Appendix to "A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo"

Lei Gong
Department of Statistics, University of California, Riverside
and
James M. Flegal
Department of Statistics, University of California, Riverside

January 9, 2015

In this appendix, we summarize the spatial Bayesian variable selection model on functional magnetic resonance imaging (fMRI) data proposed by Lee et al. (2014) and specify the model formulations for the StarPlus data set that we are interested in.

1 Spatial Bayesian variable selection model

For voxel $v = 1, \ldots, N$, let $\{y_{v,i}; i = 1, \ldots, t\}$ be the BOLD image intensities at t time points. Although other alternatives are possible, a conventional voxelwise regression analysis assumes a linear model with a balance between model complexity and computational feasibility (Friston et al., 1995; Smith and Fahrmeir, 2007),

$$y_{v,i} = z_i^T a_v + x_{v,i}\beta_v + \epsilon_{v,i}.$$

Linear combination $z_i^T a_v$ is the baseline trend to remove stimulus-independent effects. β_v is the activation amplitude and $x_{v,i}$ is the transformed stimulus (see Figure 1). In many experiments, the external stimulus $\{s_i; i = 1, \ldots, t\}$ alternates activation/inactivation in a 0-1 'boxcar' pattern. However, instead of proceeding in a 0-1 'boxcar' function, the brain produces a fairly fixed, stereotyped blood flow response with delay d_v every time a stimulus hits it, where d_v is estimated in a preprocessing step. The so-called hemodynamic response function (HRF) is used to characterizes this process. There are several formulations of HRF (see e.g. Friston et al., 1998; Glover, 1999; Gössl et al., 2001).

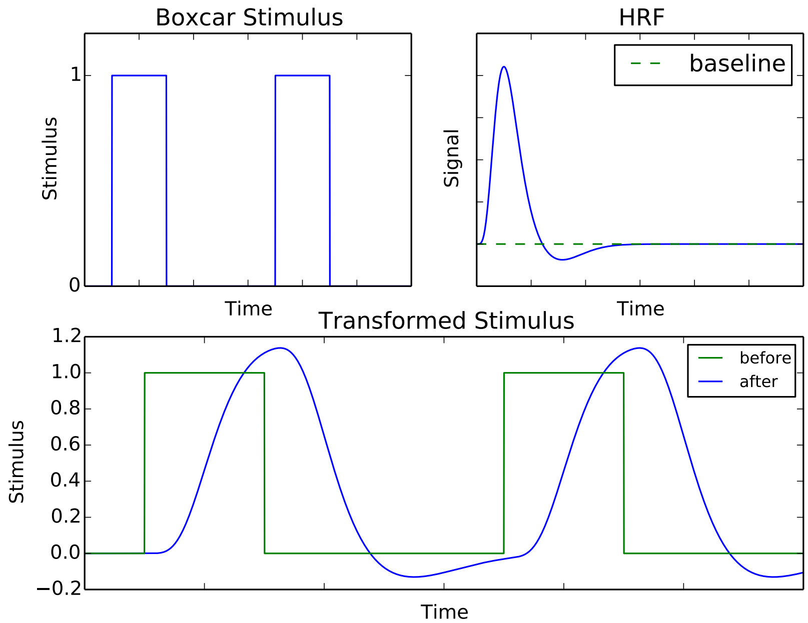

Figure 1: The transformed stimulus is obtained by convolving the original 0-1 'boxcar' stimulus and the HRF.

One approach is to use a canonical HRF consisting of a difference of two gamma functions (Lindquist et al., 2009),

$$h(t) = A\left(\frac{t^{\alpha_1-1}\beta_1^{\alpha_1}e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c\frac{t^{\alpha_2-1}\beta_2^{\alpha_2}e^{-\beta_2 t}}{\Gamma(\alpha_2)}\right),$$

where $\alpha_1 = 6$, $\alpha_2 = 16$, $\beta_1 = \beta_2 = 1$ and $c = 1/6$. The only unknown parameter, i.e. the amplitude A, is estimated in a preprocessing step. We can transform the orignal 'boxcar' stimulus by a convolution with the HRF,

$$x_{v,i} = \sum_{k=0}^{i-d_v} h(k)s_{i-d_v-k}.$$

The measurement error is denoted by $\epsilon_{v,i}$. Appropriate distributional assumptions about $\epsilon_{v,i}$ can be made to incorporate temporal correlation and specific priors can be chosen to reflect spatial dependence.

In this article, we consider the spatial Bayesian variable selection models for single subject (Lee et al., 2014). This approach is shown to incorporate temporal-spatial correlation and allow for the task-related changes in BOLD response while mitigates the computational burden. It also possesses the ability to account for anatomic prior information. A general MCMC algorithm is designed to perform the large dimensional posterior inference. Here we summarize the model formulation and estimation process from Lee et al. (2014). An interested reader is directed to their paper for more details.

Denote $\boldsymbol{y}_v = (y_{v,1}, \ldots, y_{v,t})^T$ as the BOLD image intensity at time $i = 1, \ldots, t$ for voxel $v = 1, \ldots, N$. Let X_v be a $t \times p$ design matrix of transformed stimulus and $\boldsymbol{\beta}_v = (\beta_{v,1}, \ldots, \beta_{v,p})^T$ be a vector of p regression coefficients for each voxel. We formulate a linear regression mode,

$$\boldsymbol{y}_v = X_v \boldsymbol{\beta}_v + \boldsymbol{\epsilon}_v, \quad \boldsymbol{\epsilon}_v \sim N_t\left(\mathbf{0}, \sigma_v^2 \Lambda_v\right). \tag{1}$$

Notice that the detection of voxel activation is equivalent to the identification of nonzero $\boldsymbol{\beta}_v$s. To this end, we introduce 0/1 binary indicators $\boldsymbol{\gamma}_v = (\gamma_{v,1}, \ldots, \gamma_{v,p})$, $v = 1, \ldots, N$, such that $\beta_{v,j} = 0$ if $\gamma_{v,j} = 0$ and $\beta_{v,j} \neq 0$ if $\gamma_{v,j} = 1$. The $\gamma_{v,j}$ is used to indicate whether the voxel v is activated by input stimulus j. Given $\boldsymbol{\gamma}_v$, let $\boldsymbol{\beta}_v(\boldsymbol{\gamma}_v)$ be the vector of nonzero regression coefficients and $X_v(\boldsymbol{\gamma}_v)$ be the corresponding design matrix. Then, the model (1) can be rewritten as

$$\boldsymbol{y}_v = X_v(\boldsymbol{\gamma}_v)\, \boldsymbol{\beta}_v(\boldsymbol{\gamma}_v) + \boldsymbol{\epsilon}_v.$$

Further, we assume the independence among σ_v^2 and set its prior $\pi(\sigma_v^2) \propto 1/\sigma_v^2$. Zellner's g-prior on $\boldsymbol{\beta}_v(\boldsymbol{\gamma}_v)\,|\boldsymbol{\gamma}_v$ is placed to undertake variable selection or model averaging. The parameter g is adjusted to obtain similar results with those if BIC were used,

$$\boldsymbol{\beta}_v(\boldsymbol{\gamma}_v)\,|\boldsymbol{y}_v, \sigma_v^2, \Lambda_v, \boldsymbol{\gamma}_v \sim N\left(\hat{\boldsymbol{\beta}}_v(\boldsymbol{\gamma}_v), T_v \sigma_v^2 \left[X_v^T(\boldsymbol{\gamma}_v)\, \Lambda_v^{-1} X_v(\boldsymbol{\gamma}_v)\right]^{-1}\right),$$

where

$$\hat{\boldsymbol{\beta}}_v(\boldsymbol{\gamma}_v) = \left[X_v^T(\boldsymbol{\gamma}_v)\, \Lambda_v^{-1} X_v(\boldsymbol{\gamma}_v)\right]^{-1} X_v^T(\boldsymbol{\gamma}_v)\, \Lambda_v^{-1} \boldsymbol{y}_v. \tag{2}$$

Define the corresponding sum of squares for posterior inference

$$S\left(\rho_v, \boldsymbol{\gamma}_v\right)=\left(\boldsymbol{y}_v-X_v\left(\boldsymbol{\gamma}_v\right) \hat{\boldsymbol{\beta}}_v\left(\boldsymbol{\gamma}_v\right)\right)^T \Lambda_v^{-1}\left(\boldsymbol{y}_v-X_v\left(\boldsymbol{\gamma}_v\right) \hat{\boldsymbol{\beta}}_v\left(\boldsymbol{\gamma}_v\right)\right).$$

We incorporate the temporal dependence between observations on a given voxel through the specification of the structure of Λ_v. The AR(1) dependence, i.e. $\Lambda_v(i, j)=\rho_v^{|i-j|}$, is an effective compromise between inferential efficacy and computational efficiency. We specify a point mass prior for $\boldsymbol{\rho}=\left(\rho_1, \ldots, \rho_N\right)$ at a fixed point $\hat{\boldsymbol{\rho}}$ using maximum likelihood methods.

We incorporate the spatial dependence, as well as the anatomical information, by using a binary Markov random field (MRF) prior, i.e. the Ising prior, on $\boldsymbol{\gamma}_v$. Let $\boldsymbol{\gamma}_{(j)}=\left(\gamma_{1, j}, \ldots, \gamma_{N, j}\right)^T$ be the vector of indicators for regressor j over all voxels. Then, let $w_{v, k}$ be pre-specified constants that weigh the interaction between voxels v and k and let ν_j be parameter to measure the strength of the interaction between voxels for regressor j. We denote $v \sim k$, if two voxels v and k are defined as neighbors by the user. In this article, we employ a widely used three-dimensional structure containing the six immediate neighbors: 1 above, 1 below and 4 adjacent. The weight $w_{v, k}$ is set to be the reciprocal of the Euclidean distance between voxel v and k. Then, the spatial interaction is described as $\nu_j \sum_{v=1}^N \sum_{v \sim k} w_{v, k} I\left(\gamma_{v, j}=\gamma_{k, j}\right)$, where $I(x)$ is the usual 0/1 indicator function. A linear "external field" $\sum_{v=1}^N \alpha_{v, j} \gamma_{v, j}$ is specified to incorporate anatomical prior information, where $\alpha_{v, j}$ is chosen to reflect prior knowledge.

We consider the prior on $\boldsymbol{\gamma}$ to be $\pi(\boldsymbol{\gamma} \mid \boldsymbol{\nu})=\prod_{j=1}^p \pi\left(\boldsymbol{\gamma}_{(j)} \mid \nu_j\right)$, where

$$\pi\left(\boldsymbol{\gamma}_{(j)} \mid \nu_j\right) \propto \exp \left\{\sum_{v=1}^N \alpha_{v, j} \gamma_{v, j}+\nu_j \sum_{v=1}^N \sum_{v \sim k} w_{v, k} I\left(\gamma_{v, j}=\gamma_{k, j}\right)\right\}.$$

The remaining prior to be addressed is the distribution of $\boldsymbol{\nu}=\left(\nu_1, \ldots, \nu_p\right)$. A uniform prior is placed $\pi(\boldsymbol{\nu}) \propto \prod_{j=1}^p I\left(0<\nu_j<\nu_{max}\right)$, where Moller and Waagepetersen (2003) suggests to use $\nu_{max} \leq 2.0$.

The posterior density is characterized by

$$q\left(\boldsymbol{\beta}(\boldsymbol{\gamma}), \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma}^2 \mid y\right) \propto p\left(y \mid \boldsymbol{\beta}(\boldsymbol{\gamma}), \boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \Lambda\right) \times \pi\left(\boldsymbol{\beta}(\boldsymbol{\gamma}) \mid y, \boldsymbol{\sigma}^2, \Lambda, \boldsymbol{\gamma}\right) \pi(\boldsymbol{\gamma} \mid \boldsymbol{\nu}) \pi(\boldsymbol{\rho}) \pi\left(\boldsymbol{\sigma}^2\right) \pi(\boldsymbol{\nu}).$$

We follow the two-step component-wise Metropolis-hastings algorithm designed by Lee et al.

(2014) to update $\boldsymbol{\gamma}$ and $\boldsymbol{\nu}$. Particularly, we are interested in estimating the posterior mean of $\boldsymbol{\theta} = \{\boldsymbol{\gamma}, \boldsymbol{\nu}\}$.

1.1 Model the Starplus dataset

Based on the settings of the StarPlus experiment, we rewrite the linear model (1) as

$$\boldsymbol{y}_v = \alpha_0 \boldsymbol{z}_0 + \alpha_1 \boldsymbol{z}_1 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{\epsilon}_v,$$

where $\alpha_i, \boldsymbol{z}_i$s are the baseline signal, β_is are the activation amplitude corresponding to the two tasks "Semantic" and "Symbol", respectively, The binary indicator $\gamma_v = \{1, 1, \gamma_{v,3}, \gamma_{v,4}\}$ is used in the variable selection problem described previously. Notice that we assume all α_is nonzero and set $\nu_{max} = 1.0$ as in Lee et al. (2014). Figure 2 visualize the design matrix for this linear model as we described previously.

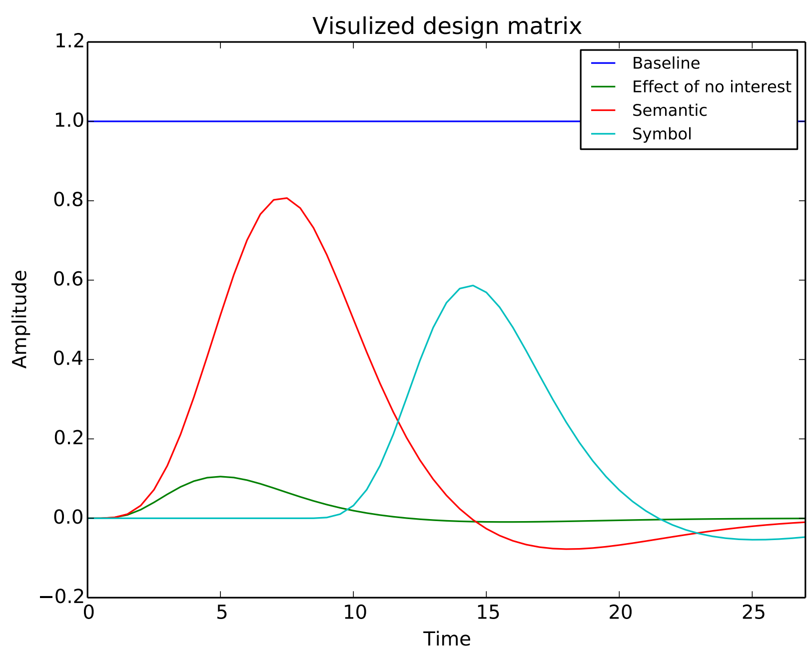

Figure 2: The visualization of the design matrix for the experimental dataset.

References

Friston, K., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., Frackowiak, R. S., et al. (1995). Spatial registration and normalization of images. *Human brain mapping*, 3(3):165–189.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1):30–40.

Glover, G. H. (1999). Deconvolution of impulse response in event-related bold fMRI. *Neuroimage*, 9(4):416–429.

Gössl, C., Auer, D. P., and Fahrmeir, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, 57(2):554–562.

Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. S. (2014). Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis*, 9(3):699–732.

Lindquist, M. A., Meng Loh, J., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage*, 45(1):S187–S198.

Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press.

Smith, M. and Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 102(478):417–431.