Useful (Continuous) Probability Distributions

...

6 February 2019

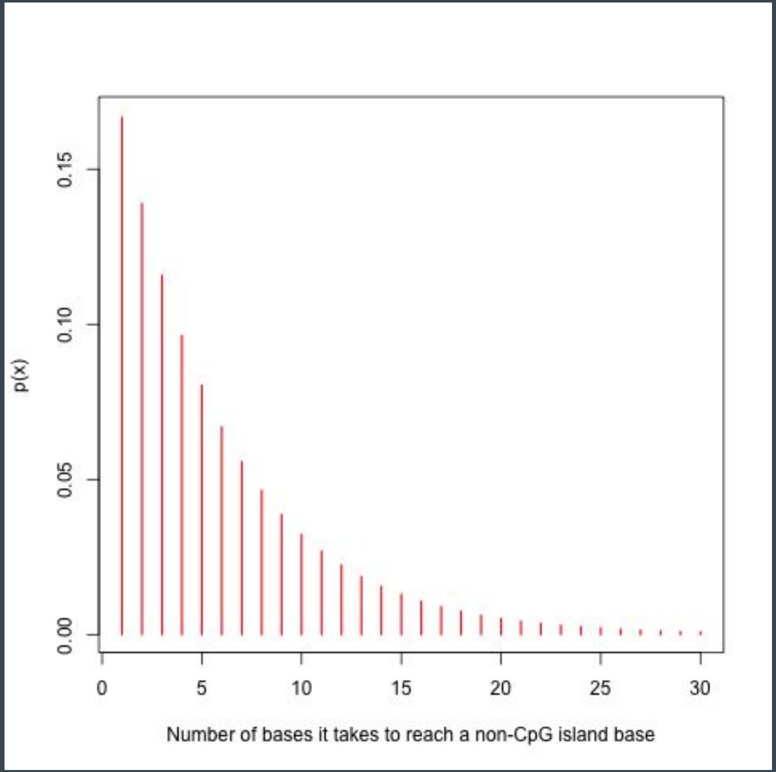
PLSC 309

Review

- We learned that data is drawn from random variables
- Random variables follow a probability distribution
  - These distributions have parameters, which change their shape
- We guess what distribution based on the process that creates the data
  - Always an *assumption*
- Discrete distributions
  - Bernoulli
  - Binomial
  - Multinomial
  - Geometric

Exponential Distribution

Geometric distribution

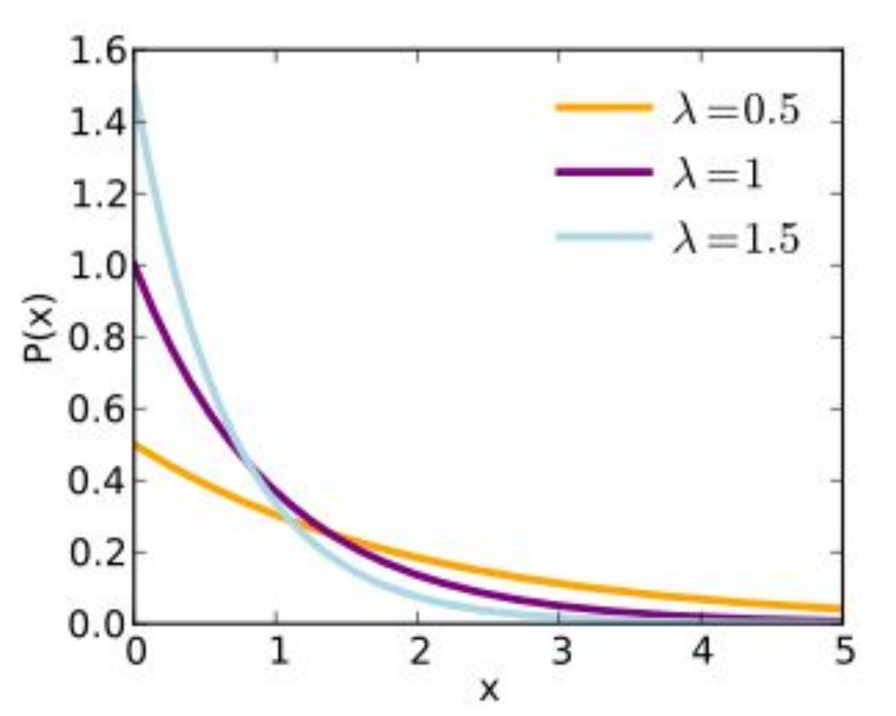


$$(1-p)^{k-1}p$$

Geometric to exponential

- How many *k* Bernoulli trials are needed before one success
- If we consider trials to be a unit of distance (i.e 2 trials is a greater distance than 1)
- We can rephrase as how much distance before the thing we're interested in happens
- When those distances are continuous instead of discrete, we use the exponential distribution

Exponential Distribution

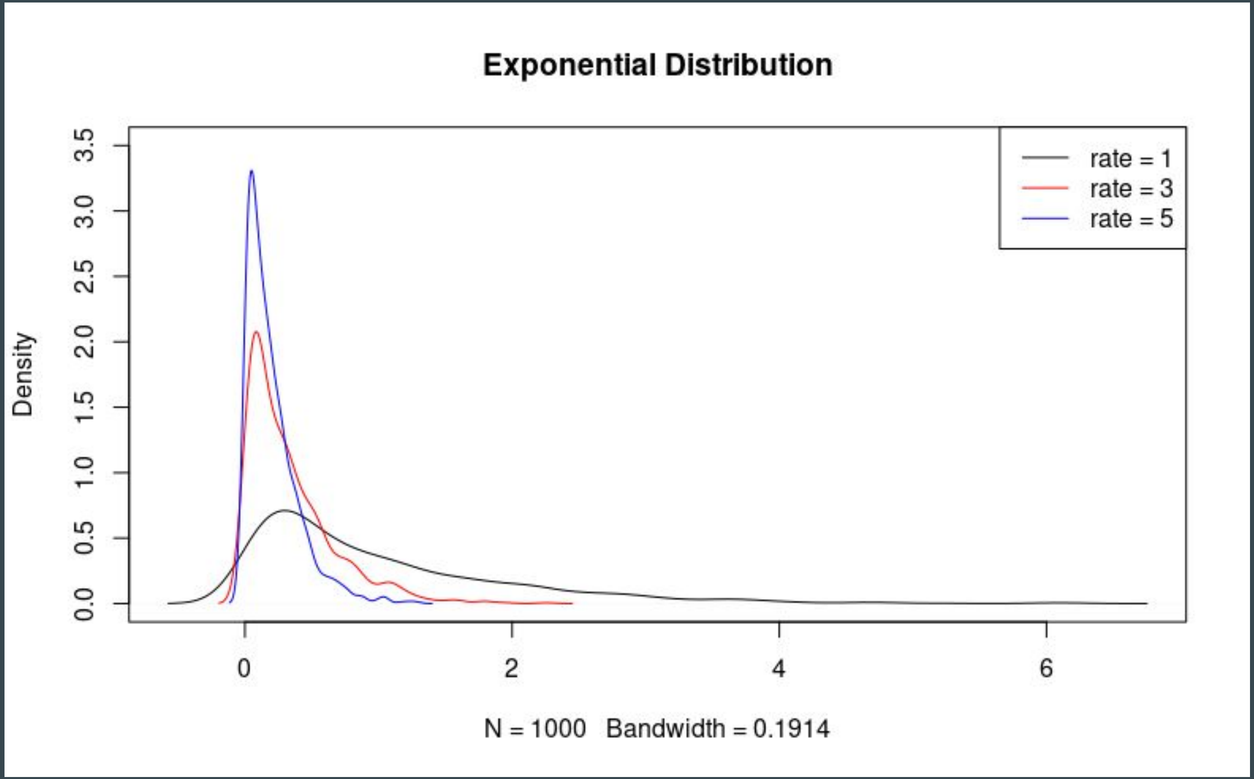


$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Exponential distribution

- Captures any question of “how long until x happens”
- Long-tailed, skewed
- λ: the rate parameter
  - Known as the “constant of proportionality”
  - Decay of function is proportionate to λ

Changes in rate parameter



Expected value and variance for exponential

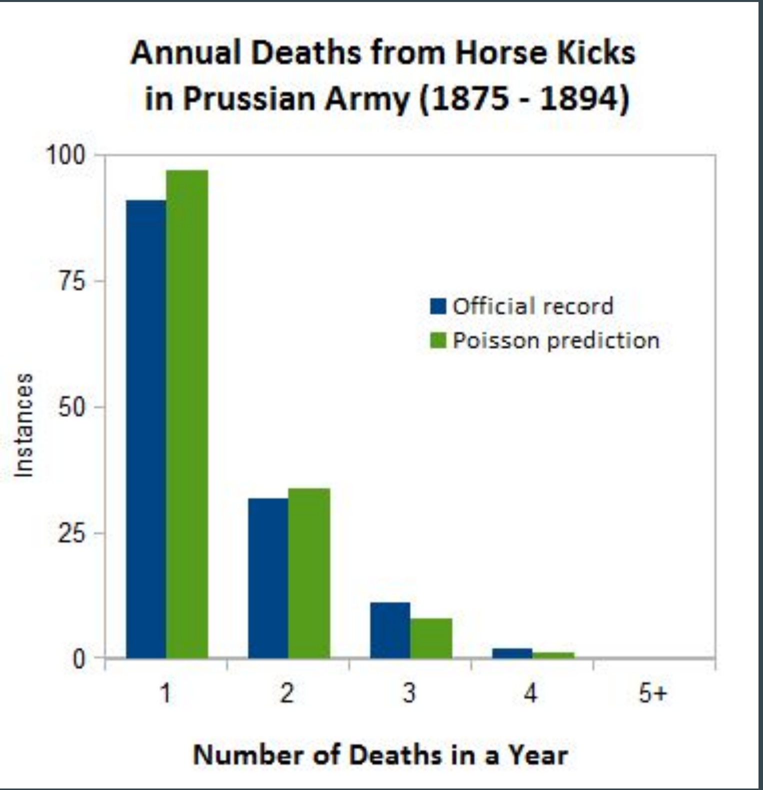
- Both expected value and variance are defined in terms of the rate parameter
- Identical to the geometric distribution
- $E(X) = 1 / \lambda$
- $Var(X) = 1 / \lambda^2$

Examples of exponential processes

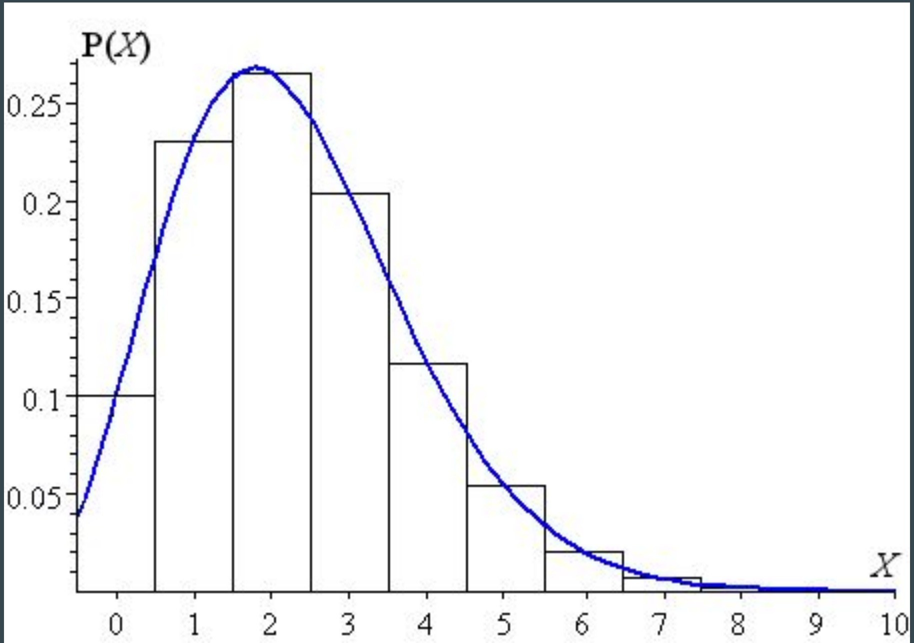
- You are waiting in line at the bank. Generally, there is a rate parameter of 2 for this type of line. How long do you wait before you are seen?
  - $E(X) = 1/2$
  - $Var(X) = 1/4$
- How far does a plane travel before needing to be refueled? The rate parameter is 3.
  - $E(X) = 1/3$
  - $Var(X) = 1/9$

Poisson Distribution

An interesting history



Poisson distribution

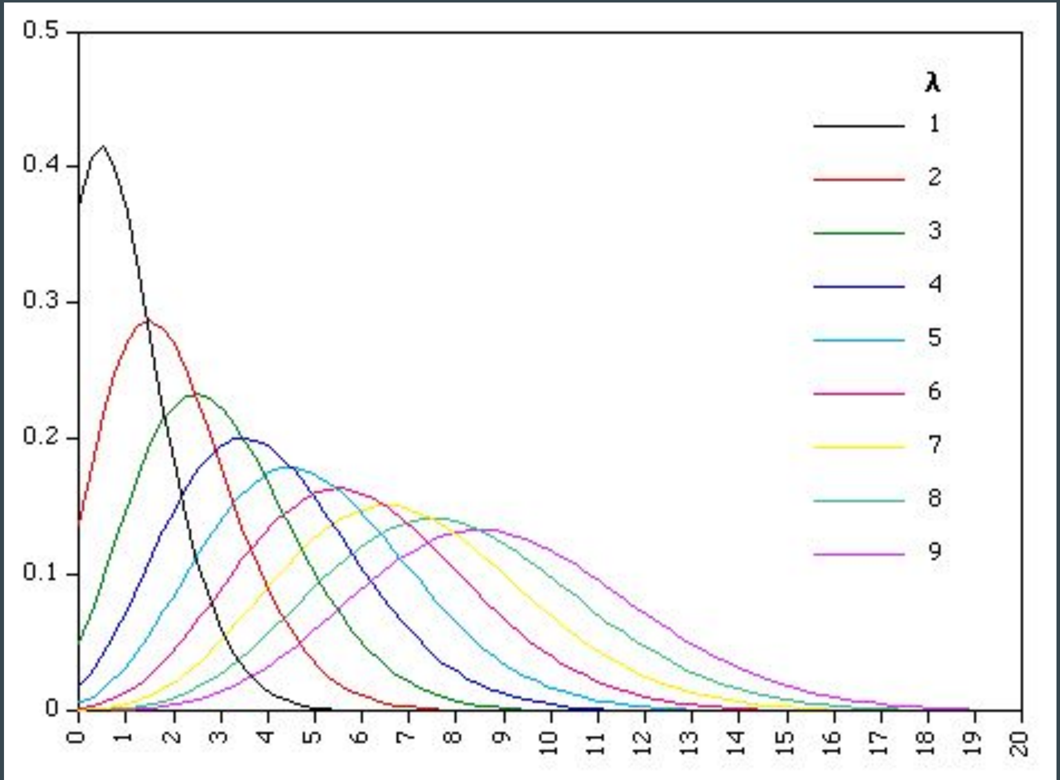


$$e^{-\lambda}\frac{\lambda^{x}}{x!}$$

A distribution of counts

- The poisson can model either discrete or continuous counts
- Similar to exponential distribution
  - Exponential measures the distance between successes
  - Poisson measures the average number of success given an interval
- The parameter λ is the average number of successes per interval

What happens as λ changes?



Expected value and variance for poisson

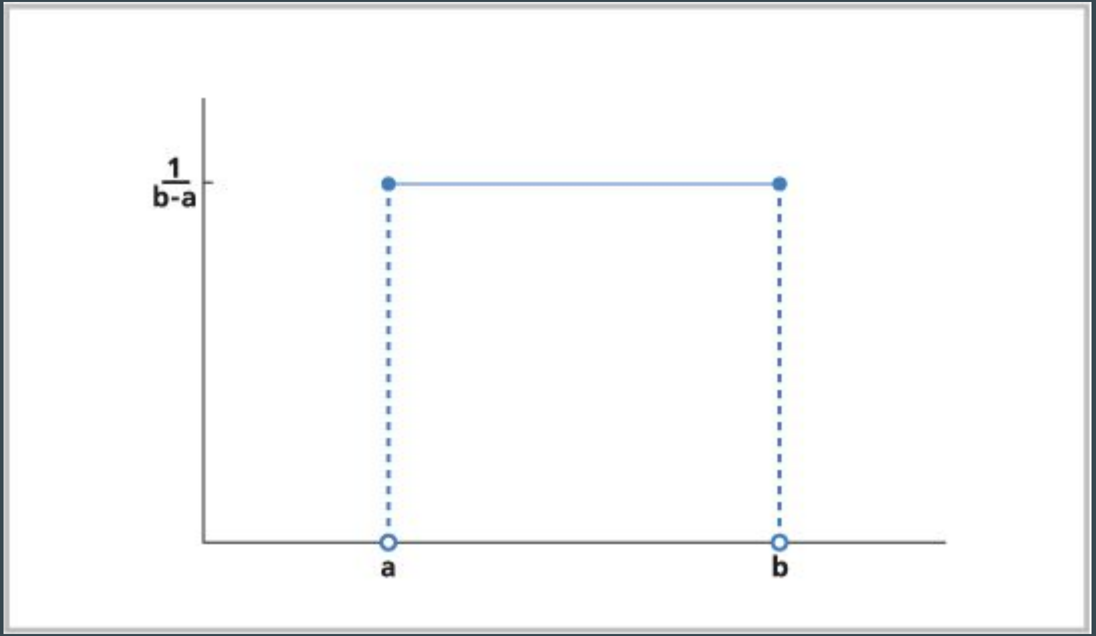
- Since λ is the average number of successes per interval, expected value is straightforward
    - $E(X) = \lambda$
- $Var(X) = \lambda$
    - Variance and mean are equal for this distribution

Examples of poisson processes

- Anything with a count!
- Predicting the number of students to show up to class. Average attendance is 28.
  - $E(X) = 28$
  - $Var(X) = 28$
- Predicting the number of new civil wars throughout the world. There is about one new civil war every year.
  - $E(X) = 1$
  - $Var(X) = 1$

Non-natural distributions

Uniform distribution



$$x \in [a, b]$$

Expected value and variance of a uniform distribution

- There are only two parameters
  - *a:* minimum value
  - *b:* maximum value
- $E(X) = ½ (a+b)$
- $Var(X) = 1/12 (a+b)^2$

What are uniform distributions used for?

- Any situation where you want the probability of X to be the same for all observations
- Simple random sampling
- Assigning treatments in an experiment

Sampling distributions

What is a statistic?

- A statistic is any number or quantity derived from data
  - Mean
  - Variance
  - Parameter estimates
- Each draw of data produces a single statistics

What is a sample?

- A sample is created everytime you draw data from the real-world
- For example, we wait at the bank to record wait times
  - Sample: the data that we collected at that particular bank over those particular times
  - Population: waiting times for all banks across the world at all times
- For example, we want to a survey on likely voters
  - Sample: the data on 100 likely voters who we go and talk to
  - Population: all people likely to vote throughout the country
- If we do our sampling well, the probability distribution from the sample will be similar to the probability distribution from the population
  - But there is always random variation!

Sampling distribution

Respondent	Age
1	20
2	18
3	22
4	23
5	17
6	21
7	20

Sample	Mean
1	20.14

Sampling distribution

Respondent	Age
8	22
9	19
10	24
11	16
12	20
13	18
14	21

Sample	Mean
1	20.14
2	20

Sampling distribution

Respondent	Age
15	18
16	22
17	19
18	21
19	19
20	23
21	20

Sample	Mean
1	20.14
2	20
3	20.28

Sampling distribution

Respondent	Age
22	23
23	16
24	18
25	22
26	23
27	17
28	19

Sample	Mean
1	20.14
2	20
3	20.28
4	19.71

Sampling distribution

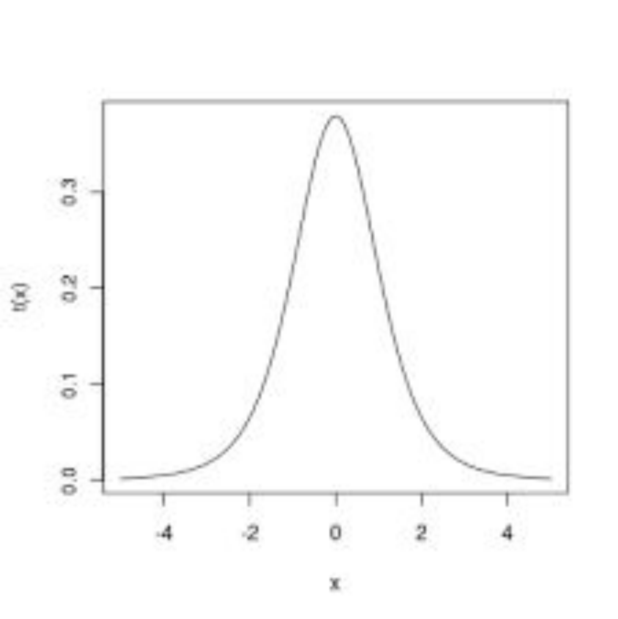
Respondent	Age
29	18
30	21
31	20
32	23
33	19
34	21
35	22

Sample	Mean
1	20.14
2	20
3	20.28
4	19.71
5	20.57

Student's T distribution

- How do you make inferences when you only have very small samples?

Student's T distribution

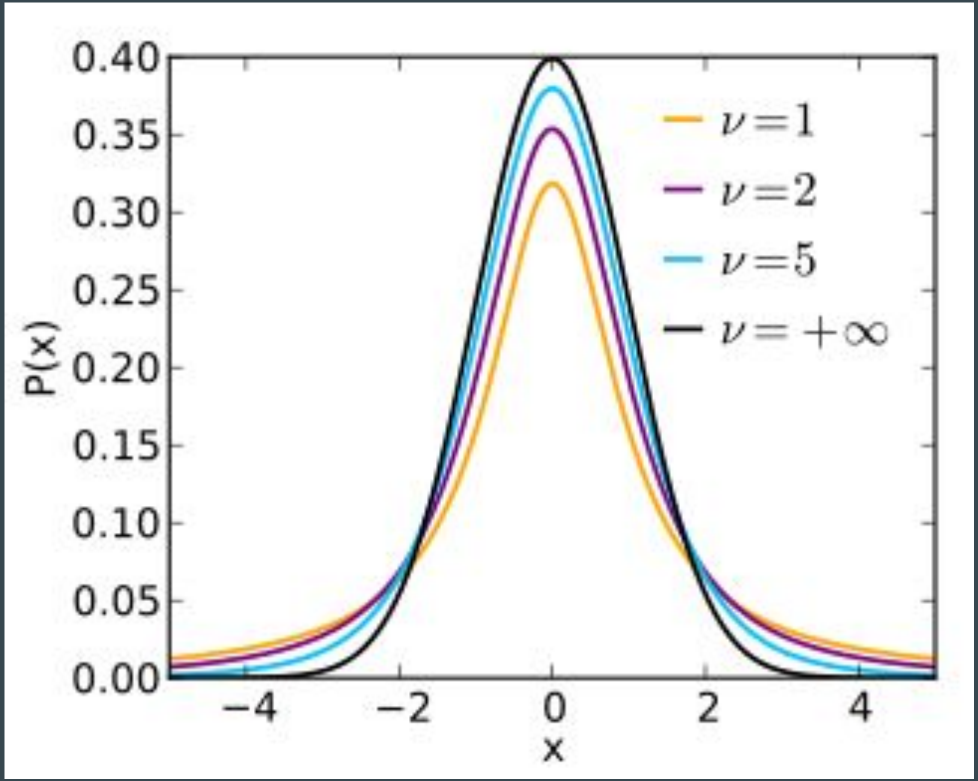


$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Student’s t distribution

- Student’s t is never used to model data directly
- Instead, it is used to compare means of different data draws (samples)
- There is one parameter
  - v: normality parameter
  - v = n-1
- In other words, as n increases, the student’s t becomes more “normal”

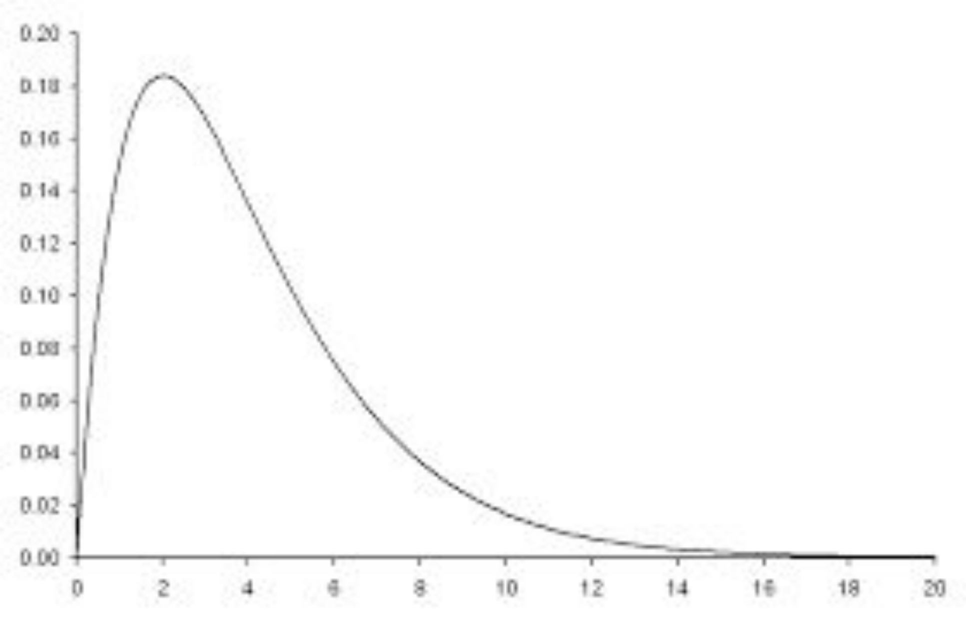
As ν increases, the distribution becomes more symmetric



Applications for the student's t

- The t-test: comparing means from two different samples
  - $E(X) = 0$
  - i.e. there is no difference between two means
- Useful when sample sizes are not very large

The Chi-squared distribution

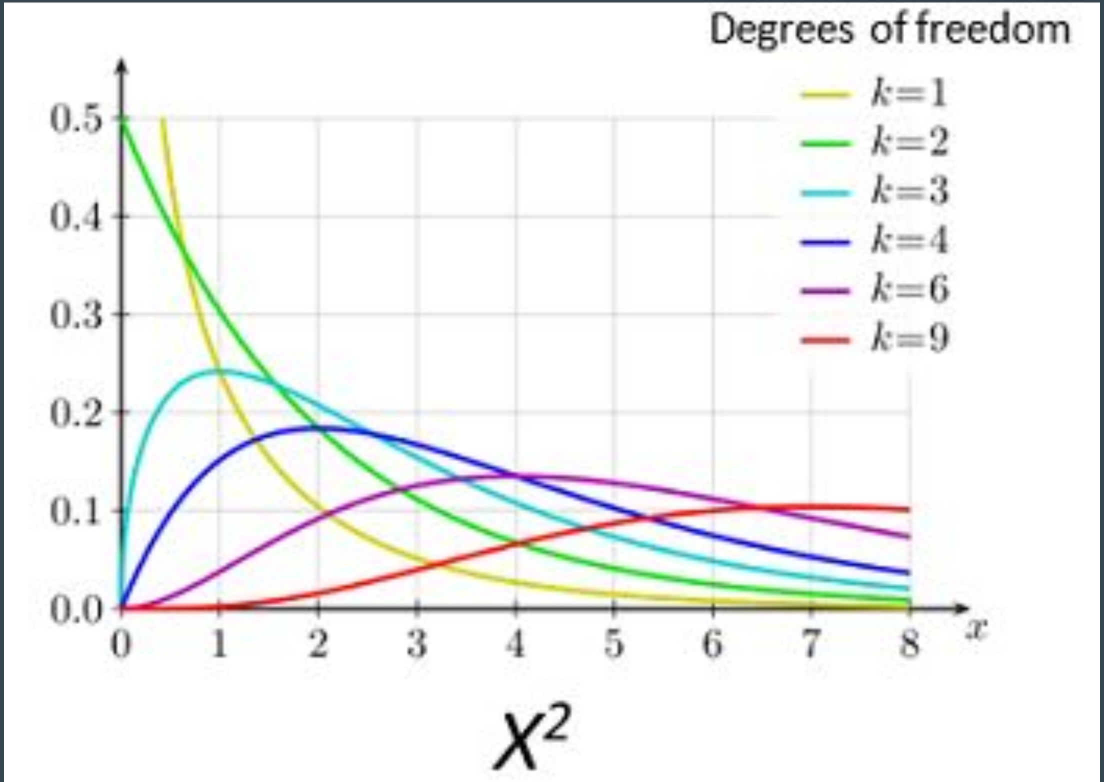


$$\frac{1}{2^{k/2}\Gamma(k/2)}\, x^{k/2-1} e^{-x/2}$$

Distribution of squared sums

- We saw that the student’s t follows a distribution of means
- Chi-squared describes any sampling distribution where the statistic is a sum of squares
  - I.e. variance
- One parameter
  - *k:* degrees of freedom

As k increases, the distribution becomes more symmetric



Applications of the chi-squared

- ANOVA
- Comparing the variances between statistical models
- $E(X) = k$
- $Var(X) = 2k$

Review

- Continuous distributions
  - Exponential: distance between successes
  - Poisson: count data
- Non-natural
  - Uniform: equal probability for all observations
- Sampling distributions
  - Student’s t: comparison of means
  - Chi-squared: comparison of variances