A Brief Introduction to Bayesian Statistics

Floyd Bullard
The North Carolina School of Science and Mathematics NCTM 2001

> *"According to its designers, Apollo was 'three nines' reliable: The odds of an astronaut's survival were 0.999, or only one chance in a thousand of being killed. [Astronaut William A.] Anders, for one, didn't believe it; the odds couldn't possibly be that good. Soon after he found out about the lunar mission he took time to ponder his chances of coming back from Apollo 8, and he made a mental tabulation of risk and reward in an effort to come to terms with what he was about to do. … If he had two chances in three of coming back—and he figured the odds were probably a good bit better than that—he was ready."*
>
> —Andrew Chaikin, in *A Man on the Moon*, describing the chances of success of Apollo 8, the first manned flight around the moon.

Outline

1. Example: comparing two models
2. $\text{Prob}\left(\text{model}\middle|\text{data}\right)$ *versus* $\text{Prob}\left(\text{data}\middle|\text{model}\right)$
3. Step-by-step analysis or single-step analysis?
4. Terminology and computational conventions
5. Multiple models
6. Subjective interpretation of probability
7. Credible interval *versus* confidence interval
8. The choice of a prior distribution and the effect of the prior distribution on the posterior distribution
9. Summary

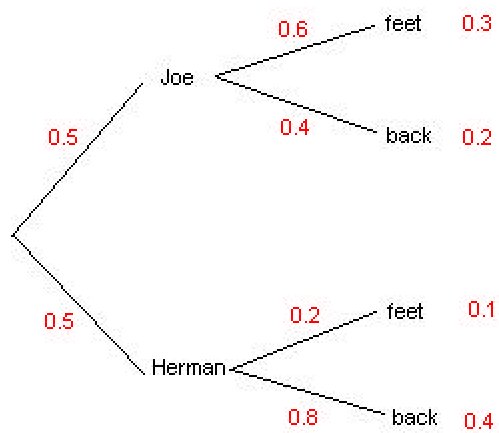
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

1. Example: comparing two models

Suppose that we have two paper frogs named Mighty Joe and Herman. We know that Mighty Joe lands on his feet 60% of the time and Herman lands on his feet only 20% of the time. Both frogs always land either on their feet or their backs, and the position in which they land on any jump is independent of the position in which they land on all other jumps. One of the frogs is red and the other one is blue. Unfortunately, the frogs' nametags have fallen off, and the only way we have of telling which one is which is by their ability to jump and land on their feet!

We pick one of the frogs—the red one—having no idea if it is Mighty Joe or Herman. It jumps, and lands on its feet. What is the probability that the red frog is Mighty Joe?

This problem may be approached in several ways. One way is through a probability tree, which is shown below.



The question of interest is this: what is the probability that we have selected Mighty Joe, given that after one jump, the frog landed on its feet? This is a conditional probability, so the formula for computing a conditional probability must be used.

$$\text{Prob}(A|B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)}$$

In this problem, the formula applies in the following way:

$$\text{Prob}(\text{Joe}|\text{feet}) = \frac{\text{Prob}(\text{Joe} \cap \text{feet})}{\text{Prob}(\text{feet})}$$

The values needed to perform this computation can all be found in the tree diagram. Note that $\text{Prob}(\text{feet})$ is the *total* probability that the frog will land on its feet. $\text{Prob}(\text{feet}) = \text{Prob}(\text{Joe} \cap \text{feet}) + \text{Prob}(\text{Herman} \cap \text{feet})$. Thus,

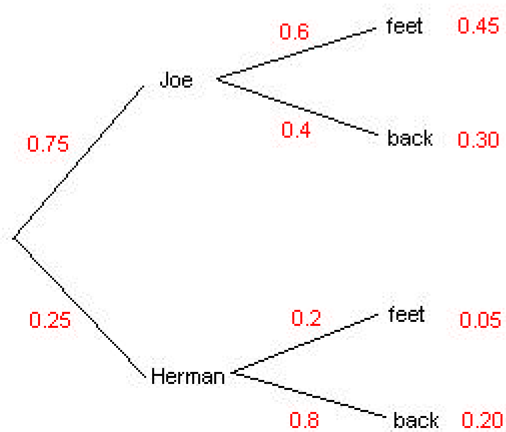
$$\text{Prob}(\text{Joe}|\text{feet}) = \frac{\text{Prob}(\text{Joe} \cap \text{feet})}{\text{Prob}(\text{feet})} = \frac{0.3}{0.3+0.1} = 0.75.$$

Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

We began with a probability of 0.50 that we had selected Mighty Joe, because we assumed it was equally likely that we had selected either frog. Then after observing the frog land on its feet, the selection of the two frogs no longer seems equally likely. Our revised estimate of the probability that we have selected Mighty Joe is 0.75.

One jump does not produce very much information. Suppose that we make the frog jump a second time, and it lands on its feet again. Now what is our updated estimate of the probability that we have selected Mighty Joe?

The problem can be approached in exactly the same manner as before. The difference is that now we begin not with a 0.50 chance of having selected Mighty Joe, but a 0.75 chance of having selected him. We go into this problem having already observed the frog land on its feet once, so our "initial" estimate is now an informed one. The new probability tree looks like this:



The computation of the updated probability estimate is essentially as before:

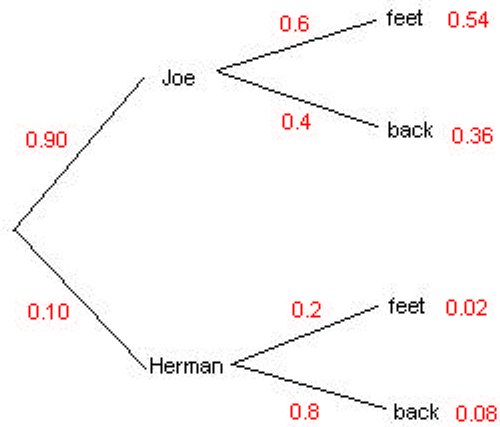
$$\text{Prob}(\text{Joe}|\text{feet}) = \frac{\text{Prob}(\text{Joe} \cap \text{feet})}{\text{Prob}(\text{feet})} = \frac{0.45}{0.45 + 0.05} = 0.90.$$

We now estimate that there is a 0.90 chance that the frog we are holding is Mighty Joe.

Finally, suppose that we jump the frog a third time and now observe that it lands on its back. What is our new estimate of the probability that we are holding Mighty Joe?

The probability tree and computations are as follows:

Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics



$$\text{Prob}(\text{Joe}|\text{back}) = \frac{\text{Prob}(\text{Joe} \cap \text{back})}{\text{Prob}(\text{back})} = \frac{0.36}{0.36 + 0.08} \approx 0.82.$$

Notice that when the frog landed on its feet and then on its feet a second time, our estimate of the probability that we had selected Mighty Joe rose from 0.50 to 0.75, and then to 0.90. When it then landed on its back, the probability decreased to about 0.82. This should not be surprising: the more the frog lands on its feet, the more likely it is that the frog we hold is Mighty Joe and the less likely it is that the frog we hold is Herman.

Suggested activity: Repeat the above process to calculate the probability that the selected frog is Herman. Assume that the observations are feet, feet, back, just as before. You should be able to verify that at each step of the way, the probability of the frog being Herman is equal to one minus the probability of it being Mighty Joe, as you would expect.

2. Prob(model|data) *versus* Prob(data|model)

Consider the interpretation of the number 0.82 that was obtained in the given example after observing a frog land on its feet, then on its feet again, then on its back. We interpret it to mean that there is an 82% chance that the frog we selected was Mighty Joe. If we let p be the probability that the selected frog lands on its feet, then we have two possible *models*: $p = 0.6$ and $p = 0.2$. (These are similar to the *hypotheses* that one studies in tests of significance using frequentist methods.) After the three observed jumps, we conclude that there is an 82% chance that the $p = 0.6$ model is the correct one.

In frequentist inference, tests of significance are performed by supposing that a hypothesis is true (the null hypothesis) and then computing the probability of observing a statistic at least as extreme as the one actually observed during hypothetical future repeated trials. (This is the *P-value*). In other words, frequentist statistics examine the probability of the data given a model (hypothesis), while Bayesian statistics examine the probability of a model given the data.

3. Step-by-step analysis or single-step analysis?

Consider the frog experiment once more. Suppose that we select a frog and make it jump three times, and we observe that it lands on its feet two of the three times and on its back once. Does it matter that we do the analysis in three steps as was done earlier, or can it be done all at once? The answer is that it does not matter; the results will be the same either way. This will not be proven here for the general case, but it will be demonstrated for the example introduced earlier.

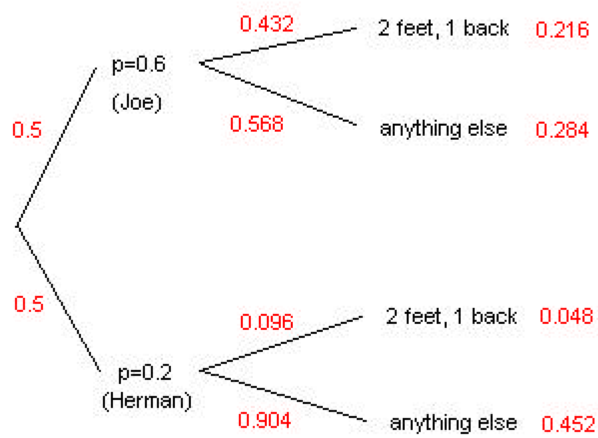
If the probability of a frog landing on its feet in a single trial is equal to $p = 0.6$, then the probability of it landing on its feet 2 out of 3 jumps can be found using the known probability associated with a binomial random variable:

$$\Pr ob\left(2F,1B \middle| p=0.6\right) = \binom{3}{2}(0.6)^2(0.4)^1 = 0.432.$$

Similarly, if $p = 0.2$, then the probability of the frog landing on its feet 2 out of 3 jumps is equal to

$$\Pr ob\left(2F,1B \middle| p=0.2\right) = \binom{3}{2}(0.2)^2(0.8)^1 = 0.096.$$

Now the appropriate probability tree looks like this:



And finally, computing the probability that we have selected Mighty Joe, given the observed 2 jumps in which the frog landed on its feet and 1 jump in which it landed on its back, we have:

$$\text{Prob}\left(\text{Joe}\middle|\text{2F,1B}\right) = \frac{\text{Prob}\left(\text{Joe} \cap \text{2F,1B}\right)}{\text{Prob}\left(\text{2F,1B}\right)} = \frac{0.216}{0.216+0.048} \approx 0.82,$$

which agrees with the earlier result. (Notice that it does not matter in what order the three observations are made.)

Suggested activity: Repeat the three-step analysis that was described in section 1 of this paper, only this time suppose that the frog landed on its back after the first jump, and then on its feet after the other two jumps. Verify that in the end, the probability that the selected frog is Mighty Joe is still 0.82.

Suggested activity: Suppose that the frog was jumped 10 times, and on 6 of the 10 jumps it landed on its feet. Verify that the probability that it is Mighty Joe is approximately 0.979.

Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

4. Terminology and computational conventions

There are certain terms that are commonly used by Bayesian statisticians. One is the *prior probability*, or simply the *prior*: this is the probability that a model is true before any data are observed. In our example, the prior probability that the frog was Mighty Joe was 0.5, because we assumed Joe and Herman were equally likely to have been chosen.

A second term is the *posterior probability*, or simply the *posterior*: this refers to the probability that a model is true after observed data have been taken into account. In our example, the posterior probability of the chosen frog being Mighty Joe was approximately 0.82.

A third very common term is *likelihood*: this is used to describe the conditional probability of the data given a particular model. In our example, the likelihood of the data given the $p = 0.6$ model was 0.432, and the likelihood of the data given the $p = 0.2$ model was 0.096.

Notice that each time a probability tree was drawn, its only branches that were used in the computation of the conditional probability were those that corresponded to what was actually observed. The branches corresponding to events not observed are not involved in the computation. This is in fact related to a key principle of Bayesian statistics: only what is actually observed is relevant in determining the probability that any particular model is true. For example, values "more extreme" than the observed ones are irrelevant.

The computations may be summarized in a table, in which each row corresponds to a model. The tabular version of the 2-out-of-3 feet-landing frog computation is shown below.

model	prior probability	likelihood	prior probability × likelihood	posterior probability
$p = 0.6$	0.5	0.432	0.216	0.82
$p = 0.2$	0.5	0.096	0.048	0.18
			total = 0.264	

The total of the values in the fourth column is computed for convenience: the posterior probabilities are computed by dividing the numbers in this fourth column by the total of all the numbers in that column. For this reason, that total is often referred to as the *normalizing constant.*

Suggested activity: Suppose that the frog jumped 10 times, and on 6 of the 10 jumps it landed on its feet. Use the tabular computation method to verify that the probability that it is Mighty Joe is approximately 0.979. This is the same suggested activity that was given in the previous section. Compare your earlier probability tree solution to the tabular solution done here. Observe where each number goes from the tree to the table; also notice what is altogether excluded from the table (as irrelevant).

bullard@ncssm.edu
The North Carolina School of Science and Mathematics

5. Multiple models

There is no reason we must be limited to only two models. Suppose that I ask all the students in my class to make paper frogs. They are tossed into a pile, and I randomly select one. I don't know what the probability is that it will land on its feet when it jumps (I'll call it p), but based on my past experience with student-constructed frogs, a reasonable guess as to the distribution of p is given in the following table.

model	probability estimate
$p = 0.0$	0.05
$p = 0.1$	0.10
$p = 0.2$	0.15
$p = 0.3$	0.19
$p = 0.4$	0.18
$p = 0.5$	0.15
$p = 0.6$	0.09
$p = 0.7$	0.05
$p = 0.8$	0.02
$p = 0.9$	0.01
$p = 1.0$	0.01

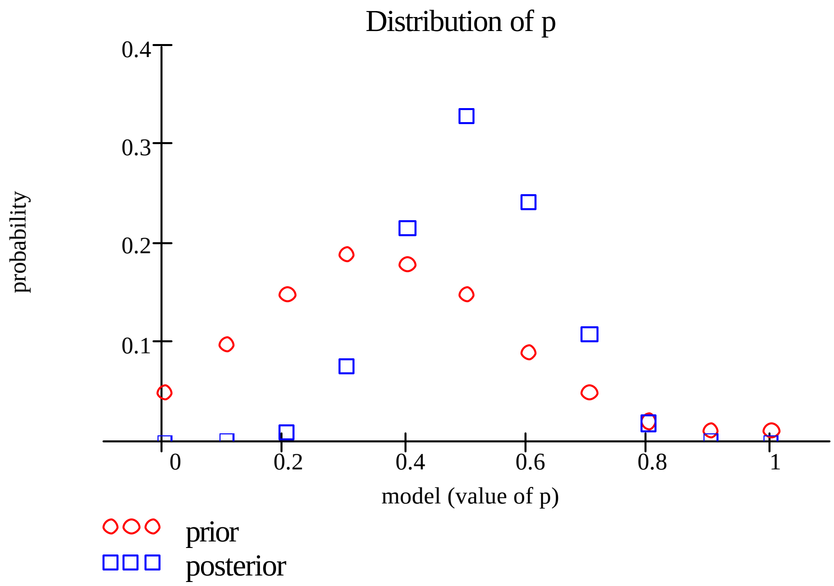
In words: my experience tells me that student-constructed frogs typically land on their feet 30 to 40 percent of the time; about 5% of the frogs are incapable of landing on their feet; the ability *always* to land on its feet occurs in only about one out of a hundred frogs; and so on. I assigned these probabilities based on my experience, and I made sure they added up to one.

Suppose now that after having randomly selected a frog from among those that my students constructed, I observe that it lands on its feet an impressive 6 out of 10 jumps. Given this additional information, what do I now estimate to be the distribution of p? The tabular computation of posterior probabilities serves for multiple models quite as well as for two models. For the case of 6 out of 10 feet-landings, the table is given below. The numbers in the likelihood column are binomial probabilities for 6 out of 10 successes, all based on the different models.

model	prior probability	likelihood	prior probability × likelihood	posterior probability
$p = 0.0$	0.05	0.0000	0.00000	0.0000
$p = 0.1$	0.10	0.0001	0.00001	0.0001
$p = 0.2$	0.16	0.0055	0.00083	0.0089
$p = 0.3$	0.19	0.0368	0.00698	0.0750
$p = 0.4$	0.18	0.1115	0.02007	0.2155
$p = 0.5$	0.15	0.2051	0.03076	0.3304
$p = 0.6$	0.09	0.2508	0.02257	0.2425
$p = 0.7$	0.05	0.2001	0.01001	0.1075
$p = 0.8$	0.02	0.0881	0.00176	0.0189
$p = 0.9$	0.01	0.0112	0.00011	0.0012
$p = 1.0$	0.00	0.0000	0.00000	0.0000
			total = 0.09310	

Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

The table above shows how the posterior probabilities are computed, but it does not give us a *feel* for how the probabilities associated with the different models have changed. For that, a graphical representation of these probabilities is more appropriate. The graph below shows, for the example just given, the prior distribution of p and the posterior distribution of p shown together.



This graph makes it clear that the prior distribution of p shifts to the right after observing 6 out of 10 feet-landings. Yet the most probable model is not $p = 0.6$ as one might expect, but $p = 0.5$. This is because the relatively small number of data are not sufficient to eliminate the influence of the prior distribution of p. A characteristic of Bayesian inference is that in the absence of much data, the prior distribution carries a lot of weight; but the more data that are observed, the less influence the prior distribution has on the posterior distribution.

Suggested activity: using the prior distribution given in this example (and using technology to compute and graph the posterior distributions) observe what the posterior distribution looks like when 12 out of 20 jumps are feet-landings. Repeat for 15 of 25 and 30 out of 50. You should observe that the posterior distribution not only becomes more centered on the model $p = 0.6$*, but also that it becomes narrower.*

6. Subjective interpretation of probability

Up to this point, I have used phrases such as “the probability that the frog we selected is Mighty Joe” and “the distribution of p”. Yet these phrases are meaningless under the frequentist interpretation of probability. AP statistics teachers may (or may not) have been bothered by my blithe use of these phrases. My guess is that those who are unhappy at this point with the language I have used first grew uncomfortable in the last section, in which I assigned probabilities to models based on my classroom experience. After all, we stress to our students that p—being a parameter—is constant, not random, and hence cannot have a distribution.

Yet a strict frequentist would have been up in arms from the very first example in this handout. Every computation and every example given here is based on a subjective interpretation of probability, even the very

Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

first example—which most readers probably accepted without difficulty. Recall the question that was asked in the first example:

"We pick one of the frogs—the red one—having no idea if it is Mighty Joe or Herman. It jumps, and lands on its feet. What is the probability that the red frog is Mighty Joe?"

According to the frequentist interpretation of probability, once a frog has been selected and is sitting in one's hand, p is no longer any more random than the frog's color. The selection of the frog (and hence the model for a value of p) has gone from being a future, potentially random event to a past, fixed event, and so it is meaningless to say something like "there is a 50% chance that I just selected Mighty Joe." Such a statement describes my degree of uncertainty about the kind of frog I hold in my hand—it is a subjective measurement.

The frequentist interpretation of probability is the long-run relative frequency with which an event will occur when a trial is repeated under "similar" conditions. Once an event is in the past, it is no longer random, so it is meaningless to discuss its probability (if you follow a frequentist interpretation of probability). That is why the interpretation of a confidence interval must be somewhat convoluted. If we say "there is a 95% chance that the parameter p lies in this confidence interval", then we are (incorrectly) assigning a probability to a fixed parameter and a fixed interval. We should instead say something like "this interval is one result of a procedure that had a 95% chance of producing an interval that would contain the parameter p."

In contrast, the subjective interpretation of probability is a subjective measure of one's degree of certainty about an event. It must obey certain mathematical rules (*e.g.*, total probability must equal unity, *etc*.), but is applicable to many more events than frequentist probability allows: non-repeatable events, past events, anything one is unsure about. It allows different people to have different estimates of probability because they may have different information, or simply different experiences.

There is a fairly well-known problem in which one estimates the probability of a randomly selected person being HIV+ given that he has tested positive for the virus.† This problem is meaningless if the frequentist interpretation of probability is used strictly. A person must be selected in order to be tested for HIV, but once he is selected, there is no longer any randomness associated with whether he is HIV positive—hence, one cannot meaningfully discuss the probability of his having the virus. Either he does or he does not. The fact that some variation on this question appears in many statistics texts, and that so many teachers discuss it in class without batting an eyelash is, I believe, a testament to the fact that people instinctively interpret probability subjectively.

Further evidence of this is the great difficulty we teachers have in getting our students to interpret confidence intervals correctly. If the frequentist interpretation were more "natural" than the subjective interpretation, then students would not have such a strong impulse to interpret an already calculated 95% confidence interval as one that contains the parameter with probability 95%.

Suggested activity: try the following with your students. Tell them you are about to flip a fair coin. Ask them to tell you what the probability is that it will land heads up. After hesitating, sure that you are up to something, they will likely answer 50%. Then flip the coin, catch it in your hand, and close your palm over the coin without looking at it. Ask your students what the probability is that it is lying heads up. I think you will find that more than half of your students will answer "50%" and the rest will say nothing, aware that something is fishy, but not knowing what exactly is wrong. Finally, peek at the coin in your hand so that you can see how it lies, but do not show it to the students. Ask them a second time: what is the probability that this coin lies face up? They will realize now how you have tricked them: for them, the probability remains 50% that the coin shows heads. For you, the probability has become either 0% or 100%, because you have information

† For example, see *Yates, Moore, and McCabe*, pp. 363-4. Also see *Yates, Moore, and McCabe*, p. 499 for a brief comment on Bayesian inference.

that they lack. Of course, that is if you choose to interpret probability subjectively. If you follow a frequentist interpretation, then even the second question was meaningless, for regardless of whether anyone had looked at the coin, the event had already occurred and as such, was no longer random.

7. Credible interval *versus* confidence interval

In frequentist statistics, one speaks of a *confidence interval* as an estimate of a parameter. The probability statement involved in a confidence interval refers to the randomness of the sampling process. In many repeated random samples, the procedure produces an interval that contains the parameter 95% of the time. (Or 90%, *etc*.) Once the sample has been taken and a particular confidence interval constructed, there is no more randomness, hence no more probability.

In Bayesian statistics, an interval may also be used to estimate a parameter, but it is called a *credible set*, a *credible interval* or a *posterior probability interval* so as to distinguish it from a frequentist confidence interval. The probability statement in a *credible interval* is the one people seem "naturally" to want: it is the probability that the parameter lies in the interval.

Consider once more the example of a frog constructed by a student that is then observed to jump 10 times, landing on its feet 6 of the 10 times. The graph on page 8 shows the posterior distribution of p for one particular prior distribution. It happens that this posterior distribution has the following property:

$$\text{Prob}(p=0.3)+\text{Prob}(p=0.4)+\text{Prob}(p=0.5)+\text{Prob}(p=0.6)+\text{Prob}(p=0.7)\approx 0.97\,.$$

Thus, we say that for the given prior distribution of p and the observed data, a 97% *credible set* estimate of p is {0.3, 0.4, 0.5, 0.6, 0.7}. The probability that p lies in that set is 97%.[‡]

Suggested activity: use the same prior distribution given in section 5 and suppose that the frog was jumped 20 times, and 12 times it landed on its feet. Show that{0.4, 0.5, 0.6, 0.7} is a 97.5% credible interval estimate of p.

[‡] It may appear that a weakness of the use of Bayesian credible sets is that we can only get rather "clunky" discrete sets rather than continuous intervals, and sets whose credibility level cannot be chosen at any arbitrary level. But these things occur in the given example only because the prior distribution allowed only for values of p that were multiples of 0.1. Using technology to perform calculations, it is not difficult to make the division of the domain [0,1] arbitrarily fine. Indeed, in the limit, the probability *mass* function becomes a probability *density* function. When a probability density function is used instead of a probability mass function, one may construct a continuous credible interval of any desired credibility level: 95% or 90% or whatever.

8. The choice of a prior distribution and the effect of the prior distribution on the posterior distribution

The most common criticism of Bayesian methods is that since there is no single correct prior distribution, then all conclusions drawn from the posterior distribution are suspect. For example, a dishonest statistician could manipulate the prior distribution to produce any desired posterior distribution. And even an honest statistician might easily produce a result that I disagree with simply because my prior beliefs do not concur with his or hers.

The criticism is quite valid.† Published research using Bayesian methods should consider a variety of prior distributions, thus allowing the reader to see the effects of different prior beliefs on the posterior distribution of a parameter. One prior distribution that in some sense eliminates personal subjectivity is one that is uniform across all possible values of the parameter—this is called a *diffuse prior distribution*. Other reasonable prior distributions are not arbitrarily determined by a single statistician, but are based on the opinions of experts (*e.g*, doctors' opinions about treatment success), on previously performed experiments, on known properties of similar populations, *etc*. Prior distributions, while they are subjective, need not be arbitrary, and should be justified in any published research as should other choices we make.

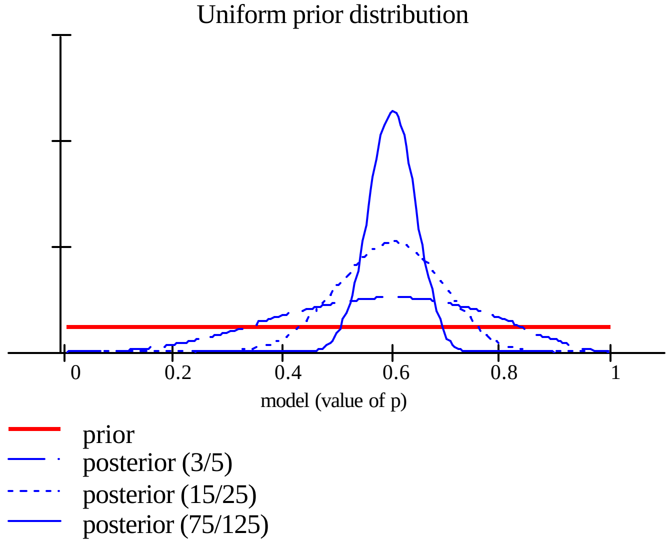
But there is one further characteristic of Bayesian inference that weakens this criticism of the prior distribution: the more data that are collected, the less influence the prior distribution has on the posterior distribution relative to the influence of the data. The graphs on the following page demonstrate this for the jumping frog scenario. Four different prior distributions are considered, which I chose for their variety: uniform, right-skewed, bimodal, and mound-shaped. For each prior distribution, the resulting posterior distributions are shown for three different data sets: 3 out of 5 jumps landing on their feet, 15 out of 25 jumps landing on their feet, and 75 out of 125 jumps landing on their feet. As you can see in the graphs, as more data are collected, the posterior distribution converges to the same distribution regardless of the prior distribution.

† However, if this criticism is valid, then we must also criticize equally subjective choices used in frequentist statistics such as cutoff values for statistical significance, choice of null and alternative hypotheses, and choice of likelihoods.

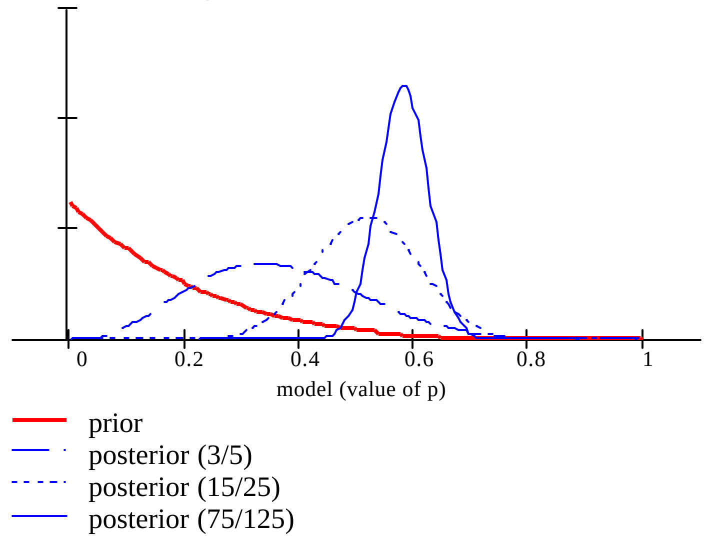
Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

The effect of the prior distribution on the posterior distribution

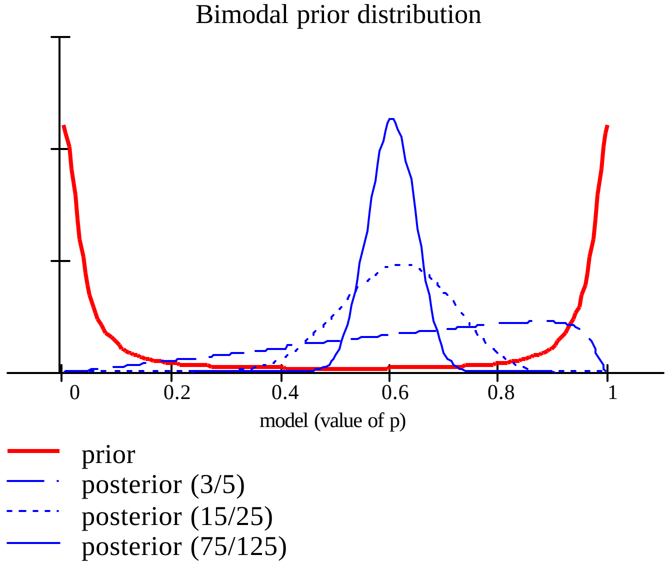
Uniform prior distribution



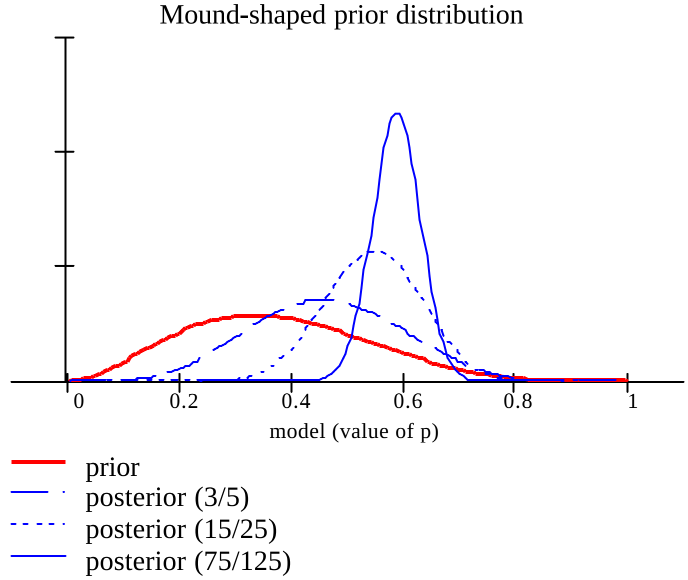
Right-skewed prior distribution



Bimodal prior distribution



Mound-shaped prior distribution



Floyd Bullard
NCTM 2001
bullard@ncssm.edu
The North Carolina School of Science and Mathematics

9. Summary

The following table contrasts frequentist and Bayesian statistical methods.

Frequentist	*Bayesian*
Probability is interpreted as the long-run relative frequency with which an event occurs in many repeated similar trials. Probability lies objectively in the world, not in the observer.	Probability is interpreted as a measure of one's degree of uncertainty about an event. Probability lies in the mind of the observer and may be different for people having different information or different past experiences.
Inference is performed by evaluating the probability of the observed data, or data more extreme, given a hypothesized model.	Inference is performed by evaluating the probability of a hypothesized model given observed data.
A 95% confidence interval is one result of a procedure that had a 95% chance of generating an interval that would contain the parameter being estimated.	The probability is 95% that the parameter being estimated lies in a 95% credible interval.
The P-value in a test of significance is the probability of getting a result at least as extreme as the one observed, given that the null hypothesis is true.	One may evaluate the probability of any particular model or set of models given observed data. Data not observed (*e.g.* "more extreme" values) are irrelevant.

It is not the purpose of this paper nor of the presentation at NCTM to argue a case for Bayesian methods over frequentist methods, but simply to introduce the very basic principles of Bayesian analysis in a way that can be understood easily by high school students with some knowledge of probability. Bayesian methods are gaining popularity in many areas such as clinical trials, genomics, marketing, environmental science, and other fields where prediction and decision making must follow from statistical analysis. Since Bayesian methods are highly computational, they are also gaining wider acceptance as technology makes analyses possible that were not feasible in the recent past.

For these reasons, I believe that a brief introduction to Bayesian methods is an excellent topic to cover in an AP statistics course after the AP exam is over.

More information about Bayesian methods is readily available on the web and in many texts. The resources on the next page are ones that I've found helpful while learning about this topic. I encourage anyone who is interested in Bayesian statistics to study it further. It really is an exciting field!

References

Statistics: A Bayesian Perspective, by Berry. Duxbury, 1996. (Undergraduate introductory textbook on Bayesian methods)
Bayesian Data Analysis, by Gelman, Carlin, Stern, and Rubin. Chapman and Hall/CRC, 1997.
Probability and Statistics 2nd edition, by Morris H. DeGroot. Addison-Wesley, 1986.
Bayesian Methods, by Leonard and Hsu. Cambridge University Press 1999.
Data Analysis: A Bayesian Tutorial, by D. S. Sivia. Clarendon Press, 1996.
Bayes and Emperical Bayes Methods for Data Analysis, 2nd edition, by Carlin and Louis. Chapman and Hall 2000.

The Practice of Statistics, by Yates, Moore, and McCabe. W. H. Freeman and Co., 1999.

A Probability/Activity Approach for Teaching Introductory Statistics, by Jim Albert. (Web site. Address: http://math-80.bgsu.edu/nsf_web/main.htm)

Activity-Based Statistics, by Schaeffer, Gnanadesikan, Watkins, and Witmer. Springer 1996.
Make an Origami Frog, University of Wisconsin Sea Grant Institute. (Web site. Address: http://www.seagrant.wisc.edu/madisonjason10/origami.html)

A Man on the Moon, by Andrew Chaiken.

I would like to thank Dalene Stangl, associate professor of statistics at Duke University, and Jeff Witmer, professor of mathematics at Oberlin College, for reading and correcting this handout. I have tried to make all the corrections they suggested. Any errors that remain are entirely my own.

bullard@ncssm.edu
The North Carolina School of Science and Mathematics