



UNIVERSITAT DE
BARCELONA

Hedge Funds: Inferencia del riesgo en un escenario real de estrés severo

Raül Martínez Buixeda

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Capítulo 2. Mixtura de dos distribuciones Normales: método de los momentos con estimadores muestrales no sesgados

1. Introducción

La mixtura de distribuciones Normales a partir de la metodología de los momentos fue introducida en 1894 por Karl Pearson. En su famoso trabajo "*Contributions to the Mathematical Theory of Evolution*", Pearson logra descomponer una distribución de probabilidad desconocida en la combinación lineal de dos Normales.

Los parámetros poblacionales que definen estas dos distribuciones Normales $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ y los pesos de su combinación lineal (p_1, p_2) , son estimados al resolver un sistema de ecuaciones basado en los cinco primeros momentos $(\hat{\mu}, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4, \hat{\mu}_5)$ de la muestra de la distribución analizada.

Expresión 1:

$$1 = p_1 + p_2$$

$$0 = \hat{\mu}_1 = p_1\theta_1 + p_2\theta_2$$

$$\hat{\mu}_2 = p_1(\theta_1^2 + \sigma_1^2) + p_2(\theta_2^2 + \sigma_2^2)$$

$$\hat{\mu}_3 = p_1(\theta_1^3 + 3\theta_1\sigma_1^2) + p_2(\theta_2^3 + 3\theta_2\sigma_2^2)$$

$$\hat{\mu}_4 = p_1(\theta_1^4 + 6\theta_1^2\sigma_1^2 + 3\sigma_1^4) + p_2(\theta_2^4 + 6\theta_2^2\sigma_2^2 + 3\sigma_2^4)$$

$$\hat{\mu}_5 = p_1(\theta_1^5 + 10\theta_1^3\sigma_1^2 + 15\theta_1\sigma_1^4) + p_2(\theta_2^5 + 10\theta_2^3\sigma_2^2 + 15\theta_2\sigma_2^4)$$

donde:

$$\theta_1 = \mu_1 - \hat{\mu}$$

$$\theta_2 = \mu_2 - \hat{\mu}$$

Para estimar los cinco primeros momentos, Pearson propuso dos opciones: "*analytically or graphically according to the accuracy or rapidity with which we wish to work*"¹⁷. Pero ambas metodologías de estimación presentan un problema con el sesgo, principalmente cuando el tamaño de la muestra es pequeño o la forma en la que la muestra se distribuye es "extrema".

¹⁷ Pearson (1894) p.78

Probablemente, el impacto del sesgo en el ejemplo¹⁸ numérico del trabajo de Pearson no fue significativo debido a la utilización de una muestra suficientemente grande “one thousand individuals from Napels” y a una distribución muestral bastante “común”. De todas formas, un enfoque más general, lo debería haber tenido en cuenta.

La cantidad de álgebra requerida para obtener y solucionar el polinomio de noveno grado generado a partir del sistema de ecuaciones, llevó a algunos autores a buscar procesos más sencillos de resolución. C. Radhakrishna Rao (1948) utilizó *cumulants* en vez de momentos, y A. Clifford Cohen (1967) desarrolló el sistema de ecuaciones inicial (expresión 1) para casuísticas específicas. En el primer caso, Rao igualó los *k statistics* de Fisher a los *cumulants*, evitando el problema del sesgo de la estimación muestral¹⁹. Pero en el segundo caso, Cohen no hizo ninguna referencia explícita al uso de estimadores muestrales insesgados²⁰. Algunos años más tarde, Richard E. Quandt (1975), tampoco hizo ninguna referencia al respecto²¹.

Lo mismo sucede con trabajos recientes que abarcan una casuística más amplia, por ejemplo, Bruce G. Lindsay and Prasanta Basak (1993) y Animashree Anandkumar, Daniel Hsu, and Sham M. Kakade (2012). En estos estudios, aunque el método de los momentos es utilizado en mixturas finitas de modelos multivariantes, las implicaciones del sesgo de los momentos muestrales no es tenida en cuenta.

¹⁸ Pearson (1894) p.85

¹⁹ Rao (1948) p.164: “Estimation by the method of moments.- The method of estimating the parameters by the method of the moments consists in equating the moments as calculated from the observations with functions of parameters representing the moments in the population. Since the expectations of calculated moments are not the same as the moments in the population this method might introduce a little bias in the estimating equations. This can be avoided by equating the calculated moments to their expected values Instead of this, one can choose the system of *k*-statistics of Fisher (defined in *Statistical Methods for Research Workers*) and equate them to their expectations which are the *cumulants* of the distribution.”

²⁰ Cohen (1967) p.17, 18: “Upon equating population moments to corresponding sample moments, it follows from (6) that $\alpha[\theta_1 - \bar{x}] + (1 - \alpha)[\theta_2 - \bar{x}] = 0$, and further that $\alpha m_1 + (1 - \alpha)m_2 = 0, \dots$ $\alpha[15\sigma_1^4 m_1 + 10\sigma_1^2 m_1^3 + m_1^5 - v_5] + (1 - \alpha)[15\sigma_2^4 m_2 + 10\sigma_2^2 m_2^3 + m_2^5 - v_5] = 0$ where \bar{x} is the sample mean and $v_i, (i = 2, 3, 4, \dots)$ is the *i*th central moment of the sample.”

²¹ Quandt (1975) p.3: “In case of two components the density function the parameters of which are to be estimated is

$$f(x; \lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{\lambda}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{1-\lambda}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \quad (2-1)$$

Equating the sample mean to the theoretical first moment of (2-1) and the second, third, fourth and fifth sample moments about the mean to the corresponding theoretical central moments we obtain five equations from which it is possible to solve for $\lambda, \mu_1, \mu_2, \sigma_1, \sigma_2$.”

2. Estimadores no sesgados de los momentos

Una estimación sesgada de los momentos de la muestra, puede distorsionar el resultado del proceso de cálculo de la construcción de la mixtura. Para evitar la distorsión, se hace imprescindible eliminar el sesgo.

2.1. Estimadores no sesgados de los cuatro primeros momentos

Las expresiones matemáticas de los estimadores no sesgados de la media y la varianza de una muestra de variables aleatorias IID, son populares en el ámbito de la estadística²²:

Expresión 2:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_i X_i$$

$$\hat{\sigma}^2 = \frac{n}{n-1} \sigma_{\bar{X}}^2$$

En cambio, las expresiones matemáticas de los estimadores no sesgados de la asimetría, \hat{S} , y la *kurtosis*, \hat{K} , son menos conocidas, y han requerido de una álgebra más elaborada por parte de sus desarrolladores²³:

Expresión 3:

$$\hat{S} = \frac{n^2}{(n-1)(n-2)} S_{\bar{X}}$$

$$\hat{K} = \frac{n}{(n-1)(n-2)(n-3)} \left((n^2 - 2n + 3) K_{\bar{X}} - 3(2n - 3) \sigma_{\bar{X}}^4 \right)^{24}$$

En general, en el campo de la estadística descriptiva, sólo los cuatro primeros momentos centrales son considerados relevantes. Pero para estimar los

²² Klemens (2008) p.9

²³ Cramer (1947) p.352; Abbasi (2008) p.693, 695; Ruiz Espejo et al. (2013) p.46, p.54, p.55, p.56

²⁴ Error en Klemens (2009) p.8. El cálculo de $\sigma_{\bar{X}}^4$ es incorrecto, lo cual lleva a una expresión incorrecta del estimador no sesgado de la kurtosis en p.9. La expresión correcta de $\sigma_{\bar{X}}^4$ debería ser:

$$\sigma_{\bar{X}}^4 = \frac{(n-1)^2}{n^2} \left(\frac{n(n-2)+3}{n(n-1)} \sigma^4 + \frac{1}{n} K \right).$$

Utilizando esta nueva fórmula, es posible obtener la expresión correcta del estimador no sesgado de la kurtosis.

parámetros de la mixtura de dos Normales, la estimación del quinto momento central también es necesaria. Es por ello, que el desarrollo del álgebra requerida para la construcción del estimador no sesgado del momento central de quinto orden es imprescindible.

2.2. Estimador no sesgado del quinto momento central

Dada una muestra de variables aleatorias IID, el quinto momento central respecto su promedio puede ser descompuesto de la siguiente forma:

Expresión 4:

$$F_{\bar{X}}(X_i) = E(X_i - \bar{X})^5 = E(X_i^5) - 5E(X_i^4\bar{X}) + 10E(X_i^3\bar{X}^2) - 10E(X_i^2\bar{X}^3) + 5E(X_i\bar{X}^4) - E(\bar{X}^5)$$

Tratando algebraicamente los sumandos dos, tres y cuatro obtenidos por la descomposición anterior para que sean función de los momentos ordinarios/centrales de las variables aleatorias que conforman la muestra, se pueden obtener las respectivas expresiones:

Expresión 5:

$$E(X_i^4\bar{X}) = \frac{1}{n}E\left(X_i^5 + X_i^4\sum_{i=2}^n X_i\right) = \dots = \frac{1}{n}E(X_i^5) + \frac{(n-1)}{n}E(X_i^4)\mu$$

Expresión 6:

$$\begin{aligned} E(X_i^3\bar{X}^2) &= \frac{1}{n^2}E\left(X_i^3\left(\binom{2}{2,0,\dots,0}\sum_i X_i^2 + \binom{2}{1,1,\dots,0}\sum_i X_i\sum_{j>i} X_j\right)\right) = \dots \\ &= \frac{1}{n^2}E(X_i^5) + \frac{(n-1)}{n^2}E(X_i^3)E(X_i^2) + \frac{2(n-1)}{n^2}E(X_i^4)\mu + \frac{(n^2-3n+2)}{n^2}E(X_i^3)\mu^2 \end{aligned}$$

Expresión 7:

$$\begin{aligned} E(X_i^2\bar{X}^3) &= \frac{1}{n^3}E\left(X_i^2\left(\binom{3}{3,0,\dots,0}\sum_i X_i^3 + \binom{3}{2,1,\dots,0}\sum_i X_i^2\sum_{j>i} X_j + \binom{3}{1,1,1,\dots,0}\sum_i X_i\sum_{j>i}\sum_{k>j} X_k\right)\right) = \dots \\ &= \frac{1}{n^3}E(X_i^5) + \frac{(n-1)}{n^3}E(X_i^2)E(X_i^3) + \frac{3(n-1)}{n^3}E(X_i^4)\mu + \frac{3(n-1)}{n^3}E(X_i^3)E(X_i^2) + \\ &+ \frac{3(n^2-n-2(n-1))}{n^3}E(X_i^2)^2\mu + \frac{3(n-1)(n-2)}{n^3}E(X_i^3)\mu^2 + \frac{(n-1)(n-2)(n-3)}{n^3}E(X_i^2)\mu^3 \end{aligned}$$

Si además se tiene en cuenta la siguiente igualdad:

Expresión 8:

$$E(X_i \bar{X}^4) = E(\bar{X}^5)$$

el quinto momento centrado en el promedio puede ser expresado como:

Expresión 9:

$$F_{\bar{X}}(X_i) = \frac{n^3 - 5n^2 + 10n - 10}{n^3} E(X_i^5) + \frac{-5n^3 + 25n^2 - 50n + 30}{n^3} E(X_i^4) \mu + \frac{10n^2 - 50n + 40}{n^3} E(X_i^3) E(X_i^2) + \\ + \frac{10n^3 - 60n^2 + 110n - 60}{n^3} E(X_i^3) \mu^2 - \frac{30n^2 - 90n + 60}{n^3} (E(X_i^2))^2 \mu - \frac{10(n-1)(n-2)(n-3)}{n^3} E(X_i^2) \mu^3 + \\ + 4E(\bar{X}^5)$$

El desarrollo individual del último sumando de la expresión anterior al incorporar la media de las variables aleatorias de la muestra, genera:

Expresión 10:

$$E(\bar{X}^5) = E\left(\left(\frac{\bar{X} - \mu}{n} + \mu\right)^5\right) = \dots = \frac{1}{n^4} R(X_i) + \frac{10(n-1)}{n^4} S(X_i) \sigma^2(X_i) + 5\mu \left(\frac{1}{n^3} (K(X_i) + 3(n-1)\sigma^4(X_i))\right) + \\ + 10\mu^2 \left(\frac{1}{n^2} S(X_i)\right) + 10\mu^3 \left(\frac{1}{n} \sigma^2(X_i)\right) + \mu^5$$

Introduciendo la expresión 10 en la expresión 9 y utilizando la descomposición tradicional de los cinco primeros momentos centrales, se obtiene:

Expresión 11:

$$F_{\bar{X}}(X_i) = \dots = \frac{n^4 - 5n^3 + 10n^2 - 10n + 4}{n^4} R(X_i) + \frac{10n^3 - 50n^2 + 80n - 40}{n^4} S(X_i) \sigma^2(X_i)$$

Finalmente, el valor esperado del producto de la *skewness* y la varianza:

Expresión 12:

$$S\sigma^2 = \left(\frac{n^4}{(n-1)(n-2)(n^2-5n+10)} S_{\bar{X}} \sigma_{\bar{X}}^2 - \frac{(n-1)}{(n^2-5n+10)} F \right)$$

permite obtener la expresión del estimador no sesgado del quinto (F) momento central:

Expresión 13:

$$\hat{F} = \frac{n^2}{(n-1)(n-2)(n-3)(n-4)} \left((n^2 - 5n + 10)F_{\bar{X}} - 10(n-2)S_{\bar{X}}\sigma_{\bar{X}}^2 \right).$$

Aunque el estimador no sesgado del quinto momento central puede ser obtenido aplicando la metodología de Dwyer²⁵ (h₅-statistics) o Heffernan²⁶ (u₅-statistics), esta expresión es mucho más compacta y permite poner de relieve algunas propiedades que resultan útiles.

3. Análisis de la sensibilidad de los estimadores no sesgados de los momentos centrales

Los estimadores no sesgados de los momentos segundo al quinto son sensibles al tamaño muestral. Cuanto mayor es la muestra, menor es su discrepancia con los valores poblacionales.

Expresión 14:

$$\hat{\mu} = \bar{X}$$

Expresión 15:

$$\hat{\sigma}^2 = \frac{n}{n-1} \sigma_{\bar{X}}^2$$

$$\left. \frac{n}{n-1} \right|_{n=2} = 2$$

$$\lim_{n \rightarrow \infty} \frac{n}{n-1} = 1 \rightarrow \lim_{n \rightarrow \infty} \hat{\sigma}^2 = \lim_{n \rightarrow \infty} \frac{n}{n-1} \sigma_{\bar{X}}^2 = \sigma^2.$$

²⁵ Dwyer (1937) p.27, 28:

$$h_5 = \frac{n(n^2-5n+10)(5)}{(n-1)^{(4)}} - 5 \frac{(n^2-5n+10)(4)(1)}{(n-1)^{(4)}} - 10 \frac{(n-2)(3)(2)}{(n-1)^{(4)}} + 10 \frac{(n^2-4n+8)(3)(1)^2}{n^{(5)}} + 15 \frac{(2n-4)(2)^2(1)}{n^{(5)}} - 10 \frac{(2)(1)^3}{(n-1)^{(4)}} + \frac{4(1)^5}{n^{(5)}}$$

²⁶ Abbasi (2008) p.691

Expresión 16:

$$\hat{S} = \frac{n^2}{(n-1)(n-2)} S_{\bar{X}}$$

$$\left. \frac{n^2}{(n-1)(n-2)} \right|_{n \rightarrow 2} = \infty$$

$$\left. \frac{n^2}{(n-1)(n-2)} \right|_{n \rightarrow 3} = 4,5$$

$$\lim_{n \rightarrow \infty} \frac{n^2}{(n-1)(n-2)} = 1 \rightarrow \lim_{n \rightarrow \infty} \hat{S} = \lim_{n \rightarrow \infty} \frac{n^2}{(n-1)(n-2)} S_{\bar{X}} = S.$$

Expresión 17:

$$\hat{K} = \frac{n}{(n-1)(n-2)(n-3)} \left((n^2 - 2n + 3)K_{\bar{X}} - 3(2n - 3)\sigma_{\bar{X}}^4 \right)$$

$$\left. \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \right|_{n \rightarrow 3} = \infty \quad \left. \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \right|_{n \rightarrow 3} = \infty$$

$$\left. \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \right|_{n \rightarrow 4} = 7,3333 \quad \left. \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \right|_{n \rightarrow 4} = 10$$

$$\lim_{n \rightarrow \infty} \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} = 1 \quad \lim_{n \rightarrow \infty} \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} = 0$$

$$\lim_{n \rightarrow \infty} \hat{K} = \lim_{n \rightarrow \infty} \frac{n}{(n-1)(n-2)(n-3)} \left((n^2 - 2n + 3)K_{\bar{X}} - 3(2n - 3)\sigma_{\bar{X}}^4 \right) = K$$

Expresión 18:

$$\hat{F} = \frac{n^2}{(n-1)(n-2)(n-3)(n-4)} \left((n^2 - 5n + 10)F_{\bar{X}} - 10(n-2)S_{\bar{X}}\sigma_{\bar{X}}^2 \right)$$

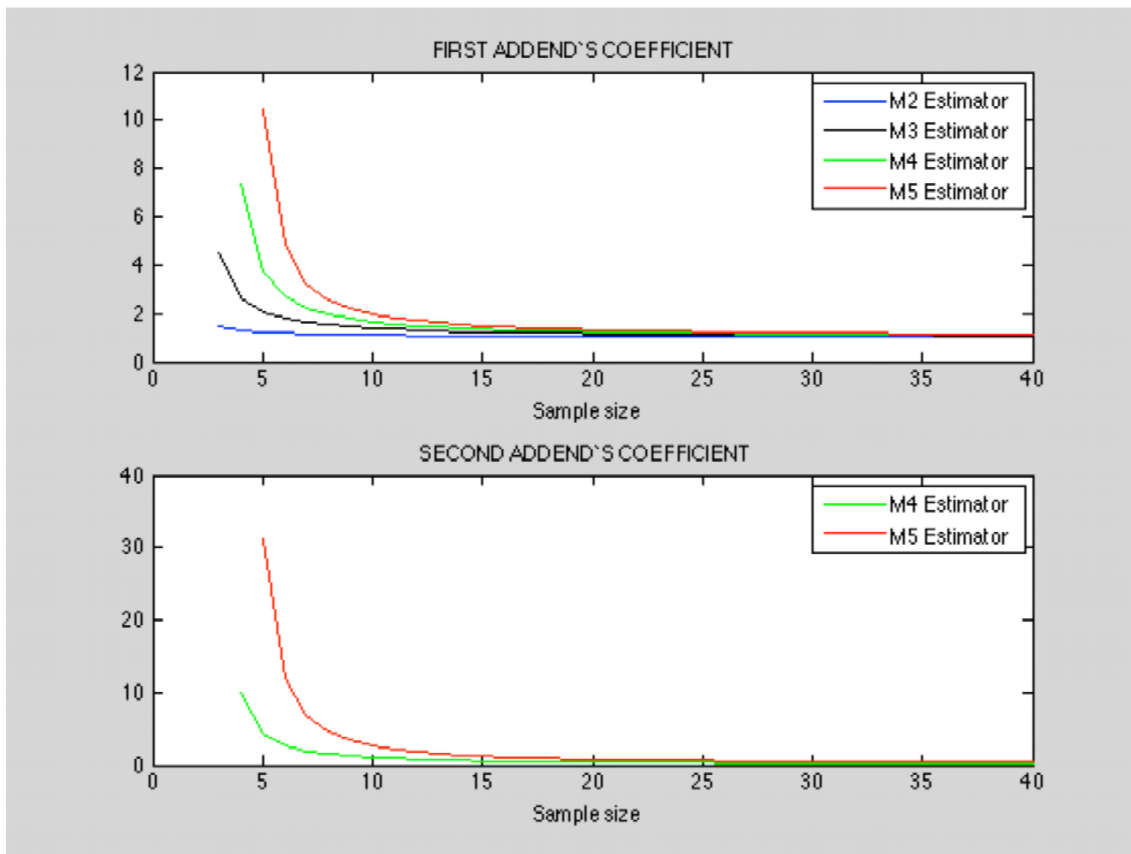
$$\left. \frac{n^2(n^2-5n+10)}{(n-1)(n-2)(n-3)(n-4)} \right|_{n \rightarrow 4} = \infty \quad \left. \frac{10n^2(n-2)}{(n-1)(n-2)(n-3)(n-4)} \right|_{n \rightarrow 4} = \infty$$

$$\left. \frac{n^2(n^2-5n+10)}{(n-1)(n-2)(n-3)(n-4)} \right|_{n \rightarrow 5} = 10,4167 \quad \left. \frac{10n^2(n-2)}{(n-1)(n-2)(n-3)(n-4)} \right|_{n \rightarrow 5} = 31,25$$

$$\lim_{n \rightarrow \infty} \frac{n^2(n^2-5n+10)}{(n-1)(n-2)(n-3)(n-4)} = 1 \quad \lim_{n \rightarrow \infty} \frac{10n^2(n-2)}{(n-1)(n-2)(n-3)(n-4)} = 0$$

$$\lim_{n \rightarrow \infty} \hat{F} = \lim_{n \rightarrow \infty} \frac{n^2}{(n-1)(n-2)(n-3)(n-4)} \left((n^2 - 5n + 10)F_{\bar{X}} - 10(n-2)S_{\bar{X}}\sigma_{\bar{X}}^2 \right) = F$$

Figura 1: Evolución de los coeficientes de los estimadores sesgados



3.1. Tamaño mínimo de la muestra

Los estimadores no sesgados de los momentos centrales de cuarto y quinto orden, \hat{K} y \hat{F} respectivamente, también están condicionados por estimadores sesgados de otros momentos centrales: $\sigma_{\bar{X}}$ y $S_{\bar{X}}$.

En el caso de la *kurtosis*, dado que:

Expresión 19

$$y_i = X_i - \bar{X}$$

$$\sum_{\substack{i \\ j>i}} (y_i^2 - y_j^2)^2 > 0 \rightarrow \frac{1}{n} \sum_i y_i^4 > \left(\frac{1}{n} \sum_i y_i^2 \right)^2 \rightarrow K_{\bar{X}} > \sigma_{\bar{X}}^4$$

sólo cuando el número de elementos es superior a 5:

Expresión 20

$$\frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \Big|_{n \rightarrow 5} = 3,75 \quad \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \Big|_{n \rightarrow 5} = 4,375$$

$$\frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \Big|_{n \rightarrow 6} = 2,7 \quad \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \Big|_{n \rightarrow 6} = 2,7$$

$$\frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \Big|_{n \rightarrow 7} = 2,2167 \quad \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \Big|_{n \rightarrow 7} = 1,9250$$

es posible asegurar la positividad del estimador no sesgado, \hat{K} . Para ser consistente con esta propiedad, sería deseable que el tamaño de la muestra fuera superior a 5.

En el caso del momento central de quinto orden, ninguna evidencia analítica puede extraerse del comportamiento de $F_{\bar{X}}$ con respecto $S_{\bar{X}}\sigma_{\bar{X}}^2$ (sólo cuando $n = 3$ y $n = 4$ puede demostrarse que $|F_{\bar{X}}| > |S_{\bar{X}}\sigma_{\bar{X}}^2|$). Pero al simular el comportamiento de $F_{\bar{X}}$ con respecto $S_{\bar{X}}\sigma_{\bar{X}}^2$, se puede comprobar que en la gran mayoría de los casos se cumple que: 1) el signo de $F_{\bar{X}}$ es el mismo que el de $S_{\bar{X}}\sigma_{\bar{X}}^2$, y 2) $|F_{\bar{X}}| > |S_{\bar{X}}\sigma_{\bar{X}}^2|$.

Si el número de elementos de la muestra es superior a 10:

Expresión 21

$$\frac{n^2(n^2-5n+10)}{(n-1)(n-2)(n-3)(n-4)} \Big|_{n \rightarrow 10} = 1,7091 \quad \frac{10n^2(n-2)}{(n-1)(n-2)(n-3)(n-4)} \Big|_{n \rightarrow 10} = 1,8182$$

$$\frac{n^2(n^2-5n+10)}{(n-1)(n-2)(n-3)(n-4)} \Big|_{n \rightarrow 11} = 1,6217 \quad \frac{10n^2(n-2)}{(n-1)(n-2)(n-3)(n-4)} \Big|_{n \rightarrow 11} = 1,5648$$

se evita que el tamaño muestral pueda ser decisivo en el signo de \hat{F} , y que este pueda estar determinado por $S_{\bar{X}}\sigma_{\bar{X}}^2$ en vez de por $F_{\bar{X}}$ aún cuando se cumplan las dos condiciones anteriores. Esta es una propiedad deseable, dado que es consistente con el comportamiento esperado de \hat{F} cuando el tamaño de la muestra aumenta.

3.2. Comportamiento del sesgo

Expresión 22:

$$Bias = \hat{K} - K_{\bar{X}} = \left(\frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} - 1 \right) K_{\bar{X}} - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \sigma_{\bar{X}}^4 = \frac{1}{(n-1)(n-2)(n-3)} \left((4n^2 - 8n + 6)K_{\bar{X}} - (6n^2 - 9n)\sigma_{\bar{X}}^4 \right)$$

donde:

$$4n^2 - 8n + 6 > 6n^2 - 9n, \text{ sólo si } n = 1.$$

Expresión 23:

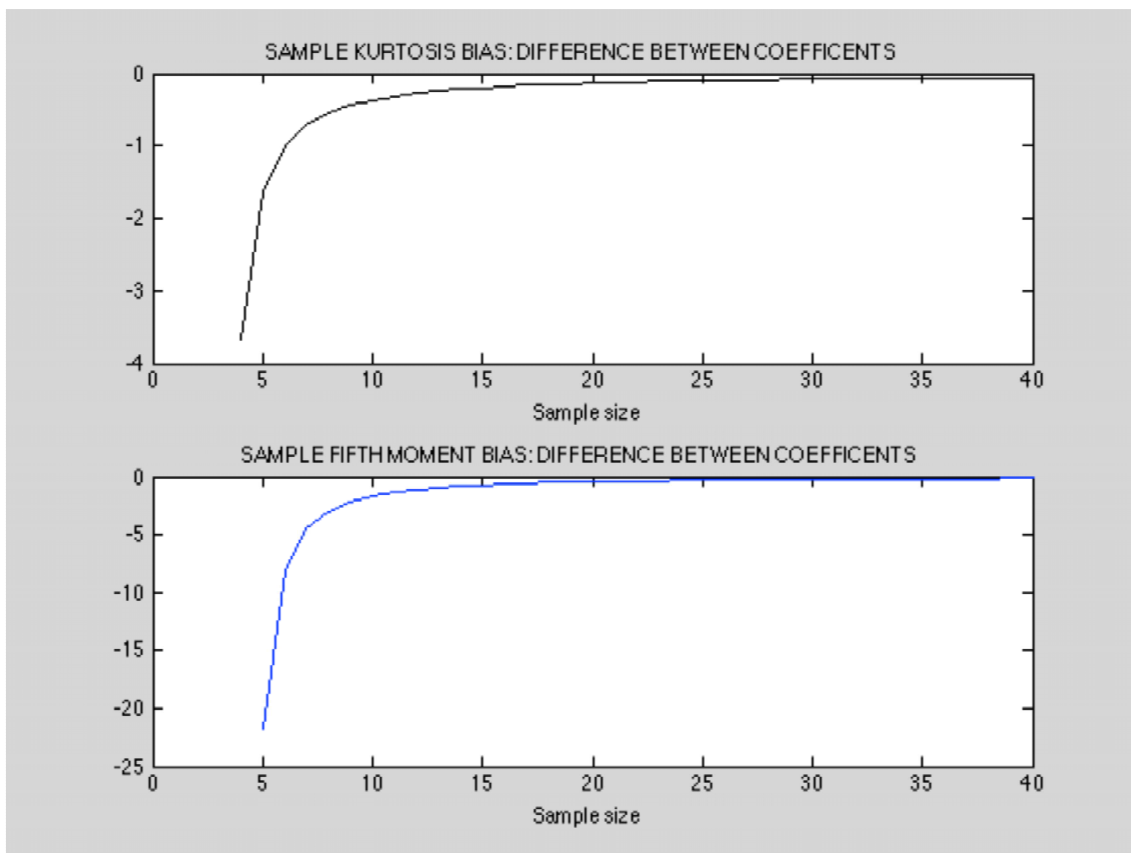
$$Bias = \hat{F} - F_{\bar{X}} = \left(\frac{n^2(n^2-5n+10)}{(n-1)(n-2)(n-3)(n-4)} - 1 \right) F_{\bar{X}} - \frac{10n^2(n-2)}{(n-1)(n-2)(n-3)(n-4)} \sigma_{\bar{X}}^4 =$$

$$\frac{1}{(n-1)(n-2)(n-3)(n-4)} \left((5n^3 - 25n^2 + 50n - 24) F_{\bar{X}} - (10n^3 - 20n^2) S_{\bar{X}} \sigma_{\bar{X}}^2 \right)$$

donde:

$$5n^3 - 25n^2 + 50n - 24 > 10n^3 - 20n^2, \text{ sólo si } n < 3$$

Figura 2: Evolución de $\frac{-2n^2+n+6}{(n-1)(n-2)(n-3)}$ y $\frac{-5n^3-5n^2+50n-24}{(n-1)(n-2)(n-3)(n-4)}$



Aunque los coeficientes de $K_{\bar{X}}$ y $F_{\bar{X}}$ son menores que los coeficientes de $\sigma_{\bar{X}}^4$ y $S_{\bar{X}} \sigma_{\bar{X}}^2$ respectivamente, el hecho que: 1) la diferencia de estos coeficientes tiende a cero muy rápidamente cuando la muestra aumenta, 2) $K_{\bar{X}} > \sigma_{\bar{X}}^4$ y 3) en general se cumple que el signo de $F_{\bar{X}}$ es el mismo que el de $S_{\bar{X}} \sigma_{\bar{X}}^2$ y $|F_{\bar{X}}| > |S_{\bar{X}} \sigma_{\bar{X}}^2|$, implica, que el sesgo es mayor cuando las muestras están distribuidas de forma "extrema".

4. Evidencia empírica de la transcendencia del sesgo

El impacto del sesgo puede ser muy significativo en la estimación de una mixtura de dos Normales. Para evidenciar este hecho, se generan, sobre las mismas muestras estresadas (el criterio de elección de las muestras se basa en el resultado del análisis de la sección anterior), dos mixturas de dos Normales. Una de las mixturas, es estimada a partir de estimadores sesgados de los momentos centrales de la muestra (mixtura sesgada), y la otra a partir de estimadores no sesgados (mixtura no sesgada). Una vez el proceso de estimación ha finalizado, se analiza el grado de similitud entre las mixturas.

4.1. Particularidades de la metodología de cálculo

Para estimar los parámetros de las mixturas, se utiliza el algoritmo 'Trust-region-dogleg' propuesto por Matlab para resolver sistemas de ecuaciones no lineales²⁷. Los valores por defecto de los principales inputs del algoritmo han sido modificados para aumentar la precisión del proceso.

Script 1

```
optimoptions('fsolve','Diagnostics','on','Display','Iter','MaxFunEvals',10000,'MaxIter',1000,'TolFun',1e-8,'TolX',1e-8);
```

4.2. Restricciones en las soluciones del sistema de ecuaciones

A pesar de que el sistema de ecuaciones inicial (expresión 1) no presenta ningún tipo de restricción sobre el dominio de los pesos de las Normales, p_1 y p_2 , la programación fuerza el proceso iterativo a la restricción $p_1, p_2 \in [-1, 1]$ para poder condicionar posteriormente las soluciones a la combinación lineal convexa entre p_1 y p_2 .

Expresión 24:

$$1 = |p_1| + |p_2|$$

$$0 = \hat{\mu}_1 = |p_1|\theta_1 + |p_2|\theta_2$$

$$\hat{\mu}_2 = |p_1|(\theta_1^2 + \sigma_1^2) + |p_2|(\theta_2^2 + \sigma_2^2)$$

$$\hat{\mu}_3 = |p_1|(\theta_1^3 + 3\theta_1\sigma_1^2) + |p_2|(\theta_2^3 + 3\theta_2\sigma_2^2)$$

$$\hat{\mu}_4 = |p_1|(\theta_1^4 + 6\theta_1^2\sigma_1^2 + 3\sigma_1^4) + |p_2|(\theta_2^4 + 6\theta_2^2\sigma_2^2 + 3\sigma_2^4)$$

$$\hat{\mu}_5 = |p_1|(\theta_1^5 + 10\theta_1^3\sigma_1^2 + 15\theta_1\sigma_1^4) + |p_2|(\theta_2^5 + 10\theta_2^3\sigma_2^2 + 15\theta_2\sigma_2^4)$$

²⁷ Mathworks: "trust-region-dogleg' is the only algorithm that is specially designed to solve nonlinear equations. The others attempt to minimize the sum of squares of the function."

Además, para evitar incoherencias de carácter conceptual, las soluciones obtenidas como estimación de las volatilidades de las dos Normales son positivizadas a posteriori.

Expresión 25:

$$\sigma_1 = |\sigma_1| \text{ y } \sigma_2 = |\sigma_2|.$$

4.3. Valores iniciales

La obtención de soluciones eficientes del sistema de ecuaciones, implica iniciar el proceso de cálculo desde diferentes orígenes. La generación de un rango suficientemente amplio de valores iniciales para los parámetros de las dos Normales se programa en base a bucles anidados, generando una matriz de cuatro dimensiones con el output del proceso.

Script 2

```
cont1 = 0;
for mu1 = -30:10:30
    cont2 = 0;
    for mu2 = -9:2:9
        cont2 = cont2+1;
        cont3 = 0;
        for vola1 = 1:3:10
            cont3 = cont3+1;
            cont4 = 0;
            for vola2 = 1:3:10
                cont4 = cont4+1;

                % . . .
                % PROCESO DE ESTIMACIÓN DE LAS MIXTURAS
                % . . .

            end
        end
    end
    cont1 = cont1+1;
end
```

4.4. Contraste de similitud entre mixturas

El grado de similitud de las mixturas obtenidas se determina a partir del p-value del test de Kolmogorov-Smirnov.

El test de Kolmogorov-Smirnov se aplica a las mixturas sesgadas y no sesgadas estimadas de la misma muestra estresada, usando cien valores distribuidos uniformemente a lo largo del rango:

Expresión 26:

$$[\min\{(-2\sigma + \mu)_{i=1,2,3,4}\}, \max\{(2\sigma + \mu)_{i=1,2,3,4}\})];$$

donde σ_i y μ_i son los valores estimados de cada una de las cuatro distribuciones Normales $i = 1,2,3,4$ asociadas a cada análisis de similitud.

Script 3

```
M = [2*volal+mu1 2*vola2+mu2 2*vola3+mu3 2*vola4+mu4];
M_AUX = max(M);
m = [-2*volal+mu1 -2*vola2+mu2 -2*vola3+mu3 -2*vola4+mu4];
m_AUX = min(m);
n = (M_AUX-m_AUX)/99;
x = m_AUX:n:M_AUX;
% . . .
% KOLMOGOROV-SMIRNOV
% . . .
```

4.5. Resultado obtenido para una muestra estresada de 40 observaciones

Tabla 1: Datos de la muestra

-0,599272	-0,058959	-0,853541	-1,853008	-0,207303
0,270378	-0,652771	0,477227	-0,071320	-0,938301
0,161364	-0,268183	-0,409873	-0,711323	0,061445
-1,846129	-0,398333	-0,543548	-2,735696	-5,958096
-9,202814	1,621899	2,927523	-0,470611	0,833398
1,918552	-0,242934	-1,622610	-3,787695	-7,331272
-2,968688	-2,486508	0,153496	-0,188180	1,346763
-1,089775	0,950106	1,324278	2,093174	-0,232220

Tabla 2: Estimaciones sesgadas y no sesgadas de los primeros cinco momentos de la muestra

	BIASED	UNBIASED	Absolute Difference	Relative Difference
Mean	-0,839734	-0,839734	0,000000	0,0000%
Variance	5,550465	5,692784	0,142320	2,5641%
Skewness	-22,183907	-23,950237	-1,766330	7,9622%
Kurtosis	197,856477	214,625120	16,768643	8,4752%
Fifth Moment	-1372,297608	-1530,396673	-158,099065	11,5208%

Figura 3: Primera solución del proceso de estimación de las mixturas

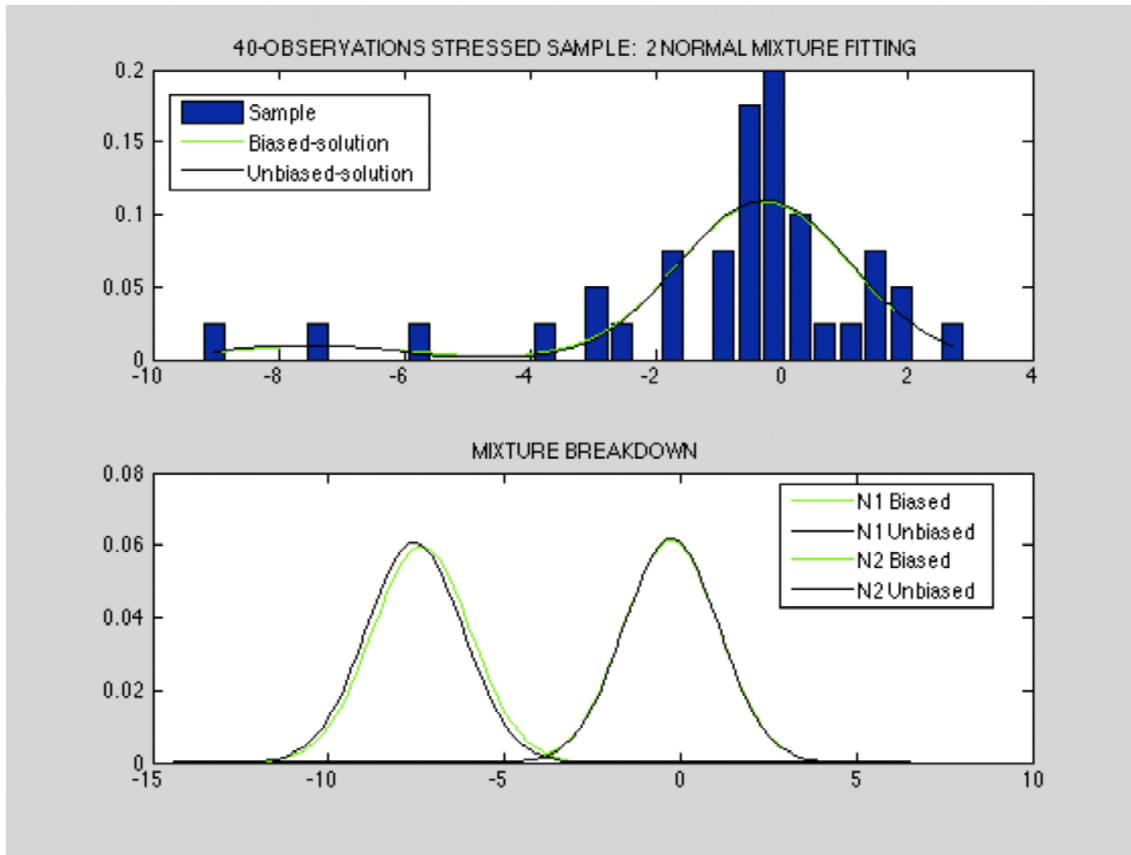


Tabla 3: Primera solución del proceso de estimación de la mixtura sesgada

	S_1	S_{58}	S_{121}	S_{126}
H_1	-10	-10	0	0
H_2	0	-2	-4	-2
H_3	1	2	2	3
H_4	1	3	2	3
Flag	1	1	3	1
P_1	8,0194%	8,0194%	91,9806%	91,9806%
P_2	91,9806%	91,9806%	8,0194%	8,0194%
$Mean_1$	-7,357479	-7,357479	-0,271480	-0,271480
$Mean_2$	-0,271480	-0,271480	-7,357479	-7,357479
$Std\ dev_1$	1,390639	1,390639	1,356146	1,356146
$Std\ dev_2$	1,356146	1,356146	1,390639	1,390639
FV_1	0	0	0	0
FV_2	0,00E+00	0,00E+00	0,00E+00	0,00E+00
FV_3	-8,88E-16	8,88E-16	0,00E+00	8,88E-16
FV_4	-7,11E-15	3,55E-15	3,55E-15	0,00E+00
FV_5	1,99E-13	-5,68E-14	-2,84E-14	0,00E+00
FV_6	-4,54747E-13	-4,54747E-13	6,82121E-13	4,54747E-13
Flag	1	1	1	1

	S_1	S_{58}	...	S_{121}	S_{126}
95th percentil	-6,919926	-6,919926	...	-6,919926	-6,919926
FV₁	-9,90E-10	-9,89E-10	...	-9,90E-10	-9,90E-10
Flag	1	1	...	1	1
90th percentil	-3,013017	-3,013017	...	-3,013017	-3,013017
FV₁	6,90E-10	6,90E-10	...	6,90E-10	6,90E-10
Flag	1	1	...	1	1
75th percentil	-1,489202	-1,489202	...	-1,489202	-1,489202
FV₁	9,06E-11	9,07E-11	...	9,06E-11	9,05E-11
Flag	1	1	...	1	1
50th percentil	-0,419963	-0,419963	...	-0,419963	-0,419963
FV₁	8,88E-16	8,88E-16	...	7,77E-16	-1,28E-15

$S_i - S_j$: Soluciones idénticas del sistema de ecuaciones, a partir de puntos iniciales diferentes en el proceso de cálculo.

H_1 : Valor inicial de la media₁.

H_2 : Valor inicial de la media₂.

H_3 : Valor inicial de la desviación estándar₁.

H_4 : Valor inicial de la desviación estándar₂.

Flag: Indicador de la "bondad" del proceso de cálculo.

P_1 : Peso de la primera distribución Normal.

P_2 : Peso de la segunda distribución Normal.

Mean₁: Media de la primera distribución Normal.

Mean₂: Media de la segunda distribución Normal.

Std deviation₁: Volatilidad de la primera distribución Normal.

Std deviation₂: Volatilidad de la segunda distribución Normal.

FV_i: Valor del i^{th} término constante del sistema de ecuaciones cuando el proceso de cálculo ha finalizado.

Tabla 4: Primera solución del proceso de estimación de la mixtura no sesgada

	S_1	S_{54}	...	S_{118}	S_{124}
H_1	-10	-20	...	0	0
H_2	-2	-2	...	-4	-2
H_3	1	2	...	3	3
H_4	3	3	...	1	3
Flag	1	1	...	1	1
P_1	7,9347%	7,9347%	...	92,0653%	92,0653%
P_2	92,0653%	92,0653%	...	7,9347%	7,9347%
Mean₁	-7,554624	-7,554624	...	-0,261007	-0,261007
Mean₂	-0,261007	-0,261007	...	-7,554624	-7,554624
Std dev₁	1,366354	1,366354	...	1,342202	1,342202
Std dev₂	1,342202	1,342202	...	1,366354	1,366354
FV₁	0	0	...	-1,11022E-16	0
FV₂	-1,11E-16	0,00E+00	...	0,00E+00	1,11E-16
FV₃	-8,88E-16	1,78E-15	...	0,00E+00	0,00E+00

	S_1	S_{54}	...	S_{118}	S_{124}
FV_4	3,55E-15	-1,07E-14	...	-3,55E-15	0,00E+00
FV_5	2,84E-14	0,00E+00	...	-2,84E-14	-5,68E-14
FV_6	4,54747E-13	-4,55E-13	...	-2,27374E-13	2,27374E-13
Flag	1	1	...	1	1
95 th percentil	-7,100685	-7,100685	...	-7,100685	-7,100685
FV_1	-5,41E-09	-5,41E-09	...	-5,41E-09	-5,41E-09
Flag	1	1	...	1	1
90 th percentil	-2,952523	-2,952523	...	-2,952523	-2,952523
FV_1	3,44E-09	3,44E-09	...	3,44E-09	3,44E-09
Flag	1	1	...	1	1
75 th percentil	-1,462440	-1,462440	...	-1,462440	-1,462440
FV_1	1,43E-08	1,43E-08	...	1,43E-08	1,43E-08
Flag	1	1	...	1	1
50 th percentil	-0,406271	-0,406271	...	-0,406271	-0,406271
FV_1	1,11E-16	2,22E-15	...	2,33E-15	0,00E+00

Tabla 5: Test de Kolmogorov-Smirnov aplicado a la primera solución de mixturas sesgadas y no sesgadas

	K-S
$H_{5\%}$	false
p-value	99,2102%
n	100

$H_{5\%}$: Contraste de la hipótesis de Kolmogorov-Smirnov con un p-value del 5%.

Figura 4: Segunda solución del proceso de estimación de las mixturas

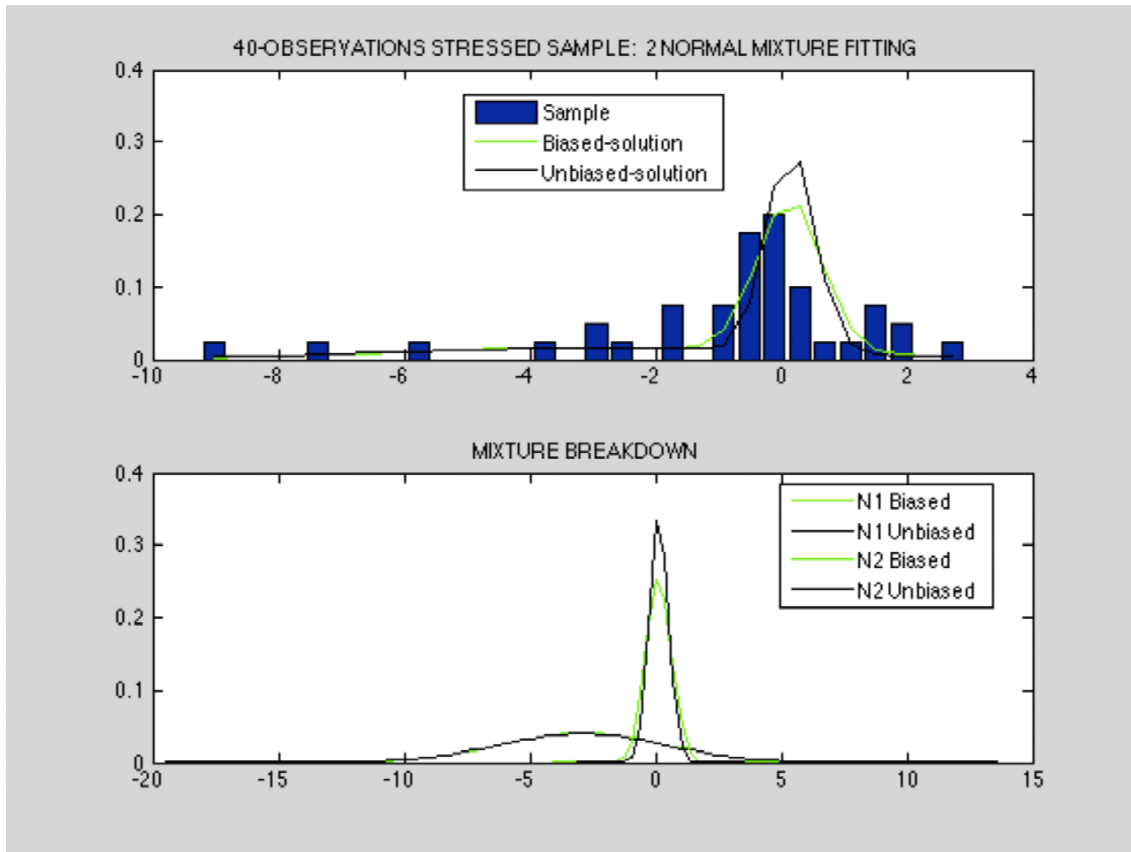


Tabla 6: Segunda solución del proceso de estimación de la mixtura sesgada

	S_{59}	S_{99}	S_{100}	S_{120}
H_1	0	0	0	0
H_2	0	0	0	-4
H_3	4	3	1	1
H_4	1	2	3	3
Flag	1	1	1	1
P_1	32,4714%	32,4714%	67,5286%	67,5286%
P_2	67,5286%	67,5286%	32,4714%	32,4714%
$Mean_1$	-2,860357	-2,860357	0,131889	0,131889
$Mean_2$	0,131889	0,131889	-2,860357	-2,860357
$Std\ dev_1$	3,239018	3,239018	0,517052	0,517052
$Std\ dev_2$	0,517052	0,517052	3,239018	3,239018
FV_1	0	0	0	0
FV_2	0,00E+00	0,00E+00	3,33E-16	-1,11E-16
FV_3	-8,88E-16	4,44E-15	6,39E-14	0,00E+00
FV_4	7,11E-15	0,00E+00	1,49E-13	-7,11E-15
FV_5	-1,14E-13	5,68E-14	7,67E-13	0,00E+00
FV_6	2,27374E-13	1,36424E-12	2,04636E-12	4,54747E-13
Flag	1	1	1	1

	S_{59}	S_{99}	S_{100}	S_{120}
95th percentil	-6,162547	-6,162547	-6,162547	-6,162547
FV₁	5,85E-08	5,85E-08	5,85E-08	5,85E-08
Flag	1	1	1	1
90th percentil	-4,485148	-4,485148	-4,485148	-4,485148
FV₁	6,92E-12	6,94E-12	6,94E-12	6,94E-12
Flag	1	1	1	1
75th percentil	-0,917078	-0,917078	-0,917078	-0,917078
FV₁	2,53E-10	2,53E-10	2,53E-10	2,53E-10
Flag	1	1	1	1
50th percentil	-0,063406	-0,063406	-0,063406	-0,063406
FV₁	9,47E-09	9,47E-09	9,47E-09	9,47E-09

Tabla 7: Segunda solución del proceso de estimación de la mixtura no sesgada

	S_{55}	S_{94}	S_{95}	S_{117}
H₁	-30	-30	0	0
H₂	0	0	-4	-2
H₃	2	2	1	3
H₄	1	2	4	4
Flag	1	1	1	1
P₁	32,3226%	32,3226%	67,6774%	67,6774%
P₂	67,6774%	67,6774%	32,3226%	32,3226%
Mean₁	-2,916120	-2,916120	0,151944	0,151944
Mean₂	0,151944	0,151944	-2,916120	-2,916120
Std dev₁	3,306558	3,306558	0,383900	0,383900
Std dev₂	0,383900	0,383900	3,306558	3,306558
FV₁	0	0	0	2,22045E-16
FV₂	-1,11E-16	0,00E+00	1,11E-16	1,11E-16
FV₃	8,88E-16	1,40E-13	5,33E-14	0,00E+00
FV₄	3,55E-14	3,66E-13	1,24E-13	0,00E+00
FV₅	8,53E-14	1,73E-12	3,69E-13	-5,68E-14
FV₆	-9,09E-13	2,04636E-12	1,81899E-12	0
Flag	1	1	1	1
95th percentil	-6,277290	-6,277290	-6,277290	-6,277290
FV₁	4,21E-07	4,21E-07	4,21E-07	4,21E-07
Flag	1	1	1	1
90th percentil	-4,561482	-4,561482	-4,561482	-4,561482
FV₁	1,34E-12	1,35E-12	1,34E-12	1,34E-12
Flag	1	1	1	1
75th percentil	-0,710508	-0,710508	-0,710508	-0,710508
FV₁	6,04E-09	6,03E-09	6,03E-09	6,03E-09
Flag	1	1	1	1
50th percentil	0,005281	0,005281	0,005281	0,005281

	S_{55}	S_{94}	S_{95}	S_{117}
FV_1	3,77E-15	3,66E-15	3,66E-15	4,44E-15

Tabla 8: Test de Kolmogorov-Smirnov aplicado a la segunda solución de mixturas sesgadas y no sesgadas

	K-S
$H_{5\%}$	false
<i>p-value</i>	9,9376%
<i>n</i>	100

4.6. Resultado obtenido para una muestra estresada de 11 observaciones

Tabla 9: Datos de la muestra

-1,393273	1,523269	1,798494	-0,116884	-0,960589
2,452549	1,470478	2,295753	-15,334837	-4,440916
0,000000				

Tabla 10: Estimaciones sesgadas y no sesgadas de los primeros cinco momentos de la muestra

	BIASED	UNBIASED	Absolute Difference	Relative Difference
Mean	-1,155087	-1,155087	0,000000	0,0000%
Variance	23,825709	26,208280	2,382571	10,0000%
Skewness	-248,431992	-334,003012	-85,571020	34,4444%
Kurtosis	3730,278218	5318,675807	1588,397588	42,5812%
Fifth Moment	-52003,675425	-82096,655792	-30092,980367	57,8670%

Figura 5: Solución del proceso de estimación de las mixturas

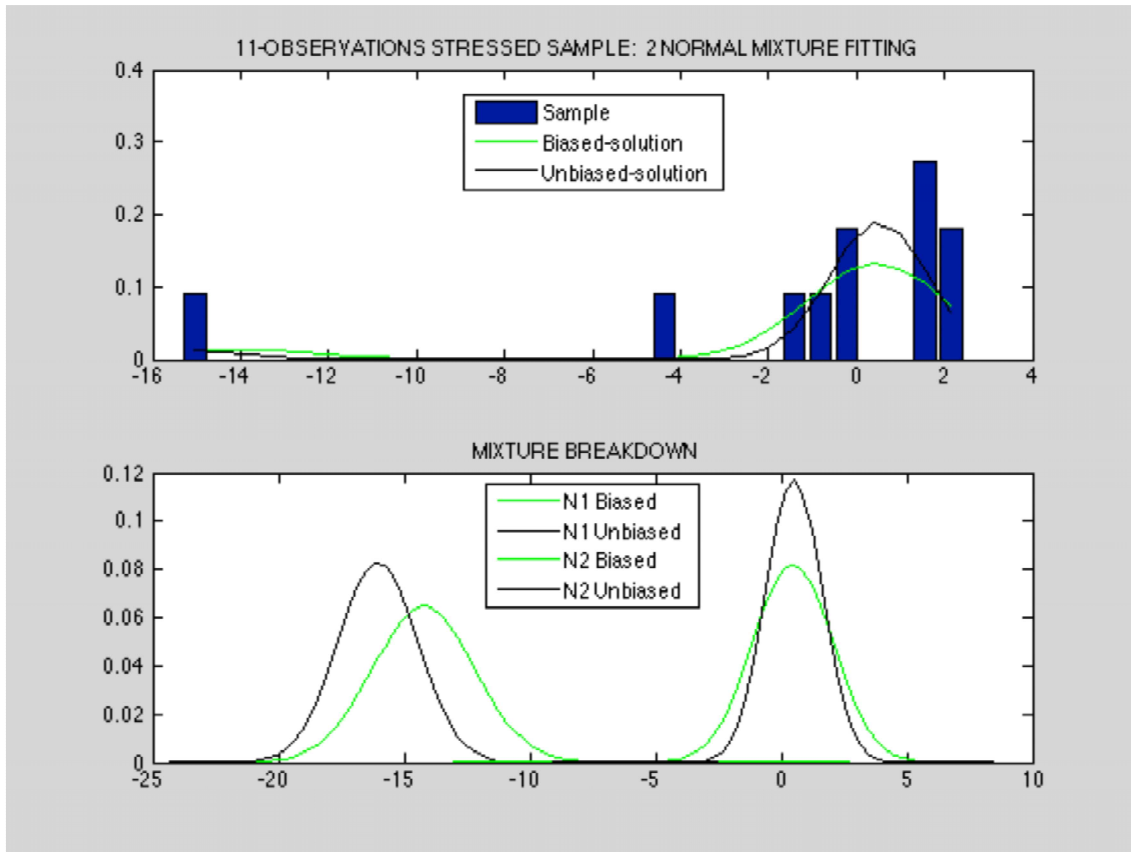


Tabla 11: Solución del proceso de estimación de la mixtura sesgada

	S_1	S_{168}
H_1	-30	0
H_2	-2	-4
H_3	1	4
H_4	2	2
Flag	3	3
P_1	11,0217%	88,9783%
P_2	88,9783%	11,0217%
$Mean_1$	-14,214781	0,462606
$Mean_2$	0,462606	-14,214781
$Std dev_1$	2,019916	1,589995
$Std dev_2$	1,589995	2,019916
FV_1	0	0
FV_2	2,22E-16	2,22E-16
FV_3	0,00E+00	3,55E-15
FV_4	8,53E-14	5,68E-14
FV_5	1,36E-12	-4,55E-13
FV_6	0	7,28E-12
Flag	1	1

	S_1	S_{168}
95th percentil	-14,449980	-14,449980
FV₁	2,75E-12	2,75E-12
Flag	1	1
90th percentil	-11,539747	-11,539747
FV₁	-6,74E-08	-6,74E-08
Flag	1	1
75th percentil	-1,137653	-1,137653
FV₁	2,98E-08	2,98E-08
Flag	1	1
50th percentil	0,214764	0,214764
FV₁	1,06E-08	1,06E-08

Tabla 12: Solución del proceso de estimación de la mixtura no sesgada

	S_1	S_{163}
H₁	-20	0
H₂	-4	0
H₃	4	2
H₄	2	3
Flag	1	3
P₁	10,0729%	89,9271%
P₂	89,9271%	10,0729%
Mean₁	-16,044701	0,512734
Mean₂	0,512734	-16,044701
Std dev₁	1,583345	1,117266
Std dev₂	1,117266	1,583345
FV₁	0	0
FV₂	2,22E-16	-2,22E-16
FV₃	-2,13E-14	0
FV₄	2,27E-13	5,68E-14
FV₅	0,00E+00	9,09E-13
FV₆	0	-1,46E-11
Flag	1	1
95th percentil	-16,059071	-16,059071
FV₁	1,23E-11	1,23E-11
Flag	1	1
90th percentil	-12,173370	-12,173370
FV₁	-5,51E-08	-5,51E-08
Flag	1	1
75th percentil	-0,571160	-0,571160
FV₁	6,78E-11	6,79E-11
Flag	1	1
50th percentil	0,355366	0,355366

	S₁	S₁₆₃
FV₁	4,80E-11	4,77E-11

Tabla 13: Test de Kolmogorov-Smirnov aplicado a solución de mixturas sesgadas y no sesgadas

	K-S
H_{5%}	true
p-value	0,0018%
n	100

4.7. Analisis de los resultados

Los resultados empíricos de la sección anterior, muestran evidencia de la transcendencia del sesgo en la estimación de la mixtura de dos distribuciones Normales. Aunque se utilicen las mismas muestras y metodología en el proceso de cálculo, tanto los gráficos, como los p-value de los test de Kolmogorov-Smirnov, como las cifras de los percentiles, muestran un alto grado de disimilitud entre las distribuciones de las mixturas si el sesgo de los estimadores de los momentos centrales de las muestras es tenido, o no tenido, en consideración.

En el primer análisis empírico (apartado 4.5.), la estimación de las mixturas utilizando una muestra estresada de 40 observaciones, es determinada por los siguientes vectores:

$$\overrightarrow{MIX_1^{biased}} = \begin{pmatrix} p_1 = 8,0194\% \\ p_2 = 91,9806\% \\ \mu_1 = -7,357479 \\ \mu_2 = -0,271480 \\ \sigma_1 = 1,390639 \\ \sigma_2 = 1,356146 \end{pmatrix} \quad \overrightarrow{MIX_1^{unbiased}} = \begin{pmatrix} p_1 = 7,9347\% \\ p_2 = 92,0653\% \\ \mu_1 = -7,554624 \\ \mu_2 = -0,261007 \\ \sigma_1 = 1,366354 \\ \sigma_2 = 1,342202 \end{pmatrix}$$

$$\overrightarrow{MIX_2^{biased}} = \begin{pmatrix} p_1 = 32,4714\% \\ p_2 = 67,5286\% \\ \mu_1 = -2,860357 \\ \mu_2 = 0,131889 \\ \sigma_1 = 3,239018 \\ \sigma_2 = 0,517052 \end{pmatrix} \quad \overrightarrow{MIX_2^{unbiased}} = \begin{pmatrix} p_1 = 32,3226\% \\ p_2 = 67,6774\% \\ \mu_1 = -2,916120 \\ \mu_2 = 0,151944 \\ \sigma_1 = 3,306558 \\ \sigma_2 = 0,383900 \end{pmatrix},$$

donde los componentes de MIX_1 and MIX_2 son los parámetros estimados de dos mixturas obtenidas como solución del sistema de ecuaciones inicial (*expresión 1*). El superíndice “sesgado” o “no sesgado” se refiere a la mixtura sesgada o a la mixtura no sesgada. Las primeras dos componentes de cada vector son los pesos estimados de la relación lineal, y los siguientes cuatro son las medias estimadas y

las desviaciones estándar estimadas de las dos distribuciones Normales, respectivamente.

El p-value de 99,2102% del test de Kolmogorov-Smirnov obtenido de las dos primeras mixturas estimadas comparables, MIX_1^{biased} and $MIX_1^{unbiased}$, presentan un alto grado de similitud.

Los percentiles 95th, 90th, 75th y 50th son también muy similares:

$$MIX_1^{biased} \text{ pcl} = \begin{pmatrix} 95^{th} = -6,919926 \\ 90^{th} = -3,013017 \\ 75^{th} = -1,489202 \\ 50^{th} = -0,419963 \end{pmatrix} \quad MIX_1^{unbiased} \text{ pcl} = \begin{pmatrix} 95^{th} = -7,100685 \\ 90^{th} = -2,952523 \\ 75^{th} = -1,462440 \\ 50^{th} = -0,406271 \end{pmatrix}.$$

En cambio, si se consideran las segundas mixturas estimadas comparables, MIX_2^{biased} and $MIX_2^{unbiased}$, el p-value obtenido del test de Kolmogorov-Smirnov es de 9,9376%. Esta cifra tan baja, muestra un alto grado de disimilitud.

Los percentiles 95th, 90th, 75th y 50th también presentan mayor discrepancia, especialmente cerca del centro de la distribución:

$$MIX_2^{biased} \text{ pcl} = \begin{pmatrix} 95^{th} = -6,162547 \\ 90^{th} = -4,485148 \\ 75^{th} = -0,917078 \\ 50^{th} = -0,063406 \end{pmatrix} \quad MIX_2^{unbiased} \text{ pcl} = \begin{pmatrix} 95^{th} = -6,277290 \\ 90^{th} = -4,561482 \\ 75^{th} = -0,710508 \\ 50^{th} = 0,005281 \end{pmatrix}.$$

Así pues, el alto grado de similitud entre las primeras soluciones comparables, MIX_1^{biased} and $MIX_1^{unbiased}$, permite considerarlas similares / iguales en términos estadísticos. Por el contrario, el importante grado de disimilitud entre las segundas soluciones comparables, MIX_2^{biased} and $MIX_2^{unbiased}$, impide poder repetir la misma afirmación, al menos de forma tan contundente.

Si cabe, una evidencia de disimilitud empírica más significativa, puede observarse en el segundo análisis (apartado 4.6.). Utilizando una muestra estresada de 11 observaciones, las discrepancias entre los estimadores sesgados y no sesgados de los momentos centrales son mayores que en el primer análisis.

Las mixturas obtenidas a partir de la muestra estresada de 11 observaciones, está determinada por los siguientes vectores solución:

$$\overrightarrow{MIX}^{biased} = \begin{pmatrix} p_1 = 11,0217\% \\ p_2 = 88,9783\% \\ \mu_1 = -14,214781 \\ \mu_2 = 0,462606 \\ \sigma_1 = 2,019916 \\ \sigma_2 = 1,589995 \end{pmatrix} \quad \overrightarrow{MIX}^{unbiased} = \begin{pmatrix} p_1 = 10,0729\% \\ p_2 = 89,9271\% \\ \mu_1 = -16,044701 \\ \mu_2 = 0,512734 \\ \sigma_1 = 1,583345 \\ \sigma_2 = 1,117266 \end{pmatrix},$$

y rechazan la hipótesis nula del test de Kolmogorov-Smirnov con un p-value de 0,0018%. Los percentiles 95th, 90th, 75th y 50th muestran una considerable mayor diferencia entre mixturas que en el análisis anterior:

$$MIX^{biased} pcl = \begin{pmatrix} 95^{th} = -14,449980 \\ 90^{th} = -11,539747 \\ 75^{th} = -1,137653 \\ 50^{th} = 0,214764 \end{pmatrix} MIX^{unbiased} pcl = \begin{pmatrix} 95^{th} = -16,059071 \\ 90^{th} = -12,173370 \\ 75^{th} = -0,571160 \\ 50^{th} = 0,355366 \end{pmatrix}.$$

Por lo tanto, en este caso, el alto grado de disimilitud entre soluciones permite considerarlas diferentes en términos estadísticos.