Probability distributions

Today's agenda:

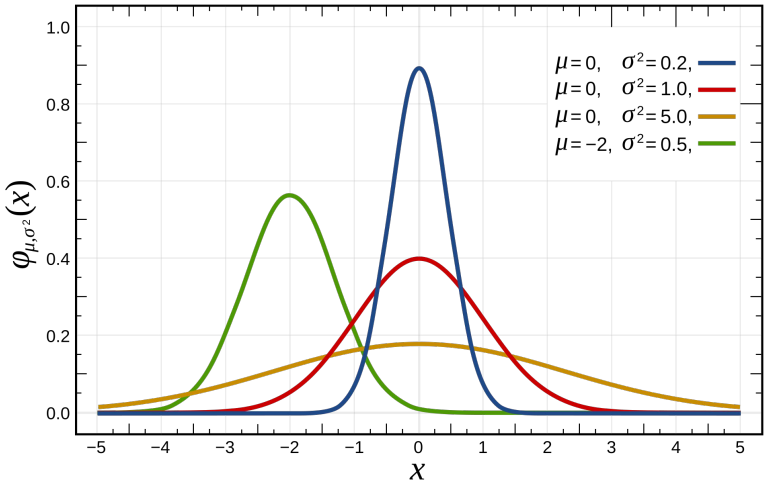
- Assignment 2
- Types of probability distributions
    - How to choose?
    - Understanding the parameters

Assignment 2

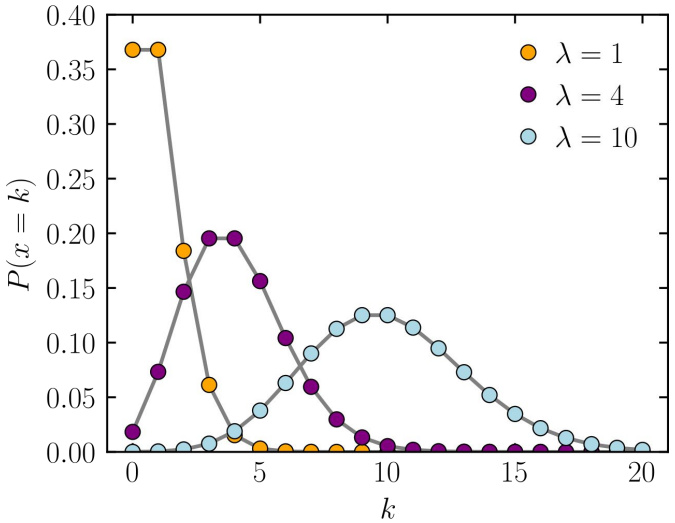
- Due October 3rd before midnight
- Focused on identifying different types of distributions
- 2 parts
  - 1) A list of different distributions to identify
  - 2) Identifying the distributions of variables of interest in your own dataset
- As before submit as .R, or .Rmd output
  - For .R files, the answers to #1 can be included as comments, for .Rmd as text.

Statistical Distributions

- Describe probabilities of different events
- Certain types of distributions have mathematical formulas describing them



Normal



Poisson

Images: Wikipedia

Picking a distribution

- Your book lists 11 different distribution types in Table 4.1
- There are many others
- How do you choose?

Picking a distribution

- Eliminate possibilities
  - Discrete vs continuous
  - Range of values (e.g., 0-1, 0 to infinity, etc.)
  - Table 4.1 is a good place to start!

- Compare the distribution’s shape to your data

- Look at what other people have done/said (including your book)

- Be able to justify your choice!

Experimental plots of distributions

All distributions are defined by some function, e.g., the normal:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

This has two parameters that control what it looks like:

μ = mean

σ = standard deviation

Experimental plots of distributions

Hard for you to understand from the equation?

Use R instead!

Experimental plots of distributions

Distributions all have functions in R (listed in 4.5)

- Help documentation tells you the parameters
- You can plot distributions with different parameter values
- Playing around like this helps understand what parameters do!

Playing with Normal distributions

First, generate a vector of values to calculate the density for

```
x_vector <- seq(from=-10, to=10, by=0.1)
```

Then, get the density using dnorm()

```
norm_density_vector <- dnorm(x = x_vector, mean = 0, sd = 1)
```

Then plot

```
plot(norm_density_vector ~ x_vector)
```

Playing with Normal distributions

You can make it easier to update by combining functions:

```
x_vector <- seq(from=-10, to=10, by=0.1)

plot(dnorm(x = x_vector, mean = 0, sd = 1) ~ x_vector)
```

Then try different values for mean and sd:

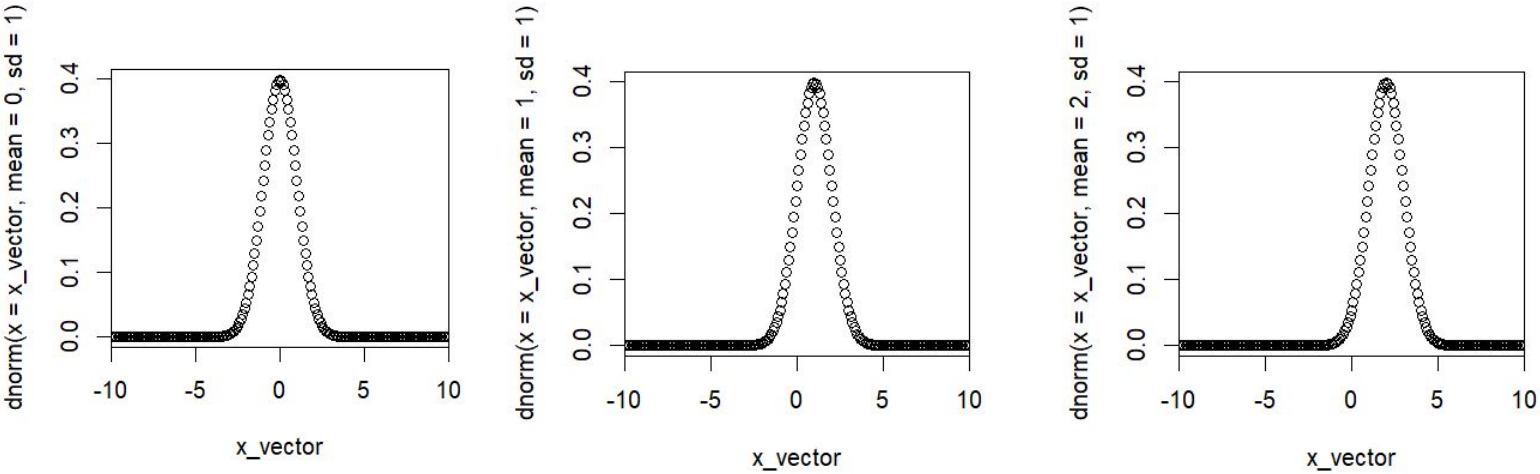
```
plot(dnorm(x = x_vector, mean = 1, sd = 1) ~ x_vector)

plot(dnorm(x = x_vector, mean = 3, sd = 1) ~ x_vector)
```

Try this out for a few values of mean and a few of sd.

What do they do?

Playing with Normal distributions



- From this, we can see how changing the mean shifts things over, but doesn't change the spread
- Trivial example, but this works for more confusing functions

Playing with the binomial distribution

```
discrete_x_vector <- seq(from=0,to=20,by=1)

binom_density_vector <- dbinom(x = discrete_x_vector,

                               size = 0, prob = 0.1)

plot(dbinom(x = discrete_x_vector, size = 10, prob = 0.1) ~
discrete_x_vector, ylim = c(0,1) )
```

Try playing with the size and prob arguments. What do they do?

Goal for today

- Play with the different probability distributions in the book
- Get a sense of what the different arguments do

Remainder of class:

- Play with distributions to understand their arguments

Before next class:

- Read 5.1 and 5.1