September 19th – Distributions

Quantitative Thinking in the Life Sciences

Today

- Distributions - probability – statistics
  - Assignment 3
  - Review to date
  - Example of where we are going
- More R fun!
  - questions and revisiting assignment code to date
  - distributions and variation

Housekeeping

- Assignment # 4 is due on Oct 3rd (two weeks) and is worth 100 points:
  - Three parts
    - Chapters 4 and 5 in R
      - Chapter 4 will be given to you today (programming)
      - Chapter 5 will be given to you next week (distributions and variability)
    - Developing and simulating distributions for your systems factors/components/variables

Assignment 3 review

- To R we go!
  - Dropbox\\Quantitative Thinking\\Sept 19 notes_AssignmentExample.R

Course overview revisited

We have discussed:

1. Thought about some of the big questions in your system
2. Looked at how probability theory links to statistics
3. Thought about some of the main variables/drivers/factors influencing your system

What is coming:

1. Think about how each of the main driver's distributions and variation inform our ability to answer our question or understand our system
2. Question how the variability in your system will influence the how, when, where and why questions of data sampling
3. Think about ways to reduce the variability to address your particular questions
4. Simulate data in your system
5. Test to see if your proposed sampling design will be able to answer your questions

Where are we going with this class?

A simplified example:

One of Josh's questions: Examining the correlation between blueberry fruit yield and percent of mycorrhizal colonization of root biomass

Examining the correlation between blueberry fruit yield and mycorrhizal colonization of root biomass

Assumptions for this example (that may realistically be … completely unrealistic):

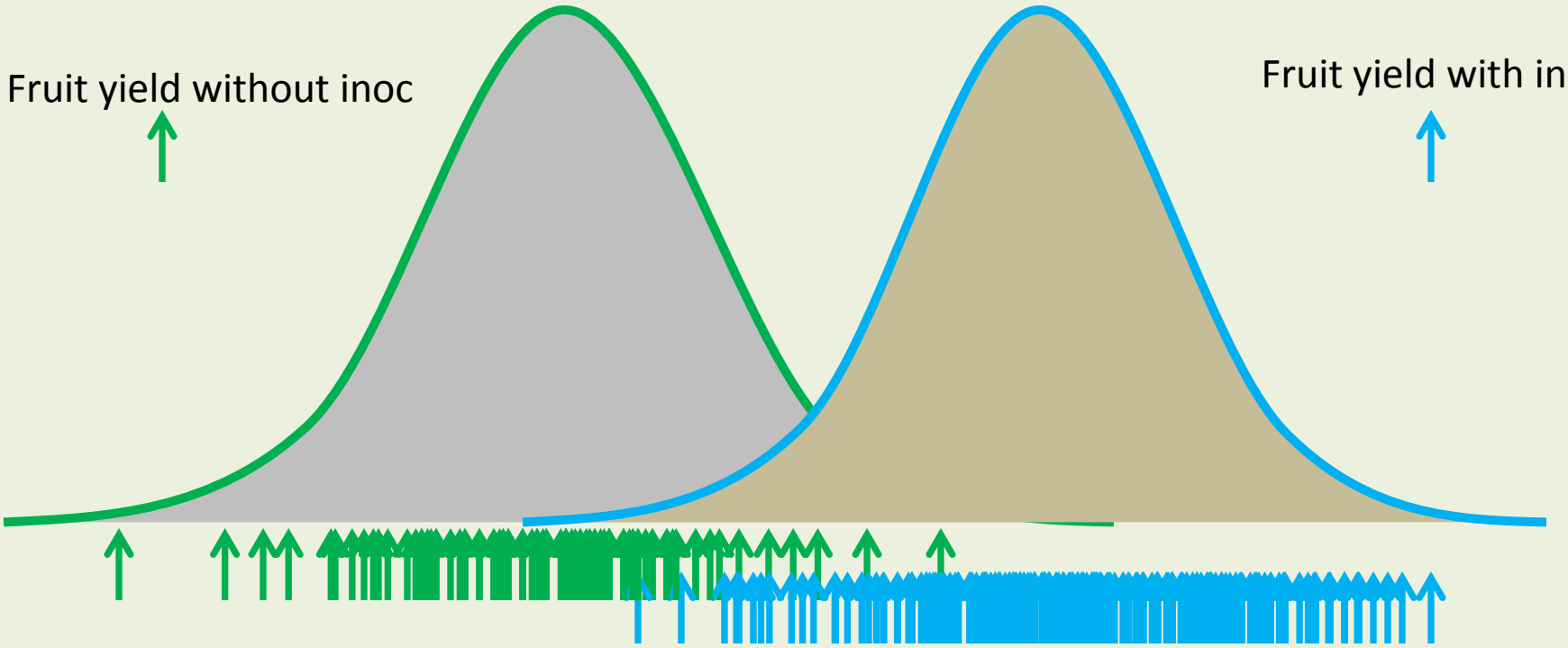
1. I know what mycorrhizal colonization is and can calculate percent colonization of root biomass given a soil/root sample
2. My sampling error is zero
3. Josh will not become upset and start throwing blueberries at me

Test 1: We want to know if there is a positive effect on blueberry fruit yield by inoculating blueberry bushes:

Frequentist approach: Test (null hypothesis) that there is no difference in fruit yield between inoculated plants and plants that were not inoculated

Examining the correlation between blueberry fruit yield and mycorrhizal colonization of root biomass

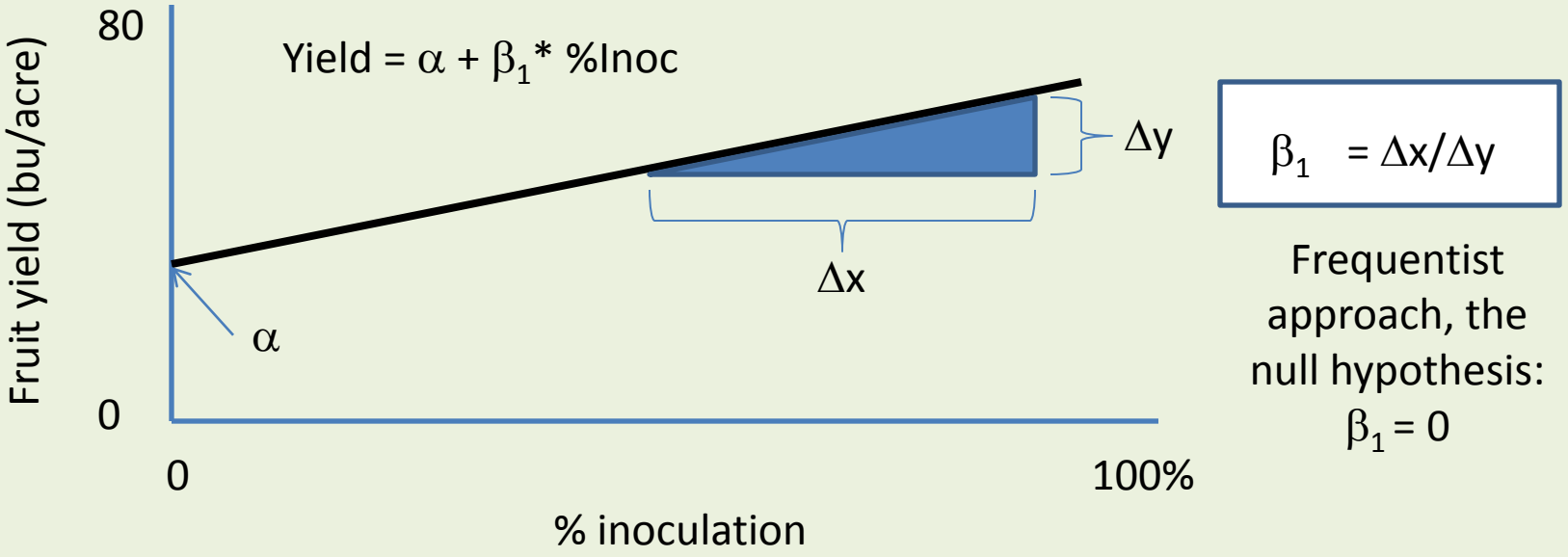
Examining the correlation between blueberry fruit yield and mycorrhizal colonization of root biomass



Test 2: We want to know if there is a positive effect of increased percent inoculation to fruit yield:

Frequentist approach: Test (null hypothesis) that there is no effect

Formula:
Yield = intercept (assumed yield with no inoculation) + effect size parameter * %Inoculation

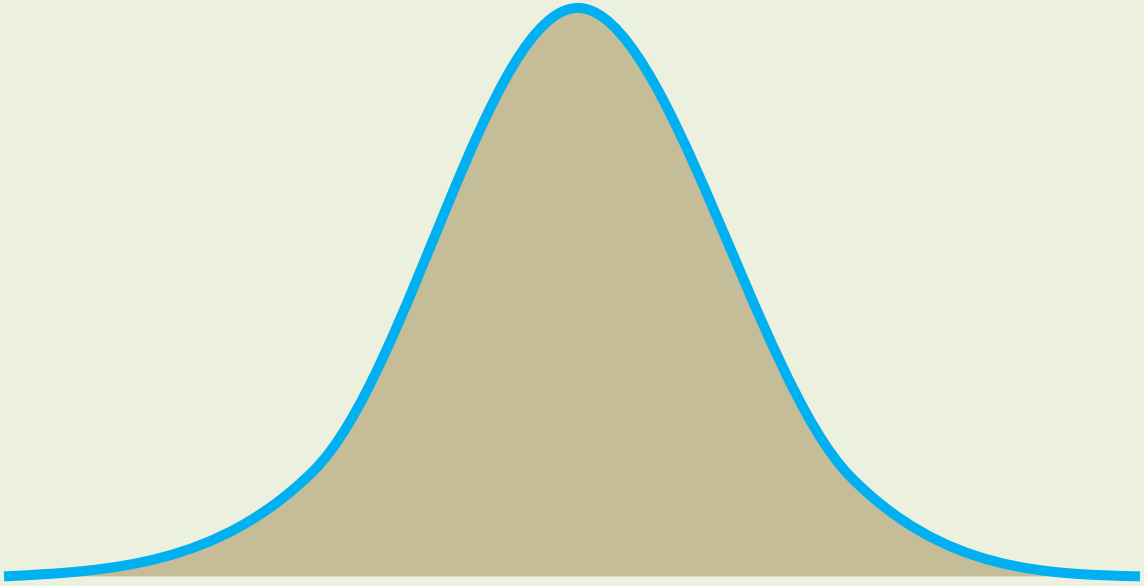


Many statistical tests assume normality (a normal distribution). Why?

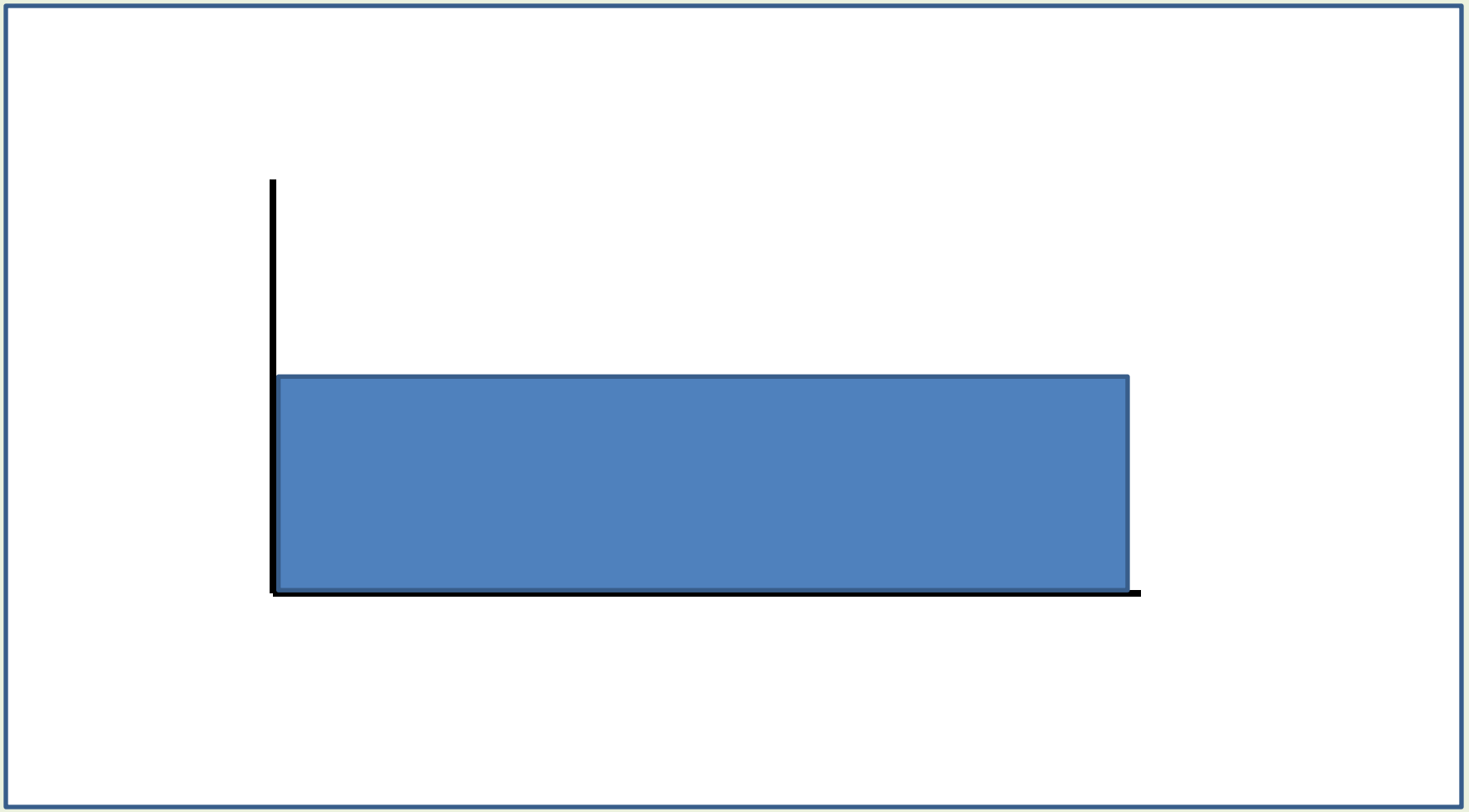
If a drunken Homer gets up, has an equal chance of stumbling 1 foot to the right, and 1 foot to the left, and stumbles (independently) 100 times, where is he likely to end up compared to his chair (assuming there are no walls and no nearby fridge full of Duff beer.

Drunken walk in R

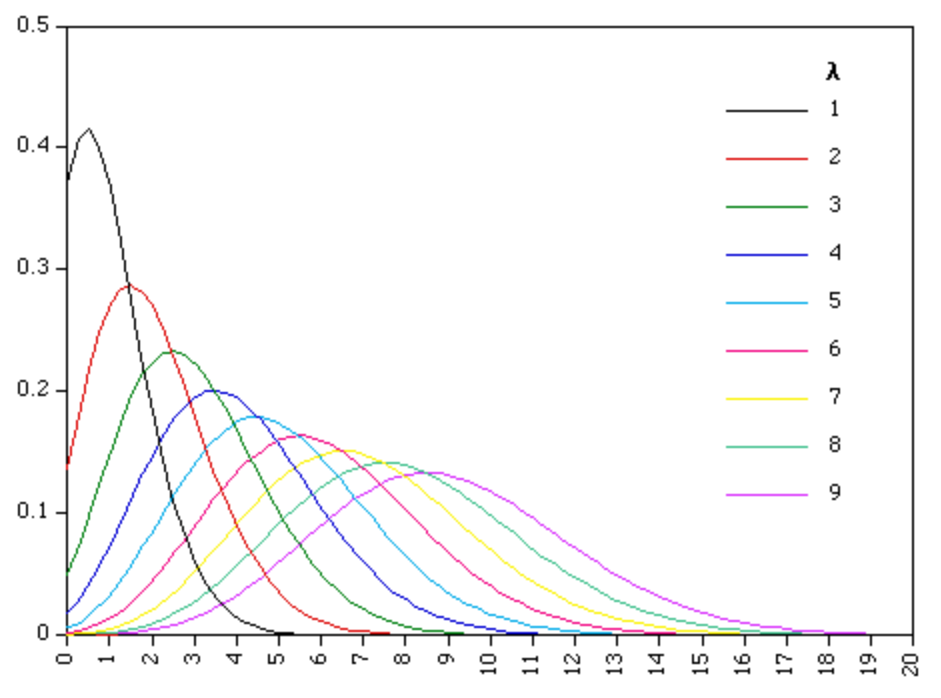
Normal probability distribution function



Uniform probability distribution function (equal probability)



Other distributions: Poisson



Other distributions: Binomial and Bernoulli

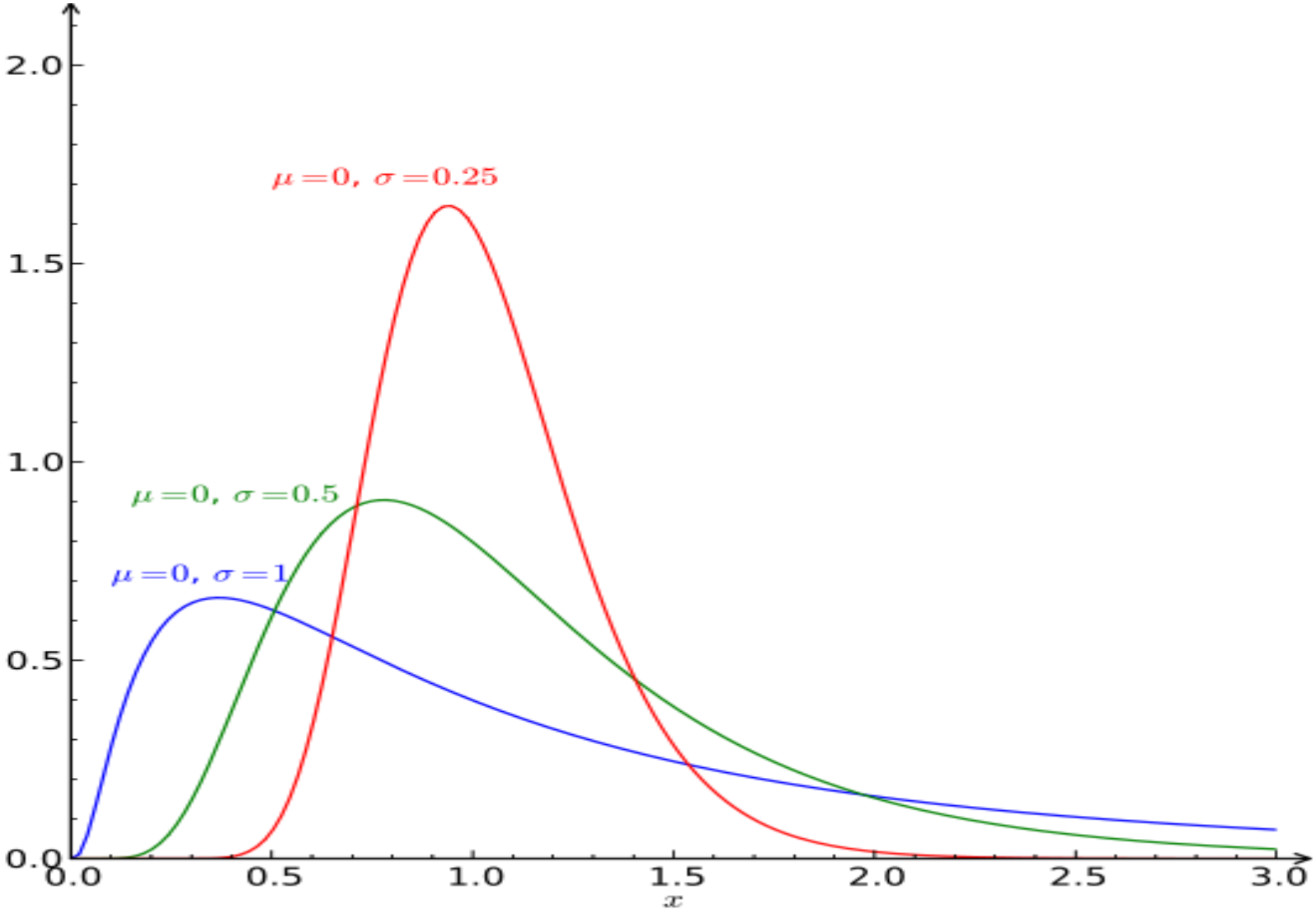
Coin flips or weighted coin flips
Bernoulli is frequently used in occupancy studies (occupied or not)

Binomial simulation = rbinom(n, size, prob)
Bernoulli simulation = rbinom(n,size=1,prob)

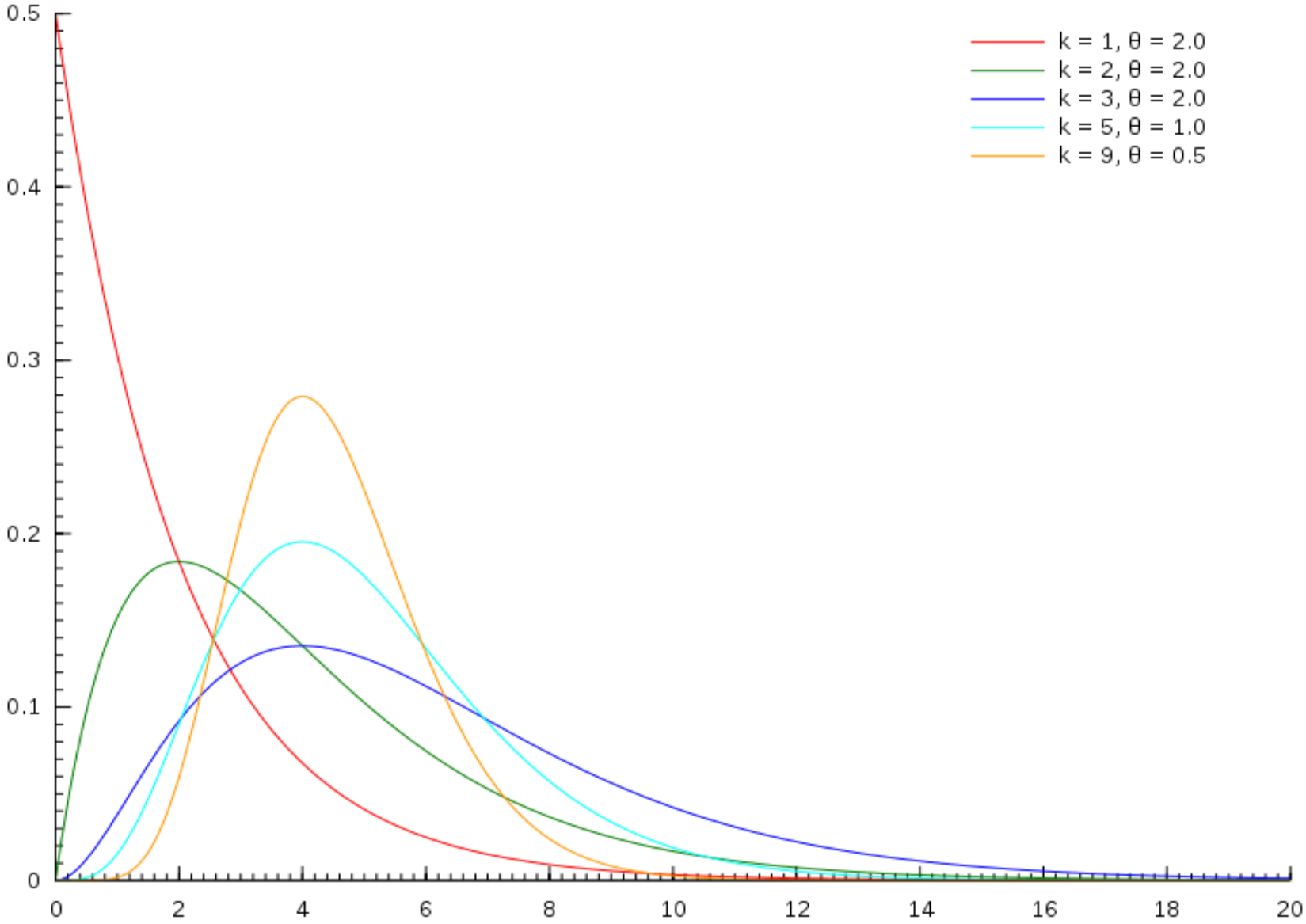
Other distributions: Multinomial

weighted dice

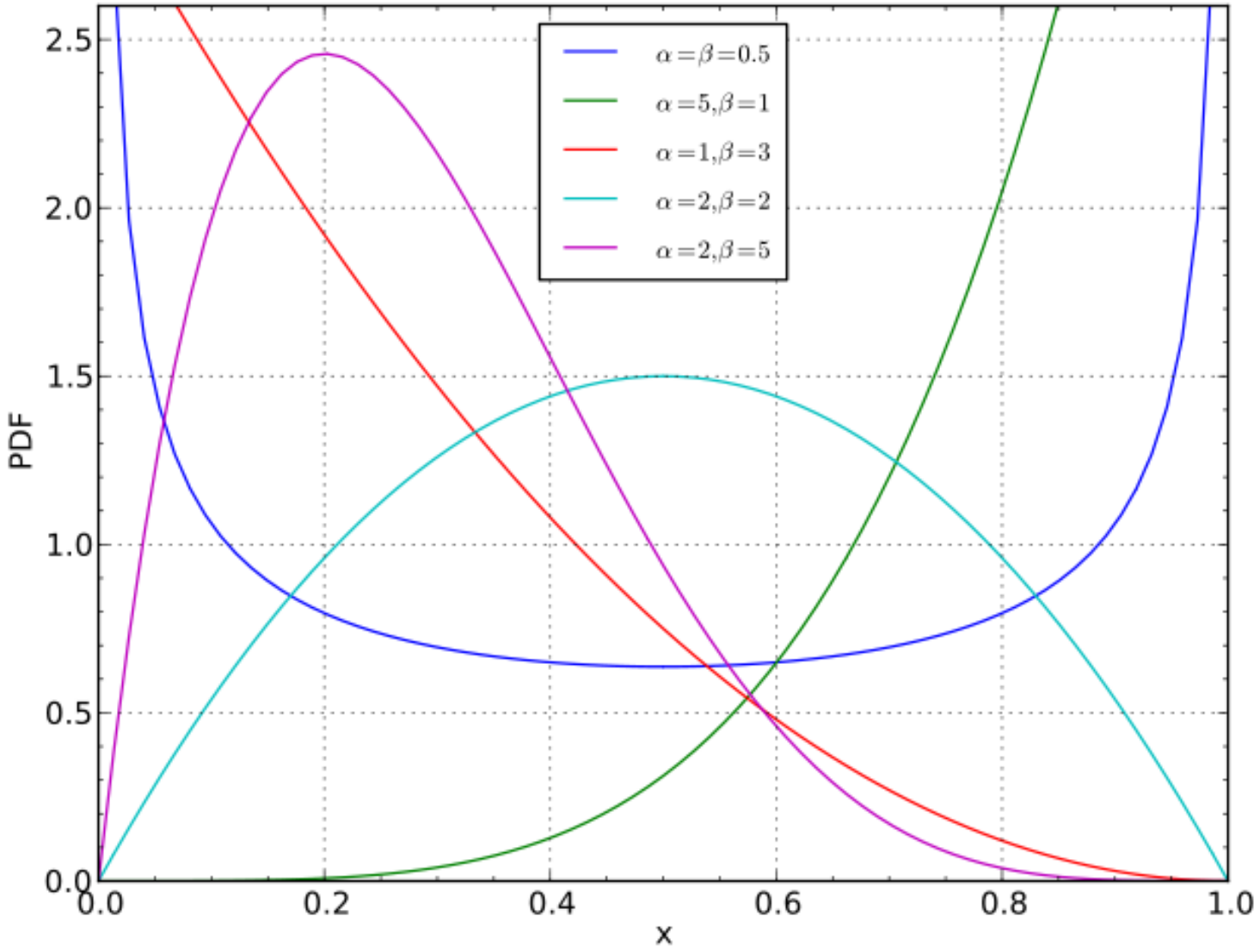
Lognormal Distribution: probability distribution function
(Always positive, its logarithm is normally distributed)



Gamma Distribution: probability distribution function
(k and θ are shape and scale parameters)



Beta distribution: probability distribution function
(α and β are shape parameters)



Getting started on your assignment

- Read pgs 36-48 in your book
- Start thinking about some of the ways your data might be distributed. Is it likely that most of your data are zeros? Will you have data that are skewed?

Endless fun with R!

- Questions from last week?
- This week – more programming!
- rm(list = ls(all=TRUE)) for your R template