Conditionally autoregressive (CAR) model

Introduction to model

Define the outcome variable, Y_i, as the alcohol sales for county i in Pennsylvania (PA), $i = 1, \ldots, n$, where n is the number of counties (67). Then, we will specify the following model,

$$Y_i = X_i^T \boldsymbol{\beta} + \theta_i + \epsilon_i, \quad i = 1, \ldots, n$$

where, X_i is a $p \times 1$ vector of county-specific covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficient parameters, θ_i is a location specific intercept and $\epsilon_i \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2\right)$.

We will control for spatial dependency through specification of the distribution of θ_i. If we define $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]^T$ then we can specify a conditionally autoregressive (CAR) prior with joint specification given by

$$f\left(\boldsymbol{\theta}|\tau^2\right) \sim CAR(\tau^2) \propto \exp\left\{-\frac{\boldsymbol{\theta}^T \mathbf{W}^* \boldsymbol{\theta}}{2\tau^2}\right\}, \tag{1}$$

where the components of $\mathbf{W}^*$ are defined as $[\mathbf{W}^*]_{ij} = -w_{ij}$ for $i \neq j$ and $[\mathbf{W}^*]_{ii} = \sum_{j=1}^n w_{ij}$, with w_{ij} being the components of the neighborhood adjacency matrix $\mathbf{W}$ (i.e. $w_{ij} = 1(i \sim j)$, where $i \sim j$ indicates locations i and j are adjacent). The CAR prior is best interpreted in its conditional formulation, however for computational purposes we will work with the joint specification. For insight, the conditional specification can be written as follows,

$$\theta_i|\theta_{\delta_i}, \tau^2 \sim \text{N}\left(\frac{\sum_{j=1}^n w_{ij}\theta_i}{\sum_{j=1}^n w_{ij}}, \frac{\tau^2}{\sum_{j=1}^n w_{ij}}\right),$$

I think θ_i should be θ_{δ_i} -BB

where δ_i is the set of all neighbors of location i. This conditional specification demonstrates that each location specific intercept, θ_i, is a weighted average of its neighbors, with variance shrinking as the number of neighbors increases.

Before proceeding, we will rewrite the model in matrix formulation,

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

where $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$, $\mathbf{X} = [X_1, \ldots, X_n]^T$ and $\boldsymbol{\epsilon} \sim \text{N}\left(\mathbf{0}_n, \sigma^2\mathbf{I}_n\right)$. This induces the following joint likelihood:

$$f\left(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2\right) \sim \text{N}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\theta}, \sigma^2\mathbf{I}_n\right).$$

We can now write the full data likelihood as

$$\begin{aligned} f\left(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \tau^2\right) &\propto f\left(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \tau^2\right) \times f\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \tau^2\right) \\ &\propto f\left(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2\right) \times f\left(\boldsymbol{\theta}|\tau^2\right) \times f\left(\boldsymbol{\beta}, \sigma^2, \tau^2\right). \end{aligned}$$

Prior specification

To complete the Bayesian framework we must specify a form for our joint prior, $f\left(\boldsymbol{\beta}, \sigma^2, \tau^2\right)$. In order to ease computation, we will specify independent priors such that $f\left(\boldsymbol{\beta}, \sigma^2, \tau^2\right) = f\left(\boldsymbol{\beta}\right) f\left(\sigma^2\right) f\left(\tau^2\right)$. Then, priors are specified as follows:

$$\boldsymbol{\beta} \sim \mathrm{N}\left(\mathbf{0}_p, \sigma_\beta^2 \mathbf{I}_p\right), \quad \sigma^2 \sim \mathrm{IG}\left(\alpha_\sigma, \beta_\sigma\right), \quad \tau^2 \sim \mathrm{IG}\left(\alpha_\tau, \beta_\tau\right).$$

Now, we can write down the full conditionals for use in MCMC.

Full conditionals

Derivation for $\boldsymbol{\theta}$

First, define $\boldsymbol{\gamma} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$. Then, we can derive the full conditional as follows,

$$\begin{aligned} f\left(\boldsymbol{\theta} | \mathbf{Y}, \boldsymbol{\beta}, \sigma^2, \tau^2\right) &\propto f\left(\mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2\right) \times f(\boldsymbol{\theta} | \tau^2) \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\boldsymbol{\gamma} - \boldsymbol{\theta}\right)^T \mathbf{I}_n \left(\boldsymbol{\gamma} - \boldsymbol{\theta}\right)}{\sigma^2} + \frac{\boldsymbol{\theta}^T \mathbf{W}^* \boldsymbol{\theta}}{\tau^2}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\theta}^T\left(\frac{\mathbf{I}_n}{\sigma^2} + \frac{\mathbf{W}^*}{\tau^2}\right)\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\left(\frac{\boldsymbol{\gamma}}{\sigma^2}\right)\right]\right\} \\ &\sim N(\mathbb{E}_\theta, \mathbb{V}_\theta). \end{aligned}$$

where $\mathbb{V}_\theta = \left(\frac{\mathbf{I}_n}{\sigma^2} + \frac{\mathbf{W}^*}{\tau^2}\right)^{-1}$ and $\mathbb{E}_\theta = \mathbb{V}_\theta\left(\frac{\boldsymbol{\gamma}}{\sigma^2}\right)$.

Derivation for $\boldsymbol{\beta}$

First, define $\boldsymbol{\gamma} = \mathbf{Y} - \boldsymbol{\theta}$. Then, we can derive the full conditional as follows,

$$\begin{aligned} f\left(\boldsymbol{\beta} | \mathbf{Y}, \boldsymbol{\theta}, \sigma^2, \tau^2\right) &\propto f\left(\mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2\right) \times f(\boldsymbol{\beta}) \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\boldsymbol{\gamma} - \mathbf{X}\boldsymbol{\beta}\right)^T \mathbf{I}_n \left(\boldsymbol{\gamma} - \mathbf{X}\boldsymbol{\beta}\right)}{\sigma^2} + \frac{\boldsymbol{\beta}^T \mathbf{I}_p \boldsymbol{\beta}}{\sigma_\beta^2}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T\left(\frac{\mathbf{X}^T\mathbf{X}}{\sigma^2} + \frac{\mathbf{I}_p}{\sigma_\beta^2}\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\left(\frac{\mathbf{X}^T\boldsymbol{\gamma}}{\sigma^2}\right)\right]\right\} \\ &\sim N(\mathbb{E}_\beta, \mathbb{V}_\beta). \end{aligned}$$

where, $\mathbb{V}_\beta = \left(\frac{\mathbf{X}^T\mathbf{X}}{\sigma^2} + \frac{\mathbf{I}_p}{\sigma_\beta^2}\right)^{-1}$ and $\mathbb{E}_\beta = \mathbb{V}_\beta\left(\frac{\mathbf{X}^T\boldsymbol{\gamma}}{\sigma^2}\right)$.

Derivation for σ^2

First, define $\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\theta}$. Then, we can derive the full conditional as follows,

$$\begin{aligned}
f\left(\sigma^2|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\beta},\tau^2\right) &\propto f\left(\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\beta},\sigma^2\right) \times f\left(\sigma^2\right)\\
&\propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{Y}-\boldsymbol{\gamma}\right)^T\left(\mathbf{Y}-\boldsymbol{\gamma}\right)\right\}\left(\sigma^2\right)^{-\alpha_\sigma-1}\exp\left\{-\frac{\beta_\sigma}{\sigma^2}\right\}\\
&\propto \left(\sigma^2\right)^{-\left(\alpha_\sigma+\frac{n}{2}\right)-1}\exp\left\{-\frac{1}{\sigma^2}\left[\beta_\sigma+\frac{\left(\mathbf{Y}-\boldsymbol{\gamma}\right)^T\left(\mathbf{Y}-\boldsymbol{\gamma}\right)}{2}\right]\right\}\\
&\sim IG\left(\alpha_\sigma+\frac{n}{2},\beta_\sigma+\frac{\left(\mathbf{Y}-\boldsymbol{\gamma}\right)^T\left(\mathbf{Y}-\boldsymbol{\gamma}\right)}{2}\right).
\end{aligned}$$

Derivation for τ^2

$$\begin{aligned}
f\left(\tau^2|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\beta},\sigma^2\right) &\propto f\left(\boldsymbol{\theta}|\tau^2\right) \times f\left(\tau^2\right)\\
&\propto \left(\tau^2\right)^{-\frac{n-G}{2}}\exp\left\{-\frac{\boldsymbol{\theta}^T\mathbf{W}^*\boldsymbol{\theta}}{2\tau^2}\right\}\left(\tau^2\right)^{-\alpha_\tau-1}\exp\left\{-\frac{\beta_\tau}{\tau^2}\right\}\\
&\propto \left(\tau^2\right)^{-\left(\alpha_\tau+\frac{n-G}{2}\right)-1}\exp\left\{-\frac{1}{\tau^2}\left[\beta_\tau+\frac{\boldsymbol{\theta}^T\mathbf{W}^*\boldsymbol{\theta}}{2}\right]\right\}\\
&\sim IG\left(\alpha_\sigma+\frac{n-G}{2},\beta_\sigma+\frac{\boldsymbol{\theta}^T\mathbf{W}^*\boldsymbol{\theta}}{2}\right).
\end{aligned}$$

Deviance

To calculate DIC we compute deviance at each scan of the MCMC sampler. At scan s, the deviance can be computed by evaluating,

$$\text{Deviance}^{(s)} = -2\log\left\{f\left(\mathbf{Y}\middle|\boldsymbol{\theta}^{(s)},\boldsymbol{\beta}^{(s)},\sigma^{2(s)}\right)\right\}.$$