

---

# TÉCNICAS BOOTSTRAP EN ANÁLISIS DISCRIMINANTE PARA SERIES TEMPORALES

David Casado de Lucas

TRABAJO DE INVESTIGACIÓN EN ESTADÍSTICA

Director: Andrés Modesto Alonso Fernández

Curso: Doctorado en Matemáticas 2002/2003. UAM

---

# Sumario

- Introducción
- Análisis Discriminante en Series Temporales
- Métodos de remuestreo
- Bootstrap en Series Temporales

---

## Introducción

La clasificación de señales, elementos u objetos es una actividad frecuente en la investigación aplicada.

Las técnicas que abordan este problema constituyen el *Análisis Discriminante*.

Las reglas de clasificación que construyamos no son perfectas. La estimación de las tasas de error de clasificación es el problema en que centraremos nuestro trabajo.

Los métodos de remuestreo no necesitan tamaños muestrales muy grandes, como los métodos asintóticos.

El método bootstrap ha sido ampliamente estudiado en A.D. para el caso de datos i.i.d., pero no para el caso de datos dependientes. Aquí es donde el presente trabajo adquiere importancia.

---

## Análisis Discriminante en series temporales

---

Tenemos las categorías  $\Pi_1, \dots, \Pi_q$  y queremos clasificar un nuevo elemento en una de ellas.

Necesitamos hacer una partición del espacio euclídeo  $\mathcal{R}^n$  ó  $\mathcal{R}^{mn}$  en  $q$  regiones disjuntas  $R_1, \dots, R_q$  tales que si  $\mathbf{X}$  pertenece a la región  $R_j$ , asignamos  $\mathbf{X}$  a la categoría  $\Pi_j$ .

Sean  $p_j(\mathbf{x})$  las densidades de probabilidad de  $\mathbf{X}$  en las categorías  $\Pi_j$  con  $j = 1, \dots, q$ , y sea

$$p(k|j) = \int_{R_k} p_j(\mathbf{x}) d\mathbf{x}$$

la probabilidad de clasificación incorrecta de un elemento de  $\Pi_j$  asignado a  $\Pi_k$ ,  $j \neq k = 1, \dots, q$ .

---

Si la probabilidad a priori de pertenecer a  $\Pi_j$  es  $\pi_j$  y el coste dado a la mala clasificación es  $c(k|j) > 0$ , entonces la pérdida esperada para la partición  $R_1, \dots, R_q$  es

$$\sum_{j=1}^q \pi_j \left\{ \sum_{k \neq j} c(k|j) p(k|j) \right\}.$$

Cuando el criterio para elegir las regiones es minimizar la pérdida, la regla discriminante asigna  $\mathbf{X}$  a  $\Pi_j$  si

$$\sum_{l \neq j} \pi_l p_l(\mathbf{X}) c(j|l) < \sum_{l \neq k} \pi_l p_l(\mathbf{X}) c(k|l) ; j \neq k = 1, \dots, q.$$

Esta regla se conoce como *regla de Bayes*.

---

## Métodos en el dominio del tiempo. Caso univariante

---

Supongamos que un proceso estacionario gaussiano  $\{X(t)\}_{t \in \mathbf{Z}}$  pertenece a una de las categorías  $\Pi_1$  ó  $\Pi_2$ . Pertenecer a la categoría  $\Pi_j$  implica que la serie temporal

$$\mathbf{X} = (X(1), \dots, X(n))', \text{ de tamaño } n \times 1,$$

es estacionaria gaussiana con vector de media

$$\mu_j = (\mu_j(1), \dots, \mu_j(n))'$$

y matriz de covarianzas

$$\Sigma_j = \{\sigma_j(t - s); s, t = 1, \dots, n\}.$$

---

La densidad de probabilidad de  $\mathbf{X}$  bajo  $\Pi_j$  es

$$p_j(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma_j|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_j)' \Sigma_j^{-1} (\mathbf{x} - \mu_j)\right\},$$

por lo que la regla de Bayes implica que asignamos  $\mathbf{X}$  a  $\Pi_1$  si

$$D(\mathbf{X}) = -\frac{1}{2} \left\{ \log \frac{|\Sigma_1|}{|\Sigma_2|} + (\mathbf{X} - \mu_1)' \Sigma_1^{-1} (\mathbf{X} - \mu_1) - (\mathbf{X} - \mu_2)' \Sigma_2^{-1} (\mathbf{X} - \mu_2) \right\}$$

es mayor que  $K = \log(\pi_2/\pi_1)$ .

---

## Métodos de remuestreo

- Datos i.i.d.
- Series temporales.

Los métodos de remuestreo evalúan los estadísticos en remuestras o submuestras, y construyen estimadores a partir de esas evaluaciones. Los más conocidos son el *jackknife* y el *bootstrap*.

Pueden dividirse en dos grupos, dependiendo de si presuponen un modelo teórico para los datos: el enfoque **basado en modelo** (BM) y el enfoque **no basado en modelo** (NBM).

El remuestreo se aplica en AD principalmente para la estimación de las probabilidades de error o probabilidades de clasificación incorrecta.

---

## Bootstrap

Se basa en el principio de analogía (*plug-in*). Sea el parámetro  $\theta = \theta(P)$ , donde  $P$  es el modelo estadístico postulado. Un estimador plug-in o análogo es  $\hat{\theta} = \theta(\hat{P})$ , donde  $\hat{P}$  es un estimador de  $P$ .

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_N)$  un conjunto de datos del modelo  $P$ , y sea  $T(\mathbf{X})$  el estadístico cuya distribución  $\mathbf{L}(T; P)$  deseamos estimar. El método bootstrap propone como estimador de  $\mathbf{L}(T; P)$  la distribución  $\mathbf{L}^*(T^*; \hat{P})$  del estadístico  $T^* = T(\mathbf{X}^*)$ , donde  $\mathbf{X}^*$  es un conjunto de datos generado por el modelo estimado  $\hat{P}$ .

---

## Bootstrap en series temporales

### Métodos basados en modelos

Su principal característica es el uso que hacen de los residuos que aparecen con el modelo  $P$  teórico propuesto para el modelo estimado  $\hat{P}$ . Por hipótesis estos residuos son independientes entre sí.

También en este caso utilizamos el procedimiento de Monte Carlo. Veamos cómo se llevaría a cabo en el caso de un modelo autorregresivo de orden  $p$ :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

donde  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  es una sucesión de v.a.i.i.d. con  $E[\varepsilon_t] = 0$  y  $E[\varepsilon_t^2] = \sigma_\varepsilon^2$ .

- 
1. Obtener estimaciones a partir de  $\mathbf{X} = (X_1, X_2, \dots, X_N)$  de los parámetros autorregresivos  $\phi = (\phi_1, \dots, \phi_p)$ .
  2. Calcular los residuos siguientes,  $\hat{\varepsilon}_t = X_t - \sum_{i=1}^p \hat{\phi}_i X_{t-i}$ , para  $t = p+1, p+2, \dots, N$ .
  3. Calcular la función de distribución empírica de los residuos centrados,  $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - (N-p)^{-1} \sum_{t=p+1}^N \hat{\varepsilon}_t$ , mediante

$$F_N^{\tilde{\varepsilon}}(x) = \frac{1}{N-p} \sum_{t=p+1}^N I(\tilde{\varepsilon}_t \leq x).$$

---

Después se deben repetir los siguientes pasos para  $b = 1, 2, \dots, B$ .

4. Obtener  $N - p$  observaciones i.i.d. según  $F_N^{\tilde{\varepsilon}}$ , que denotamos por  $(\varepsilon_{p+1}^{*(b)}, \varepsilon_{p+2}^{*(b)}, \dots, \varepsilon_N^{*(b)})$ .
5. Fijar los primeros  $p$  valores de la serie  $(X_1^{*(b)}, X_2^{*(b)}, \dots, X_p^{*(b)})$  y calcular las restantes observaciones de la remuestra, es decir,  $(X_{p+1}^{*(b)}, X_{p+2}^{*(b)}, \dots, X_N^{*(b)})$ , mediante

$$X_t^{*(b)} = \sum_{i=1}^p \hat{\phi}_i X_{t-i}^{*(b)} + \varepsilon_t^{*(b)}.$$

6. Utilizando la remuestra  $\mathbf{X}^{*(b)} = (X_1^{*(b)}, \dots, X_N^{*(b)})$  calcular los análogos bootstrap de los estimadores.

---

## Utilización en análisis discriminante

Nuestros tres primeros ejercicios utilizan un enfoque basado en modelo, pero pueden extenderse fácilmente al enfoque no basado en modelo.

Hemos utilizado como modelo teórico el de los procesos  $AR(p)$ , es decir, los autorregresivos de orden  $p$ .

Nuestro objetivo es mostrar la validez del método bootstrap. Partimos de dos poblaciones cuyas distribuciones son conocidas (el modelo teórico no es incorrecto), e intentamos clasificar en una de ellas la nueva serie temporal.

---

## Algoritmo Bootstrap:

Se parte de la observación de  $r_j$  series de tamaño  $T$  de la clase  $\Pi_j$ , para  $j = 1, 2$ :

(1)  $X_t^{(1)} = \lambda_1 + \phi_1 X_{t-1}^{(1)} + \varepsilon_t^{(1)}$  es el modelo generado en la clase  $\Pi_1$ ,

y

(2)  $X_t^{(2)} = \lambda_2 + \phi_2 X_{t-1}^{(2)} + \varepsilon_t^{(2)}$  es el modelo generado en la clase  $\Pi_2$ .

que denotemos por  $\mathcal{X}^{(1)} = (\mathbf{X}^{(1,1)}, \dots, \mathbf{X}^{(1,r_1)})$  a las  $r_1$  series de  $\Pi_1$  y por  $\mathcal{X}^{(2)} = (\mathbf{X}^{(2,1)}, \dots, \mathbf{X}^{(2,r_2)})$  a las  $r_2$  series de  $\Pi_2$ .

1. Utilizando  $\mathcal{X}^{(1)}$  y  $\mathcal{X}^{(2)}$  obtenemos los estimadores mínimos cuadrados de  $\lambda_1$  y  $\phi_1$ , y de  $\lambda_2$  y  $\phi_2$ , respectivamente. Para ello escribimos las relaciones (1) y (2) como sigue:

$$\mathbf{Y}_j = \begin{pmatrix} X_2^{(j,1)} \\ \dots \\ X_T^{(j,1)} \\ \dots \\ X_2^{(j,r_j)} \\ \dots \\ X_T^{(j,r_j)} \end{pmatrix}, \mathbf{Z}_j = \begin{pmatrix} 1 & X_1^{(j,1)} \\ \dots & \dots \\ 1 & X_{T-1}^{(j,1)} \\ \dots & \dots \\ 1 & X_1^{(j,r_j)} \\ \dots & \dots \\ 1 & X_{T-1}^{(j,r_j)} \end{pmatrix} \text{ y } \varepsilon_j = \begin{pmatrix} \varepsilon_2^{(j,1)} \\ \dots \\ \varepsilon_T^{(j,1)} \\ \dots \\ \varepsilon_2^{(j,r_j)} \\ \dots \\ \varepsilon_T^{(j,r_j)} \end{pmatrix};$$

entonces se cumple que  $\mathbf{Y}_j = \mathbf{Z}_j \cdot \begin{pmatrix} \lambda_j \\ \phi_j \end{pmatrix} + \varepsilon_j$ .

---

Los estimadores mínimos cuadrados de  $\lambda_j$ ,  $\phi_j$  y  $\varepsilon^{(j)}$  se obtienen de:

$$\begin{pmatrix} \hat{\lambda}_j \\ \hat{\phi}_j \end{pmatrix} = (\mathbf{Z}'_j \mathbf{Z}_j)^{-1} \mathbf{Z}'_j \mathbf{Y}_j \text{ y } \hat{\varepsilon}_j = \mathbf{Y}_j - \mathbf{Z}'_j \begin{pmatrix} \hat{\lambda}_j \\ \hat{\phi}_j \end{pmatrix}.$$

2. Obtenemos la regla discriminante

$$D(\mathbf{X}) = -\frac{1}{2} \left\{ \log \frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_2|} + (\mathbf{X} - \hat{\mu}_1)' \hat{\Sigma}_1^{-1} (\mathbf{X} - \hat{\mu}_1) - (\mathbf{X} - \hat{\mu}_2)' \hat{\Sigma}_2^{-1} (\mathbf{X} - \hat{\mu}_2) \right\}.$$

El cálculo de la inversa y el determinante de las matrices  $\Sigma_1$  y  $\Sigma_2$ , de tamaño  $n \times n$ , se ha efectuado utilizando los resultados de Shaman (1976) y de Galbraith y Galbraith (1974), que nos permiten estimar directamente la inversa y el determinante de esas matrices. Para un modelo autorregresivo de orden  $p$ , consisten a invertir una matriz de orden  $p \times p$ .

---

3. Se remuestrea con reemplazo sobre los residuos centrados  $\tilde{\varepsilon}^{(j)} = \hat{\varepsilon}^{(j)} - \mathbf{1}\bar{\varepsilon}^{(j)}$ , donde  $\mathbf{1}$  es un vector  $r_j \times (T - 1)$  de unos, para obtener:

$$\varepsilon^{(j)*} = (\varepsilon_2^{(j,1)*}, \dots, \varepsilon_T^{(j,1)*}, \dots, \varepsilon_2^{(j,r_j)*}, \dots, \varepsilon_T^{(j,r_j)*})'$$

4. Se construyen las  $r_1 + r_2$  series bootstrap mediante:

$$X_t^{(j,k)*} = \hat{\lambda}_j + \hat{\phi}_j X_{t-1}^{(j,k)*} + \varepsilon_t^{(j,k)*}.$$

5. Obtenemos la regla discriminante bootstrap  $D^*$  de manera similar a 2.

6. Se estiman las probabilidades de error mediante:

$$e(2|1) = \frac{1}{r_1} \sum_{i=1}^{r_1} I(D^*(\mathbf{X}^{(1,i)}) \leq K)$$

y

$$e(1|2) = \frac{1}{r_2} \sum_{i=1}^{r_2} I(D^*(\mathbf{X}^{(2,i)}) > K).$$

---

# EJERCICIO 1

Matrices de covarianza iguales  
y medias distintas

	$r = 25$		$r = 50$	
<b>Teóricos</b>				
n=25	0,3325	0,3325	0,3325	0,3325
n=50	0,2739	0,2739	0,2739	0,2739
n=100	0,2000	0,2000	0,2000	0,2000
<b>Simulación</b>				
n=25	0,3403 (0,0544)	0,3326 (0,0538)	0,3293 (0,0434)	0,3340 (0,0452)
n=50	0,2698 (0,0546)	0,2803 (0,0510)	0,2721 (0,0382)	0,2795 (0,0367)
n=100	0,1954 (0,0460)	0,2086 (0,0489)	0,2001 (0,0293)	0,2019 (0,0290)
<b>Bootstrap</b>				
n=25	0,3342 (0,0561)	0,3343 (0,0525)	0,3366 (0,0404)	0,3374 (0,0421)
n=50	0,2699 (0,0525)	0,2687 (0,0519)	0,2760 (0,0333)	0,2747 (0,0351)
n=100	0,1976 (0,0435)	0,2002 (0,0432)	0,2044 (0,0295)	0,2038 (0,0289)

---

# EJERCICIO 2

Matrices de covarianza  
distintas y medias iguales

---

$r = 25$

$r = 50$

	$r = 25$		$r = 50$	
<b>Aproximados</b>				
n=25	0,2712	0,2487	0,2712	0,2487
n=50	0,1944	0,1686	0,1944	0,1686
n=100	0,1115	0,0873	0,1115	0,0873
<b>Simulación</b>				
n=25	0,3005 (0,0477)	0,2326 (0,0373)	0,2991 (0,0338)	0,2284 (0,0299)
n=50	0,2013 (0,0411)	0,1630 (0,0313)	0,1953 (0,0305)	0,1626 (0,0267)
n=100	0,1062 (0,0286)	0,0839 (0,0241)	0,1044 (0,0216)	0,0873 (0,0154)
<b>Bootstrap</b>				
n=25	0,2936 (0,0482)	0,2322 (0,0441)	0,2903 (0,0338)	0,2340 (0,0284)
n=50	0,2036 (0,0374)	0,1684 (0,0350)	0,1985 (0,0280)	0,1597 (0,0235)
n=100	0,1036 (0,0304)	0,0867 (0,0252)	0,1029 (0,0211)	0,0873 (0,0198)

---

# EJERCICIO 3

## Matrices de covarianza y medias distintas

---

$r = 25$

$r = 50$

<b>Simulación</b>	$r = 25$		$r = 50$	
n=25	0,4012 (0,0684)	0,4169 (0,0609)	0,3870 (0,0502)	0,4098 (0,0413)
n=50	0,3479 (0,0647)	0,3736 (0,0568)	0,3442 (0,0403)	0,3627 (0,0376)
n=100	0,2881 (0,0499)	0,3157 (0,0525)	0,2955 (0,0399)	0,2991 (0,0373)
<b>Bootstrap</b>				
n=25	0,3714 (0,0557)	0,3869 (0,0482)	0,3752 (0,0464)	0,3968 (0,0398)
n=50	0,3372 (0,0522)	0,3494 (0,0525)	0,3351 (0,0415)	0,3573 (0,0393)
n=100	0,2805 (0,0460)	0,2949 (0,0463)	0,2823 (0,0353)	0,2969 (0,0350)

---

## Conclusiones del estudio de simulación

- En el caso en que conocemos las probabilidades de error, el bootstrap se aproxima a los valores exactos.
- El bootstrap se muestra igual de preciso que la simulación, incluso para tamaños muestrales pequeños.
- Como es de esperar, los resultados mejoran al aumentar el tamaño de las series,  $n$ , o el número de series,  $r$ .
- Estos tres ejercicios nos permiten afirmar que el procedimiento bootstrap propuesto ofrece resultados similares a los obtenidos por simulación.

---

## Ejemplo con datos reales. Enfoque NBM

Los datos han sido obtenidos de la página de Robert H. Shumway, en <http://anson.ucdavis.edu/~shumwaytsa.htm> y corresponden a las mediciones de las vibraciones producidas por ocho terremotos y por ocho explosiones.

Cada una de estas dieciséis series consta de 2048 datos, los 1024 primeros datos corresponden a las ondas primarias de la vibración y los siguientes a las ondas secundarias.

Se definen dos variables, y se construye un modelo bivalente. Cada una de estas variables es el logaritmo decimal de la máxima amplitud de uno de los dos tipos de ondas que componen la serie.

A continuación se describe el algoritmo que se ha utilizado para obtener las estimaciones, mediante bootstrap, de las probabilidades de error de la clasificación.

---

## Algoritmo bootstrap:

1. Se eligen al azar y con reemplazo 8 terremotos de entre los que tenemos en la muestra original; del mismo modo se sortean 8 explosiones. Es decir, la muestra original se compone de:

$$\mathbf{X}_i^{(j)} = (X_{i,1}^{(j)}, \dots, X_{i,1024}^{(j)}, Y_{i,1}^{(j)}, \dots, Y_{i,1024}^{(j)}) \quad \text{para } i = 1, \dots, 8 \quad j = 1, 2.$$

que podemos escribir en bloques de 128 datos:

$$\mathbf{X}_i^{(j)} = (P_{i,1}^{(j)}, \dots, P_{i,8}^{(j)}, S_{i,1}^{(j)}, \dots, S_{i,8}^{(j)}) \quad \text{para } i = 1, \dots, 8 \quad \text{y } j = 1, 2.$$

Obtenemos:

$$\mathbf{X}_{i_1}^{(j)} \quad \text{para } i_1 \in \{1, \dots, 8\} \quad \text{y } j = 1, 2,$$

donde la serie  $\mathbf{X}_{i_1}^{(j)}$  ha sido elegida al azar con sorteo equiprobable entre las series  $\mathbf{X}_i^{(j)}$ .

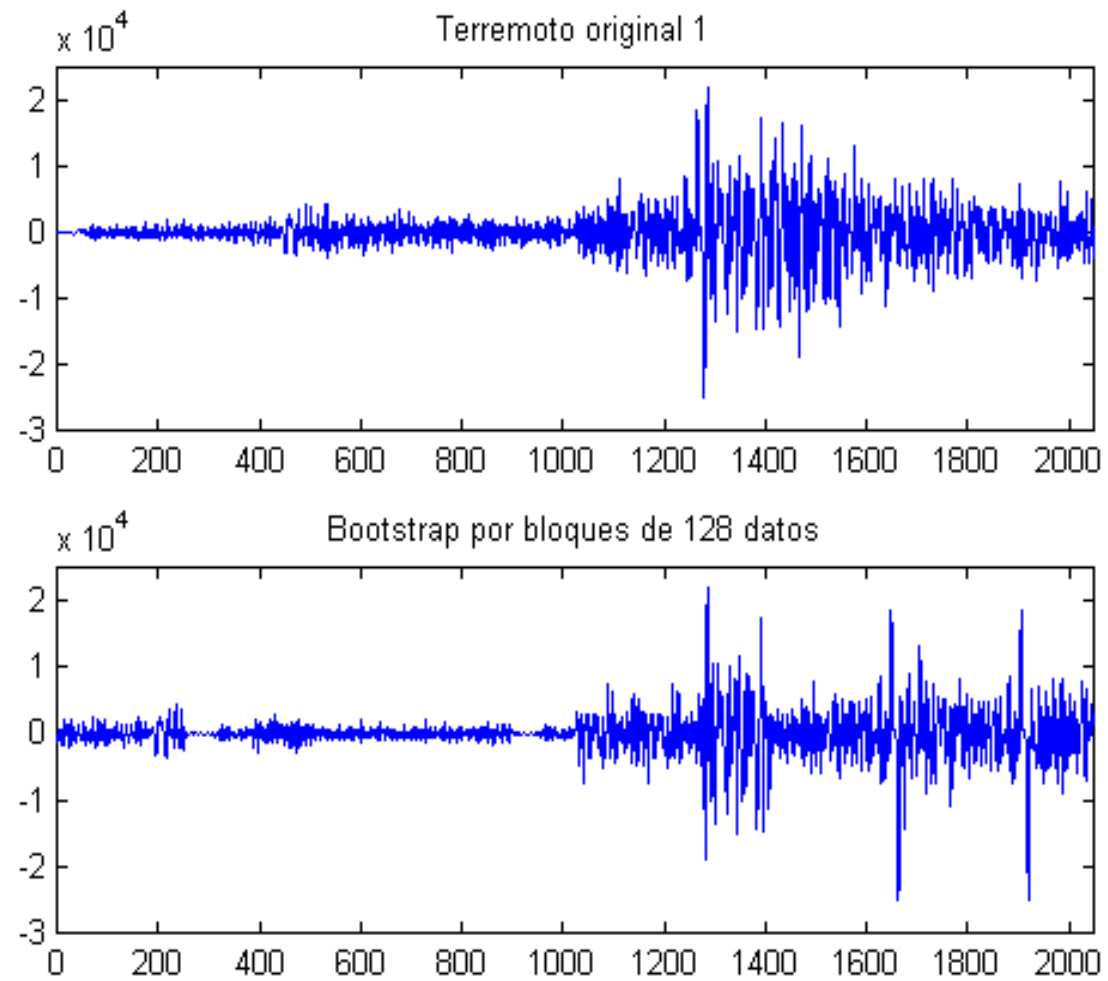
---

2. En cada una de las 16 series obtenidas en el paso 1., se remuestrea por bloques para construir una nueva serie. Este remuestreo se hace de modo que se respete la estructura en partes P y S de las ondas. Este remuestreo se hace aleatoriamente y con reemplazo usando bloques de 128 datos.

Es decir, después de este remuestreo por bloques, habremos obtenido

$$\mathbf{X}_{i_1}^{(j)*} = (P_{i,1}^{(j)*}, \dots, P_{i,8}^{(j)*}, S_{i,1}^{(j)*}, \dots, S_{i,8}^{(j)*}) \text{ para } i_1 = 1, \dots, 8 \text{ y } j = 1, 2;$$

donde cada  $P_{i,k}^{(j)*}$  es un bloque de datos consecutivos elegido entre los bloques  $P_{i,1}^{(j)}$  de la serie  $\mathbf{X}_{i_1}^{(j)}$ ; y cada  $S_{i,1}^{(j)*}$  es un bloque del mismo tipo elegido al azar entre los bloques  $S_{i,1}^{(j)}$  de la misma serie.



---

3. Tenemos ahora 8 terremotos y 8 explosiones obtenidas con el remuestreo descrito en el apartado anterior. A partir de estas 16 series construimos la tabla con los valores de las variables que hemos definido. Es decir, que construimos una tabla análoga a la tabla calculada para los datos originales.

4. A partir de los valores de esta tabla así construida, estimamos las medias y las matrices de covarianza de las dos clases: terremotos y explosiones. Tenemos así construida una regla discriminante, basada en la evaluación de la expresión vista en la teoría del apartado 2.1.2 para el caso multidimensional.

$$\widehat{D}^*(\mathbf{X}) = -\frac{1}{2} \left\{ \log \frac{|\widehat{\Sigma}_1^*|}{|\widehat{\Sigma}_2^*|} + (\mathbf{X} - \widehat{\mu}_1^*)' \widehat{\Sigma}_1^{*-1} (\mathbf{X} - \widehat{\mu}_1^*) - (\mathbf{X} - \widehat{\mu}_2^*)' \widehat{\Sigma}_2^{*-1} (\mathbf{X} - \widehat{\mu}_2^*) \right\}.$$

---

5. Aplicamos esta regla discriminante a las series originales. Estimamos las probabilidades de error mediante la proporción de los datos originales que son mal clasificados. Es decir:

$$P(2|1) = \frac{1}{8} \sum_{i=1}^8 I(\hat{D}^*(\mathbf{X}_{i_1}^{(1)}) \leq 0)$$

y

$$P(1|2) = \frac{1}{8} \sum_{i=1}^8 I(\hat{D}^*(\mathbf{X}_{i_1}^{(2)}) > 0).$$

6. Repetimos todos los pasos de 1 a 5 un número de veces  $B$ , y al final calculamos la media aritmética de todas las estimaciones anteriores.

$B$	$P(2 1)$ (std)	$P(1 2)$ (std)
<b>500</b>	0,096 (0,012)	0,019 (0.058)
<b>1000</b>	0,085 (0,012)	0,021 (0.062)

---

## Trabajos futuros

---

- El apartado 2.2 del trabajo presentaba los Métodos de Dominio de las Frecuencias del AD, y nosotros nos hemos centrado en el bootstrap aplicado a los Métodos de Dominio del Tiempo, por lo que quedaría pendiente el estudio del bootstrap para estos métodos.
- Nosotros hemos trabajado con series multivariantes en el ejemplo con datos reales que se ha presentado (apartado 3.2.3.2.). Pero, puesto que no abundan en los estudios de otros autores, son necesarios más estudios de simulación con series multidimensionales.
- Abordaremos el estudio teórico de la validez asintótica del método bootstrap en los distintos casos que hemos considerado en el presente trabajo.