A Bayesian Generalized CAR Model for Correlated Signal Detection

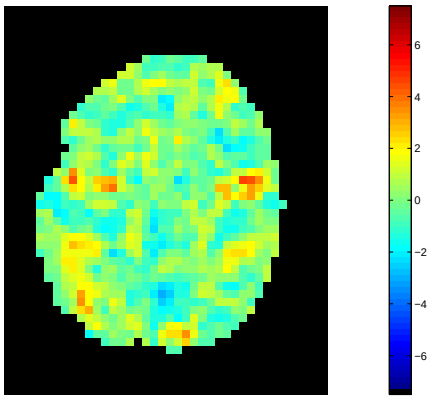
D. Andrew Brown[1], Gauri S. Datta[2], Nicole A. Lazar[2]

JSM
Chicago, IL
1 August 2016

[1] Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA
[2] Department of Statistics, University of Georgia, Athens, GA 30602, USA

Motivation

- Advances in computing technology have made possible the creation of massive datasets: $y_j,\ j = 1, \ldots, J,$ where J is large
- Microarray / RNA-seq: $y_j =$ test statistic quantifying differential expression of gene j between treatment conditions
- fMRI: $y_j =$ test statistic quantifying observed changes in BOLD signal at voxel j between stimulus and baseline conditions



B et al. (2014)

Dependence Problem

- Nontrivial dependence is known to exist in various applications
  - neuroimaging
  - syndromic surveillance
  - microarray
  - RNA sequencing
- Not accounting for it causes the variance to be poorly estimated, thereby obscuring the true effects
- Multiplicity corrections under dependence have been considered (e.g. FDR control under arbitrary dependence), but precision or power may be sacrificed
- Theoretical null distribution used to calculate the p values may be inappropriate

Benjamini and Yekutieli (2001), Genovese et al. (2002), Efron (2004, 2007), Subramanian et al. (2005), Banks et al. (2012), Lee et al. (2014), Love et al. (2014)

Two Groups Model

- Test statistics $y_j \overset{\text{indep.}}{\sim} N(\theta_j, \sigma^2), \;\; j = 1, \ldots, J$
- Each y_j belongs to the null class ($\theta_j = 0$) or the non-null class ($\theta_j \neq 0$)
- Each location j has some *a priori* probability, p, of belonging to the null class
  - $p \approx 1$: Very sparse signals
  - $p \approx 0$: Abundant non-null cases
- Data-dependent estimation of p via empirical Bayes or hierarchical Bayes induces an intrinsic multiplicity adjustment in *a posteriori* probability statements such as $P(j \text{ is an active case} \mid \mathbf{y})$ (i.e., shrinkage)

Efron (2008), Scott and Berger (2010)

Scott and Berger (2006)

- Each θ_j can be expressed as a realization from a spike-Gaussian mixture:

$$\pi(\theta_j \mid p, \tau^2) = p\delta_0(\theta_j) + (1-p)\varphi_{0,\tau^2}(\theta_j)$$

- Take $p \sim \text{Beta}(\alpha, 1)$, for suitably chosen α
  - Choice of α reflects expected sparsity in the data (i.e., higher $\alpha \Rightarrow$ more cases expected to be null cases)

- Quantities of interest are the posterior inclusion probabilities, $1 - p_j = P(\theta_j \neq 0 \mid \mathbf{y})$

Mitchell and Beauchamp (1988), George and McCulloch (1993), Chipman (1996), Barbieri and Berger (2004), Clyde and George (2004), Scott and Berger (2006), Carvalho et al. (2008), Morris et al. (2011)

Conditional Autoregressive (CAR) Model

- If $\{X_i = X(\mathbf{s}_i) : \mathbf{s}_i \in \mathcal{D} \subset \mathbb{R}^p, i = 1, \ldots, J\}$ is a process defined on an array, then the Gaussian CAR model assumes that

$$X_i \mid \mathbf{x}_{(-i)} \sim N\left(\eta_i + \sum_{j=1}^{J} c_{ij}(x_j - \eta_j), \quad \sigma_i^2\right),$$

  where $c_{ij}/\sigma_i^2 = c_{ji}/\sigma_j^2$, $c_{ii} = 0$, and $c_{ij} = 0$ except when $\mathbf{s}_i$ and $\mathbf{s}_j$ are neighbors

- The joint density is given by

$$f(\mathbf{x}) \propto \exp\left(-(1/2)(\mathbf{x} - \boldsymbol{\eta})^T \mathbf{D}^{-1}(\mathbf{I} - \mathbf{C})(\mathbf{x} - \boldsymbol{\eta})\right),$$

  where $\mathbf{C} = \{c_{ij}\}_{i,j=1}^{J}$ and $\mathbf{D} = \text{diag}\{\sigma_i^2\}$.

Besag (1974), Besag and Kooperberg (1995), Rue and Held (2005)

- SB model can be expressed by writing $\theta_j = \gamma_j \mu_j$, $\gamma_j \mid p \overset{\text{iid}}{\sim} \text{Bernoulli}(1-p)$, and $\mu_j \mid \tau^2 \overset{\text{iid}}{\sim} N(0, \tau^2)$
- Instead of IID, use a CAR structure (IAR) in the prior on $\boldsymbol{\mu}$:

$$\pi(\boldsymbol{\mu} \mid \tau^2) \propto \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\mu}^T(\mathbf{D}_w - \mathbf{W})\boldsymbol{\mu}\right),$$

where $w_{ji} \neq 0$ iff sites j and i are neighbors, $w_{j\cdot} = \sum_{i=1}^{J} w_{ji}$, $\mathbf{D}_w = \text{diag}\{w_{j\cdot},\ j = 1, \ldots, J\}$ and $\mathbf{W} = \{w_{ji}\}_{j,i=1}^{J}$.

- Prior on $\boldsymbol{\mu}$ must be proper $\rightarrow$ introduce 'propriety parameter':

$$\mathbf{D}_w - \rho\mathbf{W}$$

$\rho \in (\lambda_1^{-1}, \lambda_J^{-1}) \Rightarrow$ non-singular, where $\lambda_1 < 0$ and $\lambda_J > 0$ are the smallest and largest eigenvalues of $\mathbf{D}_w^{-1/2}\mathbf{W}\mathbf{D}_w^{-1/2}$

Besag et al. (1991), Cressie (1993), McLachlan and Peel (2000), Banerjee et al. (2015)

Including Isolated Cases

- Standard CAR models assume every observation has at least one neighbor
- In some applications, the data include observations that share neighborhoods as well as isolated cases that have no neighbors
  - Including isolated cases directly induces zero rows in the precision matrix of the joint distribution
- Accounting for isolated cases with a standard CAR means (i) excluding isolated cases from the analysis, or (ii) forcing them into the model via inappropriate neighborhood structure
- We want to include isolated cases with appropriate assumptions to retain information.

- Modify the usual IAR model by introducing d so that $c_{ji} = w_{ji}/(d + w_{j\cdot})$ and conditional variance $\tau^2/(d + w_{j\cdot})$
- Precision matrix:
$$\tau^{-2}(\mathbf{D}_w + d\mathbf{I} - \mathbf{W})$$
- $d > 0 \Rightarrow$ include isolated points in the model while maintaining propriety
- $d = 1 \Rightarrow$ isolated cases reduce to $\mu_{j'} \mid \tau^2 \sim N(0,\ \tau^2)$ independent of other cases
- This facilitates independence of μ_j while allowing all cases to share information about the variance components through the prior distribution.

How to choose d?

Inclusion of ρ is still possible when $d > 0$, provided that ρ is bounded between the reciprocals of the smallest and largest eigenvalues of $(\mathbf{D}_w + d\mathbf{I})^{-1/2}\mathbf{W}(\mathbf{D}_w + d\mathbf{I})^{-1/2}$

Fact

$d_1 < d_2 \Rightarrow \lambda_{J,1}^{-1} < \lambda_{J,2}^{-1}$, where $\lambda_{J,i} > 0$ is the maximum eigenvalue of $(\mathbf{D}_w + d_i\mathbf{I})^{-1/2}\mathbf{W}(\mathbf{D}_w + d_i\mathbf{I})^{-1/2}$.

Proof: Rayleigh quotient + Perron-Frobenius Theorem

Larger values of ρ offset larger values of $d \Rightarrow$ model is insensitive to choice of d (under positive association).

B et al. (2016)

$$y_j \mid \gamma_j,\ \mu_j,\ \sigma^2 \overset{\text{indep}}{\sim} N(\gamma_j \mu_j,\ \sigma^2),\ \ j = 1, \ldots, J$$

$$\gamma_j \mid p \overset{\text{iid}}{\sim} \text{Bern}(1-p),\ \ j = 1, \ldots, J$$

$$\mu_j \mid \boldsymbol{\mu}_{(-j)},\ \tau^2,\ \rho \sim N\left(\sum_{i=1}^{J} \frac{\rho w_{ji} \mu_i}{d + w_{j\cdot}},\ \frac{\tau^2}{d + w_{j\cdot}}\right),\ \ d \geq 0,\ \ j = 1, \ldots, J$$

$$p \sim \text{Beta}(\alpha, 1),\ \ \alpha \geq 1$$

$$\rho \sim \text{Unif}(\nu_1^{-1},\ \nu_J^{-1})$$

$$\pi_{\tau^2 \mid \sigma^2}(\tau^2 \mid \sigma^2) = \left(\frac{1}{\sigma^2}\right)\left(1 + \frac{\tau^2}{\sigma^2}\right)^{-2},\ \ \tau^2 > 0$$

$$\pi_{\sigma^2}(\sigma^2) = \frac{1}{\sigma^2},\ \ \sigma^2 > 0$$

$\nu_1, \nu_J =$ smallest and largest eigenvalues of $(\mathbf{D}_w + d\mathbf{I})^{-1/2}\mathbf{W}(\mathbf{D}_w + d\mathbf{I})^{-1/2}$

B et al. (2016)

Proposition

The posterior distribution of the proposed model is proper.

In other words,

$$\sum_{\boldsymbol{\gamma} \in \{0,1\}^J} \int_p \int_{\boldsymbol{\Theta}} f(\mathbf{y}, \boldsymbol{\Theta}, \boldsymbol{\gamma}, p) d\boldsymbol{\Theta} dp < \infty, \ \forall \mathbf{y} \ (a.e.),$$

where $\boldsymbol{\Theta} = (\boldsymbol{\mu}^T, \sigma^2, \tau^2, \rho)^T$.

Proof: Consider $f(\mathbf{y}, \boldsymbol{\mu}, \sigma^2, \tau^2, \rho \mid \boldsymbol{\gamma})$ separately for the cases in which (i) $|\{i : \gamma_i = 1\}| = J$ and (ii) $|\{i : \gamma_i = 1\}| < J$

B et al. (2016)

Illustration

Gaussian data on a 20 $\times$ 20 grid with $\mu = 0$ (null) or $\mu = 3.5$ (active). Activation pattern drawn from a Ising model (binary MRF).

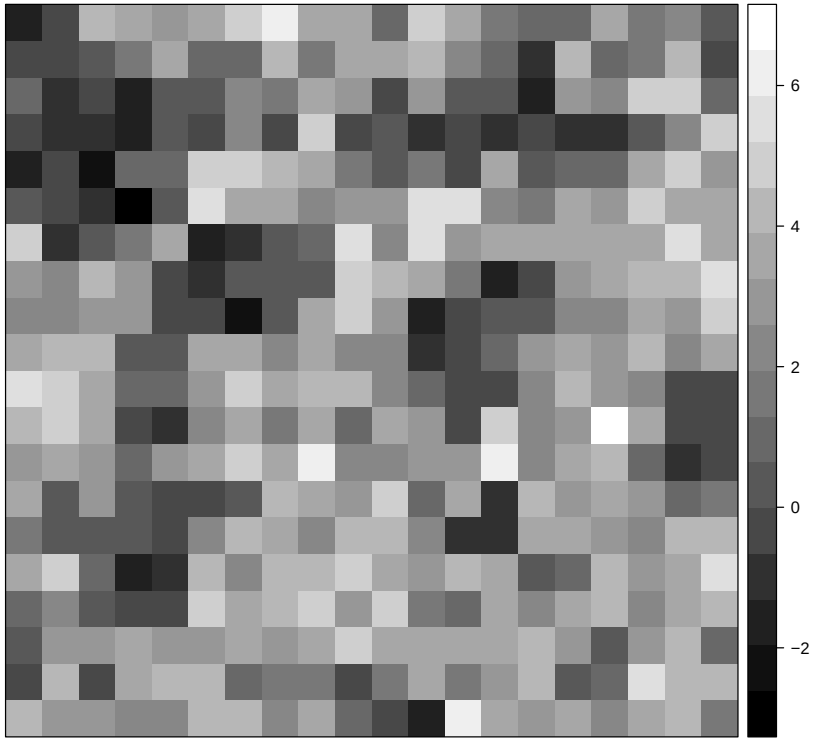
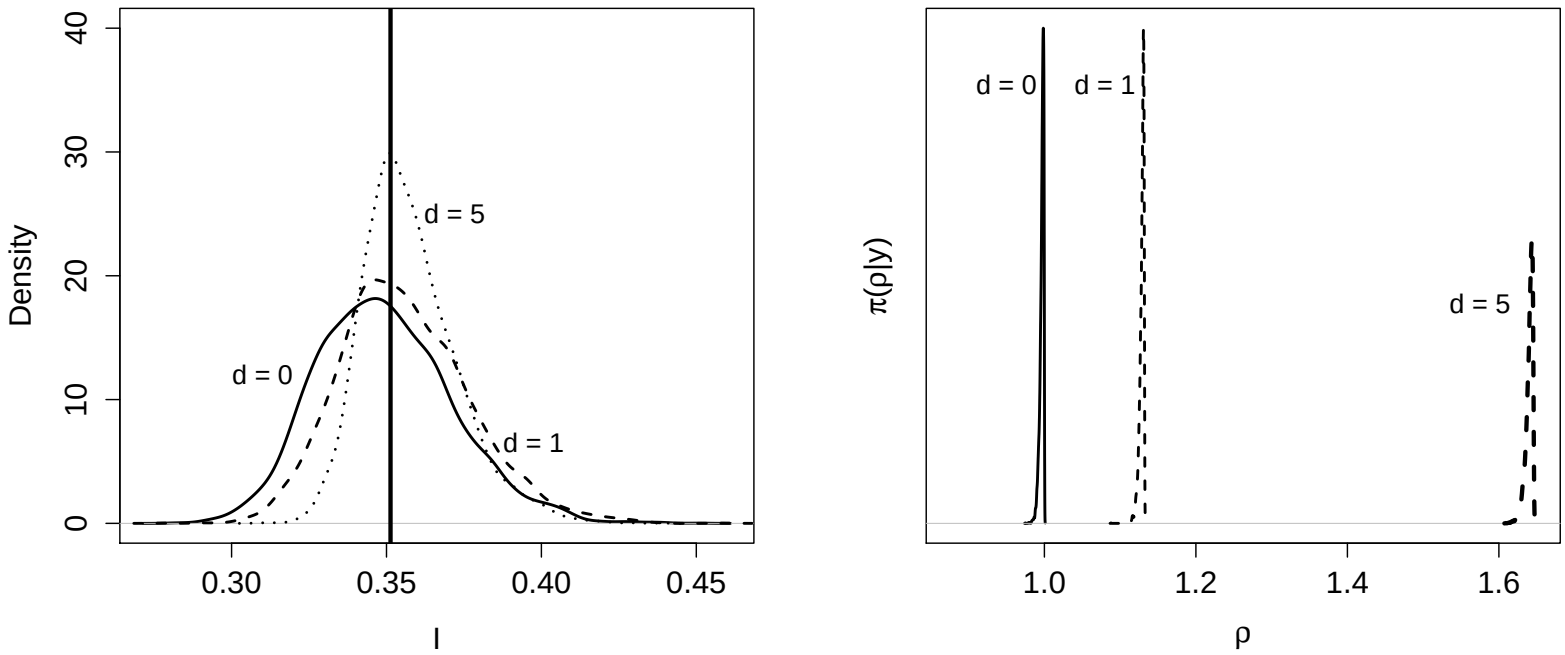


Figure : Smoothed histograms of 2,000 realizations of Moran's I from the corresponding posterior predictive distributions (left panel) and estimated marginal posterior distributions of ρ (right panel). The dark vertical line in the left panel is at the observed value, $I(\mathbf{y})$.

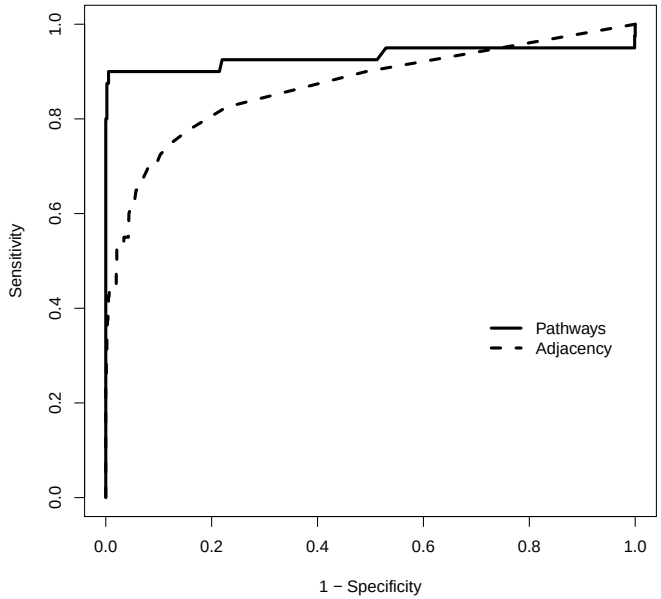


Simulation Study

- Simulate 1,000 gene expressions each for ten subjects (5 control, 5 treatment)
- j^{th} gene, subject $i = X_{ij} \sim N(\mu_{ij}, 1)$.
- Gene sets: Five groups of ten each (genes 11-20, 111-130, etc.)
- Control: $\mu_{ij} = 0,\ i = 1, \ldots, 5;\ j = 1, \ldots, 1000$
- Treatment: Sets 2 and 5 differentially expressed with $\mu = 2.5$ and $\mu = -1.5$, respectively
- Genes labeled and randomly permuted, to distinguish physical adjacency from pathway membership
- Test statistics: $y_j = \Phi^{-1}(F_T(t_j)),\ j = 1, \ldots, 1000$

Erroneously assuming physical-adjacency neighborhoods

Figure : Empirical ROC curves for the testing model using the physical-adjacency CAR model and the generalized CAR using pathways to define neighborhoods with isolated points included.



Correct neighborhoods, discarding isolated cases

Figure : Smoothed posterior estimates of p, σ^2, and τ^2 with and without the isolated cases.

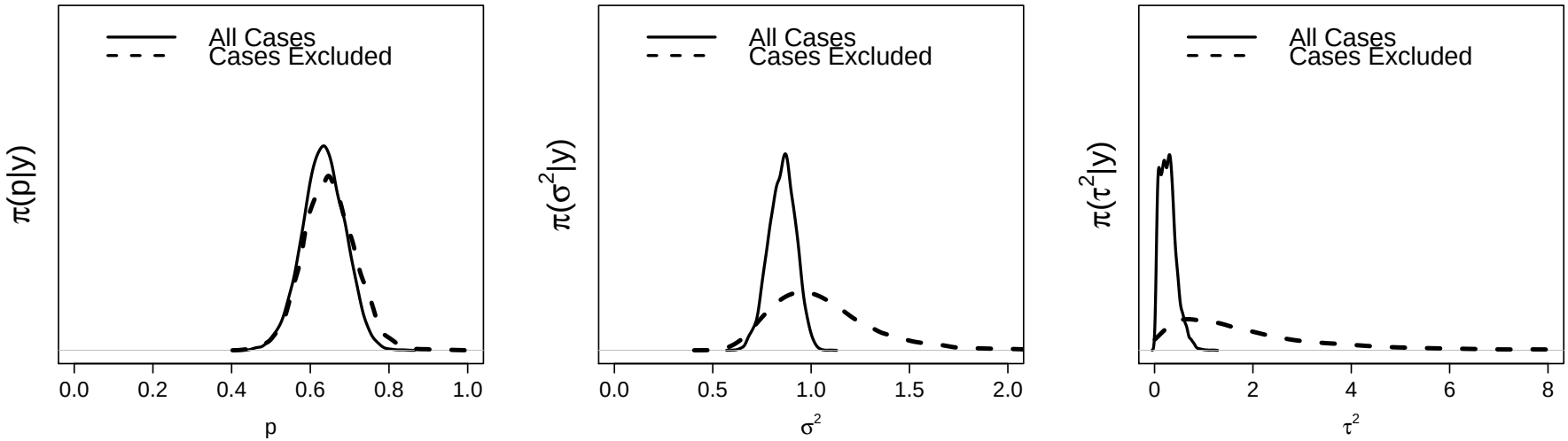
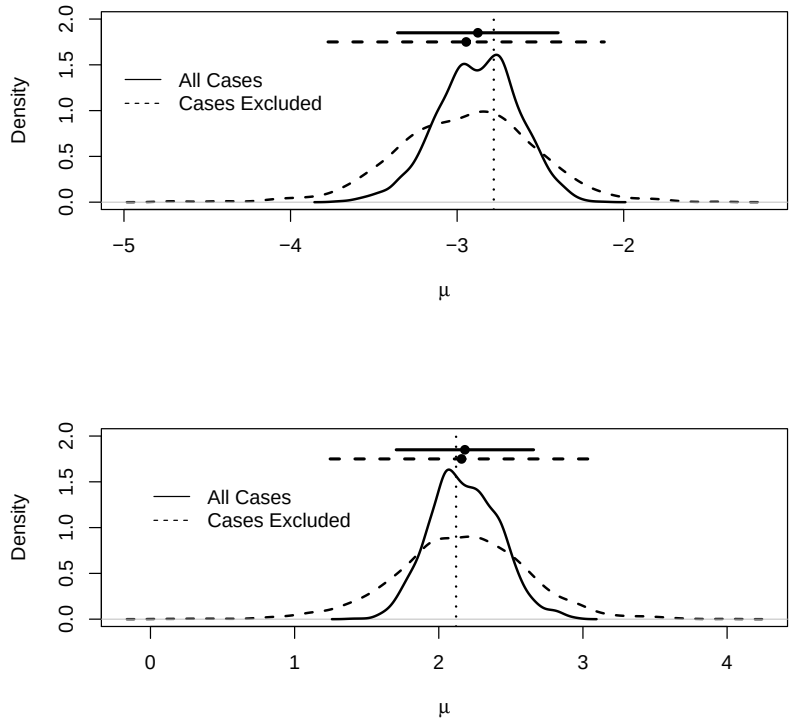


Table : Error rates for the simulated pathway example using pathway-based neighborhoods.

	With Isolated Cases	Without Isolated Cases
FNP	0.008	0.178
FDP	0.000	0.000
MCP	0.008	0.130

Figure : Smoothed approximate posterior densities of the signal strengths μ_j for two non-null cases in the simulated gene pathway example.



Summary

- Separate simulation illustrates a case where physical adjacency ($d = 0$) outperforms independence testing model and SAM procedure in terms of classification (not shown here)
- Feasibility is demonstrated by applying the proposed model to two different microarray datasets exhibiting two different types of dependence structures (not shown here)
- The proposed model offers a unified approach for simultaneous modeling of independent and dependent observations
- Inclusion of propriety parameter is recommended to reduce sensitivity to the specific weights assigned to neighbors

Subramanian et al. (2005), Scott and Berger (2006), Xiao et al. (2009), Efron (2010), **B** et al. (2016)

Brown, D. A., Datta, G. S., and Lazar, N. A. (2016), "A Bayesian generalized CAR model for correlated signal detection," *Statistica Sinica*, in press. `arXiv:1512.03769`

Brown, D. A., Lazar, N. A., Datta, G. S., Jang, W., and McDowell, J. E. (2014), "Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging," *NeuroImage*, 84, 97-112.

Thank you!