

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa I



**APLICACIÓN DE MÉTODOS NUMÉRICOS EN
INFERENCIA BAYESIANA: IMPLEMENTACIÓN DE UN
MÉTODO BAYESIANO ROBUSTO**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Javier Portela García - Miguel

Bajo la dirección del doctor

Miguel Ángel Gómez - Villegas

Madrid, 2003

ISBN: 84-669-1805-1

**UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS**

Departamento de Estadística e I. Operativa I

**APLICACIÓN DE MÉTODOS NUMÉRICOS EN
INFERENCIA BAYESIANA: IMPLEMENTACIÓN
DE UN MÉTODO BAYESIANO ROBUSTO**

Javier Portela García-Miguel

**Memoria para optar al grado de Doctor en
Ciencias Matemáticas, realizada bajo la
dirección del profesor Dr. D. Miguel Ángel
Gómez-Villegas**

Madrid, Mayo de 2003

**UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS**

Departamento de Estadística e I. Operativa I

**APLICACIÓN DE MÉTODOS NUMÉRICOS EN
INFERENCIA BAYESIANA: IMPLEMENTACIÓN
DE UN MÉTODO BAYESIANO ROBUSTO**

Javier Portela García-Miguel

Madrid, Mayo de 2003

Departamento de Estadística e Investigación Operativa I

Facultad de Ciencias Matemáticas

TESIS DOCTORAL

**APLICACIÓN DE MÉTODOS NUMÉRICOS EN
INFERENCIA BAYESIANA: IMPLEMENTACIÓN
DE UN MÉTODO BAYESIANO ROBUSTO**

Autor: Javier Portela García-Miguel

Director: Miguel Ángel Gómez Villegas

2003

Quiero agradecer la ayuda que con sus consejos, sugerencias y aportaciones me ha proporcionado el Profesor M.A. Gómez-Villegas, bajo cuya dirección ha sido realizado este trabajo.

Índice

1	Introducción	3
2	Revisión de la distribución Potencial Exponencial univariante	6
2.1	Introducción	6
2.2	Forma funcional de la distribución Potencial Exponencial	7
2.2.1	Generalización	9
2.3	Propiedades y Extensiones de la distribución.	12
2.3.1	La distribución Potencial Exponencial Asimétrica.	14
2.3.2	La distribución Potencial Exponencial Multivariante	15
2.4	Generación de muestras de la distribución Potencial Exponencial	16
2.4.1	Muestras de la distribución Potencial Exponencial	16
2.4.2	Muestras de la cola de una Potencial Exponencial	17
2.4.3	Muestras de la distribución Potencial Exponencial Asimétrica	19
2.4.4	Muestras de la distribución Potencial Exponencial Multivariante	22
2.5	Elección de las distribuciones a priori	23
2.6	Distribuciones a posteriori	27
2.6.1	Aproximaciones e Integración Numérica	29
2.6.2	Aproximaciones mediante Monte-Carlo	30
2.6.3	Generación de muestras de $p(\theta, \sigma, \beta \mathbf{y})$ a través del muestreo de Gibbs	36
2.6.4	Generación de muestras de $p(\theta, \sigma, \beta \mathbf{y})$ a través de la representación por mixturas	41
3	Medidas de discrepancia para el contraste de normalidad	49
3.1	Introducción	49
3.2	Divergencia de Kullback-Leibler como medida de discrepancia	50
3.3	Logaritmo de la razón de verosimilitudes como medida de discrepancia	52
3.4	Medida de discrepancia basada en la distribución predictiva	53
3.5	Medida de discrepancia basada en el cálculo de la región HPD.	55
3.6	Aplicación. Contrastes de bondad de ajuste	63
4	Aplicaciones en modelos bayesianos	76
4.1	Introducción	76
4.2	Aplicación al modelo lineal	77
4.2.1	Estimación de las densidades marginales mediante el método de Monte-Carlo	78
4.2.2	Extracción de muestras de las densidades a posteriori	79
4.2.3	Aplicación	82
4.3	Extensión al modelo no lineal	86
4.3.1	Estimación de las densidades marginales mediante el método Monte-Carlo	87
4.3.2	Comparación de dos modelos no lineales.	87
4.3.3	Aplicación	90

4.4	Aplicación a un Modelo lineal con datos longitudinales	97
4.4.1	Distribuciones a priori de los parámetros	98
4.4.2	Simulación de las distribuciones a posteriori	99
4.4.3	Resultados	104
5	Conclusiones	109
6	Bibliografía	111

1 Introducción

El objetivo de este trabajo es desarrollar las herramientas matemáticas y computacionales necesarias para utilizar la familia de distribuciones Potencial Exponencial univariante en el marco inferencial del tratamiento de modelos bayesianos. Se presentarán técnicas de aproximación y simulación para utilizar esta familia como generalizadora de la habitual distribución Normal, se estudiarán medidas de diagnóstico para poder comparar cuándo nos interesará tratar con esta familia en lugar de utilizar su restricción al caso particular Normal, y finalmente se estudiarán aplicaciones de esta distribución en el marco de modelos lineales, no lineales y longitudinales.

La utilización de la distribución Potencial Exponencial en el marco de inferencia bayesiana fue propiciada por (Box y Tiao, 1973), que la presentó como "una elección más amplia de la distribución de base". (O'Hagan, 1979) observó su utilidad en el estudio de Outliers. Más recientemente, otros autores como (Fernández, Osiewalski y Steel, 1995), presentan una extensión asimétrica de esta familia, (Gómez, Gómez-Villegas y Marín, 1998), la generalizan al caso multivariante, y (Walker y Gutierrez-Peña, 1999), en estudios de robustez, se han interesado por ella dentro del marco bayesiano.

La principal particularidad de la distribución es la existencia de un parámetro β que controla la curtosis, de manera que para valores positivos de este parámetro la distribución es leptocúrtica (más curtosis que la distribución Normal) y para valores negativos es platicúrtica. Aunque (Box y Tiao, 1973) ya plantearon la posibilidad de considerar β como variable aleatoria, como corresponde en el entorno bayesiano, los intentos de aproximación y utilización automática de su distribución a posteriori se vieron truncados en parte por las limitaciones técnicas y computacionales del momento, y se restringieron en general a la suposición exploratoria de varios valores fijos para β . Igualmente han procedido otros autores que han tratado la utilización de esta distribución en problemas de robustez más recientemente, como (Walker y Gutierrez-Peña, 1999), (Gómez, E. Gómez-Villegas, M.A. y Marín, J. M., 2002)

La novedad en este trabajo consiste en suponer, en consonancia con el paradigma bayesiano, la aleatoriedad de β en toda circunstancia, con las consecuencias de una mayor sofisticación en el tratamiento de todas las distribuciones a posteriori. En el primer Capítulo se sientan las bases para poder tratar con esta distribución. En particular, se presenta una generalización de la forma funcional de esta distribución, útil en su tratamiento en aplicaciones, y a continuación se plantea la utilización de esta familia dentro del paradigma bayesiano. Se presentan varias técnicas para realizar inferencias sobre las distribuciones a posteriori de los parámetros. Concretamente, se utiliza un método Monte-Carlo y dos métodos para extraer muestras de la distribución a posteriori, a través del muestreo de Gibbs. La eficacia de estos métodos se verá comprobada a través de todo el trabajo, mediante su utilización en los Capítulos 2 y 3.

Habitualmente se utiliza la distribución Normal en los modelos estadísticos, tanto bayesianos como clásicos, como modelo de distribución para el término de error, e igualmente ocurre con frecuencia para modelizar simplemente la distribución de una variable aleatoria. Si se quiere utilizar en lugar de la distribución Normal la distribución Potencial Exponencial, es necesario de alguna manera determinar si esto es pertinente, pues la complejidad de tratamiento del modelo es mayor con esta distribución.

En el Capítulo 2 se estudian y comparan algunas medidas de diagnóstico que arrojan luz sobre cuándo se debe adoptar un modelo u otro. Como el caso Normal es un caso particular de la familia Potencial Exponencial el problema podrá ser planteado como un problema de contraste de hipótesis, donde la hipótesis nula puntual es la nulidad del parámetro β , correspondiente a la Normalidad de los datos de partida. Esta hipótesis se rechazaría en favor de la hipótesis alternativa que corresponde a cualquier β distinto de cero, es decir, la distribución Potencial Exponencial con $\beta \neq 0$. El problema de contraste de hipótesis nula puntual en inferencia bayesiana ha sido tratado desde diferentes perspectivas, mediante la utilización del factor Bayes (Gómez-Villegas, Gómez, 1992), (Verdinelli, Wasserman, 1998), calibración de p-valores (Bayarri, Berger, J., 1998), o comparación de modelos (Carota et al, 1996). (Bernardo, Smith, 1994) plantean, entre otras posibilidades, abordar el problema a partir de medidas de discrepancia, que es la perspectiva adoptada en este trabajo. Se estudiarán algunas medidas de discrepancia para resolver este problema de contraste de hipótesis, haciendo énfasis en una medida de discrepancia basada en el cálculo de regiones de máxima densidad a posteriori. Se desarrollan técnicas de aproximación Monte-Carlo y simulación para el cálculo de estas medidas. Los métodos mencionados en este Capítulo 2 se verán finalmente comparados en un estudio de simulación.

La eficacia de la eventual sustitución de la distribución Normal por la distribución Potencial Exponencial debe ser observada a la luz de sus aplicaciones en modelos concretos. El Capítulo 3 presenta en primer lugar cómo adecuar las técnicas de tratamiento de las distribuciones a posteriori presentadas en el Capítulo 1 a los casos concretos del modelo lineal y del modelo no lineal. En dicho Capítulo se estudia también la posibilidad de comparar modelos desde una perspectiva de ajuste, que se tomará como referencia añadida a la medida de discrepancia estudiada en el Capítulo anterior, para poder evaluar la posible mejoría del modelo al utilizar la distribución Potencial Exponencial.

Este Capítulo 3 se centra en tres aplicaciones sobre datos reales: el estudio de un modelo lineal sobre datos químicos, la comparación de dos modelos no lineales en un experimento de agricultura y finalmente el estudio de comparación de dos modelos longitudinales sobre datos médicos de SIDA, que pone a prueba la utilidad de trabajar con esta familia de distribuciones en este tipo de modelos.

Finalmente se incluye un Capítulo de Conclusiones que recoge algunos aspectos técnicos observados en el tratamiento de esta familia de distribuciones y que pone de manifiesto

la aplicación y utilidad de esta familia en el marco de estudios de robustez en inferencia bayesiana.

2 Revisión de la distribución Potencial Exponencial univariante

2.1 Introducción

La distribución Potencial Exponencial sirvió como base a (Box y Tiao, 1973) para aportar una mayor flexibilidad respecto a la distribución normal en la modelización de datos. Tiene la particularidad de englobar a las distribuciones Normal y Laplace, pues contiene un parámetro relacionado con su curtosis, lo que la hace interesante al tratar datos con más o menos curtosis que la distribución Normal. Desde entonces ha sido utilizada sobre todo en el campo de la inferencia bayesiana. (O'Hagan, 1979) comenta su utilidad en el estudio de Outliers. (West, 1987) presenta una representación de la familia como una mixtura de escala distribuciones normales. (Gómez-Villegas, 1988) la utiliza en el contexto del contraste de hipótesis bayesiano.

Otros autores la emplean, en la línea de Box y Tiao, en estudios de robustez bayesiana: así (Fernández, Osiewalski y Steel, 1995) la presentan en el contexto de las distribuciones v -esféricas y elaboran una extensión asimétrica para esta familia. (Gómez, Gómez-Villegas y Marín 1998) y (Marín, 1998) elaboran su versión multivariante y utilizan esta versión en aplicaciones en modelos dinámicos. (Walker y Gutierrez-Peña, 1999) la utilizan en un problema de selección de modelos. Desde el punto de vista clásico, (Rahman y Gokhale, 1996) plantean cómo estimar los parámetros de esta distribución vía simulación.

Salvo en aproximaciones puntuales, como ocurre en (Box y Tiao, 1973), (Gómez-Villegas, 1988), o Marín (1998), la orientación general del tratamiento bayesiano de los autores que han tratado esta distribución es plantear el parámetro β como "fijo y desconocido", actuando en las aplicaciones de manera exploratoria, variando el valor de este parámetro y observando sus consecuencias en la modelización de los datos. Este planteamiento es en cierto modo contrario al paradigma bayesiano, pues si se consideran los parámetros de centralización θ y escala σ como aleatorios, igualmente ha de ocurrir con el parámetro de control de curtosis β , dentro de la misma distribución.

El planteamiento de β como aleatorio nos permitiría en primer lugar realizar inferencias más apropiadas sobre las restantes distribuciones a posteriori; en particular sobre la distribución a posteriori de σ . También nos permitiría elaborar hipótesis sobre cuestiones relativas a la normalidad de los datos y a su curtosis, y estudiar con más flexibilidad el tratamiento de outliers en nuestros datos. Desgraciadamente este tratamiento no es directo analíticamente, y la ausencia de técnicas para poder abordar este problema desde el punto de vista práctico ha supuesto una limitación para su tratamiento en aplicaciones.

Este trabajo pretende rellenar esta laguna, aportando y verificando técnicas que permitan tratar al parámetro β como una variable aleatoria y por lo tanto estudiar las

distribuciones a posteriori de los tres parámetros que consideramos en la distribución. En este primer Capítulo se presentarán estas técnicas. En la Sección 2.2 se expone la forma funcional usual de esta distribución, y se aporta una generalización de ésta que permitirá un tratamiento más flexible de ella en desarrollos posteriores. En la Sección 2.3 veremos una propiedad de representación por mixturas de esta familia, utilizada a lo largo del presente trabajo, y veremos la forma de sus extensiones asimétrica y multivariante. En la Sección 2.4 se elaboran técnicas de generación de muestras de esta familia, tanto para la forma usual de la función de densidad completa como para la cola de la distribución, para su versión asimétrica y para su versión multivariante. La Sección 2.5 presenta algunas posibilidades para el planteamiento de las distribuciones a priori sobre los parámetros. En la Sección 2.6 se desarrollan técnicas para el tratamiento de las distribuciones a posteriori. En particular, se estudian aproximaciones vía Monte-Carlo, generación de muestras a posteriori a través de la utilización del algoritmo de muestreo de Gibbs, y utilización de la representación por mixturas de la familia para generar muestras también a partir del mismo algoritmo.

2.2 Forma funcional de la distribución Potencial Exponencial

Esta distribución continua, que denotaremos $PE(\theta, \sigma, \beta)$, fue presentada en (Box y Tiao, 1973) bajo la forma:

$$p(y | \theta, \sigma, \beta) = w(\beta)\sigma^{-1} \exp \left[-c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{2/(1+\beta)} \right] \quad (1.1)$$

$$\text{con } -\infty < y < +\infty, \quad -\infty < \theta < +\infty, \quad -1 < \beta \leq 1, \quad \sigma > 0$$

donde

$$c(\beta) = \left\{ \frac{\Gamma[\frac{3}{2}(1+\beta)]}{\Gamma[\frac{1}{2}(1+\beta)]} \right\}^{1/(1+\beta)}$$

y

$$w(\beta) = \frac{\{\Gamma[\frac{3}{2}(1+\beta)]\}^{1/2}}{(1+\beta) \{\Gamma[\frac{1}{2}(1+\beta)]\}^{3/2}}$$

siendo los parámetros θ y σ respectivamente la media y desviación típica de la población. El parámetro β puede verse como una medida de kurtosis, y es tal que la distribución es una Normal (θ, σ) cuando $\beta = 0$, doble exponencial cuando $\beta = 1$ y Uniforme en el límite cuando $\beta \rightarrow -1$.

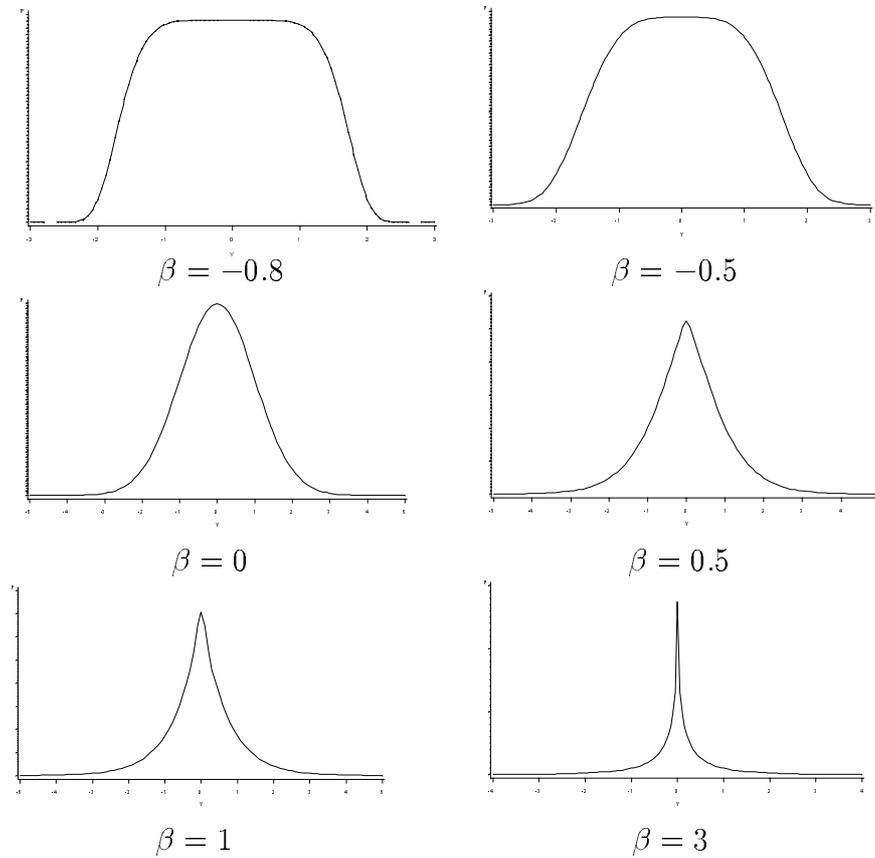


Figura 1.1. Densidad de la distribución Potencial Exponencial para diferentes valores de β , con $\theta = 0$, $\sigma = 1$.

Parametrizaciones sobre β

El parámetro de kurtosis puede ponerse en la forma $q = \frac{2}{1+\beta}$, con q perteneciente al intervalo $(1, \infty)$. El dominio de los parámetros β (y alternativamente q) puede extenderse a $(-1, \infty)$ (alternativamente, a $(0, \infty)$), dando lugar a la distribución Uniforme en ambos extremos de los intervalos citados. La justificación para adoptar la forma $\frac{2}{1+\beta}$ en $(-1, 1]$ se basa en que lleva a la distribución Uniforme y a la doble exponencial respectivamente en sus extremos, adoptando la distribución Normal en el centro del intervalo. De este modo la distribución de la población será leptocúrtica para $\beta > 0$ y platicúrtica para $\beta < 0$. Esto no impide que para casos particulares se pueda optar por unas u otras extensiones del dominio del parámetro de kurtosis.

2.2.1 Generalización

La representación de la distribución Potencial Exponencial presentada en (1.1) se puede generalizar variando $c(\beta)$ y $w(\beta)$ según la siguiente propiedad:

Teorema 1.1.

Sea $c(\beta) =$ cualquier función positiva y acotada para todo $-1 < \beta \leq 1$

Entonces, tomando $w(\beta) = \frac{c(\beta)^{\frac{1+\beta}{2}}}{2\Gamma[\frac{1}{2}(3+\beta)]}$,

1) La expresión (1.1) es la función de densidad de una variable aleatoria Y

2) La esperanza de esta variable aleatoria es $E[Y] = \theta$ y su varianza es

$$V(Y) = \frac{\Gamma[\frac{3}{2}(1+\beta)]}{\Gamma[\frac{1}{2}(1+\beta)]c(\beta)^{1+\beta}}\sigma^2$$

Demostración.

Utilizaremos para la demostración la siguiente propiedad:

Para $a > 0, p > 0, \alpha > 0$,

$$\int_0^{\infty} x^{p-1} e^{-ax^\alpha} dx = \frac{1}{\alpha} a^{-p/\alpha} \Gamma\left(\frac{p}{\alpha}\right) \quad (1.2)$$

1) Aplicando el cambio de variable $x = \frac{y - \theta}{\sigma}$ tenemos

$$\begin{aligned} & \int_{-\infty}^{+\infty} w(\beta) \sigma^{-1} \exp \left[-c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{2/(1+\beta)} \right] dy = \\ & = \int_{-\infty}^{+\infty} w(\beta) \exp \left[-c(\beta) |x|^{2/(1+\beta)} \right] dx = 2w(\beta) \int_0^{+\infty} \exp \left[-c(\beta) |x|^{2/(1+\beta)} \right] dx \end{aligned}$$

Llamando en (1.2) $p = 1$, $a = c(\beta)$ y $\alpha = 2/(1 + \beta)$, queda finalmente

$$2w(\beta) \int_0^{+\infty} \exp \left[-c(\beta) |x|^{2/(1+\beta)} \right] dx = 2w(\beta) \frac{(1+\beta)}{2} c(\beta)^{-\frac{(1+\beta)}{2}} \Gamma\left(\frac{(1+\beta)}{2}\right)$$

Para que se trate de una función de densidad esta expresión ha de ser igual a 1, lo que se cumple si y sólo si

$$w(\beta) = \frac{c(\beta)^{\frac{1+\beta}{2}}}{2\Gamma[\frac{1}{2}(3+\beta)]}$$

que es la condición impuesta en el enunciado.

$$2) E(Y) = \int_{-\infty}^{+\infty} w(\beta)\sigma^{-1}y \exp\left[-c(\beta)\left|\frac{y-\theta}{\sigma}\right|^{2/(1+\beta)}\right] dy$$

Haciendo el cambio de variable $x = \frac{y-\theta}{\sigma}$ tenemos

$$E(Y) = w(\beta)\sigma \int_{-\infty}^{+\infty} x \exp\left[-c(\beta)|x|^{2/(1+\beta)}\right] dy + \theta \int_{-\infty}^{+\infty} w(\beta) \exp\left[-c(\beta)|x|^{2/(1+\beta)}\right] dy$$

En esta expresión, la primera integral se anula y la segunda es θ multiplicado por la integral de la función de densidad de una Potencial Exponencial con $\theta = 0$ y $\sigma = 1$, con lo que $E(Y) = \theta$.

Por otra parte, haciendo el mismo cambio de variable $x = \frac{y-\theta}{\sigma}$ se tiene

$$\begin{aligned} E(Y^2) &= w(\beta)\sigma^2 \int_{-\infty}^{+\infty} x^2 \exp\left[-c(\beta)|x|^{2/(1+\beta)}\right] dy + 2\theta w(\beta) \int_{-\infty}^{+\infty} x \exp\left[-c(\beta)|x|^{2/(1+\beta)}\right] dy + \\ &+ \theta^2 \int_{-\infty}^{+\infty} w(\beta) \exp\left[-c(\beta)|x|^{2/(1+\beta)}\right] dy = \\ &= 2w(\beta)\sigma^2 \int_0^{+\infty} x^2 \exp\left[-c(\beta)|x|^{2/(1+\beta)}\right] dy + \theta^2 \end{aligned}$$

La integral se calcula con la propiedad (1.2), tomando $p = 3$, $a = c(\beta)$ y $\alpha = 2/(1+\beta)$, quedando finalmente

$$E(Y^2) = 2w(\beta)\sigma^2 \frac{(1+\beta)}{2} c(\beta)^{-\frac{3(1+\beta)}{2}} \Gamma\left(\frac{3(1+\beta)}{2}\right) + \theta^2$$

y, por lo tanto,

$$V(Y) = E(Y^2) - E(Y)^2 = \frac{\Gamma[\frac{3}{2}(1+\beta)]}{\Gamma[\frac{1}{2}(1+\beta)]c(\beta)^{1+\beta}}\sigma^2$$

◇

Utilizando esta propiedad se puede definir esta versión generalizada de la familia Potencial Exponencial:

$$p(y | \theta, \sigma, \beta) = w(\beta)\sigma^{-1} \exp \left[-c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{2/(1+\beta)} \right] \quad (1.3)$$

con $-\infty < y < +\infty$, $-\infty < \theta < +\infty$, $-1 < \beta \leq 1$, $\sigma > 0$

donde $c(\beta)$ es cualquier función de β positiva y acotada, y

$$w(\beta) = \frac{c(\beta)^{\frac{1+\beta}{2}}}{2\Gamma[\frac{1}{2}(3 + \beta)]}$$

La expresión de la distribución Potencial Exponencial presentada inicialmente es un caso particular de esta familia de distribuciones, en el que se toma

$$c(\beta) = \left\{ \frac{\Gamma[\frac{3}{2}(1 + \beta)]}{\Gamma[\frac{1}{2}(1 + \beta)]} \right\}^{1/(1+\beta)}.$$

La generalización presentada tiene el inconveniente de que la varianza de la variable aleatoria depende de los dos parámetros σ y β , además de que la variable en cuestión no tiene, en general, por qué coincidir con la distribución Normal o Laplace para ciertos valores concretos de β . Esta propiedad dependerá en cada caso de la función $c(\beta)$. Sin embargo esta generalización permite una reparametrización interesante para un problema que nos ocupará en el presente trabajo, que es reflejar el alejamiento de los datos de la distribución Normal.

Fijando concretamente $c(\beta) = \frac{1}{2}$ se tienen las siguientes propiedades :

a) La función de densidad toma la forma

$$p(y | \theta, \sigma, \beta) = w(\beta)\sigma^{-1} \exp \left[-\frac{1}{2} \left| \frac{y - \theta}{\sigma} \right|^{2/(1+\beta)} \right] \quad (1.4)$$

con $-\infty < y < +\infty$, $-\infty < \theta < +\infty$, $-1 < \beta \leq 1$, $\sigma > 0$

donde

$$w(\beta) = 2^{-\frac{3+\beta}{2}} \Gamma[\frac{1}{2}(3 + \beta)]^{-1}$$

$$\text{b) } E[Y] = \theta \text{ y } V(Y) = \frac{\Gamma[\frac{3}{2}(1 + \beta)]}{\Gamma[\frac{1}{2}(1 + \beta)]} 2^{1+\beta} \sigma^2$$

c) Cuando $\beta = 0$, $V(Y) = \sigma^2$ y además la variable aleatoria Y coincide con la distribución Normal $N(\theta, \sigma^2)$.

A la hora de trabajar con la distribución potencial exponencial con fines de inferencia, se podrá utilizar la parametrización inicial y por lo tanto el $c(\beta)$ dado en la definición (1.1) de esta familia dada por (Box y Tiao, 1973), o bien fijar otro valor constante o funcional para $c(\beta)$, como el ya indicado $c(\beta) = \frac{1}{2}$. La elección de $c(\beta)$ definirá en particular la relación de σ y β con la varianza de la distribución.

El hecho de fijar $c(\beta)$ constante simplifica los cálculos en muchos casos, manteniéndose las propiedades principales de la distribución Potencial Exponencial, es decir, el hecho de ser una distribución simétrica que incluye a la distribución Normal y que contiene un parámetro β que controla la curtosis.

En lo sucesivo, cuando sea posible se adoptará una expresión genérica de la familia, utilizando el término $c(\beta)$, dejando lugar así a la libertad de poder determinar esta función. Si en algún desarrollo o aplicación es necesario prefijar $c(\beta) = \frac{1}{2}$ para obtener resultados más directos y controlables, se indicará previamente.

2.3 Propiedades y Extensiones de la distribución.

La distribución Potencial Exponencial es simétrica respecto a θ y unimodal con moda en θ . Además es log-cóncava para $\beta < 1$.

El poder adoptar, según los valores de β , distribuciones simétricas con diferente apuntamiento, incluida la distribución Normal, permite realizar inferencias sobre los datos teniendo en cuenta, con mayor o menor sensibilidad, la presencia de datos atípicos. (O'Hagan, 1984) comenta cómo esta familia de distribuciones es sensible a outliers para $\beta < 1$ (en el sentido en que los outliers influyen en la distribución a posteriori del parámetro de centralización aunque se aumente el tamaño muestral) y, recíprocamente, insensible a outliers para $\beta > 1$ (en el sentido en que los outliers no influyen sensiblemente en la distribución a posteriori del parámetro de centralización si se aumenta el tamaño muestral), independientemente de la distribución a priori que tomemos sobre el parámetro. Estas propiedades hacen que la familia sea especialmente interesante en Inferencia Bayesiana con problemas de Robustez.

A continuación se verá una propiedad de representación de esta distribución, que se utilizará en las aplicaciones.

Teorema 1.2 (Representación de la Potencial Exponencial como mixtura de Uniforme y Gamma).

Si $U \equiv \Gamma(1 + \frac{1+\beta}{2}, 1)$ y $V \equiv U [-1, 1]$, independientes, entonces la variable aleatoria $Y = \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} V U^{\frac{1+\beta}{2}} + \theta$ se distribuye como una Potencial Exponencial $PE(\theta, \sigma, \beta)$.

Demostración.

Al ser $V \equiv U [-1, 1]$, tenemos que

$$Y | U \equiv U\left(\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} U^{\frac{1+\beta}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} U^{\frac{1+\beta}{2}}\right)$$

Bastará ver que $p(y) = \int p(y | u)p(u)du$ es la expresión de la función de densidad de una Potencial Exponencial $PE(\theta, \sigma, \beta)$.

Como

$$p(y | u) = \frac{c(\beta)^{\frac{\beta+1}{2}}}{2\sigma u^{\frac{1+\beta}{2}}}, \text{ con } \left| \frac{y - \theta}{\sigma} \right| < \left(\frac{u}{c(\beta)} \right)^{\frac{\beta+1}{2}} \iff u > c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{\frac{2}{\beta+1}}$$

y

$$p(u) = \frac{1}{\Gamma(\frac{1}{2}(3+\beta))} u^{\frac{\beta+1}{2}} e^{-u}, u \in (0, \infty)$$

tenemos que

$$\begin{aligned} \int p(y | u)p(u)du &= \int_{c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{\frac{2}{\beta+1}}}^{\infty} \frac{c(\beta)^{\frac{\beta+1}{2}}}{2\sigma u^{\frac{1+\beta}{2}}} \frac{1}{\Gamma(\frac{1}{2}(3+\beta))} u^{\frac{\beta+1}{2}} e^{-u} du = \\ &= \frac{c(\beta)^{\frac{\beta+1}{2}}}{2\sigma \Gamma(\frac{1}{2}(3 + \beta))} \int_{c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{\frac{2}{\beta+1}}}^{\infty} e^{-u} du = \frac{c(\beta)^{\frac{\beta+1}{2}}}{2\sigma \Gamma(\frac{1}{2}(3 + \beta))} \exp \left[-c(\beta) \left| \frac{y - \theta}{\sigma} \right|^{2/(1+\beta)} \right] \end{aligned}$$

que es la expresión de la densidad de una distribución Potencial Exponencial \diamond

Esta propiedad, aparte de dar lugar a un sencillo método de generación de muestras de la familia Potencial Exponencial, permite una representación de esta familia mediante mixturas de Uniformes y Gamma que toma la siguiente forma:

$$Y | U, \theta, \beta, \sigma \equiv U\left(\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} U^{\frac{1+\beta}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} U^{\frac{1+\beta}{2}}\right)$$

$$U | \beta \equiv \Gamma\left(\frac{1}{2}(3 + \beta), 1\right)$$

Se verifica por el Teorema 1.2 que $Y | \theta, \sigma, \beta \equiv PE(\theta, \sigma, \beta)$.

Esta representación permite abordar la extracción de muestras de la distribución a posteriori $p(\theta, \sigma, \beta | y)$ mediante el método de muestreo de Gibbs, como veremos en la sección 4. Esto nos permitirá realizar inferencias sobre θ, β y σ . Además, el estudio de la distribución de los valores de la variable de mixtura u nos permitiría estudiar la presencia de outliers en nuestros datos (Walker, S.G. y Gutierrez-Peña, E. (1999)).

2.3.1 La distribución Potencial Exponencial Asimétrica.

La distribución Potencial Exponencial pertenece a la familia exponencial y a su vez a la familia más amplia de las distribuciones l_q esféricas. (C. Fernández, 1995) presenta una variante de la Potencial Exponencial, la Potencial Exponencial Asimétrica, con densidad

$$p(y | \sigma, \beta) = \begin{cases} c \exp\left(-\frac{1}{2}(y | \sigma)^{2/(1+\beta)}\right) & \text{si } y \geq 0 \\ c \exp\left(-\frac{1}{2}(-\sigma y)^{2/(1+\beta)}\right) & \text{si } y < 0 \end{cases}$$

con $-\infty < y < \infty, -1 < \beta < \infty, 0 < \sigma < \infty$

y

$$c^{-1} = 2^{(1+\beta)/2} \Gamma\left(1 + \frac{1+\beta}{2}\right) \left(\sigma + \frac{1}{\sigma}\right)$$

Esta distribución coincide con la distribución Potencial Exponencial $PE(0, 1, \beta)$ si $\sigma = 1$. Para $\sigma \neq 1$ es asimétrica respecto de 0 (aunque se puede añadir también un parámetro de centralización θ y hacerla asimétrica respecto a θ). En esta familia el coeficiente de asimetría no depende sólo de β , como ocurría en la familia presentada por (Box, Tiao, 1973), sino que es también función del parámetro de escala σ . En la notación

original del autor, se fija $q = \frac{2}{1 + \beta}$.

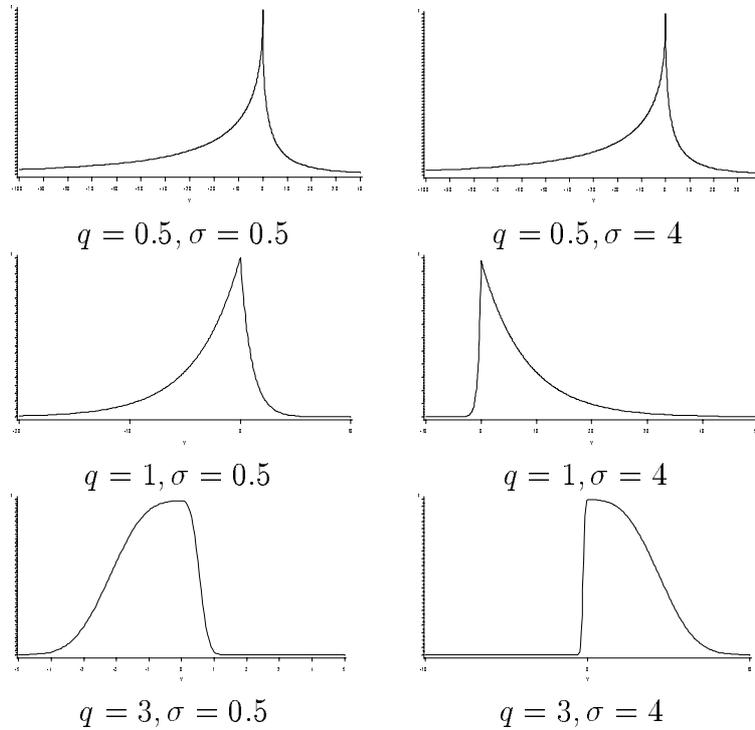


Figura 1.2. Distribución Potencial Exponencial Asimétrica. $\theta = 0$

La distribución Potencial Exponencial Asimétrica pretende extender el modelo Potencial Exponencial permitiendo modelizar también una posible asimetría en los datos. Para contener realmente a los modelos simétricos de la Potencial Exponencial sería necesario establecer un parámetro de escala adicional (en el caso $\sigma = 1$ sólo estaríamos en presencia de una Potencial Exponencial Simétrica con parámetro de escala $\sigma = 1$). Esto añadiría un esfuerzo adicional al problema de la deducción de las distribuciones a posteriori de los parámetros, difícil de por sí por la forma funcional de la densidad, que depende de la situación de y respecto de la moda, en la recta real.

2.3.2 La distribución Potencial Exponencial Multivariante

Una generalización multivariante de esta familia puede verse en Gómez, Gómez-Villegas y Marín (1998), y está definida por

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \Sigma, \beta) = \frac{n\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}\Gamma(1 + \frac{n}{2\beta})2^{1 + \frac{n}{2\beta}}} \exp(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\theta})'\Sigma(\mathbf{y} - \boldsymbol{\theta}))^\beta$$

donde \mathbf{y} es un vector n dimensional, $\boldsymbol{\theta}$ es un vector n -dimensional de parámetros, Σ es una matriz definida positiva de parámetros y β está definido en $(0, \infty)$. Para $\beta = 2$, con la notación utilizada, coincide con la distribución Normal Multivariante.

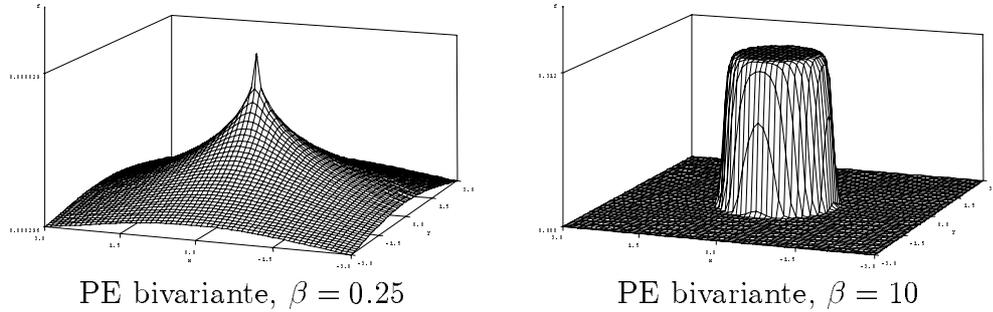


Figura 1.3. Distribución Potencial Exponencial Multivariante

2.4 Generación de muestras de la distribución Potencial Exponencial

2.4.1 Muestras de la distribución Potencial Exponencial

Utilizando el Teorema 1.2, se puede establecer un método para extraer muestras de la distribución Potencial Exponencial :

[Generación de muestras de la distribución Potencial Exponencial $PE(\theta, \sigma, \beta)$]

1. Generar V de una $\Gamma(1 + \frac{1+\beta}{2})$, U de una $U(-1, 1)$
2. $Y \leftarrow \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} UV^{\frac{1+\beta}{2}} + \theta$

La muestra obtenida y proviene de una potencial exponencial $PE(\theta, \sigma, \beta)$.

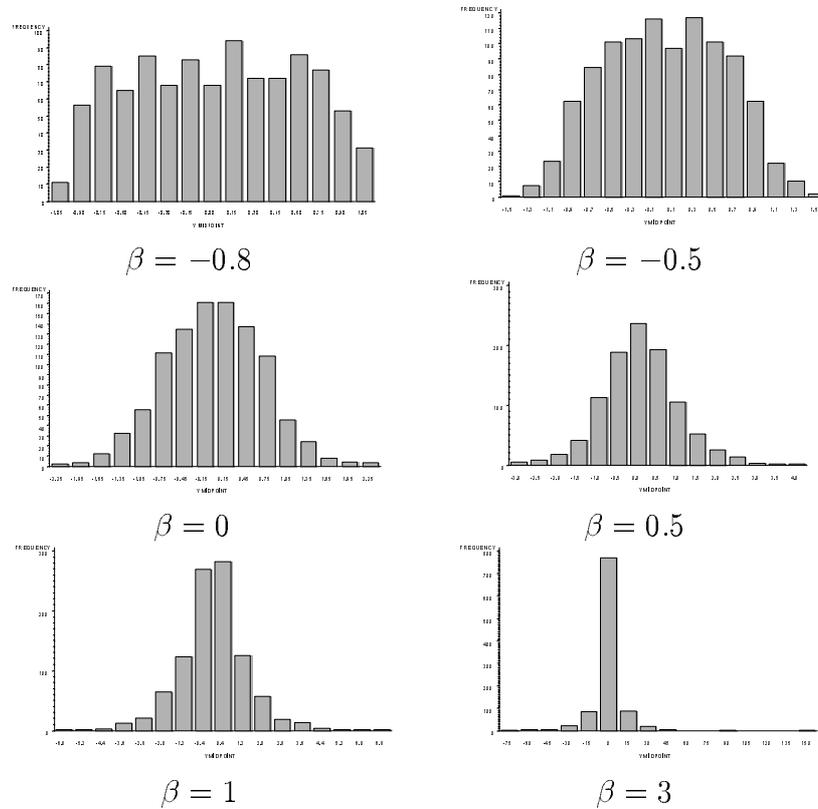


Figura 1.4. Simulación de la distribución Potencial Exponencial para diferentes valores de β , con $\theta = 0$, $\sigma = 1$.

2.4.2 Muestras de la cola de una Potencial Exponencial

En ocasiones se necesitarán extraer muestras de la cola de una Potencial Exponencial. Para ello se pueden generar valores directamente de esta distribución y rechazar aquellos que no caen en la cola, pero este modo de proceder puede ser muy ineficiente en casos en que las colas tienen una probabilidad asociada relativamente pequeña. En este apartado presentaremos un método más rápido para simular valores de la cola de una distribución Potencial Exponencial $PE(\theta, \sigma, \beta)$, es decir, valores de la Potencial Exponencial truncada en un punto a , $X \geq a$ (para simplificar supondremos $a > \theta$, y para la otra cola se razona por simetría, así como para simular de ambas colas o de un intervalo central).

Supondremos para simplificar la notación $c(\beta) = \frac{1}{2}$, aunque el método se puede aplicar con cualquier $c(\beta)$.

Para $x \geq a > \theta$, se sabe que

$$\begin{aligned} \exp\left(-\frac{1}{2} \left| \frac{x - \theta}{\sigma} \right|^{2/(1+\beta)}\right) &\leq \exp\left(-\frac{1}{2} \left| \frac{a - \theta}{\sigma} \right|^{2/(1+\beta)-1} \left| \frac{x - \theta}{\sigma} \right|\right) = \\ &= \exp(-C(x - \theta)), \text{ para } x \geq a > \theta \end{aligned}$$

$$\text{Donde } C = \frac{1}{2\sigma^{2/(1+\beta)}} |a - \theta|^{2/(1+\beta)-1}$$

Por lo tanto podemos utilizar el método de rechazo (Devroye, 1984) para generar muestras de la cola de una Potencial Exponencial. La versión más sencilla del algoritmo de rechazo es la siguiente:

Supongamos que el objetivo es extraer muestras de una distribución $f(y)$ cuya forma analítica es conocida y sabemos extraer muestras de una distribución $g(y)$ y conocemos además una constante c tal que $f(y) \leq cg(y)$ para todo y .

[Algoritmo de rechazo]

1. Extraer una muestra y de $g(y)$ y u de $U(0, 1)$
2. Hacer $t = c \frac{g(y)}{f(y)}$
3. Si $ut \leq 1$, aceptar y como una muestra de $f(y)$. Si no, rechazar y y volver al Paso 1

Para poder aplicar el algoritmo pueden conocerse solamente las formas analíticas de $f(y)$ y $g(y)$ a menos de una constante de proporcionalidad, sin que varíe el resultado del algoritmo, aunque estos casos se tiene menos control sobre la constante de rechazo. En el presente trabajo hemos utilizado esta propiedad en la mayoría de los métodos de rechazo, comprobando empíricamente el funcionamiento del algoritmo en pruebas de simulación para observar su eficiencia.

En nuestro caso aplicaremos el algoritmo generando en primer lugar muestras de la distribución Exponencial Truncada en a , y a continuación rechazando las muestras x obtenidas utilizando la razón de la función de acotación y la función acotada,

$$t = \frac{\exp\left(-\frac{1}{2} \left| \frac{a - \theta}{\sigma} \right|^{2/(1+\beta)-1} \left| \frac{x - \theta}{\sigma} \right|\right)}{\exp\left(-\frac{1}{2} \left| \frac{x - \theta}{\sigma} \right|^{2/(1+\beta)}\right)}$$

[Generación de muestras de la cola de una Potencial Exponencial]

1. Generar un valor x de la distribución exponencial $Exp(C)$ truncada en $X \geq a$
2. Generar un valor u de una $U(0, 1)$
3. Calcular t según la expresión citada. Si $ut \leq 1$, aceptar x . En caso contrario, volver al paso 1

2.4.3 Muestras de la distribución Potencial Exponencial Asimétrica

Se puede establecer también un método de simulación para la familia Potencial Exponencial Asimétrica, utilizando algunas de sus propiedades:

[Generación de muestras de la distribución Potencial Exponencial Asimétrica
PEA(θ, σ, β)]

1. Generar U de una Uniforme (0,1)
2. Generar V de una $\Gamma(\frac{1}{2}, \frac{2}{1+\beta})$
3. $Y_1 \leftarrow -\frac{1}{\sigma} V^{\frac{1}{1+\beta}}$
 $Y_2 \leftarrow \sigma V^{\frac{1}{1+\beta}}$
4. Si $U \leq \frac{1}{1 + \sigma^{\frac{2}{1+\beta}} (\frac{-Y_1}{Y_2})^{\frac{1+\beta}{2}} \exp[-\frac{1}{2}((\frac{Y_2}{\sigma})^{\frac{2}{1+\beta}} - \sigma^{\frac{2}{1+\beta}} (-Y_1)^{\frac{2}{1+\beta}})]}$

Entonces $Y \leftarrow Y_1 + \theta$

Si no, $Y \leftarrow Y_2 + \theta$

Teorema 1.3. Las muestras obtenidas con el algoritmo presentado provienen de una distribución Potencial Exponencial Asimétrica.

Demostración-. En (Devroye, 1984) se puede ver que si se sabe generar una muestra de la variable aleatoria $\psi(Y)$ con densidad h , dado que Y tiene densidad f , y $\psi(Y)$ es una función regular cuya derivada tiene, por ejemplo, signo positivo en $(0, \infty)$ y negativo en $(-\infty, 0)$, puede generarse una muestra Y con densidad f de la siguiente manera:

1. Se hallan las soluciones Y_1 y Y_2 a la ecuación $V = \psi(Y)$, de manera que $Y_1 = l(V)$ es la solución en $(-\infty, 0)$ y $Y_2 = r(V)$ es la solución en $(0, \infty)$.
2. Dada V generada con la densidad h , se toma Y_1 con probabilidad $p = \frac{f(Y_1)}{|\psi'(Y_1)|h(V)}$ e Y_2 con probabilidad $1 - p$, con

$$h(V) = \left| \frac{1}{\psi'(l(V))} \right| f(l(V)) + \left| \frac{1}{\psi'(r(V))} \right| f(r(V))$$

como densidad h de $\psi(Y)$, deducida utilizando el Teorema del cambio de Variable.

En el caso que nos ocupa, si y tiene densidad Potencial Exponencial Asimétrica, definimos

$$\begin{aligned} \psi(y) = v &= (y | \sigma)^q I_{(0, \infty)}(y) + \sigma^q (-y)^q I_{(-\infty, 0)}(y) \\ &= v_1 I_{(0, \infty)}(y) + v_2 I_{(-\infty, 0)}(y) \end{aligned}$$

Así, la densidad en v queda

$$f(v | \sigma, q) = c \exp\left(-\frac{1}{2}v\right)$$

Por el Teorema del Cambio de Variable, si $y \geq 0$,

$$v = (y | \sigma)^q, |dy/dv| = \frac{\sigma}{q} v^{\frac{1}{q}-1}$$

y

$$f(v | \sigma, q, y \geq 0) \propto v^{\frac{1}{q}-1} \exp\left(-\frac{1}{2}v\right) \equiv \Gamma\left(\frac{1}{2}, \frac{1}{q}\right)$$

Análogamente, si $y \leq 0$,

$$v = \sigma^q (-y)^q, |dy/dv| = \frac{1}{\sigma q} v^{\frac{1}{q}-1}$$

y

$$f(v | \sigma, q, y \leq 0) \propto v^{\frac{1}{q}-1} \exp\left(-\frac{1}{2}v\right) \equiv \Gamma\left(\frac{1}{2}, \frac{1}{q}\right)$$

como $v = v_1 I_{(0, \infty)}(y) + v_2 I_{(-\infty, 0)}(y)$ tenemos que

$$\begin{aligned}
p(v \mid \sigma, q) &= p(v \mid \sigma, q, y \geq 0)p(y \geq 0 \mid \sigma, q) + p(v \mid \sigma, q, y \leq 0)p(y \leq 0 \mid \sigma, q) = \\
&= \frac{\sigma^2}{\sigma^2 + 1}p(v \mid \sigma, q, y \geq 0) + \frac{1}{\sigma^2 + 1}p(v \mid \sigma, q, y \leq 0) = \\
&= \frac{\sigma^2}{\sigma^2 + 1} \frac{1}{\Gamma(\frac{1}{q})2^{\frac{1}{q}}} v^{\frac{1}{q}-1} \exp(-\frac{1}{2}v) + \frac{1}{\sigma^2 + 1} \frac{1}{\Gamma(\frac{1}{q})2^{\frac{1}{q}}} v^{\frac{1}{q}-1} \exp(-\frac{1}{2}v) = \\
&= \frac{1}{\Gamma(\frac{1}{q})2^{\frac{1}{q}}} v^{\frac{1}{q}-1} \exp(-\frac{1}{2}v) \equiv \Gamma(\frac{1}{2}, \frac{1}{q})
\end{aligned}$$

Así, $v = \psi(y)$ tiene densidad $\Gamma(\frac{1}{2}, \frac{1}{q})$. Además,

$$\psi'(y) = -q\sigma^q(-y)^{q-1} < 0 \text{ para } y \in (-\infty, 0)$$

y

$$\psi'(y) = \frac{q}{\sigma^q} y^{q-1} > 0 \text{ para } y \in (0, \infty)$$

Por otra parte, las soluciones Y_1 y Y_2 a la ecuación $V = \psi(Y)$ son

$$Y_1 = -\frac{1}{\sigma} \sqrt[q]{V} \text{ para } y \in (-\infty, 0)$$

y

$$Y_2 = \sigma \sqrt[q]{V} \text{ para } y \in (0, \infty)$$

Es fácil ver que la probabilidad p mencionada en el Algoritmo de generación propuesto es, en nuestro caso, de la forma

$$p = \frac{1}{1 + \sigma^{2q} \left(\frac{-Y_1}{Y_2}\right)^{q-1} \exp\left[-\frac{1}{2}\left(\left(\frac{Y_2}{\sigma}\right)^q - \sigma^q(-Y_1)^q\right)]}$$

con lo cual queda concluida la demostración. \diamond

Se recoge una aplicación de este algoritmo para distintos valores de q en la Figura 1.5.

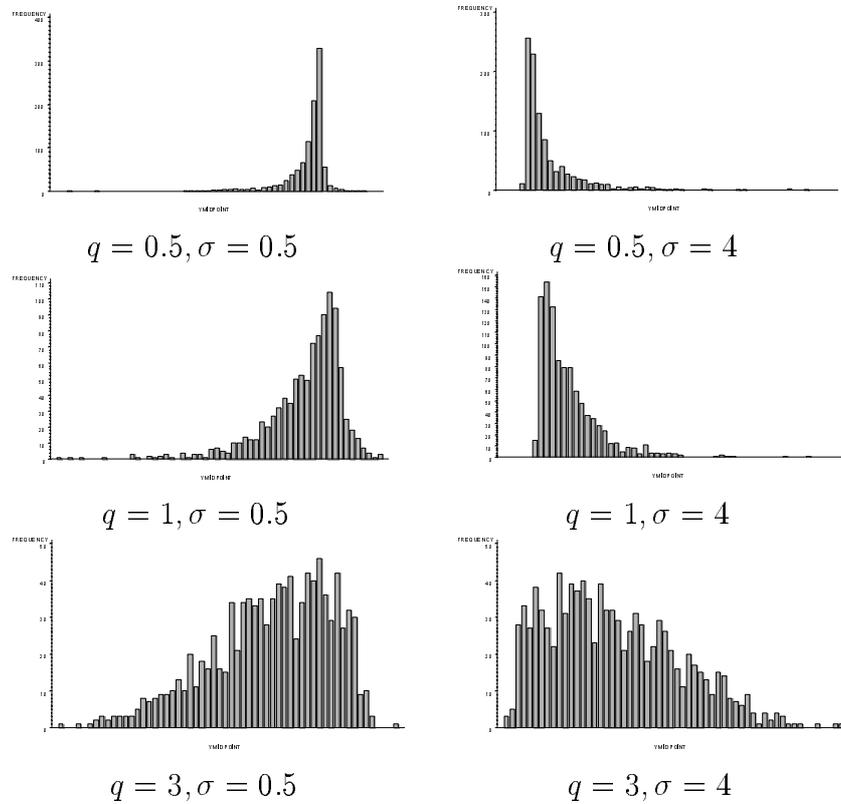


Figura 1.5. Simulación de la distribución Potencial Exponencial Asimétrica, $\theta = 0$

2.4.4 Muestras de la distribución Potencial Exponencial Multivariante

Un método para simular muestras de la distribución Potencial Exponencial Multivariante está expuesto en Gómez, Gómez-Villegas y Marín (1998). Toma la siguiente forma:

[Generación de muestras de la distribución Potencial Exponencial Multivariante $PE(\theta, \Sigma, \beta)$]

1. Calcular una matriz A tal que $A'A = \Sigma$
2. Generar U uniformemente en la esfera unidad en R^n
3. Generar V uniformemente en la esfera unidad en R^n
4. Generar S de una $\Gamma(1, \frac{n(1+\beta)}{2})$
5. $R \leftarrow (2S)^{\frac{1+\beta}{2}}$
6. $Y \leftarrow \theta + A'RU$

El vector obtenido Y proviene de una distribución Potencial Exponencial Multivariante $PE(\theta, \Sigma, \beta)$.

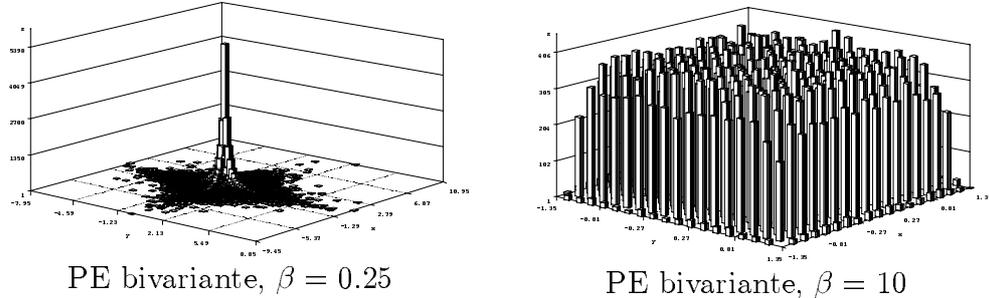


Figura 1.6. Muestras simuladas de la distribución Potencial Exponencial Multivariante

2.5 Elección de las distribuciones a priori

En (Box, Tiao, 1973) se presenta el modelo a priori con los tres parámetros independientes. Aunque la independencia a priori entre el parámetro de centralización θ y los parámetros de escala σ y β es razonable, es necesario profundizar en la independencia a priori entre los dos parámetros de escala σ y β . Una posibilidad es proponer $\log \sigma$ Uniforme ($p(\sigma) = \frac{1}{\sigma}$) e independiente de β ; es decir $p(\sigma, \beta) = \frac{1}{\sigma}p(\beta)$. Esta opción, además de ser natural en el marco de una familia de centralización y escala, pues corresponde a la obtenida mediante la regla de Jeffreys cuando se consideran fijos los demás parámetros, permite reducir sustancialmente los cálculos mediante integración en σ , y adoptar con más flexibilidad las distribuciones a priori $p(\beta)$ y $p(\theta)$.

Por lo tanto, adoptaremos a partir de ahora este criterio, tomando

$$p(\theta, \sigma, \beta) \propto \frac{1}{\sigma}p(\theta)p(\beta)$$

Distribuciones a priori sobre θ

Tanto (Box, Tiao, 1973) como (Marín, 1998) y (Fernández, 1995) (ésta última en el contexto de la extensión Asimétrica de la distribución) plantean como distribución a priori sobre θ una distribución no informativa $p(\theta) \propto cte$. Esta distribución corresponde a la distribución obtenida mediante la regla de Jeffreys cuando se consideran fijos los parámetros σ y β . Posiblemente el hecho de no disponer de distribuciones conjugadas para θ limitó el análisis Bayesiano de la distribución en (Box, Tiao, 1973), que sí adoptaron el modelo Normal a priori para el parámetro de centralización en el contexto de Normalidad de los datos. Si bien es cierto que los métodos de aproximación e Integración Numérica

aplicados en su estudio, resolvían parte de los problemas numéricos planteados en el cálculo de las distribuciones a posteriori de la Potencial Exponencial, con el método de Monte-Carlo y los métodos de simulación presentados en este trabajo es posible llegar a resultados a posteriori partiendo prácticamente de cualquier distribución $p(\theta)$, dejando el terreno abierto a cualquier aplicación práctica en la que se disponga de información más o menos precisa sobre θ .

Por lo tanto en este trabajo utilizaremos la distribución a priori genérica $p(\theta)$, continua y acotada, con dominio en $(-\infty, +\infty)$.

En aplicaciones particulares podremos plantear en algún caso una distribución no informativa para θ , restringiendo su dominio a un rango arbitrario suficientemente amplio $[-M, M]$.

Distribuciones a priori sobre β

La versatilidad del parámetro β reside en primer lugar en la posibilidad de asignarle un valor fijo, pretendiendo con ello modelizar los datos con distribuciones con colas más o menos pesadas, o distribuciones más o menos alejadas de la Normal o de la Uniforme en lo que se refiere a la curtosis. Así, (Box, Tiao, 1973) dedican parte de su trabajo al estudio de la distribución con β conocido. Aunque se pueda objetar que la interpretación sobre los valores del parámetro diferentes de 0, -1 o 1 puede ser oscura (no se sabe qué significado en términos reales tiene un valor para β como 0.63345), la posibilidad de ajustar mejor la forma de la distribución a los datos de que se dispone en casos concretos puede ser valiosa, como comenta (O'Hagan, 1984) al presentarla como un modelo interesante en cuanto a datos con Outliers.

El interpretar β como variable aleatoria permitirá contrastar y enriquecer con los datos la opinión inicial sobre β aportada a través de la distribución a priori. Esta información sobre β es, en realidad, información sobre el apuntamiento de los datos. Algunas posibilidades son:

- Suponer los datos Uniformes, lo que equivale a tomar una distribución sobre β de la forma $\frac{1}{(1+\beta)^p}$ con moda en el límite para $\beta \rightarrow -1$.
- Adoptar una distribución Uniforme (no informativa) sobre β , $p(\beta) = cte$, dejando a los datos, mediante la distribución a posteriori sobre β , la responsabilidad de aportar toda la información sobre este parámetro.
- Finalmente (Box, Tiao, 1973) proponen

$$p(\beta) = \frac{\Gamma(2a)}{[\Gamma(a)]^2 2^{2a-1}} (1 - \beta^2)^{a-1}, \quad -1 < \beta \leq 1, \quad a \geq 1$$

que es un modelo Beta-simétrico extendido al intervalo $(-1, 1]$. Es una distribución unimodal con moda en $\beta = 0$, (que corresponde a suponer una distribución Normal inicial sobre los datos) y valores $\beta \rightarrow -1$ y $\beta = 1$ en los extremos, relativos a las distribuciones Uniforme y Laplace respectivamente. El parámetro a permitirá ajustar el apuntamiento de esta distribución a priori. A medida que a aumenta la distribución es más apuntada, tendiendo hacia una función δ en $\beta = 0$ cuando $a \rightarrow \infty$, con lo cual se asumiría Normalidad de los datos independientemente de la información aportada por la muestra.

Es sencillo deducir un generador para esta distribución, cuya utilidad reside en los métodos de estimación de densidades a posteriori que se verán más adelante.

[Generación de muestras de la distribución Beta Simétrica $(-1, 1)(a)$]

1. Generar U de una Uniforme $U(0,1)$
2. Generar V de una Beta $B(\frac{1}{2}, a)$
3. Si $U \leq \frac{1}{2}$, Entonces $\beta \leftarrow \sqrt{V}$. Si no, $\beta \leftarrow -\sqrt{V}$

Teorema 1.4. Las muestras obtenidas provienen de la distribución Beta Simétrica presentada.

Demostración-. De la forma simétrica de la distribución se deduce que basta encontrar un generador para la variable restringida al intervalo $(0,1)$, sorteando después con idéntica probabilidad la muestra obtenida entre los intervalos $(-1,0)$ y $(0,1)$. Para obtener una muestra de la distribución restringida a $(0,1)$, basta realizar el cambio de variable $v = \beta^2$ en la función de densidad, lo que da lugar a una distribución $\beta(\frac{1}{2}, a)$ para V . El cambio se deshace en el algoritmo, con $\beta = \sqrt{v}$. En el algoritmo, se asigna el signo positivo o negativo con idéntica probabilidad a través de la distribución Uniforme.

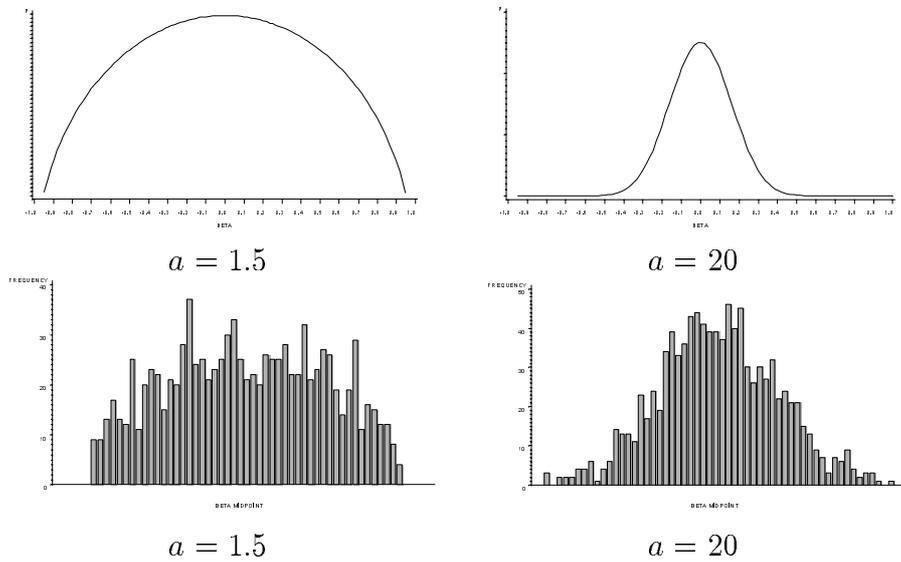


Figura 1.7. Distribución Beta Simétrica(-1,1)

Aunque la idea original de adoptar esta distribución radicaba en que la moda $\beta = 0$ estaba situada en el centro de un intervalo con dos distribuciones de referencia en $\beta \rightarrow -1$ y $\beta = 1$ el dominio de β se puede extender a $(-1, \infty)$, adoptando los datos la distribución Uniforme en el límite en ambos extremos. Para esta extensión se puede adoptar un modelo Gamma trasladado, con $\beta = \beta' - 1$, donde $\beta' \equiv \Gamma(a, b)$. Los parámetros a y b se pueden escoger de manera que la moda de β' esté en el valor $\beta' = 1$, correspondiente a $\beta = 0$ y por lo tanto a Normalidad en los datos. Esto equivale a tomar como distribución para β' $p(\beta') \equiv \Gamma(a, a + 1)$, con $a > 0$.

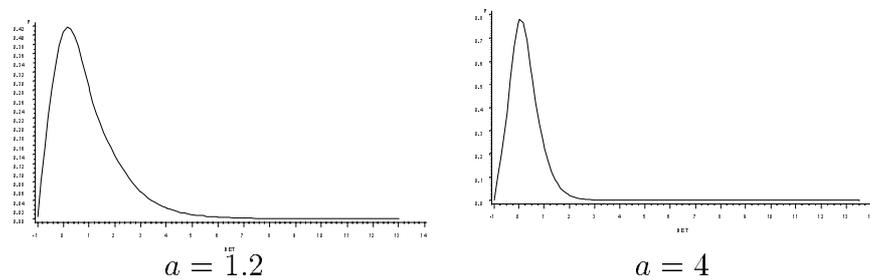


Figura 1.8. Distribución Gamma trasladada, moda en 0

2.6 Distribuciones a posteriori

En lo sucesivo supondremos que β es una variable aleatoria con distribución a priori $p(\beta)$, θ es una variable aleatoria con distribución a priori $p(\theta)$ y σ es una variable aleatoria con distribución a priori impropia $p(\sigma) \propto \frac{1}{\sigma}$, siendo las tres distribuciones sobre los parámetros independientes y por lo tanto $p(\theta, \sigma, \beta) = \frac{1}{\sigma} p(\theta) p(\beta)$.

En estas condiciones, si $\mathbf{y} = (y_1, \dots, y_n)$ es una muestra aleatoria simple de una Potencial Exponencial $PE(\theta, \sigma, \beta)$, la distribución conjunta a posteriori de (θ, σ, β) , vendrá dada, mediante

$$p(\theta, \sigma, \beta | \mathbf{y}) = \frac{1}{p(\mathbf{y})} p(\theta) p(\beta) [w(\beta)]^n \sigma^{-(n+1)} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)} \right]$$

siendo $p(\mathbf{y})$ la constante integradora, $p(\mathbf{y}) = \int p(\theta, \sigma, \beta | \mathbf{y}) d\theta d\sigma d\beta$. En lo sucesivo, utilizaremos generalmente la expresión de proporcionalidad

$$p(\theta, \sigma, \beta | \mathbf{y}) \propto p(\theta) p(\beta) [w(\beta)]^n \sigma^{-(n+1)} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)} \right] \quad (1.5)$$

Nos interesará, en particular, obtener las distribuciones a posteriori marginales $p(\theta | \mathbf{y})$, $p(\beta | \mathbf{y})$ y eventualmente $p(\sigma | \mathbf{y})$. Al no existir distribuciones conjugadas conocidas para β y θ será necesario establecer métodos numéricos que nos lleven a correctas aproximaciones de estas distribuciones.

Cuando se utiliza la densidad a priori $p(\sigma)$, la densidad a posteriori se puede reducir en σ como veremos en el siguiente resultado.

Teorema 1.5.

$$p(\theta, \beta | \mathbf{y}) = \int p(\theta, \sigma, \beta | \mathbf{y}) d\sigma \propto p(\theta) p(\beta) [w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} \Gamma(n(1+\beta)/2) (1+\beta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n\frac{(1+\beta)}{2}} \quad (1.6)$$

Demostración-

En la integral de la expresión (1.5) se aplica que:

Para $a > 0, p > 0, \alpha > 0$,

$$\int_0^{\infty} x^{-(p+1)} e^{-ax^{-\alpha}} dx = \frac{1}{\alpha} a^{-p/\alpha} \Gamma\left(\frac{p}{\alpha}\right)$$

Tomando $p = n$, $a = c(\beta) \sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)}$, y $\alpha = \frac{2}{1+\beta}$ se obtiene el resultado del enunciado \diamond

Esta reducción del parámetro de escala σ proviene del hecho de utilizar la distribución a priori $p(\sigma) = \frac{1}{\sigma}$ para σ , y permitirá abordar con más sencillez problemas de Inferencia sobre los parámetros θ y β .

Surge la pregunta de si para todas las distribuciones $p(\theta)$ y $p(\beta)$ esta distribución es propia, pues hemos integrado respecto del parámetro σ , que tenía distribución $p(\sigma) = \frac{1}{\sigma}$ impropia.

Teorema 1.6.

Si $p(\beta)$ es una distribución continua y acotada en $(-1, 1)$ y $p(\theta)$ es una distribución continua y acotada en $(-\infty, +\infty)$, la expresión $p(\theta, \beta | \mathbf{y})$ que aparece en (1.4) es integrable en θ y β y por lo tanto $p(\theta, \beta | \mathbf{y})$ es una densidad a posteriori propia.

Demostración-

$$\begin{aligned} \text{Tenemos } \int \int p(\theta, \beta | \mathbf{y}) d\beta d\theta &\propto \\ \int_{\theta \in (-\infty, +\infty)} \int_{\beta \in (-1, 1)} p(\theta) p(\beta) [w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} \Gamma(n(1+\beta)/2) (1+\beta) & \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}} \end{aligned}$$

Como función de β , $p(\theta, \beta | \mathbf{y})$ es continua en el intervalo $(-1, 1)$ para todo θ ,

$$\lim_{\beta \rightarrow -1} [w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} = \lim_{\beta \rightarrow -1} 2^{-n} \Gamma\left(\frac{3+\beta}{2}\right)^{-n} = (2\Gamma(1))^{-n} \text{ acotado,}$$

y además el término $\left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}}$ está acotado para todo θ :

$$k_1 < \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}} < k_2$$

con

$$\begin{aligned} k_1 &= \text{Min}\{(n | \text{Max}\{y_i\} - \theta|)^{-n}, (n | \text{Min}\{y_i\} - \theta|)^{-n}\} \\ k_2 &= \text{Max}\{(n | \text{Max}\{y_i\} - \theta|)^{-n}, (n | \text{Min}\{y_i\} - \theta|)^{-n}\} \end{aligned}$$

Por lo tanto su límite en $\beta \rightarrow -1$ es finito para todo θ .

Como los demás términos en β están acotados o son nulos, pues $\Gamma(n(1+\beta)/2)(1+\beta) = 0$, entonces $\lim_{\beta \rightarrow -1} p(\theta, \beta | \mathbf{y}) = 0$, de modo que la integral en $\beta \in (-1, 1)$ existe para todo θ .

Por otra parte, al ser $p(\theta)$ acotado, el término $\left[\sum_{i=1}^n |y_i - \theta|^{\frac{2}{1+\beta}} \right]^{-n \frac{(1+\beta)}{2}} \cdot p(\theta)$ está acotado, para todo β , por una función de orden $\frac{K}{\theta^n}$, con $n > 1$, que es integrable en $\theta \in (-\infty, +\infty)$. De manera que la integral en $\theta \in (-\infty, +\infty)$ existe para todo β .

Así, la integral doble existe y $p(\theta, \beta | \mathbf{y})$ es una distribución a posteriori propia. \diamond

En algunas aplicaciones se utiliza como distribución de θ $p(\theta) \propto \text{constante}$, en un entorno suficientemente grande, $[-M, M]$ de manera que su distribución es aproximadamente no informativa. En este caso el término en θ en la integral es continuo y está acotado en el intervalo acotado $[-M, M]$ con lo cual la integral doble existe igualmente.

2.6.1 Aproximaciones e Integración Numérica

Para obtener $p(\theta | \mathbf{y})$ y $p(\beta | \mathbf{y})$ hay que integrar la expresión (1.4) en β o θ respectivamente. Al no existir primitivas para estas integrales es necesario adoptar soluciones numéricas. Por ejemplo, para calcular $p(\theta | \mathbf{y})$ para varios valores de θ es necesario aproximar numéricamente la integral

$$\int p(\theta)p(\beta)[w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} \Gamma(n(1+\beta)/2)(1+\beta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}} d\beta$$

para cada valor diferente de θ para el cual queramos calcular $p(\theta | \mathbf{y})$. Esta última evaluación requeriría por ejemplo, métodos de integración numérica en β para cada valor de θ . Análogamente, si queremos calcular $p(\beta | \mathbf{y})$ deberemos evaluar numéricamente la integral

$$\int p(\theta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}} d\theta$$

para cada valor de β en el que estemos interesados,. Estas tareas de integración numérica se complican dependiendo de las distribuciones a priori que se adopten para θ y β , y del tamaño muestral n .

(Box, Tiao, 1973) se centran sobre todo en la distribución $p(\theta | \mathbf{y})$, solucionando el problema mediante integración numérica sobre el parámetro β , sin detenerse en el cálculo de la constante normalizadora $\frac{1}{p(\mathbf{y})}$. El mayor esfuerzo de su estudio se centra en el cálculo de aproximaciones a $p(\theta | \beta, \mathbf{y})$ donde β está considerado conocido. En este caso, proponen la aproximación

$$p(\theta | \beta, \mathbf{y}) \cong C \cdot \left[\sum_{i=1}^n \left| y_i - \hat{\theta} \right|^{2/(1+\beta)} + d(\theta - \hat{\theta})^2 \right]^{-\frac{1}{2}n(1+\beta)}$$

donde $\hat{\theta}$ es la moda aproximada de $p(\theta | \beta, \mathbf{y})$, y d y C son constantes calculadas apropiadamente.

En cualquier caso, en las aplicaciones prácticas que abordan comparan resultados con varios β fijos, y raramente tratan el problema de inferencia con los tres parámetros como aleatorios, enfoque que seguiremos en este trabajo.

2.6.2 Aproximaciones mediante Monte-Carlo

El método Monte-Carlo de integración permitirá abordar los problemas de integración numérica planteados en el cálculo de las distribuciones a posteriori sobre los tres parámetros, de una manera económica en cuanto a cálculos computacionales, a pesar de la necesidad de tomar un tamaño muestral suficientemente grande en las simulaciones. Dado el problema de estimar la integral

$$I(y) = \int p(y | \theta) p(\theta) d\theta$$

donde tanto y como θ pueden ser vectores, y, supuesto que se sabe extraer muestras aleatorias de $p(\theta)$, el método Monte-Carlo de integración toma la siguiente forma:

[Integración Monte-Carlo]

1. Generar una muestra $\theta_1, \theta_2, \dots, \theta_m$ de $p(\theta)$
2. Calcular $\hat{I}(y) \cong \frac{1}{m} \sum_{i=1}^m p(y | \theta_i)$

En (Geweke, 1989) se puede ver que $\hat{I}(y) \xrightarrow{c.s.} I(y)$.

Una estimación de la varianza del estimador Monte-Carlo es :

$$\widehat{V}(\hat{I}(y)) = \frac{1}{m(m-1)} \sum_{i=1}^m (p(y | \theta_i) - \hat{I}(y))^2$$

Se puede presentar un intervalo de confianza clásico para el verdadero valor de la integral :

$$(\hat{I}(y) - z_{\alpha/2} \sqrt{\widehat{V}(\hat{I}(y))}, \hat{I}(y) + z_{\alpha/2} \sqrt{\widehat{V}(\hat{I}(y))})$$

En la práctica suele ser necesario recurrir a tamaños muestrales m grandes para obtener buenas estimaciones; es necesario que los θ simulados cubran suficientemente el dominio de definición de θ (en los métodos de integración numérica las abscisas o puntos de evaluación están cuidadosamente escogidos, mientras que en el método de Monte-Carlo son valores aleatorios simulados).

Veremos a continuación una aplicación de este método al problema que nos ocupa. Partiendo de (1.4), sabemos que

$$\begin{aligned} p(\theta | \mathbf{y}) &\propto \int p(\theta, \beta | \mathbf{y}) d\beta \propto \\ &\propto \int p(\theta) p(\beta) [w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} \Gamma(n(1+\beta)/2) (1+\beta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}} d\beta \end{aligned}$$

La estimación de la integral via Monte-Carlo se reduce por lo tanto a extraer m muestras independientes de $p(\beta)$ y calcular la media del resto del integrando en esas m muestras de β .

Para el cálculo de $p(\beta | \mathbf{y})$ se emplea:

$$p(\beta | \mathbf{y}) \propto \int p(\theta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n \frac{(1+\beta)}{2}} d\theta$$

y para $p(\sigma | \mathbf{y})$:

$$p(\sigma | \mathbf{y}) \propto \int p(\mathbf{y} | \theta, \beta, \sigma) p(\theta) p(\beta) p(\sigma) d\theta d\beta$$

Si hiciera falta calcular la constante de normalización $p(\mathbf{y})$ se podrían utilizar métodos de integración numérica, o bien incluso con el mismo método de Monte-Carlo, extrayendo valores de $p(\theta, \beta)$ y estimando $\frac{1}{\hat{p}(\mathbf{y})}$ a través de la expresión

$$\frac{1}{\hat{p}(\mathbf{y})} = \int p(\theta)p(\beta)[w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} \Gamma(n(1+\beta)/2)(1+\beta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n\frac{(1+\beta)}{2}} d\theta d\beta$$

Dada la muestra aleatoria simple $\mathbf{y} = (y_1, \dots, y_n)$ y las distribuciones a priori sobre los parámetros $p(\theta)$, $p(\beta)$, y $p(\sigma) = \frac{1}{\sigma}$, un método Monte-Carlo para estimar las distribuciones a posteriori en la Potencial Exponencial es presentado a continuación. Para simplificar las expresiones, llamaremos $h(\mathbf{y}, \theta, \beta)$ a

$$h(\mathbf{y}, \theta, \beta) = [w(\beta)]^n c(\beta)^{-\frac{n(1+\beta)}{2}} \Gamma(n(1+\beta)/2)(1+\beta) \left[\sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)} \right]^{-n\frac{(1+\beta)}{2}}$$

[Estimación Monte-Carlo de distribuciones a posteriori en la distribución Potencial Exponencial]

1. Generar m muestras $(\theta, \beta)_i$ independientes de $p(\theta, \beta) = p(\theta)p(\beta)$.
2. Calcular

$$\begin{aligned} \hat{p}(\theta | y) &\propto \frac{1}{m} \sum_{i=1}^m h(y, \theta, \beta_i) p(\theta) \\ \hat{p}(\beta | y) &\propto \frac{1}{m} \sum_{i=1}^m h(y, \theta_i, \beta) p(\beta) \\ \hat{p}(\sigma | y) &\propto \frac{1}{m^2} \frac{1}{\sigma} \sum_{i=1}^m \sum_{j=1}^m p(y | \sigma, (\theta_i, \beta_j)) \\ \hat{p}(y) &= \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m h(y, \theta_i, \beta_j) \end{aligned}$$

Este método permitirá abordar los problemas de inferencia basados en la distribución Potencial Exponencial con las siguientes consideraciones:

- Generando m muestras de los pares (θ, β) obtenemos estimaciones proporcionales a las distribuciones a posteriori $p(\beta | \mathbf{y})$, $p(\theta | \mathbf{y})$ y $p(\sigma | \mathbf{y})$. Estos cálculos permiten evaluar gráficamente estas distribuciones, además de una estimación de las probabilidades obtenida tras el cálculo de la constante normalizadora por integración numérica trapezoidal en una rejilla de puntos escogidos. Se podrán realizar inferencias a través de estos resultados, así como estudios de sensibilidad del modelo tales como realizar análisis de robustez de la distribución a posteriori respecto de la distribución a priori para cada parámetro, como, por ejemplo, evaluar el efecto que puede tener sobre $p(\beta | \mathbf{y})$ el hecho de plantear una determinada distribución a priori $p(\theta)$ sobre el parámetro de centralización.
- Para poder utilizar este método es necesario conocer la forma analítica de las distribuciones a priori $p(\theta)$ y $p(\beta)$, así como saber extraer muestras de ellas. Se ha supuesto además la restricción $p(\sigma) = \frac{1}{\sigma}$. En caso de tener σ otra distribución a priori se puede extender el método. Existe también la posibilidad de utilizar aproximaciones en el caso de tomar $p(\sigma)$ como una Gamma inversa con ciertos parámetros que la hagan aproximadamente no informativa.
- Los métodos Monte-Carlo permiten aumentar la precisión de las estimaciones aumentando m , siendo conocida una medida del error estándar cometido en la estimación. Para aumentar la precisión, si en lugar de ello utilizamos técnicas de integración numérica aplicadas a nuestro problema de inferencia, es necesario aumentar el número de abscisas de evaluación en cada integral numérica a aproximar; en cambio, con Monte-Carlo, aumentar m redundará simultáneamente en una precisión mayor en todas las aproximaciones a $p(\mathbf{y})$, $p(\theta | \mathbf{y})$, $p(\beta | \mathbf{y})$, y $p(\sigma | \mathbf{y})$.

Para estudiar el tamaño m apropiado para obtener aproximaciones estables de las probabilidades estudiadas se ha realizado un estudio de simulación a partir de diferentes tamaños muestrales n provenientes de una $N(0, 1)$. Se observa en la Figura 1.9 como la reducción en la varianza del estimador es mucho menor a partir de muestras de tamaño $m^2 = 360000$.

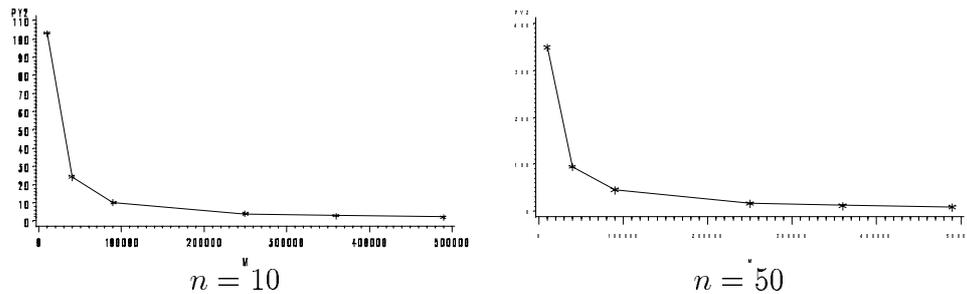


Figura1.9.Evolución de la varianza estimada de $\hat{p}(y)$ para diferentes tamaños de simulación m

Ejemplo 1.1 (Box, Tiao, 1973) aplicaron la distribución Potencial Exponencial al estudio de datos y tomados por Darwin. Cada valor y representa la diferencia en altura (octavos de pulgada) entre un par de plantas, cuidadas ambas en las mismas condiciones pero con distintos tipo de abono: $y = -67, -48, 6, 8, 14, 16, 23, 24, 28, 29, 41, 49, 56, 60, 75$.

En este caso el tamaño muestral es pequeño, $n = 15$; los valores -67 y -48 tendrán repercusiones en la distribución a posteriori del parámetro de centralización θ , en diferente medida según se adopte un modelo más o menos resistente a Outliers. El coeficiente de apuntamiento es de 1.4; cabe esperar una distribución a posteriori para β centrada en valores positivos de β .

Estamos interesados en observar las distribuciones a posteriori de los tres parámetros θ, σ, β . Para observar la influencia de las distribuciones a priori sobre el modelo final plantearemos diferentes modelos informativos y no informativos sobre θ y β . En primer lugar planteamos como distribuciones a priori $p(\sigma) = \frac{1}{\sigma}$, $p(\theta) \equiv N(15, 5)$ y $p(\beta) \equiv B(-1, 1)$ simétrica con parámetro $a = 3$; es decir, asumimos cierta información inicial sobre la centralización de los datos, y asumimos también, a través de la distribución de β , con moda en $\beta = 0$, la distribución Normal como la distribución más probable a priori en cuanto a apuntamiento se refiere.

Hemos tomado $m = 500000$ para obtener un estimador de $p(y)$ que se estabilice a lo largo de la secuencia de simulación. Para estimaciones de las diferentes distribuciones

se ha tomado un tamaño $m = 1000$, para θ y β y $m = 25000$ para σ . Los cálculos de las distribuciones a posteriori se realizan a través de una curva suavizada basada en interpolación cúbica por splines de los puntos dados en las coordenadas (x, y) .

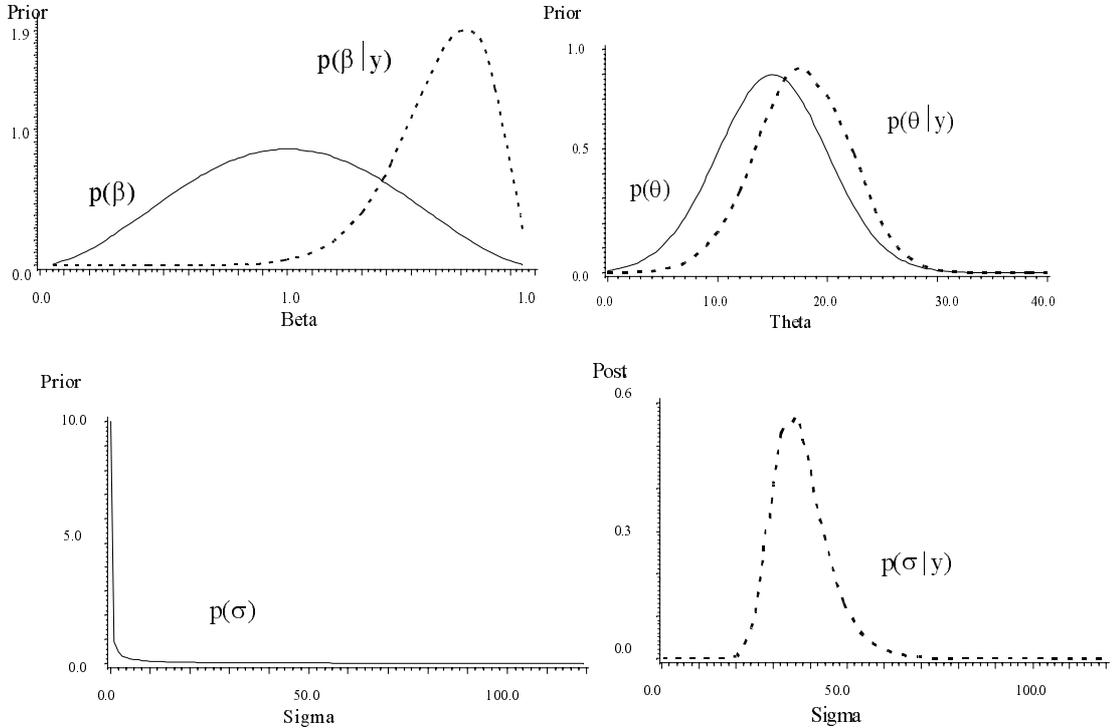


Figura 1.10. Distribuciones a posteriori $p(\beta | y)$, $p(\theta | y)$ y $p(\sigma | y)$

Se observa en la Figura 1.10 cómo la distribución a posteriori $p(\beta | y)$ se desplaza respecto de $p(\beta)$ hacia una distribución con moda en 0.7, alejándose por lo tanto de la hipótesis de Normalidad de la variable Y . La distribución a posteriori de θ se desplaza hacia la derecha; $p(\theta | y)$ no sufre una traslación mayor seguramente a causa de la fuerte información a priori planteada, respecto del tamaño de los datos, $n = 15$. Por último vemos cómo la distribución a posteriori de σ , inicialmente de la forma $\frac{1}{\sigma}$, deriva en una distribución unimodal con moda en $\sigma = 37$, que coincide con la desviación típica muestral s_y .

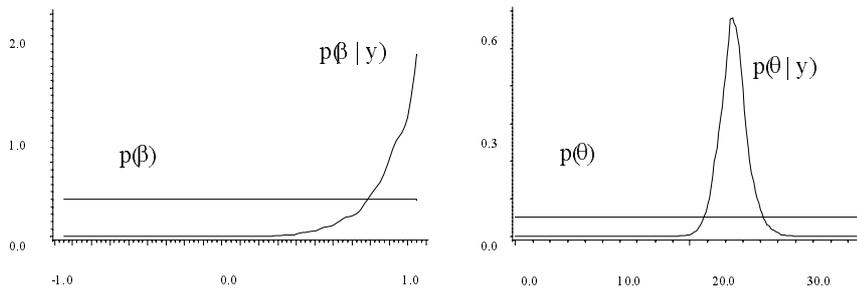


Figura 1.11. Distribuciones a posteriori $p(\beta | y)$ y $p(\theta | y)$ para distribuciones a priori no informativas

En la Figura 1.11 se han tomado diferentes distribuciones a priori no informativas sobre los parámetros θ y β . En particular es interesante observar cómo $p(\beta | \mathbf{y})$ toma la forma de una distribución con moda en el extremo $\beta = 1$, y $p(\theta | \mathbf{y})$ es una distribución unimodal con moda en $\theta = 26$, correspondiente a la media muestral $\bar{y} = 26$. El hecho de no suponer información sobre β lleva a una distribución a posteriori $p(\beta | \mathbf{y})$ muy apuntada, afectada por la curtosis de la muestra, mientras que $p(\theta | \mathbf{y})$ toma también una forma apuntada en torno a la media muestral. Siendo $p(\beta | \mathbf{y})$ muy sensible al apuntamiento de la muestra cuando se utiliza una distribución a priori sobre β no informativa, adoptaremos este modelo a priori en lo sucesivo.

2.6.3 Generación de muestras de $p(\theta, \sigma, \beta | \mathbf{y})$ a través del muestreo de Gibbs

En este apartado utilizaremos $c(\beta) = \frac{1}{2}$ con el fin de simplificar los algoritmos de simulación.

Sabemos que la distribución a posteriori $p(\theta, \sigma, \beta | \mathbf{y})$ toma la forma, cuando $c(\beta) = \frac{1}{2}$, de

$$p(\theta, \sigma, \beta | \mathbf{y}) \propto [w(\beta)]^n \sigma^{-n} \exp \left[-\frac{1}{2} \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta, \sigma, \beta)$$

donde

$$w(\beta) = \frac{1}{2^{\frac{3+\beta}{2}} \Gamma[\frac{1}{2}(3 + \beta)]}$$

Para extraer muestras de $p(\theta, \sigma, \beta | \mathbf{y})$ podemos utilizar el muestreo de Gibbs (Geman y Geman, 1984). Si el problema es por ejemplo, extraer muestras de una distribución a

posteriori $p(\theta_1, \theta_2, \theta_3, \dots | \mathbf{y})$ y se sabe como extraer muestras de la distribución condicional para cada θ_i , el método es el siguiente:

[Algoritmo de Muestreo de Gibbs]

1. Tomar valores iniciales $\theta_2^{(0)}, \theta_3^{(0)}, \dots$
2. Generar una muestra $\theta_1^{(1)}$ de $p(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \mathbf{y})$
3. Generar una muestra $\theta_2^{(1)}$ de $p(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \mathbf{y})$
4. Generar una muestra θ_3 de $p(\theta_3 | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \mathbf{y})$
5. ...
6. Continuar el algoritmo de sustitución hasta un número de iteraciones m suficientemente alto.

Las muestras obtenidas provienen de la distribución $p(\theta_1, \theta_2, \theta_3, \dots | \mathbf{y})$.

Existen muchas maneras de seleccionar las muestras obtenidas el algoritmo, piedra angular de los llamados métodos MCMC (Markov Chain Monte-Carlo). A título ilustrativo nombraremos solamente algunas. Una revisión del procedimiento puede verse en (Chen, Shao, Ibrahim, 1999):

1. Correr el algoritmo una sola vez hasta un número suficientemente alto de iteraciones m y registrar solamente las muestras obtenidas a partir de esta iteración m . Este método tiene la desventaja de que pueda existir correlación entre muestras consecutivas.
2. Correr el algoritmo k veces en paralelo hasta un número de iteraciones m y registrar la última muestra obtenida de cada secuencia. Este método es más costoso computacionalmente, y puede depender demasiado de los valores iniciales si m no es muy alto.
3. Correr el algoritmo una sola vez hasta un número muy alto de iteraciones M registrando la muestra obtenida cada k iteraciones. Aunque el salto entre registros elimina la correlación consecutiva, puede haber problemas de dependencia de la secuencia particular.

En nuestro caso particular, utilizaremos el algoritmo de Gibbs extrayendo sucesivamente muestras de $p(\theta | \sigma, \beta, \mathbf{y})$, $p(\sigma | \theta, \beta, \mathbf{y})$ y $p(\beta | \theta, \sigma, \mathbf{y})$. Para cada una de estas distribuciones será necesario disponer de un método de simulación.

Generación de muestras de $p(\sigma | \theta, \beta, \mathbf{y})$.

En este caso la extracción es directa en el entorno que hemos definido con $p(\sigma) \propto \sigma^{-1}$, pues

$$p(\sigma | \theta, \beta, \mathbf{y}) \propto \sigma^{-(n+1)} \exp \left[-\left(\frac{1}{2} \sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)}\right) \frac{1}{\sigma^{2/(1+\beta)}} \right]$$

Haciendo el cambio de variable $z = \sigma^{2/(1+\beta)}$ tenemos que $z \propto IG(a, b)$ donde $a = \frac{n(1+\beta)}{2}$ y $b = \left(\frac{1}{2} \sum_{i=1}^n |y_i - \theta|^{2/(1+\beta)}\right)$ con lo que un método directo para extraer muestras de $p(\sigma | \theta, \beta, \mathbf{y})$ será extraer una muestra x de una Gamma $\Gamma(a, b)$ y transformarla mediante $\sigma = x^{-(1+\beta)/2}$.

Generación de muestras de $p(\theta | \sigma, \beta, \mathbf{y})$.

En este caso tenemos

$$p(\theta | \sigma, \beta, \mathbf{y}) \propto \exp \left[-\frac{1}{2} \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta)$$

Se puede verificar que para $k \geq 1$, es $\sum_{i=1}^n |\theta - y_i|^k \geq n |\theta - \bar{y}|^k$ de lo cual se deduce que

$$\exp \left[-\frac{1}{2} \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta) \leq \exp \left[-\frac{n}{2} \left| \frac{\bar{y} - \theta}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta)$$

El primer término de la acotación superior toma la forma de una distribución Potencial Exponencial, con lo cual podemos emplear el método de rechazo para extraer muestras de θ . En nuestro caso aplicaremos el algoritmo: bien utilizando la distribución Potencial Exponencial como base, en caso de que ese término sea dominante, por ejemplo cuando $p(\theta)$ sea no informativa, bien utilizando $p(\theta)$ como base.

[Algoritmo de rechazo para simular muestras de $p(\theta | \sigma, \beta, \mathbf{y})$]

1. Extraer una muestra θ de una PE($\bar{y}, \frac{\sigma}{n(1+\beta)/2}, \beta$) [alternativamente., extraer una muestra θ de $p(\theta)$]
2. Tomar $\log t = -\frac{n}{2} \left| \frac{\bar{y} - \theta}{\sigma} \right|^{2/(1+\beta)} + \frac{1}{2} \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)}$

3. Extraer u de una $U(0, 1)$
4. Si $\log(u) \leq -\log t$, Aceptar θ . En caso contrario, rechazar θ y volver al paso 1

Este algoritmo de rechazo funciona relativamente bien para muestras moderadas, debido sobre todo a la similitud entre las colas de la distribución que acota y la distribución objetivo.

Generación de muestras de $p(\beta \mid \theta, \sigma, \mathbf{y})$

En este caso tendremos

$$p(\beta \mid \theta, \sigma, \mathbf{y}) \propto \frac{1}{2} \frac{1}{n(3 + \beta)} \frac{1}{[\Gamma[\frac{1}{2}(3 + \beta)]]^n} \exp \left[-\frac{1}{2} \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)} \right] p(\beta)$$

Esta expresión está acotada superiormente por

$$\frac{1}{2^{3n/2}} \frac{1}{[\Gamma(\frac{3}{2})]^n} \exp(-(\frac{n}{2} \log(2))\beta) p(\beta)$$

con lo cual podemos utilizar otra vez el método de rechazo para generar muestras de $p(\beta \mid \theta, \sigma, \mathbf{y})$, análogamente al caso anterior, utilizando alternativamente la distribución exponencial truncada o bien $p(\beta)$ como base para el algoritmo.

[Algoritmo de rechazo para simular muestras de $p(\beta \mid \theta, \sigma, \mathbf{y})$]

1. Extraer una muestra β de una distribución exponencial $\text{Exp}(\frac{n}{2} \log(2))$ truncada en $\beta \in (-1, 1]$ [o bien, alternativamente, extraer una muestra de $p(\beta)$]
2. Tomar $\log t = -n \log(\Gamma(3/2)) - \frac{3n}{2} \log(2) + \frac{1}{2} \sum_{i=1}^n \left| \frac{y_i - \theta}{\sigma} \right|^{2/(1+\beta)}$
3. Extraer u de una $U(0, 1)$
4. Si $\log(u) \leq -\log t$, Aceptar β . En caso contrario, rechazar β y volver al paso 1

Se presenta a continuación un método para generar muestras de la distribución exponencial truncada, necesario en el algoritmo anterior:

[Algoritmo para generar muestras de la distribución exponencial truncada $Exp(\lambda)_{a_1 < X < a_2}$]

1. Calcular $k = e^{-\lambda a_1} - e^{-\lambda a_2}$.
2. Generar U de una Uniforme (0,1) . Calcular $X = -\frac{1}{\lambda} \ln(e^{-\lambda a_1} - kU)$.

Teorema 1.6

Las muestras obtenidas provienen de una distribución Exponencial $Exp(\lambda)$ truncada en el intervalo (a_1, a_2) .

Demostración

El método de inversión para generar muestras de una variable aleatoria X (Devroye, 1984) requiere conocer la forma analítica de la función de distribución $F(x)$ y saber obtener una solución única a la ecuación $x = F^{-1}(y)$. El método consiste en

[Método de Inversión]

1. Simular U de una distribución Uniforme $U(0, 1)$.
2. Hacer $x = F^{-1}(u)$.

En nuestro caso, la densidad de la exponencial truncada es

$$p(x) = \frac{\exp(-\lambda x)}{\int_{a_1}^{a_2} \exp(-\lambda x) dx} \text{ para } x \in (a_1, a_2)$$

$$\text{Como } \int_{a_1}^{a_2} \exp(-\lambda x) dx = \frac{1}{\lambda} (e^{-\lambda a_1} - e^{-\lambda a_2})$$

$$F(x) = \frac{1}{\frac{1}{\lambda} (e^{-\lambda a_1} - e^{-\lambda a_2})} \int_{a_1}^x \exp(-\lambda x) dx = \frac{e^{-\lambda a_1} - e^{-\lambda x}}{(e^{-\lambda a_1} - e^{-\lambda a_2})}$$

De lo que se deduce que la solución a la ecuación $u = F(x)$ es

$$x = -\frac{1}{\lambda} \ln(e^{-\lambda a_1} - ku) \diamond$$

Utilizando el muestreo de Gibbs, es decir, simulando alternadamente muestras de $p(\theta | \sigma, \beta, \mathbf{y})$, $p(\sigma | \theta, \beta, \mathbf{y})$ y $p(\beta | \theta, \sigma, \mathbf{y})$, obtendremos una muestra de $p(\theta, \sigma, \beta | \mathbf{y})$.

2.6.4 Generación de muestras de $p(\theta, \sigma, \beta | \mathbf{y})$ a través de la representación por mixturas

Hemos visto que esta familia se puede representar de la siguiente manera:

$$y | \theta, \sigma, \beta, u \equiv U\left(\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u^{\frac{\beta+1}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u^{\frac{\beta+1}{2}}\right)$$

con

$$u | \beta \equiv \Gamma\left(\frac{1}{2}(3 + \beta), 1\right)$$

La función de verosimilitud de una muestra \mathbf{y} , dado el vector de parámetros de mixtura $\mathbf{u} = (u_1, u_2, \dots, u_n)$, será

$$p(\mathbf{y} | \theta, \sigma, \beta, \mathbf{u}) = (2\sigma)^{-n} c(\beta)^{-\frac{n(\beta+1)}{2}} \prod_{i=1}^n u_i^{-1},$$

con $y_i \in \left[\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}}\right]$.

La implementación de esta representación para obtener muestras a posteriori de $(\theta, \sigma, \beta | \mathbf{y})$ mediante el muestreo de Gibbs es complicada, pues es necesario extraer muestras de $(\beta | \theta, \sigma, \mathbf{u}, \mathbf{y})$ en un paso del algoritmo.

Plantaremos las distribuciones a priori $p(\theta)$ y $p(\beta)$, y supondremos $p(\sigma)$ como una Gamma Inversa $GI(a, b)$ con los parámetros a y b tales que sea aproximadamente no informativa, para actuar en consonancia con el planteamiento $p(\sigma) = \frac{1}{\sigma}$ de las secciones anteriores.

La implementación del muestreo de Gibbs se realiza a través de las distribuciones condicionales de los parámetros θ, β, σ , además de los parámetros de mixtura u_i . En el resultado siguiente se expresan la forma proporcional a las distribuciones de interés.

Teorema 1.7.

Las distribuciones condicionales obtenidas con la representación por mixturas toman la forma

$$\begin{aligned}
\theta \mid \sigma, \beta, \mathbf{u}, \mathbf{y} &\equiv p(\theta), \theta \in \left[\max_i \left\{ y_i - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right\}, \min_i \left\{ y_i + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right\} \right] \\
\sigma \mid \theta, \beta, \mathbf{u}, \mathbf{y} &\equiv IG(a+n, b), \sigma > \max_i \left\{ \left(\frac{c(\beta)}{u_i} \right)^{\frac{\beta+1}{2}} |y_i - \theta| \right\} \\
\beta \mid \theta, \sigma, \mathbf{u}, \mathbf{y} &\propto \frac{p(\beta)c(\beta)^{\frac{n(1+\beta)}{2}}}{\Gamma\left(\frac{1}{2}(3+\beta)\right)^n}, \beta \text{ tal que } \frac{u_i^{\frac{\beta+1}{2}}}{c(\beta)^{\frac{\beta+1}{2}}} > \left(\frac{y_i - \theta}{\sigma} \right)^2, \forall i \\
u_i \mid \theta, \sigma, \beta, \mathbf{y} &\propto Exp(1), \frac{u_i}{c(\beta)} > \left(\frac{(y_i - \theta)^2}{\sigma^2} \right)^{\frac{1}{1+\beta}}
\end{aligned}$$

Demostración

$$p(\theta \mid \sigma, \beta, \mathbf{u}, \mathbf{y}) \propto p(\mathbf{y} \mid \theta, \sigma, \beta, \mathbf{u})p(\theta) \propto p(\theta)$$

$$\text{Al ser } y_i \in \left[\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right] \implies$$

$$\implies y_i - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} < \theta < y_i + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \forall i$$

$$\implies \theta \in \left[\max_i \left\{ y_i - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right\}, \min_i \left\{ y_i + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right\} \right]$$

$$\text{Por otra parte, } p(\sigma \mid \theta, \beta, \mathbf{u}, \mathbf{y}) \propto p(\mathbf{y} \mid \theta, \sigma, \beta, \mathbf{u})p(\sigma) \propto p(\sigma)$$

$$\text{Como } y_i \in \left[\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right] \implies |y_i - \theta| < \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \implies \sigma >$$

$$\max_i \left\{ \left(\frac{c(\beta)}{u_i} \right)^{\frac{\beta+1}{2}} |y_i - \theta| \right\}$$

En el caso del parámetro β , tenemos

$$p(\beta \mid \theta, \sigma, \mathbf{u}, \mathbf{y}) \propto p(\mathbf{y} \mid \theta, \sigma, \beta, \mathbf{u})p(\beta) \propto \frac{p(\beta)c(\beta)^{\frac{n(1+\beta)}{2}}}{\Gamma\left(\frac{1}{2}(3+\beta)\right)^n}$$

Dado que $y_i \in \left[\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right] \implies \frac{u_i^{\frac{\beta+1}{2}}}{c(\beta)^{\frac{\beta+1}{2}}} > \left(\frac{y_i - \theta}{\sigma} \right)^2 \forall i$ que es la condición impuesta.

Por último, como

$$p(y_i | \theta, \sigma, \beta, u_i) = \frac{2c(\beta)^{\frac{\beta+1}{2}}}{\sigma u_i^{\frac{\beta+1}{2}}}, \text{ con } y_i \in \left(\theta - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}}, \theta + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} \right)$$

$$\implies p(u_i | \theta, \sigma, \beta, \mathbf{y}) = p(u_i | \theta, \sigma, \beta, y_i) \propto p(y_i | \theta, \sigma, \beta, u_i) p(u_i | \beta) \propto$$

$$\propto \frac{1}{u_i^{\frac{\beta+1}{2}}} \cdot u_i^{\frac{\beta+1}{2}} e^{-u_i} \propto \text{Exp}(1), \text{ con } \frac{u_i}{c(\beta)} > \left(\frac{(y_i - \theta)^2}{\sigma^2} \right)^{\frac{1}{1+\beta}} \diamond$$

La implementación de m iteraciones del algoritmo de Gibbs, dada la muestra $\mathbf{y} = (y_1, y_2, \dots, y_n)$ toma la siguiente forma:

[Algoritmo de Gibbs para generar muestras de $p(\beta, \theta, \sigma | u, y)$]

1. Extraer una primera muestra de las distribuciones a priori $\theta_0 \sim p(\theta)$, $\sigma_0 \sim p(\sigma)$, $\beta_0 \sim p(\beta)$. Hacer $j = 1$
2. Extraer una muestra del vector $u = (u_1, u_2, \dots, u_n)$ muestreando $u_i \sim p(u_i | \theta, \sigma, \beta, u_j, y)$ para $i = 1, \dots, n$
3. Hacer $j = j + 1$. Extraer $\theta_j \sim (\theta | \sigma, \beta, u, y)$, $\sigma_j \sim (\sigma | \theta, \beta, u, y)$, y $\beta_j \sim (\beta | \theta, \sigma, u, y)$
4. Si $j = m$, parar. Si no, volver al paso 2

Obtenemos con este procedimiento una muestra de tamaño m de $(\theta, \sigma, \beta | \mathbf{y})$.

Dos problemas surgen a la hora de extraer muestras de $\beta | \theta, \sigma, \mathbf{u}, \mathbf{y}$:

a) Qué valores de β son los que cumplen la condición

$$\left\{ \beta \text{ tal que } \left(\frac{u_i}{c(\beta)} \right)^{\frac{\beta+1}{2}} > \left(\frac{y_i - \theta}{\sigma} \right)^2, \forall i \right\}$$

b) Cómo extraer muestras de la densidad proporcional a $\frac{p(\beta)c(\beta)^{\frac{n(1+\beta)}{2}}}{\Gamma(\frac{1}{2}(3+\beta))^n}$, condicionado

a la región determinada en a).

Debido a que nuestro interés se centra sobre todo en las desviaciones respecto a la Normalidad, utilizaremos la reparametrización planteada al inicio, con $c(\beta) = 1/2$ constante.

Las densidades condicionales de la representación por mixturas quedan

$$\begin{aligned} \theta \mid \sigma, \beta, \mathbf{u}, \mathbf{y} &\equiv p(\theta), \theta \in [\max_i \{y_i - \sigma(2u_i)^{\frac{\beta+1}{2}}\}, \min_i \{y_i + \sigma(2u_i)^{\frac{\beta+1}{2}}\}] \\ \sigma \mid \theta, \beta, \mathbf{u}, \mathbf{y} &\equiv IG(a+n, b), \sigma > \max_i \left\{ \left(\frac{1}{2u_i} \right)^{\frac{\beta+1}{2}} |y_i - \theta| \right\} \\ \beta \mid \theta, \sigma, \mathbf{u}, \mathbf{y} &\propto \frac{p(\beta)}{[2^{\frac{(1+\beta)}{2}} \Gamma(\frac{1}{2}(3+\beta))]^n}, \beta \text{ tal que } (2u_i)^{\beta+1} > \left(\frac{y_i - \theta}{\sigma} \right)^2, \forall i \\ u_i \mid \theta, \sigma, \beta, u_j, \mathbf{y} &\propto Exp(1), 2u_i > \left(\frac{(y_i - \theta)^2}{\sigma^2} \right)^{\frac{1}{1+\beta}} \end{aligned}$$

Las condiciones a) y b) también se simplifican sustancialmente. Habría que resolver:

a) Qué valores de β son los que cumplen la condición $\left\{ \beta \text{ tal que } (2u_i)^{\beta+1} > \left(\frac{y_i - \theta}{\sigma} \right)^2, \forall i \right\}$

b) Cómo extraer muestras de la densidad proporcional a $\frac{p(\beta)}{[2^{\frac{(1+\beta)}{2}} \Gamma(\frac{1}{2}(3+\beta))]^n}$, condicionado a la región determinada en a).

Daremos respuesta a estas dos preguntas en los siguientes apartados.

Búsqueda de la región de pertenencia de β para cada u .

Se puede verificar fácilmente que la restricción depende de la región de pertenencia de u_i :

$$\left\{ \begin{array}{l} \text{Para } 2u_i \in (0, 1), \beta < \min_i \left\{ \frac{\log \left\{ \left(\frac{y_i - \theta}{\sigma} \right)^2 \right\}}{\log(2u_i)} \right\} - 1 \\ \text{Para } 2u_i \in (1, \infty), \beta > \max_i \left\{ \frac{\log \left\{ \left(\frac{y_i - \theta}{\sigma} \right)^2 \right\}}{\log(2u_i)} \right\} - 1 \end{array} \right.$$

Extracción de muestras de $(\beta \mid \theta, \sigma, u, \mathbf{y})$

Es necesario extraer muestras de la distribución proporcional a la función

$$f(\beta) = \frac{p(\beta)}{[2^{\frac{(3+\beta)}{2}} \Gamma(\frac{1}{2}(3+\beta))]^n}$$

truncada en la región indicada en a). Se verifica que

$$f(\beta) = \exp(-\frac{3n}{2} \log(2)) \exp(-(\frac{n}{2} \log(2))\beta) \frac{p(\beta)}{[\Gamma(\frac{1}{2}(3+\beta))]^n}$$

Además se puede comprobar que la función $[\Gamma(\frac{1}{2}(3+\beta))]^n$ tiene un máximo global en $\beta = 0$ para todo n , con lo que $f(\beta)$ se puede acotar superiormente por la expresión

$$\exp(-\frac{3n}{2} \log(2)) \exp(-(\frac{n}{2} \log(2))\beta) \frac{p(\beta)}{[\Gamma(\frac{3}{2})]^n}$$

con lo cual se puede aplicar el método de rechazo para extraer muestras de la densidad proporcional a $f(\beta)$, generando muestras de la distribución exponencial de parámetro $\frac{n}{2} \log(2)$, truncada en las cotas sobre beta calculadas anteriormente, o bien utilizando $p(\beta)$ como base. Usualmente, utilizaremos como distribución a priori $p(\beta) \equiv \text{Uniforme}(-1,1)$, en cuyo caso el algoritmo sería:

[Algoritmo para generar muestras de $(\beta \mid \theta, \sigma, u, \mathbf{y})$]

1. Generar $U \sim \text{Uniforme}(0,1)$ y $\beta \sim \text{Exponencial}(\frac{n}{2} \log(2))$ truncada en la región calculada en a).
2. $T = \frac{\exp(-\frac{n}{2} \log(2))}{[\Gamma(\frac{3}{2})]^n} [\Gamma(\frac{1}{2}(3+\beta))]^n$
3. Si $UT \leq 1$ parar. Si no, volver al paso 1-.

La aplicación del algoritmo de Gibbs también requiere generar muestras de varias distribuciones truncadas, para las que existen ya métodos (Devroye, 1984).

Generar muestras de la distribución a posteriori de σ plantea ciertos problemas, debido a que la región de truncamiento es a veces tan pequeña que es necesario un método

apropiado para reducir el tiempo de proceso. El método para generar una muestra X de una Gamma Inversa acotada en $[0, k]$ consiste en generar previamente una muestra Y de una Gamma acotada en $[0, \frac{1}{k}]$ y posteriormente hacer $X = \frac{1}{Y}$. Para ello se aplica un método para variables acotadas en el intervalo $[0, \frac{1}{k}]$ (Devroye, 1984) que en el caso concreto de una variable Gamma queda:

[Algoritmo para generar muestras de la distribución $\text{Gamma}(a, b)$ en $[0, \frac{1}{k}]$

1. Generar U, V Uniformes $(0,1)$.
2. Calcular $X = \frac{1}{k}V$.
3. Si $\frac{1}{k} < \frac{a-3}{b}$, $M = (\frac{1}{k})^{a-3}e^{-\frac{b}{k}}$.
4. Si $\frac{1}{k} > \frac{a-3}{b}$, $M = (\frac{a-3}{b})^{a-3}e^{-(a-3)}$
5. Si $UM \leq X^{a-3}e^{-bX}$ parar. Si no, volver al paso 1.

Ejemplo 1.2.-

Para ilustrar el método de extracción de $(\theta, \sigma, \beta \mid \mathbf{y})$ en primer lugar a través del muestreo directo de Gibbs, y en segundo lugar mediante la representación como mixturas de Gamma y Uniforme, utilizaremos los datos de Darwin ya tratados en la sección anterior. Emplearemos como distribuciones a priori $p(\sigma) = \frac{1}{\sigma}$, $p(\theta) \equiv N(15, 5)$ y $p(\beta) \equiv U(-1, 1)$ en el primer caso, cambiando $p(\sigma) \equiv GI(0.001, 0.01)$ lo que da una aproximación a la distribución no informativa $p(\sigma) = \frac{1}{\sigma}$, en el método basado en la representación por mixturas.

Se han realizado $m = 500000$ iteraciones del muestreo de Gibbs, eliminando las 2000 primeras y tomando los valores obtenidos de 5000 en 5000 para evitar dependencia en los resultados obtenidos.

Se observa en la Figura 1.12 cómo a partir de $m = 350000$ las varianzas ergódicas de la cadena de Gibbs se estabilizan para cada parámetro, con una mayor variabilidad en el caso de β debido también a haber utilizado a priori una distribución Uniforme .

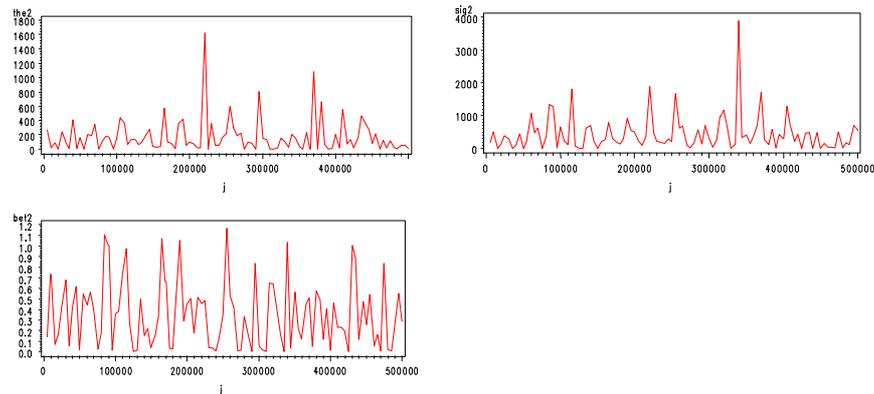


Figura 1.12. Varianzas ergódicas de los parámetros θ, β y σ .

Los histogramas obtenidos de las marginales a posteriori de θ, σ , y β son:

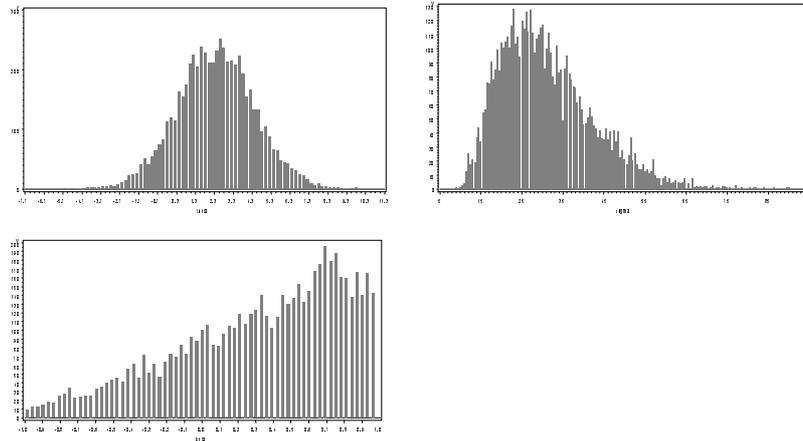


Figura 1.13. Histogramas de las marginales de los parámetros θ, σ y β .

A título indicativo, las medias muestrales a posteriori de θ y de σ son respectivamente 15.7 y 36.85, que coinciden aproximadamente con la media y desviación típica muestrales de la muestra \mathbf{y} .

La media muestral a posteriori de β es 0.18, aunque se ve en el gráfico como la moda a posteriori está situada sobre el valor 0.65, lo que nos indicaría la presencia de una curtosis mayor que la de la distribución Normal.

La utilización de la representación por mixturas también nos permite utilizar las muestras obtenidas de $\mathbf{u} = (u_1, u_2, \dots, u_n)$, para estudiar la posible presencia de outliers en nuestra muestra. Las gráficas de las distribuciones de los parámetros son similares a las obtenidas utilizando directamente el muestreo de Gibbs directo.

Ordenada la muestra original, se presenta en la Figura 1.14 el gráfico de cajas para cada u_i (cada u_i está relacionado con su y_i correspondiente) :

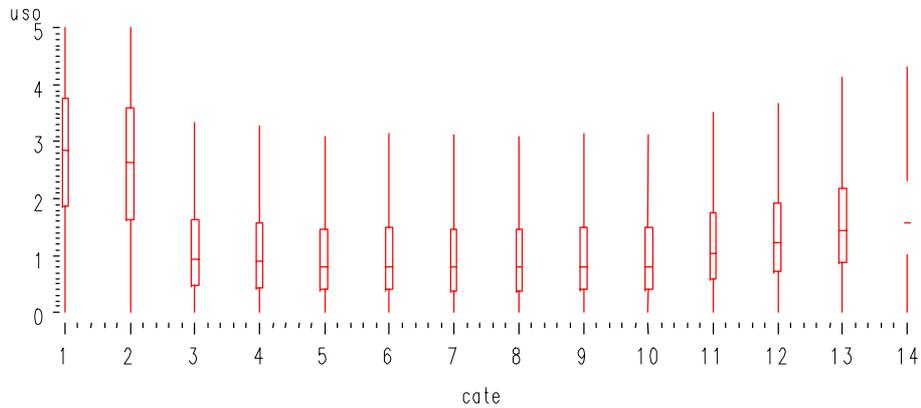


Figura 1.14. Diagrama de caja de los valores u_i .

Se ve claramente cómo las observaciones 1 y 2 (valores $y_1 = -67$ y $y_2 = -59$) son posibles outliers dentro de la familia Potencial Exponencial, conclusión a la que habían llegado también otros autores (Choy,1999) considerando modelos con β como un valor concreto prefijado.

3 Medidas de discrepancia para el contraste de normalidad

3.1 Introducción

Hemos visto que la familia Potencial Exponencial Exponencial contiene a la distribución Normal, y puede considerarse una extensión de ésta en términos de robustez para modelos en que nos interese adoptar distribuciones con una mayor o menor curtosis. En concreto, nuestro interés radica en utilizar esta propiedad para considerar la aplicación del modelo Normal a nuestros datos o rechazarla y utilizar en su lugar la familia Potencial Exponencial $PE(\theta, \sigma, \beta)$. Esta decisión será de importancia como veremos, al plantear modelos estadísticos clásicos como los modelos lineales utilizados en el último capítulo de este trabajo. El problema que vamos a abordar consiste por lo tanto en evaluar la desviación de nuestros datos de la normalidad en el marco de la distribución Potencial Exponencial.

Este problema se puede plantear como un contraste de hipótesis $H_0 : \beta = 0$ frente a $H_1 : \beta \neq 0$ puesto que para $\beta = 0$ la distribución Potencial Exponencial coincide con la Normal. El tema del contraste puntual bayesiano se ha tratado desde muchas perspectivas diferentes, como por ejemplo la utilización del Factor Bayes (Verdinelli, Wasserman, 1998), (Gómez-Villegas, Gómez, 1992); calibración de p-valores (Bayarri, Berger, J., 1998) o comparación de modelos (Carota et al, 1996), (Walker, Gutiérrez-Peña, 1999), (Gelfand y Ghosh, 1998) o (Goutis, Robert, 1998). (Bernardo, Smith, 1994) plantean el contraste que nos ocupa, en el que podemos suponer el modelo normal como una hipótesis de partida, en el marco definido como Rechazo de Modelos .

Bajo esta perspectiva, (Bernardo, Smith, 1994) establecen una medida de discrepancia $\delta(\beta)$ entre las verosimilitudes correspondientes a las dos hipótesis y calculan la esperanza a posteriori de esa medida $E_{\beta|\mathbf{y}}[\delta(\beta)]$ bajo uno de los dos modelos planteados. La medida se plantea como la diferencia de utilidades de ambos modelos, es decir

$$\delta(\beta) = U(\beta) - U(\beta_0)$$

Las conclusiones sobre el contraste se reflejarán en el cálculo de la diferencia de utilidades esperadas a posteriori:

$$E_{\beta|\mathbf{y}}[\delta(\beta)] = \int \delta(\beta)p(\beta | \mathbf{y})d\beta$$

como una medida a favor del modelo $p(\mathbf{y} | \beta)$, frente al modelo $p(\mathbf{y} | \beta_0)$.

El objetivo de este Capítulo es, dentro de nuestro problema concreto y de la perspectiva señalada de utilización de medidas de discrepancia, evaluar, el comportamiento

que pueden tener las medidas de discrepancia propuestas por varios autores, comparadas con una medida propuesta en la Sección 3.5 basada en el cálculo de regiones de máxima densidad a posteriori. En este marco, aplicaremos en la Sección 3.2 la divergencia de Kullback-Leibler como medida de discrepancia para este problema. En la Sección 3.3 se planteará la medida de discrepancia llamada "deviance" o logaritmo de la razón de verosimilitudes. En la Sección 3.4 se planteará una medida de discrepancia basada en la distribución predictiva. En la Sección 3.5 se presenta una medida basada en el cálculo de regiones HPD, y se procede a la comparación de las distintas medidas en un estudio de simulación. La Sección 3.6 presenta una aplicación de la utilización de la medida presentada al contraste de normalidad y al contraste general de bondad de ajuste, comparando los resultados obtenidos con la distribución Potencial Exponencial con los obtenidos por otros autores con otras familias de distribuciones.

3.2 Divergencia de Kullback-Leibler como medida de discrepancia

Si planteamos como utilidades de cada modelo

$$U(\beta, \theta, \sigma) = \int p(\mathbf{y} | \beta, \theta, \sigma) \log p(\mathbf{y} | \beta, \theta, \sigma) d\mathbf{y}$$

$$U(\beta_0, \theta, \sigma) = \int p(\mathbf{y} | \beta_0, \theta, \sigma) \log p(\mathbf{y} | \beta_0, \theta, \sigma) d\mathbf{y}$$

con

$$\delta(\beta, \theta, \sigma) = U(\beta, \theta, \sigma) - U(\beta_0, \theta, \sigma)$$

En este caso, la medida de discrepancia $\delta(\beta)$ será

$$\delta_{KL}(\beta) = \delta_{KL}(\beta, \theta, \sigma) = \int p(\mathbf{y} | \beta, \theta, \sigma) \log \frac{p(\mathbf{y} | \beta, \theta, \sigma)}{p(\mathbf{y} | 0, \theta, \sigma)} d\mathbf{y}$$

con

$$E_{\beta, \theta, \sigma | \mathbf{y}}[\delta_{KL}(\beta, \theta, \sigma)] = \int \delta_{KL}(\beta, \theta, \sigma) p(\beta, \theta, \sigma | \mathbf{y}) d\beta d\theta d\sigma$$

La primera integral, en el caso que nos ocupa, (distribución Potencial Exponencial) se puede desarrollar analíticamente y no depende de (θ, σ) , quedando, si aplicamos la reparametrización con $c(\beta) = \frac{1}{2}$ constante como aparece en el siguiente resultado.

Teorema 2.1

$$\delta_{KL}(\beta, \theta, \sigma) = n \frac{(1 + \beta)}{2\Gamma(\frac{1}{2}(3 + \beta))} \left[\Gamma(\frac{1+\beta}{2}) \log \frac{\Gamma(\frac{3}{2})}{2^{\frac{\beta}{2}}\Gamma(\frac{1}{2}(3 + \beta))} - \Gamma(\frac{1}{2}(3 + \beta)) + 2^\beta \Gamma(\frac{3}{2}(1 + \beta)) \right]$$

Demostración-

$$\begin{aligned} \delta_{KL}(\beta, \theta, \sigma) &= \int p(\mathbf{y} | \beta, \theta, \sigma) \log \frac{p(\mathbf{y} | \beta, \theta, \sigma)}{p(\mathbf{y} | 0, \theta, \sigma)} d\mathbf{y} = \\ &= n \int p(y | \beta, \theta, \sigma) \log \frac{p(y | \beta, \theta, \sigma)}{p(y | 0, \theta, \sigma)} dy = \end{aligned}$$

Haciendo el cambio de variable $z = \frac{y - \theta}{\sigma}$ tenemos

$$\begin{aligned} \delta_{KL}(\beta, \theta, \sigma) &= n \int_{-\infty}^{\infty} w(\beta) \exp \left[-\frac{1}{2} |z|^{2/(1+\beta)} \right] \log \frac{w(\beta) \exp \left[-\frac{1}{2} |z|^{2/(1+\beta)} \right]}{w(0) \exp \left[-\frac{1}{2} |z|^{2/(1+\beta)} \right]} dz = \\ &= 2nw(\beta) \log \frac{w(\beta)}{w(0)} \int_0^{\infty} \exp \left[-\frac{1}{2} |z|^{2/(1+\beta)} \right] - w(\beta) \int_0^{\infty} |z|^{2/(1+\beta)} \exp \left[-\frac{1}{2} |z|^{2/(1+\beta)} \right] dz + \\ &w(\beta) \int_0^{\infty} |z|^2 \exp \left[-\frac{1}{2} |z|^{2/(1+\beta)} \right] dz \end{aligned}$$

Las tres integrales se resuelven a partir de (1.2), tomando, respectivamente, $p = 1, a = 1/2, \alpha = \frac{2}{1+\beta}$ en la primera, $p = \frac{3+\beta}{1+\beta}, a = 1/2, \alpha = \frac{2}{1+\beta}$ en la segunda, y $p = 3, a = 1/2, \alpha = \frac{2}{1+\beta}$ en la tercera. Finalmente

$$\begin{aligned} \delta_{KL}(\beta, \theta, \sigma) &= 2nw(\beta) \left(\log \frac{w(\beta)}{w(0)} \right) \frac{(1 + \beta)}{2} 2^{\frac{1+\beta}{2}} \Gamma(\frac{1+\beta}{2}) - w(\beta) \frac{(1 + \beta)}{2} 2^{\frac{3+\beta}{2}} \Gamma(\frac{3+\beta}{2}) + \\ &+ w(\beta) \frac{(1 + \beta)}{2} 2^{\frac{3(1+\beta)}{2}} \Gamma(\frac{3(1+\beta)}{2}) = \\ &= n \frac{(1 + \beta)}{2\Gamma(\frac{1}{2}(3 + \beta))} \left[\Gamma(\frac{1+\beta}{2}) \log \frac{\Gamma(\frac{3}{2})}{2^{\frac{\beta}{2}}\Gamma(\frac{1}{2}(3 + \beta))} - \Gamma(\frac{1}{2}(3 + \beta)) + 2^\beta \Gamma(\frac{3}{2}(1 + \beta)) \right] \diamond \end{aligned}$$

Se puede ver la forma de esta medida asimétrica de distancia entre las distribuciones $p(\mathbf{y} | \beta, \theta, \sigma)$ y $p(\mathbf{y} | 0, \theta, \sigma)$ en la Figura 2.1.

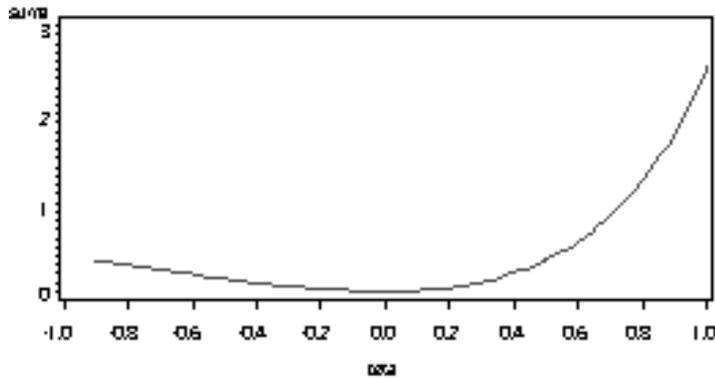


Figura 2.1. Medida de Kullback Leibler de la distribución Normal $N(\theta, \sigma)$, a la distribución Potencial Exponencial $PE(\beta, \theta, \sigma)$ para diferentes parámetros β .

Si queremos calcular la esperanza a posteriori de $\delta_{KL}(\beta, \theta, \sigma)$ el desarrollo analítico no es posible pero sí se puede realizar una estimación Monte- Carlo a partir de las muestras generadas de la distribución a posteriori $p(\beta, \theta, \sigma | \mathbf{y})$:

$$E_{\beta, \theta, \sigma | \mathbf{y}}[\delta_{KL}(\beta, \theta, \sigma)] = \int \delta_{KL}(\beta, \theta, \sigma) p(\beta, \theta, \sigma | \mathbf{y}) d\beta d\theta d\sigma \simeq \frac{1}{m} \sum_{j=1}^m \delta_{KL}(\beta_j, \theta_j, \sigma_j)$$

donde el vector $(\beta_j, \theta_j, \sigma_j)$ está generado de $p(\beta, \theta, \sigma | \mathbf{y})$ por los métodos de extracción de muestras presentados en el capítulo anterior.

3.3 Logaritmo de la razón de verosimilitudes como medida de discrepancia

(Gelfand y Ghosh, 1998) utilizan el logaritmo de la razón de verosimilitudes (la deviance) como medida de discrepancia o función de pérdida.

Esta medida es, en nuestro caso,

$$\delta_D(\beta, \theta, \sigma) = \log \frac{p(\mathbf{y} | \beta, \theta, \sigma)}{p(\mathbf{y} | 0, \theta, \sigma)}$$

Utilizar esta medida equivaldría a plantear como utilidades de ambos modelos

$$U(\beta, \theta, \sigma) = \log p(\mathbf{y} | \beta, \theta, \sigma)$$

y

$$U(0, \theta, \sigma) = \log p(\mathbf{y} \mid 0, \theta, \sigma)$$

con

$$\delta_D(\beta, \theta, \sigma) = U(\beta, \theta, \sigma) - U(0, \theta, \sigma)$$

y donde $E_{\beta|\mathbf{y}}[\delta_D(\beta, \theta, \sigma)]$ sería la diferencia de utilidades a posteriori.

La aproximación a $E_{\beta|\mathbf{y}}[\delta_D(\beta, \theta, \sigma)]$ será

$$\begin{aligned} E_{\beta|\mathbf{y}}[\delta_D(\beta, \theta, \sigma)] &\simeq \frac{1}{m} \sum_{j=1}^m \delta_D(\beta_j, \theta_j, \sigma_j) = \\ &= \frac{1}{m} \sum_{j=1}^m \log \frac{p(\mathbf{y} \mid \beta_j, \theta_j, \sigma_j)}{p(\mathbf{y} \mid 0, \theta_j, \sigma_j)} = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n \log \frac{p(y_i \mid \beta_j, \theta_j, \sigma_j)}{p(y_i \mid 0, \theta_j, \sigma_j)} \end{aligned}$$

donde el vector $(\beta_j, \theta_j, \sigma_j)$ está generado de $p(\beta, \theta, \sigma \mid \mathbf{y})$. Respecto a la medida discretizada anterior, sería necesario calibrar esta medida para saber cuándo considerarla suficientemente grande o pequeña pues sus cotas dependen de n .

3.4 Medida de discrepancia basada en la distribución predictiva

En este marco, (Walker y Gutiérrez-Peña (1999)) plantean como medida a favor del modelo $p(\mathbf{y} \mid \beta, \theta, \sigma)$ frente a $p(\mathbf{y} \mid 0, \theta, \sigma)$ la diferencia de utilidades esperadas estimadas, donde las utilidades son:

$$U(\beta, \theta, \sigma) = \frac{1}{n} \sum_{i=1}^n \log p(y_i; (\beta, \theta, \sigma))$$

y $p(y_i; (\beta, \theta, \sigma))$ es la densidad predictiva cuando se trata del modelo (β, θ, σ) :

$$p(y_i; (\beta, \theta, \sigma)) = p(y_i \mid \mathbf{y}) = \int p(y_i \mid \beta, \theta, \sigma) p(\beta, \theta, \sigma \mid \mathbf{y}) d\beta d\theta d\sigma$$

y análogamente,

$$U(0, \theta, \sigma) = \frac{1}{n} \sum_{i=1}^n \log p(y_i; (0, \theta, \sigma))$$

con

$$p(y_i; (0, \theta, \sigma)) = p(y_i \mid \mathbf{y}) = \int p(y_i \mid 0, \theta, \sigma) p(0, \theta, \sigma \mid \mathbf{y}) d\theta d\sigma$$

La diferencia de utilidades quedará

$$\delta_P(\beta, \theta, \sigma) = U(\beta, \theta, \sigma) - U(0, \theta, \sigma) = \frac{1}{n} \sum_{i=1}^n \log \frac{p(y_i; (\beta, \theta, \sigma))}{p(y_i; (0, \theta, \sigma))}$$

Si aproximamos la densidad predictiva por el estimador Monte-Carlo

$$p(y_i; (\beta, \theta, \sigma)) = \int p(y_i | \beta, \theta, \sigma) p(\beta, \theta, \sigma | \mathbf{y}) d\beta d\theta d\sigma \simeq \frac{1}{m} \sum_{j=1}^m p(y_i | \beta_j, \theta_j, \sigma_j)$$

donde el vector $(\beta_j, \theta_j, \sigma_j)$ está generado de $p(\beta, \theta, \sigma | \mathbf{y})$ y

$$p(y_i; (0, \theta, \sigma)) = \int p(y_i | \beta, \theta, \sigma) p(0, \theta, \sigma | \mathbf{y}) d\theta d\sigma \simeq \frac{1}{m} \sum_{j=1}^m p(y_i | 0, \theta_j^*, \sigma_j^*)$$

y el vector (θ_j^*, σ_j^*) está generado de $p(0, \theta, \sigma | \mathbf{y})$, se tiene entonces que

$$\delta_P(\beta, \theta, \sigma) = U(\beta, \theta, \sigma) - U(0, \theta, \sigma) \simeq \frac{1}{n} \sum_{i=1}^n \log \frac{\sum_{j=1}^m p(y_i | \beta_j, \theta_j, \sigma_j)}{\sum_{j=1}^m p(y_i | 0, \theta_j^*, \sigma_j^*)}$$

Nótese el gran parecido con el resultado obtenido con el enfoque que toma la deviance como medida de discrepancia.

La diferencia, aparte de que los sumatorios en m aparecen en la fracción, es que los (θ_j^*, σ_j^*) han sido generados de la distribución a posteriori $p(0, \theta, \sigma | \mathbf{y})$, mientras que en el caso de la deviance todos los $(\beta_j, \theta_j, \sigma_j)$ están generados de $p(\beta, \theta, \sigma | \mathbf{y})$.

Al igual que en el caso de la deviance, sería conveniente en este caso establecer una calibración de esta medida de evidencia a favor de (β, θ, σ) .

3.5 Medida de discrepancia basada en el cálculo de la región HPD.

Una medida de evidencia en contra de la Normalidad de los datos consiste en encontrar la región a posteriori

$$C' = \{\beta \mid p(\beta \mid \mathbf{y}) \geq p(0 \mid \mathbf{y})\}$$

y calcular a continuación la probabilidad a posteriori $P(C' \mid \mathbf{y})$. Cuanto menor sea $P(C' \mid \mathbf{y})$, la evidencia aportada por la muestra a favor de la hipótesis $\beta = 0$ será mayor.

Este método tiene relación directa con el cálculo de regiones de máxima densidad a posteriori (Berger, 1985).

La región de máxima densidad a posteriori C (región HPD) es

$$C = \{\beta \mid p(\beta \mid y) \geq k(\alpha)\}$$

donde $k(\alpha)$ es la mayor constante tal que $P(C \mid y) \geq 1 - \alpha$.

Teorema 2.2.- Para variables aleatorias continuas, se puede verificar que plantear un contraste via región HPD a nivel $1 - \alpha$, es decir,

1. Calculando la región C a posteriori.
2. Rechazando la hipótesis nula si $\beta = 0$ cae fuera de esa región C

es equivalente a:

1. Calcular $P(C' \mid \mathbf{y})$ con C' definida por $C' = \{\beta \mid p(\beta \mid \mathbf{y}) \geq p(0 \mid \mathbf{y})\}$
2. Rechazar H_0 si $P(C' \mid \mathbf{y}) > 1 - \alpha$.

Con lo cual existe una relación directa entre la medida de evidencia $P(C' \mid \mathbf{y})$ y el contraste por regiones de confianza .

Demostración.-

Demostraremos que se rechaza de manera equivalente en el primer método que en el segundo .

Método HPD \implies Método $P(C' | \mathbf{y})$

En el contraste vía región HPD, al ser $k(\alpha)$ tal que $P(C | y) \geq 1 - \alpha$, tenemos que $P_{\beta|y}\{\beta | p(\beta | y) \geq k(\alpha)\} \geq 1 - \alpha$.

Este contraste rechaza H_0 si $\beta = 0$ cae fuera de la región C , es decir, si $k(\alpha) > p(0 | y)$ y por lo tanto, si

$\{\beta | p(\beta | y) \geq p(0 | y)\} \supset \{\beta | p(\beta | y) \geq k(\alpha)\}$. Lo que equivale a

$$P_{\beta|y}\{\beta | p(\beta | y) \geq p(0 | y)\} > P_{\beta|y}\{\beta | p(\beta | y) \geq k(\alpha)\} \geq 1 - \alpha$$

Esto equivale a decir que se rechaza H_0 si $P_{\beta|y}\{\beta | p(\beta | y) \geq p(0 | y)\} > 1 - \alpha$

Es decir, se rechaza H_0 si $P(C' | \mathbf{y}) > 1 - \alpha$ siendo C' la región definida en el segundo método.

Método $P(C' | \mathbf{y}) \implies$ Método HPD

En el método $P(C' | \mathbf{y})$, se rechaza H_0 si $P_{\beta|y}\{\beta | p(\beta | y) \geq p(0 | y)\} > 1 - \alpha$. Como

$k(\alpha)$ es la mayor constante tal que $P_{\beta|y}\{\beta | p(\beta | y) \geq k(\alpha)\} \geq 1 - \alpha$, ha de ser por fuerza $p(0 | y) < k(\alpha)$, lo que equivale a decir que $\beta = 0$ cae fuera de la región C definida en el método HPD.

Esta equivalencia se puede verificar gráficamente en la Figura 2.2

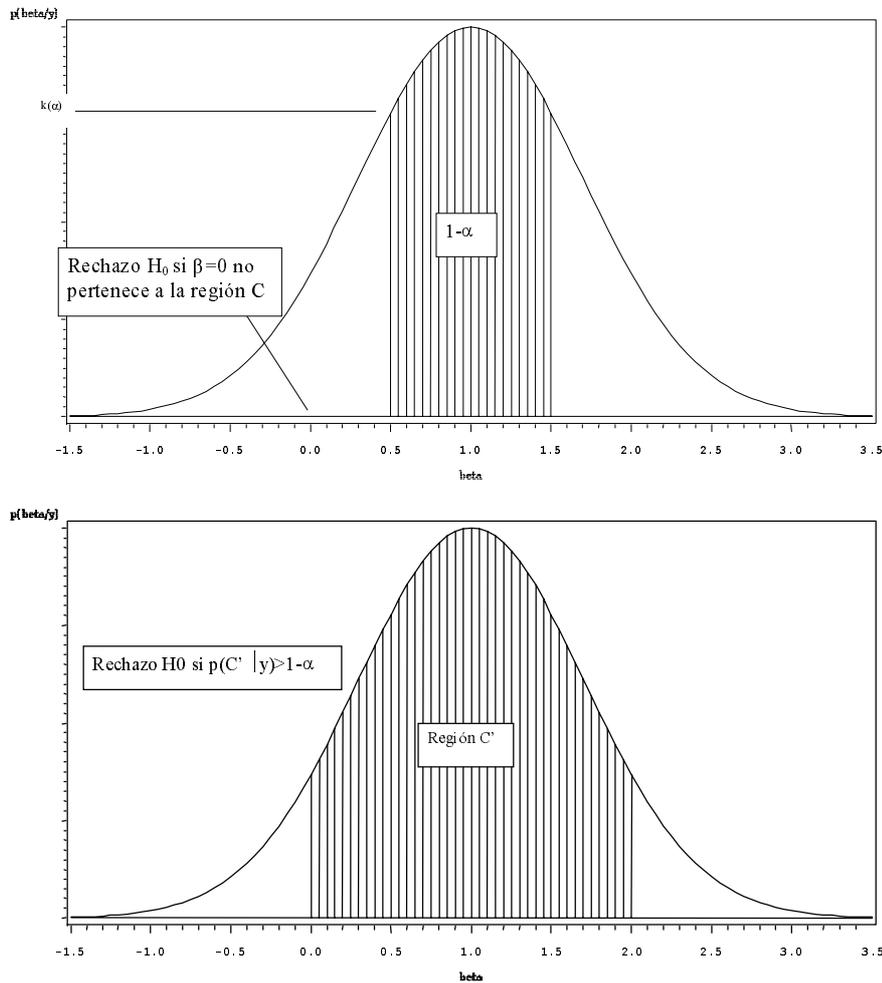


Figura 2.2. Relación entre el contraste via regiones de máxima densidad a posteriori y la medida de evidencia presentada.

Si queremos situar esta medida de evidencia dentro del marco relativo a las medidas de discrepancia, pondremos

$$P(C' | \mathbf{y}) = \int_{\beta/\beta \in C'} p(\beta, \theta, \sigma | \mathbf{y}) d\beta d\theta d\sigma =$$

$$\int \mathbf{I}\{\beta | p(\beta | \mathbf{y}) \geq p(0 | \mathbf{y})\} p(\beta, \theta, \sigma | \mathbf{y}) d\beta d\theta d\sigma$$

es decir, la medida de evidencia $P(C' | \mathbf{y})$ se puede poner como

$$P(C' | \mathbf{y}) = E_{\beta|\mathbf{y}}[\delta_{HPD}(\beta)]$$

donde

$$\begin{aligned} \delta_{HPD}(\beta) &= \mathbf{I}\{\beta \mid p(\beta \mid \mathbf{y}) \geq p(0 \mid \mathbf{y})\} = \\ &= \begin{cases} 1 & \text{si } \frac{p(\beta \mid \mathbf{y})}{p(0 \mid \mathbf{y})} \geq 1 \\ 0 & \text{si } \frac{p(\beta \mid \mathbf{y})}{p(0 \mid \mathbf{y})} < 1 \end{cases} \end{aligned}$$

Como hemos fijado $p(\beta) \propto cte$, quedaría

$$\delta_{HPD}(\beta) = \begin{cases} 1 & \text{si } \frac{\int p(\mathbf{y} \mid \beta, \theta, \sigma) p(\theta, \sigma) d\theta d\sigma}{\int p(\mathbf{y} \mid 0, \theta, \sigma) p(\theta, \sigma) d\theta d\sigma} \geq 1 \\ 0 & \text{si } \frac{\int p(\mathbf{y} \mid \beta, \theta, \sigma) p(\theta, \sigma) d\theta d\sigma}{\int p(\mathbf{y} \mid 0, \theta, \sigma) p(\theta, \sigma) d\theta d\sigma} < 1 \end{cases}$$

Una ventaja de este método, aparte de tener relación directa con los contrastes via regiones HPD, es que $E_{\beta|\mathbf{y}}[\delta_{HPD}(\beta, \theta, \sigma)] = E_{\beta|\mathbf{y}}[\delta_{HPD}(\beta)]$ está calibrada automáticamente: toma valores entre 0 y 1, pues es una probabilidad a posteriori. El planteamiento del contraste con otras medidas $\delta_{HPD}(\beta, \theta, \sigma)$ exige realizar análisis de las cotas de la esperanza a posteriori de estas medidas, cuestión difícil de tratar aunque algunos autores han establecido aproximaciones (Carota et al, 1996).

Hay que señalar que esta medida concreta $\delta_{HPD}(\beta, \theta, \sigma)$ depende de la muestra \mathbf{y} , mientras que la distancia de Kullback-Leibler, planteada como medida de discrepancia, no. Esto no parece una objeción en el sentido de que al final siempre se va a tomar la esperanza a posteriori de esta medida respecto de la muestra \mathbf{y} .

Para el cálculo de $E_{\beta|\mathbf{y}}[\delta_{HPD}(\beta, \theta, \sigma)]$ es más sencillo utilizar la aproximación Monte-Carlo a la distribución a posteriori que las muestras extraídas $(\beta_j, \theta_j, \sigma_j)$ mediante la aproximación por mixturas o Gibbs.

Para estimar $E_{\beta|\mathbf{y}}[\delta_{HPD}(\beta, \theta, \sigma)]$ a partir de las muestras $(\beta_j, \theta_j, \sigma_j)$, sería necesario previamente realizar una estimación kernel de la densidad $p(\beta \mid \mathbf{y})$ a partir de la marginal en β en esas muestras, pues necesitamos conocer el punto β^* tal que $p(\beta^* \mid \mathbf{y}) = p(0 \mid \mathbf{y})$ para poder a continuación estimar la probabilidad a posteriori.

Ejemplo 2.1

En el ejemplo siguiente simulamos 100 valores de una T-Student con 4 g. de l. Pretendemos contrastar la hipótesis de Normalidad en los datos a través del contraste

$H_0 : \beta = \beta_0$ frente a $H_1 : \beta \neq \beta_0$. La distribución a posteriori $p(\beta | y)$ obtenida, con moda aproximada en 0.7, tiene una región C' que abarca desde $\beta = 0$ hasta $\beta = 1$. Su densidad es $p(C' | y) = 0.9968$, con lo que claramente rechazaríamos la hipótesis nula de Normalidad de los datos. A modo de comparación, el p-valor obtenido para la hipótesis de normalidad de los datos con el contraste de Kolmogorov-Smirnov-Lilliefors es $p = 0.45$.

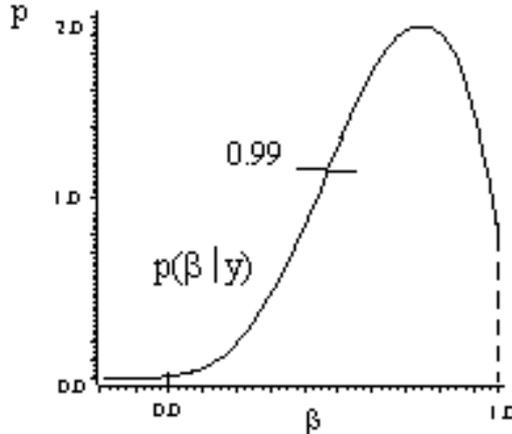


Figura 2.3. Contraste de normalidad sobre una muestra de una distribución $T - Student$ con 4 g. de l.

Ejemplo 2.2

Veremos un ejemplo de comparación de las diferentes medidas de discrepancia con datos no simulados, utilizando los datos de Darwin ya tratados.

Generando los puntos $(\beta_j, \theta_j, \sigma_j)$ para $j = 1, \dots, m$ con $m = 100000$ (y también (θ_j^*, σ_j^*) en el caso de la medida basada en la distribución predictiva) tenemos las siguientes diferencias de utilidades estimadas para cada de discrepancia utilizada:

- Kullback Leibler $U(\beta, \theta, \sigma) - U(0, \theta, \sigma) = 9.1668$
- Medida basada en HPD $U(\beta, \theta, \sigma) - U(0, \theta, \sigma) = 0.8173$
- Deviance $U(\beta, \theta, \sigma) - U(0, \theta, \sigma) = 2.97$
- Distribución predictiva $U(\beta, \theta, \sigma) - U(0, \theta, \sigma) = 0.08$

Aunque no son comparables entre sí debido a las diferencias de escala, se observa cómo en todos los casos estaríamos más cerca de adoptar modelos con $\beta \neq 0$ que de adoptar el modelo con $\beta = 0$. En las medidas Kullback Leibler, Deviance y Distribución predictiva se deduce de la magnitud de la diferencia de utilidades esperadas a favor del modelo (β, θ, σ) , aunque habría que calibrar esa magnitud en cada caso. En el caso de la Medida

basada en HPD , cuya diferencia de utilidades esperada se sitúa en una escala de 0 a 1, se ve claramente cómo se rechazaría el modelo $\beta = 0$.

Ejemplo 2.3.

Se han realizado simulaciones con el objetivo de comparar el comportamiento de las diferentes medidas de evidencia presentadas. Se simulan muestras de tamaño $n = 100$ provenientes de la distribución Potencial Exponencial con parámetro β . Se calculan los valores de las medidas de discrepancia presentadas que medirán la evidencia a posteriori en contra de que la muestra original provenga de una distribución Normal. Este proceso se repetirá 30 veces con 30 muestras de tamaño $n = 100$, simuladas de una Potencial Exponencial para cada uno de los parámetros $\beta = -0.75, -0.5, -0.25, 0(Normal), 0.25, 0.5, 0.75$.

Obtenemos por tanto, para cada parámetro β de la distribución original de los datos, una muestra de cada medida para contrastar normalidad. De estas muestras presentaremos los diagramas de caja en la Figura 2.4, ordenados según el valor del parámetro β de la muestra original simulada .

Para tener una referencia de comparación se han incluido también los p-valores de los contrastes de Kolmogorov-Smirnov-Lilliefors y el contraste χ^2 sobre las mismas muestras. A efectos de poder comparar, presentamos en estos casos la cantidad "1 - pvalor" pues nuestras medidas toman valores mayores según nos alejamos de la Normalidad.

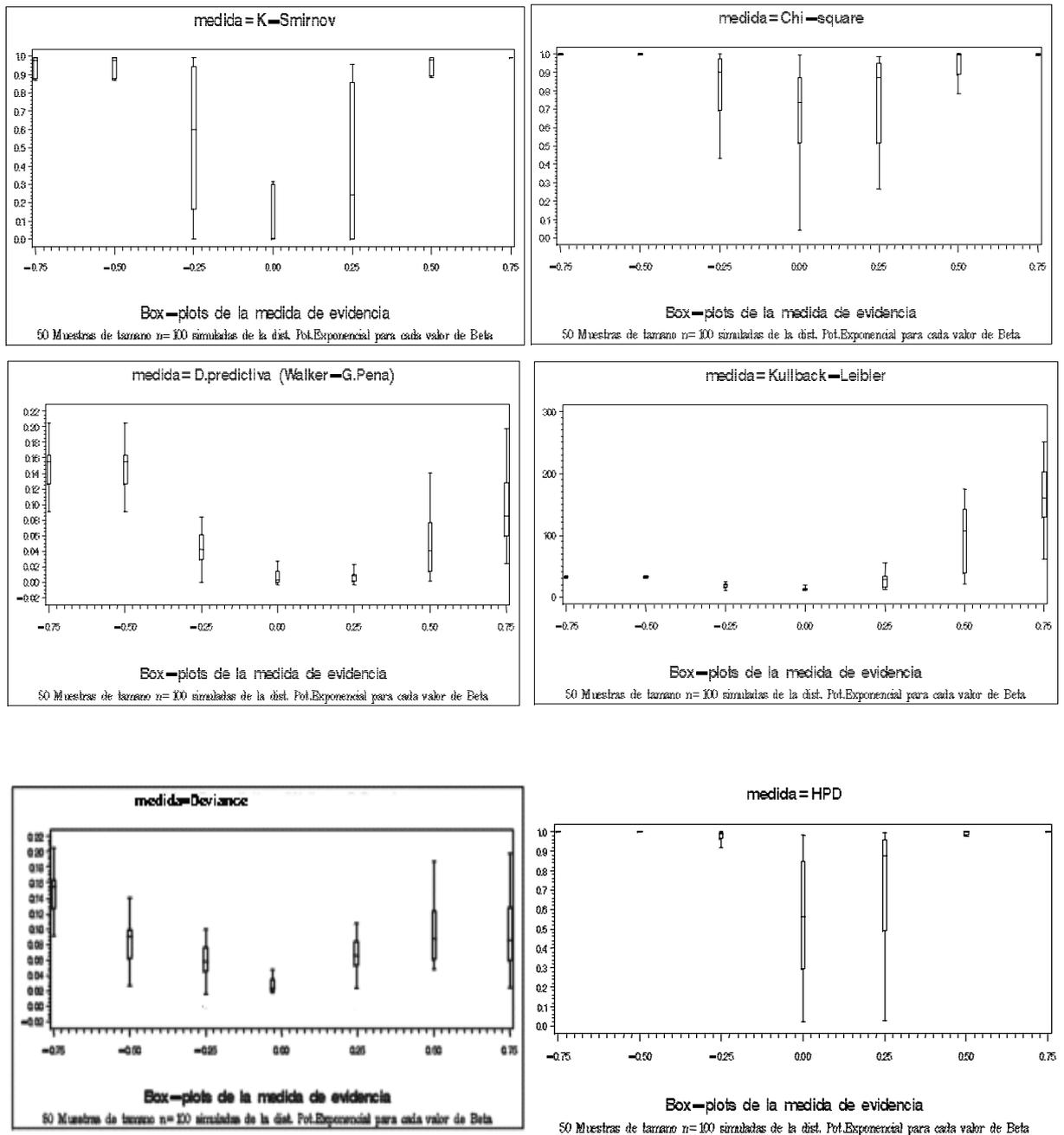


Figura 2.4. Diagramas de caja de las diferentes medidas de discrepancia obtenidas en 30 muestras de tamaño $n = 100$ de la distribución Potencial Exponencial con diferentes parámetros β .

Se observa cómo las medidas de evidencia basadas en los contrastes clásicos de Kolmogorov-Smirnov-Lilliefors y χ^2 , la deviance, la medida a posteriori *HPD* y la medida basada en la distribución predictiva funcionan correctamente, dando valores mayores a medida que las muestras simuladas se alejan de la Normal ($\beta = 0$) tanto hacia una curtosis más baja

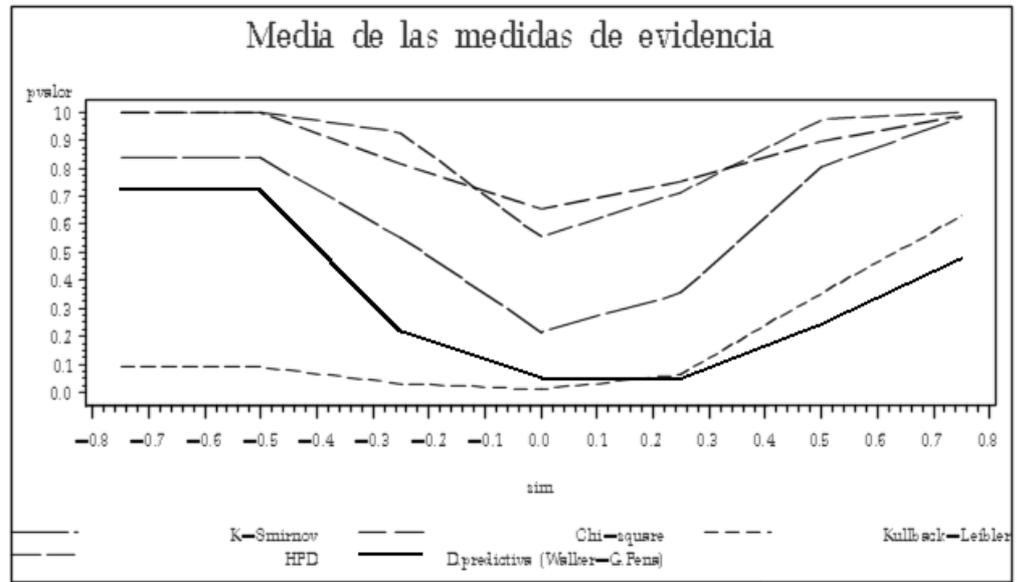


Figure 1:

de lo normal ($\beta \rightarrow -1$) como hacia una curtosis más alta de lo normal ($\beta \rightarrow 1$).

En el caso de la medida de Kullback-Leibler, se nota en el gráfico su asimetría: discriminaría mejor los casos en los que la muestra tenga una curtosis más alta de lo normal que aquellos casos en los que la curtosis sea más baja.

Al ser cada valor de la medida de discrepancia la esperanza a posteriori de la distancia de Kullback-Leibler, es lógico que la curva formada por los gráficos de caja se asemeje exactamente a la curva de la distancia de Kullback-Leibler presentada en la Figura 2.4.

En la Figura 2.5 se reescalan los valores de las medidas de Kullback-Leibler y la medida basada en la distribución predictiva al intervalo $(0, 1)$ para poder ser comparadas con los demás métodos, Kolmogorov-Smirnov, χ^2 y medida *HPD*. Se presentan líneas uniendo respectivamente la media y la mediana muestrales de la medida de evidencia obtenida para cada parámetro β .

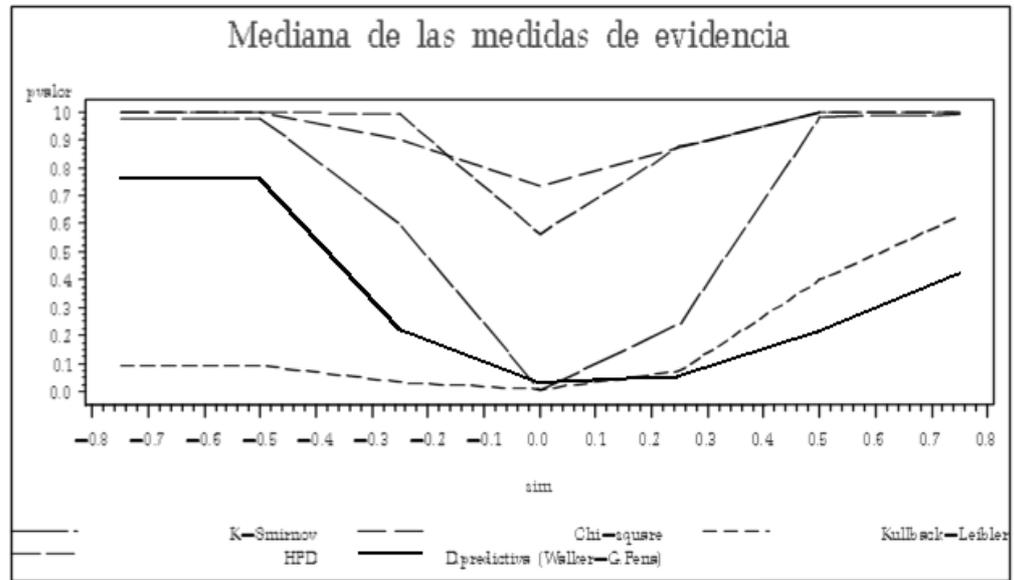


Figura 2.5. Medias y medianas de las medidas de discrepancia a posteriori para cada valor de β en las muestras simuladas.

3.6 Aplicación. Contrastes de bondad de ajuste

Es conocido que si X es una variable aleatoria con densidad $f(x)$ y función de distribución $F(x)$, la variable aleatoria $Y = F(X)$ se distribuye como una Uniforme $U(0, 1)$. Podemos contrastar la hipótesis de que la muestra x proviene de una distribución $F(x)$ contrastando la uniformidad de la muestra transformada $y = F(x)$, si la forma analítica de $F(x)$ es completamente conocida.

(Bernardo, 1982), (Girón, 1985) y (Gómez-Villegas, 1988) realizan contrastes de bondad de ajuste para cualquier distribución a partir de contrastes sobre los parámetros de, respectivamente, la familia Alfa, la familia Beta simétrica en $(0,1)$ y la familia Potencial exponencial.

En esta sección comparamos los métodos de las familias Alfa, Beta simétrica y Potencial-Exponencial para una distribución completamente conocida.

En primer lugar mostraremos una aplicación del método de la transformada por $F(x)$ utilizando la distribución Potencial Exponencial.

Ejemplo 2.4.

Hemos realizado dos contrastes recíprocos para mostrar algunas limitaciones del contraste a través de la familia Potencial Exponencial utilizando la transformada por $F(x)$.

Se sabe que la distribución Potencial Exponencial coincide con la distribución Uniforme cuando $\beta \rightarrow -1$, con lo cual realizaremos el contraste calculando

$$P(C' | \mathbf{y}) = P_{\beta|y}\{\beta | p(\beta | \mathbf{y}) \geq \lim_{\beta \rightarrow -1} p(\beta | \mathbf{y})\}.$$

En el primer caso se simulan $n = 100$ valores de una Normal $N(0,0.75)$ y se transforman mediante la función de distribución de una T de Student con 4 grados de libertad. Es decir, contrastaremos que esa muestra que en realidad proviene de la distribución Normal, proviene de una distribución T de Student con 4 grados de libertad. Como resultado de la transformación obtenemos una muestra unimodal y con baja curtosis, pero no Uniforme. La distribución a posteriori $p(\beta | y)$, tomando como priori una Uniforme, toma la forma de una distribución unimodal con moda en $\beta = -0.72$. La medida de evidencia $P(C' | y) = 0.97$ con lo cual rechazamos, adecuadamente, que los datos provienen de una t-Student con 4 grados de libertad.

El segundo ejemplo es el inverso: se simulan 100 valores de una distribución T de Student con 4 grados de libertad y se transforman mediante la función de distribución de una Normal(0,0.75). El resultado es una muestra con distribución en forma de U, que implica, como ya se ha visto, una distribución $p(\beta | y)$ con moda hacia $\beta = -1$ en el límite, y $p(C' | y) = 0$, lo que nos llevaría a aceptar, erróneamente, la normalidad de los datos.

Se puede deducir que el método, en el caso de la distribución Potencial Exponencial, depende en gran medida del resultado de la transformación tomada sobre la muestra. Si la imagen de los datos puede ser tomada como perteneciente a la familia Potencial Exponencial, las conclusiones sobre $p(\beta | y)$ son directas, mientras que sino es así, estas conclusiones son cuestionables.

(Bernardo, 1982) propone utilizar la familia Alfa(θ, ω), cuya forma analítica es

$$p(y | \theta, \omega) = \begin{cases} \theta \left(\frac{y}{\omega}\right)^{\theta-1} & \text{si } 0 \leq y < \omega \\ \theta \left(\frac{1-y}{1-\omega}\right)^{\theta-1} & \text{si } \omega \leq y \leq 1 \end{cases}$$

con $y \in [0, 1], \theta > 0, \omega \in [0, 1]$.

Si se dispone de una muestra y_1, \dots, y_n y se define

$$t(\omega) = \frac{1}{n} \left\{ \sum \log\left(\frac{\omega}{y_i}\right) I_{(0,\omega)}(y_i) + \sum \log\left(\frac{1-\omega}{1-y_i}\right) I_{(\omega,1)}(y_i) \right\}$$

Bernardo propone, utilizando la priori de referencia $p(\theta) = \frac{1}{\theta}$, la aproximación $p(\theta | y) \cong \Gamma(n, nt(\hat{\omega}))$, donde $\hat{\omega}$ es un estimador consistente de ω , calculado a partir de un estudio de los extremos relativos de $t(\omega)$.

Cuando $\theta=1$ la distribución adopta la forma de una Uniforme, lo que permitirá utilizar la técnica de contraste planteada.

Análogamente, (Girón, 1985) utiliza la familia Beta simétrica(α) en (0,1):

$$p(y | \alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2} \{y(1-y)\}^{\alpha-1}$$

con $y \in (0, 1), \alpha > 0$.

Si se dispone de una muestra y_1, \dots, y_n y se adopta una distribución a priori no informativa sobre α , la distribución a posteriori toma la forma

$$p(\alpha | \mathbf{y}) \propto \left(\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}\right)^n G^{\alpha-1}$$

$$\text{con } G = \prod_{i=1}^n y_i(1-y_i)$$

Cuando $\alpha=1$ la distribución adopta la forma de una Uniforme.

Se han calculado las distribuciones a posteriori para diferentes muestras, con los modelos Alfa, Beta simétrico y Potencial Exponencial con el objetivo de comparar los tres procedimientos .

Ejemplo 2.5.

En la Figura 2.6 contrastamos la hipótesis de normalidad $N(0,1)$ de una muestra de tamaño generada de una T de Student con 4 g.l. Aplicamos a los datos la transformación por la función de distribución $F(x)$ de la Normal para contrastar Uniformidad de la muestra transformada, con los tres métodos presentados anteriormente.

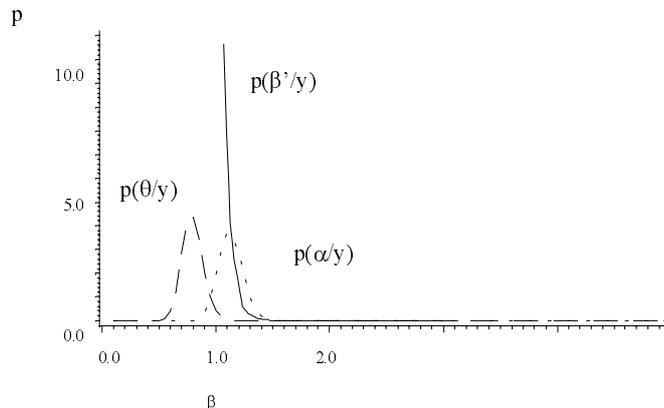


Figura 2.6. Distribuciones a posteriori de α, β, θ . Datos originales T Student con 4 g. l.

Observamos cómo las familias de distribución Alfa y Beta-simétricas parecen más sensibles a la no uniformidad que la familia Potencial Exponencial, que tiene $(-\infty, +\infty)$ por dominio para y . (Para poder comparar con las familias Alfa y Beta-simétrica hemos trasladado el plot de esta distribución con la reparametrización $\beta' = \beta + 2$). La distribución a posteriori $p(\beta | \mathbf{y})$ de la Potencial Exponencial tiene su moda cerca de $\beta = -1$, en el primer caso, dando $P(C' | \mathbf{y}) = 0$. Con lo que aceptaríamos erróneamente, la hipótesis de uniformidad. en este ejemplo tenemos $p(C' | \mathbf{y}) = 0.91$ para la familia Alfa y $p(C' | \mathbf{y}) = 0.89$ para la familia Beta simétrica, rechazando la uniformidad en ambos casos.

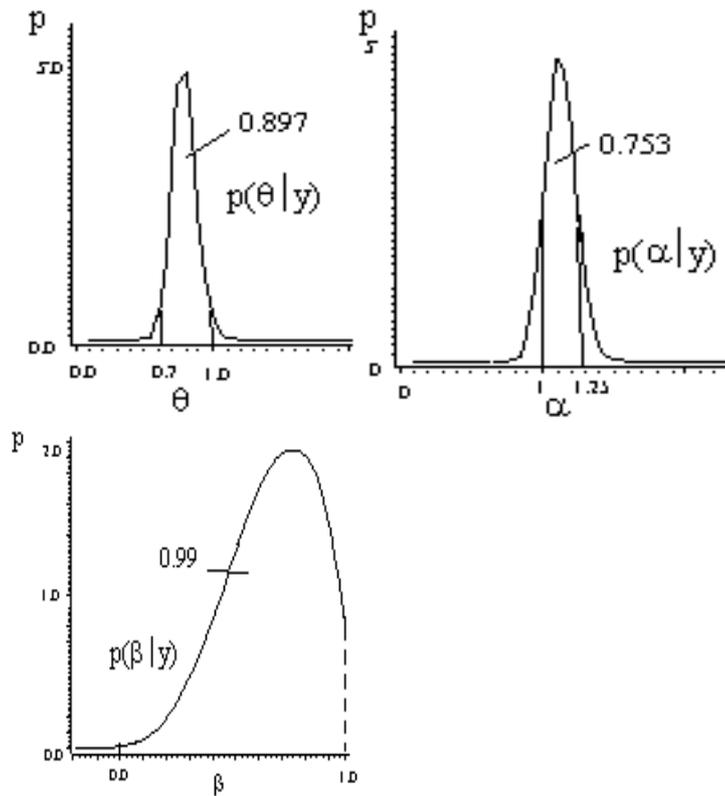


Figura 2.7. Distribuciones a posteriori de α, β, θ . Datos originales T Student con 5 g. l. transformados por $F(x)$

En la Figura 2.7 hemos generado datos de una T Student con 5 g. de l. Calculamos la medida de evidencia utilizando las familias Alfa y Beta simétrica para el contraste pero tomando los datos sin transformar para el caso concreto del contraste a través de la familia Potencial Exponencial.

Con el modelo Alfa tenemos $p(C' | \mathbf{y}) = 0.7534$, con lo que no rechazaríamos completamente normalidad original de los datos. Con la familia Beta simétrica $P(C' | \mathbf{y}) = 0.89$. Utilizando la familia Potencial Exponencial, sin usar la transformación $F(y)$, tenemos $P(C' | \mathbf{y}) = 0.9989$ con lo que rechazaríamos la hipótesis de Normalidad de los datos.

En este ejemplo particular tenemos

$$P_{Alfa}(C' | \mathbf{y}) < P_{Beta}(C' | \mathbf{y}) < P_{Pot.Expon.}(C' | \mathbf{y})$$

Podemos ver que la familia Potencial Exponencial otorga adecuadamente menos evidencia a favor de la distribución Normal que las otras dos familias.

Puede ser indicativo también el hecho de que el p-valor del contraste de Normalidad de Kolmogorov-Smirnov-Lilliefors sea $p = 0.46$.

Finalmente realizamos contrastes de Normalidad para muestras generadas de las distribuciones Normal, T de Student con 4 g. de l., Cauchy y Doble Exponencial. Generamos 10 muestras de tamaño $n = 100$ para cada distribución y calculamos la medida de evidencia para el contraste de Normalidad. Mostramos el resultado en Box-Plots de los valores de la medida de evidencia obtenida a través de cada procedimiento, y a partir de cada distribución original de los datos. Añadimos el resultado obtenido utilizando el p-valor en el contraste clásico de Kolmogorov-Smirnov.

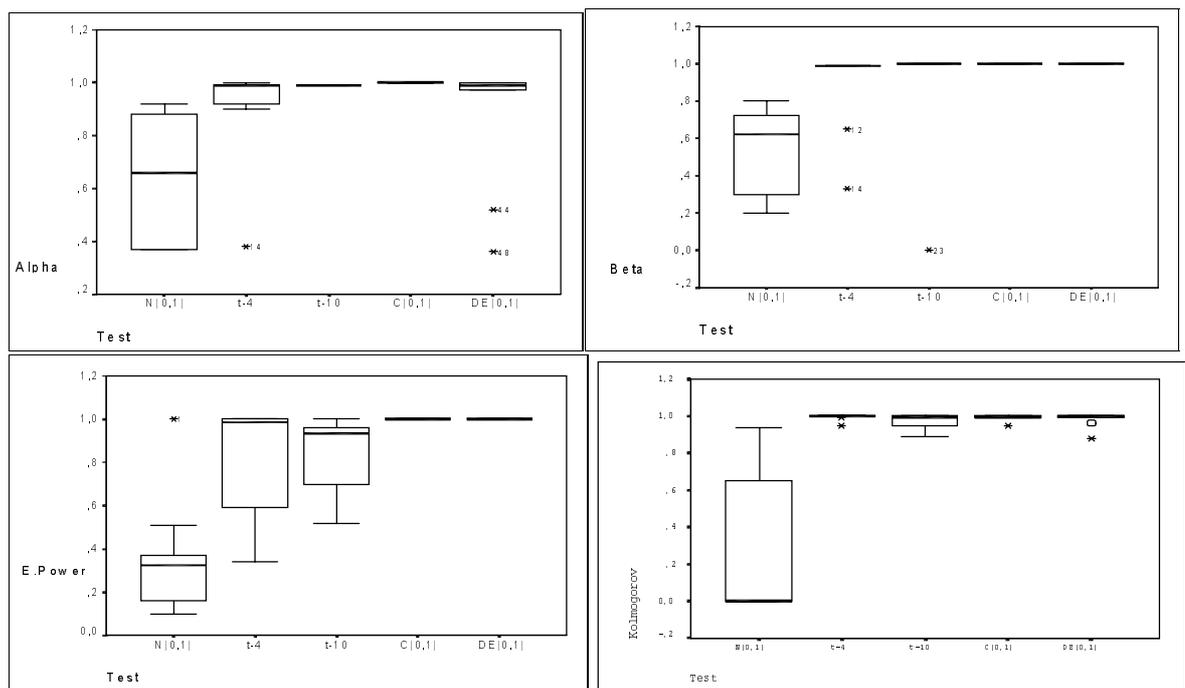


Figura 2.8. Box-plots de $P(C' | \mathbf{y})$ sobre datos simulados

El contraste Potencial-Exponencial obtiene resultados aceptables a la hora de aceptar Normalidad cuando es cierta, y también rechazando normalidad cuando los datos provienen de otras distribuciones, obteniendo en este caso ligeramente peores resultados que los contrastes Alfa y Beta simétricos. Los resultados a través de la distribución Beta

simétrica parecen ser los mejores rechazando normalidad cuando no es cierta, mientras los resultados a partir del contraste Potencial Exponencial detectan mejor cuando se da realmente Normalidad. El contraste clásico de Kolmogorov-Smirnov parece en ambos casos tener buenos resultados.

En las figuras siguientes presentaremos boxplots de las medidas obtenidas bajo diferentes tamaños muestrales y con varias distribuciones de origen. En primer lugar observamos en la Figura 2.9 para tamaños muestrales respectivos de los datos originales $n = 10, 50, 100$ cómo, cuando las muestras provienen de la distribución Potencial Exponencial, el contraste Potencial-Exponencial y los contrastes χ^2 y Kolmogorov-Smirnov discriminan correctamente a medida que nos alejamos de la Normalidad en nuestros datos y tienden a aceptar, también correctamente, a medida que el parámetro β de la distribución de origen de los datos simulados se acerca al valor $\beta = 0$. Los contrastes basados en las familias Beta Simétrica y Alpha no discriminan tan bien con este tipo de datos de origen. Se puede observar también cómo la precisión de los contrastes mejora al aumentar el tamaño de la muestra original.

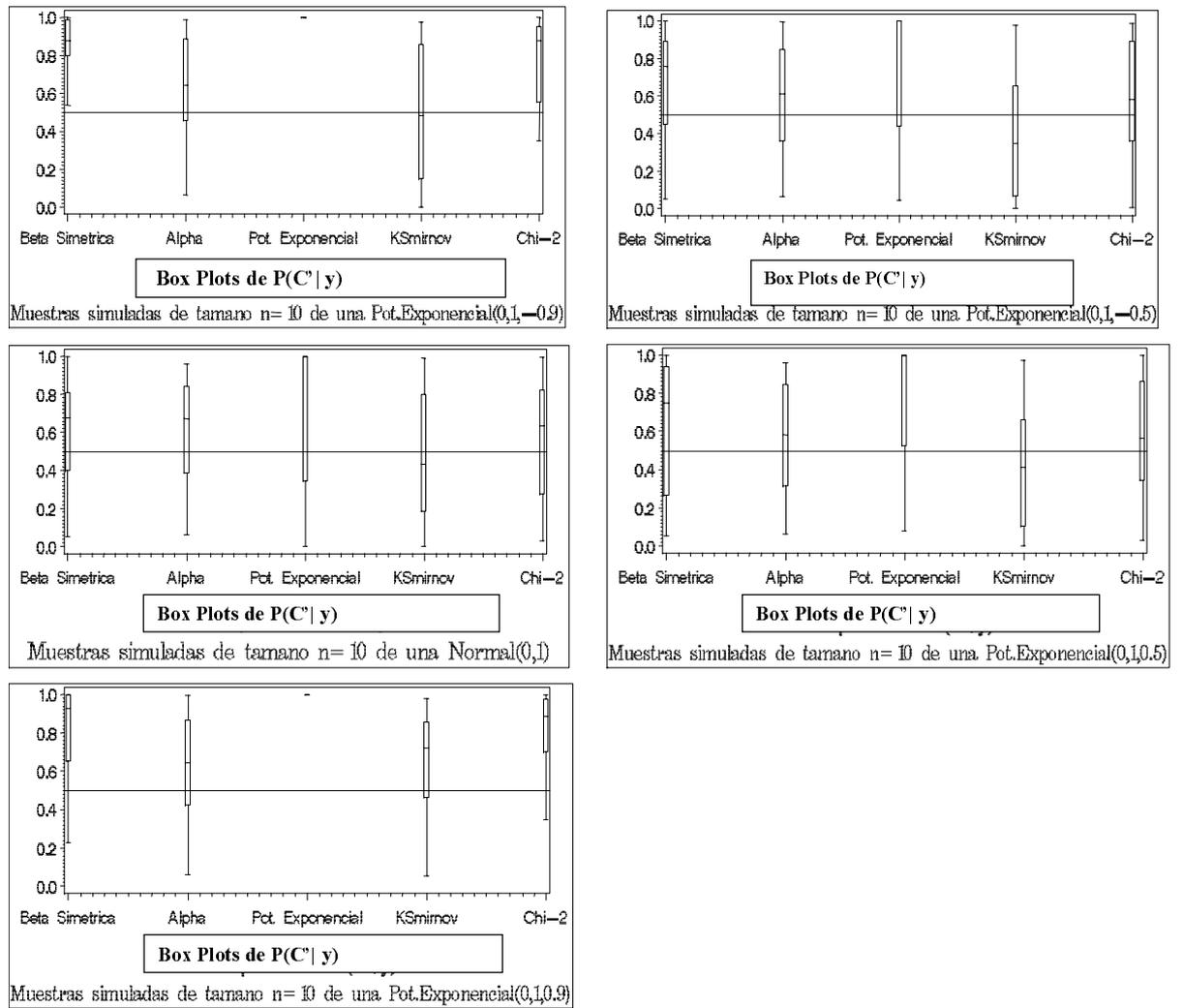


Figura 2.9

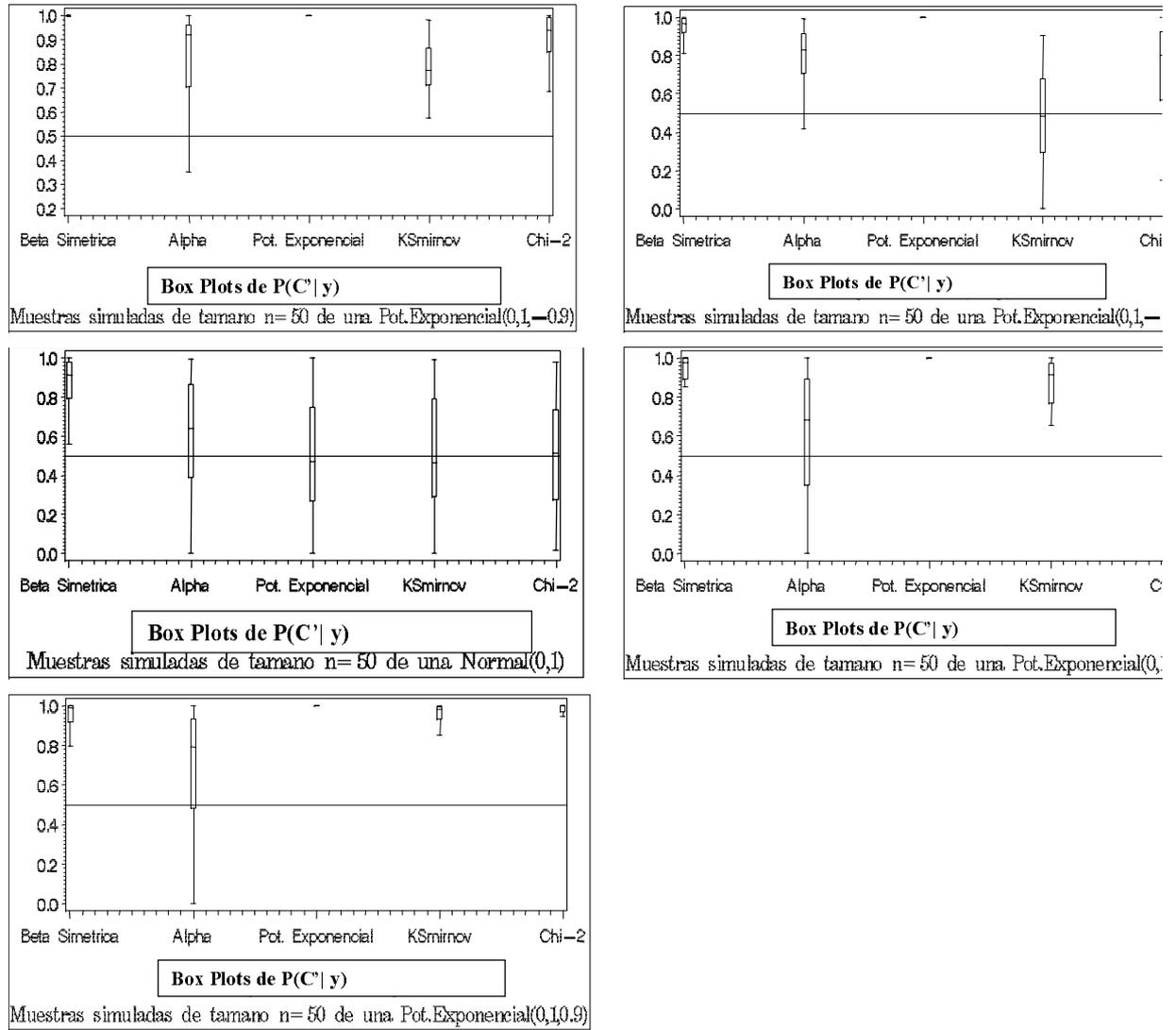


Figura 2.10.

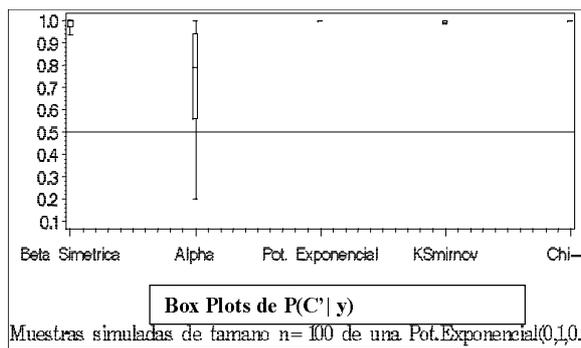
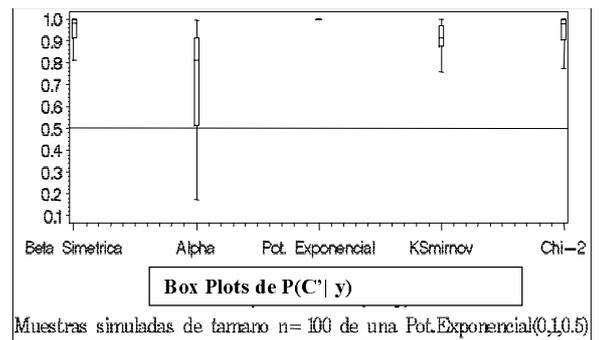
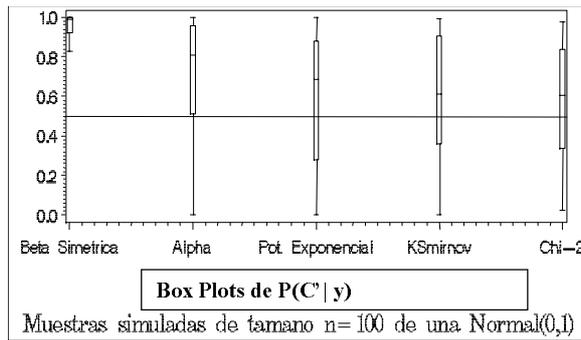
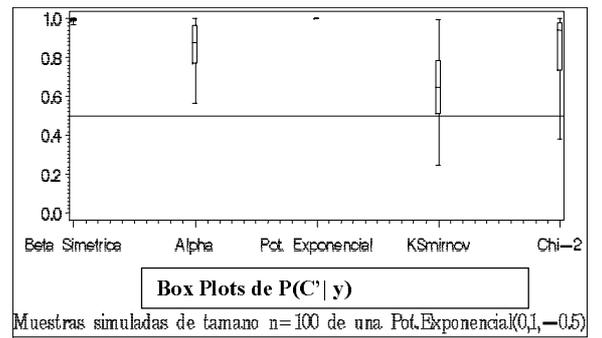
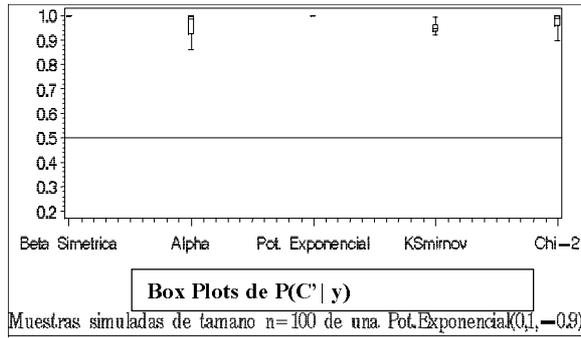


Figura 2.11

En las siguientes Figuras realizamos el contraste de Normalidad sobre muestras tomadas de la distribución t de Student con respectivamente 4,8 y 12 grados de libertad. Se observa cómo todos los contrastes tienden a aceptar normalidad a medida que aumentamos los grados de libertad de la distribución original.

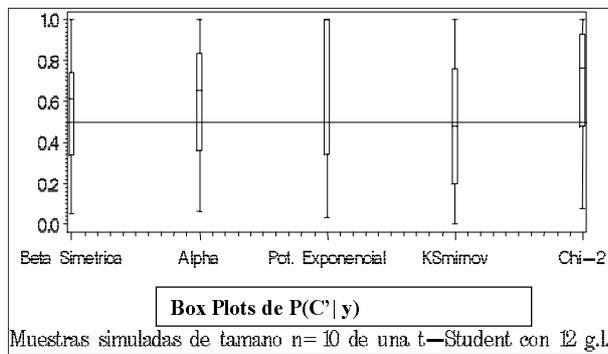
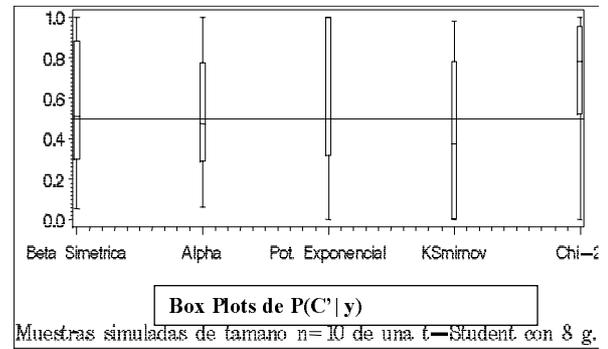
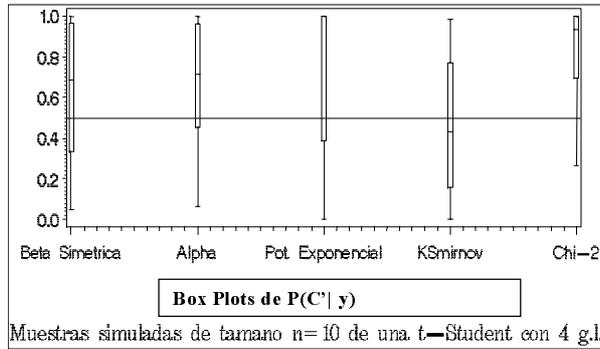


Figura 2.12

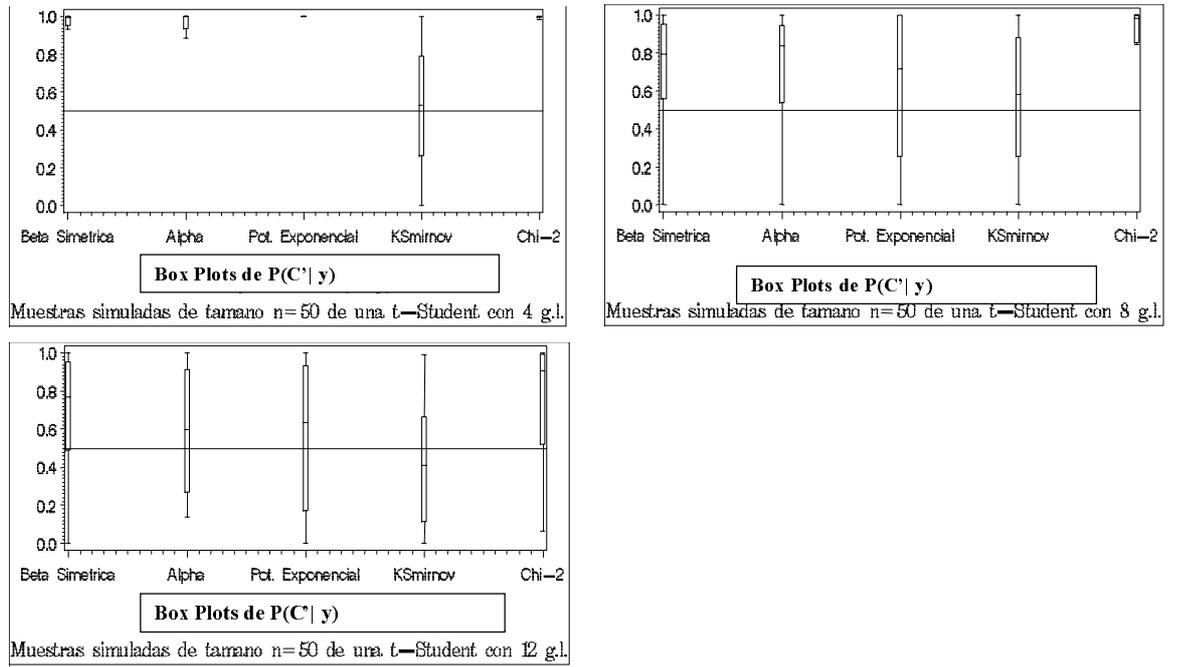


Figura 2.13

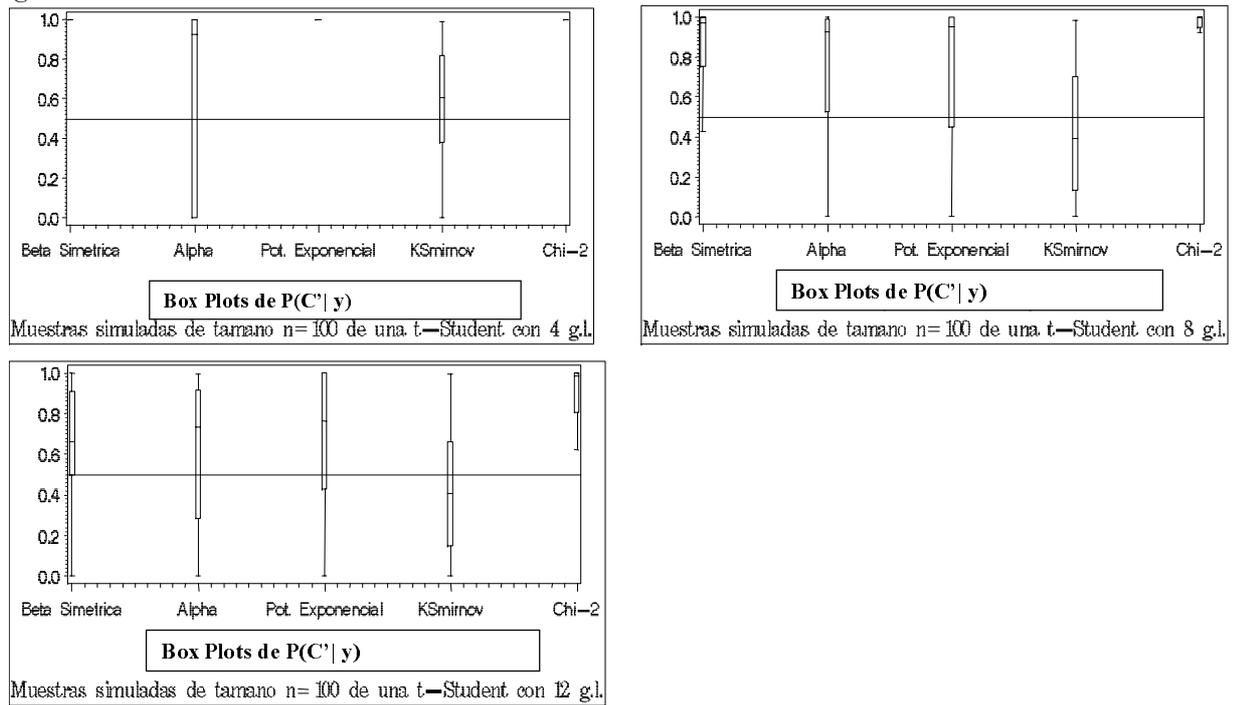


Figura 2.14

La Figura 2.15 presenta las medias de las medidas de discrepancia obtenidas a medida

que se varía el parámetro β de la distribución original de los datos simulados de una Potencial Exponencial, para $n = 100$. Se observa cómo los contrastes que mejor funcionan en este caso son el basado en la distribución Potencial-Exponencial y los contrastes clásicos de Kolmogorov-Smirnov y χ^2 . Se ve cómo la media de la medida de discrepancia desciende a medida que las muestras provienen de una distribución Normal.

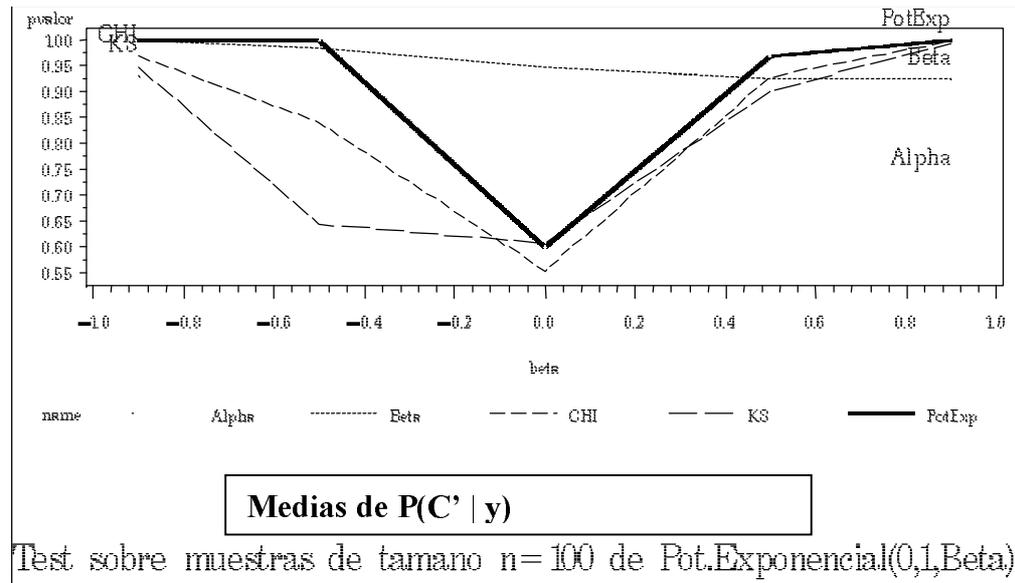


Figura 2.15.

Obtenemos un gráfico parecido para muestras originales tomadas de una T de Student, observando cómo los contrastes que mejor parecen discriminar son los de la familia Potencial-Exponencial, familia Beta simétrica y contraste χ^2 , en ese orden. Como es natural, la media de la medida de discrepancia desciende a medida que aumentan los grados de libertad.

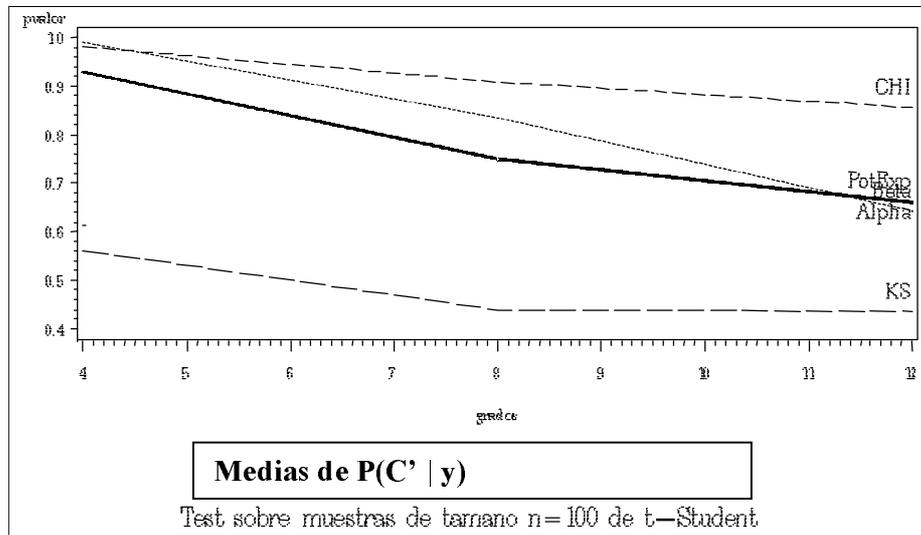


Figura 2.16

En la Figura 2.17 presentamos la media de la medida de discrepancia basada en la Potencial-Exponencial frente a los p-valores de los contrastes χ^2 y Kolmogorov-Smirnov-Lilliefors en el caso en que la distribución de los datos originales es una $\Gamma(\alpha, \beta)$, y vamos aumentando el parámetro β de $\beta = 5$ a $\beta = 20$. Se observa cómo las medidas de discrepancia tienden a aceptar correctamente Normalidad de los datos a medida que aumentamos el tamaño muestral.

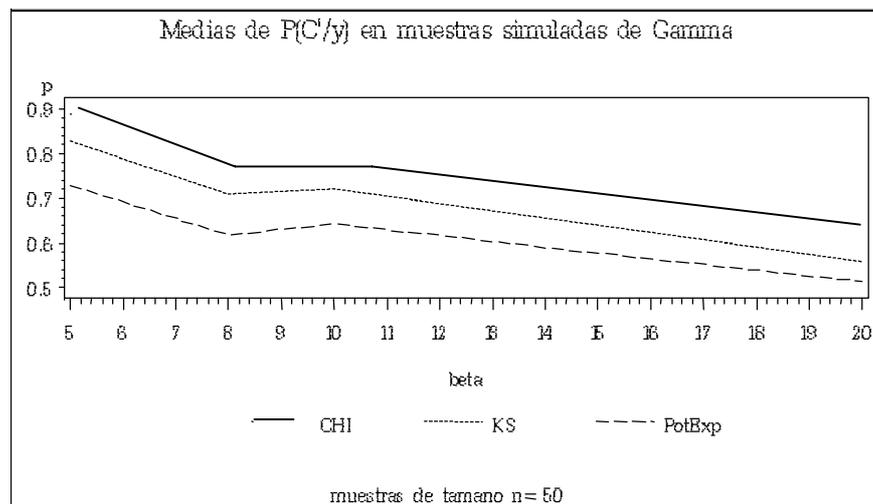


Figura 2.17

4 Aplicaciones en modelos bayesianos

4.1 Introducción

Algunos autores como (Box, Tiao, 1973) y (Marín, 1998) han presentado la posibilidad de aplicación de la distribución Potencial Exponencial en modelos lineales a través de aplicaciones concretas, poniendo de manifiesto la utilidad de un margen más flexible en cuanto a curtosis se refiere a la hora de modelizar los errores. Se profundizará ahora en esta idea, a través de la utilización de la distribución a posteriori del parámetro β , en aplicaciones que aparecen en (Gómez-Villegas, Portela, 2003).

En la Sección 4.2 se sentarán las bases para poder aplicar las técnicas de tratamiento bayesiano de la distribución Potencial Exponencial presentadas en el Capítulo 1, en el contexto concreto de los modelos lineales. Se verá una aplicación concreta en el campo de la química, en la que se comparan los modelos Normal y Potencial Exponencial.

En la Sección 4.3 se estudia cómo utilizar la distribución Potencial Exponencial en modelos no lineales, con especial énfasis en la comparación de modelos, tanto en la función lineal como en el hecho de utilizar la distribución Potencial Exponencial o la Normal para los errores. Para contrastar esta última posibilidad se utilizará la medida de discrepancia HPD enunciada en el Capítulo anterior y, debido a que el ajuste en términos de predicción puede ser una referencia alternativa de la bondad del modelo en los modelos lineales que vamos a tratar, utilizaremos una medida de este ajuste como medida complementaria. Se plantea como ejemplo una aplicación para la comparación de dos modelos no lineales en la evolución del peso de bulbos de cebolla.

Para poder evaluar la factibilidad de aplicación de la distribución Potencial Exponencial en todo tipo de modelos, se desarrolla una aplicación más compleja en la Sección 4.4, en la que se comparan dos modelos longitudinales sobre la evolución de células afectadas por el virus del SIDA. En este modelo, como en los estudiados en la Sección anterior, se pone de manifiesto cómo la utilización de la distribución Potencial Exponencial como modelo para los errores puede mejorar el modelo.

4.2 Aplicación al modelo lineal

Los estudios sobre la distribución a posteriori de los parámetros (θ, σ, β) pueden aplicarse al modelo lineal

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \varepsilon$$

donde \mathbf{y} es un vector de n observaciones, \mathbf{X} es una matriz de $n \times k$ elementos fijos, $\boldsymbol{\theta}$ es un vector de coeficientes de regresión desconocidos $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ y ε es un vector de n errores aleatorios.

Supondremos que los errores son independientes con distribución $PE(0, \sigma, \beta)$, es decir, $\forall i$,

$$p(\varepsilon_i | \sigma, \beta) = w(\beta)\sigma^{-1} \exp \left[-c(\beta) \left| \frac{\varepsilon_i}{\sigma} \right|^{2/(1+\beta)} \right]$$

Esto lleva a que

$$p(y_i - \mathbf{x}_i^t \boldsymbol{\theta} | \theta, \sigma, \beta) = w(\beta)\sigma^{-1} \exp \left[-c(\beta) \left| \frac{y_i - \mathbf{x}_i^t \boldsymbol{\theta}}{\sigma} \right|^{2/(1+\beta)} \right]$$

$\forall i, y$, por lo tanto, el modelo a posteriori para los parámetros será

$$p(\theta, \sigma, \beta | \mathbf{y}) \propto [w(\beta)]^n \sigma^{-n} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - \mathbf{x}_i^t \boldsymbol{\theta}}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta, \sigma, \beta)$$

Nuestros objetivos serán :

1. Estudiar la desviación de nuestro modelo respecto del modelo usual (errores distribuidos según una $N(0, \sigma^2)$), a través del estudio de la distribución a posteriori $p(\beta | \mathbf{y})$.
2. Realizar inferencias sobre los parámetros de regresión $\boldsymbol{\theta}$ considerando β como aleatorio en el modelo Potencial Exponencial para los errores, y comparar con los resultados obtenidos en el modelo Normal usual.

Para cumplir estos objetivos será necesario utilizar las técnicas expuestas en los anteriores apartados.

Distribuciones a priori

Continuando en la línea de plantear distribuciones a priori no informativas y suponiendo a priori las distribuciones de los parámetros independientes, supondremos las distribuciones genéricas $p(\beta)$ y $p(\boldsymbol{\theta})$.

$$p(\boldsymbol{\theta}, \sigma, \beta) = p(\boldsymbol{\theta})p(\sigma)p(\beta) \propto \sigma^{-1}$$

Donde $p(\boldsymbol{\theta}) = p((\theta_1, \dots, \theta_k))$ y supondremos estos parámetros independientes a priori, con lo que $p(\boldsymbol{\theta}) = \prod_{i=1}^k p(\theta_i)$.

Distribuciones a posteriori

Si estamos interesados en realizar inferencias sobre los parámetros $(\boldsymbol{\theta}, \sigma, \beta)$ buscaremos, bien obtener una estimación de las probabilidades marginales $p(\boldsymbol{\theta} | \mathbf{y})$, $p(\beta | \mathbf{y})$, $p(\sigma | \mathbf{y})$ o eventualmente $p(\theta_j | \mathbf{y})$, o bien obtener muestras generadas de estas distribuciones marginales. Comenzaremos con el primer caso, la estimación de las densidades a posteriori, para el cual utilizaremos la técnica de Monte-Carlo expuesta en apartados anteriores, modificada adecuadamente. El segundo objetivo, la obtención de muestras de esas densidades marginales, se abordará mediante el método de mixturas también modificado para ser aplicado al caso concreto del modelo lineal.

4.2.1 Estimación de las densidades marginales mediante el método de Monte-Carlo

Utilizando las distribuciones a priori que hemos planteado en el apartado anterior,

$$p(\boldsymbol{\theta}, \sigma, \beta | \mathbf{y}) \propto [w(\beta)]^n \sigma^{-(n+1)} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - \mathbf{x}_i^t \boldsymbol{\theta}}{\sigma} \right|^{2/(1+\beta)} \right]$$

Esta expresión se puede integrar en σ , resultando

$$p(\boldsymbol{\theta}, \beta | \mathbf{y}) \propto [\Gamma((\beta + 3)/2)]^{-n} \Gamma(n(\beta + 1)/2 + 1) \left[\sum_{i=1}^n |y_i - \mathbf{x}_i^t \boldsymbol{\theta}|^{2/(1+\beta)} \right]$$

Si deseamos obtener una estimación Monte-Carlo de las distribuciones marginales de los parámetros podemos aplicar el siguiente proceso:

Simulando muestras respectivamente de $p(\beta)$, $p((\theta_1, \dots, \theta_k))$, o $p(\theta_i)$, estimaremos el valor de la función proporcional a la densidad marginal de cada uno de los parámetros, respectivamente, mediante las expresiones

$$\begin{aligned}\hat{p}((\theta_1, \dots, \theta_k) | \mathbf{y}) &\propto \sum_{i=1}^m p((\theta_1, \dots, \theta_k), \beta_i | \mathbf{y}) \\ \hat{p}(\beta | \mathbf{y}) &\propto \sum_{i=1}^m p((\theta_1, \dots, \theta_k)_i, \beta | \mathbf{y}) \\ \hat{p}(\sigma | \mathbf{y}) &\propto \frac{1}{\sigma} \sum_{i=1}^{m^2} p((\theta_1, \dots, \theta_k)_i, \beta_i, \sigma | \mathbf{y}) \\ \hat{p}(\theta_l | \mathbf{y}) &\propto \sum_{i=1}^{m^2} p((\theta_1, \dots, \theta_k)_i^{-l}, \beta_i | \mathbf{y})\end{aligned}$$

donde el vector $(\theta_1, \dots, \theta_k)_i^{-l}$ representa el vector $(\theta_1, \dots, \theta_k)_i$ eliminada la componente θ_l .

4.2.2 Extracción de muestras de las densidades a posteriori

Método de Muestreo de Gibbs

Si deseamos obtener muestras de las densidades a posteriori podemos utilizar el método de muestreo de Gibbs, en el cual se extraen muestras de las distribuciones condicionales.

Para ello sabemos que, para $k \geq 1$, $\sum_{i=1}^n |\theta - y_i|^k \geq n |\theta - \bar{y}|^k$, con lo cual, como

$$p(\theta_l | (\theta_1, \dots, \theta_k)^{-l}, \sigma, \beta, \mathbf{y}) \propto \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - \mathbf{x}_i^t \boldsymbol{\theta}}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta_l)$$

Tenemos que

$$p(\theta_l | (\theta_1, \dots, \theta_k)^{-l}, \sigma, \beta, \mathbf{y}) \leq \exp \left[-nc(\beta) \left| \frac{\overline{(y - \mathbf{x}_{-l}^t(\theta_1, \dots, \theta_k)^{-l})} - \theta_l}{\sigma} \right|^{2/(1+\beta)} \right] p(\theta_l)$$

donde $\overline{(y - \mathbf{x}_{-l}^t(\theta_1, \dots, \theta_k)^{-l})}$ representa la media muestral de $y - \mathbf{x}_{-l}^t(\theta_1, \dots, \theta_k)^{-l}$ en las n observaciones, y \mathbf{x}_{-l}^t es el vector \mathbf{x} sin la componente l . Esta expresión de acotación se utiliza como base para implementar el método de rechazo obtenido en la sección 2.2.4.

Como $p(\sigma | \boldsymbol{\theta}, \beta, \mathbf{y}) \propto \sigma^{-(n+1)} \exp \left[-(c(\beta) \sum_{i=1}^n |y_i - \mathbf{x}_i^t \boldsymbol{\theta}|^{2/(1+\beta)}) \frac{1}{\sigma^{2/(1+\beta)}} \right]$ basta aplicar el método visto en la sección 2.2.4 sin más que cambiar $|y_i - \theta|$ por $|y_i - \mathbf{x}_i^t \boldsymbol{\theta}|$.

Del mismo modo, como

$$p(\beta | \boldsymbol{\theta}, \sigma, \mathbf{y}) \propto w(\beta) \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - \mathbf{x}_i^t \boldsymbol{\theta}}{\sigma} \right|^{2/(1+\beta)} \right] p(\beta)$$

Se utilizará un método similar al visto en 2.2.4 para extraer muestras de la distribución $p(\beta | \boldsymbol{\theta}, \sigma, \mathbf{y})$.

El método de Gibbs se desarrollará como sigue:

1. Tomar valores iniciales $\boldsymbol{\theta}^{(0)}, \sigma^{(0)}, \beta^{(0)}$
2. Extraer muestras sucesivamente de los θ_i a través de la distribución condicional $p(\theta_i | (\theta_1, \dots, \theta_k)^{-i}, \sigma^{(0)}, \beta^{(0)}, \mathbf{y})$ según el método propuesto. Con este paso se obtiene el vector $\boldsymbol{\theta}_1$ actualizado en cada iteración del algoritmo
3. Extraer $\sigma^{(1)}$ de $p(\sigma | \boldsymbol{\theta}^{(1)}, \beta^{(0)}, \mathbf{y})$
4. Extraer $\beta^{(1)}$ de $p(\beta | \boldsymbol{\theta}^{(1)}, \sigma^{(1)}, \mathbf{y})$
5. Volver al paso 2

Utilización de la representación por mixturas

Otra posibilidad es utilizar la representación de la Potencial Exponencial a través de mixturas. De acuerdo con esta representación, tenemos que

$$y | \boldsymbol{\theta}, \sigma, \beta, u, \mathbf{x} \equiv U\left(\mathbf{x}^t \boldsymbol{\theta} - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u^{\frac{\beta+1}{2}}, \mathbf{x}^t \boldsymbol{\theta} + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u^{\frac{\beta+1}{2}}\right)$$

con

$$u | \beta \equiv \Gamma\left(\frac{1}{2}(3 + \beta), 1\right)$$

Planteando como distribuciones a priori para cada parámetro θ_i del vector $(\theta_1, \dots, \theta_k)$ distribuciones uniformes, $p(\theta_i) \propto cte$, así como para el parámetro β , $p(\beta) \propto cte$ y para σ la distribución $IG(a, b)$, se pueden generar muestras de la distribución a posteriori.

Las densidades condicionales, necesarias para aplicar el muestreo de Gibbs, son:

Para $l \geq 2$,

$$\theta_i | (\theta_1, \dots, \theta_k)^{-l}, \sigma, \beta, \mathbf{u}, \mathbf{y} \equiv cte, \theta_i \in [\max(Max1, Max2), \min(Min1, Min2)]$$

donde

$$Max1 = \max_{x_i^l > 0} \left(y_i - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} - \mathbf{x}_i^t (\theta_1, \dots, \theta_k)^{-l} \right) \frac{1}{x_i^l}$$

$$Max2 = \max_{x_i^l < 0} \left(y_i + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} - \mathbf{x}_i^t (\theta_1, \dots, \theta_k)^{-l} \right) \frac{1}{x_i^l}$$

$$Min1 = \min_{x_i^l > 0} \left(y_i + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} - \mathbf{x}_i^t (\theta_1, \dots, \theta_k)^{-l} \right) \frac{1}{x_i^l}$$

$$Min2 = \min_{x_i^l < 0} \left(y_i - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} - \mathbf{x}_i^t (\theta_1, \dots, \theta_k)^{-l} \right) \frac{1}{x_i^l}$$

(denotando por x_i^l la componente l –ésima del vector \mathbf{x} en la muestra i)

Para $l = 1$,

$$\theta_l \mid (\theta_1, \dots, \theta_k)^{-l}, \sigma, \beta, \mathbf{u}, \mathbf{y} \quad \equiv cte,$$

$$\text{con } \theta_l \in \left[\max_i \left(y_i - \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} - \mathbf{x}_i^t (\theta_1, \dots, \theta_k)^{-l} \right), \min_i \left(y_i + \frac{\sigma}{c(\beta)^{\frac{\beta+1}{2}}} u_i^{\frac{\beta+1}{2}} - \mathbf{x}_i^t (\theta_1, \dots, \theta_k)^{-l} \right) \right]$$

El resto de distribuciones condicionales quedará

$$\sigma \mid \theta, \beta, \mathbf{u}, \mathbf{y} \quad \equiv IG(a + n, b), \sigma > \max \left\{ \left(\frac{c(\beta)}{u_i} \right)^{\frac{\beta+1}{2}} |y_i - \mathbf{x}^t \boldsymbol{\theta}| \right\}$$

$$\beta \mid \theta, \sigma, \mathbf{u}, \mathbf{y} \quad \propto \frac{c(\beta)^{\frac{n(1+\beta)}{2}}}{\Gamma\left(\frac{1}{2}(3 + \beta)\right)^n}, \beta \text{ tal que } \frac{u_i^{\frac{\beta+1}{2}}}{c(\beta)^{\frac{\beta+1}{2}}} > \left(\frac{y_i - \mathbf{x}^t \boldsymbol{\theta}}{\sigma} \right), \forall i$$

$$u_i \mid \theta, \sigma, \beta, u_j \quad \propto Exp(1), \frac{u_i}{c(\beta)} > \left(\frac{(y_i - \mathbf{x}^t \boldsymbol{\theta})^2}{\sigma^2} \right)^{\frac{1}{1+\beta}}$$

Con estas distribuciones condicionales se desarrolla el muestreo de Gibbs obteniendo muestras de la distribución conjunta a posteriori de los parámetros, del modo planteado en la sección 2.2.4.

4.2.3 Aplicación

(Box, Tiao,1973) estudian datos de un experimento que relaciona la tasa de una reacción química K con la temperatura absoluta T a la que se hizo el experimento. En su modelo

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i$$

donde $y_i = \log K$ y x_i es una función de la temperatura T_i prefijada en varias pruebas repetidas. En total el tamaño muestral es de $n = 20$.

Supondremos que $\varepsilon_i \equiv PE(0, \sigma, \beta)$ independientes y aplicaremos la técnica propuesta para establecer inferencias sobre σ, β , y el vector de coeficientes que en este caso es bidimensional, $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

El estudio de estos datos fue llevado a cabo por (Box, Tiao,1973) sin considerar la aleatoriedad del parámetro β , realizándose las inferencias sobre θ_1, θ_2 y σ para ciertos valores concretos de este parámetro de curtosis β . En nuestro caso, el reflejar la incertidumbre de β tendrá repercusiones en las distribuciones a posteriori de θ_1, θ_2 y σ .

Hemos implementado el método Monte-Carlo para estimar las densidades marginales respectivas $p(\beta | \mathbf{y})$, $p(\sigma | \mathbf{y})$, $p(\theta_1, \theta_2 | \mathbf{y})$, $p(\theta_1 | \mathbf{y})$ y $p(\theta_2 | \mathbf{y})$, adoptado las distribuciones a priori no informativas ya mencionadas, en dominios suficientemente amplios.

Se observa en la Figura 3.1 cómo en este caso la distribución a posteriori de β refleja prácticamente Normalidad de los datos : la moda a posteriori se sitúa en $\beta = 0.08$ y la medida de discrepancia HPD toma el valor 0.18, con lo que podríamos eventualmente utilizar el modelo Normal usual para los errores ε_i .

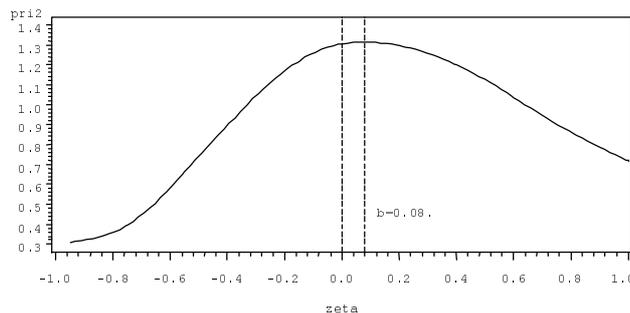


Figura 3.1. Distribución a posteriori $p(\beta | \mathbf{y})$

En la Figura 3.2 se observa el histograma de una muestra de $p(\beta | \mathbf{y})$ obtenida a través del método de mixturas:

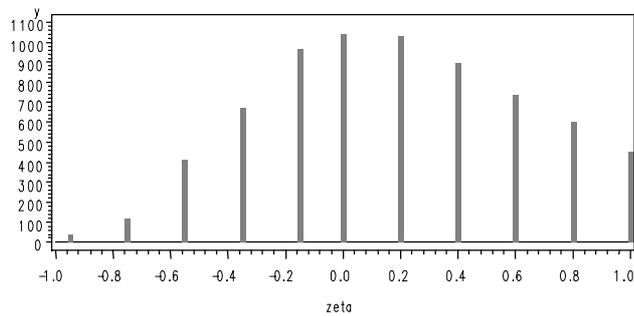


Figura 3.2. Histograma de muestras de $p(\beta | y)$

La distribución a posteriori conjunta de los parámetros de regresión (θ_1, θ_2) se refleja en las Figuras 3.3 y 3.4 donde, en concreto en el gráfico de contorno, el centro corresponde aproximadamente al punto $(\theta_1 = -4.00, \theta_2 = -0.2145)$.

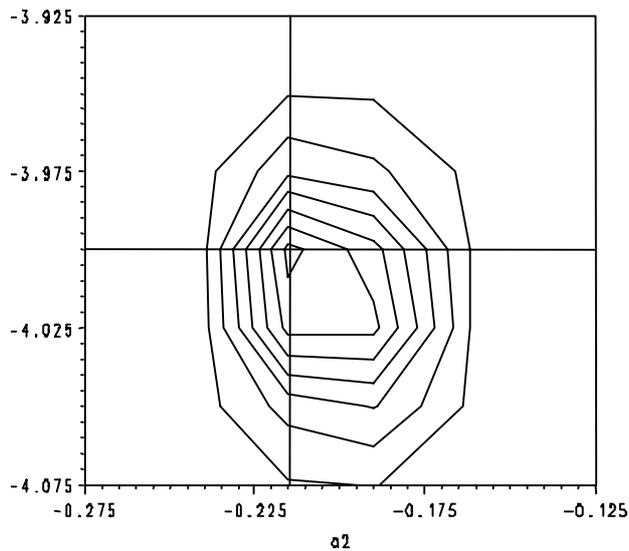


Figura 3.3. Gráfico de contorno de $p(\theta_1, \theta_2 | y)$

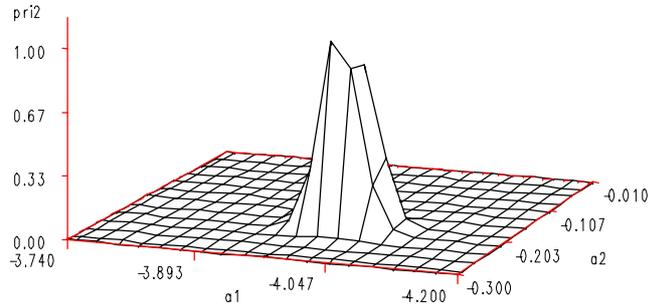


Figura 3.4. Gráfico de $p(\theta_1, \theta_2 | y)$

En las distribuciones marginales para θ_1 y θ_2 presentadas en la Figura 3.5 se puede apreciar cómo las modas respectivas a posteriori se dan en los valores $\theta_1 = -4.015$ y $\theta_2 = -0.2035$.

(Los estimadores de mínimos cuadrados, correspondientes al modelo con errores Normales, son $\hat{\theta}_1 = -4.013$ y $\hat{\theta}_2 = -0.2027$).

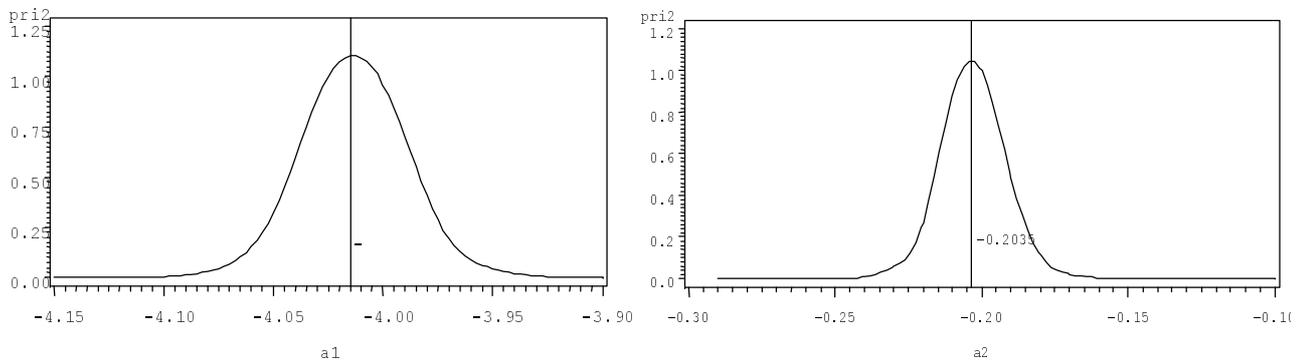


Figura 3.5. Distribuciones $p(\theta_1 | y)$ y $p(\theta_2 | y)$

La Figura 3.6 presenta los histogramas respectivos de las muestras obtenidas de $p(\theta_1 | y)$ y $p(\theta_2 | y)$. No se aprecian diferencias entre los resultados obtenidos con el método de la estimación de la densidad via Monte-Carlo.

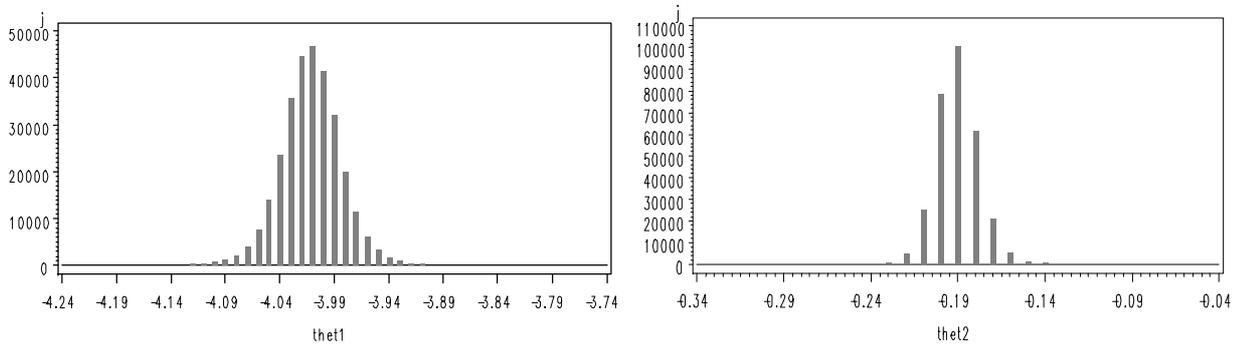


Figura 3.6. Muestras simuladas de $p(\theta_1 | y)$ y $p(\theta_2 | y)$

En el caso del parámetro de escala σ , la distribución a posteriori tiene moda aproximada en $\sigma = 0.048$.

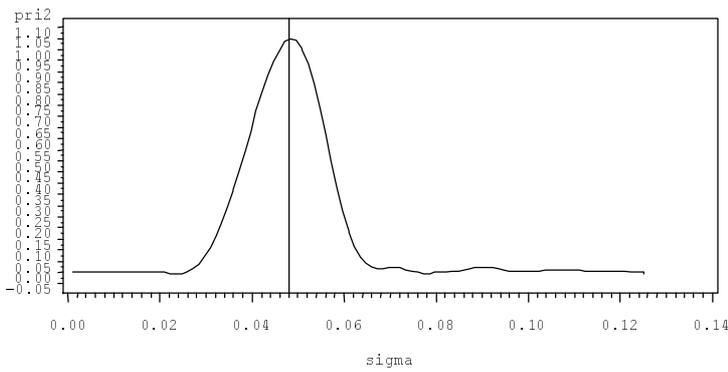


Figura 3.7. Distribución $p(\sigma | y)$

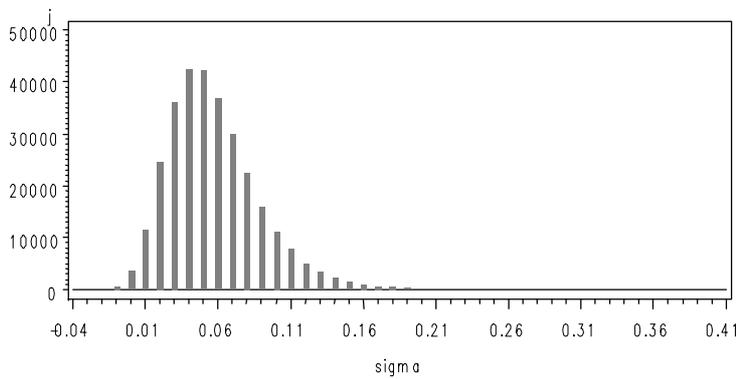


Figura 3.8. Muestras simuladas de $p(\sigma | y)$

En la Figura 3.9 presentamos la nube de puntos y las rectas ajustadas para los

parámetros θ_1 y θ_2 , siguiendo tanto el modelo de errores Normales, como el modelo con errores $PE(0, \sigma, \beta)$ y utilizando la moda a posteriori. Las rectas son respectivamente,

$$y = -4.013 - 0.2027x \text{ para el modelo con } \varepsilon \equiv N(0, \sigma^2)$$

$$y = -4.00 - 0.2145x \text{ para el modelo con } \varepsilon \equiv PE(0, \sigma, \beta)$$

Debido a que, como hemos visto a través de la medida de discrepancia y de la distribución a posteriori $p(\beta | \mathbf{y})$, no se puede asegurar una gran desviación del modelo Normal, estas dos rectas son muy parecidas.

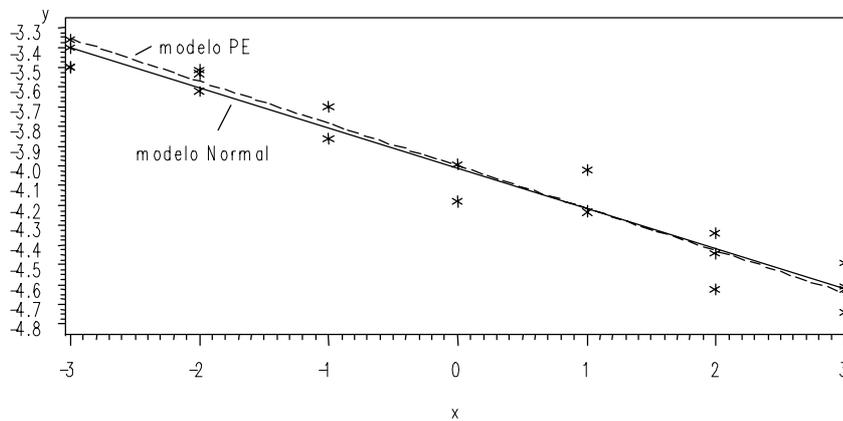


Figura 3.9. Nube de puntos y rectas ajustadas según $\hat{\theta}_1, \hat{\theta}_2 = \text{moda a posteriori}$, en el modelo $PE(\theta, \sigma, \beta)$ y en el modelo Normal.

4.3 Extensión al modelo no lineal

Si la relación entre la variable respuesta y las regresoras se expresa según cierta función g según

$$y_i = g(\mathbf{x}_i | \boldsymbol{\theta}) + \varepsilon_i$$

donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$, y $\mathbf{x}_i = (x_1, \dots, x_k)_i$.

Por ejemplo, en el modelo lineal, tenemos que $p = k + 1$ y

$$g(\mathbf{x} | \boldsymbol{\theta}) = \theta_1 + \theta_2 x_1 + \dots + \theta_{k+1} x_k$$

Análogamente al planteamiento expuesto en la sección anterior, supondremos que los errores ε_i son variables aleatorias independientes con distribución Potencial Exponencial $PE(0, \sigma, \beta)$. Planteando las mismas distribuciones a priori para los parámetros que en el modelo lineal, se obtiene que la distribución a posteriori es

$$p(\boldsymbol{\theta}, \sigma, \beta | \mathbf{y}) \propto [w(\beta)]^n \sigma^{-(n+1)} \exp \left[-c(\beta) \sum_{i=1}^n \left| \frac{y_i - g(\mathbf{x}_i | \boldsymbol{\theta})}{\sigma} \right|^{2/(1+\beta)} \right]$$

Para la estimación via Monte-Carlo de las densidades marginales a posteriori esta expresión no encierra mayores dificultades y se puede plantear de modo general para cualquier función $g(\mathbf{x}_i | \boldsymbol{\theta})$ regular. No es así sin embargo para la extracción de muestras utilizando el método de muestreo de Gibbs directo o la representación por mixturas, pues el desarrollo técnico depende en cada caso de la función $g(\mathbf{x}_i | \boldsymbol{\theta})$ utilizada.

4.3.1 Estimación de las densidades marginales mediante el método Monte-Carlo

El método Monte-Carlo utilizado tiene un desarrollo idéntico al caso del modelo lineal:

Simulando muestras de tamaño m respectivamente de $p(\beta)$, $p((\theta_1, \dots, \theta_k))$, o $p(\theta_l)$, estimaremos el valor de la función proporcional a la densidad marginal de cada uno de los parámetros, respectivamente, por las expresiones

$$\begin{aligned} \hat{p}((\theta_1, \dots, \theta_k) | \mathbf{y}) &\propto \sum_{i=1}^m p((\theta_1, \dots, \theta_k), \beta_i | \mathbf{y}) \\ \hat{p}(\beta | \mathbf{y}) &\propto \sum_{i=1}^m p((\theta_1, \dots, \theta_k)_i, \beta | \mathbf{y}) \\ \hat{p}(\sigma | \mathbf{y}) &\propto \frac{1}{\sigma} \sum_{i=1}^{m^2} p((\theta_1, \dots, \theta_k)_i, \beta_i, \sigma | \mathbf{y}) \\ \hat{p}(\theta_l | \mathbf{y}) &\propto \sum_{i=1}^{m^2} p((\theta_1, \dots, \theta_k)_i^{-l}, \beta_i | \mathbf{y}) \end{aligned}$$

donde el vector $(\theta_1, \dots, \theta_k)_i^{-l}$ representa el vector $(\theta_1, \dots, \theta_k)_i$ eliminada la componente θ_l .

4.3.2 Comparación de dos modelos no lineales.

(Gelfand, Dey y Chang, 1992) plantean la elección entre dos modelos no lineales a partir de la utilización de la distribución predictiva bajo una perspectiva de validación cruzada, es decir, denotando por $\mathbf{y}_{(r)}$ al vector \mathbf{y} sin la r -ésima observación,

$$p(y_r | \mathbf{y}_{(r)}) = \frac{p(y_r, \mathbf{y}_{(r)})}{p(\mathbf{y}_{(r)})} = \frac{p(\mathbf{y})}{p(\mathbf{y}_{(r)})}$$

Una posibilidad para comparar los modelos

$$y_i = g_1(x_i | \boldsymbol{\theta}) + \varepsilon$$

$$y_i = g_2(x_i | \boldsymbol{\theta}) + \varepsilon_i$$

consiste en utilizar una función de diagnóstico que represente el ajuste del modelo para cada observación $h(y_r)$ y calcular su esperanza bajo la distribución predictiva presentada. Es decir, calcular $d_r = E(h(Y_r) | \mathbf{y}_{(r)})$. Tendríamos así una medida intuitiva del ajuste de cada uno de los dos modelos para cada observación, que podría agregarse para obtener una medida general.

Concretamente, una posibilidad es tomar

$$h(y_r) = y_r - Y_r$$

de manera que

$$d_r = y_r - E(Y_r | \mathbf{y}_{(r)})$$

Siendo estos d_r residuos mencionados en (Geisser, 1987). Estas medidas se pueden estandarizar tomando $\sigma_r^2 = \text{Var}(Y_r | \mathbf{y}_{(r)})$ y calculando $d'_r = d_r / \sigma_r$.

La suma $\sum (d_r)^2$ puede usarse como un índice de bondad de ajuste de cada modelo. Los d'_r permitirán identificar observaciones anómalas.

Como veremos a continuación, la estimación de $E(Y_r | \mathbf{y}_{(r)})$ se puede realizar a partir de estimaciones Monte-Carlo.

En el caso que nos ocupa, si consideramos los errores independientes y distribuidos según una Potencial Exponencial, es decir,

$$\varepsilon \equiv PE(0, \sigma, \beta)$$

Tenemos que, para cada r ,

$$p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}) \equiv PE(g(x_r, \theta), \sigma, \beta)$$

Además,

$$p(Y_r | \mathbf{y}_{(r)}) = \int p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}, \mathbf{y}_{(r)}) p(\boldsymbol{\theta}, \sigma, \beta, \mathbf{x} | \mathbf{y}_{(r)}) d\theta d\sigma d\beta$$

y, por lo tanto,

$$E(Y_r | \mathbf{y}_{(r)}) = \int Y_r p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}, \mathbf{y}_{(r)}) p(\boldsymbol{\theta}, \sigma, \beta, \mathbf{x} | \mathbf{y}_{(r)}) d\boldsymbol{\theta} d\sigma d\beta dY_r$$

Suponemos independencia condicional en la distribución de Y_r , de manera que

$$p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}, \mathbf{y}_{(r)}) = p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x})$$

y por lo tanto podemos poner

$$\begin{aligned} E(Y_r | \mathbf{y}_{(r)}) &= \int \left[\int Y_r p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}) dY_r \right] p(\boldsymbol{\theta}, \sigma, \beta | \mathbf{y}_{(r)}, \mathbf{x}) d\boldsymbol{\theta} d\sigma d\beta = \\ &= \int [E(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x})] p(\boldsymbol{\theta}, \sigma, \beta | \mathbf{y}_{(r)}, \mathbf{x}) d\boldsymbol{\theta} d\sigma d\beta = \\ &= \int g(x_r, \theta) p(\boldsymbol{\theta}, \sigma, \beta | \mathbf{y}_{(r)}, \mathbf{x}) d\boldsymbol{\theta} d\sigma d\beta \end{aligned}$$

Esta expresión se puede integrar en σ , con lo que queda

$$E(Y_r | \mathbf{y}_{(r)}) = \frac{1}{p(\mathbf{y}_{(r)})} \int g(x_r, \theta) p(\mathbf{y}_{(r)} | \boldsymbol{\theta}, \beta, \mathbf{x}) p(\boldsymbol{\theta}) p(\beta) d\boldsymbol{\theta} d\beta$$

En esta última expresión puede utilizarse el método Monte-Carlo expuesto para la estimación de valores de la distribución a posteriori $p(\boldsymbol{\theta}, \sigma, \beta | \mathbf{y}_{(r)}, \mathbf{x})$, de la siguiente manera:

1. Se generan i muestras de $p(\boldsymbol{\theta})$, y $p(\beta)$
2. Se estima la constante normalizadora $\frac{1}{p(\mathbf{y}_{(r)})}$ por Monte-Carlo o bien por integración numérica en los puntos $(\boldsymbol{\theta}, \beta)^{(i)}$ calculados en la expresión $p(\mathbf{y}_{(r)} | \boldsymbol{\theta}, \beta, \mathbf{x}) p(\boldsymbol{\theta}) p(\beta)$
3. Se calcula $\hat{E}(Y_r | \mathbf{y}_{(r)}) = \frac{1}{m} \frac{1}{p(\mathbf{y}_{(r)})} \sum_{i=1}^m g(x_r, \boldsymbol{\theta}^{(i)}) p(\mathbf{y}_{(r)} | (\boldsymbol{\theta}, \beta)^{(i)}, \mathbf{x})$

Finalmente se calcula la medida

$$d_r = y_r - \hat{E}(Y_r | \mathbf{y}_{(r)})$$

Todos estos cálculos se deben repetir para cada observación r .

Por otra parte, para estandarizar estos residuos, y para calcular $\sigma_r^2 = \text{Var}(Y_r | \mathbf{y}_{(r)})$, se puede tener en cuenta, suponiendo de nuevo independencia condicional en Y_r , que

$$V(Y_r | \mathbf{y}_{(r)}) = E(Y_r^2 | \mathbf{y}_{(r)}) - [E(Y_r | \mathbf{y}_{(r)})]^2$$

y

$$E(Y_r^2 | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}, \mathbf{y}_{(r)}) = E(Y_r^2 | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}) =$$

$$\begin{aligned}
&= V(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}) + [E(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x})]^2 = \\
&= \sigma^2 + g(x_r, \boldsymbol{\theta})^2
\end{aligned}$$

Por lo tanto, para estimar la varianza $V(Y_r | \mathbf{y}_{(r)})$ utilizando el método Monte-Carlo, tenemos que

$$\begin{aligned}
V(Y_r | \mathbf{y}_{(r)}) &= \int Y_r^2 p(Y_r | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}, \mathbf{y}_{(r)}) p(\boldsymbol{\theta}, \sigma, \beta, \mathbf{x} | \mathbf{y}_{(r)}) d\boldsymbol{\theta} d\sigma d\beta dY_r - [E(Y_r | \mathbf{y}_{(r)})]^2 \\
&= \int [\sigma^2 + g(x_r, \theta)] p(\boldsymbol{\theta}, \sigma, \beta, \mathbf{x} | \mathbf{y}_{(r)}) d\boldsymbol{\theta} d\sigma d\beta - [E(Y_r | \mathbf{y}_{(r)})]^2 \\
&= \frac{1}{p(\mathbf{y}_{(r)})} \int [\sigma^2 + g(x_r, \theta)] p(\mathbf{y}_{(r)} | \boldsymbol{\theta}, \sigma, \beta, \mathbf{x}) p(\boldsymbol{\theta}) p(\beta) p(\sigma) d\boldsymbol{\theta} d\sigma d\beta - [E(Y_r | \mathbf{y}_{(r)})]^2
\end{aligned}$$

En esta expresión la primera integral se puede estimar con un método similar al presentado anteriormente para estimar $E(Y_r | \mathbf{y}_{(r)})$ y en el segundo término podemos aproximar $[E(Y_r | \mathbf{y}_{(r)})]^2$ por $[\hat{E}(Y_r | \mathbf{y}_{(r)})]^2$.

Una vez calculados $\hat{E}(Y_r | \mathbf{y}_{(r)})$ y $\hat{V}(Y_r | \mathbf{y}_{(r)})$ se construyen

$$d'_r = \frac{d_r}{\sigma_r} = \frac{y_r - \hat{E}(Y_r | \mathbf{y}_{(r)})}{\sqrt{\hat{V}(Y_r | \mathbf{y}_{(r)})}}$$

Obtendremos además por el método Monte-Carlo, estimaciones para las densidades a posteriori de los parámetros de regresión y de σ y β , con lo que podremos estudiar la posibilidad de utilizar o no el modelo con errores normales en alguno de los dos modelos.

4.3.3 Aplicación

(Gelfand, Dey y Chang, 1992) comparan los modelos siguientes

$$\text{Modelo logístico: } y = \theta_0(1 + \theta_1\theta_2^x)^{-1} + \varepsilon$$

$$\text{Modelo de Gompertz: } y = \theta_0 e^{-\theta_1\theta_2^x} + \varepsilon$$

donde la variable Y es el peso de bulbos de cebolla medido a través del tiempo creciente X .

X	1	2	3	4	5	6	7	8	9
Y	16.08	33.83	65.80	97.2	191.55	326.2	386.87	520.53	590.03
10	11	12	13	14	15				
651.92	724.93	699.56	689.86	637.56	717.41				

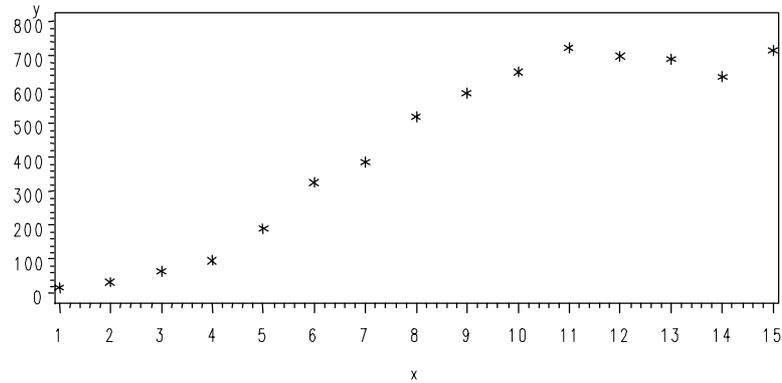


Figura 3.10

Estos autores plantean errores ε distribuidos según una Normal, utilizando distribuciones no informativas para los parámetros. En nuestro caso supondremos que los errores se distribuyen según la Potencial Exponencial, comparando ambos modelos en este marco y poniendo de relieve las diferencias de ajuste utilizando esta distribución para el término de error. Compararemos no solamente las diferencias de ajuste para los dos modelos, sino también respecto a utilizar el modelo de errores distribuidos según una Normal y según la distribución Potencial Exponencial.

En ambos modelos, logístico y de Gompertz, utilizaremos la reparametrización siguiente:

- $\theta'_1 = \log(\theta_1)$
- $\theta'_2 = \log\left(\frac{\theta_2}{1 - \theta_2}\right)$

Las distribuciones a priori que tomaremos sobre los parámetros serán independientes a priori y no informativas, de manera que

$$p(\theta_0, \theta'_1, \theta'_2, \sigma, \beta) \propto \frac{1}{\sigma}$$

Modelo Logístico. Para este modelo, se presentan las distribuciones a posteriori marginales de los parámetros en la figura siguiente.

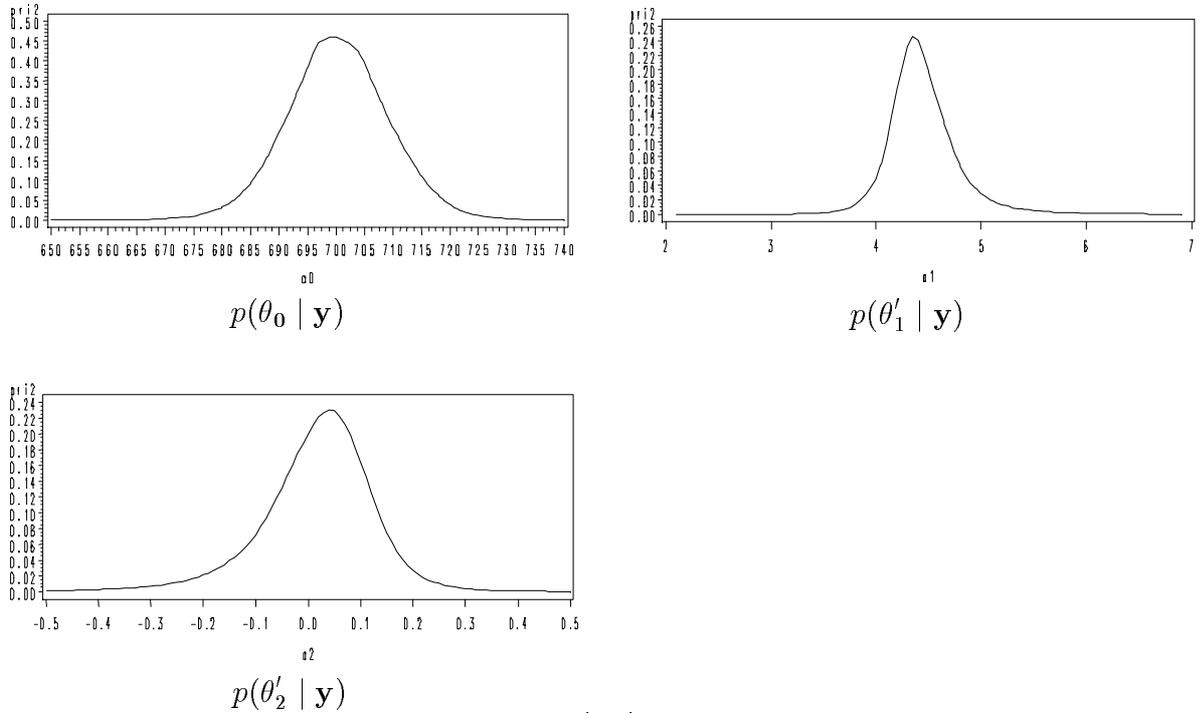


Figura 3.11. Distribuciones a posteriori de $\theta_0, \theta'_1, \theta'_2$

Las distribuciones a posteriori para los parámetros σ y β correspondientes al error ε quedan reflejadas en la siguiente figura.

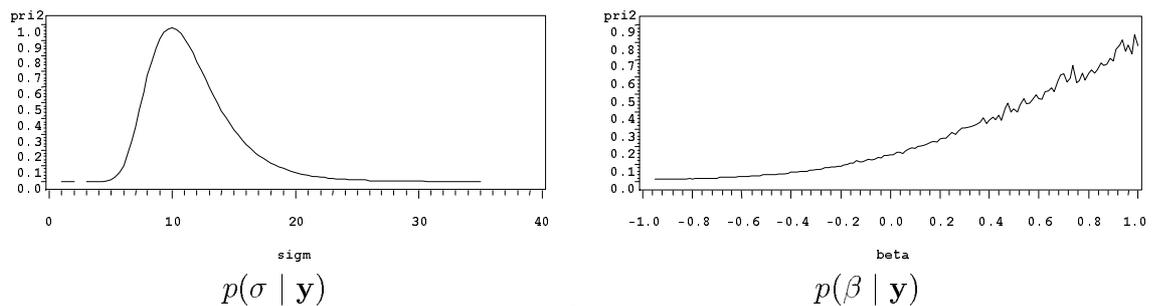


Figura 3.12. Distribuciones a posteriori de σ, β

A la vista del último gráfico, donde la distribución $p(\beta | \mathbf{y})$ tiene moda en $\beta = 1$ y por lo tanto alejada del valor $\beta = 0$ correspondiente a Normalidad, parece más apropiado utilizar el modelo Potencial Exponencial que el modelo usual con los errores normales, si, como estamos haciendo, se adopta el modelo logístico a los datos.

Se han realizado también estimaciones de las distribuciones a posteriori de los parámetros para el caso concreto de errores normales, considerando por lo tanto $\beta = 0$ constante en el modelo Potencial Exponencial. Los resultados aparecen en la Figura 3.13.

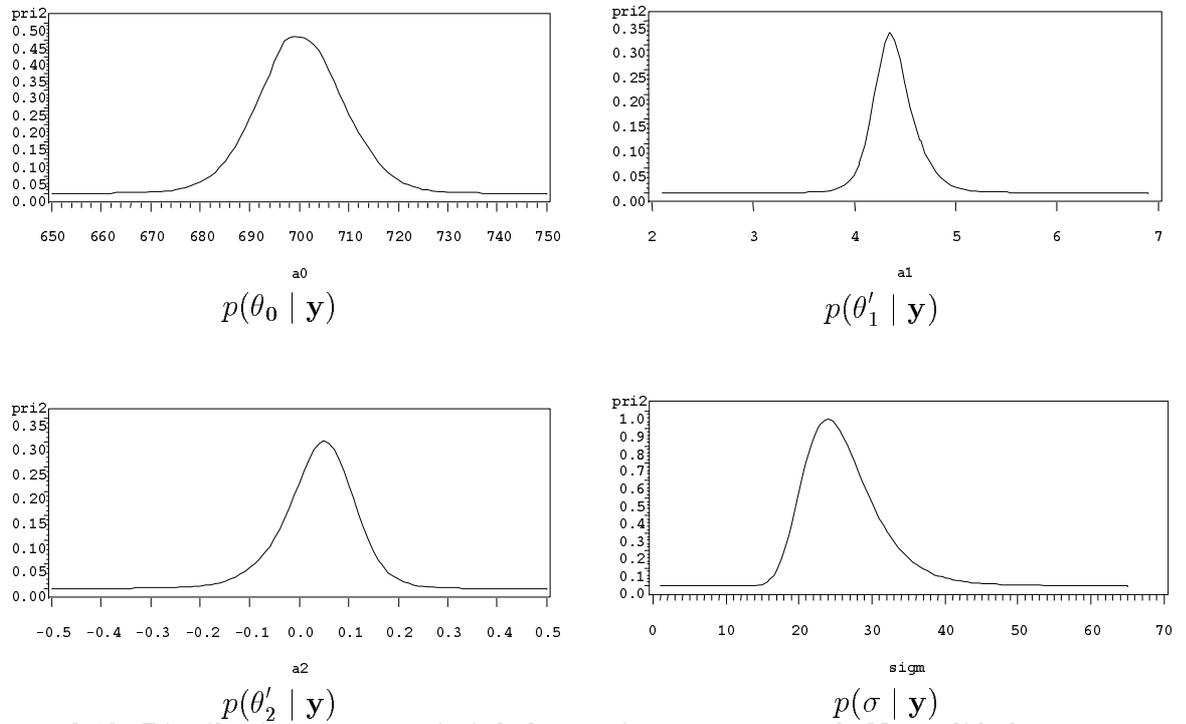


Figura 3.13. Distribuciones a posteriori de los parámetros en caso de Normalidad.

Finalmente se presenta una tabla con las medias y modas a posteriori de los parámetros de interés en ambos casos:

	Modelo Potencial Exponencial					Modelo Normal			
	θ_0	θ'_1	θ'_2	σ	β	θ_0	θ'_1	θ'_2	σ
Moda a posteriori	699.75	4.37	0.0158	11.3	1	697	4.35	0.015	25.6
Media a posteriori	702.31	4.44	0.016	13.5	0.8	702.30	4.39	0.016	27.3

Modelo Gompertz. Utilizando la misma reparametrización, y con el modelo Potencial Exponencial para los errores, tenemos las distribuciones a posteriori siguientes:

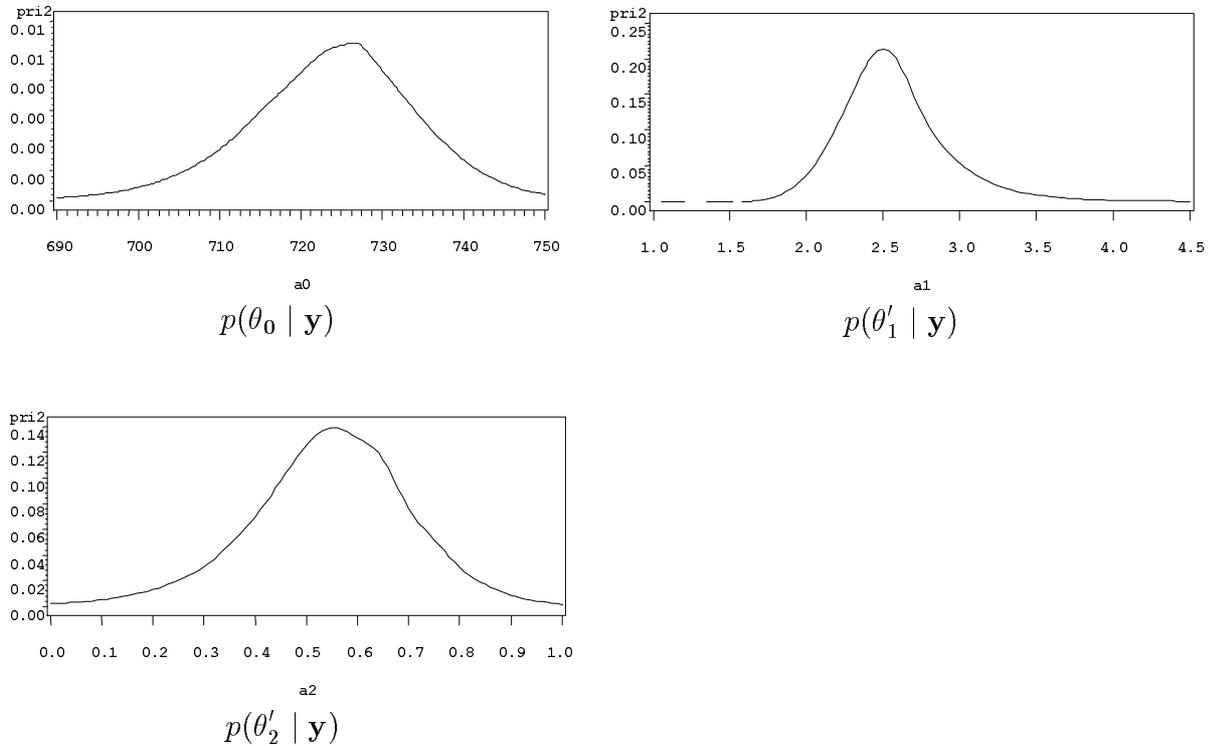


Figura 3.14. Distribuciones a posteriori de $\theta_0, \theta'_1, \theta'_2$
Las distribuciones marginales de σ y β son, en este modelo,

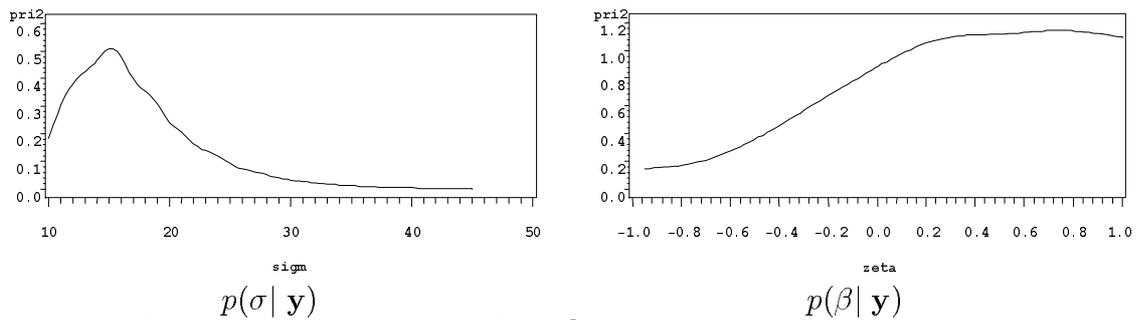


Figura 3.15. Distribuciones a posteriori de σ, β

Como ocurre con el modelo logístico, parece más adecuado plantear un modelo Potencial Exponencial, dado que la distribución a posteriori $p(\beta | \mathbf{y})$ tiene su moda muy alejada del valor $\beta = 0$ correspondiente a la distribución Normal. La medida $P(C' | \mathbf{y}) = 0.85$.

Si planteamos el modelo Normal para los errores ε , (es decir $\beta = 0$ en el modelo

Potencial Exponencial), obtenemos las distribuciones

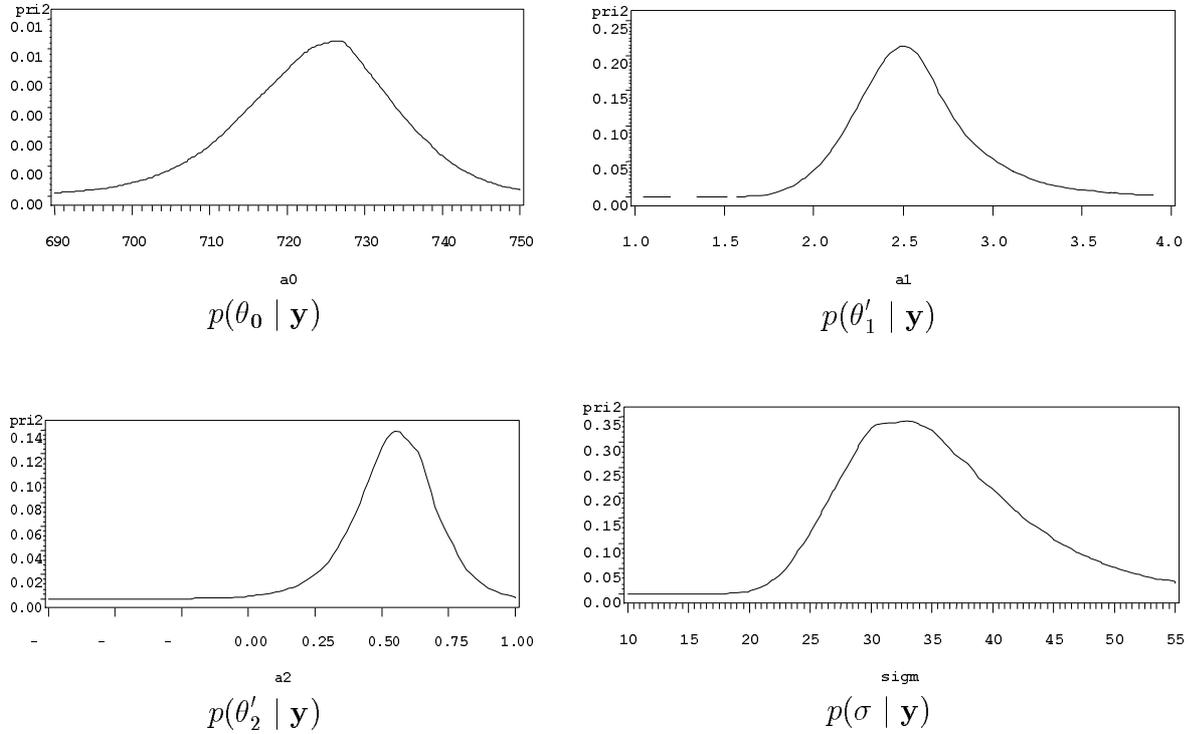


Figura 3.16. Distribuciones a posteriori de los parámetros en el caso de normalidad

La Tabla de estimaciones a posteriori para los parámetros se presenta a continuación. Las medias a posteriori se estiman por integración numérica.

	Modelo Potencial Exponencial					Modelo Normal			
	θ_0	θ'_1	θ'_2	σ	β	θ_0	θ'_1	θ'_2	σ
Moda a posteriori	726.5	2.55	0.55	15.8	0.8	721.5	2.55	0.55	33
Media a posteriori	723.8	2.567	0.543	19.5	0.75	721.37	2.57	0.54	33.8

Ajuste de los modelos. Comparación. Aplicando el método de comparación a partir de la estimación $\hat{E}(Y_r | \mathbf{y}_{(r)})$, obtenemos los siguientes resultados:

		Modelo Logístico					
		$\varepsilon \equiv$ Potencial Exponencial			$\varepsilon \equiv$ Normal		
x_r	y_r	$\hat{E}(Y_r \mathbf{y}_{(r)})$	d_r	d'_r	$\hat{E}(Y_r \mathbf{y}_{(r)})$	d_r	d'_r
1	16.08	16.72	-0.64	-0.06	17.03	-0.95	-0.04
2	33.83	32.53	1.29	0.11	32.69	1.13	0.04
3	65.80	61.92	3.87	0.33	61.47	4.33	0.17
4	97.20	113.01	-15.81	-1.35	111.19	-13.99	-0.53
5	191.55	192.21	-0.66	-0.06	188.67	2.87	0.11
6	326.20	296.60	29.59	2.62	292.29	33.90	1.30
7	386.87	409.92	-23.05	-1.93	406.58	-19.70	-0.78
8	520.53	510.93	9.59	0.80	509.42	11.11	0.42
9	590.03	586.71	3.31	0.29	586.57	3.45	0.14
10	651.92	636.44	15.47	1.35	636.87	15.04	0.58
11	724.93	666.19	58.73	5.04	666.69	58.23	2.25
12	699.56	682.95	16.60	1.37	683.34	16.21	0.62
13	689.96	692.07	-2.11	-0.18	692.32	-2.36	-0.09
14	637.56	696.93	-59.37	-5.21	697.08	-59.52	-2.28
15	717.41	699.49	17.91	1.46	699.58	17.82	0.70

		Modelo de Gompertz					
		$\varepsilon \equiv$ Potencial Exponencial			$\varepsilon \equiv$ Normal		
x_r	y_r	$\hat{E}(Y_r \mathbf{y}_{(r)})$	d_r	d'_r	$\hat{E}(Y_r \mathbf{y}_{(r)})$	d_r	d'_r
1	16.08	1.12	14.95	0.91	1.13	14.94	0.45
2	33.83	9.63	24.19	1.45	9.57	24.25	0.72
3	65.8	41.24	24.55	1.56	40.79	25.00	0.74
4	97.2	108.53	-11.33	-0.73	107.49	-10.29	-0.31
5	91.55	206.23	-14.68	-0.98	204.73	-13.18	-0.39
6	326.2	315.65	10.54	0.63	313.93	12.26	0.38
7	386.87	418.50	-31.63	-2.08	416.68	-29.81	-0.88
8	520.53	504.39	16.13	1.03	502.46	18.06	0.54
9	590.03	570.58	19.44	1.28	568.53	21.49	0.67
10	651.92	618.92	32.99	2.10	616.74	35.17	1.07
11	724.93	652.96	71.96	4.67	650.68	74.24	2.20
12	699.56	676.36	23.19	1.43	674.02	25.54	0.75
13	689.96	692.20	-2.24	-0.14	689.81	0.14	0.01
14	637.56	702.81	-65.25	-3.97	700.39	-62.83	-1.90
15	717.41	709.88	7.52	0.46	707.45	9.96	0.30

	Modelo Logístico		Modelo de Gompertz	
	Potencial Exponencial	Normal	Potencial Exponencial	Normal
$\sum (d_r)^2$	9593.30	9646.06	14632.62	14993.61
$\sum d_r $	258.06	260.68	370.67	377.23

Se ve cómo el modelo Potencial Exponencial parece ajustar ligeramente mejor en ambos modelos (logístico y de Gompertz), algo que se preveía tras presentar la distribución a posteriori de β en los dos casos. Por otra parte, el modelo logístico parece comportarse con claridad mejor que el modelo de Gompertz en el ajuste de los datos.

Por último se presenta el gráfico de los residuos estandarizados d'_r en los modelos que mejor se ajustan (modelo logístico con distribución Potencial Exponencial para los errores y modelo logístico con distribución Normal para los errores). El objetivo es ahora detectar datos anómalos. Como se ve en el gráfico, las observaciones correspondientes a $x = 11$ y $x = 14$ son datos problemáticos para ambos modelos (esto también se puede confirmar observando la tabla de valores en el modelo de Gompertz).

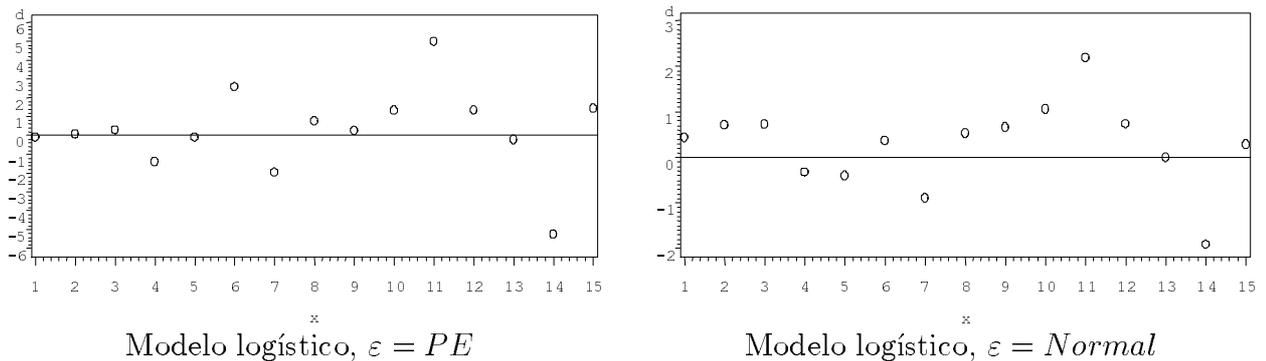


Figura 3.17. Gráficos de residuos

4.4 Aplicación a un Modelo lineal con datos longitudinales

(Han, Carlin, 2001) plantean comparar dos modelos con el objetivo de tratar datos longitudinales en un experimento clínico con pacientes con SIDA. Estos datos ya fueron tratados en otros estudios como (Carlin, Louis, 2000) o (Goldman et al, 1996). Consisten en mediciones del número de linfocitos afectados por SIDA y llamados CD4, realizadas a cada sujeto a los 0, 2, 6, 12 y 18 meses de comenzar el seguimiento. Se dividió a los pacientes en dos grupos según dos tratamientos distintos (variable $z_1 = 1$ o 0), y también se consideró si tenían un diagnóstico de SIDA positivo en $t = 0$ (variable $z_2 = 1$ o 0). Debido a múltiples factores existen muchos datos ausentes en la muestra. Se considera

conveniente transformar mediante la raíz cuadrada las medidas de CD4, con lo cual el vector respuesta será $Y_i = (Y_{i1}, \dots, Y_{i5})$, la raíz del número de CD4 para cada paciente i , tomado en cada tiempo $j = 1, \dots, 5$, equivalentes a los tiempos $t_{ij} = 0, 2, 6, 12, 18$ meses. Debido a los datos ausentes el vector Y_i tendrá dimensión variable para cada paciente. Llamaremos s_i al número de mediciones en el tiempo de cada paciente Y_i , es decir, a la dimensión de este vector.

Nos disponemos a comparar los modelos M_1 y M_2 , donde

$$M_1 : Y_i = X_i \mathbf{a} + W_i \mathbf{b}_i + \varepsilon_i$$

y

$$M_2 : Y_i = P_i \mathbf{c} + Q_i \mathbf{d}_i + \nu_i$$

En el primer modelo, W_i es una matriz $s_i \times 3$, con fila j -ésima $(1, t_{ij}, \max(0, t_{ij} - 2))$ X_i es una matriz $s_i \times 9$, tal que $X_i = (W_i \mid z_{1i} W_i \mid z_{2i} W_i)$.

En este marco, \mathbf{a} es un vector de 9 componentes (a_1, \dots, a_9) y los \mathbf{b}_i relativos a efectos aleatorios debidos a cada sujeto, son vectores de 3 componentes $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})$.

El segundo modelo es similar al primero, pero en el modelo M_2 la matriz Q_i es de dimensiones $s_i \times 2$, tal que tiene fila j -ésima $(1, t_{ij})$. En concreto, la diferencia entre los dos modelos radica en que el modelo M_1 permite un punto de cambio en la evolución lineal de la variable respuesta en el mes $t = 2$, que pretende reflejar las consecuencias inmediatas de los tratamientos, mientras que el modelo más simple M_2 aproxima la tendencia lineal de la variable respuesta respecto al tiempo, desde $t = 0$, sin considerar puntos de cambio. En el desarrollo bayesiano siguiente nos centraremos en el modelo M_1 , pues no hay diferencia sustancial en el tratamiento básico de ambos modelos.

4.4.1 Distribuciones a priori de los parámetros

Siguiendo un planteamiento similar al de (Zeger, Karim, 1992) supondremos una distribución a priori no informativa para el vector \mathbf{a} . Los vectores \mathbf{b}_i seguirán a priori una distribución Normal Multivariante $N(0, \mathbf{V})$, donde \mathbf{V}^{-1} es una matriz definida positiva que sigue a su vez una distribución de Wishart $W((\rho R)^{-1}, \rho)$ con ciertos valores iniciales que tomaremos idénticos a los valores sugeridos por (Carlin, Louis, 2000) para estos datos.

Continuando con la propuesta de modelizar los errores según una distribución Potencial Exponencial, planteamos $\varepsilon_i \propto PE(0, \sigma^2 I, \beta)$, independientes e igualmente distribuidos. La extensión multivariante no será necesaria a la hora de los desarrollos, pues trabajaremos en cada componente por separado.

Para cada observación, (paciente i y momento del tiempo j), tendremos

$$y_{ij} = \mathbf{x}_{ij}\mathbf{a} + \mathbf{w}_{ij}\mathbf{b}_i + \varepsilon_{ij}$$

es decir,

$$y_{ij} = (x_{ij}^1 a_1 + \dots + x_{ij}^9 a_9) + (w_{ij}^1 b_{i1} + \dots + w_{ij}^3 b_{i3}) + \varepsilon_{ij}$$

De este modo, la verosimilitud quedará

$$\begin{aligned} p(y_{ij} | \mathbf{a}, \mathbf{b}_i, \sigma, \beta) &\propto PE(\mathbf{x}_{ij}\mathbf{a} + \mathbf{w}_{ij}\mathbf{b}_i, \sigma, \beta) = \\ &= [w(\beta)]^{ns_i} \sigma^{-ns_i} \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} \left| \frac{y_{ij} - (\mathbf{x}_{ij}\mathbf{a} + \mathbf{w}_{ij}\mathbf{b}_i)}{\sigma} \right|^{2/(1+\beta)} \right] \end{aligned}$$

Si deseamos comparar ambos modelos utilizando la técnica del apartado anterior, se llega a

$$E(Y_{ij} | \mathbf{y}_{(ij)}) = \int (\mathbf{x}_{ij}\mathbf{a} + \mathbf{w}_{ij}\mathbf{b}_i) p(\sigma, \beta, \mathbf{a}, \mathbf{b}_i | \mathbf{y}_{(ij)}) d\sigma d\beta d\mathbf{a} d\mathbf{b}_i$$

Es decir, para estimar $E(Y_{ij} | \mathbf{y}_{(ij)})$, utilizaremos el método Monte-Carlo, extrayendo muestras de $p(\sigma, \beta, \mathbf{a}, \mathbf{b}_i | \mathbf{y}_{(ij)})$ y evaluando la función $(\mathbf{x}_{ij}\mathbf{a} + \mathbf{w}_{ij}\mathbf{b}_i)$, mediante

$$\hat{E}(Y_{ij} | \mathbf{y}_{(ij)}) = \frac{1}{m} \sum_{k=1}^m (\mathbf{x}_{ij}\mathbf{a}^{(m)} + \mathbf{w}_{ij}\mathbf{b}_i^{(m)})$$

En esta última estimación se considerarán solamente las mediciones en el tiempo concreto j existentes para el paciente i , para poder establecer la medida de ajuste $d_{ij} = y_{ij} - \hat{E}(Y_{ij} | \mathbf{y}_{(ij)})$ y posteriormente $D(\text{Modelo } M_1) = D_1 = \sum_{i=1}^n \sum_{j=1}^{s_i} d_{ij}^2$ como una medida global de ajuste. Respectivamente, para el modelo M_2 se calculará la cantidad D_2 . Finalmente se compararán ambas cantidades D_1 y D_2 .

4.4.2 Simulación de las distribuciones a posteriori

Se desean extraer muestras de $p(\sigma, \beta, \mathbf{a}, \mathbf{b} | \mathbf{y}_{(ij)})$. Para ello utilizaremos el muestreo de Gibbs, extrayendo sucesivamente muestras de $p(\sigma | \beta, \mathbf{a}, \mathbf{b}, \mathbf{y}_{(ij)})$, $p(\beta | \sigma, \mathbf{a}, \mathbf{b}, \mathbf{y}_{(ij)})$, $p(\mathbf{a} | \sigma, \beta, \mathbf{b}, \mathbf{y}_{(ij)})$ y $p(\mathbf{b} | \sigma, \beta, \mathbf{a}, \mathbf{V}, \mathbf{y}_{(ij)})$. Es necesario disponer de medios para extraer muestras de estas distribuciones, para lo que recurriremos a las técnicas expuestas en el

apartado de simulación de muestras de los parámetros de una distribución Potencial Exponencial. Para no sobrecargar la notación, haremos los desarrollos suponiendo un vector genérico de datos \mathbf{y} .

Muestras de la distribución $p(\sigma | \beta, a, b, \mathbf{y})$

Al ser

$$p(\sigma | \beta, \mathbf{a}, \mathbf{b}, \mathbf{y}) \propto \sigma^{-(\sum s_i + 1)} \exp \left[-\frac{1}{\sigma^{2/(1+\beta)}} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} |y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)|^{2/(1+\beta)} \right]$$

$$\text{Fijando } a = \frac{\sum_{i=1}^n s_i (1 + \beta)}{2} \text{ y } b = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} |y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)|^{2/(1+\beta)},$$

- 1-. Se extrae una muestra q de una Gamma $\Gamma(a, b)$
- 2-. Se calcula $\sigma = q^{-(1+\beta)/2}$

Se verifica que σ proviene de la distribución objetivo.

Muestras de la distribución $p(\beta | \sigma, a, b, \mathbf{y})$

En este caso,

$$p(\beta | \sigma, \mathbf{a}, \mathbf{b}, \mathbf{y}) \propto \frac{1}{\sum s_i (3 + \beta)} \frac{1}{[\Gamma(\frac{1}{2}(3 + \beta))]^{n \sum s_i}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} \left| \frac{y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)}{\sigma} \right|^{2/(1+\beta)} \right]$$

que está acotada por

$$\frac{1}{2^{3n/2}} \frac{1}{[\Gamma(\frac{3}{2})]^n} \exp(-(\frac{n}{2} \log(2))\beta)$$

con lo que utilizaremos el siguiente método de rechazo:

1. Extraer una muestra β de una distribución exponencial $\text{Exp}(\frac{n}{2} \log(2))$ truncada en $\beta \in (-1, 1]$
2. Tomar $\log t = -n \log(\Gamma(3/2)) - \frac{3n}{2} \log(2) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} \left| \frac{y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)}{\sigma} \right|^{2/(1+\beta)}$

3. Extraer u de una $U(0, 1)$
4. Si $\log(u) \leq -\log t$, Aceptar β . En caso contrario, rechazar β y volver al paso 1.

Muestras de la distribución $p(\mathbf{a} \mid \sigma, \beta, \mathbf{b}, \mathbf{y})$

En este caso, $\mathbf{a} = (a_1, \dots, a_9)$, con lo que utilizaremos el muestreo de Gibbs generando alternativamente muestras de las distribuciones $p(a_k \mid \sigma, \beta, \mathbf{a}_{(k)}, \mathbf{b}, \mathbf{y})$, denotando por $\mathbf{a}_{(k)}$ el vector (a_1, \dots, a_9) sin la componente k -ésima.

Para generar valores de $p(a_k \mid \sigma, \beta, \mathbf{a}_{(k)}, \mathbf{b}, \mathbf{y})$ hay que ver que, denotando por

$$z_{ij}^k = \left[\frac{y_{ij} - \mathbf{w}_{ij} \mathbf{b}_i - \sum_{l \neq k} x_{ij}^l a_l}{x_{ij}^k} \right]$$

se tiene que

$$p(a_k \mid \sigma, \beta, \mathbf{a}_{(k)}, \mathbf{b}, \mathbf{y}) \propto \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} \left| \frac{(a_k - z_{ij}^k)}{\sigma / x_{ij}^k} \right|^{2/(1+\beta)} \right] \quad (4.1)$$

Con lo que se tiene el marco apropiado para utilizar el algoritmo de rechazo usual a la hora de trabajar con la distribución Potencial Exponencial.

Definiendo las cantidades $\bar{z}^k = \sum_{i=1}^n \sum_{j=1}^{s_i} z_{ij}^k$ y $\sigma_{\max} = \frac{\sigma}{\text{Min}\{x_{ij}^k\}}$, se verifica que la expresión (4.1) está acotada por

$$\exp \left[-\frac{\sum s_i}{2} \left| \frac{a_k - \bar{z}^k}{\sigma_{\max}} \right|^{2/(1+\beta)} \right]$$

De modo que el método de rechazo para simular muestras de a_k tomará la siguiente forma:

1. Extraer una muestra a_k de una Potencial Exponencial $\text{PE}(\bar{z}^k, \frac{\sigma_{\max}}{\sum s_i^{(1+\beta)/2}}, \beta)$
2. Tomar $\log t = -\frac{\sum s_i}{2} \left| \frac{a_k - \bar{z}^k}{\sigma_{\max}} \right|^{2/(1+\beta)} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} \left| \frac{y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)}{\sigma} \right|^{2/(1+\beta)}$
3. Extraer u de una $U(0, 1)$

4. Si $\log(u) \leq -\log t$, Aceptar a_k . En caso contrario, rechazar a_k y volver al paso 1

Esta operación se realizará sucesivamente con los a_1, \dots, a_9 con un número suficiente de iteraciones para obtener una muestra final proveniente de $p(\mathbf{a} \mid \sigma, \beta, \mathbf{b}, \mathbf{y}_{(ij)})$.

Muestras de la distribución $p(b_i \mid \sigma, \beta, a, V, y)$

En este caso, tenemos que $p(\mathbf{b}_i \mid \sigma, \beta, \mathbf{a}, \mathbf{V}, \mathbf{y}) \propto p(\mathbf{y} \mid \sigma, \beta, \mathbf{a}, \mathbf{b}_i) p(\mathbf{b}_i \mid \mathbf{V})$

Es necesario obtener, para cada $i = 1, \dots, n$, una muestra del vector $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})$. Aplicando la misma idea que en la simulación del vector \mathbf{a} , y llamando $\mathbf{b}_{i(k)}$ al vector \mathbf{b}_i sin la componente k , tenemos:

$$p(b_{ik} \mid \sigma, \beta, \mathbf{a}, \mathbf{b}_{i(k)}, \mathbf{V}, \mathbf{y}) \propto \exp \left[-\frac{1}{2} \sum_{j=1}^{s_i} \left| \frac{y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)}{\sigma} \right|^{2/(1+\beta)} - \frac{1}{2} \frac{\mathbf{b}_i^t \mathbf{b}_i}{|\mathbf{V}|} \right] =$$

$$\exp \left[-\frac{1}{2} \sum_{j=1}^{s_i} \left| \frac{b_{ik} - z_{ij}^k}{\sigma / w_{ij}^k} \right|^{2/(1+\beta)} - \frac{1}{2} \frac{\mathbf{b}_i^t \mathbf{b}_i}{|\mathbf{V}|} \right]$$

donde

$$z_{ij}^k = \frac{y_{ij} + \mathbf{x}_{ij} \mathbf{a} - \sum_{l \neq k} w_{ij}^l b_l}{w_{ij}^k}$$

Definiendo las cantidades

$$\bar{z}_i^k = \sum_{j=1}^{s_i} z_{ij}^k$$

$$\sigma_{\max} = \frac{\sigma}{\text{Min}\{w_{ij}^k\}}$$

La cota es ahora

$$\exp \left[-\frac{s_i}{2} \left| \frac{b_{ik} - \bar{z}_i^k}{\sigma_{\max}} \right|^{2/(1+\beta)} \right]$$

De modo que el método de rechazo para simular muestras de b_{ik} tomará la siguiente forma:

1. Extraer una muestra b_{ik} de una Potencial Exponencial $\text{PE}(\bar{z}_i^k, \frac{\sigma_{\max}}{s_i^{(1+\beta)/2}}, \beta)$
2. Tomar $\log t = -\frac{s_i}{2} \left| \frac{b_{ik} - \bar{z}_i^k}{\sigma_{\max}} \right|^{2/(1+\beta)} + \frac{1}{2} \sum_{j=1}^{s_i} \left| \frac{y_{ij} - (\mathbf{x}_{ij} \mathbf{a} + \mathbf{w}_{ij} \mathbf{b}_i)}{\sigma} \right|^{2/(1+\beta)}$
3. Extraer u de una $U(0, 1)$
4. Si $\log(u) \leq -\log t$, Aceptar b_k . En caso contrario, rechazar b_k y volver al paso 1

Esta operación se realizará sucesivamente para cada $i = 1, \dots, n$ y dentro de cada i , con los b_{i1}, b_{i2}, b_{i3} alternativamente, con un número suficiente de iteraciones de muestreo de Gibbs para obtener una muestra final proveniente de $p(\mathbf{b}_i | \sigma, \beta, \mathbf{a}, \mathbf{y}_{(ij)})$.

Correcciones para acelerar el proceso de simulación

Las simulaciones de los vectores \mathbf{a} y \mathbf{b} por rechazo, pueden ser costosas en tiempo sobre todo en los casos en los que β es pequeño y tenemos un tamaño muestral n relativamente grande. El método se puede acelerar truncando superiormente la distribución de acotación según la Figura 3.18.

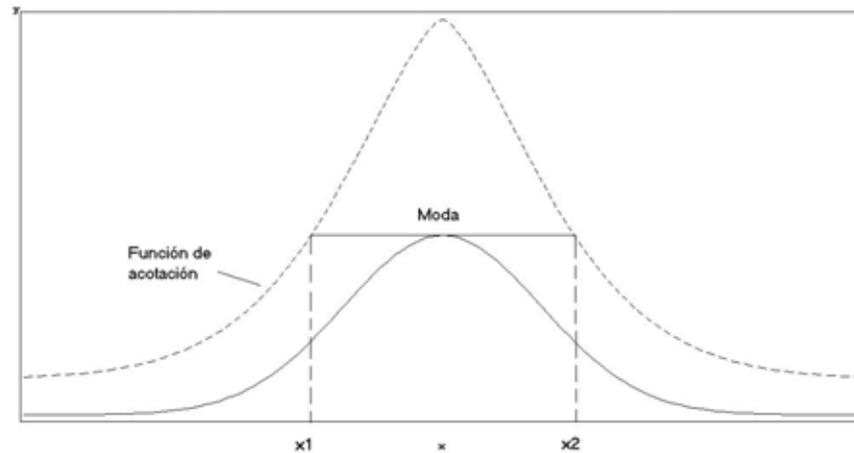


Figura 3.18. Truncamiento de la función de acotación.

El valor de la función en la moda de la función objetivo, \hat{m} , se puede estimar por métodos numéricos. Los puntos x_1 y x_2 que corresponden a las abscisas cuyo valor en la

distribución de acotación es \hat{m} se deducen analíticamente y son, por ejemplo, en el caso de cada componente a_j ,

$$\begin{aligned}x_1 &= \bar{z}^k + \sigma_{\max}\left(\frac{\hat{m}}{\sum s_i}\right)^{\frac{1+\beta}{2}} \\x_2 &= \bar{z}^k - \sigma_{\max}\left(\frac{\hat{m}}{\sum s_i}\right)^{\frac{1+\beta}{2}},\end{aligned}$$

Para simular de la nueva función de acotación y aplicar el método de rechazo se estiman las áreas de las colas y del rectángulo central y se sortea de qué parte se simulará, con probabilidades proporcionales a cada área. Seguidamente, para simular de las colas se utilizará el método planteado anteriormente para simulación de colas de una Potencial Exponencial, y para el rectángulo central según una uniforme. Finalmente se continúa con el método de rechazo como habitualmente.

Por último, existen situaciones extremas en las que el método de rechazo, debido a una variabilidad muy pequeña en la función objetivo, puede llevar demasiado tiempo. En estos casos se toma como valor simulado, con una ínfima pérdida de precisión, la moda a posteriori calculada numéricamente en el apartado anterior.

4.4.3 Resultados

Presentamos a continuación los resultados correspondientes al modelo $M = 1$.

Se han realizado 500 iteraciones del muestreo de Gibbs, descartando las 100 primeras. Se realizan 5 secuencias paralelas del muestreo de Gibbs. Se observa en los gráficos siguientes, que presentan cómo evolucionan los valores modales de a_1 , $b_{8,1}$ y σ , cómo el algoritmo parece converger antes de las 100 iteraciones, aunque la gran variabilidad de $b_{8,1}$ se aprecia en el gráfico.

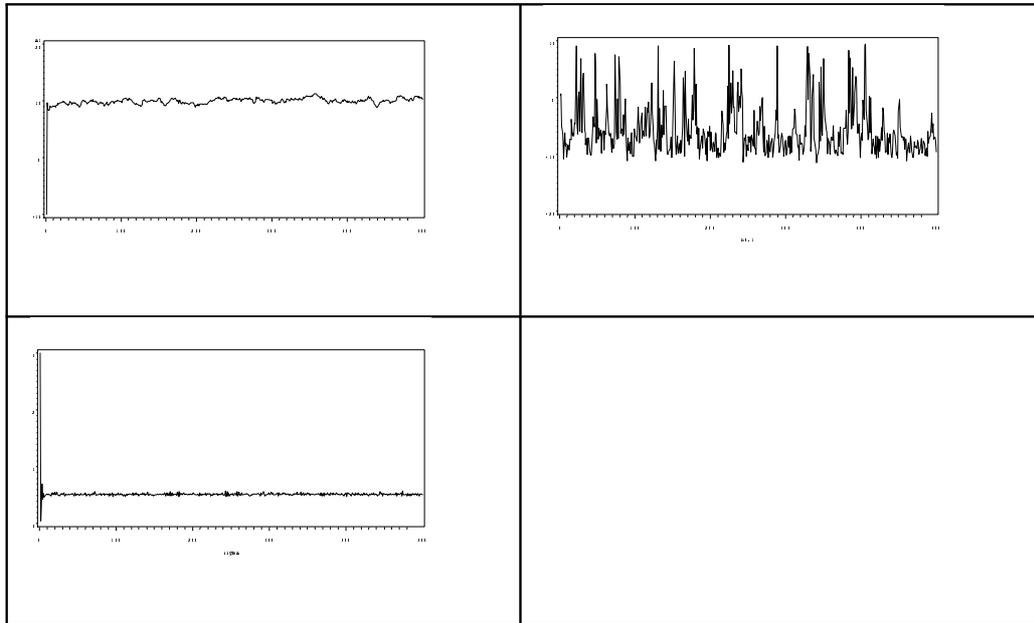


Figura 3.19. Convergencia de los valores modales de los parámetros en el algoritmo.

En la Figura 3.20 se muestran histogramas de las distribuciones a posteriori de los parámetros $a_1, a_7, b_{8,1}, b_{8,2}, \sigma$ y β . En la distribución a posteriori de β , con moda en $\beta = 1$, y $P(C' | \mathbf{y}) \simeq 1$, se aprecia cómo en este modelo parece más adecuado utilizar una distribución con mayor apuntamiento que la distribución Normal para los errores.

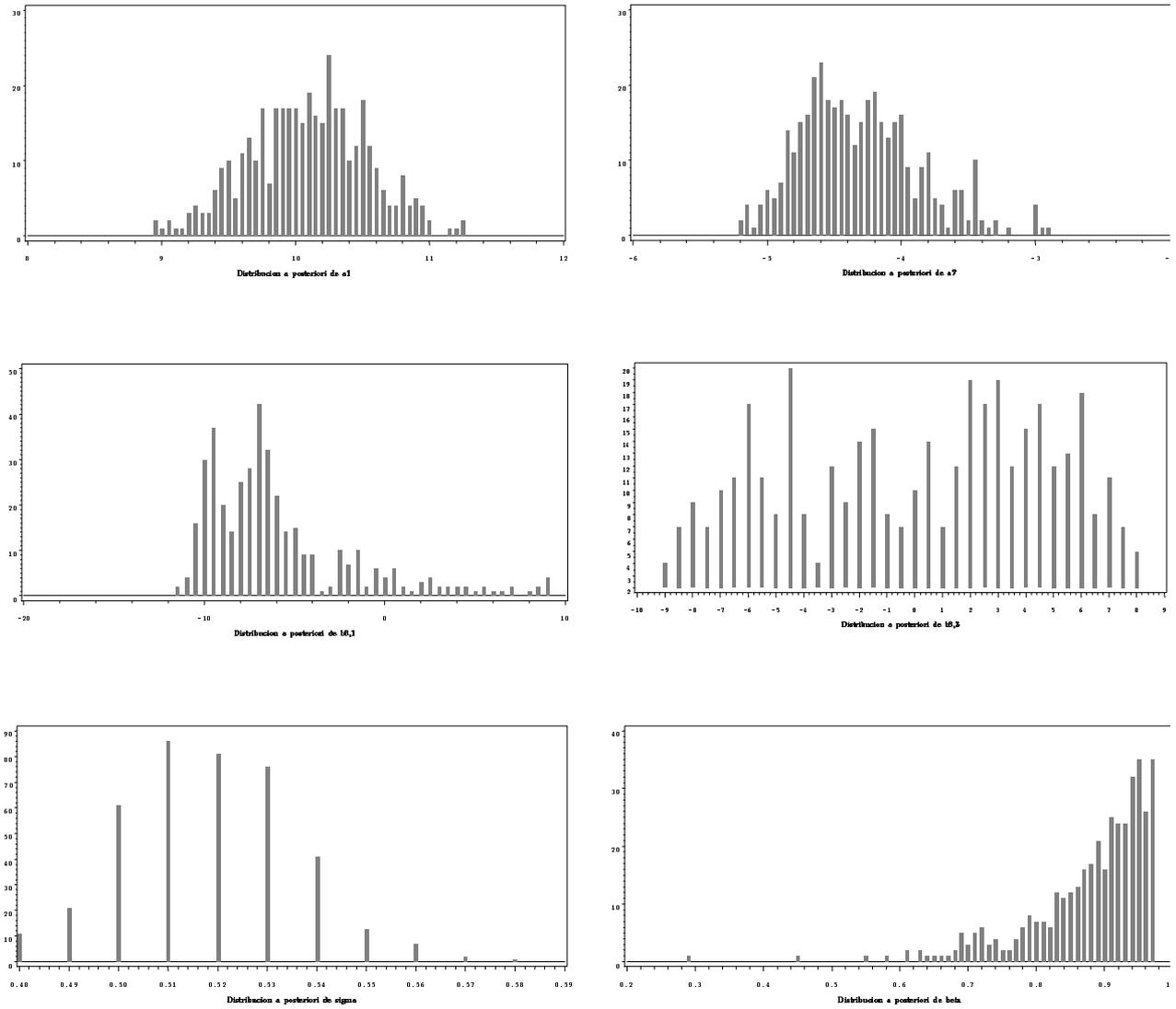


Figura 3.20. Distribuciones a posteriori de $a_1, a_7, b_{8,1}, b_{8,2}, \sigma$ y β .

En la siguiente tabla se presentan las modas a posteriori de los parámetros $a_1, \dots, a_9, b_{8,1}, b_{8,2}, b_{8,3}, \sigma$ y β . A título comparativo, se añaden los resultados obtenidos por (Carlin y Louis, 2000) con el modelo Normal (equivalente a considerar $\beta = 0$ constante).

Modelo con β aleatorio		Modelo con $\beta=0$ constante
Parámetro	Moda	Moda
a_1	10.1	9.93
a_2	-1.25	-0.04
a_3	-0.75	-0.16
a_4	0.15	0.004
a_5	1.21	0.309
a_6	-0.22	-0.34
a_7	-4.31	-4.29
a_8	-0.60	-0.32
a_9	0.50	0.35
$b_{8,1}$	-5.86	-7.5
$b_{8,2}$	-0.04	-0.5
$b_{8,3}$	-0.08	0
σ	0.51	1.83
β	1	

Los resultados se asemejan, tanto en lo que afecta a los parámetros correspondientes a efectos fijos, a_1, \dots, a_9 como a los de efectos aleatorios $b_{8,1}, b_{8,2}, b_{8,3}$, a pesar de la diferencia del enfoque propuesto y del diferente desarrollo computacional. Coherentemente con los modelos presentados anteriormente, el parámetro más afectado por el hecho de introducir β como variable aleatoria es la desviación típica σ .

Para comparar los modelos $M = 1$ y $M = 2$, en las versiones $\beta = 0$ (Modelo con errores normales) y β aleatorio (modelo con errores Potencial Exponencial) se utiliza el método de estudio de ajuste calculando $d_{ij} = y_{ij} - \hat{E}(Y_{ij} | \mathbf{y}_{(ij)})$ para los y_{ij} cuyo valor es conocido. La tabla siguiente presenta los valores predichos para las primeras observaciones con cada modelo y finalmente la medida de ajuste calculada $\sum \sum d_{ij}^2$ para cada uno de los cuatro planteamientos.

Modelo $M = 1$, β aleatorio			Modelo $M = 1$, $\beta=0$ cte	
y_{ij}	$\hat{E}(Y_{ij} \mathbf{y}_{(ij)})$	$ d_{ij} $	$\hat{E}(Y_{ij} \mathbf{y}_{(ij)})$	$ d_{ij} $
10.67	9.95	0.72	9.66	1.01
8.42	6.23	2.19	5.88	2.54
9.43	9.82	0.40	12.62	3.19
6.32	4.30	2.02	2.89	3.43
8.12	10.40	2.28	13.98	5.86
5	2.3	2.7	1.66	3.34
3.46	9.22	5.76	8.03	4.57
3.60	8.51	4.91	11.60	8
6.16	1.80	4.36	2.97	3.19
...
$\sum d_{ij}^2$		98.33		169.96

Modelo $M = 2$, β aleatorio		Modelo $M = 2$, $\beta = 0$ cte	
$\hat{E}(Y_{ij} \mathbf{y}_{(ij)})$	$ d_{ij} $	$\hat{E}(Y_{ij} \mathbf{y}_{(ij)})$	$ d_{ij} $
7.55	3.12	7.36	3.31
6.98	1.43	10.79	2.37
9.23	0.20	7.21	2.22
6.32	0.00	8.79	2.47
7.28	0.84	9.10	0.98
2.26	2.73	3.2	1.8
3.21	0.24	5.1	1.64
4.41	0.80	4.8	1.2
2.25	3.90	2.20	3.96
...
		35.88	51.61

Se observa que en general, de los modelos utilizados, el modelo $M = 2$, menos complicado, da un ajuste mucho mejor que el modelo $M = 1$. Estos resultados son coherentes con los de (Carlin, Louis, 2000) y (Han, Carlin, 2001). Dentro de cada uno de los modelos la utilización de β como parámetro aleatorio da lugar a una mayor precisión en el ajuste.

5 Conclusiones

En el primer Capítulo se han presentado técnicas para el tratamiento de la distribución Potencial Exponencial en el marco de la Inferencia Bayesiana. Los resultados en cuanto a las distribuciones a posteriori coinciden con los obtenidos por otros autores sobre los mismos datos, como el ejemplo de datos de Darwin presente en (Box, Tiao, 1973) y (Walker y Gutierrez-Peña (1999)), con la ventaja que presenta nuestra aproximación de incorporar un método alternativo a la integración numérica y técnicas de aproximación para la distribución a posteriori de β , tratadas escasamente por otros autores.

Las tres técnicas presentadas para el tratamiento de las distribuciones a posteriori presentan ventajas alternativas. Utilizando la aproximación vía Monte-Carlo se obtienen valores puntuales proporcionales al valor funcional de las distribuciones a posteriori. En algunos casos es la manera más sencilla de tratar los problemas. Se ha utilizado este método, por ejemplo, en el cálculo de valores de medidas de discrepancia y en la aplicación a modelos no lineales, donde la existencia de una función no lineal impide un desarrollo directo de las técnicas de simulación de las distribuciones a posteriori. El método Monte-Carlo empleado se ha revelado como bastante rápido y eficiente en las aplicaciones.

La utilización del método de muestreo de Gibbs para extraer muestras de las distribuciones a posteriori de la distribución Potencial Exponencial, permite una gran flexibilidad en cuanto al tratamiento de muchos modelos, pues a menudo existen estructuras jerárquicas que permiten la utilización de este método en el seno de simulaciones con un número mayor de parámetros. Esto ha quedado patente por ejemplo, en el tratamiento de la aplicación del modelo de datos longitudinales del Capítulo 3. La utilización de la representación de la distribución Potencial Exponencial a través de mixturas para la extracción de muestras de las distribuciones a posteriori tiene el inconveniente, frente al método anterior, de incorporar la extracción de muestras de la variable de mixtura u , lo que redundaría en mayor tiempo de proceso. . Esto en algunos casos puede ser una ventaja, pues para cada observación tenemos una muestra de los parámetros de mixtura u , que nos permitirá indagar hasta qué punto puede ser considerada un valor discordante.

La aplicación de estos dos últimos métodos requiere utilizar los algoritmos de rechazo presentados. Estos funcionan eficientemente para muestras moderadas, pero pueden ser lentos (rechazar demasiado a menudo) en casos de gran apuntamiento o de muestras grandes. En algunos casos, cuando el porcentaje de puntos rechazado es excesivamente alto, se recurre a técnicas más sencillas como integración numérica seguida de simulación trapezoidal, o a métodos de truncamiento de la función de rechazo como la presentada en el Capítulo 3. Los métodos basados en el muestreo de Gibbs parecen más lentos que la utilización del método Monte-Carlo, pero estos últimos no serían tan fácilmente aplicables en modelos más complejos o de índole jerárquica, con lo cual ambas técnicas son complementarias.

Se han presentado algunas alternativas para tratar el contraste de normalidad en el seno de la familia Potencial Exponencial; este contraste es equivalente a un contraste de hipótesis nula puntual. Hemos escogido como referencia la medida basada en la región de máxima densidad a posteriori, pues está automáticamente calibrada en el intervalo $(0,1)$ y responde a un simple cálculo de probabilidad a posteriori. Este cálculo se puede llevar a cabo en las aplicaciones a través del método Monte-Carlo, o bien mediante las técnicas de muestreo de Gibbs, aunque en este último caso se requiere un número alto de muestras y por consiguiente mayor tiempo de cálculo, para tener una aproximación igual de precisa que en el método de Monte-Carlo.

La sensibilidad del procedimiento de contraste de normalidad está supeditada a la procedencia de la muestra de una familia, si no Potencial Exponencial, al menos unimodal y simétrica. En caso de no ser Potencial Exponencial pero sí cumplir estas últimas condiciones (como es el caso de las simulaciones de variables T de Student) el procedimiento parece funcionar correctamente, pues la distribución Potencial Exponencial sí parece una familia capaz de aproximar gran parte de este tipo de distribuciones. Se ha visto también que cuando la muestra original proviene de una Potencial Exponencial el procedimiento de contraste funciona adecuadamente. Por otra parte, si el objetivo es realizar contrastes bayesianos de bondad de ajuste para cualquier distribución a partir de la transformación de la función de distribución, es deseable utilizar familias de distribuciones como las presentadas por otros autores, que al tener limitado su dominio en $(0,1)$ se muestran más sensibles en el contraste de uniformidad.

Se ha utilizado la distribución Potencial Exponencial como base para aplicaciones en modelos bayesianos. Con los métodos propuestos, su utilización no es excesivamente complicada, con la ventaja de representar una generalización de la distribución Normal útil en muchos casos. Se compara el modelo normal con el modelo Potencial Exponencial en todas las situaciones, y en algunas se comparan también dos modelos alternativos entre sí, situándonos así en el marco de comparación de modelos. Se han utilizado como medidas de comparación la medida de discrepancia HPD, y una medida de eficacia predictiva útil dentro de la perspectiva práctica de las aplicaciones tratadas.

En la mayor parte de las aplicaciones, se observa cómo las técnicas presentadas se aplican con cierta rapidez y las medidas de comparación permiten la decisión de emplear el modelo Normal usual, como en la primera aplicación, o bien de suponer los errores como distribuidos como una Potencial Exponencial, con beneficios en lo que corresponde a la capacidad predictiva y a la adecuación del modelo a los datos.

6 Bibliografía

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics, Berlin: Springer Verlag.

Bernardo J.M.and Smith, A.F.(1993). *Bayesian Theory*. New York: Wiley .

Bernardo, J.M. (1982). Contraste de Modelos probabilísticos desde una perspectiva Bayesiana. *Trabajos de Estadística e Investigación Operativa* **33**, 2, 16-30.

Box G.E.P. and Tiao, G. C. (1979) *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

Carota, C. Parmigiani, G. and Polson, N. (1996) . Diagnosing Measures for Model Criticism. *Journal of the American Statistical Association* **91**, 753-762.

Casella, G. and Berger, R.L. (2002). *Statistical Inference*. Spain: Duxbury Press.

Carlin, B.P. and Louis, T.A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall.

Chen, M., Shao, Q. and Ibrahim, J. (1999) *Monte Carlo Methods in Bayesian Computation*. New Yor: Springer Verlag.

Choy, S.T.B.(1999) Discussion-Robustifying Bayesian procedures. *Bayesian Statistics* **6**, 685-710. Oxford: University Press.

Devroye, L. (1984) *Non Uniform Random Variable Generation*. New York: Springer-Verlag.

Fernández, C., Osiewalski, J. and Steel, M. F. J.(1995). Modeling and Inference with v-Spherical Distributions. *Journal of the American Statistical Association*, **90**, 432, Theory and Methods.

Geman, S. and Geman, D.(1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721- 741

Gelfand, A.E. and Ghosh, S.K. (1998) Model Choice: a minimum posterior predictive loss approach.. *Biometrika* **85**, 1-11.

Gelfand, A. E., Dey, D.K.,and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics 4*. Oxford: University Press.

- Gelman, A., Carlin, J. B., Stern, H., Rubin, D.B.(2003) . *Bayesian Data Analysis* . New York: Chapman and Hall.
- Geweke, J. (1989) . Bayesian inference in Econometric problems using Monte-Carlo Integration. *Econometrica*, **57**, 1317-1339.
- Girón, F. y Ortiz, C. R. (1985). Una alternativa Bayesiana a los contrastes de Bondad de Ajuste. *Trabajos de Estadística e Investigación Operativa*, **36**, 2, 56-69.
- Gómez-Villegas, M. A. (1988). Ajuste Bayesiano a una distribución mediante la Familia Exponencial Potencial. *Estadística Española*, **30**, 118, 203-213.
- Gómez, E. Gómez-Villegas, M.A. and Marín, J. M. (1998) A multivariate generalization of the exponential family of distributions. *Communications in Statistics. Theory and Methods*, **27**, 3, 589-600.
- Gómez, E. Gómez-Villegas, M.A. and Marín, J. M. (2002). Continuous elliptical and exponential power linear dynamic models. *Journal of Multivariate Analysis* **83**, 22-36.
- Gómez, E. Gómez-Villegas, M.A. and Marín, J. M. (2002). A matrix variate generalization of the power-exponential family. *Communications in Statistics-Theory and Methods* **31**, 2, 201-217.
- Gómez-Villegas, M. A. and Gómez, E. (1992) . Bayes Factor in Testing precise hypotheses *Communications in Statistics. Theory and Methods*, **21**, 6, 1707-1715.
- Goutis, C, y Robert, C.P. (1998) Model choice in generalised linear models: A Bayesian approach via Kullback-Leibler projections. *Biometrika* **85**, 29-37.
- Han, C.y Carlin, B.P. (2001) Markov Chain Monte Carlo methods for computing bayes factors: a comparative review. *Journal of the American Statistical Association*, **96**, 1122-1132.
- Lee, P. M. (1994) . *Bayesian Statistics: an Introduction* . New York: Wiley and Sons.
- Marín, J.M. (1998) *Extensión Multivariante de la distribución Potencial Exponencial y su aplicación a los modelos lineales dinámicos*. Tesis Doctoral . Facultad de Matemáticas. Universidad Complutense de Madrid.
- Natarajan , R.and Kass, R. (2000) Reference Bayesian methods for generalised linear mixed models. *Journal of the American Statistical Association*, **95**, 227-237
- O'Hagan, A. (1979). On Outlier Rejection Phenomena in Bayes Inference. *Journal of the Royal Statistical Society, Series B*, **41**, 3, 358-367.

Portela, J. and Gómez-Villegas (2003). Implementation of a Robust Bayesian Method. *Journal of Statistical Computing and Simulation*. (Aceptado para su publicación en 2003).

Raftery, A. E. (1995) . *Hypothesis testing and model selection via posterior simulation. (Practical Markov Chain Monte Carlo)*. New York: Chapman and Hall.

Rahman, M. and Gokhale, D.V.(1996). On Estimation of Parameters of the Exponential Power Family of Distributions. *Communications in Statistics. Simulation and Computation*, **25**(2), 291-299.

Robert, C. P. (1994) . *The Bayesian Choice*. New York: Springer-Verlag.

Rubinstein, R. Y. (1981) . *Simulation and the Monte-Carlo Method*. New York: John Wiley.

SAS Institute.(1995) *SAS Language* .

SAS Institute.(1995) *SAS Guide to Macro Processing*.

Tanner, M. A. (1996). *Tools for Statistical Inference*. New York: Springer-Verlag.

Vounatsou, P. Smith, A.F.M.y Choy, S.T.B.(1998) Bayesian robustness for location and scale parameters using simulation. *Imperial College Technical Report Series TR 96-21*.

Walker, S.G.and Gutierrez-Peña, E. (1999). Robustifying Bayesian procedures. *Bayesian Statistics 6*, 685-710. Oxford: University. Press.

West, M. (1987) On scale mixtures of normal distributions. *Biometrika* **74**, 646-648.

Wright, D.E. (1985) A note on the Construction of Highest Posterior Density Intervals. *Applied Statistics*, **35**, 1, 49-53.

Zeger, S. L.and Karim, M. R. (1991) *Journal of the American Statistical Association*, **86**, 79-86