Bayes class 2

Sean van der Merwe

# The Bayesian model flow

In this class we are going to look at how Bayesian models naturally flow in a mathematical sense.

# The properties of the data

Before we collect data we already know a lot about it, often enough to build a useful model.

Suppose you plan to observe the time that people stand in line at the supermarket. Perhaps a customer complained but the manager says that the customer is lying, and head office is paying you to sort it out statistically. What can you say about the observations right now?

# Distribution

What are some distributions that might work for this problem?

There are much better options, but we are going to use a very simple example for today, specifically a [This might be appropriate if each customer has 5 people in front of them with independent serving times.]

We need the density function, list some places where you might find it, then write it out.

# Likelihood

To keep things simple, we will assume (incorrectly) that observations are independent and identically distributed.

Likelihoods usually work better on the log scale, so the procedure is:

1. Write down the **log density** given a single observation ; 2. **Sum** over

Constants are used to capture the terms that are not of interest, in this case terms that contain no parameters.

# Priors

The prior sums up everything we know about the **parameters** *before* we collect any data.

There are many kinds (categories) of priors that are used in Bayes models, can you name a few?

For this simple problem we will use an easy prior:

# Hierarchy

In Bayes we usually write models in hierarchical notation, meaning in layers:

# Posterior

The posterior distribution is proportional to the prior times the likelihood. Alternatively, the log posterior is equal to the log prior plus the log likelihood plus a constant.

Write down the log posterior for this problem now:

Which leads to . After class check that you can get to this result yourself.

# Predictions

To make good predictions for new data we must consider as much uncertainty as is practically feasible. This includes at least the uncertainty in the observations and the uncertainty in the parameters of the model. However, when making predictions *we don’t really care about the values of the parameters*, only the information and uncertainty they carry, so we integrate over all possible values of the parameters to obtain a distribution for a new observation given all the old observations.

In this case,

Note that in almost all real problems you will find that the integration cannot be performed explicitly. Luckily this integral can be easily approximated via simulation techniques. *For example, ask your lecturer to show the approximate distribution a random future customer and of the maximum wait time of 10 random future customers.*

**The beauty of the simulation approach is that you only need to simulate once and then you can answer as many questions as you want.** Suppose after presenting your results someone asks you, “What is the probability of waiting more than 12 minutes?” This question can be answered easily using the same simulations as before by counting the proportion that meet the condition.

# Improvements

As the course progresses, we will learn how to use more flexible models with more realistic assumptions, and thus make better inferences and predictions. However, the principles do not change and should be learned right now.

# Fun final thought

It might be ironic that the frequentist definition of probability is manipulated to great effect to calculate Bayesian probabilities flowing from Bayesian models because Bayesian models are generally fit via simulation and simulation fits the assumption of ‘a large number of trials’. This is in contrast to the common frequentist use of the definition, where the ‘large number of trials’ is hypothetical (often referring to an infinite replication of an experiment) and doesn’t hold in practice.