ISBA

INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

THE ISBA BULLETIN

OFFICIAL BULLETIN OF THE INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

SOFTWARE HIGHLIGHT

Luca Porta Mana

pgl@portamana.org

INFERNO
INFERence in R with Bayesian NOnparametrics

Many researchers in important fields, such as medicine, sadly still use p-values and frequentist statistics to do *population inference*, despite their intrinsic flaws. Some researchers use Bayesian methods but limit themselves to parametric ones, which make possibly unrealistic statistical assumptions.

Until a couple decades ago such practices could somehow be justified by pragmatic reasons:

- Better methods were computationally too expensive.

- The population-inference problems were low-dimensional; one could *visually* check whether the method or assumptions were appropriate to the problem, and change them otherwise, or consider the results as simply qualitative.

But today, in many cases, the reasons above cannot meaningfully be given anymore:

- Bayesian methods, even nonparametric, have become computationally feasible for many inference problems.

- Many population-inference problems today involve tens or hundreds of variates of different kinds, and are therefore very high-dimensional. It is impossible to visually check whether frequentist results or parametric assumptions are acceptable, or by how much they err. Results may therefore be affected by large errors (Draper, 1995), whose existence is often not even reported.

There is one reason that can still be given today for not using Bayesian methods, especially parametric ones, in population inference: *lack of user-friendly software*. Clinicians who'd be curious to try out Bayesian analysis of their studies simply can't, because that would require the study of Markov-chain Monte Carlo techniques, of programming languages to implement them, and of a plethora of debated methods and acquired visual skills to "assess convergence". Most clinicians don't have time to learn all this even if they wanted to.

Available packages for Bayesian nonparametrics focus moreover on the problem of *regression*, that is, the inference of a functional relationship – assumed to exist – between a set of predictor variates and a target or "predictand" variate. These packages are not appropriate to *population inference*, where no functional relationships exist a priori, and where often there is no a priori division between predictor and predictand variates. In clinical studies one may be interested in calculating the probability of an effect or symptom given a condition, and of a condition given an effect or symptom.

The R-package **inferno** was built to try to fill this lack of software.

(Walker, 2010)

References

D. Draper. Assessment and propagation of model uncertainty. *J. R. Stat. Soc. B*, 57(1):45–70, 1995. doi: 10.1111/j.2517-6161.1995.tb02015.x. See also discussion and reply in Spiegelhalter et al. (1995).

N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker, editors. *Bayesian Nonparametrics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, 2010. doi: 10.1017/CBO9780511802478.

D. J. Spiegelhalter, A. P. Grieve, D. V. Lindley, D. V. Lindley, J. B. Copas, A. J. G. Cairns, C. Chatfield, G. Box, D. Cox, J. W. Tukey, A. E. Raftery, M. Aitkin, G. A. Barnard, P. Donnelly, A. S. C. Ehrenberg, A. E. Gelfand, B. K. Mallick, A. Gelman, X.-L. Meng, C. A. Glasbey, G. J. Gibson, R. Glendinning, U. Hjorth, R. E. Kass, L. Wasserman, M. Lavine, D. Madigan, B. M. Pötscher, J. W. Pratt, A. F. de Vos, B. J. Worton, A. Zellner, and D. Draper. Discussion of the paper by Draper. *J. R. Stat. Soc. B*, 57(1):71–97, 1995. doi: 10.1111/j.2517-6161.1995.tb02016.x. See Draper (1995).

S. G. Walker. Bayesian nonparametric methods: motivation and ideas. In *Hjort et al. (2010)*, chapter ch. 1, pages 22–34. 2010. doi: 10.1017/CBO9780511802478.002.