

Facultad de Ciencias Departamento de Estadística

Taller 1 Estadística Espacial

Docente: Ramón Giraldo Henao

Autores

Gerardo Sebastian Gil Sanchez Brayan Stick Cubides Sarmiento Andrés Felipe Rache Espitia Brad Dereck Rozo Rubio

Febrero 2024

1. Representación espectral de un proceso estocástico

En estadística espacial, la representación espectral es una herramienta crucial para el análisis de datos geoespaciales y el modelado de procesos espaciales. Proporciona una forma de descomponer la varianza o la autocorrelación espacial en diferentes frecuencias, lo que permite identificar patrones espaciales significativos y entender la estructura de dependencia espacial en los datos.

1.1. Función de Autocovarianza Espacial y Función de Autocorrelación Espacial

Para un proceso estocástico espacial, la representación espectral se puede obtener a partir de la función de autocovarianza espacial o la función de autocovarianza espacial. Si $\gamma(h)$ es la función de autocovarianza espacial, entonces su Transformada de Fourier, denotada como $\Gamma(\omega)$, se define como:

$$\Gamma(\omega) = \int_{-\infty}^{\infty} \gamma(h) e^{-i\omega h} dh$$

Si $\rho(h)$ es la función de autocorrelación espacial, entonces su Transformada de Fourier, denotada como $R(\omega)$, se define como:

$$R(\omega) = \int_{-\infty}^{\infty} \rho(h)e^{-i\omega h}dh$$

1.2. Densidad Espectral de Potencia Espacial

La densidad espectral de potencia espacial (DEP) se define como la magnitud al cuadrado de la Transformada de Fourier de la función de autocovarianza espacial:

$$S(\omega) = |\Gamma(\omega)|^2$$

1.3. Interpretación en el Contexto Espacial

En la representación espectral espacial, el eje horizontal representa la frecuencia espacial, que puede ser la frecuencia angular o la frecuencia normalizada según la escala de distancia utilizada. El eje vertical representa la densidad espectral de potencia espacial, que indica cuánta varianza del proceso está presente en cada frecuencia espacial.

La representación espectral espacial es útil para identificar patrones espaciales periódicos, como variaciones espaciales estacionales o tendencias espaciales de largo plazo. También es útil para el análisis de ruido espacial y la identificación de componentes de alta frecuencia que pueden corresponder a variabilidad espacial aleatoria en los datos.

1.4. Ejemplo de Aplicación

Consideremos un conjunto de datos de temperatura registrados en diferentes ubicaciones geográficas a lo largo de un período de tiempo. Podemos calcular la función de autocovarianza espacial entre las temperaturas en diferentes ubicaciones y luego obtener su densidad espectral de potencia espacial para analizar la distribución de la varianza espacial en diferentes escalas de frecuencia.

2. Transformada de Fourier

En el campo de la estadística espacial, la Transformada de Fourier tiene aplicaciones importantes en el análisis y modelado de datos geoespaciales, así como en la detección de patrones espaciales y la interpolación espacial.

2.1. Transformada de Fourier Continua

La Transformada de Fourier continua se define como:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$$
$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega$$

donde f(t) es la función en el dominio del tiempo, $F(\omega)$ es su transformada en el dominio de la frecuencia y ω es la frecuencia angular.

La Transformada de Fourier continua se utiliza en estadística espacial cuando se trabaja con datos espaciales que están definidos en coordenadas continuas, como por ejemplo, la concentración de contaminantes en un área geográfica.

2.2. Transformada de Fourier Discreta

La Transformada de Fourier discreta se define como:

$$F(k) = \sum_{n=0}^{N-1} f(n)e^{-i2\pi kn/N}$$

$$f(n) = \frac{1}{N} \sum_{k=0}^{N-1} F(k)e^{i2\pi kn/N}$$

donde f(n) es la secuencia en el dominio discreto del tiempo, F(k) es su transformada en el dominio discreto de la frecuencia y N es el número de muestras.

La Transformada de Fourier discreta es útil en estadística espacial cuando se trabaja con datos discretizados en el espacio, como por ejemplo, la densidad de población en regiones específicas.

2.3. Aplicaciones en Estadística Espacial

La Transformada de Fourier encuentra aplicaciones en estadística espacial en:

- Análisis Espacial de Datos: Permite examinar la estructura de correlación espacial en los datos al identificar patrones de variabilidad espacial en diferentes escalas de frecuencia.
- Interpolación Espacial: Facilita la interpolación de datos espaciales al descomponer la variabilidad espacial en componentes de frecuencia y reconstruir la superficie espacial.
- Detección de Patrones Espaciales: Ayuda a detectar patrones espaciales y anomalías al analizar la distribución de energía espectral en el dominio espacial.

La Transformada de Fourier es una herramienta poderosa en estadística espacial que permite explorar y modelar la estructura espacial de los datos geoespaciales.

3. Continuidad en media cuadrática

La continuidad en media cuadrática es un concepto fundamental en estadística espacial que se refiere a una forma de continuidad que se define en términos de la convergencia de la media cuadrática de una sucesión de funciones o procesos estocásticos.

3.1. Definición

Para una secuencia de funciones $f_n(x)$ definidas en un intervalo [a, b], la secuencia converge en media cuadrática a una función f(x) si la siguiente condición se cumple:

$$\lim_{n \to \infty} \int_a^b |f_n(x) - f(x)|^2 dx = 0$$

Esta condición implica que, a medida que n tiende a infinito, el cuadrado de la diferencia entre $f_n(x)$ y f(x) integrado sobre el intervalo [a,b] tiende a cero.

3.2. Interpretación

La continuidad en media cuadrática es una noción de convergencia más débil que la convergencia puntual o la convergencia uniforme. En el contexto de estadística espacial, esta propiedad es relevante al analizar la convergencia de procesos estocásticos espaciales. Por ejemplo, si una sucesión de funciones o procesos estocásticos converge en media cuadrática, entonces la diferencia cuadrada entre las funciones o procesos converge a cero de manera integrada, lo que indica una cierta forma de continuidad en el comportamiento estadístico.

3.3. Aplicaciones

En estadística espacial, la continuidad en media cuadrática es relevante en el análisis de la convergencia de estimadores, la validez de modelos de predicción espacial y la evaluación de la calidad de interpolaciones espaciales. Por ejemplo, al evaluar la convergencia de un estimador de kriging a medida que aumenta el tamaño de la muestra, la continuidad en media cuadrática proporciona un criterio para determinar si el estimador converge en términos de sus propiedades estadísticas fundamentales.

La continuidad en media cuadrática es una herramienta importante en estadística espacial para evaluar la convergencia y la calidad de los métodos y modelos utilizados en el análisis de datos geoespaciales. Proporciona un marco teórico sólido para comprender la convergencia de procesos estocásticos y la precisión de las predicciones espaciales.

4. Descomposición de Cholesky

Sea A_{n*n} una matriz simétrica de tamaño, definida positiva (con todos sus valores propios positivos), entonces existe una matriz L triangular inferior tal que

$$A = LL^t$$

donde todos los elementos de la diagonal de L son positivos.

4.1. Procedimiento

Para la construcción de la matriz L, según los componentes a_{ij} de la matriz A, se hará componente a componente,

$$l_{ij} = \begin{cases} l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ki}^2} & \text{si } i = j; \\ l_{ij} = \frac{a_{ij} - \sum_{k=1}^{i-1} l_{ki} l_{kj}}{l_{jj}} & \text{si } i \neq j. \end{cases}$$

En particular, para i = 1, se tiene

$$l_{1j} = \begin{cases} \sqrt{a_{ii}} & \text{si } j = 1; \\ \frac{a_{1j}}{l_{11}} & \text{si } j > 1. \end{cases}$$

4.2. Propiedad Importante

- $det(A) = det(LL^t) = det(L)det(L^t) = \prod_{i=1}^n t_{ii}^2$ donde t_{ii} es la i-ésima componente de la diagonal de L
- Si no se puede obtener la descomposición de Cholesky de una matriz (por ejemplo, cuando al realizar el procedimiento presentado arriba surge una raíz cuadrada de un número negativo), esto es indicio de que la matriz simétrica no es definida positiva.

EJERCICIOS

Del libro Statistical Methods for Spatial Data Analysis. Schabenberger & Gotway

Problema 1.1

Clasifique los siguientes ejemplos de datos espaciales según su tipo de datos:

- 1. Elevaciones en las estribaciones de las montañas Allegheny.
 - Geoestadística
- 2. Elevación más alta dentro de cada estado de los Estados Unidos.
 - Análisis de datos de áreas
- 3. Concentración de un mineral en el suelo.
 - Geoestadística

- 4. Trazar los rendimientos en una prueba de uniformidad.
 - Factores Puntuales
- 5. Estadísticas de delincuencia que indican los nombres de las subdivisiones donde se produjeron robos el año anterior y los valores de las pérdidas de propiedad.
 - Análisis de datos de áreas
- 6. Igual que (5), pero en lugar de la subdivisión, la vivienda individual es identificado.
 - Factores Puntuales
- 7. Distribución de robles y pinos en un rodal forestal.
 - Geoestadística
- 8. Número de nidos de ardillas en los pinos del rodal en (7).
 - Factores Puntuales

Problema 1.2

Considerando $Y_1, Y_2, ..., Y_n$, variables aleatorias gaussianas, con media μ , varianza σ^2 , y $COV[Y_i, Y_j] = \sigma^2 \rho$, $\forall i \neq j$. ¿Es $S^2 = (n-1)^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ un estimador insesgado de σ^2 ?

Solución.

$$\begin{split} E[S^2] &= E\left[\left(n-1\right)^{-1}\sum_{i=1}^n(Y_i-\bar{Y})^2\right] \\ &= \frac{1}{n-1}E\left[\sum_{i=1}^n(Y_i^2-2Y_i\bar{Y}+\bar{Y}^2)\right] \\ &= \frac{1}{n-1}\sum_{i=1}^n\left[E(Y_i^2)-2E(Y_i\bar{Y})+E(\bar{Y}^2)\right] \\ &= \frac{1}{n-1}\sum_{i=1}^n\left[Var(Y_i)+(E(Y_i))^2-2E\left(Y_i\frac{1}{n}\sum_{j=1}^nY_j\right)+Var(\bar{Y})+(E(\bar{Y}))^2\right] \\ &= \frac{1}{n-1}\sum_{i=1}^n\left[\sigma^2+\mu^2-\frac{2}{n}E\left(Y_i^2+\sum_{j\neq i}^nY_iY_j\right)+Var\left(\frac{1}{n}\sum_{j=1}^nY_j\right)+\mu^2\right] \\ &= \frac{1}{n-1}\sum_{i=1}^n\left[\sigma^2+\mu^2-\frac{2}{n}\left(E(Y_i^2)+\sum_{j\neq i}^nE(Y_iY_j)\right)+\frac{1}{n^2}\left(\sum_{j=1}^nVar(Y_j)+\sum_{i\neq j}^nCov(Y_i,Y_j)\right)+\mu^2\right] \\ &= \frac{1}{n-1}\sum_{i=1}^n\left[\sigma^2+\mu^2-\frac{2}{n}\left(\sigma^2+\mu^2+\sum_{j\neq i}^n(Cov(Y_i,Y_j)+E(Y_i)E(Y_j))\right)+\frac{1}{n^2}\left(n\sigma^2+n(n-1)\rho\sigma^2\right)+\mu^2\right] \\ &= \frac{n}{n-1}\left[\sigma^2+\mu^2-\frac{2}{n}\left(\sigma^2+\mu^2+n(n-1)(\rho\sigma^2+\mu^2)\right)+\frac{\sigma^2}{n}+\frac{(n-1)\rho}{n}\sigma^2+\mu^2\right] \\ &= \frac{n}{n-1}\left[\sigma^2+\mu^2-\frac{2}{n}\sigma^2-\frac{2}{n}\mu^2-2(n-1)\rho\sigma^2-2(n-1)\mu^2+\frac{\sigma^2}{n}+\frac{(n-1)\rho}{n}\sigma^2+\mu^2\right] \\ &= \frac{n}{n-1}\sigma^2+\frac{n}{n-1}\mu^2-2(n-1)\sigma^2-2(n-1)\mu^2-2n\rho\sigma^2-2n\mu^2+\frac{\sigma^2}{n-1}+\rho\sigma^2+\frac{n}{n-1}\mu^2 \\ &= \left(\frac{n}{n-1}-2(n-1)-2n\rho+\frac{1}{n-1}+\rho\right)\sigma^2+\left(\frac{2n}{n-1}-2(n-1)-2n\right)\mu^2 \end{split}$$

Por lo tanto S^2 no es un estimador insesgado de σ^2 .

Problema 1.8

Considere la red simple 2 x 3 con observaciones $Z(s_1) = 5$, $Z(s_2) = -3$, $Z(s_3) = -6$, etc. Para los problemas (1)–(3) siguientes, supongamos la definición de conectividad espacial de la torre.

	Columna 1	Columna 2	Columna 3
Fila 1	5	-3	-6
Fila 2	2	4	-2

1. Derive empíricamente la media y la varianza del I de Moran bajo aleatorización, enumerando las 6! permutaciones de la red. Compare su respuesta con las fórmulas para E[I] y Var[I].

2. Calcule el p-valor empírico para la hipótesis de que no hay autocorrelación espacial. Compárelo con el p-valor basado en la aproximación gaussiana bajo aleatorización. Para este problema necesitas conocer la varianza de *I* bajo aleatorización. Puede obtenerse de la siguiente expresión, dada en Cliff y Ord (1981, cap 2):

$$E_{r}[I^{2}] = \frac{n[n^{2} - 3n + 3]S_{1} - nS_{2} + 3w^{2}... - b[(n^{2} - n)S_{1} - 2nS_{2} + 6w^{2}...}{(n - 3)(n - 2)(n - 1)w^{2}...}$$

$$S_{1} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} + w_{ji})^{2}$$

$$S_{2} = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji} \right\}^{2}$$

$$b = n \frac{\sum_{i=1}^{n} (Z(s_{i}) - \bar{Z})^{4}}{\left\{ \sum_{i=1}^{n} (Z(s_{i}) - \bar{Z})^{2} \right\}^{2}}$$

3. ¡Prepara un histograma de los 6! I valores de Moran. ¿Parecen distribución Normal?

Solución.

1. haciendo la simulación del las 6! permutaciones tenemos que $E_s(I) = 0.16$ y $sd_s(I) = 0.213$ mientras que bajo las fomulas tenemos que:

$$E(I)=-\frac{1}{n-1}=-0.2$$
y
$$sd(I)=\sqrt{\frac{n^2S_1-n^2S_2+3S_0^2}{(n^2-1)S_0^2}-\frac{1}{(n-1)^2}}=0.212$$

- 2. el p-valor empirico para la hipótesis de no autocorrelación espacial es 0,002952624 mientras que el p-valor basado en la aproximación gaussiana es 0,00331191.
- 3. A continuación se presenta el siguiente histograma con las 6! permutaciones de la estadística I de Moran.

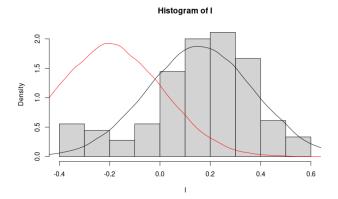


Figura 1: Histograma de permutaciones de la estadística I

La linea negra representa la curva de la función de densidad de una distribución normal con la media y varianza empiricas mientras que la curva roja representa la densidad de una normal con las formulas de la media y la varianza. Como podemos ver la curva negra mas o menos se ajusta al histograma de las permutaciones de I, por lo que los valores de I podrian tener una distribución normal.

Problema 1.9

Considerando $Y_1, Y_2, ..., Y_n$, variables aleatorias gaussianas, con media μ , varianza σ^2 , y $COV[Y_i, Y_j] = \sigma^2 \rho(|i-j|=1)$. Asumiendo que $\rho > 0$, encontrar la función de potencia de la prueba $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$, donde el estadístico de prueba está dado por

$$Z_{obs} = \frac{\bar{Y} - \mu_0}{\sqrt{Var[\bar{Y}]}}$$

Compare esta función de potencia con la función de potencia que se obtiene para $\rho=0$

Solución. para $\rho > 0$ tenemos que la función de potencia $\pi(\mu)$ es:

$$\pi(\mu) = p(\text{"Rechazar } H_0\text{"}|\text{"}H_1 \text{ es cierta"})$$

= $1 - p(\text{"Aceptar } H_0\text{"}|\text{"}H_1 \text{ es cierta"})$
= $1 - p(Z \le Z_{1-\alpha})$

donde Z tiene distribución normal con $E(Z) = \mu > \mu_0$ y $Var(Z) = \frac{\sigma^2}{n} + \frac{2(n-1)}{n^2}\rho\sigma^2$. Si $\rho = 0$ tenemos la misma función de potencia salvo que $Var(Z) = \frac{\sigma^2}{n}$. Por lo tanto si $\rho > 0$ la varianza de la distribución es mas alta que en comparación con el caso en el que $\rho = 0$.

Problema 1.12

Establecer algebraicamente la equivalencia de $p(Y_0) = \lambda' Z(s)$ y (1.25), donde λ viene dada por (1.24).

Solución.

1. Expresión 1: $p(Y_0) = \lambda' Y$

$$p(Y_0) = \lambda' Y$$

Sustituimos la expresión dada para λ' :

$$p(Y_0) = \left[c + \frac{\mathbf{1} \cdot (1 - \mathbf{1} \cdot \Sigma^{-1} \cdot c)}{\mathbf{1}^T \cdot \Sigma^{-1} \cdot \mathbf{1}}\right] \Sigma^{-1} Y \mathbf{1}$$
$$p(Y_0) = c \Sigma^{-1} Y + \frac{\mathbf{1} \cdot (1 - \mathbf{1} \cdot \Sigma^{-1} \cdot c)}{\mathbf{1}^T \cdot \Sigma^{-1} \cdot \mathbf{1}} \Sigma^{-1} Y$$

Notamos que $\frac{\mathbf{1}\cdot(1-\mathbf{1}\cdot\Sigma^{-1}\cdot c)}{\mathbf{1}^T\cdot\Sigma^{-1}\cdot\mathbf{1}}$ es un escalar, lo llamaremos α :

$$p(Y_0) = c\Sigma^{-1}Y + \alpha\Sigma^{-1}Y$$

$$p(Y_0) = (c + \alpha)\Sigma^{-1}Y$$

2. Expresión 2: $p(Y_0) = \hat{\mu} + c' \Sigma^{-1} (Y - \mathbf{1} \cdot \hat{\mu})$

$$p(Y_0) = \hat{\mu} + c' \Sigma^{-1} (Y - \mathbf{1} \cdot \hat{\mu})$$

$$p(Y_0) = c' \Sigma^{-1} Y - c' \Sigma^{-1} \mathbf{1} \cdot \hat{\mu} + \hat{\mu}$$

Agrupamos términos:

$$p(Y_0) = c' \Sigma^{-1} Y + (\hat{\mu} - c' \Sigma^{-1} \mathbf{1} \cdot \hat{\mu})$$

Notamos que $(\hat{\mu} - c'\Sigma^{-1}\mathbf{1} \cdot \hat{\mu})$ es un escalar, lo llamaremos β :

$$p(Y_0) = c' \Sigma^{-1} Y + \beta$$

Ambas expresiones son de la forma $c''\Sigma^{-1}Y$, donde c'' es una combinación lineal de c, α , c', y β , dependiendo de la expresión.

En conclusión, ambas expresiones $p(Y_0) = \lambda' Y$ y $p(Y_0) = \hat{\mu} + c' \Sigma^{-1} (Y - \mathbf{1} \cdot \hat{\mu})$ son equivalentes y pueden expresarse como $p(Y_0) = c'' \Sigma^{-1} Y$, donde c'' es una combinación lineal de c, α , c', y β .

Problema 1.13

Considere un campo aleatorio $Z(s) \sim G(\mu, \sigma^2)$ con covariograma COV[Z(s), Z(s+h)] = C(h), y sea s_0 una ubicación no observada. Si $[Z(s_0), Z(s)]'$ tienen distribución normal, conjunta, encontrar

 $E[Z(s_0)|Z(s)]$ y comparar con el valor hallado en (1.25).

Solución.

Dado que $Z(s_0)$ es una variable aleatoria de un proceso estocástico, y Z(s) es un vector aleatorio con varias variables aleatorias que describen un proceso estocástico.

Considere un campo aleatorio $Z(s) \sim G(\mu, \sigma^2)$ con covariograma C(h), y sea s_0 una ubicación no observada. Si $[Z(s_0), Z(s)]'$ tienen distribución normal, conjunta, encontrar $E[Z(s_0)|Z(s)]$ y comparar con el valor hallado anteriormente, de $\hat{\mu} = (\mathbf{1}\Sigma^{-1}\mathbf{1})^{-1}\mathbf{1}'\Sigma^{-1}Y$.

Para abordar este problema, consideremos un campo aleatorio $Z(s) \sim G(\mu, \sigma^2)$ con covariograma C(h). Además, supongamos que $[Z(s_0), Z(s)]'$ tiene una distribución conjunta normal.

Ahora, consideremos la fórmula general para la expectativa condicional:

$$E[Z(s_0)|Z(s)] = \mu_{Z(s_0)} + \Sigma_{Z(s_0),Z(s)}\Sigma_{Z(s)}^{-1}(Z(s) - \mu_{Z(s)})$$

Dado que estamos tratando con un vector aleatorio $Z(s) = [Z_1, Z_2, \dots, Z_n]'$, donde Z_i son las variables aleatorias en diferentes ubicaciones, la media $\mu_{Z(s)}$ y la matriz de covarianza $\Sigma_{Z(s)}$ se pueden expresar de la siguiente manera:

$$\mu_{\boldsymbol{Z(s)}} = [\mu, \mu, \dots, \mu]'$$

$$\Sigma_{Z(s)} = \begin{bmatrix} \sigma^2 & C(h) \\ C(h)' & \Sigma_{n-1} \end{bmatrix}$$

Donde Σ_{n-1} es la matriz de covarianza entre las ubicaciones s_i diferentes de s y s_0 .

La covarianza $\Sigma_{Z(s_0), \mathbf{Z}(s)}$ entre $Z(s_0)$ y $\mathbf{Z}(s)$ puede expresarse como:

$$\Sigma_{Z(s_0), \mathbf{Z}(s)} = [C(h), C(h), \dots, C(h)]'$$

Sustituyendo estas expresiones en la fórmula de expectativa condicional, obtenemos:

$$E[Z(s_0)|\mathbf{Z}(s)] = \mu + C(h)\Sigma_{\mathbf{Z}(s)}^{-1}(\mathbf{Z}(s) - \mu)$$

Ahora, si comparamos esto con la expresión anterior para $\hat{\mu}$:

$$\hat{\mu} = (\mathbf{1}\Sigma^{-1}\mathbf{1})^{-1}\mathbf{1}^T\Sigma^{-1}Y$$

Vemos que hay similitudes en la estructura, pero la diferencia clave radica en que la expresión para $E[Z(s_0)|Z(s)]$ incorpora la covarianza C(h) entre $Z(s_0)$ y Z(s), mientras que la expresión para $\hat{\mu}$ se basa únicamente en la estructura de covarianza dentro de Z(s) sin considerar la ubicación específica s_0 . En otras palabras, $E[Z(s_0)|Z(s)]$ utiliza información adicional sobre la relación entre s_0 y s proporcionada por la covarianza espacial C(h).