CSE 312

Foundations of Computing II

Lecture 12: Zoo of Discrete RVs

Review Variance – Properties

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

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

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

(Proof: Exercise!)

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

Review 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, $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

$P(X_1 = k_1 \wedge \cdots \wedge X = k_2) = P(X_1 = k_1) \cdot P(X_2 = k_2) \cdots P(X_n = k_n)$

Corollary. If $X_1, X_2, \ldots, X_n$ mutually independent,

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

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)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 (integer) value between integers a and b (inclusive), is uniform.

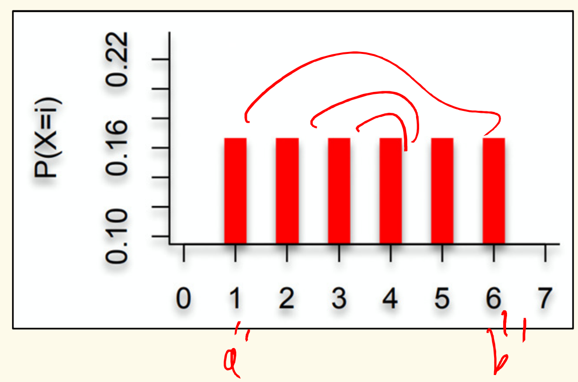
Notation: Unif(a, b)

PMF: $\frac{1}{b-a+1}$ for $i \in [a,b]$, i integer

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

Variance:

Example: value shown on one roll of a fair die



Discrete Uniform Random Variables

\sim

~ "is distributed as"

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

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

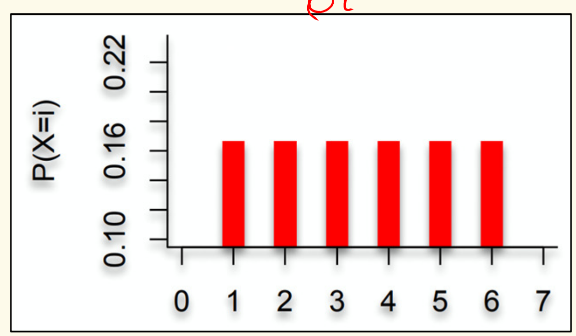
PMF: $\text{P}(X = i) = \frac{1}{b - a + 1}$

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

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

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

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



$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

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

Expectation:

Variance:

Poll:
pollev.com/paulbeame028

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

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

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

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

Examples:

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

multiple choice

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/paulbeame028
$P(X = k)$

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

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

Expectation:

Variance:

Poll:
pollev.com/paulbeame028

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

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

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

Mean, Variance of the Binomial

"i.i.d." is a commonly used phrase.
It means "independent & identically distributed"

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

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

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

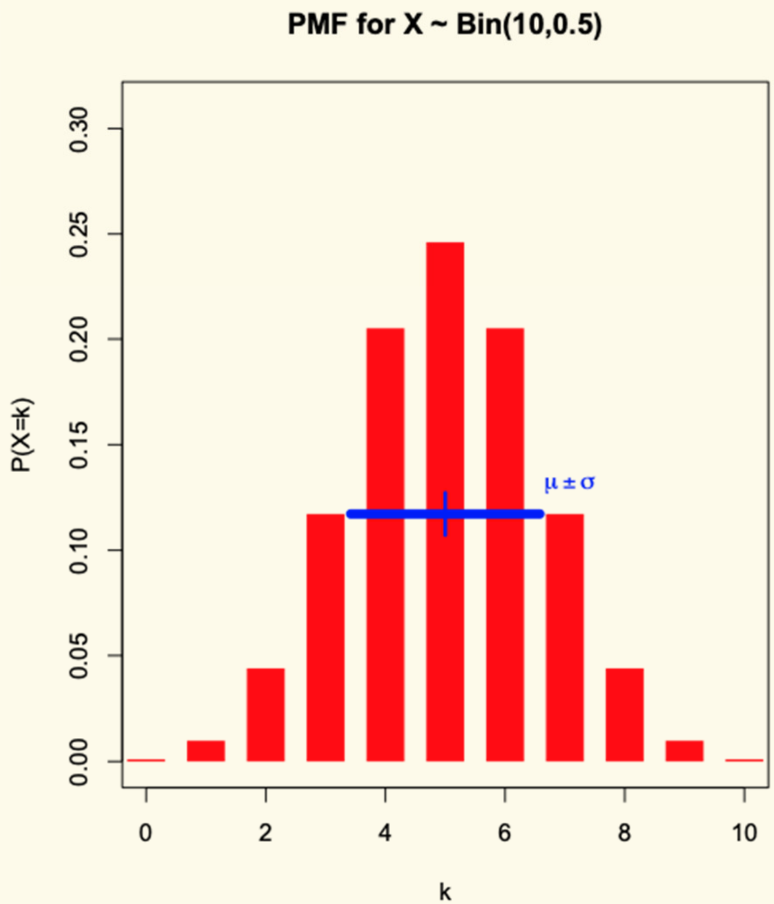
Claim $\text{Var}(X) = np(1-p)$

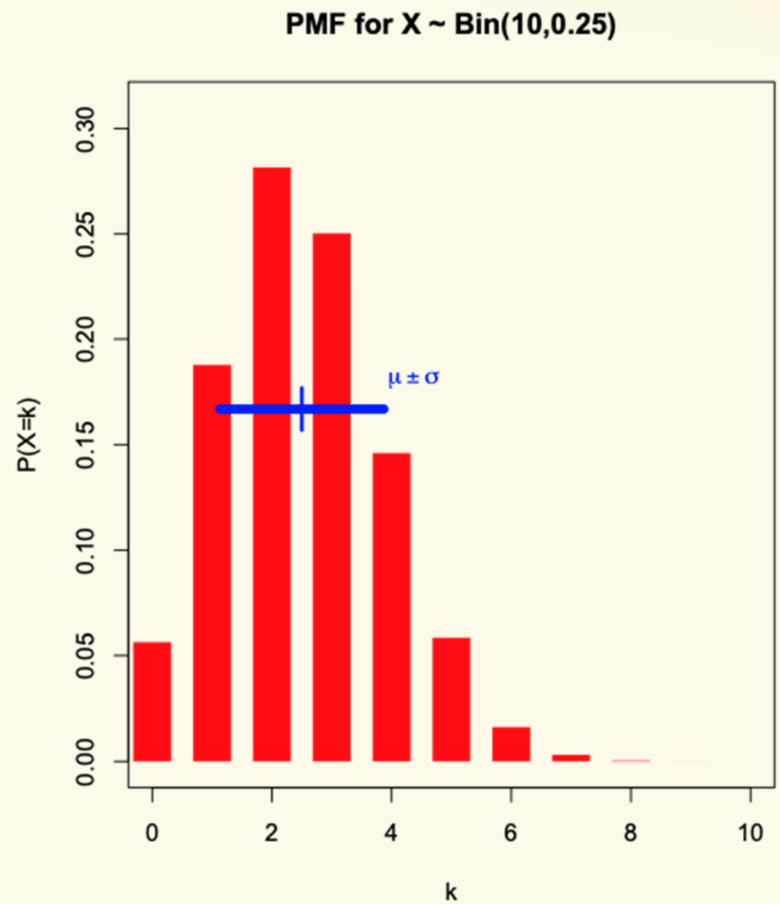
$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{Var}(Y_i) = n\text{Var}(Y_1) = np(1-p)$$

Binomial PMFs

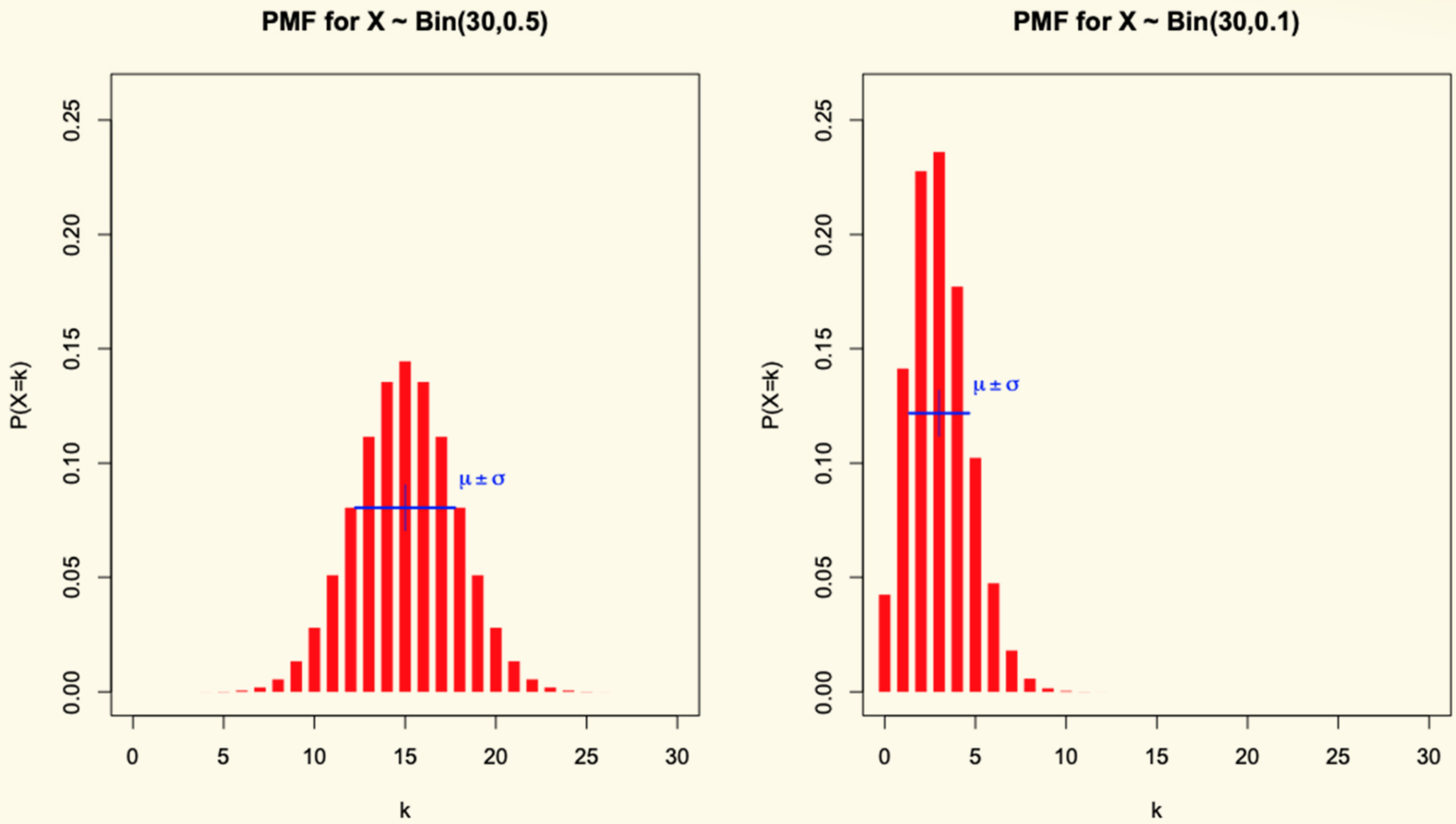
$\sigma^2 = Var(X)$

$\sigma = \sqrt{Var(X)}$





Binomial PMFs



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 $\mathbb{E}[X]$?

Poll:
pollev.com/paulbeame028

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

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

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 $Y_i \sim \mathrm{Ber}(p)$ before seeing the first success.

X is called a Geometric random variable with parameter p.

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

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

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

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 $\mathbb{E}[X]$?

prob correct $= (.999)^{1000}$

$\frac{1}{1000}$ error

$p = \left(1 - \frac{1}{1000}\right)^{1000}$ no error $\approx e^{-1}$

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

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 $\text{Z}_i \sim \text{Geo}(p)$.

X is called a Negative Binomial random variable with parameters r, p.

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

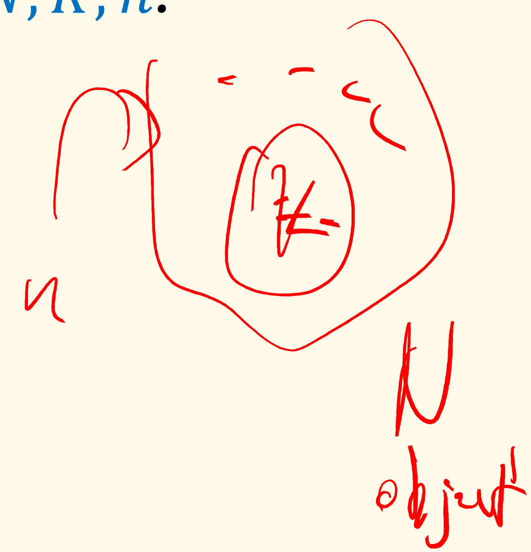
PMF: $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$

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

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

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

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

Hope you enjoyed the zoo!

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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