derive posterior distribution

Renjie Wei

2022-05-02

The suggested Bayesian model

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \epsilon_i(t)$$

where $Y_i(t)$ the wind speed at time t (i.e. 6 hours earlier), $\Delta_{i,1}(t)$, $\Delta_{i,2}(t)$ and $\Delta_{i,3}(t)$ are the changes of latitude, longitude and wind speed between t and $t+6$, and $\epsilon_{i,t}$ follows a normal distributions with mean zero and variance σ^2, independent across t.

In the model, $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{5,i})$ are the random coefficients associated the ith hurricane, we assume that

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

follows a multivariate normal distributions with mean $\boldsymbol{\mu}$ and covariance matrix Σ.

We assume the following non-informative or weak prior distributions for σ^2, $\boldsymbol{\beta}$ and Σ.

$$P(\sigma^2) \propto \frac{1}{\sigma^2}; \quad P(\boldsymbol{\mu}) \propto 1; \quad P(\Sigma^{-1}) \propto |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\Sigma^{-1})$$

d is dimension of β.

Posterior Distributions

Let $\mathbf{B} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_n^\top)^\top$, derive the posterior distribution of the parameters $\Theta = (\mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2, \Sigma)$.

Let

$$\boldsymbol{X}_i(t)\boldsymbol{\beta}_i^\top = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t)$$

where $\boldsymbol{X}_i(t) = (1, Y_i(t), \Delta_{i,1}(t), \Delta_{i,2}(t), \Delta_{i,3}(t))$, $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}, \beta_{4,i})$

then, we can find that

$$Y_i(t+6) \sim N(\boldsymbol{X}_i(t)\boldsymbol{\beta}_i^\top, \sigma^2)$$

For i^{th} hurricane, there may be m_i times of record (excluding the first observation), let

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_i(t_0+6) \\ Y_i(t_1+6) \\ \vdots \\ Y_i(t_{m_i-1}+6) \end{pmatrix}_{m_i \times 1}$$

denotes the m_i-dimensional result vector for the i^{th} hurricane Therefore, since $Y_i(t)$'s are independent across t

$$\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}_i, \sigma^2 \sim N(\boldsymbol{X}_i\boldsymbol{\beta}_i^\top, \sigma^2 I)$$

where

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & Y_i(t_0) & \Delta_{i,1}(t_0) & \Delta_{i,2}(t_0) & \Delta_{i,3}(t_0) \\ 1 & Y_i(t_1) & \Delta_{i,1}(t_1) & \Delta_{i,2}(t_1) & \Delta_{i,3}(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Y_i(t_{m_i-1}) & \Delta_{i,1}(t_{m_i-1}) & \Delta_{i,2}(t_{m_i-1}) & \Delta_{i,3}(t_{m_i-1}) \end{pmatrix}_{m_i \times d}$$

and the pdf of $\boldsymbol{Y}_i$ is

$$\begin{aligned} f(\boldsymbol{Y}_i \mid \boldsymbol{\beta}_i, \sigma^2) =& \det(2\pi\sigma^2 I_{(m_i \times m_i)})^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\sigma^2 I_{(m_i \times m_i)})^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)\} \\ =& (2\pi\sigma^2)^{-m_i/2} \exp\big\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\sigma^2 I_{(m_i \times m_i)})^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)\big\} \end{aligned}$$

Since

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Therefore

$$\pi(\boldsymbol{\beta}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top)$$

Notice that $\mathbf{B} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_n^\top)^\top$, i.e.

$$\mathbf{B} = \begin{pmatrix} \beta_{0,1} & \beta_{1,1} & \beta_{2,1} & \beta_{3,1} & \beta_{4,1} \\ \beta_{0,2} & \beta_{1,2} & \beta_{2,2} & \beta_{3,2} & \beta_{4,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{0,n} & \beta_{1,n} & \beta_{2,n} & \beta_{3,n} & \beta_{4,n} \end{pmatrix}_{n \times d}$$

So, by using Bayesian rule, we have the following posterior distribution:

$$\pi(\mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2, \Sigma \mid Y) \propto \prod_{i=1}^{n} \Big\{ (2\pi\sigma^2)^{-m_i/2} \exp\big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)\big\}\Big\}$$
$$\times \prod_{i=1}^{n} \Big\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\big\{ -\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top\big\}\Big\} \times \frac{1}{\sigma^2} \times \det(\boldsymbol{\Sigma})^{-(d+1)} \exp\big\{ -\frac{1}{2}\boldsymbol{\Sigma}^{-1}\big\}$$

To apply MCMC, we need to find conditional posterior distribution of each parameter, then we can implement Gibbs sampling on these conditional posterior distributions.

1. Finding the posterior distribution of $\mathbf{B}$

Since finding the posterior distribution of $\mathbf{B}$ is the same to find the posterior distribution of $\boldsymbol{\beta_i}$, we try to derive the conditional distribution $\pi(\boldsymbol{\beta_i}|.)$

$$\begin{aligned}
\pi(\mathbf{B}|.) &\propto L_Y(\mathbf{B}^\top, \sigma^2) \times \pi(\mathbf{B} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \prod_{i=1}^{n} f(\boldsymbol{Y}_i \mid \boldsymbol{\beta}_i, \sigma^2) \prod_{i=1}^{n} \pi(\boldsymbol{\beta}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto \prod_{i=1}^{n} \Big\{ (2\pi\sigma^2)^{-m_i/2} \exp\Big(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)\Big)\Big\} \\
&\times \prod_{i=1}^{n} \Big\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top)\Big\} \\
&\propto \prod_{i=1}^{n} \exp\{-\frac{1}{2}\Big((\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top) + (\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top)\Big)\} \\
&= \exp\{-\frac{1}{2}\Big(\boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i + \boldsymbol{\beta}_i\boldsymbol{X}_i^\top(\sigma^2 I)^{-1}\boldsymbol{X}_i\boldsymbol{\beta}_i^\top - \boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{X}_i\boldsymbol{\beta}_i^\top \\
&- \boldsymbol{\beta}_i\boldsymbol{X}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i + \boldsymbol{\beta}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i^\top + \boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^\top - \boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i^\top - \boldsymbol{\beta}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^\top\Big)\} \\
&= \exp\{-\frac{1}{2}\Big(\boldsymbol{\beta}_i\boldsymbol{V}\boldsymbol{\beta}_i^\top - \boldsymbol{M}\boldsymbol{\beta}_i^\top - \boldsymbol{\beta}_i\boldsymbol{M}_i^\top + \boldsymbol{R}\Big)\}
\end{aligned}$$

where,

$$\begin{aligned}
\boldsymbol{V} &= \boldsymbol{\Sigma}^{-1} + \boldsymbol{X}_i^\top(\sigma^2 I)^{-1}\boldsymbol{X}_i \\
\boldsymbol{M} &= \boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{X}_i + \boldsymbol{\mu}\boldsymbol{\Sigma}^{-1} \\
\boldsymbol{R} &= \boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i + \boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^\top
\end{aligned}$$

re-writing the conditional posterior distribution, and ignoring some constant terms

$$\pi(\mathbf{B}|.) \propto \prod_{i=1}^{n} \exp\{(\boldsymbol{\beta}_i^\top - \boldsymbol{V}^{-1}\boldsymbol{M})^\top\boldsymbol{V}(\boldsymbol{\beta}_i^\top - \boldsymbol{V}^{-1}\boldsymbol{M})\}$$

Hence, each $\boldsymbol{\beta}_i$ has a conditional posterior multivariate normal distribution

$$\pi(\boldsymbol{\beta}_i|.) \sim N(\boldsymbol{V}^{-1}\boldsymbol{M}, \boldsymbol{V}^{-1})$$

2. Finding the posterior distribution of $\pi(\sigma^2|.)$

$$
\begin{aligned}
\pi(\sigma^2|.) &\propto L_Y(\mathbf{B}^\top, \sigma^2) \times \pi(\boldsymbol{\sigma^2}) \\
&\propto \frac{1}{\sigma^2} \prod_{i=1}^{n} \Big\{ (2\pi\sigma^2)^{-m_i/2} \exp\Big(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top (\sigma^2 I)^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)\Big)\Big\} \\
&\propto \frac{1}{\sigma^2}^{(\frac{\sum\limits_{i=1}^{n} m_i}{2}+1)} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)\}
\end{aligned}
$$

which follows the form of pdf of inverse gamma distribution

$$
f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{1}{x}\exp\{-\frac{\beta}{x}\}
$$

in this case, x is replaced by σ^2, α is replaced by $\frac{1}{2}\sum\limits_{i=1}^{n} m_i$, β is replaced by $\frac{1}{2}\sum\limits_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)$

i.e.

$$
\pi(\sigma^2|.) \sim IG(\frac{1}{2}\sum_{i=1}^{n} m_i, \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top)^\top(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i^\top))
$$

3. Finding the posterior distribution of $\pi(\boldsymbol{\Sigma}|.)$

$$
\begin{aligned}
\pi(\boldsymbol{\Sigma}|.) \propto& \pi(\mathbf{B} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma}^{-1}) \\
\propto& \prod_{i=1}^{n} \Big\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top) \Big\} |\Sigma|^{-(d+1)} \exp(-\frac{1}{2}\Sigma^{-1}) \\
\propto& |\Sigma|^{-(n+d+1+d+1)/2} \exp\{-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top - \frac{1}{2}\Sigma^{-1}\} \\
\propto& |\Sigma|^{-(n+d+1+d+1)/2} \exp\{-\frac{1}{2}tr(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\}
\end{aligned}
$$

where

$$
\boldsymbol{S} = \boldsymbol{I} + \sum_{i=1}^{n}(\boldsymbol{\beta}_i - \boldsymbol{\mu})(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top
$$

which is the form of pdf of the inverse wishart distribution Inverse Wishart$(\mathbf{V}, \boldsymbol{S})$, where $\mathbf{V} = n + d + 1$, i.e.

$$
\pi(\boldsymbol{\Sigma}|.) \sim IW(n + d + 1,\ \boldsymbol{I} + \sum_{i=1}^{n}(\boldsymbol{\beta}_i - \boldsymbol{\mu})(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top)
$$

4. Finding the posterior distribution of $\pi(\boldsymbol{\mu}|.)$

$$
\begin{aligned}
\pi(\boldsymbol{\mu}|.) \propto & \pi(\mathbf{B} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}) \\
= & \prod_{i=1}^{n} \Big\{ \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top) \Big\} \\
\propto & \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\beta}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu})^\top\} \\
\propto & \exp\{-\frac{1}{2}\Big(\sum_{i=1}^{n}\boldsymbol{\beta}_i\ \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i^\top + n\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^\top - 2\sum_{i=1}^{n}\boldsymbol{\beta}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^\top\Big)\} \\
= & \exp\{-\frac{1}{2}\Big(\boldsymbol{\mu}\boldsymbol{V'}\boldsymbol{\mu}^\top - 2\boldsymbol{M'}\boldsymbol{\mu}^\top + \boldsymbol{R'}\Big)\} \\
\propto & \exp\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{V'}^{-1}\boldsymbol{M'})\boldsymbol{V'}(\boldsymbol{\mu} - \boldsymbol{V'}^{-1}\boldsymbol{M'})^\top\}
\end{aligned}
$$

where

$$
\boldsymbol{V'} = n\boldsymbol{\Sigma}^{-1},\ \boldsymbol{M'} = \sum_{i=1}^{n}\boldsymbol{\beta}_i\boldsymbol{\Sigma}^{-1},\ \boldsymbol{R'} = \sum_{i=1}^{n}\boldsymbol{\beta}_i\ \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i^\top
$$

Hence

$$
\pi(\boldsymbol{\Sigma}|.) \sim N(\boldsymbol{V'}^{-1}\boldsymbol{M'}, \boldsymbol{V'}^{-1})
$$

Markov Chain Monte Carlo

Because our hierarchical Bayesian Model exploited non-informative priors for four parameters, the Gibbs Sampling method would be implemented, updating parameters in the following order from their conditional posteriors distributions, $\boldsymbol{B}$, σ^2, $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$.