Regresión lineal funcional.

Alejandro Cholaquidis

CMAT-Facultad de Ciencias, UdelaR
Montevideo Uruguay

Trabajo en conjunto con: A. Cuevas and J.R. Berrendero

Seminario de Probabilidad y Estadística - Julio 2021

1. Regresión lineal

2. RKHS

3. Estimación de β
   - Un estimador consistente en L^2, basado en regularización
   - Un estimador consistente en el RKHS basado en discretización.

1. Regresión lineal

2. RKHS

3. Estimación de β
   - Un estimador consistente en L^2, basado en regularización
   - Un estimador consistente en el RKHS basado en discretización.

El modelo lineal clásico

Supongamos que Y es una v.a. real, X un elemento aleatorio a valores en un Hilbert H, y

$$Y = \alpha_0 + \Gamma(X) + \epsilon$$

donde $\alpha_0 \in \mathbb{R}$, Γ es un funcional lineal acotado en H, ϵ un ruido aleatorio, que se asume independiente de X.

El modelo lineal clásico

Supongamos que Y es una v.a. real, X un elemento aleatorio a valores en un Hilbert H, y

$$Y = \alpha_0 + \Gamma(X) + \epsilon$$

donde $\alpha_0 \in \mathbb{R}$, Γ es un funcional lineal acotado en H, ϵ un ruido aleatorio, que se asume independiente de X.

Usualmente $H = L^2[0,1]$, $\Gamma(X) = \langle X, \beta \rangle$ con $\beta \in H$, $X(t) \in L^2(\Omega)$, $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}(\epsilon^2) = \sigma^2 < \infty$, entonces

$$Y = \beta_0 + \int_0^1 \beta(s)X(s)ds + \epsilon. \tag{1}$$

El modelo lineal clásico

Supongamos que Y es una v.a. real, X un elemento aleatorio a valores en un Hilbert H, y

$$Y = \alpha_0 + \Gamma(X) + \epsilon$$

donde $\alpha_0 \in \mathbb{R}$, Γ es un funcional lineal acotado en H, ϵ un ruido aleatorio, que se asume independiente de X.

Usualmente $H = L^2[0,1]$, $\Gamma(X) = \langle X, \beta \rangle$ con $\beta \in H$, $X(t) \in L^2(\Omega)$, $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}(\epsilon^2) = \sigma^2 < \infty$, entonces

$$Y = \beta_0 + \int_0^1 \beta(s)X(s)ds + \epsilon. \tag{1}$$

Si tenemos (X_i, Y_i) iid, para $i = 1, \dots, n$, con

$$Y_i = \beta_0 + \int_0^1 \beta(s)X_i(s)ds + \epsilon_i,$$

queremos estimar β.

El modelo lineal clásico

Escribimos

$$Y - \mathbb{E}(Y) = \int_0^1 \beta(t)(X(t) - \mathbb{E}(X(t)))dt + \epsilon \tag{2}$$

El modelo lineal clásico

Escribimos

$$Y - \mathbb{E}(Y) = \int_0^1 \beta(t)(X(t) - \mathbb{E}(X(t)))dt + \epsilon \tag{2}$$

Si multiplicamos (2) por $X - \mathbb{E}(X)$, y tomamos esperanza

$$\mathbb{E}\Big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big] = \mathbb{E}\Big[(X - \mathbb{E}(X))\underbrace{\int_0^1 \beta(t)(X(s) - \mathbb{E}(X(s)))ds}_{\langle X - \mathbb{E}(X), \beta\rangle}\Big] =: K(\beta)$$

donde $K : L^2[0,1] \to L^2[0,1]$, es

$$(K(u))(t) = \mathbb{E}\Big[\langle X - \mathbb{E}(X), u\rangle \Big(X(t) - \mathbb{E}(X(t))\Big)\Big] = \int_0^1 Cov(X(s), X(t))u(s)ds$$

El modelo lineal clásico

Escribimos

$$Y - \mathbb{E}(Y) = \int_0^1 \beta(t)(X(t) - \mathbb{E}(X(t)))dt + \epsilon \tag{2}$$

Si multiplicamos (2) por $X - \mathbb{E}(X)$, y tomamos esperanza

$$\mathbb{E}\Big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big] = \mathbb{E}\Big[(X - \mathbb{E}(X))\underbrace{\int_0^1 \beta(t)(X(s) - \mathbb{E}(X(s)))ds}_{\langle X - \mathbb{E}(X), \beta\rangle}\Big] =: K(\beta)$$

donde $K : L^2[0,1] \to L^2[0,1]$, es

$$(K(u))(t) = \mathbb{E}\Big[\langle X - \mathbb{E}(X), u\rangle \Big(X(t) - \mathbb{E}(X(t))\Big)\Big] = \int_0^1 Cov(X(s), X(t))u(s)ds$$

Se puede probar

- K es autoadjunto: $\langle K(u), v\rangle = \langle u, K(v)\rangle$

El modelo lineal clásico

Escribimos

$$Y - \mathbb{E}(Y) = \int_0^1 \beta(t)(X(t) - \mathbb{E}(X(t)))dt + \epsilon \tag{2}$$

Si multiplicamos (2) por $X - \mathbb{E}(X)$, y tomamos esperanza

$$\mathbb{E}\Big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big] = \mathbb{E}\Big[(X - \mathbb{E}(X)) \underbrace{\int_0^1 \beta(t)(X(s) - \mathbb{E}(X(s)))ds}_{\langle X - \mathbb{E}(X), \beta \rangle}\Big] =: K(\beta)$$

donde $K : L^2[0,1] \to L^2[0,1]$, es

$$(K(u))(t) = \mathbb{E}\Big[\langle X - \mathbb{E}(X), u\rangle \Big(X(t) - \mathbb{E}(X(t))\Big)\Big] = \int_0^1 Cov(X(s), X(t))u(s)ds$$

Se puede probar

- K es autoadjunto: $\langle K(u), v\rangle = \langle u, K(v)\rangle$
- no negativo: $\langle K(u), u\rangle \geq 0$

El modelo lineal clásico

Escribimos

$$Y - \mathbb{E}(Y) = \int_0^1 \beta(t)(X(t) - \mathbb{E}(X(t)))dt + \epsilon \tag{2}$$

Si multiplicamos (2) por $X - \mathbb{E}(X)$, y tomamos esperanza

$$\mathbb{E}\Big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big] = \mathbb{E}\Big[(X - \mathbb{E}(X)) \underbrace{\int_0^1 \beta(t)(X(s) - \mathbb{E}(X(s)))ds}_{\langle X - \mathbb{E}(X), \beta \rangle}\Big] =: K(\beta)$$

donde $K : L^2[0,1] \to L^2[0,1]$, es

$$(K(u))(t) = \mathbb{E}\Big[\langle X - \mathbb{E}(X), u \rangle \Big(X(t) - \mathbb{E}(X(t))\Big)\Big] = \int_0^1 Cov(X(s), X(t))u(s)ds$$

Se puede probar

- K es autoadjunto: $\langle K(u), v \rangle = \langle u, K(v) \rangle$
- no negativo: $\langle K(u), u \rangle \geq 0$
- Hilbert-Schmidt: $\sum_i \|K(e_i)\| < \infty$, e_i b.o.n.

El modelo lineal clásico

Escribimos

$$Y - \mathbb{E}(Y) = \int_0^1 \beta(t)(X(t) - \mathbb{E}(X(t)))dt + \epsilon \qquad (2)$$

Si multiplicamos (2) por $X - \mathbb{E}(X)$, y tomamos esperanza

$$\mathbb{E}\Big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big] = \mathbb{E}\Big[(X - \mathbb{E}(X)) \underbrace{\int_0^1 \beta(t)(X(s) - \mathbb{E}(X(s)))ds}_{\langle X - \mathbb{E}(X), \beta \rangle}\Big] =: K(\beta)$$

donde $K : L^2[0,1] \to L^2[0,1]$, es

$$(K(u))(t) = \mathbb{E}\Big[\langle X - \mathbb{E}(X), u \rangle \Big(X(t) - \mathbb{E}(X(t))\Big)\Big] = \int_0^1 Cov(X(s), X(t))u(s)ds$$

Se puede probar

- K es autoadjunto: $\langle K(u), v \rangle = \langle u, K(v) \rangle$
- no negativo: $\langle K(u), u \rangle \geq 0$
- Hilbert-Schmidt: $\sum_i \|K(e_i)\| < \infty$, e_i b.o.n.

Por lo tanto $\overline{K(B_H(0,1))}$ es compacto, y por lo tanto no es invertible.

El modelo lineal clásico

Para resolver este problema, sea λ_i los autovalores de K, asociados a autovectores v_i. Se puede probar que

$$X(t) - \mathbb{E}(X(t)) = \sum_{i=1}^{\infty} \xi_i v_i(t) \quad \text{en } L^2(\Omega), \text{ uniforme en } t \tag{3}$$

donde ξ_i v.a. tal que $\mathbb{E}(\xi_i) = 0$, $\mathbb{E}(\xi_j^2) = \lambda_j$, $\mathbb{E}(\xi_i, \xi_j) = 0$ si $i \neq j$.

El modelo lineal clásico

Para resolver este problema, sea λ_i los autovalores de K, asociados a autovectores v_i. Se puede probar que

$$X(t) - \mathbb{E}(X(t)) = \sum_{i=1}^{\infty} \xi_i v_i(t) \quad \text{en } L^2(\Omega), \text{ uniforme en } t \tag{3}$$

donde ξ_i v.a. tal que $\mathbb{E}(\xi_i) = 0$, $\mathbb{E}(\xi_j^2) = \lambda_j$, $\mathbb{E}(\xi_i, \xi_j) = 0$ si $i \neq j$.
Si $\{v_i\}_i$ es bon, $\beta = \sum \langle \beta, v_i \rangle v_i$,

$$\left\langle \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right], v_j \right\rangle = \langle K(\beta), v_j \rangle = \langle \beta, K(v_j) \rangle = \lambda_j \langle \beta, v_j \rangle \quad j = 1, 2, \ldots .$$

El modelo lineal clásico

Para resolver este problema, sea λ_i los autovalores de K, asociados a autovectores v_i. Se puede probar que

$$X(t) - \mathbb{E}(X(t)) = \sum_{i=1}^{\infty} \xi_i v_i(t) \quad \text{en } L^2(\Omega), \text{ uniforme en } t \tag{3}$$

donde ξ_i v.a. tal que $\mathbb{E}(\xi_i) = 0$, $\mathbb{E}(\xi_j^2) = \lambda_j$, $\mathbb{E}(\xi_i, \xi_j) = 0$ si $i \neq j$.
Si $\{v_i\}_i$ es bon, $\beta = \sum \langle \beta, v_i \rangle v_i$,

$$\left\langle \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right], v_j \right\rangle = \langle K(\beta), v_j \rangle = \langle \beta, K(v_j) \rangle = \lambda_j \langle \beta, v_j \rangle \quad j = 1, 2, \ldots .$$

De (3):

$$\beta = \sum_{j=1}^{\infty} \frac{\left\langle \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right], v_j \right\rangle}{\lambda_j} v_j = \sum_{j=1}^{\infty} \frac{\mathbb{E}\left[\xi_j (Y - \mathbb{E}(Y))\right]}{\lambda_j} v_j .$$

Por lo tanto $\beta \in L^2[0, 1]$ si y sólo si $\sum_j (\mathbb{E}[\xi_j (Y - \mathbb{E}(Y))])^2 / \lambda_j^2 < \infty$.

El modelo lineal clásico

Algunas desventajas de este modelo

- $L^2[0, 1]$ es "demasiado grande"para β: se usan métodos de regularización (penalización, suavizado, etc...) para excluir funciones muy irregulares.

El modelo lineal clásico

Algunas desventajas de este modelo

- $L^2[0,1]$ es "demasiado grande"para β: se usan métodos de regularización (penalización, suavizado, etc...) para excluir funciones muy irregulares.
- La función $\{X(t) : t \in [0,1]\} \to X(t_0)$ no es un operador lineal y continuo en $L^2[0,1] \to \mathbb{R}$. Esto importa si se quiere hacer reducción de dimensión vía selección de variables:

$$\{X(t) : t \in [0,1]\} \to (X(t_1), \ldots, X(t_p))$$

El modelo lineal clásico

Algunas desventajas de este modelo

- $L^2[0,1]$ es “demasiado grande”para β: se usan métodos de regularización (penalización, suavizado, etc...) para excluir funciones muy irregulares.
- La función $\{X(t) : t \in [0,1]\} \to X(t_0)$ no es un operador lineal y continuo en $L^2[0,1] \to \mathbb{R}$. Esto importa si se quiere hacer reducción de dimensión vía selección de variables:
$$\{X(t) : t \in [0,1]\} \to (X(t_1), \ldots, X(t_p))$$
- El modelo finito $Y_i = \alpha_0 + \sum_{j=1}^{p} \beta_j X_i(t_j) + \epsilon_i \; i = 1, \ldots, n$ no es un caso particular del L^2, ya que $\{X(t) : t \in [0,1]\} \to \sum_j \beta_j X(t_j)$ no es un funcional continuo en L^2.

El modelo lineal clásico

Algunas desventajas de este modelo

- $L^2[0,1]$ es "demasiado grande"para β: se usan métodos de regularización (penalización, suavizado, etc...) para excluir funciones muy irregulares.
- La función $\{X(t) : t \in [0,1]\} \to X(t_0)$ no es un operador lineal y continuo en $L^2[0,1] \to \mathbb{R}$. Esto importa si se quiere hacer reducción de dimensión vía selección de variables:
$$\{X(t) : t \in [0,1]\} \to (X(t_1), \dots, X(t_p))$$
- El modelo finito $Y_i = \alpha_0 + \sum_{j=1}^{p} \beta_j X_i(t_j) + \epsilon_i \; i = 1, \dots, n$ no es un caso particular del L^2, ya que $\{X(t) : t \in [0,1]\} \to \sum_j \beta_j X(t_j)$ no es un funcional continuo en L^2.

RKHS

En el enfoque RKHS $\beta \in H(K) \subset L^2[0,1]$ donde $H(K)$ es el RKHS del operador K.

1. Regresión lineal

2. RKHS

3. Estimación de β
   - Un estimador consistente en L^2, basado en regularización
   - Un estimador consistente en el RKHS basado en discretización.

RKHS I

Sea $k(s,t)$ un núcleo continuo, simétrico, semidefinido positivo. Definimos

$$H_0(k) = \left\{ f : f(s) = \sum_{i=1}^{n} a_i k(s, t_i), a_i \in \mathbb{R}, t_i \in [0,1], n \in \mathbb{N} \right\},$$

RKHS I

Sea $k(s,t)$ un núcleo continuo, simétrico, semidefinido positivo. Definimos

$$H_0(k) = \left\{ f : f(s) = \sum_{i=1}^{n} a_i k(s, t_i), a_i \in \mathbb{R}, t_i \in [0,1], n \in \mathbb{N} \right\},$$

en este espacio definimos el producto interno $\langle \cdot, \cdot \rangle_k$, dado por

$$\langle f, g \rangle_k = \sum_{i,j} a_i b_j k(s_i, t_j)$$

RKHS I

Sea $k(s,t)$ un núcleo continuo, simétrico, semidefinido positivo. Definimos

$$H_0(k) = \left\{ f : f(s) = \sum_{i=1}^{n} a_i k(s,t_i), a_i \in \mathbb{R}, t_i \in [0,1], n \in \mathbb{N} \right\},$$

en este espacio definimos el producto interno $\langle \cdot, \cdot \rangle_k$, dado por

$$\langle f, g \rangle_k = \sum_{i,j} a_i b_j k(s_i, t_j)$$

El RKHS asociado a k es $H(k)$, el límite puntual de sucesiones de Cauchy en $H_0(k)$.

RKHS I

Sea $k(s,t)$ un núcleo continuo, simétrico, semidefinido positivo. Definimos

$$H_0(k) = \left\{ f : f(s) = \sum_{i=1}^{n} a_i k(s, t_i), a_i \in \mathbb{R}, t_i \in [0,1], n \in \mathbb{N} \right\},$$

en este espacio definimos el producto interno $\langle \cdot, \cdot \rangle_k$, dado por

$$\langle f, g \rangle_k = \sum_{i,j} a_i b_j k(s_i, t_j)$$

El RKHS asociado a k es $H(k)$, el límite puntual de sucesiones de Cauchy en $H_0(k)$.

Propiedad de reproducción:

$$f(t) = \langle f, k(\cdot, t) \rangle_k \quad \forall f \in H(k)$$

RKHS I

Sea $k(s,t)$ un núcleo continuo, simétrico, semidefinido positivo. Definimos

$$H_0(k) = \left\{ f : f(s) = \sum_{i=1}^{n} a_i k(s,t_i), a_i \in \mathbb{R}, t_i \in [0,1], n \in \mathbb{N} \right\},$$

en este espacio definimos el producto interno $\langle \cdot, \cdot \rangle_k$, dado por

$$\langle f, g \rangle_k = \sum_{i,j} a_i b_j k(s_i, t_j)$$

El RKHS asociado a k es $H(k)$, el límite puntual de sucesiones de Cauchy en $H_0(k)$.

Propiedad de reproducción:

$$f(t) = \langle f, k(\cdot,t) \rangle_k \quad \forall f \in H(k)$$

El operador evaluación es continuo:

$$|f_n(s) - f(s)| \leq |\langle k(s,\cdot), f_n - f \rangle_k| \leq \|k(s,\cdot)\|_k \|f_n - f\|_k = k(s,s)^{1/2} \|f_n - f\|_k.$$

RKHS I

Sea $k(s,t)$ un núcleo continuo, simétrico, semidefinido positivo. Definimos

$$H_0(k) = \left\{ f : f(s) = \sum_{i=1}^{n} a_i k(s,t_i), a_i \in \mathbb{R}, t_i \in [0,1], n \in \mathbb{N} \right\},$$

en este espacio definimos el producto interno $\langle \cdot, \cdot \rangle_k$, dado por

$$\langle f, g \rangle_k = \sum_{i,j} a_i b_j k(s_i, t_j)$$

El RKHS asociado a k es $H(k)$, el límite puntual de sucesiones de Cauchy en $H_0(k)$.

Propiedad de reproducción:

$$f(t) = \langle f, k(\cdot,t) \rangle_k \quad \forall f \in H(k)$$

El operador evaluación es continuo:

$$|f_n(s) - f(s)| \leq |\langle k(s,\cdot), f_n - f \rangle_k| \leq \|k(s,\cdot)\|_k \|f_n - f\|_k = k(s,s)^{1/2} \|f_n - f\|_k.$$

Como k es continuo, la convergencia en el RKHS implica la convergencia uniforme y por lo tanto las funciones en $H(k)$ son continuas.

RKHS II, otra construcción posible

Supongamos $X(t) \in L^2(\Omega)$, $X \in L^2[0,1]$ c.s, con función de covarianza k y función de media m. Sea

$$L_0(X) = \left\{ \sum_{i=1}^{p} a_i (X(t_i) - m(t_i)), p \in \mathbb{N}, t_i \in [0,1], a_1, \ldots, a_p \in \mathbb{R} \right\}$$

y $L(X)$ la completación $L^2(\Omega)$ de $L_0(X)$.

RKHS II, otra construcción posible

Supongamos $X(t) \in L^2(\Omega)$, $X \in L^2[0,1]$ c.s, con función de covarianza k y función de media m. Sea

$$L_0(X) = \left\{ \sum_{i=1}^{p} a_i(X(t_i) - m(t_i)), p \in \mathbb{N}, t_i \in [0,1], a_1, \ldots, a_p \in \mathbb{R} \right\}$$

y $L(X)$ la completación $L^2(\Omega)$ de $L_0(X)$. Definimos para $U \in L(X)$,

$$\Psi_X(U)(s) = \langle U, X(s) - m(s) \rangle = \mathbb{E}\big(U(X(s) - m(s))\big).$$

Es un mapa inyectivo de $L(X)$ a $L^2[0,1]$.

RKHS II, otra construcción posible

Supongamos $X(t) \in L^2(\Omega)$, $X \in L^2[0,1]$ c.s, con función de covarianza k y función de media m. Sea

$$L_0(X) = \left\{ \sum_{i=1}^{p} a_i(X(t_i) - m(t_i)), p \in \mathbb{N}, t_i \in [0,1], a_1, \ldots, a_p \in \mathbb{R} \right\}$$

y $L(X)$ la completación $L^2(\Omega)$ de $L_0(X)$. Definimos para $U \in L(X)$,

$$\Psi_X(U)(s) = \langle U, X(s) - m(s) \rangle = \mathbb{E}\big(U(X(s) - m(s))\big).$$

Es un mapa inyectivo de $L(X)$ a $L^2[0,1]$. Definimos

$$H(k) = \{\Psi_X(U) : U \in L(X)\}.$$

RKHS II, otra construcción posible

Supongamos $X(t) \in L^2(\Omega)$, $X \in L^2[0,1]$ c.s, con función de covarianza k y función de media m. Sea

$$L_0(X) = \left\{ \sum_{i=1}^{p} a_i(X(t_i) - m(t_i)), p \in \mathbb{N}, t_i \in [0,1], a_1, \ldots, a_p \in \mathbb{R} \right\}$$

y $L(X)$ la completación $L^2(\Omega)$ de $L_0(X)$. Definimos para $U \in L(X)$,

$$\Psi_X(U)(s) = \langle U, X(s) - m(s) \rangle = \mathbb{E}\big(U(X(s) - m(s))\big).$$

Es un mapa inyectivo de $L(X)$ a $L^2[0,1]$. Definimos

$$H(k) = \{\Psi_X(U) : U \in L(X)\}.$$

Es un espacio de Hilbert con el producto interno heredado de $L(X)$:

$$\langle f, g \rangle_k := \langle \Psi_X^{-1}(f), \Psi_X^{-1}(g) \rangle.$$

RKHS II, otra construcción posible

Supongamos $X(t) \in L^2(\Omega)$, $X \in L^2[0,1]$ c.s, con función de covarianza k y función de media m. Sea

$$L_0(X) = \left\{ \sum_{i=1}^{p} a_i(X(t_i) - m(t_i)), p \in \mathbb{N}, t_i \in [0,1], a_1, \ldots, a_p \in \mathbb{R} \right\}$$

y $L(X)$ la completación $L^2(\Omega)$ de $L_0(X)$. Definimos para $U \in L(X)$,

$$\Psi_X(U)(s) = \langle U, X(s) - m(s) \rangle = \mathbb{E}\big(U(X(s) - m(s))\big).$$

Es un mapa inyectivo de $L(X)$ a $L^2[0,1]$. Definimos

$$H(k) = \{\Psi_X(U) : U \in L(X)\}.$$

Es un espacio de Hilbert con el producto interno heredado de $L(X)$:

$$\langle f, g \rangle_k := \langle \Psi_X^{-1}(f), \Psi_X^{-1}(g) \rangle.$$

Por lo tanto Ψ_X es una isometría entre $L(X)$ y $H(k)$. Observar que $\Psi_X(X(t_i) - m(t_i))(s) = k(t_i, s) \in H(k)$.

RKHS vía el operador $K^{1/2}$

Dado $k(s,t)$ continuo, simétrico, semidefinido positivo, $K(f)(\cdot) = \int_0^1 k(s,\cdot)f(s)ds$ es compacto. Denotamos $K^{1/2}$ el (único) operador tal que $K^{1/2}K^{1/2} = K$. Se define

$$\mathcal{H}(K) = \left\{K^{1/2}(f) : f \in L^2[0,1]\right\}.$$

RKHS vía el operador $K^{1/2}$

Dado $k(s,t)$ continuo, simétrico, semidefinido positivo, $K(f)(\cdot) = \int_0^1 k(s,\cdot)f(s)ds$ es compacto. Denotamos $K^{1/2}$ el (único) operador tal que $K^{1/2}K^{1/2} = K$. Se define

$$\mathcal{H}(K) = \left\{K^{1/2}(f) : f \in L^2[0,1]\right\}.$$

Si $\{v_i\}_i$ es bon de $L^2[0,1]$, de autofunciones de K, con autovalores $\lambda_i > 0$,

$$\mathcal{H}(K) = \left\{ f \in L^2[0,1] : f = \sum_i \langle f, v_i\rangle v_i \quad \text{y} \quad \left\{\frac{\langle f, v_i\rangle}{\sqrt{\lambda_i}}\right\}_i \in l^2(\mathbb{N}) \right\}$$

RKHS vía el operador $K^{1/2}$

Dado $k(s,t)$ continuo, simétrico, semidefinido positivo, $K(f)(\cdot) = \int_0^1 k(s,\cdot)f(s)ds$ es compacto. Denotamos $K^{1/2}$ el (único) operador tal que $K^{1/2}K^{1/2} = K$. Se define

$$\mathcal{H}(K) = \left\{K^{1/2}(f) : f \in L^2[0,1]\right\}.$$

Si $\{v_i\}_i$ es bon de $L^2[0,1]$, de autofunciones de K, con autovalores $\lambda_i > 0$,

$$\mathcal{H}(K) = \left\{ f \in L^2[0,1] : f = \sum_i \langle f, v_i\rangle v_i \quad \text{y} \quad \left\{\frac{\langle f, v_i\rangle}{\sqrt{\lambda_i}}\right\}_i \in l^2(\mathbb{N}) \right\}$$

La condición $\{\langle f, v_i\rangle/\sqrt{\lambda_i}\}_i \in l^2(\mathbb{N})$ impone regularidad en f ya que por el Teorema de Mercer, $\sum_i \lambda_i = \int_0^1 k(s,s)ds < \infty$.

RKHS vía el operador $K^{1/2}$

Dado $k(s,t)$ continuo, simétrico, semidefinido positivo, $K(f)(\cdot) = \int_0^1 k(s,\cdot)f(s)ds$ es compacto. Denotamos $K^{1/2}$ el (único) operador tal que $K^{1/2}K^{1/2} = K$. Se define

$$\mathcal{H}(K) = \left\{K^{1/2}(f) : f \in L^2[0,1]\right\}.$$

Si $\{v_i\}_i$ es bon de $L^2[0,1]$, de autofunciones de K, con autovalores $\lambda_i > 0$,

$$\mathcal{H}(K) = \left\{ f \in L^2[0,1] : f = \sum_i \langle f, v_i\rangle v_i \quad \text{y} \quad \left\{\frac{\langle f, v_i\rangle}{\sqrt{\lambda_i}}\right\}_i \in l^2(\mathbb{N}) \right\}$$

La condición $\{\langle f, v_i\rangle/\sqrt{\lambda_i}\}_i \in l^2(\mathbb{N})$ impone regularidad en f ya que por el Teorema de Mercer, $\sum_i \lambda_i = \int_0^1 k(s,s)ds < \infty$. El producto interno en $\mathcal{H}(K)$ es

$$\langle f, g\rangle_k = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle\langle g, v_i\rangle}{\lambda_i}.$$

Los dos RKHS son el mismo, para eso

Teorema de Mercer

Sea $k : [0,1]^2 \to \mathbb{R}$ continuo, simétrico, definido positivo. Definimos

$$K(f)(t) = \int_0^1 k(s,t) f(s) ds$$

el cual resulta compacto y autoadjunto. Sea $\{v_i\}_i$ bon de $L^2[0,1]$, de autofunciones de K, con autovalores$\{\lambda_i\}_i$. Entonces para todo t, s,

$$k(t,s) = \sum_{i=1}^{\infty} \lambda_i v_i(s) v_i(t). \tag{4}$$

La convergencia es absoluta para todo $s, t \in [0,1]^2$, y uniforme en $[0,1]^2$.

El mapa $\Phi : [0,1] \to l^2(\mathbb{N})$ definido como

$$\Phi(t) = \{\sqrt{\lambda_i} v_i(t)\}_i$$

se denomina *feature map*.

RKHS

Propiedad reproductora

$$\langle f, k(s,\cdot)\rangle_k = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle \langle k(s,\cdot), v_i\rangle}{\lambda_i} = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle K(v_i)(s)}{\lambda_i} = f(s)$$

RKHS

Propiedad reproductora

$$\langle f, k(s,\cdot)\rangle_k = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle \langle k(s,\cdot), v_i\rangle}{\lambda_i} = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle K(v_i)(s)}{\lambda_i} = f(s)$$

Por el Teorema de Mercer:

- $\langle \Phi(s), \Phi(t)\rangle_{l^2} = k(s,t)$
- $\langle \Phi(s), \Phi(\cdot)\rangle_{l^2} = k(s,\cdot) \in \mathcal{H}(K)$.

$\mathcal{H}(K)$ and $H(k)$ son el mismo espacio de funciones con el mismo producto.

Dado un kernel de Mercer k, existe un único espacio de Hilbert H:

1. $k(s,\cdot) \in H$.

RKHS

Propiedad reproductora

$$\langle f, k(s,\cdot)\rangle_k = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle \langle k(s,\cdot), v_i\rangle}{\lambda_i} = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle K(v_i)(s)}{\lambda_i} = f(s)$$

Por el Teorema de Mercer:

- $\langle \Phi(s), \Phi(t)\rangle_{l^2} = k(s,t)$
- $\langle \Phi(s), \Phi(\cdot)\rangle_{l^2} = k(s,\cdot) \in \mathcal{H}(K)$.

$\mathcal{H}(K)$ and $H(k)$ son el mismo espacio de funciones con el mismo producto.

Dado un kernel de Mercer k, existe un único espacio de Hilbert H:

1. $k(s,\cdot) \in H$.
2. Para todo $f \in H, f(s) = \langle f, k(s,\cdot)\rangle_H$.

RKHS

Propiedad reproductora

$$\langle f, k(s,\cdot)\rangle_k = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle \langle k(s,\cdot), v_i\rangle}{\lambda_i} = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle K(v_i)(s)}{\lambda_i} = f(s)$$

Por el Teorema de Mercer:

- $\langle \Phi(s), \Phi(t)\rangle_{l^2} = k(s,t)$
- $\langle \Phi(s), \Phi(\cdot)\rangle_{l^2} = k(s,\cdot) \in \mathcal{H}(K)$.

$\mathcal{H}(K)$ and $H(k)$ son el mismo espacio de funciones con el mismo producto.

Dado un kernel de Mercer k, existe un único espacio de Hilbert H:

1. $k(s,\cdot) \in H$.
2. Para todo $f \in H, f(s) = \langle f, k(s,\cdot)\rangle_H$.
3. El espacio generado por $\{k(s,\cdot) : s \in [0,1]\}$ es denso en H,

RKHS

Propiedad reproductora

$$\langle f, k(s,\cdot)\rangle_k = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle \langle k(s,\cdot), v_i\rangle}{\lambda_i} = \sum_{i=1}^{\infty} \frac{\langle f, v_i\rangle K(v_i)(s)}{\lambda_i} = f(s)$$

Por el Teorema de Mercer:

- $\langle \Phi(s), \Phi(t)\rangle_{l^2} = k(s,t)$
- $\langle \Phi(s), \Phi(\cdot)\rangle_{l^2} = k(s,\cdot) \in \mathcal{H}(K)$.

$\mathcal{H}(K)$ and $H(k)$ son el mismo espacio de funciones con el mismo producto.

Dado un kernel de Mercer k, existe un único espacio de Hilbert H:

1. $k(s,\cdot) \in H$.
2. Para todo $f \in H, f(s) = \langle f, k(s,\cdot)\rangle_H$.
3. El espacio generado por $\{k(s,\cdot) : s \in [0,1]\}$ es denso en H,

Resta probar 3, que se sigue de: si $f \in \mathcal{H}(K)$ tal que $\langle f, k(s,\cdot)\rangle = 0$ para todo s, entonces $f(s) = 0$ para todo s.

La idea general

Reemplazar $\langle X, \beta \rangle$ en $Y = \langle X, \beta \rangle + \epsilon$ con $\langle X, \beta \rangle_k$ donde $\beta \in H(k)$.

La idea general

Reemplazar $\langle X, \beta\rangle$ en $Y = \langle X, \beta\rangle + \epsilon$ con $\langle X, \beta\rangle_k$ donde $\beta \in H(k)$. El problema es que en la mayoría de casos $X(\omega)(\cdot) \notin H(k)$ c.s.. Por ejemplo si X es el Browniano

$$H(k) = \left\{f \in L^2[0,1] : f \text{ abs. cont. y } f' \in L^2[0,1]\right\}.$$

La idea general

Reemplazar $\langle X, \beta \rangle$ en $Y = \langle X, \beta \rangle + \epsilon$ con $\langle X, \beta \rangle_k$ donde $\beta \in H(k)$. El problema es que en la mayoría de casos $X(\omega)(\cdot) \notin H(k)$ c.s.. Por ejemplo si X es el Browniano

$$H(k) = \left\{ f \in L^2[0,1] : f \text{ abs. cont. y } f' \in L^2[0,1] \right\}.$$

Parzen define $\langle X, \beta \rangle_k =: \Psi_X^{-1}(\beta)$. Si asumimos

$$Y = U_X + \varepsilon, \tag{5}$$

donde $U_X \in L(X)$, y $\varepsilon \in L(X)^\perp$, $\mathbb{E}(\varepsilon) = 0$. El modelo (5) es

$$Y = \Psi_X^{-1}(\alpha) + \varepsilon, \tag{6}$$

donde $\alpha \in H(k)$ y $\varepsilon \in L(X)^\perp$, además

La idea general

Reemplazar $\langle X, \beta \rangle$ en $Y = \langle X, \beta \rangle + \epsilon$ con $\langle X, \beta \rangle_k$ donde $\beta \in H(k)$. El problema es que en la mayoría de casos $X(\omega)(\cdot) \notin H(k)$ c.s.. Por ejemplo si X es el Browniano

$$H(k) = \left\{ f \in L^2[0,1] : f \text{ abs. cont. y } f' \in L^2[0,1] \right\}.$$

Parzen define $\langle X, \beta \rangle_k =: \Psi_X^{-1}(\beta)$. Si asumimos

$$Y = U_X + \varepsilon, \tag{5}$$

donde $U_X \in L(X)$, y $\varepsilon \in L(X)^{\perp}$, $\mathbb{E}(\varepsilon) = 0$. El modelo (5) es

$$Y = \Psi_X^{-1}(\alpha) + \varepsilon, \tag{6}$$

donde $\alpha \in H(k)$ y $\varepsilon \in L(X)^{\perp}$, además

$$\alpha(t) = \langle \alpha, k(\cdot, t) \rangle_k = \mathbb{E}[\psi_X^{-1}(\alpha) X(t)] = \mathbb{E}((Y - \varepsilon) X(t)) = \mathbb{E}(YX(t)).$$

El modelo finito-dimensional y el L^2 son RKHS

Finito dimensional

Si

$$Y = \Psi_X^{-1}(\alpha) + \varepsilon. \tag{7}$$

Entonces existen $X(t_1^*), \ldots, X(t_p^*)$ y $\beta_1, \ldots, \beta_p \in \mathbb{R}$ tal que

$$Y = \beta_1 X(t_1^*) + \cdots + \beta_p X(t_p^*) + \varepsilon$$

si y sólo si para todo $t \in T$, $\alpha(t) = \beta_1 k(t, t_1^*) + \cdots + \beta_p k(t, t_p^*)$.

El modelo finito-dimensional y el L^2 son RKHS

Finito dimensional

Si

$$Y = \Psi_X^{-1}(\alpha) + \varepsilon. \tag{7}$$

Entonces existen $X(t_1^*), \ldots, X(t_p^*)$ y $\beta_1, \ldots, \beta_p \in \mathbb{R}$ tal que

$$Y = \beta_1 X(t_1^*) + \cdots + \beta_p X(t_p^*) + \varepsilon$$

si y sólo si para todo $t \in T$, $\alpha(t) = \beta_1 k(t, t_1^*) + \cdots + \beta_p k(t, t_p^*)$.

El modelo L^2 es un RKHS

Si

$$Y = \int_0^1 \beta(t) X(t) dt + \varepsilon := \langle \beta, X \rangle_2 + \varepsilon, \tag{8}$$

entonces vale el modelo (7), ya que $\langle \beta, X \rangle_2 \in L(X)$.

El modelo finito-dimensional y el L^2 son RKHS

Finito dimensional

Si

$$Y = \Psi_X^{-1}(\alpha) + \varepsilon. \tag{7}$$

Entonces existen $X(t_1^*), \ldots, X(t_p^*)$ y $\beta_1, \ldots, \beta_p \in \mathbb{R}$ tal que

$$Y = \beta_1 X(t_1^*) + \cdots + \beta_p X(t_p^*) + \varepsilon$$

si y sólo si para todo $t \in T$, $\alpha(t) = \beta_1 k(t, t_1^*) + \cdots + \beta_p k(t, t_p^*)$.

El modelo L^2 es un RKHS

Si

$$Y = \int_0^1 \beta(t) X(t) dt + \varepsilon := \langle \beta, X \rangle_2 + \varepsilon, \tag{8}$$

entonces vale el modelo (7), ya que $\langle \beta, X \rangle_2 \in L(X)$.

Recíprocamente, si vale (7), y existe $\beta \in L^2[0,1]$ tal que $\alpha = K\beta$, entonces el modelo (8) vale.

Componentes principales

Para resolver el modelo L^2, usualmente $\{X(t) : t \in [0, 1]\}$ se proyecta sobre $u_1, \dots, u_p$ donde $\{u_j : 1, \dots\}$ es una b.o.n. La trayectoria se reemplaza con

$$\langle X, u_1\rangle_2 u_1 + \cdots + \langle X, u_p\rangle_2 u_p$$

Intuitivamente, esto funciona si

$$\int_0^1 X(t)\beta(t)dt \approx \sum_{j=1}^{p} \beta_j \langle X, u_j\rangle_2 \quad \text{donde } \beta_j = \langle \beta, u_j\rangle_2.$$

Componentes principales

Para resolver el modelo L^2, usualmente $\{X(t) : t \in [0,1]\}$ se proyecta sobre $u_1, \dots, u_p$ donde $\{u_j : 1, \dots\}$ es una b.o.n. La trayectoria se reemplaza con

$$\langle X, u_1\rangle_2 u_1 + \cdots + \langle X, u_p\rangle_2 u_p$$

Intuitivamente, esto funciona si

$$\int_0^1 X(t)\beta(t)dt \approx \sum_{j=1}^{p} \beta_j \langle X, u_j\rangle_2 \quad \text{donde } \beta_j = \langle \beta, u_j\rangle_2.$$

¿Cuándo no hay pérdida, y cómo tiene que ser α?.

Componentes principales

Para resolver el modelo L^2, usualmente $\{X(t) : t \in [0,1]\}$ se proyecta sobre $u_1, \ldots, u_p$ donde $\{u_j : 1, \ldots\}$ es una b.o.n. La trayectoria se reemplaza con

$$\langle X, u_1\rangle_2 u_1 + \cdots + \langle X, u_p\rangle_2 u_p$$

Intuitivamente, esto funciona si

$$\int_0^1 X(t)\beta(t)dt \approx \sum_{j=1}^{p} \beta_j \langle X, u_j\rangle_2 \quad \text{donde } \beta_j = \langle \beta, u_j\rangle_2.$$

¿Cuándo no hay pérdida, y cómo tiene que ser α?. Consideremos el modelo

$$Y = \beta_1\langle X, u_1\rangle_2 + \cdots + \beta_p\langle X, u_p\rangle_2 + \varepsilon, \tag{9}$$

donde $\beta_1, \ldots, \beta_p \in \mathbb{R}$ y $\varepsilon \in L(X)^{\perp}$.

Proposición

Si vale (9) entonces $Y = \Psi_X^{-1}(\alpha) + \varepsilon$.

Componentes principales

Para resolver el modelo L^2, usualmente $\{X(t) : t \in [0,1]\}$ se proyecta sobre $u_1, \dots, u_p$ donde $\{u_j : 1, \dots\}$ es una b.o.n. La trayectoria se reemplaza con

$$\langle X, u_1\rangle_2 u_1 + \cdots + \langle X, u_p\rangle_2 u_p$$

Intuitivamente, esto funciona si

$$\int_0^1 X(t)\beta(t)dt \approx \sum_{j=1}^{p} \beta_j \langle X, u_j\rangle_2 \quad \text{donde } \beta_j = \langle \beta, u_j\rangle_2.$$

¿Cuándo no hay pérdida, y cómo tiene que ser α?. Consideremos el modelo

$$Y = \beta_1\langle X, u_1\rangle_2 + \cdots + \beta_p\langle X, u_p\rangle_2 + \varepsilon, \tag{9}$$

donde $\beta_1, \dots, \beta_p \in \mathbb{R}$ y $\varepsilon \in L(X)^{\perp}$.

Proposición

Si vale (9) entonces $Y = \Psi_X^{-1}(\alpha) + \varepsilon$. Recíprocamente si $Y = \Psi_X^{-1}(\alpha) + \varepsilon$ y α pertenece al espacio generado por $\{Ku_1, \dots, Ku_p\}$ entonces vale el modelo (9).

1. Regresión lineal

2. RKHS

3. Estimación de β
   - Un estimador consistente en L^2, basado en regularización
   - Un estimador consistente en el RKHS basado en discretización.

Regularización de Tichonov

$\alpha(t) = \mathbb{E}(YX(t))$ sugiere usar $\tilde{\alpha}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i(t)$, pero $\mathbb{P}(\tilde{\alpha} \in H(k)) = 0$.

Regularización de Tichonov

$\alpha(t) = \mathbb{E}(YX(t))$ sugiere usar $\tilde{\alpha}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i(t)$, pero $\mathbb{P}(\tilde{\alpha} \in H(k)) = 0$.
Consideremos

$$\breve{\alpha} = \arg\min_{f \in H_k} \|\tilde{\alpha} - f\|_2^2 + \gamma_n \|f\|_k^2 = (K + \gamma_n I)^{-1} K \tilde{\alpha}.$$

y

$$\hat{\alpha} := (\widehat{K} + \gamma_n I)^{-1} \widehat{K} \tilde{\alpha}, \tag{10}$$

donde $\widehat{K}$ es el operador integral definido por el kernel $\hat{k}(s,t) = n^{-1} \sum_{i=1}^{n} X_i(s) X_i(t)$.

Regularización de Tichonov

$\alpha(t) = \mathbb{E}(YX(t))$ sugiere usar $\tilde{\alpha}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i(t)$, pero $\mathbb{P}(\tilde{\alpha} \in H(k)) = 0$.
Consideremos

$$\breve{\alpha} = \arg\min_{f \in H_k} \|\tilde{\alpha} - f\|_2^2 + \gamma_n \|f\|_k^2 = (K + \gamma_n I)^{-1} K \tilde{\alpha}.$$

y

$$\hat{\alpha} := (\widehat{K} + \gamma_n I)^{-1} \widehat{K} \tilde{\alpha}, \tag{10}$$

donde $\widehat{K}$ es el operador integral definido por el kernel $\hat{k}(s,t) = n^{-1} \sum_{i=1}^{n} X_i(s) X_i(t)$.

Teorema

Sea $\gamma_n \to 0$ tal que $\gamma_n^2 \sqrt{n} \to \infty$, supongamos que $\mathbb{E}(\|X\|_2^4) < \infty$. Entonces $\|\hat{\alpha} - \alpha\|_2 \to 0$ en probabilidad, cuando $n \to \infty$.

Regularización de Tichonov

$\alpha(t) = \mathbb{E}(YX(t))$ sugiere usar $\tilde{\alpha}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i(t)$, pero $\mathbb{P}(\tilde{\alpha} \in H(k)) = 0$. Consideremos

$$\breve{\alpha} = \arg\min_{f \in H_k} \|\tilde{\alpha} - f\|_2^2 + \gamma_n \|f\|_k^2 = (K + \gamma_n I)^{-1} K \tilde{\alpha}.$$

y

$$\hat{\alpha} := (\widehat{K} + \gamma_n I)^{-1} \widehat{K} \tilde{\alpha}, \tag{10}$$

donde $\widehat{K}$ es el operador integral definido por el kernel $\hat{k}(s,t) = n^{-1}\sum_{i=1}^{n} X_i(s) X_i(t)$.

Teorema

Sea $\gamma_n \to 0$ tal que $\gamma_n^2 \sqrt{n} \to \infty$, supongamos que $\mathbb{E}(\|X\|_2^4) < \infty$. Entonces $\|\hat{\alpha} - \alpha\|_2 \to 0$ en probabilidad, cuando $n \to \infty$.

¿y en norma RKHS?

$\hat{\alpha} \in H(\hat{k})$, pero no necesariamente $\hat{\alpha} \in H(k)$.

Regularización de Tichonov

$\alpha(t) = \mathbb{E}(YX(t))$ sugiere usar $\tilde{\alpha}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i(t)$, pero $\mathbb{P}(\tilde{\alpha} \in H(k)) = 0$.
Consideremos

$$\breve{\alpha} = \arg\min_{f \in H_k} \|\tilde{\alpha} - f\|_2^2 + \gamma_n \|f\|_k^2 = (K + \gamma_n I)^{-1} K \tilde{\alpha}.$$

y

$$\hat{\alpha} := (\widehat{K} + \gamma_n I)^{-1} \widehat{K} \tilde{\alpha}, \tag{10}$$

donde $\widehat{K}$ es el operador integral definido por el kernel $\hat{k}(s,t) = n^{-1}\sum_{i=1}^{n} X_i(s) X_i(t)$.

Teorema

Sea $\gamma_n \to 0$ *tal que* $\gamma_n^2 \sqrt{n} \to \infty$, *supongamos que* $\mathbb{E}(\|X\|_2^4) < \infty$. *Entonces* $\|\hat{\alpha} - \alpha\|_2 \to 0$ *en probabilidad, cuando* $n \to \infty$.

¿y en norma RKHS?

$\hat{\alpha} \in H(\hat{k})$, pero no necesariamente $\hat{\alpha} \in H(k)$.
Teorema: si $n\gamma_n^2 \to \infty$, $\gamma_n \to 0$, $||\breve{\alpha} - \alpha||_k \to 0$ en probabilidad.

Lema de Parzen

Lemma

Sean $\alpha \in H(k)$, $T_p = \{t_{j,p} : j = 1, \ldots, p\}$ *donde* $0 \leq t_{1,p} \leq \cdots \leq t_{p,p} \leq 1$, *tal que* $T_p \subset T_{p+1} \subset [0,1]$ *y* $\overline{\cup_p T_p} = [0,1]$. *Entonces existe* $\beta_{1,p} \ldots, \beta_{p,p}$ *tal que*

$$\left\| \alpha(\cdot) - \sum_{j=1}^{p} \beta_{j,p} k(t_{j,p}, \cdot) \right\|_k^2 \to 0, \quad \textit{cuando } p \to \infty.$$

Lema de Parzen

Lemma

Sean $\alpha \in H(k)$, $T_p = \{t_{j,p} : j = 1, \ldots, p\}$ *donde* $0 \leq t_{1,p} \leq \cdots \leq t_{p,p} \leq 1$, *tal que* $T_p \subset T_{p+1} \subset [0,1]$ *y* $\overline{\cup_p T_p} = [0,1]$. *Entonces existe* $\beta_{1,p} \ldots, \beta_{p,p}$ *tal que*

$$\left\| \alpha(\cdot) - \sum_{j=1}^{p} \beta_{j,p} k(t_{j,p}, \cdot) \right\|_k^2 \to 0, \quad \textit{cuando } p \to \infty.$$

Esto sugiere usar

$$\alpha_p(\cdot) = \sum_{j=1}^{p} \beta_{j,p} k(t_{j,p}, \cdot) \quad \text{y} \quad \hat{\alpha}_p(\cdot) = \sum_{j=1}^{p} \hat{\beta}_{j,p} k(t_{j,p}, \cdot), \tag{11}$$

donde para $j = 1, \ldots, p$, $\hat{\beta}_{j,p}$ son los estimadores por mínimos cuadrados del modelo lineal p-dimensional

$$Y_i = \sum_{j=1}^{p} \beta_{j,p} X(t_{j,p}) + e_{i,p} = \langle \alpha_p, X \rangle_k + e_{i,p}, \quad i = 1, \ldots, n. \tag{12}$$

Consistencia

Theorem

Supongamos que $Y_i = \langle X, \alpha \rangle_k + \epsilon_i$, *con* $i = 1, \ldots, n$. *Consideramos*

$$Y_i = \sum_{j=1}^{p} \beta_{j,p} X(t_{j,p}) + e_{i,p} = \langle \alpha_p, X \rangle_k + e_{i,p}, \quad i = 1, \ldots, n. \tag{13}$$

donde $p = p_n \to \infty$. *Supongamos que*

(i) $e_{i,p} = e_i$ *son iid sub-Gaussianos,* $SG(\sigma_p^2)$ *con* $\sigma_p^2 \geq C_0 > 0$ *para todo* p. [a]

(ii) $\sup_{t \in [0,1]} X(t)$ *es sub-Gaussiano.*

(iii) $n(\gamma_{p,p})^2 / (p^2 \log^3 n) \to \infty$, *donde* $\gamma_{p,p}$ *es el autovalor mas chico de* K_{T_p}, *la matriz de covarianzas de* $(X(t_{1,p}), \ldots, X(t_{p,p}))$.

Entonces

$$\nu_n \|\hat{\alpha}_p - \alpha_p\|_k^2 \to 0 \quad c.s \quad \forall \nu_n \to \infty$$

tal que $n\gamma_{p,p}/(p^2 \nu_n \log n) \to \infty$. *Como consecuencia del Lema de Parzen,*

$$\|\alpha - \hat{\alpha}_p\|_k^2 = \max\{\nu_n^{-1}, \mathcal{O}(\|\alpha - \alpha_p\|_k^2)\} \quad c.s.$$

[a] $\mathbb{P}(|e_{i,p}| > t) \leq 2\exp(-t^2/(2\sigma_p^2))$

El autovalor mas chico de K_{T_p}

Proposición

Sea $\{W(t)\}_{t\in[0,1]}$ *con incrementos estacionarios e independientes, tal que* $\mathbb{E}(W^2(t)) < \infty$ *y* $\mathbb{E}(W(t)) = 0$ *para todo* $t \in [0,1]$*, entonces para todo* $\delta > 0$*,*

$$p^{1+\delta}\gamma_{p,p} \to \infty,$$

donde $\gamma_{p,p}$ *es el autovalor mas chico de la matriz de covarianzas de* $(W(1/p), \ldots, W(1))$*.*

Para el fBM con Hurst H, $\gamma_{p,p} = \mathcal{O}(1/p^{2H})$.

- Berrendero, J.R., Bueno-Larraz, B. and Cuevas, A.(2018) An RKHS model for variable selection in functional linear regression. *Journal of Multivariate Analysis*
- Berrendero, J.R., Bueno-Larraz, B. and Cuevas, A.(2019) On functional logistic regression via RKHS?s. *To appear in...*
- Berrendero, J.R., Cuevas, A. and Torrecilla, J.L.(2017) On the use of reproducing kernel Hilbert spaces in functional classification *JASA*
- Cucker, F. and Smale J. (2002) On the mathematical foundations of learning. Bulletin of the American mathematical society, 39(1), 1-49.
- Gupta, A. and Joshi, S. (2008) Some studies on the structure of covariance matrix of discrete-time fBm. *IEEE Transactions on Signal Processing* vol. 56, no 10, p. 4635-4650.
- Parzen, E. (1959) Statistical inference on time series by Hilbert space methods, I. *Technical Report 23, Stanford University.*