NILS L. HJORT[1]

Norwegian Computing Center

1. Introduction. Although hardly the authors' intention, these papers by Diaconis and Freedman (D & F) will probably be read by many as criticism against and pessimism about Bayesian analysis in situations with high-dimensional parameter spaces. It is perhaps also easy for the statistician scanning the papers to get an impression of "just counterexamples," which would be unfair; these and earlier papers by D & F (or F & D) contain many new important statistical ideas and also useful mathematical techniques.

I will try to be (more) positive and hope to show that thinking Bayes in semi- and nonparametric models may be a worthwhile enterprise, sometimes giving additional insight into old problems, and sometimes (dare I say often?) leading to sensible Bayes procedures that also behave agreeably in the frequentist asymptotic sense. The bulk of my comments concerns a problem that is almost as old as statistics itself, that of fitting a parametric model to a data set, and that can be attacked again with ideas underlying some of the constructions of D & F. Let $X_1, \dots, X_n$ be a sample from some unknown distribution F with density f. Some (possibly crude) parametric family $\{F_\theta, f_\theta\colon \theta \in \Theta\}$ is then forced on the data. Textbooks teach us how to proceed, for example, advocating finding the maximum likelihood estimator $\hat\theta_{\mathrm{ML}}$, on the grounds of good asymptotic behavior, in particular, consistency. What very few textbooks tell us, however, is what $\hat\theta_{\mathrm{ML}}$ does when the model is wrong, i.e., there is no θ_0 with $f = f_{\theta_0}$. It is however not difficult to see that $\hat\theta_{\mathrm{ML}}$ still is a meaningful estimator in that it takes aim at the parameter value $\theta = \theta_1$ that minimises Kullback–Leibler "information distance"

$$I(f : f_\theta) = \int f \log(f/f_\theta)\, dx; \tag{1}$$

the log likelihood divided by n is a consistent estimate of $\int f \log f\, dx - I(f : f_\theta)$. Under appropriate conditions $\hat\theta_{\mathrm{ML}}$ is consistent for this "least false" parameter value. Hjort (1985a, Chapter 3) has further comments about the behavior of maximum likelihood machinery when the model is wrong.

One of the major uses of a fitted model is prediction, or probability assessments, for certain sets. Thus we could be interested in stating that approximately 90% of future Xs from a fitted normal will fall in $(\hat\mu - 1.645\hat\sigma, \hat\mu + 1.645\hat\sigma)$, or that approximately 50% of future data points from a fitted Weibull fall below $\hat\theta(\log 2)^{1/\hat\alpha}$, etc. If such statements are an important part of the statistical analysis, then there are disadvantages to using $\hat\mu_{\mathrm{ML}}$, $\hat\sigma_{\mathrm{ML}}$, resp. $\hat\theta_{\mathrm{ML}}$, $\hat\alpha_{\mathrm{ML}}$ in the case of an incorrectly specified model, and one could do better with other estimates that aimed at other versions of least false population parameters. It is the aim of the present notes to show that such least false parameters can be defined and that a suitably engineered semiparametric Bayesian setup can result in estimates that actually manage to estimate these.

[1] This work was done while the author visited Stanford University with grants from the Norwegian Computing Center and the Royal Norwegian Council for Scientific and Industrial Research.