STAT40380/STAT40390/STAT40850
Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

February 2016

Bayes' theorem

Last time we saw the version of Bayes' theorem we will be using throughout the course:

Bayes' Theorem

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$$

where:

- θ represents parameters about which we want to learn.
- $\mathbf{x}$ represents data that we have already observed.
- $p(\theta|\mathbf{x})$ is the posterior distribution.
- $p(\mathbf{x}|\theta)$ is the likelihood.
- $p(\theta)$ is the prior distribution.

But what is $p()$?

Some notes on Bayesian probability

- Bayesian statistics is often presented as controversial for two main reasons:
  1. A prior distribution *has* to be specified for our parameters.
     Where does this distribution come from?
     What if we get it wrong?
     What if we truly have no knowledge about our parameters?
  2. The definition of probability relies on *subjectivity*.
     This is an idea which is unpopular in many scientific disciplines.
- We will discuss prior distributions more in later lectures.
- Today we will explore the Bayesian definition of probability.

What do we mean by $p()$?

- I have been writing $P()$ for the probability that a discrete event occurs, and $p()$ for the probability density of a continuous observation.
- Thus $P(A)$ means the probability that event A happens.
- Similarly, $p(x)$ means the probability density that x takes at a specific value or set of values.
- More fully, if Z is a continuous random variable, $p(z) = p(Z = z)$ for any value of z we care to name. We have called this the probability density function or pdf.

Some things we can all agree on

We should all know that, for events A and B:

- $0 \leq P(A) \leq 1$.
- If $P(A) = 1$ then the event A will definitely happen.
- If $P(A) = 0$ then the event A will definitely not happen.
- If $P(A, B) = P(A) \times P(B)$ then the events are independent.
- If $P(A \text{ or } B) = P(A) + P(B)$ then the events are mutually exclusive.
- $\ldots$

All these ideas generalise to when A and B are continuous random variables.

What do we think we know about probability?

$$P(\text{Single die rolls a 6}) = \frac{1}{6}$$
$$P(\text{Coin comes up heads}) = \frac{1}{2}$$
$$P(\text{Ace of Spaces is the first dealt card}) = \frac{1}{52}$$
$$P(\text{6 on first die roll and 6 on second die roll}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Why do we think these probabilities are correct?

All probability in the real world is conditional on assumptions.
We call a collection of these assumptions a model.

- The assumptions we make when writing down our models are often not stated. For example, last time we wrote $x \sim N(\theta, \phi)$. We could have written $x|\theta, \phi \sim N(\theta, \phi)$ to be clearer.

Probability and independence

- If all probability is conditional, then there is no such thing as independence, only *conditional independence*.
- Thus if A is the event that I roll a 6 on a die on the first throw, and B is the event that I roll a 6 on a die on the second throw, these are only independent events conditional on:
  - The die is 6-sided
  - My throwing action does not influence the value of the die
  - Gravity acts on the die
  - ...
- If we don't know these things, then we might not believe that they are independent.
- Thus when we wrote $x_i \sim N(\theta, \phi)$ so that

$$p(x_1, \ldots, x_n|\theta) = p(x_1|\theta) \times \cdots \times p(x_n|\theta)$$

we were really saying that $x_1, \ldots, x_n$ are conditionally independent of each other given θ and ϕ.

Definitions of probability

- If all probability is conditional on assumptions, what does it mean to give the probability of an event a numerical value?
- We will look at two definitions of probability: the *frequentist* definition and the *Bayesian* definition.
    - Frequentist: $P(A)$ is given by the *relative frequency* of a long run of experiments testing whether A occurs or not.
    - Bayesian: $P(A)$ is a measurement of the *degree of belief* in event A taking place.
- Both definitions of probability agree with the mathematical properties in the previous slides.
- Note that the frequentist definition seems well-suited to problems such as tossing coins or dealing cards.
- However, it is not so well-suited to questions such as 'what is the probability of a 2° temperature rise by 2100?

An example

Think about the probability of a coin landing heads-up.
We might all naturally assume that the probability is $\frac{1}{2}$. Why?

1. Symmetry: there are two equal sides to the coin, and only one of these will give is 'heads'. Hence $\frac{1}{2}$. This is based on assumptions about the physical nature of the coin, the nature of the coin toss etc.
2. Frequency: we might have performed a large number of coin tosses and assumed these to occur in an identical manner, and be physically independent of each other.

Note that both of these arguments rely on some assumptions about the nature of the coin and the tossing procedure, and the semantic meaning of the terms identical and independent.

If we rely on assumptions to make probability statements, then our probabilities will be subjective.

Statistical inference - an introduction

- *Statistical inference* looks at how data should influence our beliefs about the real world.
- *Probability models* are used to describe the mechanism that gave rise to the data.
- These models are generally completely specified apart from a few unknown quantities called *parameters*.
- We assume that the model is a correct description of the mechanism that generated the data...
- although we can check this later.
- A word of warning from George Box *"All models are wrong but some models are useful."*

Main Inferential Procedures

- *Parameter estimation:* provide estimation for the values of the unknown parameters in the model.
- *Interval estimation:* provide sets of parameter values to describe the uncertainty in our estimate of the unknown parameters.
- *Hypothesis testing:* test a hypothesis about the model and its parameter values.
- *Prediction:* Given the current state of knowledge, how do we predict future observations?
- *Decision Theory:* Choose a course of action based on what we know.

Statistical paradigms

We can extend our ideas about probabilities to three paradigms for statistical inference:

- *Likelihood:* This idea adheres to the principle that all that can be learned about the parameters from the data is contained in the likelihood function.
- *Frequentist:* Statistical procedures should be chosen so that they are well behaved when averaged over an infinite sequence of independent repetitions of the experiment that produced the data.
- *Bayesian:* probabilities are used to measure our uncertainty about unknown parameters. Statistical procedures should use all knowledge that we have about the unknown parameters.

A short game

A coin is biased by a small unknown amount. What is your estimated probabiity of heads? Suppose you now have a small sample of coin throws. What is your new probability?

Different types of Bayesianism

- Most Bayesian statisticians accept that probability, at some level, is subjective.
- However, within this framework, there are several varieties of Bayesian statisticians:
  - Subjective Bayesians believe that probability is tied to an *individual's* belief about an event. Thus it makes sense to *elicit* an individual's probability distribution and use that as a prior distribution.
  - Objective Bayesians believe that combining information and probabilities using Bayes' theorem actually makes them *more objective* because more information is used and assumptions are made explicitly.
  - Linear Bayesians, like subjective Bayesians, believe that probability is tied to the individual. However, they argue that full probability distributions cannot be elicited from them, and instead focus on modelling *means and variances* only.

What does all this mean for us?

- *Both the prior distribution and the likelihood will be, to some degree, subjective choices.*
- When we are creating Bayesian models we should be aware of the assumptions we are making and state them explicitly.
- When we choose a prior distribution, we should be careful as to which distribution we choose, and whose belief it can be applied to.

Without realising it, you have always been a Bayesian!

Some further reading

Jeffries, H. (1961), *Theory of probability*, third edition. Oxford University Press.

Jaynes, E.T. (1996), *Probability theory: the logic of science*, Cambridge University Press.

de Finetti, B. (1974), *Theory of probability*, Wiley.

Berger, J.O. and Wolpert, R. (1984), *The likelihood principle*, Hayward, California Institute of Mathematical Statistics.

O'Hagan, A. (1988), *Probability: methods and measurement*, New York, Chapman and Hall.