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

1   Population inference and Bayesian nonparametrics

A very important kind of inference in research fields such as medicine is *population inference*, often also called "density inference" or "density regression". Its general goal is to infer the frequency distribution of some variates in a population. This is different, for instance, from *functional regression*, where the goal is to infer the functional relationship – assumed to exist – between a set of predictor variates and a target or "predictand" variate. In population inference the existence of a functional relation cannot be assumed. In fact there may not even be a clear distinction between predictor and predictand variates. A typical goal is the inference of frequency distributions within particular sub-populations or sub-groups, thus all kinds of conditional probabilities are required. A clinician may be interested in the statistics and probability of a medical condition given a symptom, but also that of a symptom given a medical condition; and maybe only within subjects of a given sex or age. De Finetti's theorem (see e.g. Bernardo and Smith, 2000, §§ 4.2, 4.3, 4.6) lies at the heart of population-inference methods; a particularly brilliant discussion is given by Lindley & Novick (1981).

Sadly many researchers still approach population-inference problems by means of p-values or other frequentist practices, which only give limited, coarse, and not seldom misleading results about a population's frequency distribution. Some researchers adopt Bayesian methods but limit themselves to *parametric* ones, which make very restrictive and possibly unrealistic assumptions about the population's distribution; as opposed to *nonparametric* ones, which don't.

Until a couple decades ago the use of parametric methods, and maybe even of frequentist practices, was somehow justified by pragmatic reasons. Better methods were computationally too costly or unfeasible. Population-inference problems were low-dimensional; one could *visually* check whether the assumptions were appropriate to the problem and the results reasonable. Today the reasons cannot be earnestly be given, however. Bayesian nonparametric methods have become computationally feasible for many kinds of population inference. And many population-inference problems today involve tens, hundreds, or thousands of variates of different kinds; being so high-dimensional, it is impossible to visually check whether frequentist practices 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.

One reason can still be given today for the avoidance of Bayesian nonparametric methods for population inferences: *lack of user-friendly software*. Clinicians who'd be curious to try out a Bayesian nonparametric analysis of their studies simply can't: that would require the study of Markov-chain Monte Carlo techniques, of programming languages to implement the latter, and of a plethora of debated practices to "assess convergence". Most clinicians don't have time to learn all this even if they wanted to. Available packages for Bayesian nonparametrics are not quite suited to population inference. Some focus on functional regression, which as discussed above is not an appropriate assumption. Some make an a priori distinction between predictor and predictand variates, limiting the range of useful inferences. Most still require non-statistical technical expertise.

The R-package **inferno**[1] was built to try to remedy the lack of this kind of software.

2 Features and application range

The package

(Walker, 2010)

References

J.-M. Bernardo and A. F. Smith. *Bayesian Theory*. Wiley series in probability and mathematical statistics. Wiley, New York, repr. edition, 2000. doi: 10.1002/9780470316870. First publ. 1994.

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. V. Lindley and M. R. Novick. The role of exchangeability in inference. *Ann. Stat.*, 9(1):45–58, 1981. doi: 10.1214/aos/1176345331.

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 1, pages 22–34. 2010. doi: 10.1017/CBO9780511802478.002.

---

[1] https://pglpm.github.io/inferno/