SELECTION PARADOXES OF BAYESIAN INFERENCE

By A. P. Dawid

University College London

When the inference to be made is selected after looking at the data, the classical statistical approach demands — as seems intuitively sensible — that allowance be made for the bias thus introduced. From a Bayesian viewpoint, however, no such adjustment is required, even when the Bayesian inference closely mimics the unadjusted classical one. In this paper we examine more closely this seeming inadequacy of the Bayesian approach. In particular, it is argued that conjugate priors for multivariate problems typically embody an unreasonable determinism property, at variance with the above intuition.

1. Introduction. A great deal of statistical practice involves, explicitly or implicitly, a two stage analysis of the data. At the first stage, the data are used to identify a particular parameter on which attention is to focus; the second stage then attempts to make inferences about the selected parameter.

Perhaps one of the most important intuitions which the discipline of Statistics has to offer — and perhaps one of the hardest to put across to the outside world — is the inappropriateness, in such circumstances, of a "face-value" approach, in which the second stage proceeds just as if the quantities identified at stage one had been chosen before the experiment. Instead, the selection or optimization applied at the first stage introduces a bias which needs to be allowed for — either by explicit modelling of the whole two-stage process, or by some general de-biasing technique such as cross-validation (Stone, 1974) or prequential analysis (Dawid, 1992).

In this paper we contrast the Bayesian approach to this problem with the above intuition. Since Bayesian posterior distributions are already fully conditioned on the data, the posterior distribution of any quantity is the same, whether it was chosen in advance or selected in the light of the data: that is, for a Bayesian, the face-value approach is fully valid, and no further adjustment for selection is required. This seeming paradox is all the more striking in problems where the face-value Bayesian inference closely mimics (in form, if

---