CSE 312

Foundations of Computing II

Lecture 12: Zoo of Discrete RVs

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

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

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

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

(Proof: Exercise!)

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

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,

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

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)

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

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

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

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

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

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

$$E[X] = p$$

$$Var(X) = p(1 - p)$$

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

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

$$E[X] = np$$

$$Var(X) = np(1 - p)$$

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

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

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

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

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

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

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

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

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

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

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

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

Agenda

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

Discrete Uniform Random Variables

$X(\Omega) = \{a, a+1, \ldots, b-1, b\}$

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

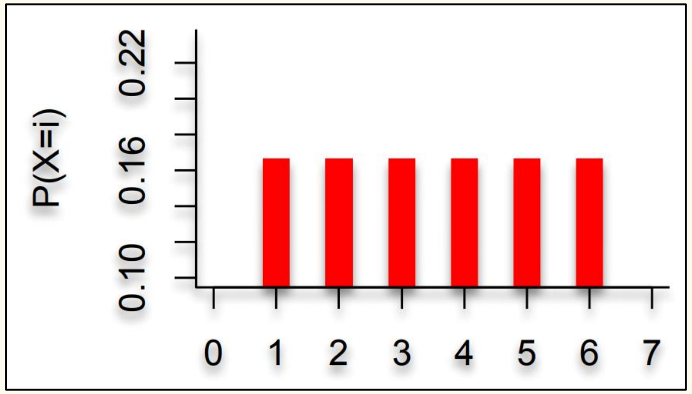
Notation: $X \sim Unif(a,b)$

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

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

Variance:

Example: value shown on one roll of a fair die $Unif(1,6)$



Discrete Uniform Random Variables

22-sided die : $X \sim Unif(1,22)$

$Pr(X=i) = 1/22$

$E[X] = \frac{23}{2}$, $Var(X) = \frac{21 \cdot 23}{12}$

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

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

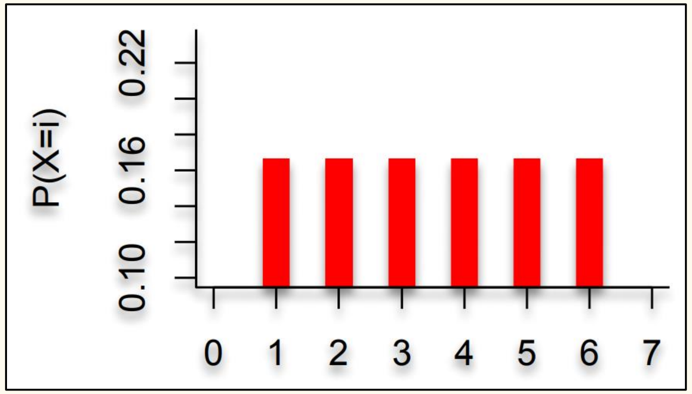
PMF: $\Pr(X = i) = \frac{1}{b - a + 1}$

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

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

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

Indicator RVs

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: $E[X] = 1 \cdot p + 0 \cdot (1-p) = p$

Variance: $Var(X) = E[X^2] - E[X]^2 = p - p^2$

$= p(1-p)$

$E[X^2] = 1^2 \cdot p + 0 \cdot (1-p)$

$= p$

Poll:
pollev.com/hunter312

	Mean	Variance
a.	p	p
b.	p	$1 - p$
c.	p	$p(1 - p)$
d.	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: $X \sim \mathrm{Ber}(p)$

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

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

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

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

iid = Independent and Identically Distributed

$Y_1, Y_2, \ldots, Y_n \sim Ber(p)$ iid

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

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

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

$Y_i \sim Ber(p)$

$E[Y_i] = p$

$Var(Y_i) = p(1-p)$

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:

$E[X] = E[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} E[Y_i] = \sum_{i=1}^{n} p = np$

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

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 $Y_i \sim \mathrm{Ber}(p)$. X is a Binomial random variable where $X = \sum_{i=1}^{n} Y_i$

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

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

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

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

Mean, Variance of the Binomial

If $Y_1, Y_2, \ldots, Y_n \sim \mathrm{Ber}(p)$ and independent (i.i.d), then
$X = \sum_{i=1}^{n} Y_i, \quad X \sim \mathrm{Bin}(n, p)$

Claim $E[X] = np$

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

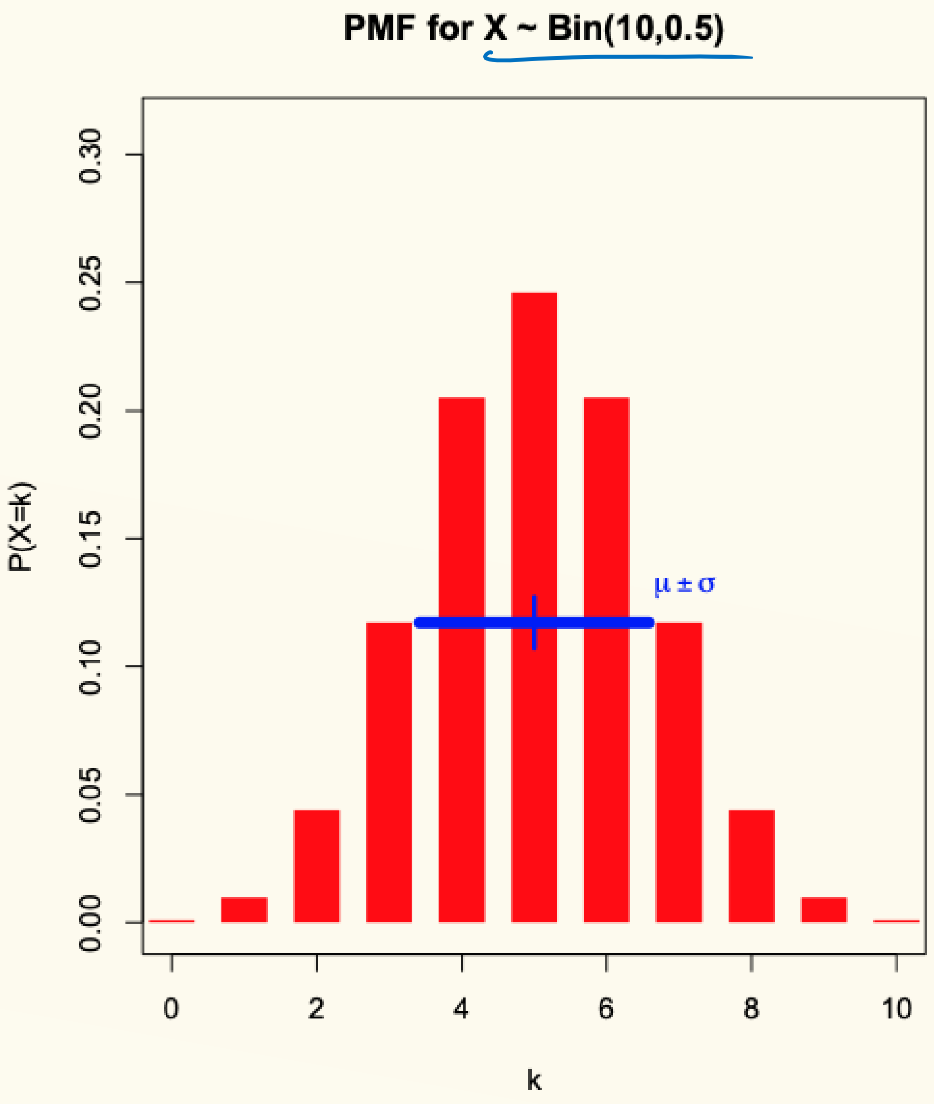
Claim $Var(X) = np(1-p)$

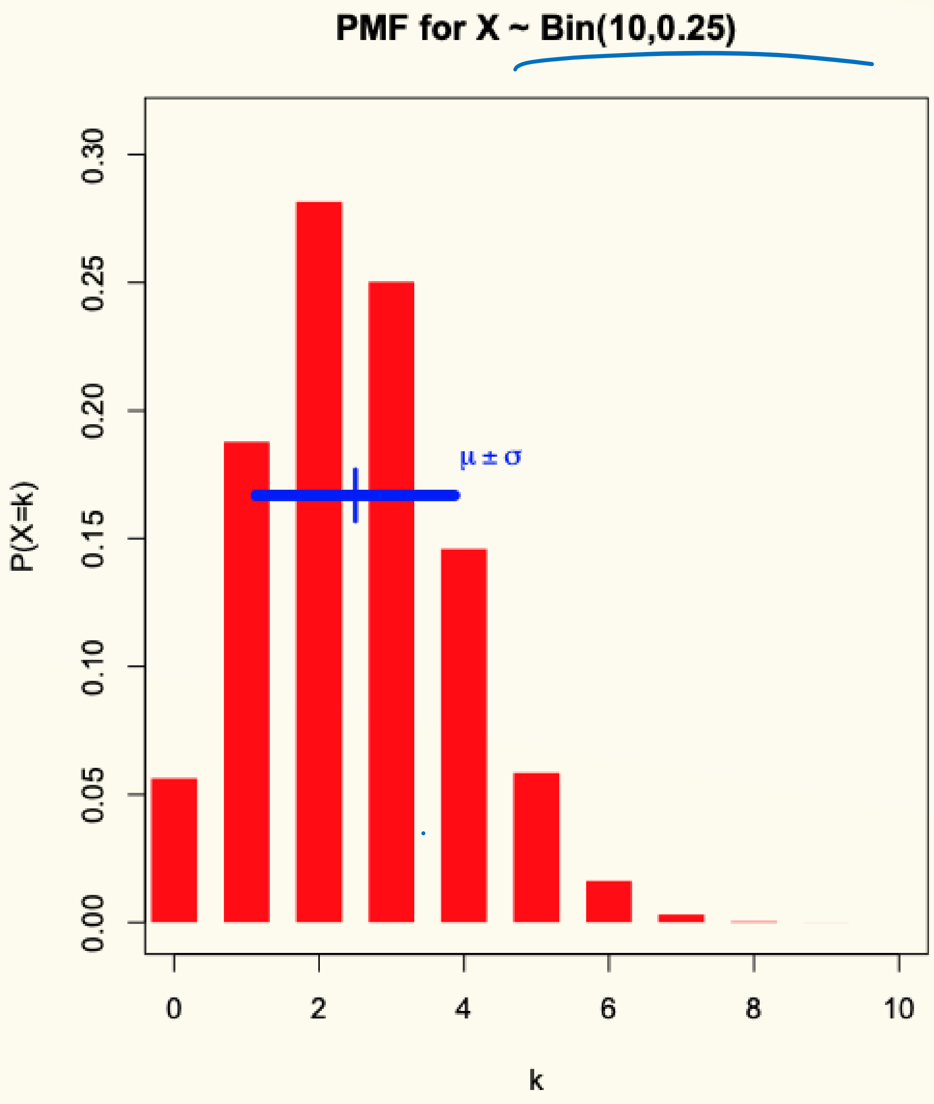
$$Var(X) = Var\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} Var(Y_i) = nVar(Y_1) = np(1-p)$$

Binomial PMFs

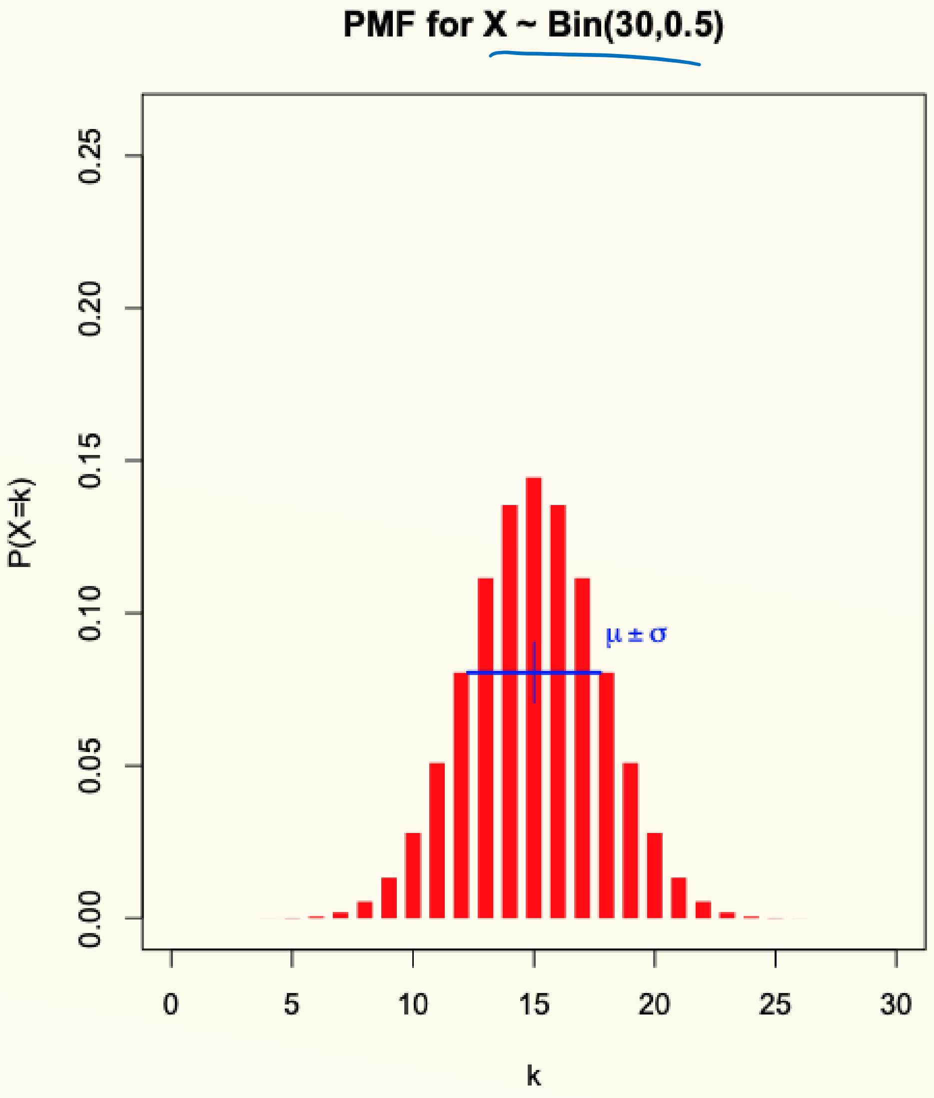
$\mu = E[X]$

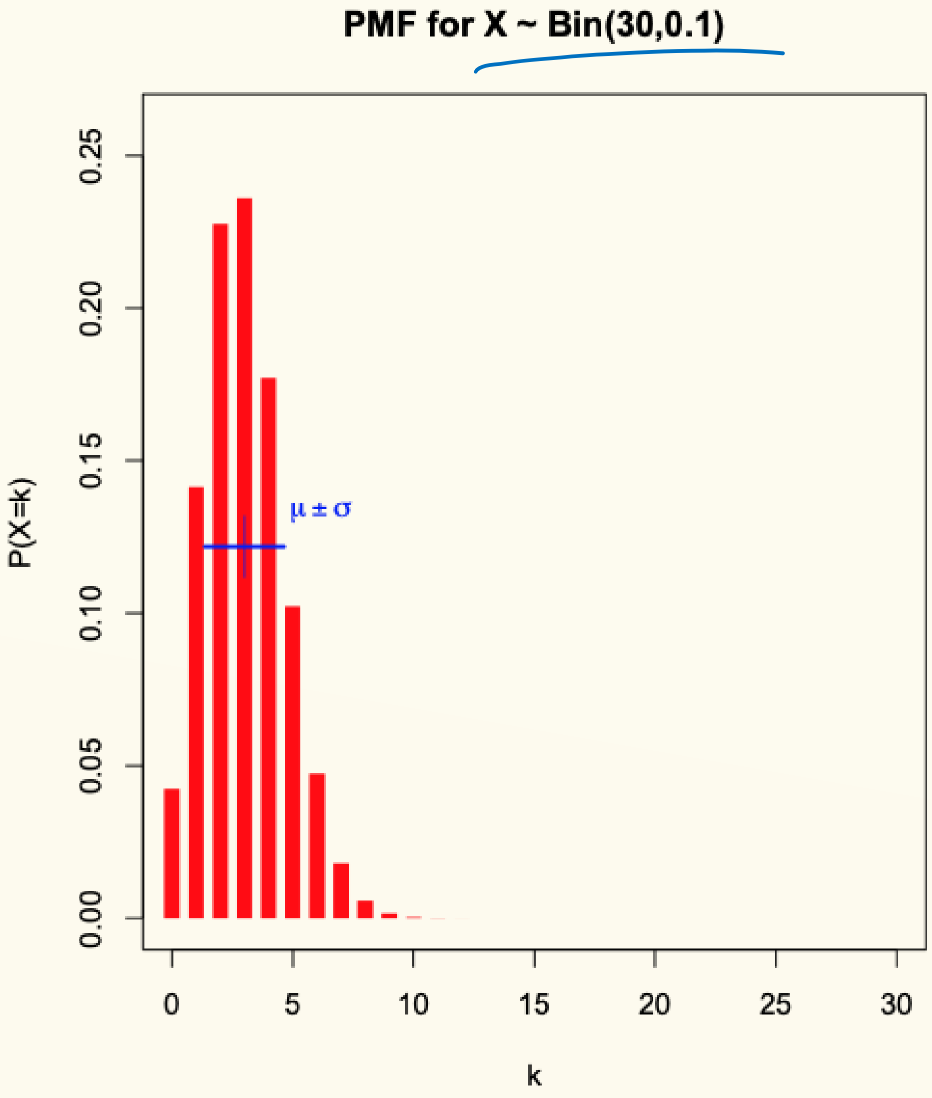
$\sigma = \sqrt{Var(X)} \Rightarrow \sigma^2 = Var(X)$





Binomial PMFs





Example

0101100011...

↑

success w/ prob 0.999, fail w/ prob 0.001

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 $\mathrm{E}[X]$?

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

$E[X] = 1024 \cdot 0.001 = 1.024$

↑ From last slide

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

trials until first success

If X=k, then

$Y_1 = Y_2 = \ldots = Y_{k-1} = 0 \wedge Y_k = 1$

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

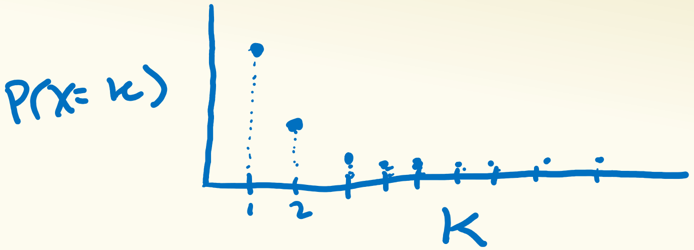
Expectation: $E[X] = 1/p$ (proved in L9)

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 $Y_i \sim \mathrm{Ber}(p)$ before seeing the first success. X is called a Geometric random variable with parameter p.

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

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

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

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

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 $\mathrm{E}[X]$?

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

$\Downarrow$

$X \sim \mathrm{Geo}(0.3677)$

$E[X] = 1/0.3677 \approx 2.72$

p is prob of playing song correctly

Y = # correct notes in one play of the song

$Y \sim \mathrm{Bin}(1000, 0.999)$

$p = \Pr(Y=1000) = \binom{1000}{1000} 0.999^{1000} \cdot 0.001^{0}$

≈ 0.3677

Negative Binomial Random Variables

A discrete random variable X that models the number of independent trials $Y_i \sim Ber(p)$ 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: $\text{E}[X] = \frac{r}{p}$

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

Similar to Geo(p) but waiting for r successes

$X = \sum_{i=1}^{r} Z_i \quad , \; Z_i \sim \text{Geo}(p)$

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

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

Example: Balls in bin
3 red, 2 blue, 4 green

X = # red balls drawn in 4 draws w/o replacement

$X \sim \text{HypGeo}(9, 3, 4)$

Hope you enjoyed the zoo!

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

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

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

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

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

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

$$E[X] = p$$

$$Var(X) = p(1 - p)$$

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

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

$$E[X] = np$$

$$Var(X) = np(1 - p)$$

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

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

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

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

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

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

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

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

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

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

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

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

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