

CSE 312

Foundations of Computing II

Lecture 12: Zoo of Discrete RVs



Rachel Lin, Hunter Schafer

Slide Credit: Based on Stefano Tessaro's slides for 312 19au incorporating ideas from Alex Tsun's and Anna Karlin's slides for 312 20su and 20au

Variance – Properties

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

$$\text{Var}(X) = \mathbb{E} \left[(X - \mathbb{E}(X))^2 \right] = \sum_x \mathbb{P}_X(x) \cdot (x - \mathbb{E}(X))^2$$

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

(Proof: Exercise!)

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

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,

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

Motivation: “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)



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

$$P(X = k) = \frac{1}{b - a + 1}$$
$$E[X] = \frac{a + b}{2}$$
$$\text{Var}(X) = \frac{(b - a)(b - a + 2)}{12}$$

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

$$P(X = 1) = p, P(X = 0) = 1 - p$$
$$E[X] = p$$
$$\text{Var}(X) = p(1 - p)$$

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

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$
$$E[X] = np$$
$$\text{Var}(X) = np(1 - p)$$

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

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

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

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$$
$$E[X] = \frac{r}{p}$$
$$\text{Var}(X) = \frac{r(1 - p)}{p^2}$$

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

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$
$$E[X] = n \frac{K}{N}$$
$$\text{Var}(X) = n \frac{K(N-K)(N-n)}{N^2(N-1)}$$

Agenda

- Discrete Uniform Random Variables ◀
- Bernoulli Random Variables
- Binomial Random Variables
- Geometric Random Variables
- Applications

Discrete Uniform Random Variables

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

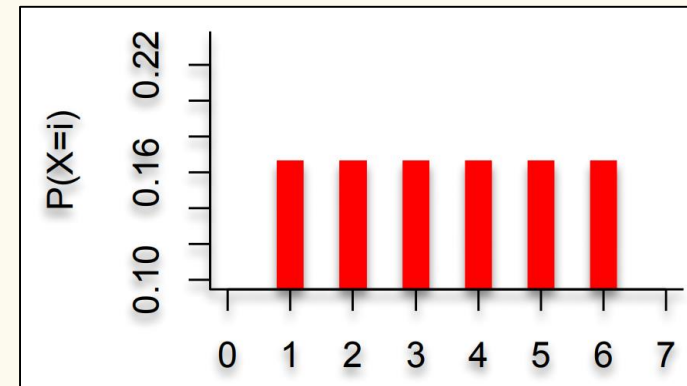
Notation:

PMF:

Expectation:

Variance:

Example: value shown on one roll of a fair die



Discrete Uniform Random Variables

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

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

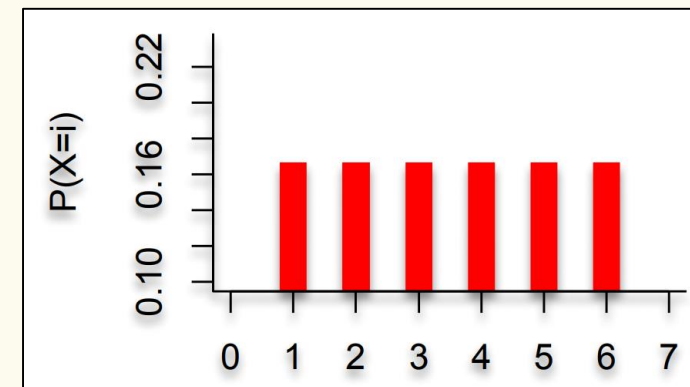
PMF: $\Pr(X = i) = \frac{1}{b - a + 1}$

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

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

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

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



Agenda

- Discrete Uniform Random Variables
- **Bernoulli Random Variables** ◀
- Binomial Random Variables
- Geometric Random Variables
- Applications

Bernoulli Random Variables

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: $\Pr(X = 1) = p, \Pr(X = 0) = 1 - p$

Expectation:

Variance:

Poll:

pollev.com/hunter312

	Mean	Variance
<i>a.</i>	p	p
<i>b.</i>	p	$1 - p$
<i>c.</i>	p	$p(1 - p)$
<i>d.</i>	p	$p(1 - p)$

Bernoulli Random Variables

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: $\Pr(X = 1) = p, \Pr(X = 0) = 1 - p$

Expectation: $E[X] = p$ Note: $E[X^2] = p$

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

Examples:

- Coin flip
- Randomly guessing on a MC test question
- A server in a cluster fails

Agenda

- Discrete Uniform Random Variables
- Bernoulli Random Variables
- **Binomial Random Variables** ◀
- Geometric Random Variables
- Applications

Binomial Random Variables

A discrete random variable X that is the number of successes in n independent random variables $Y_i \sim \text{Ber}(p)$. X is a **Binomial random variable** where $X = \sum_{i=1}^n Y_i$

Examples:

- # of heads in n 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

Poll:

pollev.com/hunter312

$\Pr(X = k)$

a. $p^k(1-p)^{n-k}$

b. np

c. $\binom{n}{k}p^k(1-p)^{n-k}$

d. $\binom{n}{n-k}p^k(1-p)^{n-k}$

Binomial Random Variables

A discrete random variable X that is the number of successes in n independent random variables $Y_i \sim \text{Ber}(p)$. X is a **Binomial random variable** where $X = \sum_{i=1}^n Y_i$

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

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

Expectation:

Variance:

Poll:

pollev.com/hunter312

	Mean	Variance
a.	p	p
b.	np	$np(1 - p)$
c.	np	np^2
d.	np	n^2p

Binomial Random Variables

A discrete random variable X that is the number of successes in n independent random variables $Y_i \sim \text{Ber}(p)$. X is a **Binomial random variable** where $X = \sum_{i=1}^n Y_i$

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

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

Expectation: $E[X] = np$

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

Mean, Variance of the Binomial

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 $E[X] = np$

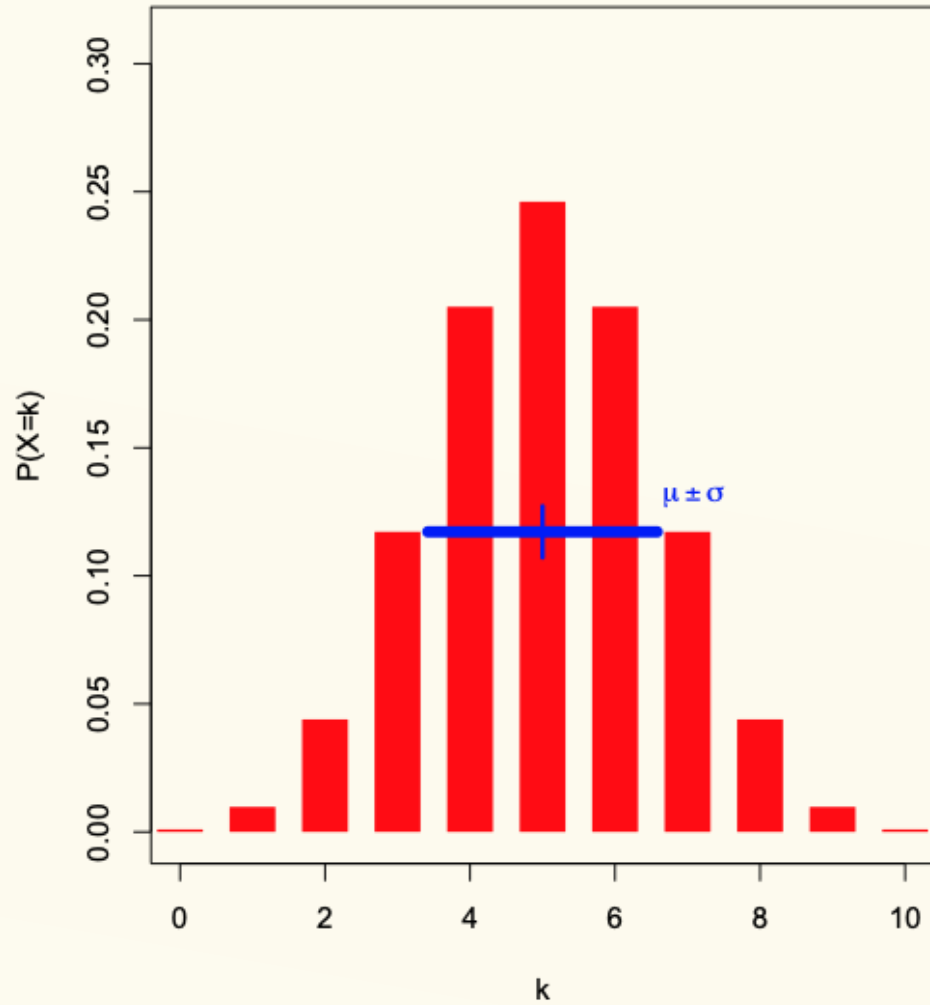
$$E[X] = E\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n E[Y_i] = nE[Y_1] = np$$

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

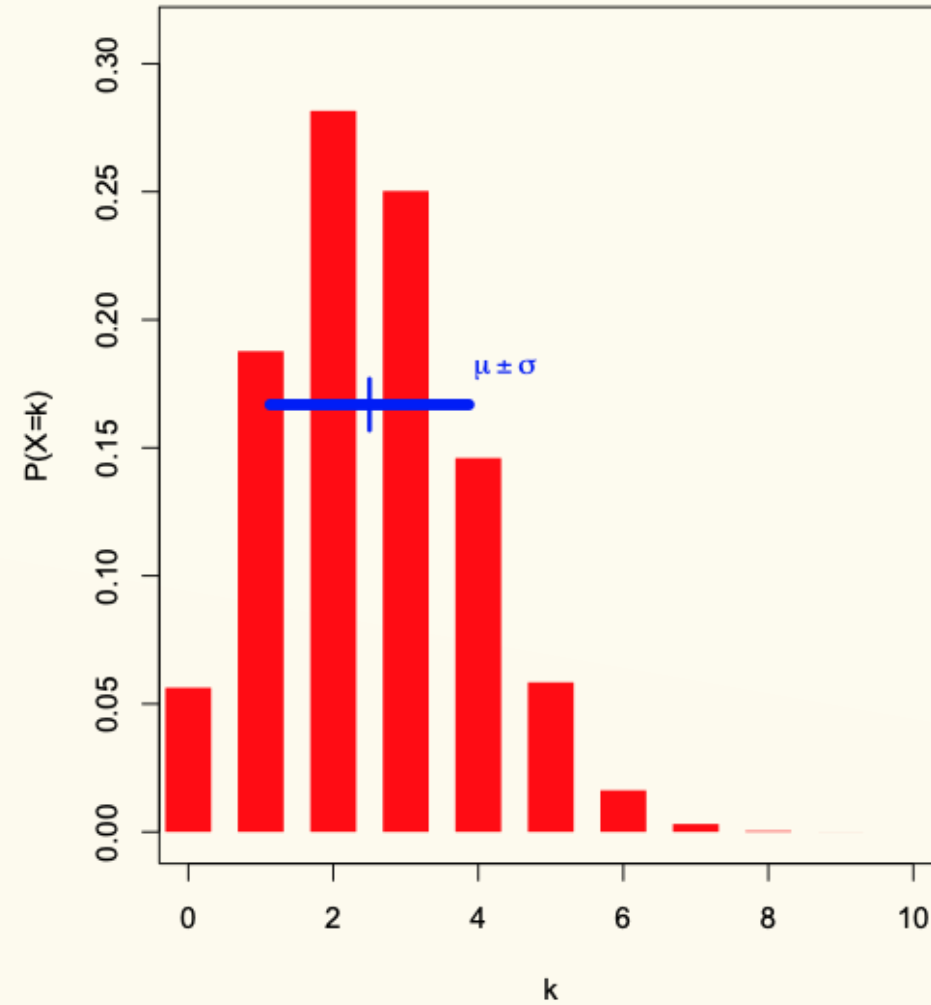
$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \text{Var}(Y_i) = n\text{Var}(Y_1) = np(1 - p)$$

Binomial PMFs

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

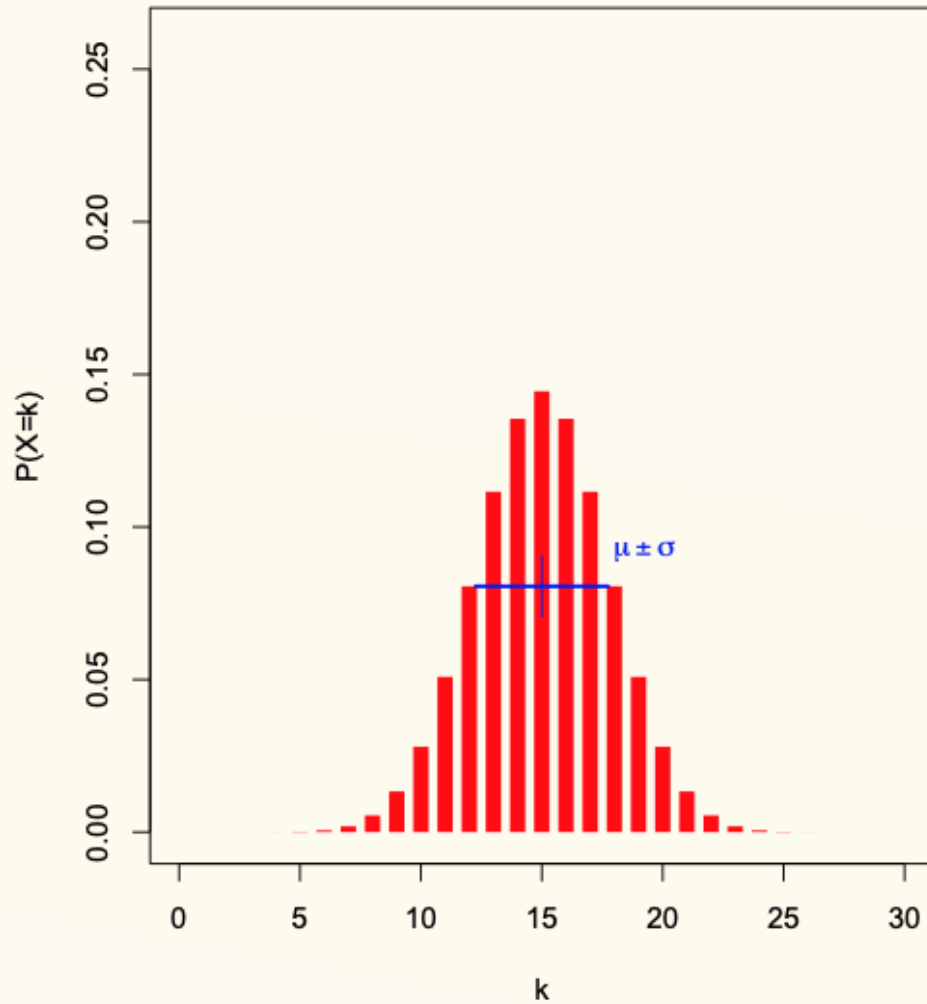


PMF for $X \sim \text{Bin}(10, 0.25)$

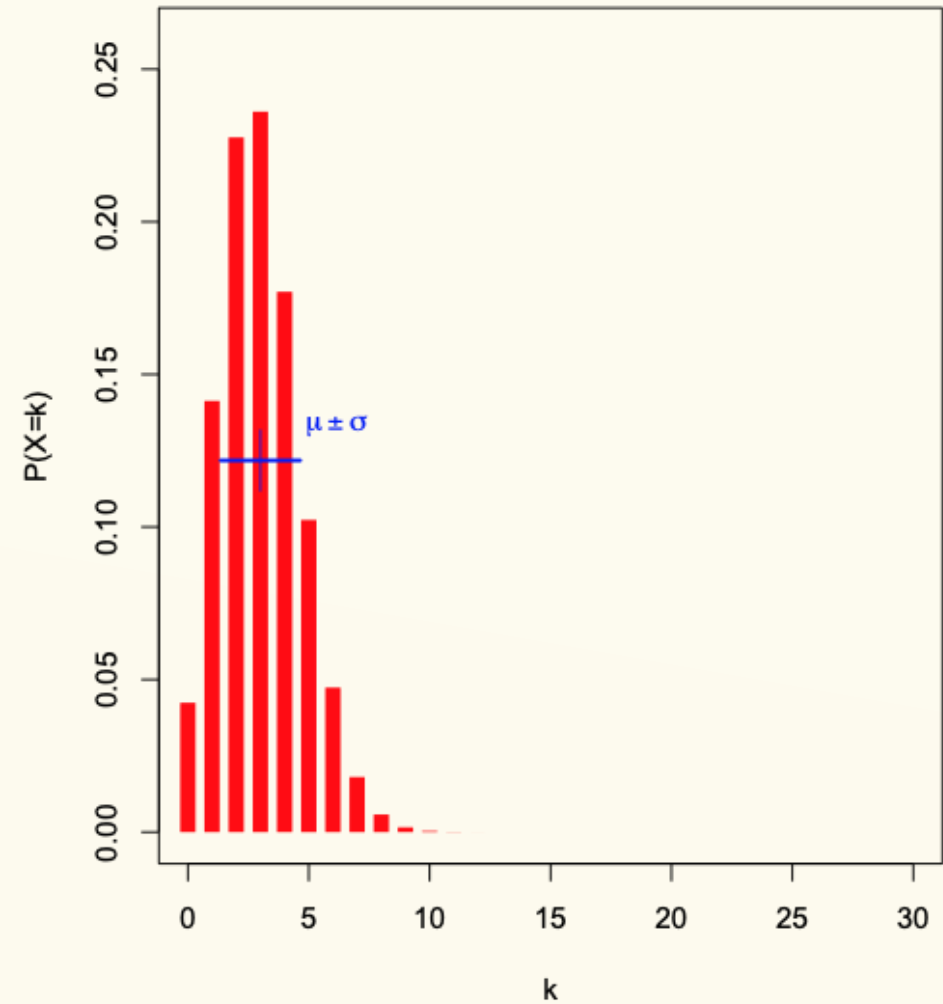


Binomial PMFs

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



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



Example

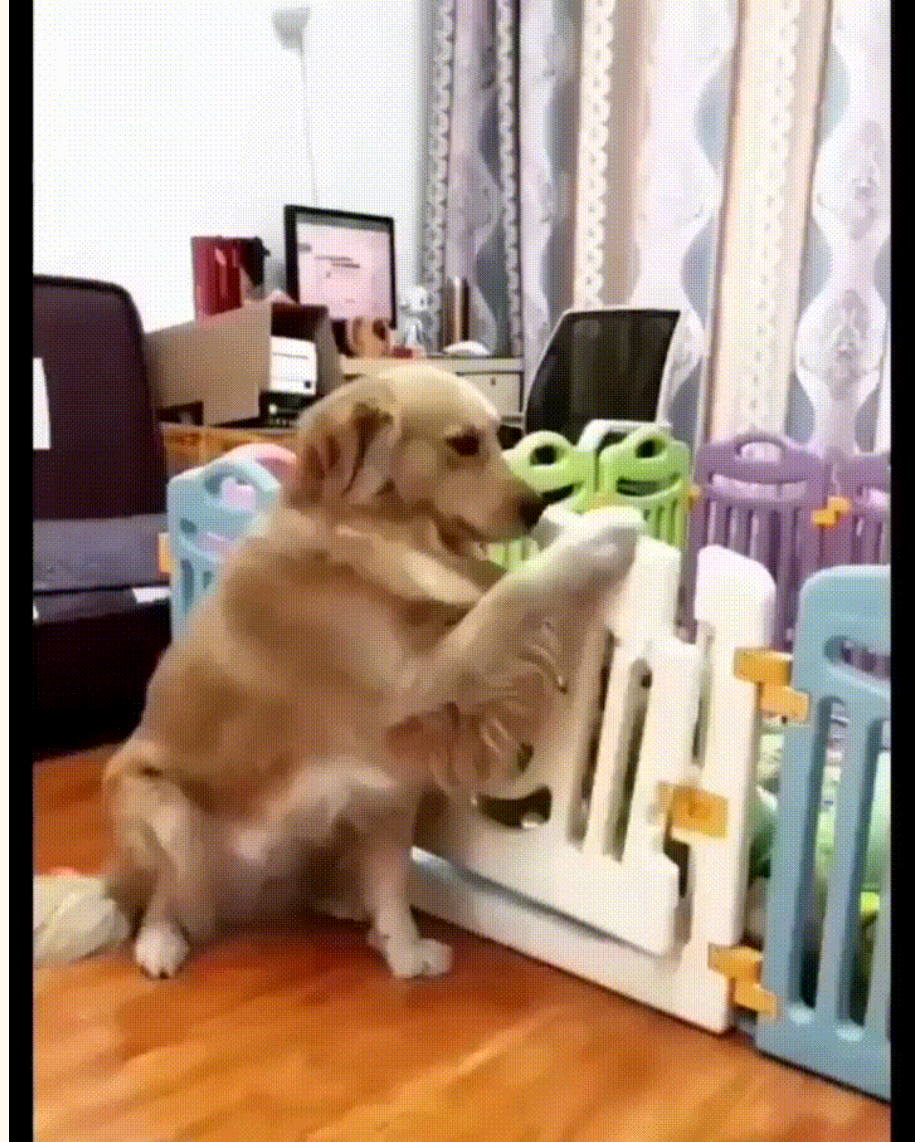
Sending a binary message of length 1024 bits over a network with probability 0.999 of correctly sending each bit in the message without corruption (independent of other bits). Let X be the number of corrupted bits. What is $E[X]$?

Poll:

pollev.com/hunter312

- a. 1022.99
- b. 1.024
- c. 1.02298
- d. 1
- e. Not enough information to compute

Brain Break



Agenda

- Discrete Uniform Random Variables
- Bernoulli Random Variables
- Binomial Random Variables
- Geometric and other Random Variables ◀

Geometric Random Variables

A discrete random variable X that models the number of independent trials $Y_i \sim \text{Ber}(p)$ before seeing the first success. X is called a **Geometric random variable** with parameter p .

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

PMF:

Expectation:

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

Geometric Random Variables

A discrete random variable X that models the number of independent trials $Y_i \sim \text{Ber}(p)$ before seeing the first success. X is called a **Geometric random variable** with parameter p .

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

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

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

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

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 is $E[X]$?

Negative Binomial Random Variables

A discrete random variable X that models the number of independent trials $Y_i \sim \text{Ber}(p)$ before seeing the r^{th} success. Equivalently, $X = \sum_{i=1}^r Z_i$ where $Z_i \sim \text{Geo}(p)$. X is called a **Negative Binomial random variable** with parameters r, p .

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

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

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

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

Hypergeometric Random Variables

A discrete random variable X that models the number of successes in n draws (without replacement) from N items that contain K successes in total. X is called a **Hypergeometric RV** with parameters N, K, n .

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

PMF: $\Pr(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$

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

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

Hope you enjoyed the zoo!

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

$$P(X = k) = \frac{1}{b - a + 1}$$
$$E[X] = \frac{a + b}{2}$$
$$\text{Var}(X) = \frac{(b - a)(b - a + 2)}{12}$$

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

$$P(X = 1) = p, P(X = 0) = 1 - p$$
$$E[X] = p$$
$$\text{Var}(X) = p(1 - p)$$

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

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$
$$E[X] = np$$
$$\text{Var}(X) = np(1 - p)$$

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

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

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

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$$
$$E[X] = \frac{r}{p}$$
$$\text{Var}(X) = \frac{r(1 - p)}{p^2}$$

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

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$
$$E[X] = n \frac{K}{N}$$
$$\text{Var}(X) = n \frac{K(N-K)(N-n)}{N^2(N-1)}$$

Preview: Poisson

Model: # events that occur in an hour

- Expect to see 3 events per hour (but will be random)
- The expected number of events in t hours, is $3t$
- Occurrence of events on disjoint time intervals is independent