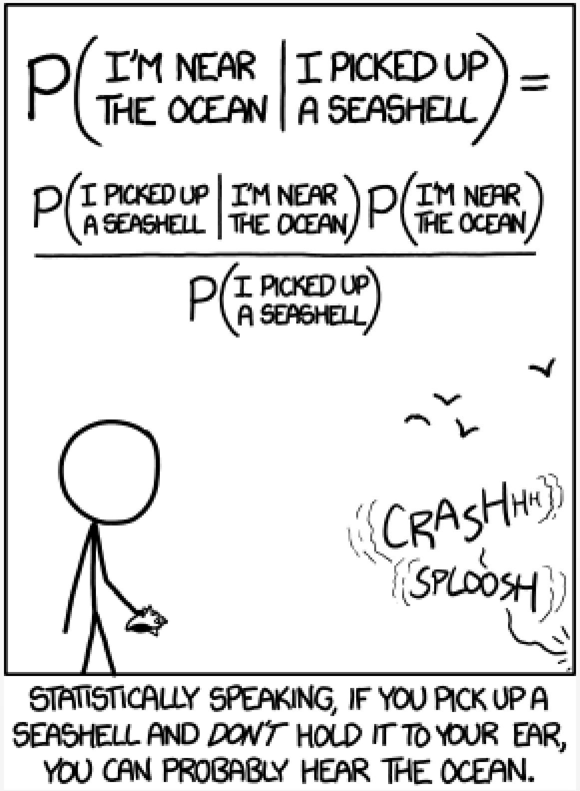
STATS 331

Image credit: Randall Monroe, xkcd



Introduction to Bayesian Statistics
Semester 2, 2016

Lecture 4

Bigger Bayes' Boxes
and
Intro to Discrete Parameter Estimation

A Medical Testing Scenario

- There is a disease *frequentitus* and your prior probability of having it is 0.01, given that you have the symptoms.
- The test is 95% accurate
- The test is also sensitive to the more common disease *wrongiosus*, and you have a 0.05 prior probability of having that.

Image is in public domain

- *D* = The test comes back positive.
- How plausible is it that you have *Frequentitus? Wrongiosus? Both? Neither?*
- *Oh look, four mutually exclusive hypotheses!*

Bayes' Box

- On document camera, with the help of R (for doing the arithmetic)

Bayes' Rule for Sets of Hypotheses

- Compare this to the phone question from the 2012 exam/the lecture notes

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{P(D)}$$

- Compare this to Q2 in Assignment 1

$$P(D) = \sum_{i=1}^{N} P(H_i)P(D|H_i).$$

- Remember, doing a Bayes' Box == Using Bayes' Rule

Discrete Parameter Estimation

Image is in Public domain

Parameter Estimation

- Parameter estimation is a *really important* part of statistics. A parameter is really just an *unknown quantity* whose value we would like to know

- We will describe our uncertainty by a **probability distribution** (discrete today, continuous later)
- The probability distribution will get **updated** (from a prior distribution to a posterior distribution) as we get more information!

Parameter Estimation

- Bayes' rule can be applied to a set of hypotheses about the value of an unknown parameter

- A hypothesis might be "$\theta = 1$"
- Another might be "$\theta = 2$"
- Suppose we observed data "$x = 3$"

Bayes' Rule Lots of Times

$$P(\theta = 1 | x = 3) = \frac{P(\theta = 1)P(x = 3 | \theta = 1)}{P(x = 3)}$$

$$P(\theta = 2 | x = 3) = \frac{P(\theta = 2)P(x = 3 | \theta = 2)}{P(x = 3)}$$

$$\cdots$$

$$P(\theta = 10 | x = 3) = \frac{P(\theta = 10)P(x = 3 | \theta = 10)}{P(x = 3)}$$

Bayes' Rule Lots of Times

$$P(\theta = 1 | x = 3) = \frac{P(\theta = 1) P(x = 3 | \theta = 1)}{P(x = 3)}$$

$$P(\theta = 2 | x = 3) = \frac{P(\theta = 2) P(x = 3 | \theta = 2)}{P(x = 3)}$$

$$\cdots$$

$$P(\theta = 10 | x = 3) = \frac{P(\theta = 10) P(x = 3 | \theta = 10)}{P(x = 3)}$$

Posterior Distribution **Prior Distribution**

Bayes' Rule Lots of Times

$$P(\theta = 1|x = 3) = \frac{P(\theta = 1)\boxed{P(x = 3|\theta = 1)}}{\boxed{P(x = 3)}}$$

$$P(\theta = 2|x = 3) = \frac{P(\theta = 2)\boxed{P(x = 3|\theta = 2)}}{\boxed{P(x = 3)}}$$

$$\dots$$

$$P(\theta = 10|x = 3) = \frac{P(\theta = 10)\boxed{P(x = 3|\theta = 10)}}{\boxed{P(x = 3)}}$$

Green are likelihoods. Orange is a **common** normalisation constant – the marginal likelihood

Bayes' Rule Lots of Times

Bayes' Rule Lots of Times == Bayes' Box

Bayes' Boxes work well for parameter estimation!

Bayes' Box for Parameter Estimation

Possible Values θ	Prior p(θ)	Likelihood p(x|θ)	Prior x Likelihood p(θ)p(x|θ)	Posterior p(θ|x)
1.0	0.0909			
1.2	0.0909			
1.4	0.0909			
1.6	0.0909			
1.8	0.0909			
2.0	0.0909			
2.2	0.0909			
2.4	0.0909			
2.6	0.0909			
2.8	0.0909			
3.0	0.0909			
Totals	1		p(x)	1

NOT SO DIFFERENT HEY?

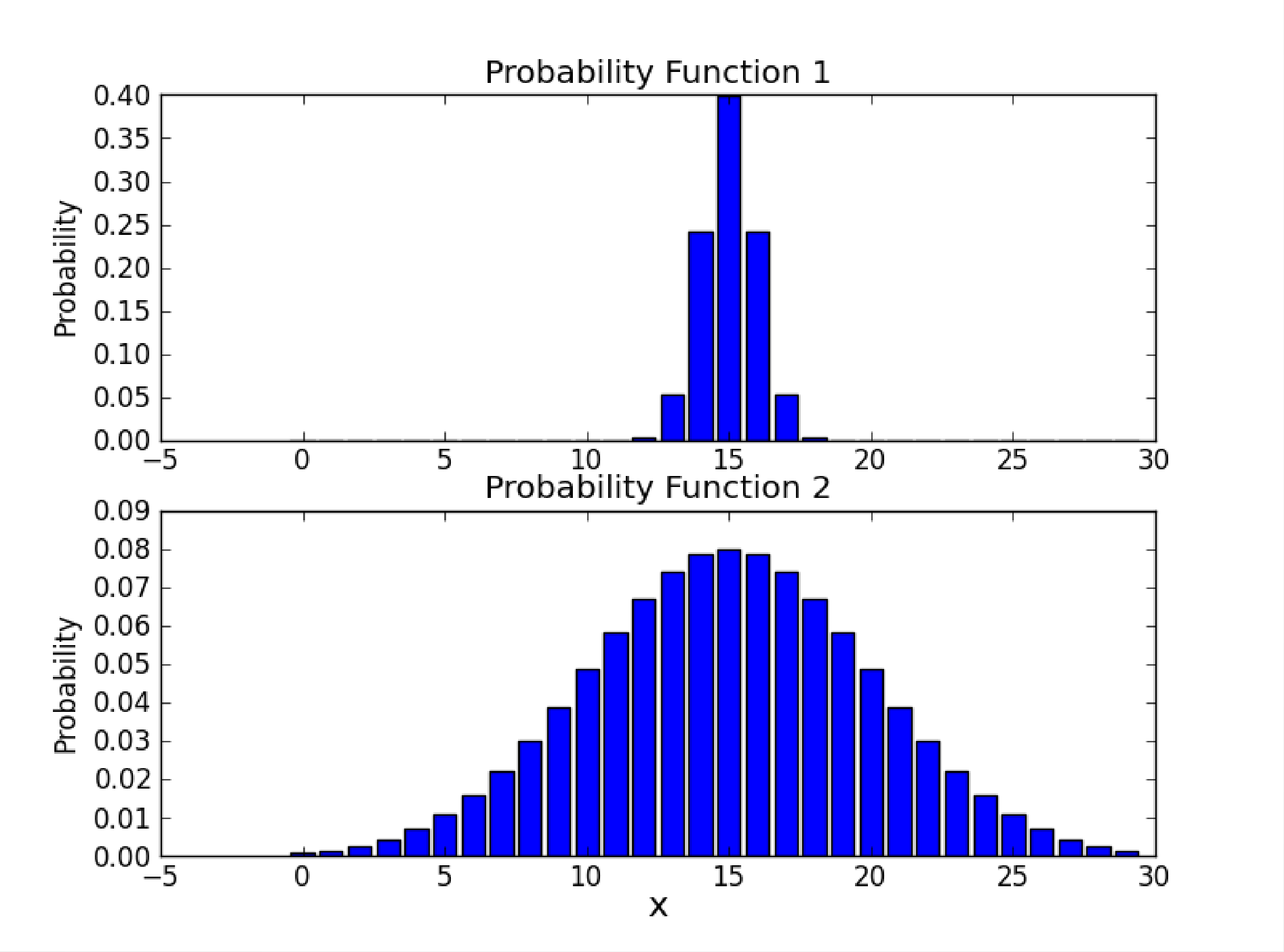
Bayes' Rule for Parameter Estimation

$$
\begin{aligned}
p(\theta|x) &= \frac{p(\theta)p(x|\theta)}{p(x)} \\
p(\theta|x) &\propto p(\theta)p(x|\theta) \\
\texttt{posterior} &\propto \texttt{prior} \times \texttt{likelihood}.
\end{aligned}
$$

This works for discrete and continuous distributions

Posterior and Prior Distribution

- We hope to achieve something like this



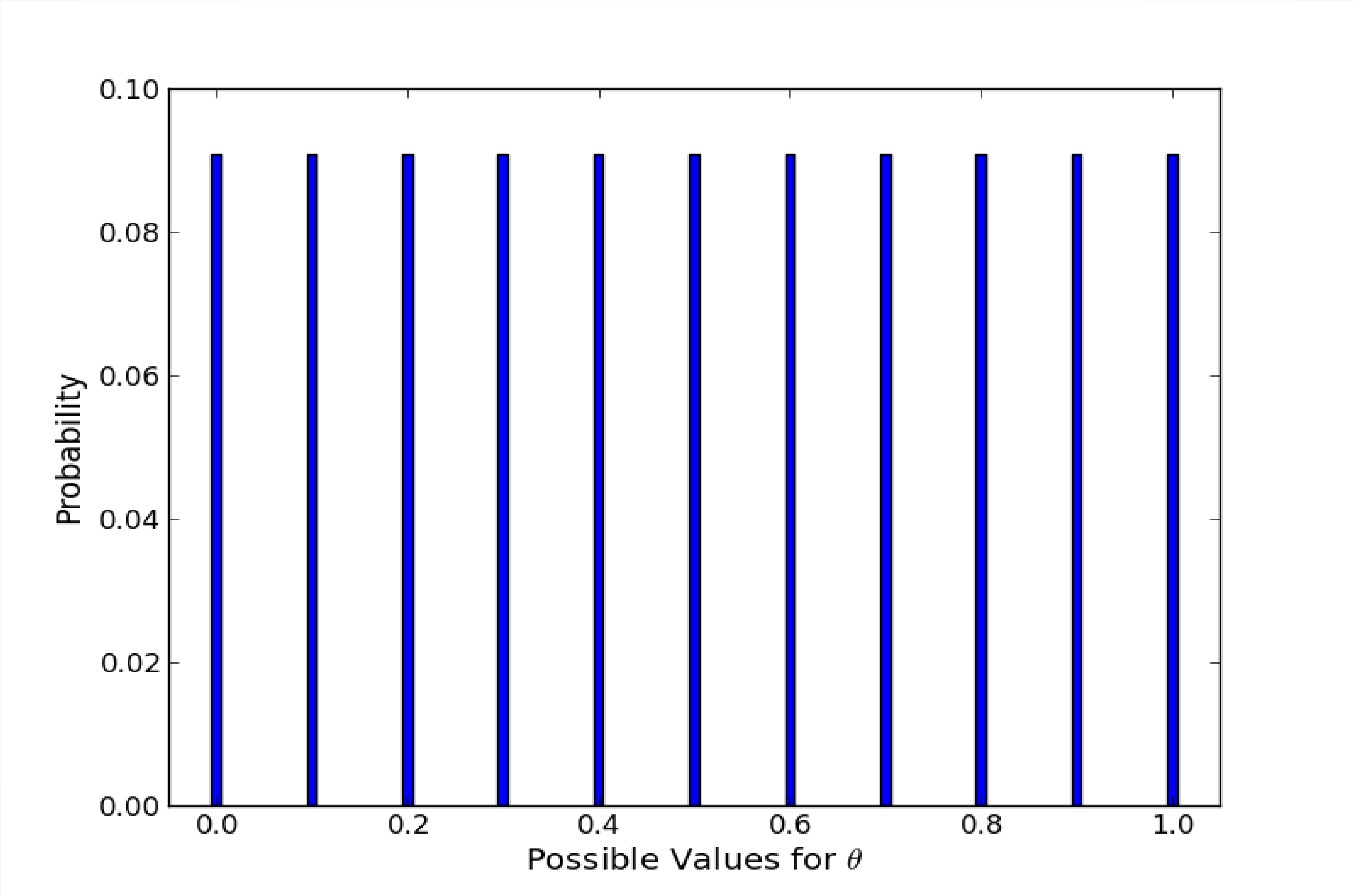
Estimating a Proportion

- We will now solve the famous problem of *estimating a proportion* using a Bayes' Box.
- Suppose there is an election coming up, with two major parties. 10 people are called to take a poll. This is a very small poll!
- Results (data): {1, 1, 0, 0, 1, 0, 1, 0, 1, 1}
- What fraction θ of the entire population support Party A?

Bayes Box: Possible Values

Possible Values θ	Prior $p(\theta)$	Likelihood $p(x\|\theta)$	Prior x Likelihood $p(\theta)p(x\|\theta)$	Posterior $p(\theta\|x)$
0	0.0909			
0.1	0.0909			
0.2	0.0909			
0.3	0.0909			
0.4	0.0909			
0.5	0.0909			
0.6	0.0909			
0.7	0.0909			
0.8	0.0909			
0.9	0.0909			
1.0	0.0909			
Totals	1		$p(x)$	1

Uniform Prior



Likelihood for Our Problem

- We called 10 people. Results are 0s and 1s (don't support or do support party A).

$P(1 \mid \theta) = \theta$

$P(0 \mid \theta) = 1-\theta$

$P(\{1,1,0\} \mid \theta) = \theta\theta(1-\theta) = \theta^2(1-\theta)$

$P(\{1, 1, 0, 0, 1, 0, 1, 0, 1, 1\}|\theta) = \theta^6(1-\theta)^4$

Likelihood in general

- The most awkward part of the Bayes' Box (for beginners) is usually the likelihood column.
- Hint: Imagine you knew the true hypothesis/parameter value, but hadn't gotten your data yet, and write down the probability of your data

Likelihood from a sampling distribution

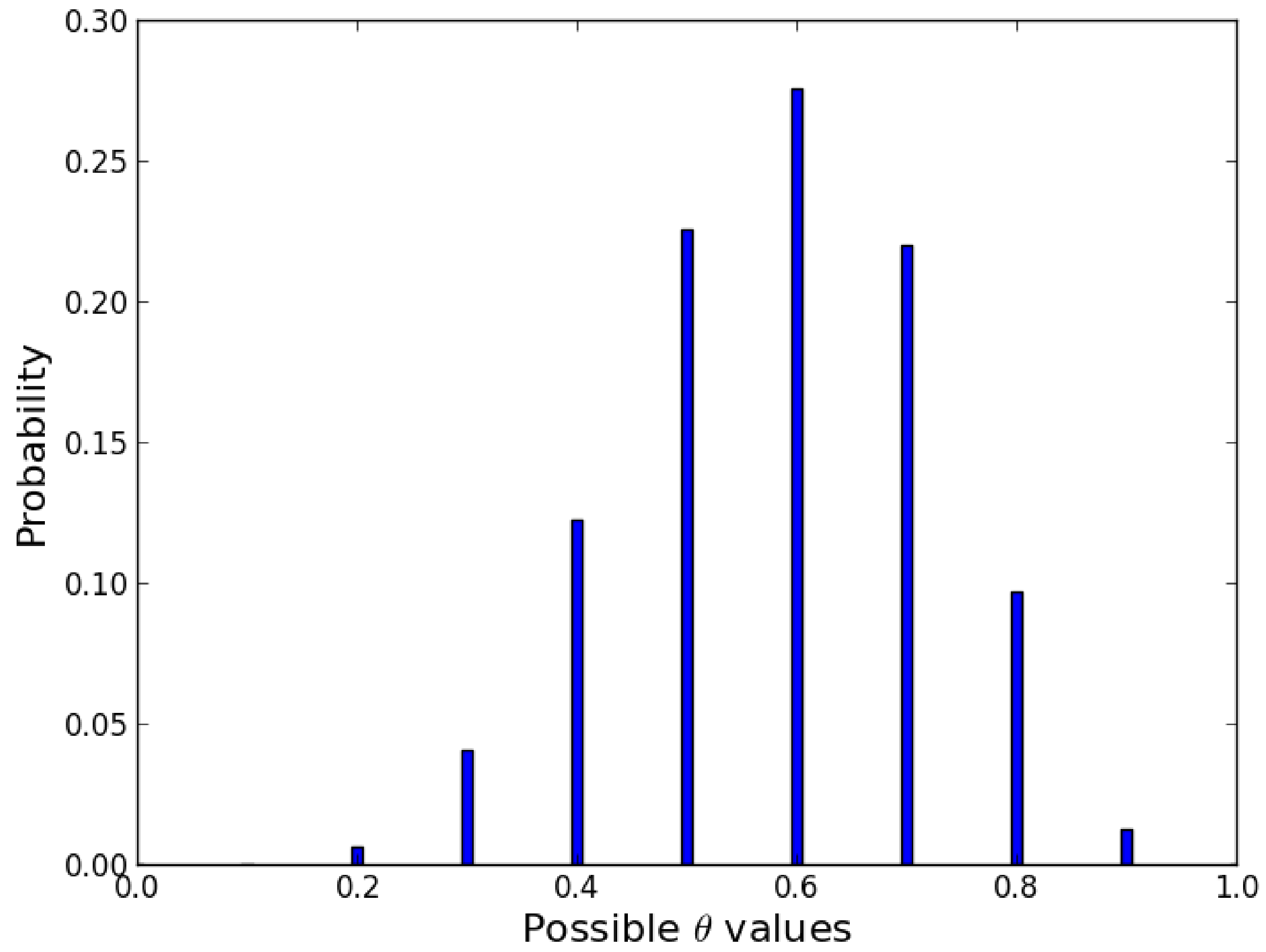
- Later, we will usually get likelihoods by first writing down a “sampling distribution” p(D|θ) – a probability distribution **for the data given the parameters**
- Some people think of this as “the distribution the data was drawn from”
- **Plug in the observed data --> then p(D|θ) tells us how the probability of the observed data varies as a function of θ. i.e. the likelihood**

Solution in R

```
theta = seq(0, 1, by=0.1)
prior = rep(1/11, 11)
lik = theta^6*(1 - theta)^4

h = prior*lik
Z = sum(h)
posterior = h/Z
```

The Posterior!



A Note on Constants in the Likelihood

- The bus problem in the notes is very closely related to this. It uses the binomial distribution for the likelihood.
- This is appropriate if you only know about the 6 successes (out of the 10 trials), rather than the full sequence of successes and failures
- HOWEVER, the only difference in the formula for the likelihood is some constants out the front that do not depend on the parameter
- THE POSTERIOR IS EXACTLY THE SAME. THIS IS THE LIKELIHOOD PRINCIPLE. Sorry for shouting.

The Posterior is the Answer

- In Bayesian stats, the full posterior distribution is the *complete* answer to a parameter estimation problem

- It describes how confident we should be about each of the possible values of the parameter

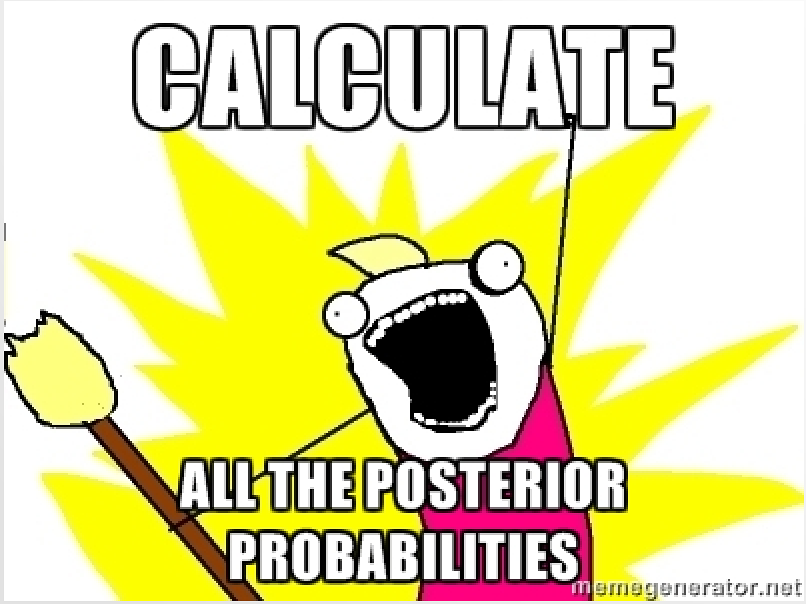


Image from Hyperbole and a Half
By Allie Brosh