Bayes class 8

Sean van der Merwe

# Frequentist Parameter Estimation

In the frequentist setting parameters have a fixed value, we observe one sample out of a hypothetical infinite set of possible samples, and we use the idea of sampling variation to try to say something related to the fixed value of the parameter.

For example, based on a sample I obtain a 95% confidence interval for the population mean as (1.2, 3.4). What does this interval say?

# Bayesian Parameter Description

In the Bayesian setting it is the sample that is fixed and we try to obtain as much information about the parameters as we reliably can from the sample. This is done by finding posterior distributions for the parameters. Once we have a distribution for a parameter we can create parameter ‘estimates’ in many ways, because there are many ways to describe or estimate the location of a distribution.

Why is ‘estimates’ in inverted commas here?

Consider the general problem of explaining the location or central tendency of a sample and realise that estimating a parameter value from a posterior is actually the same sort of problem, especially when working with simulations. Give a few measures of central tendency you know:

Some measures are ideal under particular loss functions. Connect 3 measures of central tendency of a posterior to their related loss functions:

# Bayesian Credibility Intervals

In the Bayesian setting the interpretation of an interval is much simpler: A 95% credibility interval covers 95% of a posterior distribution.

It does not say “that there is a 95% probability of the parameter value being in the interval” because in Bayes there is no “the parameter value”.

Credibility intervals can be created in many ways. The simplest is the ‘symmetric’ interval, which is symmetric in probability. If we want an interval covering proportion$ w$ of the posterior then we estimate the intervals using quantiles $\frac{1-w}{2}$ (think $^{α}/\_{2}$) and $\frac{1+w}{2}$ (think $\frac{1-α}{2}$). A 95% interval would then have 2.5% of the probability outside the left limit and 2.5% outside the right limit, leaving 95% in the inside of the interval.

An alternative is the Highest Density Interval (HDI) / Highest Posterior Density (HPD) interval / shortest interval. A shorter interval with the same coverage carries more information, so in a sense we can say that the best 95% interval is the shortest 95% interval. In theory this can be tough to find, but with a simulated sample it is straightforward: consider every viable 95% interval and use the shortest one.

# Objective Prior Derivation

In the absence of informative or expert prior information, it is often desired to appear as objective as possible in the choice of prior distributions. One might choose a prior based on a theoretically desirable property, rather than just convenience.

The first set of properties relate to parameter estimation. A prior might be preferred if it offered superior abilities to come close to parameter values and desired coverage in a simulated scenario that is similar to the real data scenario. For example, a 95% interval should ideally cover a selected value 95% of the time.

The second set of properties relate to information theory. An objective prior should minimise the information provided by the prior, and maximise the information provided by the sample. In the case of i.i.d. distributed observations we can formally derive such priors, can you give two examples with their algorithms?