



Centro de Investigación en Matemáticas, A.C.

---

# Selección del ancho de banda en la estimación por núcleo de funciones de covarianza

TESIS

Que para obtener el grado de

**Maestro en Ciencias con especialidad en  
Probabilidad y Estadística**

Presenta:

**Zitlalli Salas Gutiérrez**

Director de tesis:

**Dr. Rolando J. Biscay Lirio**

# Agradecimientos

Quiero agradecer al Dr. Rolando J. Biscay Lirio por permitirme realizar esta tesis bajo su tutela, por su paciencia y orientación a lo largo de este trabajo. También quiero agradecer a mis sinodales, Dra. Leticia Ramírez Ramírez y Dr. Inder R. Tecuapetla-Gomez, su apoyo fué indispensable.

A mis profesores, personas talentosas que también fueron ejemplo de disciplina.

Al Centro de Investigación en Matemáticas (CIMAT) por darme una educación de calidad y por permitirme usar sus instalaciones y recursos para concluir este trabajo. Agradezco también al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado en estos dos años de maestría.

A mis amigos y compañeros con quienes compartí momentos inigualables.

A mi familia por creer en mí, brindarme su cariño y motivarme para mejorar día a día. Gracias mamá y hermanas por compartir conmigo su bondad y carisma.

# Resumen

Estimar la función de covarianza de un campo aleatorio es un problema fundamental en muchos dominios de aplicación; por ejemplo, en geoestadística, análisis de datos funcionales, finanzas, epidemiología y neurociencias, por mencionar algunos. En esta tesis se propone un estimador de funciones de covarianza sobre la base de observaciones discretas. Este se basa en una estimación por núcleo caracterizada por la elección de su ancho de banda a través de un nuevo criterio. Se demuestra que este criterio estima insesadamente el riesgo asociado a la norma de Frobenius como función de pérdida. El estimador es válido para funciones de covarianza definidas sobre un dominio de cualquier dimension, y no requiere del supuesto de estacionariedad ni de que el investigador especifique alguna base de funciones conveniente. Se demuestra que el estimador es una función simétrica y definida no negativa (*i.e.*, efectivamente es una función de covarianza), satisface además una desigualdad de concentración que implica buen comportamiento de su riesgo para toda muestra finita, y es asintóticamente óptimo desde el punto de vista del riesgo cuadrático. Se presentan estudios de simulación que muestran que el estimador propuesto es factible de implementar con bajo costo computacional, y presenta buen comportamiento en la práctica, incluso para tamaños de muestra moderados.

# Índice

<b>Resumen</b>	<b>II</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Método de estimación por núcleo de funciones de covarianza</b>	<b>5</b>
2.1. Marco de trabajo . . . . .	5
2.2. Estimador por núcleo . . . . .	6
2.3. Criterio insesgado de selección del ancho de banda . . . . .	10
<b>3. Simulaciones</b>	<b>23</b>
3.1. Campo con función de covarianza estacionaria . . . . .	23
3.2. Campo con función de covarianza no estacionaria . . . . .	29
3.3. Comparación con procedimientos alternativos . . . . .	33
3.3.1. Suavizando los datos con un valor de $h$ . . . . .	33
3.3.2. Suavizando los datos con distintos valores de $h$ . . . . .	38
<b>4. Conclusiones y posibles líneas de trabajo a futuro</b>	<b>41</b>
<b>Referencias</b>	<b>43</b>
<b>Apéndice: Álgebra de Matrices</b>	<b>46</b>

# Índice de figuras

3.1.	Comparación del Riesgo Teórico vs. su Estimador Inssegado, considerando $n = 50$ , $m = 15$ y $a = 0.3$ . La línea roja indica el valor $h_0$ de $h$ que minimiza $R(h)$ , y la línea azul indica el valor $\hat{h}$ que minimiza $\hat{R}(h)$ .	26
3.2.	Comparación del Riesgo Teórico vs. su Estimador Inssegado, considerando $n = 500$ , $m = 15$ y $a = 0.3$ . La línea roja indica el valor $h_0$ de $h$ que minimiza $R(h)$ , y la línea azul indica el valor $\hat{h}$ que minimiza $\hat{R}(h)$ .	27
3.3.	Comparación de la función de covarianza teórica vs. sus estimaciones con los anchos de banda $\hat{h}$ y $h_0$ ( $n = 50$ , $m = 15$ y $a = 0.3$ ). La línea negra indica la función de covarianza teórica, y las líneas azul y roja indican sus estimaciones utilizando los anchos de banda $\hat{h}$ y $h_0$ .	28
3.4.	Comparación de la función de covarianza teórica vs. sus estimaciones con los anchos de banda $\hat{h}$ y $h_0$ ( $n = 500$ , $m = 15$ y $a = 0.3$ ). La línea negra indica la función de covarianza teórica, y las líneas azul y roja indican sus estimaciones utilizando los anchos de banda $\hat{h}$ y $h_0$ .	29
3.5.	Comparación del Riesgo Teórico vs. su Estimador Inssegado, considerando $n = 50$ , $m = 15$ y $a = 0.3$ . La línea roja indica el valor $h_0$ de $h$ que minimiza $R(h)$ , y la línea azul indica el valor $\hat{h}$ que minimiza $\hat{R}(h)$ .	30
3.6.	Comparación del Riesgo Teórico vs. su Estimador Inssegado, considerando $n = 500$ , $m = 15$ y $a = 0.3$ . La línea roja indica el valor $h_0$ de $h$ que minimiza $R(h)$ , y la línea azul indica el valor $\hat{h}$ que minimiza $\hat{R}(h)$ .	31
3.7.	Función de covarianza teórica evaluada en $m = 15$ puntos.	32

---

3.8. Estimador de función de covarianza con $h_0$ , considerando $n = 50$ y $m = 15$ . . . . .	32
3.9. Estimador de función de covarianza con $\hat{h}$ , considerando $n = 50$ y $m = 15$ . . . . .	32
3.10. Estimador de función de covarianza con $h_0$ , considerando $n = 500$ y $m = 15$ . . . . .	33
3.11. Estimador de función de covarianza con $\hat{h}$ , considerando $n = 500$ y $m = 15$ . . . . .	33
3.12. Curvas de función de covarianza teórica $\sigma$ y sus estimaciones $\hat{\sigma}_{\hat{h}}$ y $\hat{\sigma}_{h_{cv}}$	36
3.13. Comparación de valores de $h$ obtenidos por el método propuesto $\hat{h}^b$ y el método alternativo $\mathbf{h}_{cv}^b$ . . . . .	39
3.14. Curvas de función de covarianza teórica $\sigma$ y sus estimaciones $\hat{\sigma}_{\hat{h}}$ y $\hat{\sigma}_{h_{cv}}$	40
3.15. Curvas de función de correlación teórica $\rho$ y sus estimaciones $\hat{\rho}_{\hat{h}}$ y $\hat{\rho}_{h_{cv}}$	40

---

---

---

# CAPÍTULO 1

---

## Introducción

Estimar la función de covarianza de un campo aleatorio (definido sobre un dominio  $d$ -dimensional) es un problema fundamental con muchas aplicaciones. Por ejemplo, en geoestadística, finanzas, epidemiología y neurociencias, por mencionar algunas (ver *e.g.*, [9, 30], y sus referencias). En particular, la función de covarianza desempeña un rol clave en la construcción de predictores lineales como el "kriging"([9, 14, 31]) y en el estudio de las estructuras de dependencias (*e.g.*, [30, 19, 16]).

La estimación de funciones de covarianzas se ha abordado mediante métodos paramétricos, semiparamétricos y no paramétricos.

Los métodos paramétricos han sido extensamente estudiados en la literatura de estadística espacial ([9]). Sin embargo, estos métodos imponen una estructura restrictiva a la función de covarianza que en muchas ocasiones no es realista.

Los métodos semiparamétricos proporcionan una metodología más flexible dado que tienen componentes paramétricos y no paramétricos; de esta manera no son tan restrictivos como los paramétricos y conservan cierta facilidad de interpretación. Por ejemplo,

en [21] y [12] se estiman las funciones de covarianza de datos longitudinales mediante un modelo parcialmente lineal.

Los métodos no paramétricos estiman la función de covarianza desconocida dentro de una clase (no paramétrica) suficientemente grande de funciones. En las últimas tres décadas estos métodos han recibido una notable atención.

Una vertiente dentro de esta última metodología requiere de ciertos supuestos distribucionales. Por ejemplo, en [27] y [33] utilizan un enfoque bayesiano que presupone Gaussianidad de los datos.

Otra vertiente de la estimación no paramétrica de funciones de covarianza se enfoca al caso específico de covarianzas estacionarias (ver *e.g.*, [8, 10, 11, 15, 24, 27]). Cuando se supone estacionariedad, usualmente basta una sola realización del campo aleatorio sobre una rejilla creciente de puntos para construir un estimador consistente.

En la literatura de datos funcionales ([13, 18, 26, 25, 28, 34]) la estimación no paramétrica general de funciones de covarianza, sin requerir estacionariedad ni supuestos distribucionales específicos, es un problema central. En este contexto, típicamente se considera que el dominio del campo aleatorio es un intervalo y que se observa (o se aproxima mediante un preprocesamiento) su valor sobre todo el dominio.

En contraste, el presente trabajo se centra en estimar funciones de covarianza a partir de observaciones discretas del campo aleatorio; *i.e.*, sobre un conjunto finito de puntos  $t_j$  de un dominio  $T$ , se observan los valores  $Y_i(t_j)$  de réplicas  $Y_i$  de un campo aleatorio  $Y$ .

Hay métodos para estimar funciones de covarianza más generales, que no requieren supuestos de estacionariedad ni distribucionales. Algunos de estos métodos se describen a continuación.

Una de las principales dificultades en este marco de trabajo es imponer que el estimador tenga la propiedad de ser definido no negativo (*d.n.n.*). Esta propiedad es fun-

---



damental para diversos objetivos; por ejemplo, para interpretar el estimador como una covarianza y para construir predictores lineales óptimos a partir de él (e.g. [15, 4]).

Un enfoque no paramétrico que garantiza estimaciones *d.n.n* de la función de covarianza consiste en: a) aproximar la función de covarianza mediante un desarrollo en productos tensoriales asociado a una base de funciones, y b) determinar los coeficientes del desarrollo considerando como función de riesgo una norma matricial de Frobenius conveniente (ver e.g., [2, 3, 4, 5, 6, 7, 17]).

En particular, en [7] se utiliza una base spline y se minimiza una versión empírica de dicho riesgo. Este enfoque tiene la dificultad de requerir seleccionar la cantidad y posición de los nodos de la base.

De manera más general, en los trabajos [2],[3],[4],[5] y [17] la estimación de funciones de covarianza se aborda como un problema de modelo lineal matricial asociado a un desarrollo en base de funciones, y la determinación del número de funciones  $q$  se aborda como un problema estadístico de selección. Específicamente, en [2],[3],[4] y [5] esto se realiza a partir de una función de riesgo penalizada. Alternativamente, en [17] se utiliza una versión insesgada del riesgo. Un inconveniente de este enfoque es especificar una base de funciones adecuada para la situación práctica de interés.

En esta tesis se construye un estimador de la función de covarianza que no requiere especificar una base de funciones. Este estimador se basa en una aproximación por núcleo, seleccionando su ancho de banda  $h$  mediante una estimación insesgada de un riesgo cuadrático. Este método tiene la ventaja de permitir estimar funciones de covarianza generales (sin supuestos sobre estacionariedad ni distribucionales) definidas sobre un dominio  $d$ -dimensional. Además, es computacionalmente simple ya que solo depende de un parámetro escalar  $h$ .

Más específicamente, se desarrollan los siguientes objetivos.

**Objetivo general:** Construir un estimador de la función de covarianza  $\sigma$  basado en núcleo, incorporando una selección adecuada de su ancho de banda  $h$ .

---

**Objetivos específicos:**

- i) Definir un estimador por núcleo  $\hat{\sigma}_h$  de la función de covarianza  $\sigma$ .
- ii) Proponer un criterio para seleccionar el ancho de banda  $h$  del estimador.
- iii) Estudiar propiedades teóricas del estimador  $\hat{\sigma}_h$ .
- vi) Evaluar el comportamiento del criterio de selección y del estimador a través de simulaciones.

La estructura de la tesis es la siguiente. En el Capítulo 2 se describe la construcción del estimador de  $\sigma$ , así como sus propiedades teóricas. También se define un criterio insesgado para la selección del ancho de banda  $h$  y se estudian algunas de sus características. A lo largo del Capítulo 3, mediante diferentes escenarios de simulación, se evalúa el desempeño del estimador de  $\sigma$  y del criterio de selección de  $h$ . Por último, en el Capítulo 4 se presentan conclusiones sobre los resultados obtenidos, y posibles líneas abiertas de investigación.

---

---

---

## CAPÍTULO 2

---

# Método de estimación por núcleo de funciones de covarianza

En el presente capítulo se describe el desarrollo de un estimador de funciones de covarianza basado en núcleo; la construcción de un criterio para la selección de su ancho de banda y algunas de sus propiedades teóricas.

### 2.1. Marco de trabajo

Sea  $Y = (Y(t))_{t \in T}$  un campo aleatorio, indexado por  $T \subset \mathbb{R}^d$  ( $d \in \mathbb{N}$ ), que toma valores en  $\mathbb{R}$ . Se supone que tiene segundo momento finito, con función de media  $\mu(t) = \mathbb{E}(Y(t)) = 0$  y función de covarianza  $\sigma(s, t) = \mathbb{E}(Y(s)Y(t))$  para todo  $s, t \in T$ . Cabe destacar que no se supone Gaussianidad ni estacionariedad.

El tipo de datos que se consideran son los valores  $Y_i(t_j)$  para  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ , donde los puntos fijos  $t_1, \dots, t_m \in T$  componen una rejilla de  $T$ , y  $Y_1, \dots, Y_n$  son copias independientes del campo  $Y$ .

En la siguiente sección el campo aleatorio  $Y$  se aproximará mediante un estimador de Nadaraya-Watson.

## 2.2. Estimador por núcleo

El estimador de Nadaraya Watson es un estimador no paramétrico que se basa en estimar una función  $f : \mathbb{R} \rightarrow \mathbb{R}$ , en este caso  $Y$ , mediante un promedio ponderado localmente. Explícitamente,

$$\tilde{Y}(t) = \sum_{r=1}^m Y(t_r) g_r(t),$$

donde

$$g_r(t) = \frac{K_h(t - t_r)}{\sum_{l=1}^m K_h(t - t_l)}. \quad (2.1)$$

Aquí,  $K_h : T \rightarrow \mathbb{R}$  usualmente es una función simétrica tal que  $K_h \geq 0$  y  $\int_T K_h(t) = 1$ ; y  $h > 0$  es un parámetro conocido como ancho de banda.

La idea central del estimador de covarianza que a continuación se propone es que, sobre la base de la aproximación  $Y(t) \approx \tilde{Y}(t)$ , es natural considerar que

$$\sigma(s, t) \approx \tilde{\sigma}(s, t) = Cov(\tilde{Y}(s), \tilde{Y}(t)).$$

De este modo, resulta que

$$\begin{aligned} \tilde{\sigma}(s, t) &= Cov \left( \sum_{r=1}^m Y(t_r) g_r(s), \sum_{l=1}^m Y(t_l) g_l(t) \right) \\ &= \sum_{r=1}^m \sum_{l=1}^m g_r(s) g_l(t) Cov(Y(t_r), Y(t_l)) \\ &= \sum_{r=1}^m \sum_{l=1}^m g_r(s) g_l(t) \sigma(t_r, t_l), \\ \sigma(s, t) &\approx \sum_{r=1}^m \sum_{l=1}^m g_r(s) g_l(t) \sigma(t_r, t_l) \end{aligned}$$


---

donde  $g$  esta definida por la ecuación (2.1). Lo anterior sugiere definir el siguiente estimador que se propone en este trabajo:

$$\hat{\sigma}(s, t) = \sum_{r=1}^m \sum_{l=1}^m g_r(s) g_l(t) s_{rl}, \quad (2.2)$$

donde

$$\mathbf{S} = (s_{rl})_{1 \leq r, l \leq m} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^T,$$

y

$$\mathbf{y}_i = \begin{bmatrix} Y_i(t_1) \\ \vdots \\ Y_i(t_m) \end{bmatrix}.$$

Notemos que tal estimador  $\hat{\sigma}$  depende del ancho de banda  $h$ , parámetro implícito en  $g_r(t)$ ; es decir,  $\hat{\sigma} = \hat{\sigma}_h$ . Se necesita un criterio de elección del parámetro  $h$  para que el estimador quede completamente especificado. Dicho criterio se desarrolla en la sección 2.3. A continuación se estudian algunas propiedades básicas del estimador  $\hat{\sigma}_h$  para un valor  $h$  fijo.

### Representaciones equivalentes del estimador

**Lema 2.2.1** (Representaciones alternativas). *El estimador (2.2) también puede representarse de las siguientes formas:*

i) (Forma escalar)

$$\hat{\sigma}_h(s, t) = \sum_{r=1}^m \sum_{l=1}^m g_r(s) g_l(t) s_{rl}.$$

ii) (Forma vectorial)

$$\hat{\sigma}_h(s, t) = \mathbf{g}^T(s) \mathbf{S} \mathbf{g}(t),$$

$$\text{donde } \mathbf{g}(s) = \begin{bmatrix} g_1(s) \\ \vdots \\ g_m(s) \end{bmatrix}.$$

iii) (Forma matricial)

$$\hat{\Sigma}_h = (\hat{\sigma}_h(t_J, t_r))_{1 \leq J, r \leq m} = \mathbf{G}^T \mathbf{S} \mathbf{G}, \quad (2.3)$$

donde

$$\mathbf{G} = \begin{bmatrix} g_1(t_1) & \cdots & g_1(t_m) \\ \vdots & \ddots & \vdots \\ g_m(t_1) & \cdots & g_m(t_m) \end{bmatrix}. \quad (2.4)$$

Estas representaciones serán de gran utilidad en la demostración de resultados posteriores.

### Condición de ser función definida no negativa

**Proposición 2.2.1.1.** *El estimador  $\hat{\sigma}_h$  de la covarianza  $\sigma$  es una función simétrica y definida no negativa; por lo tanto, es una función de covarianza.*

*Demostración.*

- i) Trivialmente  $\hat{\sigma}_h$  es función simétrica.
- ii) Para  $u_1, \dots, u_p \in T$ ,  $p \in \mathbb{N}$  y  $a_1, \dots, a_p \in \mathbb{R}$  arbitrarios,

$$\begin{aligned} \sum_{j=1}^p \sum_{q=1}^p a_j a_q \hat{\sigma}_h(u_j, u_q) &= \sum_{j=1}^p \sum_{q=1}^p a_j a_q \sum_{r=1}^m \sum_{l=1}^m g_r(u_j) g_l(u_q) s_{rl} \\ &= \sum_{r=1}^m \sum_{l=1}^m \left( \sum_{j=1}^p a_j g_r(u_j) \right) \left( \sum_{q=1}^p a_q g_l(u_q) \right) s_{rl}, \end{aligned}$$

donde  $\alpha_r = \left( \sum_{j=1}^p a_j g_r(u_j) \right)$  y  $\alpha_l = \left( \sum_{q=1}^p a_q g_l(u_q) \right)$

$$= \sum_{r=1}^m \sum_{l=1}^m \alpha_r \alpha_l s_{rl} \geq 0.$$

Esta última desigualdad se cumple ya que toda covarianza muestral es función definida no negativa.

De aquí que

$$\sum_{j=1}^p \sum_{q=1}^p a_j a_q \hat{\sigma}_h(u_j, u_q) \geq 0,$$

y por lo tanto la función  $\hat{\sigma}_h$  es definida no negativa.

■

### Relación con Modelo Lineal Matricial

Dado que,  $\mathbb{E}(\mathbf{y}_i \mathbf{y}_i^T) = \Sigma = (\sigma(t_J, t_r))_{1 \leq J, r \leq m}$ , el proceso que se observa puede ser reescrito dentro de un marco de regresión matricial de la siguiente manera:

$$\mathbf{y}_i \mathbf{y}_i^T = \Sigma + \mathbf{U}_i,$$

donde  $\mathbf{U}_i$  es una matriz aleatoria con  $\mathbb{E}(\mathbf{U}_i) = \mathbf{0}$ . Para simplificar la notación, sea

$$\mathbf{W}_i = \mathbf{y}_i \mathbf{y}_i^T.$$

Así, el modelo queda de la forma

$$\mathbf{W}_i = \Sigma + \mathbf{U}_i. \quad (2.5)$$

Consideremos la siguiente estructura lineal para  $\Sigma$

$$\mathbf{W}_i = \mathbf{G}^T \Psi \mathbf{G} + \mathbf{U}_i. \quad (2.6)$$

donde  $\mathbf{G}$  esta dada por (2.4), y  $\Psi \in \mathbb{R}^{m \times m}$  es una matriz simétrica de parámetros desconocidos.

**Lema 2.2.2.** Si  $(\mathbf{G}\mathbf{G}^T)^- \mathbf{G} = \mathbf{G}^T (\mathbf{G}\mathbf{G}^T)^- = \mathbf{I}_m$ , entonces  $\hat{\Sigma}_h$  coincide con el estimador por mínimos cuadrados (MC) del modelo (2.6), con respecto al parámetro  $\Psi$ , es decir,

$$\hat{\Psi} = \mathbf{S} = \underset{\Psi^T = \Psi \in \mathbb{R}^{m \times m}}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{W}_i - \mathbf{G}^T \Psi \mathbf{G}\|_F^2.$$

La demostración de este resultado es consecuencia inmediata del Teorema 3.2 que se encuentra en [2].

**Lema 2.2.3.** *El modelo lineal matricial (2.6) puede escribirse como modelo lineal multivariado,*

$$\mathbf{Z}_i = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_i,$$

donde  $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n \in \mathbb{R}^p$  son vectores aleatorios con  $\mathbb{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}$ ,  $\mathbf{Z}_i = \text{vec}(\mathbf{W}_i)$ ,  $\mathbf{X} = \mathbf{G} \otimes \mathbf{G}$ ,  $\boldsymbol{\beta} = \text{vec}(\boldsymbol{\Psi})$ .

*Demostración.* Vectorizando el modelo (2.6)

$$\text{vec}(\mathbf{W}_i) = \text{vec}(\mathbf{G}^T \boldsymbol{\Psi} \mathbf{G}) + \text{vec}(\mathbf{U}_i),$$

$$\text{vec}(\mathbf{G}^T \boldsymbol{\Psi} \mathbf{G}) = (\mathbf{G} \otimes \mathbf{G}) \text{vec}(\boldsymbol{\Psi}).$$

■

Notemos que no se tiene un modelo lineal multivariado estándar, ya que la media  $\mathbf{X}\boldsymbol{\beta}$  es combinación de todas las componentes  $\boldsymbol{\beta}$  para diferentes elementos de  $\mathbf{Z}$ .

## 2.3. Criterio insesgado de selección del ancho de banda

El método que se propone para seleccionar el ancho de banda  $h$  para estimar la función de covarianza, consiste en determinar el valor de  $h$  que minimice una función de riesgo. Usaremos la función de riesgo definida por el valor esperado del error cuadrático (MSE):

$$R(h) = \mathbb{E} \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}_h\|_F^2. \quad (2.7)$$

Notar que estimador  $\hat{\boldsymbol{\Sigma}}_h$  (ver fórmula (2.3)) es equivalente a calcular la covarianza muestral de las aproximaciones  $\tilde{\mathbf{y}}_i$  de las realizaciones  $\mathbf{y}_i$ , esto es

$$\frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T = \frac{1}{n} \sum_{i=1}^n \mathbf{G}^T \mathbf{y}_i \mathbf{y}_i^T \mathbf{G} = \mathbf{G}^T \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^T \mathbf{G} = \mathbf{G}^T \mathbf{S} \mathbf{G}.$$



Notar también que el criterio de riesgo (2.7) para la selección de  $h$  difiere de la selección de un ancho de banda particular  $h_i$  para cada aproximación  $\tilde{y}_i$ ; además es un criterio que toma en cuenta la estructura de covarianza.

Sin embargo, no podemos usar directamente tal función de riesgo en una situación real, ya que ésta involucra parámetros desconocidos. En lugar de la función de riesgo, se pudiera usar como un criterio el riesgo empírico o función de pérdida cuadrática:

$$\tilde{R}(h) = \|\mathbf{S} - \hat{\Sigma}_h\|_F^2$$

Pero el criterio  $\tilde{R}(h)$  tiene la desventaja de no ser un estimador insesgado de  $R(h)$ . A continuación construiremos un criterio  $\hat{R}(h)$  que es un estimador insesgado de  $R(h)$  salvo una constante aditiva. Esta estrategia es semejante a lo que sigue en la construcción del llamado criterio  $C_p$  (ver *e.g.*, [20, 22, 23, 29, 32]); pero hay que notar que nuestro caso se trata de un modelo lineal matricial no estándar.

**Teorema 2.3.1** (Estimador insesgado de la función de riesgo). *Sea*

$$\hat{R}(h) = \tilde{R}(h) + B(h), \quad (2.8)$$

donde

$$B(h) = \frac{2}{n(n-1)} \left( \sum_{i=1}^n \text{tr} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G} \} \right).$$

Entonces, se cumple que

$$\mathbb{E}(\hat{R}(h)) = R(h) + \mathbb{E}\|\mathbf{S} - \Sigma\|_F^2.$$

O sea, salvo una constante aditiva,  $\hat{R}(h)$  es un estimador insesgado del riesgo  $R(h)$ .

Notar que el adjetivo insesgado se debe a que, en promedio, el valor de  $h$  que minimiza el riesgo  $R(h)$  es igual al que minimiza el estimador  $\hat{R}(h)$ .

*Demostración.*

$$\begin{aligned}
\mathbb{E}(\tilde{R}(h)) &= \mathbb{E}\|\mathbf{S} - \Sigma + \Sigma - \hat{\Sigma}_h\|_F^2 \\
&= \mathbb{E}\|\mathbf{S} - \Sigma\|_F^2 + \mathbb{E}\|\Sigma - \hat{\Sigma}_h\|_F^2 + 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)(\Sigma - \hat{\Sigma}_h)\right\}\right) \\
&= \mathbb{E}\|\mathbf{S} - \Sigma\|_F^2 + R(h) + 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)(\Sigma - \hat{\Sigma}_h)\right\}\right)
\end{aligned}$$

Reescribiendo el último término de la igualdad anterior

$$\begin{aligned}
2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)(\Sigma - \hat{\Sigma}_h)\right\}\right) &= 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)(\Sigma - \mathbf{G}^T\mathbf{S}\mathbf{G})\right\}\right) \\
&= 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)(\Sigma - \mathbf{G}^T\Sigma\mathbf{G} - [\mathbf{G}^T\mathbf{S}\mathbf{G} - \mathbf{G}^T\Sigma\mathbf{G}])\right\}\right) \\
&= 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)(\Sigma - \mathbf{G}^T\Sigma\mathbf{G})\right\}\right) - 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)\mathbf{G}^T(\mathbf{S} - \Sigma)\mathbf{G}\right\}\right) \\
&= 2\text{tr}\left\{\mathbb{E}\left((\mathbf{S} - \Sigma)(\Sigma - \mathbf{G}^T\Sigma\mathbf{G})\right)\right\} - 2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)\mathbf{G}^T(\mathbf{S} - \Sigma)\mathbf{G}\right\}\right) \\
&= -2\mathbb{E}\left(\text{tr}\left\{(\mathbf{S} - \Sigma)\mathbf{G}^T(\mathbf{S} - \Sigma)\mathbf{G}\right\}\right)
\end{aligned}$$

sustituyendo a  $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^T$

$$\begin{aligned}
&= -2\mathbb{E}\left(\text{tr}\left\{\left(\frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^T - \Sigma\right) \mathbf{G}^T \left(\frac{1}{n} \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j^T - \Sigma\right) \mathbf{G}\right\}\right) \\
&= -2\mathbb{E}\left(\text{tr}\left\{\frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}^T \frac{1}{n} \sum_{j=1}^n (\mathbf{y}_j \mathbf{y}_j^T - \Sigma) \mathbf{G}\right\}\right) \\
&= -\frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\left(\text{tr}\left\{(\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}^T (\mathbf{y}_j \mathbf{y}_j^T - \Sigma) \mathbf{G}\right\}\right) \\
&= -\frac{2}{n^2} \sum_{i=1}^n \text{tr}\left\{\mathbb{E}\left((\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}\right)\right\} \\
&= -\frac{2}{n} \text{tr}\left\{\mathbb{E}\left((\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}\right)\right\}
\end{aligned}$$

$$\implies \mathbb{E}(\tilde{R}(h)) = R(h) + \mathbb{E}\|\mathbf{S} - \Sigma\|_F^2 - \frac{2}{n} \text{tr}\left\{\mathbb{E}\left((\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}\right)\right\}$$

Si

$$B(h) = \frac{2}{n(n-1)} \left( \sum_{i=1}^n \text{tr}\left\{(\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}\right\} \right),$$

es un estimador insesgado de

$$\frac{2}{n} \text{tr} \{ \mathbb{E} (\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \Sigma) \mathbf{G} \},$$

se cumple lo enunciado en el teorema. Para verificar que efectivamente  $B(h)$  es insesgado, calculamos su esperanza

$$\begin{aligned} \mathbb{E}(B(h)) &= \frac{2}{n(n-1)} \mathbb{E} \left( \sum_{i=1}^n \text{tr} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G} \} \right) \\ &= \frac{2}{n(n-1)} \text{tr} \left( \sum_{i=1}^n \mathbb{E} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G} \} \right) \\ &= \frac{2}{n(n-1)} \text{tr} \left( \sum_{i=1}^n \mathbb{E} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \} \mathbf{G} \right) \\ &= \frac{2}{n(n-1)} \text{tr} (n \mathbb{E} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \} \mathbf{G}) \\ &= \frac{2}{(n-1)} \text{tr} (\mathbb{E} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \} \mathbf{G}) \end{aligned}$$

Sean  $\mathbf{C} = \mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}$  y  $\mathbf{e}_i = [0 \dots 1 \dots 0]$  el  $i$ -ésimo elemento de la base canónica.

$$(\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) = \mathbf{C} \mathbf{G}^T \mathbf{C} = [\mathbf{e}_w \mathbf{C} \mathbf{G}^T \mathbf{C} \mathbf{e}_j^T]_{1 \leq w, j \leq m}$$

Específicamente

$$\begin{aligned} \mathbf{e}_w \mathbf{C} \mathbf{G}^T \mathbf{C} \mathbf{e}_j^T &= [c_{w1} \ c_{w2} \ \dots \ c_{wm}] \mathbf{G}^T \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix} \\ &= [c_{w1} \ c_{w2} \ \dots \ c_{wm}] \begin{bmatrix} g_1(t_1) & \cdots & g_m(t_1) \\ \vdots & \ddots & \vdots \\ g_1(t_m) & \cdots & g_m(t_m) \end{bmatrix} \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix} \\ &= \left[ \sum_{l=1}^m c_{wl} g_1(t_l) \quad \sum_{l=1}^m c_{wl} g_2(t_l) \quad \dots \quad \sum_{l=1}^m c_{wl} g_m(t_l) \right] \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix} \\ &= \sum_{r=1}^m \sum_{l=1}^m c_{wl} g_r(t_l) c_{rj} = \sum_{r=1}^m \sum_{l=1}^m (Y_i(t_w) Y_i(t_l) - s_{wl}) g_r(t_l) (Y_i(t_r) Y_i(t_j) - s_{rj}) \end{aligned}$$


---

entonces

$$\begin{aligned}
& \mathbb{E} \left\{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}) \right\} \\
&= \mathbb{E} \left\{ \sum_{r=1}^m \sum_{l=1}^m g_r(t_l) (Y_i(t_w) Y_i(t_l) - s_{wl}) (Y_i(t_r) Y_i(t_j) - s_{rj}) \right\}_{1 \leq w, j \leq n} \\
&= \sum_{r=1}^m \sum_{l=1}^m g_r(t_l) \mathbb{E} \left\{ (Y_i(t_w) Y_i(t_l) - s_{wl}) (Y_i(t_r) Y_i(t_j) - s_{rj}) \right\}
\end{aligned}$$

Sustituyendo  $s_{ij}$  en la expresión anterior se procede a calcular la esperanza

$$\begin{aligned}
& \mathbb{E} \left( Y_i(t_w) Y_i(t_l) - \frac{1}{n} \sum_{p=1}^n Y_p(t_w) Y_p(t_l) \right) \left( Y_i(t_r) Y_i(t_j) - \frac{1}{n} \sum_{q=1}^n Y_q(t_r) Y_q(t_j) \right) \\
&= \mathbb{E} \left\{ Y_i(t_w) Y_i(t_l) Y_i(t_r) Y_i(t_j) - (Y_i(t_w) Y_i(t_l)) \frac{1}{n} \sum_{q=1}^n Y_q(t_r) Y_q(t_j) \right. \\
&\quad \left. - (Y_i(t_r) Y_i(t_j)) \frac{1}{n} \sum_{p=1}^n Y_p(t_w) Y_p(t_l) \right. \\
&\quad \left. + \frac{1}{n^2} \left( \sum_{p=1}^n Y_p(t_w) Y_p(t_l) \right) \left( \sum_{q=1}^n Y_q(t_r) Y_q(t_j) \right) \right\} \\
&= \mathbb{E} (Y_i(t_w) Y_i(t_l) Y_i(t_r) Y_i(t_j)) - \frac{1}{n} \sum_{q=1}^n \mathbb{E} (Y_q(t_r) Y_q(t_j) Y_i(t_w) Y_i(t_l)) \\
&\quad - \frac{1}{n} \sum_{p=1}^n \mathbb{E} (Y_p(t_w) Y_p(t_l) Y_i(t_r) Y_i(t_j)) \\
&\quad + \frac{1}{n^2} \mathbb{E} \left( \sum_{p=1}^n Y_p(t_w) Y_p(t_l) Y_p(t_r) Y_p(t_j) + \sum_{p \neq q} Y_p(t_w) Y_p(t_l) Y_q(t_r) Y_q(t_j) \right) \\
&= \mathbb{E} (Y_i(t_w) Y_i(t_l) Y_i(t_r) Y_i(t_j)) - \frac{2}{n} \mathbb{E} (Y_i(t_r) Y_i(t_j) Y_i(t_w) Y_i(t_l)) \\
&\quad - \frac{1}{n} \sum_{q \neq i} \mathbb{E} (Y_q(t_r) Y_q(t_j)) \mathbb{E} (Y_i(t_w) Y_i(t_l)) - \frac{1}{n} \sum_{p \neq i} \mathbb{E} (Y_p(t_w) Y_p(t_l)) \mathbb{E} (Y_i(t_r) Y_i(t_j)) \\
&\quad + \frac{1}{n^2} \sum_{p=1}^n \mathbb{E} (Y_p(t_w) Y_p(t_l) Y_p(t_r) Y_p(t_j)) + \frac{1}{n^2} \sum_{p \neq q} \mathbb{E} (Y_p(t_w) Y_p(t_l)) \mathbb{E} (Y_q(t_r) Y_q(t_j))
\end{aligned}$$


---

Como las replicas  $Y_1, \dots, Y_n$  son iid

$$\begin{aligned}
&= \mathbb{E}(Y_i(t_w)Y_i(t_l)Y_i(t_r)Y_i(t_j)) - \frac{2}{n}\mathbb{E}(Y_i(t_r)Y_i(t_j))Y_i(t_w)Y_i(t_l) \\
&- \frac{n-1}{n}\mathbb{E}(Y_q(t_r)Y_q(t_j))\mathbb{E}(Y_i(t_w)Y_i(t_l)) - \frac{n-1}{n}\mathbb{E}(Y_p(t_w)Y_p(t_l))\mathbb{E}(Y_i(t_r)Y_i(t_j)) \\
&+ \frac{n}{n^2}\mathbb{E}(Y_p(t_w)Y_p(t_l)Y_p(t_r)Y_p(t_j)) + \frac{n(n-1)}{n^2}\mathbb{E}(Y_p(t_w)Y_p(t_l))\mathbb{E}(Y_q(t_r)Y_q(t_j)) \\
&= \mathbb{E}(Y_i(t_w)Y_i(t_l)Y_i(t_r)Y_i(t_j)) - \frac{2}{n}\mathbb{E}(Y_i(t_r)Y_i(t_j))Y_i(t_w)Y_i(t_l) \\
&- \frac{n-1}{n}\sigma(t_r, t_j)\sigma(t_w, t_l) - \frac{n-1}{n}\sigma(t_w, t_l)\sigma(t_r, t_j) \\
&+ \frac{1}{n}\mathbb{E}(Y_p(t_w)Y_p(t_l)Y_p(t_r)Y_p(t_j)) + \frac{n-1}{n}\sigma(t_w, t_l)\sigma(t_r, t_j) \\
&= \left(1 - \frac{2}{n} + \frac{1}{n}\right)\mathbb{E}(Y_i(t_w)Y_i(t_l)Y_i(t_r)Y_i(t_j)) - \left(2\frac{n-1}{n} - \frac{n-1}{n}\right)\sigma(t_w, t_l)\sigma(t_r, t_j) \\
&= \left(\frac{n-1}{n}\right)\mathbb{E}(Y_i(t_w)Y_i(t_l)Y_i(t_r)Y_i(t_j)) - \left(\frac{n-1}{n}\right)\sigma(t_w, t_l)\sigma(t_r, t_j) \\
&= \frac{n-1}{n}\left(\mathbb{E}(Y_i(t_w)Y_i(t_l)Y_i(t_r)Y_i(t_j)) - \sigma(t_w, t_l)\sigma(t_r, t_j)\right)
\end{aligned}$$

Entonces

$$\mathbb{E}\left\{(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{S})\mathbf{G}^T(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{S})\right\} = \frac{n-1}{n}\mathbb{E}\left\{(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{\Sigma})\mathbf{G}^T(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{\Sigma})\right\}$$

y

$$\begin{aligned}
\mathbb{E}(B(h)) &= \frac{2}{(n-1)}tr\left(\mathbb{E}\left\{(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{S})\mathbf{G}^T(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{S})\right\}\mathbf{G}\right) \\
&= \frac{2}{n}tr\left\{\mathbb{E}\left(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{\Sigma}\right)\mathbf{G}^T\left(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{\Sigma}\right)\mathbf{G}\right\}
\end{aligned}$$

Entonces  $B(h)$  es un estimador insesgado de

$$\frac{2}{n}tr\left\{\mathbb{E}\left(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{\Sigma}\right)\mathbf{G}^T\left(\mathbf{y}_i\mathbf{y}_i^T - \mathbf{\Sigma}\right)\mathbf{G}\right\}.$$

Por lo tanto

$$\mathbb{E}(\hat{R}(h)) = R(h) + \mathbb{E}\|\mathbf{S} - \mathbf{\Sigma}\|_F^2.$$

■

El método de selección del ancho de banda  $h$  que se propone en este trabajo consiste en determinar el valor  $\hat{h}$  que minimiza a  $R(h)$ .

De manera análoga que en ([17]) el siguiente teorema establece una desigualdad que asegura que el riesgo del estimador seleccionado al minimizar  $\hat{R}(h)$ , no es demasiado grande con respecto al riesgo que se obtiene al minimizar  $R(h)$ .

**Teorema 2.3.2** (Acotación del riesgo del estimador). *Para toda  $c > 0$  se cumple:*

$$R(\hat{h}) \leq (1 + c^{-1}) \inf_{h>0} R(h) + (4 + c) \mathbb{E} \|\mathbf{S} - \boldsymbol{\Sigma}\|_F^2,$$

donde

$$\hat{h} \in \operatorname{argmin}_{h>0} \hat{R}(h),$$

y

$$R(\hat{h}) = \mathbb{E} \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}_{\hat{h}}\|_F^2$$

Notar que el término  $\mathbb{E} \|\mathbf{S} - \boldsymbol{\Sigma}\|_F^2$  es pequeño, específicamente  $\mathbb{E} \|\mathbf{S} - \boldsymbol{\Sigma}\|_F^2 = O(\frac{1}{n})$ .

*Demostración.* Como  $B(h) \geq 0$ ,

$$R(\hat{h}) = \mathbb{E} \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}_{\hat{h}}\|_F^2 \leq \mathbb{E} \|\boldsymbol{\Sigma} - \mathbf{S}\|_F^2 + \mathbb{E} \left[ \|\mathbf{S} - \hat{\boldsymbol{\Sigma}}_{\hat{h}}\|_F^2 + B(\hat{h}) \right] + 2tr \left( \left[ \mathbf{S} - \hat{\boldsymbol{\Sigma}}_{\hat{h}} \right] \left[ \boldsymbol{\Sigma} - \mathbf{S} \right] \right)$$

Sea  $h_0 \in \operatorname{argmin}_{h>0} R(h)$ . Por definición de  $\hat{h}$  se cumple

$$\|\mathbf{S} - \hat{\boldsymbol{\Sigma}}_{\hat{h}}\|_F^2 + B(\hat{h}) \leq \|\mathbf{S} - \hat{\boldsymbol{\Sigma}}_{h_0}\|_F^2 + B(h_0)$$

Entonces

$$\mathbb{E} \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}_{\hat{h}}\|_F^2 \leq \mathbb{E} \|\boldsymbol{\Sigma} - \mathbf{S}\|_F^2 + \mathbb{E} \left[ \|\mathbf{S} - \hat{\boldsymbol{\Sigma}}_{h_0}\|_F^2 + B(h_0) \right] + 2tr \left( \left[ \mathbf{S} - \hat{\boldsymbol{\Sigma}}_{\hat{h}} \right] \left[ \boldsymbol{\Sigma} - \mathbf{S} \right] \right)$$

utilizando el resultado del Teorema 2.3.1 se tiene

$$\mathbb{E} \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}_{\hat{h}}\|_F^2 \leq \mathbb{E} \|\boldsymbol{\Sigma} - \mathbf{S}\|_F^2 + \left[ \mathbb{E} \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}_{h_0}\|_F^2 + \mathbb{E} \|\boldsymbol{\Sigma} - \mathbf{S}\|_F^2 \right] + 2tr \left( \left[ \mathbf{S} - \hat{\boldsymbol{\Sigma}}_{\hat{h}} \right] \left[ \boldsymbol{\Sigma} - \mathbf{S} \right] \right)$$

Además utilizando la desigualdad de Cauchy-Schwarz (1)

$$\text{tr} \left( \left[ \mathbf{S} - \hat{\Sigma}_{\hat{h}} \right] [\Sigma - \mathbf{S}] \right) \leq \|\mathbf{S} - \hat{\Sigma}_{\hat{h}}\| \|\Sigma - \mathbf{S}\|$$

Y usando nuevamente esta desigualdad con la representación de integral

$$\mathbb{E} \left[ \|\mathbf{S} - \hat{\Sigma}_{\hat{h}}\| \|\Sigma - \mathbf{S}\| \right] \leq \sqrt{\mathbb{E} \left[ \|\mathbf{S} - \hat{\Sigma}_{\hat{h}}\|_F^2 \right]} \sqrt{\mathbb{E} \left[ \|\Sigma - \mathbf{S}\|_F^2 \right]}$$

se tiene

$$\begin{aligned} \mathbb{E} \left[ \text{tr} \left( \left[ \mathbf{S} - \hat{\Sigma}_{\hat{h}} \right] [\Sigma - \mathbf{S}] \right) \right] &\leq \sqrt{\mathbb{E} \left[ \|\mathbf{S} - \hat{\Sigma}_{\hat{h}}\|_F^2 \right]} \sqrt{\mathbb{E} \left[ \|\Sigma - \mathbf{S}\|_F^2 \right]} \\ &\leq \sqrt{\mathbb{E} \left[ \|\mathbf{S} - \hat{\Sigma}_{\hat{h}}\|_F^2 + B(\hat{h}) \right]} \sqrt{\mathbb{E} \left[ \|\Sigma - \mathbf{S}\|_F^2 \right]} \end{aligned}$$

Por los argumentos ya mencionados obtenemos

$$\begin{aligned} \mathbb{E} \left[ \text{tr} \left( \left[ \mathbf{S} - \hat{\Sigma}_{\hat{h}} \right] [\Sigma - \mathbf{S}] \right) \right] &\leq \sqrt{\mathbb{E} \|\Sigma - \hat{\Sigma}_{h_0}\|_F^2 + \mathbb{E} \|\mathbf{S} - \Sigma\|_F^2} \sqrt{\mathbb{E} \|\Sigma - \mathbf{S}\|_F^2} \\ &\leq \sqrt{\mathbb{E} \|\Sigma - \hat{\Sigma}_{h_0}\|_F^2} \sqrt{\mathbb{E} \|\Sigma - \mathbf{S}\|_F^2 + \mathbb{E} \|\mathbf{S} - \Sigma\|_F^2} \end{aligned}$$

Entonces

$$\mathbb{E} \|\Sigma - \hat{\Sigma}_{\hat{h}}\|_F^2 \leq \mathbb{E} \|\Sigma - \hat{\Sigma}_{h_0}\|_F^2 + 4\mathbb{E} \|\mathbf{S} - \Sigma\|_F^2 + 2\sqrt{\mathbb{E} \|\Sigma - \hat{\Sigma}_{h_0}\|_F^2} \sqrt{\mathbb{E} \|\Sigma - \mathbf{S}\|_F^2}$$

Con la siguiente desigualdad que se mantiene para toda  $a, b \in \mathbb{R}$  y para toda  $c > 0$

$$2ab \leq a^2 c^{-1} + b^2 c$$

se obtiene que para toda  $c > 0$ :

$$\mathbb{E} \|\Sigma - \hat{\Sigma}_{\hat{h}}\|_F^2 \leq \mathbb{E} \|\Sigma - \hat{\Sigma}_{h_0}\|_F^2 (c^{-1} + 1) + \mathbb{E} \|\Sigma - \mathbf{S}\|_F^2 (c + 4)$$

Es decir,

$$R(\hat{h}) \leq R(h_0) (c^{-1} + 1) + \mathbb{E} \|\Sigma - \mathbf{S}\|_F^2 (c + 4).$$

■

**Corolario 2.3.3** (Optimalidad Asintótica).

$$\frac{R(\hat{h})}{\inf_{h>0} R(h)} \xrightarrow[n]{} 1.$$

*Demostración.*

$$\begin{aligned} 1 \leq \frac{R(\hat{h})}{\inf_{h>0} R(h)} &\leq \frac{(1 + c_n^{-1}) \inf_{h>0} R(h) + (4 + c_n)O\left(\frac{1}{n}\right)}{\inf_{h>0} R(h)} \\ &= 1 + c_n^{-1} + 4O\left(\frac{1}{n}\right) + c_n O\left(\frac{1}{n}\right) \end{aligned}$$

Si  $\frac{c_n}{n} \rightarrow 0$  y  $c_n^{-1} \rightarrow 0$  por el teorema de estricción se cumple que

$$\frac{R(\hat{h})}{\inf_{h>0} R(h)} \xrightarrow[n]{} 1$$

Esto se cumple por ejemplo con  $c_n = \log(n)$ . ■

Entonces se ha obtenido un estimador de la función de covarianza  $\sigma$  mediante un procedimiento de selección del ancho de banda que tiene buen comportamiento para muestras finitas (ver Teorema 2.3.2) y es asintóticamente óptimo (ver Corolario 2.3.3).

Hasta el momento se ha hecho el supuesto de que la media del campo es cero,  $\mathbb{E}(Y(t)) = 0$ , es equivalente a que la media del campo  $\mu(t)$  es conocida. Sin embargo, en la práctica comúnmente la media es desconocida. Una forma simple para resolver esta dificultad es utilizar el criterio (2.8) sustituyendo la matriz  $\mathbf{S}$  por la matriz de covarianza muestral de los datos centrados por su media muestral  $\tilde{\mathbf{S}}$ .

El siguiente resultado muestra que esta forma de proceder es equivalente a la que se seguía cuando la media  $\mu(t)$  se conocía salvo un error  $O_p\left(\frac{1}{n}\right)$

---



**Teorema 2.3.4** (Error por corrección). *Al corregir los datos por la media muestral se cumple que*

$$R^*(h) = \hat{R}(h) + O_p\left(\frac{1}{n}\right),$$

donde

$$\hat{R}(h) = \|\mathbf{S}(\mu) - \mathbf{G}^T \mathbf{S}(\mu) \mathbf{G}\|_F^2 + \frac{2}{n(n-1)} \sum_{i=1}^n \text{tr} \left\{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G} \right\},$$

$$\mathbf{S}(\mu) = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^T,$$

$$R^*(h) = \|\tilde{\mathbf{S}} - \mathbf{G}^T \tilde{\mathbf{S}} \mathbf{G}\|_F^2$$

$$+ \frac{2}{n(n-1)} \sum_{i=1}^n \text{tr} \left\{ \left( (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G}^T \left( (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G} \right\},$$

$$\tilde{\mathbf{S}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T,$$

y

$$\bar{\mathbf{y}} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n Y_i(t_1) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n Y_i(t_m) \end{bmatrix}.$$

*Demostración.* Primero se mostrará que se cumple

$$\tilde{\mathbf{S}} = \mathbf{S}(\mu) + O_p\left(\frac{1}{n}\right).$$

$$\begin{aligned}
\mathbf{S}(\boldsymbol{\mu}) &= \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^T \\
&= \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \boldsymbol{\mu})(\mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \boldsymbol{\mu})^T \\
&= \frac{1}{n} \sum_{i=1}^n [(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T + (\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T \\
&\quad + (\mathbf{y}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T + (\bar{\mathbf{y}} - \boldsymbol{\mu})(\mathbf{y}_i - \bar{\mathbf{y}})^T] \\
&= \frac{n-1}{n} \tilde{\mathbf{S}} + \frac{1}{n} \sum_{i=1}^n (\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T \\
&\quad + \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T + (\bar{\mathbf{y}} - \boldsymbol{\mu}) \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^T \\
&= \tilde{\mathbf{S}} + \frac{1}{n} \tilde{\mathbf{S}} + (\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T
\end{aligned}$$

Recordemos que  $\tilde{\mathbf{S}} = O_p(1)$  y  $(\bar{\mathbf{y}} - \boldsymbol{\mu}) = O_p\left(\frac{1}{\sqrt{n}}\right)$ . Entonces

$$\mathbf{S}(\boldsymbol{\mu}) = \tilde{\mathbf{S}} + O_p\left(\frac{1}{n}\right).$$

A continuación, partiendo de la definición de  $R^*(h)$ , se desarrollaran sus componentes

$$\begin{aligned}
R^*(h) &= \|\tilde{\mathbf{S}} - \mathbf{G}^T \tilde{\mathbf{S}} \mathbf{G}\|_F^2 \\
&\quad + \frac{2}{n(n-1)} \sum_{i=1}^n \text{tr} \left\{ \left( (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G}^T \left( (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G} \right\},
\end{aligned}$$


---

Por un lado

$$\begin{aligned}
\|\tilde{\mathbf{S}} - \mathbf{G}^T \tilde{\mathbf{S}} \mathbf{G}\|_F^2 &= \left\| \mathbf{S}(\mu) + O_p\left(\frac{1}{n}\right) - \mathbf{G}^T \left( \mathbf{S}(\mu) + O_p\left(\frac{1}{n}\right) \right) \mathbf{G} \right\|_F^2 \\
&= \left\| \mathbf{S}(\mu) - \mathbf{G}^T \mathbf{S}(\mu) \mathbf{G} + O_p\left(\frac{1}{n}\right) - \mathbf{G}^T O_p\left(\frac{1}{n}\right) \mathbf{G} \right\|_F^2 \\
&= \|\mathbf{S}(\mu) - \mathbf{G}^T \mathbf{S}(\mu) \mathbf{G}\|_F^2 + \left\| O_p\left(\frac{1}{n}\right) - \mathbf{G}^T O_p\left(\frac{1}{n}\right) \mathbf{G} \right\|_F^2 \\
&\quad + 2tr \left\{ \left( \mathbf{S}(\mu) - \mathbf{G}^T \mathbf{S}(\mu) \mathbf{G} \right) \left( O_p\left(\frac{1}{n}\right) - \mathbf{G}^T O_p\left(\frac{1}{n}\right) \mathbf{G} \right) \right\} \\
&= \|\mathbf{S}(\mu) - \mathbf{G}^T \mathbf{S}(\mu) \mathbf{G}\|_F^2 + O_p\left(\frac{1}{n^2}\right) + O_p\left(\frac{1}{n}\right)
\end{aligned}$$

Por otro lado

$$\begin{aligned}
&\sum_{i=1}^n tr \left\{ \left( (\mathbf{y}_i - \bar{\mathbf{y}}) (\mathbf{y}_i - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G}^T \left( (\mathbf{y}_i - \bar{\mathbf{y}}) (\mathbf{y}_i - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G} \right\} \\
&= \sum_{i=1}^n tr \left\{ \left( (\mathbf{y}_i - \mu + \mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu + \mu - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G}^T \right. \\
&\quad \left. \left( (\mathbf{y}_i - \mu + \mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu + \mu - \bar{\mathbf{y}})^T - \tilde{\mathbf{S}} \right) \mathbf{G} \right\} \\
&= \sum_{i=1}^n tr \left\{ \left[ (\mathbf{y}_i - \mu) (\mathbf{y}_i - \mu)^T - \tilde{\mathbf{S}} + (\mu - \bar{\mathbf{y}}) (\mu - \bar{\mathbf{y}})^T + (\mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu)^T + (\mathbf{y}_i - \mu) (\mu - \bar{\mathbf{y}})^T \right] \right. \\
&\quad \left. \mathbf{G}^T \left[ (\mathbf{y}_i - \mu) (\mathbf{y}_i - \mu)^T - \tilde{\mathbf{S}} + (\mu - \bar{\mathbf{y}}) (\mu - \bar{\mathbf{y}})^T + (\mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu)^T + (\mathbf{y}_i - \mu) (\mu - \bar{\mathbf{y}})^T \right] \mathbf{G} \right\}
\end{aligned}$$

Desarrollando, reagrupando los productos matriciales y sustituyendo a  $\tilde{\mathbf{S}}$  por  $\mathbf{S}(\mu) + O_p\left(\frac{1}{n}\right)$  se obtiene

$$\begin{aligned}
&= \frac{2}{n(n-1)} \sum_{i=1}^n tr \left\{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G} \right. \\
&\quad - \left[ (\mathbf{y}_i - \mu) (\mathbf{y}_i - \mu)^T - \tilde{\mathbf{S}} \right] \mathbf{G}^T \left[ (\mu - \bar{\mathbf{y}}) (\mu - \bar{\mathbf{y}})^T + (\mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu)^T + (\mathbf{y}_i - \mu) (\mu - \bar{\mathbf{y}})^T \right] \mathbf{G} \\
&\quad - \left[ (\mu - \bar{\mathbf{y}}) (\mu - \bar{\mathbf{y}})^T + (\mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu)^T + (\mathbf{y}_i - \mu) (\mu - \bar{\mathbf{y}})^T \right] \mathbf{G}^T \left[ (\mathbf{y}_i - \mu) (\mathbf{y}_i - \mu)^T - \tilde{\mathbf{S}} \right] \mathbf{G} \\
&\quad + \left[ (\mu - \bar{\mathbf{y}}) (\mu - \bar{\mathbf{y}})^T + (\mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu)^T + (\mathbf{y}_i - \mu) (\mu - \bar{\mathbf{y}})^T \right] \mathbf{G}^T \\
&\quad \left. \left[ (\mu - \bar{\mathbf{y}}) (\mu - \bar{\mathbf{y}})^T + (\mu - \bar{\mathbf{y}}) (\mathbf{y}_i - \mu)^T + (\mathbf{y}_i - \mu) (\mu - \bar{\mathbf{y}})^T \right] \mathbf{G} \right\} \\
&= \frac{2}{n(n-1)} \sum_{i=1}^n tr \left\{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G} \right\} + O_p\left(\frac{1}{n}\right)
\end{aligned}$$


---

Entonces

$$\begin{aligned} R^*(h) &= \|\mathbf{S}(\mu) - \mathbf{G}^T \mathbf{S}(\mu) \mathbf{G}\|_F^2 + \frac{2}{n(n-1)} \sum_{i=1}^n \text{tr} \{ (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G}^T (\mathbf{y}_i \mathbf{y}_i^T - \mathbf{S}(\mu)) \mathbf{G} \} \\ &\quad + O_p\left(\frac{1}{n^2}\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{n}\right) \\ &= \hat{R}(h) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{n^2}\right) \end{aligned}$$

Por lo tanto

$$R^*(h) = \hat{R}(h) + O_p\left(\frac{1}{n}\right).$$

■

---

---

# CAPÍTULO 3

---

## Simulaciones

En este capítulo se presentan dos escenarios en los que, a través de simulaciones, se analiza el comportamiento del estimador insesgado  $\hat{R}(h)$  de la función de riesgo  $R(h)$ , así como el comportamiento del estimador  $\hat{\sigma}_h$  de la función de covarianza  $\sigma$ . Además, se propone un procedimiento para comparar la eficiencia del estimador propuesto con la de un estimador alternativo, frecuentemente utilizado en la literatura de Análisis de Datos Funcionales.

### 3.1. Campo con función de covarianza estacionaria

El proceso estocástico  $(Y(t))_{t \in T}$  que se considera en este escenario es un campo Gaussiano con media  $\mathbb{E}(Y(t)) = 0$ ,  $T = [0, 1]$  y función de covarianza estacionaria

$$\sigma(s, t) = \text{cov}(Y(s), Y(t)) = \exp\left(-\left(\frac{t-s}{a}\right)^2\right).$$

Se generan datos normales multivariados independientes  $\mathbf{y}_i \sim N_m(\mathbf{0}, \Sigma)$ ,  $i = 1, \dots, n$ , donde  $\Sigma = (\sigma(t_I, t_J))_{1 \leq I, J \leq m}$ , sobre la rejilla uniforme  $\{t_1, t_2, \dots, t_m\}$ ,  $t_i \in T$ ,  $m =$

15.

En las estimaciones se utiliza el núcleo Gaussiano  $K_h(t) = \exp\left(-\frac{1}{h}t^2\right)$ .

En el caso Gaussiano se puede obtener una expresión explícita para la función de riesgo  $R(h)$  como se muestra en la siguiente proposición.

**Proposición 3.1.0.1.** *Sea  $(Y(t))_{t \in T}$ ,  $T = [0, 1]$ , un campo Gaussiano con media  $\mathbb{E}(Y(t)) = 0$  y función de covarianza  $\sigma$ . Entonces la función de riesgo  $R(h)$  del estimador  $\hat{\Sigma}_h$  (ver ecuaciones (2.7) y (2.3)) tiene la forma:*

$$R(h) = \mathbb{E}\|\Sigma\|_F^2 - 2tr(\Sigma \mathbf{G}^T \Sigma \mathbf{G}) + \left[ \sum_{r=1}^m \sum_{l=1}^m a_{lr} \mathbb{E}(s_{wl} s_{rj}) \right]_{1 \leq w, j \leq m}, \quad (3.1)$$

donde  $a_{lr}$  esta dada por  $A = (a_{lr}) = \mathbf{G} \mathbf{G}^T$ , y

$$\begin{aligned} \mathbb{E}(s_{wl} s_{rj}) &= \frac{n}{(n-1)^2} [\sigma(t_w, t_l) \sigma(t_r, t_j) + \sigma(t_w, t_r) \sigma(t_l, t_j) + \sigma(t_w, t_j) \sigma(t_l, t_r)] \\ &\quad + \frac{n}{n-1} (\sigma(t_w, t_l) \sigma(t_r, t_j)). \end{aligned}$$

*Demostración.*

$$\begin{aligned} R(h) &= \mathbb{E}\|\Sigma - \hat{\Sigma}_h\|_F^2 = \mathbb{E}\|\Sigma\|_F^2 - 2tr(\Sigma \mathbb{E}(\hat{\Sigma}_h)) + \mathbb{E}\|\hat{\Sigma}_h\|_F^2 \\ &= \mathbb{E}\|\Sigma\|_F^2 - 2tr(\Sigma \mathbf{G}^T \Sigma \mathbf{G}) + \mathbb{E}\|\hat{\Sigma}_h\|_F^2. \end{aligned}$$

Explicitamente,

$$\begin{aligned} \mathbb{E}\|\hat{\Sigma}_h\|_F^2 &= \mathbb{E}\|\mathbf{G}^T \mathbf{S} \mathbf{G}\|_F^2 = \mathbb{E}(tr\{\mathbf{G}^T \mathbf{S} \mathbf{G} \mathbf{G}^T \mathbf{S} \mathbf{G}\}) \\ &= \mathbb{E}(tr\{\mathbf{S} \mathbf{G} \mathbf{G}^T \mathbf{S} \mathbf{G} \mathbf{G}^T\}) = tr(\mathbb{E}\{\mathbf{S} \mathbf{G} \mathbf{G}^T \mathbf{S}\} \mathbf{G} \mathbf{G}^T). \end{aligned}$$

Por otro lado,

$$\mathbf{S} \mathbf{A} \mathbf{S} = [e_w \mathbf{S} \mathbf{A} \mathbf{S} e_j^T]_{1 \leq w, j \leq m},$$

donde

$$\begin{aligned}
\mathbf{e}_w \mathbf{S} \mathbf{A} \mathbf{S} \mathbf{e}_j^T &= [s_{w1} \ s_{w2} \ \dots \ s_{wm}] \mathbf{A} \begin{bmatrix} s_{1j} \\ \vdots \\ s_{mj} \end{bmatrix} \\
&= [s_{w1} \ s_{w2} \ \dots \ s_{wm}] \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} \end{bmatrix} \begin{bmatrix} s_{1j} \\ \vdots \\ s_{mj} \end{bmatrix} \\
&= \left[ \sum_{l=1}^m s_{wl} a_{l1} \quad \sum_{l=1}^m s_{wl} a_{l2} \quad \dots \quad \sum_{l=1}^m s_{wl} a_{lm} \right] \begin{bmatrix} s_{1j} \\ \vdots \\ s_{mj} \end{bmatrix} = \sum_{r=1}^m \sum_{l=1}^m s_{wl} a_{lr} s_{rj}.
\end{aligned}$$

Por tanto,

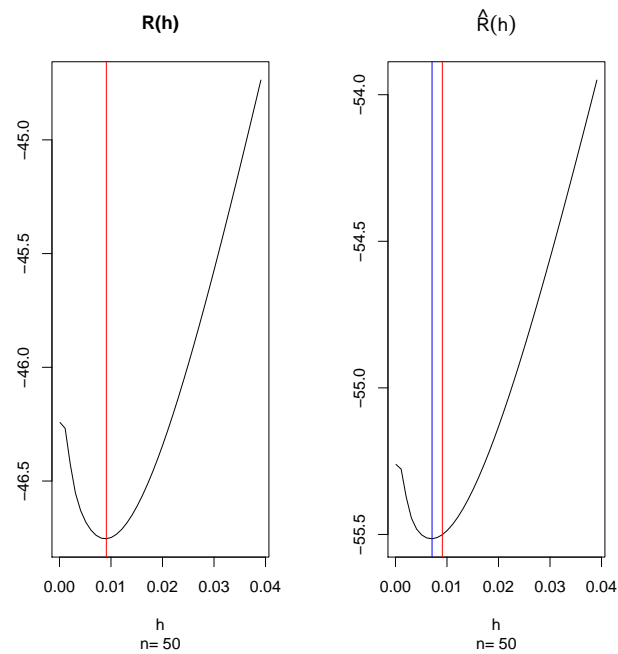
$$\mathbb{E}[\mathbf{S} \mathbf{A} \mathbf{S}] = \left[ \mathbb{E} \left( \sum_{r=1}^m \sum_{l=1}^m s_{wl} a_{lr} s_{rj} \right) \right]_{1 \leq w, j \leq m} = \left[ \sum_{r=1}^m \sum_{l=1}^m a_{lr} \mathbb{E}(s_{wl} s_{rj}) \right]_{1 \leq w, j \leq m},$$

donde utilizando las conocidas fórmulas de momentos de la distribución Normal ([1]), se obtiene que

$$\begin{aligned}
\mathbb{E}(s_{wl} s_{rj}) &= \frac{1}{(n-1)^2} \mathbb{E} \left( \sum_{p=1}^n Y_p(t_w) Y_p(t_l) \sum_{q=1}^n Y_q(t_r) Y_q(t_j) \right) \\
&= \frac{n}{(n-1)^2} \mathbb{E}(Y_p(t_w) Y_p(t_l) Y_p(t_r) Y_p(t_j)) + \frac{n}{n-1} \mathbb{E}(Y_p(t_w) Y_p(t_l)) \mathbb{E}(Y_q(t_r) Y_q(t_j)) \\
&= \frac{n}{(n-1)^2} [\sigma(t_w, t_l) \sigma(t_r, t_j) + \sigma(t_w, t_r) \sigma(t_l, t_j) + \sigma(t_w, t_j) \sigma(t_l, t_r)] \\
&\quad + \frac{n}{n-1} (\sigma(t_w, t_l) \sigma(t_r, t_j)).
\end{aligned}$$

■

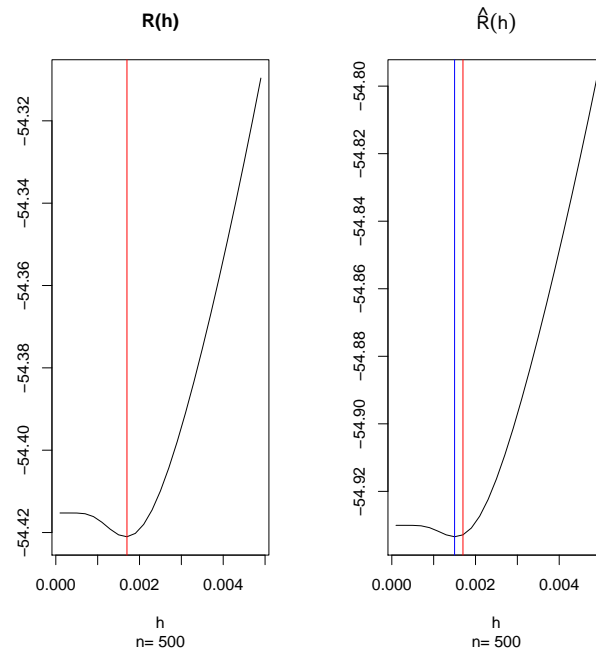
El siguiente gráfico muestra la comparación entre el Riesgo (Teórico)  $R(h)$  dado por (2.7) y su Estimador Insesgado  $\hat{R}(h)$  dado por (2.8) para un tamaño de muestra moderado  $n = 50$ .



**Figura 3.1:** Comparación del Riesgo Teórico vs. su Estimador Insesgado, considerando  $n = 50$ ,  $m = 15$  y  $a = 0.3$ . La línea roja indica el valor  $h_0$  de  $h$  que minimiza  $R(h)$ , y la línea azul indica el valor  $\hat{h}$  que minimiza  $\hat{R}(h)$ .

Un comportamiento similar se obtiene con tamaños de muestra mayores, como se observa en el siguiente gráfico

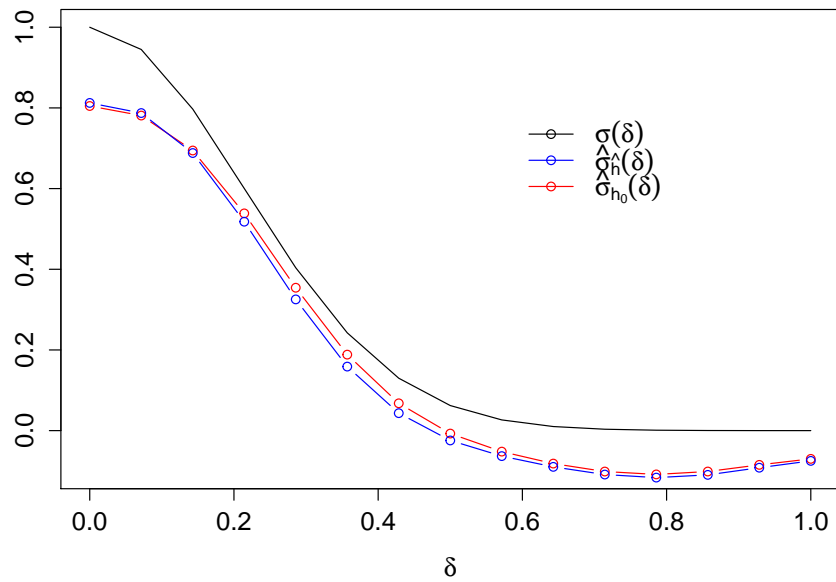




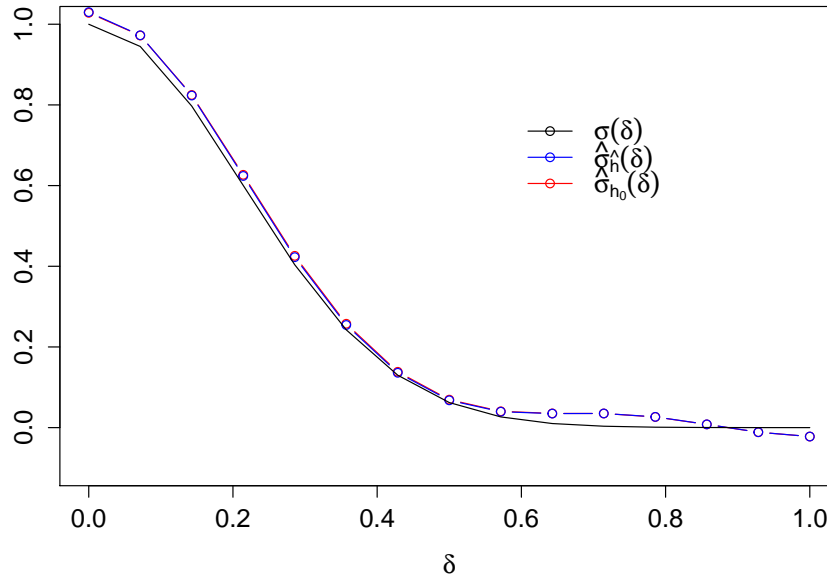
**Figura 3.2:** Comparación del Riesgo Teórico vs. su Estimador Insesgado, considerando  $n = 500$ ,  $m = 15$  y  $a = 0.3$ . La línea roja indica el valor  $h_0$  de  $h$  que minimiza  $R(h)$ , y la línea azul indica el valor  $\hat{h}$  que minimiza  $\hat{R}(h)$ .

Las figuras 3.1 y 3.2 muestran que el estimador  $\hat{R}(h)$  tiene un buen desempeño, es decir, que el valor de  $h$  que minimiza este riesgo es muy cercano al que minimiza el riesgo teórico  $R(h)$ . Por lo tanto, el procedimiento de selección de  $h$  es adecuado, en el sentido de que se comporta como si el verdadero riesgo estuviera a la mano.

El siguiente gráfico muestra la función de covarianza teórica y sus estimadores por núcleo utilizando los anchos de banda  $\hat{h}$  y  $h_0$ .



**Figura 3.3:** Comparación de la función de covarianza teórica vs. sus estimaciones con los anchos de banda  $\hat{h}$  y  $h_0$  ( $n = 50$ ,  $m = 15$  y  $a = 0.3$ ). La línea negra indica la función de covarianza teórica, y las líneas azul y roja indican sus estimaciones utilizando los anchos de banda  $\hat{h}$  y  $h_0$ .



**Figura 3.4:** Comparación de la función de covarianza teórica vs. sus estimaciones con los anchos de banda  $\hat{h}$  y  $h_0$  ( $n = 500$ ,  $m = 15$  y  $a = 0.3$ ). La línea negra indica la función de covarianza teórica, y las líneas azul y roja indican sus estimaciones utilizando los anchos de banda  $\hat{h}$  y  $h_0$ .

De las figuras 3.3 y 3.4 podemos ver que  $\hat{\sigma}_h$  estima adecuadamente a  $\sigma$ , incluso cuando el tamaño de muestra  $n = 50$  es moderado.

Resultados análogos a los mostrados en las figuras anteriores se obtienen al cambiar los parámetros  $a$  y  $m$ , pero siempre considerando que  $m \leq n$ .

## 3.2. Campo con función de covarianza no estacionaria

El proceso estocástico  $(Y(t))_{t \in T}$  que se considera en este escenario es un campo Browniano, *i.e.*, Gaussiano con media  $\mathbb{E}(Y(t)) = 0$ ,  $T = [0, 1]$  y función de covarianza

no estacionaria

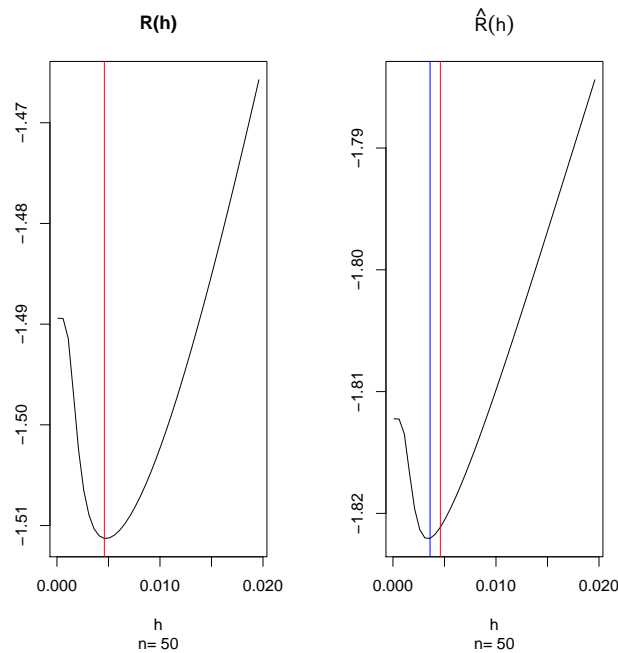
$$\sigma(s, t) = \text{cov}(Y(s), Y(t)) = \min(s, t) - st.$$

Igual que en el ejemplo anterior se generan datos normales multivariados independientes  $\mathbf{y}_i \sim N_m(\mathbf{0}, \Sigma)$ ,  $i = 1, \dots, n$ , donde  $\Sigma = (\sigma(t_I, t_J))_{1 \leq I, J \leq m}$ , sobre la rejilla uniforme  $\{t_1, t_2, \dots, t_m\}$ ,  $t_i \in T$ ,  $m = 15$ .

En las estimaciones se utiliza el núcleo Gaussiano  $K_h(t) = \exp(-\frac{1}{h}t^2)$ .

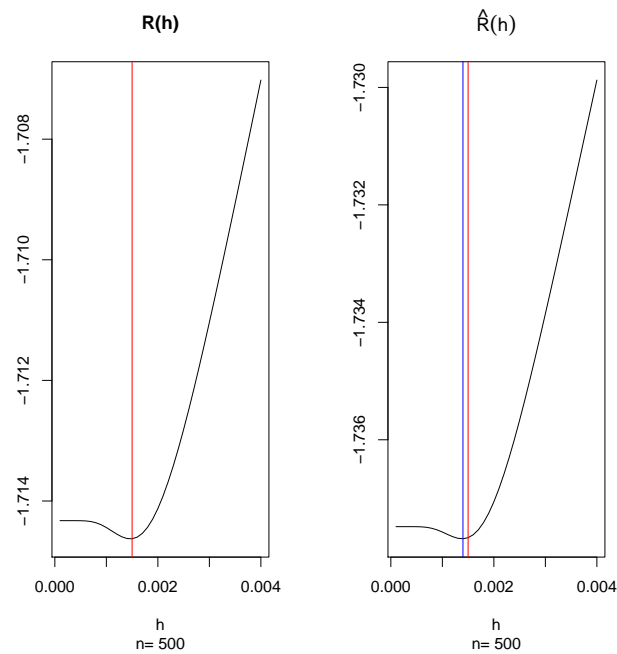
Debido que este campo es Gaussiano, el riesgo teórico  $R(h)$  esta dado por (3.1).

El siguiente gráfico muestra la comparación entre el Riesgo (Teórico)  $R(h)$  dado por (2.7) y su Estimador Insesgado  $\hat{R}(h)$  dado por (2.8) para un tamaño de muestra moderado  $n = 50$ .



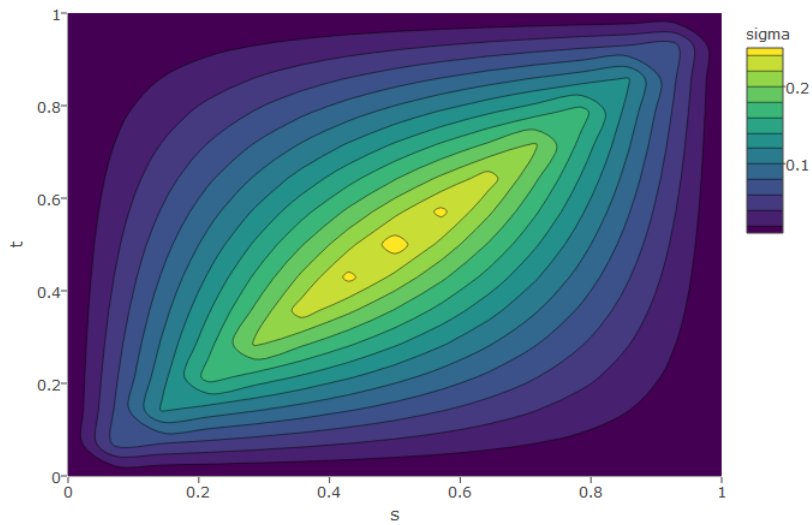
**Figura 3.5:** Comparación del Riesgo Teórico vs. su Estimador Insesgado, considerando  $n = 50$ ,  $m = 15$  y  $a = 0.3$ . La línea roja indica el valor  $h_0$  de  $h$  que minimiza  $R(h)$ , y la línea azul indica el valor  $\hat{h}$  que minimiza  $\hat{R}(h)$ .

Análogamente al ejemplo anterior, se tiene un comportamiento similar con tamaños de muestra mayores, como se observa en el siguiente gráfico.

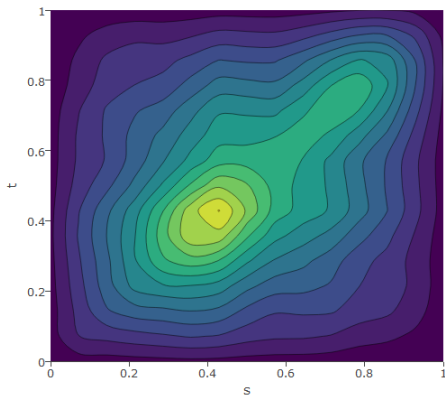


**Figura 3.6:** Comparación del Riesgo Teórico vs. su Estimador Insesgado, considerando  $n = 500$ ,  $m = 15$  y  $a = 0.3$ . La línea roja indica el valor  $h_0$  de  $h$  que minimiza  $R(h)$ , y la línea azul indica el valor  $\hat{h}$  que minimiza  $\hat{R}(h)$ .

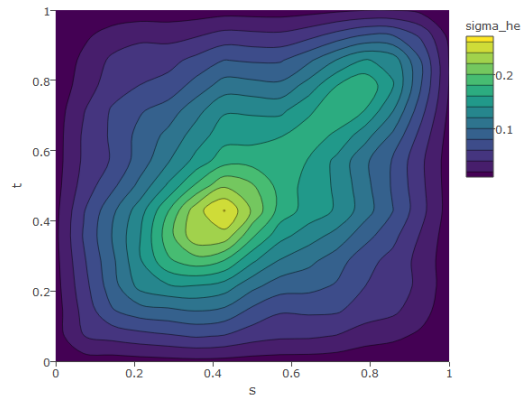
En las siguientes gráficos, Figuras 3.7-3.10, se muestra la función de covarianza teórica y sus estimadores por núcleo utilizando los anchos de banda  $\hat{h}$  y  $h_0$ , para  $n = 50, 500$ .



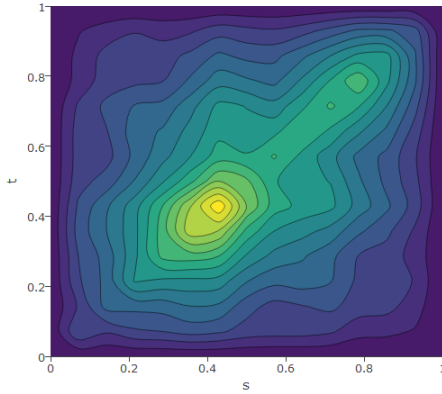
**Figura 3.7:** Función de covarianza teórica evaluada en  $m = 15$  puntos.



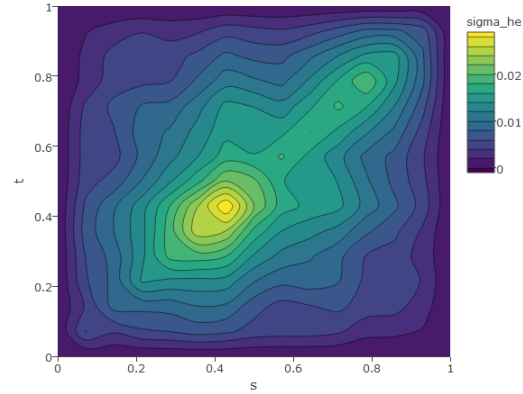
**Figura 3.8:** Estimador de función de covarianza con  $h_0$ , considerando  $n = 50$  y  $m = 15$ .



**Figura 3.9:** Estimador de función de covarianza con  $\hat{h}$ , considerando  $n = 50$  y  $m = 15$ .



**Figura 3.10:** Estimador de función de covarianza con  $h_0$ , considerando  $n = 500$  y  $m = 15$ .



**Figura 3.11:** Estimador de función de covarianza con  $\hat{h}$ , considerando  $n = 500$  y  $m = 15$ .

### 3.3. Comparación con procedimientos alternativos

Hasta ahora hemos visto que con el método propuesto en esta tesis, la estimación de una función de covarianza  $\sigma$  muestra buen comportamiento; pero podría surgir la duda de si tal estimación es mejor comparada con procedimientos alternativos de uso común en la literatura de datos funcionales.

#### 3.3.1. Suavizando los datos con un valor de $h$

Un método dentro del enfoque de estos últimos procedimientos se puede describir de la siguiente manera. En primer lugar, a partir de la observación discreta  $y_i$  de cada replica del campo se construye una aproximación (suavizamiento)  $\tilde{y}_i$  de la curva  $y_i$  mediante núcleo, pero utilizando el mismo valor  $h_{cv}$  de  $h$  en todas las replicas. Esto permite obtener una estimación de la media del campo mediante la media muestral de las aproximaciones:

$$\tilde{\mu}(t) = \frac{1}{n} \sum_{i=1}^n \tilde{y}_i. \quad (3.2)$$

El valor  $h_{cv}$  se determina minimizando el siguiente criterio CV ( de validación cruzada):

$$h_{cv} \in \underset{h>0}{\operatorname{argmin}} CV(h)$$

donde

$$CV(h) = \sum_{i=1}^n \|\mathbf{y}_i - \tilde{\mathbf{y}}_i(-i)\|_F^2,$$

y

$$\tilde{\mathbf{y}}_i(-i) = \left( \sum_{j \neq i} Y_i(t_j) \frac{K_h(t_r - t_j)}{\sum_{l \neq i} K_h(t_r - t_l)} \right)_{1 \leq r \leq m}.$$

Esta estimación de la media es similar a la desarrollada en [28], a diferencia de que se utiliza núcleo en lugar de spline.

Finalmente, el estimador de la función de covarianza se obtiene calculando la función de covarianza empírica de las curvas aproximadas  $\tilde{\mathbf{y}}_i$  con corrección por (3.2). Esto equivale a tomar como estimador de  $\Sigma$  a  $\hat{\Sigma}_{h_{cv}}$ .

Para comparar las estimaciones  $\hat{\Sigma}_{\hat{h}}$  y  $\hat{\Sigma}_{h_{cv}}$  se aproximan sus funciones de riesgo de la siguiente manera:

- i) Se generan  $B$  muestras de tamaño  $n$  del campo  $\mathbf{y}_1^b, \dots, \mathbf{y}_n^b$  ( $b = 1, \dots, B$ ).
- ii) Para cada muestra se calculan las estimaciones  $\hat{\Sigma}_{\hat{h}^b}$ ,  $\hat{\Sigma}_{h_{cv}^b}$  y se calculan  $R^z$  y  $R^{cv}$ , donde

$$R^z = \frac{1}{B} \sum_{b=1}^B \|\Sigma - \hat{\Sigma}_{\hat{h}^b}\|_F^2.$$

y

$$R^{cv} = \frac{1}{B} \sum_{b=1}^B \|\Sigma - \hat{\Sigma}_{h_{cv}^b}\|_F^2,$$

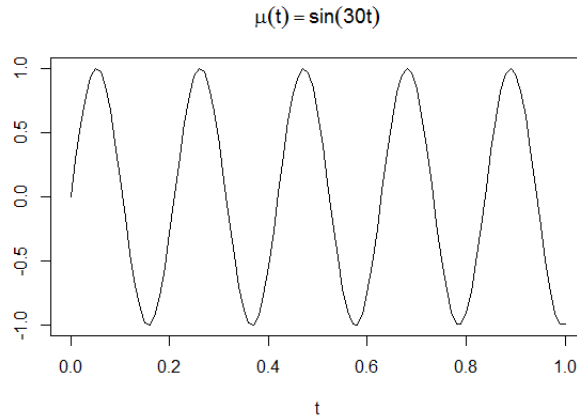
Se considera que  $Y(t) = \mu(t) + e(t)$ , donde  $t \in T = [0, 1]$ ,  $\mu(t) = \sin(30t)$  y  $e(t)$  es un campo Gaussiano con media cero y función de covarianza  $\sigma(s, t) = \exp\left(-\left(\frac{t-s}{0.3}\right)^2\right)$ .

---

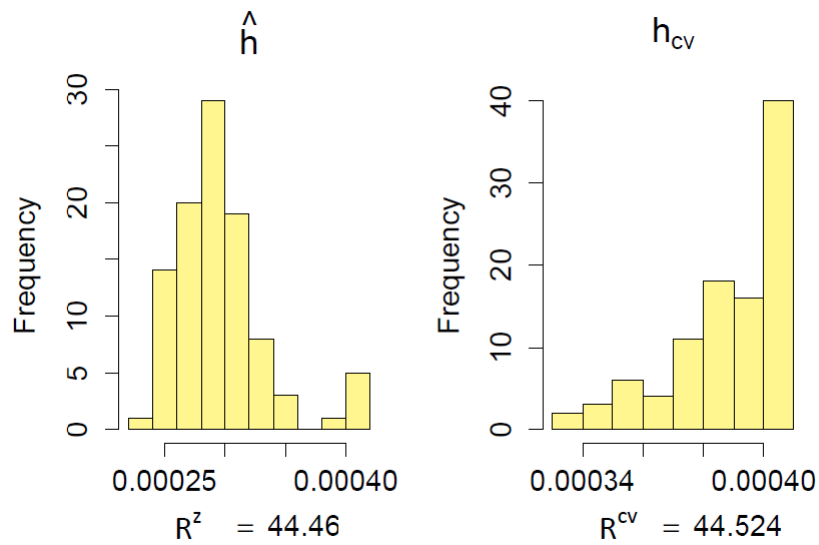


El vector de observaciones discretas  $y \sim N_m(\mu(t), \Sigma)$  en cada réplica del campo se toman sobre una rejilla uniforme de  $m = 50$ . Aquí  $\Sigma = (\sigma(s, t))$ .

En el siguiente gráfico se muestra la función media  $\mu(t)$ .

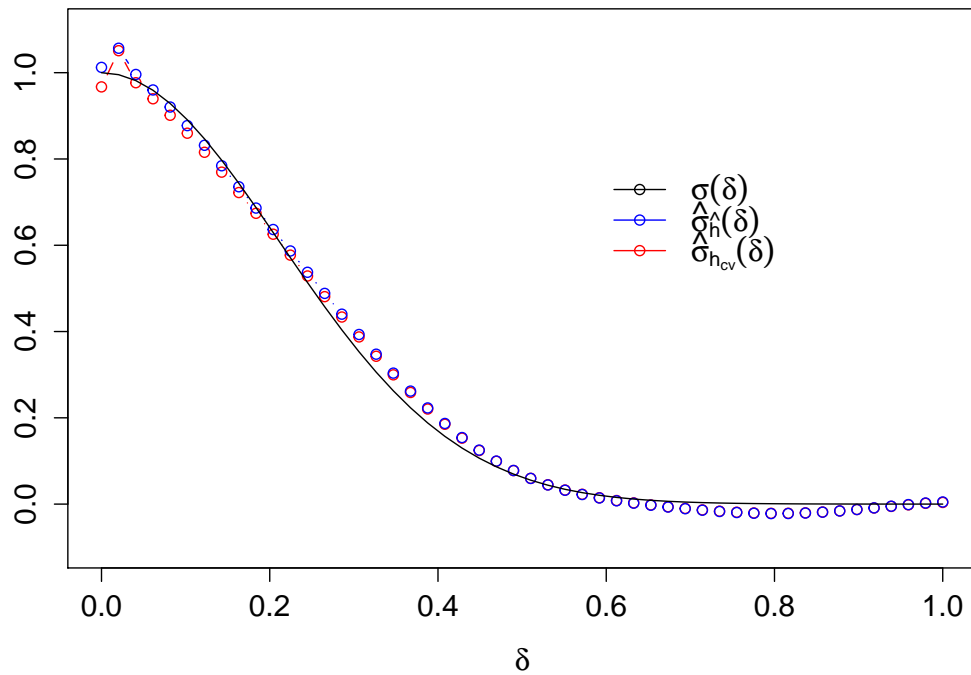


El siguiente gráfico muestra los histogramas de los valores de  $h_{cv}^b$  y  $\hat{h}^b$  para un tamaño de muestra  $n = 70$ , utilizando  $B = 100$  muestras. También se muestran los valores del riesgo de cada método.



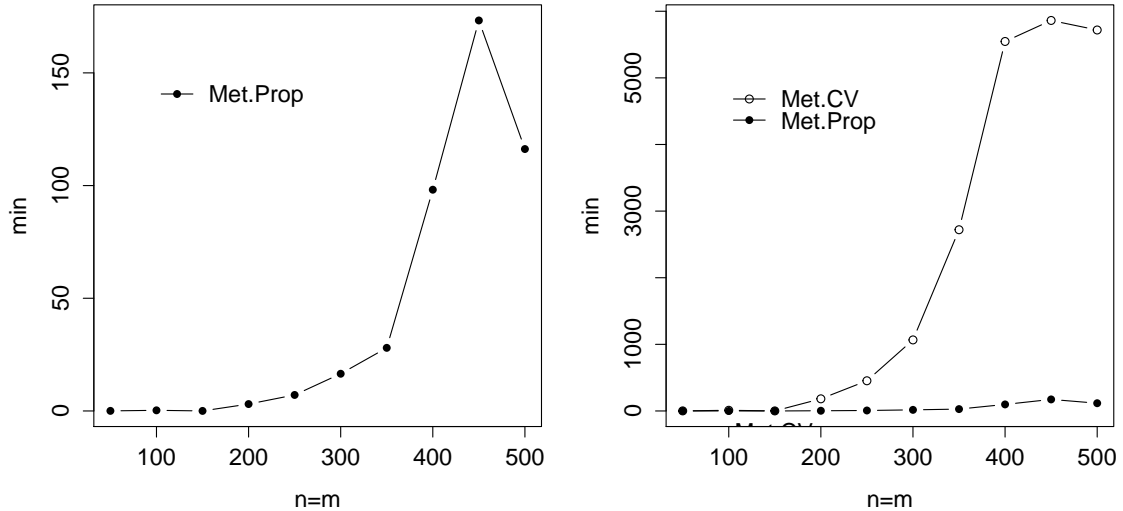
Se obtiene que  $\bar{h} = 0.00029041$  y  $\bar{h}_{cv} = 0.00038371$ .

En la siguiente gráfica se comparan las estimaciones de la función de covarianza  $\sigma$  con los anchos de banda promedios  $\bar{h}$  y  $\bar{h}_{cv}$ .



**Figura 3.12:** Curvas de función de covarianza teórica  $\sigma$  y sus estimaciones  $\hat{\sigma}_{\bar{h}}$  y  $\hat{\sigma}_{\bar{h}_{cv}}$

En la siguiente gráfica se compara el tiempo que tarda en implementarse el método propuesto en comparación con el alternativo para distintos tamaños de muestra y  $n = m$ .



Se puede observar que en este ejemplo resulta un riesgo ligeramente menor utilizando el método propuesto en comparación con el alternativo. Sin embargo, el método propuesto tiene un tiempo de implementación considerablemente menor (para cada tamaño de muestra) con respecto al alternativo. Este es un resultado parcial, ya que en otros escenarios pudiesen haber diferencias más evidentes. En general, el método alternativo no determina  $h$  tomando en cuenta la función covarianza sino la función de media; esto pudiese conducir a valores de  $h$  inadecuados en ciertos escenarios donde los grados de suavidad de la media y la covarianza sean muy distintos. Esto amerita más estudios en futuro.

### 3.3.2. Suavizando los datos con distintos valores de $h$

Otro método común en la literatura de datos funcionales se describe de la siguiente manera. En primer lugar, para cada observación discreta  $y_i$  de cada replica del campo se construye una aproximación (suavizamiento)  $\tilde{y}_i$  de la curva  $y_i$  mediante núcleo, utilizando un  $h_i$  obtenido por

$$h_i \in \underset{h>0}{\operatorname{argmin}} CV(h_i)$$

donde

$$CV(h_i) = \sum_{j=1}^m \left( Y_i(t_j) - \tilde{Y}_i(-j) \right)^2,$$

y

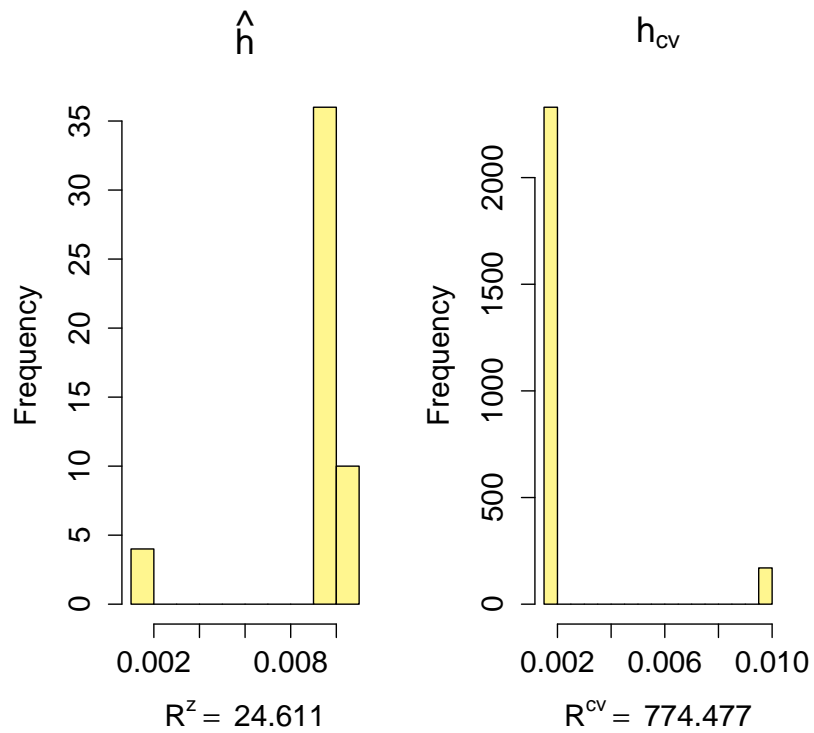
$$\tilde{Y}_i(-j) = \left( \sum_{l \neq j} Y_i(t_l) \frac{K_h(t_j - t_l)}{\sum_{r \neq j} K_h(t_j - t_r)} \right)_{1 \leq r \leq m}.$$

El estimador de la función de covarianza se obtiene calculando la función de covarianza empírica de las curvas aproximadas  $\tilde{y}_i$  con el respectivo  $h_i$ .

Como en la subsección anterior, se considera que  $Y(t) = \mu(t) + e(t)$ , donde  $t \in T = [0, 1]$ ,  $\mu(t) = \sin(30t)$  y  $e(t)$  es un campo Gaussiano con media cero y función de covarianza  $\sigma(s, t) = \exp\left(-\left(\frac{t-s}{0.3}\right)^2\right)$ .

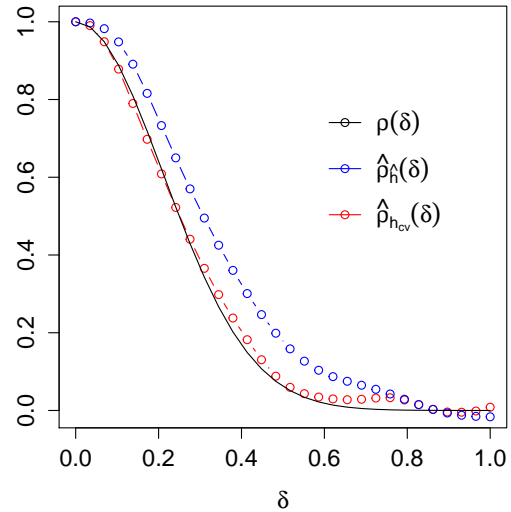
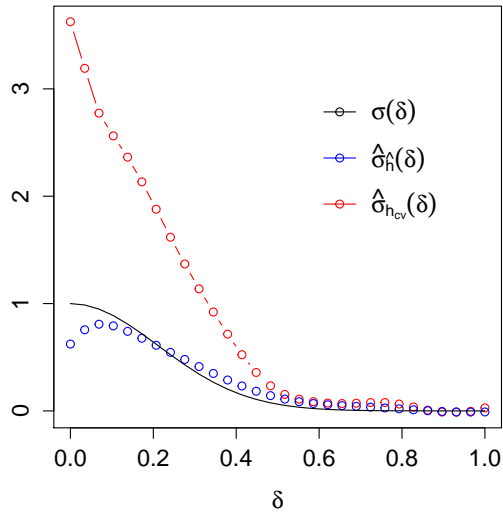
El vector de observaciones discretas  $y \sim N_m(\mu(t), \Sigma)$  en cada réplica del campo se toman sobre una rejilla uniforme de  $m = 50$ . Aquí  $\Sigma = (\sigma(s, t))$ .

El siguiente gráfico muestra los histogramas de los valores de  $\hat{h}^b$  y los vectores  $\mathbf{h}_{cv}^b$  para un tamaño de muestra  $n = 70$ , utilizando  $B = 50$  muestras. También se muestran los valores del riesgo de cada método.



**Figura 3.13:** Comparación de valores de  $h$  obtenidos por el método propuesto  $\hat{h}^b$  y el método alternativo  $h_{cv}^b$

En la siguiente gráfica se comparan las estimaciones de la función de covarianza  $\sigma$  y de la función de correlación  $\rho$  con los anchos de banda promedios  $\bar{\hat{h}}$  y  $\bar{h}_{cv}$ .



**Figura 3.14:** Curvas de función de covarianza teórica  $\sigma$  y sus estimaciones  $\hat{\sigma}_h$  y  $\hat{\sigma}_{h_{cv}}$

**Figura 3.15:** Curvas de función de correlación teórica  $\rho$  y sus estimaciones  $\hat{\rho}_h$  y  $\hat{\rho}_{h_{cv}}$

A diferencia del método anterior, en este ejemplo resulta un riesgo mucho menor utilizando el método propuesto en comparación con el alternativo. En las Figuras 3.14 y 3.15 se muestra que método alternativo estima bien las covarianzas pero sobreestima la varianza de los datos. En la Figura 3.13 se observa que los valores de  $h$  obtenidos con el método alternativo son más pequeños que con el método propuesto, esto se debe a que este método tampoco determina  $h$  tomando en cuenta la función covarianza implicando la selección de valores de  $h$  inadecuados. Sin embargo, esto también amerita más estudios en futuro.

---

---

## CAPÍTULO 4

---

# Conclusiones y posibles líneas de trabajo a futuro

### Conclusiones

1. El criterio propuesto para la selección del ancho de banda del estimador por núcleo de funciones de covarianza es un estimador insesgado del riesgo cuadrático.
2. El estimador propuesto de la función de covarianza, con el ancho de banda seleccionado de acuerdo al criterio insesgado, cumple las siguientes propiedades teóricas deseables:
  - i) Es una función simétrica y definida no negativa, o sea, es una función de covarianza.
  - ii) Satisface una desigualdad de concentración que implica buen comportamiento de su riesgo para toda muestra finita.
  - iii) Es asintóticamente óptimo desde el punto de vista del riesgo cuadrático cuando el tamaño de muestra crece.

3. Resultados de simulación, en una variedad de escenarios que incluyen covarianzas estacionarias y no estacionarias, muestran que el criterio de selección propuesto tiene buen comportamiento, conduciendo a estimaciones del ancho de banda cercanas al valor óptimo, incluso para tamaños de muestra moderados.
4. Estudios de simulación muestran que el estimador propuesto de la función de covarianza tiene buen comportamiento en la práctica, incluso para tamaños de muestra moderados.
5. El estimador propuesto es fácil de implementar, y tiene la ventaja (en comparación con otros enfoques) de no requerir al investigador especificar alguna base de funciones adecuada ni resolver numéricamente un problema de autovalores y autovectores.

## **Trabajo a futuro**

Las siguientes son algunas líneas de posible trabajo a futuro.

- a) Estudio asintótico del estimador propuesto cuando no solo el tamaño de muestra  $n$  crece sino además el número  $m$  de puntos en la rejilla en que se observa el campo aleatorio.
  - b) Estudiar el comportamiento de predictores lineales, por ejemplo de tipo “kriging” construidos sobre la base de la función de covarianza estimada por el método propuesto. Particularmente esto es relevante en aplicaciones de interés actual.
-



# Referencias

- [1] Balakrishnan, N. (2006). *Continuous multivariate distributions*. Wiley Online Library.
- [2] Bigot, J., Biscay, R., Loubes, J.-M., Muniz-Alvarez, L., et al. (2010). Nonparametric estimation of covariance functions by model selection. *Electronic Journal of Statistics*, 4, 822–855.
- [3] Bigot, J., Biscay, R. J., Loubes, J.-M., y Muñoz-Alvarez, L. (2011). Group lasso estimation of high-dimensional covariance matrices. *Journal of Machine Learning Research*, 12(Nov), 3187–3225.
- [4] Biscay, R. J., L. H. y Loubes, J. M. (2012). Adaptive covariance estimation with model selection. *Mathematical Methods of Statistics*, 21, 283–297.
- [5] Biscay, R. J., Camejo, D. G., Loubes, J.-M., y Alvarez, L. M. (2014). Estimation of covariance functions by a fully data-driven model selection procedure and its application to kriging spatial interpolation of real rainfall data. *Statistical Methods & Applications*, 23(2), 149–174.
- [6] Biscay, R. J., Rodríguez, L. M., y Díaz-Frances, E. (1997). Cross-validation of covariance structures using the Frobenius matrix distance as a discrepancy function. *Journal of Statistical Computation and Simulation*, 58(3), 195–215.
- [7] Cao, G., Wang, L., Li, Y., y Yang, L. (2016). Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Statistica Sinica*, 26, 359–383.

- 
- [8] Choi, I., Li, B., y Wang, X. (2013). Nonparametric estimation of spatial and space-time covariance function. *Journal of agricultural, biological, and environmental statistics*, 18(4), 611–630.
- [9] Cressie, N. (1991). *Statistics for Spatial Data*. Wiley series in probability and mathematical statistics.
- [10] Diggle, P. J. y Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, 401–415.
- [11] Elogne, S. N., Perrin, O., y Thomas-Agnan, C. (2008). Non parametric estimation of smooth stationary covariance functions by interpolation methods. *Statistical Inference for Stochastic Processes*, 11(2), 177–205.
- [12] Fan, J., Huang, T., y Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102(478), 632–641.
- [13] Ferraty, F. y Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.
- [14] Galán, L., Biscay, R., Valdés, P., Neira, L., y Virues, T. (1994). Multivariate statistical brain electromagnetic mapping. *Brain topography*, 7(1), 17–28.
- [15] Hall, P., Fisher, N. I., y Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *The Annals of Statistics*, 22, 2115–2134.
- [16] He, B. J., Snyder, A. Z., Zempel, J. M., Smyth, M. D., y Raichle, M. E. (2008). Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. *Proceedings of the National Academy of Sciences*, 105(41), 16039–16044.
- [17] Hélène Lescornel, Jean-Michel Loubes, C. C. (2014). Unbiased risk estimation method for covariance estimation. *ESAIM-PROBABILITY AND STATISTICS*, 18, 251–264.
-

- 
- [18] Horváth, L. y Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.
- [19] Jiménez, J. C., Biscay, R., y Montoto, O. (1995). Modeling the electroencephalogram by means of spatial spline smoothing and temporal autoregression. *Biological cybernetics*, 72(3), 249–259.
- [20] Kou, S. y Efron, B. (2002). Smoothers and the  $c_p$ , generalized maximum likelihood, and extended exponential criteria: A geometric approach. *Journal of the American Statistical Association*, 97(459), 766–782.
- [21] Lin, H. y Pan, J. (2013). Nonparametric estimation of mean and covariance structures for longitudinal data. *Canadian Journal of Statistics*, 41(4), 557–574.
- [22] Mallows, C. L. (1973). Some comments on  $C_p$ . *Technometrics*, 15(4), 661–675.
- [23] Mallows, C. L. (1995). More comments on  $C_p$ . *Technometrics*, 37(4), 362–372.
- [24] Masry, E. (1983). Non-parametric covariance estimation from irregularly-spaced data. *Advances in Applied Probability*, 15, 113–132.
- [25] Ramsay, J. O., Hooker, G., y Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.
- [26] Ramsay J., S. B. (2005). *Functional Data Analysis*. Springer Series in Statistics.
- [27] Reich, B. J. y Fuentes, M. (2012). Nonparametric bayesian models for a spatial covariance. *Statistical methodology*, 9(1), 265–274.
- [28] Rice, J. A. y Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 233–243.
- [29] Sparks, R., Coutsourides, D., y Troskie, L. (1983). The multivariate  $C_p$ . *Communications in Statistics-Theory and Methods*, 12(15), 1775–1793.
-

- [30] Sporns, O. (2012). *Discovering the human connectome*. MIT press.
- [31] Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Berlin; New York: Springer-Verlag Inc.
- [32] Yanagihara, H. y Satoh, K. (2010). An unbiased Cp criterion for multivariate Ridge regression. *Journal of Multivariate Analysis*, 101(5), 1226–1238.
- [33] Yang, J., Zhu, H., Choi, T., Cox, D. D., et al. (2016). Smoothing and mean-covariance estimation of functional data with a Bayesian hierarchical model. *Bayesian Analysis*, 11(3), 649–670.
- [34] Zhang, J.-T. (2013). *Analysis of variance for functional data*. CRC Press.
-

# Álgebra de Matrices

## La Traza

La traza de una matriz  $A = [a_{ij}]$  de dimension  $m \times m$  es definida como

$$tr(A) = a_{11} + a_{22} + \cdots + a_{mm} = \sum_{i=1}^m a_{ii}$$

### Resultados generales

1.  $A, B(m \times m) : tr(A \pm B) = tr(A) \pm tr(B)$
2.  $A(m \times m), c \in \mathbb{C} : tr(cA) = tr(A) \pm ctr(A)$
3.  $A, B(m \times m), c_1, c_2 \in \mathbb{C} : tr(c_1A \pm c_2B) = c_1tr(A) \pm c_2tr(B)$
4.  $A(m \times m) : tr(A^T) = tr(A)$
5.  $A(m \times n), B(n \times m) : tr(AB) = tr(BA)$
6.  $A, B(m \times n) :$

a) Desigualdad de Cauchy-Schwarz

$$(tr(A^T B))^2 \leq tr(A^T A)tr(B^T B) \quad (1)$$

b)  $(tr(A^T B))^2 \leq tr(A^T A)tr(B^T B)$

c)  $(tr(A^T B))^2 \leq tr(AA^T)tr(BB^T)$

## Producto Kronecker

Sean  $A = [a_{ij}] (m \times n)$ ,  $B = [b_{ij}] (n \times p)$ . El producto Kronecker de estas dos matrices es definido como

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} (pm \times nq)$$

Resultados generales

1.  $A(m \times n), B(p \times q)$  :  $A \otimes B \neq B \otimes A$  en general.
2.  $A(m \times n), B, C(p \times q)$  :  $A \otimes (B \pm C) = A \otimes B \pm A \otimes C$
3.  $A(m \times n), B(p \times q), C(r \times s)$  :

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C = A \otimes B \otimes C$$

## Operador vectorización

$vec$  denota el operador vectorización por columna, el cual concatena las columnas de una matriz en un vector columna, esto es, para una matriz  $A = [a_{ij}]$ ,

$$vec(A) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \\ a_{12} \\ \vdots \\ a_{m2} \\ \vdots \\ a_{mn} \end{bmatrix} (mn \times 1)$$

Resultados generales

1.  $A, B(m \times n)$  :  $vec(A \pm B) = vec(A) \pm vec(B)$

$$2. A(m \times n), c \in \mathbb{C} : \quad \text{vec}(cA) = c \text{vec}(A)$$

$$3. A(m \times n), B(n \times r), C(r \times s) : \quad \text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$$

## Norma de Frobenius

La norma matricial de Frobenius proporciona una métrica significativa para comparar matrices de covarianza, ampliamente utilizado en el análisis multivariado, en particular en la teoría sobre el análisis de componentes principales. Sea  $A \in \mathbb{R}^{m \times n}$ , la norma de Frobenius se define como sigue

$$\|A\|_F = +\sqrt{\text{tr}(A^T A)}$$

### Resultados generales

$$1. A, B(n \times m) : \quad \|A + B\|_F^2 = \text{tr}(A^T A) + 2\text{tr}(AB) + \|B\|_F^2$$