11.5 What a Sampling Distribution is Good For

Zongyi Liu

2023-06-24

11.5 What a Sampling Distribution is Good For

Sampling distribution aren't as useful as posterior distributions for making inferences about hypotheses from a set of observed data; the reason is that sampling distributions tell us the probabilities of possible data if we run an intended experiment given a particular hypothesis, rather than the believabilities of possible hypotheses given that we have a particular set of data.

11.5.1 Planning an Experiment

When planning research, we have some hypothesis about how the world might be, and we want to gather data that will inform us about the viability of that hypothesis. Typically we have some notion already about the experimental treatments or observational settings, and we want to plan how many observations we'll probably need to make, or how long we'll need to run the study, in order to have reasonably reliable evidence one way or the other.

11.5.2 Exploring Model Predictions

A Bayesian analysis only indicates the relative veracities of the various parameter values or models under consideration.

One way to evaluate whether the least unbelievable parameter values are any good is via a **posterior predictive check**. A posterior predictive check is an inspection of patterns in simulated data that are generated by typical posterior parameters values. This use of the posterior predictive check is suspiciously like null hypothesis significance testing: We start with a hypothesis (i.e., the least unbelievable parameter values), and we generate simulated data as if we were repeating our intended experiment over and over. Then we see if the actual data are typical or atypical in the space of simulated data. If we were to go further, and determine critical values for false alarm rates and then reject the model if the actual data fall in its extreme tails, then we would indeed be doing NHST.