Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 4, 734–744
https://doi.org/10.1214/19-BJPS435

The limiting distribution of the Gibbs sampler for the intrinsic conditional autoregressive model

Marco A. R. Ferreira
Virginia Tech

Abstract. We study the limiting behavior of the one-at-a-time Gibbs sampler for the intrinsic conditional autoregressive model with centering on the fly. The intrinsic conditional autoregressive model is widely used as a prior for random effects in hierarchical models for spatial modeling. This model is defined by full conditional distributions that imply an improper joint "density" with a multivariate Gaussian kernel and a singular precision matrix. To guarantee propriety of the posterior distribution, usually at the end of each iteration of the Gibbs sampler the random effects are centered to sum to zero in what is widely known as centering on the fly. While this works well in practice, this informal computational way to recenter the random effects obscures their implied prior distribution and prevents the development of formal Bayesian procedures. Here we show that the implied prior distribution, that is, the limiting distribution of the one-at-a-time Gibbs sampler for the intrinsic conditional autoregressive model with centering on the fly is a singular Gaussian distribution with a covariance matrix that is the Moore–Penrose inverse of the precision matrix. This result has important implications for the development of formal Bayesian procedures such as reference priors and Bayes-factor-based model selection for spatial models.

1 Introduction

The intrinsic conditional autoregressive (ICAR) model is widely used as a prior for random effects in spatial hierarchical models. These models allow incorporation of neighborhood-based spatial dependence and have been used in diverse fields such as for example, disease mapping (Clayton and Kaldor, 1987, Bell and Broemeling, 2000, Goicoa et al., 2016), image restoration (Besag, York and Mollié, 1991), statistical analysis of fluid flow through porous media (Lee et al., 2002, Ferreira et al., 2003, Ferreira and Lee, 2007), and neuroimaging (Liu et al., 2016). Usually, analysis for such hierarchical models is implemented with Gibbs samplers based on full conditional distributions of the unknown parameters (Gelfand and Smith, 1990, Robert and Casella, 2005, Gamerman and Lopes, 2006). However, the ICAR model is defined by full conditional distributions that imply an improper joint "density" that has a multivariate Gaussian kernel with a singular precision matrix. By improper, we mean that the density does not integrate to one.

Key words and phrases. Areal data, ICAR models, Markov random fields, spatial data.
Received April 2018; accepted February 2019.

To guarantee propriety of the posterior distribution, usually at the end of each iteration of the Gibbs sampler the spatial random effects are centered to sum to zero in what is widely known as centering on the fly. While this works well in practice, this informal computational way to recenter the random effects does not make it explicit the mathematics behind the procedure. This lack of explicit mathematics obscures the implied prior distribution and prevents the development of formal Bayesian procedures. Here we show that the implied prior distribution, that is, the limiting distribution of the one-at-a-time Gibbs sampler for the intrinsic conditional autoregressive model with centering on the fly is a singular Gaussian distribution with a covariance matrix that is the Moore–Penrose inverse of the precision matrix.

This result is intuitive but it is not straightforward. The difficulty arises because of an apparent contradiction in the usual implementation of the one-at-a-time Gibbs sampler for the ICAR model with centering on the fly. Specifically, the one-at-a-time Gibbs sampler for the ICAR model is usually implemented in two steps. In the first step, the elements of an auxiliary vector $\boldsymbol{\phi}_t^*$ are simulated one at a time from the full conditional distributions implied by the ICAR model *but ignoring the sum-zero constraint*. We note that each such full conditional distribution depends only on the neighboring regions, which are typically of much lower cardinality than the total number of regions. In the second step, the sum-zero constraint is imposed by centering $\boldsymbol{\phi}_t^*$ so that the elements of the resulting vector of random effects sum to zero. Hence, when the sum-zero constraint is taken into account, each element of the vector of random effects actually depends on all the other elements. That is, when the sum-zero constraint is considered the actual full conditional distribution for a spatial random effect should depend on all the other spatial random effects. Hence, the usual implementation of the one-at-a-time Gibbs sampler for the ICAR model with centering on the fly does not actually simulate from the full conditional distributions of the random effects. Therefore, the conditions for the design of Gibbs samplers are not satisfied and, as a result, the limiting distribution of the algorithm is unclear.

Keefe, Ferreira and Franck (2018) have proposed a formal specification of a sum-zero constrained ICAR model. Let $\tau\mathbf{H}$ be the singular precision matrix of the ICAR model, $\tau > 0$ be a precision parameter, and $\mathbf{K}$ be a symmetric positive semi-definite matrix for which the sum of its elements is positive. To obtain their specification, Keefe, Ferreira and Franck (2018) perform three steps. First, they start with a proper CAR model that has a positive definite precision matrix equal to $\tau(\lambda\mathbf{K} + \mathbf{H})$, where $\lambda > 0$ is a scalar. Second, they project the vector of proper conditional autoregressive spatial random effects onto a subspace where the projected vector is constrained to sum to zero. Finally, they take the limit when λ approaches zero and the proper conditional autoregressive model approaches the ICAR model. Keefe, Ferreira and Franck (2018) show that, for any matrix $\mathbf{K}$ as describe above, the resulting distribution does not depend on the matrix $\mathbf{K}$. Specifically, the distribution they obtain for the sum-zero constrained ICAR model is

the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$. Here we show that this distribution proposed by Keefe, Ferreira and Franck (2018) is the limiting distribution of the one-at-a-time Gibbs sampler for the intrinsic conditional autoregressive model with centering on the fly.

The fact that the limiting distribution is the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$ has important implications for the development of formal Bayesian procedures for spatial hierarchical models with ICAR random effects. The development of formal Bayesian procedures for these models has been prevented because the ICAR "density" is improper and not fully specified. In contrast, the density of the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$ is proper, it explicitly includes the sum-zero constraint, and it has a known constant of proportionality. This knowledge of the entire density is of fundamental importance for the development of formal objective Bayesian methodology.

Formal objective Bayesian methods such as Bayesian model selection and the development of reference Bayesian analysis require the integrated likelihood function that is obtained by integrating out the spatial random effects. Further, we note that the previous lack of clarity about the ICAR density prevented the computation of such integrated likelihood function. In contrast, in Gaussian hierarchical models with ICAR random effects the result that the limiting distribution is the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$ leads to an explicit expression for the integrated likelihood function. Such expression then can be used for the development of formal objective Bayesian methods. For example, for a Gaussian hierarchical model with ICAR random effects, Keefe, Ferreira and Franck (2018) have developed formal Bayesian model selection, and Keefe, Ferreira and Franck (2019) have developed formal objective Bayesian estimation.

The remainder of the article is organized as follows. Section 2 presents the intrinsic conditional autoregressive model. Section 3 presents the one-at-a-time Gibbs sampler for the ICAR model with centering on the fly and shows that the sequence of simulated vector of random effects follows a time series vector autoregressive model of order 1. Section 4 presents the theorem that establishes that the limiting distribution is a singular Gaussian distribution with a covariance matrix that is the Moore–Penrose inverse of the precision matrix. Section 4 also presents a proof for the theorem. Section 5 concludes with a brief discussion and possible avenues for future research. Finally, the Appendix presents proofs of auxiliary lemmas.

2 Intrinsic conditional autoregressive model

Consider a geographical region of interest that is partitioned into n disjoint subregions. In addition, assume a neighborhood structure such that N_j, $j = 1, \ldots, n$, represents the set of subregions that are neighbors of subregion j. Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$ be a vector of random effects corresponding to a process of interest

over the n subregions. Assume that $\boldsymbol{\phi}$ follows an intrinsic conditional autoregressive model (Besag and Kooperberg, 1995) with joint density

$$p(\boldsymbol{\phi}) \propto \exp\left\{-\frac{\tau}{2}\boldsymbol{\phi}'\mathbf{H}\boldsymbol{\phi}\right\}, \tag{1}$$

where $\tau > 0$ is a precision parameter and $\tau\mathbf{H}$ is the precision matrix. The matrix $\mathbf{H}$ is usually assumed to be known and depends on the neighborhood structure such that

$$(H)_{ij} = \begin{cases} h_i & \text{if } i = j, \\ -g_{ij} & \text{if } i \in N_j, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $g_{ij} \geq 0$ is a measure of similarity of subregions i and j, $g_{ij} = g_{ji}$, and $h_i = \sum_{j \neq i} g_{ij}$. For example, a common choice is $g_{ij} = 1$ if subregions i and j are neighbors, and $g_{ij} = 0$ otherwise. Note that $g_{ij} = g_{ji}$ implies $\mathbf{H}$ is symmetric.

Further, $h_i = \sum_{j \neq i} g_{ij}$ implies $\mathbf{H}$ is positive semidefinite. We assume that any two subregions are connected by a path. As a consequence, $\mathbf{H}$ has only one eigenvalue equal to 0 with corresponding normalized eigenvector $n^{-1/2}\mathbf{1}$, where $\mathbf{1}$ is the n-dimensional vector of ones. Because $\mathbf{H}$ is not positive definite, the covariance matrix of $\boldsymbol{\phi}$ and the constant of proportionality in Equation (1) are not well defined and the implied distribution for $\boldsymbol{\phi}$ is improper. Thus, the use of the ICAR definition given by Equations (1) and (2) for spatial random effects may lead to an improper posterior distribution.

To guarantee posterior propriety, implementation of Gibbs samplers for models that have ICAR random effects usually use *centering on the fly*. That is, at the end of each iteration of the Gibbs sampler, the vector of random effects is re-centered to impose a sum-zero constraint so that $\boldsymbol{\phi}'\mathbf{1} = 0$. While this works well in practice, centering on the fly is a computational hack that does not make it explicit the mathematics behind the procedure. This lack of mathematical clarity has prevented previous literature from obtaining the prior distribution for the random effects $\boldsymbol{\phi}$ implied by centering on the fly.

3 One-at-a-time Gibbs sampler centering on the fly

Consider the one-at-a-time Gibbs sampler with centering on the fly (OGC) to simulate the vector of random effects $\boldsymbol{\phi}$ from the ICAR model given in Equation (1). Let $\boldsymbol{\phi}_t^* = (\phi_{t1}^*, \ldots, \phi_{tn}^*)'$ be an auxiliary vector simulated at the tth iteration of the algorithm. And, after centering on the fly, let $\boldsymbol{\phi}_t = (\phi_{t1}, \ldots, \phi_{tn})'$ be the vector of random effects simulated at iteration t of the algorithm. Without loss of generality, assume that the elements of $\boldsymbol{\phi}_t^*$ are simulated sequentially from ϕ_{t1}^* to ϕ_{tn}^*.

Each iteration of the one-at-a-time Gibbs sampler with centering on the fly proceeds in two steps. In the first step, the elements of the auxiliary vector $\boldsymbol{\phi}_t^*$ are simulated one at a time from the full conditional distributions implied by the ICAR

model given in Equation (1) but ignoring the sum-zero constraint. After that, a centering on the fly step is applied to obtain $\boldsymbol{\phi}_t$, meaning that the elements of $\boldsymbol{\phi}_t^*$ are centered so that $\boldsymbol{\phi}'\mathbf{1} = 0$. Let $\mathbf{I}$ be the n by n identity matrix. Then, a mathematical way to represent the centering on the fly step is by writing $\boldsymbol{\phi}_t = \mathbf{P}\boldsymbol{\phi}_t^*$, where $\mathbf{P}$ is the projection matrix $\mathbf{P} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$. Specifically, the one-at-a-time Gibbs sampler with centering on the fly is as follows.

Algorithm 3.1 (One-at-a-time Gibbs sampler with centering on the fly).

1. *Initialize* $\boldsymbol{\phi}_0$ *and set iteration counter to* $t = 1$.
2. *For* $i = 1, \ldots, n$, *simulate* ϕ_{ti}^* *from the full conditional distribution*

$$N\left(\frac{\sum_{j=1}^{i-1} g_{ij}\phi_{tj}^* + \sum_{j=i+1}^{n} g_{ij}\phi_{t-1,j}}{h_i}, \frac{1}{h_i\tau}\right).$$

3. *Center on the fly*: *Compute* $\boldsymbol{\phi}_t = \mathbf{P}\boldsymbol{\phi}_t^*$.
4. *Set* $t = t + 1$ *and return to Step* (2) *until desired number of iterations is reached.*

Let $\mathbf{D} = \text{diag}(h_1, \ldots, h_n)$. In addition, consider the lower triangular matrix

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 0 \\ g_{21} & 0 & \cdots & 0 & 0 & 0 \\ g_{31} & g_{32} & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ g_{n-1,1} & g_{n-1,2} & \cdots & g_{n-1,n-2} & 0 & 0 \\ g_{n1} & g_{n2} & \cdots & g_{n,n-2} & g_{n,n-2} & 0 \end{pmatrix}.$$

Thus, the simulation of the auxiliary vector $\boldsymbol{\phi}_t^*$ in the OGC algorithm can be represented with the equation

$$\boldsymbol{\phi}_t^* = \mathbf{E}\boldsymbol{\phi}_t^* + \mathbf{F}\boldsymbol{\phi}_{t-1} + \tau^{-1/2}\mathbf{D}^{-1/2}\boldsymbol{\epsilon}_t,$$

where $\mathbf{E} = \mathbf{D}^{-1}\mathbf{L}$, $\mathbf{F} = \mathbf{D}^{-1}\mathbf{L}'$, and $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I})$.

Note that the matrix $\mathbf{I} - \mathbf{E}$ is diagonally dominant and, thus, has an inverse. Therefore, we can write the simulation of $\boldsymbol{\phi}_t$ as

$$\boldsymbol{\phi}_t = \mathbf{A}\boldsymbol{\phi}_{t-1} + \tau^{-1/2}\mathbf{B}\boldsymbol{\epsilon}_t, \tag{3}$$

where $\mathbf{A} = \mathbf{P}(\mathbf{I} - \mathbf{E})^{-1}\mathbf{F}$ and $\mathbf{B} = \mathbf{P}(\mathbf{I} - \mathbf{E})^{-1}\mathbf{D}^{-1/2}$.

We note from Equation (3) that $\boldsymbol{\phi}_t$ follows a time series vector autoregressive model of order 1 (VAR(1)). This relationship between the VAR(1) model and the OGC algorithm provides several properties of the OGC algorithm. First, the number of lags in the VAR model is just one, which makes intuitive sense because of the Markovian property that is implicit in the very definition of the OGC algorithm. Second, the vector of random effects simulated in one iteration of the OGC

algorithm can be written as a linear function of the random effects from the previous iteration plus a random vector of innovations. Finally, and most important for our purposes in this article, the limiting distribution of the OGC algorithm can be obtained through an analysis of the VAR(1) model given in Equation (3) (Reinsel, 1997, Prado and West, 2010).

4 Limiting distribution

As described in the Introduction section, Keefe, Ferreira and Franck (2018) have proposed a formal specification of a sum-zero constrained ICAR model. The distribution they obtain for this model is the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$. This distribution has density

$$p(\boldsymbol{\phi}) = (2\pi)^{(n-1)/2}\tau^{(n-1)/2}\left(\prod_{i=1}^{n-1} s_i\right)^{1/2} \exp\left\{-\frac{\tau}{2}\boldsymbol{\phi}'\mathbf{H}\boldsymbol{\phi}\right\}\mathbb{1}(\mathbf{1}_n'\boldsymbol{\phi} = 0), \qquad (4)$$

where $s_1 \geq \cdots \geq s_{n-1} > s_n = 0$ are the ordered eigenvalues of $\mathbf{H}$ and $\mathbb{1}(\cdot)$ is the indicator function.

Note that the density given in Equation (4) above is proportional to the density given in Equation (1). However, in contrast to Equation (1), the density in Equation (4) is proper, explicitly includes the sum-zero constraint, and has a know constant of proportionality. In particular, knowledge of the entire density is of fundamental importance for the development of formal objective Bayesian methodology (Keefe, Ferreira and Franck 2018, 2019). Specifically, formal objective Bayesian methods require the integrated likelihood function that is obtained by integrating out the spatial random effects. Further, we note that the lack of clarity in previous literature about the ICAR density prevented the computation of such integrated likelihood function.

In contrast, in Gaussian hierarchical models the knowledge that the limiting distribution of the ICAR random effects is the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$ leads to an explicit expression for the integrated likelihood function. Such expression can then be used for the development of formal objective Bayesian methods for Gaussian hierarchical models with ICAR random effects, such as for example formal Bayesian model selection (Keefe, Ferreira and Franck, 2018) and formal objective Bayesian estimation (Keefe, Ferreira and Franck, 2019).

The following theorem establishes that the limiting distribution of the OGC algorithm is indeed the distribution proposed by Keefe, Ferreira and Franck (2018).

Theorem 4.1. *The limiting distribution of the one-at-a-time Gibbs sampler with centering on the fly for the spatial conditional autoregressive model* (*Algorithm* 3.1) *is the singular Gaussian distribution* $N(\mathbf{0}, \tau^{-1}\mathbf{H}^{+})$.

To prove Theorem 4.1, let us consider the spectral decomposition of the precision matrix $\mathbf{H}$ given by $\mathbf{H} = \mathbf{QSQ}'$, where $\mathbf{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$ is a matrix comprised of columns which are the normalized eigenvectors of $\mathbf{H}$ and $\mathbf{S} = \text{diag}(s_1, \ldots, s_n)$ are the ordered eigenvalues of $\mathbf{H}$. Further, to prove Theorem 4.1 we need two auxiliary facts and seven lemmas stated below.

The following two auxiliary facts have appeared before in Ferreira and De Oliveira (2007) and De Oliveira and Ferreira (2011).

Auxiliary Fact 4.1 (Ferreira and De Oliveira, 2007, De Oliveira and Ferreira, 2011). *The ordered eigenvalues of* $\mathbf{H}$ *are such that* $s_1 \geq \cdots \geq s_{n-1} > s_n = 0$,

Auxiliary Fact 4.2 (Ferreira and De Oliveira, 2007, De Oliveira and Ferreira, 2011). *The eigenvector corresponding to* $s_n = 0$ *is* $\boldsymbol{q}_n = n^{-1/2}\mathbf{1}_n$.

The seven lemmas below provide auxiliary results about matrix operations involving the precision matrix $\mathbf{H}$, its generalized inverse, the projection matrix $\mathbf{P}$, and the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{L}$ that appear in the definition of the VAR(1) model given in Equation (3). We provide proofs of the seven lemmas in the Appendix.

Lemma 4.1. $\mathbf{H}\mathbf{H}^{+} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$.

Lemma 4.2. $\mathbf{L}'\mathbf{H}^{+} = \mathbf{D}\mathbf{H}^{+} - \mathbf{L}\mathbf{H}^{+} - \mathbf{I} + n^{-1}\mathbf{1}\mathbf{1}'$.

Lemma 4.3. $(\mathbf{D} - \mathbf{L})^{-1}\mathbf{L}'\mathbf{1} = \mathbf{1}$.

Lemma 4.4. $\mathbf{P}\mathbf{1} = \mathbf{0}$.

Lemma 4.5. $\mathbf{P}(\mathbf{D} - \mathbf{L})^{-1}\mathbf{L}'\mathbf{1} = \mathbf{0}$.

Lemma 4.6. $\mathbf{P}\mathbf{H}^{+}\mathbf{P} = \mathbf{H}^{+}$.

Lemma 4.7. $\mathbf{A} = \mathbf{P}(\mathbf{D} - \mathbf{L})^{-1}\mathbf{L}'$ *and* $\mathbf{B} = \mathbf{P}(\mathbf{D} - \mathbf{L})^{-1}\mathbf{D}^{1/2}$.

Lemmas 4.2, 4.4, 4.5, 4.6, and 4.7 are directly used in the proof of Theorem 4.1. Auxiliary Facts 4.1 and 4.2 as well as Lemmas 4.1 and 4.3 are used in the proofs of the other lemmas.

We now provide the proof of Theorem 4.1.

Proof of Theorem 4.1. For the VAR(1) model given in Equation (3), because the eigenvalues of the matrix $\mathbf{A}$ are less than one in absolute value and the errors are Gaussian with mean vector $\mathbf{0}$, there is a unique limiting distribution that is multivariate Gaussian with mean vector $\mathbf{0}$. Further, the covariance matrix $\mathbf{V}$ of this limiting distribution is the unique solution of the equation $\mathbf{V} = \mathbf{AVA} + \tau^{-1}\mathbf{BB}'$

(Reinsel, 1997). We now check that $\mathbf{V} = \mathbf{H}^+$ is the solution. That is, we show that $\mathbf{M} = \mathbf{A}\mathbf{H}^+\mathbf{A} + \tau^{-1}\mathbf{B}\mathbf{B}' = \mathbf{H}^+$.

First, we apply Lemmas 4.2 and 4.7 to obtain:

$$\begin{aligned}
\mathbf{M} &= \mathbf{A}\mathbf{H}^+\mathbf{A} + \tau^{-1}\mathbf{B}\mathbf{B}' \\
&= \mathbf{P}(\mathbf{D}-\mathbf{L})^{-1}\mathbf{L}'\mathbf{H}^+\mathbf{L}(\mathbf{D}-\mathbf{L}')^{-1}\mathbf{P} + \mathbf{P}(\mathbf{D}-\mathbf{L})^{-1}\mathbf{D}(\mathbf{D}-\mathbf{L}')^{-1}\mathbf{P} \\
&= \mathbf{P}(\mathbf{D}-\mathbf{L})^{-1}\big[\{(\mathbf{D}-\mathbf{L})\mathbf{H}^+ - \mathbf{I} + n^{-1}\mathbf{1}\mathbf{1}'\}\mathbf{L} + \mathbf{D}\big](\mathbf{D}-\mathbf{L}')^{-1}\mathbf{P} \\
&= \mathbf{P}\big[\mathbf{H}^+\mathbf{L}(\mathbf{D}-\mathbf{L}')^{-1} + (\mathbf{D}-\mathbf{L})^{-1}(-\mathbf{I} + n^{-1}\mathbf{1}\mathbf{1}')\mathbf{L}(\mathbf{D}-\mathbf{L}')^{-1} \\
&\quad + (\mathbf{D}-\mathbf{L})^{-1}\mathbf{D}(\mathbf{D}-\mathbf{L}')^{-1}\big]\mathbf{P} \\
&= \mathbf{P}\big[\mathbf{H}^+\mathbf{L}(\mathbf{D}-\mathbf{L}')^{-1} + (\mathbf{D}-\mathbf{L})^{-1}(\mathbf{D}-\mathbf{L})(\mathbf{D}-\mathbf{L}')^{-1} \\
&\quad + n^{-1}(\mathbf{D}-\mathbf{L})^{-1}\mathbf{1}\mathbf{1}'\mathbf{L}(\mathbf{D}-\mathbf{L}')^{-1}\big]\mathbf{P}.
\end{aligned}$$

Next, we apply Lemmas 4.2, 4.4, 4.5, and 4.6 to obtain the main result:

$$\begin{aligned}
\mathbf{M} &= \mathbf{P}\big[\mathbf{H}^+\mathbf{L}(\mathbf{D}-\mathbf{L}')^{-1} + (\mathbf{D}-\mathbf{L}')^{-1}\big]\mathbf{P} \\
&= \mathbf{P}\big[\{\mathbf{H}^+(\mathbf{D}-\mathbf{L}') - \mathbf{I} + n^{-1}\mathbf{1}\mathbf{1}'\}(\mathbf{D}-\mathbf{L}')^{-1} + (\mathbf{D}-\mathbf{L}')^{-1}\big]\mathbf{P} \\
&= \mathbf{P}\big[\{\mathbf{H}^+ + n^{-1}\mathbf{1}\mathbf{1}'(\mathbf{D}-\mathbf{L}')^{-1}\big]\mathbf{P} \\
&= \mathbf{H}^+
\end{aligned}$$

□

5 Discussion

The result that the limiting distribution of the one-at-a-time Gibbs sampler with centering on the fly for ICAR models is the singular Gaussian distribution $N(\mathbf{0}, \tau^{-1}\mathbf{H}^+)$ has many important consequences. In particular, the fact that the $N(\mathbf{0}, \tau^{-1}\mathbf{H}^+)$ distribution has a known and fully specified density opens up the possibility of development of formal Bayesian statistical methods that use the sum-zero constrained ICAR model as prior for random effects. For example, for a Gaussian hierarchical model with ICAR random effects, Keefe, Ferreira and Franck (2018) have developed formal Bayesian model selection, and Keefe, Ferreira and Franck (2019) have developed formal objective Bayesian estimation.

Current research includes the development of analyses for observations in the exponential family of distributions. In that regard, we are currently working on the development of formal objective Bayesian model selection and estimation for spatial generalized linear mixed effects models with sum-zero constrained ICAR priors for the spatial random effects. The explicit expression and the propriety of the prior density of the sum-zero constrained ICAR random effects guarantee that we can integrate out the ICAR spatial random effects to obtain the corresponding

marginal likelihood. However, with the exception of the case of Gaussian observations, the marginal likelihood that results from such integration does not have an explicit analytic expression. Thus, for the case of non-Gaussian observations the solutions will be approximate. We are currently exploring different ways to approximate the marginal likelihood such as for example, by using Laplace approximations which have been successfully used recently for model selection for non-spatial GLMs (Wu, Ferreira and Gompper, 2016, Wu et al., 2019).

Appendix: Auxiliary facts and lemmas

Proof of Lemma 4.1. From Auxiliary Fact 4.1, the Moore–Penrose inverse of $\mathbf{H}$ is $\mathbf{H}^+ = \mathbf{Q}\mathbf{S}^+\mathbf{Q}$, where $\mathbf{S}^+ = \text{diag}(s_1^{-1}, \ldots, s_{n-1}^{-1}, 0)$. Hence, $\mathbf{H}\mathbf{H}^+ = \mathbf{Q}\mathbf{S}\mathbf{Q}'\mathbf{Q}\mathbf{S}^+\mathbf{Q} = \mathbf{Q}\mathbf{S}\mathbf{S}^+\mathbf{Q} = \mathbf{Q}\text{diag}(1, \ldots, 1, 0)\mathbf{Q} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$, where the last equality follows from Auxiliary Fact 4.2. □

Proof of Lemma 4.2. Note that $\mathbf{H} = \mathbf{D} - \mathbf{L} - \mathbf{L}'$. Hence, by Lemma 4.1, $\mathbf{I} + n^{-1}\mathbf{1}\mathbf{1}' = \mathbf{H}\mathbf{H}^+ = (\mathbf{D} - \mathbf{L} - \mathbf{L}')\mathbf{H}^+$. Therefore, $\mathbf{L}'\mathbf{H}^+ = \mathbf{D}\mathbf{H}^+ - \mathbf{L}\mathbf{H}^+ - \mathbf{I} + n^{-1}\mathbf{1}\mathbf{1}'$. □

Proof of Lemma 4.3. Note that $(\mathbf{D} - \mathbf{L})$ is a diagonally dominant matrix and therefore it is invertible. Hence, $\mathbf{x} = (\mathbf{D} - \mathbf{L})^{-1}\mathbf{L}'\mathbf{1}$ is the unique solution to $(\mathbf{D} - \mathbf{L})\mathbf{x} = \mathbf{L}'\mathbf{1}$. Further, note that $\mathbf{D} = \text{diag}\{(\mathbf{L} + \mathbf{L}')\mathbf{1}\}$ and $\mathbf{D}\mathbf{1} = (\mathbf{L} + \mathbf{L}')\mathbf{1}$. Therefore, $\mathbf{x} = \mathbf{1}$. □

Proof of Lemma 4.4. $\mathbf{P}\mathbf{1} = (\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}')\mathbf{1} = \mathbf{1} - n^{-1}\mathbf{1}\mathbf{1}'\mathbf{1} = \mathbf{0}$. □

Proof of Lemma 4.5. : This follows directly from Lemmas 4.3 and 4.4. □

Proof of Lemma 4.6. Using the spectral decomposition of $\mathbf{H}$ and Auxiliary Fact 4.1, we can write $\mathbf{H}^+ = \sum_{i=1}^{n-1} s_i^{-1}\mathbf{q}_i\mathbf{q}_i'$. In addition, because of Auxiliary Fact 4.2, $\mathbf{q}_1, \ldots, \mathbf{q}_{n-1}$ are orthogonal to $\mathbf{1}$. Thus, for $i = 1, \ldots, n-1$, $\mathbf{P}\mathbf{q}_i = (\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}')\mathbf{q}_i = \mathbf{q}_i$. Therefore, $\mathbf{P}\mathbf{H}^+\mathbf{P} = \mathbf{P}\sum_{i=1}^{n-1} s_i^{-1}\mathbf{q}_i\mathbf{q}_i'\mathbf{P} = \sum_{i=1}^{n-1} s_i^{-1}\mathbf{P}\mathbf{q}_i\mathbf{q}_i'\mathbf{P} = \sum_{i=1}^{n-1} s_i^{-1}\mathbf{q}_i\mathbf{q}_i' = \mathbf{H}^+$. □

Proof of Lemma 4.7. Using the definitions of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{F}$ we have $\mathbf{A} = \mathbf{P}(\mathbf{I} - \mathbf{E})^{-1}\mathbf{F} = \mathbf{P}(\mathbf{I} - \mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{D}^{-1}\mathbf{L}' = \mathbf{P}\{\mathbf{D}^{-1}(\mathbf{D} - \mathbf{L})\}^{-1}\mathbf{D}^{-1}\mathbf{L}' = \mathbf{P}(\mathbf{D} - \mathbf{L})^{-1}\mathbf{L}'$ and $\mathbf{B} = \mathbf{P}(\mathbf{I} - \mathbf{E})^{-1}\mathbf{D}^{-1/2} = \mathbf{P}\{\mathbf{D}^{-1}(\mathbf{D} - \mathbf{L})\}^{-1}\mathbf{D}^{-1/2} = \mathbf{P}(\mathbf{D} - \mathbf{L})^{-1}\mathbf{D}^{1/2}$. □

Acknowledgments

We gratefully acknowledge the constructive comments and suggestions made by an anonymous referee and the Associate Editor that led to a substantially improved article.

References

Bell, B. S. and Broemeling, L. D. (2000). A Bayesian analysis for spatial processes with application to disease mapping. *Statistics in Medicine* **19**, 957–974.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746. MR1380811

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–20. MR1105822

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–681.

De Oliveira, V. and Ferreira M. A. R. (2011). Maximum likelihood and restricted maximum likelihood estimation for a class of Gaussian Markov random fields. *Metrika* **74**, 167–183. MR2822154

Ferreira, M. A. R., Bi, Z., West, M., Lee, H. and Higdon, D. (2003). Multi-modelling of 1-D permeability fields. In *Bayesian Statistics 7* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 519–527. Oxford: Oxford University Press. MR2003196

Ferreira M. A. R. and De Oliveira, V. (2007). Bayesian reference analysis for Gaussian Markov random fields. *Journal of Multivariate Analysis* **98**, 789–812. MR2322129

Ferreira, M. A. R. and Lee, H. K. H. (2007). *Multiscale Modeling: A Bayesian Perspective*. *Springer Series in Statistics*. New York: Springer. MR2348732

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Boca Raton: CRC Press. MR2260716

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409. MR1141740

Goicoa, T., Ugarte, M. D., Etxeberria, J. and Militino, A. F. (2016). Age–space–time CAR models in Bayesian disease mapping. *Statistics in Medicine* **35**, 2391–2405. MR3513520

Keefe, M. J., Ferreira, M. A. R. and Franck, C. T. (2018). On the formal specification of sum-zero constrained intrinsic conditional autoregressive models. *Spatial Statistics* **24**, 54–65. MR3795058

Keefe, M. J., Ferreira, M. A. R. and Franck, C. T. (2019). Objective Bayesian analysis for Gaussian hierarchical models with intrinsic conditional autoregressive priors. *Bayesian Analysis* **14**, 181–209. MR3910043

Lee, H. K. H., Higdon, D., Bi, Z., Ferreira, M. A. R. and West, M. (2002). Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media. *Technometrics* **44**, 230–241. MR1951639

Liu, Z., Berrocal, V. J., Bartsch, A. J. and Johnson, T. D. (2016). Pre-surgical fMRI data analysis using a spatially adaptive conditionally autoregressive model. *Bayesian Analysis* **11**, 599–625. MR3472004

Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*. Boca Raton: CRC Press. MR2655202

Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*, 2nd ed. Berlin: Springer. MR1451875

Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer. MR1707311

Wu, H. H., Ferreira, M. A. R., Elkhouly, M. and Ji, T. (2019). Nonlocal priors for variable selection in generalized linear models. *Sankhya Series A*. DOI:10.1007/s13171-018-0151-9. To appear.

Wu, H. H., Ferreira, M. A. R. and Gompper, M. E. (2016). Consistency of hyper-g-prior-based Bayesian variable selection for generalized linear models. *Brazilian Journal of Probability and Statistics* **30**, 691–709. MR3582395

Department of Statistics
Virginia Tech
Hutcheson Hall, RM 406-A
250 Drillfield Drive
Blacksburg, Virginia 24061
USA
E-mail: marf@vt.edu