EPG3343 - Seminario de Estadística III
Clase 7

Jonathan Acosta

Pontificia Universidad Católica de Chile

Segundo Semestre, 2022

Esquema

Esquema

1. Modelos para Datos de Área
   - Modelo CAR
   - Modelo SAR

Modelos para Datos de Área

- Con los datos regionales (área), una especificación directa de la matriz de varianza $\boldsymbol{\Sigma}$ limita la medida de proximidad espacial a las distancias entre las ubicaciones de los puntos que se supone representan cada región (por ejemplo, las distancias intercentroides).

Modelos para Datos de Área

- Con los datos regionales (área), una especificación directa de la matriz de varianza $\mathbf{\Sigma}$ limita la medida de proximidad espacial a las distancias entre las ubicaciones de los puntos que se supone representan cada región (por ejemplo, las distancias intercentroides).

- En vez de utilizar un modelo geoestadístico para describir la dependencia espacial, es más apropiado, definir dicha dependencia a través de modelos de regresión espacial que pueden incorporar las estructuras de vecindad.

Modelos para Datos de Área

- Con los datos regionales (área), una especificación directa de la matriz de varianza $\boldsymbol{\Sigma}$ limita la medida de proximidad espacial a las distancias entre las ubicaciones de los puntos que se supone representan cada región (por ejemplo, las distancias intercentroides).
- En vez de utilizar un modelo geoestadístico para describir la dependencia espacial, es más apropiado, definir dicha dependencia a través de modelos de regresión espacial que pueden incorporar las estructuras de vecindad.
- El análogo espacial (con respecto a series de tiempo) representa los datos en ubicaciones como una combinación lineal de valores vecinos.

Modelos para Datos de Área

- Con los datos regionales (área), una especificación directa de la matriz de varianza $\boldsymbol{\Sigma}$ limita la medida de proximidad espacial a las distancias entre las ubicaciones de los puntos que se supone representan cada región (por ejemplo, las distancias intercentroides).
- En vez de utilizar un modelo geoestadístico para describir la dependencia espacial, es más apropiado, definir dicha dependencia a través de modelos de regresión espacial que pueden incorporar las estructuras de vecindad.
- El análogo espacial (con respecto a series de tiempo) representa los datos en ubicaciones como una combinación lineal de valores vecinos.
- Esta autorregresión induce la dependencia espacial de los datos.

- Ahora bien no existe una única forma de inducir la dependencia espacial a través de los vecinos, en esta sección veremos los métodos más comunes.

- Ahora bien no existe una única forma de inducir la dependencia espacial a través de los vecinos, en esta sección veremos los métodos más comunes.

- Una alternativa es un modelo tipo regresión donde

$$Y_j = f(Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n) + \epsilon_j, \quad j = 1, \ldots, n$$

- Ahora bien no existe una única forma de inducir la dependencia espacial a través de los vecinos, en esta sección veremos los métodos más comunes.

- Una alternativa es un modelo tipo regresión donde

$$Y_j = f(Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n) + \epsilon_j, \quad j = 1, \ldots, n$$

- Recordemos, que la matriz de contiguidad $\boldsymbol{W}$ es 1 si los sitios están conectados y vale 0 si no lo están.

Modelos para Datos de Área

- Ahora bien no existe una única forma de inducir la dependencia espacial a través de los vecinos, en esta sección veremos los métodos más comunes.
- Una alternativa es un modelo tipo regresión donde

$$Y_j = f(Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n) + \epsilon_j, \quad j = 1, \ldots, n$$

- Recordemos, que la matriz de contiguidad $\boldsymbol{W}$ es 1 si los sitios están conectados y vale 0 si no lo están.
- Así el término $\boldsymbol{WY}$ se denomina lag-espacial.

- Por ejemplo, para $n = 4$, suponga que

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{WY} = \begin{pmatrix} Y_2 \\ Y_1 + Y_3 + Y_4 \\ Y_2 + Y_4 \\ Y_2 + Y_3 \end{pmatrix}$$

Un posible modelo corresponde a

$$\boldsymbol{Y} = \rho \boldsymbol{WY} + \boldsymbol{\epsilon}$$

donde ρ es un parámetro adecuado.

Modelos para Datos de Área

- Por ejemplo, para $n = 4$, suponga que

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{W}\boldsymbol{Y} = \begin{pmatrix} Y_2 \\ Y_1 + Y_3 + Y_4 \\ Y_2 + Y_4 \\ Y_2 + Y_3 \end{pmatrix}$$

  Un posible modelo corresponde a

$$\boldsymbol{Y} = \rho \boldsymbol{W}\boldsymbol{Y} + \boldsymbol{\epsilon}$$

  donde ρ es un parámetro adecuado.

- Notar que Y_1 es solo explicado por Y_2, mientras que Y_2 es explicado por Y_1, Y_3, e Y_4. Para evitar esto se propone estandarizar la matriz $\boldsymbol{W}$.

- **Estandarización por fila**: Consiste en dividir cada elemento por la suma de la fila a la que pertenece, es decir,

$$\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$$

- **Estandarización por fila**: Consiste en dividir cada elemento por la suma de la fila a la que pertenece, es decir,

$$\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$$

- **Doble Estandarización**: Consiste en dividir cada elemento por la suma de todos los elementos, es decir,

$$\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}$$

Modelos para Datos de Área

- **Estandarización por fila**: Consiste en dividir cada elemento por la suma de la fila a la que pertenece, es decir,

$$\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$$

- **Doble Estandarización**: Consiste en dividir cada elemento por la suma de todos los elementos, es decir,

$$\widetilde{w}_{ij} = \frac{w_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}$$

- **Estabilización de varianza**: Consiste en dos etapas

$$\text{1st:} \quad w_{ij}^{*} = \frac{w_{ij}}{\sqrt{\sum_{j=1}^{n} w_{ij}^{2}}}; \qquad \text{2nd:} \quad \widetilde{w}_{ij} = \frac{n w_{ij}^{*}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{*}}$$

Surge en el contexto del Índice de Moran para datos Heterocedásticos.

Modelos para Datos de Área

Obs.: En R tradicionalmente se denota por B, W, C, y S para la matriz de pesos Binaria, estandarizada por fila, global y estabilización de varianza, respectivamente.

Obs.: En R tradicionalmente se denota por B, W, C, y S para la matriz de pesos Binaria, estandarizada por fila, global y estabilización de varianza, respectivamente.

Con esta nueva notación y la conexión de vecinos del ejemplo anterior

$$\boldsymbol{B} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

- **Estandarización por fila**:

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{WY} = \begin{pmatrix} Y_2 \\ \dfrac{Y_1 + Y_3 + Y_4}{3} \\ \dfrac{Y_2 + Y_4}{2} \\ \dfrac{Y_2 + Y_3}{2} \end{pmatrix}$$

Modelos para Datos de Área

- **Doble Estandarización**: Consiste en dividir cada elemento por la suma de todos los elementos, es decir,

$$\boldsymbol{C} = \begin{pmatrix} 0 & 1/8 & 0 & 0 \\ 1/8 & 0 & 1/8 & 1/8 \\ 0 & 1/8 & 0 & 1/8 \\ 0 & 1/8 & 1/8 & 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{CY} = \begin{pmatrix} \dfrac{Y_2}{8} \\ \dfrac{Y_1 + Y_3 + Y_4}{8} \\ \dfrac{Y_2 + Y_4}{8} \\ \dfrac{Y_2 + Y_3}{8} \end{pmatrix}$$

Modelos para Datos de Área

- **Doble Estandarización**: Consiste en dividir cada elemento por la suma de todos los elementos, es decir,

$$\boldsymbol{C} = \begin{pmatrix} 0 & 1/8 & 0 & 0 \\ 1/8 & 0 & 1/8 & 1/8 \\ 0 & 1/8 & 0 & 1/8 \\ 0 & 1/8 & 1/8 & 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{CY} = \begin{pmatrix} \dfrac{Y_2}{8} \\ \dfrac{Y_1 + Y_3 + Y_4}{8} \\ \dfrac{Y_2 + Y_4}{8} \\ \dfrac{Y_2 + Y_3}{8} \end{pmatrix}$$

- **Estabilización de varianza**: Consiste en dos etapas

$$\boldsymbol{S} = \frac{8}{\sqrt{3} + 2\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/\sqrt{3} & 0 & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}$$

- Para definir un modelo estadístico asociado a este tipo de dependencia se debe tener presente que no basta con conocer información (densidad) marginal de cada Y_i.

Modelos para Datos de Área

- Para definir un modelo estadístico asociado a este tipo de dependencia se debe tener presente que no basta con conocer información (densidad) marginal de cada Y_i.

- Para recuperar la densidad conjunta es imprescindible contar con la información de las densidades marginales

Modelos para Datos de Área

- Para definir un modelo estadístico asociado a este tipo de dependencia se debe tener presente que no basta con conocer información (densidad) marginal de cada Y_i.

- Para recuperar la densidad conjunta es imprescindible contar con la información de las densidades marginales

- La información de la condicionales debe ser consistente, por ejemplo, si $n = 2$, suponga que

$$Y_1 \mid Y_2 = y_2 \sim \mathcal{N}(\beta_0 + \beta_1 y_2, \sigma^2) \quad \wedge \quad Y_2 \mid Y_1 = y_1 \sim \mathcal{N}(\alpha_0 + \alpha_1 y_1^2, \sigma^2)$$

Entonces

$$\begin{aligned}
\mathbb{E}[Y_1] &= \mathbb{E}[\mathbb{E}[Y_1 \mid Y_2]] = \mathbb{E}[\beta_0 + \beta_1 Y_2] = \beta_0 + \beta_1 \mathbb{E}[Y_2] \\
\mathbb{E}[Y_2] &= \mathbb{E}[\mathbb{E}[Y_2 \mid Y_1]] = \mathbb{E}[\alpha_0 + \alpha_1 Y_1^2] = \alpha_0 + \alpha_1 \mathbb{E}[Y_1^2]
\end{aligned}$$

La cual solo tienen solución trivial, significando que las medias son incompatible.

Lemma (de Brook)

La densidad conjunta se puede obtener a partir de las densidades marginales mediante

$$\begin{aligned} p(y_1,\ldots,y_n) &= \frac{p(y_1 \mid y_2,\ldots,y_n)}{p(y_{10} \mid y_2,\ldots,y_n)} \cdot \frac{p(y_2 \mid y_{10},y_3,\ldots,y_n)}{p(y_{20} \mid y_{10},y_3,\ldots,y_n)} \cdot \\ & \quad \cdots \frac{p(y_n \mid y_{10},\ldots,y_{n-1,0})}{p(y_{n0} \mid y_{10},\ldots,y_{n-1,0})} p(y_{10},\ldots,y_{n0}) \end{aligned}$$

donde $\boldsymbol{y}_0 = (y_{10},\ldots,y_{n0})$ *es un punto fijo en el soporte de* $p(y_1,\ldots,y_n)$

Modelo CAR

Definición (Conditional Autorregressive Models (Caso Gaussiano))

Suponga que

$$Y_i|y_j \overset{j\neq i}{\sim} \mathcal{N}\left(\sum_j b_{ij}y_j, \tau_i^2\right), \quad i = 1, \ldots, n.$$

Estas condicionales completas son compartibles, así por el lema de Brook se tiene

$$p(y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{D}^{-1}(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{y}\right\},$$

donde $\boldsymbol{B} = (b_{ij})$ y $\boldsymbol{D} = \text{diag}(\tau_1^2, \ldots, \tau_n^2)$

Modelo CAR

Definición (Conditional Autorregressive Models (Caso Gaussiano))

Suponga que

$$Y_i|y_j \overset{j\neq i}{\sim} \mathcal{N}\left(\sum_j b_{ij}y_j, \tau_i^2\right), \quad i = 1, \ldots, n.$$

Estas condicionales completas son compartibles, así por el lema de Brook se tiene

$$p(y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{D}^{-1}(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{y}\right\},$$

donde $\boldsymbol{B} = (b_{ij})$ y $\boldsymbol{D} = \text{diag}(\tau_1^2, \ldots, \tau_n^2)$

Pregunta: ¿Tiene $\boldsymbol{Y}$ una distribución normal multivariada?

Modelo CAR

- Si $\boldsymbol{\Sigma}_y = (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{D}$ fuese la varianza de $\boldsymbol{Y}$ entonces $\boldsymbol{\Sigma}_y^\top = \boldsymbol{\Sigma}_y$, obteniendo las condiciones
$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2} \quad \forall\ i, j \tag{1}$$
Claramente $\boldsymbol{B}$ no necesita ser simétrica.

Modelo CAR

- Si $\boldsymbol{\Sigma}_y = (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{D}$ fuese la varianza de $\boldsymbol{Y}$ entonces $\boldsymbol{\Sigma}_y^\top = \boldsymbol{\Sigma}_y$, obteniendo las condiciones
$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2} \quad \forall\ i, j \tag{1}$$
Claramente $\boldsymbol{B}$ no necesita ser simétrica.

- Volviendo a la matriz de proximidad $\boldsymbol{W}$ (asumida simétrica), considere
$$b_{ij} = \frac{w_{ij}}{w_{i+}} \quad \wedge \quad \tau_i^2 = \frac{\tau^2}{w_{i+}}.$$
Sea $D_w = \text{diag}(w_{1+}, \ldots, w_{n+})$, entonces
$$\begin{aligned} Y_i|y_j \quad &\overset{j\neq i}{\sim} \quad \mathcal{N}\left(\frac{1}{w_{i+}}\sum_j w_{ij}y_j, \frac{\tau^2}{w_{i+1}}\right), \\ p(y_1, \ldots, y_n) \quad &\propto \quad \exp\left\{-\frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{D}_w - \boldsymbol{W})\boldsymbol{y}\right\}. \end{aligned}$$

Modelo CAR

- Notar que la ultima expresión puede ser escrita como

$$p(y_1,\ldots,y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i\neq j} w_{ij}(y_i - y_j)^2\right\}$$

Modelo CAR

- Notar que la ultima expresión puede ser escrita como

$$p(y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i \neq j} w_{ij}(y_i - y_j)^2\right\}$$

- Sin embargo, como $(\boldsymbol{D}_w - \boldsymbol{W})\mathbf{1} = 0$ entonces $\boldsymbol{\Sigma}_y^{-1}$ es singular.

Modelo CAR

- Notar que la ultima expresión puede ser escrita como

$$p(y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i\neq j} w_{ij}(y_i - y_j)^2\right\}$$

- Sin embargo, como $(\boldsymbol{D}_w - \boldsymbol{W})\mathbf{1} = 0$ entonces $\boldsymbol{\Sigma}_y^{-1}$ es singular.
- Esto significa que $\boldsymbol{\Sigma}$ no existe por lo que $p(y_1, \ldots, y_n)$ es impropia.

Modelo CAR

- Notar que la ultima expresión puede ser escrita como

$$p(y_1, \ldots, y_n) \propto \exp\left\{ -\frac{1}{2\tau^2} \sum_{i \neq j} w_{ij}(y_i - y_j)^2 \right\}$$

- Sin embargo, como $(\boldsymbol{D}_w - \boldsymbol{W})\mathbf{1} = 0$ entonces $\boldsymbol{\Sigma}_y^{-1}$ es singular.
- Esto significa que $\boldsymbol{\Sigma}$ no existe por lo que $p(y_1, \ldots, y_n)$ es impropia.
- Este modelo en realidad es conocido como IAR (intrinsically auto-regresive)

Modelo CAR

- Notar que la ultima expresión puede ser escrita como

$$p(y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i\neq j} w_{ij}(y_i - y_j)^2\right\}$$

- Sin embargo, como $(\boldsymbol{D}_w - \boldsymbol{W})\mathbf{1} = 0$ entonces $\boldsymbol{\Sigma}_y^{-1}$ es singular.
- Esto significa que $\boldsymbol{\Sigma}$ no existe por lo que $p(y_1, \ldots, y_n)$ es impropia.
- Este modelo en realidad es conocido como IAR (intrinsically autoregresive)
- La impropiedad puede ser removida incluyendo un parámetro adicional, ρ y redefiniendo

$$\boldsymbol{\Sigma}_y^{-1} = \boldsymbol{D}_w - \rho\boldsymbol{W}$$

Modelo CAR

- Para garantizar que $\boldsymbol{\Sigma}_y^{-1}$ sea no singular se tiene que

$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$

donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{D}_w^{-1/2} \boldsymbol{W} \boldsymbol{D}_w^{-1/2}$.

Modelo CAR

- Para garantizar que $\boldsymbol{\Sigma}_y^{-1}$ sea no singular se tiene que

$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$

donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{D}_w^{-1/2} \boldsymbol{W} \boldsymbol{D}_w^{-1/2}$.

- Reemplazando la matriz de adyacencia $\boldsymbol{W}$ por la matriz estandarizada por filas, denotada por, $\widetilde{W}$, entonces $\boldsymbol{\Sigma}_y^{-1}$ puede ser escrito como

$$\boldsymbol{\Sigma}_y^{-1} = \boldsymbol{M}^{-1}(\boldsymbol{I} - \alpha \widetilde{\boldsymbol{W}})$$

Así $\boldsymbol{\Sigma}_y^{-1}$ es invertible si $|\alpha| < 1$ teniendo un comportamiento similar a una correlación.

Modelo CAR

- Para garantizar que $\boldsymbol{\Sigma}_y^{-1}$ sea no singular se tiene que

$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$

donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{D}_w^{-1/2} \boldsymbol{W} \boldsymbol{D}_w^{-1/2}$.

- Reemplazando la matriz de adyacencia $\boldsymbol{W}$ por la matriz estandarizada por filas, denotada por, $\widetilde{W}$, entonces $\boldsymbol{\Sigma}_y^{-1}$ puede ser escrito como

$$\boldsymbol{\Sigma}_y^{-1} = \boldsymbol{M}^{-1}(\boldsymbol{I} - \alpha \widetilde{\boldsymbol{W}})$$

Así $\boldsymbol{\Sigma}_y^{-1}$ es invertible si $|\alpha| < 1$ teniendo un comportamiento similar a una correlación.

- Este último aspecto se aborda mejor en el modelo SAR.

Modelo CAR

- Para garantizar que $\mathbf{\Sigma}_y^{-1}$ sea no singular se tiene que

$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$

donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{D}_w^{-1/2} \boldsymbol{W} \boldsymbol{D}_w^{-1/2}$.

- Reemplazando la matriz de adyacencia $\boldsymbol{W}$ por la matriz estandarizada por filas, denotada por, $\widetilde{W}$, entonces $\mathbf{\Sigma}_y^{-1}$ puede ser escrito como

$$\mathbf{\Sigma}_y^{-1} = \boldsymbol{M}^{-1}(\boldsymbol{I} - \alpha \widetilde{\boldsymbol{W}})$$

Así $\mathbf{\Sigma}_y^{-1}$ es invertible si $|\alpha| < 1$ teniendo un comportamiento similar a una correlación.

- Este último aspecto se aborda mejor en el modelo SAR.

- El modelo inicial puede ser escrito como

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Y} + \boldsymbol{\epsilon} \quad \equiv \quad (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{Y} = \boldsymbol{\epsilon}$$

Modelo CAR

- Si $p(\boldsymbol{y})$ es propia, entonces $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{D}$ de donde

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B})^{\top}.$$

Modelo CAR

- Si $p(\boldsymbol{y})$ es propia, entonces $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{D}$ de donde

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B})^{\top}.$$

- Es decir, $\boldsymbol{Y}$ induce una distribución de los $\boldsymbol{\epsilon}$.

Modelo CAR

- Si $p(\boldsymbol{y})$ es propia, entonces $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{D}$ de donde

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B})^{\top}.$$

- Es decir, $\boldsymbol{Y}$ induce una distribución de los $\boldsymbol{\epsilon}$.

- Notar que los componentes de $\boldsymbol{\epsilon}$ no son independientes y $\mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{Y}) = \boldsymbol{D}$.

Modelo CAR

- Si $p(\boldsymbol{y})$ es propia, entonces $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{D}$ de donde

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}(\boldsymbol{I}-\boldsymbol{B})^{\top}.$$

- Es decir, $\boldsymbol{Y}$ induce una distribución de los $\boldsymbol{\epsilon}$.
- Notar que los componentes de $\boldsymbol{\epsilon}$ no son independientes y $\mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{Y}) = \boldsymbol{D}$.
- El modelo SAR resuelve esta situación, asumiendo una distribución en el error y a partir de esta se induce una distribución a $\boldsymbol{Y}$.

Modelo CAR

- Si $p(\boldsymbol{y})$ es propia, entonces $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{D}$ de donde

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{B})^{\top}.$$

- Es decir, $\boldsymbol{Y}$ induce una distribución de los $\boldsymbol{\epsilon}$.

- Notar que los componentes de $\boldsymbol{\epsilon}$ no son independientes y $\operatorname{cov}(\boldsymbol{\epsilon}, \boldsymbol{Y}) = \boldsymbol{D}$.

- El modelo SAR resuelve esta situación, asumiendo una distribución en el error y a partir de esta se induce una distribución a $\boldsymbol{Y}$.

- Finalmente, es posible incluir un componente de regresión de la forma $\boldsymbol{x}_i^{\top}\boldsymbol{\beta}$ de la forma:

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Y} + \boldsymbol{X}\beta + \boldsymbol{\epsilon}$$

  Sin embargo, veremos el modelo CAR solo como una distribución para el efecto aleatorio espacial.

Modelo SAR

Definición (Simultaneous Autorregressive Models)

Suponga que

$$Y_i = \sum_j b_{ij} y_j + \epsilon_i, \quad i = 1, \ldots, n, \quad \text{con} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

Si $(\boldsymbol{I} - \boldsymbol{B})$ es de rango completo, entonces

$$\boldsymbol{Y} \sim \mathcal{N}\left(\mathbf{0}, (\boldsymbol{I} - \boldsymbol{B})^{-1} \widetilde{D} ((\boldsymbol{I} - \boldsymbol{B})^{-1})^\top\right)$$

donde $\widetilde{\boldsymbol{D}} = \text{diag}(\sigma_i^2)$.

Modelo SAR

Definición (Simultaneous Autorregressive Models)

Suponga que

$$Y_i = \sum_j b_{ij} y_j + \epsilon_i, \quad i = 1, \ldots, n, \quad \text{con} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

Si $(\boldsymbol{I} - \boldsymbol{B})$ es de rango completo, entonces

$$\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{0}, (\boldsymbol{I} - \boldsymbol{B})^{-1} \widetilde{D}((\boldsymbol{I} - \boldsymbol{B})^{-1})^{\top}\right)$$

donde $\widetilde{\boldsymbol{D}} = \text{diag}(\sigma_i^2)$.

Obs.: $(\boldsymbol{I} - \boldsymbol{B})$ debe ser de rango completo pero no necesariamente simétrica.

Modelo SAR

- Existen dos elecciones para $\boldsymbol{B}$ se forma frecuente en la literatura.

Modelo SAR

- Existen dos elecciones para $\boldsymbol{B}$ se forma frecuente en la literatura.
- La primera es $\boldsymbol{B} = \rho \boldsymbol{W}$, donde $\boldsymbol{W}$ es la matriz de contigüidad.

Modelo SAR

- Existen dos elecciones para $\boldsymbol{B}$ se forma frecuente en la literatura.
- La primera es $\boldsymbol{B} = \rho \boldsymbol{W}$, donde $\boldsymbol{W}$ es la matriz de contigüidad.
- En este caso
$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$
donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{W}$.

Modelo SAR

- Existen dos elecciones para $\boldsymbol{B}$ se forma frecuente en la literatura.
- La primera es $\boldsymbol{B} = \rho \boldsymbol{W}$, donde $\boldsymbol{W}$ es la matriz de contigüidad.
- En este caso
$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$
donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{W}$.
- La segunda elección es $\boldsymbol{B} = \alpha \widetilde{\boldsymbol{W}}$, donde $\widetilde{\boldsymbol{W}}$ es la matriz estocástica.

Modelo SAR

- Existen dos elecciones para $\boldsymbol{B}$ se forma frecuente en la literatura.
- La primera es $\boldsymbol{B} = \rho \boldsymbol{W}$, donde $\boldsymbol{W}$ es la matriz de contigüidad.
- En este caso
$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$
donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{W}$.
- La segunda elección es $\boldsymbol{B} = \alpha \widetilde{\boldsymbol{W}}$, donde $\widetilde{\boldsymbol{W}}$ es la matriz estocástica.
- En este caso $\alpha \in (-1, 1)$ ya que todos los valores propios de $\widetilde{\boldsymbol{W}}$ son menores a 1.

Modelo SAR

- Existen dos elecciones para $\boldsymbol{B}$ se forma frecuente en la literatura.
- La primera es $\boldsymbol{B} = \rho \boldsymbol{W}$, donde $\boldsymbol{W}$ es la matriz de contigüidad.
- En este caso

$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}} \right)$$

  donde $\lambda_{(1)} < \cdots < \lambda_{(n)}$ son los valores propios ordenados de $\boldsymbol{W}$.
- La segunda elección es $\boldsymbol{B} = \alpha \widetilde{\boldsymbol{W}}$, donde $\widetilde{\boldsymbol{W}}$ es la matriz estocástica.
- En este caso $\alpha \in (-1, 1)$ ya que todos los valores propios de $\widetilde{\boldsymbol{W}}$ son menores a 1.
- Ambos ρ y α se denominan parámetro de autocorrelación espacial.

Bibliografia

Banerjee S., Carlin B., and Gelfand A. (2015) *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman Hall/CRC.

Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.

Matheron, G., (1965) Les variables régionalisés et leur estimation. Masson, Paris.

Shabenberger, O., Gotway, C. A. (2005) *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

¿Alguna Consulta?