BAYESIAN STATISTICS 9,
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

Bayesian variable selection for random intercepts models: a discussion

Nicolas Chopin and Omiros Papaspiliopoulos
CREST (ENSAE), Paris and Pompeu Fabra, Barcelona
nicolas.chopin@ensae.fr, pmiros.Papaspiliopoulos@upf.edu

It is our experience that applied (micro-)econometricians are not very keen on "modern" statistics, including non-Bayesian methods such as the Lasso. We commend the authors on demonstrating the importance of concepts such as sparsity to a problem of great practical interest for econometricians (and others), namely taking into account individual heterogeneity in longitudinal (a.k.a. panel) data. Building a bridge between the simplistic standard regression model and the over-parametrized regression model with a random intercept that differs for each individual is certainly an appealing approach.

We'd like to make a few specific comments. In Section 2, see (3), the authors mention the usual reformulation of a model with random intercepts as a regression model with indicator functions, but recommend against using a g-prior under this representation, because of the "information inbalance" between α (the vector of regression coefficients) and β (the vector of random intercepts); instead, they assume prior independence, $p(\alpha, \beta) = p(\alpha)p(\beta)$. We do not entirely understand this line of reasoning. (a) Imagine that one covariate x_{it} is constant over time, and is one for a few individuals, zero otherwise. Should we do the same, i.e. treating this covariate separately, and assuming an independent prior for the corresponding coefficient? (b) Could we use the following justification instead for treating α and β separately? An appealing property of g-priors is invariance through linear transformation of the design matrix. But, in the reformulated model, where random intercepts are regression coefficients in front of indicator functions (=1 if individual is i, zero otherwise), not all linear transformations of the complete design matrix are meaningful: e.g. creating a new covariate as a linear combination of covariates and some "individual" indicator functions seems of no practical interest. So it would make sense to consider a restricted form of invariance, where only linear combinations of the genuine covariates (excluding the indicator functions) would be allowed. (c) We wonder whether α and β should not have a common scale, through a common hierarchy: for instance, if one multiplies all the y_{it} by 2, then both α and β should be multiplied by 2 as well.

N. Chopin is supported by the 2007–2010 grant ANR-07-BLAN-0237-01 "SP Bayes".

In Section 3, the authors discuss several "shrinkage" prior, but seem to use theterm "shrinkage" in an unconventional way, i.e. to mean that the posterior mode of a given coefficient may be zero with positive probability. Is this such an appealing property? (a) MAP estimation has no clear decision-theoretical justification; (b) a joint inference of the variables to be selected and the corresponding coefficients is already provided by the Dirac-spike-and-slab approach; and (c) the authors only compute posterior expectations in their simulations anyway. More generally, what would be the authors recommend as a reasonable "default choice"? possibly the Dirac-spike-normal-gamma-slab prior?

In the conclusion, the authors suggest to use a similar framework to detect changes in dynamic models: i.e. the random intercept becomes a function of time, δ_t, not of the individual. However, if a constant probability of change is assumed (following the logic behind the spike and slab prior), then the periods between change would follow a geometric distribution (possibly conditional on hyper-parameters). In our experience, such a prior is not always flexible enough, especially if long periods between changes are expected; see e.g. Chopin (2007), Fearnhead (2006) and Koop and Potter (2007).

REFERENCES

Chopin, N. 2007. Dynamic detection of change points in long time series. *Ann. of the Inst. of Stat. Math.* 59(2): 349–366.

Fearnhead, P. 2006. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and computing* 16(2): 203–213.

Koop, G. and S. M. Potter. 2007. Forecasting and Estimating Multiple Change-point models with an Unknown Number of Change-points. *Review of Economic Studies* 74: 763 – 789.