Biometrical Journal **xx** (2014) yy, zzz–zzz / DOI: 10.1002/bimj.200100000

Supporting information for the Biometrical Journal publication:

Bayesian hierarchical modelling of continuous non-negative longitudinal data with a spike at zero: an application to a study of birds visiting gardens in winter

Ben T. Swallow[*,1], **Stephen T. Buckland** [1], **Ruth King** [1], and **Mike P. Toms** [2]

[1] Centre for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, University of St Andrews, KY16 9LZ, UK

[2] British Trust for Ornithology, The Nunnery, Thetford, Norfolk, IP24 2PU, UK

1 Sequential tuning of random walk Metropolis algorithm proposal variance

The random walk Metropolis algorithm is a frequently used Markov chain Monte Carlo algorithm which requires the definition of a distribution $q(.)$ to generate proposal values for the parameters of interest. Although arbitrary, the efficiency of the chain is highly dependent on the specification of this distribution. To improve the mixing and efficiency of the algorithm, we follow Sherlock *et al.* (2010), using Algorithm 2 of the sequential tuning approach outlined. We denote θ_i the parameters to be updated via the random walk Metropolis algorithm and $\theta_i^{(\xi)}$ the value of the parameter at the current iteration. The variances λ_i^2 of the proposal distribution $q(\theta_i^{(\xi)}, \lambda_i^2)$ are tuned independently for each parameter. For a Gaussian proposal and target distribution, the optimum acceptance probability is between 0.4 and 0.45. The target distributions in this application are, however, not Gaussian so we found that tuning the proposals to give acceptance probabilities between 0.1 and 0.8 gave the best mixing. In general, the sequential tuning algorithm used here works on the basis that if the proposal variance is too small, acceptance probabilities tend to be high (and hence the chain will move slowly around parameter space). Conversely, variances that are too large will produce acceptance probabilities that are smaller and hence the chain will get stuck at the same value for longer periods.

During the burn-in phase, for parameter θ_i:

1. At iteration ξ of the MCMC algorithm, with current parameter value $\theta_i^{(\xi)}$, generate new parameter value ϕ from $q(\theta_i^{(\xi)}, \lambda_i^{2(\xi)})$.

2. Calculate the acceptance probability,

$$\alpha(\theta_i^{(\xi)}, \phi) = \min\left(1, \frac{\pi(\phi|.)q(\theta_i^{(\xi)}, \lambda_i^{2(\xi)})}{\pi(\theta_i^{(\xi)}|.)q(\phi, \lambda_i^{2(\xi)})}\right).$$

   where $\pi(\phi|.)$ is the posterior conditional distribution of parameter ϕ, coinditional on the other parameters in the model and the data.

*Corresponding author: e-mail: bts3@st-andrews.ac.uk

3. Set $\theta_i^{(\xi+1)} = \phi$ with probability $\alpha(\theta_i^{(\xi)}, \phi)$.
4. If $\alpha(\theta_i^{(\xi)}, \phi) < 0.1$ set $\lambda_i^{2(\xi+1)} = \lambda_i^{2(\xi)}/2$.
5. If $\alpha(\theta_i^{(\xi)}, \phi) > 0.8$ set $\lambda_i^{2(\xi+1)} = 1.5\ \lambda_i^{2(\xi)}$.
6. Set $\xi = \xi + 1$ and return to step 1 until $\xi =$ burn-in.

2 Reversible jump algorithm

The reversible jump algorithm used in the paper can be defined as follows. We denote $\boldsymbol{\theta}_m = \{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ for all β_j and γ_k in model m. At each iteration, a single regression parameter is proposed to be added or removed from the model, depending on whether or not it is in model m.

1. Suppose that at iteration ξ the Markov chain is in model m. Update all regression parameters in $\boldsymbol{\theta}_m$ and additional variance parameters (ϕ, p and $\boldsymbol{\epsilon}$) using the random walk Metropolis algorithm and σ^2 via the inverse-gamma Gibbs algorithm, conditional on model m.
2. Select one of the regression parameters at random. Propose to move to a new neighbouring model m' with probability $p(m'|m)$ and associated regression parameter vector $\boldsymbol{\theta}'_m$.

   If the proposed parameter is not currently in the model, then $\boldsymbol{\theta}_m = \{\boldsymbol{\delta}\}$ and $\boldsymbol{\theta}'_m = \{\boldsymbol{\delta}', \kappa'\}$,

   i. Set $\boldsymbol{\delta}' = \boldsymbol{\delta}$ and $\kappa' = u$, where $u \sim q(u)$.
   ii. Calculate acceptance probability in min$(1, A)$, where,

   $$A = \frac{\pi(\boldsymbol{\theta}'_m, m'|.)p(m|m')}{\pi(\boldsymbol{\theta}_m, m|.)p(m'|m)q(u)} \left| \frac{\partial(\boldsymbol{\delta}', \kappa')}{\partial(\boldsymbol{\delta}, u)} \right|.$$

   iii. With probability min$(1, A)$ set $(\boldsymbol{\theta}_m^{\xi+1}, m^{\xi+1}) = (\{\boldsymbol{\delta}', \kappa'\}, m')$, else $(\boldsymbol{\theta}_m^{\xi+1}, m^{\xi+1}) = (\boldsymbol{\delta}, m)$.

   Else if we propose the reverse move from $\boldsymbol{\theta}'_m = \{\boldsymbol{\delta}', \kappa'\}$ and $\boldsymbol{\theta}_m = \{\boldsymbol{\delta}\}$,

   i. Remove κ' from the model (equivalently set $u = \kappa'$) and calculate A as above.
   ii. With probability min$(1, A^{-1})$ set $(\boldsymbol{\theta}_m^{\xi+1}, m^{\xi+1}) = (\boldsymbol{\delta}, m)$, else $(\boldsymbol{\theta}_m^{\xi+1}, m^{\xi+1}) = (\{\boldsymbol{\delta}', \kappa'\}, m)$.

Using the identity function as the bijective function means that the Jacobian,

$$\left| \frac{\partial(\boldsymbol{\delta}', \kappa')}{\partial(\boldsymbol{\delta}, u)} \right| = \begin{vmatrix} \frac{\partial \boldsymbol{\delta}'}{\partial \boldsymbol{\delta}} & \frac{\partial \boldsymbol{\delta}'}{\partial u} \\ \frac{\partial \kappa'}{\partial \boldsymbol{\delta}} & \frac{\partial \kappa'}{\partial u} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

Equal prior probabilities are specified on the covariates being present or absent i.e. $p(m|m') = p(m'|m) = 0.5$ and hence A can be simplified to,

$$A = \frac{\pi(\boldsymbol{\theta}', m'|.)}{\pi(\boldsymbol{\theta}, m|.)q(u)}.$$

The parameters of the proposal distribution $q(u) = N(\hat{\theta}_i, \hat{\lambda}_i^2)$ are estimated from a previous analysis conducted in the saturated model. The model was run for 20,000 iterations and the first 5,000 were discarded as burn-in. Posterior means and variances $\hat{\theta}_i$ and $\hat{\lambda}_i^2$ correpsond to the posterior means and variances of the parameters respectfully.

References

Sherlock, C., Fearnhead, P. and Roberts, G. O. (2010) The Random Walk Metropolis: Linking Theory and Practice Through a Case Study. *Statistical Science* **25**, 172-190

