

Excursion 4 Tour II: Rejection Fallacies: Whose Exaggerating What?

We begin with the *Mountains out of Molehills Fallacy* (large n problem): The fallacy of taking a (P-level) rejection of H_0 with larger sample size as indicating greater discrepancy from H_0 than with a smaller sample size. (4.3). The Jeffreys-Lindley paradox shows with large enough n , a .05 significant result can correspond to assigning H_0 a high probability .95. There are family feuds as to whether this is a problem for Bayesians or frequentists! The severe tester takes account of sample size in interpreting the discrepancy indicated. A modification of confidence intervals (CIs) is required.

It is commonly charged that significance levels overstate the evidence against the null hypothesis (4.4, 4.5). What's meant? One answer considered here, is that the P-value can be smaller than a posterior probability to the null hypothesis, based on a lump prior (often .5) to a point null hypothesis. There are battles between and within tribes of Bayesians and frequentists. Some argue for lowering the P-value to bring it into line with a particular posterior. Others argue the supposed exaggeration results from an unwarranted lump prior to a wrongly formulated null. We consider how to evaluate reforms based on Bayes factor standards (4.5). Rather than dismiss criticisms of error statistical methods that assume a standard from a rival account, we give them a generous reading. Only once the minimal principle for severity is violated do we reject them. Souvenir R summarizes the severe tester's interpretation of a rejection in a statistical significance test. At least 2 benchmarks are needed: reports of discrepancies (from a test hypothesis) that are, and those that are not, well indicated by the observed difference.

Keywords

significance test controversy, mountains out of molehills fallacy, large n problem, confidence intervals, P-values exaggerate evidence, Jeffreys-Lindley paradox, Bayes/Fisher disagreement, uninformative (diffuse) priors, Bayes factors, spiked priors, spike and slab, equivocating terms, severity interpretation of rejection (SIR)