Objective Bayesian Analysis for Gaussian Hierarchical Models with Intrinsic Conditional Autoregressive Priors

Marco A. R. Ferreira[1]
Joint work with M.J. Keefe[2] and C.T. Franck[1]

[1]Virginia Tech; [2]Walt Disney Company

Spring Lecture Series – 2019

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

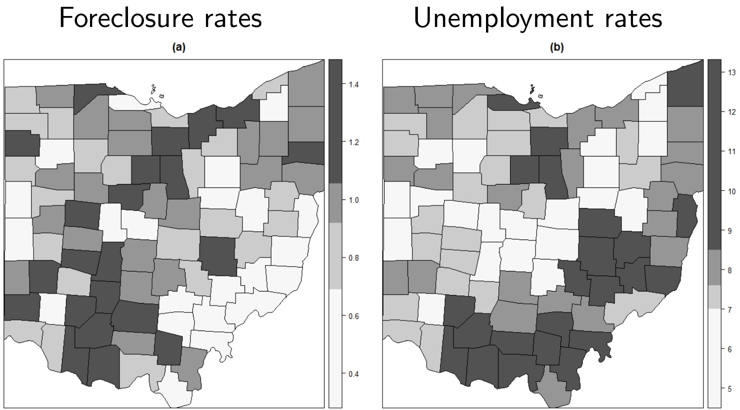
Objective priors

Simulation Study

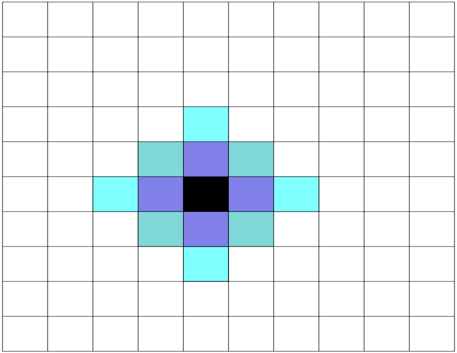
Application: foreclosures and unemployment in OH in 2012

Conclusions

Motivating example: foreclosures and unemployment in OH in 2012



CAR Neighborhood Structure



Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

Hierarchical model with ICAR prior

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\theta} + \boldsymbol{\phi},$$

where

- $\boldsymbol{Y}$ is the $n \times 1$ vector containing the response variable.
- X is a $n \times p$ matrix of covariates.
- $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients.
- $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$ is a vector of unstructured random effects such that $\theta_1, \ldots, \theta_n$ iid $N(0, \sigma^2)$.
- $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$ is a vector of spatial random effects following a sum-zero constrained intrinsic CAR prior.
- $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are assumed independent a priori.

Technical difficulty

- To obtain a reference prior for a spatial hierarchical model, we have to first integrate out the spatial random effects.
- This integral operation can only be performed when the spatial random effects have a well-defined distribution.
- Usual improper intrinsic CAR random effects do not have a well-defined distribution.
- That is probably the reason why a reference prior for this widely used spatial hierarchical model has not been published to date.

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

Brief literature review

- CAR specifications for modeling areal data (Besag, 1974).
- Intrinsic CAR model as a prior for spatial random effects (Besag, York and Mollié, 1991).
- Comprehensive coverage of Gaussian Markov random fields (Rue and Held, 2005).
- Comprehensive coverage of hierarchical models for spatial data (Banerjeee, Carlin and Gelfand, 2014).

Brief literature review

- OBayes for proper CAR to model observed areal data (Ferreira and De Oliveira, 2007; De Oliveira, 2012; Ren and Sun, 2013).
- OBayes analysis for geostatistical models (Berger et al, 2001; De Oliveira, 2007).
- OBayes for hierarchical models for areal data with proper CAR priors (Ren and Sun, 2014).

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

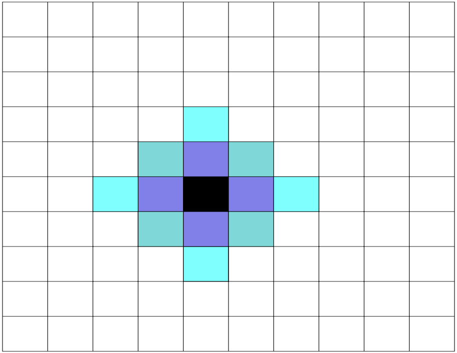
Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

CAR Neighborhood Structure



Improper Intrinsic CAR Models

The intrinsic CAR model for a vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)^T$ is specified by its conditional distributions

$$p(\omega_i | \boldsymbol{\omega}_{\sim i}) \propto \exp\left\{ -\frac{\tau_\omega}{2} \left[\sum_{i=1}^{n} \omega_i^2 h_i - 2 \sum_{i<j} \omega_i \omega_j g_{ij} \right] \right\},$$

where

- $\boldsymbol{\omega}_{\sim i}$ is the vector of the CAR elements for all subregions except subregion i;
- $\tau_\omega > 0$ is a precision parameter;
- $g_{ij} \geq 0$ is a measure of how similar subregions i and j are, $g_{ij} = g_{ji}$, and $h_i = \sum_{j=1}^{n} g_{ij}$.

Improper Intrinsic CAR Models

Alternatively, we may write the joint density for $\boldsymbol{\omega}$ as

$$p(\boldsymbol{\omega}) \propto \exp\left\{-\frac{\tau_\omega}{2}\boldsymbol{\omega}^T H \boldsymbol{\omega}\right\},$$

where H is a symmetric, positive semi-definite precision matrix defined as

$$(H)_{ij} = \begin{cases} h_i, & \text{if } i = j \\ -g_{ij}, & \text{if } i \in N_j \\ 0 & \text{otherwise.} \end{cases}$$

Sum-zero constrained ICAR prior

- Start with a vector of proper CAR random effects.
- Project this vector of random effects onto the subspace of $\mathbb{R}^n$ that is orthogonal to the subspace spanned by the vector $n^{-1/2}\mathbf{1}_n$.
- Take the limit to get a sum-zero constrained ICAR.

Proper CAR (Ferreira and De Oliveira, 2007)

We use a signal-to-noise ratio parametrization to express the proper CAR as

$$\phi^{**} \sim N\left(\mathbf{0}, \frac{\sigma^2}{\tau_c}\Sigma_\lambda\right),$$

where

- σ^2 and $\tau_c > 0$ are unknown parameters,
- $\lambda > 0$ is a propriety parameter,
- $\Sigma_\lambda^{-1} = \lambda I_n + H$,
- When $\lambda \to 0$ we get the improper intrinsic CAR.

Sum-Zero Constrained Proper CAR

- We define $\phi^* = P\phi^{**}$, where $P = \left(I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^T\right)$ is a centering (projection) matrix.
- The distribution of ϕ^* is given by

$$\phi^* \sim N\left(\mathbf{0}, \frac{\sigma^2}{\tau_c}\Sigma_{\phi\lambda}\right),$$

  where $\Sigma_{\phi\lambda} = P\Sigma_\lambda P^T$.
- By construction, $\sum_{i=1}^n \phi_i^* = 0$.

Sum-Zero Constrained ICAR

- Now we take the limit as $\lambda \to 0$ to obtain the final distribution for ϕ given by

$$\phi \sim N\left(\mathbf{0}, \frac{\sigma^2}{\tau_c}\Sigma_\phi\right),$$

- $\Sigma_\phi = \lim_{\lambda \to 0} \Sigma_{\phi\lambda} = QMQ^T$,
- $M = \text{diag}\left(d_1^{-1}, \ldots, d_{n-1}^{-1}, 0\right)$,
- $d_1 \geq \cdots \geq d_{n-1} > d_n = 0$ are the ordered eigenvalues of H
- Thus $\Sigma_\phi = H^+$ is the Moore-Penrose generalized inverse of H.

Sum-zero constrained ICAR prior (Keefe et al, 2018)

$$p(\phi) = (2\pi)^{(n-1)/2} \tau_c^{(n-1)/2} \left(\prod_{i=1}^{n-1} d_i \right)^{1/2} \exp\left\{ -\frac{\tau_c}{2\sigma^2} \phi' H \phi \right\} \mathbb{1}(\phi' \mathbf{1} = 0)$$

Compare to the Improper Intrinsic CAR

$$p(\boldsymbol{\omega}) \propto \exp\left\{-\frac{\tau_\omega}{2}\boldsymbol{\omega}^T H \boldsymbol{\omega}\right\}$$

Facts about the Sum-Zero Constrained ICAR

- Keefe et al (2018) consider proper CAR models with $\Sigma_{\lambda}^{-1} = \lambda K + H$.
- Let K be a symmetric positive semi-definite matrix for which the sum of its elements is positive.
- Keefe et al (2018) shows that for any matrix K in this class, the sum-zero constrained intrinsic CAR model does not depend on K and is, therefore, unique.

Facts about the Sum-Zero Constrained ICAR

- The singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}H^{+})$ is the stationary distribution of a one-at-a-time Gibbs sampler applied to the improper ICAR prior with centering on the fly (Ferreira, 20XXa).
- Centering spatial random effects $\boldsymbol{\omega}$ simulated from the full conditional distribution implied by the improper ICAR is equivalent to simulating from the full conditional distribution for ϕ implied by the $N(\mathbf{0}, \tau^{-1}H^{+})$ prior (Ferreira, 20XXb).
- Therefore, our reference prior is directly applicable to Gaussian hierarchical models with intrinsic CAR priors widely used in practice.

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

Objective priors

All of the objective priors we have derived fall into a class of priors of the form

$$\pi(\boldsymbol{\beta}, \sigma^2, \tau_c) \propto \frac{\pi(\tau_c)}{(\sigma^2)^a},$$

where $a \in \mathbb{R}$ is a hyperparameter and $\pi(\tau_c)$ is referred to as the marginal prior for τ_c.

Reference prior

- Let $G^* = I_n - X(X'X)^{-1}X'$.
- Let Q^* be the matrix with columns that are the normalized eigenvectors corresponding to the non-zero eigenvalues of G^*.
- Let $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_{n-p} > 0$ be the ordered eigenvalues of $Q^{*\prime} H^+ Q^*$.

Reference prior

The reference prior has $a = 1$ and

$$\pi^R(\tau_c) \propto \frac{1}{\tau_c}\left[\sum_{j=1}^{n-p}\left(\frac{\xi_j}{\tau_c+\xi_j}\right)^2 - \frac{1}{n-p}\left\{\sum_{j=1}^{n-p}\left(\frac{\xi_j}{\tau_c+\xi_j}\right)\right\}^2\right]^{1/2}.$$

(Keefe, Ferreira, and Franck, 2019)

Posterior propriety

(i) The posterior $\pi^R(\boldsymbol{\beta}, \sigma^2, \tau_c | \boldsymbol{y}, X)$ resulting from the reference prior $\pi^R(\tau_c)$ is proper.

(ii) The k^{th} moment of the marginal reference posterior $\pi^R(\tau_c | \boldsymbol{y}, X)$ does not exist for $k \geq 1$.

(iii) The independence Jeffreys prior leads to an improper posterior distribution.

(iv) The Jeffreys-rule prior leads to an improper posterior distribution.

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

Comparison of Priors

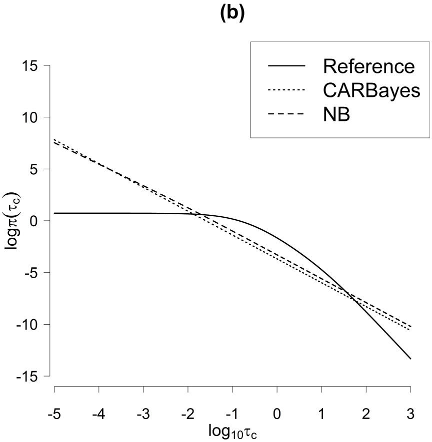
Competing priors:

- `CARBayes` package (version 4.0) has implemented gamma(0.001, 0.001) prior distributions as the default for both precision parameters.
- Best et al (1999) has used gamma(0.001, 0.001) and gamma(0.1, 0.1) prior distributions for the precisions of the unstructured and spatial random effects, respectively.

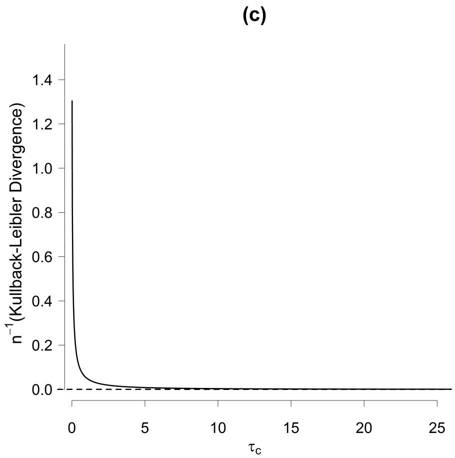
Performance is assessed using:

- frequentist coverage of credible interval.
- average interval length (IL).
- mean squared error (MSE).

Comparison of Priors



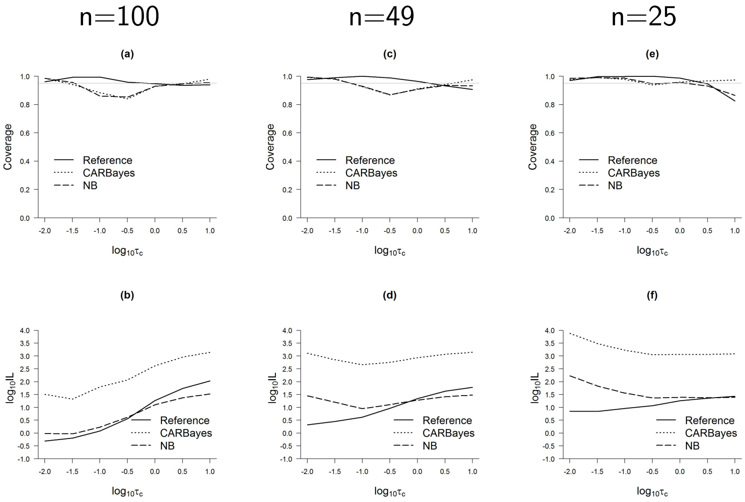
KL divergence between the spatial model and the independent model



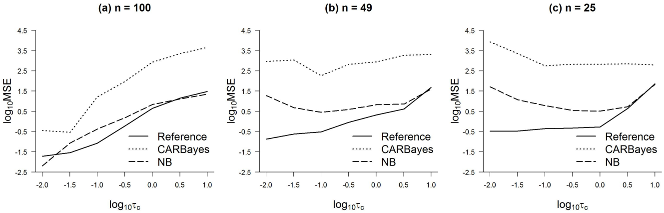
Simulation Design

- Square regions with sample sizes $n = 5^2$, 7^2, 10^2.
- $\sigma^2 = 2$.
- $\tau_c = 0.01, 0.032, 0.1, 0.32, 1, 3.2, 10$.
- First- and second-order neighborhood structure.
- $p = 1$ (intercept only) with $\boldsymbol{\beta} = 1$;
  and $p = 6$ with $\boldsymbol{\beta} = (-3, -2, -1, 1, 2, 3)'$.
- All covariates are generated from a normal distribution with mean 0 and variance 1.
- We generate results based on 1,000 simulated data sets for each combination of these levels of n, τ_c, neighborhood structure, and p.

Frequentist coverage and mean interval length for τ_c



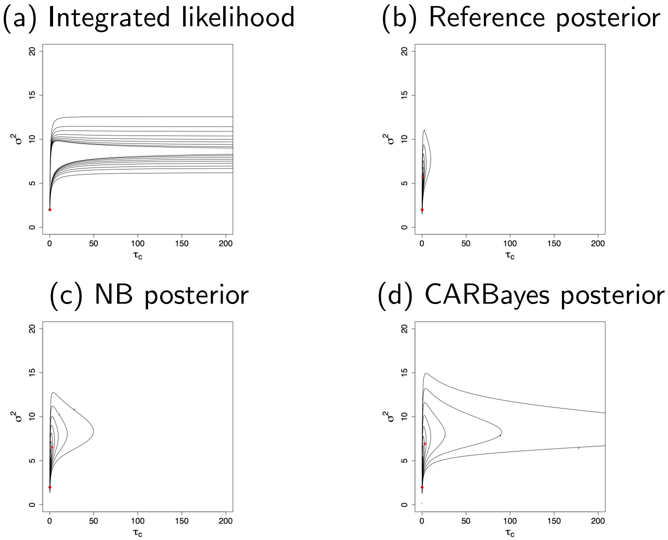
MSE for the posterior median of τ_c



One simulated dataset

- Dataset simulated with a first-order neighborhood structure, square regular grid, $n = 49$, and parameters $\beta = 1, \sigma^2 = 2$, and $\tau_c = 0.1$.
- The posterior medians of τ_c are equal to 1.0, 2.2, and 3.2 for the reference, NB, and CARBayes analyses, respectively.
- Contours correspond to HPD regions with credible levels equal to 10%, 20%, ..., 90%.

One simulated dataset



One simulated dataset

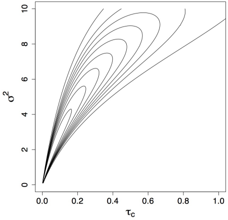
- We have found empirically that the probability of such a bad integrated likelihood behavior increases as the Moran-I statistic decreases.
- For this particular dataset, the Moran-I statistic is equal to 0.203 with a corresponding p-value of 0.0181 (null hypothesis is of no spatial dependence).
- This is not an extreme dataset; under the conditions used to simulate this dataset, the probability of the Moran-I statistic being less than 0.203 is about 0.12.

Another simulated dataset

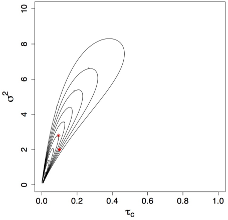
- We now consider a dataset simulated under the same conditions as the previous dataset, but with a Moran-I statistic of moderate size.
- Specifically, for this dataset the Moran-I statistic is equal to 0.363 with a corresponding p-value of 0.00017 (null hypothesis is of no spatial dependence).
- This Moran-I statistic is close to the median of the sampling distribution of the Moran-I statistics.

Another simulated dataset

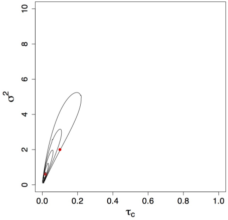
(a) Integrated likelihood



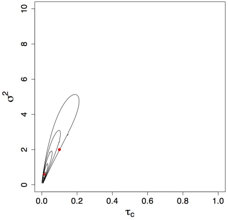
(b) Reference posterior



(c) NB posterior



(d) CARBayes posterior



Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

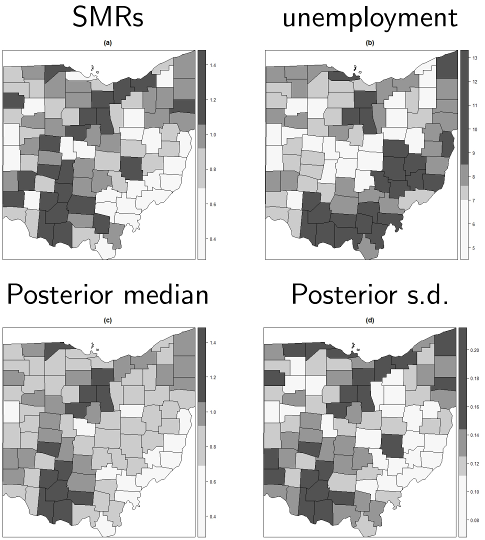
Application: foreclosures and unemployment in OH in 2012

Conclusions

Application: foreclosures and unemployment in OH in 2012

- We consider a data set containing foreclosure rates as a proportion of all housing transactions for each of the 88 counties in the state of Ohio for the year 2012.
- Let $\text{SMR}_i = O_i/E_i$, where O_i is the observed number of foreclosures in county i and E_i is the expected number of foreclosures in county i.
- The expected counts are calculated by $E_i = n_i \left(\sum_j O_j / \sum_j n_j \right)$, where n_i is the total number of housing transactions in county i.
- We consider $y_i = \log(\text{SMR}_i)$ as the response variable and unemployment rate as a covariate.

Foreclosures and unemployment in OH in 2012



Foreclosures and unemployment in OH in 2012

Prior	Parameter	Estimate	95% Credible Interval
Reference	β_0 (intercept)	-0.5561	(-0.9388, -0.1710)
	β_1 (unemployment rate)	0.0488	(0.0008, 0.0967)
	τ_c	0.2519	(0.0017, 0.9781)
	σ^2	0.0432	(0.0046, 0.0854)
CARBayes	β_0 (intercept)	-0.5612	(-0.9482, -0.1708)
	β_1 (unemployment rate)	0.0496	(0.0009, 0.0977)
	τ_c	0.1959	(0.0009, 1.0231)
	σ^2	0.0385	(0.0013, 0.0819)
NB	β_0 (intercept)	-0.5687	(-0.9516, -0.1802)
	β_1 (unemployment rate)	0.0504	(0.0020, 0.0985)
	τ_c	0.1589	(0.0010, 0.6290)
	σ^2	0.0345	(0.0019, 0.0755)

Foreclosures and unemployment in OH in 2012

Table: Foreclosures and unemployment in OH in 2012 – FBF-based posterior model probabilities

Model	Posterior Probability
Spatial Model intercept only	0.3603
Spatial Model with unemployment rate	0.6372
Independent Model intercept only	0.0019
Independent Model with unemployment rate	0.0006

Foreclosures and unemployment in OH in 2012

- Say we want to identify counties with risk higher than predicted by the regressors.
- A common decision rule would be to flag counties for which $P(\phi_i > 0|\boldsymbol{Y}) > 0.95$.
- For the OH dataset, the reference, CARBayes, and NB analyses identify 3, 4, and 9 counties respectively.

Outline

Motivating example

Hierarchical model with ICAR prior

Literature review

Sum-zero constrained ICAR prior

Objective priors

Simulation Study

Application: foreclosures and unemployment in OH in 2012

Conclusions

Conclusions

- Objective Bayesian analysis for a widely used Gaussian hierarchical model with ICAR prior for spatial data.
- Compared to two commonly used priors, the reference prior leads to a combination of favorable frequentist coverage, average interval length, and mean squared error.
- R package ref.ICAR available on CRAN (Porter et al., 2019).
- The reference prior proposed here leads to a proper posterior distribution and lets the data speak for themselves.