Instituto Tecnológico Autónomo de México

Inferencia Bayesiana en Modelos de Mezclas Gaussianas en Media-Varianza

Tesis

que para obtener el título

Licenciado en Actuaría

presenta

David Edgardo Castillo Rodríguez

Asesor: Juan Carlos Martínez Ovando

Ciudad de México 2017

Índice general

1. Introducción al paradigma bayesiano de inferencia **9**

1.1. Antecedentes 9

1.2. Proceso de aprendizaje 11

1.3. El proceso de aprendizaje en la inferencia estadística 12

1.4. Predicción 13

1.5. Estimadores a través de la teoría de decisión 15

1.5.1. Función de pérdida cuadrática 17

1.5.2. Función de pérdida valor absoluto 18

1.5.3. Función de pérdida $0-1$ 19

2. Distribuciones de probabilidad relevantes **21**

2.1. Distribución gaussiana 22

2.2. Distribución gaussiana p-variada 25

2.3. Distribución gamma 30

2.4. Distribución gamma inversa 33

2.5. Distribución gaussiana inversa 36

2.6. Distribución gaussiana inversa generalizada 38

2.7. Distribución Wishart 42

2.8. Distribución Wishart inversa 43

3. Mezclas de distribuciones gaussianas **45**
3.1. Distribuciones tipo mezcla 45
3.2. Mezcla gaussiana en media 47
3.3. Mezcla gaussiana en varianza 52
3.4. Mezcla gaussiana en media-varianza 56
3.5. Distribución hiperbólica generalizada 60

4. Inferencia bayesiana en distribuciones tipo mezcla **65**
4.1. Verosimilitud y verosimilitud extendida 65
4.2. Verosimilitud extendida 66
4.3. Muestreador de Gibbs 69
4.4. Algoritmo para la distribución normal gaussiana inversa . . 71
4.5. Distribuciones condicionales completas 74
4.5.1. Condicional completa para μ 74
4.5.2. Condicional completa para β 76
4.5.3. Condicional completa para Σ 77
4.5.4. Condicional completa para λ 78
4.5.5. Condicional completa para Ψ 79
4.5.6. Condicional completa para u_i 80
4.6. Alternativa para la matriz de varianza-covarianza 81
4.6.1. Condicional completa para s_i 82
4.6.2. Condicional completa para r_{ij} 83

5. Ejercicio de simulación **85**
5.1. Modelo 1 86
5.2. Modelo 2 89
5.3. Resultados y consideraciones 90

6. Conclusiones **97**

A. Estimación por máxima verosimilitud y algoritmo EM **101**
A.1. Estimación por máxima verosimilitud 101
A.2. Algoritmo EM 102

B. Distribuciones tipo mezcla: Resultados adicionales **105**
B.1. Simulación de distribuciones tipo mezcla gaussiana en esperanza varianza 105
B.2. Distribución tipo mezla gaussiana p-variada 106
B.3. Covarianza de un vector aleatorio p-variado 106
B.4. Covarianza de una distribución tipo mezcla 106
B.5. Kernel de una distribución gaussiana p variada 108
B.6. Kernel de una distribución Wishart 108
B.7. Kernel del producto de n distribuciones gaussianas p variadas con mismo vector de medias y misma matriz de varianza covarianza 109
B.8. Kernel del vector de medias de una distribución gaussiana p variada multiplicada por la distribución del vector de medias 110
B.9. Probabilidad condicional 111

C. Integración Monte Carlo y Slice Sampler **113**
C.1. Integración Monte Carlo 113
C.2. Slice Sampler 115

D. Descomposición $diag(S)Rdiag(S)$ **119**

E. Gráficos de convergencia y estabilización de varianza **123**

Bibliografía **144**

Introducción

Con la finalidad de modelar algún fenómeno de interés se han propuesto modelos estocásticos que requieren del uso de alguna distribución de probabilidad. En tales modelos muchas veces se utiliza la distribución de probabilidad normal (o normal multivariada), como sucede en la selección de un portafolio de mínima varianza (Markowitz, 1952), por ejemplo. En muchas ocasiones, la selección de esta distribución podría no ser conveniente, puesto que no captura adecuadamente la información concerniente a los valores extremos y a una alta curtosis, que es muy común observar en las distribuciones empíricas de datos de retornos financieros de algún activo (Peiró, 1998).

Dicho lo anterior, la presente tesis fue motivada por la siguiente pregunta: ¿qué hacer cuando hay evidencia de que un conjunto de datos no sigue una distribución de probabilidad normal multivariada? Varios autores han dado respuesta a esta pregunta, ya sea mediante la distribución T, o la incorporación de un sesgo a la distribución normal (o a la distribución T), o mediante el uso de mezclas normales (Javier Mencía, 2009), o mediante el uso de la distribución hiperbólica generalizada (Barndorff-Nielsen, 1977), por ejemplo. El tema central de esta tesis es cómo estimar una distribución normal gaussiana inversa (caso particular de la distribución hiperbólica generalizada).

El problema de cómo estimar una distribución normal gaussian inversa se realizará mediante distribuciones tipo mezcla gaussiana en esperanza-varianza, ya que de esta forma se podrá trabajar con algunos resultados de la distribución normal y, además, dada la naturaleza de las distribuciones tipo mezcla, se puede incorporar naturalmente el enfoque bayesiano de inferencia, y así aprovechar la información inherente a los datos. Al final se obtendrá una muestra de la distribución de cada parámetro de interés de la mezcla ajustada, con la cual se obtendrán los estimadores deseados con la ayuda de la función de pérdida cuadrática, por ejemplo.

La presente tesis se divide en seis capítulos. En el capítulo uno se introduce de manera breve el paradigma bayesiano de inferencia, el cual es necesario para el tipo de estimación que se realizará. En el capítulo dos se expone la notación a usar, así como las distribuciones necesarias para la estimación; también se habla brevemente de algunas características de las distribuciones mencionadas. En el capítulo tres se introducen las distribuciones tipo mezcla gaussiana y se mencionan brevemente algunas de sus propiedades, para así poder caracterizar a la distribución hiperbólica. En el capítulo cuatro se construyen la función de verosimilitud y las densidades marginales completas necesarias para la estimación. En el capítulo cinco se presenta un caso de simulación, así como los resultados y conclusiones obtenidas. Y por último, en el capítulo seis se habla de las conclusiones generales. Es importante mencionar que para la obtención de resultados se utilizó el software estadístico *R*. En el CD adjunto a esta tesis se encuentran los códigos utilizados.

Capítulo 1

Introducción al paradigma bayesiano de inferencia

En este capítulo se hablará brevemente sobre los antecedentes del paradigma bayesiano de inferencia, así como de algunas de las ideas principales que dieron origen a éste. Se concluye cómo se realiza la estimación e inferencia bayesiana en general, la cual es fundamental en este trabajo.

1.1. Antecedentes

En 1763, dos años después de la muerte de Thomas Bayes ($1702-1761$), se publicó uno de sus ensayos (Bayes (1763), *An Essay towards solving a Problem in the Doctrine of Chances*), en el cual resolvía un problema de información inversa planteado por Jacob Bernoulli . El problema de Bernoulli consistía en obtener información sobre la realización de experimentos de variables aleatorias independientes distribuidas Bernoulli; para ello, Bayes propuso un método que consistió en suponer la posibilidad de que un evento que tenga que ocurrir, algunas veces ocurra con éxito y otras veces no ocurra (o sea, un fracaso).

En este punto de la historia, las ideas de Bayes carecían de claridad; diez años más tarde se esclarecieron debido a que Laplace las retomó, dándole así forma al paradigma bayesiano de inferencia. Dichas ideas se consolidaron en su libro *Theoriè analytique des probabilitès* (Laplace (1812)). Pero el trabajo de Laplace (concerniente al paradigma Bayesiano) aún carecía de rigor teórico.

El paradigma bayesiano de inferencia continuó eclipsado hasta que Jeffreys (1939) (*Theory of Probability*), y Finetti (1937) (*La Prévision, ses Lois Logiques, ses Sources Subjectives*) desarrollaron y dieron sustento teórico al paradigma. Harold Jeffreys mantuvo una postura objetiva dentro del paradigma bayesiano, ya que él decía que la información inicial era capturada objetivamente mediante el uso de distribuciones no informativas. Por otro lado, de Finetti mantuvo una postura subjetiva, y en su obra se desarrolla la teoría de la probabilidad refiriéndose a esta como un grado de creencia.

Hasta la fecha, el paradigma bayesiano sigue adoptándose en distintos campos de la ciencia, ya que es una gran herramienta para la inferencia y para la predicción, lo cual es de vital interés en ciertos ámbitos del conocimiento, como en finanzas, econometría, actuaría. En Savage, 1972 o Lindley, 1957 se puede encontrar una explicación más profunda del paradigma bayesiano de inferencia.

Teniendo un contexto sobre la evolución del paradigma bayesiano, veamos cuáles son sus principios técnicos y la idea detrás de estos.

1.2. Proceso de aprendizaje

Supongamos que tenemos un conjunto de eventos A_1, A_2, . . ., A_n mutuamente excluyentes, sobre los cuales nos gustaría hacer inferencia de acuerdo a una función de probabilidad P (referente a estos eventos), y con un nivel de información o evidencia disponible, el cual denotaremos por B. Con estos elementos podemos definir nuestro estado de información actual sobre los eventos A_1, A_2, . . ., A_n, dada nuestra información inicial B; esto queda expresado, para cualquier conjunto A_i, como $P(A_i|B)$, donde $\sum_{i=1}^{n} P(A_i|B) = 1$.

Lo que nos interesa es actualizar nuestro nivel de información actual dada nueva información, digamos C, lo cual lo conseguimos de la siguiente manera utilizando la definición de probabilidad condicional B.9,

$$\begin{aligned} P(A_i|C \cap B) &= \frac{P(A_i \cap C \cap B)}{P(C \cap B)} \\ &= \frac{P(A_i \cap B)}{P(B)} \frac{P(B)}{P(C \cap B)} \frac{P(C \cap A_i \cap B)}{P(A_i \cap B)} \\ &= P(A_i|B)P(C|A_i \cap B)/P(C|B). \end{aligned} \tag{1.1}$$

La ecuación anterior representa la idea del proceso de aprendizaje, es decir, dado el nivel de evidencia inicial denotado por B, mediante la incorporación de nueva información C, podemos actualizar nuestras creencias que ya teníamos con respecto al evento A_i; es decir, pasar de tener $P(A_i|B)$ a $P(A_i|B \cap C)$.

1.3. El proceso de aprendizaje en la inferencia estadística

Aplicando las ideas de la sección anterior, podemos incorporar nueva información en una distribución de probabilidad parametrizada, cuyo parámetro es desconocido para nosotros. Este falta de información acerca del verdadero valor del parámetro la incorporamos en una distribución de probabilidad propia para el parámetro desconocido. Nos referimos a la distribución de probabilidad del parámetro como distribución inicial o ditribución a priori, y la denotamos con $\Pi(\theta)$, donde θ es el parámetro desconocido perteneciente a un espacio parametral Θ.

Dicho lo anterior, dada una variable aleatoria X con distribución de densidad de probabilidad parametrizada por algún parámetro θ, tenemos la distribución $f_X(x|\theta)$, que se lee como distribución de X dado θ. Después, incorporando nuestro desconocimiento de θ, con $\Pi(\theta)$ podemos formular la siguiente expresión:

$$\begin{aligned} \Pi(\theta|X) &= \frac{f(x \cap \theta)}{f(x)} \\ &= \frac{f(x \cap \theta)}{f(x)} \frac{\Pi(\theta)}{\Pi(\theta)} \\ &= f(x|\theta) \frac{\Pi(\theta)}{f(x)}. \end{aligned} \tag{1.2}$$

De la ecuación (1.2) podemos deducir que la distribución de $\Pi(\Theta|X)$ es proporcional a $f(x|\theta)\Pi(\theta)$, lo cual expresamos mediante $\Pi(\theta|X) \propto f(x|\theta)\Pi(\theta)$. La parte derecha de la expresión anterior se conoce como distribución a posteriori de θ. Esta se puede interpretar como la actualización del nivel de información que teniamos de θ (dicho nivel de información está considerado en $\Pi(\theta)$), al incluir nueva información proveniente de la realización (o

realizaciones) de la variable aleatoria X.

La información contenida en la variable aleatoria X queda capturada por su función de densidad condicional dado θ, es decir, $f(x|\theta)$; también a esta densiadad condicionada la conocemos como función de verosimilitud en función de Θ. El símbolo $\propto$ nos indica que la expresión de la izquierda es proporcional a la de la derecha, es decir, sólo difieren por una constante que no depende del argumento de la parte izquierda.

La expresión (1.2) también nos da un algoritmo recursivo de aprendizaje, el cual nos permite mejorar nuestra información del parámetro θ conforme más información obtengamos de la variable aleatoria X, es decir, si consideramos una nueva ralización de $X = x^*$, entonces nuestras nueva distribución a priori o inicial es $\Pi(\theta|x)$, mientras que la distribució a posteriori (o actualizada) queda dada como $\Pi(\theta|x \cap x^*)$, por lo que mediante un proceso análogo a (1.2) tendriamos la siguiente expresión:

$$\Pi(\theta|x \cap x^*) \propto f(x \cap x^*|\theta)\Pi(\theta|x \cap x^*). \tag{1.3}$$

1.4. Predicción

Hasta ahora hemos visto cómo incorporar nueva información de la variable aleatoria X para actualizar la información disponible del parámetro θ. Pero, también podemos formularnos las siguientes preguntas ¿Qué pasa con el próximo valor de X dados los valores x_i que ya fueron observados? ¿Hay forma de que la información previa de X sea incorporada para mejorar el nivel de información de futuras realizaciones? La respuesta a las preguntas anteriores tiene connotación positiva, y queda representada por

la siguiente expresión:

$$\begin{aligned} f(x*|\overline{x}) &= \int_{S_\theta} f(x*\cap\theta|\overline{x})d\theta \\ &= \int_{S_\theta} f(x*|\theta\cap\overline{x})\Pi(\theta)d\theta, \end{aligned} \tag{1.4}$$

donde $x*$ representa la realización futura de X, $\overline{x}$ representa un conjunto de realizaciones observadas de X, y θ el parámetro de la distribución de probabilidad de X.

También es importante notar que podemos actualizar nuestros niveles de información (capturados por la distribución inicial $\Pi(\theta)$), del parámetro θ, a niveles de información posteriores de haber observado los valores $\overline{x}$. Lo anterior se resume en la siguiente ecuación:

$$\begin{aligned} f_{\Theta|\overline{x}}(\theta) &= \int_{S_x} f(\theta\cap\overline{x})dx \\ &= \int_{S_x} f(\overline{x}|\theta)\Pi(\theta)dx. \end{aligned} \tag{1.5}$$

A la ecuación (1.5) se le conoce como distribución predictiva a priori del parámetro θ; en esta vemos actualizadas nuestras creencias iniciales de θ, una vez ocurridas las realizaciones de la variable aleatoria X.

Es de gran importancia notar que el paradigma bayesiano nos permite obtener la distribución completa de los parámetros Θ de interés, con lo cual es natural preguntarse acerca de qué valor debe tomar θ, si es posible obtener un estimador de este dada la función a posteriori. La respuesta a la pregunta anterior es afirmativa, y corresponde a un problema de decisión en un ambiente de incertidumbre.

1.5. Estimadores a través de la teoría de decisión

De manera muy resumida, la teoría de decisión trata acerca de qué elección tomar, dado un conjunto de posibilidades; esta teoría ha sido adaptada a una gran cantidad de disciplinas, como en economía, ciencia política, estadística, por ejemplo. Para los fines de este trabajo bastará con hablar brevemente en cómo se adecua esta teoría en el paradigma bayesiano, para así obtener los estimadores de interés.

De manera intuitiva un problema de teoría de decisión se plantea de la siguiente manera:

Supongamos que contamos con dos espacios de estados Ω y $\mathcal{A}$. Donde Ω representa el espacio de todos los posibles sucesos que pudieran ocurrir en algún experimento de interés. Mientras que el espacio $\mathcal{A}$ contiene todas nuestras posibles acciones o elecciones dado a que se presente algún estado de Ω.

Entonces podemos formar trayectorias de decisión (α, ω), con ω en Ω, y α en $\mathcal{A}$. Donde cada dupla representa la elección α tomada dado el valor ω observado. Las acciones tomadas llevarán a resultados distintos, de los cuales algunos son prefereibles que otros, por lo cual es necesario contar con alguna medida que nos permita identificar que elecciones α tomar tal que nuestras acciones sean coherentes y consistentes. La forma en la que podemos medir qué acciones son mejores que otras puede ser mediante una función de utilidad, la cual se define como $u : \mathcal{A} \times \Omega \to R_+$.

Adaptando lo anterior al enfoque bayesiano, podemos intentar medir nuestros niveles de credibilidad del estado ω con una función de probabilidad $P(\cdot|B)$ donde B representa la información relevante disponible al

momento en que pudiera ocurrir ω, tal que $\Omega = \mathcal{A}$. Dicho lo anterior, el problema Bayesiano consiste en encontrar la acción α óptima que maximice la utilidad esperada Bernardo and Smith (2008). Lo cual se resume en resolver la siguiente ecuación:

$$\Theta^* = argmax E_{\Theta|x}[u(\Theta, \Theta^*|x)] \tag{1.6}$$

Teniendo en cuenta que una función de utilidad puede verse como una función de pérdida, y que un problema de maximización sobre una función $f(.)$ se puede definir como un problema de minimización sobre la función $-f(.)$, tenemos que la ecuación (1.6) también puede plantearse como:

$$\Theta^* = argmin E_{\Theta|x}[l(\Theta, \Theta^*|x)], \tag{1.7}$$

donde $l = -u$ es una función de pérdida.

En resumen, con el enfoque bayesiano, podemos encontrar estimadores puntuales de Θ, siempre y cuando contemos con la distribución a posteriori de Θ, y una función de pérdida. De aquí que sea de esperar que cada función de pérdida produzca un estimador diferente, por lo que el uso de la función de pérdida dependerá del contexto del problema.

A continuación se hablará brevemente de tres de las funciones de pérdida más utilizadas en la literatura, y del estimador que produce cada una de ellas, para un análisis mucho más detallado de las funciones de pérdida es recomendable consultar Bernardo and Smith (2008).

1.5.1. Función de pérdida cuadrática

La función de pérdida cuadrática se define como la diferencia del cuadrado de Θ, y su estimador $\widehat{\Theta}$, o sea $(\Theta - \widehat{\Theta})^2$. Esta función de pérdida produce como estimador de Θ a $\widehat{\Theta} = E[\Theta|X]$, lo cual se puede verificar de la siguiente forma usando la ecuación (1.7):

$$\begin{aligned} E[(\Theta - \widehat{\Theta})^2] &= \int_{S_\Theta} (\Theta - \widehat{\Theta})^2 f(\Theta|x) d\Theta \\ &= \int_{S_\Theta} (\Theta^2 - 2\Theta\widehat{\Theta} + \widehat{\Theta}^2) f(\Theta|x) d\Theta \\ &= \widehat{\Theta}^2 \int_{S_\Theta} f(\Theta|x) d\Theta - 2\Theta \int_{S_\Theta} \widehat{\Theta} f(\Theta|x) d\Theta + \int_{S_\Theta} \widehat{\Theta}^2 f(\Theta|x) d\Theta \\ &= \widehat{\Theta}^2 - \widehat{\Theta} E[\Theta|x] + E[\Theta|x], \end{aligned} \tag{1.8}$$

luego con el críterio de primera derivada para encontrar valores óptimos tenemos que:

$$\frac{d}{d\widehat{\Theta}} E[(\Theta - \widehat{\Theta})^2] = 2\widehat{\Theta} - 2E[\Theta|x] \tag{1.9}$$

Igualando a cero tenemos que:

$$2\widehat{\Theta} - 2E[\Theta|x] = 0 \tag{1.10}$$

Luego, despejando a $\widehat{\Theta}$, tenemos que:

$$\widehat{\Theta} = E[\Theta|x] \tag{1.11}$$

Que efectivamente es un valor mínimo ya que $\frac{d^2}{d\widehat{\Theta}} E[(\Theta - \widehat{\Theta})^2] > 0$.

Por lo tanto:

$$MinE_{\Theta|x}[(\Theta - \widehat{\Theta}|x)^2] = E[\Theta|x], \tag{1.12}$$

1.5.2. Función de pérdida valor absoluto

La función de pérdida de valor absoluto se define como la diferencia del valor absoluto de Θ, y su estimador $\widehat{\Theta}$, o sea $|(\Theta - \widehat{\Theta})|$. Esta función de pérdida produce como estimador de Θ a la mediana de la distribución a posteriori de Θ, lo cual se puede verificar de la siguiente forma usando la ecuación (1.7):

$$\begin{aligned}
E[|\Theta - \widehat{\Theta}|] &= \int_{S_\Theta} |\Theta - \widehat{\Theta}| f(\Theta|x) d\Theta \\
&= \int_{-\infty}^{\widehat{\Theta}} (\Theta - \widehat{\Theta}) f(\Theta|x) d\Theta + \int_{\widehat{\Theta}}^{\infty} (\Theta - \widehat{\Theta}) f(\Theta|x) d\Theta \\
&= \int_{-\infty}^{\widehat{\Theta}} \Theta f(\Theta|x) d\Theta - \int_{-\infty}^{\widehat{\Theta}} \widehat{\Theta} f(\Theta|x) d\Theta \\
&+ \int_{\widehat{\Theta}}^{\infty} \Theta f(\Theta|x) d\Theta - \int_{\widehat{\Theta}}^{\infty} \widehat{\Theta} f(\Theta|x) d\Theta
\end{aligned} \tag{1.13}$$

Luego, derivando la ecuación(1.13), y aplicando el segundo teorema fundamental del cálculo, tenemos que:

$$
\begin{aligned}
\frac{d}{d\widehat{\Theta}}E[|\Theta-\widehat{\Theta}|] &= \int_{-\infty}^{\widehat{\Theta}} f(\Theta|x)d\Theta + \widehat{\Theta}f(\widehat{\Theta}|x) \\
&- \widehat{\Theta}f(\widehat{\Theta}|x) - \widehat{\Theta}f(\widehat{\Theta}|x) \\
&- \int_{\widehat{\Theta}}^{\infty} f(\Theta|x)d\Theta + \widehat{\Theta}f(\widehat{\Theta}|x)
\end{aligned} \tag{1.14}
$$

Después, usando el criterio de primera derivada tenemos que:

$$
\int_{-\infty}^{\widehat{\Theta}} f(\Theta|x)d\Theta - \int_{\widehat{\Theta}}^{\infty} f(\Theta|x)d\Theta = 0 \tag{1.15}
$$

$$
\int_{-\infty}^{\widehat{\Theta}} f(\Theta|x)d\Theta = \int_{\widehat{\Theta}}^{\infty} f(\Theta|x)d\Theta \tag{1.16}
$$

$$
F(\widehat{\Theta}|x) = 1 - F(\widehat{\Theta}|x) \tag{1.17}
$$

$$
F(\widehat{\Theta}|x) = \frac{1}{2} \tag{1.18}
$$

Con lo que llegamos a que el valor $\widehat{\Theta}$ que satisface la ecuación(1.18) es la mediana de la distribución a posteriori de Θ, que en efecto es el valor que minimiza la función de pérdida, pues $\frac{d^2}{d\widehat{\Theta}}E[|\Theta-\widehat{\Theta}|] = 2f(\widehat{\Theta}|x) > 0$.

1.5.3. Función de pérdida $0-1$

La función de pérdida de $0-1$ se define como una variable dicotómica, donde $l(\Theta,\widehat{\Theta})$ toma el valor de cero cuando $\Theta = \widehat{\Theta}$, y 1 en caso contrario. Entonces se define a la esperanza de la función de pérdida $0-1$ como:

$$
\begin{aligned}
E[l(\Theta, \widehat{\Theta})] &= 0P(\Theta = \widehat{\Theta}) + 1P(\Theta \neq \widehat{\Theta}) \\
&= P(\Theta \neq \widehat{\Theta}).
\end{aligned}
\tag{1.19}
$$

Minimizar la función (1.5.3) resulta sencillo si tomamos en cuenta que $P(\Theta \neq \widehat{\Theta}) = 1 - P(\Theta = \widehat{\Theta})$. Entonces el problema cambia a maximizar $P(\Theta = \widehat{\Theta})$, lo cual se obtiene cuando $\widehat{\Theta}$ es la moda de la distribución a posteriori de Θ.

En resumen, en este capítulo se habló brevemente de la historia e ideas concernientes al paradigma bayesiano. Luego se introdujeron de forma muy resumida algunos conceptos básicos de teoría de decisión, así como la implementación de estos dentro del paradigma bayesiano, para así concluir cómo obtener estimadores bayesianos. En este capítulo no se habló de alguna distribución de probabilidad en particular de la cual se deseen obtener estimadores de sus parámetros, esto será visto en el capítulo 4.4 para la distribución normal gaussiana inversa. Pero antes de ello es necesario hablar de algunas distribuciones de probabilidad que nos ayudarán a estudiar nuestra distribución de interés como una mezcla gaussiana en esperanza-varianza, las cuales se verán en el siguiente capítulo.

Capítulo 2

Distribuciones de probabilidad relevantes

En este capítulo se expone de manera breve la notación y algunas características de las distribuciones de probabilidad gaussiana, gaussiana p variada, gamma, gamma inversa, Wishart, Wishart inversa, gaussiana inversa generalizada, gaussiana inversa, las cuales se usarán en capítulos posteriores, la referencia a este capítulo es Abramowitz and Stegun (1964).

La notación a emplear adoptará la usada en el enfoque bayesiano, para ser consistente con el tipo de modelación que se realizará. Por lo que si la variable o vector aleatorio, X, tiene una función de densidad parametrizada por algún vector de parámetros, Θ, entonces la función de densidad de X se denotará por $f(x|\Theta)$, que se lee e interpreta como: X se distribuye $f(\cdot)$ dado Θ. La función de densidad de Θ será $\Pi(\theta|\theta_0)$ para algún vector de hiperparámetros θ_0. Esta función de densidad es conocida como distribución a priori, y se interpreta de la misma manera que la distribución de X, salvo que representa un grado subjetivo de creencia respecto a Θ. Para la función de densidad de Θ condicional en X usaremos la notación $f(\theta|x)$,

la cual se conoce como densidad a posteriori (Bernardo and Smith, 2008).

2.1. Distribución gaussiana

Se dice que una variable aleatoria X, que toma valores en los números reales, se distribuye gaussiana con media μ, y desviación estándar σ, si su función de densidad de probabilidad es de la siguiente manera,

$$f_x(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} I_{(-\infty,\infty)}(x). \tag{2.1}$$

En la distribución gaussiana, dos parámetros caracterizan la forma de la distribución, $\mu = E[X]$, y $\sigma^2 = Var[X]$, de aquí que el parámetro μ nos da información sobre la localización de la distribución, mientras que σ^2 de la dispersión. Para referirnos a que X se distribuye gaussiana con parámetros μ y σ usaremos la notación $X|\mu,\sigma \sim N(x|\mu,\sigma)$.

Como se mencionó anteriormente, el parámetro μ es una medida de localización, mientras que el parámetro σ es una medida de dispersión de la distribución gaussiana, lo cual se ilustra en las siguientes gráficas.

En la gráfica 2.1 se pueden observar cuatro funciones de densidad gaussianas, etiquetadas como (a), (b), (c) y (d), con el mismo parámetro de dispersión $\sigma = 1$, pero diferentes parámetros de localización, es decir, $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 4$ y $\mu_4 = 6$ respectivamente.

Por otro lado, en la gráfica 2.2, se pueden observar cuatro funciones de densidad gaussianas, etiquetadas como (a), (b), (c) y (d), con parámetros de localización $\mu_1 = 0$, $\mu_2 = 0$, $\mu_3 = 0$ y $\mu_4 = 0$, y parámetros de dispersión $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 4$ y $\sigma_4 = 6$, respectivamente; adviértase cómo incrementan las colas de la distribución gaussiana entre mayor es el parámetro de dispersión.

Gráfica 2.1: Densidades gaussianas con el mismo parámetro de dispersión, pero diferentes parámetros de localización.

Gráfica 2.2: Densidades gaussianas con diferente parámetro de dispersión.

Como caso particular de la distribución gaussiana tenemos la distribución gaussiana estándar, que se obtiene cuando el parámetro $\mu = 0$, y $\sigma = 1$. También podemos llegar a esta distribución mediante un proceso de normalización de la distribución (2.1), es decir, si X se distribuye $N(x|\mu, \sigma)$, y si consideramos la transformación $Y = \frac{X-\mu}{\sigma}$, entonces Y se distribuye $N(y|0, 1)$.

Por último, la función generadora de momentos y característica de una distribución gaussiana es de la siguiente manera:

$$m_x(t) = \exp\{\mu t + \frac{t^2\sigma^2}{2}\}, \tag{2.2}$$

$$\varphi_x(it) = \exp\{i\mu t - \frac{t^2\sigma^2}{2}\}, \tag{2.3}$$

respectivamente.

2.2. Distribución gaussiana p-variada

Se dice que un vector aleatorio X de dimensión p, con parámetros μ y Σ, que toma valores tanto positivos como negativos en cada entrada, tiene una distribución gaussiana p-variada si su función de densidad es de la siguiente forma:

$$f_x(X|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\{-1/2(X-\mu)'\Sigma^{-1}(X-\mu)\} I_{R^p}(x),$$

donde $\mu = E[X]$, de dimensión p, es el vector de medias, y $\Sigma = Cov[X]$ de dimensión $p \times p$, matriz simétetrica positiva definida, es la matriz de varianza-covarianza. Análogamente a la distribución gaussiana, el vector μ es un vector de posición, el cual indica dónde está centrada la distribución del vector aleatorio X. Mientras que la matriz Σ está formada

por la varianza de cada variable X_i en las entradas diagonales, es decir en la (i,i)-entrada, y por las covarianzas $Cov[X_{i,j}]$ en la (i,j)-entrada de la matriz. De aquí que la matriz de varianzas-covarianzas indique la dispersión del vector aleatorio X alrededor del vector de medias μ. En este caso, si X es un vector aleatorio de dimensión p, con vector de medias μ, y matriz ve varianza-covarianza Σ, usaremos la notación X se distribuye $N_p(x|\mu,\Sigma)$. Como se mencionó anteriormente el vector μ es un parámetro de localización, de la distribución gaussiana p-variada, lo cual se ilustra a continuación para el caso bivariado. En la gráfica 2.3 se muestran tres funciones de densidad gaussianas (a), (b) y (c); todas con la misma matriz de varianza-covarianza I, con $\sigma_1 = \sigma_2 = 1$ y $\sigma_{1,2} = \sigma_{2,1} = 0$; pero con vectores de medias $\mu_1 = (-2,-2)$, $\mu_2 = (0,0)$ y $\mu_3 = (2,2)$, respectivamente.

Gráfica 2.3: Densidad gaussiana p-variada con diferentes vectores de medias μ.

Ahora si nos enfocamos en la matriz de varianza-covarianza Σ, entre mayor sea la varianza del elemento x_i del vector aleatorio X, mayor será la dispersión de la distribución de X en el eje coordenado correspondiente a x_i.

Por otro lado, entre mayor o menor sea la covarianza entre algún par de elementos x_i, x_j pertenecientes a X, la distribución de X tendrá una notable forma elíptica en los ejes coordenados i, j, la cual está ligada a la interpretación de la correlación lineal entre X_i y X_j, pues si $Cov[X_i, X_j]$ es mayor a cero, la distribución de X tendrá una foma elíptica centrada en el

vector de medias μ, pero rotada entre 0 y 90 grados, lo que indica que hay una dependencia lineal positiva entre el i-ésimo y j-ésimo elemento de X. Mientras que, sí la covarianza entre el i-ésimo y j-ésimo elemento de X es menor a cero, la distribución de X en los ejes coordenados i, j tendrá una forma elíptica rotada entre 90 y 180 grados. Los dos puntos anteriores se ilustran a continuación para el caso bivariado.

En la gráfica 2.4 (a) se observa una distribución gaussiana bivariada con vector de medias $\mu = (0, 0)$, y con matriz de varianza-covarianza, Σ, con las entradas $Var[X] = 1$, $Var[Y] = 5$ y $Cov[X, Y] = 0$. En (b) con vector de medias $\mu = (0, 0)$, $Var[X] = 1$, $Var[Y] = 1$ y $Cov[X, Y] = 0$. Y en (c) con vector de medias $\mu = (0, 0)$, $Var[X] = 5$, $Var[Y] = 1$ y $Cov[X, Y] = 0$.

En la gráfica 2.5 (a) se observa una distribución gaussiana bivariada con vector de medias $\mu = (0, 0)$, y con matriz de varianza-covarianza, Σ, con las entradas $Var[X] = 2$, $Var[Y] = 2$ y $Cov[X, Y] = 0.75$. En (b) con vector de medias $\mu = (0, 0)$, $Var[X] = 5$, $Var[Y] = 5$ y $Cov[X, Y] =$ -0.75.

Gráfica 2.4: Densidad gaussiana bivariada con diferentes varianzas y covarianzas

Análogamente a la distribución gaussiana estándar, si X se distribuye $N_p(x|\mu, \Sigma)$, y consideramos la transformación $Y = A^{-1}(X - \mu)$, con $\Sigma = AA^{'}$, entonces el vector aleatorio Y tendrá una distribución $N_p(y|0, I)$, donde 0 es el vector cero de dimensión p, mientras que I es la matriz identidad de dimensión $p \times p$. A esta distribución se le conoce como gaussiana estándar p-variada.

Por último, la función generadora de momentos y característica de la

Gráfica 2.5: Densidad gaussiana bivariada con diferentes valores de covarianza

distribución normal p-variada son las siguientes:

$$m_x(t) = \exp\{\mu t' + \frac{t'\Sigma t}{2}\}, \tag{2.4}$$

$$\varphi_x(it) = \exp\{\mu i t' - \frac{t'\Sigma t}{2}\}, \tag{2.5}$$

respectivamente.

2.3. Distribución gamma

Se dice que una variable aleatoria X, que toma valores en los números reales positivos (sin incluir al cero), se distribuye gamma con parámetros

λ y α, si su función de densidad de probabilidad es de la siguiente manera:

$$f_x(x|\lambda,\alpha) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\lambda x\} I_{(0,\infty)}(x). \tag{2.6}$$

En la distribución gamma dos parámetros caracterizan la forma de la distribución, λ y α. El parámetro λ toma valores mayores a cero, mientras que α puede tomar valores mayores o iguales a cero. El parámetro λ también es conocido como parámetro de escala e influye en el tamaño de la densidad respecto al eje y. Por otro lado, el parámetro α es conocido como parámetro de forma, e influye en la forma de la distribución.

Para referirnos a que X se distribuye gamma con parámetros λ y α, usaremos la notación $X|\lambda,\alpha$ se distribuye $\Gamma(x|\lambda,\alpha)$.

Como se mencionó anteriormente, el parámetro λ es un parámetro de escala, mientras que el parámetro α es un parámetro de forma, lo cual se ilustra a continuación.

En la gráfica 2.6 se muestran cuatro funciones de densidad gamma con mismo parámetro de forma $\alpha = 1$, pero con distintos parámetros de escala, es decir, con $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$ y $\lambda_4 = 4$, respectivamente.

En la gráfica 2.7 se muestran cuatro funciones de densidad gamma con al mismo parámetro de escala $\lambda = 1$, pero con distintos parámetros de forma, es decir, con $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$ y $\alpha_4 = 4$, respectivamente. Adviértase que conforme α es menor el sesgo de la distribución aumenta a la derecha, mientras que si es mayor disminuye a la derecha y aumenta a la izquierda.

Gráfica 2.6: Densidad gamma con diferente parámetro λ

Si la variable aleatoria X se distribuye gamma con parámetros λ, α, entonces:

$$\begin{aligned} E[x] &= \frac{\alpha}{\lambda}, \\ Var[x] &= \frac{\alpha}{\lambda^2}, \\ m_x(t) &= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \\ \varphi_x(it) &= \left(\frac{\lambda}{\lambda - it}\right)^{\alpha}. \end{aligned} \tag{2.7}$$

Por último, es importante mencionar que la distribución gamma es la distribución a priori conjugada de la distribución gaussiana, para el parámetro de escala, la cual se introduce más adelante.

Gráfica 2.7: Densidad gamma con diferente parámetro α

2.4. Distribución gamma inversa

La distribución gamma inversa resulta de aplicar la transformación $X = 1/Y$ a la variable aleatoria Y, donde Y se distribuye gamma con parámetros λ y α. Dicho lo anterior se tiene la siguiente definición.

Se dice que una variable aleatoria X, que toma valores en los números reales positivos, se distribuye gamma inversa con parámetros λ y α, si su función de densidad de probabilidad es de la siguiente manera:

$$f_x(x|\lambda, \alpha) = \frac{\lambda^\alpha x^{1-\alpha}}{\Gamma(\alpha)} \exp\{-\frac{\lambda}{x}\} I_{(0,\infty)}(x). \tag{2.8}$$

La distribución gamma inversa hereda sus dos parámetros de la distribución gamma, por lo que tanto λ como α tienen las mismas restricciones que en la distribución gamma, y también caracterizan la forma de la distribución. Al igual que en la distribución gamma, el parámetro λ es conocido

como parámetro de escala e influye en el tamaño de la densidad respecto al eje y. Mientras que el parámetro α influye en la forma de la distribución.

Para referirnos a que X se distribuye gamma inversa con parámetros λ y α, usaremos la notación $X|\lambda, \alpha \sim Gamma - Inv(x|\lambda, \alpha)$.

Como se mencionó anteriormente, el parámetro λ es un parámetro de escala, mientras que el parámetro α es un parámetro de forma, lo cual se ilustra a continuación.

En la gráfica 2.8 se muestran cuatro funciones de densidad gamma inversa con el mismo parámetro de forma $\alpha = 1$, pero con distintos parámetros de escala, es decir, con $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$ y $\lambda_4 = 4$, respectivamente.

En la gráfica 2.9 se muestran cuatro funciones de densidad gamma inversa con al mismo parámetro de escala $\lambda = 1$, pero con distintos parámetros de forma, es decir, con $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$ y $\alpha_4 = 4$, respectivamente. Adviértase que conforme α es mayor la curtosis de la distribución aumenta, mientras que si es menor disminuye.

Gráfica 2.8: Densidad gamma inversa con diferente parámetro λ

Si la variable aleatoria X se distribuye gamma inversa con parámetros λ, α, entonces:

$$\begin{aligned} E[x] &= \frac{\lambda}{\alpha - 1}, \\ Var[x] &= \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}, \\ \varphi_x(it) &= \left(\frac{2(-\lambda it)^{\alpha/2}}{\gamma(\alpha)}\right)\kappa(\sqrt{-4\lambda it}). \end{aligned} \tag{2.9}$$

Gráfica 2.9: Densidad gamma inversa con diferente parámetro α

2.5. Distribución gaussiana inversa

Se dice que la variable aleatoria X tiene una distribución gaussiana inversa si su función de densidad es de la siguiente forma:

$$f_x(X|\lambda,\psi) = \sqrt{\frac{\lambda}{2\pi x^3}}\exp\{-\frac{\lambda(x-\Psi)^2}{2\Psi^2 x}\},$$

para λ, ψ y X mayores a cero. Como notación, si X se distribuye gausiana inversa con parámetros λ, Ψ, diremos que $X|\lambda,\psi$ se distribuye $GI(x|\lambda,\Psi)$.

Ahora veamos cómo es la función característica de esta distribución. Si X se distribuye $GI(x|\lambda,\Psi)$, entonces su función característica es:

$$
\begin{aligned}
\varphi_x(it) &= \int_0^\infty \exp\{itx\}\sqrt{\frac{\lambda}{2\pi x^3}}\exp\{-\frac{\lambda(x-\Psi)^2}{2\Psi^2 x}\}dx \\
&= \int_0^\infty \sqrt{\frac{\lambda}{2\pi x^3}}\exp\{-\frac{\lambda}{2\Psi^2}(x-2\Psi-(itx2\Psi^2/\lambda)+\Psi^2 x)\}dx.
\end{aligned} \tag{2.10}
$$

Trabajando únicamente con el exponente de la expresión anterior, y factorizando el término $(1-\frac{it2\Psi^2}{\lambda})$ tenemo que:

$$
\begin{aligned}
&-\frac{\lambda}{2\Psi^2}(1-\frac{it2\Psi^2}{\lambda})(x-2\Psi/(1-\frac{it2\Psi^2}{\lambda})+\Psi^2/(1-\frac{it2\Psi^2}{\lambda}) \\
&= -\frac{\lambda}{2\Psi^2}(1-\frac{it2\Psi^2}{\lambda})(x-2\Psi/(1-\frac{it2\Psi^2}{\lambda})+\Psi^2/x(1-\frac{it2\Psi^2}{\lambda}) \\
&+2\Psi/\sqrt{1-\frac{it2\Psi^2}{\lambda}}-2\Psi/\sqrt{1-\frac{it2\Psi^2}{\lambda}}) \\
&= \frac{\lambda}{\Psi}-\frac{\lambda}{\Psi}\sqrt{(1-\frac{it2\Psi^2}{\lambda})}-\frac{\lambda}{2x\Psi^2}(1-\frac{it2\Psi^2}{\lambda})(x-\Psi/\sqrt{1-2it\Psi^2/\lambda})^2.
\end{aligned} \tag{2.11}
$$

Por lo que la función característica queda de la siguiente forma:

$$
\begin{aligned}
\varphi_x(it) &= (\exp\{\frac{\lambda}{\Psi}-\frac{\lambda}{\Psi}\sqrt{(1-\frac{it2\Psi^2}{\lambda})}\}) \\
&\quad (\int_0^\infty \sqrt{\frac{\lambda}{2\pi x^3}}\exp\{\frac{\lambda}{2x\Psi^2}(1-\frac{it2\Psi^2}{\lambda})(x-\Psi/\sqrt{1-2it\Psi^2/\lambda})^2\}dx) \\
&= \exp\{\frac{\lambda}{\Psi}-\frac{\lambda}{\Psi}\sqrt{(1-\frac{it2\Psi^2}{\lambda}}\}.
\end{aligned}
$$

El resultado anterior se sigue de que la integral previa es una distribución gaussiana inversa con parámetros λ, $\Psi/\sqrt{1-(2it\Psi^2)/\lambda}$ integrada

sobre su soporte, por lo cual vale 1, y por lo tanto:

$$\varphi_x(it) = \exp\{\frac{\lambda}{\Psi} - \frac{\lambda}{\Psi}\sqrt{(1 - \frac{2it\Psi^2}{\lambda})}\}. \tag{2.12}$$

Con el resultado anterior se puede probar que si X se distribuye $GI(x|\lambda, \Psi)$, entonces:

$$m_x(t) = \exp\{\frac{\lambda}{\Psi} - \frac{\lambda}{\Psi}\sqrt{(1 - \frac{2t\Psi^2}{\lambda})}\}, \qquad E[X] = \Psi,$$
$$Var[X] = \frac{\Psi^3}{3}, \qquad Sesgo[X] = \frac{2\Psi^4}{\lambda} + \frac{3\Psi^5}{\lambda^2}.$$

Entoces, al igual que la distibución gaussiana, la distribución gaussiana inversa tiene un parámetro de localización, es decir, Ψ. Es importante notar como afecta el parámetro λ a la dispersión de la variable aleatoria X, y además, esta distribución siempre tiene un sesgo positivo ya que Ψ y λ son mayores a cero. Por lo que para valores muy pequeños de λ o muy grandes de Ψ la distribución tendrá más peso en la cola.

2.6. Distribución gaussiana inversa generalizada

Se dice que la variable aleatoria X, que toma valores en los números reales positivos, tiene una distribución gaussiana inversa generalizada, denotada por $GIG(x|\lambda, \xi, \Psi)$, si su función de densidad es de la siguiente forma:

$$f(x) = \frac{\xi^{-\lambda}\sqrt{\xi\Psi}^{\lambda}x^{\lambda-1}\exp\{\frac{-1}{2}(\xi x^{-1} + \Psi x)\}}{2\kappa_\lambda(\sqrt{\xi\Psi})}I_{(0,\infty)}(x),$$

donde $\kappa_{\lambda(.)}$ es una función modificada de Bessel de tercer tipo (Abramowitz and Stegun, 1964) , y si $\lambda < 0$, entonces $\xi > 0$, $\Psi \geq 0$; si $\lambda = 0$, entonces $\xi > 0$, $\Psi > 0$, si $\lambda > 0$, entonces $\xi \geq 0$, $\Psi > 0$.

En la distribución gaussiana inversa generalizada tres parámetros caracterizan la forma de la distribución. Tanto el parámetro ξ como el parámetro Ψ influyen en la escala de la distribución, mientras que el parámetro λ en la forma. El punto anterior se ilustra a continuación.

En la gráfica 2.10 se muestran cuatro funciones de densidad GIG, todas con los mismos parámetros $\lambda = 1$ y $\xi = 1$, pero con diferentes parámetros $\psi_1 = 0.5$, $\psi_2 = 1$, $\psi_3 = 2$ y $\psi_4 = 4$, respectivamente. Es importante notar que entre menor es el valor de ψ, menor es la curtosis, pero mayor es el sesgo a la derecha de la distribución.

Gráfica 2.10: Densidad gaussiana inversa generalizada con diferentes valores del parámetro ψ.

En la gráfica 2.11 se observan cuatro funciones de densidad GIG, todas con prámetros $\lambda = 1$, $\psi = 1$, pero con diferentes parámetros $\xi_1 = 0.5$, $\xi_2 = 1$, $\xi_3 = 2$ y $\xi_4 = 4$, respectivamente. Es importante notar que entre mayor es el valor del parámetro ξ, mayor es el sesgo a la derecha de la distribución.

Gráfica 2.11: Densidad gaussiana inversa generalizada con diferentes valores del parámetro ξ

En la gráfica 2.12 se observan cuatro funciones de densidad GIG, todas con parámetros $\psi = 1$, $\xi = 1$, pero con diferentes parámetros λ_1 = -0.5, $\lambda_2 = 0$, $\lambda_3 = 0.5$ y $\lambda_4 = 1$, respectivamente. En este caso el parámetro λ también influye en la curtosis de la distribución, pero sobre todo influye en la cola de esta. El parámetro λ determina la familia paramétrica, pues si la variable aleatoria X se distribuye GIG, con parámetro $\lambda = 0$ entonces se obtiene la distribución hiperbólica, mientras que si λ = -0.5 se obtiene

la distribución gaussiana inversa de la cual se hablará posteriormente.

Gráfica 2.12: Densidad gaussiana inversa generalizada con diferentes valores de λ.

Ahora veamos algunos resultados de la distribución GIG. Si X se distribuye $GIG(x|\lambda, \xi, \Phi)$, entonces su función generadora de momentos es:

$$\begin{aligned}
\varphi(it) &= \int_{-\infty}^{\infty} \exp\{itx\} \frac{\xi^{-\lambda}\sqrt{\xi\Psi}^{\lambda} x^{\lambda-1} \exp\{\frac{-1}{2}(\xi x^{-1} + \Psi x)\}}{2\kappa_\lambda(\sqrt{\xi\Psi})} dx \\
&= c \int_{-\infty}^{\infty} \xi^{-\lambda} \frac{\sqrt{\xi(2it+\Psi)}^{\lambda}}{2\kappa_\lambda(\sqrt{\xi(2it+\Psi)})} \exp\{\frac{-1}{2}(\xi x^{-1} + (2it+\Psi)x)\} dx,
\end{aligned}$$

con $c = \frac{\sqrt{\xi\Psi}^{\lambda}}{2\kappa_\lambda(\sqrt{\xi\Psi})} \frac{2\kappa_\lambda(\sqrt{\xi(2it+\Psi)})}{\sqrt{\xi(2it+\Psi)}^{\lambda}}$. La última integral vale uno por ser una

densidad $GIG(\lambda, \xi, 2it + \psi)$ integrada sobre su soporte, por lo que:

$$\varphi(it) = \frac{\sqrt{\xi\Psi}^{\lambda}}{\kappa_\lambda(\sqrt{\xi\Psi})} \frac{\kappa_\lambda(\sqrt{\xi(2it + \Psi)})}{\sqrt{\xi(2it + \Psi)}^{\lambda}}.$$

Si X se distribuye $GIG(x|\lambda, \xi, \Phi)$, entonces su r-ésimo momento es:

$$E[X^r] = \int_{-\infty}^{\infty} x^r \frac{\xi^{-\lambda}\sqrt{\xi\Psi}^{\lambda} x^{\lambda-1} \exp\{\frac{-1}{2}(\xi x^{-1} + \Psi x)\}}{2\kappa_\lambda(\sqrt{\xi\Psi})} dx.$$

$$= \frac{x^r 2\kappa_{\lambda+r}(\sqrt{\xi\Psi})}{\sqrt{\xi\Psi}^{r} 2\kappa_\lambda(\sqrt{\xi\Psi})} \int_{-\infty}^{\infty} \frac{\xi^{-\lambda-r}\sqrt{\xi\Psi}^{\lambda+r} x^{\lambda+r-1} \exp\{\frac{-1}{2}(\xi x^{-1} + \Psi x)\}}{2\kappa_{\lambda+r}(\sqrt{\xi\Psi})} dx,$$

por lo tanto:

$$E[X^r] = (\frac{\xi}{\Psi})^{\frac{r}{2}} \frac{\kappa_{\lambda+r}(\sqrt{\xi\Psi})}{\sqrt{\kappa_\lambda(\sqrt{\xi\Psi})}}.$$

2.7. Distribución Wishart

La distribución Wishart es utilizada como distribución conjugada de la matriz de varianza-covarianza de vectores aleatorios normales de dimensión p, y se deriva de la siguiente manera:

Supongamos que tenemos n vectores aleatorios X de dimensión p que se distribuyen $N_p(x|0, \Sigma)$, luego un estimador de la matriz de varianza-covarianza es $S = \Sigma_{i=1}^{n} x_i x_i^{'}/n$, que es una matriz positiva y simétrica de dimensión $p \times p$, por lo que para los distintos valores que tomen los vectores aleatorios X, S también tomará un valor diferente, de quí que sea natural preguntarse por la distribución de S, con lo cual se llega a la siguiente caracterización.

Se dice que una matriz S de dimensión $p \times p$, simétrica y positiva definida, se distribuye Wishart con n grados de libertad, si su función de densidad es de la siguiente forma:

$$f_S(S|\Sigma, n, p) = c\frac{|S|^{(n-p-1)/2}}{|\Sigma|^{n/2}} \exp\{-\frac{1}{2}tr(\Sigma^{-1}S)\},$$

donde $c = \left(2^{np/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\gamma(\frac{n+1-i}{2})\right)^{-1}$, Σ es una matriz simétrica y positiva definida de dimensión $p \times p$, y se le conoce como matriz de escala, n es el número de vectores disponibles, y es mayor a la dimensión de los vectores, es decir, mayor que p. En este caso usaremos la notación S se distribuye $W(S|\Sigma, n, p)$.

Algunas características numéricas de la distribución Wishart son las siguientes:

- $E[S] = n\Sigma$.

- Si S_1 se distribuye $W(S_1|\Sigma, n, p)$, y S_2 se distribuye $W(S_2|\Sigma, m, p)$, entonces $S_1 + S_2$ se distribuye $W(S_1 + S_2|\Sigma, n + m, p)$.

- La distribución marginal del i-ésimo componente de la distribución Wishart se distribuye $\Gamma(\frac{1}{2}, \frac{1}{2})$.

2.8. Distribución Wishart inversa

Se dice que una matriz G de dimensión $p \times p$, simétrica y positiva definida, se distribuye wishart inversa con n grados de libertad, si su función

de densidad es de la siguiente forma:

$$f_G(G) = c\frac{|K|^{(n-p-1)/2}}{|G|^{n/2}} \exp\{-\frac{1}{2}tr(G^{-1}K)\},$$

donde $c = \left(2^{(n-p-1)p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\gamma(\frac{n-p-i}{2})\right)^{-1}$, K es una matriz simétrica y definida positiva, que se conoce como matriz de escala. En este caso usaremos la notación G se distribuye $IW(G|K, n, p)$.

Una propiedad importante que liga a la distribución Wishart y Wishart inversa, es la siguiente.

Si una matriz aleatoria Σ de dimensión $p \times p$ se distribuye wishart con matriz de escala S, y con n grados de libertad, entonces su matriz inversa, Σ^{-1}, se distribuye wishart inversa con matriz de escala S^{-1}, y con $n+p+1$ grados de libertad.

Capítulo 3

Mezclas de distribuciones gaussianas

En este capítulo primero se hablará en términos generales sobre qué es una distribución tipo mezcla. Después se hablará de un caso particular de las distribuciones tipo mezcla continua, es decir, de las distribuciones tipo mezcla normal (McNeil et al., 1967). Para después caracterizar a la distribución hiperbólica generalizada a partir de una distribución tipo mezcla. Por último, se aprovecharán algunas propiedades de las distribuciones tipo mezcla normal, para obtener algunas características de interés de la distribución hiperbólica generalizada.

3.1. Distribuciones tipo mezcla

Este tipo de distribución consiste en ponderar un conjunto discreto de distribuciones de vectores aleatorios X_i (discretos) con el mismo soporte y la misma dimensión con un conjunto de pesos w_i, positivos y menores a uno, w_i, donde i proviene de un conjunto I de índices numerable, y a su vez $\sum_{i \in I} w_i = 1$. El índice i puede ser generado por una distribución de

probabilidad discreta tal que $P(i) = w_i$; por ello a los pesos también se les conoce como probabilidades. Entonces la función de densidad es de la siguiente forma:

$$f_x(x) = \sum_{i \in I} w_i f_{X_i}(x|w_i),$$

donde $0 < w_i < 1$, $\sum_{i \in I} w_i = 1$, y $f_{X_i}(x|w_i)$ es un vector aleatorio discreto de dimensión p para todo i en I.

Es común encontrar este tipo de distribuciones en un conjunto de datos que provienen de dos o más distribuciones diferentes, lo cual podría resultar en una distribución multimodal.

Si el conjunto de vectores aleatorios provienen de la misma distribución, entonces se dice que es una mezcla de familias paramétricas, y si el vector aleatorio tiene una distribución continua se le conoce como densidad tipo mezcla continua.

Una densidad tipo mezcla continua resulta de considerar que los pesos w están dados por una función de densidad de probabilidad continua $f_w(w)$ con soporte en R_+; a w se le conoce como variable de mezcla. Ahora, el vector aleatorio X queda parametrizado por (w), por lo que la familia de vectores aleatorios queda determinada por la familia no numerable X dado w, con w distribuido $f_w(w)$. Entonces, la densidad de X está dada por:

$$f_x(x) = \int_0^\infty f(x|w) f_w(w) dw.$$

Si w es una variable aleatoria con distribución paramétrica, es común que la distribución de X dependa de los parámetros de w, y además ciertas

características como la dispersión o la cola de la distribución de X también dependan de las características de la variable de mezcla w.

3.2. Mezcla gaussiana en media

Se dice que un vector aleatorio X, de dimensión p, es una mezcla en media si X dado u tiene una distribución gaussiana con vector de medias $u\mu$, y matriz de varianza-covarianza Σ. La variable aleatoria u está definida en R_+. Entonces la probabilidad de que X pertenezca a algún subconjunto $S_p \subset R^p$ es:

$$P(X \in S_p) = \int_0^\infty N_p(x \in S_p | u\mu, \Sigma) f_u(u) du, \tag{3.1}$$

donde $\mu \in R^p$, $\Sigma \in M_{p\times p}$, u es la variable de mezcla y $X = (X_1, X_2, ..., X_p)$.

Si nos enfocamos en la distribución marginal de X dado u podemos notar que el valor de u influye en donde está centrada la distribución, pues ahora el vector de medias pertenece al segmento dirigido $L(\mu) = (u\mu | u \in R_+, \mu \in R^p)$. Entonces, si consideramos que u ha tomado el valor de u_i, y una muestra aleatoria de X dado u_i, tendremos una muestra centrada sobre el segmento dirigido $L(\mu)$ en el punto $u_i\mu$, por lo que para valores de u cercanos a uno, la muestra permanece centrada alrededor de μ, mientras que para valores grandes de u, la muestra se aleja en dirección del vector de medias μ.

En la gráfica 3.1 se ilustra la simulación de una distribución tipo mezcla gaussiana con vector de medias $\mu = (9, 11)$, matriz de varianza-covarianza Σ, con entradas $\sigma_1 = \sigma_2 = 5$, y $\sigma_{1,2} = \sigma_{2,1} = 0.1$, y con variable de mezcla

distribuida exponencialmente, para tres diferentes valores de u. Obsérvese que para cada valor que tomó la variable de mezcla u, las realizaciones del vector aleatorio X dado u se concentrán al rededor de algún punto del vector de medias $u\mu$.

Gráfica 3.1: Simulaciones de una distribució tipo mezcla gaussiana en media.

Gráfica 3.2: Contornos de una distribución tipo mezcla gaussiana en media.

En la gráfica 3.2 se observa como tanto la función de densidad, como los contornos del vector aleatorio X dado u, se ven afectados por la variable de mezcla u, para tres diferentes valores de u.

Ahora veamos analíticamente cómo influyen algunas de las caractarísticas numéricas de la variable de mezcla u en la dispersión del vector aleatorio

X, por lo que:

$$\begin{aligned}
E[X] &= \int_0^\infty \left(\int_{D_X} x N_p(x|u\mu, \Sigma) f_u(u) du \right) dx \\
&= \int_0^\infty f_u(u) \left(\int_{D_X} x N_p(x|u\mu, \Sigma) f_u(u) dx \right) du \\
&= \int_0^\infty f_u(u) E[X|u] du \\
&= \int_0^\infty f_u(u) u\mu du \\
&= \mu \int_0^\infty u f_u(u) du \\
&= \mu E[u]
\end{aligned} \tag{3.2}$$

Mientras que por el apéndice B.4 se tiene que:

$$\begin{aligned}
Cov(X) &= E[Cov(X|u)] + Cov[E[X|u]] \\
&= E[\Sigma] + Cov[u\mu] \\
&= \Sigma + \mu \Sigma_u \mu',
\end{aligned} \tag{3.3}$$

donde Σ_u es la matriz diagonal, con todas sus entradas diagonales iguales a $var(u)$.

Así, la variable de mezcla u influye en la localización y dispersión del vector aleatorio X; intuitivamente podemos considerar al producto de u con el vector de medias μ como un sesgo adicional que influye en la realización del vector aleatorio X. Este tipo de mezcla nos permite modelar conjuntos de datos que sigan alguna tendencia con dispersión constante, como se ilustra en la gráfica 3.1. Por otra parte , una de las mayores difi-

cultades es que, salvo en algunas excepciones, no siempre existe una forma cerrada para la densidad del vector aleatorio X, por lo que tendría que ser aproximada mediante algún método númerico.

Mediante un proceso de simulación (apéndice B.1) se generaron realizaciones del vector aleatorio X a través de distribuciones tipo mezcla normal, lo cual se ilustra en la gráfica 3.3. La gráfica (a) proviene de una variable de mezcla exponencial con parámetro $\lambda = 1.5$, la (b) proviene de una variable de mezcla gamma con parámetros $\alpha = 20$, $\beta = 5$, la (c) provine de una variable de mezcla pareto con parámetros $\theta = 5$, $\alpha = 2$, y la (d) proviene de una variable de mezcla ji-cuadrada con 20 grados de libertad (el vector aleatorio X dado u se distribuye de la misma manera que en la gráfica 3.1).

Gráfica 3.3: Simulación del vector aleatorio X a través de distribuciones tipo mezcla gaussiana en media con diferente variable de mezcla.

Como se describió en la gráfica 3.1, las simulaciones del vector aleatorio X parecen seguir una tendencia, y además es importante notar que la dispersión alrededor del vector de tendencia de la gráfica (a) o (b) es mayor

a la de la gráfica (c) o (d). También es importante notar que en la gráfica (c) hay valores aparentemente grandes, mientras que en la gráfica (d) casi todos los valores parecen concentrarse en algún intervalo de la posible recta de tendencia. Por lo que, intuitivamente la cola de la distribución de u influye en la distribución de X. La gráfica (c) proviene de considerar una variable de mezcla distribuida Pareto, la cual tiene cola pesada, y como consecuencia podemos observar gran dispersión sobre el vector de tendencia. Mientras que la gráfica (d) proviene de considerar una variable de mezcla u distribuida ji cuadrado, que es de cola ligera, y como consecuencia las simulaciones del vector X no están tan dispersas sobre el vector de tendencia (lo mismo podemos decir de la gráfica (a) y (b)).

3.3. Mezcla gaussiana en varianza

Ahora consideremos un modelo similar al anterior, pero con la diferencia que la variable de mezcla u solamente afecta a la matriz de varianza-covarianza Σ. Luego, la distribución del vector aleatorio X de dimensión p está dado por:

$$P\left(X \in S_p\right) = \int_0^\infty N_p(x \in S_p|\mu, u\Sigma) f_u(u) du, \tag{3.4}$$

donde $\mu \in R^p$, $\Sigma \in M_{p\times p}$, u es la variable de mezcla y $X = (X_1, X_2, ..., X_p)$.

En la distribución condicional de X dado u, podemos notar que u influye en la dispersión de X, por lo que entre mayor sea el valor que toma u, mayor será la dispersión al rededor del vector de medias μ. Como ejemlo, en la gráfica 3.4 se presenta la simulación de una variable tipo mezcla normal en varianza, con vector de medias $\mu = (9, 11)$, matriz de varianza-covarianza $\sigma_1 = \sigma_2 = 5$, y $\sigma_{1,2} = \sigma_{2,1} = 0.1$, y variable de mezcla u

distribuida Exponencial, con parámetro $\lambda = 1$.

Gráfica 3.4: Simulación de una distribución tipo mezcla gaussiana en varianza.

En la gráfica (3.5) se observa cómo se modifican tanto la densidad como los contornos de la densidad de X dado u del ejemplo anterior. Claramente se observa que entre mayor sea el valor de u, mayor dispersión tendrá el vector X dado u.

Gráfica 3.5: Contornos de una distribución tipo mezcla gaussiana en varianza.

Ahora considerando algunas características numéricas de X, y usando el apéndice B.4, es fácil probar que si X se puede expresar como una distribución tipo mezcla normal en varianza, entonces:

$$\begin{aligned} E[X] &= \mu, \\ COV[X] &= COV_u(E_x(X|u)) + E_x(COV_u(X|u)) \\ &= E(u)\Sigma E(u)'. \end{aligned} \tag{3.5}$$

Por lo que podemos conocer la esperanza y matriz de varianza-covarianza del vector X con tan sólo conocer la matriz de varianza-covarianza de la distribución de X dado u. Además, la distribución del vector X estará centrada en el mismo vector de medias que X dado u para cualquier valor de u, mientras que la matriz de varianza-covarianza será proporcional a la matriz de varianza-covarianza de X dado u.

En la gráfica 3.6 se presentan simulaciones de realizaciones del vector aleatorio X a través de distribuciones tipo mezcla normal en varianza, para ello se utilizaron las mismas distribuciones de mezcla que en la gráfica 3.3. Adviértase que, a diferencia del condicionamiento en media, los datos solo se concentran alrededor del vector de medias μ, pues $E[X] = E[X|u]$, perdiendo así la trayectoria que induce el vector de medias; intuitivamente podemos pensar que $u\mu$ del modelo de mezcla en media genera un vector de trayectoria para las observaciones de X, y que además induce un sesgo de las observaciones del vector X sobre el vector de tendencia μ.

Gráfica 3.6: Simulación de distribuciones tipo mezcla gaussiana en varianza.

3.4. Mezcla gaussiana en media-varianza

Se dice que un vector aleatorio $X \in R^p$ es un vector tipo mezcla p-dimensional, si X dado u se ditribuye Normal con vector de medias $\mu + u\beta$, y matriz de varianza-covarianza $u\Sigma$, donde u es la variable de mezcla con soporte en R_+, μ, β, $\in R^p$ y $\Sigma \in M_{p\times p}$ matriz de varianza-covarianza.

Así pues, la distribución del vector X está dada por:

$$f_X(x) = \int_{S_u} \frac{1}{(2\pi)^{n/2}|u\Sigma|^{n/2}} \exp\{-\frac{1}{2}(x-\mu-uB)'u\Sigma^{-1}(x-\mu-uB)\} f(u)du.$$

Si nos concentramos en la distribución marginal de X dado u, notamos que ahora, para cada realización u_i, la densidad de X dado u_i estará

centrada en el vector $\mu + u_i\beta$, y conforme u_i tome valores más grandes, la densidad de X dado u_i se desplazará en dirección del vector $u_i\beta$ cada vez con mayor dispersión.

En la gráfica 3.7 se puede observar una variable tipo mezcla en media-varianza, con parámetros $\beta = (9, 11)$, $\mu = (5, 10)$, y matriz de varianza-covarianza Σ con $\sigma_z = \sigma_y = 5$, y $COV(Z, Y) = 0.1$, donde $X = (Z, Y)$ es un vector aleatorio de dimensión 2. La variable de mezcla u se distribuye exponencial y tomó los valores de 1, 3 y 6 respectivamente.

Gráfica 3.7: Simulación de una variable tipo mezcla gaussiana en esperanza-varianza.

En la distribución de X podemos considerar a μ como un vector de posición, mientras que $u\beta$ es un vector de dirección o tendencia alrededor del cual estarán las observaciones de X. Los datos conservarán la estructura de correlación inherente a Σ, pero además estarán dispersos en proporción a la esperanza de la variable de mezcla u, más alguna constante que depende de u. Siguiendo la idea intuitiva del modelo de mezcla en media, el producto de la variable de mezcla u y el vector β inducen un sesgo en cada realización del vector aleatorio X. Entonces, considerando el apéndice B.4, la esperanza y covarianza del vector X tienen la forma:

$$\begin{aligned} E(X) &= E_u(E_x(x|u)) \\ &= E_u(\mu + u\beta) \\ &= \mu + E(u)\beta, \\ COV(X) &= COV_u(E_x(X|u)) + E_u(COV_x(X|u)) \\ &= COV_u(\mu + u\beta) + E_u(u\Sigma) \\ &= Var(u)\beta\beta' + E(u)\Sigma. \end{aligned} \tag{3.6}$$

Y además, como se verá a continuación, la función generadora de momentos $M_x(t)$ del vector aleatorio X es proporcional a $M_u(t)$, la función generadora de momentos de la variable de mezcla u. Y la función característica $\varphi_x(it)$ es proporcinal a la función característica de la variable de mezcla u, $\varphi_u(it)$.

La función característica de X es:

$$\varphi_X(t) = \exp\{it\mu'\}\varphi_u(it\beta' - \frac{1}{2}t'\Delta t)$$

Donde $\varphi_u(.)$ es la función característica de la variable de mezcla u.

$$
\begin{aligned}
\varphi_X(t) &= \int_x \exp\{it\} f(x)dx \\
&= \int_x \exp\{it\}(\int_u f(x|u)f(u)du)dx \\
&= \int_x \int_u exp\{it\} f(x|u)f(u)dudx \\
&= \int_u \int_x exp\{it\} f(x|u)f(u)dudx \\
&= \int_u exp\{it(\mu + u\beta) + \frac{1}{2}itu\Delta it\} f(u)du,
\end{aligned}
$$

ya que la función generadora de momentos de una distribución Normal p-variada es $exp(t'\mu + t\Sigma t)$, y además X dado u se distribuye $N(\mu + u\beta, u\Delta)$.

Factorizando el exponenete de la exponencial se tiene que:

$$
\begin{aligned}
\int_u exp\{it(\mu + u\beta) &+ \frac{1}{2}itu\Delta it\} f(u)du = \\
&= \int_u \exp\{it'\mu\} exp\{u(it'\beta - \frac{1}{2}t'u\Delta t)\} f(u)du \\
&= exp\{it'\mu\}\varphi_u(it'\beta - \frac{1}{2}t\Delta t).
\end{aligned} \tag{3.7}
$$

Por lo tanto:

$$\begin{aligned} \varphi_X(t) &= exp\{it'\mu\}\varphi_u(it'\beta + \frac{i^2}{2}t'\Delta t) \\ &= exp\{it'\mu\}\varphi_u(it'\beta - \frac{1}{2}t'\Delta t). \end{aligned} \tag{3.8}$$

Para comprobar que la función generadora de X es:

$$M_x(t) = \exp\{t'\mu\}M_u(t'\beta + \frac{1}{2}t'\Delta t),$$

basta con sustituir t por it en el resultado anterior.

El resultado anterior nos indica que podemos obtener la función generadora de momentos o característica del vector aleatorio X con tan solo conocer la función característica o generadora de la variable de mezcla u, lo cual sería de utilidad si sólo nos interesan algunos momentos, o también nos ayudaría a identificar la distribución de X, en caso de reconocer la función característica o generadora de momentos. Inversamente, el resutado anterior indica que si la función característica o generadora de algún vector aleatorio Z se puede descomponer como el producto de $\exp\{it'\mu\}$, y alguna $M_y(t)$, entonces Z es una variable normal de mezcla en esperanza-varianza, con distribución de mezcla Y.

3.5. Distribución hiperbólica generalizada

La familia paramétrica hiperbólica generalizada se obtiene a partir de definir una mezcla de distribuciones gaussianas en media-varianza, donde el vector aleatorio X de dimensión p condicionado en u tiene una distribución gaussiana con vector de medias $\mu + u\beta$, y matriz de varianza covarianza

$u\Sigma$, mientras que la variable de mezcla u se distribuye gaussiana inversa generalizada con parámetros Ψ, ξ y λ, por lo que formalmente la función de densidad del vector aleatorio X queda de la siguiente forma:

Se dice que el vector aleatorio X p-variado tiene una distribución hiperbólica generalizada si su densidad es de la siguiente manera:

$$f_X(x|\lambda,\xi,\Psi,\mu,\Sigma,\beta) = \frac{\sqrt{\xi + (x-\mu)'\Sigma^{-1}(x-\mu)(\Psi+\beta\Sigma^{-1}\beta)}}{\xi + (x-\mu)'\Sigma^{-1}(x-\mu)(\Psi+\beta'\Sigma^{-1}\beta)^{\frac{d}{2}-\lambda}} \tag{3.9}$$

$$\times c\kappa_{\lambda-d/2}\exp\{(x-\mu)'\Sigma^{-1}\beta\},$$

donde $c = \frac{\sqrt{\xi\lambda}^{-\lambda}\Psi^{\lambda}(\Psi+\beta\Sigma^{-1}\beta)^{\frac{d}{2}-\lambda}}{(2\Pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\kappa_{\lambda(\sqrt{\xi\Psi})}}$.

Si el vector aleatorio X tiene una distribución hiperbólica generalizada usaremos la notación X se distribuye $GH_p(x|\lambda,\xi,\Psi,\mu,\Sigma,\beta)$.

La distribución hiperbólica generalizada se caracteriza a partir de una mezcla gaussiana en esperanza y varianza de la siguiente manera: sea X dado u un vector aleatorio de dimensión p que se distribuye $N(x|\mu+u\beta,u\Sigma)$, y sea u la variable de mezcla distribuida $GIG(u|\lambda,\xi,\Psi)$, entonces:

$$f_X(x|\mu,u,\beta,\Sigma,\lambda,\xi,\Psi) = \int_{-\infty}^{\infty} \frac{\exp\{(x-\mu-u\beta)'(u\Sigma)^{-1}(x-\mu-u\beta)\}}{(2\Pi)^{d/2}|\Sigma|^{1/2}u^{d/2}} f(u)du, \tag{3.10}$$

donde:

$$f(u) = \frac{\xi^{-\lambda}\sqrt{\xi\Psi}^{\lambda}u^{\lambda-1}\exp\{\frac{-1}{2}(\xi u^{-1}+\Psi u)\}}{2\kappa_{\lambda}(\sqrt{\xi\Psi})}.$$

Para verificar que efectivamente de está forma podemos obtener la distribución hiperbólica generalizada mostrada anteriormente, sea $a = (x - \mu)'(\Sigma)^{-1}(x - \mu)$, y $b = \beta\Sigma^{-1}\beta$, entonces de aquí se sigue que:

$$\begin{aligned} f(x) &= \frac{\exp\{(x-\mu)'\Sigma^{-1}\beta\}\xi^{-\lambda}\sqrt{\xi\Psi}^{\lambda}}{(2\Pi)^{d/2}|\Sigma|^{1/2}2k_{\lambda}(\sqrt{\xi\Psi})} \\ & \int_{-\infty}^{\infty} u^{\lambda-d/2-1}\exp\{-1/2(au^{-1}+bu+\xi u^{-1}+\Psi u)\}du, \end{aligned}$$

trabajando solamente con el integrando se tiene que:

$$\begin{aligned} & \int_{-\infty}^{\infty} u^{\lambda-d/2-1}\exp\{-1/2(au^{-1}+bu+\xi u^{-1}+\Psi u)\}du \\ = & \int_{-\infty}^{\infty} u^{\lambda-d/2-1}\exp\{-1/2((a+\xi)u^{-1}+(b+\Psi)u)\}du, \end{aligned}$$

ahora, sea $\lambda' = \lambda - d/2$, $\xi' = a + \xi$, $\Psi' = b + \Psi$, entonces:

$$\begin{aligned} & \int_{-\infty}^{\infty} u^{\lambda-d/2-1}\exp\{-1/2((a+\xi)u^{-1}+(b+\Psi)u\}du \\ = & \int_{-\infty}^{\infty} u^{\lambda'-1}\exp\{-1/2(\xi' u^{-1}+\Psi' u)\}du \end{aligned}$$

luego,

$$\int_{-\infty}^{\infty} u^{\lambda'-1} \exp\{-1/2(\xi' u^{-1} + \Psi' u)\} du = \frac{2k_{\lambda'}(\sqrt{\xi'\Psi'})}{\xi'^{-\lambda'}(\sqrt{\xi'\Psi'})^{\lambda'}} \int_{-\infty}^{\infty} \frac{\xi^{-\lambda'}(\sqrt{\xi\Psi})^{\lambda'}}{2k_{\lambda'}(\sqrt{\xi'\Psi'})}$$

$$u^{\lambda'-1} \exp\{-1/2(\xi' u^{-1} + \Psi' u)\} du$$

Donde la última integral vale uno por ser una distribución $GIG(\lambda', \xi', \Psi')$ integrada sobre su soporte, por lo que tenemos que:

$$f(x) = \frac{\exp\{(x-\mu)'\Sigma^{-1}\beta\}\xi^{-\lambda}\sqrt{\xi\Psi}^{\lambda}}{(2\Pi)^{d/2}|\Sigma|^{1/2}2k_{\lambda}(\sqrt{\xi\Psi})} \frac{2k_{\lambda'}(\sqrt{\xi'\Psi'})}{\xi^{-\lambda'}(\sqrt{\xi'\Psi'})^{\lambda'}}$$

Por último, sustituyendo los valores de a, b, λ', ξ', Ψ', y agrupando algunos términos tenemos que efectivamente X se distribuye $GH_p(\lambda, \xi, \Psi, \mu, \Sigma, \beta)$.

Ahora para encontrar los momentos de la distribución GH_p, como puede ser expresada como una variable de mezcla en esparanza-varianza, la ecuación (3.8) nos dice que $\varphi_X(t) = \exp\{it\beta'\mu\}\varphi_u(it\beta' - \frac{1}{2}t\Sigma t')$, donde $\varphi_u(.)$ es la función característica del la variable de mezcla u con distribución $GIG(u|\lambda, \xi, \Psi)$.

La función característica nos permite calcular fácilmente la distribución de una transformación lineal del vector p-variado X, pues si ahora consideramos el vector aleatorio $Y = aX + b$, donde $a \in M_{rxp}$, $b \in R^p$ entonces:

$$
\begin{aligned}
\varphi_Y(t) &= exp(bit)\Phi_X(at) \\
&= \exp(bit + it\beta\mu')\Phi_u(iat\beta' - \frac{1}{2}at\Sigma at') \\
&= \exp(it(b + \beta\mu)')\Phi_u(iat\beta' - \frac{1}{2}at\Sigma at')
\end{aligned}
$$

Lo cual nos indica que Y se distribuye $GH_p(\lambda, \xi, \Psi, b + \beta\mu, a\Sigma a', a\beta)$.

De la distribución hiperbólica generalizada podemos obtener como caso particular a la distribución normal gaussiana inversa, esto es cuando de la expresión (3.9) λ toma el valor de menos un medio. Esta distribución se emplea extensivamente en McNeil et al., (2015).

En (3.10) podemos observar que λ solo aparece en la variable de mezcla u la cual se distribuye $GIG(u| - \frac{1}{2}, \psi, \xi)$ que como se mencionó en 2.6 es una distribución $GI(\lambda|\psi, \xi)$.

En el siguiente capítulo estudiamos el uso del paradigma bayesiano de inferencia para estimar la distribución normal gaussiana inversa p-dimensional.

Capítulo 4

Inferencia bayesiana en distribuciones tipo mezcla

En este capítulo se verá cómo es la función de verosimilitud de una distribución tipo mezcla continua, y la noción de estimar sus respectivos parámetros, para así plantear la necesidad de extender la verosimilitud. Por último, se mencionara como obtener la distribución final de los parámetros empleando la verosimilitud extendida.

4.1. Verosimilitud y verosimilitud extendida

Ahora, en términos generales, nos centraremos en construir la función de verosimilitud de una distribución tipo mezcla continua, para que después abordemos el problema de cómo estimar los parámetros de interés.

Supongamos que tenemos una muestra aleatoria independiente e idénticamente distribuida de tamaño n del vector aleatorio $X \in R^p$ parametrizado con algún vector de parámetros Θ, y que además la distribución de X puede ser estudiada a partir de una distribución tipo mezcla, con variable

de mezcla u, de esta manera la función de verosimilitud está dada de la siguiente forma:

$$\begin{aligned} P(X_1 &= x_1, ..., X_n = x_n|\Theta) \\ &= \prod_{i=1}^{n} P(X_i = x_i|\Theta) \\ &= \prod_{i=1}^{n} \int_u f(X_i = x_i|\Theta, u) f_u(u|\alpha) du, \end{aligned}$$

donde α son los parámetros de la distribución de mezcla u. La última expresión obtenida parece complicada a simple vista, pues hay un proceso de integración y de multiplicación en la verosimilitud de interés. Esto nos lleva a buscar una alternativa para manipular la función de verosimilitud, lo cual se logra a través de extender la verosimilitud.

4.2. Verosimilitud extendida

El proceso de extender la verosimilitud consiste en tratar al vector aleatorio X p-dimensional como una variable tipo mezcla. Este método pretende encontrar los mejores estimadores para la variable tipo mezcla u, y a su vez que ésta variable otorgue la más verosimil probabilidad de haber observado al vector X.

Por ejemplo, supongamos que X es una variable aleatoria Pareto(1,β), de la cual tenemos una realización X_0, y que nos interesa estimar el valor de β; supongamos además que X dado u y λ se distribuyen exponencial con parámetro λ y β respectivamente. Entonces, X se puede expresar como una mezcla de distribuciones exponenciales como sigue:

$$f_X(x) = \int_0^\infty f_{x|\lambda}(x) f_\lambda(\lambda) d\lambda = \int_0^\infty \lambda \exp\{-\lambda x\} \beta \exp\{-\beta\lambda\} d\lambda = \frac{\beta}{(x+\beta)^2}.$$

Ahora, si consideramos la distribución de X dado λ_i, para cada λ_i tendríamos una densidad diferente. En la siguiente figura se ilustra la densidad de X dado λ, para algunos valores de λ.

Gráfica 4.1: Densidades exponenciales con parámetro $\lambda_1 = 5$, $\lambda_2 = 10$, $\lambda_3 = 15$ y $\lambda_4 = 20$ respectivamente.

De aquí nace la siguiente pregunta: dada una observación x_0, ¿cuál sería la λ_i que le asigne mayor probabilidad o densidad a la ocurrencia de esta observación? Por ejemplo, de la gráfica anterior se puede observar que si $x_0 = 0.3$, el mejor valor de lambda sería $\lambda_1 = 5$, mientras que si $x_0 = 0$, sería $\lambda_4 = 20$. Siguiendo la misma línea de pensamiento, también sería acertado preguntarse por el mejor estimador de β, donde la probabilidad de que ocurra λ_i (que maximiza la probabilidad de ocurrencia de x_0) sea máxima. Entonces, la función de verosimilitud para la observación x_0 es:

$$Lik(\beta|X) = f_x(x) = \frac{\beta}{(x+\beta)^2} = \int_0^\infty f_{x|\lambda}(x) f(\lambda) d\lambda.$$

Y con el planteamiento anterior, en lugar de buscar el mejor estimador de β en $Lik(\beta|X_0)$, podríamos preguntarnos por la mejor λ_i inducida por β tal que β maximice la ocurrencia de λ_i, y a su vez, λ_i maximice la probabilidad de haber observado X_0. Entonces con este enfoque, la función de verosimilitud para X_0 queda expresada de la siguiente manera:

$$Lik(\lambda_i, \beta|X_0) = f_{x|\lambda_i}(x|\beta, \lambda_i) f(\lambda i).$$

Siguiendo las ideas planteadas previamente, la verosimilitud extendida se define de la siguiente forma:

Sean $\{X_i\}_{i=1}^n$ un conjunto de vectores aleatorios independientes e idénticamente distribuidos de dimensión p con algún vector de parámetros θ, y sea u una variable aleatoria con algún vector de parámetros β y con soporte en R_+, tal que X_i se puede expresar como una distribución tipo mezcla con variable de mezcla u. Entonces $f_{X_i}(x_i) = \int_0^\infty f_{x_i}(x_i|u_i) f_{u_i}(u_i) du_i$ para toda i, con X_i dado u_i condicionalmente independiente a X_j dado u_j, y u_i dado β independiente a u_j dado β para toda i diferente de j, por lo que la

verosimilitud extendida es:

$$Lik(u_i, ..., u_n, \theta, \beta | X_1, ..., X_n) = \prod_{i=1}^{n} f_{X_i}(x|\theta, u_i, \beta) f_{u_i}(u_i|\beta),$$

donde u_i es una variable latente.

Trabajar con la verosimilitud extendida podría parecer más complicado al tener que estimar una variable latente para cada observación de X, pero en ocasiones resulta más sencillo optimizar la función $\prod_{i=1}^{p} f_{x_i|u_i}(x_i) f(u_i)$ que $f(x)$ y, además, existen algoritmos computacionales, vía cadenas de Markov, o el algoritmo EM (Anexo A.2), que permiten encontrar los estimadores de los parámetros deseados. Por lo que ahora se hablará del muestreador de Gibbs, el cual será implementado posteriormente para obtener los estimadores de interés.

4.3. Muestreador de Gibbs

El muestreador de Gibbs es un un caso particular del método de simulación Monte Carlo vía cadenas de Markov (MCMC). Los métodos de simulación MCMC permiten obtener muestras de alguna densidad de probabilidad f_x mediante la construcción de una cadena de Markov que converja a dicha densidad.

El muestreador de Gibbs permite obtener una muestra de una distribución conjunta a partir de las distribuciones marginales, por lo que no es necesario conocer la distribución conjunta, basta con construir una cadena de Markov adecuada con las distribuciones marginales, las cuales convergerán a una muestra aleatoria de la distribución conjunta.

De forma intuitiva, el muestreador de Gibbs se plantea de la siguien-

te manera; supongamos que tenemos una distribución de la forma $f(x, \theta)$ donde X es un vector aleatorio y Θ algún vector de parámetors del cual se desearía tener una muestra mediante una cadena de Markov, es decir, obtener una muestra de $f(\theta^k|x)$ a partir de una sucesión de parámetros $\{\theta\}_{k=1}^{\infty}$, tal que la probabilidad de transición de esta sucesión converja a la distribución de la cual se desearía tener una muestra, esto es, $\lim_{k->\infty} P(\theta^{k+1}|\theta^k, x) = P(\theta|x)$. Dada la descripción del modelo anterior, la cadena debe ser estacionaria, irreducible, aperiódica y recurrente positiva. La propiedad de estacionariedad asegurar que $P(\theta|x)$ no cambie cuando k tienda a infinito; irreducible para no descartar ningún valor posible de θ^k en cada iteración; así como también, aperiodica para que la cadena no tenga ciclos, y así permanecer en el estado estacionario. Por último, recurrente positiva, para que se pueda visitar cualquier estado un número infinito de veces y así asegurar que la distribución límite tomó toda la información disponible en consideración.

Ahora bien, para encontrar una cadena que nos permita obtener las caractrísticas anteriores en Stuart Geman (1984) se demuestra que $\lim_{k->\infty}$ $P(\theta^k|x) = P(\theta|x)$, donde:

$$P(\theta^k|x) = \prod_{j=1}^{t} P(\theta_j|x, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_n),$$

y θ_j dado $X, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_n$ es condicionalmente independiente para toda j, para alguna medida de probabilidad arbitraria $P(.)$. Entonces si θ_j^k se distribuye $f(\theta_j^k|X, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_n)$, de aquí se puede obtener facilmente una muestra de θ_j^k para toda j, si su densidad tiene una forma cerrada, y así se construiría una cadena de Markov con las características deseadas, $\{\theta^k\}_1^{\infty}$, con probabilidad de transición $P(\theta^k|x)$.

Entonces el Muestreador de Gibbs se implementa de la siguiente forma:

1. Dar un conjunto de valores iniciales Θ_0, no importa cuáles valores se den, la propiedad de estacionareidad, asegurará la convergencia al estado estacionario. El vector θ puede ser particionado en d subconjuntos excluyentes, tal que d sea menor o igual que n, con la finalidad de simplificar los cálculos. Convenientemente se pueden seleccionar los sobconjuntos de modo que los que tienen alta correlación queden juntos.

2. Para cada θ_j obtener su distribución condicionada

$$f(\theta_j|X, ..., \theta_{0,j-1}, \theta_{0,j+1}, ..., \theta_{0,n})$$

   , y realizar una simulación de esta (a esta distribución se le conoce como distribución condicional completa). El valor de θ_j obtenido reemplazará a los valores iniciales. En caso de que no se pueda tener la distribución de $\theta|_{X,\theta\ \theta_j}$ se realiza un procedimiento de aceptación-rechazo para obtener una realización de θ_j.

3. Una vez obtenidos todos los valores de $\theta|_{X,\theta,\ \theta_j}$ se reemplazan por los valores iniciales y se repite el procedimiento hasta que la cadena se estabilice.

4.4. Algoritmo para la distribución normal gaussiana inversa

Supongamos que tenemos una colección de r_n datos, donde cada $r_i \in R^p$, de los cuales tenemos motivos para creer que fueron generados por una

distribución normal gaussiana inversa p-variada. Entonces nos interesaría inferir los valores de los parámetros de dicha distribución, pero antes de ello sería conveniente expresar a nuestra distribución de interés como una distribución tipo mezcla gaussiana, con variable de mezcla u distribuida gaussiana inversa (GI), por lo que así podemos replantear el problema, enfocándonos en buscar los parámetros de la distribución de mezcla, μ, β, Σ.

Ahora con un enfoque bayesiano, podemos plantear la verosimilitud de la distribución *a posteriori* de μ, β y Σ, es decir:

$$f(\mu, \beta, \Sigma|(r_i)_{i=1}^n) = \prod_{i=1}^n f_{r_i}(r_i|\mu, \beta, \Sigma)\Pi(\mu, \beta, \Sigma)$$

Donde $\Pi(\mu, \beta, \Sigma)$ es la distribución conjunta a *apriori* de μ, β y Σ, y además asumiremos que tanto μ, β, Σ son independientes, y que μ se distribuye $N(\mu_0, \Sigma_0)$, β se distribuye $N(\beta_0, \Sigma_\beta)$, y Σ se distribuye $IW(S, n, p)$. Así, nuestra función de verosimilitud adquiere la siguiente forma:

$$f(\mu, \beta, \Sigma|(r_i)_{i=1}^n) = \prod_{i=1}^n f_{r_i}(r_i|\mu, \beta, \Sigma)\Pi(\mu|\mu_0, \Sigma_0)\Pi(\beta|\mu_\beta, \Sigma_\beta)\Pi(\Sigma|S),$$

la cual denotaremos como:

$$Lik(\mu, \mu_0, \beta, \beta_0, \Sigma, \Sigma_0, \Sigma_\beta, S|(r_i)_{i=1}^n)$$

Luego, expresando a nuestra distribución de interés $f_{r_i}(r_i)$ como una distribución tipo mezcla gaussiana en esperanza-varianza, con variable de mezcla u, la función de verosimilitud tiene la sigueinte forma:

$$\prod_{i=1}^n \int_0^\infty N_p(r_i|\mu, u\beta, u\Sigma)GI(u|\lambda_0, \Psi_0)g(\mu, \beta, \Sigma)du,$$

donde $g(\mu, \beta, \Sigma) = N_p(\mu|\mu_0, \Sigma_0)N_p(\beta|\beta_0, \Sigma_\beta)IW_{pxp}(\Sigma|S)$.

Por lo que ahora necesitamos conocer también los valores de λ y Ψ propios de la distribución gaussiana inversa, para ello supondremos que λ y ψ son independientes, y además gracias a la gran flexibilidad de la distribución gamma, asumiremos que cada uno de ellos proviene de una distribución gamma con parámetros α_λ, β_λ, α_Ψ y β_Ψ, respectivamente, por lo que la distribución *apriori* para λ y Ψ está dada de la siguiente manera:

$$\Pi(\lambda, \Psi) = \Pi(\lambda|\alpha_\lambda, \beta_\lambda)\Pi(\Psi|\alpha_\Psi, \beta_\Psi)$$

Ahora según lo expuesto en 4.2, es conveniente extender la verosimilitud, esto es que:

$$Lik(\mu, \beta, \Sigma, (u_i)_{i=1}^n, \lambda, \Psi|(r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi) =$$

$$\prod_{i=1}^{n} N_p(r_i|\mu, u_i\beta, u_i\Sigma)GI(u_i|\lambda_0, \Psi_0)\Pi(\lambda|\alpha_\lambda, \beta_\lambda)\Pi(\Psi|\alpha_\Psi, \beta_\Psi)$$

$$N_p(\mu|\mu_0, \Sigma_0)N_p(\beta|\mu_\beta, \Sigma_\beta)W_{pxp}^{-1}(\Sigma|S, n, p)$$

Una vez teniendo nuestra función de verosimilitud, ya estamos en condiciones de proponer algún método numérico para encontrar los parámetros de interés; para ello, por un lado se utilizará el muestreador de Gibbs para obtener una muestra de los parámetros buscados; por otro lado se optará por un enfoque bayesiano para tratar la verosimilitud, ya que gracias a este método, podemos aprovechar la información implícita de los datos.

Entonces, con un enfoque bayesiano, será necesario encontrar las distribuciones condicionales completas de los parámetros de interés, para que una vez obtenidas se pueda implementar el muestreador de Gibbs.

Por lo que ahora necesitamos conocer las distribuciones condicionales completas de:

$$\Pi(\mu|\beta, \Sigma, (u_i)_{i=1}^n, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$
$$\Pi(\beta|\mu, \Sigma, (u_i)_{i=1}^n, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$
$$\Pi(\Sigma|\mu, \beta, (u_i)_{i=1}^n, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$
$$\Pi(u_i|\mu, \beta, \Sigma, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$
$$\Pi(\lambda|\mu, \beta, \Sigma, (u_i)_{i=1}^n, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$
$$\Pi(\Psi|\mu, \beta, \Sigma, (u_i)_{i=1}^n, \lambda, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$

4.5. Distribuciones condicionales completas

4.5.1. Condicional completa para μ

Para encontrar la distribución marginal de $\mu|\beta, \Sigma, (u_i)_{i=1}^n, (r_i)_{i=1}^n$ a partir de la función de verosimilitud, podemos tomar las densidades donde aparece μ e identificar su kernel, es decir, de:

$$\prod_{i=0}^{n} N_p(r_i|\mu, u_i\beta, u_i\Sigma)N_p(\mu|\mu_0, \Sigma_0) =$$
$$\prod_{i=1}^{n} \frac{exp\{-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)\}}{(2u_i\Pi)^{p/2}|\Sigma|^{1/2}}$$
$$\frac{exp\{-\frac{1}{2}(\mu - \mu_0)'\Sigma_0(\mu - \mu_0)\}}{(2\Pi)^{p/2}|\Sigma|^{1/2}}$$

tomando el cambio de variable $r_i\acute{} = r_i - u_i\beta$, nos quedamos con:

$$\prod_{i=1}^{n} exp\{-\frac{1}{2}(\mu - r_i')'(u_i\Sigma)^{-1}(\mu - r_i')\}exp\{-\frac{1}{2}(\mu - \mu_0)'\Sigma_0(\mu - \mu_0)\},$$

después, trabajando únicamente con:

$$\begin{aligned}
&\prod_{i=1}^{n} exp\{-\frac{1}{2}(\mu - r_i')'(u_i\Sigma)^{-1}(\mu - r_i')\} = \\
&exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mu - r_i')'(u_i\Sigma)^{-1}(\mu - r_i')\} = \\
&exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mu'(u_i\Sigma)^{-1}\mu - 2\mu(u_i\Sigma)^{-1}r_i' + r_i'(u_i\Sigma)^{-1}r_i')\}\alpha \\
&exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mu'(u_i\Sigma)^{-1}\mu - 2\mu(u_i\Sigma)^{-1}r_i')\} = \\
&exp\{-\frac{1}{2}(\mu'(U\Sigma)^{-1}\mu - 2\mu(U\Sigma)^{-1}R)\},
\end{aligned}$$

donde:

$$U = (\sum_{i=1}^{n}\frac{1}{u_i})^{-1}, R = \sum_{i=1}^{n}\frac{r_i}{u_i}U$$

Entonces, según el anexo (B.5), a esta parte del kernel de μ le corresponda una distribución normal p variada con vector de medias R, y matriz de varianza covarianza $U\Sigma$. Luego, multiplicando el Kernel anterior por la densidad de μ, y por el anexo (B.8), se tiene que μ se distribuye normal con vector de medias:

$$(U\Sigma^{-1} + \Sigma_0^{-1})^{-1}((U\Sigma)^{-1}R + \Sigma_0^{-1}\mu_0)$$

y matriz de varianza covarianza:

$$((U\Sigma)^{-1} + (\Sigma_0)^{-1})^{-1}$$

4.5.2. Condicional completa para β

Mediante un procedimiento similar, encontraremos el Kernel de β dado las demás variables de interés. Entonces, de la función de verosimilitud, nos quedamos con la parte donde sólo aparece β, es decir:

$$\prod_{i=0}^{n} N_p(r_i|\mu, u_i\beta, u_i\Sigma)N_p(\beta|\mu_\beta, \Sigma_\beta)$$

Luego trabajando con el producto de las n variables gaussianas p variadas correspondientes a r_i, considerando los términos que únicamente dependen de β, y con el cambio de variable $\mu_i = r_i - \mu$, tenemos que:

$$\prod_{i=1}^{n} exp\{-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)\} =$$

$$exp\{-\frac{1}{2}\sum_{i=1}^{n}((u_i\beta - \mu_i)'(u_i\Sigma)^{-1}(u_i\beta - \mu_i))\}$$

Nuevamente desarrollando el exponente de la función exponencial, y quedándonos únicamente con la parte que depende de β tenemos que:

$$exp\{-\frac{1}{2}\sum_{i=1}^{n}((u_i\beta)'(u_i\Sigma)^{-1}(u_i\beta) - 2(u_i\beta)'(u_i\Sigma)^{-1}\mu_i)\} =$$

$$exp\{-\frac{1}{2}(\beta(V\Sigma)^{-1}\beta - 2\beta(V\Sigma)^{-1}VM)\}$$

Donde:

$$V = \frac{1}{\sum_{i=1}^{n} u_i}, M = \sum_{i=1}^{n}(r_i - \mu)$$

Por lo tanto, según el anexo (B.5), el Kernel del producto de las n variables normales p-variadas correspondientes a β corresponde a una distribución gaussiana con vector de medias VM, y matriz de varianza covarianza $V\Sigma$. Luego, según el anexo (B.8), juntando el Kernel obtenido anteriormente con la densidad de β, se tiene que β se distribuye normal p variado con vector de medias:

$$((V\Sigma)^{-1} + \Sigma_\beta^{-1})^{-1}((V\Sigma)^{-1}VM + \Sigma_\beta^{-1}\mu_\beta)$$

Y matriz de varianza covarianza:

$$((V\Sigma)^{-1} + (\Sigma_\beta)^{-1})^{-1}$$

4.5.3. Condicional completa para Σ

Para encontrar la distribción marginal de Σ a partir de la función de verosmilitud, podemos tomar las densidades donde aparece σ, y a partir de ahí, identificar el Kernel correspondiente, es decir, tomando:

$$\prod_{i=0}^{n} N_p(r_i|\mu, u_i\beta, u_i\Sigma)W_{p\times p}^{-1}(\Sigma|S, n, p) \tag{4.1}$$

Y luego trabajando con el producto de las n variables normales p-variadas correspondientes a r_i, considerando los términos que únicamente dependen

de Σ, y con el cambio de variable $(x_i = r_i - \mu - \beta)/\sqrt{u_i}$, tenemos que:

$$
\begin{aligned}
& \prod_{i=1}^{n} exp\{-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)\} \\
& \frac{\exp\{-\frac{1}{2}tr(\Sigma^{-1}S)\}}{|\Sigma|^{n/2}} \\
= & \frac{\exp\{-\frac{1}{2}\sum_{i=1}^{n} x_i\Sigma^{-1}\sum_{i=1}^{n} x_i\}}{|\Sigma|^{\frac{n}{2}}} \frac{\exp(-\frac{1}{2}tr(\Sigma^{-1}S))}{|\Sigma|^{\frac{n}{2}}} \\
= & \frac{\exp\{-\frac{1}{2}(X\Sigma^{-1}X + tr(\Sigma^{-1}S))\}}{|\Sigma|^{n}} \\
= & \frac{\exp\{-\frac{1}{2}(tr(\Sigma^{-1}XX') + tr(\Sigma^{-1}S)))\}}{|\Sigma|^{n}} \\
= & \frac{\exp\{-\frac{1}{2}(tr(\Sigma^{-1}(XX' + S)))\}}{|\Sigma|^{n}}
\end{aligned}
$$

Haciendo $X = \sum_{i=1}^{n} x_i$, y ya que $X\Sigma^{-1}X = tr(\Sigma^{-1}(XX'))$. La última expresión corresponde al kernel de una distribución Whishart con matriz de escala $XX' + S$, y $2n$ grados de libertad.

4.5.4. Condicional completa para λ

Para encontrar la distribción marginal de λ a partir de la función de verosmilitud, podemos tomar las densidades donde aparece λ, y a partir de ahí, identificar el Kernel correspondiente; es decir, de la sigueinte expresión:

$$\prod_{i=1}^{n} GI(u_i|\lambda, \Psi)\Pi(\lambda|\alpha_\lambda, \beta_\lambda)$$

nos quedamos con:

$$\prod_{i=1}^{n} \sqrt{\lambda} \exp\{-\frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\} \lambda^{\beta_\lambda - 1} \exp\{-\lambda\alpha_\lambda\}\alpha$$

$$\lambda^{n/2+\beta_\lambda - 1} \exp\{-\lambda X\}$$

donde $X = \sum_{i=1}^{n} \frac{(u_i - \Psi)^2}{2\Psi^2 u_i} + \alpha_\lambda$. Por lo que tenemos que el kernel correponde a una distribución gamma, concluimos que λ se distribuye gamma con parámetros $(\frac{n}{2} + \beta_\lambda, X + \alpha_\lambda)$.

4.5.5. Condicional completa para Ψ

Para encontrar la distribción marginal de Ψ a partir de la función de verosmilitud, podemos tomar las densidades donde aparece Ψ, y a partir de ahí identificar el Kernel correspondiente; es decir, de la sigueinte expresión:

$$\prod_{i=1}^{n} GI(u_i|\lambda, \Psi)\Pi(\Psi|\alpha_\Psi, \beta_\Psi)$$

nos quedamos con:

$$\prod_{i=1}^{n} \exp\{-\frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\} \Psi^{\beta_\Psi - 1} \exp\{-\Psi\alpha_\Psi\}\alpha$$

$$\exp\{-\sum_{i=1}^{n} \frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\} \Psi^{\beta_\Psi - 1} \exp\{-\Psi\alpha_\Psi\}.$$

El Kernel obtenido resulta difícil de identificar, y tal vez sea complicado obtener la densidad condicional de Ψ necesaria para implementar el muestreador de GIBBS; por lo que se atacará este problema por medio de una herramienta útil en simulación, llamada slice sampler (apéndice C.2); la

ventaja de utilizar el slice sampler es que nos permite obtener una muestra aleatoria de algúna distribución de probabilidad, con tan solo conocer su Kernel.

4.5.6. Condicional completa para u_i

Para encontrar la distribción marginal de u_i a partir de la función de verosmilitud, podemos tomar las densidades donde aparece u_i, y a partir de ahí identificar el Kernel correspondiente, es decir, de la sigueinte expresión:

$$\prod_{i=1}^{n} N_p(r_i|\mu, u_i\beta, u_i\Sigma)GI(u_i|\lambda, \Psi)$$

nos quedamos con:

$$\frac{exp\{-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)\}}{(u_i)^{p/2}}\sqrt{\frac{1}{u_i^3}}$$
$$\exp\{-\frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\}.$$

EL kernel anterior resulta difícil de identificar, e intentar encontrar la forma analítica de la distribución condicionada de u_i puede resultar muy difícil, si no es que imposible, por lo que optaremos por obtener una muestra aleatoria de u_i por medio del slice sampler (apéndice C.2).

Una vez descrita la metodología a emplear para inferir los parámetros de nuesta muestra de interés, ya estamos en condiciones de programar el muestreador de Gibbs y emplearlo para estudiar un conjunto de datos simulados, lo cual se llevará acabo en el siguiente capítulo.

4.6. Alternativa para la matriz de varianza-covarianza

En la sección anterior, por terminos prácticos, supusimos que la distribución a priori de la matriz de varianza-covarianza Σ sigue una distribución Whishart inversa. Pues de esta manera podemos utilizar el hecho de que la distribución gaussiana p variada y la distribución Whishart inversa son distribuciones conjugadas, es decir, la distribución posterior pertenece a la misma familia que la distribución a priori. Lo que tal vez no sea muy adecuado, pues como se menciona en el anexo D, existen varias complicaciones al trabajar con esta distribución. Dicho lo anterior, se propone como alternativa al supuesto de utilizar la distribución Whishart inversa como distribución a priori de la matriz de varianza-covarianza, utilizar la descomposición $diag(S)Rdiag(S)$, como en John Barnard (2000). De esta manera podemos obtener una distribución marginal completa para cada elemento de la matriz de varianza-covarianza de la siguiente manera:

Sustituyendo la distribución a priori inversa Wishart de 4.1 por $\Pi(R)$ $\Pi(S)$ D, donde $\Pi(R)$ tiene la distribución D.2, y $log(S)$ se distribuye gaussiana con vector de medias μ_s , y con matriz de varianza covarianza la matriz identidad de dimensión p; tenemos que el kernel que nos interesa es el siguiente:

$$\prod_{i=0}^{n} N_p(r_i|\mu, u_i\beta, u_i diag(S)Rdiag(S))\Pi(R)N(S|\mu_s, I_{p\times p}),$$

de donde podemos obtener las respectivas densidades marginales como se describe a continuación.

4.6.1. Condicional completa para s_i

Si de la expresión 4.6 tomamos los términos en donde sólo aparece S nos quedamos con:

$$\prod_{i=1}^{n} \frac{exp\{-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)\}}{(2u_i\Pi)^{p/2}|\Sigma|^{1/2}}$$
$$\frac{exp\{-\frac{1}{2}(log(S) - \mu_s)'I_{p\times p}(log(S) - \mu_s)\}}{(2\Pi)^{p/2}|I_{p\times p}|^{1/2}}$$
,

Luego por independencia podemos tomar el i-ésimo componente de la distribución de S, por lo que tenemos que la expresión anterior es proporcional a:

$$\frac{1}{s_i^n} exp\{-\frac{1}{2}\sum_{i=1}^{n} x_i'(diag(S)Rdiag(S))^{-1}x_i\}exp\{-\frac{1}{2}(log(s_i) - \mu_{s_i})^2\}, \quad (4.2)$$

donde $x_i = (r_i - \mu - u_i\beta)/\sqrt{ui}$, y ya que:

$$\begin{aligned} |\Sigma| &= |diag(S)Rdiag(S)| \quad (4.3) \\ &= |diag(S)||R||diag(S)| \\ &\propto |diag(S)|^2 \\ &= (\prod_{k=1}^{p} s_k)^2 \\ &\propto s_i^2 \end{aligned}$$

De la expresión 4.2 resulta difícil identificar si el kernel pertenece a alguna distribución conocida, por lo que se utilizará el slice sampler (apéndice C.2) para obtener una muestra aleatoria de esta distribución.

4.6.2. Condicional completa para r_{ij}

Si de la expresión 4.6 tomamos los términos en donde sólo aparece R nos quedamos con:

$$\prod_{i=1}^{n} \frac{exp\{-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)\}}{(2u_i\Pi)^{p/2}|\Sigma|^{1/2}}$$
$$|R|^{-1/2(v-(p-j)-1)(j-1)-1}(\Pi_{i=1}^{j}|R_{ii}^{n}|)^{-v/2}.$$

Si usamos el cambio de variable $x_i = (r_i - \mu - u_i\beta)/\sqrt{ui}$, y el hecho de que $|\Sigma| \propto |R|$, la expresión anterior es proporcional a:

$$\frac{1}{|R|^{n/2}} exp\{-\frac{1}{2}\sum_{i=1}^{n} x_i'(diag(S)Rdiag(S))^{-1}x_i\} \tag{4.4}$$
$$|R|^{-1/2(v-(p-j)-1)(j-1)-1}(\Pi_{i=1}^{j}|R_{ii}^{n}|)^{-v/2}.$$

Si la expresión 4.4 se ve como función de r_{ij}, se puede obtener una muestra aleatoria para todo i diferente de j gracias al slice sampler.

En este capítulo se habló de cómo construir el modelo y algoritmo que nos permitirá obtener una muesta de los parámetros de interés. Mediante la obtención de la muestra se pueden calcular los estimadores deseados al proponer una función de pérdida. En este trabajo se utilizará la función de

pérdida cuadrática, y mediante la expresión C.2 obtendrémos los estimadores puntuales bayesianos.

En el siguiente capítulo haremos un ejercicio de simulación donde se aplique lo visto hasta ahora. Para ello, se propondrán dos modelos, uno donde se aplique la descomposición $Diag(S)RDiag(S)$ (John Barnard (2000)), y otro donde se use la distribución condicional completa para Σ, (4.5.3).

Capítulo 5

Ejercicio de simulación

En este capítulo se muestran los resultados de aplicar el algoritmo descrito en 4.4, a un conjunto de mil datos simulados provenientes de una distribución normal gaussiana inversa. Estos datos fueron simulados mediante el uso de una distribución tipo mezcla normal en esperanza-varianza (B.1), donde X es un vector aleatorio de dimensión 3, con distribución normal con vector de medias $\mu + u\beta$, y matriz de varianza-covarianza $u\Sigma$; con variable de mezcla u distribuida gaussiana inversa con parámetros λ, Φ, donde:

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} .0001 & .00000009999 & .00000009999 \\ .00000009999 & .0001 & .0000001 \\ .00000009999 & .0000001 & .0001 \end{pmatrix}, \qquad (5.1) \\
\mu &= (\text{-}0.0001127,\ 0.0005493,\ 0.0002965), \\
\beta &= (\text{-}0.0000064, \text{-}0.00003.646, \text{-}.00001876), \\
\lambda &= 14.34, \\
\Phi &= 1.073.
\end{aligned}
$$

Con fines experimentales se realizaron dos ejercicios. Primero se supuso que no se conocia nada acerca de Σ, por lo que se utilizó la descomposición $Diag(S)RDiag(S)$ descrita en el capítulo 4.6 (Modelo 1). Después se hizo el supuesto de que la distribución a priori de la matriz de varianzas covarianzas Σ fue inversa Wishart (Modelo 2).

5.1. Modelo 1

Para este modelo se supuso que las distribuciones a priori fueron las siguientes:

$$
\begin{aligned}
r_i &\sim N(r_i|\mu + u_i\beta, u_i\Sigma), \\
\mu &\sim N(\mu|(.00001, .00001, .00001), \Sigma_0), \\
\beta &\sim N(\beta|(.00001, .00001, .00001), \Sigma_0), \\
u_i &\sim GI(u_i|\lambda_0, \Phi_0), \\
\lambda_0 &\sim \Gamma(\lambda_0|0, 1), \\
\Phi_0 &\sim \Gamma(\Phi_0|0, 1),
\end{aligned} \tag{5.2}
$$

donde $\Sigma_0 = \begin{pmatrix} .00001 & 0 & 0 \\ 0 & .00001 & 0 \\ 0 & 0 & .00001 \end{pmatrix}$.

Para Σ se utilizó la descomposición $Diag(S)RDiag(S)$, donde r_{ij} es la correlación de la variable i con la variable j, y su distribución conjunta es proporcional a:

$$
f_3(r_{12}, r_{13}, r_{23}|1000) \propto \frac{(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})^{998}}{((1 - r_{12}^2)(1 - r_{13}^2)(1 - r_{23}^2))^{500}}
$$

Mientras que s_i es la desviación estándar de la variable i, y su logaritmo se distribuye como $log(s_i) \sim N(s_i|0, 1)$.

Luego aplicando el muestreador de Gibbs a las distribuciones marginales completas obtenidas en 4.4, y mediante la función de pérdida cuadrática descrita en 1.5.1, obtenemos que los respectivos estimadores para μ, β, Σ, λ y Φ son:

$$
\begin{aligned}
\widehat{\Sigma} &= \begin{pmatrix} .0001 & .0000001005 & .0000001004 \\ .0000001005 & .0001 & .0000001008 \\ .0000001004 & .0000001008 & .0001 \end{pmatrix}, \qquad (5.3) \\
\widehat{\mu} &= (0.0002288,\ \text{-}0.0002307,\ 0.0007267), \\
\widehat{\beta} &= (\text{-}0.0000687, \text{-}0.00007141,\ \text{-}.00003133), \\
\widehat{\lambda} &= 7.889, \\
\widehat{\Phi} &= 1.108.
\end{aligned}
$$

Para cada estimador, mediante el muestreador de Gibbs, se obtuvieron muestras de quince mil simulaciones, donde no se tomo en consideración las primeras dos mil quinientas. Pues para efectos prácticos, como el proceso no converge en las primeras simulaciones, es natural creer que tales valores no pertenecen aún a la muestra de la cual se desea obtener un estimador. Después mediante el método de integración montecarlo (apéndice C) se obtuvo un estimado de la esperanza de la función de pérdida cuadrática, es decir, del estimador de interés.

En los gráficos $G1-G5$ del apéndice E se puede observar el valor de diez ejercios diferentes de los promedios ergódicos de cada parámetro a estimar.

Los promedios ergódicos se definen como los promedios aritméticos parciales, es decir, como:

$$\widehat{X}_m = \frac{\sum_{i=1}^{m} x_i}{m}. \qquad (5.4)$$

Es importante notar que para cada valor de m los promedios ergódicos representan un estimador construido con m observaciones, por lo que cuando m crece, podemos observar la convergencia del parámetro a algún valor. En nuestro caso el valor al que converge el promedio ergódico es nuestro

estimador de interés. Por simplicidad se ha denotado a m_i como el i-ésimo elemento del vector de medias μ, a b_i como el i-ésimo elemento del vector de tendencia β, a λ y Φ como a los parámetros de la distribución gaussiana inversa, a s_i como la varianza de la i-ésima variable, y a r_{ij} como a la covarianza de la variable i con la variable j.

En los gráficos $G6 - G10$ del apéndice E se puede observar la estabilización de varianza (apéndice C.4)de los estimadores obtenidos.

5.2. Modelo 2

Para este modelo se supuso que las distribuciones a priori fueron las siguientes:

$$\begin{aligned}
r_i &\sim N(r_i|\mu + u_i\beta, u_i\Sigma), \\
\mu &\sim N(\mu|(.00001, .00001, .00001), \Sigma_0), \\
\beta &\sim N(\beta|(.00001, .00001, .00001), \Sigma_0), \\
u_i &\sim GI(u_i|\lambda_0, \Phi_0), \\
\lambda_0 &\sim \Gamma(\lambda_0|0, 1), \\
\Phi_0 &\sim \Gamma(\Phi_0|0, 1), \\
\Sigma &\sim WI(\Sigma|\Sigma_0, 2000, 3),
\end{aligned} \tag{5.5}$$

donde $\Sigma_0 = \begin{pmatrix} .00001 & 0 & 0 \\ 0 & .00001 & 0 \\ 0 & 0 & .00001 \end{pmatrix}$.

Luego aplicando el muestreador de Gibbs a las distribuciones marginales completas obtenidas en 4.4, y mediante la función de pérdida cuadrática descrita en 1.5.1, obtenemos que los respectivos estimadores para μ, β, Σ, λ y Φ son:

$$
\begin{aligned}
\widehat{\Sigma} &= \begin{pmatrix} 0.0004876 & 0.00001052 & 0.00001404 \\ 0.00001052 & 0.0005002 & \text{-}0.000001617 \\ 0.00001404 & \text{-}0.000001617 & 0.0004817 \end{pmatrix} \qquad (5.6) \\
\widehat{\mu} &= (0.0001277,\ \text{-}0.0003272,\ 0.0000726), \\
\widehat{\beta} &= (.00001368, 0.000001996, \text{-}0.000009617), \\
\widehat{\lambda} &= 0.7985, \\
\widehat{\Phi} &= 74.31.
\end{aligned}
$$

Para cada estimador, mediante el muestreador de Gibbs, se obtuvieron muestras de quince mil simulaciones, donde se tiraron las primeras dos mil quinientas. Pues para efectos prácticos, como el proceso no converge en las primeras simulaciones, es natural creer que tales valores no pertenecen aún a la muestra de la cual se desea obtener un estimador. Después mediante el método de integración montecarlo (apéndice C) se obtuvo un estimado de la esperanza de la función de pérdida cuadrática, es decir, del estimador de interés.

Análogamente al Modelo 1, en los gráficos $G11 - G15$ del apéndice E se observa el valor de los promedios ergódicos de los parámetros del modelo 2, mientras que en los gráficos $G16 - G20$ se pueden observar los gráficos de estabilización de varianza.

5.3. Resultados y consideraciones

En la siguiente tabla se muestran los parámetros originales con los que fueron simulados los datos (Valores), así como los resultados obtenidos por

los dos modelos.

Cuadro 5.1

Parámetros	**Valores**	**Modelo** 1	**Modelo** 2
μ_1	-0.0001127	0.0002288	0.0001277
μ_2	0.0005493	-0.0002307	-0.0003272
μ_3	0.0002965	0.0007267	0.0000726
β_1	-0.0000064	-0.0000687	0.00001368
β_2	-0.00003646	-0.00007141	0.000001996
β_3	-0.00001876	-0.00003133	-0.000009617
λ	14.34	7.889	0.7985
Φ	1.073	1.008	74.31
σ_1^2	0.0001	0.0001	0.0004876
σ_2^2	0.0001	0.0001	0.0005002
σ_3^2	0.0001	0.0001	0.0004817
σ_{12}	0.00000009999	0.0000001004	0.00001052
σ_{13}	0.0000000999	0.0000001004	0.00001404
σ_{23}	0.0000001	0.0000001008	-0.000001617

Para cada modelo obtenido generamos una muestra de mil simulaciones de un vector aleatorio X, de dimensión 3, con distribución tipo mezcla gaussiana en esperanza-varianza.

En la gráfica 5.1 se presentan los diagramas de densidad marginales de las simulaciones del vector $X = (x_1, x_2, x_3)$. Mientras que en las gráficas 5.2-5.4 se presentan los diagramas de dispersión.

En la gráfica 5.1 se puede observar que marginalmente el modelo 1 simula mejor los datos originales que el modelo 2.

En los gráficos de dispersión (5.2-5.4) podemos observar que los valores simulados del modelo 2 están más dispersos que los datos del modelo original. También se puede notar mayor semejanza de los datos simulados

con el modelo 1 con el modelo original.

Si nos enfocamos únicamente en las distribuciones marginales del vector X el modelo 1 simula mucho mejor a los datos que el modelo 2. Pero esto no significa que el modelo dos sea ineficiente, pues hay que considerar que este se basa en el supuesto de que la distribución a priori de la matriz de varianza-covarianza sigue una distribución inversa Wishart, lo cual en nuestro caso es un supuesto que afecta significativamente al resultado. Pero, en algún otro contexto este supuesto podría ser más acertado que suponer desconocimiento alguno en esta distribución apriori.

También, en nuestro ejercicio de simulacón resultó mejor opción el modelo 1, sin embargo faltaría contar con algún criterio que nos permita evaluar la calidad de los estimadores obtenidos.

Otro punto a considerar es que dependiendo de la naturaleza de los datos se deben ajustar los supuestos acerca de los parámetros de las distribuciones apriori.

40
30
20
10
0
-0.025
0.000
0.025
0.050
X1
Modelo
1
2
Original
40
30
20
10
0
-0.025
0.000
0.025
0.050
X2
Modelo
1
2
Original
40
30
20
10
0
-0.025
0.000
0.025
0.050
X3
Modelo
1
2
Original

Gráfica 5.1

Modelo original
0.050
0.025
0.000
-0.025
-0.050
X2
-0.025
0.000
0.025
0.050
X1
Modelo 1
0.050
0.025
0.000
-0.025
-0.050
X2
-0.025
0.000
0.025
0.050
X1
Modelo 2
0.050
0.025
0.000
-0.025
-0.050
X2
-0.025
0.000
0.025
0.050
X1

Gráfica 5.2

Modelo original
0.050
0.025
0.000
-0.025
-0.050
X3
-0.025
0.000
0.025
0.050
X1
Modelo 1
0.050
0.025
0.000
-0.025
-0.050
X3
-0.025
0.000
0.025
0.050
X1
Modelo 2
0.050
0.025
0.000
-0.025
-0.050
X3
-0.025
0.000
0.025
0.050
X1

Gráfica 5.3

Modelo original
0.050
0.025
0.000
-0.025
-0.050
X3
-0.025
0.000
0.025
0.050
X2
Modelo 1
0.050
0.025
0.000
-0.025
-0.050
X3
-0.025
0.000
0.025
0.050
X2
Modelo 2
0.050
0.025
0.000
-0.025
-0.050
X3
-0.025
0.000
0.025
0.050
X2

Gráfica 5.4

Capítulo 6

Conclusiones

A lo largo de esta tesis se introdujeron de manera breve los principios, notación y la metodología concerniente a la obtención de estimadores puntuales bayesianos, pues se optó por este enfoque para llegar a los resultados deseados. De igual manera su hubiera podido trabajar con un enfoque frecuentista mediante la sustitución de los métodos MCMC utlizados, por el método EM. Este segundo enfoque no se implementó, por lo que podría ser interesante ver cómo son los estimadores obtenidos bajo este, y que simulitudes o diferencias tendrían respecto a los estimadores bayesianos.

En esta tesis estudiamos los modelos de mezcla gaussiana en esperanza varianza, y en cómo expresar la distribución hiperbólica generalizada a través de estos. Para después trabajar con un la distribución normal gaussian inversa, la cual resulta ser un caso particular de la distibución hiperbólica generalizada. Se pueden obtener otros casos partículares de la distribución hiperbólica generalizada, al considerar algún parámetro de esta constante (se puede ahondar más de estos casos en McNeil et al., 1967), por lo que también podría ser interesante trabajar con algún otro caso particular de esta distribución, y comparar los resultados con los obtenidos a

lo largo de este trabajo.

Una vez que mostramos cómo expresar la distribución normal gaussiana inversa en términos de una distribución tipo mezcla gaussiana en esperanza-varianza, pasamos a construir el modelo para la estimación de sus parámetros. Para ello nos encontramos con la dificultad de un proceso de integración inherente a la función de verosimilitud de la distribución tipo mezcla construida. Tal problema fue resuelto mediante la incorporación de variables latentes, las cuales se adaptaron de manera natural a la función de verosimilitud planteada.

Con respecto a la matriz de varianza covarianza propia del modelo construido, primero se utilizó la distribución inversa Wishart para su modelación, esta alternativa tiene la desventaja de que pondera por igual la variabilidad de cada variable. Por lo que, con fines de tener mayor libertad en la modelación de la matriz de varianza covarianza se aplicó la descomposición $diag(S)Rdiag(S)$ John Barnard (2000), ya que de esta manera se pueden hacer supuestos sobre la variabilidad de cada variable, e incorporar está variablilidad mediante algúna función a priori. Aunque en el ejercicio de simulación el modelo que considera la descomposición $diag(S)Rdiag(S)$ parece simualr los datos de manera eficiente, en este trabajo no nos inclinamos a creer que algún modelo es mejor que el otro, sino que depende completamente de la naturaleza del problema.

Un problema que suele aparecer al simular la distribución hiperbólica generalizada mediante distribuciones tipo mezcla gaussiana en esperanza-varianza, es el problema de identificabilidad, el cual puede solucionarse al hacer igual a algúna constante el determinante de la matriz de varianza covarianza, para ello se propone alguna reparametrización conveniente, como dividir la matriz de varianza covarianza entre su determinante. El proble-

ma de identificabilidad no ha sido tratado en el presente trebajo, pero en caso de presentarse puede solucionarse mediante la reparametrización de la matriz de varianza-covarianza.

Apéndice A

Estimación por máxima verosimilitud y algoritmo EM

A.1. Estimación por máxima verosimilitud

El método de máxima verosimilitud tiene un enfoque frecuentista, en el que los datos son generados por una función de densidad de probabilidad $f_x(.)$, con esta función es posible construir una función de verosimilitud, la cual tiene como argumento los parámetros a estimar, y a su vez depende de los datos ya observados. Por lo regular la función de verosimilitud se construye de acuerdo al evento, concerniente a los datos de interés, por lo que muchas veces es la misma función de densidad, pues se considera el evento en que la variable aleatoria X_i tome el valor x_i. De manera intuitiva, la función de verosimilitud representa la probabilidad de haber observado la variable aleatoria X bajo el modelo dado, la cual se pretende maximizar con respecto al parámetro de interés para así obtener qué tan probable es que el modelo dado haya generado los datos.

En términos generales los estimadores máximo verosimiles se construyen de la siguiente forma:

1. Construir la función de verosimilitud, la cual llamaremos $LIK(\Theta|X)$.
2. Plantear el problema de maximización:

$$max_{\Theta}(LIK(\Theta|X)).$$

3. La solucion $\Theta*$ del problema de maximización es el estimador buscado.

Aunque el algoritmo para encontrar estimadores máximo verosimiles parece sencillo, no siempre lo es, pues la dificultad del proceso de maximización aumenta conforme más variables se tienen, lo cual puede llevar a procesos computacionales poco eficientes. Estas dificultades incentivan el uso del algoritmo EM.

A.2. Algoritmo EM

El algoritmo Expectation Maximization (EM) fue formalizado por Dempster et al., (1977); este algoritmo tiene un enfoque frecuentista, y tiene como objetivo encontrar los estimadores máximo verosímiles de funciones de densidad con datos no observados, lo cual es ideal para manejar distribuciones tipo mezcla, pues la variable de mezcla toma el lugar de los datos no observados.

Para implementar el algoritmo EM se requiere un conjunto de datos u observaciones x_i, de las cuales se conoce su función paramétrica de densidad, el parámetro de dicha distribución, Θ, es desconocido y es lo que se pretende estimar; también se requiere la distribución, parametrizada por Θ, de los datos no observados y finalmente una suposición inicial, Θ^0, sobre

el parámetro Θ.

De manera intuitiva, el algorimo EM se desarrolla de la siguiente forma: llamemos a la función de densidad asociada a x_i como $f(x_i|\Theta)$, después supongamos que existen datos no observados z, y una función determinista $H(.)$ que tiene como dominio los datos no observados y como imagen los datos observados. Esto es, que para cada dato no observado se cumple que $H(Z_s) = x_i$, para Z_s un subconjunto de Z. Luego nos interesaría encontrar el valor de Θ que maximiza la probabilidad de haber obtenido los datos no observados dado Θ, pero justamente no conocemos los datos no observados Z, por lo que la probabilidad anterior la ponderamos por la probabilidad de haber obtenido los datos no observados dados los datos sí observados y la suposición inicial del parámetro de interés, luego nos interesa maximizar $f(Z|\Theta)f(Z|X,\Theta^0)$, lo cual no nos libra del problema de los datos no observados. Entonces, nos fijamos en un promedio de todas las posibles ocurrencias de Z, por lo que nuestra función a maximizar con respecto a Θ se convierte en:

$$E_Z[f(Z|\Theta)] = \int_Z f(Z|\Theta)f(Z|X,\Theta^0)dz$$

Por último, si se define a la esperanza descrita previamente como una función de Θ, es decir, que $Q(\Theta|\Theta^0) = E_Z[f(Z|\Theta)]$, entonces el problema planteado se resume en dado un valor Θ^0, maximizar con respecto a Θ la función $Q(\Theta|\Theta^0)$. De esta forma se crea un proceso iterativo, donde el valor Θ^* que maximiza la función $Q(.)$ se convierte ahora en Θ^0, y el procedimiento se repite de nuevo; de esta forma se asegura que la esperanza converge a la verosimilitud buscada, y a su vez, Θ^* converge al estimador máximo verosímil.

En términos generales, el estimador máximo verosímil a través del al-

goritmo EM se construye de la siguiente manera:

1. Dar un valor inicial Θ^0.

2. Obtener las funciones de densidad $f(Z|\Theta)$, $f(Z|X, \Theta^0)$.

3. Calcular $Q(\Theta|\Theta^0) = E_Z[f(Z|\Theta)]$, algunas veces resulta más sencillo calcular $E_Z[log(f(Z|\Theta))]$.

4. Maximizar $Q(\Theta|\Theta^0)$.

5. Una vez obtenido Θ^* sustituir por Θ^0, y repetir los pasos 3, 4 y 5, hasta que el algoritmo converja.

Apéndice B

Distribuciones tipo mezcla: Resultados adicionales

B.1. Simulación de distribuciones tipo mezcla gaussiana en esperanza varianza

Para simular distribuciones tipo mezcla gaussiana en esperanza varianza basta con seguir los siguientes pasos:

1. Proponer valores para el vector de medias μ, el vector de sesgo β, y la matriz de varianza-covarianza Σ.

2. Simular una realización de una variable aleatoria con soporte en los reales positivos, esta será nuestra variable de mezcla u.

3. Realizar una simulación de X, donde X se distribuye gaussiana con vector de medias $\mu + u\beta$ y matriz de varianza-covarianza Σ.

4. Realizar los pasos (2) y (3) hasta obtener la muestra deseada.

B.2. Distribución tipo mezla gaussiana p-variada

Se dice que el vector aleatorio $X \in R^p$ con densidad $f(x)$ puede ser expresado como una distribución de mezcla gaussiana $N(x|\mu, \beta, \Sigma)$ con variable de mezcla u con densidad $f(u)$ y soporte en R^+, si:

$$f(X|\mu, \beta, \Sigma) = \int_0^\infty N_p(X|\mu, u\beta, u\Sigma) f_u(u) du,$$

donde $N_p(x|\mu, u\beta, u\Sigma)$ es una distribución gaussiana p-variada con vector de medias $\mu + u\beta$, y matriz de varianza-covarianza $u\Sigma$.

B.3. Covarianza de un vector aleatorio p-variado

Se define la covarianza de un vector aleatorio p variado como:

$$COV(X) = E((X - \mu)(X - \mu)')$$

Siempre y cuando el vector de medias (del vector aleatorio X) μ exista.

B.4. Covarianza de una distribución tipo mezcla

Sea X dado u una distribución tipo mezcla de dimensión p, y u una variable de mezcla con soporte en R_+, entonces la covarianza de X se puede calcular como:

$$COV(X) = E_u(COV(X|u)) - COV_u(E(X|u))$$

De la definición de $COV(X)$ tenemos que:

$$
\begin{aligned}
COV(X) &= E_x((X-\mu)(X-\mu)') \\
&= E_x(XX' - 2X\mu + \mu\mu') \\
&= E_x(XX') - 2E(X)\mu + \mu\mu' \\
&= E_x(XX') - E_x(X)E_x(X)' \\
&= E_x(XX') - E(X)_x E_x(X)' + E(X)_x E(X)'_x \\
&- 2E(X)_x E(X)'_x + E(X)_x E(X)_x \\
&= E_u(E_x(XX'|u) - E_x(X|u)E_x(X|u)') \\
&+ E_u(E_x(X|u)E_x(X|u)') - 2E_u(E_x(X|u)E(X)') \\
&+ E_u(E(X|u))E_u(E(X|U)) \\
&= E_u(E_x(XX'|u) - 2E_x(X|u)E_x(X|u)' \\
&+ E_x(X|u)E_x(X|u)') + E_u(E_(X|u)E_x(X|u)' \\
&- 2E_x(X|u)E_u(E_x(X|u)) \\
&+ E_u E_x(X|u)E_u E_x(X|u)') \\
&= E_u(E_x(XX' - 2XE_x(X|u) \\
&+ E_x(X|u)E_x(X|U)) \\
&+ E_u((E_x(X|u) - E_u(E_x(X|u))) \\
&\quad (E_x(X|u) - E_u(E_x(X|u)))') \\
&= E_u((E_x(X) - E_x(X|u))(E_x(X) - E_x(X|u))') \\
&+ COV_u(E_x(X|u)) \\
&= E_u(COV_x(X|u)) + COV_u(E_x(X|u)).
\end{aligned}
$$

B.5. Kernel de una distribución gaussiana p variada

El kernel un vector aleatorio es la parte de la función de densidad que únicamente depende de dicho vector, y a su vez nos permite identificar de que familia proviene la distribución. Por ejemplo, en el caso de la distribución gaussiana p variada

$$\begin{aligned} f(x|\mu,\Sigma) &= \frac{1}{(2\Pi)^{p/2}|\Sigma|^{1/2}} \\ & \quad exp(-\frac{1}{2}(x'\Sigma^{-1}x - x'\Sigma^{-1}\mu - \mu'\Sigma^{-1}x + \mu'\Sigma^{-1}\mu)) \end{aligned}$$

Tenemos que el Kernel correspondiente es:

$$exp(-\frac{1}{2}(x'\Sigma^{-1}x - 2x'\Sigma^{-1}\mu))$$

donde $\mu'\Sigma^{-1}x = x'\Sigma^{-1}\mu$, ya que es una forma cuadrática.

B.6. Kernel de una distribución Wishart

Análogamente al caso gaussiana p variado, si nos concentramos en la parte de la densidad Wishart, con matriz de escala S y con n grados de libertad, que únicamente depende de Σ, tenemos que el correspondiente kernel es:

$$\frac{\exp(-\frac{1}{2}tr(\Sigma^{-1}S))}{|\Sigma|^{\frac{n}{2}}}$$

B.7. Kernel del producto de n distribuciones gaussianas p variadas con mismo vector de medias y misma matriz de varianza covarianza

Supongamos que X_i se distribuye gaussiana p variada con vector de media μ y matriz de varianza covarianza Σ, entonces la función de interés es de la siguiente manera:

$$\prod_{i=1}^{n} N(X_i|\mu,\Sigma) = \prod_{i=1}^{n} \frac{1}{(2\Pi)^{p/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(X_i-\mu)'\Sigma^{-1}(X_i-\mu)),$$

de aquí, trabajando únicamente con el exponente tenemos que:

$$\begin{aligned}
&\prod_{i=1}^{n} exp(-\frac{1}{2}(X_i-\mu)'\Sigma^{-1}(X_i-\mu)) = \\
&exp(-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)'\Sigma^{-1}(X_i-\mu)) = \\
&exp(-\frac{1}{2}\sum_{i=1}^{n}(X_i'\Sigma^{-1}X_i - 2X_i'\Sigma^{-1}\mu + \mu'\Sigma^{-1}\mu)) = \\
&exp(-\frac{1}{2}(X'\Sigma^{-1}X - 2X'\Sigma^{-1}\mu + n\mu\Sigma^{-1}\mu))
\end{aligned}$$

Donde la jésima coordenada del vector X es $\sum_{i=1}^{n} X_{i,j}$ Por último,como sólo nos interesan los términos donde aparece X_i, llegamos a que el kernel es:

$$exp(-\frac{1}{2}(X'\Sigma^{-1}X - 2X'\Sigma^{-1}\mu))$$

B.8. Kernel del vector de medias de una distribución gaussiana p variada multiplicada por la distribución del vector de medias

En este caso tenemos que X se ditribuye gaussiana con vector de medias μ y matriz de varianza covarianza Σ, mientras que μ se distribuye gaussiana con vector de medias μ_0 y matriz de varianza covarianza Σ_0. Ahora nos interesa conocer el Kernel correspondiente a μ, entonces tendríamos que el producto de las funciones de densidad es:

$$f_X(X|\mu,\Sigma)f_\mu(\mu|\mu_0,\Sigma_0) =$$

$$\frac{exp(-1/2(X-\mu)'\Sigma^{-1}(X-\mu))}{(2\Pi)^{p/2}|\Sigma|^{1/2}}\frac{exp(-1/2(\mu-\mu_0)'\Sigma_0^{-1}(\mu-\mu_0))}{(2\Pi)^{p/2}|\Sigma_0|^{1/2}}$$

Luego de cada función de densidad tomamos lo que dependa de μ, para así obtener sus respectivos kerneles según el anexo 4,lo cual implica que:

$$\begin{aligned}
&Kernel = \\
&exp(-\frac{1}{2}(\mu'\Sigma^{-1}\mu - 2x'\Sigma^{-1}\mu))exp(-\frac{1}{2}(\mu'\Sigma_0^{-1}\mu - 2\mu_0'\Sigma_0^{-1}\mu)) = \\
&exp(-\frac{1}{2}(\mu'\Sigma^{-1}\mu - 2x'\Sigma^{-1}\mu + \mu'\Sigma_0^{-1}\mu - 2\mu_0'\Sigma_0^{-1}\mu)) = \\
&exp(-\frac{1}{2}(\mu'(\Sigma^{-1}+\Sigma_0^{-1})\mu - 2(X'\Sigma^{-1}+\mu_0'\Sigma_0^{-1})\mu) = \\
&exp(-\frac{1}{2}(\mu'(\Sigma^{-1}+\Sigma_0^{-1})\mu - 2(X'\Sigma^{-1}+\mu_0'\Sigma_0^{-1})(\Sigma^{-1}+\Sigma_0^{-1})^{-1} \\
&(\Sigma^{-1}+\Sigma_0^{-1})\mu)).
\end{aligned}$$

De aquí se tiene que μ se distribuye gaussiana p variada con matriz de varianza covarianza $\Sigma + \Sigma_0$, y vector de medias $(X'\Sigma^{-1} + \mu_0'\Sigma_0^{-1})(\Sigma^{-1} +$

$\Sigma_0^{-1})^{-1}$

B.9. Probabilidad condicional

La probabilidad condicional se define de la siguiente manera Ross (1976): Si $P(F > 0)$, entonces:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Apéndice C

Integración Monte Carlo y Slice Sampler

C.1. Integración Monte Carlo

El método de integración Monte Carlo permite estimar integrales mediante un proceso de simulación. Este método es adaptado en el enfoque bayesiano de la siguiente manera:

Supongamos que tenemos la función de densidad de un vector aleatorio Θ, es decir $\Pi(\Theta)$, y que para una función $f(\theta)$ nos gustaría encontrar:

$$\int_{S_\theta} f(\theta)\Pi(\theta)d\theta \tag{C.1}$$

Como Π es una función de densidad, entonces la ecuación (C.1) es la esperanza de $f(\theta)$ bajo la ley de probabilidad de Θ, es decir $E[f(\theta)]$.

Ahora, el método de integración montecarlo se basa en que es posi-

ble generar una muestra aleatoria de Θ, es decir, que es posible generar $\{\Theta_i\}_{i=1}^n$. Y de esta forma obtener un estimador de $E[f(\theta)]$ de la siguiente manera:

$$\widehat{E}[f(\theta)] = \frac{\sum_{i=1}^n f(\theta_i)}{n} \tag{C.2}$$

Como podemos observar, el estimador montecarlo es el promedio aritmético de la muestra generada evaluada en la función $f(.)$.

También podemos notar que $\widehat{E}[f(\theta)] = \frac{\sum_{i=1}^n f(\theta_i)}{n}$ es un estimador insesgado, pues:

$$\begin{aligned} E[\widehat{E}[f(\theta)]] &= E[\frac{\sum_{i=1}^n f(\theta_i)}{n}] \\ &= \frac{\sum_{i=1}^n E[f(\theta_i)]}{n} \\ &= n\frac{E[f(\theta)]}{n} \\ &= E[f(\theta)] \end{aligned} \tag{C.3}$$

Se puede probar que el estimador montecarlo converge casi seguramente al valor de la integral de interés, e inclusive cuando es finita la esperanza de $f(\theta)^2$ bajo la ley de probabilidad de Θ, la convergencia de (C.2) puede medirse con ayuda de la varianza teórica con la siguiente expresión:

$$var[\widehat{E}[f(\theta)]] = \frac{1}{N}\int_{S_\Theta} (f(\theta) - E[f(\theta)])^2 \Pi(\theta) d\theta,$$

a la expresión anterior también es posible aplicarle el método de integración montecarlo para obtener un estimador de esta, con lo cual llegamos al

siguiente resultado:

$$\widehat{var}[\widehat{E}[f(\theta)]] = \frac{\sum_{i=1}^{n}(f(\theta_i) - \widehat{E}[f(\theta_i)])^2}{N^2}. \tag{C.4}$$

C.2. Slice Sampler

Los métodos de simulación slice sampler consisten en generar una sucesión que converge a la distribución que se desea muestrear. La idea intuitiva de estos métodos consiste en suponer que el vector aleatorio bivariado (X, Y) se distribuye uniforme en la región que está por debajo de la función de densidad, de la cual obtendríamos una muestra (X_0, Y_0), y de ésta sólo nos quedaríamos con X_0, siendo ésta nuestra variable de interés.

Ahora, para generar una muestra uniforme en dicha región, se utiliza un algoritmo recursivo con el fin de tener una sucesión que converge a la muestra deseada.

En términos ilustrativos el procedimiento es como sigue:

1) Supongamos que tenemos el kernel o la densidad de una variable aleatoria univarada X, de la cual nos interesaría obtener una muestra aleatoria, y que además la gráfica de esta densidad es por ejemlo:

2) Luego definimos el vector aleatorio bivariado (X, Y), el cual suponemos que se distribuye uniforme en la región que está por debajo de la gráfica anterior, y a esta región la denotamos como U, por lo que U sería el conjunto de parejas (x, y) con la propiedad de que x es menor a y, siendo $y = f(x)$.

Gráfica C.1

Entonces la región de la cual nos interesa obtener una muestra se ve de la siguiente forma:

Gráfica C.2

3) Ahora, proponemos un valor inicial x_0, si un vector aleatorio bivariado se distribuye uniforme en alguna región, entonces para cada valor que tome la variable X, la distribución marginal de Y dado $X = x_0$ se distribuye uniforme dentro del segmento de recta $L(y) = ((y, x_0)|y \in R)$,

donde $L(y)$ está dentro de U. De esta manera, podríamos generar una realización de Y dado x_0, así obtendríamos algún valor y_0, después fijamos éste valor y_0, y tendríamos ahora que X dado y_0 se distribuye uniforme en $U \cap L(x) = ((x, y_0)|x \in R)$, así obtendríamos una realización de X dado y_0 de donde obtendríamos nuevamente un valor x_0. Así repetiríamos el proceso, pues de esta manera tenemos una cadena de Markov que converge a la realización de haber obtenido una muestra uniforme sobre U, donde sólo nos interesa el valor de X.

En la siguiente gráfica se ilustra el punto 3), por simplicidad se supuso que el primer valor que tomó la variable X fue 0, luego Y dado x igual a cero se distribuye uniforme en el intervalo $(0, 1200)$, de donde ahora obtenemos una realización de Y dado x igual a cero, donde Y tomó ahora el valor de 500, con esta nueva información actualizamos el valor de Y, entonces obtenemos una realización de X dado Y igual a 500, y así seguimos actualizando los valores de X y Y hasta obtener la muestra deseada.

Gráfica C.3

Apéndice D

Descomposición $diag(S)Rdiag(S)$

Simular matrices de varianza-covarianza puede resultar complicado, pues existen varias restricciones necesarias en sus componentes, para que efectivamente se tenga una matriz de varianza covarianza. Por ello es conveniente analizar tales matrices en partes, y para ello en John Barnard (2000) se propone la descomposición $\Sigma = diag(S)Rdiag(S)$, donde Σ es la matriz de varianza-covarianza de dimensión $p \times p$, $diag(S)$ es la matriz diagonal formada con el vector de desviaciones estándar $S = (\sigma_1, ..., \sigma_p)$ de dimensión p y R es la matriz de correlaciones de dimensión $p \times p$.

A continuación se hablará brevemente de cómo trabajar con la descomposición $diag(S)Rdiag(S)$ y la distribución a priori de Σ, conforme se menciona en John Barnard (2000). Para ello expresemos a la disribución a priori de Σ de la siguiente manera:

$$
\begin{aligned}
P(\Sigma) &= P(S,R) \\
&= P(R|S)P(S).
\end{aligned}
\tag{D.1}
$$

Es decir, podemos tener una suposición inicial acerca del vector S, y dado éste actualizar nuestro conocimiento acerca de R.

Después el problema consiste en proponer adecuadamente las distribuciones para $(R|S)$ y $P(S)$. Para la distribución de S basta con proponer una distribución p-variada que cumpla con las característica de ser una distribución para las desviaciones estándar $\sigma_1, ..., \sigma_p$.

Una posible elección para S es una distribución lognormal tal que $log(S) \sim N(\xi, \Delta)$. Luego, si Δ es una matriz diagonal, se tiene que las desviaciones estándar son independientes entre sí.

Para el caso de la distribución de la matriz de correlaciones R, se puede hacer el supuesto de que R y S son independientes, de esta manera se puede tratar a cada una de estas distribuciones por separado, y a su vez, a falta de información se evita usar erróneamente alguna distribución que ligue a R con S. De esta manera, la distribución a priori de Σ queda como $P(\Sigma) = P(R)P(S)$.

Ahora, teniendo en mente que la distribución de R no depende de S, a partir de la distribución Whishart inversa se puede encontrar una distribución conjunta en el espacio de matrices de dimensión $p \times p$, tal que todas las densidades marginales de las correlaciones $r_{i,j}$ sean uniformes.

Entonces si se parte de que Σ tiene una distribución Wishart inversa, es decir, $W^{-1}(\Sigma|I, v, p)$, donde I es la matriz identidad de las matrices de dimensión $p \times p$, v son los grados de libertad, y p la dimensión del espacio de matrices. Y mediante la transformación $\Sigma = diag(S)Rdiag(S)$, y el supuesto de que S y R son independientes se obtiene entonces que la distribución de R es proporcional a la siguiente función:

$$f_k(R|v) \propto |R|^{-1/2(v-1)(k-1)-1}(\Pi_{i=1}^{p}|R_{ii}|)^{-v/2}, \tag{D.2}$$

donde R_{ii} es la i-ésima submatriz de R.

También se pueden obtener las densidades marginales de las submatrices de correlación de dimensión $j \times j$, con ayuda de la propiedad de marginalización de la distribución Wishart inversa. Si R^n es una submatriz de correlación de R, con distribución wishart inversa con parámetros I_j, y dimensión $j \times j$, esta tiene una distribución wishart inversa $W^{-1}(R^j|I_j, v-(p-j), j)$. Entonces si se sustituye v por $v-(p-j)$, y k por j en la ecuación D.2 se tiene que:

$$f_j(R^n|v-(p-j)) \propto |R|^{-1/2(v-(p-j)-1)(j-1)-1}(\Pi_{i=1}^{j}|R_{ii}^n|)^{-v/2}.$$

Si en la ecuación anterior se sustituye j por 2 se tiene que:

$$f_2(r_{ij}|v-(p-2)) \propto (1-r_{ij}^2)^{(v-p-1/2)},$$

mientras que si j se sustituye por 3 tendríamos que:

$$f_3(r_{12}, r_{13}, r_{23}|v-(p-3)) \propto \frac{(1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23})^{v-p+1}}{((1-r_{12}^2)(1-r_{13}^2)(1-r_{23}^2))^{(v-p+3)/2}}.$$

Apéndice E

Gráficos de convergencia y estabilización de varianza

Gráfica E.1: Convergencia de μ

Gráfica E.2: Convergencia de β

Gráfica E.3: Convergencia de λ y ϕ

Gráfica E.4: Convergencia de σ_i^2

Gráfica E.5: Convergencia de r_{ij}

Gráfica E.6: Estabilización de varianza de μ

Gráfica E.7: Estabilización de varianza de β

Gráfica E.8: Estabilización de varianza de λ y ϕ

Gráfica E.9: Estabilización de varianza de σ_i^2

Gráfica E.10: Estabilización de varianza de r_{ij}

Gráfica E.11: Convergencia de μ

Gráfica E.12: Convergencia de β

Gráfica E.13: Convergencia de λ y ϕ

Gráfica E.14: Convergencia de σ_i^2

Gráfica E.15: Convergencia de r_{ij}

Gráfica E.16: Estabilización de varianza de μ

Gráfica E.17: Estabilización de varianza de β

Gráfica E.18: Estabilización de varianza de λ y ϕ

Gráfica E.19: Estabilización de varianza de σ_i^2

Gráfica E.20: Estabilización de varianza de r_{ij}

Bibliografía

Abramowitz, M. and I. A. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: US National Bureau of Standards.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London 353(1674), 401–419.*

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. International Statistical Review 50(2), 370–418.

Bernardo, J. and Smith (2008). *Bayesian Theory.* Chichester: Jhon Wiley and Sons.

Finetti, B. D. (1937). La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré 1(17), 1–68.*

Javier Mencía, E. S. (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. Journal of Econometrics 1(153), 105–121.

Jeffreys, S. H. (1939). *Theory of probability.* OUP UK: Oxford Classic Texts in the Physical Sciences.

John Barnard, Robert McCulloch, X.-L. M. (2000). Modeling covariance

matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica 10, 1281–1311.*

Laplace, P. S. (1812). Théorie analytique des probabilités. *Paris: Paris, Ve Courcier, 1812.*

Lindley, D. V. (1957). A statistical paradox. Biometrika, 44(1-2)*, 187–192.*

Markowitz, H. (1952). Portfolio selection. The Journal of Finance 7(1), 77–91.

McNeil, A. J., R. Frey, and P. Embrechts (1967). *Quantitative Risk Management: Concepts, Techniques and Tools.* New Jersey: Princeton University Press.

Peiró, A. (1998). Skewness in financial returns. *23(6), 847–862.*

Ross, S. (1976). A First Course in Probability. *New York: Pearson/Prentice Hall.*

Savage, L. (1972). The Foundations of Statistics. *Dover Publications, New York.*

Stuart Geman, D. G. (1984). Stochastic relaxation, gibbs distributions, the bayesian restoration of images. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 6(6), 721–741.