CSE 312

Foundations of Computing II

Lecture 12: Zoo of Discrete RVs

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Anna R. Karlin

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

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) 🦍🐘🦁🐅🦓🐪🦒

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

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

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

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

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

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

$$E[X] = p$$

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

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

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

$$E[X] = np$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

PMF: $\Pr(X=k) = \frac{1}{b-a+1}$, $k \in \{a, a+1, \ldots, b\}$

Expectation: $\sum_{k=a}^{b} k \frac{1}{b-a+1} = \frac{a+b}{2}$

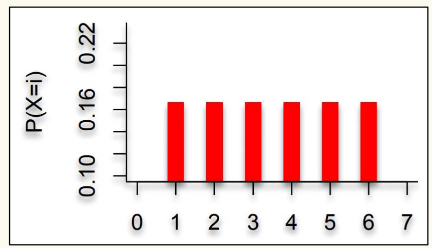
Variance:

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

$E(X^2) = \sum_{k=a}^{b} k^2 \cdot \frac{1}{b-a+1}$

Example: value shown on one roll of a fair die

$X \sim \text{Unif}(1, 6)$



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

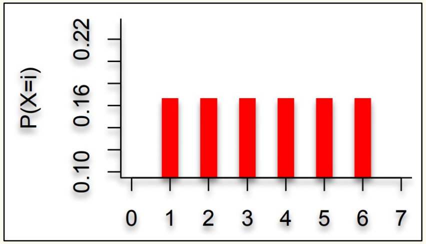
PMF: $\Pr(X = i) = \frac{1}{b - a + 1}$

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

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

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

Bernoulli Random Variables

Indicator

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: p.

Variance: $p - p^2 = p(1-p)$

https://pollev.com/ annakarlin185

Poll:

	Mean	Variance
a.	p	p
b.	p	$1 - p$
c.	p	$p(1-p)$
d.	p	$p(1-p)$

$E(X^2) = 1^2 p + 0^2 (1-p) = p$

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

n, 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 $X = \sum_{i=1}^{n} Y_i$

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:

https://pollev.com/ annakarlin185

$\Pr(X = k) =$

a. $p^k(1-p)^{n-k}$
b. np
c. $\binom{n}{k}p^k(1-p)^{n-k}$
d. $\binom{n}{n-k}p^k(1-p)^{n-k}$

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

$E(Y_i) = p$

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

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: $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \dots, n$

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

$Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n)$

if X_i's mutually indep.

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 $Y_1, Y_2, \dots, Y_n \sim \text{Ber}(p)$ 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 $Var(X) = np(1-p)$

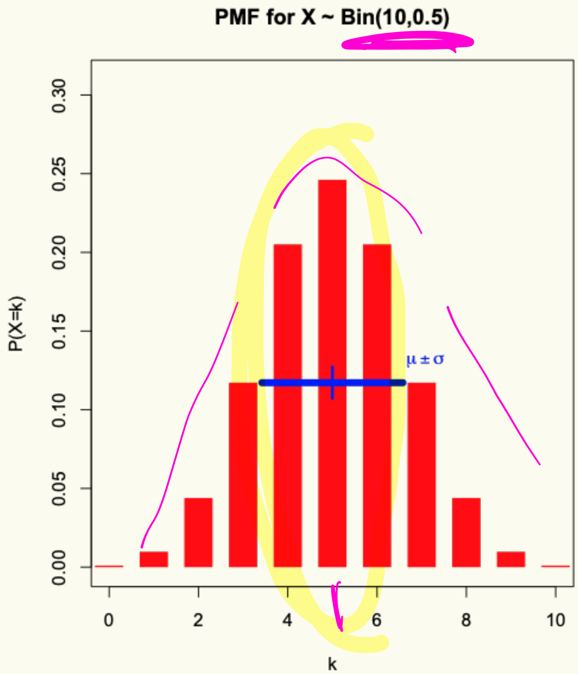
$$Var(X) = Var\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} Var(Y_i) = nVar(Y_1) = np(1-p)$$

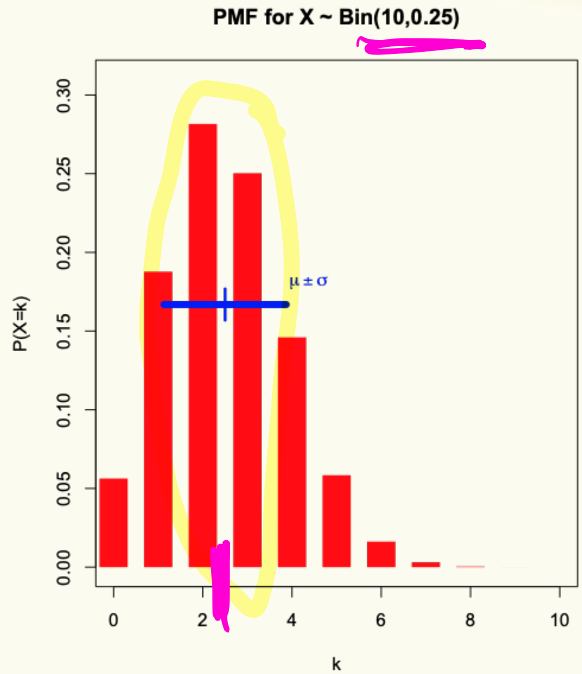
by independence

mean = expectation.

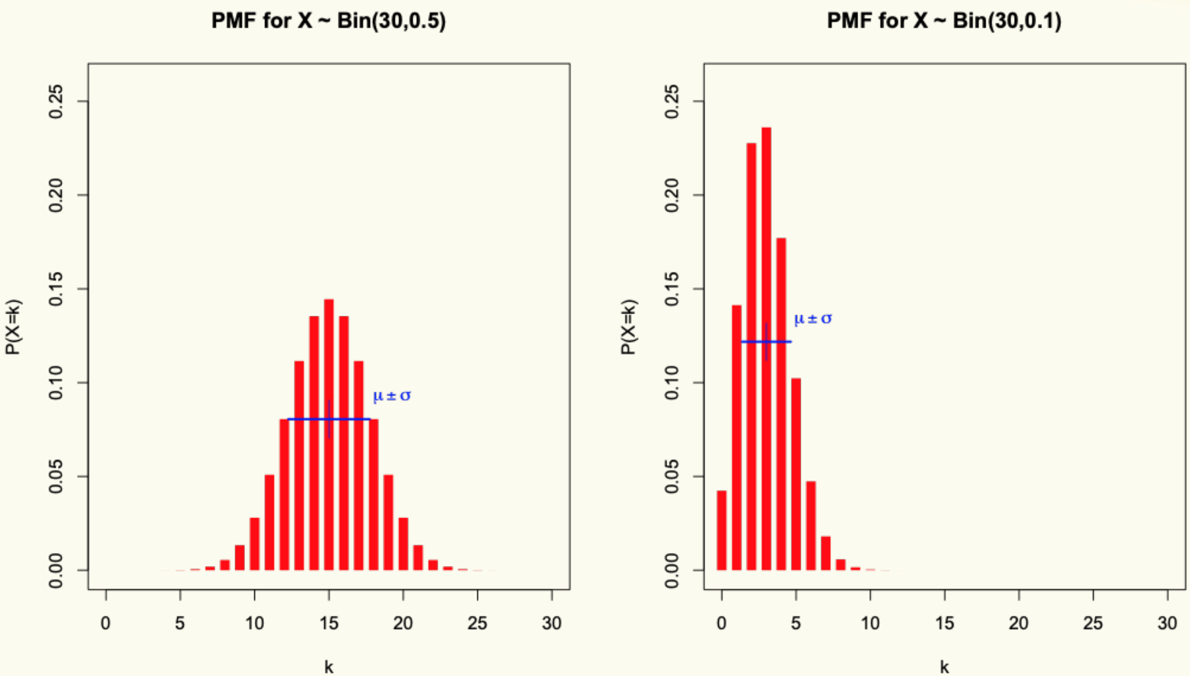
Binomial PMFs

p = 0.25





Binomial PMFs



Example

X ~ Bin(1024, 0.001)
n, p

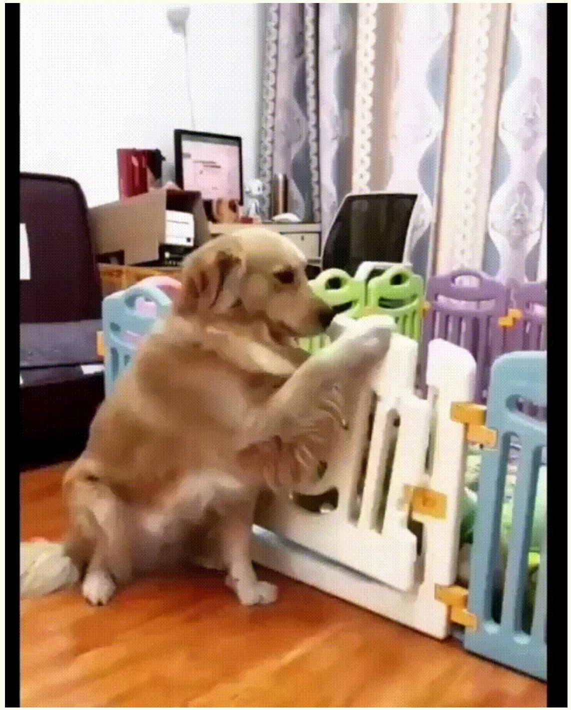
Sending a binary message of length 1024 bits over a network with probability 0.999 of correctly sending each bit in the message without corruption (independent of other bits). Let X be the number of corrupted bits. What is $\mathrm{E}[X]$?

Poll:

https://pollev.com/ annakarlin185

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

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

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

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

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

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 $\mathrm{E}[X]$?

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

0.001 Prob make mistake on a single note.

Pr(successfully play all the way thru) $= (0.999)^{1000}$

Y: # notes I play until I mess up a note.

$\Pr(Y=10) = 0.999^{9}\, 0.001$

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:

Expectation:

Variance:

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 $\mathrm{Z}_i \sim \mathrm{Geo}(p)$. X is called a Negative Binomial random variable with parameters r, p.

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

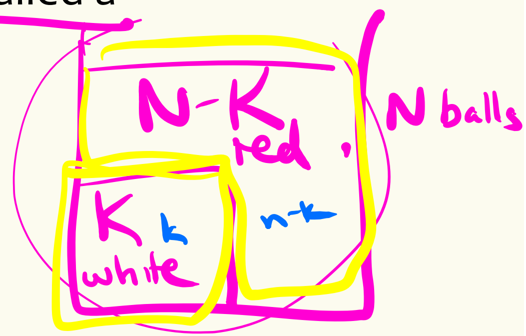
PMF: $\Pr(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$

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

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

Hypergeometric Random Variables

A discrete random variable X that measures the number of white balls you draw when you draw n balls uniformly at random from a total of N of which K are white and the rest are black. 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:

X is # of white balls you pull out when you pull out set of size n at random

k white
n-k red

Hypergeometric Random Variables

A discrete random variable X that measures the number of white balls you draw when you draw n balls uniformly at random from a total of N of which K are white and the rest are black. 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: $\mathrm{Var}(X) = n\frac{K(N-K)(N-n)}{N^2(N-1)}$

Hope you enjoyed the zoo! 🦍🐘🦁🐅🦓🐪🦒

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

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

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

$$P(X = 1) = p, P(X = 0) = 1 - p$$
$$E[X] = p$$
$$Var(X) = p(1 - p)$$

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

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

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

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

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

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

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

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