



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

# Modelos no lineales de regresión

Jose Antonio Cribeiro Ramallo

2019/2020

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



GRAO DE MATEMÁTICAS

**Traballo Fin de Grao**

# Modelos no lineales de regresión

Jose Antonio Crieiro Ramallo

2019/2020

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA





# Trabajo propuesto

**Área de Coñecemento: Estítstica e Investigación Operativa**

**Título: Modelos no lineales de regresión**

**Breve descripción do contido**

Los modelos de regresión estudian el efecto de ciertas variables explicativas sobre una variable de interés, también llamada variable respuesta. Los primeros modelos de regresión son lineales, en el sentido de que los parámetros desconocidos son coeficientes que multiplican a las variables explicativas o a funciones de ellas. Sin embargo, hay modelos muy conocidos de la Física, la Química o la Biología, como la ley de enfriamiento de Newton, los modelos de crecimiento o los modelos periódicos, en los cuales no es posible expresar los parámetros como coeficientes. En este trabajo se revisarán los métodos de inferencia en modelos no lineales de regresión. Mediante simulaciones, se estudiarán las propiedades de los estimadores bajo condiciones diversas, como normalidad o no de la distribución del error, homocedasticidad o heterocedasticidad, u otras circunstancias. Asimismo, se ilustrará el trabajo aplicando modelos no lineales de regresión en datos reales.

**Recomendacións**

**Outras observacións**

# Índice general

<b>Resumen</b>	<b>VII</b>
<b>Introducción</b>	<b>IX</b>
<b>1. Inferencia en modelos no lineales</b>	<b>1</b>
1.1. Presentación del modelo no lineal . . . . .	1
1.2. Estimadores del modelo . . . . .	2
1.2.1. Estimador de mínimos cuadrados . . . . .	2
1.2.2. La varianza del error . . . . .	3
1.3. Comportamiento asintótico de los estimadores . . . . .	3
1.4. Intervalos de confianza y contrastes basados en la distribución asintótica . .	14
1.5. Intervalos de confianza basadas en el perfil de RSS . . . . .	14
<b>2. Simulación</b>	<b>17</b>
2.1. Motivación y Objetivos . . . . .	17
2.2. Sesgo . . . . .	18
2.3. Varianza asintótica . . . . .	19
2.4. Intervalos de confianza . . . . .	20
2.5. Contrastos . . . . .	24
2.6. Distribuciones . . . . .	25
2.6.1. Efecto de la distribución del error . . . . .	28
2.7. Conclusión . . . . .	29
<b>Código de R empleado</b>	<b>31</b>
<b>Bibliografía</b>	<b>47</b>





## **Resumen**

En este trabajo se ha hecho una revisión de los modelos no lineales. Se han estudiado las propiedades de consistencia y de distribución asintótica de los estimadores, y se han presentado los procedimientos para la inferencia sobre los parámetros del modelo. Además, se ha hecho un estudio por simulación en donde se han verificado varias propiedades teóricas de los estimadores y de los métodos inferenciales expuestos en la teoría, como intervalos de confianza y contrastes de hipótesis, acompañado de una breve conclusión de los resultados obtenidos

## **Abstract**

In the following paper, we have revised the methods for the non linear regression models. We have studied the properties of consistency and the asymptotic distribution of the model estimator, on top of having presented a range of tools for the parameter inference. Furthermore, a simulation study has been taken where we have verified the multiple theoretical properties of the estimators and the inferential methods exposed previously, like the contrast statistics and confidence intervals, followed by a brief conclusion of the results that have been obtained.



# Introducción

En 1805 Legendre publicó un trabajo, que más adelante fue desarrollado más en profundidad por Gauss, que se conoce como la primera forma de regresión lineal documentada. Desde entonces el modelo lineal de regresión ha sido ampliamente estudiado y expandido, debido en gran parte a su gran variedad de aplicaciones y la fácil interpretación que gozan los resultados obtenidos a través de él.

Sin embargo, la regresión lineal no siempre es capaz de explicar todas las formas de regresión observables. Por ejemplo, en la Ecología el modelo Weibull intenta describir las curvas de crecimiento de una especie a través del tiempo mediante una función no lineal de sus parámetros:

$$m(x, (\theta_1, \theta_2, \theta_3, \theta_4)) = \theta_1 - (\theta_1 - \theta_2) \exp\left(-e^{\theta_3} x^{\theta_4}\right),$$

en donde  $m(x, \theta)$  la conocemos como la función de regresión (en este caso describe la curva de crecimiento de una especie a través del tiempo),  $x$  es la variable explicativa (en este caso corresponde al tiempo) y  $\theta_1, \dots, \theta_4$  son los coeficientes de este modelo (donde  $\theta_1$  es el tamaño máximo,  $\theta_2$  el tamaño mínimo, y  $\theta_3$  y  $\theta_4$  el ritmo y la forma de crecimiento, respectivamente). En la gráfica 1 podemos ver una representación del modelo para ciertos valores de los parámetros y de la variable explicativa.

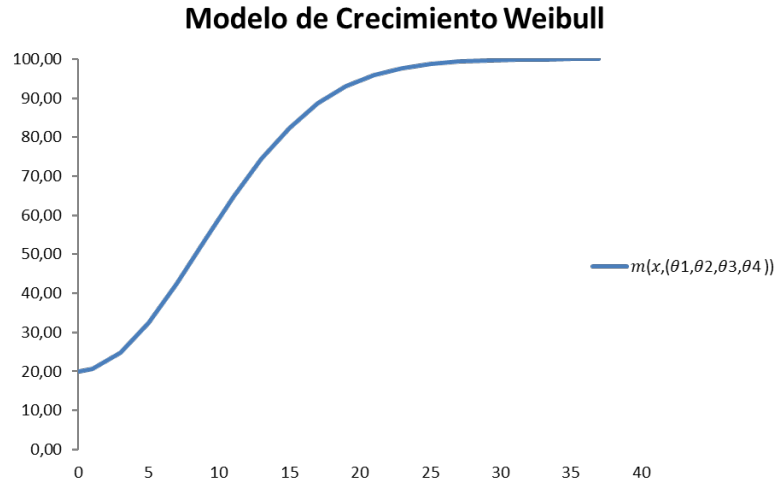
No solo en la Ecología ocurre esto, sino que muchos modelos químicos o físicos siguen ecuaciones no lineales. En el ámbito de la Química encontramos el modelo de Michaelis-Menten, el cual describe la velocidad inicial en una reacción enzimática, en función del sustrato:

$$m(x, (V_m, K)) = \frac{V_m x}{K + x},$$

siendo  $x$  la concentración del sustrato,  $V_m$  la velocidad máxima y  $K$  es la concentración para la cual se alcanza la mitad de la velocidad máxima. En la práctica  $K$  y  $V_m$  son desconocidos, siendo nuestro objetivo aproximarlos.

En la Física, tenemos un modelo muy estudiado, la Ley de enfriamiento de Newton:

$$m(t, \theta) = T_0 + (T - T_0)e^{-\theta t},$$

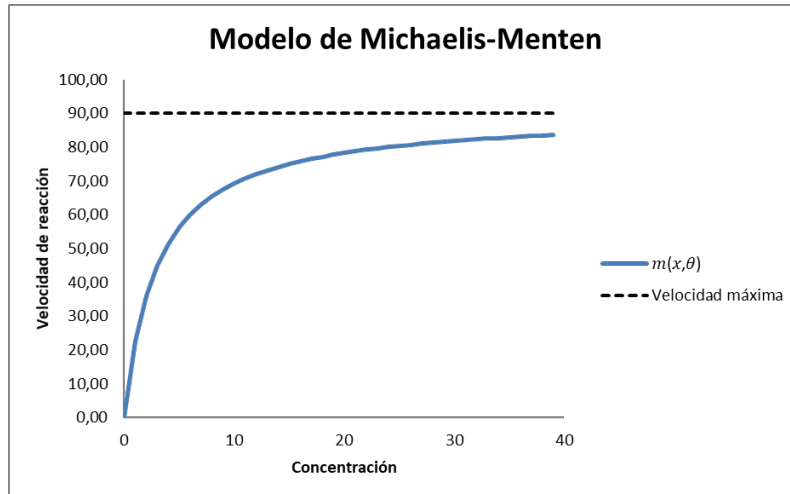


Gráfica 1: Modelo Weibull con  $\theta_1 = 100$ ,  $\theta_2 = 20$ ,  $\theta_3 = -5$  y  $\theta_4 = 2$ .

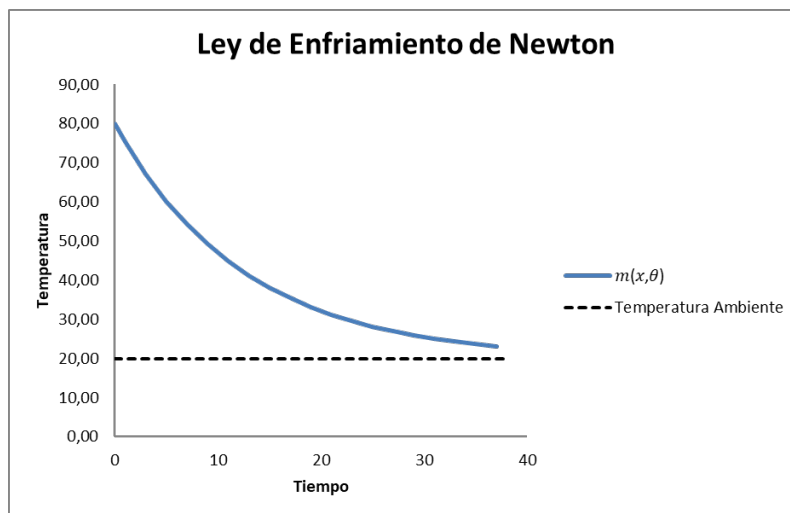
donde  $T_0$  es la temperatura ambiente,  $T$  la temperatura en el instante inicial,  $t$  el tiempo y  $\theta$ , que es desconocido, un parámetro que depende de varios factores (entre ellos la superficie de contacto del objeto y el calor específico del material). Nuestro interés en la práctica será estimar este parámetro  $\theta$ . En las gráficas 2 y 3 podemos ver una representación del modelo de Michaelis-Menten y de la Ley de Newton respectivamente.

Viendo estos ejemplos, es claro que existe un interés por desarrollar un maco teórico que nos permita modelar estos sucesos no lineales desde ya tiempos de Newton, sin embargo, debido a la no linealidad del modelo, el cálculo explícito del estimador no es posible, siendo necesario el uso de algún algoritmo iterativo. Fue el propio Gauss quien propició un algoritmo para el cálculo de mínimos cuadrados no lineales que aun hoy en día es relevante (algoritmo de Gauss-Newton). Gracias al avance de la computación, en estos tiempos es mucho más sencillo que antes hacer todos estos cálculos, permitiéndonos analizar modelos que en la época de Gauss no hubieran sido posibles de estudiar.

Nuestro objetivo será estudiar las propiedades teóricas de los estimadores y exponer una serie de métodos inferenciales para los modelos no lineales. A su vez, haremos una verificación por simulación de todas estas propiedades, en donde buscaremos replicar los resultados expuestos en la teoría.



Gráfica 2: Modelo de Michaelis-Menten con  $K = 3$ ,  $V = 90$  y  $x = 0, 1, \dots, 39$ .



Gráfica 3: Ley de enfriamiento de Newton con  $T = 80$ ,  $T_0 = 20$  y  $\theta = 0,08$ .



# Capítulo 1

## Inferencia en modelos no lineales

### 1.1. Presentación del modelo no lineal

Poseemos  $n$  observaciones de variables explicativas con sus respuestas

$$(x_1, y_1), \dots, (x_n, y_n);$$

tales que

$$y_i = m(x_i, \theta) + \varepsilon_i,$$

donde cada  $y_i$  corresponde a la respuesta del individuo  $i$ ,  $x_i$  corresponde al valor de la variable explicativa y  $\varepsilon_i$  denota al error. Sobre estos últimos tenemos que suponer que:

*i)* La media de  $\varepsilon_i$ ;  $\forall i = 1, \dots, n$  es 0:

$$\mathbb{E}[\varepsilon_i] = 0$$

*ii)* La varianza de  $\varepsilon_i$ ;  $\forall i = 1, \dots, n$  es  $\sigma^2$

$$\text{Var}(\varepsilon_i) = \sigma^2$$

*iii)* El vector  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$  es un vector de variables aleatorias independientes.

*Observación 1.1.* Nótese que, a diferencia de los modelos lineales, el error no tiene que ser necesariamente normal. Más adelante podremos ver la demostración formal de por qué no es necesario suponer normalidad para los procesos de inferencia en modelos no lineales. Una idea intuitiva sería suponer que mientras  $n$ , el tamaño de la muestra, se hace cada vez

más grande, el estimador converge a su valor verdadero y las tangentes se vuelven cada vez mejor aproximación local. Es por esto que el estimador tendrá un comportamiento cada vez más simétrico, y cada vez más parecido al de un estimador lineal dado por la tangente, tendiendo así a un comportamiento normal.

El propósito será estimar el vector de parámetros  $\theta \in \mathbb{R}^p$  y la varianza del error  $\sigma^2$  a partir de una muestra, de las variables explicativas y la respuesta, de tamaño  $n$ .

## 1.2. Estimadores del modelo

### 1.2.1. Estimador de mínimos cuadrados

Tal y como hacíamos en la regresión lineal, será conveniente tomar el estimador que nos garantice la menor suma residual de cuadrados:

$$RSS_n(\theta) = \sum_{i=1}^n (y_i - m(x_i, \theta))^2 \quad (1.1)$$

Vamos a denotar a este estimador por  $\hat{\theta}$ , tal que:

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} RSS_n(\theta)$$

es decir,  $\hat{\theta}_n$  sería el  $\theta \in \Theta$  tal que la derivada de (1.1) es cero:

$$\sum_{i=1}^n (y_i - m(x_i, \theta)) \frac{\partial m}{\partial \theta}(x_i, \theta) = 0. \quad (1.2)$$

*Observación 1.2.* Si tomamos  $\theta \in \mathbb{R}^p$ , la derivada de (1.1) sería un vector gradiente.

Al sistema de  $p$  ecuaciones y  $p$  incógnitas (1.2) lo vamos a conocer como las ecuaciones normales de regresión (al igual que en el caso lineal.)

A diferencia que en el caso lineal, las ecuaciones normales de regresión son, en este caso, no lineales, por lo que no vamos a poder obtener una expresión explícita del estimador. Por lo tanto, la única forma de obtener el estimador de mínimos cuadrados será aproximándolo mediante algún algoritmo iterativo.

*Observación 1.3.* Recordemos que, en el modelo lineal, que la matriz  $X'X$  tuviera autovalores pequeños generaba problemas al invertir la matriz, suponiendo un incremento en la varianza de los estimadores (lo llamábamos problema de colinealidad).

En el caso no lineal volvemos a tener este problema, y es que la inversión de la matriz hessiana  $\frac{\partial^2 RSS}{\partial^2 \theta}$  (simétrica y semidefinida positiva) juega un papel crucial en algunos procesos



iterativos (como el método de Newton). Por lo que si  $\frac{\partial^2 RSS}{\partial^2 \theta}$  presenta autovalores pequeños, esta inversión se dificulta; llegando a, incluso, hacer que el algoritmo no converja.

El problema de la convergencia se puede resolver seleccionando correctamente los valores iniciales, utilizando métodos iterativos que tratan de paliar este problema (como el de Levenberg-Marquardt) o incluso tratar con otros métodos en donde no se emplean las derivadas.

Las consecuencias sobre la varianza de los estimadores solo se pueden paliar modificando el modelo.

### 1.2.2. La varianza del error

De forma idéntica al modelo lineal, nos encontramos con que

$$\hat{\sigma}^2 = \frac{RSS(\hat{\theta}_n)}{n-p} = \frac{\sum_{i=1}^n (y_i - m(x_i, \hat{\theta}_n))^2}{n-p},$$

es un estimador de la varianza del error que asintóticamente es una ji-cuadrado de  $n-p$  grados de libertad (lo veremos más adelante).

## 1.3. Comportamiento asintótico de los estimadores

Es importante destacar que no hemos supuesto al principio que los errores del modelo tengan que ser normales para demostrar la normalidad asintótica de los estimadores en el modelo no lineal.

Esto no es algo único del modelo no lineal, si no que también ocurre en la regresión lineal. Eicker (1963) y Grenander y Rosenblatt (1957) dieron condiciones sobre la función de regresión para la normalidad asintótica cuando los errores no son idénticamente distribuidos y condiciones para la eficiencia asintótica.

Nuestro objetivo va a ser demostrar la normalidad asintótica de los estimadores de mínimos cuadrados bajo ciertas hipótesis. La primera de ellas será:

### Hipótesis 1.4.

$$y_i = m(x_i, \theta_0) + \varepsilon_i = f_i(\theta_0) + \varepsilon_i, \quad i = 1, 2, \dots;$$

con  $f_i$  funciones continuas conocidas en el subconjunto compacto  $\Theta$  de un espacio Euclideo y  $\varepsilon_i$  errores iid de media 0 y varianza  $\sigma^2 > 0$  ( $\theta_0$  y  $\sigma$  son desconocidos).

*Observación 1.5.* Durante el resto de esta sección vamos a considerar  $\Theta$  como en la hipótesis 1.4, un subconjunto compacto de un espacio Euclideo.

*Observación 1.6.* Recordemos que estamos trabajando con una única variable explicativa y una única respuesta. Gallant[2] habla sobre el caso multivariante en su artículo, en donde establece una analogía entre el modelo no lineal multivariante asintótico con el lineal multivariante.

Vamos a llamar a cualquier vector  $\hat{\theta} \in \Theta$  que minimice

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - f_i(\theta))^2$$

como estimador de mínimos cuadrados de  $\theta_0$  basado en los  $n$  primeros valores de  $y$ .

Nuestro objetivo es obtener la distribución límite del estimador de los parámetros  $\hat{\theta}$ , para ello seguiremos la demostración dada por Jenrich [4].

Empezamos considerando una serie de conceptos que vamos a necesitar a lo largo de la demostración.

**Definición 1.7.** Sean  $x = (x_i)$  e  $y = (y_i)$  dos sucesiones de números reales, y sea

$$\langle x, y \rangle_n = n^{-1} \sum_{i=1}^n x_i y_i$$

el producto interior  $n$ -ésimo de  $x$  e  $y$ .

Si  $\langle x, y \rangle_n$  converge cuando  $n \rightarrow \infty$  a un número  $\langle x, y \rangle$ , vamos a conocer a este límite como el producto de colas (*tail product*) de  $x$  e  $y$ .

**Definición 1.8.** Sean  $g = (g_i)$ ,  $h = (h_i)$  sucesiones de funciones de  $\Theta$ , tal que  $g(\theta) = (g_i(\theta))$ ,  $h(\theta) = (h_i(\theta))$  (es decir,  $g$  y  $h$  las podemos considerar como funciones de  $\Theta$  al espacio de sucesiones de elementos de  $\Theta$ ,  $\Theta^{\mathbb{N}}$ ).

Consideremos que  $\langle g(\alpha), h(\beta) \rangle_n \rightarrow \langle g(\alpha), h(\beta) \rangle$  uniformemente  $\forall \alpha, \beta \in \Theta$ . Llamaremos producto cruzado de colas (*tail cross product*) de  $g$  y  $h$  a la función:

$$[g, h] : (\alpha, \beta) \in \Theta \times \Theta \mapsto \langle g(\alpha), h(\beta) \rangle \in \mathbb{R}$$

Nótese que si  $g$  y  $h$  son funciones continuas, entonces el producto cruzado de colas de ambas es una función continua.

*Observación 1.9.* Denotaremos a  $\langle x, x \rangle_n$  como:

$$\|x\|_n^2 = \langle x, x \rangle_n$$

y a su límite como:

$$\|x\|^2 = \langle x, x \rangle,$$

que lo llamaremos la norma de cola (*tail norm*) de  $x$ .

Notese que si  $[g, g]$ ,  $[h, h]$ , y  $[g, h]$  existen, entonces:

$$\|g(\alpha) - h(\beta)\|_n \longrightarrow \|g(\alpha) - h(\beta)\|$$

uniformemente para todo  $\alpha, \beta \in \Theta$ .

*Observación 1.10.* Se deduce facilmente que  $\|a + b\|_n^2 = \|a\|_n^2 + 2\langle a, b \rangle_n + \|b\|_n^2$

Consideremos ahora dos teoremas que serán útiles a la hora de probar la existencia de  $[g, h]$  para dos funciones acotadas y continuas en  $\mathbb{X} \times \Theta$ , con  $\mathbb{X}$  un espacio Euclidiano.

**Teorema 1.11.** Sean  $\mathbb{X}$  un espacio Euclidiano y  $g$  una función continua y acotada en  $\mathbb{X} \times \Theta$ .

Sea  $(F_n)$  una sucesión de distribuciones en  $\mathbb{X}$  que convergen a una distribución  $F$

Entonces:

$$\int g(x, \theta) dF_n(x) \longrightarrow \int g(x, \theta) dF(x).$$

**Teorema 1.12.** Sea  $g$  una función en  $\mathbb{X} \times \Theta$  con  $\mathbb{X}$  un espacio Euclidiano.

Sea  $g(x, \theta)$  una función continua de  $\theta$  para cada  $x$  y una función medible de  $x$  para cada  $\theta$ , tal que:

$$|g(x, \theta)| \leq h(x); \text{ para todo } x, \theta.$$

con  $h$  integrable respecto a  $F$ , una distribución en  $\mathbb{X}$ .

Si  $x_1, x_2, \dots$  son una muestra aleatoria de  $F$ , entonces, para casi toda sucesión  $(x_n)$  de muestras:

$$n^{-1} \sum_{i=1}^n g(x_i, \theta) \longrightarrow \int g(x, \theta) dF(x)$$

uniformemente, para todo  $\theta \in \Theta$ .

*Observación 1.13.* La demostración de estos teoremas se encuentra en Jenrich [4, p. 635-636]

De las definiciones anteriores, extraemos un resultado que vamos a necesitar más adelante:

Sea  $\varepsilon = (\varepsilon_i)$ , con  $\varepsilon_i$  bajo la hipótesis 1.4.

**Teorema 1.14.** Si existe el producto de cola de una sucesión  $x$  de números reales se verifica que:

$$\langle x, \varepsilon \rangle_n \longrightarrow 0,$$

para casi todo  $\varepsilon$ .

Este resultado se puede extender como:

**Teorema 1.15.** *Bajo la hipótesis 1.4, si  $g$  es una sucesión de funciones de  $\Theta$  y si  $[g, g]$  existe, tenemos que, para casi todo  $\varepsilon$ :*

$$\langle g(\theta), \varepsilon \rangle_n \longrightarrow 0$$

*uniformemente para todo  $\theta \in \Theta$ .*

*Demostración.* Como  $g$  es continua y  $\|g(\alpha) - g(\beta)\|_n$  converge uniformemente en  $\alpha$ ; existe, para todo  $\delta > 0$  y para todo  $\beta \in \Theta$ , un entorno  $N$  de  $\beta$  tal que:

$$\|g(\alpha) - g(\beta)\|_n < \delta,$$

para todo  $\alpha \in N$  y para todo  $n \in \mathbb{N}$  suficientemente grande.

Luego, usando el Teorema anterior en la inecuación:

$$|\langle g(\alpha), \varepsilon \rangle_n| \leq \|g(\alpha) - g(\beta)\|_n \|\varepsilon\|_n + |\langle g(\beta), \varepsilon \rangle_n|,$$

obtenida a partir de la inecuación de Cauchy, deducimos que para casi todo  $\varepsilon$ ,  $\Theta$  es recubierto por entornos  $N$  tales que  $|\langle g(\theta), \varepsilon \rangle_n| < \delta$  para todo  $\theta \in N$ , cuando  $n \in N$  es lo suficientemente grande.

De esta forma, como  $\Theta$  es compacto, hay una colección finita de estos entornos que recubren a  $N$ , por lo que para casi todo  $\varepsilon$ :

$$|\langle g(\theta), \varepsilon \rangle_n| < \delta,$$

para todo  $\theta \in \Theta$  con  $n$  lo suficientemente grande.

Es decir,

$$|\langle g(\theta), \varepsilon \rangle_n - 0| < \delta \implies \langle g(\theta), \varepsilon \rangle_n \longrightarrow 0.$$

□

Vamos a establecer ahora una forma del Teorema Central del Límite que emplea las nuevas definiciones que hemos introducido y nos será de utilidad más adelante:

**Teorema 1.16.** *Sea  $\varepsilon$  según la hipótesis 1.4 y supongamos que la norma de cola de  $x$ , una sucesión de números reales, existe. Entonces:*

$$\mathcal{L}(\sqrt{n} \langle x, \varepsilon \rangle_n) \longrightarrow N(0, \|x\|^2 \sigma^2)$$

Este resultado puede ser extendido al caso multilinear como:

**Corolario 1.17.** *Sea  $\varepsilon$  bajo las mismas hipótesis que en el Teorema anterior.*

*Si existe cada producto de cola posible entre las sucesiones  $x_1, \dots, x_p$  de números reales, entonces, la distribución conjunta de:*

$$\sqrt{n} (\langle x_1, \varepsilon \rangle_n, \dots, \langle x_p, \varepsilon \rangle_n)$$

*converge a*

$$N(0, \sigma^2 X),$$

*siendo  $X$  la matriz tal que el elemento  $(i, j)$ -ésimo es  $\langle x_i, x_j \rangle$ .*

El siguiente resultado es de vital importancia, ya que nos garantiza que la hipótesis 1.4 es suficiente para demostrar la existencia de estimadores de mínimos cuadrados medibles en el caso no lineal.

**Lema 1.18.** *Sea  $Q$  una función real en  $\Theta \times \mathbb{Y}$ , con  $\mathbb{Y}$  un espacio medible. Para cada  $\theta \in \Theta$ , sea  $Q(\theta, y)$  una función medible de  $y$ , y para cada  $y \in \mathbb{Y}$  una función continua de  $\theta$ .*

*Entonces, existe*

$$\hat{\theta} : \mathbb{Y} \longrightarrow \Theta$$

*medible, tal que, para todo  $y \in \mathbb{Y}$ :*

$$Q(\hat{\theta}(y), y) = \inf_{\theta \in \Theta} Q(\theta, y)$$

*Demostración.* Sea  $(\Theta_n)$  una sucesión creciente de subconjuntos finitos de  $\Theta$ , cuyo límite es denso en  $\Theta$ .

Por el Th. de Weierstrass, para cada  $n$  hay una función medible,

$$\tilde{\theta}_n : \mathbb{Y} \longrightarrow \Theta_n$$

tal que:

$$Q(\tilde{\theta}_n(y), y) = \inf_{\theta \in \Theta_n} Q(\theta, y)$$

Sea  $\tilde{\theta}_{n_1}$  la primera componente de  $\tilde{\theta}_n$  y  $\hat{\theta}_1 = \lim_{n \rightarrow \infty} \tilde{\theta}_n$  su límite (nótese que  $\hat{\theta}_1$  es medible por ser límite de medibles).

Para cada  $y \in \mathbb{Y}$  hay una subsucesión  $(\tilde{\theta}_{n_i}(y))$  de  $(\tilde{\theta}_n(y))$ , que converge a un punto  $\bar{\theta} = (\hat{\theta}_1(y), \bar{\theta}_2, \dots, \bar{\theta}_p)$ .

Ahora, como  $\lim_{n \rightarrow \infty} \Theta_n$  es denso en  $\Theta$ :

$$\inf_{(\theta_1, \dots, \theta_n) \in \Theta} Q \left( \left( \hat{\theta}_1(y), \theta_2, \dots, \theta_p \right), y \right) \leq Q(\bar{\theta}, y) \quad (1.3)$$

$$= \lim_{i \rightarrow \infty} Q \left( \tilde{\theta}_{n_i}(y), y \right) \quad (1.4)$$

$$= \lim_{i \rightarrow \infty} \inf_{\theta \in \Theta_{n_i}} Q(\theta, y) \quad (1.5)$$

$$= \inf_{\theta \in \Theta} Q(\theta, y). \quad (1.6)$$

Por lo tanto:

$$\inf_{(\theta_1, \dots, \theta_n) \in \Theta} Q \left( \left( \hat{\theta}_1(y), \theta_2, \dots, \theta_p \right), y \right) = \inf_{\theta \in \Theta} Q(\theta, y); \quad \forall y \in \mathbb{Y}.$$

Sea ahora,

$$Q'((\theta_1, \dots, \theta_p), y) = Q \left( \left( \hat{\theta}_1(y), \theta_2, \dots, \theta_p \right), y \right),$$

luego,  $Q'(\theta, y)$  es una función continua de  $\theta \forall y \in \mathbb{Y}$ , y una función medible de  $y \forall \theta \in \Theta$ .

Con el mismo razonamiento que hicimos para  $Q$ , encontramos que

$$\inf_{(\theta_1, \dots, \theta_n) \in \Theta} Q \left( \left( \hat{\theta}_1(y), \hat{\theta}_2(y), \theta_3, \dots, \theta_p \right), y \right) = \inf_{\theta \in \Theta} Q(\theta, y); \quad \forall y \in \mathbb{Y}.$$

De esta forma, repitiendo esto encontramos las funciones medibles de  $\mathbb{Y}$  a  $\mathbb{R}$ :  $\hat{\theta}_1(y), \dots, \hat{\theta}_p(y)$ ; tales que:

$$Q \left( \left( \hat{\theta}_1(y), \dots, \hat{\theta}_p(y) \right), y \right) = \inf_{\theta \in \Theta} Q(\theta, y); \quad \forall y \in \mathbb{Y}.$$

□

El Lema 1.18 junto con el Th. del valor medio nos otorga el siguiente resultado, que no es más que una versión del Th. del valor medio para funciones medibles.

**Proposición 1.19.** *Sea  $f : \mathbb{Y} \times \Theta \rightarrow \mathbb{R}$ , con  $\mathbb{Y}$  un espacio medible y  $\Theta$  convexo.*

*Para cada  $\theta \in \Theta$ ,  $f(y, \theta)$  es una función medible de  $y$  y para cada  $y \in \mathbb{Y}$  una función continuamente diferenciable de  $\theta$ .*

*Sean  $\theta_1, \theta_2$  funciones medibles de  $\mathbb{Y}$  a  $\Theta$ .*

*Entonces, existe una función  $\bar{\theta} : \mathbb{Y} \rightarrow \Theta$  medible, tal que:*

$$i) \quad f(y, \theta_1(y)) - f(y, \theta_2(y)) = \frac{\partial}{\partial \theta} f(y, \bar{\theta}(y)) (\theta_1(y) - \theta_2(y)).$$

$$ii) \quad \hat{\theta}(y) \text{ está en el segmento que une } \theta_1(y) \text{ con } \theta_2(y), \text{ para todo } y \in \mathbb{Y}.$$

*Demostración.* Sea  $D(y, \theta)$  la distancia de  $\theta$  al segmento que une  $\theta_1(y)$  y  $\theta_2(y)$ , y sea

$$P(\theta, y) = |f(y, \theta_1(y)) - f(y, \theta_2(y)) - \frac{\partial}{\partial \theta} f(y, \theta) (\theta_1(y) - \theta_2(y))| + D(y, \theta).$$

Si vemos a  $\theta_1$  y  $\theta_2$  como los límites puntuales de una sucesión de funciones simples y a  $\frac{\partial}{\partial \theta} f(y, \theta)$  como:

$$\frac{\partial}{\partial \theta} f(y, \theta) = \lim_{h \rightarrow 0} \frac{f(y, \theta + h) - f(y, \theta)}{h},$$

es claro que  $P$  es una función medible de  $y$  para todo  $\theta \in \Theta$ , y una función continua de  $\theta$  para todo  $y \in \mathbb{Y}$ . Por esto último, en virtud del Lema 1.18, encontramos que  $\bar{\theta} : \mathbb{Y} \rightarrow \Theta$  medible, tal que  $\bar{\theta}(y)$  minimiza a  $P(\theta, y)$  para todo  $y \in \mathbb{Y}$  como función de  $\theta$ .

Además, por el Th de valor de medio, cuando  $\theta = \bar{\theta}(y)$  el primer y el segundo término de  $P(\theta, y)$  son 0, luego:

$$i) |f(y, \theta_1(y)) - f(y, \theta_2(y)) - \frac{\partial}{\partial \theta} f(y, \bar{\theta}(y)) (\theta_1(y) - \theta_2(y))| = 0 \implies f(y, \theta_1(y)) - f(y, \theta_2(y)) = \frac{\partial}{\partial \theta} f(y, \bar{\theta}(y)) (\theta_1(y) - \theta_2(y))$$

$$ii) D(y, \bar{\theta}(y)) = 0 \implies \bar{\theta}(y) \in L(\theta_1(y), \theta_2(y))$$

□

De ahora en adelante vamos a necesitar considerar una nueva condición sobre las  $f_i$  de la hipótesis 1.4:

**Hipótesis 1.20.**

- $\exists [f, f]$ , con  $f = (f_i) : \Theta \longrightarrow \Theta^{\mathbb{N}}$
- La función  $Q(\theta) = \|f(\theta) - f(\theta_0)\|^2$  tiene un único mínimo en  $\theta = \theta_0$

*Observación 1.21.* Las hipótesis 1.4 y 1.20 son suficientes para probar la continuidad de  $Q(\theta)$

*Observación 1.22.* Notese que la  $Q_n$  que definimos al principio coincide con

$$Q_n(\theta) = \|f(\theta_0) + \varepsilon - f(\theta)\|_n^2$$

Esta hipótesis es necesaria para demostrar una propiedad muy importante de los estimadores: la consistencia.

**Teorema 1.23.** Sea  $(\hat{\theta}_n)$  una sucesión de EMC.

Bajo las hipótesis 1.4 y 1.20,  $\hat{\theta}_n$  y  $\hat{\sigma}_n^2 = Q_n(\hat{\theta}_n)$  son estimadores fuertemente consistentes.

*Demostración.* Consideremos primero que  $\|\varepsilon\| = \sigma$  y que:

$$\langle f(\theta) - f(\theta_0), \varepsilon \rangle_n \longrightarrow 0$$

uniformemente, para todo  $\theta \in \Theta$ , gracias a la Ley fuerte de los grandes números y al Th. 1.15 respectivamente.

$Q_n(\theta)$  se puede escribir como:

$$\begin{aligned} Q_n(\theta) &= \|f(\theta_0) + \varepsilon - f(\theta)\|_n^2 \\ &= \|f(\theta_0) - f(\theta)\|_n^2 + 2 \langle f(\theta_0) - f(\theta), \varepsilon \rangle_n + \|\varepsilon\|_n^2, \end{aligned}$$

por lo tanto, encontramos que:

$$Q_n(\theta) \longrightarrow \|f(\theta_0) - f(\theta)\|^2 + \sigma^2 = Q(\theta) + \sigma^2, \text{ uniformemente para todo } \theta \in \Theta.$$

Sean  $\theta_n = \hat{\theta}_n(f(\theta_0) + \varepsilon)$ ,  $\theta'$  el límite de la sucesión  $(\theta_n)$  y  $(\theta_{n_i})$  una subsucesión cualquiera que converge a  $\theta'$ .

Por la continuidad de  $Q$  y la convergencia uniforme de  $Q_n$  a  $Q + \sigma^2$ ,

$$Q_{n_i}(\theta_{n_i}) \xrightarrow{i \rightarrow \infty} Q(\theta') + \sigma^2$$

Como  $\sigma_{n_i}$  es un EMC,

$$Q_{n_i}(\theta_{n_i}) \leq Q_{n_i}(\theta_0) := \|\varepsilon\|_{n_i}^2.$$

Si  $i \rightarrow \infty$ , encontramos que:

$$Q(\theta') + \sigma^2 \leq \sigma^2 \implies Q(\theta') = 0$$

$$\iff \theta' = \theta_0 \text{ (esto lo deducimos ya que el mínimo de } Q \text{ es único y } Q(\theta) \geq 0).$$

En conclusión, lo que tenemos es que:

$$\theta_n \longrightarrow \theta_0, \text{ para casi todo } \varepsilon \implies \hat{\theta} \xrightarrow{cs} \theta_0 \text{ y } \hat{\sigma}_n^2 = Q_n(\hat{\sigma}_n^2) \xrightarrow{cs} Q(\theta_0) + \sigma^2 = \sigma^2.$$

□

*Observación 1.24.* Consideremos la notación:

$$f'_{i_j}(\theta) = \frac{\partial}{\partial \theta_j} f_i(\theta); \quad f''_{i_{jk}}(\theta) = \frac{\partial^2}{\partial \theta_j \partial \theta'_k} f_i(\theta),$$

de tal forma que:

$$f'_j = (f'_{i_j}); \quad f''_{jk} = (f''_{i_{jk}}).$$

Además:

$$[a_n(\theta)]_{j_1, j_2} = \langle f'_{j_1}(\theta), f'_{j_2}(\theta) \rangle_n,$$



tal que  $[a_n(\theta)]_{j_1, j_2}$  es la componente  $(j_1, j_2)$ -ésima de la matriz  $a_n(\theta)$ :

$$a_n(\theta) = \begin{pmatrix} [a_n(\theta)]_{1,1} & \cdots & [a_n(\theta)]_{1,p} \\ \vdots & & \vdots \\ [a_n(\theta)]_{p,1} & \cdots & [a_n(\theta)]_{p,p} \end{pmatrix}$$

Consideremos también:

$$[a(\theta)]_{j_1, j_2} = \langle f'_{j_1}(\theta), f'_{j_2}(\theta) \rangle$$

en donde  $[a(\theta)]_{j_1, j_2}$  es la componente  $(j_1, j_2)$ -ésima de la matriz  $a(\theta)$ :

$$a(\theta) = \begin{pmatrix} [a(\theta)]_{1,1} & \cdots & [a(\theta)]_{1,p} \\ \vdots & & \vdots \\ [a(\theta)]_{p,1} & \cdots & [a(\theta)]_{p,p} \end{pmatrix},$$

y el producto de colas de  $f'_{j_1}(\theta)$  y  $f'_{j_2}(\theta)$ .

Vamos a considerar ahora las dos últimas hipótesis que necesitamos para poder demostrar la normalidad asintótica de los estimadores.

Sea  $f'_{j_1} = (f'_{i_{j_1}})$  y  $f''_{j_1 j_2} = (f''_{i_{j_1} j_2})$  sucesiones de funciones, tales que:  $f'_{j_1}, f''_{j_1 j_2} : \Theta \rightarrow \Theta^{\mathbb{N}}$ .

**Hipótesis 1.25.**  $f'_{j_1}, f''_{j_1 j_2}$  existen, son continuas en  $\Theta$  y todos los productos cruzados de colas posibles de  $f, f'_{j_1}, f''_{j_1 j_2}$  existen.

**Hipótesis 1.26.** El vector de los verdaderos parámetros,  $\theta_0$ , es un punto interior de  $\Theta$  y  $a(\theta_0)$  es no-singular.

**Teorema 1.27.** Sea  $(\hat{\theta}_n)$  una sucesión de EMC de  $\theta_0$ . Bajo las hipótesis 1.4, 1.20, 1.25 y 1.26;

$$\mathcal{L} \left( \sqrt{n} (\hat{\theta}_n - \theta_0) \right) \rightarrow N(0, \sigma^2 a^{-1}(\theta)).$$

Además,  $a_n(\hat{\theta}_n)$  es un estimador fuertemente consistente de  $a(\theta_0)$ .

*Demostración.* Ya que  $\hat{\theta}_n \rightarrow \theta_0$  en casi todo punto, hay una sucesión  $(\tilde{\theta}_n)$  que es *tail equivalent* a  $(\hat{\theta}_n)$  (es decir,  $\|(\tilde{\theta}_n)\|^2 = \|(\hat{\theta}_n)\|^2$ ) tal que cada  $\tilde{\theta}_n$  toma valores en un entorno convexo compacto de  $\theta_0$  en  $\dot{\Theta}$ .

Sean  $\tilde{\theta}_n = (\tilde{\theta}_{n_1}, \dots, \tilde{\theta}_{n_p})$  y  $\theta_0 = (\theta_{0_1}, \dots, \theta_{0_p})$ , por la proposición 1.19 (tomando  $f(y, \theta) = \langle f'_j(\theta), y - f(\theta) \rangle_n$ ) existe una función medible  $\bar{\theta}_n : \mathbb{Y} \rightarrow \Theta$  tal que:

$$\begin{aligned} \langle f'_j(\tilde{\theta}_n), y - f(\tilde{\theta}_n) \rangle_n - \langle f'_j(\theta_0), y - f(\theta_0) \rangle_n &= \sum_{k=1}^p \langle f''_{jk}(\bar{\theta}_n), y - f(\bar{\theta}_n) \rangle_n (\tilde{\theta}_{n_k} - \theta_{0_k}) \\ &\quad - \langle f'_j(\bar{\theta}_n), f'_k(\bar{\theta}_n) \rangle_n (\tilde{\theta}_{n_k} - \theta_{0_k}), \end{aligned}$$

donde  $|\bar{\theta}_n - \theta_0| \leq |\tilde{\theta}_n - \theta_0|$

Multiplicando por  $\sqrt{n}$ , usando la hipótesis 1.4 y despejando:

$$\sqrt{n} \langle f'_j(\theta_0), y - f(\theta_0) \rangle_n = \sum_{k=1}^p \langle f'_j(\bar{\theta}_n), f'_k(\bar{\theta}_n) \rangle_n \sqrt{n} (\tilde{\theta}_{n_k} - \theta_{0_k}) \quad (1.7)$$

$$- \langle f''_{jk}(\bar{\theta}_n), y - f(\bar{\theta}_n) \rangle_n \sqrt{n} (\tilde{\theta}_{n_k} - \theta_{0_k}) \quad (1.8)$$

$$+ \frac{\sqrt{n}}{p} \langle f'_j(\tilde{\theta}_n), y - f(\tilde{\theta}_n) \rangle_n. \quad (1.9)$$

Como  $(\tilde{\theta}_n)$  y  $(\hat{\theta}_n)$  son *tail equivalents*, por la definición de producto de colas:

i)  $\tilde{\theta}_n = \hat{\theta}_n$  de forma casi segura, para un  $n$  lo suficientemente grande.

Además, también encontramos que:

ii)  $\frac{\partial}{\partial \theta_i} Q_n(\hat{\theta}_n) = 0$  cuando  $\hat{\theta}_n$  es un punto interior de  $\Theta$ .

También, como:

$$\frac{\partial}{\partial \theta_i} Q_n(\theta) = \frac{\partial}{\partial \theta_i} \|y - f(\theta)\|_n^2 = 4\|y - f(\theta)\|_n \langle f'_i(\theta), f(\theta) - y \rangle,$$

encontramos que:

iii) (1.9) es producto de  $\frac{\partial}{\partial \theta_j} Q_n(\tilde{\theta}_n)$ .

Así, mediante:

- (i), (ii), (iii),
- $\hat{\theta}_n \xrightarrow{cs} \theta_0$ ,
- Th. 1.15,

encontramos que:

$$(1.7) \quad \left\langle f'_j(\bar{\theta}_n), f'_k(\bar{\theta}_n) \right\rangle_n \xrightarrow{cs} [a(\theta)]_{jk},$$

$$(1.8) \quad \left\langle f''_{jk}(\bar{\theta}_n), y - f(\bar{\theta}_n) \right\rangle_n (\bar{\theta}_{n_k} - \theta_{0_k}) \xrightarrow{cs} 0,$$

$$(1.9) \quad \left\langle f'_j(\tilde{\theta}_n), y - f(\tilde{\theta}_n) \right\rangle_n \xrightarrow{cs} 0.$$

Por otro lado, por el corolario 1.17 encontramos que:

$$\mathcal{L} \left( \sqrt{n} \left( \langle f'_1(\theta_0), \varepsilon \rangle_n, \dots, \langle f'_p(\theta_0), \varepsilon \rangle_n \right) \right) \longrightarrow N(0, \sigma^2 a(\theta_0)).$$

De esta manera, por definición de producto de matrices y vectores, y considerando que  $y - f(\theta_0) = \varepsilon$ , tenemos que:

$$\sqrt{n} \langle f'_j(\theta_0), \varepsilon \rangle_n = [a_n(\theta_0)]_{\cdot j} \sqrt{n} (\hat{\theta}_n - \theta_0) + o(1), \quad \text{casi seguro,}$$

donde  $[a_n(\theta_0)]_{\cdot j}$  es la columna  $j$ -ésima de la matriz  $a_n(\theta_0)$  y  $o(1)$  denota un término que converge a 0 de forma casi segura (en este caso recoge a los términos (1.8) y (1.9)). Así, es directo que:

$$\sqrt{n} \left( \langle f'_1(\theta_0), \varepsilon \rangle_n, \dots, \langle f'_p(\theta_0), \varepsilon \rangle_n \right) = a_n(\theta_0) \sqrt{n} (\hat{\theta}_n - \theta_0) + o(1), \quad \text{casi seguro,}$$

y concluimos que:

$$\mathcal{L} \left( a_n(\theta_0) \sqrt{n} (\hat{\theta}_n - \theta_0) \right) \longrightarrow N(0, \sigma^2 a(\theta_0));$$

por lo tanto:

$$\mathcal{L} \left( \sqrt{n} (\hat{\theta}_n - \theta_0) \right) \longrightarrow N(0, \sigma^2 a^{-1}(\theta_0)).$$

Finalmente, la segunda parte del Teorema se deduce de que  $a_n \rightarrow a$  uniformemente y de que  $\hat{\theta}_n \xrightarrow{cs} \theta_0$

□

*Observación 1.28.* Para mantener una notación coherente con la bibliografía empleada, vamos a considerar  $V(\theta) = \frac{\sigma^2}{n} a^{-1}(\theta)$ , de tal forma que:

$$\hat{\theta} \xrightarrow{d} N(\theta_0, V(\theta_0)).$$

Donde  $V(\theta_0)$  será conocida como la matriz de varianzas y covarianzas de la distribución asintótica de los estimadores, y  $V(\theta_1), \dots, V(\theta_p)$ , con  $\theta_0 = (\theta_1, \dots, \theta_p)$ , como las varianzas de la distribución asintótica de los estimadores  $\theta_1, \dots, \theta_p$ .

Como vimos que  $\hat{\theta}$  es un estimador fuertemente consistente de  $\theta_0$ , es directo que  $V(\hat{\theta})$  es un estimador fuertemente consistente de  $V(\theta_0)$ .

Es importante aclarar que en la práctica rara vez vamos a conocer el valor de  $\sigma^2$  y la expresión de  $a(\theta)$  para  $V(\hat{\theta})$ . En ambos casos, bastará con sustituir  $\sigma$  por  $\hat{\sigma}_n$ , el cual hemos visto en el Teorema 1.23 que es un estimador fuertemente consistente de  $\sigma$ , y  $a(\theta)$  por  $a_n(\theta)$ , ya que conocemos que  $a_n(\hat{\theta})$  es un estimador fuertemente consistente de  $a(\theta_0)$ .

#### 1.4. Intervalos de confianza y contrastes basados en la distribución asintótica

Los estimadores asintóticos que hemos presentado en la sección 1.3 los podemos utilizar para construir intervalos de confianza y hacer contrastes de hipótesis tal y como hacíamos en la regresión lineal. Mas concretamente, obtenemos:

$$ICA(\hat{\theta}_j) = \left( \hat{\theta}_j - t_{n-p, \frac{\alpha}{2}} SD(\hat{\theta}_j), \quad \hat{\theta}_j + t_{n-p, \frac{\alpha}{2}} SD(\hat{\theta}_j) \right),$$

un intervalo de confianza del parámetro  $\hat{\theta}_j$ , en donde  $SD(\hat{\theta}_j)$  denota una estimación del error típico de  $\hat{\theta}_j$ .

Para los contrastes usaremos la misma metodología que en el caso lineal. Por ejemplo, para contrastes bilaterales:

$$\begin{cases} H_0 : \theta_j = \theta_j^*, \\ H_1 : \theta_j \neq \theta_j^*, \end{cases} \quad (1.10)$$

utilizaremos el estadístico

$$T = \frac{\hat{\theta}_j - \theta_j^*}{SD(\hat{\theta}_j)} \sim T_{n-p}$$

#### 1.5. Intervalos de confianza basadas en el perfil de RSS

El que los estimadores  $\hat{\theta}$  ya no sean normales va a suponer que estos ya no tienen porque ser simétricos. Esto nos hace preguntarnos por unos intervalos que mantengan la asimetría de los estimadores.

Para esta tarea, nos vamos a apoyar en el concepto, propio de la Inferencia Estadística, de la verosimilitud. De esta manera, proponemos un intervalo de confianza basándonos en la verosimilitud de los valores de  $\theta_j$  al minimizar lo que conocemos como perfil de  $RSS$ , de tal manera que será más verosímil aquel valor de  $\theta_j$  que perjudique menos a la  $RSS$ .

Definimos el perfil  $RSS$  (profile  $RSS$ ) de  $\theta_j$  como:

$$pRSS(\theta_j) = \min_{\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p} RSS(\theta_1, \dots, \theta_p);$$

es decir, fijado el valor de  $\theta_j$ , minimizamos la suma residual de cuadrados según los demás parámetros.

*Observación 1.29.* El haber fijado el valor de  $\theta_j$  hace que empeore la  $RSS$ :

$$pRSS(\theta_j) > pRSS(\hat{\theta}_j)$$

Ahora, para construir un intervalo vamos a tomar:

$$\tau(\theta_j) = \text{signo}(\theta_j - \hat{\theta}_j) \frac{\sqrt{pRSS(\theta_j)} - \sqrt{\hat{\theta}_j}}{\hat{\sigma}};$$

obteniendo el intervalo:

$$\left\{ \theta_j : -t_{n-p, \frac{\alpha}{2}} < \tau(\theta_j) < t_{n-p, \frac{\alpha}{2}} \right\}$$

*Observación 1.30.* Mientras en el caso lineal la función  $RSS(\theta)$  describía elipsoides, en el caso no lineal no tiene porque ocurrir, ya que los estimadores no tienen ni porque ser simétricos.



## Capítulo 2

# Simulación

### 2.1. Motivación y Objetivos

Vistas las propiedades teóricas de los estimadores, veamos si somos capaces de reproducirlas, así como de ver su comportamiento en diferentes contextos, mediante simulación de Montecarlo.

Consideremos el modelo de cinética enzimática de Michaelis-Menten:

$$m(x, (K, V)) = \frac{Vx}{K + x}, \quad (2.1)$$

del que extraemos una muestra aleatoria simple  $(x_1, y_1), \dots, (x_n, y_n)$  que nos permitirá estimar  $K$  y  $V$ . El objetivo de la simulación será comprobar las propiedades asintóticas de los estimadores, así como los intervalos de confianza y tests que hemos expuesto.

Para ello, vamos a:

- i)* Verificar que  $\hat{\theta}_n = (\hat{K}_n, \hat{V}_n)$  es asintóticamente insesgado.
- ii)* Comprobar que el valor  $[V(\theta_0)]_{11} = V(K) = \left( \frac{\partial m(x, (K, V))}{\partial K} \right)^2$  es asintóticamente la varianza de  $\hat{K}_n$  (respectivamente  $[V(\theta_0)]_{22} = V(V)$  la de  $\hat{V}_n$ ).
- iii)* Estudiar el comportamiento de los intervalos dados por el perfil RSS frente a los asintóticos.
- iv)* Analizar los test de contrastes unilaterales que nos proporciona R.
- v)* Constatar que el estimador converge en distribución a una normal.

## 2.2. Sesgo

Hemos visto que  $\hat{\theta}$  es un estimador fuertemente consistente de  $\theta$ , que además es asintóticamente insesgado, es decir,  $\mathbb{E}(\hat{\theta}_n)$  converge a  $\theta_0$  cuando  $n \rightarrow \infty$ .

Con el objetivo de estudiar experimentalmente esta convergencia tenemos que considerar que el valor de  $\mathbb{E}(\hat{\theta}_n)$  ha de ser aproximado, ya que carecemos de una expresión analítica. Por este motivo, vamos a simular  $S$  muestras del modelo (2.1) con  $\sigma = 5$ ,  $K = 20$ ,  $V = 120$ ; de tal forma que vamos a tener  $S$  estimadores obtenidos cada uno de ellos con una de las  $S$  muestras:

$$\begin{pmatrix} (x_1^1, y_1^1) & \dots & (x_n^1, y_n^1) \\ \vdots & & \vdots \\ (x_1^s, y_1^s) & \dots & (x_n^s, y_n^s) \end{pmatrix} \leftarrow \begin{matrix} \hat{\theta}_{n_1} = (\hat{K}_1, \hat{V}_1) \\ \vdots \\ \hat{\theta}_{n_s} = (\hat{K}_{n_s}, \hat{V}_{n_s}) \end{matrix},$$

en donde hemos tomado  $x_1, \dots, x_n$  de manera equiespaciada en  $[10, 110]$ ,  $y_i = m(x_i, \theta) + \varepsilon_i$  para todo  $i$ , y con  $\varepsilon_1, \dots, \varepsilon_n$  normales de media cero y desviación típica  $\sigma$ .

Estos  $S$  estimadores simulados conforman una muestra de variables aleatorias independientes e idénticamente distribuidas ( $\hat{\theta}_1 \stackrel{d}{=} \dots \stackrel{d}{=} \hat{\theta}_s \stackrel{d}{=} \hat{\theta}$ ) a la que llamaremos  $\mathcal{T} = (\mathcal{K}, \mathcal{V})$ . De esta muestra simulada conocemos que, dado  $n$  el tamaño muestral y  $\hat{\theta}_i$  el estimador obtenido en la  $i$ -ésima muestra simulada, en virtud del Teorema central del límite:

$$\frac{1}{s} \sum_{i=1}^s \hat{\theta}_i = \bar{\mathcal{T}} \longrightarrow N \left( \mathbb{E}(\hat{\theta}_n), \frac{V(\hat{\theta}_n)}{S} \right),$$

Cuando  $S$  tiende a infinito.

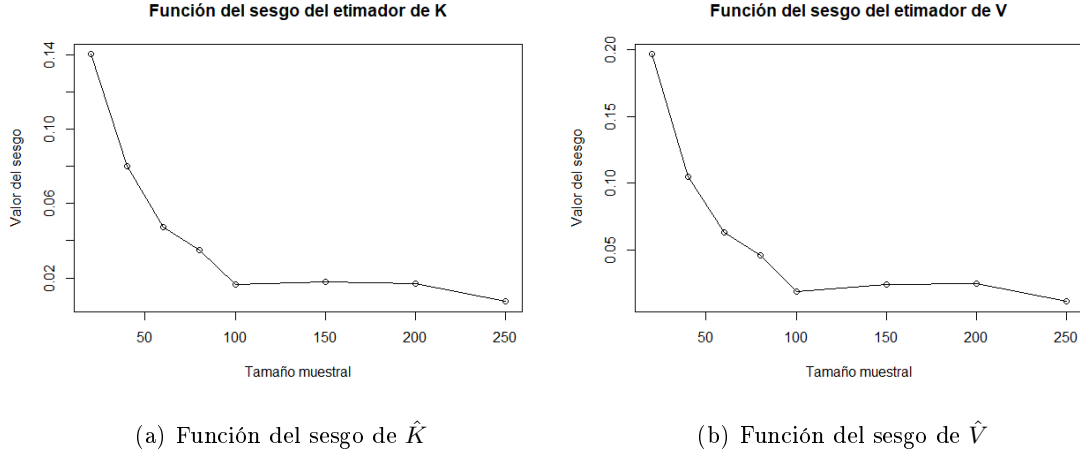
Luego, tomando la media muestral de todos los  $\hat{\theta}_i$  simulados (llamemosle  $\bar{\mathcal{T}} = (\bar{\mathcal{K}}, \bar{\mathcal{V}})$ ), para un  $S$  lo suficientemente grande como para minimizar los efectos de la varianza, obtendríamos una buena aproximación de  $\mathbb{E}(\hat{\theta}_n)$ .

En las gráficas 2.1 podemos apreciar como el sesgo disminuye cada vez más a medida de que aumenta el tamaño muestral, llegando a estar cerca del cero.

*Observación 2.1.* En estas mismas gráficas (las 2.1) podemos ver cómo a partir de  $n = 100$ , las dos parecen sugerir pequeños aumentos y disminuciones en el sesgo para tamaños muestrales mayores. Esto se debe a la desviación típica de  $\bar{\mathcal{K}}$  y de  $\bar{\mathcal{V}}$  en relación al tamaño del sesgo. Por ejemplo, en  $n = 100$  la desviación típica de  $\bar{\mathcal{K}}$  sería aproximadamente  $\sqrt{\frac{1,35}{S}} = 0,0082$ , mientras que el sesgo vale 0,02.

Teniendo en cuenta que estamos trabajando con sesgos tan pequeños, es normal que tengamos pequeños errores de aproximación



Gráfica 2.1: Sesgo de  $\hat{K}$  y  $\hat{V}$  obtenida a partir de 20.000 simulaciones

### 2.3. Varianza asintótica

Tal y como hicimos con el sesgo, necesitamos encontrar aproximaciones de la varianza de  $\hat{K}_n$  y  $\hat{V}_n$  para poder comprobar mediante simulación que asintóticamente se aproximan a los valores teóricos de  $V(K)$  y  $V(V)$  respectivamente.

Dados los estimadores simulados  $\mathcal{T}$ , conocemos que  $Sc(\mathcal{K}) = \frac{1}{S-1} \sum_{i=1}^S (\hat{K}_i - \bar{K})^2$ , la cuasivarianza de los valores de  $\hat{K}_n$ s, es una aproximación por Montecarlo de la varianza verdadera de  $\hat{K}_n$ . Por lo tanto, con un  $S$  lo suficientemente grande como para minimizar el error de simulación, deberíamos obtener que la cuasivarianza se aproxima a la varianza verdadera de  $\hat{K}_n$ ,  $Var(\hat{K}_n)$ . Por otro lado, tenemos la varianza de la distribución asintótica de  $\hat{K}_n$ , que denotaremos  $AVar(\hat{K})$  y que por el Teorema 1.27 vale  $AVar(\hat{K}) = V(K) = \frac{\sigma^2}{n} a^{-1}(\theta)$ . Ahora bien, como no conocemos la expresión de  $a(\theta)$ , vamos a aproximarla por  $a_n(\theta)$ , de tal forma que para calcular  $AVar(\hat{K}_n)$ , vamos a tomar:

$$AVar(\hat{K}) = \frac{\sigma^2}{n} a_n^{-1}(K),$$

con  $\sigma$  y  $K$  conocidos.

En concreto, queremos establecer que la razón de las varianzas converge a 1:

$$\frac{Sc(\mathcal{K})}{AVar(\hat{K}_n)} \rightarrow 1,$$

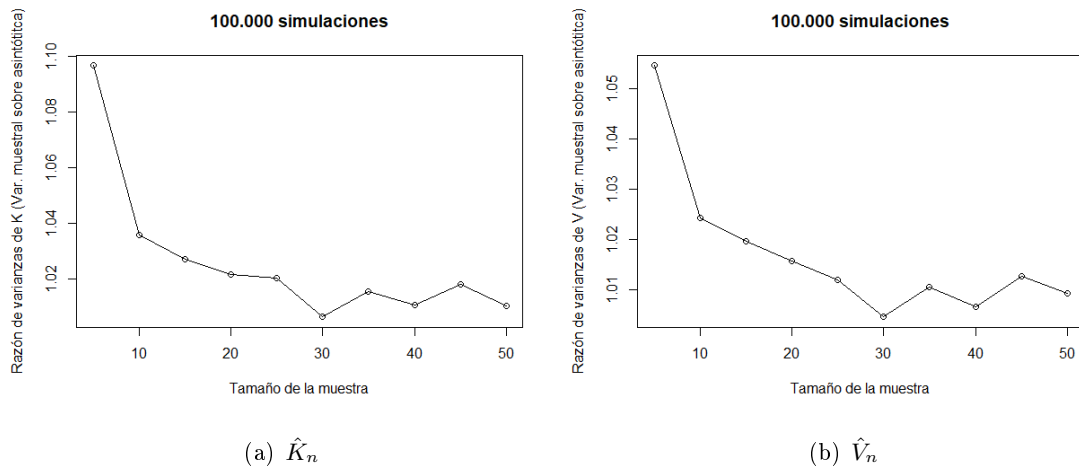
cuando  $n \rightarrow \infty$ .

De forma similar, establecemos que:

$$\frac{Sc(\mathcal{V})}{AVar(\hat{V})} \rightarrow 1,$$

cuando  $n \rightarrow \infty$ .

En las gráficas 2.2 podemos observar un comportamiento tal como el esperado para una muestra simulada de tamaño  $S = 100000$ .



Gráfica 2.2: Razón de varianzas muestrales y asintóticas de los estimadores

*Observación 2.2.* Nótese que, al igual que pasaba con el sesgo, las gráficas presentan comportamientos no deseados a partir de  $n=25$ . Esto es debido, nuevamente, a que la variación de  $Sc(\mathcal{K})$  es relativamente grande en comparación a lo cerca que está de  $AVar(\hat{K}_n)$ . En el cuadro 2.1, podemos ver una comparación entre los valores de  $Sc(\mathcal{K})$  y  $AVar(\hat{K}_n)$ , y el intervalo de confianza para la varianza de  $\mathcal{K}$  para reflejar el error de aproximación. Vemos como, incluso, para valores pequeños de  $n$   $AVar(\hat{K})$  no está en los intervalos.

## 2.4. Intervalos de confianza

A nivel teórico hemos visto dos tipos de intervalos para los parámetros del modelo: uno asintótico, asumiendo normalidad del estimador por tener un  $n$  lo suficientemente grande, y otro mediante el perfil de  $RSS$ .

Por lo tanto, sería de interés comprobar mediante simulación la cobertura de estos intervalos para diferentes tamaños muestrales.

$n$	$Sc(\mathcal{K})$	$AVar(\hat{K})$	Intervalo de confianza
5	22.83	20.82	(22.62, 23.02)
10	12.34	11.91	(12.23, 12.44)
15	8.53	8.30	(8.45, 8.60)
20	6.51	6.37	(6.45, 6.56)
25	5.27	5.19	(5.22, 5.32)
30	4.38	4.35	(4.33, 4.41)
35	3.81	3.75	(3.77, 3.84)
40	3.34	3.30	(3.30, 3.36)
45	3.00	2.95	(2.97, 3.02)
50	2.69	2.66	(2.66, 2.71)

Cuadro 2.1: Valores de  $Sc(\mathcal{K})$ ,  $AVar(\hat{K})$ , y el intervalo elemental basado en la  $\mathcal{X}_{S-1}^2$  sobre la muestra  $\mathcal{K}$ ; para diferentes tamaños muestrales

Para cada par  $(\hat{K}_{n_i}, \hat{V}_{n_i})$ , calcularemos los intervalos de confianza y de perfil, y contaremos en cuántos de ellos están los verdaderos valores de  $K$  y  $V$ .

Con un número de simulaciones  $S = 20000$  y  $\sigma = 15$ , para intervalos con nivel de confianza del 95%, vemos que para tamaños muestrales pequeños los intervalos de perfil tienen una cobertura más próxima al nivel que los asintóticos, mientras que los intervalos asintóticos tienen coberturas por debajo del nivel. Para muestras mayores, los intervalos asintóticos comienzan a comportarse mejor (ver cuadro 2.2).

$n$	$ICA$		$pRSS$	
	$K$	$V$	$K$	$V$
15	0.9390	0.9467	0.9471	0.9537
30	0.9426	0.9473	0.9485	0.9482

Cuadro 2.2: Cobertura de los intervalos de confianza asintóticos y de perfil de  $RSS$  ( $ICA$  y  $pRSS$ ) para  $K$  y  $V$ , con  $\sigma = 15$  y 20000 simulaciones

Si considerásemos  $\sigma = 5$ , conseguimos un resultado diferente. Tal y como vemos en el cuadro 2.3, los intervalos asintóticos tienen más cobertura que el nivel nominal del 5% para  $n = 15$

Vamos a estudiar ahora la longitud de los intervalos. En el cuadro 2.4 observamos que la

$n$	$ICA$		$pRSS$	
	$K$	$V$	$K$	$V$
15	0.9647	0.9660	0.9470	0.949
30	0.9570	0.9576	0.9490	0.948

Cuadro 2.3: Cobertura de los intervalos de confianza asintóticos y de perfil de  $RSS$  ( $ICA$  y  $pRSS$ ) para  $K$  y  $V$ , con  $\sigma = 5$  y 20000 simulaciones

longitud de los intervalos asintóticos es menor que los de perfil. Notese que estos intervalos tienen menor cobertura, como veíamos en el cuadro 2.2.

n	$ICA$		$pRSS$	
	$K$	$V$	$K$	$V$
15	12.45 ; 20.270		12.570 ; 20.296	
30	8.544 ; 13.990		8.5650 ; 13.900	

Cuadro 2.4: Longitud de los intervalos de confianza asintóticos ( $ICA$ ) y de perfil  $RSS$  ( $pRSS$ ) para  $K$  y  $V$ , con  $\sigma = 5$  y 20000 simulaciones

Otra característica de los intervalos dados por el perfil  $RSS$  es que son asimétricos. Así pues, dada una muestra única de datos, vamos a representar los intervalos sobre la función perfil de los parámetros y veamos como se construyen.

Consideremos, entonces:

$$pRSS(K) = \min_V RSS(K, V), \quad pRSS(V) = \min_K RSS(K, V);$$

las funciones perfil de  $K$  y de  $V$ .

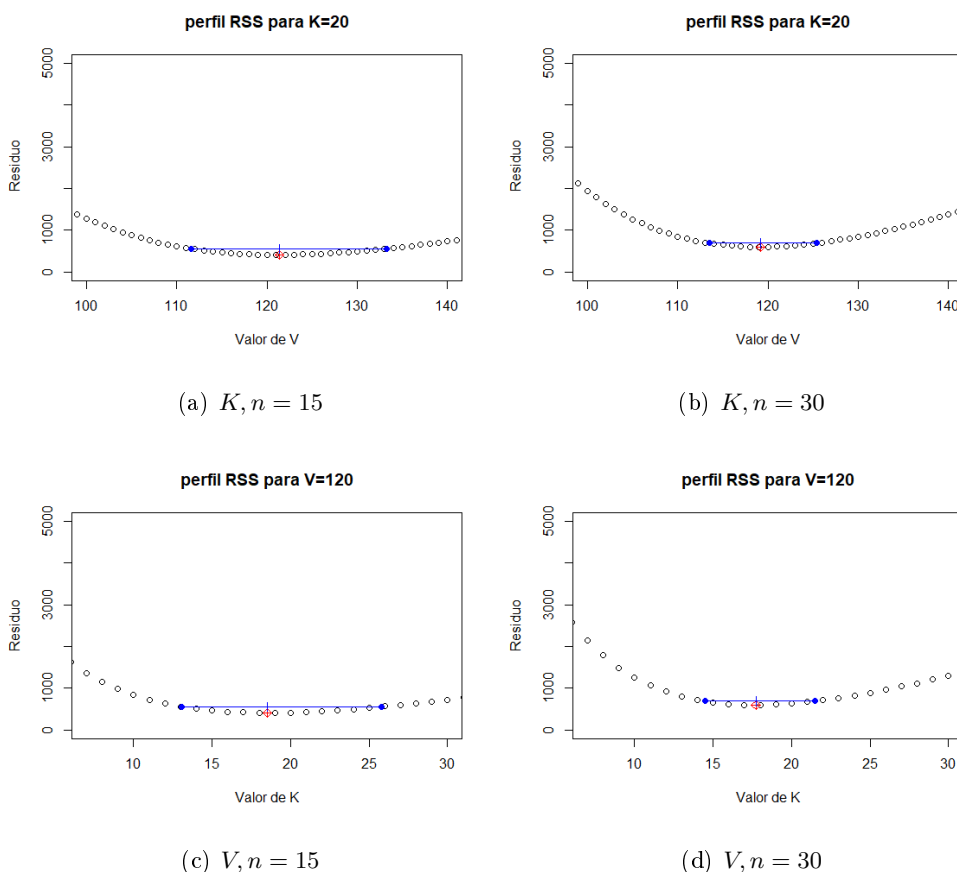
Podemos ver en la gráfica 2.3 que, como era de esperar, no encontramos una asimetría muy grande en los intervalos para  $\sigma = 5$  debido a que para desviaciones pequeñas los estimadores de nuestro modelo convergen más rápido a una normal. Si tomamos  $\sigma = 15$  podemos apreciar mucho mejor la asimetría respecto al estimador, tal y como se ve en la gráfica 2.4.

Para una mejor visualización, en el cuadro 2.5 se han dado las razones de las asimetrías de los intervalos, es decir:

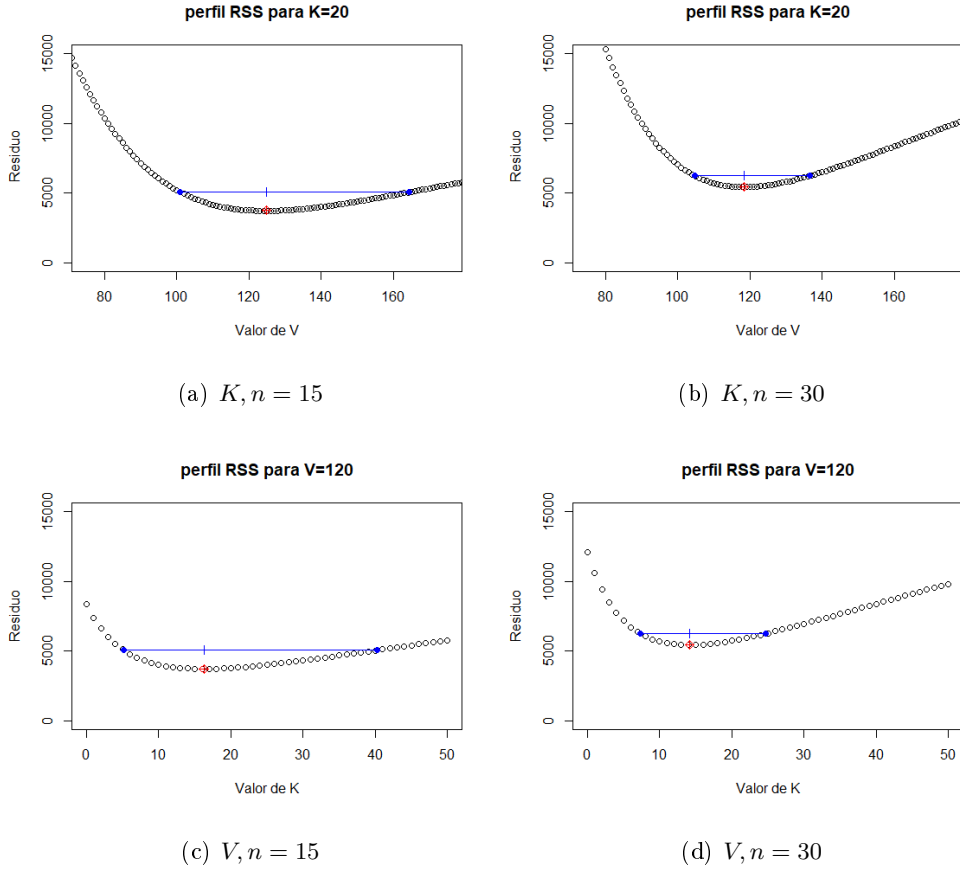
$$\frac{\text{Estimador} - \text{Extremo izquierdo del intervalo}}{\text{Extremo derecho del intervalo} - \text{Estimador}}$$

	$\hat{K}_n$	$\hat{V}_n$
$\sigma = 15, n = 15$	2.16	1.66
$\sigma = 15, n = 30$	1.54	1.33
$\sigma = 5, n = 15$	1.33	1.21
$\sigma = 5, n = 30$	1.17	1.11

Cuadro 2.5: Razón de asimetrías de los intervalos de perfil.



Gráfica 2.3: Función perfil de ambos parámetros para  $n = 15$  y  $n = 30$ , con  $\sigma = 5$ . En rojo el valor del estimador y en azul los intervalos.



Gráfica 2.4: Función perfil de ambos parámetros para  $n = 15$  y  $n = 30$ , con  $\sigma = 15$ . En rojo el valor del estimador y en azul los intervalos.

## 2.5. Contrastes

Planteamos los siguientes contrastes unilaterales sobre el valor de los parámetros:

$$\begin{cases} H_0 : K = 20; \\ H_1 : K \neq 20. \end{cases} \quad (2.2)$$

$$\begin{cases} H_0 : V = 120; \\ H_1 : V \neq 120. \end{cases} \quad (2.3)$$

Hemos visto con anterioridad que usaríamos para estos test los estadísticos:

$$T_{K_{ni}} = \frac{\hat{K}_{ni} - K_0}{\sqrt{AVar(\hat{K}_n)}}, \quad T_{V_{ni}} = \frac{\hat{V}_{ni} - V_0}{\sqrt{AVar(\hat{V}_n)}} \sim T_{n-2} \quad ;$$

asumiendo normalidad de los estimadores, y en donde hemos considerado esta vez  $\sigma$ , y  $K$  y  $V$  desconocidos, de tal forma que hemos calculado las varianzas de sus distribuciones

asintóticas como:  $AVar(\hat{K}_n) = \frac{\hat{\sigma}_n^2}{n} a_n(\hat{K}_n)$  y  $AVar(\hat{V}_n) = \frac{\hat{\sigma}_n^2}{n} a_n(\hat{V}_n)$ .

Obviamente, los estimadores no siempre son normales, por lo que cabría preguntarse como evoluciona el nivel de estos tests en función del tamaño muestral.

Dada la muestra simulada de  $S$  estimadores, vamos a calcular cada  $T_{K_{n_i}}$  y  $T_{V_{n_i}}$  para así poder averiguar con cuánta frecuencia los estadísticos no verifican que:

$$-t_\alpha \leq \hat{T} \leq t_\alpha,$$

es decir, cuantos estadísticos caen en la región de rechazo, siendo  $t_\alpha$  un valor de una T de Student de  $n - 2$  grados de libertad, que deja a su derecha una región de probabilidad  $\alpha$ . Podemos ver en los cuadros 2.6 y 2.7 los diferentes porcentajes de rechazo de los contrastes para distintos valores de  $\sigma$ .

$n$	$H_0 : K = 20$	$H_0 : V = 129$
15	6.5	5.1
30	5.6	4.9

Cuadro 2.6: Porcentajes de rechazo de los dos contrastes según el tamaño muestral con nivel de significación del 5%,  $\sigma = 15$  y 100000 simulaciones

$n$	$H_0 : K = 20$	$H_0 : V = 129$
15	5.1	4.9
30	5.1	5.0

Cuadro 2.7: Porcentaje de rechazo de los dos contrastes según el tamaño muestral con nivel de significación del 5%,  $\sigma = 5$  y 100000 simulaciones

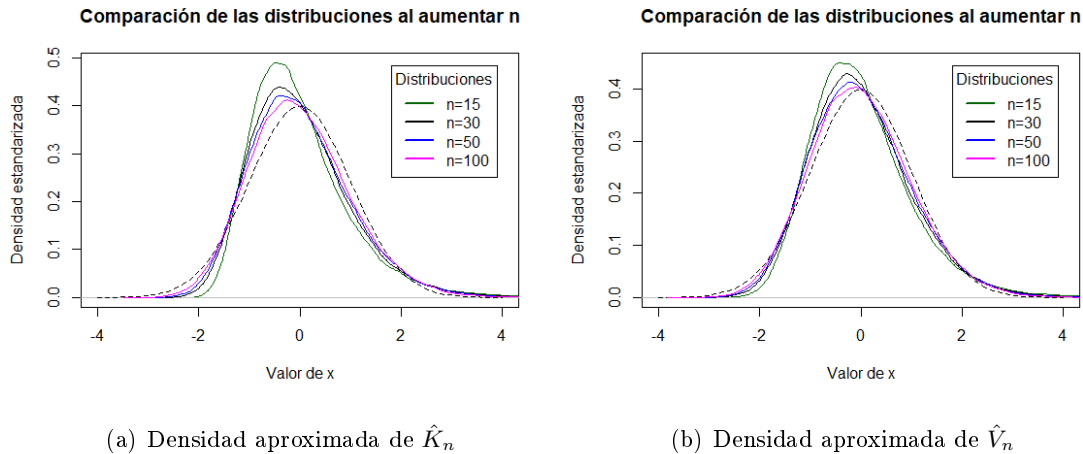
Estos resultados coinciden con lo obtenido en la sección 2.4 : al reducir la varianza, las aproximaciones asintóticas mejoran.

## 2.6. Distribuciones

Durante todo momento hemos considerado que la distribución de  $\hat{\theta}_n$  se aproxima por una normal de media  $\theta$  y varianza  $AVar(\hat{\theta}_n)$ . Hemos, incluso, construido intervalos y tests de contraste basados en esto, cuyos niveles mejoraban al aumentar el tamaño muestral  $n$ .

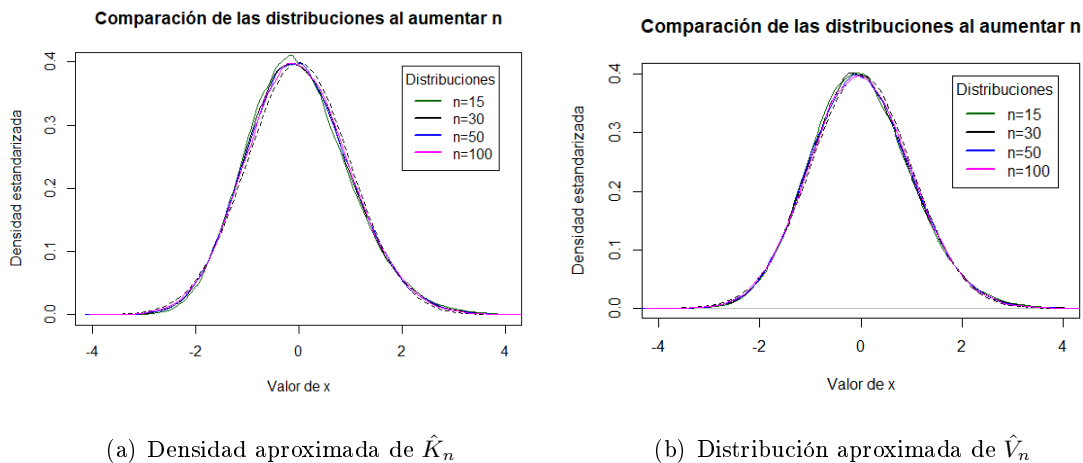
Consideremos  $\mathcal{T} = (\mathcal{K}, \mathcal{V})$  la muestra de estimadores simulados. Para estudiar la densidad de éstos vamos a tomar la estandarización de las muestras  $\mathcal{K}$  y  $\mathcal{V}$ , de tal forma que de seguir  $\mathcal{K}$  y  $\mathcal{V}$  una distribución normal, la muestra estandarizada tendrá que seguir una normal de

media 0 y varianza 1; y vamos a utilizar un estimador de la función de densidad tipo núcleo.



Gráfica 2.5: Funciones de densidad de  $\hat{K}_n$  y  $\hat{V}_n$ , con  $\sigma = 15$  y 100000 simulaciones.

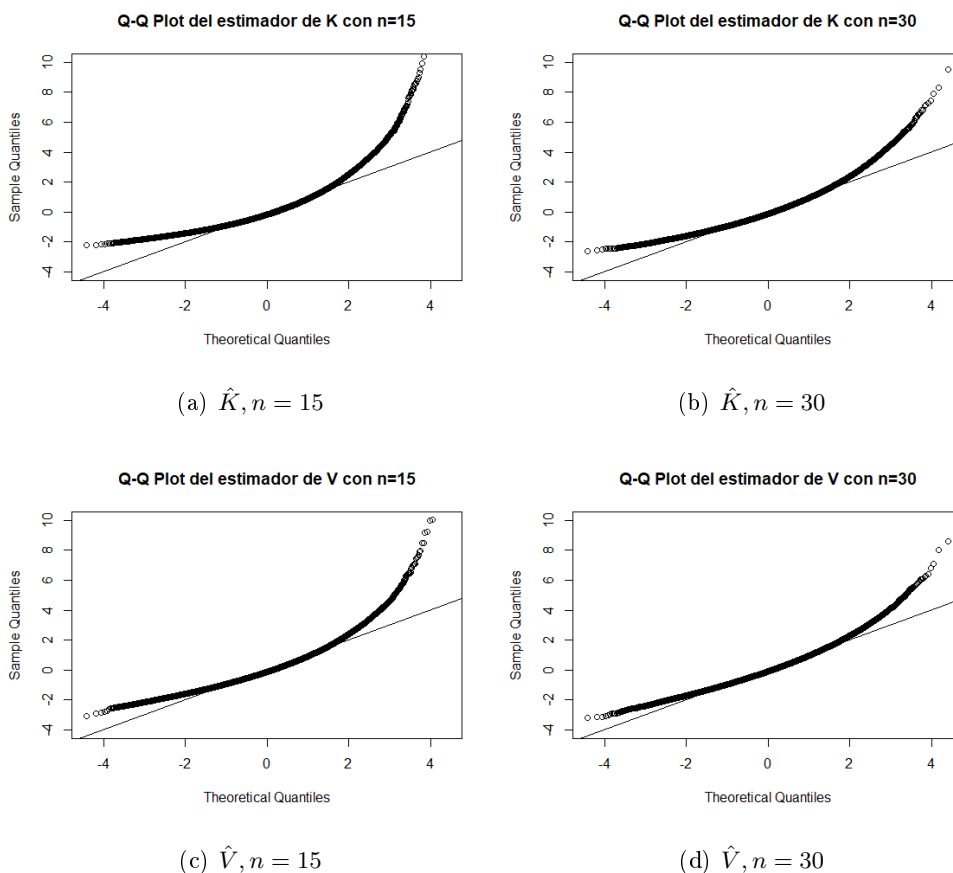
En la gráfica 2.5 podemos ver cierta tendencia a una distribución normal al aumentar el tamaño muestral, algo coherente con los resultados obtenidos en las secciones 2.4 y 2.5. Ahora bien, considerando una varianza menor, los resultados obtenidos en apartados anteriores sugerirían una mayor normalidad de los estimadores, por lo que sería interesante estudiar estas densidades, pero para un  $\sigma$  menor. Los resultados vistos en la gráfica 2.6



Gráfica 2.6: Funciones de densidad de  $\hat{K}_n$  y  $\hat{V}_n$ , con  $\sigma = 5$  y 100000 simulaciones.

sugieren que una disminución de la varianza de los errores conlleva una convergencia más



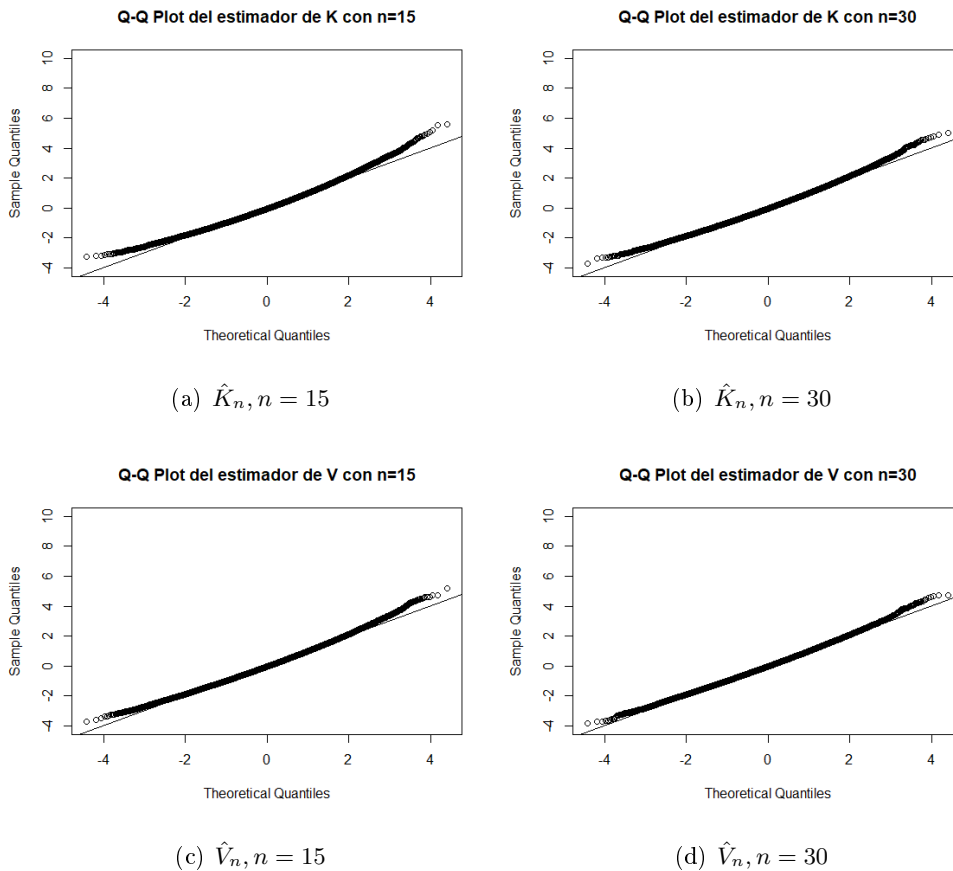


Gráfica 2.7: Q-Q Plot de ambos estimadores para  $n = 15$  y  $n = 30$ , con  $\sigma = 15$ .

rápida a una normal. Podemos ver estos resultados con más claridad mirando los gráficos cuantil-cuantil de los diferentes estimadores en la gráfica 2.7 (a mas alejados de la diagonal, mayor discrepancia respecto de la normal).

Cabe destacar que para  $\sigma = 5$ , tal y como se ve en la gráfica 2.8, los cuantiles más grandes y los más pequeños se acercan mucho más a su valor teórico, indicándonos que los estimadores son cada vez más simétricos.

*Observación 2.3.* En las gráficas Cuantil-Cuantil de una muestra de tamaño  $n$  (también llamadas Q-Q Plots) representamos  $n + 1$  puntos, dados por los  $(n + 1)$ -cuantiles de la muestra con los correspondientes cuantiles de la distribución de comparación (en este caso la normal de media 0 y varianza 1). De esta manera, si la distribución de la muestra fuese normal, sus cuantiles tendrían que coincidir con los teóricos, acercándose a la diagonal.



Gráfica 2.8: Q-Q Plot de ambos estimadores para  $n = 15$  y  $n = 30$ , con  $\sigma = 5$ .

### 2.6.1. Efecto de la distribución del error

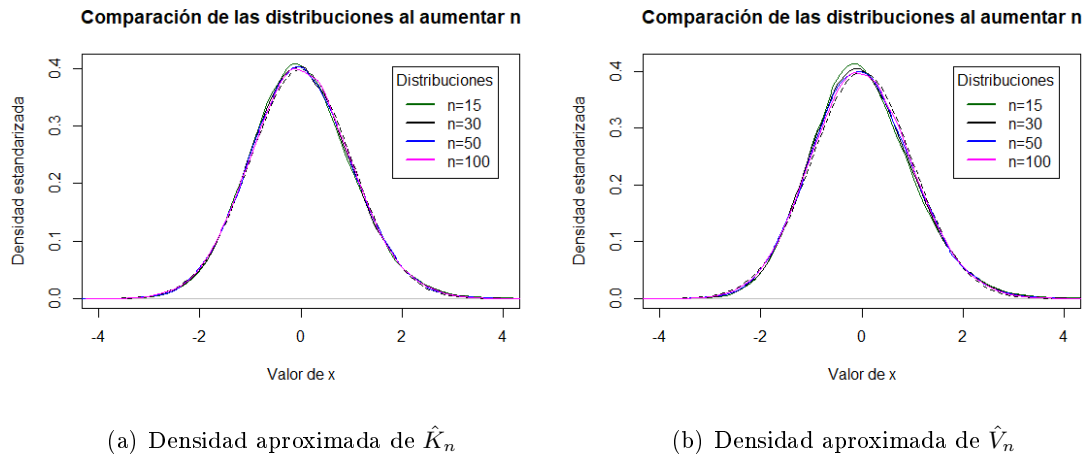
Uno de los resultados que probamos durante la teoría fue que no es necesario la normalidad de los errores para garantizar la normalidad asintótica de los estimadores. Sin embargo, en todos los ejemplos hemos considerado siempre los errores normales por su sencillez.

Para ilustrar mejor esta propiedad vamos a considerar los mismos datos, pero esta vez con errores bajo una distribución Ji-cuadrado centrada de 13 grados de libertad (para tener una varianza similar a casos anteriores).

Bajo estos valores, encontramos unos resultados que se ajustan a los resultados teóricos: fijándonos en las gráfica 2.9, en donde representamos las distribuciones de los estimadores para varios  $n$ , encontramos resultados similares, incluso mejores para  $n$  pequeños, a cuando teníamos un error normal para la distribución de  $K$  con una desviación típica similar (en este caso es aproximadamente 5.1). Este resultado puede ser confuso, ya que la no normalidad de los errores intuitivamente parece ser algo que dificulte la convergencia a

una normal, sin embargo, este resultado nos indica lo contrario.

Esto se debe a que la asimetría que produce el proceso de regresión debido a la no linealidad de la función, corrige la propia asimetría que presenta una Ji-cuadrado con 13 grados de libertad.



Gráfica 2.9: Funciones de densidades de  $\hat{K}_n$  y  $\hat{V}_n$ , con  $\sigma = 5.1$  y 100000 simulaciones.

## 2.7. Conclusión

Analizando los resultados obtenidos, hemos visto cómo para nuestro modelo se han verificado todas las propiedades teóricas de los estimadores, además de mostrar una relación entre el aumento de varianza y la pérdida de normalidad de los estimadores, y la convergencia en distribución cuando  $n \rightarrow \infty$ . Además, también encontramos que no necesariamente la normalidad de los errores implica una convergencia más rápida hacia una normal.



# Código de R empleado

El siguiente código aquí expuesto ha sido el empleado en el apartado de simulación. Las gráficas presentadas en la Introducción provienen de los apuntes para la asignatura Regresión Generalizada y Modelos Mixtos del Máster en técnicas estadísticas[6], mientras que otras partes del código pueden ser encontradas también en Ritz[1].

Cabe destacar que el código que se presenta contiene una gran cantidad de datos simulados, y en ciertas ocasiones puede llevar un tiempo de ejecución muy elevado.

## Librerías

Las siguientes librerías fueron usadas al menos en algún momento del código y son necesarias para su correcta ejecución.

```
library(Hmisc)
library(ggplot2)
library(TeachingDemos)

#Para instalar los paquetes:

#install.packages(Hmisc)
#install.packages(ggplot2)
#install.packages(TeachingDemos)
```

## Sesgo

```
#####CALCULO DEL SESGO EN 2.2#####
Med=matrix(0,nrow=8,ncol=2)
colnames(Med)=c('K','V')
ns=100000
V=120
```

```

K=20
sigma=5
x=seq(10,110,length=n)

#Simulaciones para n=20,...,100
set.seed(123456)
for (i in 1:5){
n=20*i # Tamano muestral
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
Med[i,]=colMeans(theta)

}

#Simulacion de los estimadores para n=150,...,250
for (i in 1:3){
n=100+i*50 # Tamano muestral
V=120
K=20
sigma=5
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)

for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
Med[i+5,]=colMeans(theta)

}

Index=seq(20,100,20)
Index=c(Index,seq(150,250,50)) #Vector de los tamanos muestrales

Sesgo=cbind(matrix(20,8,1),matrix(120,8,1))
Sesgo=Med-Sesgo

```

```

plot(Sesgo[,1]~Index, xlab='Tamano_muestral',ylab='Valor_del_sesgo',
main='Funcion_del_sesgo_del_estimador_de_K')
lines(Index,Sesgo[,1])

plot(Sesgo[,2]~Index, xlab='Tamano_muestral',ylab='Valor_del_sesgo',
main='Funcion_del_sesgo_del_estimador_de_V')
lines(Index,Sesgo[,2])

```

## Varianza

```

#####CALCULO DE LA VARIANZA EN 2.3#####

Var=matrix(0,nrow=8,ncol=2)
Med=matrix(0,nrow=8,ncol=2)
colnames(Med)=c('K','V')
colnames(Var)=c('K','V')
ns=100000
V=120
K=20
sigma=5 #Varianza
x=seq(10,110,length=n)
VAK=matrix(0,nrow=8,ncol=1)
VAV=matrix(0,nrow=8,ncol=1)
set.seed(123456)
#Calculo de la varianza teorica de la distribucion asymptotica de los estim

m=10
VAK=matrix(0,nrow=m,ncol=1)
VAV=matrix(0,nrow=m,ncol=1)
VAC=matrix(0,nrow=m,ncol=1)
for (i in 1:m){
n=5*i

x=seq(10,110,length=n)

Mk=-V*x/(K+x)^2
Mv=x/(K+x)

```

```

Sigma=cbind(c(sum(Mk^2), sum(Mv*Mk)), c(sum(Mv*Mk), sum(Mv^2)))
Sigma=solve(Sigma)
Sigma=Sigma*sigma^2

VAK[i,]=Sigma[1,1]
VAV[i,]=Sigma[2,2]
VAC[i,]=Sigma[1,2]
}

#Calculo de la varianza aproximada por simulacion

v=matrix(0,ncol=2,nrow=m)
int=matrix(0,ncol=2,nrow=m)
for (i in 1:m){
set.seed(123456)
n=5*i
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
int[i,]=sigma.test(theta[,1])$conf.int[1:2]
v[i,]=apply(theta, 2, var)
}

#Representacion de las graficas
v/cbind(VAK,VAV)
par(mfrow=c(1,2))
plot(v[,1]/VAK~Index[1:10],xlab='Tamano de la muestra',ylab='Razon de var',
lines(v[,1]/VAK~Index[1:10])
plot(v[,2]/VAV~Index[1:10],xlab='Tamano de la muestra',ylab='Razon de var',
lines(v[,2]/VAV~Index[1:10])

#Intervalos de confianza para la varianza de theta[,1]
int

```

## Intervalos de Confianza



```
##### COBERTURA DE LOS INTERVALOS pRSS Y ASINTOTICO en 2.4#####
```

```
set.seed(123456)

ns=20000
V=120
K=20
m=2
theta=matrix(0,nrow=ns,ncol=2)
pRSS_IC=matrix(0,nrow=m,ncol=2)
ICA=matrix(0,nrow=m,ncol=2)
contador=matrix(0,nrow=ns,ncol=2)
contador2=matrix(0,nrow=ns,ncol=2)
difeICA=matrix(0,ncol=2,nrow=m)
difeRSS=matrix(0,ncol=2,nrow=m)
#para sigma=5
sigma=5
for (i in 1:m){
n=15*i# Tamano muestral
x=seq(10,110,length=n)

pRSS_ICs=matrix(0,nrow=ns,ncol=4) #Intervalos mediante el pRSS
ICAs=matrix(0,nrow=ns,ncol=4) #Intervalos asintoticos

for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
pRSS_ICs[is,]=c(confint(m)[1,],confint(m)[2,]) #pRSS
ICAs[is,]=c(theta[is,1]+qt(0.05/2,n-2)*summary(m)$coefficients[1,2],
theta[is,1]-qt(0.05/2,n-2)*summary(m)$coefficients[1,2],
theta[is,2]+qt(0.05/2,n-2)*summary(m)$coefficients[2,2],theta[is,2]-
qt(0.05/2,n-2)*summary(m)$coefficients[2,2])

if (20 >= pRSS_ICs[is,1] & 20 <= pRSS_ICs[is,2]){
contador[is,1]=1
```

```

}
if (120 >= pRSS_ICs[is,3] & 120 <= pRSS_ICs[is,4]){
  contador[is,2]=1
}
if (20 >= ICAs[is,1] & 20 <= ICAs[is,2]){

  contador2[is,1]=1
}
if (120 >= ICAs[is,3] & 120 <= ICAs[is,4]){
  contador2[is,2]=1
}

}

}

pRSS_IC[i,]=colSums(contador)
ICA[i,]=colSums(contador2)

difeRSS[i,1]=(pRSS_ICs[,2]-pRSS_ICs[,1])/ns
difeICA[i,1]=(ICAs[,2]-ICAs[,1])/ns

difeRSS[i,2]=(pRSS_ICs[,4]-pRSS_ICs[,3])/ns
difeICA[i,2]=(ICAs[,4]-ICAs[,3])/ns
}
pRSS_IC
ICA
difeRSS
difeICA

#para sigma=15
sigma=15
for (i in 1:m){
n=15*i# Tamano muestral
x=seq(10,110,length=n)

pRSS_ICs=matrix(0,nrow=ns,ncol=4) #Intervalos mediante el pRSS
ICAs=matrix(0,nrow=ns,ncol=4) #Intervalos asintoticos

```

```

for (is in 1:ns){
y=V*x/(K+x)+rnorm(n, sd=sigma)
m=nls(y~V*x/(K+x), start=list(K=20, V=120))
theta[is,]=coef(m)
pRSS_ICs[is,]=c(confint(m)[1,], confint(m)[2,]) #pRSS
ICAs[is,]=c(theta[is,1]+qt(0.05/2,n-2)*summary(m)$coefficients[1,2],
theta[is,1]-qt(0.05/2,n-2)*summary(m)$coefficients[1,2], theta[is,2]+
qt(0.05/2,n-2)*summary(m)$coefficients[2,2], theta[is,2]-qt(0.05/2,n-2)*
summary(m)$coefficients[2,2])

if (20 >= pRSS_ICs[is,1] & 20 <= pRSS_ICs[is,2]){
contador[is,1]=1
}
if (120 >= pRSS_ICs[is,3] & 120 <= pRSS_ICs[is,4]){
contador[is,2]=1
}
if (20 >= ICAs[is,1] & 20 <= ICAs[is,2]){

contador2[is,1]=1
}
if (120 >= ICAs[is,3] & 120 <= ICAs[is,4]){
contador2[is,2]=1

}

}

pRSS_IC[i,]=colSums(contador)
ICA[i,]=colSums(contador2)

difeRSS[i,1]=(pRSS_ICs[,2]-pRSS_ICs[,1])/ns
difeICA[i,1]=(ICAs[,2]-ICAs[,1])/ns

difeRSS[i,2]=(pRSS_ICs[,4]-pRSS_ICs[,3])/ns
difeICA[i,2]=(ICAs[,4]-ICAs[,3])/ns
}
pRSS_IC
ICA

```

```
difeRSS
difeICA
```

## Representación de los perfiles y las razones de asimetría

```
#####pRSS en 2.4#####
Asim=matrix(0,ncol=2,nrow=4)
vn=c(15,30)
vs=c(15,5)
j=0
for (sigma in vs){
  for (n in vn){
    set.seed(123456)
    j=j+1
    p=50
    q=250
    x=seq(10,110,length=n)
    K=20
    V=120
    y=V*x/(K+x)+rnorm(n,sd=sigma)
    m=nls(y~V*x/(K+x),start=list(K=20,V=120))
    theta=matrix(0,nrow=1,ncol=2)
    theta=coef(m)
    RSSK=matrix(0,ncol=1,nrow=q-59)
    RSSV=matrix(0,ncol=1,nrow=p+1)
    for (i in 60:q){
      V=i

      RSSK[i-59]=deviance(nls(y~V*x/(K+x),start=list(K=20)))

    }

    V=120

    for (i in 0:p){
      K=i

      RSSV[i+1]=deviance(nls(y~V*x/(K+x),start=list(V=120)))

    }
  }
}
```

```

K=20

#Representemos la funcion
IndexK=60:q
IndexV=0:p

plot(RSSK~IndexK, xlab='Valor de V', ylab='Residuo', main='perfil RSS para V')

#Valor del estimador de V
V=coef(m)[2]

points(c(coef(m)[2]), c(deviance(nls(y~V*x/(K+x), start=list(K=20)))),
col="red", pch=9)

#Intervalo de V sobre la funcion
V=confint(m)[2,]
y1=deviance(nls(y~V[1]*x/(K+x), start=list(K=20)))
y2=deviance(nls(y~V[2]*x/(K+x), start=list(K=20)))

points(confint(m)[2,], c(y1,y2), col="blue", pch=19)
lines(confint(m)[2,], c(y1,y2), col="blue", type='o')

#Recta
points(c(coef(m)[2]), c(y1), col="blue", pch=3)

plot(RSSV~IndexV, xlab='Valor de K', ylab='Residuo', main='perfil RSS para V')

#Valor del estimador de K
K=coef(m)[1]

points(c(coef(m)[1]), c(deviance(nls(y~V*x/(K+x), start=list(V=120)))),
col="red", pch=9)

#Intervalo de K sobre la funcion
K=confint(m)[1,]
y1=deviance(nls(y~V*x/(K[1]+x), start=list(V=120)))
y2=deviance(nls(y~V*x/(K[2]+x), start=list(V=120)))

points(confint(m)[1,], c(y1,y2), col="blue", pch=19)

```

```

lines(confint(m)[1,],c(y1,y2),col="blue",type='o')

#Asimetrías
Asim[i,]=c((coef(m)[1]-K[1])/(K[2]-coef(m)[1]),(coef(m)[2]-confint(m)[2,1]))

#Recta
points(c(coef(m)[1]),c(y1),col="blue",pch=3)

}
}

```

## Contrastes

```

#####CONTRASTES DE 2.5#####
set.seed(123456)
par(mfrow=c(1,1))
pval=matrix(0,ncol=2,nrow=ns)
ns=100000
theta=matrix(0,nrow=ns,ncol=2)
K=20
V=120
niveisd5=matrix(0,ncol=2,nrow=2)
niveisd15=matrix(0,ncol=2,nrow=2)
contador=matrix(0,nrow=ns,ncol=2)
ThK=matrix(0,ncol=1,nrow=ns)
ThV=matrix(0,ncol=1,nrow=ns)

sigma=5
i=0
for (n in vn){
i=i+1
set.seed(123456)
x=seq(10,110,length=n)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
ThK=(summary(m)$coef[,1][1] - 20)/(summary(m)$coef[,2][1])
ThV=(summary(m)$coef[,1][2] - 120)/(summary(m)$coef[,2][2])
}
}

```

```

pval[is,]=c(2*pt(-abs(ThK),n-2),2*pt(-abs(ThV),n-2))
}

nivelsd5[i,1]=length(pval[1:is,1][pval[1:is,1]<=0.05])/(is)
nivelsd5[i,2]=length(pval[1:is,2][pval[1:is,2]<=0.05])/(is)
}

i=0
sigma=15
for (n in vn){
i=i+1
set.seed(123456)
x=seq(10,110,length=n)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
ThK=(summary(m)$coef[,1][1] - 20)/(summary(m)$coef[,2][1])
ThV=(summary(m)$coef[,1][2] - 120)/(summary(m)$coef[,2][2])
pval[is,]=c(2*pt(-abs(ThK),n-2),2*pt(-abs(ThV),n-2))
}

nivelsd15[i,1]=length(pval[1:is,1][pval[1:is,1]<=0.05])/(is)
nivelsd15[i,2]=length(pval[1:is,2][pval[1:is,2]<=0.05])/(is)
}

```

## Distribuciones para errores normales y sus Q-Q Plot

```

#####DISTRIBUCIONES DE 2.6#####
set.seed(123456)
ns=100000
vs=c(5,15)

for (sigma in vs){
n=15 # Tamano muestral
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)
theta15=matrix(0,nrow=ns,ncol=2)

```

```

for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
theta15=theta
aux=theta15[,1]
auxe=(aux-mean(aux))/sd(aux)
plot(density(auxe),col='darkgreen',type='l',lwd=1,xlim=c(-4,4),
xlab="Valor_de_x",
ylab="Densidad_estandarizada",
main="Comparacion_de_las_distribuciones_al_aumentar_n")

x <- seq(-4, 4, length=100)
hx <- dnorm(x)

degf <- c(1, 3, 8, 30)
colors <- c("red", "blue", "darkgreen", "gold", "black")
labels <- c("df=1", "df=3", "df=8", "df=30", "normal")

lines(x, hx, type="l", lty=2, xlab="Valor_de_x",
ylab="Densidad_estandarizada", main="Comparacion_de_las_distribuciones_")

set.seed(123456)
n=30 # Tama no muestral
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)
theta30=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
aux=theta[,1]
auxe=(aux-mean(aux))/sd(aux)
lines(density(auxe))
theta30[,1]=theta[,1]
theta30[,2]=theta[,2]

```



```

set.seed(123456)
n=50 # Tama no muestral
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
aux=theta[,1]
auxe=(aux-mean(aux))/sd(aux)
lines(density(auxe),col=4,type='l')

set.seed(123456)
n=100 # Tama no muestral
x=seq(10,110,length=n)
theta=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rnorm(n,sd=sigma)
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
aux=theta[,1]
auxe=(aux-mean(aux))/sd(aux)
lines(density(auxe),col=6)

colors=c('darkgreen','black',4,6)
labels=c('n=15','n=30','n=50','n=100')
legend("topright", inset=.05, title="Distribuciones",
labels, lwd=2, lty=c(1, 1, 1, 1, 2), col=colors)

#####QQ PLOTS#####

aux=theta15[,2]
auxe=(aux-mean(aux))/sd(aux)
qqnorm(auxe,main='Q-QPlot del estimador de V con n=15',ylim=c(-4,10))
abline(a=0,b=1)

```

```

aux=theta30[,2]
auxe=(aux-mean(aux))/sd(aux)
qqnorm(auxe,main='Q-QPlot del estimador de V con n=30',ylim=c(-4,10))
abline(a=0,b=1)

aux=theta15[,1]
auxe=(aux-mean(aux))/sd(aux)
qqnorm(auxe,main='Q-QPlot del estimador de K con n=15',ylim=c(-4,10))
abline(a=0,b=1)

aux=theta30[,1]
auxe=(aux-mean(aux))/sd(aux)
qqnorm(auxe,main='Q-QPlot del estimador de K con n=30',ylim=c(-4,10))
abline(a=0,b=1)
}

```

## Distribución con errores no normales

```
##### REPRESENTACION DE LAS DENSIDADES #####
```

```

ns=100000
set.seed(123456)

```

```

n=15 # Tamano muestral
x=seq(10,110,length=n) # Diseno fijo
theta=matrix(0,nrow=ns,ncol=2) #Estimadores simulados
theta15=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rchisq(n,13)-13 #Michaelis
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
} #calcula por metodos numericos 10000 estimadores
theta15=theta
aux=theta[,2]
auxe=(aux-mean(aux))/sd(aux)
plot(density(auxe),col='darkgreen',type='l',lwd=1,xlim=c(-4,4),xlab="Val
ylab="Densidad estandarizada", main="Comparacion de las distribuciones al

```

```

x <- seq(-4, 4, length=100)
hx <- dnorm(x)

degf <- c(1, 3, 8, 30)
colors <- c("red", "blue", "darkgreen", "gold", "black")
labels <- c("df=1", "df=3", "df=8", "df=30", "normal")

lines(x, hx, type="l", lty=2, xlab="Valor de x",
ylab="Densidad estandarizada", main="Comparacion de las distribuciones al

set.seed(123456)
n=30 # Tamano muestral
x=seq(10,110,length=n) # Diseno fijo
theta=matrix(0,nrow=ns,ncol=2) #Estimadores simulados
theta30=matrix(0,nrow=ns,ncol=2)
for (is in 1:ns){
y=V*x/(K+x)+rchisq(n,13)-13 #Michaelis
m=nls(y~V*x/(K+x), start=list(K=20,V=120))
theta[is,]=coef(m)
} #calcula por metodos numericos 10000 estimadores
aux=theta[,2]
auxe=(aux-mean(aux))/sd(aux)
lines(density(auxe))

set.seed(123456)
n=50 # Tamano muestral
x=seq(10,110,length=n) # Diseno fijo
theta=matrix(0,nrow=ns,ncol=2) #Estimadores simulados
for (is in 1:ns){
y=V*x/(K+x)+rchisq(n,13)-13 #Michaelis
m=nls(y~V*x/(K+x), start=list(K=20,V=120))
theta[is,]=coef(m)
}
aux=theta[,2]

```

```
auxe=(aux-mean(aux))/sd(aux)
lines(density(aux),col=4,type='l')

set.seed(123456)
n=100 # Tamano muestral
x=seq(10,110,length=n) # Diseno fijo
theta=matrix(0,nrow=ns,ncol=2) #Estimadores simulados
for (is in 1:ns){
y=V*x/(K+x)+rchisq(n,13)-13#Michaelis
m=nls(y~V*x/(K+x),start=list(K=20,V=120))
theta[is,]=coef(m)
}
aux=theta[,2]
auxe=(aux-mean(aux))/sd(aux)
lines(density(aux),col=6)

colors=c('darkgreen','black',4,6)
labels=c('n=15','n=30','n=50','n=100')
legend("topright", inset=.05, title="Distribuciones",
labels, lwd=2, lty=c(1, 1, 1, 1, 2), col=colors)
```

# Bibliografía

- [1] Ritz, C. and Streibig J. C., *Nonlinear Regression with R*, Springer, New York, 2008.
- [2] Gallant, A. R., *Nonlinear Regression*, The American Statistician, Vol. 29, No. 2 (May 1975), pp. 73-81.
- [3] Bates, D.M. y Watts, D.G., *Nonlinear regression analysis and its applications* John Wiley & sons, 1988.
- [4] Jenrich, R.I., *Asymptotic properties of non-linear least squares estimators*, Ann. Math. Statist. 40 633-643.
- [5] Subba Rao, S., *Advanced Statistical Inference*, Department of Statistics Texas A&M, 2017.
- [6] Crujeiras Casais, R.M y Sánchez Sello, C.A., *Regresión Generalizada y Modelos Mixtos-Máster en técnicas estadísticas*, Departamento de estadística e investigación operativa USC, 2019-2020.