STAT 4444/5444G

Applied Bayesian Statistics

Unit 1

Basics of Bayesian Statistics

Hamdy F. F. Mahmoud, PhD
Collegiate Assistant Professor
Statistics Department @ VT
ehamdy@vt.edu

Spring 2021

This Lecture Unite covers

- A bit of Bayesian History
- Steps of scientific method
- Differences between frequentist and Bayesian approaches
- Motivation example
  - Quantifying uncertainty using probabilities
  - Models and prior probabilities
  - Data
  - Likelihood and posterior probabilities
  - Bayesian sequential analysis
- More Practice
- What is to come?

A bit of Bayesian History

- Great fire of London 1666 that destroyed over 13,000 houses, 89 churches, and dozens of public buildings was the sparkle of the philosophy and methods of Bayesian statistics. By the early eighteenth century, the idea of life insurance was taking hold in England.

- Thomas Bayes who was born in London in 1702 applied his mind to the questions urgently raised by the insurers and laid out his resulting theory of probability in his Essay towards solving a problem in the doctrine of chances. After Bayes' death in 1761, his friend Richard Price sent the article to the Royal Society of London journal and the article were published in 1764.

A bit of Bayesian History

- Before1990, applying Bayesian methods to statistical analyses in real-world problems was very limited because the necessary mathematical computations could be done analytically only for very simple models.

- In the early 1990s, the increasing accessibility of powerful computers, along with the development of new computing algorithms for fitting Bayesian models, opened the door to the use of Bayesian methods in complex, real-world applications. As a result, Bayesian methods applications spread in many fields, such as astrophysics, weather forecasting, health-care policy, biology, criminal justice, engineering, ….. Etc.

- Today, Bayesian statistics is widely used to guide learning and decision-making in business and industry as well as in science. For example, software using Bayesian analysis guides Google's driverless robotic cars (McGrayne 2011a).

Steps of scientific method

Statistics is central to steps 2, 3, and 4. *Bayesian* statistics is central to steps 2 and 4.

Differences between frequentist and Bayesian approaches

The differences between the two frameworks come from the way the concept of *probability* itself is interpreted.

❑ Parameter estimation

Consider this example. We want to estimate the average height of adult females. First, we assume that height has a normal distribution. Second, we assume that the standard deviation is available and we don't need to estimate it. Therefore, the only thing we need to estimate is the mean of the distribution.

The frequentist way:

The frequentist approach would reason as follows: "*I don't know what the mean female height is. However, I know that its value is fixed (not random). Therefore, I cannot assign probabilities to the mean being equal to a certain value or being less than/greater than some other value. The most I can do is collect data from a sample of the population and estimate its mean as the value which is most consistent with the data.*" That estimated value is known as the maximum likelihood estimate.

Differences between frequentist and Bayesian approaches

The Bayesian way:

A Bayesian, on the other hand, would reason differently: *I agree that the mean is a fixed and unknown value, but I see no problem in representing the uncertainty probabilistically. I will do so by defining a probability distribution over the possible values of the mean and use sample data to update this distribution*.

- In a Bayesian setting, the newly collected data makes the probability distribution over the parameter *narrower*.

Differences between frequentist and Bayesian approaches

❑ Data prediction

Here, the difference between the frequentist and Bayesian approaches is analogous to their difference in parameter estimation. Again, **frequentists** don't assign probabilities to possible parameter values and they use (maximum likelihood) point estimates of unknown parameters to predict new data points.

Bayesians, on the other hand, have a full posterior distribution over the possible parameter values which allows them to take into account the uncertainty in the estimate by integrating the full posterior distribution, instead of basing the prediction just on the most likely value.

Differences between frequentist and Bayesian approaches

❑ P-values and confidence intervals

Frequentists don't treat the uncertainty in the true parameter value probabilistically. However, that doesn't magically eliminate uncertainty. The maximum likelihood estimate could still be wrong and, in fact, most of the time it is wrong! When you assume that a particular estimate is the correct one, but in reality it isn't, you will make an error. Consequently, this has lead to the development of two mathematical techniques for quantifying and limiting long-term error rates:

- Null hypothesis significance testing and the related concept of p-values.
- Confidence intervals

Differences between frequentist and Bayesian approaches

❑ P-values and confidence intervals

The general idea is to make an estimate, then assume something about the estimate only under certain conditions. You choose these conditions in a way that limits the long-term error rate by some number, usually 5% or lower.

- What is the meaning of p-value?

- What is Type I and Type II errors?

Differences between frequentist and Bayesian approaches

❑ P-values and confidence intervals

Confidence intervals are the frequentist way of doing parameter estimation that is more than a point estimate.

If you choose a population with a fixed mean, collect sample data, and finally calculate the 95% confidence interval, 95% of the time the interval you calculated will cover the true mean.

Once you've calculated a confidence interval, it's incorrect to say that it covers the true mean with a probability of 95% (this is a common misinterpretation). You can only say, in advance, that in the long-run, 95% of the confidence intervals you've generated by following the same procedure will cover the true mean.

Differences between frequentist and Bayesian approaches

❑ P-values and confidence intervals

Click on the image below to start a short animation that illustrates graphically the procedure that is described above:

Motivating Example: does a friend have breast cancer?

The National Cancer Institute recommends that women aged 40 and above should have mammograms every one to two years. A mammogram produces x-ray images of tissues and structure inside the breast and "**may**" help detect and identify cancerous tumors, benign cysts, and other breast conditions.

Assume that a friend was referred by her physician for her first screening mammogram. Assume this friend does not have a family history of breast cancer, and before being referred for the screening mammogram, she had given no thought whatsoever to breast cancer as something that could conceivably happen to her.

- **What is her chances of being diagnosed with breast cancer?** - step 1 of the scientific method.
- Bayes' rule within the scientific method can be used to answer this research question!

Before we answer this question, let us discuss how do we quantify uncertainty and the properties of probabilities!

Quantifying uncertainty using probabilities

In science, business, and daily life, people quantify uncertainty in the form of probabilities. The weather forecaster says there is a 30% probability of precipitation today; the seismologist says that there is a 21% chance of a major earthquake along the San Andreas fault by the year 2032. There are two interpretation of probabilities:

1) **The long run frequency interpretation**
2) **Subjective probability**

Quantifying uncertainty using probabilities

1) **The long run frequency interpretation**

   Saying that the probability of a fair coin coming up heads is 0.5 means that we expect to get a head about half the time if we flip the coin a huge number of times under exactly the same conditions.

- This interpretation is useful and the frequentist approach is based on it, but in many situations it can not be used. For example, in 2007-2008, editorial writers were assessing the probability that the United States economy was headed for a major recession. It is obvious that this question could not be couched as the proportion of the time that countries facing exactly the economic, social, and political conditions that then existed in the USA would go into recession.

Quantifying uncertainty using probabilities

2) **Subjective probability**

A probability of an event or of the truth of a statement is a number between 0 and 1 that quantifies a particular person's subjective opinion as to how likely that event is to occur (or to have already occurred) or how likely the statement is to be true.

Suppose that I tell you that I have carried out a study consisting of 6 trials. Each trial has only two possible outcomes, which I choose to call "success" and "failure." I believe that the probability of success was the same for each trial and that the six trials were independent. The outcome was six successes in six trials. What is your prediction of the outcome of the 7th trial? Or what is the probability of getting the 7th trial a success?

Properties of probabilities

When you assess probability of an event, keep in your mind two of the elementary properties of probability, regardless of which interpretation of probability is being used, must hold:

1. Probabilities must not be negative. If A is any event and P(A) denotes "the probability that A occurs," then $P(A) \geq 0$.
2. All possible outcomes of an experiment or random phenomenon, taken together, must have probability 1.

Back to our motivating example: Models and Prior Probabilities

For the friend, there are two possible true states of the world:

- She has breast cancer.
- She does not have breast cancer.

- Before obtaining any data specifically about her own breast cancer status, it would be rational for the friend to figure that her chance of being discovered to have breast cancer is similar to that of a randomly selected person in the population of all women who undergo screening mammograms.
- Based on published studies, the proportion who are diagnosed with breast cancer within 1 year is about 0.0045 (4.5 per 1,000). Therefore, the friend assigns the following prior probabilities to the two models.

Model	Prior probability
Breast cancer	0.0045
No breast cancer	0.9955

Note that this is step 2 in the scientific method. Prior probabilities refer to probabilities assessed before new data are gathered in step 3.

Data

The information in step 2 is not enough, new information needs to be collected. By having the mammogram, we go on to step 3.

The following are the possible results of a mammogram:

- A "positive" mammogram indicates possible or likely cancer and results in a recommendation of further diagnostic procedures,
- A "negative" mammogram does not give any evidence of cancer.

Let

- C+ (C−) the event that a person has (does not have) breast cancer
- M+ (M−) the event that the person has a positive (negative) mammogram result.

For a perfect test, P(M+|C+) would equal 1, and P(M+|C−) would equal 0. However, perfect tests generally do not exist.
A study of tens of thousands of people who received screening mammograms found that, if a person does have breast cancer, then the probability that the screening mammogram will be positive is about 0.724, P(M +|C+) = 0.724, and also found P(M+|C−)=0.027. The probabilities of a negative test under the two models are P(M−|C−)=0.973 and P(M−|C+) = 0.276.

Data

The available information so far is summarized in the following table:

Models, prior probabilities, and conditional probabilities of outcomes

Model	Prior Probabilities	P(M+ | Model)	P(M− | Model)
Breast cancer	0.0045	0.724	0.276
No breast cancer	0.9955	0.027	0.973

- **What is her chance of being diagnosed with breast cancer given all this information?**

Likelihoods and Posterior Probabilities

Bayes' rule is the mathematical rule for using data to update one's probabilities about models in a rational, systematic way. The likelihood associated with a particular model is the probability of the observed data under that model. Bayes' rule combines the prior probabilities with the likelihood to compute posterior probabilities for the respective models. In its simplest form, Bayes' rule states:

$$P(\text{Model} \mid \text{Data}) \propto P(\text{Model}) \times P(\text{Data} \mid \text{Model})$$

Likelihoods and Posterior Probabilities

The entries in the "Prior × likelihood" column are not probabilities. The probabilities must sum to 1. Bayes' rule says the posterior probabilities are proportional to, rather than equal to. so, we must normalize them — divide each one by the sum of the two.

Model	Prior Probabilities	Likelihood for M+	Prior × Likelihood	Posterior Probabilities
Breast cancer	0.0045	0.724	0.0033	0.107
No breast cancer	0.9955	0.0274	0.0273	0.893

The information from the mammogram increased the probability that she had breast cancer by a factor of 24. Although the actual probability was still fairly small (a bit over 1/10), it was far too large to ignore.

Bayesian Sequential Analysis

At her doctor's recommendation, the friend underwent a procedure called stereotactic core needle biopsy (SCNB). SCNB is a procedure in which mammographic methods are used to guide the placement of a hollow needle into the suspicious part of the breast. Cells are drawn out through the needle and are examined under a microscope.

If cancerous cells are observed, the test is considered positive (S+) and surgery is indicated. If no abnormal cells are observed, the test is negative (S−).

According to Bauer et al. (1997), $Pr(S+|C+)=0.89$ and $Pr(S-|C-)=0.94$. In addition, $Pr(S+|C-)=0.06$ and $P(S-|C+)= 0.11$. In deciding to undergo SCNB, the friend was returning to step 2 of the scientific method, determining that the current information was insufficient.

Bayesian Sequential Analysis

- After the mammogram, the friend's current probabilities were the posterior probabilities based on the mammogram. These became her prior probabilities with respect to the new data obtained from the SCNB.
- A Bayesian analysis in which accumulating information is incorporated over time, with the posterior probabilities from one step becoming the prior probabilities for the next step, is called a Bayesian sequential analysis.
- **SCNB for the friend was negative**. After incorporating the data from the SCNB, the posterior probability that the friend had breast cancer was only 0.014. Her doctor recommended no immediate surgery or other treatment but scheduled her for a follow-up mammogram in a year.

Model	Prior Probabilities	Likelihood for S-	Prior × Likelihood	Posterior Probabilities
Breast cancer	0.107	0.11	0.012	0.014
No breast cancer	0.893	0.94	0.839	0.986

More Practice

Assume that the SCNB result was positive, what is the posterior probability of having cancer?

What is to come?

This unit has provided a brief introduction to the scientific method, subjective probability, and how the Bayesian approach uses data to update people's knowledge.

In unit 2, basic concepts of probability will be reviewed. In the remainder units, we will build on these foundations in order to understand Bayesian modeling and inference for increasingly realistic problems.