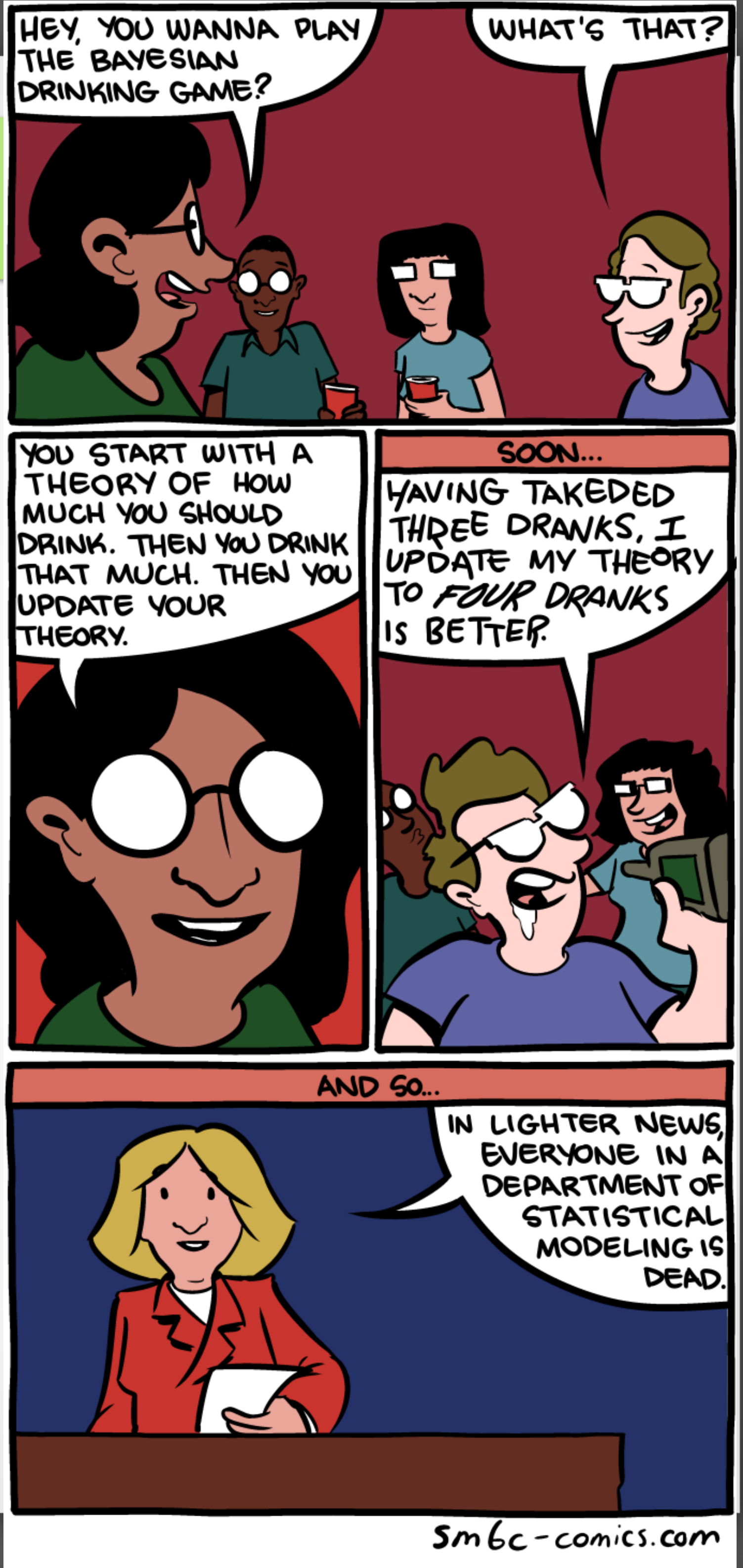
STATS 331, Lecture 2

Introduction to
Bayesian Statistics

Semester 2, 2016



Maths, Probability, and **R**

- To understand and use Bayesian statistics, we will require some mathematics and some **R** programming skills

- In this lecture we will review some of the concepts that we will need

- If you are a bit rusty, there will be plenty of opportunity to brush up. If you are already a pro, great!

Probability

Probability Theory 1

- If A and B are two “events” (things that can either occur or not occur)
- Or two “propositions” (statements that are either true or false)
- The **product rule** gives the probability that A **and** B occur/are true

$$P(A, B) = P(A)P(B|A)$$

$$P(A, B) = P(B)P(A|B)$$

COMMA MEANS “AND”!
“|” MEANS “GIVEN”

Conditional Probability

The product rule $P(A, B) = P(A)P(B|A)$ holds *conditional* on any other statement.

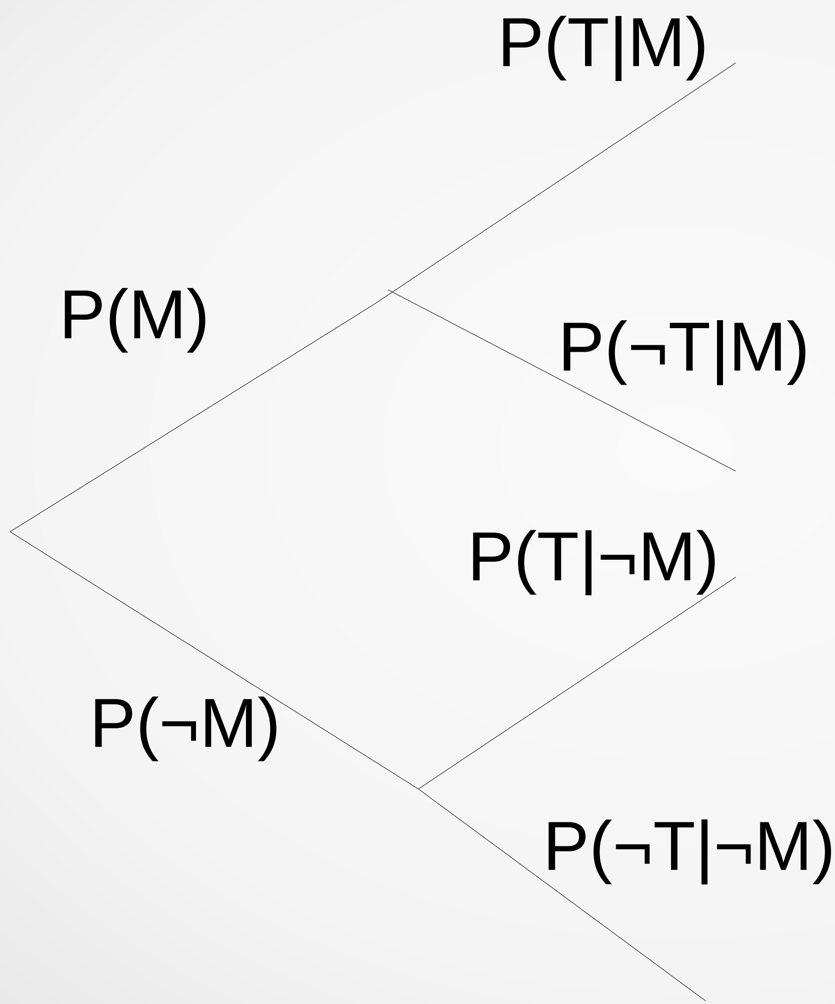
$$P(A, B \mid C) = P(A \mid C)P(B \mid A, C)$$

In Bayesian statistics, probability has a specific interpretation and is used in a specific way. More details in the next lecture!

Product Rule Question

- Suppose the probability that a person is male is 50%
- Suppose the probability that a male is taller than 6 feet is 30%
- What is the probability that a person is both male **and** taller than 6 feet?

Tree diagram from high school



Notation

P = probability
T = tall
M = male
| = given
¬ = not

Probability Theory 2

- The sum rule

$$P(A \lor B) = P(A) + P(B) - P(A, B)$$

- v means “**or**”

If A and B are mutually exclusive i.e. $P(A, B) = 0$

then $P(A \lor B) = P(A) + P(B)$

We use this a lot in 331

Conditional Probability

- As with the product rule, the sum rule also holds when a specific proposition is “given” throughout the whole equation.

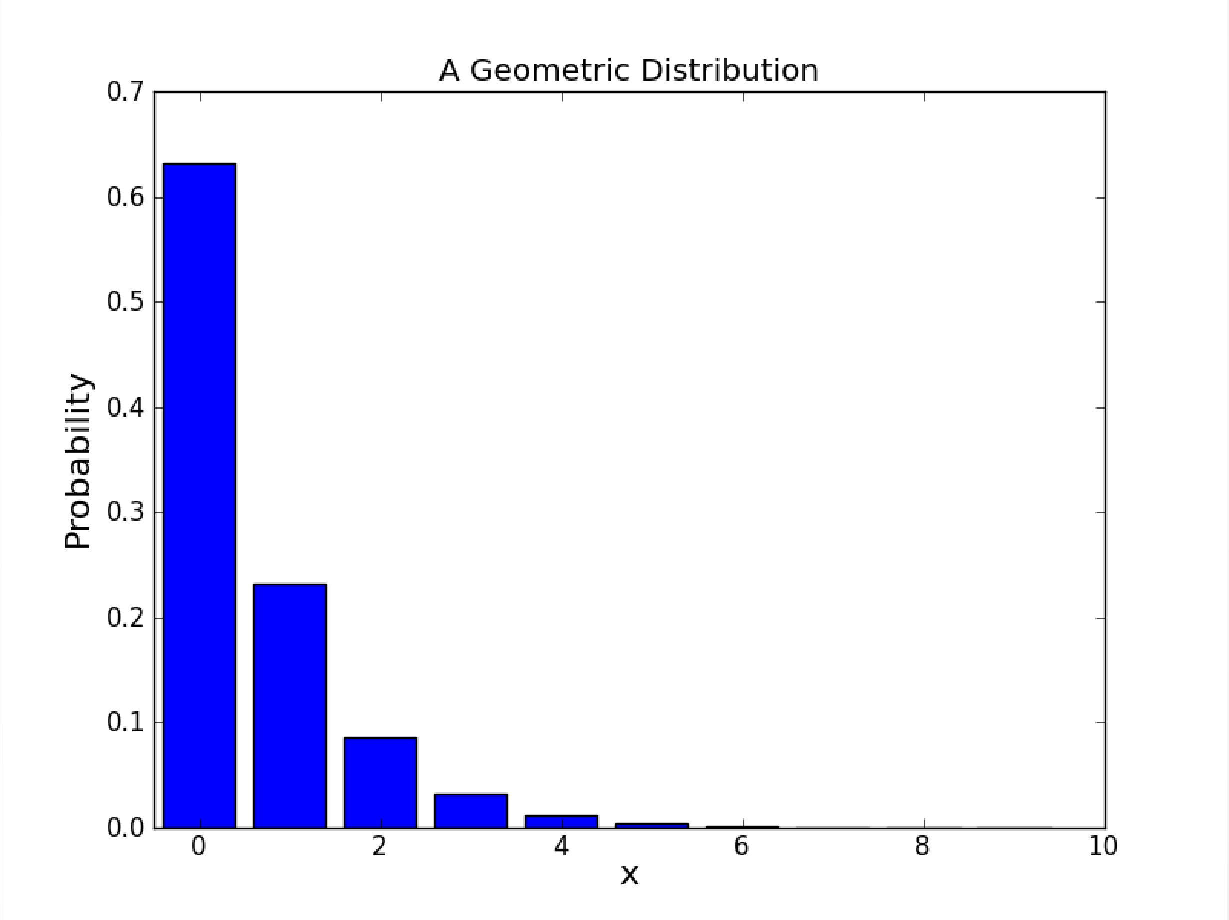
$$P(A \vee B \mid C) = P(A|C) + P(B|C) - P(A, B|C)$$

Random Variables

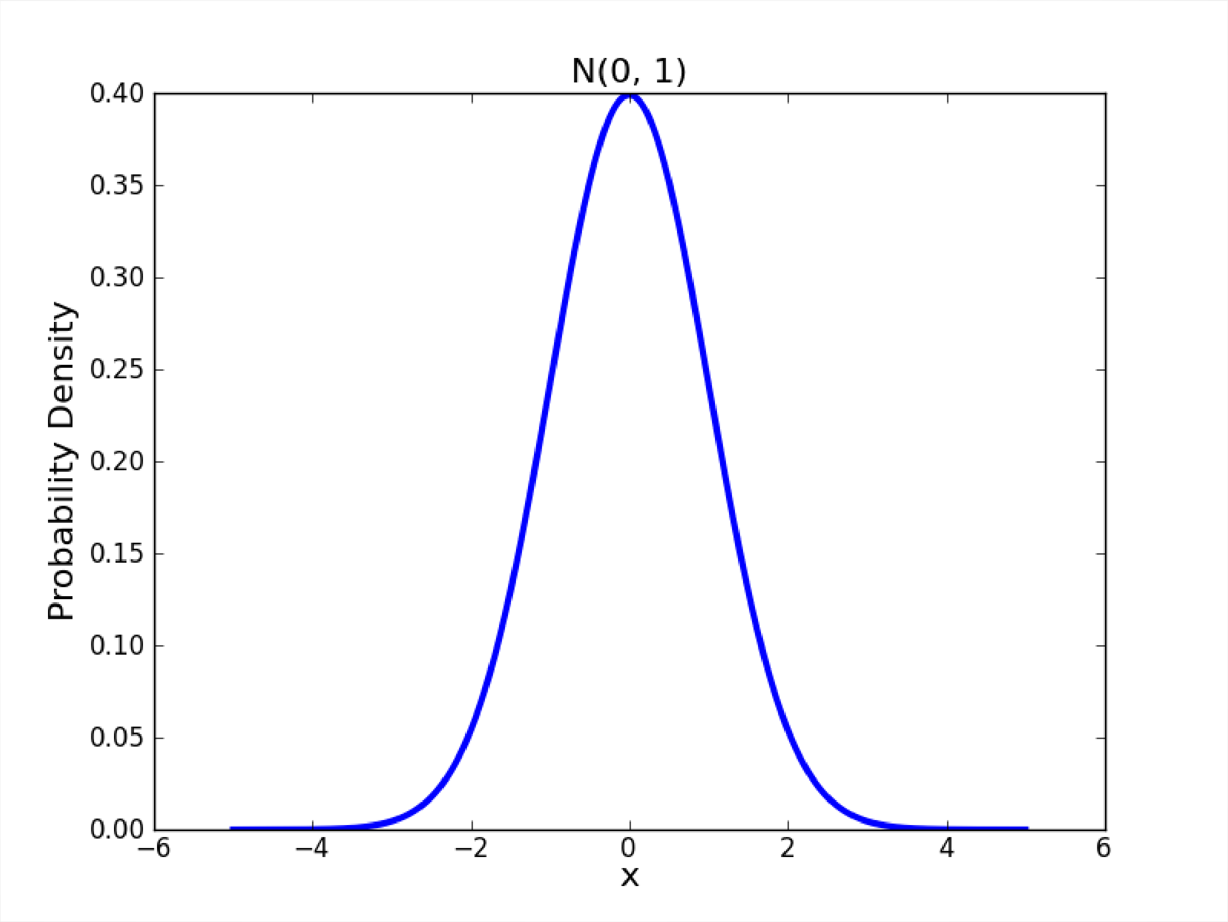
- A random variable is a quantity that has an associated **probability distribution**
- Discrete R.V.s have a probability mass function
- Continuous R.V.s have a probability density function

- Note: the word **random** has frequentist connotations – it implies variability. In Bayesian statistics, a probability distribution describes **uncertainty** about a **fixed but unknown** quantity.

Discrete RVs



Continuous RVs



Maths

Maths: Integration

The bad news: Integration happens *all the time* in Bayesian stats

The good news: You don't have to know how to do all different kinds of integrals

But you need to know what integrals mean and how they relate to probability density functions (PDFs)

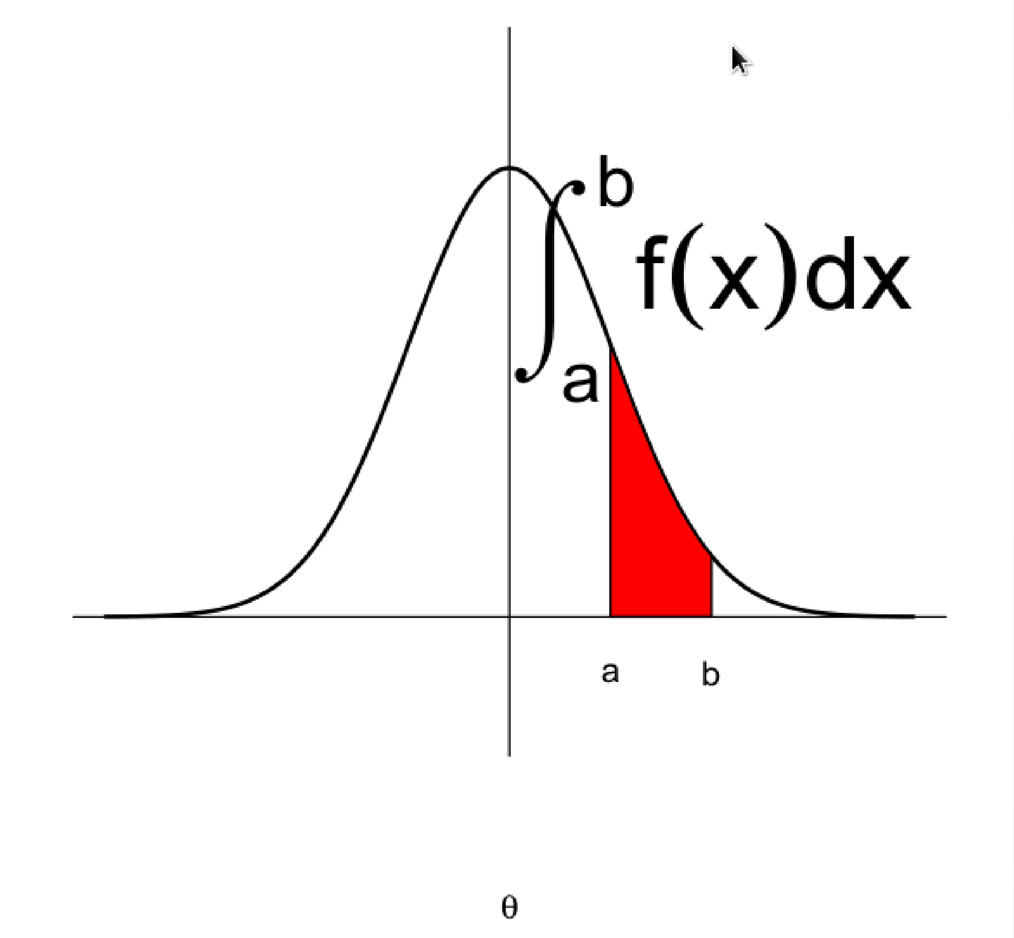
Integration

- Suppose we have a function *f(x)*
- The integral between two points a and b is the area under the curve *y=f(x)*
- Negative parts of the curve count as negative area...**but in this course we won't have anything negative because we will only ever integrate *probability density functions* (PDFs)**
- Integrals are just like sums but when you add up “an infinite number of infinitely small quantities” (really it's a limit)

Integration and PDFs

- When *f(x)* is a PDF, it is non-negative everywhere and integrates to 1
- The integral of *f(x)* between two points is the probability that the random variable is in that interval

Integration and PDFs



The probability that a quantity is between *a* and *b* is the intregral of the PDF from *a* to *b*

This is a special case of the **sum rule** of probabilities

Integration ↔ Summation Duality

- In the case of discrete parameters or data (countable number of possibilities), many formulas in Bayesian stats involve sums
- For continuous parameters or data (infinite number of possible values), **the formulas are exactly the same, just with an integral instead of a sum**!!

Continuous problem $$\int \equiv \sum$$ Discrete problem

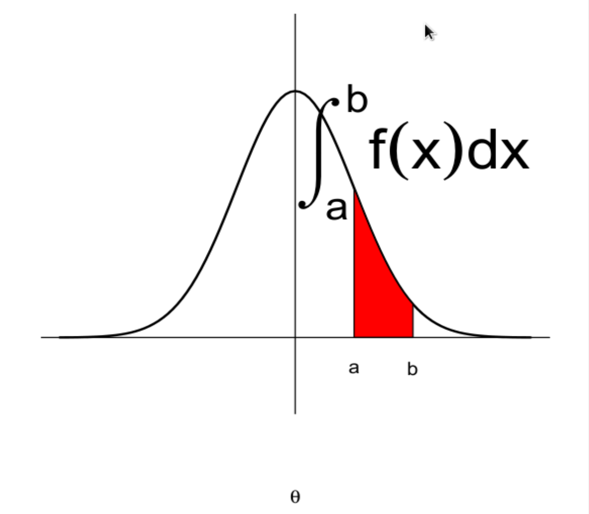
Example Question 1

- Suppose X ~ N(0, 1)
- The probability density is

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- What is the value of the following integral?

$$\int_0^{\infty} p(x)\, dx$$



Example Question 2

Y can be either 1, 2, 3, or 4,

with probabilities 0.1, 0.1, 0.3, 0.5 respectively.

What's the probability that Y is either 2 or 3?

Expected Value

- The expected value, (also called the *expectation* or sometimes just the *mean*) is defined as

$$\mathbb{E}(X) = \sum xp(x)$$
$$\mathbb{E}(X) = \int xp(x)\,dx$$

- It's a single number that tells you where the “centre” of a probability distribution is

Variance and Standard Deviation

- Variance and standard deviation are measures of how “wide” a probability distribution is.
- Standard deviation is usually more intuitive.

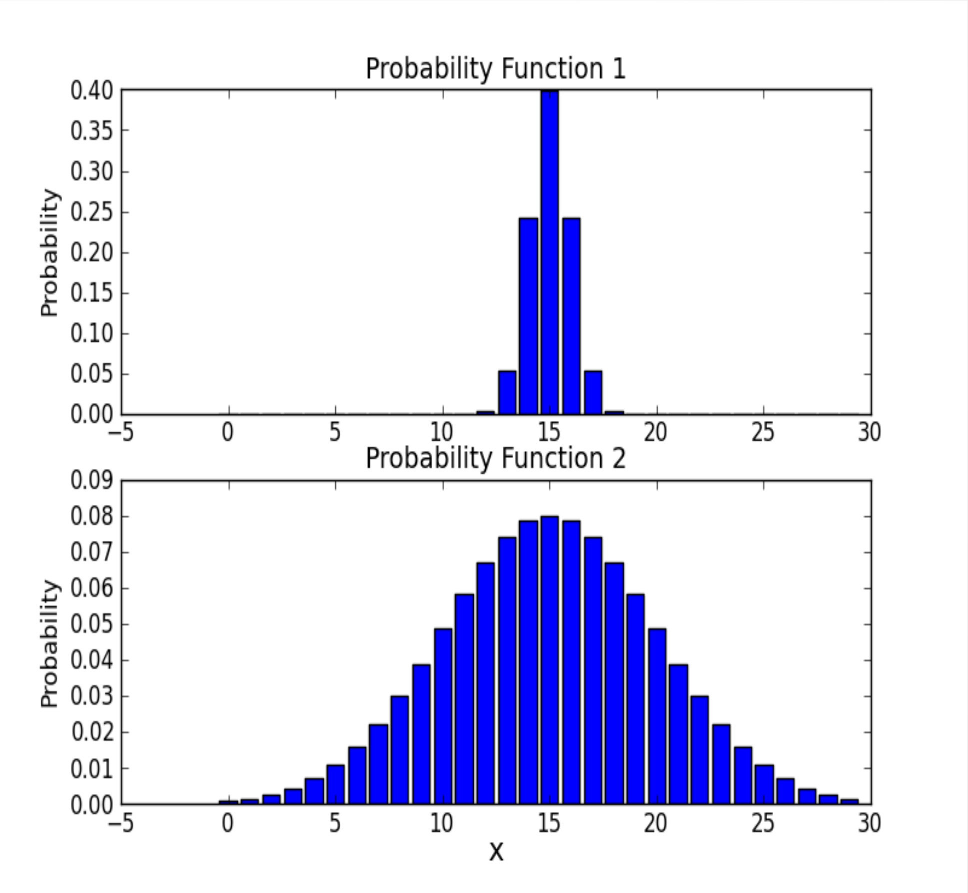
$$
\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
\mathrm{sd}(X) &= \sqrt{\mathrm{Var}(X)}
\end{aligned}
$$

Which probability mass function has a larger expected value?

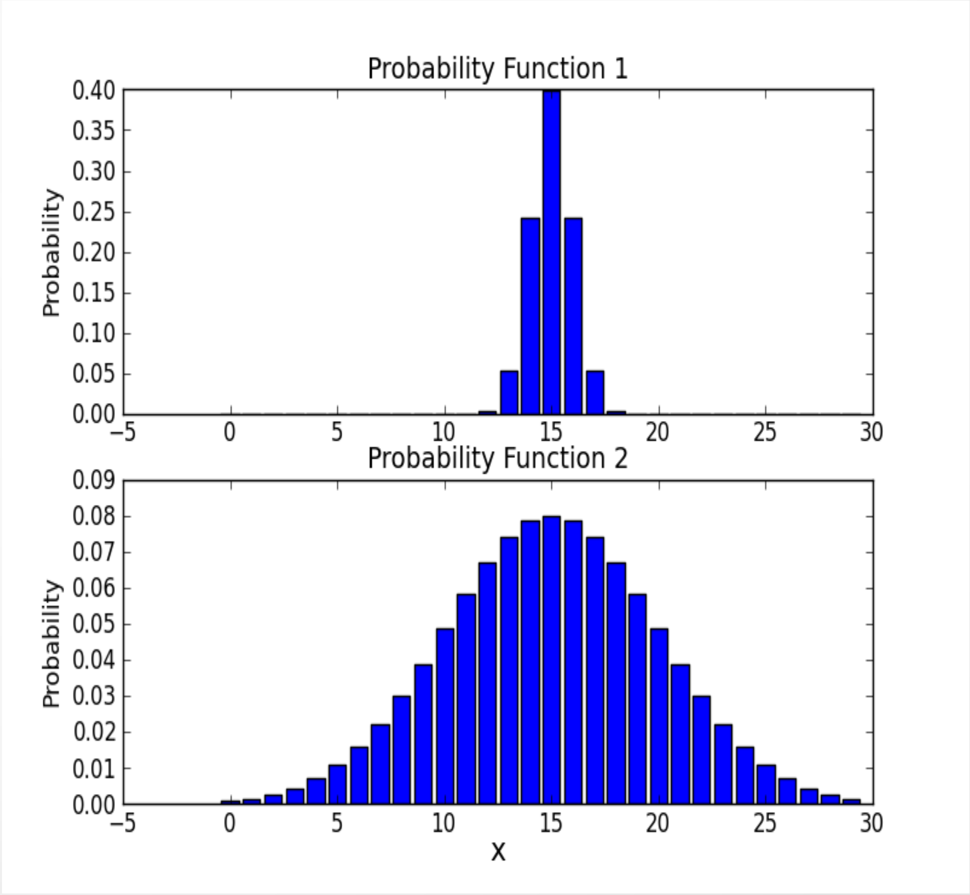
a) Probability mass function 1

b) Probability mass function 2

c) About the same



Which probability mass function has a larger standard deviation?



a) Probability Mass Function 1

b) Probability Mass Function 2

c) About the same

Log and exp

Logs and exponentials will come up from time to time. Be familiar with them.

$e^a e^b = e^{a + b}$

$(e^a)^b = e^{ab}$

$\log(ab) = \log(a) + \log(b)$

$\log(a^b) = b \log(a)$

Note: $e \sim 2.71828$ and if I write log, I mean natural log.

Basics of **R**

- Creating and using variables
- The initial > is the R console prompt, not part of the R code

```
> x = 5
> y = 3.2
> health = x + 3*y
> health
[1] 14.6
```

R Vectors

A collection of variables of the same type (for us, usually numeric floating point values)

```
> x = c(5, 3, -3.3)
> x
[1]  5.0  3.0 -3.3
```

The `c` function (`c` for *combine*) is one way of creating vectors.

R Vectors

Accessing subsets of vectors

```
> y = c(5, 3, -3.3, 7.2)
> y[3]
[1] -3.3
> y[1:2]
[1] 5 3
> y > 2
[1]  TRUE  TRUE FALSE  TRUE
> y[y > 2]
[1] 5.0 3.0 7.2
```

Useful functions related to vectors

```
> a_sequence = seq(1, 2, by=0.2)
> a_sequence
[1] 1.0 1.2 1.4 1.6 1.8 2.0
> boring = rep(2, 4)
> boring
[1] 2 2 2 2
> sum(boring)
[1] 8
> length(boring)
[1] 4
```

Probability Distributions in R

- Many probability distributions are built in to **R**
- **Two common tasks**: Generating random numbers from a distribution, evaluating the PDF.

Example with Uniform(0, 1) distribution:

```
> runif(3)
[1] 0.5155841 0.6827213 0.2204015
> dunif(c(0.3, 0.7, 1.5))
[1] 1 1 0
```

Probability Distributions in R

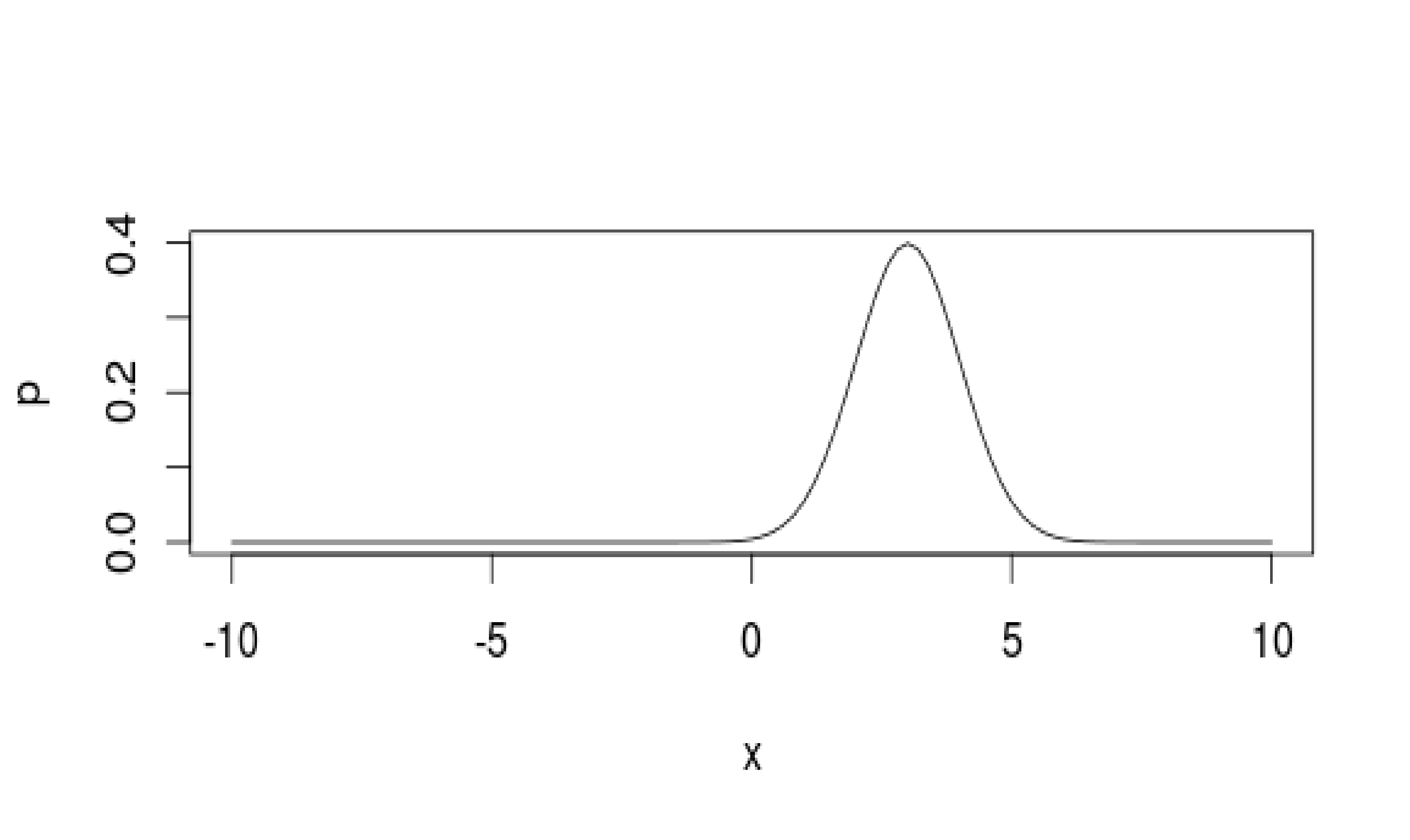
```
> rnorm(3)
[1] -0.4229854 -0.6851893 -1.1867885
> dunif(c(0.3, 0.7, 1.5))
[1] 1 1 0
> dnorm(c(1, -5, 10))
[1] 2.419707e-01 1.486720e-06
 7.694599e-23
```

Plotting Probability Distributions

- e.g. Normal(3, 1)

```
x = seq(-10, 10, length=101)
p = dnorm(x, mean=3, sd=1)
plot(x, p, type='l')
```

- Note use of optional arguments to functions.



p
0.0
0.2
0.4
-10
-5
0
5
10
x

Help on Built-in **R** Functions

- For help on the function `rnorm`, do this:

```
> ?rnorm
```

- If you don't know what function you're looking for, search with ??. eg:

```
> ??poisson
```

Recap

- Maths: integration (the concept, and simple integrals, nothing too fancy)
- Stats: Probability, random variables, expectation value, variance and standard deviation
- Working with **R** vectors
- How to use the various probability distributions that are built-in in **R**

See you next week!