Report on Entropy 2563693
Adaptive MCMC for Bayesian variable selection in generalised linear models and survival models
by Xitong Liang, Samuel Livingstone and Jim Griffin

This paper introduces a novel adaptive MCMC scheme for Bayesian variable selection (BVS) in glm and survival models. The main contribution is presented in Sections 4.2 and 4.3 where the PARNI proposal is coupled with an adaptive ALA proposal. Specifically, a novel adaptive learning scheme of the PARNI parameters A_j and B_j, $j = 1, \ldots, p$ is presented.

The paper is well written and, in general, easy to follow although sometimes there is a lack of details. Nonetheless, I congratulate the Authors to have summarised almost 20 years of research in BVS in a few pages. Many more references could have been included but I noticed the bibliography is already quite long. Many of these references have been published recently in the last few years.

Overall, the motivation of the paper is clear, and the main novelty is well presented. I do not have many comments regarding the general methodology and the proposed solutions, so I'll try to provide feedback that could improve the presentation of the paper. The only less positive comment is the quality of the numerical experiments, both the simulated and real examples, compared to the rest of the paper.

***

- P(age)2-L(ine)47: Strictly speaking what is presented, and what I presume the Authors refer to, are not "clinical problems" since, besides clinical outcomes, there should be also clinical predictors. I would call them "problems in genetics/genomics". Also, "fine mapping" refers to a very specific task to find "causal" genetic variants from a dense set of highly correlated variants in a small region spanning a few Kbp. In "fine mapping" the problem is not only variable selection but how to deal with the resulting high level of collinearity

(or LD) among the predictors. Thus, I would call it simply "genetic mapping", to avoid confusion (please change also P3-L129, and throughout the paper).

- P2-L58: I disagree. The problem is not the inclusion/derivation of a trans-dimensional move, but to devise trans-dimensional moves that are efficient and mixed well (Jasra et al., 2007). Please rephrase the sentence.

- P2-L84: The reference to the Hamming Ball sampler (HBS) (Titsias and Yau, 2017) is relevant. It would be great if, in the simulated experiments and the real data logistic regression examples, the HBS could also be included - to reassure the reader that the actual results of the proposed BVS algorithm (and in particular the preferred PARNI-aALA-CPM) make sense and the proposed method is competitive with alternative methods. Moreover, when PARNI neighbourhoods $\mathcal{N}_r$ are introduced on P10, a short description of the similarities and differences with HBS's neighbourhoods $\mathcal{N}$ would be very informative (perhaps providing an example). This could highlight the benefits of the proposed adaptive PARNI algorithm.

- P2-L93: The reference to the weights is difficult to follow, and many details are missing. For instance, it is not clear what t or $g(\cdot)$ are. I would try to simplify this paragraph and provide more details of $g(\cdot)$ specifically for the PARNI algorithm, extending the paragraph on P11-L263.

- P3-L103: Shotgun Stochastic Search (SSS) (Hans et al., 2007) approach should also be included in the simulation experiments and real data analysis. SSS software is freely available for both logistic and Weibull regression (although it targets $\pi_{LA}(y)$). The description of the SSS neighbourhoods $\mathcal{N}$, and the differences with PARNI neighbourhoods should be also included on P10.

- P5-L164: Please define explicitly what γ is. At the end of eq. (5), a full stop is missing.

- P5-L164: In eq. (7), p_γ in I_{p_γ} is not defined.

- P5-L167: I think the Authors want to generalise the prior structure to include also (generalised) g-prior. However, its use is quite controversial since it retains much of the OLS problems for ill-conditioned design matrices (Brown et al., 2002) which are ubiquitous in genetic mapping due to the LD structure. I would remove the remark altogether as well as the reference to (generalised) g-prior. In any case, please specify what $W(\hat{\beta}_\gamma)$ is.

- P5-L167: As far as I remember, Kohn et al. (2001) first introduced the flexible Beta-Binomial prior $p(\gamma)$ with the elicitation of a and b. Please correct the reference.

- P5-L183: I'm not sure I understood the MCMC scheme to update the parameter k of the Weibull distribution. It seems that for the Weibull regression, the Authors suggest working with a conditional (on k) marginal likelihood for the update of γ. However, I thought all the machinery works because either $p(\gamma|y)$ is known analytically or (an unbiased) $\hat{p}(\gamma|y)$ can be estimated. Could you please clarify this point, adding also details of the MH step with an adaptive random walk proposal to update $k|y, \gamma$?

- P8-L196: In eq. (21), please add that $a > 0$ and $b > 0$. Just after eq. (22), please specify what W is.

- P9-L206: Could you please specify (or add a reference) whether $p_{LA}(\gamma|y)$ is an unbiased estimator of the true marginal likelihood $p(\gamma|y)$?

- P9-L238: Would be also possible to say the same regarding $p_{ALA}(\gamma|y)$? Is it unbiased?

- P11-L264: In eq. (34), I think there is a typo. It should be $\hat{p}(y|\gamma')$ and $\hat{p}(y|\gamma)$, to keep it consistent with eq. (19), although when the PG is used in the logistic regression, the marginal likelihood is known analytically.

- P12-L270: The variable η_i in the derivatives is not defined (it is defined later as the linear predictor). I'm also wondering at which value $\eta_i^{(0)}$ the derivatives are

evaluated. Does the choice of $\eta_i^{(0)}$ impact the warm-start approximation $\tilde{\pi}_j^{(0)}$?

- P12-L271: I'm not sure I understood the sentence: "If the data contains many highly correlated covariates, ...". Could the Authors explain it?

- P14-L339: Is it possible to see how the different instances of the proposed method perform in terms of support recovery? Mixing is important (perhaps with an evaluation of the accumulation of the posterior mass as long as the algorithm proceeds, see for instance Hans et al. (2007)), but it is also important to check how the algorithm (and its instances) perform in practice. Comparisons with SSS (Hans et al., 2007) and HBS (Titsias and Yau, 2017) will shed light in this respect.

- P16-L384: I've several comments regarding how the comparisons in the real data analysis have been performed. Is it fair to compare instances of the proposed BVS algorithm which have different computational complexity? PARNI-DA-DA is more parsimonious than PARNI-aALA-CPM because to estimate CPM in eq. (19) data need to be touched N times. So, running both instances for 1 hour is perhaps not a fair comparison. Could the Authors think of how to make the comparison more fair? Also, I'm not sure why the Authors declared PARNI-aALA-CPM (ran for 12 hours) the " gold standard". I would have chosen PARNI-DA-DA (ran as well for 12 hours) as the gold standard since the marginal likelihood is available in a closed-form expression. Finally, comparisons with SSS and HBS should be also provided when evaluating the support recovery and the out-out-sample prediction. In my view, the baseline ADS is too simple and many different approaches have been proposed to overcome the problems of slow mixing due to the multi-modality of the ultra-large model space.

***

References

Brown, P. J., M. Vannucci, and T. Fearn (2002). Bayes Model Averaging with selection of regressors. *J. R. Stat. Soc. B 64*(3), 519–536.

Hans, C., A. Dobra, and M. West (2007). Shotgun stochastic search for "large p" regression. *J. Am. Stat. Assoc. 102*(478), 507–516.

Jasra, A., D. A. Stephens, and C. C. Holmes (2007). Population-based reversible jump Markov chain Monte Carlo. *Biometrika 94*, 787–807.

Kohn, R., M. Smith, and D. Chan (2001). Nonparametric regression using linear combinations of basis functions. *Stat. Comput. 11*(4), 313–322.

Titsias, M. K. and C. Yau (2017). The Hamming ball sampler. *J. Am. Stat. Assoc. 112*(520), 1598–1611.