

# MÉTODOS BOOTSTRAP: Principios, Teoría y su aplicación al Procesamiento Digital de Señales

José Ismael de la Rosa Vargas\*

Universidad Autónoma de Zacatecas - Unidad Académica de Ingeniería Eléctrica,  
Lab. de Procesamiento Digital de Señales, Av. López Velarde No. 821, Col. Centro,  
C.P. 98068, Zacatecas, Zacatecas.

Versión 1.2

30 de junio de 2003

## Resumen

El propósito de este curso es de introducir algunas técnicas de remuestreo que han sido explotadas recientemente para llevar a cabo la caracterización estadística de sistemas físicos (cálculo de intervalos de confianza, de los primeros momentos estadísticos, de la densidad de probabilidad completa o densidades completas que interactúan en un sistema bajo estudio, etc.). Lo anterior se hace con la finalidad de clasificar los estimadores adaptados a un objetivo particular de modelado, o bien para llevar a cabo la selección de modelos puestos en competencia para representar algún sistema físico predeterminado. En nuestro caso, nos interesamos en los sistemas que tienen que ver con el Procesamiento Digital de Señales (PDS). La caracterización estadística se lleva a cabo suponiendo que no se conoce la densidad de probabilidad o densidades de probabilidad que interactúan en el sistema bajo estudio, por ejemplo, la densidad de los errores de observación o de adquisición  $\varphi(\mathbf{e})$  es supuesta desconocida. Se presenta un panorama general de los métodos existentes y las nuevas tendencias, y bajo que condiciones pueden ser utilizados. Se brindan varios ejemplos de aplicación tanto en instrumentación como en PDS en general.

**Palabras clave:** Procesamiento Digital de Señales, Bootstrap, MCMC, Estimación Paramétrica, Estimación No-paramétrica, Medidas Indirectas, Estimación Robusta.

---

\* Autor de los apuntes, e-mail:ismaelrv@ieee.org



## Índice

<b>1. Definiciones propias de la estadística y algunas notaciones</b>	<b>5</b>
1.1. Notaciones generales . . . . .	5
1.2. Notaciones utilizadas por los métodos bootstrap . . . . .	6
<b>2. Introducción General</b>	<b>7</b>
<b>3. Bootstrap (Marco teórico)</b>	<b>9</b>
3.1. Convergencia y consistencia de la DDP bootstrap . . . . .	11
3.2. Bootstrap caso no-paramétrico . . . . .	14
3.3. Bootstrap caso paramétrico . . . . .	15
3.4. Métodos de Jackknife . . . . .	16
3.5. Ejemplos caricaturales . . . . .	20
<b>4. Aplicaciones en instrumentación</b>	<b>20</b>
4.1. Regresión lineal . . . . .	21
4.2. Regresión no-lineal . . . . .	22
4.3. Ecuación de medición . . . . .	23
4.4. Densidad bootstrap para los parámetros . . . . .	23
4.5. Densidad bootstrap de las mediciones . . . . .	25
<b>5. Mejoras en la convergencia de bootstrap</b>	<b>26</b>
5.1. Bootstrap aplicado a los residuos . . . . .	26
5.2. Otros esquemas bootstrap en problemas regresivos . . . . .	28
5.3. Bootstrap ponderado iterativo . . . . .	29
5.4. Desarrollo en series de Edgeworth para el esquema bootstrap ponderado iterativo . . . . .	31
<b>6. Robustez, optimalidad y eficacia de bootstrap</b>	<b>33</b>
6.1. Optimalidad del estimador de mínimos cuadrados (MC) (Nonlinear Least Squares–NLS) . . . . .	33
6.2. Robustez inducida pour un esquema bootstrap sesgado . . . . .	34
6.3. Robustez vs. eficacia . . . . .	36
<b>7. Aplicación a un sistema de mediciones</b>	<b>37</b>

<b>8. Aplicaciones en PDS</b>	<b>45</b>
8.1. Modelos AR, MA, ARMA, ARN, etc. . . . .	47
8.2. Bootstrap en el caso de hipótesis de correlación . . . . .	50
8.3. Señales de voz (identificación, síntesis, etc.) . . . . .	51
8.4. Modelado de imágenes . . . . .	53
8.4.1. Descripción de los algoritmos y formulación del problema . . . . .	54
<b>9. Relación con la selección de modelos y construcción de límites de confianza</b>	<b>56</b>
9.1. Selección de modelos o del orden de un modelo anidado . . . . .	56
9.2. Estimación de límites de confianza . . . . .	57
<b>10. Resumen y comentarios</b>	<b>57</b>
<b>A. Simulación por medio de Monte-Carlo</b>	<b>61</b>
A.1. Solución de integrales haciendo uso de Monte-Carlo . . . . .	62
A.2. Técnicas de reducción de variancia (TRV) . . . . .	64
<b>B. Desarrollo de ecuaciones recursivas QP y NSHP</b>	<b>65</b>
<b>Referencias</b>	<b>69</b>

## 1. Definiciones propias de la estadística y algunas notaciones

**Definición 1.1 (Consistencia)** *Todo estimador  $\hat{\boldsymbol{\theta}}$  converge asintóticamente ( $n \rightarrow \infty$ ) hacia el valor real de  $\bar{\boldsymbol{\theta}}$ .* □

**Definición 1.2 (Normalidad)** *El error de estimación  $e_e = \sqrt{n} (\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})$  converge asintóticamente en distribución hacia una ley  $\mathcal{N}(0, \mathbf{F}(\bar{\boldsymbol{\theta}})^{-1})$ .* □

**Definición 1.3 (Eficacidad)** *La matriz de variancia/covariancia de  $\bar{\boldsymbol{\theta}}$ :  $\mathbf{F}(\bar{\boldsymbol{\theta}})^{-1}$ , converge asintóticamente hacia el límite inferior de Cramer-Rao,  $\mathbf{F}(\bar{\boldsymbol{\theta}})^{-1}$  representa la inversa de la matriz de información de Fisher.* □

### 1.1. Notaciones generales

---

$E \{ \cdot \}$	Operador para la esperanza matemática
$\text{tr} \{ \cdot \}$	Operador para la traza
$P \{ \cdot \}$	Operador de probabilidad
$\mathcal{I} \{ \cdot \}$	Función índice
$a, \mathbf{a}, \mathbf{A}$	Escalar, vector y matriz
$\mathbf{a}^\top, \mathbf{A}^\top$	Vector transpuesto y matriz transpuesta
$\{a_i\}_{i=1}^k$	Vector $\mathbf{a}$ de dimensiones $1 \times k$ , $i = 1, \dots, k$
$y_i, \mathbf{y}$	Vector de observaciones (datos brutos)
$x_i, \mathbf{x}$	Protocolo experimental o parámetros instrumentales
$f(\mathbf{x}, \boldsymbol{\theta}), f(\cdot)$	Función modelo parametrada por $\mathbf{x}$ y $\boldsymbol{\theta}$
$\boldsymbol{\theta}$	Vector de parámetros desconocidos
$\mathbb{M}$	Espacio a priori de la medición
$\Theta$	Espacio a priori de los parámetros
$g(\boldsymbol{\theta}), g(\cdot)$	Funcional sobre los parámetros para la medición
$G_k(\boldsymbol{\theta}), G_k(\cdot)$	Relación biyectiva entre los parámetros y la medición
$f(0)$	Función modelo evaluada en cero
$f'(x), f''(x), f^{(n)}(x)$	Derivada primera, segunda, et n-ésima de la función modelo

$m_P$	Medición de la pendiente en el origen
$m_I$	Medición de la integral sobre los datos observados
$\mathcal{J}(\boldsymbol{\theta})$	Criterio o función de costo sobre $\boldsymbol{\theta}$
$\wp(\boldsymbol{\theta})$	Distribución de los parámetros
$\mathbf{W}_\theta$	Matriz de covariancia de los parámetros
$\mathbf{b}_\theta, \mathbf{b}(\boldsymbol{\theta})$	Vector de sesgos de los parámetros
$\wp(m), \wp(\mathbf{m})$	Distribución de la medición
$\mathbf{b}(m)$	Sesgo de la medición
$\wp(\mathbf{e})$	Distribución de los errores
$\Sigma$	Matriz de covariancia de los errores
$\Sigma^{-1}$	Inversa de la matriz de covariancia de los errores
$\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}$	Estimadores de los parámetros y parámetros reales
$\hat{\boldsymbol{\theta}}_L, \hat{\boldsymbol{\theta}}_{NL}$	Estimadores para un modelo lineal y un modelo no-lineal en los parámetros
$\hat{\boldsymbol{\theta}}_{MC}, \hat{\boldsymbol{\theta}}_{MV}, \hat{\boldsymbol{\theta}}_{MAP}$	Estimadores de mínimos cuadrados, de máximo de verosimilitud y de máximo a posteriori
$\mathbf{F}(\boldsymbol{\theta})$	Matriz de información de Fisher
$p$	Dimensiones del vector $\boldsymbol{\theta}$
$n$	Dimensiones de los vectores $\mathbf{y}$ , $\mathbf{x}$ y $\mathbf{e}$
$N$	Número de simulaciones según los métodos de tipo Monte-Carlo

## 1.2. Notaciones utilizadas por los métodos bootstrap

---

$Z$	Variable aleatoria auxiliar
$z_i, \mathbf{z}$	Vector de variables aleatorias i.i.d.
$R(Z, F_z)$	Estimador de una estadística cualquiera de $z_i$ , según $F_z$
$F_z,$	Distribución de $z_i$
$F_z(\hat{R})$	Distribución parametrada de $z_i$
$\hat{F}_z,$	Distribución empírica de $z_i$
$z_i^*$	Vector de remuestras bootstrap
$R^*(Z, F_z)$	Estimador bootstrap de una estadística cualquiera de $z_i$ , según $\hat{F}_z$

$\hat{e}_i, \hat{\mathbf{e}}$	Vector de residuos
$\hat{\rho}_e, \hat{\rho}(\mathbf{e})$	Distribución empírica de los residuos
$e_i^*, \mathbf{e}^*$	Vector de residuos bootstrap
$y_i^*, \mathbf{y}^*$	Datos fictivos bootstrap
$\hat{\boldsymbol{\theta}}^*$	Estimadores paramétricos bootstrap
$\hat{m}^*, \hat{\mathbf{m}}^*$	Estimador bootstrap de la medición
$\varphi(\hat{m}^*), \varphi(\hat{\mathbf{m}}^*)$	Distribución bootstrap de la medición
$\varphi(\hat{\boldsymbol{\theta}}^*)$	Distribución bootstrap de los parámetros
$h_i, w_i$	Vectores de pesos para el bootstrap externo y el bootstrap sesgado
$\xi(\varepsilon)$	Función de punto de ruptura
$\gamma(w_i)$	Medida de la dispersión de los datos
$\Delta_H(\cdot)$	Distancia de Hellinger
$B$	Número de simulaciones según los métodos bootstrap
$B_1, B_2$	Número de simulaciones para el bootstrap interno y externo

## 2. Introducción General

En este curso presentamos una familia de técnicas de remuestreo fundamentadas en la técnica conocida como “bootstrap.” Entre los ejemplos presentados, comentamos su aplicación a un sistema de interés llamado “sistema de medición” el cual se desea caracterizar estadísticamente (referirse a la sección § 4). La metodología bootstrap se encuentra definida en la literatura de manera muy genérica, se dice que: “Los métodos bootstrap tienen la capacidad de simular un proceso un gran número de veces, lo cual desearía realizar un experimentador en la práctica (en una gran cantidad de problemas esto no es posible)” [14], [56], [65]. Durante los últimos 15 años, dichos métodos han sido bastante explotados, puesto que representan una herramienta de simulación altamente potente y se aplican de manera simple para realizar inferencias estadísticas sobre sistemas físicos complejos.

La publicación del primer artículo de Bradley Efron en 1979 [21] sobre las técnicas bootstrap, causó un gran revuelo en la comunidad de estadísticos, su trabajo sintetiza las ideas de su época sobre los métodos de remuestreo y establece un nuevo marco de referencia sobre el análisis estadístico basado en los métodos de simulación (simulación estocástica). La idea de reemplazar aproximaciones complicadas y con frecuencia inexactas del sesgo (bias), la variancia (variance) y otro tipo de medidas de incertidumbre captó la atención de estadísticos teóricos y experimentistas. Un primer paso, fué vencer el escepticismo de los teóricos, quienes después de cierto tiempo dedicaron sus esfuerzos a la demostración

e investigación matemática de las condiciones bajo las cuales convergen dichos métodos de simulación ([32], [33], [56], entre algunos de ellos).

La implementación práctica de los métodos bootstrap implica la utilización de la simulación estocástica y el uso del remuestreo al mismo tiempo. La simulación estocástica permite en todo caso obtener medidas empíricas de incertidumbre como los son: los intervalos de confianza, los primeros momentos estadísticos, la densidad de probabilidad (DDP) (entre otros) de estimadores  $\hat{\theta}$  (para ejemplificar), y permite formalizar las nociones de inyección de información apriori, de modelado probabilístico, de verosimilitud, de errores estandar, etc. Los métodos bootstrap son técnicas muy utilizadas en contextos donde la inferencia estadística obtenida mediante técnicas analíticas convencionales es invalidada (en problemas complejos, este tipo de aproximaciones pueden ser tediosas y con resultados potencialmente inapropiados debido a suposiciones o simplificaciones erróneas, otra causa puede ser que la dimensión  $n$  del vector muestra es muy pequeña). Así, los métodos de remuestreo se substituyen por las derivaciones teóricas necesarias para aplicar los métodos analíticos de la estadística tradicional (por ejemplo, la linealización utilizando un desarrollo en series de Taylor). El remuestreo del vector muestra (o de muestras) original es por otro lado una operación extremadamente simple. Como ya se comentó con anterioridad, Efron fué el primero en proponer la idea de construir una población de muestras fictiva (una reacción inicial en el caso no-paramétrico describía que esto era un fraude, pero no es así) de manera empírica [21]. El nombre de “bootstrap” proviene del mismo principio de regeneración de muestras<sup>1</sup>. En fin, ir más allá de los primeros momentos y obtener una aproximación de la DDP completa, de los momentos de orden superior o de los límites de confianza de un estimador puede reflejarse necesario en algunas aplicaciones. Finalmente, los métodos bootstrap pueden ser vistos como técnicas para mejorar la precisión y el tiempo de convergencia de los métodos de tipo Monte-Carlo [63].

La primer sección establece algunas definiciones y conceptos básicos de la terminología utilizada a lo largo de todo este documento. En la tercera sección de este curso (§ 3), establecemos el marco teórico de los métodos bootstrap, así como la división de los diferentes métodos en términos generales (bootstrap paramétrico y bootstrap no-paramétrico) y su relación con los métodos conocidos como “Jackknife”. La introducción de un esquema reciente llamado *bootstrap ponderado* mejora la respuesta del esquema clásico de bootstrap no-paramétrico (ver la sección § 5). En la cuarta sección, nos hemos interesado en presentar la caracterización estadística de estimadores paramétricos descritos según los

---

<sup>1</sup>El procedimiento descrito por el Baron Munchausen [14], quien cuenta que un día se encontraba bloqueado en el fondo de un lago, y pudo salir tirando él mismo sobre las cintas de sus botas “bootstraps”. *La Historia Cómica que contiene los Estados e Imperios de la Luna* (1657) de Cyrano de Bergerac retraza una historia análoga en donde el héroe alcanza la luna tirando él mismo sobre las lengüetas de sus botas [54].

modelos regresivos lineal y no-lineal, es decir  $\hat{\theta}$  (ver las ecuaciones (27) y (30)), los modelos regresivos pueden ser vinculados a un estimador al cual llamaremos estimador de la medición o cantidad de interés  $\hat{m} = g(\hat{\theta})$  [15], [16] (en un marco aplicativo dentro de la instrumentación). En quinta sección, presentamos un esquema propuesto en [16], [17] llamado bootstrap ponderado iterativo, buscando mejorar la convergencia del esquema bootstrap reduciendo el número de simulaciones sin descuidar el nivel de aproximación (precisión) dominando el error entre las DDP de interés y la DDP estimada. Por otro lado, el esquema de remuestreo de los residuos ponderados no es robusto con respecto a hipótesis de heteroscedasticidad, sin embargo, es posible mediante otros esquemas diferentes asegurar una robustez en los cambios de variancia que se presentan entre los mismos datos o muestras analizadas (sexta sección). En fin, otro tipo de esquemas recientes conocidos como *bootstrap intencionalmente sesgado*, propuestos por Hall [28]–[30], nos brindan ciertas propiedades de robustez pero en el caso concerniente a la presencia de datos o muestras aberrantes (ver la sección § 6). Ejemplificamos la aplicación de algunos de los métodos bootstrap en un sistema de medición en la séptima sección.

En la octava sección, presentamos nuestro interés en la caracterización estadística de estimadores más generales que pueden establecerse según un modelado/síntesis de señales de voz y de imágenes (aquí presentamos algunas referencias de la aplicación potencial de los métodos bootstrap en PDS: [14], [10], [48], [50]–[52], [65]–[71]). Finalmente, en las secciones nueve y diez se presentan algunos aspectos adicionales concernientes a la caracterización bootstrap como la selección de modelos y algunos comentarios.

### 3. Bootstrap (Marco teórico)

El problema de la caracterización de incertidumbres sobre modelos o sistemas físicos, suscita dos subproblemas. El primer problema es el de la construcción de una estadística de interés cualquiera cuando la dimensión del vector muestra es pequeña y por consiguiente el uso de métodos analíticos de tipo asintótico no nos brinda soluciones. El segundo problema se presenta cuando el cálculo analítico de la distribución de la estadística de interés es prácticamente imposible. En algunas aplicaciones industriales o de laboratorio, la repetición de una experiencia  $N$  veces puede ser altamente costosa o imposible, en este tipo de casos, el uso de los métodos bootstrap se muestra muy ventajoso por supuesto. En fin, como ya se comentó, los métodos bootstrap son similares a los diversos métodos de simulación, como los métodos de permutación, los métodos de jackknife [56] y a todos los métodos que hacen uso del remuestreo.

En los libros de texto de Kleijnen [37], y de Robert [54], y en el trabajo de tesis de Keramat [36], se introduce la idea fundamental sobre los métodos de Monte-Carlo (simulation estocástica)<sup>2</sup>. En general, los métodos de Monte-Carlo requieren del conocimiento previo de la ley  $F_z$  de las variables estocásticas conocidas como variables aleatorias de entrada  $Z$  que después de  $n$  realizaciones forman un vector aleatorio  $\mathbf{z} = \{z_1, \dots, z_n\}$ . En la práctica, el conocimiento previo de la ley de las variables  $z_i$  es una tarea en la mayoría de veces imposible o difícil de llevar a cabo, puesto que en algunas situaciones no se conoce la naturaleza estadística de las variables estocásticas dichas de entrada. Con el fin de compensar esta falta de conocimiento, Efron [21], Freedman [25], Hinkley [33], Wu [62], entre otros, han propuesto el uso de las técnicas de tipo jackknife y bootstrap con la finalidad de caracterizar estadísticamente las variables  $z_i$ .

El primero que propone la idea de utilizar los métodos bootstrap es obviamente Efron [21], y las diferentes variantes de éstos métodos han sido desarrolladas principalmente en la estimación de intervalos de confianza [20], [22], [66]<sup>3</sup>. Las estadísticas que en el caso paramétrico se busca aproximar pertenecen a una clase de estadísticas que son dependientes de la población de parámetros desconocidos, en dicho caso, la aproximación por la vía de los métodos bootstrap podría ser mejorada sensiblemente utilizando algunas técnicas diferentes, entre las existentes se pueden mencionar: *bootstrap iterativo y el uso simultaneo de pre-pivoteo* [22], [33], *los métodos de corrección del sesgo (biased bootstrap)* [22], [28]–[30] y *bootstrap ponderado* [34], [55], [56], [62].

El principio de los métodos bootstrap es entonces, la simulación de varias propiedades importantes de un procedimiento estadístico suponiendo que se tiene un mínimo de información acerca del modelo que será simulado. La evaluación de ciertas propiedades estadísticas para un estimador  $R(Z, F_z)$ , suponiendo que la distribución  $F_z$  de las variables  $z_i$  no se conoce, puede llevarse a cabo utilizando la distribución empírica  $\hat{F}_z$ . En donde la función  $R(\cdot)$  representará una estadística de interés cualquiera, la variancia, por ejemplo.

Muchas de las aproximaciones asintóticas propuestas en la literatura son poco realistas cuando se trata de analizar las propiedades estadísticas de un vector muestra de dimensión limitada. En dicho caso, los métodos bootstrap han demostrado que brindan buenos resultados en términos aproximativos de las estadísticas del vector muestra  $\mathbf{z}$ , así como de su densidad de probabilidad. Entonces debemos

---

<sup>2</sup>Ver un pequeño repaso en el Anexo A.

<sup>3</sup>Los métodos bootstrap han sido desarrollados particularmente para una clase de estadísticas cuyas distribuciones asintóticas son independientes de la población de parámetros desconocidos. Dicha clase de estadísticas se encuentra, luego que se busca estimar los intervalos de confianza de una variable aleatoria, dichas estadísticas son conocidas como: estadísticas asintóticamente pivotaes (en inglés “statistics asymptotically pivotal”).

tomar conciencia de que en la práctica, cuando se observa un vector aleatorio  $\{z_i\}_{i=1}^n$  de dimensión  $n$  a la salida de un proceso estocástico, generalmente su densidad de probabilidad  $F_z$  es desconocida,

$$z_i \sim F_z, \quad i = 1, \dots, n, \quad (1)$$

en donde  $F_z$  se encuentra definida como distribución positiva sobre la recta de los reales:  $\mathbb{R}^2$ . En fin, el objetivo principal es entonces obtener una característica estadística específica  $R(z, F_z)$ , la cual depende de las variables aleatorias  $z_i$  y de la distribución  $F_z$  la cual no se conoce. El primer paso a seguir, es estimar la distribución de la muestra  $R(\cdot)$  partiendo del conocimiento de las variables de entrada  $z_i$  que ya han sido observadas.

En la primera introducción de los métodos bootstrap, Efron [21] describe tres metodologías para calcular la distribución  $F_z$ :

**Método 1:** describe como llevar a cabo el cálculo teórico directo,

**Método 2:** se obtiene una aproximación de la distribución bootstrap por medio de los métodos de tipo Monte-Carlo,

**Método 3:** se realiza un desarrollo en series de Taylor para obtener una aproximación de la media y de la variancia de la distribución bootstrap (como en el caso de los trabajos sobre modelos regresivos no-lineales publicados por Box [7] y Clarke [12], [13]).

En aplicaciones prácticas, los métodos de tipo Monte-Carlo permiten la repetición de las realizaciones de  $z_i^*$  generadas a partir de las muestras aleatorias de dimensión  $n$ , gracias al conocimiento de la densidad empírica  $\hat{F}_z$ . Por lo tanto las realizaciones  $\{z^{*1}, \dots, z^{*B}\}$ , así como el histograma de los valores correspondientes a  $R(z^{*1}, \hat{F}_z), \dots, R(z^{*B}, \hat{F}_z)$  son considerados como aproximaciones de la verdadera distribución bootstrap. Es necesario remarcar que haremos una distinción entre  $N$  y  $B$ , puesto que  $N$  designará en general, el número de simulaciones necesarias según el método de Monte-Carlo, mientras que  $B$  designará el número de simulaciones necesarias según los métodos bootstrap. Es evidente que en la práctica, los métodos bootstrap son una mezcla de dos técnicas: el principio de remplazamiento y la aproximación por medio de Monte-Carlo.

### 3.1. Convergencia y consistencia de la DDP bootstrap

En general se considera el problema de la estimación de la función de distribución (DDP) de una variable aleatoria  $Z$ , y cuando se dispone de un número  $n$  de realizaciones, estas forman un vector aleatorio

$\mathbf{z}$  de dimensión  $n$ . Sea  $R(Z)$  una estadística cualquiera de  $Z$ , y  $\Gamma(z_i, F_z)$ ,  $i = 1, \dots, n$  la función de distribución de  $\mathbf{z}$  (Shao [56]), de tal manera que:

$$H_F(z) = P \{ \Gamma(z_i, F_z) \leq z \}, \quad (2)$$

donde  $z_i$  representa las muestras independientes e idénticamente distribuidas (i.i.d.) obtenidas a partir de  $F_z$ . En este caso, hacemos resaltar que  $H_F(z)$  depende de la dimensión  $n$  del vector muestra, pero por razones de simplicidad, omitiremos el índice  $n$ . Para llevar a cabo la estimación de la distribución de  $R(Z)$  se puede establecer simplemente la siguiente ecuación:

$$\Gamma(z_i, F_z) = R(Z). \quad (3)$$

Cuando  $R(Z)$  es utilizada para indicar la construcción de un intervalo de confianza para un parámetro  $\beta$  desconocido ligado a  $F_z$ , se utiliza la función  $R(Z) - \beta$  o una versión de Student de  $R(Z) - \beta$ , es decir:  $\frac{R(Z) - \beta}{S}$ , en donde  $S$  es un estimador de la desviación estandar de  $R(Z)$ . Entonces, en dicho caso la distribución sera:

$$\Gamma(z_i, F_z) = \frac{R(Z) - \beta}{S}. \quad (4)$$

En las aproximaciones estadísticas clásicas, se busca obtener una forma analítica de  $H_F(z)$ , ya sea de manera exacta, o bien aproximada. En el caso de la aproximación se reemplazan los parámetros desconocidos de la expresión teórica por sus respectivos estimadores (por ejemplo, los dos primeros momentos de una ley normal). En el caso de la aproximación bootstrap de  $H_F(z)$ . Primero se reemplaza  $F_z$  por  $\hat{F}_z$ , el cual es un estimador de  $F_z$ , enseguida se obtiene:

$$H_b(z) = H_{\hat{F}_z}(z) = P^* \left\{ \Gamma(z_i^*, \hat{F}_z) \leq z \mid Z_1, \dots, Z_n \right\}, \quad (5)$$

en donde  $z_i^*$  son las v.a.i.i.d. obtenidas a partir de  $\hat{F}_z$ ,  $P^* \{ \cdot \mid Z_1, \dots, Z_n \}$  denota la probabilidad condicional para cada una de las  $z_i$ . Si  $H_b(z)$  es una expresión explícita de las variables  $z_i$ , dicha expresión es el estimador bootstrap de  $H_F(z)$ . De otra manera, se puede utilizar el método de Monte-Carlo (Método 2 descrito por Efron en la sección anterior), para aproximar  $H_b(z)$ , es decir:

$$H_b^{(B)} = \frac{1}{B} \sum_{j=1}^B \mathcal{I} \left\{ \Gamma(z_i^{*j}, \hat{F}_z) \leq z \right\}, \quad (6)$$

en donde  $\{z_i^{*j}\}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, B$  son las muestras bootstrap i.i.d. obtenidas a partir de  $\hat{F}_z$ , y  $\mathcal{I}\{\cdot\}$  es la función índice.

La consistencia de  $H_b(z)$  como estimador de  $H_F(z)$  puede ser definida de manera general como se muestra a continuación:

**Definición 3.1 (Shao y Tu [56])** Sea  $\rho$  una medida métrica sobre  $\mathcal{F}_{\mathbb{R}^n} = \{ \text{Todas las distribuciones contenidas en el espacio } \mathbb{R}^n \}$ .  $H_b(z)$  es dicha  $\rho$ -consistente (o debilmente  $\rho$ -consistente) si  $\rho(H_b, H_F) \rightarrow 0$  en probabilidad, mientras que  $n \rightarrow \infty$ , y  $H_b(z)$  es dicha fuertemente  $\rho$ -consistente si  $\rho(H_b, H_F) \rightarrow 0$  con gran margen de seguridad, mientras que  $n \rightarrow \infty$ .  $\square$

La medida métrica  $\rho$  comunmente utilizada es  $\rho_\infty$ , es decir que la distancia generada es:

$$\rho(H_b, H_F) = \sup_z |H_F(z) - H_b(z)|. \quad (7)$$

Cuando se utiliza la distancia  $\rho_\infty$ , entonces se habla de  $\rho_\infty$ -consistencia. De manera práctica, se pueden utilizar una diversidad de distancias afines como la distancia de Hellinger utilizada en un ejemplo que se verá más adelante. Considerando la media del vector muestra como una estadística de interés, se obtienen más resultados sobre la consistencia de  $H_b(z)$  (teoremas 3.1 y 3.2 [56] y referencias citadas en este libro.).

**Teorema 3.1** Suponga que  $E \{ \|Z_1\|^2 \} < \infty$ ,  $R(Z_n) = g(\bar{Z}_n)$  y  $g$  es continuamente diferenciable en  $\mu = E \{ Z_1 \}$  con  $\nabla g(\mu) \neq 0$ . Entonces el estimador bootstrap  $H_b$  es fuertemente consistente para  $H_F$ , es decir que  $\rho_\infty(H_b, H_F) \rightarrow 0$  con gran margen de seguridad, mientras que  $n \rightarrow \infty$ .

El resultado sobre la consistencia de los estimadores bootstrap de la media del vector muestra provee de fundamentos para estudiar la consistencia de estimadores bootstrap para estadísticas más generales que pueden ser aproximadas por el mismo procedimiento que la media del vector muestra. Para mantener la consistencia del estimador bootstrap, la condición establecida es que el momento de segundo orden sea finito ( $E \{ Z_1^2 \} < \infty$ ).

**Teorema 3.2** Sea  $Z_1, \dots, Z_n$  un vector de variables aleatorias i.i.d. Si existe una función medible  $A_n$  de  $Z_1, \dots, Z_n$ , una secuencia de números estrictamente crecientes  $a_n \rightarrow \infty$ , y una función de distribución  $G(z)$  tal que

$$P^* \left\{ \frac{1}{a_n} \sum_{i=1}^n Z_i^* - A_n \leq z \right\} \rightarrow G(z)$$

con gran margen de seguridad, y entonces tendremos que  $a_n/\sqrt{n} \rightarrow 1$  y  $E \{ Z_1^2 \} < \infty$ .

En los casos en donde no existe regularidad sobre la consistencia de los estimadores bootstrap, la consistencia de  $H_b(z)$  puede ser asumida realizando algunas modificaciones durante el procedimiento de remuestreo, por ejemplo en el marco de la selección de funciones modelo, se propone variar la dimensión del vector muestra bootstrap (submuestreo). Shao [55], [56] introduce la justificación correspondiente

de la consistencia en algunos casos donde no existe regularidad (ver sección § 7.6 de Shao [56]), como en la selección de modelos. Según sus resultados, los métodos bootstrap consideran y utilizan una función de ponderación o suavizado (también conocido como “bootstrap ponderado”). Otros resultados sobre la consistencia del estimador  $H_b(z)$  en el caso de análisis de datos con modelos regresivos se brindan en las secciones § 7.5.3 y § 7.5.5 de Shao [56]. Si se asegura la consistencia del estimador  $H_b(z)$ , tal que  $\rho(H_b, H_F) \rightarrow 0$  al menos en probabilidad, entonces  $H_b(z)$  estará lo mas cercanamente posible de  $H_F(z)$ , y el nivel de convergencia y precisión del estimador  $H_b(z)$  dependerá de la dimensión del vector muestra y del número de simulaciones realizadas.

### 3.2. Bootstrap caso no-paramétrico

El método clásico de bootstrap, es mejor conocido como bootstrap no-paramétrico (conocido también como bootstrap uniforme). Este método será descrito enteramente en el parrafo siguiente, es un algoritmo bastante sencillo (al menos en principio):

- (a) el primer paso a seguir es la construcción de la función de probabilidad empírica  $\hat{F}_z$  a partir del vector muestra observado  $z_i$ , durante dicha construcción se asume que cada componente  $z_1, \dots, z_n$  tiene el mismo peso  $n^{-1}$  (probabilidad uniforme),
- (b) una vez que se conoce o se tiene la densidad empírica  $\hat{F}_z$ , el siguiente paso es simular u obtener un vector muestra de dimensión  $n$ , de manera aleatoria a partir de  $\hat{F}_z$ . El procedimiento anterior puede ser descrito de la siguiente manera:

$$z_i^* \sim \hat{F}_z, \quad i = 1, \dots, n, \quad (8)$$

en donde  $z_i^* = \{z_1^*, \dots, z_n^*\}$  es el vector muestra bootstrap. Es necesario mencionar que no se esta tomando una permutación del vector  $z_i$ , los valores de  $z_i^*$  son seleccionados por medio de remplazo en el conjunto  $\{z_1, \dots, z_n\}$ .

- (c) Finalmente, se obtiene la aproximación de la distribución de  $R(z, F_z)$  utilizando la distribución bootstrap de:

$$R^* = R(z^*, \hat{F}_z), \quad (9)$$

así, la distribución bootstrap de  $R^*$  se obtiene mediante el mecanismo aleatorio descrito por Efron como segundo método (aproximación por Monte-Carlo), consideramos que  $\hat{F}_z$  ha sido fijada según los valores observados de  $z_i$ .

**Comentario 3.1** Hacemos la remarca de que el punto mas importante es que la distribución de  $R^*$ , puede ser calculada teóricamente de manera exacta una vez que el vector  $z_i$  ha sido observado. Entonces se lleva a cabo la igualación de la distribución de  $R^*$  con la distribución que se quiere estimar de  $R$ , puesto que  $F_z = \widehat{F}_z$ . Un estimador no-paramétrico de la distribución de  $R$  brinda una buena estimación, si consideramos que no se tiene ningun conocimiento apriori sobre  $F_z$ . Asi pues, se espera que los resultados seran razonablemente precisos, ya que  $F_z = \widehat{F}_z$ , y por lo tanto  $R = R^*$  (según los resultados de consistencia presentados en la sección anterior).  $\square$

### 3.3. Bootstrap caso paramétrico

Por otro lado, se puede contar con información parcial acerca de la distribución  $F_z$ . En cuyo caso, se podría aprovechar el hecho de contar con dicha información apriori e inyectarla en el esquema de modelado y remuestreo bootstrap. Bajo estos términos, el remuestreo se llevará a cabo de forma paramétrica, como ejemplo, se podría contar con información sobre la pertenencia de  $F_z$  a una cierta familia parametrada de funciones de probabilidad, es decir que  $F_z = F_z(R)$  (cuando se utilizan los métodos bootstrap de esta manera, se les conoce como bootstrap paramétrico). Entonces, podemos simular vectores muestra de dimensión  $n$ , a partir de la estimación paramétrica de  $F_z$ , dicho en otros términos:

$$F_z(\widehat{R}) \longrightarrow \{z_1^*, \dots, z_n^*\}, \quad \text{entonces podemos obtener } z_i^* \sim F_z(\widehat{R}). \quad (10)$$

En las aplicaciones en donde se utilizan los métodos bootstrap de manera paramétrica, estos son susceptibles de brindar una mejor precisión en términos de resultados con respecto a los métodos bootstrap no-paramétricos. Evidentemente, lo anterior es válido si la información apriori que se tiene sobre  $F_z$  es correcta [56], [65].

En cualquiera de los casos de remuestreo, que este se realice utilizando métodos bootstrap paramétrico o no-paramétrico, algunas de las medidas de incertidumbre (momentos) a estimar se obtienen ayudandose de la simulación estocástica y de los estimadores:

#### Variación bootstrap

$$v^{*(B)} = \frac{1}{B} \sum_{j=1}^B (R_j^* - \overline{R}^*)^2, \quad \text{en donde } \overline{R}^* = \frac{1}{B} \sum_{j=1}^B R_j^*$$

#### Desviación estandar bootstrap

$$s^{*(B)} = \left( v^{*(B)} \right)^{1/2}$$

### Sesgo (bias) bootstrap

$$b^{*(B)} = \bar{R}^* - R$$

### Coefficiente de asimetría (skewness) bootstrap

$$sk^{*(B)} = \frac{\frac{1}{B} \sum_{j=1}^B (R_j^* - \bar{R}^*)^3}{(v^{*(B)})^{3/2}}.$$

Si existe la convergencia de la distribución bootstrap de  $R(z, F_z)$ , también existe la convergencia y consistencia de los momentos de  $R(z, F_z)$  (ver algunos resultados para el caso de la variancia en la sección § 3.2.2 de Shao y Tu [56]). Una condición de suficiencia para que exista la integrabilidad uniforme de la secuencia de variables aleatorias  $Z_1, \dots, Z_n$ , es que  $\sup_n E \{|Z_n|^{1+\delta}\} < \infty$  para  $\delta > 0$ . Para el caso de la consistencia de la variancia bootstrap se tiene que cumplir lo siguiente:

$$E \left\{ |\sqrt{n} (R^* - R)|^{2+\delta} \right\} = O_p(1) \quad (11)$$

para  $\delta > 0$  y  $n \rightarrow \infty$ .

### 3.4. Métodos de Jackknife

En esta sección presentamos de manera breve los métodos jackknife, dichos métodos son especialmente eficaces y consistentes para estimar los dos primeros momentos estadísticos. Al igual que los métodos bootstrap, para que la consistencia de los métodos jackknife sea respetada, el vector muestra deberá tener dimensión  $n$  tendiente al infinito. Sin embargo, según la teoría desarrollada para los métodos jackknife [56], [62] se ha demostrado que los estimadores jackknife de los dos primeros momentos son más precisos que sus equivalentes bootstrap, lo cual podría ser de uso ventajoso cuando se utiliza un esquema bootstrap paramétrico. En fin, el éxito de la consistencia y precisión de los estimadores jackknife se debe a la relación existente entre la estadística de interés  $R_n(\cdot)$  y su regularidad (smoothness), la cual puede ser caracterizada por la diferenciabilidad de la función que ha generado  $R_n(\cdot)$ .

Los métodos jackknife fueron introducidos en 1949 por Quenouille [53], quien propuso un estimador del sesgo eliminando un dato perteneciente al vector original  $\{z\}_i^n$  y recalculando dicho estimador una vez eliminado el  $j$ -ésimo dato  $z_{-j}$ . Sea  $R_n = R_n(Z_1, \dots, Z_n)$  el estimador de una estadística cualquiera  $R$ , el sesgo de  $R_n$  definido por:

$$b(R_n) = E \{R_n\} - R \quad (12)$$

Sea  $R_{n-1,j} = R_{n-1}(Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_n)$ ,  $j = 1, \dots, n$ , la estadística estimada pero basada en  $n - 1$  observaciones de  $\mathbf{z}$ . Entonces, el estimador del sesgo jackknife propuesto es:

$$b_{JACK}(R_n) = (n - 1)(\bar{R}_n - R_n) \quad \text{en donde} \quad \bar{R}_n = \frac{1}{n} \sum_{j=1}^n R_{n-1,j}. \quad (13)$$

Esto conduce al estimador jackknife de sesgo reducido de  $R$ , es decir

$$R_{JACK} = R_n - b_{JACK} = nR_n - (n - 1)\bar{R}_n \quad (14)$$

Los estimadores  $b_{JACK}(R_n)$  y  $R_{JACK}$  pueden ser justificados de manera heurística como sigue, sea

$$b(R_n) = \frac{a}{n} + \frac{b}{n^2} + O_p\left(\frac{1}{n^3}\right) \quad (15)$$

en donde los parámetros  $a$  y  $b$  son desconocidos, pero no dependen de  $n$ . Ahora, puesto que  $R_{n-1,j}$  son i.i.d., entonces

$$b(R_n) = \frac{a}{n-1} + \frac{b}{(n-1)^2} + O_p\left(\frac{1}{(n-1)^3}\right) \quad (16)$$

y el sesgo  $b(\bar{R}_n)$  tiene la misma expresión que (16). Entonces tendremos que

$$\begin{aligned} E\{b_{JACK}\} &= (n-1)[b(\bar{R}_n) - b(R_n)] \\ &= (n-1)\left[\left(\frac{1}{n-1} - \frac{1}{n}\right)a + \left(\frac{1}{(n-1)^2} - \frac{1}{n^2}\right)b + O_p\left(\frac{1}{n^3}\right)\right] \\ &= \frac{a}{n} + \frac{(2n-1)b}{n^2(n-1)} + O_p\left(\frac{1}{n^2}\right), \end{aligned}$$

lo cual significa que como estimador del sesgo de  $R_n$ ,  $b_{JACK}$  es corregido hasta un orden de  $n^{-2}$ . También se tiene que

$$b(R_{JACK}) = b(R_n) - E\{b_{JACK}\} = -\frac{b}{n(n-1)} + O_p\left(\frac{1}{n^2}\right),$$

y el sesgo de  $R_{JACK}$  es del orden de  $n^{-2}$ . El estimador jackknife produce un estimador de sesgo reducido eliminando el término de primer orden en la expresión de  $b(R_n)$ . Esta primera versión del método jackknife fué más popularizada desde que Tukey [59] demostró que el método jackknife podía ser utilizado para construir estimadores de la variancia. La justificación heurística es la siguiente, puesto que  $R_{JACK}$  puede escribirse como

$$R_{JACK} = \frac{1}{n} \sum_{j=1}^n \{nR_n - (n-1)R_{n-1,j}\}$$

Tukey define la siguiente expresión,

$$\tilde{R}_{n,j} = nR_n - (n-1)R_{n-1,j}, \quad j = 1, \dots, n, \quad (17)$$

esta expresión es denominada como pseudovalores de jackknife y asume dos conjeturas:

(A) Los pseudovalores  $\tilde{R}_{n,j}$  pueden ser tratados como si fuesen v.a.i.i.d.

(B) Los  $\tilde{R}_{n,j}$  tienen aproximadamente la misma variancia que  $\sqrt{n}R_n$ .

Considerando las conjeturas (A) y (B), es mas natural poder estimar  $\text{Var}(\sqrt{n}R_n)$  utilizando la aproximación de la variancia de la muestra  $\tilde{R}_{n,1}, \dots, \tilde{R}_{n,n}$ , es decir:

$$\begin{aligned} v_{JACK} &= \frac{1}{n(n-1)} \sum_{i=1}^n \left( \tilde{R}_{n,i} - \frac{1}{n} \sum_{j=1}^n \tilde{R}_{n,j} \right)^2 \\ &= \frac{(n-1)}{n} \sum_{i=1}^n \left( \tilde{R}_{n-1,i} - \frac{1}{n} \sum_{j=1}^n \tilde{R}_{n-1,j} \right)^2. \end{aligned} \quad (18)$$

Este estimador es conocido como *estimador jackknife de la variancia* de  $R_n$  donde se elimina un dato a la vez (delete-1).

Otro método sucesor del método anterior y más preciso aún (consistente) cuando la función que produce  $R_n$  es irregular (en términos de suavizado), es el método jackknife donde se eliminan  $d$  datos a la vez (delete- $d$ ) de un vector  $\{z\}_i^n$ . Este método además de ofrecer un mejor desempeño, también permite obtener una mejor aproximación de la densidad empírica de  $R_n$ , lo cual es difícil con el método simple introducido por Quenouille. Los resultados teóricos que fundamentan los dos métodos jackknife descritos en esta sección, se pueden consultar en el texto de Shao y Tu [56], por el momento presentamos de manera breve algunos estimadores obtenidos según el método de jackknife con eliminación- $d$ , donde se calculan y utilizan de manera repetitiva las estadísticas de la forma:

$$R_{r,\mathbf{s}} = R_r(Z_j, j \in \mathbf{s}^c), \quad (19)$$

en donde  $\mathbf{s}$  es un subconjunto de  $\{1, \dots, n\}$  de dimensión  $d$ ,  $\mathbf{s}^c$  es el complemento de  $\mathbf{s}$  y  $d$  es un entero que está dado en función de  $n$  de tal manera que:  $1 \leq d \leq n$ , y  $r = n - d$ . Notamos que para una cierta estadística  $R_n$ ,  $R_{r,\mathbf{s}}$  es la misma estadística pero basada en  $r$  observaciones en lugar de  $n - 1$ , en donde se eliminan  $\{Z_j, j \in \mathbf{s}^c\}$  datos del vector original. Se puede decir que esta es una versión generalizada de jackknife en donde  $d = 1$ . La consistencia del estimador  $v_{JACK}$  según este segundo método se asegura cuando  $d \rightarrow \infty$  y  $n \rightarrow \infty$ . De manera práctica, el número de datos a eliminar depende de algunas medidas sobre la irregularidad de  $R_n$ . El estimador jackknife con eliminación- $d$  de la variancia  $v_{JACK-d}$  esta dado por:

$$v_{JACK-d} = \frac{r}{dN} \sum_{\mathbf{s} \in \mathcal{S}} \left( R_{r,\mathbf{s}} - \frac{1}{N} \sum_{\mathbf{s} \in \mathcal{S}} R_{r,\mathbf{s}} \right)^2, \quad (20)$$

en donde  $\mathcal{S}$  es la colección de todos los subconjuntos de  $\{1, \dots, n\}$  de dimensión  $d$  y  $N = \binom{n}{d}$  es el número total de subconjuntos en  $\mathcal{S}$ . Donde se demuestra fácilmente que cuando  $d = 1$ ,  $v_{JACK-d} = v_{JACK}$ . El cálculo de  $v_{JACK-d}$  cuando  $d$  es grande puede causar problemas, una forma de resolver lo anterior es utilizando dos métodos de submuestreo, como en el caso de bootstrap, los métodos son: *submuestreo balanceado*,

$$v_{BJACK-d} = \frac{r}{dm} \sum_{\mathbf{s} \in \mathcal{B}} \left( R_{r,\mathbf{s}} - \frac{1}{m} \sum_{\mathbf{s} \in \mathcal{B}} R_{r,\mathbf{s}} \right)^2, \quad (21)$$

en donde  $n \leq m \leq N$ , y  $\mathcal{B}$  es un bloque de diseño balanceado e incompleto, y *submuestreo aleatorio*,

$$v_{SJACK-d} = \frac{r}{dm} \sum_{t=1}^m \left( R_{r,\mathbf{s}_t} - \frac{1}{m} \sum_{k=1}^m R_{r,\mathbf{s}_k} \right)^2, \quad (22)$$

en donde se considera que se obtienen muestras aleatorias de  $\{\mathbf{s}_1, \dots, \mathbf{s}_m\}$ , el muestreo se puede hacer ya sea con remplazo o sin remplazo. Otra ventaja que ofrece el método de jackknife con eliminación- $d$  es la obtención de la distribución empírica o también llamado *histograma jackknife* de  $R_n$ . Hagamoslo para el caso más general del estimador  $R_n$  de cualquier estadística, donde se considera

$$H_n(z) = H_{n,F}(z) = P \left\{ \sqrt{n} (R_n - R) \leq z \right\} \quad (23)$$

y

$$\begin{aligned} H_{JACK}(z) &= P \left\{ \sqrt{\frac{nr}{d}} (R_{r,\mathbf{s}} - R_n) \leq z \right\} \\ &= \frac{1}{N} \sum_{\mathbf{s} \in \mathcal{S}} \mathcal{I} \left\{ \sqrt{\frac{nr}{d}} (R_{r,\mathbf{s}} - R_n) \leq z \right\} \end{aligned} \quad (24)$$

donde se nota que el estimador de la variancia  $v_{JACK-d}$  es equivalente a la variancia del histograma jackknife. El siguiente teorema avala la consistencia del histograma jackknife.

**Teorema 3.3** *Se asume que  $R_n$  admite la expansión en términos de*

$$R_n = R + \frac{1}{n} \sum_{i=1}^n \phi_F(Z_i) + T_n, \quad \text{donde } T_n = O_p(n^{-1/2}), \quad (25)$$

y donde  $R$  es el parámetro de interés desconocido,  $\phi_F(Z_1)$  tiene media cero y variancia  $\sigma_F^2 > 0$  y  $T_n$  es el término restante de la expansión. Si  $d$  es seleccionado según:  $d/n \geq \epsilon_0$  para algún  $\epsilon_0 > 0$  y  $r \rightarrow \infty$ , entonces

$$\|H_{JACK} - H_n\|_{\infty} \rightarrow 0, \quad (26)$$

en probabilidad, en donde  $H_n$  y  $H_{JACK}$  están dados por (23) y (24), respectivamente.

La prueba de este resultado se encuentra en [56] en la sección § 2.3.2, otros resultados aplicables en PDS se brindan por Politis y Romano [52].

### 3.5. Ejemplos caricaturales

Los ejemplos que están introducidos en el bootstrap toolbox...

## 4. Aplicaciones en instrumentación

En muchos problemas de instrumentación es necesario llevar a cabo un análisis de datos observados por medio de un modelado regresivo, los modelos regresivos paramétricos lineales y no-lineales son una opción que necesita de la estimación paramétrica. Para cuantificar la incertidumbre sobre los estimadores de los parámetros es necesario obtener una aproximación de la distribución de dichos estimadores  $\hat{\theta}$ , para obtener dicha aproximación se propone recurrir a los métodos bootstrap. La población de estimadores  $\hat{\theta}$ , se encuentra ligada en este caso a un conjunto de valores diferentes de datos  $\mathbf{y}$ , a los cuales se supone se han sumado un conjunto de variables aleatorias de entrada llamadas errores o también ruido (donde se tiene la equivalencia  $F_z \triangleq \varphi(e_i)$ ,  $i = 1, \dots, n$ ) afectando el sistema de instrumentación bajo estudio de manera completa.

En la literatura consultada se puede evidenciar una gran cantidad de aplicaciones en donde intervienen los métodos bootstrap y el uso del modelado regresivo. Mas concretamente, algunos métodos bootstrap fueron extendidos en una aplicación instrumental denominada sistema de mediciones [18], esta propuesta de sistemas nació en el Servicio (laboratorio) de mediciones de SUPELEC [15], [16], [17]. En el contexto del sistema de mediciones, se supone que se cuenta con una relación (generalmente no-lineal) existente entre los parámetros y la cantidad de interés denominada medición (referirse a la ecuación (31)). En fin, se dió preferencia del uso de los métodos bootstrap sobre los métodos de tipo jackknife, puesto que en general Efron demuestra [21], que los métodos jackknife son una aproximación lineal de los métodos bootstrap y además por medio de los métodos bootstrap es posible obtener una mejor aproximación de la distribución de los estimadores a ser caracterizados estadísticamente. Por otro lado, es cierto que luego que se utilizan los métodos jackknife es posible obtener estimaciones mas precisas del sesgo y la variancia, pero no la DDP completa. Cuando se dispone de información apriori y es posible utilizar un esquema de bootstrap paramétrico, sugerimos utilizar las dos aproximaciones de manera conjunta como lo propone también Shao y Tu [56].

El uso de los métodos bootstrap presenta otra ventaja muy práctica [61], puesto que algunos de estos métodos de remuestreo de tipo no-paramétrico consideran que se cuenta únicamente con v.a. de entrada  $e_i$  i.i.d., pero no se necesita el conocimiento apriori de la ley de dichas v.a. Esto significa que la DDP de los errores (ver la ecuación (27)) no es necesariamente conocida, sin embargo el problema

inverso puede ser resuelto en un marco probabilista. Evidentemente, en el caso de análisis regresivo paramétrico,  $R(\cdot)$  juega el rol de la estadística de interés sobre  $\hat{\boldsymbol{\theta}}$ , por ejemplo, el sesgo y la variancia.

#### 4.1. Regresión lineal

Muchas aplicaciones de la estadística se basan en un análisis de modelos estadísticos. Uno de los modelos más utilizados es el modelo regresivo lineal (simple o multivariable):

$$y_i = x_i^\top \boldsymbol{\theta} + e_i, \quad i = 1, \dots, n, \quad (27)$$

donde  $y_i$  representa la  $i$ -ésima respuesta (datos observados o censados, o bien, variables a explicar),  $x_i$  es el vector de variables explicativas de dimensión  $p$  que está asociado a  $y_i$  ( $y_i, x_i^\top$ ),  $\boldsymbol{\theta}$  representa el vector de dimensión  $p$  de parámetros desconocidos (a ser estimados). Los errores que modelizan el sistema como un proceso aleatorio están dados por  $e_i$ . Los modelos regresivos lineales brindan un gran número de grados de libertad, sin embargo esto significa que el número de parámetros a estimar podría llegar a ser considerablemente grande y esto a su vez aumentaría el número de ecuaciones a resolver y por lo tanto el tiempo de cálculo [4] (en algunas ocasiones la solución del sistema de ecuaciones sufre de inestabilidad numérica o peor aún, existen más incógnitas que ecuaciones). Por otro lado,  $\Theta_L \subset \mathbb{R}^p$  representa el subconjunto admisible de parámetros  $\boldsymbol{\theta}$  y  $\mathbb{X}_L \subset \mathbb{R}^n$  representa el dominio admisible por las variables explicativas.

En fin, cuando las variables explicativas  $x_i$  son determinísticas, generalmente se asume que los errores  $e_i$  son independientes de media 0 y variancias desconocidas  $\sigma_i^2$ . Cuando las variables explicativas  $x_i$  son aleatorias, el modelo regresivo es denominado modelo de correlación. En el modelo de correlación, los pares  $(y_i, x_i^\top)$  son supuestos como v.a.i.i.d. con momentos de segundo orden finitos y  $E\{y_i|x_i\} = x_i^\top \boldsymbol{\theta}$ ; donde  $\sigma_i^2$  denotará las variancias condicionales de  $y_i$  dado  $x_i$ , la esperanza condicional  $E\{\cdot|x_i\}$  y la variancia condicional  $\text{Var}(\cdot|x_i)$  normalmente se denotan como  $E\{\cdot\}$  y  $\text{Var}(\cdot)$ . Los parámetros de principal interés son el vector  $\boldsymbol{\theta}$ , conocidos mejor como parámetros de regresión aún y cuando el problema tratado no sea propiamente regresivo. El problema principal entonces trata de estimar los parámetros desconocidos  $\boldsymbol{\theta}$  para adecuar el modelo  $x_i^\top \boldsymbol{\theta}$  con respecto a  $y_i$ , también puede llevarse a cabo para propósitos de selección de modelos, predicción, etc.

En fin, los estimadores más comunes de  $\boldsymbol{\theta}$  son:

**Estimador de mínimos cuadrados:** Este estimador es óptimo cuando la variancia de los  $e_i$  es igual,

es decir  $\sigma_i^2 = \sigma^2$  (propiedad de homoscedasticidad) y los errores están centrados, entonces:

$$\hat{\boldsymbol{\theta}}_{MC} = \left( \mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{y}, \quad (28)$$

en donde  $\mathbf{X}^\top = [x_1, \dots, x_n]$ , siendo  $\mathbf{X}$  una matriz de rango  $p$  y  $\mathbf{y} = [y_1, \dots, y_n]^\top$ . El estimador  $\hat{\boldsymbol{\theta}}_{MC}$  es asintóticamente insesgado y equivalente al estimador de máximo de verosimilitud, es decir que  $E \left\{ \hat{\boldsymbol{\theta}}_{MC} \right\} = \boldsymbol{\theta}$ , es también normal y eficaz con matriz de variancia/covariancia dada por:

$$\text{Var}(\hat{\boldsymbol{\theta}}_{MC}) = \sigma^2 \left( \mathbf{X}^\top \mathbf{X} \right)^{-1}$$

y el estimador asintóticamente insesgado de la variancia está dado por:

$$\text{Var}(\hat{\boldsymbol{\theta}}_{MC}) = \left( \frac{1}{n-p} \sum_{i=1}^n r_i^2 \right) \left( \mathbf{X}^\top \mathbf{X} \right)^{-1}, \quad \text{en donde } r_i = y_i - x_i^\top \boldsymbol{\theta}.$$

**Estimador de mínimos cuadrados ponderado:** Este estimador considera que la variancia de  $e_i$  no es constante (propiedad de heteroscedasticidad), entonces  $\mathbf{e} \sim \mathcal{N}(0, \sigma_i^2 \mathbb{I}_n)$  y se asume que los  $e_i$  estan centrados, el estimador asociado es:

$$\hat{\boldsymbol{\theta}}_{MCP} = \left( \mathbf{X}^\top \mathbf{W} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}, \quad (29)$$

y es mas eficiente que el estimador  $\hat{\boldsymbol{\theta}}_{MC}$ ,  $\mathbf{W}$  es una matriz diagonal de pesos  $w_i$  proporcionales a la estimación de  $\sigma_i^{-2}$ .

**Estimadores robustos:** Estos estimadores consideran que existen algunos datos  $y_i$  aberrantes, entre ellos se pueden mencionar los M-estimadores propuestos por Huber [35], los estimadores de norma  $L_1$  [4], [14], [35], etc.

Notamos finalmente, que los dos primeros estimadores se resuelven encontrando generalmente una expresión cerrada.

## 4.2. Regresión no-lineal

Una extensión del modelo regresivo lineal, es el modelo regresivo no-lineal. Al igual que el modelo anterior este caracteriza el modelado de los datos observados  $y_i$ , pero ahora se lleva a cabo por medio de una función  $f(\cdot)$  parametrada (ver la ecuación (30)). Dicha función es no-lineal con respecto a los parámetros, y ha sido considerada porque en muchas de las aplicaciones del modelado en instrumentación y control se ha demostrado que los sistemas pueden ser modelados más optimamente haciendo uso de este tipo de funciones, además se reduce el número de grados de libertad (de parámetros a estimar)

obteniendo en ocasiones tiempos de cálculo reducidos. Las perturbaciones que interactúan en el sistema analizado son también modeladas como procesos estocásticos. Entonces, las componentes determinista  $f(\cdot)$  y aleatoria  $e_i$  son modeladas por:

$$y_i = f(x_i, \boldsymbol{\theta}) + e_i, \quad i = 1, \dots, n \quad (30)$$

donde  $\mathbf{y} = [y_1, \dots, y_n]^\top$  es el vector de observaciones. El vector  $\mathbf{x} = [x_1, \dots, x_n]^\top$  corresponde a los parámetros instrumentales (variables explicativas), este puede representar un protocolo experimental (por ejemplo, los instantes de observación del instrumento, la posición de sensores [24], frecuencias utilizadas en corrientes de Foucault [17], etc.).  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^\top$  es el vector de parámetros funcionales que tendrán que ser estimados por algún proceso de identificación. El vector  $\mathbf{e} = [e_1, \dots, e_n]^\top$  resume los errores de adquisición de datos y aquellos de modelado. Como en el caso del modelo regresivo lineal, denotamos  $\Theta \subset \mathbb{R}^p$  de manera general el subconjunto de parámetros  $\boldsymbol{\theta}$  admisible a priori y  $\mathbb{X} \subset \mathbb{R}^n$  el dominio experimental admisible. Finalmente, una diferencia con respecto al estimador de mínimos cuadrados lineal, es que el álgebra lineal aplicada para obtener una expresión cerrada del estimador, no se puede utilizar de manera exacta y por efectos de linealización sólo se obtienen expresiones aproximadas (uso general de métodos iterativos pertenecientes al conjunto de métodos de optimización).

### 4.3. Ecuación de medición

Regresando al marco aplicativo instrumental denominado sistema de mediciones [18], se dice que una medición  $m$  puede ser formalizada como una funcional  $\mathcal{G}$  sobre el objeto  $f(\cdot)$  (por ejemplo, integración ( $m_I$ ), derivación ( $m_P$ ), interpolación, extrapolación, etc.). De lo anterior resulta, que la medición esta dada como una función  $g = \mathcal{G}(f)$  de los parámetros  $\boldsymbol{\theta}$ . Se pueden construir entonces, tantas ecuaciones de medición las cuales sean necesarias hasta formar un vector de mediciones  $\mathbf{m}$ :

$$m_\ell = g_\ell(\boldsymbol{\theta}), \quad \ell = 1, \dots, r, \quad \text{y de forma vectorial} \quad \mathbf{m} = \{m_\ell\}_{\ell=1}^r. \quad (31)$$

Suponemos naturalmente que la función modelo  $f(\cdot)$  es capaz de parametrizar la medición de interés, lo anterior se traduce estableciendo una dependencia de cada medición de interés con respecto a al menos un parámetro, es decir:  $\forall \ell, \exists k$  tal que  $\partial g_\ell / \partial \theta_k \neq 0$ . Como en el caso de los parámetros, notaremos  $\mathbb{M}$  el conjunto o subconjuntos admisibles a priori de mediciones tal que  $\mathbf{m} \in \mathbb{M} \subset \mathbb{R}^r$ .

### 4.4. Densidad bootstrap para los parámetros

Los métodos bootstrap se utilizan en análisis regresivo paramétrico, para caracterizar algunas medidas de incertidumbre sobre los parámetros desconocidos. Entonces, la meta principal es conocer u obtener

la DDP  $\widehat{\varphi}_{\widehat{\theta}}$  de la población de estimadores  $\widehat{\theta}$ . Efron [21], Freedman [25], Hinkley [33] y Wu [62] han introducido muchos de los conceptos y desarrollos teóricos que justifican el uso de los métodos bootstrap aplicados al modelado regresivo. Davison y Hinkley [14], y Shao y Tu [55], [56] presentan de manera más general el uso de los métodos bootstrap en problemas de modelado regresivo no-lineal y no-paramétrico, así mismo exponen algunos desarrollos de la teoría y aplicaciones, que justifican la asintoticidad de los métodos bootstrap en un marco general.

Retomando nuevamente la ecuación (30) y la función no-lineal  $f(\cdot)$  de estructura conocida (establecida según un proceso de selección de modelos), con un vector de parámetros desconocidos  $\theta$ . Entonces, iniciamos la construcción de la densidad empírica  $\widehat{\varphi}_e$ , primero utilizamos el proceso de remuestreo de los residuos. Dichos residuos se calculan en función de los datos  $y_i$  y del modelo  $f(\cdot)$  seleccionado según el objetivo final de estimación de la medición. Así pues, se obtiene el estimador  $\widehat{\theta}$  según<sup>4</sup>:

$$\widehat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i - f(x_i, \theta))^2. \quad (32)$$

Una vez obtenida la estimación  $\widehat{\theta}$ , la etapa siguiente será la obtención de los residuos  $\widehat{e}_i$ :

$$\widehat{e}_i = y_i - f(x_i, \widehat{\theta}), \quad (33)$$

la distribución empírica de los residuos  $\widehat{\varphi}_e$  es luego aproximada haciendo uso del método de Monte-Carlo. Puesto que los residuos se obtienen según  $\widehat{e}_i \sim \widehat{\varphi}_e$  y que conocemos la distribución empírica (DDP)  $\widehat{\theta}$ , entonces, la construcción de los vectores muestra bootstrap  $e_i^*$  y  $y_i^*$ , será la etapa siguiente,

$$y_i^* = f(x_i, \widehat{\theta}) + e_i^*, \quad e_i^* \sim \widehat{\varphi}_e. \quad (34)$$

Los parámetros  $\widehat{\theta}$  serán considerados como si estos fuesen los parámetros *reales o verdaderos*,  $\bar{\theta}$  (desconocidos en la realidad), a cada realización de  $R(e_i^*, \widehat{\varphi}_e)$ , se produce una realización de  $\widehat{\theta}^*$ , siempre y cuando se utilice el mismo proceso de minimización que se ha utilizado para obtener  $\widehat{\theta}$ , por ejemplo en este caso MC

$$\widehat{\theta}^* = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i^* - f(x_i, \theta))^2. \quad (35)$$

Se repiten  $B$  veces las réplicas bootstrap para  $\widehat{\theta}^*$ , y así finalmente se obtienen las realizaciones estocásticas  $\widehat{\theta}^{*1}, \dots, \widehat{\theta}^{*B}$ , las cuales permitirán el acceso a la estimación de la distribución bootstrap de los parámetros  $\widehat{\theta}^*$  ( $R_{\theta}(\widehat{\theta}^{*1}, F_{\theta}^*), \dots, R_{\theta}(\widehat{\theta}^{*B}, F_{\theta}^*)$ ), y de obtener simultáneamente la DDP aproximada de

<sup>4</sup>Remarcamos que se puede utilizar una técnica cualquiera para la estimación de  $\theta$ , puede ser por ejemplo por Mínimos Cuadrados (MC), Máximo de Verosimilitud (MV), M-estimador [35], etc.

$\hat{\theta}$  ( $\varphi(\hat{\theta}^*) \triangleq \varphi(\hat{\theta})$ ), y si además se considera  $\varphi(\hat{\theta}) \triangleq \varphi(\theta)$ , entonces  $\varphi(\hat{\theta}^*) \triangleq \varphi(\theta)$  asintóticamente. Una vez presentada la aproximación bootstrap para los modelos regresivos, ésta se extiende y se aplica a los diferentes contextos del sistema de mediciones.

#### 4.5. Densidad bootstrap de las mediciones

El problema de la estimación de la DDP de una medición fué previamente tratado proponiendo un estudio hecho en un marco no-lineal gaussiano [8], en dicho trabajo se presentó la extensión de las aproximaciones analíticas introducidas por Box [7], Clarke [12], [13] y Pázman [49]. Todas estas propuestas se hacen en un marco de análisis regresivo paramétrico no-lineal. La aproximación propuesta por Box se enfoca a obtener una aproximación lineal de primer orden de los dos primeros momentos del estimador  $\hat{\theta}$ , luego ésta se extiende a la medición. En lo referente a la aproximación propuesta por Clarke, ésta va más allá de la aproximación de primer orden de los dos primeros momentos, obteniendo así, estimadores del sesgo y de la matriz de variancia/covariancia más precisos. Posteriormente se propone construir una distribución parametrada con los dos momentos y se le llamó aproximación normal corregida. Por su parte, Pázman propone una aproximación analítica directa de la distribución conjunta de los estimadores de MC de  $\hat{\theta}$ , la cual fué inspirada en un análisis geométrico. Sin embargo, existen algunas desventajas de las tres aproximaciones, las primeras dos brindan una aproximación sólo en términos de momentos, pero no de la distribución completa, después se propone una estimación de dicha distribución y esto utilizando una aproximación normal, la cual en el caso de una no-linealidad fuerte entre los parámetros sería irrealista. En lo que concierne a la propuesta de Pázman, ésta es más precisa en términos de la distribución de  $\hat{\theta}$  y por lo tanto de  $\widehat{m}$ , el problema aquí, es la marginalización de la DDP conjunta la cual se dificulta cuando  $p > 2$ . La marginalización es necesaria, primero para poder construir las estadísticas de cada parámetro por separado y en seguida para poder hacer la propagación hasta cada una de las mediciones de interés  $m_1, \dots, m_r$ .

Así, la propuesta de utilizar los métodos bootstrap nos permite resolver varios de los problemas presentados en las aproximaciones analíticas anteriores y además nos permite generalizar las hipótesis formuladas sobre la distribución de los errores, relajando así la información apriori necesaria para caracterizar estadísticamente tanto  $\hat{\theta}$  como  $\widehat{m}$ . El costo se verá reflejado sobre todo en el tiempo de cálculo de la estimación de la DDP la cual se lleva a cabo en un marco más general (no-lineal y cuando la ley de los errores no se conoce). Una vez que se ha aproximado la distribución de los parámetros, y bajo el conocimiento de que la medición de interés depende de al menos uno de estos parámetros, el paso final es aproximar la DDP de  $m$  utilizando para ello el mapeo no-lineal entre parámetros y

medición, y disponiendo de las distribuciones marginales de los parámetros.

Se plantea de manera analítica que para acceder a la DDP de la medición de interés, en un primer paso se obtiene la medición a partir de un vector extendido, sea  $\mathbf{m}_e = [\theta_1, \dots, m, \dots, \theta_p]^\top$ . Luego, la relación no-lineal entre los parámetros y la medición se redefine por medio de la siguiente expresión:  $\mathbf{m}_e = G_k(\boldsymbol{\theta})$ . Luego, la distribución usual de  $\mathbf{m}_e$  se obtiene mediante

$$\wp(\mathbf{m}_e) = \frac{\wp(\boldsymbol{\theta})}{|\det(J)|} \Big|_{\boldsymbol{\theta}=G_k^{-1}(\mathbf{m}_e)}$$

en donde  $J = [\partial m_\ell / \partial \theta_j]$  es la matriz Jacobiana con cambio de variable. Así, la DDP de la medición de interés  $m$  (escalar) se obtiene según la expresión

$$\wp(m) = \int \int_{\Theta_{-k}} \wp_{\boldsymbol{\theta}}(G_k^{-1}(\mathbf{m}_e)) \left| \frac{\partial g}{\partial \theta_k} \right|^{-1} d\boldsymbol{\theta}_{-k} \quad (36)$$

en donde  $\boldsymbol{\theta}_{-k} = [\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_p]^\top$ .

Cuando el espacio de parámetros crece y asimismo el espacio de mediciones, las manipulaciones de expresiones analíticas como (36) se hacen complejas. Por el contrario, en el marco de la simulación estocástica, no se tiene la necesidad de calcular  $\mathbf{m}_e$ , puesto que una vez que el estimador  $\widehat{\boldsymbol{\theta}}^*$  ha sido calculado, la medición bootstrap inducida por  $g(\cdot)$  se obtiene de manera directa,

$$\widehat{m}^* = g(\widehat{\boldsymbol{\theta}}^*). \quad (37)$$

De la misma manera en que las DDPs de los parámetros fueron estimadas  $\widehat{\wp}(\boldsymbol{\theta}) \triangleq \wp(\widehat{\boldsymbol{\theta}}^*)$  de forma completa y marginal vía Monte-Carlo, la distribución de la medición o mediciones de interés  $\wp(m)$  será aproximada por su análogo bootstrap  $\wp(\widehat{m}^*)$ . Esta última se deduce de  $\widehat{m}^{*1}, \dots, \widehat{m}^{*B}$ , utilizando la población de realizaciones  $\widehat{\boldsymbol{\theta}}^{*1}, \dots, \widehat{\boldsymbol{\theta}}^{*B}$  en la relación no-lineal  $g(\boldsymbol{\theta})$  y sacando provecho de la integración estocástica. Entonces,  $R(\widehat{m}^{*1}, F_m^*), \dots, R(\widehat{m}^{*B}, F_m^*)$  se induce por el mecanismo  $g(R_{\boldsymbol{\theta}}^*)$ , y finalmente  $\wp(m) \triangleq \wp(m^*)$ , convergerá si  $\lim_{B \rightarrow \infty} |\wp(m^*) - \wp(m)| = 0$  (según la ley de los grandes números). La DDP o las DDPs de las mediciones se obtienen también de manera marginal,

$$\wp(\widehat{m}_\ell^*) = \wp(g_\ell(\widehat{\boldsymbol{\theta}}^*)) = \int_{\mathbb{M}} \wp(\widehat{\mathbf{m}}^*) d\widehat{\mathbf{m}}_{-\ell}^*, \quad (38)$$

en donde  $d\widehat{\mathbf{m}}_{-\ell}^* = d\widehat{m}_1^* \cdots d\widehat{m}_{\ell-1}^* d\widehat{m}_{\ell+1}^* \cdots d\widehat{m}_r^*$ .

## 5. Mejoras en la convergencia de bootstrap

### 5.1. Bootstrap aplicado a los residuos

El desempeño y la aproximación de los métodos bootstrap pueden mejorarse cuando se inyecta información adicional que se tiene acerca de los residuos, es decir, que es posible mejorar el nivel sobre la

precisión de la distribución y las estadísticas las cuales se busca aproximar, bajo las “hipótesis” adicionales de un ruido o errores de ley desconocida, pero simétrica (centrada). Comenzamos enunciando el método más simple:

- (a) obtención de los *residuos* según la aproximación clásica (referirse a la ecuación (33)),
- (b) realizar un *centrado de los residuos*: si una de las componentes del vector  $\boldsymbol{\theta}$  es un parámetro de traslación de la función  $f(\cdot)$ , entonces tenemos que  $\widehat{\varphi}_e$  es de media cero. Si no, de cualquier manera se puede modificar la traslación para obtener el centrado de los residuos [21], [25] (procedimiento similar al centrado y después Monte-Carlo),

$$\begin{aligned}\widehat{r}_i &= y_i - f(x_i, \widehat{\boldsymbol{\theta}}), \\ \bar{r} &= \frac{1}{n} \sum_{i=1}^n \widehat{r}_i, \\ \widehat{e}_i &= \widehat{r}_i - \bar{r},\end{aligned}\tag{39}$$

- (c) y el cálculo de los *residuos modificados* (o bien residuos ponderados): los residuos modificados brindan una versión suavizada (smoothed) o ponderada de los residuos, la cual puede conducir a estimadores consistentes [28], [55], por ejemplo en un marco de selección de modelos cuando se utiliza el submuestreo. Según estudios recientes, los residuos ponderados también pueden considerarse como un vector sesgado des residuos bootstrap [28], [29]:

$$\widehat{e}_i = \frac{\sqrt{\frac{n}{q}}(\widehat{r}_i - \bar{r})}{\sqrt{1 - \frac{p}{n}}}.\tag{40}$$

El procedimiento de remuestreo de los residuos centrados es modificado en el marco del análisis regresivo lineal como se ilustra a continuación: la dimensión del vector muestra bootstrap  $\widehat{e}_i^*$ ,  $i = 1, \dots, q$ , se modifica de tal manera que  $q < n$  [6] (un vector submuestra). Sin embargo, en el caso del análisis regresivo no-lineal, el valor de  $q$  se modifica de tal manera que  $q < n$  y respetando que la relación  $q/n \rightarrow 0$  se cumpla, pero la dimensión del vector muestra bootstrap permanece sin modificarse ( $i = 1, \dots, n$ , por lo que no hay submuestreo implícito). Por otro lado, la única condición que deberá cumplirse es que  $q$  deberá ser seleccionada de tal suerte que la relación  $p/q$  sea suficientemente pequeña.

**Comentario 5.1** *En el caso de aplicaciones prácticas se puede especificar un valor de  $q$  para un valor fijo de  $n$ , y entonces la selección óptima de  $q$  dependerá de la dimensión  $p$  del vector de parámetros. Por*

*ejemplo, se podría comenzar por  $q = n$  y después se disminuye el valor hasta que se tenga una relación con valor aproximado a  $q/n = 0,5$ , si  $n$  es suficientemente grande (tal que  $q/n \rightarrow 0$ ), en dicho caso se podrá considerar que  $n \rightarrow \infty$ , mientras que  $q \rightarrow \infty$  un poco mas lentamente.*  $\square$

Se hace notar que las ecuaciones (39) y (40) de los paréntesis (b) y (c) se reemplazan por la ecuación (33) de los *residuos clásicos*. Ya que la hipótesis clásica consiste en considerar que los errores siguen una ley gaussiana (por razones de simplicidad), y en la realidad dicha ley no se conoce. Entonces, el uso de los métodos bootstrap permite tratar los casos más generales en donde la distribución de los errores es desconocida, pero para la cual se asume un mínimo de hipótesis como la de simetría [15], de esta forma, el método de los residuos ponderados se muestra como el método bootstrap que brinda un mejor desempeño en términos de precisión y de tiempo de cálculo de la distribución (se puede reducir el número de simulaciones).

## 5.2. Otros esquemas bootstrap en problemas regresivos

Puesto que el remuestreo de los residuos ponderados no es robusto con respecto a la dispersión no uniforme de los datos (conocida también como propiedad de heteroscedasticidad:  $\sigma_i^2 \neq cte$ ), bajo ciertas condiciones complementarias se pueden construir esquemas más robustos. Por ejemplo, i) un esquema que considere que las variables explicativas  $x_i$  son variables aleatorias (modelo de correlación) o bien ii) utilizar un esquema conocido como bootstrap externo.

- (i) *Bootstrap por pares* o en inglés “Paired bootstrap”  $\{x_i, y_i\}$ : este procedimiento es una versión natural de remuestreo en paralelo (referirse también a los métodos de errores en las variables), en este caso las variables  $\{x_i\}$  son consideradas como v.a.i.i.d. ( $x_i \sim F_x$ ). La identificación del modelo  $f(\cdot)$  se lleva a cabo bajo el conocimiento de la distribución conjunta del par de variables  $\{x_i, y_i\}$ , la cual será aproximada por la DDP empírica  $\widehat{F}_{(x_i, y_i)}$  considerando también un peso  $n^{-1}$  por cada realización de los  $\{x_i, y_i\}$ . Los vectores muestra bootstrap  $\{x_i^*, y_i^*\}$  serán generados de manera similar a los procedimientos anteriores (ver las ecuaciones (34) – (37)). Para asegurar la consistencia de este nuevo procedimiento, Shao [55], [56] considera también una modificación similar a la que se describe en la sección anterior (ver inciso (c) variación de la dimensión según  $q$  o submuestreo).
- (ii) *Bootstrap externo* o en inglés “The external bootstrap”: ésta versión del método bootstrap fué propuesta por Wu [62]. Las realizaciones  $e_i^*$  son supuestas como v.a.i.i.d. centradas y de variancia

$\sigma^2 = 1$ . Luego, los vectores muestra bootstrap son generados a partir de:

$$y_i^* = f(x_i, \hat{\boldsymbol{\theta}}) + \frac{|r_i|}{\sqrt{1 - h_i}} e_i^*, \quad e_i^* \sim \mathcal{N}(0, 1) \quad (41)$$

en este caso los pesos se calculan usando

$$h_i = \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left[ \sum_{i=1}^n \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right]^{-1} \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top},$$

puesto que la distribución de los  $e_i^*$  es independiente del vector de residuos originales  $e_i$ , el procedimiento bootstrap es denominado externo, ver también el término en inglés “wild bootstrap” [55], en donde los  $e_i^*$  deben respetar la siguiente condición:  $E \{ e_i^{*3} \} = 1$ .

En general, los estimadores del sesgo y de la variancia de  $\hat{\boldsymbol{\theta}}$  obtenidos utilizando estos dos últimos procedimientos bootstrap, son también consistentes y robustos. Shao [55] ha demostrado, que en términos de robustez éstos dos procedimientos son mejores que los métodos bootstrap aplicados a los residuos, sin embargo, en términos de eficacia el remuestreo de residuos es el mejor método. Por otro lado, en el caso del procedimiento bootstrap externo, los pesos  $h_i$  pueden ser seleccionados por medio de un vector de ponderación sencilla, lo cual significará una alternativa ventajosa si es que se desea evitar el cálculo a veces complejo de  $\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  [62].

### 5.3. Bootstrap ponderado iterativo

En la literatura existente sobre los métodos bootstrap, es común ver que el número de simulaciones  $B$  toma aproximadamente el mismo valor que el de la dimensión del vector muestra:  $n$ . Para el caso de modelos paramétricos suficientemente regulares, y para los cuales la distribución limitada de la estadística de interés no depende de los parámetros desconocidos, el método bootstrap brinda una aproximación suficiente de dicha distribución cuando  $B = n$ . Lo anterior se establece en términos del orden de aproximación de la probabilidad (primer orden, segundo orden, etc.). En realidad, el error resultante inducido por la simulación Monte-Carlo domina con frecuencia el error de aproximación bootstrap, provocando una estimación de la distribución poco precisa. Cuando la distribución de una estadística de interés depende de los parámetros desconocidos, los errores de aproximación y de simulación pueden ser reducidos iterando el procedimiento inicial bootstrap dentro de otro procedimiento bootstrap (bootstrapping the bootstrap). La idea principal del método bootstrap iterado (conocido también como bootstrap simulado), es de construir un procedimiento que se podría emplear para reducir el error debido a la simulación y paralelamente reducir el tiempo de cálculo, con la finalidad de obtener una aproximación más precisa de la distribución bootstrap o de las estadísticas de interés.

Para poder minimizar el error estocástico producido por la simulación bootstrap, se debe satisfacer al menos la siguiente desigualdad:  $B \geq n^2$ . Por otro lado, el hecho de utilizar un esquema bootstrap iterativo, indica que el número de réplicas realizadas en el procedimiento bootstrap interno (lazo interno- $B_1$ ) deberá satisfacer la siguiente desigualdad:  $B_1 \geq n^2$ , y el número de replicas en el procedimiento bootstrap externo (lazo externo- $B_2$ ) deberá satisfacer:  $B_2 \geq n^3$ . Entonces, el número total de simulaciones  $B = B_1 \times B_2$  deberá alcanzar el orden de  $B = n^5$ , con el fin de asegurar que los errores estocástico (debido a la simulación) y el de aproximación bootstrap no sean considerablemente grandes. La condición anterior de cálculo puede ser exageradamente grande, aún y cuando se tengan que simular modelos simples (de pocos parámetros) de dimensión  $n$  moderada. En trabajos recientes desarrollados por Brown [9] se propone una aproximación que hace uso de una variable aleatoria controlada, lo anterior es con el fin de mejorar la aproximación de la DDP, cuando se utiliza el esquema bootstrap iterativo. Este nuevo esquema utiliza una función principal de la estadística de interés, la cual conduce a un desarrollo asintótico y a su esperanza matemática y esto conduce a su vez a una nueva función que tiene la misma esperanza, pero una variancia un poco más pequeña que la variancia de la estadística original. El número de simulaciones necesarias para el procedimiento bootstrap interno se reduce a  $B_1 \geq n$ , y para el procedimiento bootstrap externo se reduce a  $B_2 \geq n^2$ . Finalmente, el número total de simulaciones disminuye sensiblemente, pasa del orden de  $B = n^5$  a  $B = n^3$ .

Partiendo del esquema bootstrap iterativo propuesto por Brown [9], nosotros proponemos su extensión al contexto de sistemas de medición, agregando además la etapa de estimación según el método bootstrap ponderado (caso de bootstrap no-paramétrico para los residuos modificados). Aunque la aproximación por variable aleatoria controlada fué desarrollada en un marco donde se supone conocida de forma paramétrica la distribución de los errores (o residuos), se puede demostrar que en el caso no-paramétrico del método bootstrap ponderado, dicha aproximación es capaz de brindar una mejora sobre la precisión de la distribución bootstrap con respecto a la que se obtiene utilizando el método bootstrap clásico, asimismo y paralelamente se reduce significativamente el número de simulaciones. De hecho, las mejoras que se obtienen son claramente apreciables desde el punto de vista del tiempo de cálculo, puesto que el tiempo del esquema bootstrap iterativo clásico es mucho más grande.

La aproximación por variable aleatoria controlada fué introducida para los casos de análisis de modelos paramétricos y semi-paramétricos como el análisis regresivo no-lineal [9], en donde se hacen hipótesis de normalidad sobre  $\varphi(\mathbf{e})$ . Por nuestra parte, proponemos de extender el método a un caso en donde se hacen un mínimo de hipótesis como es el caso de remuestreo no-paramétrico y dentro de las hipótesis consideradas están: las suposiciones de independencia de las  $x_i$ , la media cero, distribución

simétrica, y variancia uniforme de los residuos [16]. Según los métodos bootstrap no-paramétricos, la forma de obtener las realizaciones a partir de la distribución empírica  $e_i^{(j)} \sim \widehat{\varphi}(\mathbf{e})$ , no permite acceder directamente a la función  $F_\omega(\boldsymbol{\beta}, t)$ , en donde  $\boldsymbol{\beta} = [\boldsymbol{\theta}, \mathbf{m}]^\top$ . Existen algunas complicaciones para hacer posible el cálculo de dicha función, sin embargo, es posible obtener de manera indirecta una muy buena aproximación de la esperanza de la función de la variable controlada, luego de llevar a cabo un número adicional de simulaciones Monte-Carlo, puesto que se sabe que  $\widehat{\varphi}(\mathbf{e})$  converge de manera consistente (según el método bootstrap ponderado). La aproximación de la función DDP se detalla con mayor énfasis en la siguiente sección (se demuestra gracias a un desarrollo en series de Edgeworth), en donde se presenta la extensión (aplicando el teorema 2 establecido por Brown [9]) al caso de la estimación de una medición, puesto que ésta también puede ser considerada como una función semi-paramétrica que depende del análisis regresivo no-lineal.

#### 5.4. Desarrollo en series de Edgeworth para el esquema bootstrap ponderado iterativo

En esta sección presentamos un desarrollo en series de Edgeworth con la intención de mostrar el orden sobre la precisión que se obtiene sobre la convergencia de la densidad de probabilidad de los parámetros y de la medición, las cuales fueron aproximadas utilizando el esquema bootstrap ponderado iterativo introducido en la sección anterior.

Sea  $\boldsymbol{\beta}$  un vector extendido de parámetros, tal que  $\boldsymbol{\beta} = [\boldsymbol{\theta}, \mathbf{m}]^\top$ , en donde el subvector  $\mathbf{m} = \{g_\ell(\boldsymbol{\theta})\}_{\ell=1}^r$ . En el marco del análisis regresivo de sistemas de medición,  $\widehat{\boldsymbol{\theta}}$  y  $\widehat{\mathbf{m}}$  están definidos como estimadores semi-paramétricos. La distribución  $\widehat{\varphi}_e(\boldsymbol{\theta})$  es una aproximación de la DDP de los errores obtenida utilizando el método bootstrap sobre los residuos modificados la cual es consistente (Shao [55]).

En este caso, la función de distribución de una estadística cualquiera  $R(\boldsymbol{\beta})$  de  $\boldsymbol{\beta}$  será redefinida por  $\Gamma(\boldsymbol{\beta}, R(\boldsymbol{\beta})) = E_\beta \{\mathcal{I}\{R(\boldsymbol{\beta}) \leq t\}\}$  [9], y la función de interés será  $H(R(\boldsymbol{\beta}), t) = \mathcal{I}\{R(\boldsymbol{\beta})\}$ . Para asegurar que  $R(\boldsymbol{\beta})$  tiene un comportamiento limitado, y bajo el conocimiento de que se cuenta con un vector aleatorio de dimensión  $n$  finita, es necesario poder establecer una función principal sobre la estadística de interés. Dicha función se puede obtener mediante una aproximación establecida según el desarrollo en series de Edgeworth siguiente:

$$R(\boldsymbol{\beta}) = \omega(e_i; \boldsymbol{\beta}) + O_p\left(n^{-q/2}\right), \quad i = 1, \dots, n, \quad (42)$$

en donde  $\omega(\boldsymbol{\beta}) = \omega(e_i; \boldsymbol{\beta})$  es una estadística que aproxima la estadística  $R(\boldsymbol{\beta})$  según un orden de precisión predeterminado. Si se cuenta con realizaciones o vectores muestra de los errores aleatorios  $e_i \sim \widehat{\varphi}_e(\boldsymbol{\theta})$  obtenidos a partir de una densidad parametrada o una estimación consistente de dicha

densidad, entonces  $F_\omega(\boldsymbol{\beta}, t) = E_\beta \{ \mathcal{I} \{ \omega(\boldsymbol{\beta}) \leq t \} \}$  podrá obtenerse utilizando el mismo procedimiento de estimación de los parámetros  $\boldsymbol{\beta}$ . La estadística  $F_\omega(\boldsymbol{\beta}, t)$  se denomina pivotal en un sentido general, pero en el caso particular cuando  $q = 1$  ésta función  $F_\omega(\boldsymbol{\beta}, t) = F_\omega(t)$ , y no depende de  $\boldsymbol{\beta}$ . El conocimiento previo de  $R(\boldsymbol{\beta})$  y  $H(R(\boldsymbol{\beta}), t)$  es indispensable para poder llevar a cabo el control de la variable aleatoria y así poder eliminar la función estocástica principal de  $H(T(\boldsymbol{\beta}), t)$  en donde se va simular el término restante  $F_\omega(\boldsymbol{\beta}, t)$ .

Dados  $\boldsymbol{\beta}$  y  $\omega(\boldsymbol{\beta})$ , se pueden generar los vectores  $e_i^{(j)} \sim \widehat{\varphi}_e(\boldsymbol{\theta})$ ,  $j = 1, \dots, B_1$ , con lo cual se puede obtener  $R^{(j)}(\boldsymbol{\beta})$  y  $\omega^{(j)}(\boldsymbol{\beta}) = \omega(e_i^{(j)}; \boldsymbol{\beta})$ , y finalmente acceder a:

$$\tilde{H} \left( R^{(j)}(\boldsymbol{\beta}), \omega^{(j)}(\boldsymbol{\beta}), t \right) = \mathcal{I} \left\{ R^{(j)}(\boldsymbol{\beta}) \right\} - \mathcal{I} \left\{ \omega^{(j)}(\boldsymbol{\beta}) \right\} + F_\omega(\boldsymbol{\beta}, t) \quad (43)$$

después el valor promedio bootstrap obtenido por Monte-Carlo se calcula como sigue:

$$\begin{aligned} \tilde{\Gamma}_{B_1}(\hat{\boldsymbol{\beta}}, t) &= \frac{1}{B_1} \sum_{j=1}^{B_1} \tilde{H} \left( R^{(j)}(\hat{\boldsymbol{\beta}}), \omega^{(j)}(\hat{\boldsymbol{\beta}}), t \right) \\ \tilde{\Gamma}_{B_1}(\hat{\boldsymbol{\beta}}, t) &= \Gamma(\boldsymbol{\beta}_0, t) + O_p \left( n^{-1(k+1)/2} \right) + O_p \left( n^{-q/2} B_1^{-1/2} \right) \end{aligned} \quad (44)$$

Para estar seguro de que el error inducido por el procedimiento de Monte-Carlo esta siendo controlado, el número de simulaciones  $B_1$  del procedimiento bootstrap interno deberá satisfacer la desigualdad:  $B_1 \geq n^{k+1-q}$ . De igual manera, el procedimiento bootstrap externo, requiere un número de simulaciones  $B_2$ , en este caso la función de interés es la siguiente:  $\tilde{R}_1(\boldsymbol{\beta}) = \tilde{\Gamma}_{B_1}(\hat{\boldsymbol{\beta}}, R(\boldsymbol{\beta}))$ , con  $\tilde{R}_1(\boldsymbol{\beta}) = \Gamma(\boldsymbol{\beta}, R(\boldsymbol{\beta})) + O_p(n^{-(k+1)/2})$ , y su función de distribución está definida por:

$$\tilde{\Gamma}_1 = E_\beta \left\{ \mathcal{I} \left\{ \tilde{R}_1(\boldsymbol{\beta}) \leq t_1 \right\} \right\}.$$

Utilizando las ideas del párrafo anterior, también se puede definir:

$$\tilde{H}_1 \left( \tilde{R}_1^{(l)}(\boldsymbol{\beta}), \omega^{(l)}(\boldsymbol{\beta}), t_1 \right) = \mathcal{I} \left\{ \tilde{R}_1^{(l)}(\boldsymbol{\beta}) \right\} - \mathcal{I} \left\{ F_\omega \left( \omega^{(l)}(\boldsymbol{\beta}) \right) \right\} + t_1 \quad (45)$$

y así el valor promedio bootstrap obtenido por Monte-Carlo se calcula como sigue:

$$\begin{aligned} \tilde{\Gamma}_{1B_2}(\hat{\boldsymbol{\beta}}, t) &= \frac{1}{B_2} \sum_{l=1}^{B_2} \tilde{H}_1 \left( \tilde{R}_1^{(l)}(\hat{\boldsymbol{\beta}}), \omega^{(l)}(\hat{\boldsymbol{\beta}}), t_1 \right) \\ \tilde{\Gamma}_{1B_2}(\hat{\boldsymbol{\beta}}, t) &= \tilde{\Gamma}_1(\boldsymbol{\beta}_0, t_1) + O_p \left( n^{-1(k+2)/2} \right) + O_p \left( n^{-q/2} B_2^{-1/2} \right) \end{aligned} \quad (46)$$

En general,  $B_2$  deberá satisfacer la desigualdad:  $B_2 \geq n^{k+2-q}$ . Como en el caso propuesto el valor de  $k = q = 1$ , entonces  $B_1 \geq n$  y  $B_2 \geq n^2$ . Y finalmente el número total de simulaciones deberá satisfacer la desigualdad:  $B \geq n^3$  ( $B = n^3$  como mínimo).

## 6. Robustez, optimalidad y eficacia de bootstrap

En la sección § 4 se introdujeron algunos estimadores utilizados comunmente en analisis regresivo lineal y cuyos conceptos son facilmente extendibles al analisis regresivo en general. Los estimadores más utilizados en la estadística clásica, son aquellos que se establecen como una función de los residuos. En esta sección, retomaremos el aspecto concerniente a la robustez y eficacia de los estimadores, sobre todo de los estimadores de mínimos cuadrados y de máximo de verosimilitud. Por otro lado, presentamos una propuesta de estimador robusto fundamentado en los métodos bootstrap intencionalmente sesgados, dicha propuesta se extendió al análisis regresivo no-lineal y al contexto del sistema de medición. Los métodos bootstrap intencionalmente sesgados fueron introducidos por Hall y Presnell [28], [29], [30], en un marco general de estimación robusta para funciones lineales y de estimación no-paramétrica. Nuestro interés sobre dichos métodos obedece al hecho de que los esquemas de remuestreo basados en los residuos no son robustos con respecto a los datos aberrantes (como con respecto al cambio de variancia entre los diferentes datos (heteroscedasticidad)), y además porque las pocas hipótesis formuladas sobre los errores no se pueden verificar siémpre de manera efectiva.

### 6.1. Optimalidad del estimador de mínimos cuadrados (MC) (Nonlinear Least Squares–NLS)

El estimador clásico más utilizado en análisis estadístico es el de mínimos cuadrados. Dicho estimador es equivalente al estimador de máximo de verosimilitud cuando las hipótesis que se hacen sobre los errores son gaussianas, de media cero y de variancia constante. Si dichas hipótesis son verificables, el estimador MC es óptimo y eficaz [35], [61], [56]. Por el contrario, si dichas hipótesis son falsas entonces el estimador MC es sub-óptimo, y para el caso en el cual la variancia no es constante o bien en presencia de datos aberrantes, dicho estimador no es robusto. A pesar de este inconveniente, el estimador MC es utilizado en muchas aplicaciones debido a su simplicidad y rapidez (referirse a la eficacia normal).

Por otro lado, en el análisis regresivo no-lineal, el estimador MC no tiene una solución analítica (explícita). Nos debemos conformar con el hecho de obtener aproximaciones iterativas que brindan con frecuencia optimizaciones de tipo local:

$$\mathbf{s}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})] \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = 0, \quad \boldsymbol{\theta} \in \Theta$$

cuyo equivalente bootstrap está dado por:

$$\mathbf{s}^*(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n [y_i^* - f(x_i, \hat{\boldsymbol{\theta}})] \frac{\partial f(x_i, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}^\top} = 0, \quad \hat{\boldsymbol{\theta}} \in \Theta. \quad (47)$$

Bajo el conocimiento de que el remuestreo de los residuos no brinda una aproximación robusta, se propone estudiar la posibilidad de obtener un estimador simultáneamente robusto y eficaz. Shao y Tu [56] han demostrado que el uso de procedimientos bootstrap por pares y bootstrap externo permiten una estimación más robusta cuando la variancia de los errores no es uniforme, sin embargo la eficacia de dichos estimadores no se garantiza. En lo concerniente a la robustez cuando hay algunos datos aberrantes, Hall y Presnell [29] proponen una aproximación robusta en el marco de análisis regresivo lineal, esta nueva clase de estimadores fué llamada *estimadores bootstrap intencionalmente sesgados*. A continuación se presenta la adaptación de las ideas anteriores en el marco de análisis regresivo no-lineal, y en el contexto del sistema de medición.

## 6.2. Robustez inducida pour un esquema bootstrap sesgado

Recientemente, Hall y Presnell [29] propusieron utilizar una versión de un esquema bootstrap intencionalmente sesgado para poder obtener una estimación robusta en los casos en que se tienen datos aberrantes y en donde las variancias de los errores no son para nada uniformes. Se trata de un nuevo método empírico, que identifica y sub-pondera los datos que ejercen una influencia inadecuada sobre los estimadores estadísticos. Los pesos utilizados para sub-ponderar se seleccionan a partir de probabilidades de remuestreo, lo cual nos coloca nuevamente frente a una versión bootstrap ponderada. Se puede considerar también que ésta última versión bootstrap equivale a la versión uniforme (clásica) sesgada.

Los métodos bootstrap intencionalmente sesgados no necesitan ni la estimación de la DDP, ni la especificación de una familia de distribuciones parametradas. Estos nuevos métodos requieren principalmente dos argumentos de entrada: la medida de una distancia entre la distribución bootstrap uniforme y la distribución bootstrap sesgada, y que se cumplan algunas condiciones impuestas, dichas condiciones están dadas por la medida de emplazamiento y la medida de dispersión de los datos que serán analizados.

La distancia utilizada para medir la distancia entre distribuciones (hablando en términos de la potencia de divergencia) es descrita por Hall y Presnell en el Anexo A de [29], y está dada por:  $D_\rho = \xi$ , en donde  $\xi$  se establece según una función conocida como “función de punto de ruptura” de la siguiente forma:

$$\xi(\varepsilon) = \begin{cases} [\rho(1-\rho)]^{-1} \{1 - (1-\varepsilon)^{1-\rho}\} & \text{si } \rho \neq 1, \\ -\log(1-\varepsilon) & \text{si } \rho = 1, \end{cases} \quad (48)$$

Según algunas sugerencias de Hall y Presnell [29], parece razonable seleccionar un punto inicial de ruptura  $\varepsilon$  dentro del intervalo  $0 < \varepsilon < 1/2$ . Por otro lado,  $\rho$  representa el exponente de la potencia de divergencia, la cual puede variar en el intervalo  $0 < \rho \leq 1$ , donde se incluyen la distancia de Hellinger y la distancia de Kullback-Leibler.

**Comentario 6.1** *Para aplicaciones prácticas y para una implementación más sencilla, Hall y Presnell sugieren acotar aún más el intervalo del punto de ruptura  $\varepsilon \in [0,01, 0,15]$ .*  $\square$

En lo que concierne a las condiciones impuestas, éstas están dadas por la medida del emplazamiento y la medida de la dispersión de los datos bajo análisis. La estimación del emplazamiento multivariable para una pareja de muestras  $\{X_i, Y_i\}$ , se obtiene haciendo uso de la ecuación (49):

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n w_i \varphi(e_i), \quad (49)$$

en donde  $w_i$  son los pesos positivos y la función  $\varphi(e_i)$  está dada por:

(a) la función de la norma cuadrática de los errores (Bootstrap Cuadrático Sesgado–BCS):

$$\varphi_1(e_i) = (y_i - f(x_i, \theta))^2, \quad (50)$$

(b) u otra función similar a la propuesta por Huber [35], dicha función fué propuesta por Brahim [8]. La función llamada cuasi-Huber fué retomada con la intención de garantizar la robustez en casos de cambios fuertes de la variancia y cuando existe un conocimiento limitado a cerca de la distribución de los errores (en donde la eficacia podría verse disminuida). La función está dada por la siguiente ecuación (Bootstrap cuasi-Huber Sesgado–BHS):

$$\varphi_2(e_i) = \frac{\Delta^2}{2} \left( \sqrt{1 + \frac{4\varphi_1(e_i)}{\Delta^2}} - 1 \right), \quad (51)$$

en donde el valor de  $\Delta$  debe ser estrictamente positivo,  $\Delta > 0$  y constante.

Por otro lado, la medida de dispersión está dada por la siguiente expresión:

$$\gamma(w_i) = \inf_{\theta \in \Theta} \sum_{i=1}^n w_i \varphi(e_i). \quad (52)$$

La distribución empírica usual se obtiene mediante el método bootstrap uniforme considerando que todos los pesos tienen la misma masa (o valor puntual) para todos los datos, es decir que:  $\mathbf{w}_{unif} = (1/n, \dots, 1/n)$  para un vector muestra de dimensión  $n$ . Los valores iniciales estimados según el emplazamiento y la dispersión se obtienen mediante:  $\hat{\theta} = \hat{\theta}(\mathbf{w}_{unif})$ , así que para  $\hat{\gamma} = \hat{\gamma}(\mathbf{w}_{unif})$ . Después, será

calibrado (o minimizado) el nivel de dispersión utilizando la versión sesgada de los pesos  $w_i$  (aquí se lleva a cabo la sub-ponderación). La primera función propuesta por Hall y Presnell [29], será retomada para el caso de análisis regresivo no-lineal; la función tiene la siguiente estructura:

$$w_i = \frac{1}{n} h_{[\xi]}(X_i, Y_i | \hat{\boldsymbol{\theta}}, \hat{\tau}, \hat{\lambda}) \quad (53)$$

en donde  $h_{[\xi]}(\cdot)$  es propuesta de manera más general por

$$h_{[\xi]}(\cdot) = \begin{cases} \left(1 + (\rho - 1) \left[\rho\xi - \hat{\lambda} \{\varphi(e_i) - \hat{\tau}\}\right]\right)^{1/(\rho-1)} & \text{si } \rho \neq 1, \\ \exp\left(\xi - \hat{\lambda} \{\varphi(e_i) - \hat{\tau}\}\right) & \text{si } \rho = 1, \end{cases}$$

Finalmente, los valores obtenidos para  $\hat{\boldsymbol{\theta}}$ ,  $\hat{\tau}$ , y  $\hat{\lambda}$  deben satisfacer las ecuaciones:

1. la minimización de un criterio, tal que:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left( \sum_{i=1}^n w_i \varphi(\hat{e}_i) \right) = 0,$$

2. el cálculo del estimador:

$$\hat{\tau} = \sum_{i=1}^n w_i \varphi(\hat{e}_i),$$

en donde  $\hat{\tau}$  se encuentra en el intervalo  $0 < \hat{\tau} \leq \hat{\gamma}$ ,

3. y  $\hat{\lambda}$  la cual es seleccionada de tal suerte que:

$$\sum_{i=1}^n w_i = 1$$

sea satisfecha. Por ejemplo, se puede obtener por medio de un método de dicotomía, sabiendo que ésta debe ser estrictamente positiva,  $\lambda > 0$ .

**Comentario 6.2** *Si la distribución de los errores  $e_i$  es simétrica y unimodal, entonces los valores verdaderos o reales de  $\boldsymbol{\theta}$  no cambiarán por desplazamiento. En el caso de distribuciones asimétricas, los valores de  $\boldsymbol{\theta}$  podrán ser alterados, sin embargo, éstos permanecen bien definidos como soluciones de  $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ .* □

### 6.3. Robustez vs. eficacia

El compromiso entre la robustez y la eficacia es crítico para los modelos cuya dimensión del vector de parámetros es considerablemente grande. La sensibilidad de los modelos no-lineales al análisis de datos

aberrantes es puesta a prueba en el ejemplo de aplicación de la próxima sección, en donde se utilizan varios estimadores robustos en un problema de medición. En los sistemas de medición en donde existen ciertas probabilidades de censar datos aberrantes, es preferible resolver el problema de estimación por medio de una aproximación robusta (en este caso se da privilegio a la precisión). Sin embargo, si el tiempo de cálculo es vital, entonces el interés se inclina hacia los estimadores eficaces, tal es el caso comentado con anterioridad del uso del estimador MC. Si el tiempo de cálculo puede ser dividido por un cálculo en paralelo, se podría preferir una aproximación de tipo robusto, la cual brindará una mejora en términos de precisión tomando un tiempo de cálculo razonable. En principio, el estimador sugerido por Hall y Presnell [29] él cual utiliza una función de costo cuadrática, podría permitir la optimización del tiempo de cálculo en este tipo de situaciones, preservando a su vez las características de robustez. Sin embargo, hacemos la remarca de que es necesario llevar a cabo una buena selección entre los diversos hiperparámetros como el caso de  $\rho$  y  $\varepsilon$  para minimizar al máximo la influencia de los datos indeseables. Lo anterior puede hacerse según los conocimientos apriori que se tienen sobre las propiedades del sistema bajo estudio.

## 7. Aplicación a un sistema de mediciones

A continuación presentamos algunos resultados obtenidos en el contexto de un problema de medición simulado, dichos resultados se obtuvieron utilizando los diferentes esquemas bootstrap propuestos en el caso del análisis regresivo. La meta principal es de comparar la DDP aproximada según los métodos bootstrap contra la DDP asintótica obtenida haciendo uso del método de Monte-Carlo Primitivo (MCP) el cual requiere del conocimiento completo de las DDPs que interactúan en el sistema bajo estudio. De manera práctica y para propósitos de validación, se consideran hipótesis gaussianas y no-gaussianas sobre los errores (distribuciones normal y uniforme centradas). En este ejemplo creemos que es importante preservar un marco uniforme en la simulación, para poder juzgar adecuadamente sobre el desempeño de los métodos bootstrap. Comenzamos por presentar el modelo no-lineal utilizado:

$$f(x_i, \boldsymbol{\theta}) = \theta_1 \tanh(\theta_2 x_i),$$

en donde el vector experimental está dado por  $\{x_i\}_{i=1}^{16} = [0, 1, \dots, 15]^\top$ , el vector de los parámetros considerados como *reales* es  $\bar{\boldsymbol{\theta}} = [2, 18, 0, 49]^\top$  ( $p = 2$ ) y los errores siguen una distribución normal  $e_i \sim \mathcal{N}(0, \sigma^2)$ , en donde la varianacia  $\sigma^2 = 0,0625$ . Entonces el vector de observaciones a ser analizadas  $\mathbf{y}$  se obtiene utilizando la ecuación (30) para luego obtener la estimación de la medición de interés pasando por supuesto por la estimación de los parámetros. La relación entre la medición de interés y

los parámetros podría expresarse por (medición de la pendiente en el origen  $\partial f_j / \partial x(0, \boldsymbol{\theta})$ ):

$$m_P = g(\boldsymbol{\theta}) = \theta_1 \theta_2,$$

o bien por (medición de la integral sobre los datos observados):

$$m_I = g(\boldsymbol{\theta}) = \int_{x_0}^{x_n} f(x, \boldsymbol{\theta}) dx.$$

En fin, retendremos únicamente la medición de la pendiente en el origen del modelo  $f(\cdot)$ . Notamos que se puede calcular de manera analítica el valor *real* de la pendiente  $\bar{m}_P = 1,0682$ . Por otro lado, hacemos la remarca de que el modelo utilizado en este ejemplo se seleccionó de manera simplista, pensando en obtener cálculos relativamente rápidos. De hecho, el modelo está muy bien adaptado como ejemplo académico puesto que éste nos permite obtener de manera precisa los datos  $y_i$  (simulados) y una medición de referencia  $m_P$ . Se trata pues, de un ejemplo gracias al cual se podrá mostrar de manera clara una buena comparación, en un marco de simulación bien definido en donde se supondrá que todo se conoce bien (por conveniencia). Entonces, la evaluación del desempeño de los métodos bootstrap con respecto al método de MCP cuando  $e_i \sim \mathcal{N}(0, \sigma^2)$ , podrá llevarse a cabo bajo un marco de simulación correcto tanto para los estimadores paramétricos como para el estimador de la medición  $\hat{m}$ , asimismo se podrá apreciar el grado de aproximación y precisión aún cuando los errores son considerados como no-gaussianos (notemos que los métodos bootstrap no asumen ninguna hipótesis sobre la ley de los errores  $e_i$ ).

(a) Primer caso: métodos bootstrap vs. MCP cuando este último esquema utiliza hipótesis de normalidad sobre los errores. Primeramente se generan datos  $\mathbf{y}$  sucesivamente bajo la hipótesis de que  $e_i \sim \mathcal{N}(0, \sigma^2)$ , en donde  $\sigma^2 = 0,0625$ . Las DDPs asintóticas de referencia para los parámetros y la medición son posteriormente calculadas después de  $N = 50,000$  simulaciones o iteraciones según el método MCP. Por otro lado, la aproximación de las DDPs obtenidas utilizando los métodos bootstrap se lleva a cabo para  $B = 3000$  y  $B = 5000$  simulaciones del esquema de residuos ponderados, en este caso sólo se toma en cuenta el primer juego generado de datos  $\mathbf{y}$ . Los resultados de ambos esquemas puestos en comparación se presentan en la tabla 1, en donde se muestra que los mejores resultados obtenidos son los del método bootstrap ponderado (o modificado, referirse a la ecuación (40)). Por otro lado, la figura 1 presenta una comparación visual entre las DDPs obtenidas con el método bootstrap ponderado y aquellas obtenidas con el método de Monte-Carlo Primitivo (la cual se considera como DDP de referencia).

(b) Métodos bootstrap en un caso más general en donde las hipótesis sobre los errores son non-gaussianas, por ejemplo, se propone como segundo caso una mezcla de dos leyes simétricas, la normal más la uniforme, ambas son supuestas centradas y de varianacia constante. En este segundo caso se generaron datos utilizando  $e_i \sim \mathcal{N}(0, \sigma^2) + \mathcal{U}(-\alpha/2, \alpha/2)$ , avec  $\sigma^2 = 0,0625$  sucesivamente y en donde se considero que  $\alpha = \sigma = 0,25$ . La DDP asintótica de referencia se obtuvo después de  $N = 10^5$  simulaciones del método MCP<sup>5</sup>. Por otro lado, la aproximación obtenida por medio del método bootstrap ponderado fué también calculada para diferentes valores de  $B$ , con la finalidad de vislumbrar la precisión y el nivel de convergencia de dicho método bootstrap en contraste con los resultados asintóticos de referencia obtenidos por MCP. La figura 2 muestra los datos (observaciones) generados por el modelo (sin errores) y los datos generados por el mismo modelo cuando se ha sumado la mezcla de los ruidos propuestos (datos simulados contaminados). Por otra parte, se sabe que el estimador paramétrico no-lineal se encuentra en general sesgado, por

---

<sup>5</sup>En este caso se aumentó el número de simulaciones para aumentar aún más la precisión sobre la estimación de la DDP.

Tabla 1: Algunas estadísticas bootstrap comparadas con las estadísticas obtenidas utilizando el método MCP y la medición de referencia. Las estadísticas fueron obtenidas mediante el esquema bootstrap ponderado en donde MCP considera por su parte que  $e_i \sim \mathcal{N}(0, 0,0625)$ .

	Media ( $\theta_1$ )	Mediana ( $\theta_1$ )	Sesgo ( $\theta_1$ )	Var ( $\theta_1$ )	dest ( $\theta_1$ )
MCP 5e4	2.1833	2.1827	0.0033	0.0057	0.0754
BOOT 3000	2.1845	2.1835	0.0045	0.0061	0.0782
BOOT 5000	2.1836	2.1836	0.0036	0.0063	0.0796
	Media ( $\theta_2$ )	Mediana ( $\theta_2$ )	Sesgo ( $\theta_2$ )	Var ( $\theta_2$ )	dest ( $\theta_2$ )
MCP 5e4	0.5051	0.4908	0.0151	0.0110	0.1046
BOOT 3000	0.5040	0.4898	0.0140	0.0111	0.1052
BOOT 5000	0.5061	0.4907	0.0161	0.0113	0.1065
	Media ( $m_P$ )	Mediana ( $m_P$ )	Sesgo ( $m_P$ )	Var ( $m_P$ )	dest ( $m_P$ )
MCP 5e4	1.0995	1.0713	0.0313	0.0456	0.2136
BOOT 3000	1.0975	1.0650	0.0293	0.0461	0.2148
BOOT 5000	1.1013	1.0717	0.0331	0.0465	0.2156

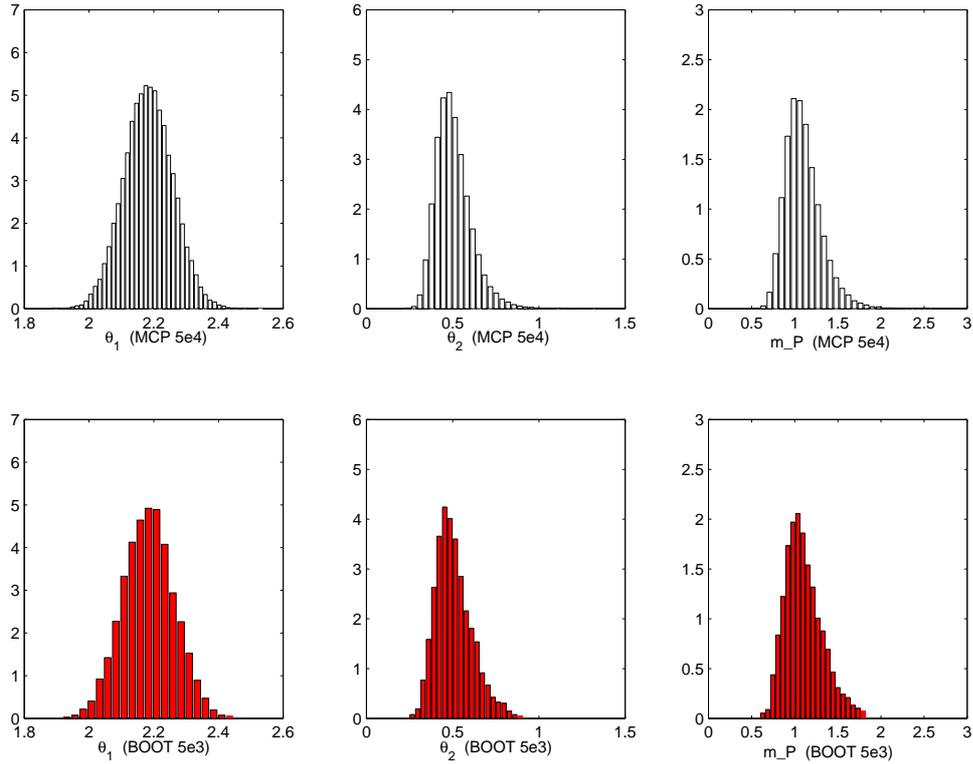


Figura 1: Histogramas marginales según el método bootstrap ponderado (abajo) contra los histogramas MPC para  $5e4$  simulaciones ( $e_i \sim \mathcal{N}(0, 0,0625)$ ).

lo que se espera que el estimador de la medición esté igualmente sesgado. Aquí se muestran las expresiones clásicas del sesgo para ambos estimadores:

$$\mathbf{b}(\theta) = E \left\{ \hat{\theta} \right\} - \bar{\theta}, \quad (54)$$

y

$$\mathbf{b}(m) = E \left\{ \hat{m} \right\} - \bar{m}, \quad (55)$$

suponiendo por supuesto que se conocen los valores *reales* de  $\theta$ :  $\bar{\theta}$ , y de  $m_\ell$ :  $\bar{m}_\ell$ . La tabla 2 muestra por otro lado algunas estadísticas bootstrap para el caso del uso de bootstrap ponderado y diferentes valores para  $B$ , se puede hacer claramente una comparación entre las estadísticas bootstrap y las de referencia (MCP). En general, las estadísticas las más precisas se obtienen cuando se realizan 5000 simulaciones del esquema bootstrap ponderado.

Los valores de  $p$  y  $q$  correspondientes según la ecuación (40) son  $p = 2$  y  $q = n - 2$ , se pueden apreciar en las figuras 3 y 4 los niveles de convergencia y precisión de los estimadores del sesgo y el de la variancia. En ambos casos, las estadísticas bootstrap obtenidas están ligeramente próximas

de los valores de referencia, sin duda si hacemos cambios en el valor de  $q$ , la precisión podría aumentar aún más. Por otro lado, la figura 5 muestra el nivel de convergencia de forma más completa (se tiene un error de 0.08, debido al número limitado de simulaciones) utilizando las DDPs bootstrap en comparación contra las DDPs de referencia (MPC). La figura 3 por su parte muestra el sesgo bootstrap ponderado estimado, se visualiza que éste converge hacia el sesgo de referencia MCP. Sabiendo que el nivel de convergencia puede ser mejorado, se utilizó el esquema bootstrap ponderado, que además converge con mayor rapidez. Dicho de otra manera, es posible obtener una muy buena aproximación hacia las DDPs de referencia disminuyendo a la vez el número de simulaciones necesarias.

La tabla 2 muestra dos resultados diferentes en donde se consideraron dos cantidades diferentes de simulaciones del método bootstrap ponderado, se ve claramente que cuando  $B = 5000$  simulaciones, se obtiene la mejor aproximación (comparando  $B = 3000$  contra  $B = 5000$  simulaciones). En lo referente al grado de precisión de la variancia, se puede verificar en la tabla 2 y en la figura 4 que los resultados obtenidos por el esquema bootstrap ponderado se aproximan a los valores de referencia con una precisión menos buena que los resultados obtenidos para el caso de

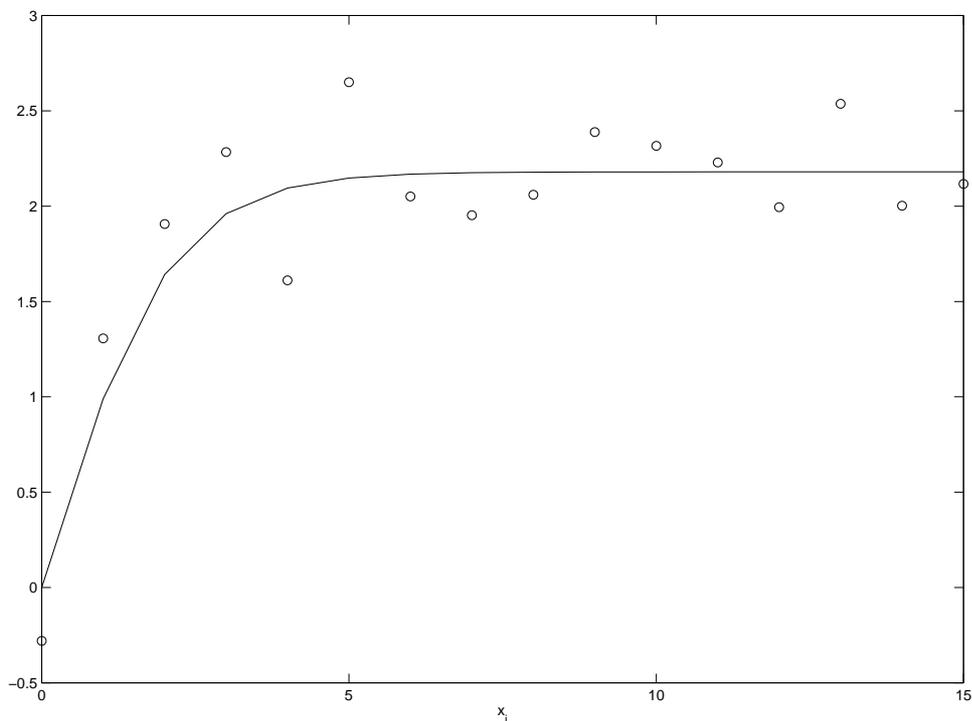


Figura 2: Representación de datos: (línea sólida) datos obtenidos por el modelo  $f(\cdot)$  y (o) datos observados  $y_i$  obtenidos por el modelo contaminado por  $e_i$ .

la estimación del sesgo. En lo que concierne al método bootstrap ponderado, para el caso en que se realizan  $B = 5000$  simulaciones se puede remarcar una mejora con respecto a la convergencia de la variancia con respecto a la estimada cuando se realizan  $B = 3000$  simulaciones.

Regresando nuevamente a la evaluación en términos de convergencia, se puede también comparar el método bootstrap ponderado con respecto al método bootstrap ponderado iterativo (BPI). En la tabla 2, se muestran algunos resultados de los cuales se puede concluir que para el método BPI con  $B = B_1 \times B_2 = 18^3$  simulaciones se obtienen muy buenos resultados en términos de precisión, por lo que el método BPI realmente mejora de forma significativa la aproximación obtenida por medio del método bootstrap ponderado. En general, en ambos casos hemos verificado que el número de simulaciones  $B$  (iteraciones) es mucho más pequeño que el número  $N$  utilizado para el método MCP, es así que se puede aseverar que ambos métodos bootstrap convergen rápidamente. Si se utiliza el esquema BPI sabiendo que la dimensión el vector muestra es  $n$ , se pueden realizar un poco más de  $B = n^3$  simulaciones con la finalidad de obtener una muy buena precisión sobre la distribución de interés  $\varphi(m)$ .

Por otro lado, nos gustaría llevar a cabo una evaluación de la distancia entre la DDP de referencia  $\varphi_r(m)$  (MCP) y la DDP bootstrap  $\varphi_b(m^*)$ . Como ya se vió en la sección anterior § 6.2, existen varios criterios para evaluar la distancia entre DDPs. Una aproximación propuesta por Shibata [57] considera un marco de selección de modelos haciendo uso de los métodos bootstrap, dicha aproximación se fundamenta en un mecanismo para acceder a la información conocida como información de Kullback-Leibler (IKL). Por su lado, Brahim [8] recomienda utilizar de preferencia una distancia métrica sobre una medida de divergencia como lo es la IKL [44], puesto que dicha medida es extremadamente sensible al comportamiento que tienen algunas DDPs de colas largas. Varias funciones han sido propuestas en la literatura, ver por ejemplo Trosset y Sets [58]. De entre tales funciones proponemos utilizar para este ejemplo la distancia de Hellinger, en particular se ha adoptado ésta distancia al caso de la aproximación vía los métodos bootstrap:

$$\Delta_H(\varphi_r(\cdot), \varphi_b(\cdot)) = \left( \int \left( \sqrt{\varphi_r(m)} - \sqrt{\varphi_b(m^*)} \right)^2 dm \right)^{1/2}. \quad (56)$$

La distancia de Hellinger podría ser utilizada de manera general con la finalidad de evaluar el grado de aproximación de la DDP de la medición bootstrap con respecto a la DDP de la medición de referencia (ver la figura 5). Por supuesto, es necesario contar con un conocimiento previo de la DDP de referencia para así poder utilizar la distancia métrica anterior. También podemos consultar otro tipo de medidas de distancia entre DDPs, la tesis de Gibbs [26] reporta una buena

comparación entre varias medidas de interés general<sup>6</sup>.

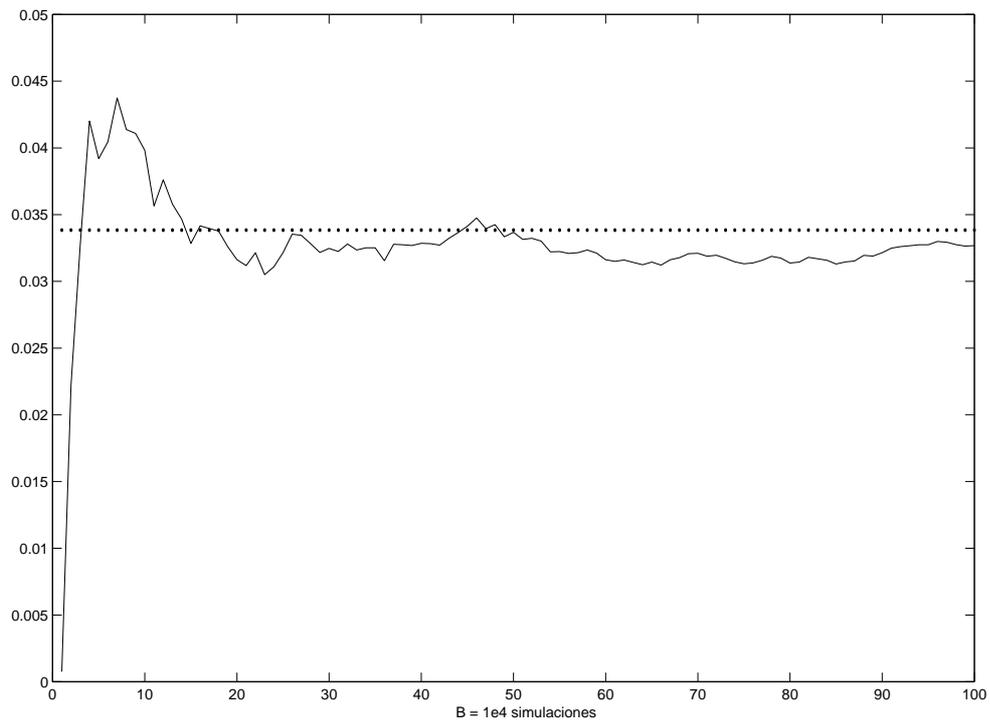


Figura 3: Convergencia del sesgo según el método bootstrap ponderado con respecto al sesgo de referencia MCP (línea punteada) para diferentes simulaciones  $B$ .

- (c) Estimación robusta de la medición: En éste tercer caso, se muestran los resultados obtenidos utilizando las ideas introducidas en la sección § 6.2 sobre estimadores bootstrap intencionalmente sesgados. Nuevamente se trata de generar datos simulados bajo las hipótesis de un ruido uniforme de ley no-centrada:  $e_i \sim \mathcal{U}(0, \alpha)$  y a tal modelo contaminado se suman algunos datos aberrantes para los cuales no se conoce su distribución (observar en la figura 6:  $y(8) = 1$ ,  $y(9) = 1,2$ ,  $y(10) = 1,4$ ,  $y(11) = 1$  et  $y(12) = 0$ ). La tabla 3 muestra algunas estadísticas obtenidas para la medición de interés cuando se utilizan los estimadores MC, bootstrap intencionalmente sesgado con norma cuadrática (BCS) y bootstrap sesgado utilizando la la función de costo cuasi-Huber (BHS). Durante la simulación se implemento el esquema bootstrap ponderado sin considerar el centrado de los vectores muestra (se descartó el uso de la media, puesto que en éste caso se trata de un ruido no-centrado) los resultados se contrastan con los resultados obtenidos por medio del esquema MCP.

<sup>6</sup>En el capítulo 6, Alison Gibbs presenta una comparación en términos teóricos y establece al mismo tiempo una relación entre las funciones presentadas.

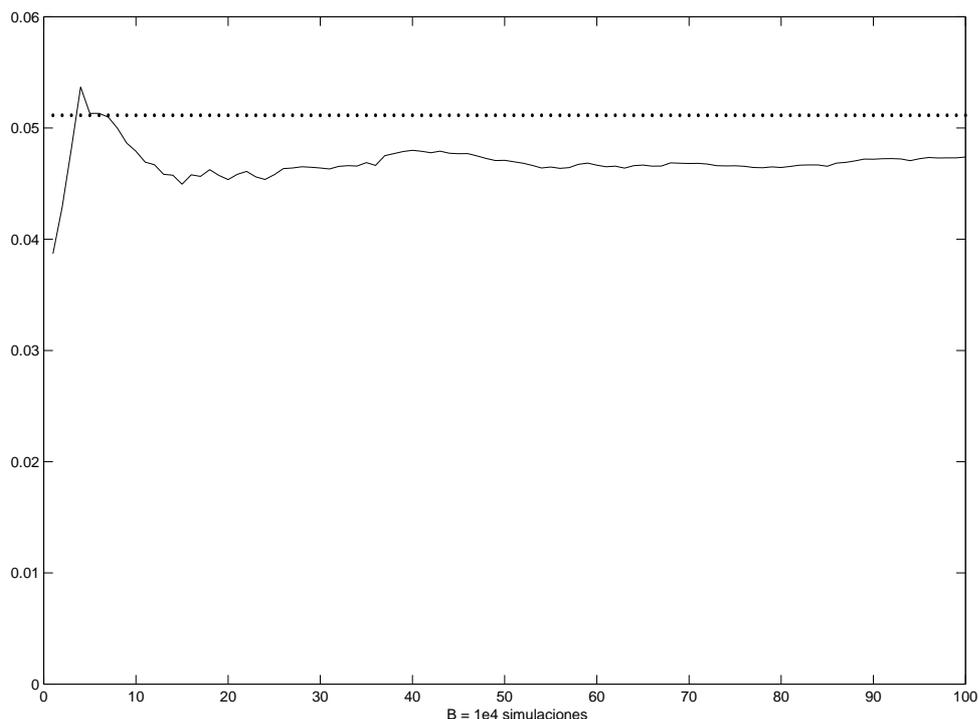


Figura 4: Convergencia de la variancia según el método bootstrap ponderado con respecto a la variancia de referencia MCP (línea punteada) para diferentes simulaciones  $B$ .

La figura 6 presenta los datos reproducidos por el modelo  $f(\cdot)$ , cuando se utilizan los valores promedio de los estimadores MC, de BCS y de BHS. De igual manera se puede remarcar que la aproximación que brinda los resultados más pobres es aquella en donde se utiliza el estimador MC. Mientras que por el contrario, la mejor aproximación es aquella en donde se utiliza el estimador BHS. No obstante, se sabe que el grado de aproximación del esquema BCS podría mejorar aún buscando optimizar al máximo la selección de los hiperparámetros  $\rho$  y  $\varepsilon$ , como ya se describió en la sección § 6.2. Para este ejemplo, los valores para  $\Delta$ ,  $\rho$ ,  $\lambda$  y  $\varepsilon$  que se seleccionaron son: 0.25, 1, 0.25 y 0.0165 respectivamente.

La figura 7 expone los diferentes histogramas obtenidos para la medición de interés. Se puede ver que el valor promedio obtenidos con el esquema BOOT-BHS se encuentra ubicado cerca del valor *verdadero* de la medición. Se ejemplifica de manera evidente que los datos aberrantes ejercen una fuerte influencia y precisamente por esta razón que la media estimada haciendo uso de los diferentes estimadores se encuentra algo alejada del valor *verdadero*. De igual manera, este ejemplo permite ver como es que el hecho de utilizar la sub-ponderación permite el desplazamiento de la DDP estimada y por lo tanto del valor promedio de la medición hacia la medición *verdadera*.

También se pueden considerar otros casos en donde los datos son obtenidos a partir de un proceso químico y en donde podría haber diferentes objetivos de medición, para este tipo de casos Bates y Watts [4] han estudiado toda una colección de modelos paramétricos bien adaptados. En lo concerniente al tiempo de convergencia, el método bootstrap ponderado utilizado en (a), como ejemplo, necesitó un tiempo de cálculo: 5.34 min para un número de simulaciones  $B = 5000$  (la plataforma de simulación utilizada fué MATLAB Ver. 5.2, Estación de trabajo SUN Sparc 5, 330 MHz, 256 Moctets de RAM), mientras que por su lado el método bootstrap ponderado iterativo requirió de un tiempo un poco más corto (según nuestra experiencia podría requerir para algunos casos del mismo tiempo).

## 8. Aplicaciones en PDS

La introducción de los métodos bootstrap como herramienta para el procesamiento digital de señales y sus aplicaciones se dió con algunos de los trabajos de Beadle y Djurić [5] y Zoubir [48], [65] – [71]. En donde se enuncian aplicaciones que van desde el modelado de señales y selección de modelos hasta la

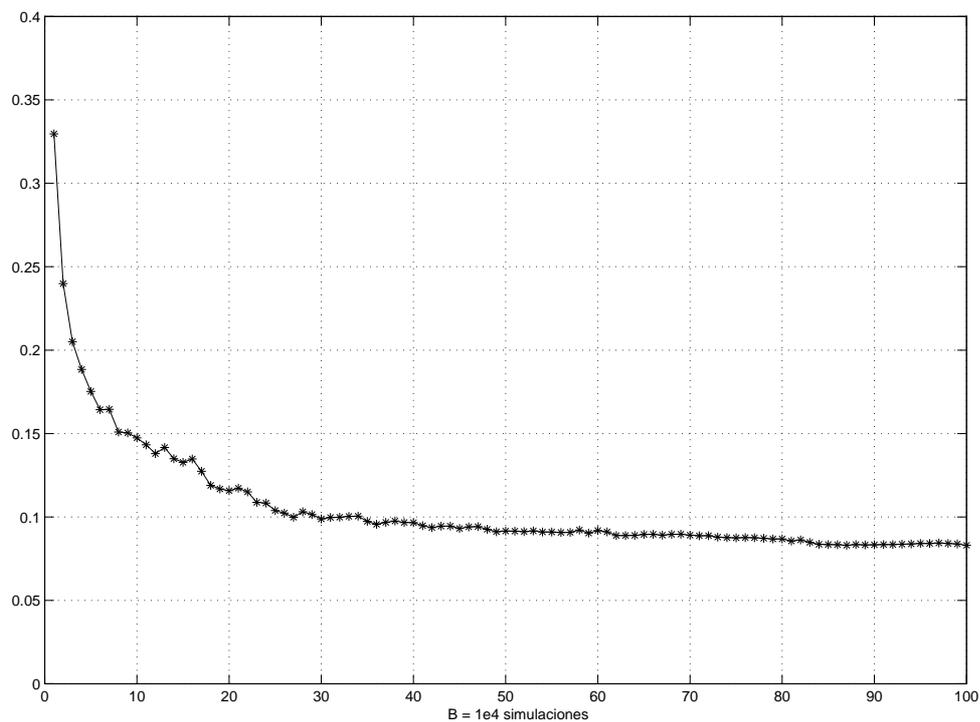


Figura 5: Distancia  $\Delta_H(\varphi_r(\cdot), \varphi_b(\cdot))$  entre los histogramas aproximados por el método bootstrap ponderado (bootstrap modificado  $\varphi_b(\hat{m}^*)$ ), y el histograma asintótico de referencia MCP ( $\varphi_r(m)$ ) para el caso de la medición.

Tabla 2: Bootstrap ponderado y bootstrap iterativo con respecto a las estadísticas de referencia MCP, para  $\theta$  y  $m_P$ , se utilizó el estimador MC (en este caso sub-óptimo).

	Media ( $\theta_1$ )	Mediana ( $\theta_1$ )	Sesgo ( $\theta_1$ )	Var ( $\theta_1$ )	dest ( $\theta_1$ )
MCP 1e5	2.1835	2.1827	0.0035	0.0059	0.0768
BOOT 3000	2.1832	2.1808	0.0032	0.0058	0.0762
BOOT 5000	2.1831	2.1808	0.0031	0.0056	0.0749
	Media ( $\theta_2$ )	Mediana ( $\theta_2$ )	Sesgo ( $\theta_2$ )	Var ( $\theta_2$ )	dest ( $\theta_2$ )
MCP 1e5	0.5064	0.4905	0.0164	0.0123	0.1110
BOOT 3000	0.5056	0.4857	0.0156	0.0111	0.1054
BOOT 5000	0.5062	0.4858	0.0162	0.0112	0.1058
	Media ( $m_P$ )	Mediana ( $m_P$ )	Sesgo ( $m_P$ )	Var ( $m_P$ )	dest ( $m_P$ )
MCP 1e5	1.1020	1.0700	0.0338	0.0511	0.2260
BOOT 3000	1.1007	1.0612	0.0325	0.0464	0.2154
BOOT 5000	1.1019	1.0614	0.0337	0.0471	0.2170
n = 14					
BPI 2744	1.0993	1.0644	0.0311	0.0448	0.2116
n = 18					
BPI 5832	1.1048	1.0629	0.0366	0.0481	0.2193

Tabla 3: Estadísticas de la medición según los diferentes estimadores bootstrap en contraste con el estimador de MC.

$\bar{m}_P = 1.0682$	Media ( $m_P$ )	Mediana ( $m_P$ )	Sesgo ( $m_P$ )	Var ( $m_P$ )	dest ( $m_P$ )
MCP-MC	1.5367	1.5261	0.4685	0.0115	0.1073
MCP-BHS	1.2813	1.2809	0.2131	0.0049	0.0700
BOOT-BCS	1.4296	1.4216	0.3614	0.0088	0.0938
BOOT-BHS	1.2708	1.2690	0.2026	0.0067	0.0821

manera de optimizar la posición de transductores.

### 8.1. Modelos AR, MA, ARMA, ARN, etc.

El análisis de señales temporales o series en el tiempo es una tarea que se realiza comunmente en muchas aplicaciones del procesamiento de señales. Las series temporales pueden ser considerados procesos estocásticos que se modifican con respecto al tiempo y bajo algunas condiciones se pueden considerar como señales estacionarias. Existe una gran cantidad de modelos paramétricos capaces de modelar señales estocásticas dependientes del tiempo, tales modelos son los modelos: Autoregresivo o en inglés “Autoregressive linear (AR)”, Autoregresivo no-lineal o en inglés “Autoregressive non-linear (ARN)”, Autoregresivo con promedio móvil o en inglés “Autoregressive Moving Average (ARMA), Autoregresivo con variables exógenas o en inglés “Autoregressive linear with exogenius variables (ARX)”, etc. Todos estos modelos han sido desarrollados y estudiados ampliamente dentro del análisis espectral moderno y aplicados al procesamiento y caracterización de señales de voz, en análisis de modelos economicos, en ingeniería sísmica, en medicina, en química, en geofísica, etc.

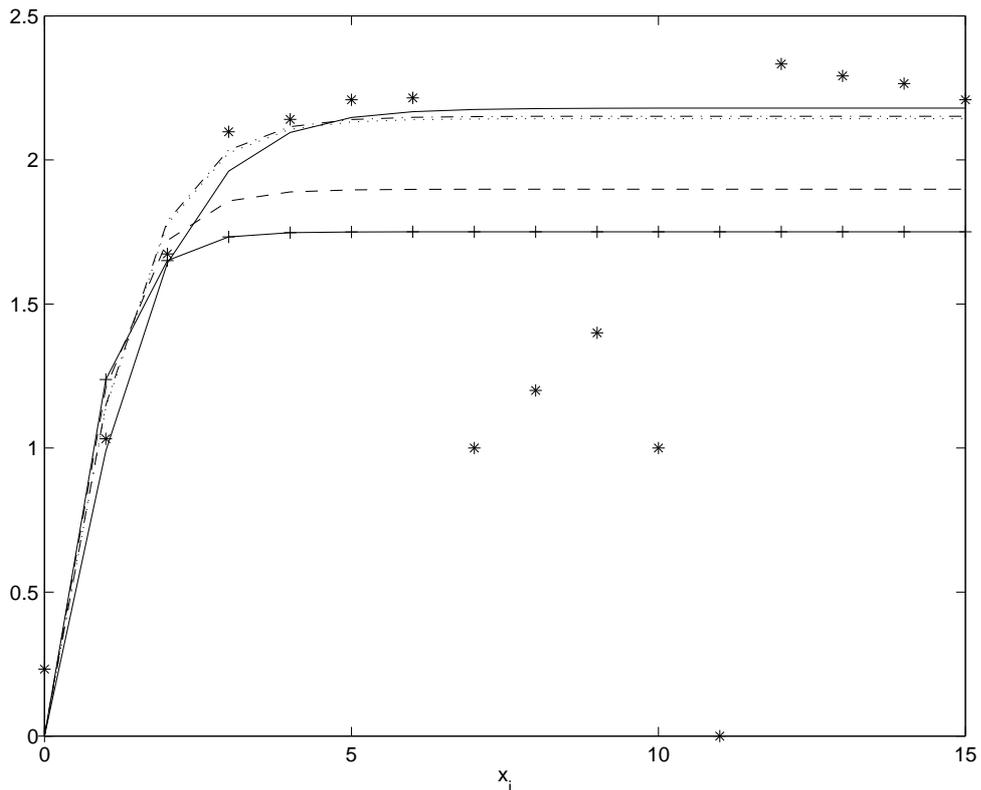


Figura 6: Comparación entre los diferentes estimadores paramétricos, se utiliza su valor promedio. Generación de datos utilizando únicamente el modelo (—), Datos contaminados por un ruido uniforme más algunos datos aberrantes (\*), MCP-MC (-+), MCP-BHS (- - -), Bootstrap-BCS (- - -) y Bootstrap-BHS (· · ·).

El modelo AR se introdujo con el propósito de mejorar la estimación espectral clásica, y se trata del modelo más simple para representar una señal. Mucha de la literatura existente sobre este modelo está consagrada a una cantidad muy grande de aplicaciones de tipo industrial. Lo anterior se debe a que la clase de señales autoregresivas permite modelar una gran gama de señales estocásticas que se pueden encontrar en la práctica (señales de voz, sonar, de vibraciones, de textura en las imágenes, etc.). La determinación de los parámetros AR conduce a la resolución de sistemas lineales, para los cuales existe una gran cantidad de métodos de resolución que pueden ser adaptados a la particularidad de algunos sistemas. Por otro lado, la naturaleza intrínseca que tienen los modelos AR de codificar la señales permite obtener una compresión notable de las señales cuando éstas poseen una cierta redundancia. De hecho, el modelo AR fué justificado por una consideración teórica que consiste en Maximizar la Entropía de una señal AR. Lo cual convierte al modelo AR en un modelo a mínimo de información, el cual respeta el conocimiento de los  $p$  primeros coeficientes de autocorrelación de una señal.

La ecuación de diferencia o relación de recurrencia temporal de una señal autoregresiva de orden

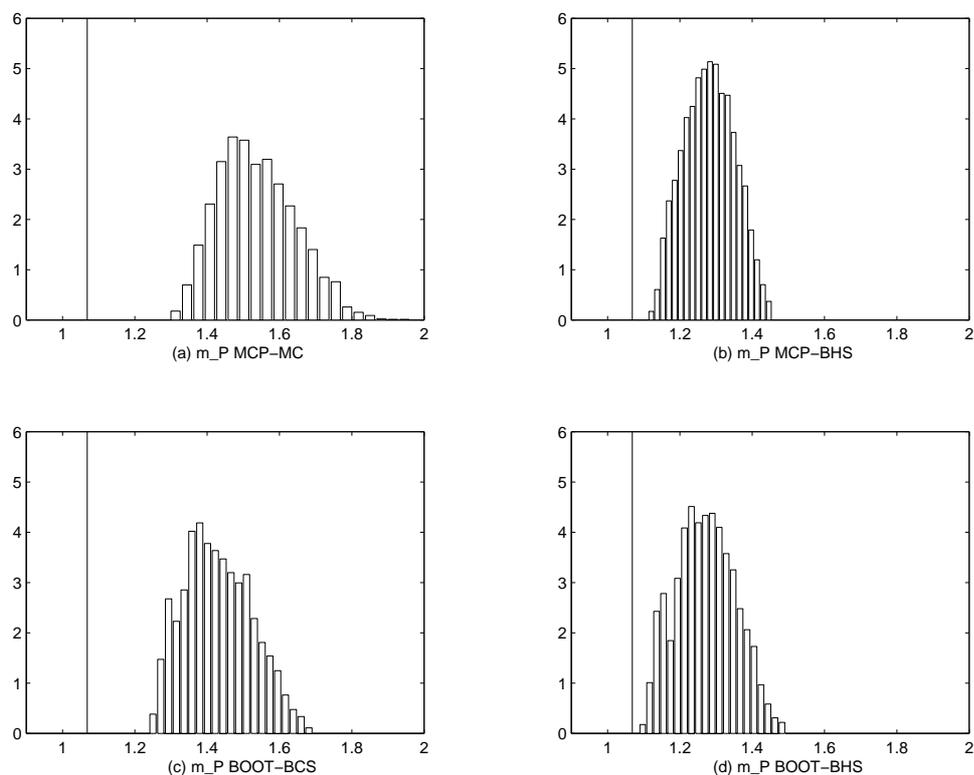


Figura 7: Comparación entre la medición de referencia o verdadera (—) y los diferentes histogramas: MCP-MC (a), MCP-BHS (b), Bootstrap-BCS (c) y Bootstrap-BHS (d).

$p$  se puede escribir según el modelo:

$$y_k = \sum_{i=1}^p a_i y_{k-i} + \sigma u_k \quad (57)$$

en donde  $\{a_i\}$  son coeficientes o parámetros desconocidos,  $\sigma$  es un factor multiplicativo y  $u_k \sim \mathcal{N}(0, \sigma^2)$  es un ruido blanco aditivo de media cero y variancia constante. La expresión anterior también representa un filtro inverzo cuya expresión general esta dada por:  $H(z) = \sigma/A(z)$  en donde  $A(z) = 1 - a_1 z^{-1} - \dots - a_p z^{-p}$  (referirse también a los filtros de respuesta al impulso infinito – RII). Una señal AR es *estacionaria* de media cero y con coeficientes de correlación  $r_k = E\{y_i y_{i-k}\}$ . En lo que concierne a los estimadores de los parámetros, estos pueden ser tan clásicos como los vistos para los modelos regresivos lineales si se trata de una estimación fuera de línea (no se requiere cálculo en tiempo real) o bien métodos recursivos como mínimos cuadrados recursivos o el método del gradiente estocástico si se requiere de una estimación en línea.

El modelo de promedio móvil (también conocido como promedio ajustado) o en inglés “Moving Average” (MA) permite modelar señales las cuales tienen un espectro que posee mínimos de energía bastante marcados, esto último es difícil de realizar con los modelos AR. Por el contrario, los modelos MA no pueden representar señales las cuales tienen un espectro con bandas bastante estrechas (referirse también a los filtros de respuesta al impulso finito – RIF), y es por esto que su uso no es tan extendido puesto que es más frecuente tratar señales cuyo espectro posee frecuencias preponderantes que señales con espectros caracterizados por ausencias de energía a ciertas frecuencias. El modelo MA tiene la siguiente estructura,

$$y_k = \sum_{j=0}^q b_j u_{k-j} \quad (58)$$

en donde  $\{b_j\}$  son coeficientes o parámetros desconocidos.

El modelo autoregresivo de promedio móvil (ARMA) es un modelo generalizado del AR en donde se toman en cuenta las señales que también tienen un espectro que posee mínimos de energía al rededor de algunas frecuencias. Se dice que este modelo permite tomar en cuenta de manera concreta los polos y los ceros del espectro, la ecuación en términos de diferencias de una señal autoregresiva de promedio móvil de orden  $p$  y  $q$  se puede escribir según el modelo:

$$\sum_{i=1}^p a_i y_{k-i} = \sum_{j=0}^q b_j u_{k-j} \quad \text{con } a_0 = 1. \quad (59)$$

Los estimadores que se utilizan para la aproximación de los parámetros desconocidos son construidos de tal suerte que se considera la estimación de un modelo AR con vector de parámetros extendido, y un método óptimo en muchas ocasiones conocido como método iterativo de Durbin (referirse al algoritmo de Levinson-Durbin).

Otros modelos de tiempo discreto que son modelos generalizados según los antes presentados varían desde el modelo ARX, ARARX, ARMAX, ARARMAX, CARMA, NARMAX, CARIMA, etc. [61], estos modelos se utilizan con mayor frecuencia en aplicaciones de control.

## 8.2. Bootstrap en el caso de hipótesis de correlación

Hasta ahora solo se habían considerado casos en donde se consideran v.a.i.i.d., pero existen otros casos sobre todo cuando se lleva a cabo un análisis de datos variantes en el tiempo o no estacionarios, en donde las hipótesis más comunes sobre los errores toman en cuenta el tipo de dependencia conocida como correlación entre v.a. En estos últimos casos, los esquemas bootstrap anteriormente presentados no brindan resultados adecuados ya que no están correctamente adaptados a la situación bajo estudio, en los casos de correlación es necesario poder capturar la dependencia existente entre las variables aleatorias, o bien conocer el proceso generador de dicha dependencia. En esta sección, presentamos de manera breve un estado del arte de algunas aproximaciones bootstrap que toman en cuenta la correlación de los datos que serán modelados. Nos interesa evidenciar dichos métodos con la finalidad de poderlos explotar en problemas de procesamiento digital de señales [65], [66], [70] en donde los errores pertenecientes al modelado se suponen correlacionados (Autoregressive linear modelling (AR), Autoregressive non-linear modelling (ARN), Autoregressive Moving Average modelling (ARMA), Autoregressive linear with exogenous variables (ARX), etc.).

Como ya se vió con anterioridad, una implementación relativamente simple de un proceso generativo de variables dependientes se lleva a cabo utilizando un modelo paramétrico, como el modelo autoregresivo de promedio móvil (ARMA). Se trata de un modelo que transforma el proceso de generación de v.a.i.i.d. a v.a.d.i.d. (variables aleatorias dependientes e idénticamente distribuidas). De hecho, se trata de un ejemplo clásico de generación de datos correlacionados, la serie de datos  $y_k$  es generada por:

$$A(q^{-1}, \mathbf{a})y_k = B(q^{-1}, \mathbf{b})u_k \quad (60)$$

en donde  $A$  y  $B$  son funciones conocidas,  $q^{-1}$  es el operador de retardo,  $\mathbf{a}$  y  $\mathbf{b}$  son los vectores de parámetros desconocidos de dimensión  $p$  y  $q$  respectivamente, y  $u_k$  es una secuencia de v.a.i.i.d. ( $u_k \sim$

$\mathcal{N}(0, \sigma^2)$ ). O bien en su forma de diferencias

$$\sum_{i=1}^p a_i y_{k-i} = \sum_{j=0}^q b_j u_{k-j} \quad \text{con } a_0 = 1.$$

Para considerar nuevamente los métodos de remuestreo bootstrap, es necesario realizar algunas pequeñas modificaciones con la finalidad de tener la capacidad de detectar la dependencia entre las variables aleatorias. Se pueden citar algunos trabajos como [40], [50], en los cuales se han desarrollado algunas aproximaciones fundamentadas en los métodos bootstrap no-paramétricos con la finalidad de modelar la dependencia para el caso de un cierto número de densidades de probabilidad. Este otro tipo de técnicas bootstrap pueden ser implantadas dividiendo los vectores de variables aleatorias en bloques y dichos bloques serán remustreandos con remplazo. Algunas técnicas propuestas en la literatura son: *el método bootstrap por desplazamiento de bloques* (en inglés “the moving blocks bootstrap”), *el método bootstrap por bloques circulares* (en inglés “the circular blocks bootstrap”), *el método bootstrap por bloques de bloques* (ou en inglés “the blocks of blocks bootstrap”) y *el método bootstrap estacionario* (en inglés “the stationary bootstrap”). Para información más detallada sobre la teoría de dichos métodos, se pueden consultar las referencias antes citadas y otros trabajos como: [47], [51], [52].

### 8.3. Señales de voz (identificación, síntesis, etc.)

Durante el procesamiento digital de señales de voz (PDV) frecuentemente se utiliza un modelado para simplificar las tareas de análisis, de síntesis, de identificación de locutor, etc. La selección de modelos es un proceso entonces necesario y en muchas aplicaciones dicha selección se ha llevado a cabo de manera heurística, debido a los tiempos requeridos para el procesamiento, aunque a veces se presenta una deficiencia. Las deficiencias pueden ser debidas a varios factores y no únicamente a una mala selección de modelo (u orden del modelo), tal vez se deben también a una mala selección del proceso de minimización y por lo tanto de la estimación paramétrica, puesto que se trata de modelos paramétricos los cuales representan lo más cercanamente posible el sistema de habla humano. El proceso de producción de voz, por ejemplo, es modelado incorporando algunos elementos importantes que intervienen en el sistema físico generador (posiblemente no se incorporan todos los elementos que se deberían, puesto que se trata de un proceso muy complejo y no completamente comprendido). En fin, ya sea que se trate de realizar un análisis o síntesis de voz, los modelos paramétricos pueden ser utilizados (puesto que se supone que el sistema de habla ha sido modelado) y éstos se auxilian frecuentemente vía un análisis por predicción lineal o en inglés “Linear Prediction Coding” (LPC) para llevar a cabo la identificación de los coeficientes de reflexión y sucesivamente los coeficientes de predicción lineal. Existen varias razones

que justifican el uso del análisis LPC, éstas son a) se llega a reproducir bastante bien la voz, b) por lo que la precisión del análisis LPC es bastante adecuada y c) se puede contar con una alta velocidad durante el cálculo de los coeficientes.

La idea principal que se encuentra detras de la predicción lineal es que una muestra de voz  $s(n)$  puede ser predecida a partir de una suma lineal de  $p$  muestras previamente ponderadas, es decir,

$$\widehat{s}(n) = \sum_{i=1}^p a_i s(n-i) \quad (61)$$

en donde  $\{a_i\}$  son los coeficientes de predicción y pueden ser detreminados minimizando el error cuadrático medio entre las muestras de voz actuales y las muestras linealmente predecidas.

Un modelo paramétrico o filtro digital generalmente utilizado para la producción de voz tiene la función de transferencia,

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 + \sum_{i=1}^p a_i z^{-i}} = \frac{G}{A(z)} \quad (62)$$

la cual produce una señal de voz  $s(n)$  dada una señal de excitación de entrada  $u(n)$  (que puede ser un tren de impulsos o ruido blanco), y en donde  $G$  representa la ganancia del filtro. Esta función equivale a un modelo conocido como *modelo todo polos*, en donde las raices del polinomio del denominador  $A(z)$  son los polos del sistema ( $A(z)$  es también conocido como filtro inverzo). Entonces, el modelo básico de producción de voz está dado por la siguiente ecuación de diferencia en el dominio temporal,

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (63)$$

note que se trata de un modelo autoregresivo de la señal de voz, y la similitud con la ecuación de  $\widehat{s}(n)$  (si se eliminara el segundo término correspondiente a la señal de excitación serían idénticas). Para encontrar la señal de voz estimada  $\widehat{s}(n)$  a partir de la señal muestra  $s(n)$  es necesario minimizar el error cuadrático medio de un segmento corto de la señal de voz (considerar la estacionariedad del modelo AR), en donde el error total de la señal esta dado por

$$e(n) = s(n) - \widehat{s}(n) = s(n) - \sum_{i=1}^p a_i s(n-i) \quad (64)$$

y el error cuadrático medio,

$$ECM_n = \sum_m [e_n(m)]^2 \quad (65)$$

sobre un rango  $m$  a ser especificado. Puesto que se trata de obtener  $\partial ECM_n / \partial a_i = 0$ , el siguiente paso es el de resolver el sistema de ecuaciones normales dado por las ecuaciones

$$\sum_{i=1}^p a_i \sum_m s_n(m-k) s_n(m-i) = \sum_m s_n(m-k) s_n(m), \quad k = 1, \dots, p \quad (66)$$

para la resolución del sistema se recurre a dos métodos, el método de autocorrelación (referirse también al método de máximo de entropía) y el de covariancia. En fin, en lo que respecta al orden del modelo, de manera heurística se ha recomendado utilizar modelos de orden 13 o 14 considerando un rango de 10 kHz y para el cual se tendrán al menos las formantes principales de la secuencia analizada (los modelos de orden 8 al 14 son utilizados algunas veces en aplicaciones de codificación). En fin, en lo referente a este último problema, existen aún vías de investigación abiertas en donde los métodos bootstrap han sido utilizados, pero no propiamente para aplicaciones en PDV, pero si para el caso del modelado AR, ARMA, etc. de otras señales. Una propuesta futura es trabajar en nuevos esquemas de selección partiendo de las investigaciones de Shao [55], de Zoubir [68], [70], [71] y esquemas propuestos en un marco Bayesiano en el trabajo de tesis de Jaco Vermaak [60] (donde incluso se proponen modelos más completos para las señales de voz, se toma en cuenta la no-estacionariedad de dichas señales y que han sido afectadas por ruido).

#### 8.4. Modelado de imágenes

En lo que concierne al procesamiento digital de imágenes (PDI), Cho *et al* [10] son de los pocos que hasta ahora han propuesto la utilización de los métodos bootstrap. Quizá esto se deba a las grandes cantidades de tiempo de cálculo empleadas cuando se tiene que obtener una gran cantidad de parámetros, aún y cuando se utilice un modelo representativo de la secuencia de imágenes como se propone en el trabajo de tesis de Asswad [2]. Aunque podrían buscarse argumentos de más peso para considerar su uso, puesto que por su parte los métodos de Monte-Carlo mediante cadenas de Markov (en inglés “Markov Chain Monte Carlo” – MCMC) si han logrado captar la atención desde hace tiempo en ésta misma sub-disciplina del PDS. También se trata por supuesto de métodos que requieren grandes cantidades de cálculo, entre los trabajos recientes aplicados a imágenes se encuentran el de Gibbs [26].

Por el momento retomemos el trabajo de Asswad [2], en donde se estudió la posibilidad de implementar esquemas de reconstrucción en tiempo real de una señal a dos dimensiones y con muestras erróneas o faltantes. Dicho trabajo se enfocó propiamente a la reconstrucción de imágenes y se proponen algoritmos de identificación u estimación paramétrica local, la señal bidimensional es modelada por un proceso AR-2D (a 2 dimensiones).

Se supone que la señal bidimensional es barrida de derecha a izquierda y de arriba hacia abajo, un modelo general clásico es representado por la siguiente ecuación de diferencias,

$$\sum_k \sum_l b(k, l) y(i - k, j - l) = \sum_i \sum_j a(m, n) x(i - m, j - n) \quad (67)$$

en donde  $x(\cdot, \cdot)$  representa la señal de entrada y  $y(\cdot, \cdot)$  la señal de salida, y la región de soporte de los coeficientes  $a(\cdot, \cdot)$  y  $b(\cdot, \cdot)$  determina un nivel de clasificación del filtro o sistema recursivo, puesto que se supone que las condiciones de estabilidad del filtro se satisfacen, es decir que la respuesta al impulso del sistema es sumable y que

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |h(i, j)| < \infty \quad (68)$$

El modelado e identificación de señales bidimensionales muestreadas de manera clásica ha suscitado una enorme cantidad de investigaciones. Algunas de éstas conciernen a los métodos de estimación paramétrica adaptativa, ya sea para mejorar la calidad de las imágenes (en terminos de la relación señal a ruido) [27], [43], [64] o con la finalidad de compresión de imágenes para así mejorar los desempeños de los codificadores DPCM [1], [11].

En trabajos anteriores al de Asswad [45], [46], [41], se desarrollaron métodos adaptables para la estimación de señales bidimensionales por medio de modelos AR/ARMA, en dichos trabajos se consideran esquemas de muestreo no periódico de señales unidimensionales (signaux à échantillonnage non périodique – SENP). Los resultados obtenidos eran bastante prometedores puesto que mostraban un buen comportamiento del método propuesto aún cuando el porcentaje de muestras faltantes es importante. Entonces, el trabajo [2] se enfoca básicamente al problema de identificación de señales AR-2D considerando que existen un cierto porcentaje de muestras faltantes y dicha identificación se lleva a cabo ayudándose de algoritmos adaptables como los métodos de mínimos cuadrados (LMS) que permiten realizar un tratamiento en tiempo real (las causalidades del primer cuadrante del plano (QP) y de medio plano asimétrico (NSHP) permiten la construcción pertinente de algoritmos adaptables que pueden ser implementados en tiempo real, ver la figura 8).

#### 8.4.1. Descripción de los algoritmos y formulación del problema

Se supone que las muestras de la señal a 2D  $y(i, j)$  se encuentran sobre una matriz regular en un plano descrito por dos ejes ortogonales y que los instantes de llegada de las muestras son múltiplos de un periodo de muestreo oculto o desconocido, es decir que algunas muestras faltan o se perdieron (referirse a la figura 8).

Para un modelo AR-2D, el soporte de predicción  $S$  está definido por el conjunto de puntos con índices  $(m, n)$  tal que:

$$S = \{0 \leq m \leq M - 1 \quad , \quad 0 \leq n \leq N - 1 \quad , \quad m + n > 0\} \quad (69)$$

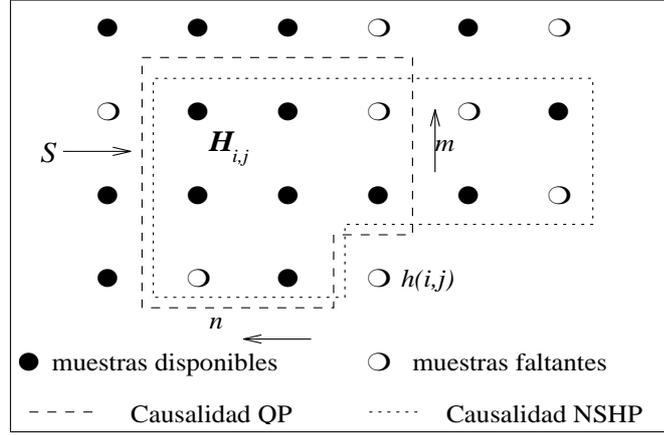


Figura 8: Representación de datos en dos dimensiones

para una causalidad QP y por otro lado:

$$S = \{0 < m \leq M - 1 \quad , \quad -N + 1 \leq n \leq N - 1\} \cup \{m = 0 \quad , \quad 0 < n \leq N - 1\} \quad (70)$$

para una causalidad NSHP. La pareja  $(M, N)$  corresponde al orden del modelo. También se introduce la matriz  $\mathbf{H}_{i,j}$  (referirse a la figura 8) la cual representa los valores de los pixeles sobre el soporte de predicción  $S$ .

$$\forall h(k, l) \in \mathbf{H}_{i,j}, \quad h(k, l) = \begin{cases} y(k, l) & \text{si la muestra está disponible} \\ \hat{y}(k, l) & \text{si no lo está} \end{cases} \quad (71)$$

$$\text{con} \quad \hat{y}(k, l) = \sum_{(m,n) \in S} \theta(m, n) h(k - m, l - n) \quad (72)$$

Lo cual significa que las muestras faltantes serán reemplazadas por sus estimaciones  $\hat{y}(k, l)$  obtenidas a partir de un predictor AR-2D adaptable según la ecuación (72).

Los parámetros  $\theta(m, n)$  del modelo son frecuentemente representados por una matriz  $\Theta$  que se encuentra sobre el soporte de predicción. El criterio  $\mathcal{J}(i, j)$  se puede definir únicamente en los instantes de llegada de las muestras. Según lo mencionado en la referencia [43] y con la finalidad de poder obtener expresiones analíticas, se selecciona un criterio cuadrático sobre el error de estimación:

$$\mathcal{J}(i, j) = [e(i, j)]^2 = [y(i, j) - \hat{y}(i, j)]^2 = [y(i, j) - \text{tr}(\mathbf{H}_{i,j} \Theta^T)]^2 \quad (73)$$

La actualización de los parámetros se lleva a cabo gracias a la siguiente formula de optimisation de tipo gradiente:

$$\Theta_{i,j} = \Theta_{i,j-1} + \mu \mathbf{G}_{i,j} \quad (74)$$

en donde:

$$\mathbf{G}_{i,j} = -\frac{1}{2} \frac{\partial \mathcal{J}(i,j)}{\partial \Theta} = e(i,j) \frac{\partial h(i,j)}{\partial \Theta} \quad (75)$$

$e(i,j)$  es el error de predicción y  $\mu$  es el paso de adaptación que controla la estabilidad, la convergencia y las características del error residual durante la fase de adquisición del predictor. La selección de dicho parámetro a sido bastante estudiada en algunos trabajos, como referencias tenemos [11], [42] y [38]. En fin, en lo referente al desarrollo de los algoritmos recursivos QP y NSHP, en el anexo B se describen de manera breve. Un trabajo aplicativo de los métodos bootstrap en el caso del filtrado recursivo o adaptable en sistemas dinámicos [5] podría servirnos como referencia para su extensión al procesamiento de imágenes en tiempo real, como lo es el caso de los filtros por partículas (en inglés “particle filters”).

Por otro lado, otro tipo de aplicaciones del modelado de imágenes, podría no necesitar del cálculo en linea, en estos casos también existen modelos relativamente sencillos de tipo regresivo y autoregresivo en donde podrían ser utilizados otro tipo de estimadores.

## 9. Relación con la selección de modelos y construcción de límites de confianza

### 9.1. Selección de modelos o del orden de un modelo anidado

Uno de los problemas quizá definido parcialmente hasta ahora es la selección eficaz del orden del filtro o modelo, de entre los métodos existentes se enuncian: el método del error de predicción final (en inglés “Final Prediction Error” – FPE) de Akaike,

$$\text{FPE}(p) = \frac{n + (p + 1)}{n - (p + 1)} \hat{\sigma}_p^2 \quad (76)$$

en donde  $\hat{\sigma}_p^2$  es un estimador de la variancia del ruido generador. Otro método desarrollado por Akaike según el principio de MV es el criterio de información de Akaike (en inglés “An Information Criterion”

– AIC), y hace hipótesis de gaussianidad,

$$\text{AIC}(p) = \frac{2p}{n} + \log \hat{\sigma}_p^2 \quad (77)$$

los dos métodos son asintóticamente equivalentes cuando  $n \rightarrow \infty$ . Una variante del criterio AIC fué introducida por Rissanen, ésta es llamada descripción por longitud mínima (en inglés “Minimum Description Length” – MDL),

$$\text{MDL}(p) = \frac{p \log n}{n} + \log \hat{\sigma}_p^2 \quad (78)$$

el término corrector  $p \log n$  crece mas rápidamente que  $p$  y se piensa que debería proporcionar un orden más óptimo que el obtenido con AIC. En fin, otro método propuesto por Parzen es llamado criterio de transferencia autoregresiva (en inglés “Criterion Autoregressive Transfer” – CAT),

$$\text{CAT}(p) = \frac{1}{n} \sum_{i=1}^p \left( \frac{n-i-1}{n} \frac{1}{\hat{\sigma}_i^2} \right) - \frac{n-p-1}{n} \frac{1}{\hat{\sigma}_p^2} \quad (79)$$

el problema de los métodos anteriores es que son poco efectivos cuando  $n$  es de dimensión limitada (se trata de métodos asintóticos), y por lo tanto regresando a la aplicación de los métodos bootstrap se vislumbra nuevamente su potencialidad. Algunos trabajos han presentado resultados en donde efectivamente los métodos bootstrap ofrecen un mejor desempeño [68], [70] y [71].

## 9.2. Estimación de límites de confianza

Percentiles, percentiles t, estabilidad de variancia.

Trabajo futuro sobre el análisis de estimadores de la media.

## 10. Resumen y comentarios

Los métodos bootstrap representan una herramienta potente, por medio de la cual se puede llevar a cabo un análisis estadístico completo de ciertos procesos aleatorios. Se ha demostrado que su carácter empírico permite la aplicabilidad de dicha herramineta en problemas aplicativos reales. Además, la implementación de estos métodos es extremadamente sencilla. Gracias al principio de remuestreo bootstrap y los métodos de simulación Monte-Carlo es posible obtener varios estimadores estadísticos a la vez, hacemos la remarca notoria sobre los estimadores de la mediana, los cuales brindan resultados aún más precisos que los estimadores de la media o del valor promedio.

- Ésta herramienta está particularmente adaptada para los casos en los cuales solo se cuenta con un número limitado de datos u observaciones. Los conocimientos apriori que se tienen sobre el

comportamiento estadístico de los errores pueden tomarse en cuenta con la finalidad de mejorar la convergencia, la eficacia y la consistencia de los estimadores bootstrap. Por ejemplo, si se verifica que la hipótesis de gaussianidad es aproximadamente válida, fácilmente se puede parametrizar una ley normal, se estima el valor promedio y la variancia del ruido, utilizando métodos precisos como el estimador de MV corregido o bien los estimadores de jackknife (bootstrap paramétrico).

- En el caso en donde los conocimientos sobre los errores o el ruido son muy limitados, los métodos bootstrap no-paramétricos pueden ser ampliamente utilizados. Además, se vió en la sección § 7 que cuando las hipótesis hechas sobre los errores son gaussianas de manera legítima, el método bootstrap ponderado brinda resultados bastante cercanos a los de referencia obtenidos utilizando el esquema MCP, pero con un costo de tiempo menor. La mejora sensible aportada haciendo una modificación a los residuos, asegura asimismo la consistencia y la convergencia de la DDP de la medición de interés. De hecho, existen varias formas de mejorar la convergencia y la consistencia como se describió en la sección § 5.2, en donde el método bootstrap propuesto presenta una cierta robustez intrínseca.
- Por otro lado, el uso de un esquema bootstrap ponderado iterativo nos demuestra que si los errores tienen una variancia constante y si dichos errores siguen una clase de ley simétrica, entonces, el nivel de precisión podría ser mejorado de manera significativa. Los algoritmos bootstrap ponderado y bootstrap ponderado iterativo convergen hacia la DDP de interés de manera rápida y reduciendo el número necesario de simulaciones (ésto con respecto al método de MCP). El número de iteraciones podrá establecerse de manera casi conocida, sabiendo que se tiene conocimiento aproximativo de  $B$  según la desigualdad:  $B \geq n^3$  (en el caso de un vector aleatorio de dimensión  $n$  moderada).
- En lo concerniente a la robustez con respecto a la existencia de datos aberrantes, se verificó que la aproximación propuesta por Hall y Presnell [30] permite realmente construir estimadores robustos, los cuales a su vez pueden ser extendidos al contexto de la medición y más aún, a un marco de aplicaciones más general como el de DSP. La introducción de una función próxima a la de Huber, pero en donde ésta última es fácilmente diferenciable permite proporcionar una redundancia durante la sub-ponderación (cuando ésta es necesaria). Sin embargo, como en el caso de los M-estimadores la eficacia se verá reducida y por otra parte el tiempo de simulación aumenta. En todo caso, se podría buscar establecer un compromiso entre la eficacia y la robustez, según los conocimientos apriori que se tengan sobre el sistema a ser modelado y las necesidades

del experimentador concernientes al tiempo de cálculo y de la implementación de los algoritmos.

- Los métodos bootstrap brindan además algunas respuestas cuando los datos a analizar son dependientes entre ellos (correlación). Finalmente, se puede decir que los métodos bootstrap representan una opción atractiva si se desea caracterizar estadísticamente sistemas de tipo complejo, también se ha constatado que éstas herramientas se utilizan de manera creciente en términos aplicativos en PDS, en control, etc., aunque también en términos de investigación fundamental. Un problema interesante al cual aún se le busca solución posible es a la mejora de los métodos bootstrap cuando la ley de las variables aleatorias no es simétrica y para mejorar el desempeño de los estimadores de la media.



## A. Simulación por medio de Monte-Carlo

Los métodos de tipo Monte-Carlo<sup>7</sup> se han explotado bastante estos últimos años gracias a la revolución de la informática, y a la construcción de computadoras u ordenadores cada vez más potentes. El método de Monte-Carlo se define en general como “una técnica general utilizada para resolver problemas analíticamente complejos (determinísticos o estadísticos) haciendo uso de variables aleatorias o pseudo-aleatorias” [37], [23], [54] (algunos autores la califican como una definición poco concisa). La generación de variables aleatorias es sin lugar a duda un procedimiento bastante importante, puesto que la definición del método de Monte-Carlo descansa sobre ésta. La naturaleza de las variables aleatorias o estocásticas está definida según una ley de distribución. Las variables aleatorias que son las más simples de generar son aquellas que se obtienen aplicando una expresión algebraica determinista, y por consecuencia éstas son las más convenientes para las aplicaciones que son de nuestro interés. En efecto, es la misma computadora quien se encarga de generar variables aleatorias ayudándose de una expresión dada. El método de Monte-Carlo se aplica en diferentes dominios de la investigación (ingeniería, medicina, física, economía, etc.) para resolver problemas de:

1. *simulación*, la simulación implica la experimentación, y ésta puede ser definida como *la experimentación con un modelo varias veces*. Sin embargo, en lugar de experimentar con un objeto del mundo real (lo cual es rara vez posible, y cuando se puede el costo es probablemente bastante elevado), dicho objeto se substituye por un modelo que emula de manera aproximativa un sistema real. Entonces, el comportamiento del objeto de interés, será estudiado por medio de un análisis del modelo en el transcurso de varias experimentaciones. Aunque la definición de simulación no implica necesariamente el uso de variables aleatorias, en el contexto que nos interesa la simulation si hace uso de procesos estocásticos [37]. Efectivamente, cuando se pretende realizar experimentaciones con un modelo estocástico, éstas se obtienen gracias a un muestreo de variables aleatorias tiradas a partir de las distribuciones que modelan el sistema físico bajo estudio,
2. *tipo determinístico*, en este caso, se trata de resolver un problema planteado según una expresión formal similar a la expresión de un proceso estocástico y para la cual no existe una aproximación numérica sencilla (integración de una función  $h(z)$ , resolución de una ecuación de derivadas

---

<sup>7</sup>Fué Metropolis a quien se le ocurrió nombrar métodos de “Monte-Carlo” a los métodos de simulación estocástica. Lo anterior fué durante la realización del proyecto “Manhattan” de la segunda guerra mundial. Metropolis se inspiró de del interés de Ulam por el juego de “poker”, y de la similitud existente entre la simulación estocástica y los juegos de azar. Y como en aquella época la capital de Monaco éra el centro privilegiado de dichos juegos, fué que él seleccionó dicho nombre.

parciales, etc.),

3. *estimación de la distribución de un conjunto de muestras*, en muchos casos es necesario buscar una aproximación de la distribución de un conjunto de muestras o variables estocásticas observadas a la salida de un sistema (puede también tratarse de algunos parámetros estadísticos que pertenecen a la distribución de interés). Las variables estocásticas de salida se expresan en términos de una función conocida de una o varias de las variables aleatorias de entrada (en donde se supone que las distribuciones de dichas variables son conocidas).

La revolución de la informática a estado acompañada sin duda, por una evolución paralela del análisis estadístico. La disponibilidad de sistemas de cálculo poderosos ha permitido el desarrollo de nuevas teorías fundamentadas en el conocimiento empírico. Se acepta entonces, que la resolución analítica de problemas complejos no siempre puede llevarse a cabo con tanto rigor matemático. La resolución de dichos problemas por medio de la simulación estocástica permite obtener aproximaciones altamente precisas, las cuales justifican el uso de los métodos de tipo Monte-Carlo [14], [56], [23], [54].

### A.1. Solución de integrales haciendo uso de Monte-Carlo

En muchas de las aplicaciones en donde el método de Monte-Carlo es utilizado, el proceso físico se simula de manera directa, sin que exista la necesidad de deducir las ecuaciones diferenciales que describen el comportamiento físico del sistema. Sin embargo, el sistema (físico o matemático) debe estar bien descrito o especificado por medio de funciones de densidad de probabilidad. Una vez que las DDPs se han especificado bien de manera paramétrica, la simulación por Monte-Carlo se lleva a cabo muestreando a partir de dichas DDPs a cada iteración del proceso de simulación. Para poder terminar con el proceso de simulación, se debe obtener una población asintótica de realizaciones (en inglés “trials” o “histories”), y por lo tanto se debe efectuar un gran número de simulaciones o iteraciones ( $N \rightarrow \infty$ ). En lo que corresponde a la justificación de la convergencia de los métodos de simulación estocástica, ésta se establece gracias al teorema del límite central [19] y a la ley sobre los grandes números [54].

Podría pensarse de manera natural, que los métodos de Monte-Carlo no pueden ser utilizados para resolver problemas de tipo determinista, puesto que éstos son de naturaleza estocástica. Por ejemplo, la evaluación de una integral definida sobre un intervalo  $[a, b]$ , la inversión de un sistema lineal o no-lineal de ecuaciones no parecerían ser de naturaleza estocástica. Sin embargo, sí es posible calcular los problemas anteriores en términos de las DDPs. Consideremos el ejemplo sencillo de la evaluación de

una integral definida:

$$I = \int_a^b h(z) dz, \quad (80)$$

en donde se supone que la función  $h(z)$  es real y con valores en el intervalo  $[-\infty, \infty]$ . La figura 9 brinda una representación gráfica de la evaluación de la integral en un intervalo  $[a, b]$ . La evaluación de esta integral por medio del método de Monte-Carlo se lleva a cabo introduciendo una segunda función definida en el mismo intervalo  $[a, b]$ ,

$$u(z) = \begin{cases} \frac{1}{(b-a)} & \text{si } a \leq z \leq b \\ 0 & \text{de otra manera} \end{cases} \quad (81)$$

entonces, se obtiene la siguiente ecuación:

$$I = (b-a) \int_a^b h(z)u(z) dz \quad (82)$$

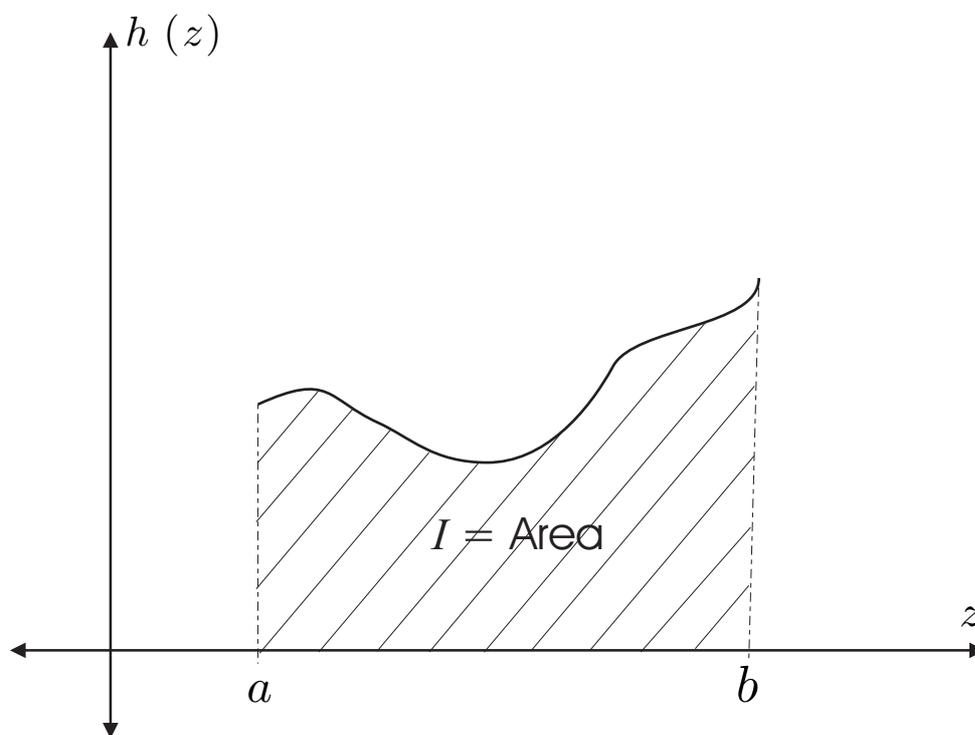


Figura 9: Integración por medio de Monte-Carlo.

Se hace notar que  $u(z)$  es una función equivalente a una ley uniforme en el intervalo  $[a, b]$ . Entonces, como  $u(z)$  es una DDP, la solución a la integral sobre un intervalo de  $[0, 1]$  está simplemente definida por la esperanza matemática de  $h(z)$ , dicho de otra forma,

$$I = E\{h(z)\} = \int_0^1 h(z)u(z) dz. \quad (83)$$

Para la aproximación es necesario tirar  $z_n$  muestras a partir de la DDP  $u(z)$ , y en cada tirada de  $z_n$  se lleva a cabo la evaluación de  $h(z_n)$  y entonces se obtiene una estimación del valor promedio  $\hat{h}$ , es decir,

$$\hat{h} = \frac{1}{N} \sum_{i=1}^N h(z_{in}), \quad (84)$$

y por lo tanto la integral definida sobre el intervalo  $[0, 1]$  vale:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N h(z_{in}) \quad (85)$$

De hecho, se trata de una estimación del valor *verdadero* de la integral  $I$ , efectuada a partir del valor promedio de  $N$  observaciones de la función  $h(z)$  en donde  $z$  fué muestreada uniformemente sobre el intervalo  $[0, 1]$  (intervalo finito). De la misma manera, los métodos de Monte-Carlo pueden ser extendidos a la evaluación de integrales de dimensión múltiple, en donde los grados de libertad necesarios para dicha evaluación y el tiempo de cálculo son tales que es posible pensar en el uso de dichos métodos para aplicaciones a problemas reales.

## A.2. Técnicas de reducción de variancia (TRV)

La versión más simple de los métodos de tipo Monte-Carlo, es el método de Monte-Carlo Primitivo (MCP) o en inglés “the Primitive Monte Carlo (PMC)”. Las variables estocásticas de entrada son generadas directamente o de forma cruda a partir de la DDP  $\varphi(z)$  (en el caso de la solución de la integral definida,  $F_z \triangleq u(z)$ ) en donde también se considera que las v.a. son independientes e idénticamente distribuidas (i.i.d.):

$$z \sim F_z.$$

Las técnicas de reducción de variancia (TRV) implican que el muestreo que se lleva a cabo de forma directa  $F_z$  (de manera primitiva) será reemplazado por un muestreo más sofisticado utilizando una DDP diferente,  $\varphi_g(z)$ . Las TRV intentan reducir la variancia de un estimador reemplazando el procedimiento original de muestreo por un procedimiento nuevo, el cual converge al mismo valor de la esperanza matemática,  $E\{h(z)\}$ , pero obteniendo una variancia más pequeña. De entre las técnicas TRV existentes, existen algunas que modifican sensiblemente el muestreo original. Aquí mencionamos algunas de las TRV más conocidas:

- (i) *muestreo ponderado* (o en inglés “Importance Sampling (IS)”) [36], [37], [54].
- (ii) *muestreo por estratos* (o en inglés “Stratified Sampling (SS)”) [37].

(iii) *muestreo por hipercubos Latinos* (o en inglés “Latin Hypercube Sampling (LHS)”) [36].

(iv) *muestreo por hipercubos Latinos modificado* (o en inglés “Modified Latin Hypercube Sampling (MHLS)”), este método es una variante del método LHS [36].

En fin, la meta principal de la simulación no se resume únicamente al cálculo del valor promedio, podría también ser utilizada para obtener, por ejemplo, la estimación de un percentil (la estimación de la probabilidad de que una respuesta es más grande que una cantidad fija). La eficiencia de una TRV se mide usualmente considerando dos aspectos:

(a) la reducción de la variancia del estimador de la esperanza matemática  $E\{h(z)\}$ , y

(b) la reducción del tiempo de cálculo de dicho estimador.

La optimalidad de las TRV es relativa al tiempo de cálculo, puesto que se debe considerar un tiempo de cálculo adicional para las nuevas integrales que se encuentran ligadas al proceso de muestreo cuando se utilizan conjuntamente las leyes  $F_z$  y  $\varphi_g(z)$ .

## B. Desarrollo de ecuaciones recursivas QP y NSHP

Para el desarrollo de las ecuaciones recursivas de los predictores propuestos por Asswad [2], se consideraron dos tipos de causalidades QP y NSHP. En el caso de la causalidad QP, se requiere obtener una ecuación recursiva según (74). La dificultad en dicha ecuación reside en que ciertos elementos de la matriz  $\mathbf{H}_{i,j}$  pueden depender directamente de  $\Theta$ . Entonces, para tomar en cuenta dicha dependencia después del cálculo del gradiente, se sugiere partir del punto  $(i, j)$  e interesarse por el valor predicho por (72)  $h(i, j) = \hat{y}(i, j)$  de la muestra  $y(i, j)$ , la cual se supone faltante o perdida.

La derivada de  $h(i, j)$  con respecto a  $\Theta$  está dada por la siguiente ecuación,

$$\frac{\partial h(i, j)}{\partial \Theta} = \begin{bmatrix} \frac{\partial h(i, j)}{\partial \theta(M-1, N-1)} & \frac{\partial h(i, j)}{\partial \theta(M-1, N-2)} & \cdots & \frac{\partial h(i, j)}{\partial \theta(M-1, 0)} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \frac{\partial h(i, j)}{\partial \theta(1, 1)} & \frac{\partial h(i, j)}{\partial \theta(1, 0)} \\ \frac{\partial h(i, j)}{\partial \theta(0, N-1)} & \cdots & \frac{\partial h(i, j)}{\partial \theta(0, 1)} & 0 \end{bmatrix} \quad (86)$$

en donde el valor cero del último elemento de la matriz se debe a que  $\theta(0, 0)$  no forma parte del soporte de predicción. La matriz, que depende de las muestras estimadas a partir de los parámetros del modelo, no permite establecer una expresión lineal analítica directa (como en el caso clásico de los estimadores

de MC (LMS)). Por el contrario, se puede obtener gracias a las relaciones recursivas que se desarrollan en seguida. La derivada con respecto a  $\theta(k, l)$  está dada por:

$$\begin{aligned} \frac{\partial h(i, j)}{\partial \theta(k, l)} &= \frac{\partial}{\partial \theta(k, l)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \theta(m, n) h(i-m, j-n) \\ &= h(i-k, j-l) + \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \theta(m, n) \frac{\partial h(i-m, j-n)}{\partial \theta(k, l)} \quad , \quad (m, n) \neq (k, l) \end{aligned} \quad (87)$$

De donde se deduce que la ecuación (86) se puede reescribir de la siguiente manera:

$$\frac{\partial h(i, j)}{\partial \Theta} = \mathbf{H}_{i,j} + d\mathbf{H}_{i,j} \quad (88)$$

en donde  $d\mathbf{H}_{i,j}$  es una matriz de dimensiones  $(M \times N)$  dada por:

$$d\mathbf{H}_{i,j} = \begin{bmatrix} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \theta(m, n) \frac{\partial h(i-m, j-n)}{\partial \theta(M-1, N-1)} & \cdots & \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \theta(m, n) \frac{\partial h(i-m, j-n)}{\partial \theta(M-1, 0)} \\ \vdots & \ddots & \vdots \\ \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \theta(m, n) \frac{\partial h(i-m, j-n)}{\partial \theta(0, N-1)} & \cdots & 0 \end{bmatrix} \quad (89)$$

Con la finalidad de establecer una relación de recurrencia se permite el cálculo del segundo término de la ecuación (88), para lo cual se introduce el cálculo siguiente:

$$\mathcal{H}_{i,j} = \begin{bmatrix} \frac{\partial h(i-M+1, j-N+1)}{\partial \Theta} & \cdots & \frac{\partial h(i-M+1, j)}{\partial \Theta} \\ \vdots & \ddots & \vdots \\ \frac{\partial h(i, j-N+1)}{\partial \Theta} & \cdots & \mathbf{0}(M, N) \end{bmatrix} \quad (90)$$

en donde  $\mathbf{0}(M, N)$  es una matriz de dimensiones  $(M, N)$ , para la cual todos los elementos valen cero.

La relación entre cada elemento de las matrices  $d\mathbf{H}_{i,j}$  y  $\mathcal{H}_{i,j}$  es la siguiente:

$$dH_{i,j}(k, l) = \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} \mathcal{H}_{i,j}(rM+k, sN+l) \theta(r, s) \quad (91)$$

El algoritmo obtenido considerando la causalidad QP se resume de la siguiente manera (algoritmo #1 QP):

$$\begin{aligned} h(i, j) &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \theta(m, n) h(i-m, j-n) \quad , \quad \theta(0, 0) = 0 \\ \mathbf{G}_{i,j} &= -\frac{1}{2} \frac{\partial J(i, j)}{\partial \Theta} = e(i, j) \frac{\partial h(i, j)}{\partial \Theta} \\ \frac{\partial h(i, j)}{\partial \Theta} &= \mathbf{H}_{i,j} + d\mathbf{H}_{i,j} \end{aligned}$$

en donde  $\mathbf{dH}_{i,j}$  se calcula a partir de  $\mathcal{H}_{i,j}$  la cual se obtiene por recurrencia. Si la muestra  $y(i, j + 1)$  es faltante:

$$\Theta_{i,j+1} = \Theta_{i,j}$$

Por el contrario, si la muestra  $y(i, j + 1)$  se encuentra disponible, se tiene que  $\partial h(i, j + 1)/\partial \Theta = 0$  y:

$$\Theta_{i,j+1} = \Theta_{i,j} + \mu \mathbf{G}_{i,j}$$

En un marco fuertemente no estacionario, la información inyectada por el término  $\mathbf{dH}_{i,j}$  en la ecuación (88) será errónea, y por lo tanto dicha información puede ser desechada. De lo anterior se obtiene el algoritmo simplificado (algoritmo #2 QP):

$$h(i, j) = \sum_{(m,n) \in S} \theta(m, n) h(i - m, j - n) \quad , \quad \theta(0, 0) = 0$$

$$\mathbf{G}_{i,j} = e(i, j) \mathbf{H}_{i,j}$$

$$\Theta_{i,j+1} = \Theta_{i,j} + \mu \mathbf{G}_{i,j}$$

Para el caso en el cual se considera la causalidad de tipo NSHP, éste puede ser visto como el caso de la causalidad QP excepto que el orden del modelo será  $(M, 2N - 1)$ , para las matrices  $\mathbf{H}_{i,j}$ ,  $\Theta$ , y  $\frac{\partial h(i, j)}{\partial \Theta}$ , los  $N$  últimos elementos de la última línea son cero. En el caso de la matriz  $\mathcal{H}_{i,j}$  (90), los  $N$  últimos bloques de la última línea son igualmente cero. Por consecuencia, las ecuaciones necesarias para el funcionamiento adecuado del algoritmo deben ser modificadas. Una versión simple del algoritmo puede obtenerse de la misma manera que en el caso de la causalidad QP.



## Referencias

- [1] Alexander S. T., y Rajala S. A., "Image Compression Results Using the LMS Adaptive Algorithm," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, Vol. ASSP-33, No. 3, pp. 712–714, June 1985.
- [2] Asswad K., *Échantillonnage et identification paramétriques adaptatifs pour des signaux bidimensionnels à échantillons manquants*, Tesis doctoral, Université de Paris XI-SUPELEC, Jun. 2002.
- [3] Athreya K. B., "Bootstrap of the mean in the infinite variance case," *The Annals of Statistics*, Vol. 15, No. 2, pp. 724–731, 1987.
- [4] Bates D. M., y Watts D. G., *Nonlinear regression analysis and its applications*, John Wiley and Sons, New York, 1988.
- [5] Beadle E. R., y Djurić P. M., "A fast weighted Bayesian Bootstrap filter for nonlinear model state estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 33, No. 1, pp. 338–343, Jan. 1997.
- [6] Bickel P. J., Götze F., y Van Zwet W. R., "Resampling fewer than  $n$  observations: gains, losses, and remedies for losses," *Empirical Bayes, sequential analysis and related topics in statistics and probability (New Brunswick, NJ, 1995)*, *Statistica Sinica*, Vol. 7, No. 1, pp. 1–31, 1997.
- [7] Box M. J., "Bias in nonlinear estimation," *Journal of the Royal Statistical Society, Series B*, Vol. 33, pp. 171–201, 1971.
- [8] Brahim-Belhouari S., *Choix de structures de modèles pertinentes pour un problème d'inversion*, Tesis doctoral, Université de Paris XI-SUPELEC, Sept. 2000.
- [9] Brown B. W., "Simulation variance reduction for Bootstrapping," *NSF Bootstrap Symposium*, Econometrics Laboratory, University of California, Berkeley 1996.  
  
URL : [//emlab.berkeley.edu/nsf96wp.html](http://emlab.berkeley.edu/nsf96wp.html)
- [10] Cho K., Meer P., y Cabrera J., "Performance assessment through Bootstrap," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 11, pp. 1185–1198, Nov. 1997.
- [11] Chung Y.S., y Kanefsky M., "On 2-D Recursive LMS Algorithms Using ARMA Prediction for ADPCM Encoding of Images," *IEEE Transactions on Image Processing*, Vol. 1, No. 3, pp. 416–421, July 1992.

- [12] Clarke G. P. Y., "Approximate confidence limits for a parameter function in nonlinear regression," *Journal of the American Statistical Association*, Vol 82, 397, pp. 221–230, 1987.
- [13] Clarke G. P. Y., "Moments of the least squares estimators in a non-linear regression model," *Journal of the Royal Statistical Society, Series B*, Vol. 42, No. 2, pp. 227–237, 1980.
- [14] Davison A. C., y Hinkley D. V., *Bootstrap methods and their application*, Cambridge University Press, (Cambridge UK), 1997-1999.
- [15] De la Rosa J. I., y Fleury G., "Bootstrap density approximation for measurement purposes," *Proc. of the IASTED Signal Processing and Communications*, pp. 357–361, Marbella (Spain), 19-22 Sept. 2000.
- [16] De la Rosa J. I., y Fleury G., "Bootstrap methods applied to indirect measurement," *Proc. of the Physics in Signal and Image Processing*, pp. 57–62, Marseille (France), 23-24 Jan. 2001.
- [17] De la Rosa J. I., y Fleury G., "Bootstrap methods for a measurement estimation problem," sometido a: *IEEE Transactions on Instrumentation and Measurement*, July 2001 (última revisión 28 / 08 / 2002).
- [18] De la Rosa J. I., *Estimation de la densité de probabilité dans un cadre non-linéaire, non-gaussien*, Tesis doctoral, Université de Paris XI-SUPELEC, Nov. 2002.
- [19] Demoment G., "Modélisation des incertitudes, inférence logique, et traitement des données expérimentales," *Bases de la théorie DEA - ATS*, Université Paris XI, Faculté des Sciences d'Orsay, 2000.
- [20] DiCiccio N., y Romano J. P., "A review of Bootstrap confidence intervals (with discussion)," *Journal of the Royal Statistical Society, Series B*, Vol. 50, No. 3, pp. 338–370, 1988.
- [21] Efron B., "Bootstrap methods: Another look at the Jackknife," *The Annals of Statistics*, Vol. 7, pp. 1–26, 1979.
- [22] Efron B., "Better Bootstrap confidence interval," *Journal of the American Statistical Association*, Vol. 82, pp. 171–185, 1987.
- [23] Fishman G. S., *Monte Carlo concepts, algorithms, and applications*, Springer Series in Operations Research, Springer-Verlag New York, 1995–1999.

- [24] Fleury G., "Non-intrusive time-of-flight flowmeter-parametric estimation and optimization," *Sensors and Actuators*, Vol. A 46–47, pp. 364–368, 1995.
- [25] Freedman D. A., "Bootstrapping regression models," *The Annals of Statistics*, Vol. 9, pp. 1218–1228, 1981.
- [26] Gibbs A. L., *Convergence of Markov Chain Monte Carlo algorithms with applications to image restoration*, Tesis doctoral, Department of Statistics, University of Toronto, 2000.  
URL : [www.utstat.toronto.edu](http://www.utstat.toronto.edu)
- [27] Hadhoud M. M., y Thomas D. W., "The Two-Dimensional Adaptive LMS (TDLMS) Algorithm," *IEEE Transactions on Circuits and Systems*, Vol. 35, No. 5, pp. 485–494, May 1988.
- [28] Hall P., y Presnell B., "Applications of intentionally biased Bootstrap methods," *Documenta Mathematica*, Extra Vol. ICM III, pp. 257–266, 1998.
- [29] Hall P., y Presnell B., "Intentionally biased Bootstrap methods," *Journal of the Royal Statistical Society*, Series B, Vol. 61, No. 1, pp. 143–158, 1999.
- [30] Hall P., y Presnell B., "Biased Bootstrap methods for reducing the effects of contamination," *Journal of the Royal Statistical Society*, Series B, Vol. 61, No. 3, pp. 661–680, 1999.
- [31] Hall P., Presnell B., y Turlach B. A., "Reducing bias without prejudicing sign," *Technical Report*, To appear in *The Annals of the Institute of Statistical Mathematics*.  
URL : [www.stat.ufl.edu/~presnell/](http://www.stat.ufl.edu/~presnell/)
- [32] Hall P., *The bootstrap and Edgeworth Expansion*, Springer Verlag, New York, 1992.
- [33] Hinkley D. V., "Bootstrap methods," *Journal of the Royal Statistical Society*, Series B, Vol. 50, No. 3, pp. 321–337, 1988.
- [34] Horowitz J. L., "Bootstrap methods in econometrics: theory and numerical performance," *NSF Bootstrap Symposium*, Econometrics Laboratory, University of California, Berkeley 1996.  
URL : [//emlab.berkeley.edu/nsf96wp.html](http://emlab.berkeley.edu/nsf96wp.html)
- [35] Huber P., *Robust statistics*, John Wiley, New York, 1981.
- [36] Keramat M., *Analyse statistique et optimisation du rendement de fabrication des circuits électroniques*, Tesis doctoral, Université de Paris XI-SUPELEC, Jan. 1998.

- [37] Kleijnen J. P. C., *Statistical techniques in simulation*, Statistics : Textbooks and Monographs, Marcel Dekker, 1974.
- [38] Kotropoulos C., y Pitas I., "Adaptive LMS L-Filters for Noise Suppression in Images," *IEEE Transactions on Image Processing*, Vol. 5, No. 12, pp. 1596–1609, December 1996.
- [39] Kukreja S. L., Kearney R.E., y Galiana H. L., "Structure detection of nonlinear Dynamic systems using Bootstrap methods," *Proc. of the IEEE Engineering in Medicine and Biology Society*, Vol. 20, No. 6, pp. 3020–3023, 1998.
- [40] Künsch H. R., "The jackknife and the bootstrap for general stationary observations," *The Annals of Statistics*, Vol. 17, No. 3, pp. 1217–1241, 1989.
- [41] Lahalle E., Fleury G., y Oksman J., "Real-Time ARMA Identification in the Case of Missing Observations," *IEEE International Symposium on Information Theory, ISIT 2000*, pp. 443, Sorrento, Italy 25-30 June 2000.
- [42] Lin J. N., Nie X., y Unbehauen R., "Two-Dimensional LMS Adaptive Filter Incorporating a Local-Mean Estimator for Image Processing," *IEEE Transactions on Circuits and Systems*, Vol. 40, No. 7, pp. 417-428, July 1993.
- [43] Makoto O., y Sumihisa H., "Two-Dimensional LMS Adaptive Filter", *IEEE Transactions on Consumer Electronics*, Vol. 37, No.1, February 1991, pp. 66–73.
- [44] Meyer M. E., y Gokhale D. V., "Kullback-Leibler information measure for studying convergence rates of densities and distributions," *IEEE Transactions on Information Theory*, Vol. 39, No. 4, pp. 1401–1404, July 1993.
- [45] Mirsaidi S., Fleury G., y Oksman J., "An ADPCM-like System Based on Non Uniform Signal Transmission," *1997 International workshop on sampling theory and applications*, Aveiro, pp. 139–143, 16-19 June 1997.
- [46] Mirsaidi S., Fleury G., y Oksman J., "LMS-like AR Modeling in the Case of Missing Observation," *IEEE Transactions on Signal Processing*, Vol. 45, No. 6, pp. 1–10, June 1997.
- [47] Naik-Nimbalkar U. V., y Rajarshi M. B., "Validity of blockwise Bootstrap for empirical processes with stationary observations," *The Annals of Statistics*, Vol. 22, No. 2, pp. 980–994, 1994.

- [48] Ong H.-T., y Zoubir A. M., “Robust signal detection using the Bootstrap,” *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1197–1200, Phoenix, March 1999.
- [49] Pázman A., “Probability distribution of the multivariate nonlinear least squares estimates,” *Ki-bernetika (Prague)*, 20, pp. 209–230, 1984.
- [50] Politis D. N., y Romano J. P., “A general resampling scheme for triangular arrays of  $\alpha$ -mixing random variables with application to the problem of spectral density estimation,” *The Annals of Statistics*, Vol. 20, No. 4, pp. 1985–2007, 1992.
- [51] Politis D. N., Romano J. P., y Lai T.-L., “Bootstrap confidence bands for spectra and cross-spectra,” *IEEE Transactions on Signal Processing*, Vol. 40, No. 5, pp. 1206–1215, May 1992.
- [52] Politis D. N., y Romano J. P., “Large sample confidence regions based on subsamples under minimal assumptions,” *The Annals of Statistics*, Vol. 22, No. 4, pp. 2031–2050, 1994.
- [53] Quenouille M., “Approximation tests of correlation in time series,” *Journal of the Royal Statistical Society, Series B*, Vol. 11, pp. 18–84, 1949.
- [54] Robert C. P., *Méthodes de Monte Carlo par Chaînes de Markov*, Statistique Mathématique et Probabilité, Economica, 1996.
- [55] Shao J., “Bootstrap model selection,” *Journal of the American Statistical Association*, Vol. 91, 434, pp. 655–665, 1996.
- [56] Shao J., y Tu D., *The jackknife and Bootstrap*, Springer Series in Statistics, Springer-Verlag, New York, 1996.
- [57] Shibata R., “Bootstrap estimate of Kullback-Leibler information for model selection,” *Statistica Sinica*, Vol. 7, No. 2, pp. 375–394, 1997.
- [58] Trosset M. W. y Sets B.Ñ., “On the choice of a discrepancy functional for model selection,” *Communications and Statistical Theory Methods*, Vol. 24, No. 11, pp. 2841–2863, 1995.
- [59] Tukey J., “Bias and confidence in not quite large samples,” *Annals of Mathematical Statistics*, Vol. 29, pp. 614, 1958.

- [60] Vermaak J., Niranjana M., y Godsill S. J., “An improved speech production model for voiced speech utilising a seasonal AR-AR model and MCMC simulation,” *Technical Report CUED/F-INFENG/TR.325*, Signal Processing Group, Cambridge University, June 1998.
- URL : [www.statslab.cam.ac.uk/~mcmc/pages/list.html](http://www.statslab.cam.ac.uk/~mcmc/pages/list.html)
- [61] Walter E., y Pronzato L., *Identification de modèles paramétriques à partir de données expérimentales*, Masson, Paris, 1994.
- [62] Wu C. F. J., “Jackknife, Bootstrap and other resampling methods in regression analysis,” *The Annals of Statistics*, 14, pp. 1261–1295, 1986.
- [63] Yang Z. R., Zwolinski M., y Chalk C. D., “Bootstrap, an alternative to Monte Carlo simulation,” *IEE Electronics Letters*, Vol. 34, No. 12, June 1998.
- [64] Youlal H., Janati-Idrissi M., y Najim M., *Modélisation paramétrique en traitement d’image*, Masson, 1994, Paris.
- [65] Zoubir A. M., “The Bootstrap and its application in signal processing,” *IEEE Signal Processing Magazine*, Vol. 15, No. 1, pp. 56–76, 1998.
- [66] Zoubir A. M., “The Bootstrap : A powerful tool for statistical signal processing with small sample sets,” *Proc. of the IEEE International Conference in Acoustics, Speech and Signal Processing (Tutorial)*, 15 March, 1999.
- URL : [www.atr.curtin.edu.au/csp](http://www.atr.curtin.edu.au/csp)
- [67] Zoubir A. M., “Identification of quadratic Volterra systems driven by non-Gaussian processes,” *IEEE Transactions on Signal Processing*, Vol. 43, No. 5, pp. 1302–1306, May 1995.
- [68] Zoubir A. M., “Model selection: a Bootstrap approach,” *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1377–1380, Phoenix, April 1999.
- [69] Zoubir A. M., y Böhme J. F., “Bootstrap multiple tests applied to sensor location,” *IEEE Transactions on Signal Processing*, Vol. 43, No. 6, pp. 1386–1396, 1995.
- [70] Zoubir A. M., y Iskander D. R., “Bootstrap model selection for polynomial phase signals,” *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 2229–2232, Seattle, May 1998.

- [71] Zoubir A. M., Ralston J. C., y Iskander D. R., "Optimal selection of model order for nonlinear system identification using the Bootstrap," *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 3945–3948, Munich, April 1997.