%$$

CC BY NC SA

Derivation of Poisson RV

Lecture 20, CS70 Fall 2026

Poisson Random Variables

- **Derivation of Poisson RV**

Jointly Distributed RVs

- Joint Distributions and Marginal Distributions
- Independent RVs, Conditional Probability and RVs
- The Sum of Two Poisson RVs

Linearity of Expectation

CC BY NC SA

Poisson Distribution

Now that we have some basic familiarity with the Poisson RV, let's:

- Derive the Poisson random variable from first principles.
- Show that the sum of two independent Poisson random variables is also Poisson.

For this derivation, we need to know that if $\mathrm{X} \sim \mathrm{Binomial}(n, \mathrm{p})$, then $\mathrm{E}[\mathrm{X}] = \mathrm{np}$.

- Will show this to be true later today. For now, just assume $\mathrm{E}[\mathrm{X}] = \mathrm{np}$.

CC BY NC SA

Poisson Distribution

The distribution of a Poisson random variable is given by:

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

Unlike our other RVs, we never derived this expression. Let's talk about how this RV can be derived from an infinite series of coin flips.

- Note: We'll use the fact that $\lim_{n \to \infty} \left(1 - \frac{x}{n}\right)^n = e^{-x}$. We won't prove this. Sorry.

CC BY NC SA

Modeling Calls with a Binomial Distribution

Motivation: As a thought experiment, suppose we are the operators of a rural phone network, and want to model the number of phone calls we're likely to receive per minute.

Suppose we know that our network receives an average of four phone calls per minute.

Let's try to model the number of calls per minute as a binomial random variable.

- We'll divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.

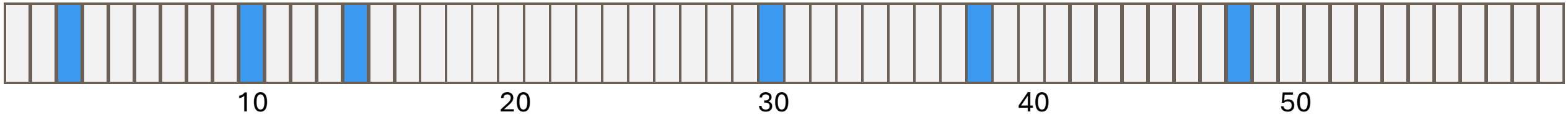
CC BY NC SA

Modeling Calls with a Binomial Distribution

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.

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



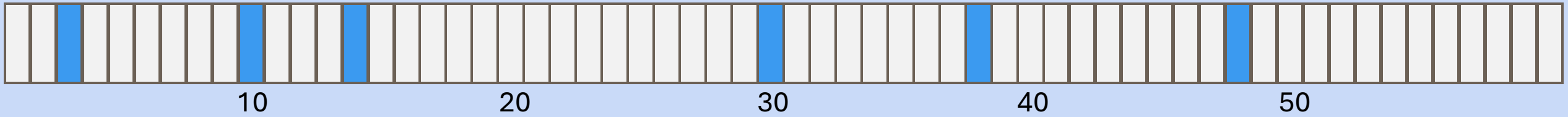
This specific outcome is one of $\binom{60}{6}$ outcomes that comprise the event $X = 6$

Modeling Calls with a Binomial Distribution

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.

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



What is the n parameter of our binomial random variable?

- $\mathrm{X} \sim \mathrm{Binomial}(n = ??, \mathrm{p} = ??)$

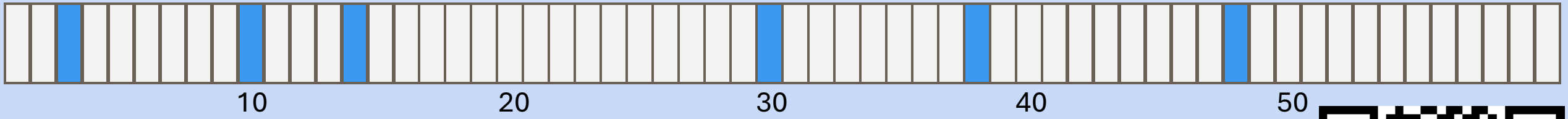
https://joshh.ug/F

Modeling Calls with a Binomial Distribution

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average.

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



What is the p parameter of our binomial random variable?

- $\mathrm{X} \sim \mathrm{Binomial}(n = 60, \mathrm{p} = ??)$

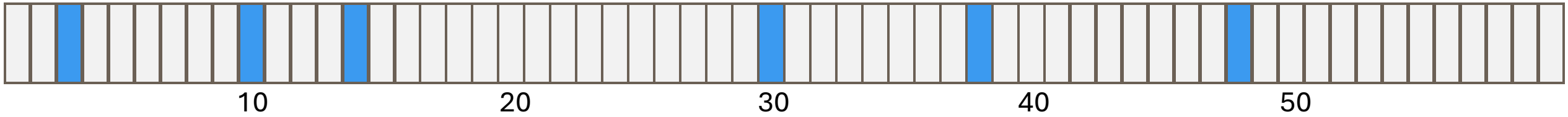
https://joshh.ug/G

Modeling Calls with a Binomial Distribution

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average, i.e. $\mathrm{E[X]} = 4 = \mathrm{np} = 60\mathrm{p}$

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



What is the p parameter of our binomial random variable?

- $\mathrm{X} \sim \mathrm{Binomial}\left(n = 60, \mathrm{p} = \frac{4}{60}\right)$

CC BY NC SA

Modeling Calls with a Binomial Distribution

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average, i.e. $\mathrm{E[X]} = 4 = \mathrm{np} = 60\mathrm{p}$

Resulting random variable is $\mathrm{X} \sim \mathbf{Binomial}\left(60, \frac{4}{60}\right)$

Can now compute quantities of interest e.g. $\mathrm{P(X = 0)}$.

- What does this quantity mean?
- How do we compute it?

CC BY NC SA

Modeling Calls with a Binomial Distribution

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average, i.e. $\mathrm{E[X]} = 4 = \mathrm{np} = 60\mathrm{p}$

Resulting random variable is $\mathrm{X} \sim \mathbf{Binomial}\left(60, \frac{4}{60}\right)$

Can now compute quantities of interest e.g. $\mathrm{P}(\mathrm{X} = 0)$.

- What does this quantity mean? Probability we get no calls in a minute.
- How do we compute it? $\mathrm{P}(\mathrm{X} = 0) = (1 - p)^n = \left(\frac{56}{60}\right)^{60} \approx 1.6\%$

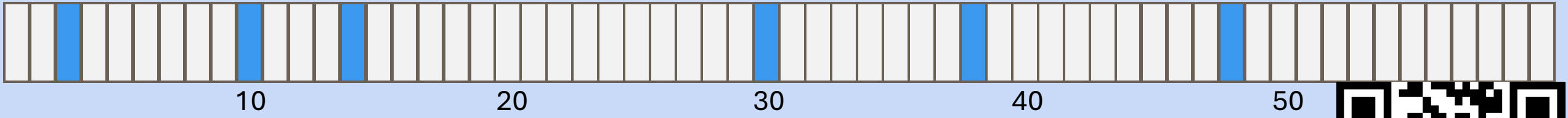
CC BY NC SA

A Fundamental Limitation of the Binomial Model

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average, i.e. $\mathrm{E[X]} = 4 = \mathrm{np} = 60\mathrm{p}$

Resulting random variable is $\mathrm{X} \sim \mathrm{Binomial}\left(60, \frac{4}{60}\right)$



Why might a Binomial random variable not work for this modeling proces

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

https://joshh.ug/H

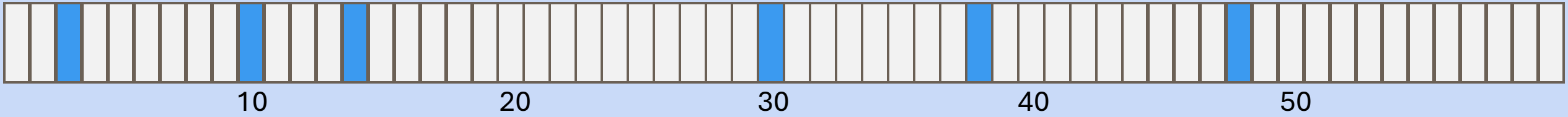
CC BY NC SA

A Fundamental Limitation of the Binomial Model

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average, i.e. $\mathrm{E[X]} = 4 = \mathrm{np} = 60\mathrm{p}$

Resulting random variable is $\mathrm{X} \sim \mathbf{Binomial}\left(60, \frac{4}{60}\right)$



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?

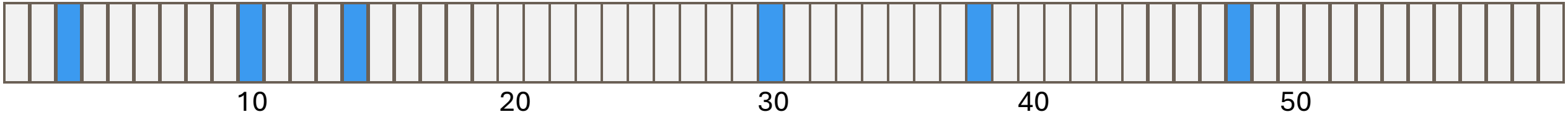
CC BY NC SA

A Fundamental Limitation of the Binomial Model

Let X be a binomial random variable modeling the number of calls per minute:

- Divide each minute into n slices, e.g. if we pick $n = 60$, each slice is one second.
- We'll "flip a coin" for each slice. If it comes up "heads", that means a call was made in that slice. If "tails", no call was made in that slice.
- We know that we get 4 calls per minute on average, i.e. $\mathrm{E[X]} = 4 = \mathrm{np} = 60\mathrm{p}$

Resulting random variable is $\mathrm{X} \sim \mathbf{Binomial}\left(60, \frac{4}{60}\right)$



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? Shorten our time interval.

Adjusting Time Intervals

Assume that we know that we receive λ calls per minute on average and we divide each minute into n slices.

- Example: For $n = 60$ and $\lambda = 4$, binomial model was $X \sim \text{Binomial}\left(60, \frac{4}{60}\right)$

If we split each minute into n slices, what is the resulting distribution?

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

CC BY NC SA

Adjusting Time Intervals

Assume that we know that we receive λ calls per minute on average and we divide each minute into n slices.

If we split each minute into n slices, what is the resulting distribution?

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

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

- Example: $\lambda = 4$ calls per minute, and $n = 60{,}000{,}000$, then probability of a call in any microsecond is $p = \frac{4}{60{,}000{,}000}$.
  - Example: $\text{P}(\text{X} = 0) = (1-p)^n = \left(1 - \frac{4}{60{,}000{,}000}\right)^{60{,}000{,}000} \approx 1.83\%$
- Idea: Pick n so large that chance of two events in one interval is negligible.

CC BY NC SA

One Second Time intervals vs. One Microsecond Time Intervals

The ground truth is that we receive on average $\lambda = 4$ calls per minute.

Two choices of model (we've discussed):

- 1 second slices: $\mathrm{P}(\mathrm{X} = 0) = (1 - p)^n = \left(\frac{56}{60}\right)^{60} \approx 1.6\%$
- 1 microsecond slices: $\mathrm{P}(\mathrm{X} = 0) = (1 - p)^n = \left(1 - \frac{4}{60{,}000{,}000}\right)^{60{,}000{,}000} \approx 1.83\%$

The latter is more accurate.

- That is, if you gather experimental data, the second model is likely to give a more accurate result.
- Why?

CC BY NC SA

One Second Time intervals vs. One Microsecond Time Intervals

The ground truth is that we receive on average $\lambda = 4$ calls per minute.

Two choices of model (we've discussed):

- 1 second slices: $\mathrm{P}(\mathrm{X} = 0) = (1 - p)^n = \left(\frac{56}{60}\right)^{60} \approx 1.6\%$
- 1 microsecond slices: $\mathrm{P}(\mathrm{X} = 0) = (1 - p)^n = \left(1 - \frac{4}{60{,}000{,}000}\right)^{60{,}000{,}000} \approx 1.83\%$

The latter is more accurate.

- That is, if you gather experimental data, the second model is likely to give a more accurate result.
- Why? You're picking an n so large that the probability of overlap (two calls in one slice) is very small.

cc BY NC SA

Computation is Awkward with 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 $\lambda = 4$ calls per minute, then the probability that the number of calls in a given minute is equal to 17 is given by the binomial distribution:

$$P(X = 17) = \binom{60{,}000{,}000}{17} \left(\frac{4}{60{,}000{,}000}\right)^{17} \left(1 - \frac{4}{60{,}000{,}000}\right)^{60{,}000{,}000-17}$$

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 very large n limit, this expression is simpler (after we go through algebra hell)

CC BY NC SA

The Large n Limit

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

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

The limit as $n \to \infty$ of blue subexpression is:

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

The limit as $n \to \infty$ of blue subexpression is:

$$\left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) = \left(n \cdot (n-1) \cdot (n-2) \cdots (n-i+1)\right) \cdot \frac{1}{n^i} = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{(n-i+1)}{n} \to 1$$

Example: $\frac{10!}{6!} \cdot \frac{1}{10^4} = (10 \cdot 9 \cdot 8 \cdot 7) \cdot \frac{1}{10^4} = \frac{10}{10} \cdot \frac{9}{10} \cdot \frac{8}{10} \cdot \frac{7}{10}$

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: Cluster together terms that will simplify when n is very large.

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

The limit of blue subexpression is 1.

What is the limit of the gold?

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

We won't prove this, you are welcome to do so.

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

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

What is the limit of the purple?

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

CC BY NC SA

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: Cluster together terms that will simplify when n is very large.

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

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:

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

CC BY NC SA

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

CC BY NC SA

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 large n limit, 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 $X \sim \text{Binomial}\left(60{,}000{,}000, \frac{4}{60{,}000{,}000}\right)$

$$P(X = 17) = \binom{60{,}000{,}000}{17} \left(\frac{4}{60{,}000{,}000}\right)^{17} \left(1 - \frac{4}{60{,}000{,}000}\right)^{60{,}000{,}000-17}$$

If instead we model as $X \sim \text{Poisson}(4)$, we have:

$$P(X = 17) = \frac{4^{17}}{17!} e^{-4}$$

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

CC BY NC SA

Summary of the Poisson Distribution

A Poisson random variable tells you the number of events that occur in a (fixed duration) time interval.

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 time interval is 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 $\lambda = np$.

CC BY NC SA

Joint Distributions and Marginal Distributions

Lecture 20, CS70 Fall 2026

Poisson Random Variables

- Derivation of Poisson RV

Jointly Distributed RVs

- **Joint Distributions and Marginal Distributions**
- Independent RVs, Conditional Probability and RVs
- The Sum of Two Poisson RVs

Linearity of Expectation

CC BY NC SA

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

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

CC BY NC SA

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each outcome 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:

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

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		

CC BY NC SA

Multiple Random Variables on the Same Sample Space

Suppose $X(\omega)$ maps each outcome 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:

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

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 outcome 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:

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

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

https://joshh.ug/B

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

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

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	1/4	TTH, THT
1	1		

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

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 outcome 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:

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

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

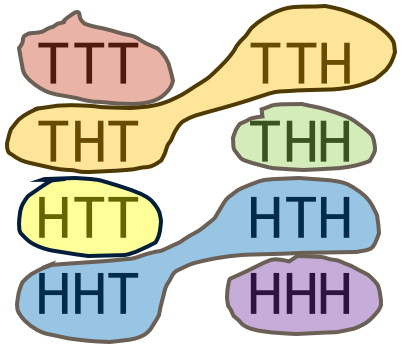
Example for $X_1 = X$, $X_2 = Z$:

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

Suppose we want to know $P(X = 1)$. How would we compute this from the table below?

$$\Omega = \begin{matrix} \text{TTT} & \text{TTH} \\ \text{THT} & \text{THH} \\ \text{HTT} & \text{HTH} \\ \text{HHT} & \text{HHH} \end{matrix}$$

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 (Warmup Example)

Suppose we want to know $P(X = 1)$. How would we compute this from the table below?

- Just add $\boldsymbol{P}(X = 1, Z = 0) + \boldsymbol{P}(X = 1, Z = 1) = 1/4 + 1/8 = 3/8$

$$\Omega = \begin{matrix} \text{TTT} & \text{TTH} \\ \text{THT} & \text{THH} \\ \text{HTT} & \text{HTH} \\ \text{HHT} & \text{HHH} \end{matrix}$$

This is called a "marginal probability", i.e. the probability of a single event (here $X = 1$), without consideration of any related events.

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.

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

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		
2		
3		

CC BY NC SA

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		
2		
3		

CC BY NC SA

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		
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.

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

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		
2		
3		

CC BY NC SA

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.

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

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	3/8	THH, HTH, HHT
3	1/8	HHH

Independent RVs, Conditional Probability and RVs

Lecture 20, CS70 Fall 2026

Poisson Random Variables

- Derivation of Poisson RV

Jointly Distributed RVs

- Joint Distributions and Marginal Distributions
- **Independent RVs, Conditional Probability and RVs**
- The Sum of Two Poisson RVs

Linearity of Expectation

CC BY NC SA

Independence and Events (review)

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

What does it mean for two events to be independent? Select all that apply:

- The events share no outcomes, i.e. the Venn diagram has no overlap.
- $\boldsymbol{P}(A_1 \cap A_2) = \boldsymbol{P}(A_1) \times \boldsymbol{P}(A_2)$
- Neither of these.

https://joshh.ug/C

$$\Omega = \begin{matrix} \text{TTT} & \text{TTH} \\ \text{THT} & \text{THH} \\ \text{HTT} & \text{HTH} \\ \text{HHT} & \text{HHH} \end{matrix}$$

(We covered this in lecture 18)

Independence and Events (review)

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

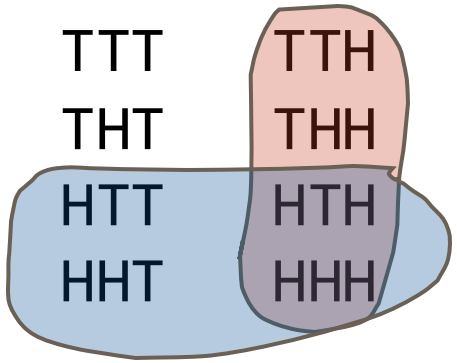
What does it mean for two events to be independent? Select all that apply:

- ~~The events share no outcomes, i.e. the Venn diagram has no overlap.~~
- $\boldsymbol{P(A_1 \cap A_2) = P(A_1) \times P(A_2)}$ - This is how we defined independent.
- ~~Neither of these.~~

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

$$\boldsymbol{P(A_1 \cap A_2) = P(A_1) \times P(A_2)}$$

$\Omega =$



Suppose coins are independent and heads comes up 1/3 of the time:

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

Independent Random Variables

We can define independent random variables in a similar manner to our definition for events.

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 in the range of the respective random variables.

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

CC BY NC SA

Example 1: Rolling two dice

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

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

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

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

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

CC BY NC SA

Example 2: Rolling Two Dice (again)

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

Are these independent?

- If so: Why?
- If not: Can you find a pair of values a and b such that $\boldsymbol{P}(X = a, S = b) \neq \boldsymbol{P}(X = a) \cdot \boldsymbol{P}(S = b)$?

https://joshh.ug/D



CC BY NC SA

Example 2: Rolling Two Dice (again)

Example: Rolling two fair six sided dice where X is the result of the 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 the event $\boldsymbol{P}(X = a, S = b)$ is not equal to the product of the probabilities of the events $\boldsymbol{P}(X = a) \cdot \boldsymbol{P}(S = b)$.

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

- Thus, these two events are not independent, and therefore X and S are not independent.

CC BY NC SA

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:

$$\boldsymbol{P}(S = 5 \mid X_1 = 1)$$

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

- Recall that $S = 5$ and $X_1 = 1$ are just subsets of the sample space Ω.

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

CC BY NC SA

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: $\boldsymbol{P}(S = 5, X_1 = 1) = \boldsymbol{P}(S = 5 \mid X_1 = 1) \cdot \boldsymbol{P}(X_1 = 1)$

- $\boldsymbol{P}(S = 5, X_1 = 1) = p^5$
- $\boldsymbol{P}(S = 5 \mid X_1 = 1) = p^4$
- $\boldsymbol{P}(X_1 = 1) = p$

CC BY NC SA

Mutual Independence

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

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

You can generalize this definition to any number of random variables, though we will not do so.

CC BY NC SA

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.

CC BY NC SA

Example with i.i.d. random variables

Suppose I throw 3 balls into n bins, where each ball is equally likely to land in any of the bins, regardless of how the other balls landed. Let X_i be the bin that ball i landed in.

The X_i are i.i.d. - they're independent since each ball will land in a bin regardless of how the other balls land, and they have the same distribution: possible values are 1 through n, each with probability 1/n.

To find the probability that $X_i = 2$ for all i (every ball lands in bin 2), we can utilize the fact that the X_i are i.i.d.

$$P[X_1=2, X_2=2, X_3=2] = P[X_1=2]\,P[X_2=2]\,P[X_3=2] \quad \text{(Mutual Independence)}$$

$$= (1/n)\,(1/n)\,(1/n) \quad \text{(Identically distributed)}$$

$$= 1/n^3$$

CC BY NC SA

The Sum of Two Poisson RVs

Lecture 20, CS70 Fall 2026

Poisson Random Variables

- Derivation of Poisson RV

Jointly Distributed RVs

- Joint Distributions and Marginal Distributions
- Independent RVs, Conditional Probability and RVs
- **The Sum of Two Poisson RVs**

Linearity of Expectation

CC BY NC SA

The Poisson Random Variable

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

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

What if we add two of them together?

CC BY NC SA

The Sum of Independent Poisson RVs

Earlier today, we mentioned that if we have independent RVs $X \sim \text{Poisson}(\lambda_1)$ 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, let me remind you of 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: $(a+b)^4 = \binom{4}{0}a^4 + \binom{4}{1}a^3b + \binom{4}{2}a^2b^2 + \binom{4}{3}ab^3 + \binom{4}{4}b^4$

CC BY NC SA

The Sum of Independent Poisson RVs

Earlier today, we mentioned that if we have independent RVs $X \sim \text{Poisson}(\lambda)$ 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:

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

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

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

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}$$

CC BY NC SA

Linearity of Expectation

Lecture 20, CS70 Fall 2026

Poisson Random Variables

- Derivation of Poisson RV

Jointly Distributed RVs

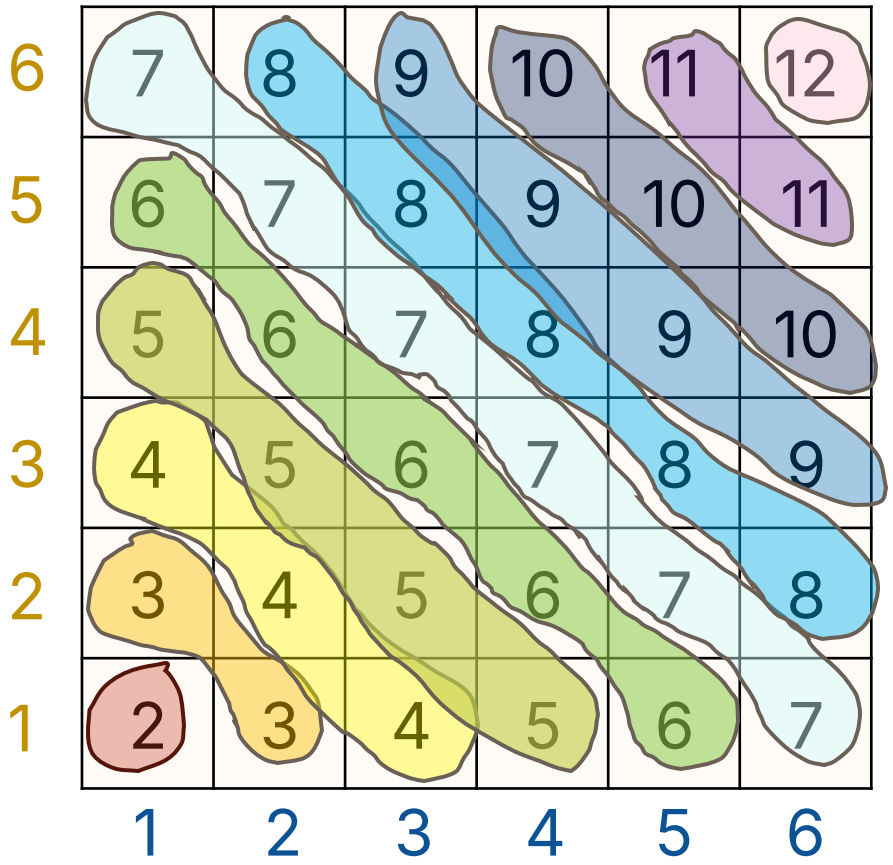
- Joint Distributions and Marginal Distributions
- Independent RVs, Conditional Probability and RVs
- The Sum of Two Poisson RVs

Linearity of Expectation

CC BY NC SA

Expectation of Two Dice

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

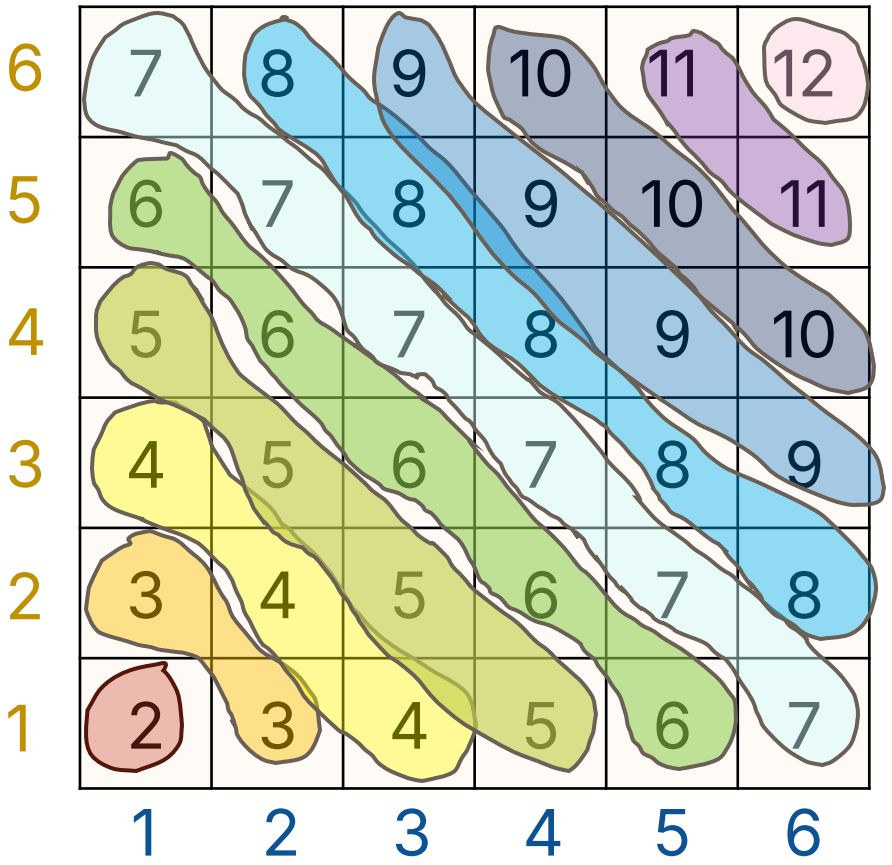


a	$P(S = a)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Expectation of Two Dice

Let $S = X + Y$ be the result after rolling two six sided 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, 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.

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

CC BY NC SA

Prelude: The Mega-Verbose Form for Expectation

Before we write our proof for $E[X+Y] = E[X] + E[Y]$, let's consider an alternate form for the definition of expectation:

$$E[X] = \sum_{a \in \text{range}(X)} a \times P(X = a)$$

We know that:

- $P(X = a)$ is the sum of $P(\omega)$ for all $\omega \in X = a$, where $X = a$ is an event, namely the set of all samples which X maps to a.
- Since X is a function, each outcome belongs to exactly one event $X = a$, i.e. the random variable partitions the sample space.

Thus, we can rewrite the formula for expectation as the mega-verbose:

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

CC BY NC SA

Prelude: The Mega-Verbose Form for Expectation

This mega-verbose form for expectation is annoying to use on real world examples.

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

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 convenient.

CC BY NC SA

Proof of Linearity of Expectation

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

Using the mega-verbose form, if $S = X + Y$, we can write $E[S]$ as:

CC BY NC SA

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:

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

Important: Expectation is linear even if X and Y are not independent!

CC BY NC SA

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. Anyone see a better way? What if we could have a RV for each trial?

CC BY NC SA

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:

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

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 $\text{B}_\text{i} \sim \text{Bernoulli}(p)$.

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

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

CC BY NC SA

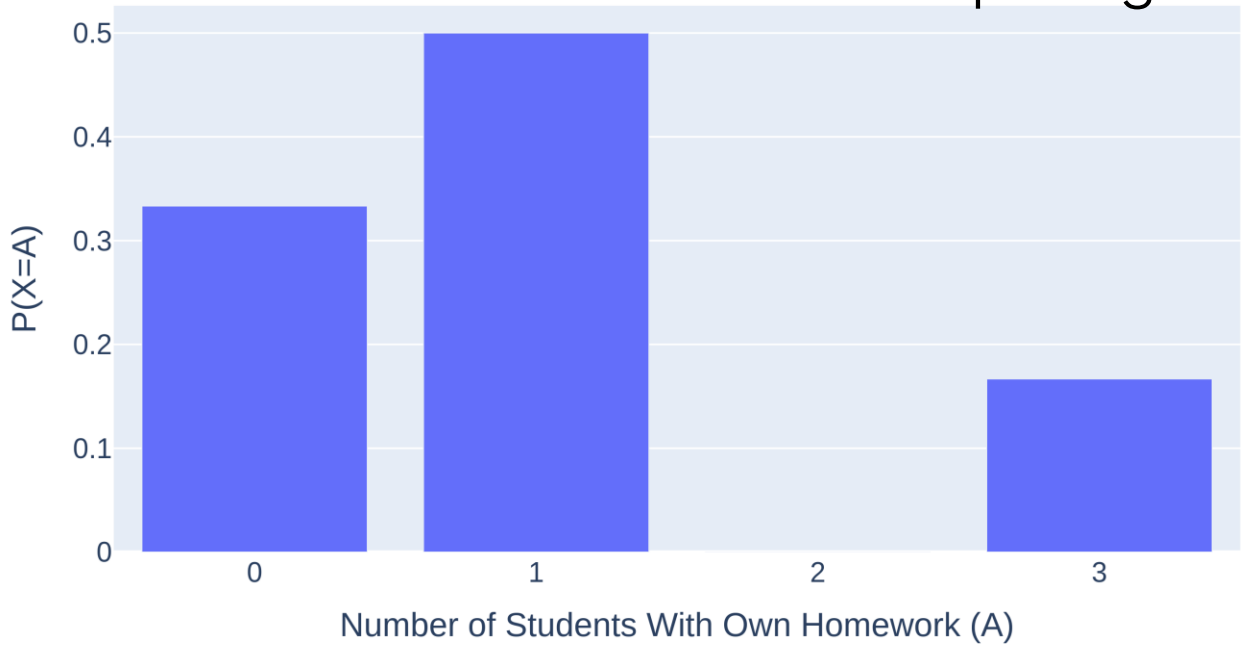
Example 3: Fixed Points

In the previous lecture, we showed the distribution for shuffling 3 students' homework and handing it back to them randomly 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.

- We will do this without ever computing the probabilities for the distribution.



Better than the naïve way where we'd need to compute $\mathrm{P}(\mathrm{X} = 0)$, $\mathrm{P}(\mathrm{X} = 1)$, …, $\mathrm{P}(\mathrm{X} = 100)$, …, $\mathrm{P}(\mathrm{X} = n)$!

Example 3: Fixed Points

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

Clever idea: Let I_i be a random variable which is 1 if student i 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 0 otherwise.

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

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

CC BY NC SA

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 student i 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 $E[I_i]$?

https://joshh.ug/E

CC BY NC SA

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 student i 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 $\mathrm{I_i} \sim \mathbf{Bernoulli}(1/n)$. So $E[I_i] = 1/n$
- Since $E[X_n] = E[I_1] + E[I_2] + \cdots + E[I_n]$, we have $E[X_n] = n \times 1/n = 1$

CC BY NC SA

Summary

Today was a bit of a grab bag:

- Poisson RVs
  - Showed how they arise from Binomials (including their expectation).
  - Showed how the sum of two Poisson RVs is also Poisson.
- Jointly distributed random variables, incl. independent random variables.
- Linearity of Expectation
  - Made Binomial distribution expectation a lot easier.
  - Made Fixed Point distribution expectation a lot lot easier.

CC BY NC SA