

## **Foundations of Statistics – Frequentist and Bayesian**

Mary Parker, University of Texas at Austin  
and Austin Community College, 2005  
parker@math.utexas.edu

“Statistics is the science of information gathering, especially when the information arrives in little pieces instead of big ones.” – Bradley Efron

Typical Conclusions  
Frequentist

Bayesian

Estimation I have 95% confidence that the population mean is between 12.7 and 14.5 mcg/liter.

There is a 95% probability that the population mean is in the interval 136.2 g to 139.6 g.

Hypothesis  
Testing

If  $H_0$  is true, we would get a result as extreme as our data only 3.2% of the time. Since that is smaller than 5%, we would reject  $H_0$  at the 5% level. These data provide significant evidence for the alternative hypothesis.

The odds in favor of  $H_0$  against  $H_A$  are 1 to 3.

So why don't we all do Bayesian statistics?

Short answer: There are two reasons.

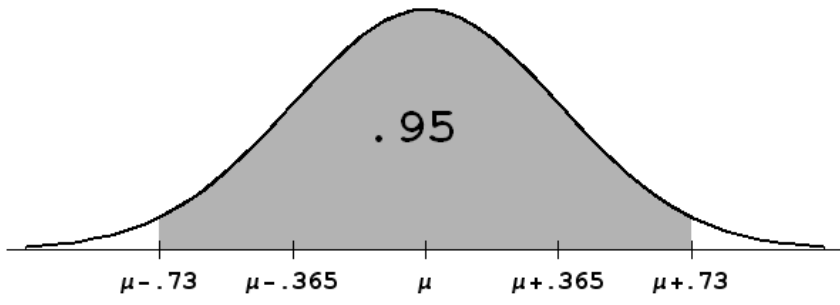
1. The calculations needed for Bayesian statistics can be overwhelming.
2. The structure requires a "prior distribution" on the parameter of interest. If you use a different prior you will obtain different results and this "lack of objectivity" makes some people uncomfortable.

## Frequentist Estimation – An Informal Introduction

**Example:** Mothers' breast milk includes calcium. Some of that calcium comes from the food the mothers eat and some comes from the mothers' bones. Researchers measured the percent change in calcium content in the spines of a random sample of 47 mothers during the three months of breast-feeding immediately after the baby was born. We want to use these data to estimate the mean (average) percent of change in the bone calcium level for the population of nursing mothers during the first three months of breast feeding.

From the data, we find that the sample mean is  $\bar{X} = -3.6$  and the standard deviation is 2.5.

Distribution of  $\bar{X}_{47}$



To obtain an interval estimate for the population mean,  $\mu$ , we make a conceptual shift (that is often rather murky for our students.)

$\mu \pm 0.73$  includes 95% of the possible  $\bar{X}_{47}$ 's  
(a probability statement about the dist'n of  $\bar{X}_{47}$  )

shifts to

one particular  $\bar{X}_{47} \pm 0.73$   
gives a 95% confidence interval for  $\mu$

The computational formula we see in our textbooks is  $\bar{X} \pm 2.0 \cdot \frac{s}{\sqrt{n}}$ .

When we plug in the values for our data, we have  
 $-3.6 \pm 2.0 \cdot 0.365 = -3.6 \pm 0.73$ .

We interpret that statement: “We have 95% confidence that the actual mean  $\mu$  of the percent change in blood calcium level for nursing mothers is in the interval between  $-4.33\%$  and  $-2.87\%$ .”

**From the “typical conclusions” on page 1, you may have noticed the following:**

- 1. The frequentist conclusions are based on  $P(\text{data}|\text{parameter})$**
- 2. The Bayesian conclusions are based on  $P(\text{parameter}|\text{data})$**

In the two methods, we observe the data in the same way, and model the distribution of the data in the same way. The differences lie in what we do with the distributions after that. We use Bayes Theorem.

## **Introduction to Bayes Theorem (The “audience participation part of this talk!”)**

Bayes Theorem enables us to “turn around” conditional probabilities.

$$P(\theta | X) = \frac{P(X | \theta) \cdot P(\theta)}{P(X)}$$

**Example:**

Enzyme immunoassay (EIA) tests are used to screen blood for the presence of antibodies to HIV, the virus that causes AIDS. Antibodies indicate the presence of the virus. The test is quite accurate – about 99% accurate. Specifically, it has 99.85% probability of detecting antibodies when they are present and 99.4% probability of a negative result when antibodies are not present. It is also known that about 1% of the overall population has the antibodies.

**When a person has a positive result on this test, how likely is it that they actually carry the antibodies?**

**Solution:** Call  $A$  the condition of having the antibodies and  $+$  and  $-$  the conditions of having a positive and negative result, respectively.

$$\begin{array}{lll} P(+|A) = & P(+|not A) = & P(A) = \\ P(-|A) = & P(-|not A) = & P(not A) = \end{array}$$

Write the question as a conditional probability.  $P(A|+)$

Use Bayes Theorem to “turn around” the probabilities.

$$P(A|+) =$$

$$P(A|+) =$$

$$P(A|+) =$$

Most people find that this is a surprising result. It's not even close to 99%, which was the (correctly) reported accuracy of the tests.

**Solution:** Call  $A$  the condition of having the antibodies and  $+$  and  $-$  the conditions of having a positive and negative result, respectively.

$$\begin{array}{lll} P(+ | A) = 0.9985 & P(+ | \text{not } A) = 0.0060 & P(A) = 0.01 \\ P(- | A) = 0.0015 & P(- | \text{not } A) = 0.9940 & P(\text{not } A) = 0.99 \end{array}$$

Write the question as a conditional probability.  $P(A|+)$

Use Bayes Theorem to “turn around” the probabilities.

$$P(A|+) = \frac{P(+|A) \cdot P(A)}{P(+)}$$

$$P(A|+) = \frac{P(+|A) \cdot P(A)}{P(+|A) \cdot P(A) + P(+|not A) \cdot P(not A)}$$

$$P(A|+) = \frac{0.9985 \cdot 0.01}{0.9985 \cdot 0.01 + 0.0060 \cdot 0.99} = 0.627$$

Most people find that this is a surprising result. It's not even close to 99%, which was the (correctly) reported accuracy of the tests.

## Bayesian Estimation – An Informal Introduction

### Example:

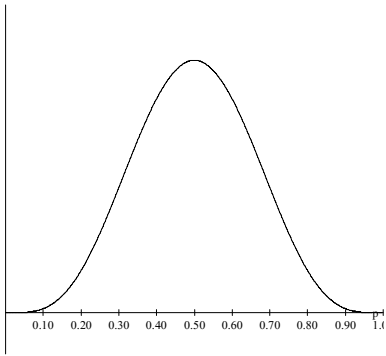
I take a coin out of my pocket and I want to estimate the probability of heads when it is tossed. I am only able to toss it 10 times. When I do that, I get seven heads. I ask three statisticians to help me decide on an estimator of  $p$ , the probability of heads for that coin.

Case 1. Sue, a frequentist statistician, used  $\hat{p} = \frac{X}{10} = 0.7$ .

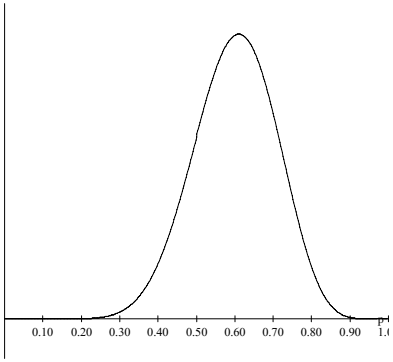
Case 2. Jose, who doesn't feel comfortable with this estimator, says that he already has an idea that  $p$  is close to 0.5, but he also wants to use the data to help estimate it. How can he blend his prior ideas and the data to get an estimate? Answer: Bayesian statistics.

Jose makes a sketch of his prior belief about  $p$ . He thinks it is very unlikely that  $p$  is 0 or 1, and quite likely that it is somewhere pretty close to 0.5. He graphs his belief.

Jose's drawing:

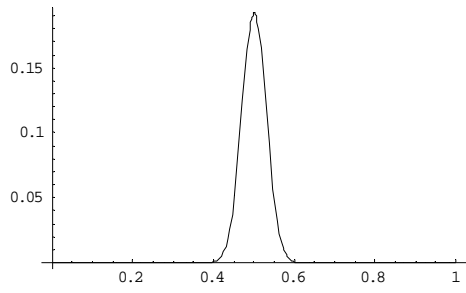


Jose's posterior distribution      Beta(12,8)



Case 3: Vicki, who is very sure that coins are unbiased, has a prior distribution like Jose's, but much narrower. There's a much higher probability on values very close to 0.5. She graphs her belief.

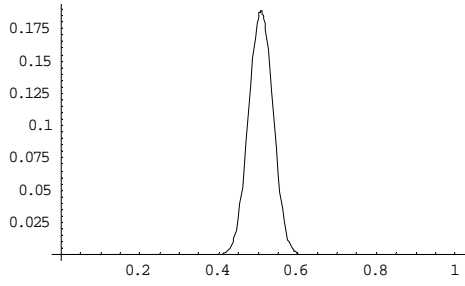
Vicki's prior distribution:



She notices that this corresponds to a particular probability distribution, which is  $\text{Beta}(138,138)$ , so that is her prior distribution of  $p$ . So the mean is 0.5, the variance is 0.0009, and the standard deviation is 0.03. Notice that her standard deviation is much smaller than Jose's.

Now, she also uses Bayes Theorem to combine the data and her prior distribution and finds that her posterior distribution of  $p$  is a  $\text{Beta}(145,141)$ . So her posterior mean is 0.507, variance is 0.0008709, and standard deviation is 0.0295.

Vicki's posterior distribution.  $\text{Beta}(145, 141)$



Jose and Vicki are both doing Bayesian estimation. Both of them decided to use the mean of the posterior distribution of the parameter as their estimator.

Summary:

Sue's estimate of the probability of heads: 0.700

Jose's estimate of the probability of heads: 0.600

Vicki's estimate of the probability of heads: 0.507

Now, Jennifer offers you a bet. You pick one of these values. She chooses one of the other two. An impartial person tosses the coin 1000 times, and get a sample proportion of heads. If that sample proportion is closer to Jennifer's value, you pay her \$25. If it is closer to yours, she pays you \$25. Which value would you choose?

## **Overview:**

What are some of the difficulties in the Bayesian approach?

1. Quantifying prior beliefs into probability distributions is not simple.
2. We might not agree with colleagues on the prior distribution.
3. Even if we can find a formula for the distribution describing our prior beliefs about the parameter, actually doing the probability calculations to find the posterior distribution using Bayes Theorem may be more complex than we can do in closed form.

The Bayesian statisticians have some answers for most of this, which will become clearer as we develop the theory.

**Mathematical Theory:**

Using Bayes Theorem, we get

$$h(p|data) = \frac{f(data|p) \cdot g(p)}{k(data)}$$

And all we actually need is

$$h(p | data) \propto f(data | p) \cdot g(p)$$

**Back to the example:**

$$\text{Here, } L(p) = f(\text{data} | p) \propto \prod_{i=1}^{10} p^{x_i} (1-p)^{(1-x_i)} = p^7 (1-p)^3$$

which is from a binomial distribution, in “unnormalized form”, i.e. without the constant multiple, considering it as a function of  $p$  and defined on  $p \in [0,1]$ .

Also, for Jose, his prior distribution is

$$g(p) = \frac{\Gamma(10)}{\Gamma(5) \cdot \Gamma(5)} p^{5-1} (1-p)^{5-1} \propto p^{5-1} (1-p)^{5-1}$$

$$\text{so } h(p|data) = f(data|p) \cdot g(p) \propto p^{7+5-1} (1-p)^{3+5-1}$$

Since the domain of  $h$  is all real numbers between 0 and 1, it is a continuous distribution, so it fits the form of a Beta distribution, with parameters 12 and 8. Thus, Jose has a posterior distribution on  $p$  which is Beta(12,8), as we said before, and is used to calculate his estimator, which is the mean of the posterior distribution on  $p$ .

We can use these posterior distributions on  $p$  to make probability statements about  $p$  as well. Typically Bayesians would compute a HPD (highest posterior density) interval with the required probability.

## **Back to the Theory:**

The major conceptual difference between Bayesian statistics and frequentist statistics is that, in Bayesian statistics, we consider the parameters to be random, so one consequence of that is that we can write probability intervals for parameters.

**Free the Parameters!**

-- Bayesian Liberation Front

**Back to the Overview:** (How they chose their prior distributions)

- Conjugate prior distributions.
- Choosing the parameters of those distributions from your prior knowledge of mean and standard deviation.

Notice that, when we look at the forms of the data distribution and the prior distribution, they are similar, and so Jose's prior distribution was the equivalent to starting your problem with 4 heads and 4 tails. Vicki's was equivalent to starting with 137 heads and 137 tails.

**Exercise:**

Your own prior? Calculate your posterior distribution.

Absolutely sure it is 0.5? Use a “point prior.” (Why is there no computation to do?)

No prior opinion – use a “flat prior” here.

## **Why aren't we all Bayesians?**

Subjectivity is a bad word in some situations.

(When is it not a bad word? When you're trying to figure out how to save money. The insurance companies were using Bayesian analyses a long time ago – before they were very easy to compute!)

## **How do Bayesians deal with these issues?**

Objective Bayesian analysis  
Subjective Bayesian analysis  
Robust Bayesian analysis  
Empirical Bayesian analysis

**Computation:**

Conjugate priors make the math easy, but what if our prior information doesn't fit that?

Aren't computers wonderful!?!?!?

## **What does the statistics community say?**

Learn to use both frequentist and Bayesian techniques and chose what is appropriate for the particular problem.

## **Additional Resources:**

These are generally organized, within categories, by the order in which I think you might want to pursue them, with the easiest first! (Go to the website listed at the top of your handout for links to the URLs below.)

### **Overview of Bayesian statistics:**

Stangl, Dalene, *An Introduction to Bayesian Methods for Teachers in Secondary Education*. 1999.

<http://www.isds.duke.edu/~dalene/talks/ncssm/sld001.htm>

Yudkowsky, E. *An Intuitive Explanation of Bayesian Reasoning*.

<http://yudkowsky.net/bayes/bayes.html>

### **Teaching Statistics:**

Consortium for the Advancement of Undergraduate Statistics Education:

<http://www.causeweb.org/> (many resources)