

Zoo of random variables

CSE 312 Spring 26

Lecture 12

Variance - summary

Definition. The **variance** of a (discrete) RV X is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \Omega_X} p_X(x) \cdot (x - \mathbb{E}[X])^2$$

Standard deviation: $\sigma(X) = \sqrt{\text{Var}(X)}$

Recall $\mathbb{E}[X]$ is a **constant**, not a random variable itself.

Intuition: Variance (or standard deviation) is a quantity that measures, in expectation, how “far” the random variable is from its expectation.

Variance – Properties (3)

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x) \cdot P(X = x)$$

Definition. The **variance** of a (discrete) RV X is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \Omega_X} p_X(x) \cdot (x - \mathbb{E}[X])^2$$

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

Theorem. For any $a, b \in \mathbb{R}$, $\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X)$

Multiple Random Variables and Independence

X, Y indep \implies so are $f(X), g(Y)$

Comma is shorthand for AND

Definition. Two random variables X, Y are **(mutually) independent** if for all x, y ,

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

Intuition: Knowing X doesn't help you guess Y and vice versa

Definition. The random variables X_1, \dots, X_n are **(mutually) independent** if for all x_1, \dots, x_n ,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

Note: No need to check for all subsets, but need to check for all outcomes!

Important Facts about Independent Random Variables

Theorem. If X, Y independent, $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Theorem. If X, Y independent, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Corollary. If X_1, X_2, \dots, X_n mutually independent,

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_i \text{Var}(X_i)$$

Agenda

- Recap
- Zoo of Discrete RVs, Part I
 - Uniform Random Variables
 - Bernoulli Random Variables
 - Binomial Random Variables
 - Geometric Random Variables

Motivation for “Named” Random Variables

Random Variables that show up all over the place.

- Easily solve a problem by recognizing it’s a special case of one of these random variables.

Each RV introduced today will show:

- A general situation it models
- Its name and parameters
- Its PMF, Expectation, and Variance
- Example scenarios you can use it

Welcome to the Zoo! (Preview)



$X \sim \text{Unif}(a, b)$

$$P(X = k) = \frac{1}{b - a + 1}$$

$$\mathbb{E}[X] = \frac{a + b}{2}$$

$$\text{Var}(X) = \frac{(b - a)(b - a + 2)}{12}$$

$X \sim \text{Ber}(p)$

$$P(X = 1) = p, P(X = 0) = 1 - p$$

$$\mathbb{E}[X] = p$$

$$\text{Var}(X) = p(1 - p)$$

$X \sim \text{Bin}(n, p)$

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

$$\mathbb{E}[X] = np$$

$$\text{Var}(X) = np(1 - p)$$

$X \sim \text{Geo}(p)$

$$P(X = k) = (1 - p)^{k-1} p$$

$$\mathbb{E}[X] = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

$X \sim \text{NegBin}(r, p)$

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$$

$$\mathbb{E}[X] = \frac{r}{p}$$

$$\text{Var}(X) = \frac{r(1 - p)}{p^2}$$

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

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

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

$$\text{Var}(X) = n \frac{K(N-K)(N-n)}{N^2(N-1)}$$

Zoo of Discrete RVs, Part I

- Uniform Random Variables
- Bernoulli Random Variables
- Binomial Random Variables
- Geometric Random Variables

Discrete Uniform Random Variables

$$\Omega_X = \{a, a+1, \dots, b\}$$

A discrete random variable X **equally likely** to take any (integer) value between integers a and b (inclusive), is **uniform**.

Notation: $\text{Unif}(a, b)$

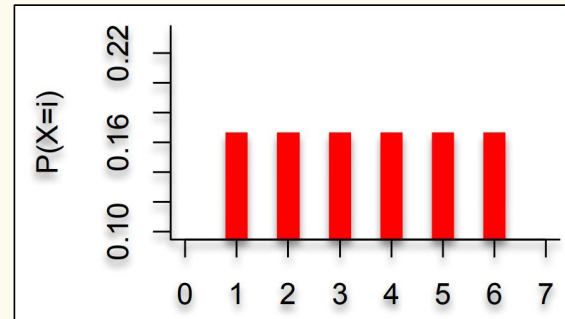
PMF:
$$P(X=i) = \begin{cases} \frac{1}{b-a+1} & i \in \{a, \dots, b\} \\ 0 & \text{o.w.} \end{cases}$$

Expectation: $\frac{a+b}{2}$

Variance:

Example: value shown on one roll of a fair die is $\text{Unif}(1,6)$:

- $P(X = i) = 1/6$
- $\mathbb{E}[X] = 7/2$
- $\text{Var}(X) = 35/12$



Discrete Uniform Random Variables (2)

A discrete random variable X **equally likely** to take any (integer) value between integers a and b (inclusive), is **uniform**.

Notation: $X \sim \text{Unif}(a, b)$

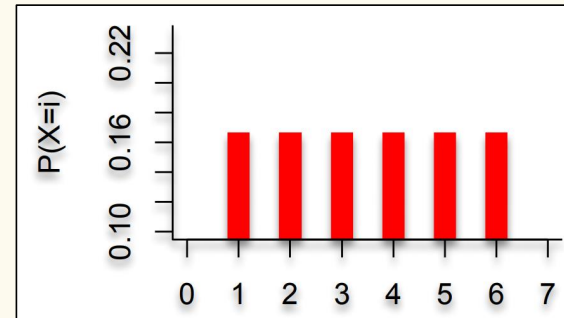
PMF: $P(X = i) = \frac{1}{b - a + 1}$

Expectation: $\mathbb{E}[X] = \frac{a+b}{2}$

Variance: $\text{Var}(X) = \frac{(b-a)(b-a+1)}{12}$

Example: value shown on one roll of a fair die is $\text{Unif}(1,6)$:

- $P(X = i) = 1/6$
- $\mathbb{E}[X] = 7/2$
- $\text{Var}(X) = 35/12$



Zoo (2)

- Zoo of Discrete RVs
 - Uniform Random Variables
 - Bernoulli Random Variables
 - Binomial Random Variables
 - Geometric Random Variables

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

Bernoulli Random Variables (aka Indicator r.v.s)

A random variable X that takes value 1 (“Success”) with probability p , and 0 (“Failure”) otherwise. X is called a **Bernoulli random variable**.

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

PMF: $P(X = 1) = p, P(X = 0) = 1 - p$

Expectation:

Variance:

Poll:

<https://pollev.com/annakarlin185>

	Mean	Variance
A.	p	p
B.	p	$1 - p$
C.	p	$p(1 - p)$
D.	p	p^2

Bernoulli Random Variables (2)

A random variable X that takes value 1 (“Success”) with probability p , and 0 (“Failure”) otherwise. X is called a **Bernoulli random variable**.

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

PMF: $P(X = 1) = p, P(X = 0) = 1 - p$

Expectation: $\mathbb{E}[X] = p$ Note: $\mathbb{E}[X^2] = p$

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

Examples:

- Coin flip
- Randomly guessing on a MC test question
- A server in a cluster fails
- Whether or not a share of a particular stock pays off or not
- Any indicator r.v.

Zoo (3)

- Zoo of Discrete RVs, Part I
 - Uniform Random Variables
 - Bernoulli Random Variables
 - Binomial Random Variables
 - Geometric Random Variables

Binomial Random Variables

$$p \cdot p \cdot \dots \cdot p_n$$

A discrete random variable $X = \sum_{i=1}^n Y_i$ where each $Y_i \sim \text{Ber}(p)$.

Counts number of successes in n independent trials, each with probability p of success.

X is a **Binomial random variable**

$$n, p$$

$$\mathcal{L}_X = \{0, 1, 2, \dots, n\}$$

Examples:

- # of heads in n indep coin flips
- # of 1s in a randomly generated n bit string
- # of servers that fail in a cluster of n computers
- # of bit errors in file written to disk
- # of elements in a bucket of a large hash table
- # of n different stocks that “pay off”

Poll:

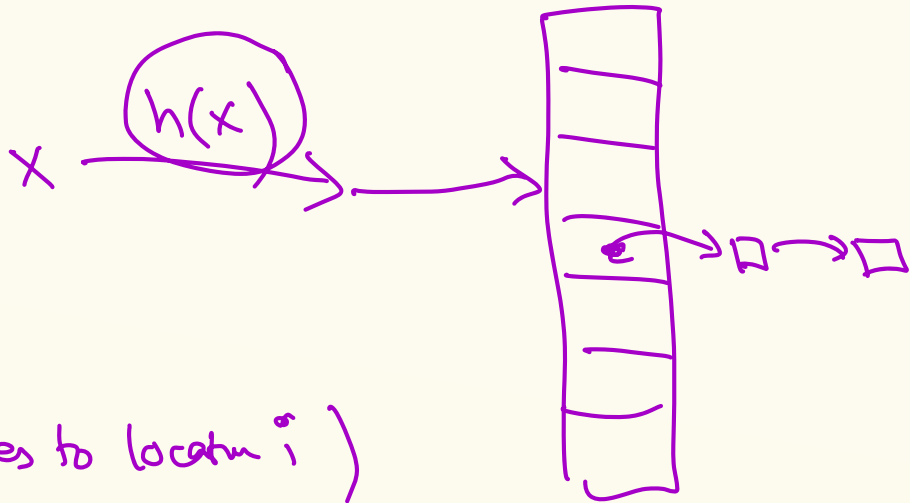
<https://pollev.com/annakarlin185>

$$P(X = k) =$$

- A. $p^k(1-p)^{n-k}$
- B. np
- C. $\binom{n}{k} p^k (1-p)^{n-k}$
- D. $\binom{n}{n-k} p^k (1-p)^{n-k}$

hashtable - size m

hashing n elts into table



$$P(x \text{ hashes to locatn } i) = \frac{1}{m}$$

X to be $\#$ of elts that hash to locatn i

$$= Y_1 + Y_2 + \dots + Y_n$$

$$\binom{n}{k} = \binom{n}{n-k}$$

hash fn distributes elts randomly

hash fn maps each elt to uniformly random locatn in table

diff elts hashed independently

$$X \sim \text{Bin}(n, \frac{1}{m})$$

$$Y_j = \begin{cases} 1 & \text{elt } j \text{ hashes to locatn } i \\ 0 & \text{ow.} \end{cases}$$

Binomial Random Variables (2)

A discrete random variable $X = \sum_{i=1}^n Y_i$ where each $Y_i \sim \text{Ber}(p)$.

Counts number of successes in n independent trials, each with probability p of success.

X is a **Binomial random variable**

Notation: $X \sim \text{Bin}(n, p)$

PMF: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

Expectation:

Variance:

Poll:

<https://pollev.com/annakarlin185>

	Mean	Variance
A.	p	p
B.	np	$np(1 - p)$
C.	np	np^2
D.	np	n^2p

Binomial Random Variables (3)

A discrete random variable $X = \sum_{i=1}^n Y_i$ where each $Y_i \sim \text{Ber}(p)$.

Counts number of successes in n independent trials, each with probability p of success.

X is a **Binomial random variable**

Notation: $X \sim \text{Bin}(n, p)$

PMF: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

Expectation: $\mathbb{E}[X] = np$

Variance: $\text{Var}(X) = np(1 - p)$

Mean, Variance of the Binomial

“i.i.d.” is a commonly used phrase.

It means “independent & identically distributed”

If $Y_1, Y_2, \dots, Y_n \sim \text{Ber}(p)$ and independent (i.i.d.), then

$$X = \sum_{i=1}^n Y_i, \quad X \sim \text{Bin}(n, p)$$

Claim $\mathbb{E}[X] = np$

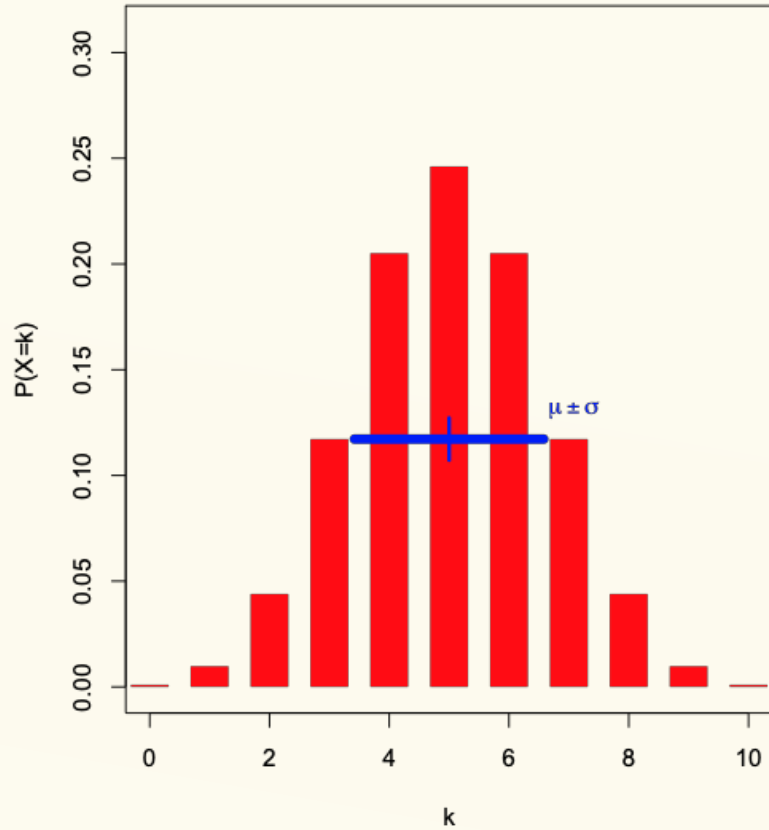
$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n Y_i\right] \stackrel{\text{LOF}}{=} \sum_{i=1}^n \mathbb{E}[Y_i] = n\mathbb{E}[Y_1] = np$$

Claim $\text{Var}(X) = np(1 - p)$, by independence of r.v.'s

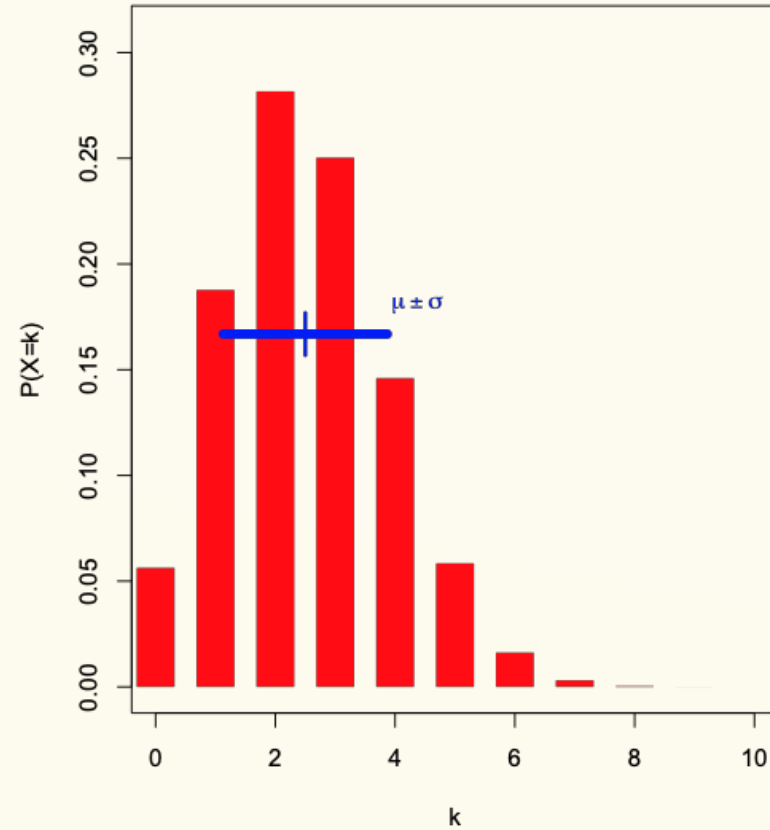
$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \text{Var}(Y_i) = n\text{Var}(Y_1) = np(1 - p)$$

Binomial PMFs

PMF for $X \sim \text{Bin}(10, 0.5)$

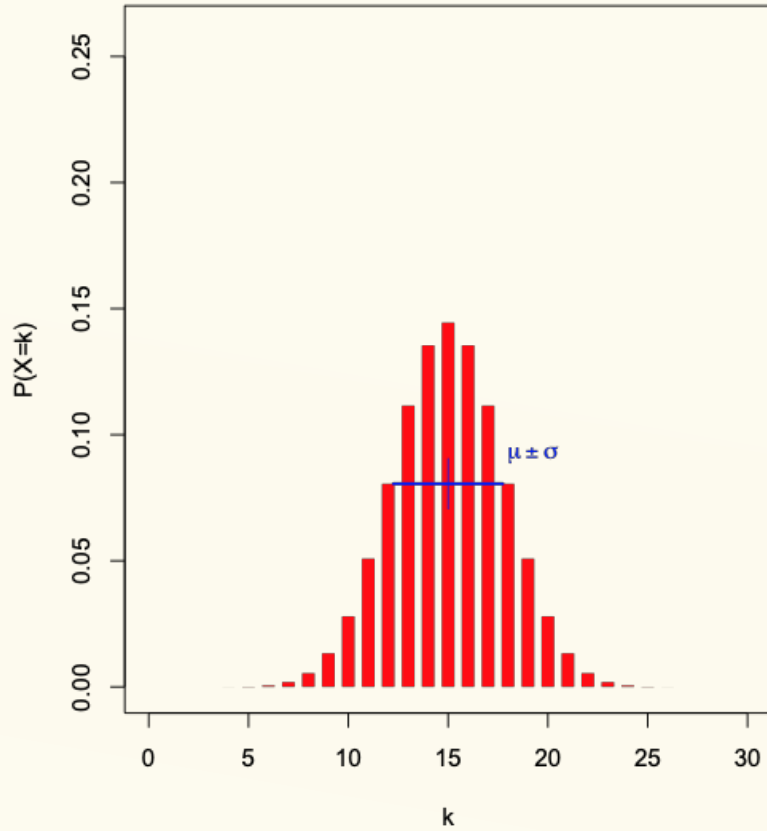


PMF for $X \sim \text{Bin}(10, 0.25)$

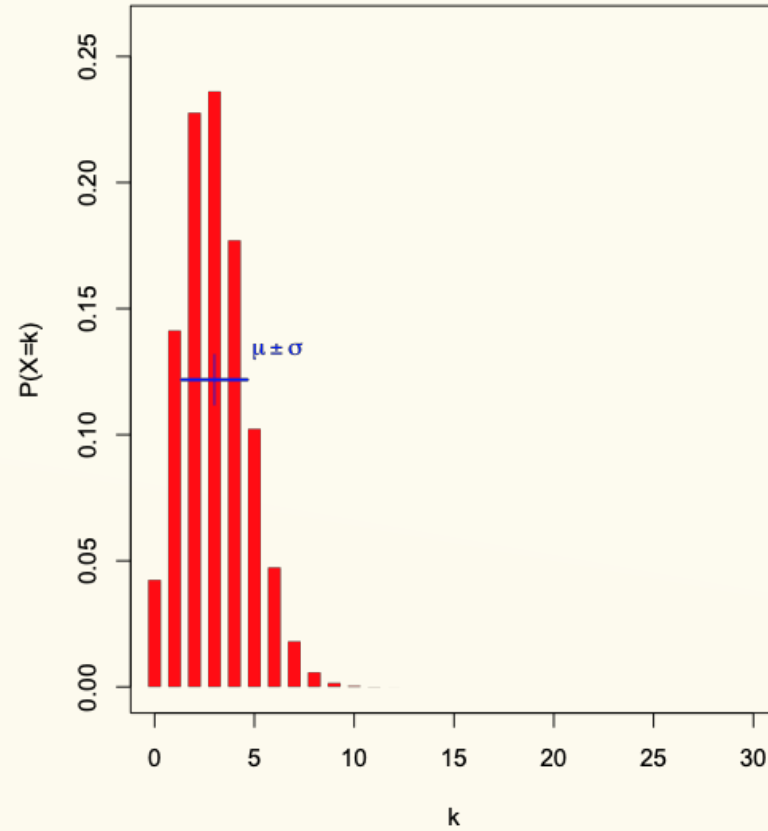


Binomial PMFs (2)

PMF for $X \sim \text{Bin}(30, 0.5)$



PMF for $X \sim \text{Bin}(30, 0.1)$



Zoo (4)

- Zoo of Discrete RVs, Part I
 - Uniform Random Variables
 - Bernoulli Random Variables
 - Binomial Random Variables
 - Geometric Random Variables

Geometric Random Variables

$$\Omega_X = \{1, 2, 3, \dots\}$$

A discrete random variable X that models the number of independent trials $Y_i \sim \text{Ber}(p)$ up until and including the first success.

X is called a **Geometric random variable** with parameter p .

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

PMF: $P(X = k) = (1-p)^{k-1} p$ $k \in \Omega_X$

Expectation: $E(X) = \frac{1}{p}$

Variance:

Examples:

- # of coin flips until first head
- # of random guesses on MC questions until you get one right
- # of random guesses at a password until you hit it

$$E(X) = \sum_{k=1}^{\infty} k P(X=k) = \sum_{k=1}^{\infty} k (1-p)^{k-1} p$$

$$\frac{d}{dx} \left(\sum_{k=0}^{\infty} x^k \right) = \frac{d}{dx} \left(\frac{1}{1-x} \right)$$

$$0 < x < 1$$

$$\sum_{k=0}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$$

Geometric Random Variables (2)

A discrete random variable X that models the number of independent trials $Y_i \sim \text{Ber}(p)$ up until and including the first success.

X is called a **Geometric random variable** with parameter p .

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

PMF: $P(X = k) = (1 - p)^{k-1}p$

Expectation: $\mathbb{E}[X] = \frac{1}{p}$

Variance: $\text{Var}(X) = \frac{1-p}{p^2}$

Examples:

- # of coin flips until first head
- # of random guesses on MC questions until you get one right
- # of random guesses at a password until you hit it
- # hash trials until a miner successfully mines a Bitcoin

Zoo (5)

- Zoo of Discrete RVs, Part I
 - Uniform Random Variables
 - Bernoulli Random Variables
 - Binomial Random Variables
 - Geometric Random Variables
 - More examples

Example 1

Sending a binary message of length **1024** bits over a network with probability 0.999 of correctly sending each bit in the message without corruption (independent of other bits).

Let X be the number of corrupted bits.

What kind of random variable is this and what is $\mathbb{E}[X]$?

$$X \sim \text{Bin}(1024, 0.001)$$

$$\mathbb{E}(X) = np = 1024 \cdot 0.001$$

Poll:

<https://pollev.com/annakarlin185>

A 1022.99

B 1.024

C 1.02298

D. 1

Example

Sending a binary message of length 1024 bits over a network with probability 0.999 of correctly sending each bit in the message without corruption (independent of other bits).

Let X be the number of corrupted bits.

What kind of random variable is this and what is $\mathbb{E}[X]$?

Binomial (1024, 0.001)

Therefore $\mathbb{E}[X] = np = 1024 \cdot 0.001 = 1.024$

Example: Music Lessons

Your music teacher requires you to play a 1000 note song without mistake. You have been practicing, so you have a probability of 0.999 of getting each note correct (independent of the others). If you mess up a single note in the song, you must start over and play from the beginning. Let X be the number of times you have to play the song from the start. What kind of random variable is this and what is $E[X]$?

$$X \sim \text{Geo} \left(0.999^{1000} \right)$$

$$E(X) = \frac{1}{0.999^{1000}}$$

Example: Music Lessons - solution

Your music teacher requires you to play a 1000 note song without mistake. You have been practicing, so you have a probability of 0.999 of getting each note correct (independent of the others). If you mess up a single note in the song, you must start over and play from the beginning. Let X be the number of times you have to play the song from the start. What kind of random variable is this and what is $E[X]$?

Probability that you play whole song without a mistake is 0.999^{1000}

Therefore X is a Geometric random variable with parameter $p = 0.999^{1000}$

So its expectation is $\frac{1}{0.999^{1000}}$

Zoo (6)

- Zoo of Discrete RVs
 - Uniform Random Variables
 - Bernoulli Random Variables
 - Binomial Random Variables
 - Geometric Random variables
 - Poisson Distribution

Preview: Poisson

Model: X is # events that occur in an hour

- Events occur at a fixed rate (3 per hour), but at random times.
- The expected number of events in t hours, is $3t$
- Occurrence of events on disjoint time intervals is independent

Example – Modelling car arrivals at an intersection

X = # of cars passing through a light in 1 hour

How to model?

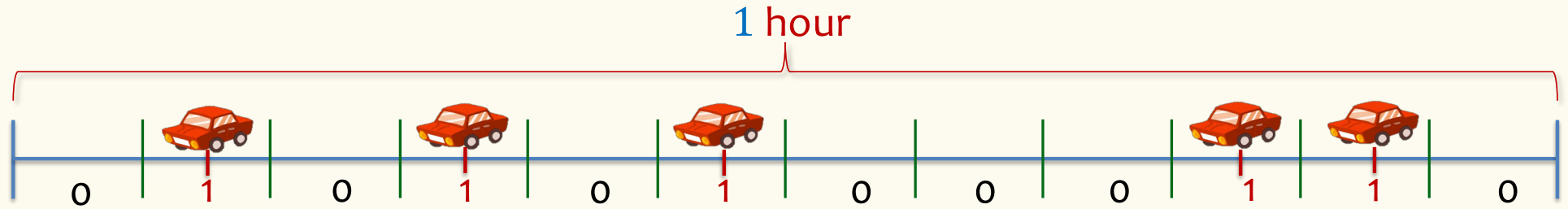
$$np = 3$$

X = # cars passing through a light in 1 hour.

Constant rate of arrival.

Disjoint time intervals are independent.

Know: $\mathbb{E}[X] = 3$



$\frac{1}{n}$ of an hour

p - Prob car arrives in one of these intervals

$$X \sim \text{Bin}(n, p)$$

$$\text{Bin}\left(n, \frac{3}{n}\right)$$

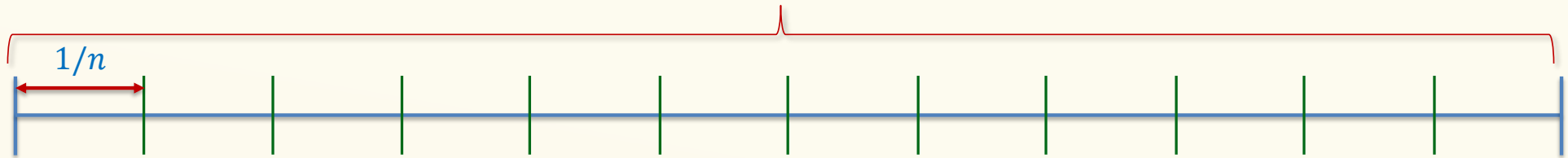
$$p = \frac{3}{n}$$

Example – Model the process of cars passing through a light in 1 hour

X = # cars passing through a light in 1 hour. $\mathbb{E}[X] = 3$

Assume: Occurrence of events on disjoint time intervals is independent

Approximation idea: Divide hour into n intervals of length $1/n$



This gives us n independent intervals

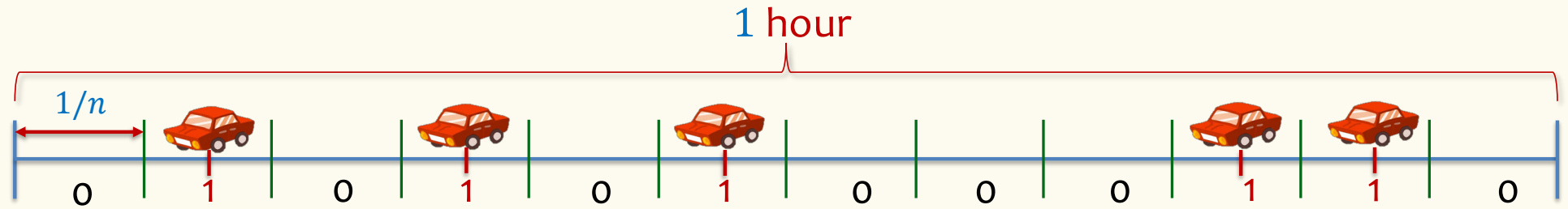
Assume either zero or one car per interval

p = probability car arrives in a single interval of length $1/n$

What should p be?

$X = \#$ cars passing through a light in 1 hour. Disjoint time intervals are independent.

Know: $\mathbb{E}[X] = 3$



This gives us n independent intervals

Assume either zero or one car per interval

$p =$ probability car arrives in an interval

Model as $\text{Bin}(n, p)$

What should p be?

<https://pollev.com/annakarlin185>

A. $3/n$

B. $3n$

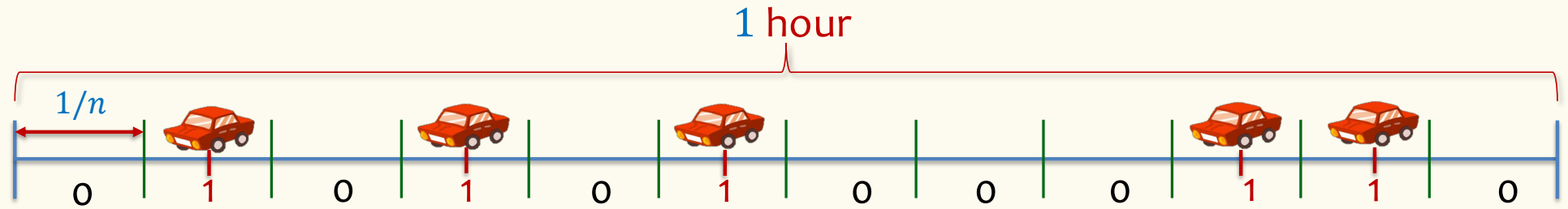
C. 3

D. $3/60$

Model as Binomial

X = # cars passing through a light in 1 hour. Disjoint time intervals are independent.

Know: $\mathbb{E}[X] = \lambda$ for some given $\lambda > 0$



Discrete version: n intervals, each of length $1/n$.

In each interval, there is a car with probability $p = \lambda/n$ (assume ≤ 1 car can pass by)

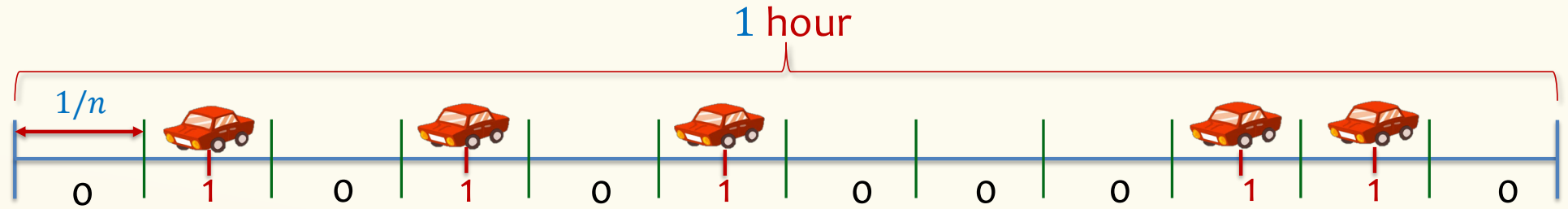
Each interval is Bernoulli: $X_i = 1$ if car in i^{th} interval (0 otherwise). $P(X_i = 1) = \lambda/n$

$$X = \sum_{i=1}^n X_i$$

Model as Binomial

X = # cars passing through a light in 1 hour. Disjoint time intervals are independent.

Know: $\mathbb{E}[X] = \lambda$ for some given $\lambda > 0$



Discrete version: n intervals, each of length $1/n$.

In each interval, there is a car with probability $p = \lambda/n$ (assume ≤ 1 car can pass by)

Each interval is Bernoulli: $X_i = 1$ if car in i^{th} interval (0 otherwise). $P(X_i = 1) = \lambda/n$

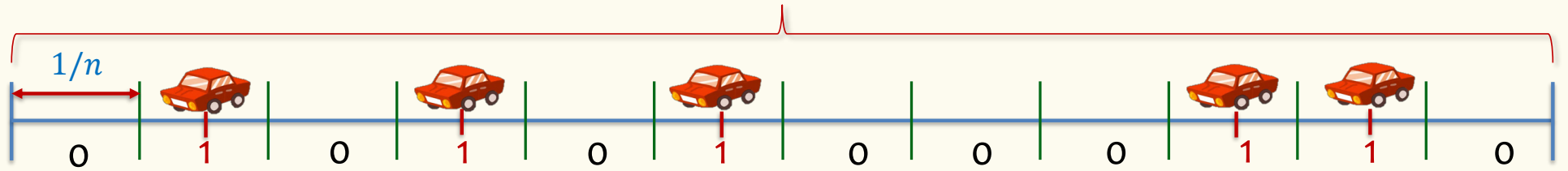
$$X = \sum_{i=1}^n X_i \quad X \sim \text{Bin}(n, p)$$

$$P(X = i) = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

indeed! $\mathbb{E}[X] = pn = \lambda$

Don't like discretization

$$X \text{ is binomial } P(X = i) = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$



We want now $n \rightarrow \infty$

$$P(X = i) = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

