Análisis espaciotemporal con modelos Bayesianos jerárquicos: fundamentos

¿Qué es un análisis espaciotemporal?

- Modelar componentes espaciales y temporales mediante el análisis de datos geolocalizados mediante modelos estadísticos.
- ¿Por qué analizar con modelos estadísticos y no modelos físicos o biológicos de los procesos involucrados?

Metas del análisis espaciotemporal estadístico

1. Determinar un patrón en espacio y tiempo.
2. Inferencia de parámetros poblacionales.
3. Pronósticos en el tiempo.

Componente espacial

- Primer principio de la Geografía (Tobler, 1970):
  "Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes."
- Es la base para la **correlación espacial**.

Correlación espacial

- La idea es que la cercanía física entre los puntos o elementos en el espacio tiende a generar una mayor similitud o interdependencia entre ellos, lo que se observa como una **dependencia espacial**.

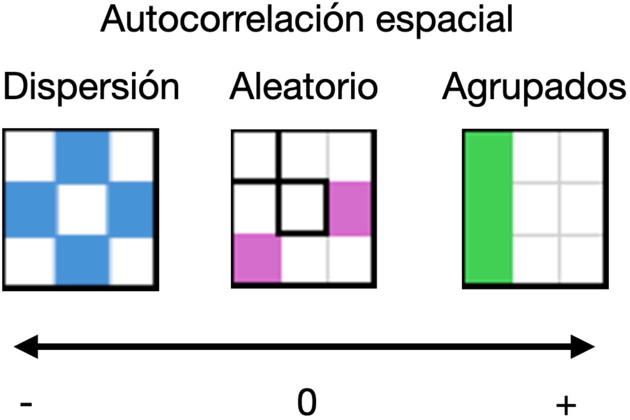
Conceptos para el patrón espacial

- **Dependencia espacial:** Proceso que determina que valores cercanos en el espacio pueden estar correlacionados a una cierta distancia (p.e., Indice de Moran)
- **Heterogeneidad espacial:** Procesos que ocurren en diferentes partes del espacio pueden tener diferentes comportamientos y patrones.

Otros aspectos

- **Autocorrelación espacial:** correlación espacial de una variable consigo misma, positiva indica valores similares que tienden a agruparse, y negativa implica que valores diferentes se encuentran cercanos unos de otros.
- **Aleatoriedad espacial:** si los datos no muestran correlación espacial, el patrón espacial es estacionario.

Autocorrelación espacial



Dependencia espacial vs. autocorrelación espacial

La dependencia espacial $=>$ implica variables exógenas a la variable de interés, i.e.,

$$y_j = \mu_y + f(\text{co-variables}_j) + \epsilon_j$$

Autocorrelación espacial $=>$ efectos endógenos de la variable de interés, i.e.,

$$y_j = \mu_y + \sum_i f(y_i - \mu_y) + \epsilon_j$$

Contiene: error *espacialmente estructurado* y una fracción independiente (i.e., ϵ_j).

Correlación temporal

Se basa en la estructura de las series de tiempo, si una variable no es estacionaria implica que la media y la varianza cambia en el tiempo.

Autocorrelación temporal: En series de tiempo univariadas, valores cercanos en el tiempo son más similares entre sí que valores lejanos en el tiempo.

Modelos estadísticos jerárquicos

- Incluyen un modelo para los datos, un modelo de los procesos, y un modelo para los parámetros.
- Análogo a los modelos en el espacio de los estados de series de tiempo.

La estructura jerárquica

- Si los parámetros son tratados con distribución *prior* en la más baja jerarquía entonces se tiene un modelo Bayesiano jerárquico.

1. Modelo de los datos: [datos|procesos, parámetros]
2. Modelo de los procesos: [procesos|parámetros]
3. Parámetros del modelo: [parámetros]

Probabilidad condicional

[proceso, parms|datos] $\propto$
[datos|procesos,parms]x[procesos|parms]x[parms]

Al dividir el lado derecho por [datos], se debe aplicar la regla de probabilidad condicional de Bayes.

Modelos complejos requieren evaluación numérica del posterior, p.e. metodos MCMC.

Modelos estadísticos espacio-temporales jerárquicos

Dos paradigmas:

$$[\text{observaciones}] = [\text{verdadero proceso}] + [\text{error de observación}]$$

$$[\text{verdadero proceso}] = [\text{regresión}] + [\text{proceso aleatorio dependiente}]$$

El segundo enfoque es un enfoque dinámico, mientras que el primero es un enfoque descriptivo.

¿Qué es INLA?

- Integrated Nested Laplace Approximation
- Introducido por Rue (2009), donde la estimación del Posterior utiliza aaproximación numérica.
- Marco conceptual unificado para analizar modelos latentes Gausianos.
- Exactitud y computacionalmente superior a métodos MCMC
- Disponible utilizando la interface R-INLA Project

Tutoriales

Gómez-Rubio, Virgilio (2020). Bayesian Inference with INLA. Chapman & Hall/CRC Press. Boca Raton, FL.

Moraga, Paula (2023). Spatial Statistics for Data Science: Theory and Practice with R

Otros disponibles en:

R-INLA Proyect - Books

Modelos latentes Gausiano

$$\eta_i = g(\mu_i) = \alpha + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f_k(c_{ki}) + \epsilon_i$$

donde: $g(.)$ es la función de enlace, y

α : intercepto β : efectos lineales de las covariables z_j $f_k(.)$: efectos no-lineales de las covariables c_k ϵ : efectos aleatorios iid

Modelos latentes Gausiano (cont.)

- Consider todos los parámetros (variables aleatorias) del predictor lineal en un campo latente, i.e.,

$$\mathbf{x} = [\alpha, \beta, f_k(.), \eta]$$

- Un modelo latente Gausiano se obtiene asignando prior Gausiano a todos los elementos de $\mathbf{x}$
- Flexible debido a las muchas formas diferentes de las funciones desconocidas $f_k(.)$ (incluye covariables temporales y/o espaciales)
- Los *hiperparámetros* dan cuenta de la variablidad y grado de dependencia.

Ejemplos de modelos latentes Gausianos

- GLMM y GAMM
- Regresión semiparamétrica
- Análisis de sobrevivencia
- Log-Gaussian Cox-processes (útil para muestreo preferencial)
- Modelos geoestadísticos
- **Modelos espaciales y espaciotemporales**
- Modelos lineal dinámicos
- Modelos en el espacio de los estados

Modelo espaciotemporal

1. Ocurrencia:

$$Z_{s,t} \sim \mathsf{Ber}(\pi_{s,t})$$

$\pi_{s,t}$ es la occurrence.

s_i son localidades (lances), donde $i = 1, 2, ..., n$.

t_j indica tiempo, donde $j = 1, 2, ..., t$.

Predictor lineal para la ocurrencia

$$\text{logit}(\pi_{s,t}) = \alpha_z + f(D) + f(T) + V^z_{s,t}$$

α_z: intercepto.

$f(D)$: efectos no-lineal aleatorio de la profundidad de fondo.

$f(T)$: efectos temporal aleatorio no estructurado (iid).

$V_{s,t} = w_{s,t}$: campo geoestadístico con función de covarianza dada por el rango y la desviación estándard marginal del campo espacial ($w_{s,t}$).

Campo espacial

$$\mathbf{W} \sim \mathsf{N}(0, Q(\kappa, \tau))$$

$\mathbf{Q}$ es una "sparse precision matrix", defined por "Stochastic Partial Differential Equations (SPDE)" para representar una aproximación del campo Gausiano continuo (GF) a través de un "Gaussian Markov Random Field (GMRF)".

Stochastic Partial Differential Equations

- La aproximación SPDE (Lindgren et al., 2011) contiene los parámetros κ y τ del campo espacial, i.e.,

$$w_s = (\kappa - \Delta)^{\alpha/2} \tau X(s)$$

$\kappa > 0$ es el parámetro de escala

τ controla la varianza del campo Gausiano

$X(s)$ es el campo Gaussian aleatorio

w_s is ruido estacionario.

Correlación espacial de Matern

$$\rho_s(X_s) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa \parallel s_i - s_j \parallel)^{\nu} K_{\nu}(\kappa \parallel s_i - s_j \parallel)$$

- Los parámetros relacionados al rango y escala de la estructura de covarianza Mátern, $\kappa > 0$ y parámetro $(\nu > 0)$,

$\parallel s_i - s_j \parallel$ denota distancia Euclidean, y K_ν is la función de Bessel de segundo orden modificada.

El modelo SPDE es una function de una trama y un parámetro α del campo espacial.

Modelo espaciotemporal densidad

2. Hurdle o Delta-Gamma: CPUA condicional a la presencia, i.e.,

$$Y_{s,t} = \begin{cases} 0 & \text{if } Z_{s,t} = 0 \\ \text{Gamma}(\mu_{s,t}, c\phi) & \text{if } Z_{s,t} = 1 \end{cases}$$

$\text{E}(Y) = \mu$

$\text{Var}(Y) = \mu^2/(c\phi)$

c es un parámetro de escalay $1/\phi$ es un parámetro de dispersión.

Nota: INLA considera precision como (1/variance), y para Gamma es equivalente a ϕ.

Predictor lineal

$$\log(\mu_{s,t}) = \alpha_y + f_1(D) + f_2(T) + V^y_{s,t}$$

Efectos no-lineales

Los efectos aleatorios no-lineales $(f(.))$, pueden ser modelados como:

- Modelos "iid"
- Modelos con caminanta aleatoria simple RW1 o doble RW2
- Modelos con autocorrelación AR(1)

Las variables aleatorias continuas como la profundidad pueden ser divididas en clases o grupos.

Modelos estucturales para el campo espacial

Campo espacial constante:

$$V_{s,t} = W_s$$

Campo espacial oportunista (replicas independientes en el tiempo):

$$V_{s,t} = W_{s,t}$$

Campo espacial progresivo (con autocorrelación temporal):

$$V_{s,t} = W_{s,t} + \sum_{k=1}^{K} \rho_k V_{s(t-k)}$$

Selección de modelos

- DIC
- WAIC

Grilla de predicción

INLA utiliza ajustes para la estimación diferente de la predicción.

- Grilla de predicción: necesaria para la predicción de las variables de interés y la obtención de índices de abundancia, i.e.,

Indice de biomasa

$$\hat{I}_t = \sum_{j=1}^{n_j} a_j \hat{y}_{j,t}$$

$\hat{I}_t$ biomasa integrada en el año t

a_j superficie de cada celda de la grilla (km^2)

Estimación del modelo

$$\hat{y}_{j,t} = \exp(\hat{Y}_{j,t}) \times \exp(\hat{Z}_{j,t})/(1 + \exp(\hat{Z}_{j,t}))$$

$\hat{Y}_{j,t}$ es la estimación Gamma,

$\hat{Z}_{j,t}$ is la ocurrencia.

PAUSA

CAFE