

Gibbs' Sampler: Dynamic CAR Model

Elliot Shannon

shann125@msu.edu

Michigan State University

Setting

We consider data observed over a spatial domain which is partitioned into a fixed number of areal units. The data are observed for each areal unit across discrete time steps. We aim to model these data using spatially and temporally explicit covariates, as well as a spatially-varying intercept term which is allowed to evolve dynamically over time. We will employ conditional autoregressive (CAR) spatial structures for these random effects. Examples of data such as described here include annual disease prevalence for US counties or quarterly average home values for census tracts in Michigan.

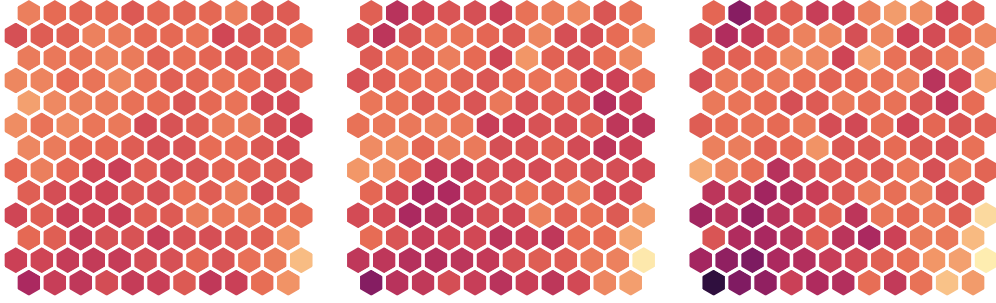


Figure 1: Example of a CAR spatial effect evolving over time (from left to right).

Model

For areal units $j = 1, \dots, J$ and time points $t = 1, \dots, T$, let

- $y_{j,t} \in \mathbb{R}$ be the response value.
- $\mathbf{x}_{j,t} = (1, x_{1,j,t}, \dots, x_{P-1,j,t}) \in \mathbb{R}^P$ be the vector of covariates (predictor variables).
- $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_P \end{pmatrix} \in \mathbb{R}^P$ be the vector of regression coefficients.
- $u_{j,t} \in \mathbb{R}$ be the spatio-temporally varying random intercept.
- $\varepsilon_{j,t} \in \mathbb{R}$ be the residual error term.

For areal unit j at time t , the proposed model/priors is then

$$\begin{aligned} y_{j,t} &= u_{j,t} + \mathbf{x}_{j,t}\boldsymbol{\beta} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \sim N(0, \sigma^2), \\ u_{j,t} &= u_{j,t-1} + w_{j,t}, \quad u_{\cdot,0} \equiv 0 \\ \sigma^2 &\sim IG(a_\sigma, b_\sigma), \quad \boldsymbol{\beta} \sim MVN(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \end{aligned} \tag{1}$$

where $\mathbf{w}_t = (w_{1,t}, \dots, w_{J,t})^\top$ follows from a conditional autoregressive (CAR) spatial structure as

$$\begin{aligned}\mathbf{w}_t &\sim MVN(0, \tau_t^2 \mathbf{Q}(\rho_t)), \quad \mathbf{Q}(\rho_t) = (\mathbf{D} - \rho_t \mathbf{W})^{-1} \\ \tau_t^2 &\sim IG(a_t, b_t), \quad \rho_t \sim U(c_t, d_t).\end{aligned}\tag{2}$$

Here, τ_t^2 is the spatial variance parameter, $\mathbf{Q}(\rho_t)$ is the CAR correlation matrix depending on spatial dependence parameter ρ_t , \mathbf{W} is a $J \times J$ binary spatial adjacency matrix, and \mathbf{D} is a $J \times J$ diagonal matrix whose elements correspond to the row sums of \mathbf{W} . We assume \mathbf{W} and \mathbf{D} to be fixed over time, but allow for unique variance and spatial dependence parameters for each time step.

Stacking

Stacking these terms for each time $t = 1, \dots, T$, we can also write them in the form

$$\begin{aligned}\bullet \mathbf{y}_t &= \begin{pmatrix} y_{1,t} \\ \vdots \\ y_{J,t} \end{pmatrix} \in \mathbb{R}^J \\ \bullet \mathbf{X}_t &= \begin{pmatrix} \mathbf{x}_{1,t} \\ \vdots \\ \mathbf{x}_{J,t} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1,t} & \cdots & x_{P-1,1,t} \\ & & \vdots & \\ 1 & x_{1,J,t} & \cdots & x_{P-1,J,t} \end{pmatrix} \in \mathbb{R}^{J \times P} \\ \bullet \mathbf{u}_t &= \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{J,t} \end{pmatrix} \in \mathbb{R}^J \\ \bullet \boldsymbol{\varepsilon}_t &= \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{J,t} \end{pmatrix} \in \mathbb{R}^J\end{aligned}$$

and the model for time t is then written as

$$\begin{aligned}\mathbf{y}_t &= \mathbf{u}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_J), \\ \mathbf{u}_t &= \mathbf{u}_{t-1} + \mathbf{w}_t, \quad \mathbf{u}_0 \equiv \mathbf{0} \\ \sigma^2 &\sim IG(a_\sigma, b_\sigma), \quad \boldsymbol{\beta} \sim MVN(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\ \mathbf{w}_t &\sim MVN(0, \tau_t^2 \mathbf{Q}(\rho_t)), \quad \mathbf{Q}(\rho_t) = (\mathbf{D} - \rho_t \mathbf{W})^{-1} \\ \tau_t^2 &\sim IG(a_t, b_t), \quad \rho_t \sim U(c_t, d_t).\end{aligned}\tag{3}$$

where \mathbf{I}_J is a $J \times J$ identity matrix with diagonal elements equal to 1, and $\mathbf{0}$ is a vector of 0's.

Here, $N(\mu, \sigma^2)$ is the normal distribution with mean $\mu \in \mathbb{R}$, variance $\sigma^2 > 0$, and pdf

$$p(x \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)\tag{4}$$

$MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^N$, covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$, and pdf

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-N/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),\tag{5}$$

$IG(a, b)$ is the Inverse Gamma distribution with shape parameter $a > 0$, scale parameter $b > 0$, and pdf

$$p(\sigma^2 | a, b) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp\left(\frac{-b}{\sigma^2}\right), \quad (6)$$

and $U(c, d)$ is the continuous Uniform distribution with support $[c, d]$ and pdf

$$p(\rho | c, d) = \begin{cases} \frac{1}{d-c} & \text{if } c \leq \rho \leq d \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Given this completed model specification, we would like to conduct inference on the model parameters given data \mathbf{y}_t and \mathbf{X}_t , $t = 1, \dots, T$. These parameters include $\beta, \mathbf{u}_1, \dots, \mathbf{u}_T, \tau_1^2, \dots, \tau_T^2, \rho_1, \dots, \rho_T, \sigma^2$. We assume that prior parameters are given (specified by the user).

Posterior

We know that the posterior distribution in Bayesian inference is proportional to the likelihood (data distribution) times the prior. Here, our likelihood is the distribution of the data \mathbf{y}_t , $t = 1, \dots, T$. Since the distribution of the residuals is MVN , we will have a MVN likelihood, written as

$$\prod_{t=1}^T MVN(\mathbf{y}_t | \mathbf{u}_t + \mathbf{X}_t \beta, \sigma^2 \mathbf{I}_J). \quad (8)$$

We will also have priors (some of which are induced) for each parameter for which we seek inference. Specifically, we have prior distributions of the form

$$\begin{aligned} & \prod_{t=1}^T MVN(\mathbf{u}_t | \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times MVN(\beta | \mu_\beta, \Sigma_\beta) \times IG(\sigma^2 | a, b) \\ & \times \prod_{t=1}^T IG(\tau_t^2 | a_t, b_t) \times \prod_{t=1}^T U(\rho_t | c_t, d_t) \end{aligned} \quad (9)$$

Given (8) and (9), their product will be proportional to the posterior distribution, and is given as

$$\begin{aligned} & \prod_{t=1}^T MVN(\mathbf{y}_t | \mathbf{u}_t + \mathbf{X}_t \beta, \sigma^2 \mathbf{I}_J) \times \prod_{t=1}^T MVN(\mathbf{u}_t | \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times MVN(\beta | \mu_\beta, \Sigma_\beta) \\ & \times IG(\sigma^2 | a, b) \times \prod_{t=1}^T IG(\tau_t^2 | a_t, b_t) \times \prod_{t=1}^T U(\rho_t | c_t, d_t) \end{aligned} \quad (10)$$

Moving forward, we only work with this proportional posterior (10).

Working it out

We will now go through each of the parameters in model (1) and derive their full conditional distributions. Moving forward, we will use different colors to help keep track of different prior and likelihood terms as we work through the derivations. This will help us see where each component of the full conditional distribution is coming from.

Update β

Looking at (10), we see that β appears in $T + 1$ terms, namely in the likelihood of $\mathbf{y}_t, t = 1, \dots, T$ and in its own prior distribution. The product of these terms looks like

$$\prod_{t=1}^T MVN(\mathbf{y}_t \mid \mathbf{u}_t + \mathbf{X}_t \beta, \sigma^2 \mathbf{I}_J) \times MVN(\beta \mid \mu_\beta, \Sigma_\beta).$$

This is the product of multivariate normal distributions, so we know through conjugacy that the full conditional distribution will also be multivariate normal. Therefore, we simply need to identify the mean and covariance matrix of this full conditional distribution, which can be obtained by *incompleting the square*.

Incompleting the square

See section 4.2 in James Clark's book 'Models for Ecological Data' for more details.

Following (5), we know that the kernel of the MVN distribution is of the form

$$p(\mathbf{x} \mid \mu, \Sigma) \propto \exp \left(-\frac{1}{2} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right)$$

where \propto mean "proportional to". This structure gives us some insight into the placement of the parameters in the pdf. To see this further, we can focus on the terms inside the exponential, ignoring the $-\frac{1}{2}$ coefficient. Expanding this term gives us:

$$(\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) = \mathbf{x}^\top \Sigma^{-1} \mathbf{x} - \mathbf{x}^\top \Sigma^{-1} \mu - \mu^\top \Sigma^{-1} \mathbf{x} + \mu^\top \Sigma^{-1} \mu \quad (11)$$

Now comes the trick. Writing the mean as a function of the covariance matrix, we can define

$$\begin{aligned} \mu &= \mathbf{V} \mathbf{v} \\ \Sigma &= \mathbf{V} \end{aligned}$$

This allows us to rewrite (11) as

$$\begin{aligned} \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x} - \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{V} \mathbf{v} - \mathbf{v}^\top \mathbf{V}^\top \mathbf{V}^{-1} \mathbf{x} + \mathbf{v}^\top \mathbf{V}^\top \mathbf{V}^{-1} \mathbf{V} \mathbf{v} \\ = \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x} - \mathbf{x}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{x} + \mathbf{v}^\top \mathbf{V} \mathbf{v} \\ = \mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x} - \mathbf{x}^\top \mathbf{v} \dots \end{aligned}$$

This is super helpful, because when we are mashing together normal priors and likelihoods, we can combine like terms and look for \mathbf{V}^{-1} and \mathbf{v} to get the mean and variance of the full conditional MVN . They will always be in the same place! \mathbf{V}^{-1} will be sandwiched between the variable of interest (i.e. the $\mathbf{x}^\top \mathbf{V}^{-1} \mathbf{x}$ term), and \mathbf{v} will be attached to the negative transpose of the variable of interest (i.e. $-\mathbf{x}^\top \mathbf{v}$) (in this example, \mathbf{x} is the variable of interest). Once we identify \mathbf{V}^{-1} and \mathbf{v} , we can sample from the full conditional distribution as $MVN(\mathbf{V} \mathbf{v}, \mathbf{V})$.

Looking back to the full conditional for β , we will find \mathbf{V} and \mathbf{v} in the following terms of the product

$$\beta^\top \mathbf{V}^{-1} \beta - \beta^\top \mathbf{v} \dots$$

So, all we have to do is multiply $T + 1$ multivariate normal pdfs (see equation 5), combine like terms in β , and identify \mathbf{V}^{-1} and \mathbf{v} from where they appear above. Since we are incompleting the square, we don't actually need to multiply the complete pdfs, just the exponential parts where \mathbf{V} and \mathbf{v} will appear. Ignoring the other components of the pdfs, we have

$$\begin{aligned} & \prod_{t=1}^T MVN(\mathbf{y}_t \mid \mathbf{u}_t + \mathbf{X}_t \beta, \sigma^2 \mathbf{I}_J) \times MVN(\beta \mid \mu_\beta, \Sigma_\beta) \\ & \propto \prod_{t=1}^T \exp \left((\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \beta)^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \beta) \right) \\ & \quad \times \exp \left((\beta - \mu_\beta)^\top \Sigma_\beta^{-1} (\beta - \mu_\beta) \right) \\ & = \exp \left(\sum_{t=1}^T \left((\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \beta)^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \beta) \right) + (\beta - \mu_\beta)^\top \Sigma_\beta^{-1} (\beta - \mu_\beta) \right) \\ & = \exp \left(\sum_{t=1}^T \left(\mathbf{y}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{y}_t - \mathbf{y}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{u}_t - \mathbf{y}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t \beta \right. \right. \\ & \quad \left. \left. - \mathbf{u}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{y}_t + \mathbf{u}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{u}_t + \mathbf{u}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t \beta \right. \right. \\ & \quad \left. \left. - \beta^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{y}_t + \beta^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{u}_t + \beta^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t \beta \right) \right. \\ & \quad \left. + \beta^\top \Sigma_\beta^{-1} \beta - \beta^\top \Sigma_\beta^{-1} \mu_\beta - \mu_\beta^\top \Sigma_\beta^{-1} \beta + \mu_\beta^\top \Sigma_\beta^{-1} \mu_\beta \right) \\ & \propto \exp \left(\beta^\top \left(\sum_{t=1}^T \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t + \Sigma_\beta^{-1} \right) \beta - \beta^\top \left(\sum_{t=1}^T \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t) + \Sigma_\beta^{-1} \mu_\beta \right) - \dots \right) \end{aligned}$$

So we can identify

$$\begin{aligned} \mathbf{V}^{-1} &= \sum_{t=1}^T \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t + \Sigma_\beta^{-1} \\ \mathbf{v} &= \sum_{t=1}^T \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t) + \Sigma_\beta^{-1} \mu_\beta \end{aligned}$$

and we can update β from its full conditional as $MVN(\mathbf{V}\mathbf{v}, \mathbf{V})$.

Update $\mathbf{u}_t, t = 1, \dots, T - 1$

Looking at (10), we see that the term \mathbf{u}_t appears in 3 terms, namely the [likelihood for \$\mathbf{y}_t\$](#) , the [prior for \$\mathbf{u}_t\$](#) , and in the [prior for \$\mathbf{u}_{t+1}\$](#) . The product of these terms is the product of multivariate normal distributions,

$$MVN(\mathbf{y}_t \mid \mathbf{u}_t + \mathbf{X}_t \beta, \sigma^2 \mathbf{I}_J) \times MVN(\mathbf{u}_t \mid \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times MVN(\mathbf{u}_{t+1} \mid \mathbf{u}_t, \tau_{t+1}^2 \mathbf{Q}(\rho_{t+1})).$$

Again, we will *incomplete the square* to identify the mean and covariance matrix of the resulting multivariate normal full conditional distribution for \mathbf{u}_t . Specifically, we will find \mathbf{V} and \mathbf{v} in the following terms of the product

$$\mathbf{u}_t^\top \mathbf{V}^{-1} \mathbf{u}_t - \mathbf{u}_t^\top \mathbf{v} \dots$$

We then have

$$\begin{aligned} & MVN(\mathbf{y}_t \mid \mathbf{u}_t + \mathbf{X}_t \boldsymbol{\beta}, \sigma^2 \mathbf{I}_J) \times MVN(\mathbf{u}_t \mid \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times MVN(\mathbf{u}_{t+1} \mid \mathbf{u}_t, \tau_{t+1}^2 \mathbf{Q}(\rho_{t+1})) \\ & \propto \exp \left((\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \boldsymbol{\beta})^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \boldsymbol{\beta}) \right) \times \exp \left((\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} (\mathbf{u}_t - \mathbf{u}_{t-1}) \right) \\ & \quad \times \exp \left((\mathbf{u}_{t+1} - \mathbf{u}_t)^\top (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} (\mathbf{u}_{t+1} - \mathbf{u}_t) \right) \\ & = \exp \left((\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \boldsymbol{\beta})^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t \boldsymbol{\beta}) + (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} (\mathbf{u}_t - \mathbf{u}_{t-1}) \right. \\ & \quad \left. + (\mathbf{u}_{t+1} - \mathbf{u}_t)^\top (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} (\mathbf{u}_{t+1} - \mathbf{u}_t) \right) \\ & = \exp \left(\mathbf{y}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{y}_t - \mathbf{y}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{u}_t - \mathbf{y}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t \boldsymbol{\beta} \right. \\ & \quad - \mathbf{u}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{y}_t + \mathbf{u}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{u}_t + \mathbf{u}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t \boldsymbol{\beta} \\ & \quad - \boldsymbol{\beta}^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{y}_t + \boldsymbol{\beta}^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{u}_t + \boldsymbol{\beta}^\top \mathbf{X}_t^\top (\sigma^2 \mathbf{I}_J)^{-1} \mathbf{X}_t \boldsymbol{\beta} \\ & \quad + \mathbf{u}_t^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} \mathbf{u}_t - \mathbf{u}_t^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} \mathbf{u}_{t-1} - \mathbf{u}_{t-1}^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} \mathbf{u}_t + \mathbf{u}_{t-1}^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} \mathbf{u}_{t-1} \\ & \quad + \mathbf{u}_{t+1}^\top (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \mathbf{u}_{t+1} - \mathbf{u}_{t+1}^\top (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \mathbf{u}_t \\ & \quad \left. - \mathbf{u}_t^\top (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \mathbf{u}_{t+1} + \mathbf{u}_t^\top (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \mathbf{u}_t \right) \\ & \propto \exp \left(\mathbf{u}_t^\top \left((\sigma^2 \mathbf{I}_J)^{-1} + (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} + (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \right) \mathbf{u}_t \right. \\ & \quad \left. - \mathbf{u}_t^\top \left((\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) + (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} \mathbf{u}_{t-1} + (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \mathbf{u}_{t+1} \right) - \dots \right) \end{aligned}$$

So we can identify

$$\begin{aligned} \mathbf{V}^{-1} &= (\sigma^2 \mathbf{I}_J)^{-1} + (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} + (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \\ \mathbf{v} &= (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) + (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} \mathbf{u}_{t-1} + (\tau_{t+1}^2 \mathbf{Q}(\rho_{t+1}))^{-1} \mathbf{u}_{t+1} \end{aligned}$$

and we can update \mathbf{u}_t from its full conditional as $MVN(\mathbf{V}\mathbf{v}, \mathbf{V})$.

Update \mathbf{u}_T

To derive the full conditional for \mathbf{u}_T , it will be the same as above, except the terms for \mathbf{u}_{t+1} 's prior will not be present, and the full conditional mean and covariance terms are of the form

$$\begin{aligned} \mathbf{V}^{-1} &= (\sigma^2 \mathbf{I}_J)^{-1} + (\tau_T^2 \mathbf{Q}(\rho_T))^{-1} \\ \mathbf{v} &= (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_T - \mathbf{X}_T \boldsymbol{\beta}) + (\tau_T^2 \mathbf{Q}(\rho_T))^{-1} \mathbf{u}_{T-1} \end{aligned}$$

and we can update \mathbf{u}_T from its full conditional as $MVN(\mathbf{V}\mathbf{v}, \mathbf{V})$.

★ Alternative update for $u_{j,t}, j = 1, \dots, J, \quad t = 1, \dots, T$

Previously, we have jointly updated \mathbf{u}_t as a multivariate normal distribution of dimension J . However, if J is large, this can be computationally infeasible. One advantage of the CAR model structure is that it can be equivalently represented as the joint distribution of J many univariate normal random variables, where the distribution of each variable is conditional on the remaining $J-1$ variables. Better still, the distribution of a given $u_{j,t}$ only depends on its adjacent neighbors (hence the name, ‘conditional autoregressive model’), so we can recursively update each $u_{j,t}$ quite efficiently.

To do this, let \mathcal{J}_j be the set of neighbor indices of areal unit j . We assume that \mathcal{J}_j does not change over time. Similarly, let d_j be the number of neighbors belonging to areal unit j . Then we have

$$u_{j,t} | u_{k,t}, k \neq j \sim N \left(u_{j,t-1} + \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}, \frac{\tau_t^2}{d_j} \right). \quad (12)$$

The intuition behind equation (12) is that the distribution of $u_{j,t}$ given all other $u_{k,t}, k \neq j$ is normally distributed with a mean that is a weighted average of its neighbors centered on $u_{j,t-1}$, and with scaled variance which decreases as d_j increases. Using this form, we can derive the full conditional distribution of each spatio-temporal random effect $u_{j,t}$ in turn, leading to $J + T$ many 1-dimensional operations rather than T -many J -dimensional operations. In practice, this may be slower for small J , but when J is large, the computational savings can be substantial.

Writing (10) in terms of univariate distributions, we have

$$\begin{aligned} & \prod_{t=1}^T \prod_{j=1}^J N(y_{j,t} | u_{j,t} + \mathbf{x}_{j,t}\boldsymbol{\beta}, \sigma^2) \times \prod_{t=1}^T \prod_{j=1}^J N \left(u_{j,t} | u_{j,t-1} + \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}, \frac{\tau_t^2}{d_j} \right) \\ & \times MVN(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \times IG(\sigma^2 | a, b) \times \prod_{t=1}^T IG(\tau_t^2 | a_t, b_t) \times \prod_{t=1}^T U(\rho_t | c_t, d_t). \end{aligned} \quad (13)$$

We can then see that $u_{j,t}$ appears in 3 terms of (13), namely the **likelihood of $y_{j,t}$** , the **prior for $u_{j,t}$** , and the **prior for $u_{j,t+1}$** . We do not consider where $u_{j,t}$ appears in the summation terms of its neighbor distributions, because for these distributions the value of $u_{j,t}$ is being conditioned on, it is treated as fixed.

The product of these three distributions is

$$\begin{aligned} & N(y_{j,t} | u_{j,t} + \mathbf{x}_{j,t}\boldsymbol{\beta}, \sigma^2) \times N \left(u_{j,t} | u_{j,t-1} + \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}, \frac{\tau_t^2}{d_j} \right) \\ & \times N \left(u_{j,t+1} | u_{j,t} + \rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}, \frac{\tau_{t+1}^2}{d_j} \right). \end{aligned}$$

Since all three terms are normally distributed, we know that their product will also be normally distributed. Hence, we can find the mean and variance of the resulting full conditional normal distribution by *incompleting the square*. Thus far, we have discussed incompleting the square for products of multivariate normal distributions, but the same technique will work for products of univariate normals.

Specifically, the resulting full conditional distribution will be distributed as $N(Vv, V)$, where V and v will appear in the product above in the form $u_{j,t}^2 V^{-1} - 2u_{j,t}v + Vv^2$. We then have

$$\begin{aligned}
& N(y_{j,t} \mid u_{j,t} + \mathbf{x}_{j,t}\beta, \sigma^2) \times N\left(u_{j,t} \mid u_{j,t-1} + \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}, \frac{\tau_t^2}{d_j}\right) \\
& \times N\left(u_{j,t+1} \mid u_{j,t} + \rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}, \frac{\tau_{t+1}^2}{d_j}\right) \\
& \propto \exp\left(\frac{1}{\sigma^2}(y_{j,t} - u_{j,t} - \mathbf{x}_{j,t}\beta)^2\right) \times \exp\left(\frac{d_j}{\tau_t^2}\left(u_{j,t} - u_{j,t-1} - \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}\right)^2\right) \\
& \times \exp\left(\frac{d_j}{\tau_{t+1}^2}\left(u_{j,t+1} - u_{j,t} - \rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}\right)^2\right) \\
& \propto \exp\left(\frac{1}{\sigma^2}(y_{j,t} - u_{j,t} - \mathbf{x}_{j,t}\beta)^2 + \frac{d_j}{\tau_t^2}\left(u_{j,t} - u_{j,t-1} - \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}\right)^2\right. \\
& \quad \left.+ \frac{d_j}{\tau_{t+1}^2}\left(u_{j,t+1} - u_{j,t} - \rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}\right)^2\right) \\
& \propto \exp\left(\frac{1}{\sigma^2}(u_{j,t}^2 - 2u_{j,t}y_{j,t} + 2u_{j,t}\mathbf{x}_{j,t}\beta) + \frac{d_j}{\tau_t^2}\left(u_{j,t}^2 - 2u_{j,t}u_{j,t-1} - 2u_{j,t}\rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}\right)\right. \\
& \quad \left.+ \frac{d_j}{\tau_{t+1}^2}\left(u_{j,t}^2 - 2u_{j,t}u_{j,t+1} + 2u_{j,t}\rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}\right)\right) \\
& = \exp\left[u_{j,t}^2\left(\left(\frac{1}{\sigma^2}\right) + \left(\frac{d_j}{\tau_t^2}\right) + \left(\frac{d_j}{\tau_{t+1}^2}\right)\right)\right. \\
& \quad - 2u_{j,t}\left(\left(\frac{1}{\sigma^2}\right)(y_{j,t} - \mathbf{x}_{j,t}\beta) + \left(\frac{d_j}{\tau_t^2}\right)\left(u_{j,t-1} + \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}\right)\right. \\
& \quad \left.+ \left(\frac{d_j}{\tau_{t+1}^2}\right)\left(u_{j,t+1} - \rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}\right)\right) + \dots\left. \right]
\end{aligned}$$

So we can identify

$$\begin{aligned}
V^{-1} &= \left(\frac{1}{\sigma^2}\right) + \left(\frac{d_j}{\tau_t^2}\right) + \left(\frac{d_j}{\tau_{t+1}^2}\right) \\
v &= \left(\frac{1}{\sigma^2}\right)(y_{j,t} - \mathbf{x}_{j,t}\beta) + \left(\frac{d_j}{\tau_t^2}\right)\left(u_{j,t-1} + \rho_t \sum_{k \in \mathcal{J}_j} \frac{u_{k,t} - u_{k,t-1}}{d_j}\right) \\
& \quad + \left(\frac{d_j}{\tau_{t+1}^2}\right)\left(u_{j,t+1} - \rho_{t+1} \sum_{k \in \mathcal{J}_j} \frac{u_{k,t+1} - u_{k,t}}{d_j}\right)
\end{aligned}$$

and we can update $u_{j,t}$ via it's full conditional as $N(Vv, v)$.

For $t = T$, we simply ignore the terms corresponding to the prior distribution of $u_{j,t+1}$, and we have

$$V^{-1} = \left(\frac{1}{\sigma^2} \right) + \left(\frac{d_j}{\tau_T^2} \right)$$

$$v = \left(\frac{1}{\sigma^2} \right) (y_{j,T} - \mathbf{x}_{j,T}\boldsymbol{\beta}) + \left(\frac{d_j}{\tau_T^2} \right) \left(u_{j,T-1} + \rho_T \sum_{k \in \mathcal{J}_j} \frac{u_{k,T} - u_{k,T-1}}{d_j} \right)$$

and we can update $u_{j,T}$ via it's full conditional as $N(Vv, v)$.

Update σ^2

Looking at (10), we see that the term σ^2 appears in $T + 1$ terms, namely the likelihood for $\mathbf{y}_t, t = 1, \dots, T$ and it's own prior distribution. The product of these terms is takes the form,

$$\prod_{t=1}^T MVN(\mathbf{y}_t \mid \mathbf{u}_t + \mathbf{X}_t\boldsymbol{\beta}, \sigma^2 \mathbf{I}_J) \times IG(\sigma^2 \mid a, b).$$

We know that the inverse gamma distribution (6) is conjugate for the multivariate normal distribution (5), and hence the resulting full conditional distribution will also be distributed as Inverse Gamma. Our goal then is to multiply these densities and identify the resulting expression as an Inverse Gamma distribution for σ^2 . Specifically,

$$\begin{aligned} & \prod_{t=1}^T MVN(\mathbf{y}_t \mid \mathbf{u}_t + \mathbf{X}_t\boldsymbol{\beta}, \sigma^2 \mathbf{I}_J) \times IG(\sigma^2 \mid a, b) \\ &= \prod_{t=1}^T (2\pi)^{-JT/2} |\sigma^2 \mathbf{I}_J|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t\boldsymbol{\beta})^\top (\sigma^2 \mathbf{I}_J)^{-1} (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t\boldsymbol{\beta}) \right) \\ & \quad \times \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp \left(\frac{-b}{\sigma^2} \right) \\ & \propto (\sigma^2)^{\frac{-JT}{2} - a - 1} \exp \left(\frac{-\frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t\boldsymbol{\beta})^\top (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t\boldsymbol{\beta}) - b}{\sigma^2} \right) \end{aligned}$$

which we can identify as in the form of an Inverse Gamma pdf (6), with parameters

$$\tilde{a} = a + \frac{JT}{2}$$

$$\tilde{b} = b + \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t\boldsymbol{\beta})^\top (\mathbf{y}_t - \mathbf{u}_t - \mathbf{X}_t\boldsymbol{\beta})$$

and we can update σ^2 via it's full conditional as $IG(\tilde{a}, \tilde{b})$.

Update τ_t^2 , $t = 1, \dots, T$

When viewing the full posterior (10), we see that τ_t^2 appears in two terms, namely in the prior distribution of \mathbf{u}_t and in its own prior distribution, which is distributed as inverse gamma. The product of these distributions will be

$$MVN(\mathbf{u}_t \mid \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times IG(\tau_t^2 \mid a_t, b_t)$$

which will form a Inverse Gamma full conditional distribution by conjugacy. As before, we multiply these densities to obtain

$$\begin{aligned} & MVN(\mathbf{u}_t \mid \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times IG(\tau_t^2 \mid a_t, b_t) \\ &= (2\pi)^{-J/2} |\tau_t^2 \mathbf{Q}(\rho_t)|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} (\mathbf{u}_t - \mathbf{u}_{t-1}) \right) \\ & \quad \times \frac{b_t^{a_t}}{\Gamma(a_t)} (\tau_t^2)^{-a_t-1} \exp \left(\frac{-b_t}{\tau_t^2} \right) \\ & \propto (\tau_t^2)^{-\frac{J}{2}-a_t-1} \exp \left(\frac{-\frac{1}{2} (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\mathbf{u}_t - \mathbf{u}_{t-1}) - b_t}{\tau_t^2} \right) \end{aligned}$$

which we can identify as in the form of an Inverse Gamma pdf (6), with parameters

$$\begin{aligned} \tilde{a}_t &= a_t + \frac{J}{2} \\ \tilde{b}_t &= b_t + \frac{1}{2} (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\mathbf{u}_t - \mathbf{u}_{t-1}) \end{aligned}$$

and we can update τ_t^2 via its full conditional as $IG(\tilde{a}_t, \tilde{b}_t)$.

Update ρ_t , $t = 1, \dots, T$

Until now we have worked directly with full conditional distributions for model parameters via conjugate priors. This has made things easier, as we could directly determine the form of the full conditional distribution. Now, we introduce a *Metropolis step* to update $\rho_t, t = 1, \dots, T$. The metropolis step is essentially a process whereby you propose a new value for your parameter, and accept or reject that proposed value based on a proposal distribution. In (1), we chose a Uniform prior for ρ_t as $U(c_t, d_t)$. Importantly, the Uniform distribution has bounded support, meaning our proposal distribution (which is typically Gaussian) may propose values outside the support of ρ_t . Hence, we will introduce a *transformation* for ρ_t , which will give it support on the real line \mathbb{R} . This way, we can use a Gaussian proposal distribution, evaluate the acceptance probability of this transformed variable value, and then back transform our sample upon acceptance. We will outline the general framework here.

First, consider the full posterior (10). As we have done in all steps, we identify the terms in which ρ_t appears, namely in the prior for \mathbf{u}_t and in its own prior distribution, which is Uniform on (c_t, d_t) . The product of these densities is

$$\begin{aligned}
& MVN(\mathbf{u}_t \mid \mathbf{u}_{t-1}, \tau_t^2 \mathbf{Q}(\rho_t)) \times U(\rho_t \mid c_t, d_t) \\
&= (2\pi)^{-J/2} |\tau_t^2 \mathbf{Q}(\rho_t)|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} (\mathbf{u}_t - \mathbf{u}_{t-1}) \right) \\
&\quad \times \frac{1}{d_t - c_t} \mathbf{1}_{c_t < \rho_t < d_t}
\end{aligned}$$

where $\mathbf{1}$ is the indicator function. Next, we employ a transformation on the Uniform variable ρ_t of the form

$$g(\rho_t) = \log \left(\frac{\rho_t - c_t}{d_t - \rho_t} \right).$$

To get the distribution of this newly transformed variable $g(\rho_t)$, we require the Jacobian. First, we find the inverse function of the transformation, which is

$$g^{-1}(g(\rho_t)) = \frac{d_t e^{g(\rho_t)} + c_t}{e^{g(\rho_t)} + 1}.$$

Next, we take the derivative of this function with respect to $g(\rho_t)$, which gives

$$\frac{e^{g(\rho_t)}(d_t - c_t)}{(e^{g(\rho_t)} + 1)^2},$$

and plugging the expression for $g(\rho_t)$ into this equation yields

$$\frac{\frac{\rho_t - c_t}{d_t - \rho_t}(d_t - c_t)}{\left(\frac{\rho_t - c_t}{d_t - \rho_t} + 1 \right)^2} \propto \frac{\frac{\rho_t - c_t}{d_t - \rho_t}}{\left(\frac{d_t - \rho_t + \rho_t - c_t}{d_t - \rho_t} \right)^2} \propto \frac{\frac{\rho_t - c_t}{d_t - \rho_t}}{\frac{1}{(d_t - \rho_t)^2}} = (\rho_t - c_t)(d_t - \rho_t).$$

With this jacobian adjustment, we can now determine the *log target density* to which we will add the Jacobian adjustment. We have

$$\begin{aligned}
& \log \left[(2\pi)^{-J/2} |\tau_t^2 \mathbf{Q}(\rho_t)|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} (\mathbf{u}_t - \mathbf{u}_{t-1}) \right) \right. \\
& \quad \left. \times \frac{1}{d_t - c_t} \mathbf{1}_{c_t < \rho_t < d_t} \times (\rho_t - c_t)(d_t - \rho_t) \right] \\
& \propto -\frac{1}{2} \log |\tau_t^2 \mathbf{Q}(\rho_t)| - \frac{1}{2} (\mathbf{u}_t - \mathbf{u}_{t-1})^\top (\tau_t^2 \mathbf{Q}(\rho_t))^{-1} (\mathbf{u}_t - \mathbf{u}_{t-1}) + \log(\rho_t - c_t) + \log(d_t - \rho_t)
\end{aligned}$$

We can then proceed by generating a proposed value of $g(\rho_t)$ from some proposal distribution (usually Gaussian), evaluating the log target density plus Jacobian adjustment, and then accepting or rejecting the back-transformed $g^{-1}(g(\rho_t))$ based on this acceptance probability.