Project 2: GMRF:s with Non-Gaussian Data

Ida Lunde Hygum and Khadijah Habib

January 13, 2025

1 Introduction

In this report we model non-Gaussian count data by employing a Poisson distribution alongside latent spatial models such as Conditional Autoregressive (CAR) and Simultaneous Autoregressive (SAR) models. The task encompasses theoretical derivations, parameter estimation and the reconstruction of latent fields using accident count data from British coal mines spanning the years 1750 to 1980. The spatial field is one dimensional since we only consider the temporal correlation between years.

1.1 Theory

In the theory section we will first describe the CAR and SAR models we can use to estimate the number of accidents and subsequently derive the approximate log-likehood used to estimate the parameters in the precision matrix Q .

1.2 CAR and SAR models

Two commonly used methods in spatial statistics are Conditional (CAR) or Simultaneous (SAR) autoregressive models, which consider the spatial dependence in errors, where both are Markov models. In CAR models a value at an unknown location depends only on it's nearest neighbors, while a SAR model may include a larger neighborhood, a SAR model can therefore be written as a unique CAR model [VHH18]. Since the neighborhood in the CAR and SAR models is relatively small, the precision matrix Q will be sparse, making computations efficient. CAR and SAR models can be used to estimate the precision matrix Q:

$$Q_{CAR} = \tau(\kappa^2 C + G) \tag{1}$$

$$Q_{SAR} = \tau(\kappa^4 C + 2\chi^2 G + G_2) \tag{2}$$

The derivation of the C and G matrices will be illustrated with an example below.

1.2.1 Computations of C and G Matrices

Matrix C

The elements of the matrix C are computed as:

$$C_{ij} = \int_{-\infty}^{\infty} \psi_i(t)\psi_j(t)\, dt.$$

For the given 1D grid with tent basis functions and spacing $h = 1$, the matrix C is diagonal because the tent functions overlap only with their immediate neighbors. Thus, $C_{ij} = 0$ for $i \neq j$.

For the diagonal elements:

$$C_{ii} = \int_{i-1}^{i} \psi_i(t) \cdot 1\, dt + \int_{i}^{i+1} \psi_i(t) \cdot 1\, dt.$$

The total area under each basis function is equal to the grid spacing h. Therefore:

$$C_{ii} = h = 1.$$

The final C matrix is:

$$C = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \end{bmatrix}.$$

Matrix G

The elements of the matrix G are computed as:

$$G_{ij} = \int_{-\infty}^{\infty} \psi_i'(t)\psi_j'(t)\,dt.$$

For tent basis functions, the derivative $\psi_i'(t)$ is constant over each interval. Specifically:

$$\psi_i'(t) = \begin{cases} 1/h & \text{for } t \in [i-1, i], \\ -1/h & \text{for } t \in [i, i+1]. \end{cases}$$

For diagonal terms ($i = j$):

$$G_{ii} = \int_{i-1}^{i} \left(\frac{1}{h}\right)^2 dt + \int_{i}^{i+1} \left(-\frac{1}{h}\right)^2 dt.$$

With $h = 1$:

$$G_{ii} = \int_{i-1}^{i} 1\,dt + \int_{i}^{i+1} 1\,dt = 1 + 1 = 2.$$

For off-diagonal terms ($i = j \pm 1$):

$$G_{i,i+1} = G_{i+1,i} = \int_{i}^{i+1} \left(\frac{1}{h}\right)\left(-\frac{1}{h}\right) dt.$$

With $h = 1$:

$$G_{i,i+1} = -1.$$

For non-adjacent terms ($|i - j| > 1$), $G_{ij} = 0$ since the basis functions do not overlap. The final G matrix is:

$$G = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

1.2.2 Effect of Grid Spacing h

If the grid spacing is h instead of 1, the scaling factors change as follows:

- For C, the area under each basis function scales proportionally to h:

$$C_{ii} \propto h.$$

- For G, the derivative scaling introduces a factor of $1/h^2$:

$$G_{ij} \propto \frac{1}{h^2}.$$

The matrices C and G must be adjusted accordingly to ensure consistency with the new grid spacing.

1.3 Approximate log-likelihood

For predicting the number of accidents we need to go through some of the theory to derive the approximate log-likelihood function ($logp(\theta|y)$), which is then used to estimate the parameters (θ) of the distribution ($p(\tilde{x}|\theta)$) and the parameters in the precisions matrix Q (τ and κ^2). This is done in a number of steps. First we compute the observation log-density $logp(y_i|z_i)$ and it's derivatives with respect to z_i. We have Poisson distributed observations (count data):

$$p(y_i|z_i) = \frac{\exp^{y*z_i}}{y_i!} * \exp(-e^{z_i}) \tag{3}$$

The log density of the poison distribution is:

$$log\ p(y_i|z_i) = log(\frac{\exp^{y*z_i}}{y_i!}) * \exp(-e^{z_i}) = log(\exp^{y*z_i}) - log(y_i!) + log(\exp(-e^{z_i})) = \tag{4}$$

$$y * z_i - log(y_i!) - e^{z_i} = y * z_i - e^{z_i} \tag{5}$$

where the term $-log(y_i!)$ drops out because it does not depend on z_i. Taking the first derivative of the $log\ p(y_i|z_i)$ wrt. z_i gives:

$$\frac{d}{dz_i}(y * z_i - e^{z_i}) = y - e^{z_i} \tag{6}$$

$$\frac{d^2}{dz_i^2}(y * z_i - e^{z_i}) = -e^{z_i} \tag{7}$$

In order to estimate the mode of x we compute the log posterior of $\tilde{x}$. The posterior of $\tilde{x}$ is:

$$p(\tilde{x}|y,\theta) = \frac{p(\tilde{x},y,\theta)}{p(y,\theta)} \propto p(y|\tilde{x},\theta)p(\tilde{x},\theta) \tag{8}$$

The probability $logp(y|\tilde{x},\theta)$ is equal to $p(y_i|z_i)$ since z is equal to $z = \tilde{A}\tilde{x}$ and the distribution $p(\tilde{x}|\theta)$ is a normal distribution with $\mu = 0$:

$$p(x|\theta) = \frac{1}{2\pi^{N/2} * |Q^{-1}|^{1/2}} \exp(-\frac{1}{2}\tilde{x}^T Q\tilde{x}) \tag{9}$$

The precision matrix Q is either a Q_{CAR} or a Q_{SAR} precision matrix, depending on which model is chosen. An expression for the log posterior of $\tilde{x}$ can be expressed as, removing terms that do not depend on x:

$$log(p(\tilde{x}|y,\theta) \propto yz_i - e^{z_i} - \frac{1}{2}\tilde{x}^T Q\tilde{x} \tag{10}$$

For estimating the mode of $\hat{x}$ ($\hat{x}^{(0)}$) we maximize the $log(p(\tilde{x}|y,\theta)$ of x:

$$\hat{x}^0 = argmax(log(p(\tilde{x}|y,\theta)) \tag{11}$$

In order to estimate the mode $\hat{x}^{(0)}$ we find the first- and second-order derivatives of the posterior, which are related to the mean and the variance, respectively, which are the central parameters for estimating the approximate loglikelihood θ. We estimate the derivatives of $log(p(\tilde{x}|y,\theta)$ with respect to $\tilde{x}$, and replace z_i with $\tilde{A}\tilde{x}$:

$$\frac{d}{d\tilde{x}}(yz_i - e^{z_i} - \frac{1}{2}(log(|Q^{-1}| + \tilde{x}^T Q\tilde{x})) = \tilde{A}^T y - \tilde{A}^T e^{\tilde{A}\tilde{x}} - Q\tilde{x} \tag{12}$$

$$\frac{d^2}{d\tilde{x}^2}(yz_i - e^{z_i} - \frac{1}{2}(log(|Q^{-1}| + \tilde{x}^T Q\tilde{x}) = Q - \tilde{A}^T H_f^{(0)}\tilde{A} \tag{13}$$

The H corresponds to the Hessian of the $logp(y|\tilde{x},\theta)$ with respect to $\tilde{x}$. We estimate the approximate log-likelihood as a function of θ:

$$logp(\theta|y) \propto logp(y|\hat{x},\theta) + log|Q_x| - \frac{1}{2}\hat{x}^T Q_x\hat{x} - log|Q_{x|y}| + \frac{1}{2}(\hat{x} - \mu_{x|y})^T Q_{x|y}(\hat{x} - \mu_{x|y}) \tag{14}$$

We pick $\hat{x}$ at the mode, which means for a Gaussian distribution the mode is equal to the mean so we get that $\hat{x} = \mu_{x|y}$ and then cancels out:

$$logp(\theta|y) \propto logp(y|\hat{x}, \theta) + log|Q_x| - \frac{1}{2}\hat{x}^T Q_x \hat{x} - log|Q_{x|y}| \tag{15}$$

To estimate the θ we do a nested optimization where we first estimate $\hat{x}^{(0)}$, which is then used to maximize:

$$\theta \approx argmax\ p(\theta|y) \tag{16}$$

This is a nested approximation, since the new θ values are then used to estimate a new $\hat{x}^{(0)}$, which again are used to estimate new θ parameters. The parameters of the precision matrix Q and the probability distribution of $p(\tilde{x}|\theta)$ are related by:

$$\theta_1 = log(\tau) \tag{17}$$

$$\theta_2 = log(\kappa^2) \tag{18}$$

1.4 Reconstruction of the Latent Field

The latent field was reconstructed as the mode of the posterior distribution:

$$E(\mathbf{x}|\mathbf{y}, \hat{\theta}) = \mathbf{x}_{\text{mode}}$$

This reconstruction provides the conditional expectation of the log-intensity, $\mathbf{z}$, given the observed data and estimated parameters:

$$\mathbf{z} = A\mathbf{x}_{\text{mode}} + B\beta$$

1.5 Posterior Variance and Confidence Intervals

The posterior precision matrix, $Q_{\mathbf{x}|\mathbf{y}}$, was calculated using a Taylor expansion around the mode:

$$Q_{\mathbf{x}|\mathbf{y}} = Q - \tilde{A}^T H_f^{(0)} \tilde{A}$$

The posterior variance can be obtained either by inverting $Q_{\mathbf{x}|\mathbf{y}}$, and deriving the confidence intervals as:

$$\text{Lower Bound} = \mathbf{z} - 1.96 \cdot \sqrt{\text{diag}(Q_{\mathbf{x}|\mathbf{y}}^{-1})}$$

$$\text{Upper Bound} = \mathbf{z} + 1.96 \cdot \sqrt{\text{diag}(Q_{\mathbf{x}|\mathbf{y}}^{-1})}$$

For larger Q matrices where inverting is not feasible one can use simulation instead where samples are generated from the posterior distribution $p(\tilde{x}|\theta, y)$ of the latent field $\tilde{x}$ given the data y and the parameters θ:

$$X_{sample} = u_{x|y} + R_{x|y}^{-1} * \varepsilon \tag{19}$$

The matrix $R_{x|y}$ is the Cholesky of the posterior $Q_{x|y}$ matrix and ε is a random error term ($\varepsilon \sim N(0, I)$).

2 Results

2.1 Estimation of Parameters

The parameters of the SAR model were estimated using a Poisson log-likelihood framework, as the data represents count observations. The global mode was determined as described in the theory section. The parameter optimization was conducted using `fminsearch`, where the log-posterior is minimized. Both the SAR and The CAR models where tested and the parameters recorded. The estimated parameters for the CAR model were:

- τ_{CAR}: $e^{\hat{\theta}_1} = 41.558$
- κ^2_{CAR}: $e^{\hat{\theta}_2} = 9.76 * 10^{-6}$

The estimated parameters for the SAR model were:

- τ_{SAR}: $e^{\hat{\theta}_1} = 2.045 * 10^4$
- κ^2_{SAR}: $e^{\hat{\theta}_2} = 2.6218 * 10^{-4}$

Figure 1 shows the training and validation data over the period 1750-1980. The validation data was selected as 5 random decades and corresponds to 50 predictions in total.

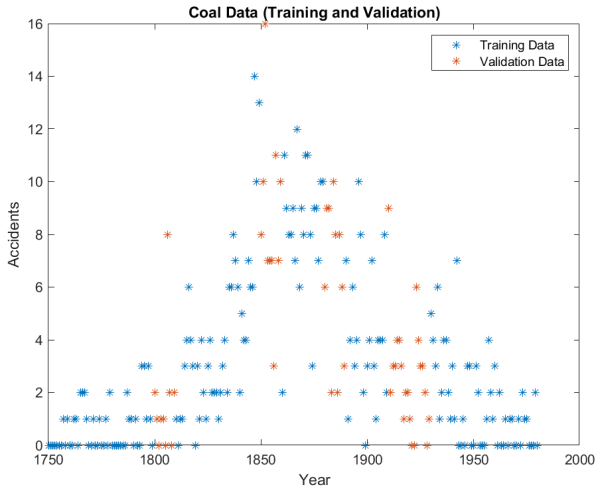

Figure 1: Coal Data (Training and Validation). Blue is training and orange is validation data.

For illustration of the sparsity of the C, G G2 and the precision matrix (in this case a SAR precision matrix), these have been plotted in figure 2. The plots show that the matrices have data only on the diagonal and centered around the diagonal (Q).

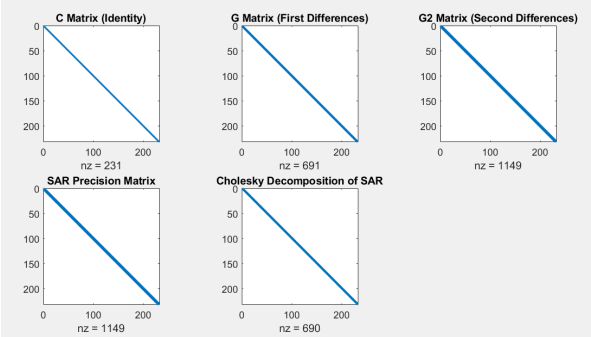

Figure 2: Sparsity Patterns of Precision Matrices

Both the CAR and SAR models were used to predict the number of accidents for the validation data and the confidence bounds were simulated using the method described in the theory section. Figure 3 and 4 shows the validation data (*) and the predicted data (o) with confidence bounds, equal to two standard deviations.

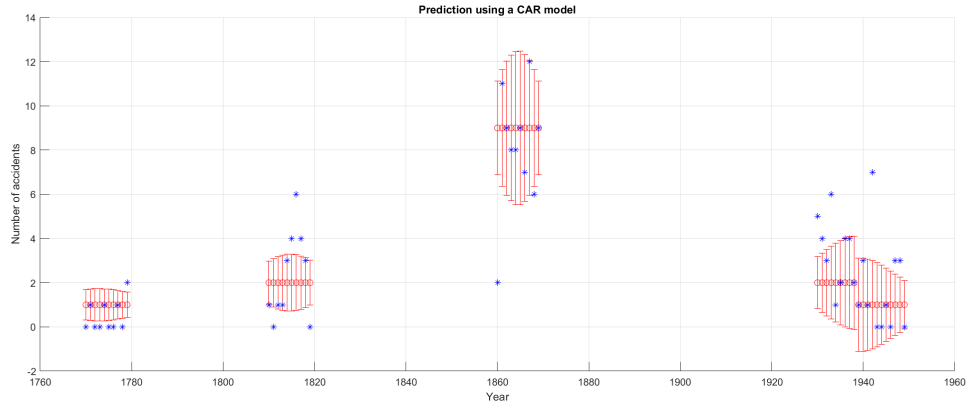

Figure 3: CAR model for prediction of accidents.

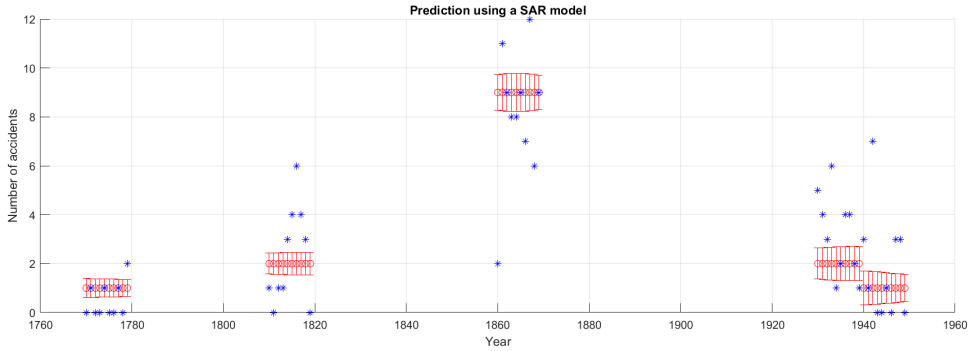

Figure 4: SAR model for prediction of accidents.

The CAR model performs overall better than the SAR model, since more of the validation data falls inside the confidence bounds, but both models performs in general poorly. The CAR model performs in general better for the central data. Different simulations were made with changing the initial conditions of x_{mode} or the initial guess for the θ values but this did not yield better results.

References

[VHH18] Jay M. Ver Hoef, Ephraim M. Hanks, and Mevin B. Hooten. On the relationship between conditional (car) and simultaneous (sar) autoregressive models. *Spatial Statistics*, 25:68–85, 2018.