# Section 8: Sums of independent r.v. and Binomial, Geometric distributions

STOR 155.02, Spring ’21

updated 2021-03-14

# What you will learn

- Variance for sums of independent random variables
- Binomial and geometric distribution basics

# Resources

- Textbook: ch 4.3-4.4

# Rest of the semester

## Discrete and continuous distributions

Binomial, geometric, normal, uniform

## Law of large numbers and central limit theorem

## Models and estimation of parameters

## Hypothesis testing

## Final data project instead of exam

# Sums of independent random variables

# Recap: X is Bernoulli(p)

## Outcomes of X are?

## Example?

**tell me what X is**

**and what p is**

## Expectation

$$E(X) = p$$

## Variance

A useful formula.

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

$$= E\left[X^2 - 2XE(X) + E(X)^2\right]$$

$$= E\left[X^2\right] - 2E(X)E(X) + E(X)^2$$

$$= E(X^2) - (E(X))^2$$

**For Bernoulli X this is**

$$= p(1 - p)$$

# Example: Sum of N independent random variables

Each week, Binky looks for ticks on her household pets: A cat, a dog and a hamster. If she finds at least one, her parents will pay her in ice cream.

- ticks on dog is worth 1 pint
- ticks on cat is worth 1/2 pint
- ticks on hamster is worth 1/4 pint

$$X_1 = 1\{\text{at least one tick on dog this week}\} \text{ is Bernoulli}(1/3)$$

$$X_2 = 1\{\text{ticks on cat}\} \text{ is Bernoulli}(1/4)$$

$$X_3 = 1\{\text{ticks on hamster}\} \text{ is Bernoulli}(1/6)$$

# Call Y the random variable representing Binky's ice cream earned this week

$$Y = \text{ ? (in terms of X\_1, X\_2, X\_3)}$$

$$E(Y) = \text{ ?}$$

# What about the variance of Y?

**remember: this tells us how spread out random variable's values are on average, relative to the mean**

**can be hard to calculate for sums of random variables in general**

# Rule: Constants multiply variance

If $X$ is any random variable and $a$ is a constant

$$Var(aX) = a^2 Var(X)$$

# Rule: Variance for sum of independent r.v.

If $Z_1 \dots Z_N$ are independent

$$Var\left(\sum_{i=1}^{N} Z_i\right) = \sum_{i=1}^{N} Var(Z_i)$$

# Example: Binky's ice cream pay

**assume X_1 ... X_3 independent**

**Use the two rules above and the representation of $Y$ in terms of $X_1 \ldots X_3$**

$$Var(Y) = \ ?$$

## Extended example:

**Binky checks for ticks for a period of 5 weeks.**

**call $Y_i$ the ice cream payout for week $i = 1 \ldots 5$**

**How would you calculate the variance of Binky's total ice cream pay over the entire 5-week period?**

# Binomial distribution

# Intuitive description: Counting successes in independent 'trials'

Related to a past data homework:

- you randomly sample $N = 10$ rows **with replacement** from the Beauty and the Beast script dataset
- call a sample a 'success' if you drew a row representing one of Belle's lines
- each draw has probability $p = 0.2$ chance of success

then

> > the random variable representing the number of successes in $N$ independent trials has the **Binomial(N, p) distribution**, where $p$ is the chance of success on each individual trial

# Mathematical definition

$$X_1, X_2 \ldots X_N \text{ are independent Bernoulli(p)}$$

$$X = \sum_{i=1}^{N} X_i \quad \text{is Binomial(N, p)}$$

Expectation?

Variance?

In words, what is X/N?

Expectation of X/N?

Variance of X/N?

# Probabilities of Binomial(N, p)

$$X \text{ can take values } 0, 1, \dots N$$

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

$$= (\text{number of ways to get k successes}) \times (\text{probability of each way})$$

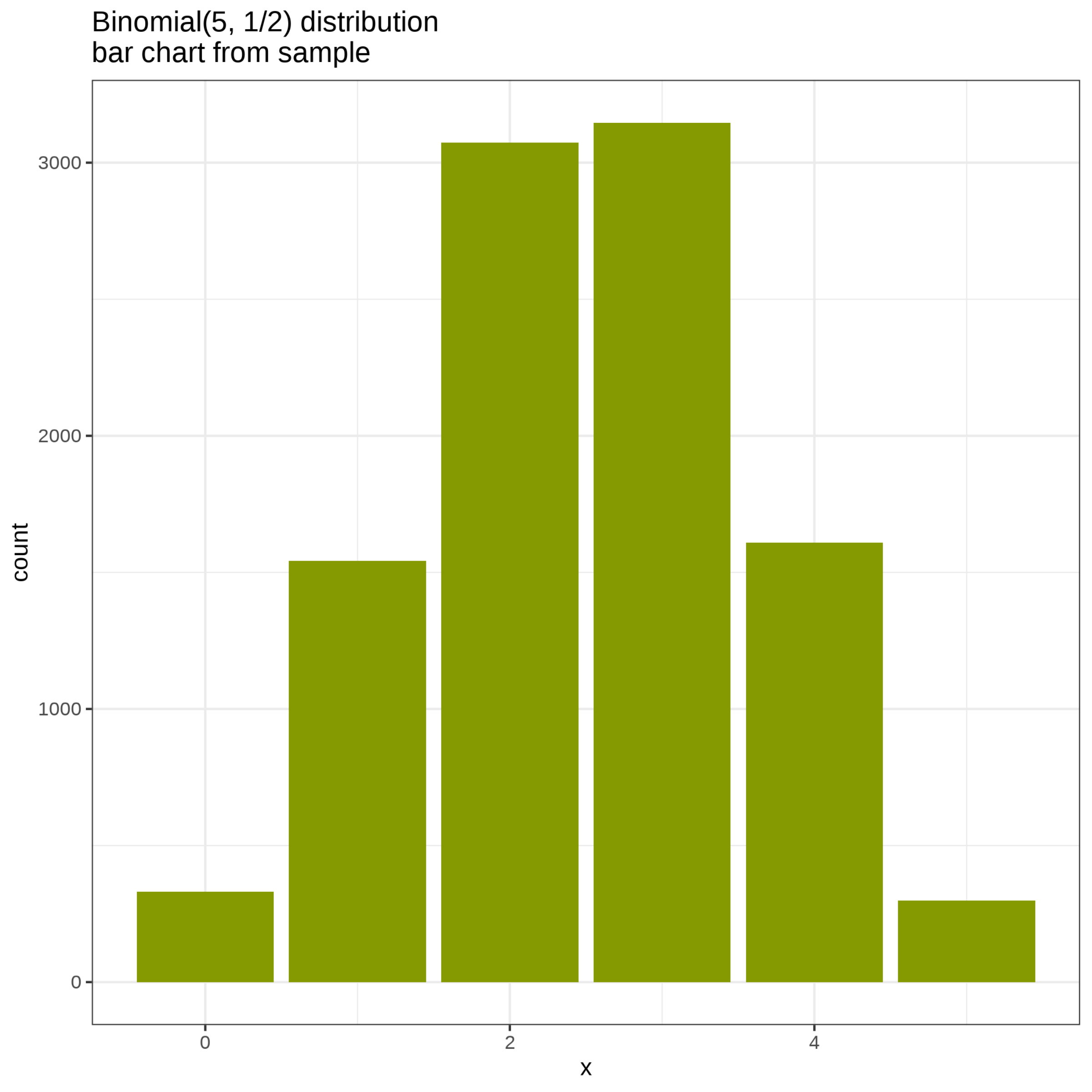
$$\binom{N}{k} = \text{N choose k} = \frac{N!}{(N-k)!k!} = \frac{N(N-1)\dots(N-k+1)}{k(k-1)\dots 1}$$

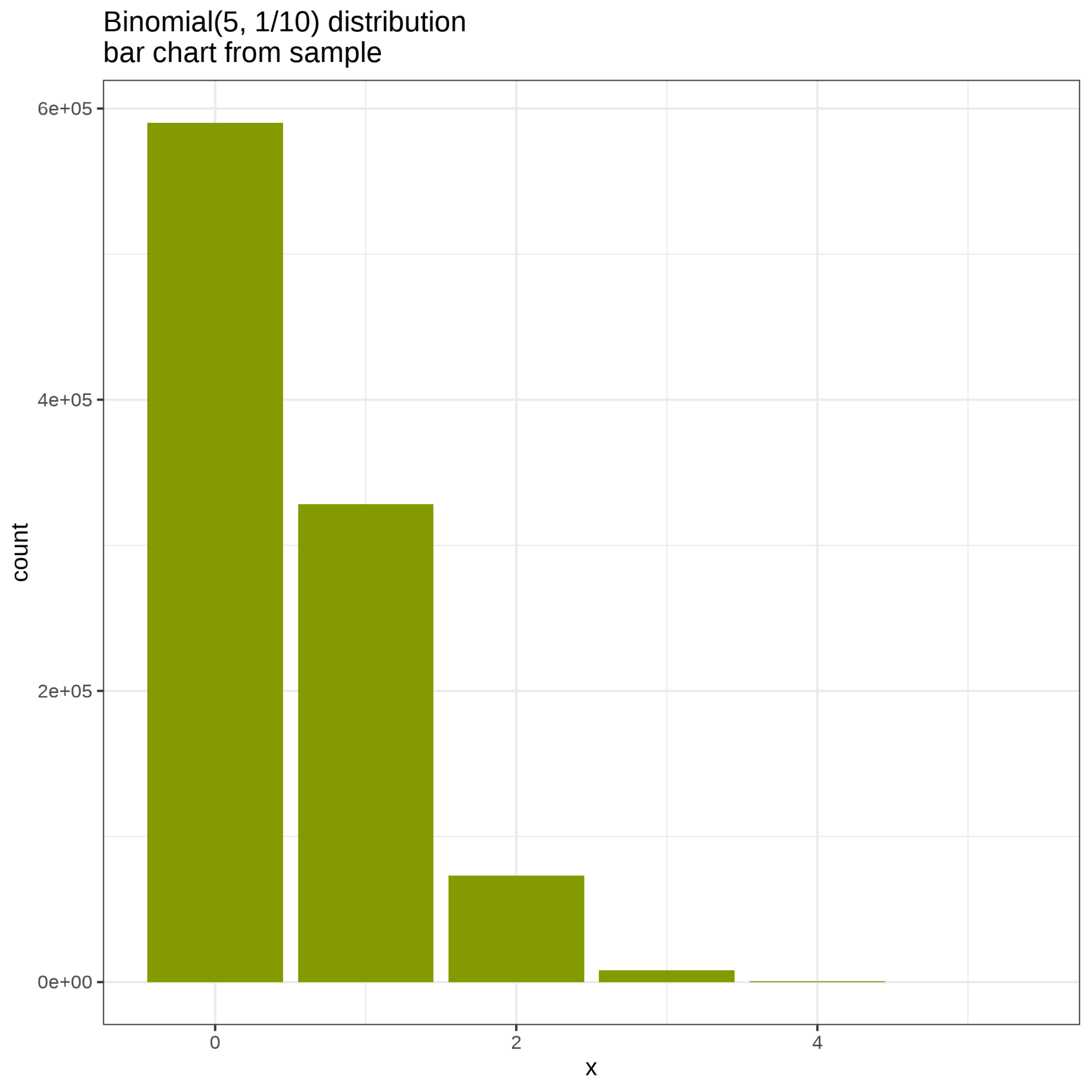
# This makes sense because

If I have $2$ successes, then I have $N-2$ failures. Since independent,

$$P(X_1 = 1 \text{ and } X_2 = 1 \text{ and } X_3 = 0 \dots X_N = 0) = p^2(1-p)^{N-2}$$

That is only one way to get exactly 2 successes. Each way has the same probability and is disjoint of other ways, so you add the same number up $\binom{N}{2}$ times.


Binomial(5, 1/2) distribution
bar chart from sample
count
3000
2000
1000
0
0
2
4
x


Binomial(5, 1/10) distribution
bar chart from sample
6e+05
4e+05
2e+05
0e+00
count
0
2
4
x

# Geometric distribution

# Intuitive description: Waiting for success in independent 'trials'

- you randomly sample rows **with replacement** from the Beauty and the Beast script dataset
- call a sample a 'success' if you drew a row representing one of Belle's lines
- each draw has probability $p = 0.2$ chance of success

then

> > the random variable representing the **number of the first successful trial** in $N$ independent trials has the **Geometric(p) distribution**, where $p$ is the chance of success on each individual trial

**for example, this random variable would be 6 if you first drew 5 non-Belle lines then on the 6th drew one of Belle's lines**

# Properties

$X$ is Geometric(p)

# Probabilities

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

# Q: Why does this make sense?

Use the intuitive description

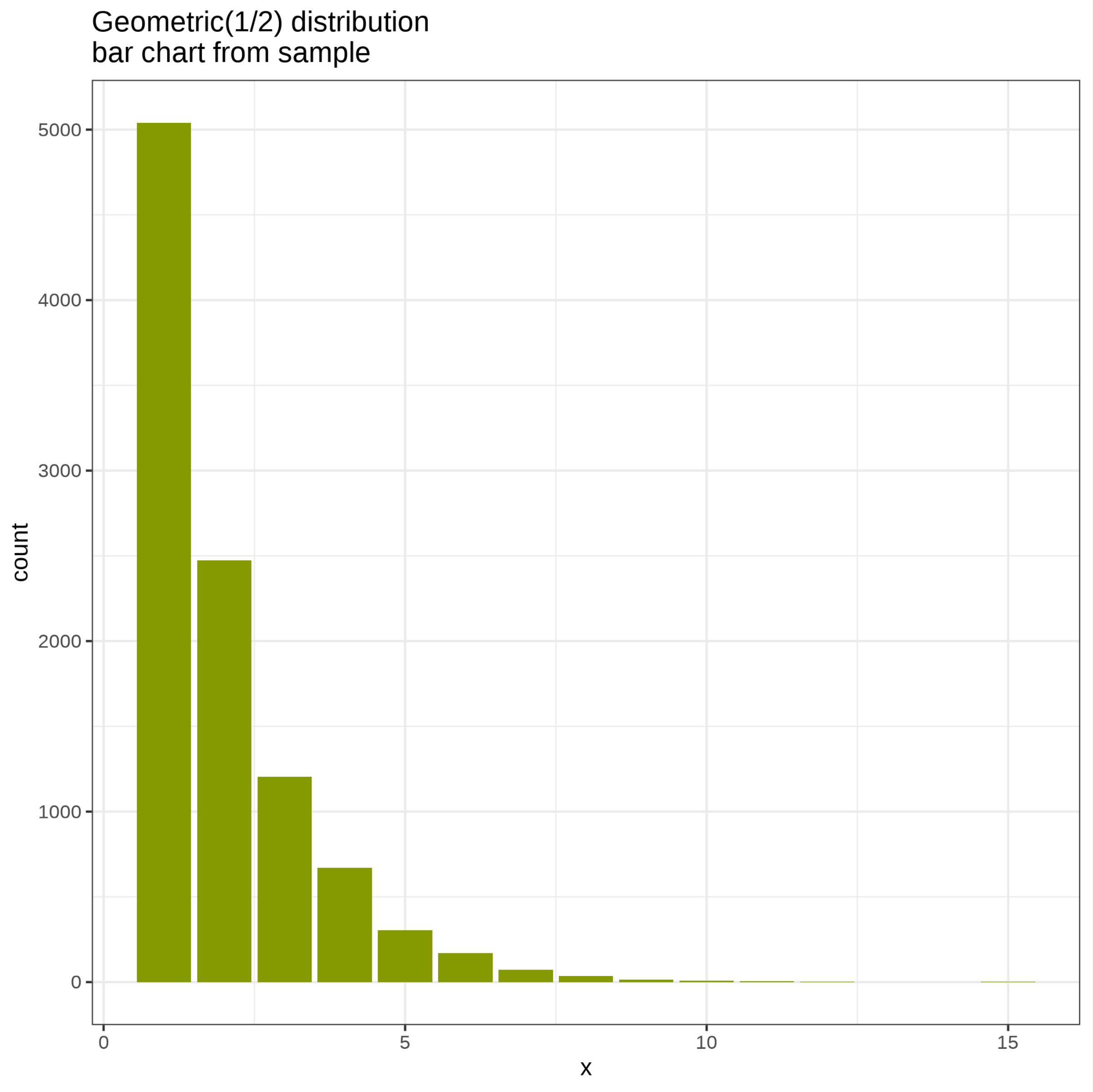
# Expectation

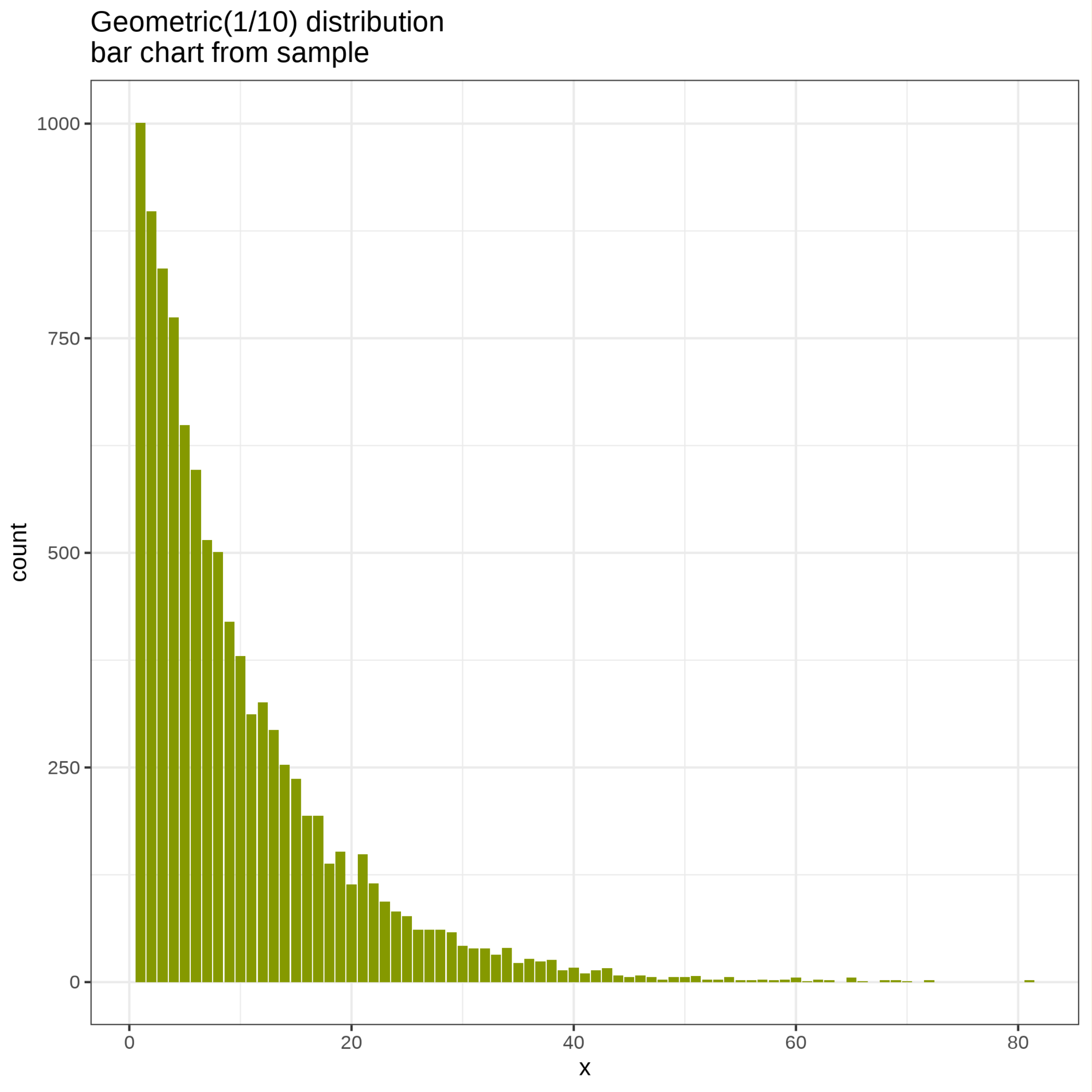
$$E(X) = \sum_{i=1}^{\infty} iP(X = i) = \frac{1}{p}$$

# Variance

$$E(X) = \sum_{i=1}^{\infty} (i - 1/p)^2 P(X = i)$$

$$= \frac{1-p}{p^2}$$


Geometric(1/2) distribution
bar chart from sample
count
5000
4000
3000
2000
1000
0
0
5
10
15
x


Geometric(1/10) distribution
bar chart from sample
count
1000
750
500
250
0
0
20
40
60
80
x

PollEv.com/brendanbrown849

poll closes at __________


five more minutes