Project 3 MCMC

Derive posterior distribution

Yi Huang

2023-04-26

The suggested Bayesian model

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

$Y_i(t)$ 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 $t+6$,

$\mathbf{X}_i = (x_{i,1}, x_{i,2}, x_{i,3})$ are covariates with fixed effect γ, where $x_{i,1}$ be the month of year when the i-th hurricane started, $x_{i,2}$ be the calendar year of the i hurricane, and $x_{i,3}$ be the type of the i-th hurricane.

and $\epsilon_{i,t}$ follows a normal distributions with mean zero and variance σ^2, independent across t.

In the model, $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{4,i})$ 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 $\boldsymbol{\beta}$ and covariance matrix Σ.

Bayesian Model

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

and $\epsilon_{i,t}$ independent across t,

$$\epsilon_{i,t} \sim N(0, \sigma^2)$$

in the model,

$$\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{4,i})$$

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 $\boldsymbol{\beta}$ and covariance matrix Σ.

Prior distributions

1. $\boldsymbol{\mu}$: a normal distribution with mean vector $\boldsymbol{0}$ and variance-covariance matrix $\boldsymbol{V}$, reflecting the prior knowledge that the mean coefficients should be centered around zero but allowing for some variability across hurricanes. The variance-covariance matrix $\boldsymbol{V}$ can be set to a diagonal matrix with large variances on the diagonal and small covariances off-diagonal, reflecting the prior knowledge that the coefficients may have some correlation but are largely independent across hurricanes.

$$P(\boldsymbol{\mu}) = \frac{1}{\sqrt{2\pi}|\boldsymbol{V}|}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{V}^{-1}\boldsymbol{\mu}\} \propto |\boldsymbol{V}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{V}^{-1}\boldsymbol{\mu}\} \propto 1 \tag{2}$$

where $|\boldsymbol{V}|$ is the deteriminant of $\boldsymbol{V}$

2. $\boldsymbol{\Sigma}$: an inverse-Wishart distribution with degrees of freedom ν and scale matrix $\boldsymbol{S}$, reflecting the prior knowledge that the covariance matrix of the coefficients should be positive definite and have some structure. The degrees of freedom ν can be set to a small value (e.g., 5) to reflect a relatively weak prior, while the scale matrix $\boldsymbol{S}$ can be set to a diagonal matrix with large variances on the diagonal and small covariances off-diagonal, reflecting the prior knowledge that the covariance matrix should be diagonal or nearly diagonal.

$$P(\Sigma^{-1}) \propto |\Sigma|^{-\frac{(\nu+d+1)}{2}}\exp(-\frac{1}{2}tr(S\Sigma^{-1})) \tag{3}$$

where d is dimension of scale matrix $\boldsymbol{S}$, $\nu = 5$

3. All the fixed effects $\gamma \sim N(0, 0.05^2)$

$$P(\gamma) \propto exp(-\frac{\gamma^2}{2*(0.05)^2}) = e^{-200\gamma^2} \tag{4}$$

4. σ: a half-Cauchy distribution with scale parameter 10, reflecting the prior knowledge that the residual variance should be positive and large enough to account for any unexplained variability in the wind speed data.

$$\sigma^2 \sim HC(\alpha = 10)$$
$$P(\sigma^2) = \frac{2\alpha}{\pi + \alpha^2} \propto \frac{1}{\sigma^2 + \alpha^2} \tag{5}$$

Posterior Distributions

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

Let

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

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

and $\boldsymbol{X}_i = (x_{i,1}, x_{i,2}, x_{i,3})$ are covariates with fixed effect γ

- $x_{i,1}$ be the month of year when the i-th hurricane started
- $x_{i,2}$ be the calendar year of the i hurricane
- $x_{i,3}$ be the type of the i-th hurricane

and

$\epsilon_{i,t} \sim N(0, \sigma^2)$ independent across t

then, we can find that

$$\boldsymbol{Y_i} \sim MVN(\boldsymbol{Z}_i\boldsymbol{\beta}_i, \sigma^2 I) \tag{6}$$

where $\boldsymbol{Z_i}$ is the $n_i \times d$ covariate matrix for hurriance

and the likelihood for our data is

$$\begin{aligned} L(\boldsymbol{Y}_i \mid \boldsymbol{B}_i, \sigma^2) =& \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{\beta}_i - \boldsymbol{X_i}\boldsymbol{\gamma}_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{\beta}_i - \boldsymbol{X_i}\boldsymbol{\gamma}_i)\} \\ \propto& (\sigma^{-\frac{N}{2}}) \prod_{i=1}^{n} \exp\Big\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\beta}_i - \boldsymbol{X_i}\boldsymbol{\gamma}_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\beta}_i - \boldsymbol{X_i}\boldsymbol{\gamma}_i)\Big\} \end{aligned} \tag{7}$$

Prior

The prior for $\boldsymbol{B}$ is, $A = \Sigma^{-1}$

$$\begin{aligned} P(\boldsymbol{B}|\mu, \Sigma^{-1}) \propto& \prod_{i=1}^{n} |\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\sum_{i}^{n}(\beta_i - \mu)^\top \Sigma^{-1}(\beta_i - \mu)\} \\ =& |A|^{-\frac{N}{2}} \exp\{-\frac{1}{2}\sum_{i}^{n}(\beta_i - \mu)^\top A(\beta_i - \mu)\} \end{aligned} \tag{8}$$

$$P(\sigma^2) \propto \frac{1}{\sigma^2 + \alpha^2} = \frac{1}{\sigma^2 + 10^2}$$

Joint Posrerior

$$\begin{aligned}
P(\Theta|Y) =&P(B,\mu,\sigma^2,A,\gamma|Y)\\
&\propto L(Y|B,\sigma^2)p(\sigma^2)p(\mu)p(A)p(\gamma)\\
&\propto \frac{1}{\sigma^{\frac{N}{2}}(\sigma^2+10^2)^N}\prod_{i=1}^{n}\exp\Big\{-\frac{1}{2}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\beta}_i-\boldsymbol{X_i}\boldsymbol{\gamma}_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\beta}_i-\boldsymbol{X_i}\boldsymbol{\gamma}_i)\Big\} \qquad (9)\\
&\times\prod_{i=1}^{n}\exp\{-\frac{1}{2}\sum_{i}^{n}(\beta_i-\mu)^\top A(\beta_i-\mu)\}|A|^{\frac{N+d+6}{2}}\exp\{-\frac{1}{2}tr(SA)\}
\end{aligned}$$

MCMC Algorithm

Implement Gibbs sampler to generate samples from $f(\Theta)$, $\Theta=(B,\mu,\sigma^2,A,\gamma)$ through the posterior distribution $P(B,\mu,\sigma^2,A,\gamma|Y)$.

1. Generate B_i from $f(B_t|\boldsymbol{\mu}_{t-1},\sigma^2_{t-1},A_{t-1},\gamma_{t-1})$

$$\begin{aligned}
f(\mathbf{B}|\boldsymbol{\mu},\sigma^2,A,\gamma) &\propto L_Y(\mathbf{B},\sigma^2)\times P(\mathbf{B}\mid\boldsymbol{\mu},\boldsymbol{\Sigma})\\
&\propto\prod_{i=1}^{n}\exp\Big\{-\frac{1}{2}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\beta}_i-\boldsymbol{X_i}\boldsymbol{\gamma}_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\beta}_i-\boldsymbol{X_i}\boldsymbol{\gamma}_i)\Big\}\exp\{-\frac{1}{2}(\beta_i-\mu)^\top A(\beta_i-\mu)\}\\
&\propto\prod_{i=1}^{n}\exp\Big\{-\frac{1}{2}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\beta}_i-\boldsymbol{X_i}\boldsymbol{\gamma}_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\boldsymbol{\beta}_i-\boldsymbol{X_i}\boldsymbol{\gamma}_i)\Big\}+(\beta_i-\mu)^\top A(\beta_i-\mu)\}\\
&\propto\prod_{i=1}^{n}\exp\{-\frac{1}{2}\Big(\boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i-\boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Z}_i\boldsymbol{\beta}_i^\top-\boldsymbol{Y}_i^\top(\sigma^2 I)^{-1}\boldsymbol{X}_i\gamma-\boldsymbol{\beta}_i\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i\\
&+\boldsymbol{\beta}_i\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Z}_i\boldsymbol{\beta}_i^\top+\boldsymbol{\beta}_i\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{X_i}\gamma-\gamma\boldsymbol{X_i}^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i+\gamma\boldsymbol{X}_i(\sigma^2 I)^{-1}\boldsymbol{Z}_i\beta_i^\top\\
&+\gamma\boldsymbol{X}_i^\top(\sigma I)^{-1}\boldsymbol{Z}_i^\top\beta_i^\top+\boldsymbol{\beta}_i^\top\boldsymbol{A}\boldsymbol{\beta}_i-\boldsymbol{\beta}_i^\top\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{\mu}^\top\boldsymbol{A}\boldsymbol{\mu}-\boldsymbol{\mu}\boldsymbol{A}\boldsymbol{\beta}_i^\top\}\\
&\propto\prod_{i=1}^{n}\exp\Big\{-\frac{1}{2}(\boldsymbol{\beta}_i(\boldsymbol{A}+\boldsymbol{Z}_i^\top(\sigma I)^{-1}Z_i)\boldsymbol{\beta}_i^\top-\gamma^2(\boldsymbol{X}_i^\top)-2(\boldsymbol{Y}_i^\top(\sigma I)^{-1}\boldsymbol{Z}_i+\gamma\boldsymbol{X}_i^\top(\sigma I)^{-1}\boldsymbol{Z}_i+\boldsymbol{\mu}\boldsymbol{A})\boldsymbol{\beta}_i^\top\\
&-2(\boldsymbol{Y}_i^\top(\sigma I)^{-1}\boldsymbol{X}_i)\gamma\}\\
&\propto\prod_{i=1}^{n}\exp\Big\{-\frac{1}{2}[\boldsymbol{\beta}_i-(A+\boldsymbol{Z}_i(\sigma I)^{-1}\boldsymbol{Z}_i)^{-1}(\boldsymbol{Y}_i(\sigma I)^{-1}\boldsymbol{Z}_i+\gamma\boldsymbol{X}_i(\sigma I)^{-1}\boldsymbol{Z}_i+\boldsymbol{\mu}_i^\top\boldsymbol{A})^\top]^\top(A+\boldsymbol{Z}_i(\sigma I)^{-1}\boldsymbol{Z}_i)\\
&[\boldsymbol{\beta}_i-(A+\boldsymbol{Z}_i(\sigma I)^{-1}\boldsymbol{Z}_i)^{-1}(\boldsymbol{Y}_i(\sigma I)^{-1}\boldsymbol{Z}_i+\gamma\boldsymbol{X}_i(\sigma I)^{-1}\boldsymbol{Z}_i+\boldsymbol{\mu}_i^\top\boldsymbol{A})^\top]\Big\}
\end{aligned}$$

2. Generate μ_t from $f(\mu_t|B_t,\sigma^2_{t-1},A_{t-1},\gamma_{t-1})$
3. Generate σ^2_t from $f(\sigma_t|B_t,\mu_t,A_{t-1},\gamma_{t-1})$
4. Generate A_t from $f(A_t|B_t,\mu_t,\sigma^2_t,\gamma_{t-1})$
5. Generate γ from $f(\gamma|B_t,\mu_t,\sigma^2_t,A_t)$