Contraste de hipótesis y Análisis de Multirresolución de Sistemas Wavelets

Henry Aquino Mosquera

25 de octubre de 2024

1. Objetivo y justificación
   - Objetivos
   - Justificación
2. Análisis Multiresolución
3. Estimación
4. Contraste de hipótesis

Objetivos de la investigación

Objetivo General: Propuesta de método de contraste de hipótesis para comparar curvas: test con propiedades estadísticas óptimas para un contraste de prueba de hipótesis funcional de la hipótesis nula expresada heurísticamente como sigue:

$$H_0 : \text{“Dos o más grupos de curvas son iguales”.}$$

Objetivos específicos

- Controlar parámetros del test: controlar los datos atípicos de una serie temporal de datos para un contraste de prueba de hipótesis funcional.
- Elegir el grado de aproximación de una serie temporal de datos para un contraste de prueba de hipótesis funcional.
- implementación del método en el lenguaje R.

Justificación

Justificación:

- Los parámetros utilizados en las ecuaciones incluyen valores promedios, covarianzas entre y dentro de grupos, y otros datos puntuales, como el número de datos o grupos.
- Los parámetros de los test ANOVA/MANOVA no son controlables, ya que dependen completamente de la calidad de la información disponible. Esto puede llevar a evaluaciones inadecuadas, ya que los datos atípicos presentes en la evidencia empírica pueden distorsionar las conclusiones y resultar en resultados sesgados. Por lo tanto, es crucial considerar estos aspectos al realizar el análisis e interpretar los resultados.

1. Objetivo y justificación
2. Análisis Multiresolución
   - Sistema intuitivo
   - Función de escala
3. Estimación
4. Contraste de hipótesis

Idea intuitiva del AMR

El requisito fundamental del Análisis de Multiresolución (AMR) es que los espacios generados por V_j estén anidados de la siguiente manera:

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots \subset L^2 \tag{1}$$

A partir de (1), podemos comenzar en cualquier V_j, por ejemplo, en $j = 0$, y escribir

$$V_{j_0} \subset V_1 \subset V_2 \subset \cdots \subset L^2$$

Visto de otra manera, se puede generar la siguiente expresión

$$V_1 = V_{j_0} \oplus W_0$$

Que se extiende a

$$V_2 = V_{j_0} \oplus W_0 \oplus W_1$$

En general esto da

$$L^2 = V_{j_0} \oplus W_0 \oplus W_1 \oplus \cdots \tag{2}$$

Sistema intuitivo

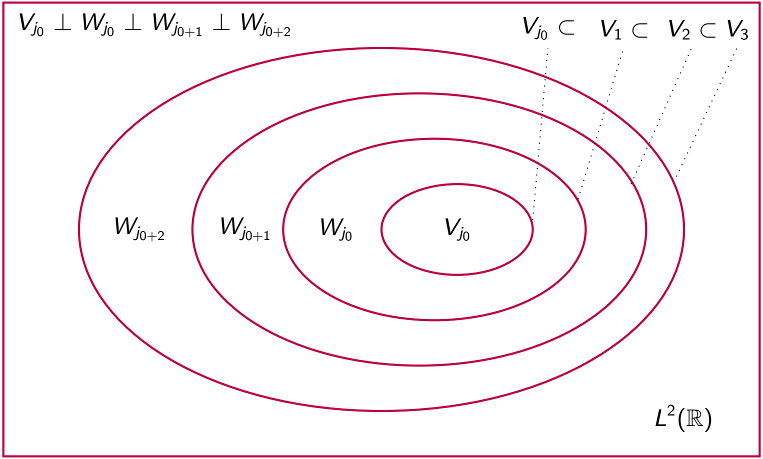

Figura 1: Función de escala y Espacio de Vectores Wavelet

1. Objetivo y justificación

2. Análisis Multiresolución
   - Sistema intuitivo
   - Función de escala

3. Estimación

4. Contraste de hipótesis

Función de escala

Se genera una familia de parametrización de funciones a partir de la función de escala básica $\varphi(t) \in L^2(\mathbb{R})$, cuyo escalado y traslación genera

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k); j, k \in \mathbb{Z}. \tag{3}$$

Para un j fijo, el espacio generado V_j sobre k es

$$V_j = \overline{\text{Span}}\left\{\varphi_{jk}(t) : k \in \mathbb{Z}\right\}$$

Función de escala

En base a la ecuación (2) se define la función wavelet

$$\psi(t) = \sum_k h_k \varphi(2^j t - k); \tag{4}$$

Puedo generar otro tipo de funciones

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k); \tag{5}$$

$$W_j = \overline{\text{Span}}\{\psi_{jk}(t) : k \in \mathbb{Z}\}$$

para el caso de combinaciones finitas o infinitas, lo cual nos lleva

$$L^2(\mathbb{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_{0+1}} \dots \tag{6}$$

de ahí que

$$f(t) = \sum_{k=1}^{2^{j_0}-1} c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j \geq j_0}^{+\infty} \sum_{k=0}^{2^{j_0}-1} d_{j,k}\psi_{j,k}(t) \tag{7}$$

donde $f(t) \in L^2(\mathbb{R})$

Secuencia de funciones

Sea una sucesión de funciones desconocida:

$$f_1, f_2, \ldots, f_r; f_i \in L^2(\mathbb{R}), i = 1, 2, \ldots, r$$

de donde puedo observar el siguiente proceso.

$$Y_i(t) = \int_0^t f_i(s)ds + \varepsilon W_i(t), i = 1, 2, \ldots, r \tag{8}$$

Donde ε es > 0

Tomando la derivada la ecuación anterior, se obtiene la siguiente ecuación diferencial (*modelo funcional*)

$$\frac{dY_i(t)}{dt} = \frac{df_i(t)}{dt} + \frac{dW_i(t)}{dt} \tag{9}$$

Donde cada función f_i será representada de la siguiente forma:

$$f_i(t) = m_0 + \mu(t) + a_i + g_i(t), i = 1, 2, \ldots, r \tag{10}$$

donde $\mu(t)$ es la función media y $g_i(t)$ la función que distingue a cada función.

Estructura funcional

Cuando los datos son observables se tiene una secuencia:

$$t_1, t_2, \ldots, t_n$$

donde

$$Y_i(t_1), Y_i(t_2), \ldots, Y_i(t_n); i = 1, 2, \ldots, r$$

Además, el tamaño de la serie para el caso discreto que permita elegir un nivel de resolución será (técnica de espejamiento):

$$J : n = 2^J$$

Por lo tanto, para un J fijo, el nivel de resolución máximo será finito:

$$Y(t) = \sum_k c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j \geq j_0}^{J} \sum d_{j,k}\psi_{j,k}(t) \tag{11}$$

Aproximación de la series

Propuesta de estimador $f(t)$

$$\widehat{f}(t) = \sum_{k} c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j \geq j_0}^{j_1} \sum \widehat{d}_{j,k}\psi_{j,k}(t) \tag{12}$$

donde

$$0 < j_0 < j_1 < J \tag{13}$$

Tener en cuenta

Metodología de estimación **dada**

- Criterio de elección de j_0, j_1
- Estimación de $\widehat{d}_{j,k}$

Por lo tanto,

- Mientras que en el universo funcional uno busca estimar los $y_i(t_n), (i \neq n)$ de cada función,
- en el universo de wavelets, se busca estimar coeficientes $d_i(j,k)$, siendo que $d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)dt$.

Tener en cuenta

Modelo básico

$$dY_i(t) = f_i(t)dt + dW_i(t)$$

Estimación de $f(t)$

$$\widehat{f}(t) = \sum_k c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j\geq j_0}^{j_1}\sum \widehat{d}_{j,k}\psi_{j,k}(t)$$

Los coeficientes son:

$$d_{j,k} = \langle f(t), \psi_{j,k}(t)\rangle = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)dt.$$

donde $d_{j,k}$ es un parámetro - **transformada de wavelet discreta de** $f(t)$.

Transformada aleatoria

A partir del proceso aleatorio $Y(t)$ tenemos que

$$D_{j,k} = \langle Y(t), \psi_{j,k}(t) \rangle = \int \psi_{j,k}(t) dY(t) = \int \psi_{j,k}(t) f(t) dt + \varepsilon \int \psi_{j,k}(t) dW(t)$$

Donde $\int \psi_{j,k}(t) dW(t)$ es la integral estocástia de ITO.

$$D_{j,k} = d_{j,k} + \varepsilon \int \psi_{j,k}(t) dW(t); \int \psi_{j,k}(t) dW(t) \sim \mathcal{N}(0,1)$$

Metodología

- Para $H_0 : \mu(t) = 0$, el modelo $\bar{Y}(t) \longrightarrow$ ruido blanco Gausseano.
- Para $H_0 : g_i(t) = 0$, el modelo $Y_i - Y(t) \longrightarrow$ ruido blanco Gausseano.

Metodología

Caso $H_0 : \mu(t)$

$$\bar{Y}(t) = \frac{1}{r}\sum : i = 1^r Y(t) = m_0 + \mu(t) + \varepsilon \bar{W}(t)$$

$$d\bar{Y}(t) = (m_0 + \mu(t))dt + \varepsilon d\bar{W}(t) \tag{14}$$

Si $\mu(t) = 0$ (constante):

$\bar{D}_{j,k} = \int \psi_{j,k}(t)\mu(t)dt + \varepsilon \int \psi_{j,k}(t)d\bar{W}(t)$

$\bar{D}_{j,k} = \bar{d}_{j,k} + \varepsilon\bar{\xi}_{j,k}$;

$\bar{d}_{j,k} = 0$, si $\mu(t) = 0$

Metodología

Para $H_0 : g_i(t)$

$$d(Y_i - \bar{Y}(t) = (a_i + g_i(t))dt + \varepsilon(W_i - \bar{W})(t)$$

$$\widetilde{D}_{j,k} = \int \psi_{j,k}(t) g_i(t) dt + \varepsilon \bar{\xi}_{j,k}$$
$$\widetilde{D}_{j,k} = \widetilde{d}_{j,k} + \varepsilon \bar{\xi}_{j,k};$$
$$g_i(t) = 0;\ \forall_i$$
$$\widetilde{d}_{j,k} = 0;\ \forall_{j,k}$$

Tests

Test para $H_0 : \mu(t) = 0$

$$X_{j_0}, \ldots, X_{j_1}$$

$$X_j = (D_1(j,1), \ldots, D_1(j, 2^j - 1), \ldots, D_r(j,1), \ldots, D_r(j, 2^j - 1))$$

Por tanto, $E(X_j) = \mu_j$:

$$H_0 : \mu_{j0} = \cdots = \mu_{j1} = 0$$

$$H_1 : \exists_j,\ \mu_j \neq 0$$

Test

Para $H_0 : g_i(t) = 0$, construimos las matrices

$$\mathbf{X}_{j_0}, \ldots, X_{j_1} \mathbf{X}_{j_0}, \ldots, X_{j_1} \mathbf{x}_{j_0}, \ldots, x_{j_1} \mathbf{x}_{j_0}, \ldots, x_{j_1}$$

X_j cuya columna

$$\mathbf{X}_{ji} \ \mathbf{X}_{ji} \ \mathbf{x}_{ji} \ \mathbf{x}_{ji} = (D_i(j,k), \ldots, D_i(j, 2^j - 1)^T) \textbf{para} i = 1, \ldots, r$$

Sea

Gracias!