22S:138
Bayesian Statistics

What is Bayesian Statistics?

Lecture 1
Aug. 25, 2003

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

The Scientific Method[1]

(But it's not just for "science"...)

1. Ask a question or pose a problem.
2. Assemble and evaluate the relevant information.
   - (Take stock of what is already known.)
3. Based on current information, design an investigation or experiment (or perhaps no experiment) to address the question posed in step 1.
   - Consider costs and benefits of the available experiments, including the value of any information they may contain.
   - Recognize that step 6 is coming.
4. Carry out the investigation or experiment.
5. Use the evidence from step 4 to update the previously available information; draw conclusions, if only tentative ones.
6. Repeat steps 3 through 5 as necessary.

[1] as stated by Don Berry

Where does statistics fit in?

- Central to steps 2, 3, and 5
- May help with step 1
  - can help show that a question is inappropriate
  - may show that answering the question will be difficult or impossible
- *Bayesian* statististics is particularly well-suited for steps 2 and 5.

Who started it all? Thomas Bayes

Born: 1702 in London, England
Died: 17 April 1761 in Tunbridge Wells, Kent, England

- ordained "Nonconformist" minister in Engalnd
- *Essay towards solving a problem in the doctrine of chances*
  - set out Bayes's theory of probability
  - published in the Philosophical Transactions of the Royal Society of London in 1764
  - The paper was sent to the Royal Society by Richard Price, a friend of Bayes', who wrote:

    > I now send you an essay which I have found among the papers of our deceased friend Mr Bayes, and which, in my opinion, has great merit... In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it

but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

- Bayes's conclusions were accepted by Laplace in a 1781 memoir, rediscovered by Condorcet (as Laplace mentions), and remained unchallenged until Boole questioned them in the Laws of Thought. Since then Bayes' techniques have been subject to controversy.
- elected a Fellow of the Royal Society in 1742 despite the fact that at that time he had no published works on mathematics. Indeed none were published in his lifetime under his own name.

Some settings in which Bayesian statistics is used today

- economics and econometrics
- marketing
- social science
- education
- health policy
- medical research
  - more common in England than in US
  - but FDA has approved some new medical devices based on Bayesian analysis and is pushing the use of Bayesian methods in device testing
- weather
- the law
- etc., etc.

Simple inference using Bayes' rule
Example: Do you have a rare disease?

- Your friend is diagnosed with a rare disease that has no obvious symptoms.
- You wish to determine how likely it is that you, too, have the disease.

  That is, you are *uncertain* about your true disease status.
- Your friend's doctor has told her that
  - The proportion of people in the general population who have the disease is .001.
  - The disease is not contagious.
- A blood test exists for this disease, but it sometimes gives incorrect results.

Quantifying uncertainty using probabilities

The long-run frequency definition of the probability of an event

> The probability of an event is the proportion of the time it would occur in a long sequence of observations (i.e. as the number of trials tends to infinity).

- example: when we say that the probability of getting a head on a toss of a fair coin is .5, we mean that we would expect to get a head half the time if we flipped the coin a huge number of times under exactly the same conditions
- requires a sequence of repeatable experiments
- no frequency interpretation possible for probabilities of many kinds of events
  - including the event that you have the rare disease

Probability as degree of belief

The subjective definition of probability is

> A probability of an event is a number between 0 and 1 that measures a particular person's subjective opinion as to how likely that event is to occur (or to have occurred).

- applies whenever the person in question has an opinion about the event
  - if we count ignorance as an opinion, always applies!
- Different people may have different subjective probabilities regarding the same event.
- The same person's subjective probability may change as more information comes in.
  - where Bayes' rule comes in

Properties of probabilities

These properties apply to probability whichever definition is being used.

- Probabilities must not be negative. If A is any event, then

$$P(A) \geq 0$$

- All possible outcomes together must have probability 1.

  If S is the *sample space* in a probability model then

$$P(S) = 1$$

Back to the example

- two possible events or *models*
  1. you have the disease
  2. you don't have the disease
- *before* taking any blood test, you think your chance of having the disease is similar to that of a randomly selected person in the population
  - so you assign the following *prior probabilities* to the two models

MODEL	PRIOR
Have disease	.001
Don't have disease	.999

Data

- You decide to take the blood test.
  - the new information that you obtain to learn about the different models is called *data*
  - the different possible data results are called *observations* or *outcomes*
  - the data in this example is the result of the blood test
- The two possible observations are
  - a "positive" blood test (+) — suggests you have the disease
  - a "negative" blood test (-) — suggest you don't have the disease

Likelihoods

- The probabilities of the two possible test results are different depending on whether you have the disease or not.
- these probabilities are called *likelihoods* — the probabilities of the different data outcomes *conditional on* each possible model.

		LIKELIHOODS	
MODEL	PRIOR	P(+ | MODEL)	P(- | MODEL)
Have disease	.001	.95	.05
Don't have disease	.999	.05	.95

Using Bayes' rule to update probabilities

- Bayes' rule is the formula for updating your probabilities about the models given the data.
- enables you to compute *posterior probabilities* given the observed data
  - *posterior* means *after*

Bayes' rule (simplest form)

$$P(MODEL \mid DATA) \propto P(MODEL) \times P(DATA \mid MODEL)$$
$$\textit{posterior} \propto \textit{prior} \times \textit{likelihood}$$

Bayes' rule applied to the example

You take the blood test and the result is positive (+). This is the data or observation.

MODEL	Prior	Like for +	Product	Posterior
Have disease	.001	.95	.00095	.019
Don't have disease	.999	.05	.04995	.981
			.05090	1

- Are the entries in the "Product" column probabilities?
- How do we convert them into probabilities?

What have you learned from the blood test?

- The probability of your having the disease has increased by a factor of 19.
- But the actual probability is still small ($< .02$).
- You decide to obtain more information by taking the blood test again.

Updating the probabilities again

- (We will assume that, conditional on your true disease status, the results from two blood tests are independent.)
- Your current probabilities are the *posterior* probabilities from after the first test.
- These will become your *prior* probabilities with respect to the second test.
- The second test result is also positive.

MODEL	Prior	Like for +	Product	Posterior
Have disease	.019	.95	.01805	.269
Don't have disease	.981	.05	.04905	.731
			.0671	1

What if the second test had been negative?

That is, the second observations was (-).

MODEL	Prior	Like for -	Product	Posterior
Have disease	.019			
Don't have disease	.981			