



# COLEGIO DE POSTGRADUADOS

---

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

**CAMPUS MONTECILLO**

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

**Modelación del ingreso en México con un enfoque  
Bayesiano**

Fredy Yair Montes Rivera

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2012

---

---

La presente tesis titulada: **Modelación del ingreso en México con un enfoque Bayesiano**, realizada por el alumno: **Fredy Yair Montes Rivera**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

## MAESTRO EN CIENCIAS

### SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

#### CONSEJO PARTICULAR

CONSEJERO



Dr. Paulino Pérez Rodríguez

ASESOR

  
Dr. Miguel Ángel Martínez Damián

ASESOR

  
Dr. José Aurelio Villaseñor Alva

ASESOR

  
Dr. Sergio Pérez Elizalde

# Modelación del ingreso en México con un enfoque Bayesiano

Fredy Yair Montes Rivera

Colegio de Postgraduados, 2011

En el presente trabajo se proponen tres distribuciones (Pareto, Lognormal y Dagum) para modelar el ingreso de la población mexicana, lo anterior se realizó desde un enfoque bayesiano dado que en nuestro país no existe evidencia alguna de un trabajo con este enfoque. Una vez obtenidos los diferentes parámetros de manera bayesiana de cada distribución, se opta trabajar con la distribución Dagum ya que ésta muestra mejor ajuste a los datos de ingreso en México. Ya seleccionado el modelo se estiman los parámetros  $(a, b, p)$  de ésta distribución con el uso del algoritmo “t-walk” el cual fue desarrollado por [Christen y Fox \(2010\)](#), esto se realizó para tres diferentes años (1998, 2002 y 2008), lo cual permitió realizar una comparación acerca de como se ha comportado la distribución de la riqueza en México medido a través del ingreso reportado por la ENIGH la cual es elaborada por el [INEGI \(1998, 2002, 2008\)](#), para este análisis se utilizó el coeficiente de Gini y la curva de Lorenz las cuales son calculadas directamente con los parámetros estimados de la distribución Dagum.

Uno de los principales resultados obtenidos es que; la asignación del ingreso en México para el periodo de análisis ha venido siendo cada vez más inequitativa, ya que para el año de 1998 el coeficiente de Gini fue de 0.6225 mientras que para 2008 éste fue de 0.6464, lo cual muestra un incremento de 7.87 por ciento en un lapso de 10 años, conclusión muy diferente a la hecha por el [CEFP \(2008\)](#)<sup>1</sup>, él cual afirma de que además de ser menor la desigualdad en términos de distribución de la riqueza, ésta ha disminuido (coeficiente de Gini para el año 2000 igual a 0.456 mientras que en 2006 éste fue de 0.448).

**Palabras clave:** Distribución Pareto, Lognormal, Dagum, algoritmo “t-walk”, coeficiente de Gini, curva de Lorenz.

---

<sup>1</sup>Centro de Estudios de las Finanzas Públicas de la H. Cámara de Diputados.

# Modeling the income of Mexico; a Bayesian approach

Fredy Yair Montes Rivera

Colegio de Postgraduados, 2011

In this work, three distributions are proposed (Pareto, Lognormal y Dagum) to model the income of Mexican population, it was done through a Bayesian approach due to in Mexico there is not any work with this approach. When the different parameters are estimated, it shows that the Dagum model is better due to the fact that it best describes the data at hand. The posterior distribution of the quantities of interest is obtained by using a new Markov Chain Monte Carlo (MCMC) method called “t-walk” developed by [Christen y Fox \(2010\)](#). The analysis was done by using data from three diferents periods of time: 1998, 2000 and 2008. The analyzed data come from ENIGH, which is reported by [INEGI \(1998, 2002, 2008\)](#). The obtained results allow us to estimate the Gini’s coefficient and the Lorenz’s curve.

The results obtained shown that the distribution of Mexican income is quite unequal. The Gini’s coefficient for the year 1998 was 0.6225 while for 2008 it was 0.6464, that represents an increase of 7.87% in only 10 years. The result is quite different to that reported by [CEFP \(2008\)](#), which shows that the Gini’s coefficient has not changed at all through a long period of time, for instance for the year 2000 it was 0.456 and for 2006 it was 0.448.

**Key words:** Pareto distribution, Lognormal, Dagum, algorithm “t-walk”, Gini coefficient, Lorenz curve.

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado durante la realización de mis estudios.

Al Colegio de Postgraduados, por haberme brindado la oportunidad de seguir mi formación profesional.

A los integrantes de mi Consejo Particular:

Al Dr. Paulino Pérez Rodríguez, por el tiempo e interés mostrado en la realización de éste trabajo, pero principalmente por que durante este tiempo se volvió un gran amigo.

Al Dr. Sergio Pérez Elizalde por su orientación, apoyo y colaboración en el presente trabajo.

Al Dr. Miguel Ángel Martínez Damián y al Dr. José Aurelio Villaseñor Alva, por sus recomendaciones y por dedicar parte de su tiempo en la revisión de este trabajo.

Al Dr. Humberto Vaquera Huerta, por sus atinados consejos y apoyo durante mi estadía en esta institución.

Al M.C. Armando Ramírez Román por confiar siempre en mi, pero principalmente por ser un buen amigo.

A mis amigos Gregorio, Ivan, Osmar, Emmanuel, Enrique, Miguel y todo Kame House.

## DEDICATORIA

A mi esposa e hija ...

A mis madres, hermano y demás familia.

# Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>3</b>
2.1. Objetivos Generales . . . . .	3
2.2. Objetivos Particulares . . . . .	3
<b>3. Marco Teórico</b>	<b>4</b>
3.1. El Papel de la Estadística . . . . .	4
3.2. El modelo Bayesiano . . . . .	5
3.2.1. Distribución a priori . . . . .	5
3.2.2. Densidad a priori invariante no informativa de Jeffreys . . . . .	6
3.2.3. Funciones impropias . . . . .	7
3.2.4. Teorema Bayesiano Central del Límite . . . . .	7
3.2.5. Distribución posteriori . . . . .	8
<b>4. Monitoreo de convergencia</b>	<b>9</b>
4.1. Cadenas de Markov Monte Carlo . . . . .	9
4.2. Pruebas de Convergencia . . . . .	10

# Índice

---

4.2.1. Prueba de convergencia de Gelman y Rubin . . . . .	10
4.2.2. Prueba de convergencia de Raftery y Lewis . . . . .	12
<b>5. Algoritmos de Muestreo</b>	<b>13</b>
5.1. Algoritmo del muestreador de Gibbs . . . . .	13
5.2. Algoritmo ARS (Muestreo de Aceptación y Rechazo Adaptativo) . . . . .	14
5.2.1. Implementación . . . . .	14
5.2.2. Implementación del algoritmo ARS . . . . .	15
5.3. Algoritmo Metropolis Hasting . . . . .	16
5.3.1. Implementación del algoritmo Metropolis Hasting . . . . .	16
5.4. Algoritmo “t-walk” . . . . .	17
<b>6. Distribución Pareto</b>	<b>18</b>
6.1. Antecedentes . . . . .	18
6.2. Definición . . . . .	19
6.3. Momentos y otras propiedades básicas . . . . .	19
6.4. Estimadores de máxima verosimilitud de la distribución Pareto . . . . .	20
6.5. Estimadores Bayesianos para la distribución Pareto . . . . .	20
6.6. Implementación del algoritmo de muestreo Gibbs . . . . .	22
6.6.1. Metodología . . . . .	22
<b>7. Distribución log-normal</b>	<b>23</b>
7.1. Antecedentes . . . . .	23
7.2. Definición . . . . .	25



# Índice

---

7.3. Momentos y otras propiedades básicas . . . . .	25
7.3.1. Propiedades . . . . .	25
7.3.2. Momentos de bajo orden . . . . .	26
7.3.3. Estimadores de máxima verosimilitud para la distribución Lognormal	27
7.4. Estimadores Bayesianos para la distribución lognormal . . . . .	27
7.5. Implementación del ARS dentro del muestreador de Gibbs . . . . .	28
7.6. Metodología . . . . .	29
<b>8. Distribución Dagum</b>	<b>30</b>
8.1. Antecedentes . . . . .	30
8.2. Definición . . . . .	31
8.3. Momentos y otras propiedades básicas . . . . .	32
8.4. Estimadores de máxima verosimilitud de la distribución Dagum . . . . .	32
8.5. Estimadores Bayesianos para la distribución Dagum . . . . .	34
<b>9. Metodología</b>	<b>36</b>
9.1. Base de datos . . . . .	36
9.1.1. Antecedentes . . . . .	36
9.1.2. Diseño conceptual de la ENIGH . . . . .	37
9.1.3. Diseño estadístico de la ENIGH . . . . .	38
9.2. Selección del Modelo . . . . .	38
9.3. Análisis del ingreso en México con la Distribución Dagum . . . . .	41
9.3.1. Prueba de convergencia en la distribución Dagum . . . . .	42
9.3.2. Distribuciones marginales posteriores . . . . .	43

# Índice

---

9.3.3. Ajuste de la distribución Dagum a los datos del ingreso, para los años 1998, 2002 y 2008 . . . . .	46
<b>10.Resultados</b>	<b>50</b>
10.1. Definición Coeficiente de Gini . . . . .	50
10.2. Relación entre la curva de Lorenz y el coeficiente de Gini . . . . .	50
10.3. Coeficiente de Gini para la distribución Dagum . . . . .	51
10.4. Cálculo del Coeficiente de Gini y Curva de Lorenz para México en los años 1998, 2002 y 2008 . . . . .	52
<b>11.Conclusiones</b>	<b>56</b>
<b>Referencias</b>	<b>57</b>
<b>Anexos</b>	<b>61</b>
Anexo A: . . . . .	61

# Índice de tablas

9.1. Criterio de Akaike. . . . .	39
9.2. Prueba de convergencia de Gelman y Rubin para el año de corte 1998. . .	42
9.3. Prueba de convergencia de Gelman y Rubin para el año de corte 2002. . .	42
9.4. Prueba de convergencia de Gelman y Rubin para el año de corte 2008. . .	43
10.1. Coeficiente de Gini. . . . .	52
10.2. Intervalo de máxima densidad al 95 % de probabilidades para el Coeficiente de Gini. . . . .	53

# Índice de figuras

9.1. Comparación entre las funciones de densidad Lognormal y Dagum, con datos del ingreso de 1998. . . . .	40
9.2. Comparación entre las funciones de densidad Lognormal y Dagum, con datos del ingreso de 2002. . . . .	40
9.3. Comparación entre las funciones de densidad Lognormal y Dagum, con datos del ingreso de 2008. . . . .	41
9.4. Distribuciones posteriores de los parámetros en la distribución Dagum, para datos del ingreso de 1998. . . . .	44
9.5. Distribuciones posteriores de los parámetros en la distribución Dagum, para datos del ingreso de 2002. . . . .	45
9.6. Distribuciones posteriores de los parámetros en la distribución Dagum, para datos del ingreso de 2008. . . . .	46
9.7. Comparación entre la densidad del ingreso en México y la densidad ajustada con la distribución Dagum. . . . .	47
9.8. Comparación entre la función de distribución empírica y teórica. . . . .	48
10.1. Histogramas del Coeficiente de Gini para los años 1998, 2002 y 2008. . . . .	54
10.2. Curvas de Lorenz para los años 1998, 2002 y 2008. . . . .	55

# Capítulo 1

## Introducción

Uno de los principales temas en el análisis económico es la distribución del ingreso, dado que es un indicador del dinamismo y de la eficiencia de un sistema económico para asignar la producción entre los individuos. La desigualdad en la distribución del ingreso es uno de los principales factores que limitan el desarrollo social y económico; el principio básico es que una sociedad menos igualitaria limita las posibilidades de los individuos para desarrollarse y generar valor. En el agregado, cuando unos pocos concentran una gran proporción del ingreso el consumo de la mayoría es escaso y, por ello, la demanda de bienes de mayor valor agregado por parte del grueso de la población es limitada; ya que la mayor proporción del ingreso disponible se destina al consumo corriente o de subsistencia. En consecuencia, el ahorro, la inversión y la producción de bienes duraderos son mínimos a partir de lo cual se restringe la capacidad de desarrollo económico nacional (CEFP, 2008).

Por ello se realiza un análisis del ingreso en la población mexicana con el objetivo de encontrar un instrumento estadísticamente significativo que permita medir el grado de desigualdad que presenta la población mexicana en términos de distribución de la riqueza. Además de comprobar si esta desigualdad ha disminuido o aumentado en un periodo de 10 años, con lo cual se pretende mostrar que a pesar de que ha aumentado el gasto público y por lo tanto el producto interno bruto, la asignación de estos recursos provenientes del erario público no se ha dado de manera eficiente en términos de mitigar la pobreza y beneficiar a los que menos ingresos tienen.

El instrumento seleccionado para éste análisis es el coeficiente de Gini, el cual será calculado de manera paramétrica, ya que se le asignara una distribución de probabilidades a los datos de ingreso. La estimación de los parámetros será de manera Bayesiana, la cual hasta hace unas décadas era bastante complicada e incluso en algunos casos imposible por lo cual esta técnica de estimación no era muy socorrida por la comunidad científica a pesar de las ventajas que presenta. La manera más común de realizar estimación bayesiana era proponiendo distribuciones conjugadas, las cuales permiten el manejo matemático

de estas al realizar el procedimiento para encontrar los estimadores deseados, aunque en algunas situaciones estas distribuciones no estén justificadas desde un punto de vista teórico, es decir que no tienen relación alguna con el parámetro a estimar.

A finales del siglo pasado el desarrollo de la tecnología informática creció de manera exponencial dando así la pauta para la creación de nuevos programas y algoritmos de muestreo, los cuales permiten resolver de manera más rápida y eficiente el problema de la estimación bayesiana, el cual regularmente se complica al tratar de reconocer el kernel de las distribuciones marginales posteriores. Dentro de estos algoritmos que funcionan a través del método Monte Carlo vía Cadenas de Markov (MCMC) se encuentran el muestreador de Gibbs explicado en [Casella y George \(1992\)](#), algoritmo Metropolis-Hastings propuesto por [Metropolis \*et al.\* \(1953\)](#), aceptación y rechazo adaptativo publicado por [Gilks y Wild \(1992\)](#), “t-walk” propuesto por [Christen y Fox \(2010\)](#), entre otros más.

Dependiendo de la naturaleza del problema a resolver, se selecciona el algoritmo a usar. Existen casos en los cuales se pueda aplicar más de un algoritmo para resolver el problema, en estos casos elegir el mejor método dependerá de los diferentes criterios del investigador. Una vez estimados los parámetros de las distribuciones de interés se procede a calcular el coeficiente de Gini.

En el desarrollo del trabajo se tratan los puntos antes mencionados de la siguiente manera. En el capítulo 2 se plantean los objetivos generales y específicos, los cuales ayudarán a no perder y cumplir la finalidad de esta investigación. En el capítulo 3 se aborda el marco teórico partiendo del papel de la estadística en la economía, abordando posteriormente todo lo referente a la estimación Bayesiana. El capítulo 4 explica los diferentes métodos de convergencia a usar en el presente trabajo. Después en el capítulo 5 se explica lo referente a los algoritmos de muestreo, que serán de ayuda al momento de muestrear de las distribuciones finales o posteriores. Los capítulos 6, 7 y 8 abordan lo referente a la importancia de las distribuciones Pareto, Lognormal y Dagum respectivamente, ya que éstas son nuestras distribuciones objetivo. El capítulo 9 discute la metodología utilizada en el presente trabajo, además de incluir un caso práctico (con datos de la ENIGH para los años 1998, 2002 y 2008) el cual muestra una aplicación del mismo. Así con lo obtenido en el apartado anterior, en el capítulo 10 se realiza un análisis económico sobre la distribución del ingreso de la población mexicana para los mismos años antes mencionados, los principales indicadores utilizados en este análisis son el coeficiente de Gini y la curva de Lorenz. Finalmente el capítulo 11 muestra las conclusiones más importantes obtenidas en esta investigación, además de que se anexa el código en [R Development Core Team \(2011\)](#) utilizado en el desarrollo del trabajo.

# Capítulo 2

## Objetivos

### 2.1. Objetivos Generales

- Modelar el ingreso de la población mexicana con tres distribuciones probabilísticas de cola pesada (Pareto, Lognormal, Dagum), para futuras inferencias sobre la misma.

### 2.2. Objetivos Particulares

- Encontrar qué distribución probabilística se ajusta mejor a los ingresos de la población mexicana.
- Realizar inferencias sobre el ingreso en México en tres cortes de tiempo transversal para concluir si la cantidad de personas con cierto ingreso ha aumentado o disminuido.

# Capítulo 3

## Marco Teórico

### 3.1. El Papel de la Estadística

El papel a desarrollar por la Estadística en la cada vez más estrecha interacción entre la Economía, Actuaría y las Finanzas es fundamental. Los modelos utilizados en el presente, y a utilizarse en el futuro próximo, necesitan de la estimación de parámetros mediante técnicas estadísticas establecidas o a desarrollarse. Dentro del área de riesgos catastróficos destaca la teoría de valores extremos y de distribuciones de colas pesadas.

La estimación de valores en riesgo es equivalente a la estimación de cuantiles. La cuantificación de la dependencia entre distintos activos financieros juega un papel fundamental en los modelos de finanzas y en el cual la Estadística debe ser una herramienta primordial. Otra área de gran importancia será la de probar la validez de los supuestos de los modelos empleados a partir de nuevas técnicas de pruebas de hipótesis que permitan introducir las características peculiares de los datos financieros.

El área de “teoría de credibilidad” dentro de Actuaría tiene su fundamento en modelos de Estadística Bayesiana por lo que este campo de la Estadística es fundamental en el desarrollo de la investigación actuarial. Las fórmulas comúnmente utilizadas en el cálculo de las primas de credibilidad en seguro de daños esencialmente utilizan principios y métodos Bayesianos.

La aplicación de técnicas de simulación y estimación Monte Carlo resultan fundamentales en la aplicación de modelos tan complejos como aquellos utilizados en la Actuaría y Finanzas modernas.

En seguro de daños y de vida se han empezado a introducir modelos espaciales y espacio-temporales para el cálculo de primas para distinguir éstas por región geográfica y en el tiempo.



## 3.2. El modelo Bayesiano

---

Varios modelos en Actuaría y Finanzas requieren del pronóstico de valores futuros de ciertas variables. En esta área, la Estadística ha contribuido con modelos para series de tiempo no lineales (GARCH, EGARCH, S(E)TAR, de volatilidad estocástica, entre otros) y en la aplicación de métodos de selección de modelos, también modelos no lineales, en la teoría de redes neuronales. También es importante destacar el desarrollo de pruebas de hipótesis para confirmar la predictibilidad de datos financieros (Pruebas de caminata aleatoria, de martingala, etc.).

La Estadística no puede quedarse ajena al desarrollo vertiginoso de la Actuaría y Finanzas en las décadas próximas más bien debe de convertirse en un participante activo que contribuya con su tradición en la construcción de modelos, estimación de parámetros, cuantificación de errores y validación de modelos que permita establecer un vínculo entre los datos empíricos y los modelos teóricos a desarrollarse que evite el mal uso de estos modelos (Fernández, 2005).

## 3.2. El modelo Bayesiano

Un modelo Bayesiano está basado en la especificación de un modelo de probabilidad para los datos observados  $\mathbf{X}$ , dado un vector de parámetros desconocidos  $\boldsymbol{\theta}$ , encabezado por la función de verosimilitud  $L(\mathbf{X} | \boldsymbol{\theta})$ . Se asume que  $\boldsymbol{\theta}$  es aleatorio y tiene una distribución inicial o distribución a priori denotada por  $\pi(\boldsymbol{\theta})$ . Cualquier inferencia sobre  $\boldsymbol{\theta}$  está basada en su distribución final o distribución posterior, la cual, por el teorema de Bayes está dada por:

$$\pi(\boldsymbol{\theta} | \mathbf{X}) = \frac{L(\mathbf{X} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Omega} L(\mathbf{X} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (3.1)$$

Donde  $\Omega$  denota el espacio paramétrico de  $\boldsymbol{\theta}$  y  $L(\mathbf{X} | \boldsymbol{\theta})$  es la función de verosimilitud.

De (3.1) es claro que  $\pi(\boldsymbol{\theta} | \mathbf{X}) \propto L(\mathbf{X} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})$  y así, esto involucra una contribución de los datos observados a través de  $L(\mathbf{X} | \boldsymbol{\theta})$ , y una contribución de la información inicial cuantificada a través de  $\pi(\boldsymbol{\theta})$ . La cantidad  $m(\mathbf{X}) = \int_{\Omega} L(\mathbf{X} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$  es la constante de normalización de  $\pi(\boldsymbol{\theta} | \mathbf{X})$ , y es llamada frecuentemente la distribución marginal de los datos o la distribución predictiva inicial (Carlin y Louis, 2000).

### 3.2.1. Distribución a priori

Puesto que generalmente no se cuenta con la ayuda de un experto en el fenómeno que se está estudiando, uno de los mayores problemas para la ejecución de un análisis Ba-

## 3.2. El modelo Bayesiano

---

yesiano, es el de elegir convenientemente una distribución a priori para los parámetros que se presumen desconocidos. Hasta hace pocos años, el inconveniente computacional de los estudios Bayesianos había sido tan grande que, por mucho tiempo, los tipos de distribuciones a priori más usadas fueron las conjugadas por permitir cálculos explícitos (formas cerradas) de las distribuciones posteriores. Frecuentemente no existe información a priori confiable acerca del vector de parámetros  $\boldsymbol{\theta}$ , o solamente se desea una inferencia basada en los datos. Al utilizar una distribución a priori no informativa, la información resultante en la distribución posterior  $f(\boldsymbol{\theta} \mid \mathbf{X})$  surge de los datos, y en consecuencia la inferencia resultante será completamente objetiva, en lugar de subjetiva (Carlin y Louis, 2000).

### 3.2.2. Densidad a priori invariante no informativa de Jeffreys

La regla de Jeffreys es una técnica para obtener densidades a priori y es útil cuando se tiene ignorancia de las propiedades de la familia de densidades  $f(\mathbf{X} \mid \boldsymbol{\theta})$ , (Hartigan, 1964).

Datta y Ghosh (1996) mencionan que la técnica desarrollada por Jeffreys ofrece una alternativa relativamente fácil de calcular, la cual es invariante bajo transformaciones uno a uno, pero no trabaja satisfactoriamente en la presencia de parámetros de estorbo (nuisance parameters, en inglés). Kass y Wasserman (1996) muestran una expresión para obtener la priori de Jeffreys en el caso multivariado, la cual está dada por:

$$f^J(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}, \quad (3.2)$$

donde  $|\cdot|$  denota el determinante, y  $I(\boldsymbol{\theta})$  es la matriz de Información de Fisher, que tiene como elemento  $i, j$  a:

$$I(\boldsymbol{\theta})_{ij} = -E_{\mathbf{x}|\boldsymbol{\theta}} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(\mathbf{X} \mid \boldsymbol{\theta}) \right). \quad (3.3)$$

Es decir, la densidad a priori de Jeffreys es proporcional a la raíz cuadrada positiva del determinante de la matriz de Información de Fisher. Aunque (3.2) proporciona un procedimiento general para obtener distribuciones a priori no informativas, en ocasiones puede ser problemático usarla en altas dimensiones. Carlin y Louis (2000) plantean obtener una distribución a priori no informativa para cada parámetro individualmente, y entonces formar la distribución a priori conjunta simplemente como el producto de las prioris individuales.

## 3.2. El modelo Bayesiano

---

### 3.2.3. Funciones impropias

Una propiedad básica de una función de densidad  $f(x)$  es que integra (o suma) sobre su rango admisible a 1. Ahora, si  $h(x)$  es una función uniforme en la línea real, esto es:

$$h(x) = c, \quad -\infty < x < \infty, \quad c > 0 \quad (3.4)$$

entonces es una función impropia, ya que la integral

$$\int_{-\infty}^{\infty} h(x) dx = c \int_{-\infty}^{\infty} dx$$

no existe, no importa que tan pequeña sea  $c$ . Las funciones de este tipo se emplean frecuentemente para representar el comportamiento local de la distribución a priori en la región donde la verosimilitud es apreciable, pero no sobre su rango admisible completo. Frecuentemente, una función a priori impropia, se puede combinar con la verosimilitud para dar una densidad posterior que es propia. Esta acción es válida en el sentido de que para un tamaño de muestra grande, la verosimilitud domina la densidad a priori (Box y Tiao, 1973). En este sentido, Carlin y Louis (2000) proporcionan el Teorema Bayesiano Central del Límite.

### 3.2.4. Teorema Bayesiano Central del Límite

Suponga que  $X_1, X_2, \dots, X_n \sim i.i.d f_{X_i}(X_i | \boldsymbol{\theta})$ , de tal manera que  $L(\mathbf{X} | \boldsymbol{\theta}) = \prod_{i=1}^n f_{X_i}(X_i | \boldsymbol{\theta})$ . Suponga además que la densidad a priori  $\pi(\boldsymbol{\theta})$  y la verosimilitud  $L(\mathbf{X} | \boldsymbol{\theta})$  son positivas y dos veces diferenciables cerca de  $\hat{\boldsymbol{\theta}}$ ; se asume también que la moda posterior de  $\boldsymbol{\theta}$  existe. Entonces bajo condiciones de regularidad, la distribución posterior  $\pi(\boldsymbol{\theta} | \mathbf{X})$  para  $n$  grande puede aproximarse por una distribución normal con media igual a la moda posterior y matriz de covarianzas igual a menos la inversa de la matriz de segundas derivadas de la log posterior evaluada en la moda.

Lee (1997) menciona que es comúnmente razonable analizar datos científicos en la suposición que la verosimilitud domina la densidad a priori por dos razones principales. Primero, en la ausencia de información a priori, simplemente se debe emplear la función de verosimilitud. Segundo, una investigación científica usualmente no se comienza a menos que la información que proporcione la investigación probablemente sea más precisa que la información disponible; si este es el caso, entonces presumiblemente la verosimilitud dominará la distribución a priori.

## 3.2. El modelo Bayesiano

---

### 3.2.5. Distribución posteriori

Con el uso del Teorema de Bayes se obtuvo la expresión general para la densidad posterior de  $\boldsymbol{\theta}$  dada en (3.1); esta expresión permite notar que la densidad posterior involucra una contribución de los datos observados a través de  $L(\mathbf{X} | \boldsymbol{\theta})$ , y una contribución de la información a priori cuantificada a través de  $\pi(\boldsymbol{\theta})$ .

La expresión  $m(\mathbf{X}) = \int_{\Omega} L(\mathbf{X} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ , como ya se mencionó antes es la constante de normalización de la densidad  $f(\mathbf{X} | \boldsymbol{\theta})$ , y comúnmente se le llama la distribución marginal de los datos o la distribución predictiva a priori. Se ha expresado que un análisis Bayesiano basa sus inferencias en la densidad posterior, sin embargo, en muchos modelos ocurre que  $m(\mathbf{X})$  no tiene una forma analítica cerrada; este dilema conduce a la pregunta; ¿cómo muestrear de la distribución multivariable posterior  $\pi(\boldsymbol{\theta} | \mathbf{X})$ , cuando ésta no tiene una forma cerrada?; esta pregunta a la vez conduce a una basta literatura sobre métodos computacionales para muestrear  $\pi(\boldsymbol{\theta} | \mathbf{X})$  y métodos para estimar  $m(\mathbf{X})$ . Dentro de los métodos computacionales más conocidos se encuentran el muestreador de Gibbs, el método de aceptación y rechazo, el método de aceptación y rechazo adaptativo, el algoritmo de Metropolis-Hastings, entre otros más.

Las inferencias respecto al vector  $\boldsymbol{\theta}$  se realizan a través de la estimación de  $E[h(\boldsymbol{\theta})]$ , donde  $h(\boldsymbol{\theta})$  es una función del vector  $\boldsymbol{\theta}$  y la esperanza calculada con respecto a la distribución final. Con  $E[h(\boldsymbol{\theta})]$  se obtienen inferencias para:

a) la media de  $\boldsymbol{\theta}$ , si  $h(\boldsymbol{\theta}) = \boldsymbol{\theta}$ .

b) la covarianza de  $\theta_i$  y  $\theta_j$ , si  $h(\boldsymbol{\theta}) = \theta_i - E(\theta_i | \mathbf{X})\theta_j - E(\theta_j | \mathbf{X})$ ,

donde  $E(\theta_j | \mathbf{X}) = \int_{\Omega} \theta_j \pi(\boldsymbol{\theta} | \mathbf{X}) d\boldsymbol{\theta}$ , y la integral con respecto al resto de las otras componentes  $\theta_j$ .

c) la densidad predictiva, si  $h(\boldsymbol{\theta}) = f(z | \boldsymbol{\theta})$ , donde  $f(z | \boldsymbol{\theta})$  es la densidad predictiva dado el vector de parámetros.

d) la probabilidad de un conjunto  $A$ , si  $h(\boldsymbol{\theta}) = \mathbf{I}_A(\boldsymbol{\theta})$ .

Otras cantidades que pueden obtenerse con  $E_{\pi_{\boldsymbol{\theta}}}[h(\boldsymbol{\theta})]$  son: las densidades marginales de cada uno o de un conjunto de los componentes del vector  $\boldsymbol{\theta}$ , las constantes de normalización, los factores de Bayes, densidades finales, cuantiles e intervalos. Para más detalles ver [Chen et al. \(2000\)](#).

# Capítulo 4

## Monitoreo de convergencia

### 4.1. Cadenas de Markov Monte Carlo

Una sucesión de vectores aleatorios,  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  es una cadena de Markov, si para cada  $t \geq 0$ ,  $P(\mathbf{x}_{t+1} \mid \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t)$  depende solamente del estado actual  $\mathbf{x}_t$  de la cadena (Gilks *et al.*, 1996), i.e.,  $P(\mathbf{x}_{t+1} \mid \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t) = P(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$ . A  $P(\cdot \mid \cdot)$  se le conoce como el kernel de transición de la cadena. En esta definición se asume que  $P(\cdot \mid \cdot)$  no depende de  $t$ .

Una cadena de Markov converge a una distribución estacionaria  $\varphi(\cdot)$ , si  $P^t(\cdot \mid \mathbf{x}_0)$  converge eventualmente a  $\varphi(\cdot)$ . Es decir la cadena olvida gradualmente su estado inicial y no depende de  $t$ .

Para que la distribución de  $\mathbf{x}_t$  converja a una distribución estacionaria, la cadena debe tener tres propiedades importantes (Gilks *et al.*, 1996):

- La primera es que la cadena sea irreducible, es decir, que para toda  $i, j$  exista  $t \geq 0$  tal que  $P^t(\mathbf{x}_t = j \mid \mathbf{x}_0 = i) > 0$ .
- La segunda es que la cadena debe ser no periódica, es decir, si para cada  $i$  se cumple que el máximo común divisor de  $\{t > 0 : P^t(\mathbf{x}_t = i \mid \mathbf{x}_0 = i) > 0\}$  es igual a 1.
- La tercera y más importante es que la cadena debe ser recurrente positiva. La cadena es recurrente si para toda  $i, j$  se cumple que:

$$\sum_t P^t(\mathbf{x}_t = j \mid \mathbf{x}_0 = i) = \infty$$

## 4.2. Pruebas de Convergencia

---

y es positiva si existe una distribución  $\varphi(\cdot)$ , tal que:

$$\sum_t \varphi(i) P^t(\mathbf{x}_t = j \mid \mathbf{x}_0 = i) = \varphi(j), \forall j, t \geq 0.$$

Estas tres propiedades son referidas como condiciones de regularidad de una cadena de Markov.

Para que un algoritmo que utiliza Cadenas de Markov de Monte Carlo sea exitoso, debe cumplir dos condiciones. La primera es construir cadenas a lo largo del tiempo, de tal manera que su distribución estacionaria  $\varphi(\cdot)$  sea precisamente la distribución de interés  $\pi(\cdot)$ , y la segunda es lograr una razón de convergencia rápida hacia la distribución estacionaria.

## 4.2. Pruebas de Convergencia

Una manera de monitorear la convergencia de una cadena de Markov, consiste en generar varias cadenas con la misma longitud pero con diferentes puntos de inicio, y verificar en qué momento las cadenas han “olvidado” sus puntos de inicio. Hay muchas formas posibles de comparar cadenas paralelas, en esta tesis se describirán dos métodos de diagnóstico de convergencia, el de [Gelman y Rubin \(1992\)](#) y el de [Raftery y Lewis \(1992\)](#).

### 4.2.1. Prueba de convergencia de Gelman y Rubin

En 1992, Gelman y Rubin propusieron una aproximación cuantitativa para monitorear la convergencia de una MCMC; el método se basa en monitorear por separado la convergencia de todas las cantidades de interés de una distribución y cuando la varianza entre las diferentes cadenas no es más grande que la varianza dentro de cada cadena individual, se declara convergencia.

Sea  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)'$  un vector de parámetros con cierta distribución final, supóngase que se está interesado en cada uno de los componentes  $\theta_i$ ,  $i = 1, 2, \dots, p$ . El monitoreo de la convergencia se realiza por separado para cada uno de los componentes  $\theta_i$ ; sin pérdida de generalidad sea  $\theta$  cualquier  $\theta_i$  y supóngase que se tienen  $m$  cadenas paralelas, cada una con longitud  $n$ . Para cada  $\theta$  se etiquetan las  $m$  cadenas paralelas de longitud  $n$  con  $(\theta_{ij})$ ,  $j = 1, 2, \dots, n$ ;  $i = 1, 2, \dots, m$  y se obtiene la varianza  $B$ , entre las cadenas y la varianza  $W$ , dentro de las cadenas como sigue:

## 4.2. Pruebas de Convergencia

---

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{\theta}_i - \bar{\theta}_{..})^2,$$

donde  $\bar{\theta}_i = \frac{1}{n} \sum_{j=1}^n \theta_{ij}$ ,  $\bar{\theta}_{..} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$ . y

$$W = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (\theta_{ij} - \bar{\theta}_i)^2. \quad (4.1)$$

La varianza  $var(\theta | y)$  de la marginal de  $\theta$  que se estima como el promedio ponderado de  $W$  y  $B$ , i.e.,

$$\widehat{var}(\theta | y) = \frac{n-1}{n} W + \frac{1}{n} B, \quad (4.2)$$

es un estimador insesgado bajo estacionariedad, pero que sobre estima la  $var(\theta | y)$  asumiendo que la distribución de inicio está apropiadamente sobredispersada. Así, si se tiene una sobredispersión, entonces  $\widehat{var}(\theta | y)$  es un estimador conservador de la varianza de  $\theta$ . Mientras que para alguna  $n$  finita, la varianza dentro de las cadenas,  $W$  subestima la varianza de  $\theta$ , ya que cada cadena tiene menos variabilidad que la distribución de todas las cadenas en el límite, cuando  $n \rightarrow \infty$ ,  $\widehat{var}(\theta | y)$  y  $W$  aproximan a  $var(\theta | y)$  en direcciones opuestas.

Se puede monitorear la convergencia de la cadena de Markov, estimando  $R$ , el factor por el cual la escala de la distribución actual de  $\theta$  puede ser reducido si las simulaciones son continuadas indefinidamente. Esta reducción potencial de escala puede ser estimada usando (4.1) y (4.2), de la siguiente manera:

$$\widehat{R} = \sqrt{\frac{\widehat{var}(\theta)}{W}}$$

y es denominado “estimador de reducción potencial de escala”.

Cuando la cadena converge,  $\widehat{R}$  se acerca a 1, significando que las cadenas de Markov están esencialmente sobrepuestas y provienen de la misma distribución. Si  $\widehat{R}$  es mayor que 1, entonces tendrá razón para creer que procediendo con más simulaciones se puede mejorar las inferencias sobre  $\theta$ .

Se recomienda calcular el estimador de reducción potencial de escala para todas las cantidades de interés, descartando la primera mitad de simulaciones de cada cadena. Si

## 4.2. Pruebas de Convergencia

---

$\widehat{R}$  estimada no esta cercano a 1 para todos los componentes de interés, una buena idea es continuar la simulación. En la práctica se corren simulaciones hasta que los valores  $\widehat{R}$  son todos menores que 1.1 ó 1.2.

Una vez que  $\widehat{R}$  este cercano a 1 para todas las cantidades de interés, las inferencias se realizan con la muestra obtenida al mezclar las simulaciones de las segundas mitades de las  $m$  cadenas paralelas. La estimaciones son confiables si están basadas en cadenas múltiples con punto de inicio sobredispersos. Las estimaciones obtenidas antes de la convergencia serán conservadoras y una vez que la convergencia ha sido alcanzada, éstas serán más precisas. Para más detalle ver también [Gelman \*et al.\* \(2004\)](#).

Para poder aplicar ésta prueba se necesita como mínimo 2 Cadenas de Markov Monte Carlo, las cuales como ya se mencionó anteriormente deben de tener diferentes puntos de inicio.

### 4.2.2. Prueba de convergencia de Raftery y Lewis

Este método de convergencia, introducido por [Raftery y Lewis \(1992\)](#), se basa en un criterio de exactitud de la estimación del cuantil  $q$  de una función real del vector  $\theta$ . Se determina el número de iteraciones  $N$  que se requiere para estimar el cuantil  $q$  dentro de una exactitud de  $\pm r$  con probabilidad  $s$ . También se obtiene  $M$ ,  $N_m$  e  $I$ , donde  $M$  es el número de iteraciones iniciales que deben ser descartadas (“burn-in”),  $N_m$  es el tamaño de muestra mínimo basado en una muestra aleatoria e  $I$  es el factor de dependencia que estima la magnitud en la cual la autocorrelación incrementa el tamaño de muestra requerido, y se calcula como  $I = (M + N)/N_m$ . Valores de  $I$  mayores que 5 indican una fuerte autocorrelación, la cual puede ser resultado de una pobre elección de valores de inicio de la cadena o una alta correlación entre los componentes del vector  $\theta$ , relativa a su distribución final.

El tamaño de muestra estimada para una variable  $U$  está basada en el proceso  $Z_t = d(U_t \leq u)$  donde  $d$  es la función indicadora y  $u$  es el cuantil  $q$  de  $U$ . El proceso  $Z_t$  es derivado de la cadena de Markov de los datos por marginalización y truncamiento, pero no es por sí misma una cadena de Markov. Sin embargo,  $Z_t$  puede comportarse como una cadena de Markov si ésta está suficientemente dispersa.

El criterio para determinar el número de iteraciones “burn-in”  $m$ , es que la distribución condicional de  $Z_t^k$  dado  $Z_0$  debe estar dentro de la distribución de equilibrio de la cadena  $Z_t^k$ .



# Capítulo 5

## Algoritmos de Muestreo

### 5.1. Algoritmo del muestreador de Gibbs

Supóngase que se desea una muestra aleatoria de la distribución final  $\pi(\boldsymbol{\theta} \mid \mathbf{X})$ , donde  $\mathbf{X}$  denota los datos en cuestión y  $\boldsymbol{\theta}$  un vector de parámetros el cual es difícil o imposible de muestrear de  $\pi(\boldsymbol{\theta} \mid \mathbf{X})$  dado que no se reconoce el kernel de la distribución final.

El muestreo de Gibbs, introducido por [Geman y Geman \(1984\)](#) en el contexto de reparación de imágenes, provee un método para generar muestras de  $\pi(\boldsymbol{\theta} \mid \mathbf{X})$ .

Sea  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)'$ , donde  $\theta_i$  es unidimensional o multidimensional y supóngase que se puede muestrear de las densidades condicionales finales  $\pi(\boldsymbol{\theta}_i \mid \mathbf{X}, \boldsymbol{\theta}_j, j \neq i)$ . El muestreador de Gibbs genera una cadena de Markov como sigue. Se comienza con  $\boldsymbol{\theta}_0 = (\theta_1^0, \theta_2^0, \dots, \theta_q^0)'$  en el soporte de  $\pi(\boldsymbol{\theta} \mid \mathbf{X})$ . En el tiempo  $t$ , la realización  $\boldsymbol{\theta}_t = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_q^{(t)})'$  se obtiene de la siguiente manera:

1. Se hace  $t = 1$ .
2. Se genera  $\theta_1^{(t)}$  de  $p(\theta_1 \mid \mathbf{X}, \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_q^{(t-1)})$ .
3. Se genera  $\theta_2^{(t)}$  de  $p(\theta_2 \mid \mathbf{X}, \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_q^{(t-1)})$ .
4. Se genera  $\theta_q^{(t)}$  de  $p(\theta_q \mid \mathbf{X}, \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{q-1}^{(t)})$ .

Bajo condiciones de regularidad, la cadena de Markov  $\{\theta_0, \theta_1, \theta_2, \dots, \theta_t, \dots\}$  tiene distribución estacionaria  $\pi(\boldsymbol{\theta} \mid \mathbf{X})$  ([Gelfand y Smith, 1990](#)). La razón de convergencia depende de las correlaciones entre los componentes del vector  $\boldsymbol{\theta}$  ([Gilks \*et al.\*, 1996](#)).

## 5.2. Algoritmo ARS (Muestreo de Aceptación y Rechazo Adaptativo)

La técnica de muestreo de ARS es un método propuesto por [Gilks y Wild \(1992\)](#) para muestrear eficientemente funciones log-cóncavas univariadas. Las funciones log-cóncavas forman una familia que incluye una amplia variedad de funciones de densidades, ver [Gilks y Wild \(1992\)](#) y [Devroye \(1986\)](#). El algoritmo ARS está basado en el hecho de que cuando una función log-cóncava es evaluada en un conjunto de puntos de su soporte, entonces las tangentes definen una función lineal que la acota superiormente y las cuerdas una función lineal que la acota inferiormente. Estas funciones son utilizadas para desarrollar un muestreo de rechazo.

### 5.2.1. Implementación

Supóngase que se desea muestrear de la función de densidad  $f(x)$  con soporte  $\Psi$ , y proporcional a  $g(x)$  donde  $h(x) = \log g(x)$  es cóncava, es decir, es dos veces diferenciable y la segunda derivada es negativa en todo  $\Psi$ . Sean  $x_1, x_2, \dots, x_k$  puntos ordenados del conjunto  $\Psi$ , tal que la moda de  $h(x)$  se encuentra entre  $x_1$  y  $x_k$ , y  $S_k := \{x_i\}_{i=1}^k$  y sean las funciones envolventes inferior  $l_k(x)$  y superior  $u_k(x)$  definidas para cualquier punto  $x \in [x_i, x_{i+1}]$ :

$$l_k(x) = h(x_i) + \frac{x_{i+1} - x_i}{h(x_{i+1}) - h(x_i)}(x - x_i),$$

$$u_k(x) = \begin{cases} h(x_i) + h'(x_i)(x - x_i), & x_i \leq x \leq z_i, \\ h(x_{i+1}) + h'(x_{i+1})(x - x_{i+1}), & z_i < x \leq x_{i+1}, \end{cases}$$

donde  $h'(\cdot)$  es la función derivada de  $h(\cdot)$  y  $z_i$  el punto de intercepción entre las tangentes en los puntos  $(x_i, h(x_i))$  y  $(x_{i+1}, h(x_{i+1}))$ , está dado por:

$$z_i = x_i + \frac{h(x_i) - h(x_{i+1}) + h'(x_{i+1})(x_{i+1} - x_i)}{h'(x_{i+1}) - h'(x_i)}.$$

Nótese que  $x_i < z_i < x_{i+1}$ ,  $i = 1, \dots, k - 1$ , y que  $z_i < z_{i+1}$ ,  $i = 1, \dots, k - 2$ , donde  $z_0 = 0$  y  $z_k = \infty$ .

Sean  $G_k(x)$  y  $F_{G_k}(\cdot)$  la función exponencial normalizada de  $u_k(x)$  a lo largo del intervalo  $[x_i, x_k]$  y la función de distribución acumulada de la densidad exponencial  $G_k(\cdot)$

## 5.2. Algoritmo ARS (Muestreo de Aceptación y Rechazo Adaptativo)

---

respectivamente,

$$G_k(x) := \frac{\exp(u_k(x))}{c_u}, x_i \leq x \leq x_k,$$

$$F_{G_K}(z_i) = \frac{1}{c_u} \sum_{j=0}^i \frac{[\exp(u_k(z_{j+1})) - \exp(u_k(z_j))]}{h'(x_{j+1})},$$

donde

$$c_u = \int_{x_1}^{x_k} \exp(u_k(x)) dx = \sum_{i=0}^{k-1} \frac{[\exp(u_k(z_{i+1})) - \exp(u_k(z_i))]}{h'(x_{i+1})}.$$

Así su valor  $x$  muestreado de la densidad  $G_k(x)$  es obtenido por

$$x = z_{[i]} + \frac{1}{h'(x_{i+1})} \log \left[ 1 + \frac{c_u h'(x_{i+1})(u - F_{G_K}(z_{[i]}))}{\exp(u_k(z_{[i]}))} \right], \quad (5.1)$$

donde  $u$  es una realización de la distribución uniforme en  $(0,1)$  y

$$z_{[i]} = \max \{z_i, i = 1, \dots, k \mid F_{G_K}(z_i) \leq u\}. \quad (5.2)$$

### 5.2.2. Implementación del algoritmo ARS

Para obtener una muestra aleatoria de la densidad  $f$  usando este algoritmo, se realizan los siguiente pasos.

1. Elegir puntos  $x_1, x_2, \dots, x_k$  de  $\Psi$ , tal que la moda de  $h(x)$  se encuentra entre  $x_1$  y  $x_k$  e inicializar el conjunto  $S_k = \{x_i\}_{i=1}^k$ .
2. Muestrear  $u$  de la distribución uniforme en  $(0,1)$
3. Determinar  $z_{[i]}$  en (5.2).
4. Obtener  $x$  en (5.1).

### 5.3. Algoritmo Metropolis Hasting

---

5. Muestrear  $u$  en la distribución uniforme en  $(0,1)$ .
6. Realizar la siguiente prueba: si  $u \leq \exp(l_k(x) - u_k(x))$  o si  $u \leq \exp(h(x) - u_k(x))$  entonces aceptar  $x$ , de otro modo rechazar  $x$ .
7. Incluir  $x$  al conjunto  $S_k$ .
8. Repetir  $B$  veces (número muestras deseado) los pasos 2 al 7.

Para más detalles sobre al algoritmo ARS, ver [Gilks y Wild \(1992\)](#).

### 5.3. Algoritmo Metropolis Hasting

El algoritmo de Metropolis-Hastings es un método Monte Carlo vía Cadenas de Markov que ha sido ampliamente usado en la física matemática y restauración de imágenes por muchos años, pero recientemente descubierto por los estadísticos. El método en su forma más simple lo publica [Metropolis \*et al.\* \(1953\)](#). Dada la distribución a posteriori de un parámetro  $\boldsymbol{\theta}$  es  $\pi(\boldsymbol{\theta} | \mathbf{X})$ , se escoge una función auxiliar  $J(\theta_t | \theta_t)$ , tal que  $J(\cdot | \theta_t)$  es una función de densidad de probabilidad para toda  $\theta_t$ , además es simétrica, es decir:

$$J(\theta^* | \theta) = J(\theta | \theta^*).$$

La función  $J$  es llamada distribución generadora de candidatos. Entonces el algoritmo de Metropolis genera una cadena de Markov,  $(\theta_0, \theta_1, \dots, \theta_k, \dots)'$  donde la probabilidad de transición de  $\theta_t$  a  $\theta_{t+1}$  ya no depende de  $\theta_0, \theta_1, \dots, \theta_{t-1}$ , y cuya distribución de transición es:

$$P(\theta_{t+1} | \theta_t) = \alpha(\theta_{t+1}, \theta_t) J(\theta_{t+1}, \theta_t).$$

Así la cadena se aproxima a una distribución estacionaria y las muestras del vector  $(\theta_{k+1}, \theta_{k+2}, \dots, \theta_{k+n})'$  serán muestras de la densidad objetivo  $\pi(\boldsymbol{\theta} | \mathbf{X})$ .

#### 5.3.1. Implementación del algoritmo Metropolis Hasting

Pasos:

1. Se elige arbitrariamente un valor inicial  $\boldsymbol{\theta}^{(0)}$  de la densidad a posteriori de  $\pi(\boldsymbol{\theta} | \mathbf{X})$  para el cual  $\pi(\boldsymbol{\theta}^{(0)} | \mathbf{X}) > 0$ .

## 5.4. Algoritmo “t-walk”

---

2. Para  $t = 1, 2, \dots, N$ .

a) Se elige un candidato  $\boldsymbol{\theta}^*$  de una distribución de salto  $J_t(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{t-1})$ . La distribución de salto debe ser simétrica, es decir  $J_t(u | v) = J_t(v | u)$ , para todo  $u, v$  y  $t$ .

b) Se calcula la razón o el cociente de importancia de las densidades,

$$r = \frac{f(\boldsymbol{\theta}^* | X)}{f(\boldsymbol{\theta}^{t-1} | X)}.$$

c) Si  $r > 1$  entonces  $\boldsymbol{\theta}^t = \boldsymbol{\theta}^*$ , ir al paso 2.

d) Si  $r < 1$  entonces generar  $u \sim Unif(0, 1)$ , si  $u < r$  entonces  $\boldsymbol{\theta}^t = \boldsymbol{\theta}^*$ .

e) Ir al paso 2.a hasta que  $t = N$ .

## 5.4. Algoritmo “t-walk”

El algoritmo “t-walk” ([Christen y Fox, 2010](#)) es un método de Cadenas de Markov Monte Carlo que permite muestrear de distribuciones continuas con estructuras de correlación arbitrarias, construido como un algoritmo de Metropolis-Hastings (por partes) en el espacio producto. El algoritmo, que trabaja como una caja negra, requiere una función objetivo,  $p(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta} \in \mathfrak{R}^m$  que puede ser una distribución a posterior o cualquier otra distribución de la que se desea muestrear y dos puntos  $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$  tales que  $p(\boldsymbol{\theta}_1) > 0$ ,  $p(\boldsymbol{\theta}_2) > 0$ , a partir de los cuales se generan movimientos en varios sentidos para generar nuevos candidatos. El algoritmo parece trabajar de manera adecuada ya que los resultados obtenidos en el ejemplo práctico son coherentes con la realidad, en el escrito de los mencionados autores se encuentran varios ejemplos de su uso, quienes además implementan el algoritmo en el programa [R Development Core Team \(2011\)](#) y C++, [Stroustrup \(1963\)](#).

# Capítulo 6

## Distribución Pareto

### 6.1. Antecedentes

La distribución Pareto ha tenido un auge durante las últimas dos décadas debido a la aplicación que ha tenido para modelar riqueza e ingreso. Esta distribución se le atribuye a [Pareto \(1895\)](#), y análogamente con la curva de Lorenz son el pilar de las distribuciones estadísticas de ingreso.

[Pareto \(1895\)](#) observó una relación lineal decreciente entre el logaritmo de los ingresos y el logaritmo de  $N_x$ , el número de contribuyentes con ingreso mayor o igual a  $x$ ,  $x \geq x_0$ , cuando se analizan ingresos declarados para efecto del impuesto sobre la renta. Así, él especifica:

$$\log N_x = A - \alpha \log x,$$

que es lo mismo que,

$$N_x = e^A x^{-\alpha},$$

donde  $A, \alpha > 0$ . Normalizando por el número de personas que reciben ingreso  $N : N_{x_0}$ , uno obtiene:

$$\frac{N_x}{N} = 1 - F(x) = \left(\frac{x}{x_0}\right)^{-\alpha}, x \geq x_0 > 0.$$

## 6.2. Definición

---

Casi inmediatamente, el interés público se despertó y otros economistas comenzaron a criticar la idea de una forma universal con un parámetro de forma única que permite comparaciones inadecuadas entre las sociedades. Se hizo evidente muy pronto que la distribución de Pareto sólo es una buena aproximación de los ingresos por encima de un determinado umbral y que la misteriosa y notable  $\alpha$  no está siempre cerca de 1.5 como Pareto inicialmente creía. Sin embargo más tarde [Davis \(1941\)](#) considera el valor de  $\alpha = 1.5$  para hacer una línea divisoria entre sociedades igualitarias ( $\alpha > 1.5$ ) y las desiguales ( $\alpha < 1.5$ ). [Kakwani \(1980b.\)](#), en una reminiscencia de [Mandelbrot \(1997\)](#) se refiere a la distribución de Pareto con un  $\alpha = 1.5$  como la más fuerte ley de Pareto.

## 6.2. Definición

La distribución de Pareto clásica está definida en términos de la siguiente función de distribución acumulada (f.d.a.):

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha}, \quad x \geq x_0 > 0,$$

donde  $\alpha > 0$  es el parámetro de forma (también medida de la cola derecha) y  $x_0$  el parámetro de escala.

## 6.3. Momentos y otras propiedades básicas

La densidad de Pareto tiene una cola derecha más pesada conforme  $\alpha$  es más pequeña, implicando que solo existan momentos de menor orden. En particular, el  $k$ -ésimo momento de la distribución de Pareto existe solo si  $k < \alpha$ , en ese caso se tiene:

$$E(X^k) = \frac{\alpha x_0^k}{\alpha - k},$$

con media:

$$E(X) = \frac{\alpha x_0}{\alpha - 1},$$

y varianza igual a:

## 6.4. Estimadores de máxima verosimilitud de la distribución Pareto

---

$$\text{Var}(X) = \frac{\alpha x_0^2}{\alpha(\alpha - 1)^2(\alpha - 2)}, \alpha > 2.$$

Así el coeficiente de variación esta dado por:

$$CV = \frac{1}{\sqrt{\alpha(\alpha - 2)}}, \alpha > 2.$$

## 6.4. Estimadores de máxima verosimilitud de la distribución Pareto

Los Estimadores de máxima verosimilitud (EMV) son los siguientes:

$$\hat{\tau} = \frac{1}{\min(x_1, x_2, \dots, x_n)},$$

$$\hat{\alpha} = \frac{n}{(n \log \hat{\tau} + \sum_{i=1}^n \log x_i)}$$

## 6.5. Estimadores Bayesianos para la distribución Pareto

Suponga se que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria i.i.d con función de distribución Pareto, con parámetro de forma o parámetro de desigualdad  $\alpha$  y parámetro de precisión (el inverso del parámetro de escala  $x_0$ )  $\tau$ , entonces la función de verosimilitud puede ser escrita de la siguiente manera:

$$L_{\mathbf{X}}(\mathbf{X}; \alpha, \tau) = \prod_{i=1}^n \tau \alpha (\tau x_i)^{-(\alpha+1)} I(\tau x_i > 1) = \alpha^n \tau^{-n\alpha} \left( \prod_{i=1}^n x_i \right)^{-(\alpha+1)} I(\tau x_{1:n} > 1),$$

lo anterior puede ser reescrito de la siguiente forma:

$$L_{\mathbf{X}}(\mathbf{X}; \alpha, \tau) = \exp \left[ n \log \alpha - n\alpha \log \tau - \left( \sum_{i=1}^n \log x_i \right) (\alpha + 1) \right] I(\tau x_{1:n} > 1).$$

En éste caso en el que ambos parámetros  $\alpha, \tau$  son desconocidos e independientes, se usan



## 6.5. Estimadores Bayesianos para la distribución Pareto

---

las distribuciones a prioris propuestas por [Arnold et al. \(1998\)](#):

$(\alpha | \tau) \sim \text{Gamma}$ , mientras que  $(\tau | \alpha) \sim \text{Pareto}$

Suponiendo que las distribuciones a priori para  $\alpha$  y  $\tau$  son independientes, entonces la distribución conjunta es el producto de sus distribuciones a priori correspondiente, quedando;

$$f(\alpha, \tau) = \exp [b \log \tau + m_{12} \log \alpha \log \tau] \exp [a_1 \alpha + a_2 \log \alpha + m_{11} \alpha \log \tau] I(\tau c > 1).$$

Donde ambos parámetros del lado derecho de la ecuación anterior se refieren a los hiperparámetros, los cuales no son afectados por los datos, pero los datos si serán afectados por estos. Con lo anterior se puede llegar a que las densidades condicionales posteriores tienen distribución Pareto y Gamma.

Así, la densidad condicional de  $(\alpha | \tau)$  tiene una distribución Gamma con parámetro de forma  $\gamma(\tau)$  y parámetro de escala  $\lambda(\tau)$ , i.e.:

$$f(\alpha | \tau) \propto \alpha^{\gamma(\tau)-1} e^{-\lambda(\tau)\alpha},$$

donde su media y varianza son:

$$E(\alpha | \tau) = \frac{-(1 + a_2 + m_{12} \log \tau)}{(a_1 + m_{11} \log \tau)}$$

$$\text{Var}(\alpha | \tau) = \frac{(1 + a_2 + m_{12} \log \tau)}{(a_1 + m_{11} \log \tau)^2}$$

Mientras que la densidad condicional de  $(\tau | \alpha)$  tiene una distribución Pareto con parámetro de forma  $\delta(\alpha)$  y parámetro de precisión  $\nu(\alpha)$ , i.e.

$$f(\tau | \alpha) \propto \nu(\alpha) \delta(\alpha) [\nu(\alpha) \tau]^{-(\delta(\alpha)+1)} I(\nu(\alpha) \tau > 1),$$

donde:

$$\delta(\alpha) = -(1 + b + m_{11} \alpha + m_{12} \log \alpha),$$

$$\nu(\alpha) = c.$$

## 6.6. Implementación del algoritmo de muestreo Gibbs

---

Cabe mencionar que si en algún análisis realizado con estas distribuciones propuestas por [Arnold et al. \(1998\)](#), se desea trabajar con densidades a priori propias, es necesario mencionar las siguientes restricciones a los hiperparámetros;  $m_{12} > 0; m_{11} < 0; a_1 < m_{11} \log c; a_2 > m_{12} \log c - 1; c > 0; -b + m_{12}(1 - \log(-m_{12}/m_{11})) > 0$ , para que los siguientes parámetros  $\gamma(\tau)$ ,  $\lambda(\tau)$ ,  $\nu(\alpha)$  de las densidades posteriores condicionales sean siempre positivos y  $\delta(\tau) > -1$ .

## 6.6. Implementación del algoritmo de muestreo Gibbs

El problema a resolver como se menciona en el planeamiento del mismo, es obtener los estimadores bayesianos (EB) de las distribución Pareto, por lo que en esta densidad el método a usar para encontrar los estimadores posteriores de  $\hat{\alpha}$  y  $\hat{\tau}$  es el muestreador de Gibbs ya que no se reconoce el kernel de la densidad posterior pero si se conocen las densidades condicionales posteriores de ambos parámetros.

Objetivo: obtener una muestra aleatoria de los EB de  $f(\boldsymbol{\theta} | \mathbf{X})$ , donde  $\boldsymbol{\theta} = (\alpha, \tau)'$ .

Se sabe que  $(\alpha | \tau) \sim \text{Gamma}$ , mientras que  $(\tau | \alpha) \sim \text{Pareto}$ .

Valores de inicio de  $\boldsymbol{\theta}_0 = (\alpha_0, \tau_0)'$  serán sus respectivos EMV, mientras que los valores de las constantes  $a_1, a_2, b, m_{11}, m_{12}, c$ , serán los propuestos por [Arnold et al. \(1998\)](#).

### 6.6.1. Metodología

**Muestreo de la distribución conjunta posterior mediante el algoritmo del Gibbs**

1. Se comienza con los valores iniciales de  $(\alpha_0, \tau_0)'$ .
2. Hacer  $t = 1$ .
3. Generar de  $(\alpha_t | \tau_{t-1})$ .
4. Generar de  $(\tau_t | \alpha_t)$ .
5. Hacer  $t = t + 1$ .
6. Repetir los pasos 3, 4 y 5  $k$  veces, con lo cual se obtendrá la secuencia  $(\alpha_0, \tau_0)', (\alpha_1, \tau_1)', \dots, (\alpha_k, \tau_k)'$ .

Así se obtiene una muestra posterior conjunta de los parámetros  $\alpha$  y  $\tau$  de tamaño  $k$ .

# Capítulo 7

## Distribución log-normal

### 7.1. Antecedentes

En la economía cuantitativa y financiera la distribución lognormal esta altamente relacionada con el movimiento browniano geométrico el cual en términos generales se refiere al movimiento aleatorio de partículas microscópicas que se hallan en un medio fluido (ejemplo, polen en una gota de agua). Este recibe el nombre en honor al escocés [Brown \(1828\)](#), biólogo y botánico el cual descubrió éste fenómeno en 1827 y observó los movimientos aleatorios aparentemente sin razón alguna de las pequeñas partículas de polen.

Este término fue adoptado por la economía, así en 1900 Louis Bachelier presenta su tesis doctoral “Théorie de la speculation” en la cual se adelanta varias décadas presentando conceptos tales como el movimiento Browniano como un modelo para el precio de una acción a través del tiempo y, la valuación de opciones del tipo europeo. Dando preámbulo a la teoría de valores extremos y de distribuciones de colas pesadas, dentro de la cual la distribución lognormal es de las más socorridos en estas disciplinas.

Aunque en lo que a distribuciones de ajuste económico concierne, el estudio pionero en el uso de la distribución lognormal fue la tesis de [Gibrat \(1931\)](#), el cual afirmó que el ingreso de un individuo (o una empresa) puede considerarse como el efecto conjunto de un largo número de causas independientes entre sí que han trabajado durante un largo periodo de tiempo. En un cierto tiempo  $t$  determinado se supone que el cambio en alguna variable  $X$  es una proporción de una función  $g(X_{t-1})$  de el valor  $X_{t-1}$  ya alcanzado. Así, la ley fundamental de movimiento es

$$X_t - X_{t-1} = \epsilon_t g(X_{t-1}),$$

## 7.1. Antecedentes

---

donde los  $\epsilon_t$ 's son mutuamente independientes de  $X_{t-1}$ . En el caso especial donde  $g(X) \equiv X$ , es la llamada ley del del efecto proporcional resultante, significando que el cambio en la variable aleatoria en cualquier paso del proceso es una proporción aleatoria del valor previo de la variable, así,

$$X_t - X_{t-1} = \epsilon_t X_{t-1},$$

lo cual puede ser reescrito como

$$\frac{X_t - X_{t-1}}{X_{t-1}} = \epsilon_t,$$

aplicando la sumatoria respecto a  $t$  en ambos términos

$$\sum_{t=1}^n \frac{X_t - X_{t-1}}{X_{t-1}} = \sum_{t=1}^n \epsilon_t.$$

Asumiendo que el efecto en cada paso es pequeño, se obtiene

$$\sum_{t=1}^n \frac{X_t - X_{t-1}}{X_{t-1}} \approx \int_{X_0}^{X_n} \frac{dX}{X} = \log X_n - \log X_0,$$

despejando

$$\log X_n = \log X_0 + \sum_{t=1}^n \epsilon_t.$$

Si se supone un adecuado teorema central del límite (TCL), el cual dice que  $X_n$  tendrá una distribución asintóticamente lognormal. En la derivación original los  $\epsilon_t$  se asumen i.i.d. para los cuales el TCL de Lindeberg–Lévy es suficiente para la solución del problema, aunque en el que caso de que los  $\epsilon_t$  sean heterogéneos se podría usar un TCL más general, ver [Hart \(1973\)](#).

## 7.2. Definición

Si una variable aleatoria  $X$  tiene una distribución lognormal, su f.d.p. está dada por la siguiente expresión:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\}, x > 0.$$

En consecuencia, la densidad surge como la distribución de una variable aleatoria  $X = \exp(Y)$ , donde  $Y \sim N(\mu, \sigma^2)$ .

Así, su f.d.a. está dada por:

$$F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), x > 0,$$

donde  $\Phi(\cdot)$  denota la f.d.p. de la distribución normal estándar.

## 7.3. Momentos y otras propiedades básicas

### 7.3.1. Propiedades

En ésta distribución claramente se observa la cercana relación con la distribución normal, por lo que muchas propiedades de la distribución lognormal provienen directamente de la distribución normal. Por ejemplo muestras aleatorias de una población lognormal son comúnmente generados de números aleatorios normales via exponencial. La principal diferencia es el nuevo rol de los parámetros:  $\exp(\mu)$  se convierte en el parámetro de escala, mientras que  $\sigma$  es ahora el parámetro de forma. Otra de las principales atracciones de la distribución normal, son las propiedades estables bajo adición. Así la suma de variables aleatorias normales independientes también tienen distribución normal con nuevos parámetros.

Lo anterior se traslada a la siguiente propiedad que se establece bajo producto en la distribución lognormal:

Para  $X_i \sim LN(\mu_i, \sigma_i^2)$  independientes, donde  $i = 1, 2$

### 7.3. Momentos y otras propiedades básicas

---

$$X_1 X_2 \sim LN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2),$$

en particular para una muestra de variables aleatorias lognormales con media geométrica de tamaño  $n$  i.i.d,

$$\left( \prod_{i=1}^n X_i \right)^{1/n} \sim LN \left( \mu, \frac{\sigma^2}{n} \right).$$

Sin embargo, la suma de variables aleatorias no son muy tratables, como fue notado por [Mandelbrot \(1997\)](#).

La estructura de la distribución lognormal la hace muy conveniente para poder expresar sus momentos. Los cuales son obtenidos de la función generatriz de momentos  $m_Y(\cdot)$  de la distribución normal.

$$E(X^k) = E(e^{kY}) = \exp \left( k\mu + \frac{1}{2} k^2 \mu^2 \right).$$

Así la distribución lognormal es una de las pocas distribuciones bien conocidas que posee momentos de todos los ordenes, positivo, negativo y fraccional. Sin embargo la secuencia de momentos enteros no caracteriza a la distribución lognormal.

#### 7.3.2. Momentos de bajo orden

La media de la distribución lognormal está dada por:

$$E(X) = \exp \left( \frac{\mu + \sigma^2}{2} \right),$$

y su varianza está dada por la siguiente expresión:

$$var(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) = e^{2\mu} \omega (\omega - 1),$$

donde  $\omega = e^{\sigma^2}$ , así su coeficiente de variación es:

$$CV(X) = \sqrt{\exp(\sigma^2) - 1} = \sqrt{\omega - 1}$$

### 7.3.3. Estimadores de máxima verosimilitud para la distribución Lognormal

Dado que existe una relación muy cercana con la distribución normal, los EMV para la distribución lognormal quedan de la siguiente manera:

$$\hat{\mu} = \frac{1}{n} \sum_i^n \log x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i^n (\log x_i - \overline{\log x})^2$$

## 7.4. Estimadores Bayesianos para la distribución lognormal

Supóngase que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria i.i.d los cuales siguen una función de distribución Lognormal, con parámetro de escala  $\mu$  y parámetro de forma  $\sigma$ , entonces su función de verosimilitud puede ser escrita de la siguiente manera:

$$L_{\mathbf{X}}(\mathbf{x}; \mu, \sigma) = \prod_i^n \frac{1}{x_i \sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_i^n (\log x_i - \mu)^2 \right\}.$$

Se consideran las siguientes distribuciones a prioris no informativas, debido a que se carece de información previa de los parámetros.

A priori no informativa de  $\pi(\mu) \propto 1$ , y a priori de  $\pi(\sigma) \propto \frac{1}{\sigma}$ , dado que se supone independencia entre parámetros, su distribución a priori conjunta es el producto de ambas, quedando la siguiente expresión:

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma}.$$

## 7.5. Implementación del ARS dentro del muestreador de Gibbs

---

Así el producto de la distribución a priori conjunta y la verosimilitud, proporciona la distribución a posteriori conjunta, quedando:

$$\pi(\mu, \sigma | \mathbf{x}) \propto \left(\frac{1}{\sigma}\right)^{n+1} \left\{ \prod_i^n \frac{1}{x_i} \right\} \exp \left[ -\frac{1}{2\sigma^2} \sum_i^n (\log x_i - \mu)^2 \right].$$

Así usando la ecuación anterior se obtiene las siguientes distribuciones condicionales posteriores de los parámetros.

$$\pi(\mu | \sigma, \mathbf{x}) \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_i^n (\log x_i - \mu)^2 \right] \quad (7.1)$$

$$\pi(\sigma | \mu, \mathbf{x}) \propto \left(\frac{1}{\sigma}\right)^{n+1} \exp \left[ -\frac{1}{2\sigma^2} \sum_i^n (\log x_i - \mu)^2 \right]. \quad (7.2)$$

De la ecuación (7.1), se puede demostrar que para  $\mu$  la función es log cóncava y dado eso, se puede hacer uso del método ARS para obtener una muestra aleatoria de esta distribución condicional posterior. Por otra parte de la ecuación (7.2) respecto a  $\sigma$ , ha sido demostrado por [Upadhyay, Vasishta y Smith \(2000\)](#) que  $\lambda = 1/\sigma$  tiene una distribución Gamma con parámetro de forma  $n/2$  y parámetro de escala  $2 / [\sum_i^n \{\log(x_i) - \mu\}^2]$ .

## 7.5. Implementación del ARS dentro del muestreador de Gibbs

El problema a resolver como se menciona en el planeamiento del mismo, es obtener los EB de las distribución Lognormal, por lo que en esta densidad el método a usar para encontrar los estimadores posteriores de  $\hat{\mu}$  y  $\hat{\sigma}$  es el algoritmo ARS dentro del muestreador Gibbs, ya que no se reconoce el kernel de la densidad posterior pero si se conoce la densidad condicional posterior del parámetro  $\mu$  y se sabe que la función condicional posterior del parámetro  $\sigma$  es log cóncava.

**Objetivo:** obtener una muestra aleatoria de los parámetros  $\pi(\boldsymbol{\theta} | \mathbf{X})$ , donde  $\boldsymbol{\theta} = (\mu, \sigma)'$ .

Se sabe que  $\pi(\mu | \sigma)$  es una función log cóncava, mientras que  $\pi(\sigma | \mu) \sim 1/\text{Gamma}$ . Y sus valores de inicio de  $\mu, \sigma$  serán sus respectivos EMV.



## 7.6. Metodología

Muestreo de la distribución conjunta posterior mediante el algoritmo del Gibbs y ARS

1. Se desea obtener una muestra de tamaño  $k$  de  $\pi(\mu, \sigma \mid \mathbf{X})$ .
2. Se comienza con los valores iniciales de  $(\mu_0, \sigma_0)$ .
3. Hacer  $t = 1$ .
4. Generar de  $(\mu_t \mid \sigma_{t-1})$  usando el comando *ars* del paquete adaptado por [Paulino Perez Rodriguez \(2009\)](#), que se encuentra disponible en el programa [R Development Core Team \(2011\)](#).
5. Generar  $(\sigma_t \mid \mu_t)$ .
6. Hacer  $t = t + 1$ .
7. Repetir los pasos 4, 5 y 6  $k$  veces, con lo cual se obtendrá la secuencia  $(\mu_0, \sigma_0)'$ ,  $(\mu_1, \sigma_1)'$ , ...,  $(\mu_k, \sigma_k)'$ .

Así se obtiene una muestra posterior conjunta de los parámetros  $\mu, \sigma$  de tamaño  $k$ .

# Capítulo 8

## Distribución Dagum

### 8.1. Antecedentes

Aunque presentada como una distribución del ingreso sólo un año después del modelo Singh-Maddala, la distribución de Dagum es menos conocida. Presumiblemente, esto debido al hecho de que el trabajo de Camilo Dagum fue publicado en la revista francesa “Economie Appliquée”, mientras que el artículo de Singh-Maddala apareció en una de las más leídas “Econometrica”. Sin embargo, en los últimos años hay indicios de que la distribución Dagum es en realidad, una elección más apropiada en muchas aplicaciones.

La distribución Dagum ha sido descubierta en varias ocasiones en diferentes campos de la ciencias. Aparentemente esto ocurrió por primera vez con Burr (1942) como el tercer ejemplo de la solución de su sistema de ecuaciones. Por lo que esta puede ser conocida como la distribución Burr III, también esta distribución se puede ver como una particular situación de la distribución Singh-Maddala, específicamente:

$$X \sim D(a, b, c) \iff \frac{1}{X} \sim SM(a, \frac{1}{b}, p).$$

Este resultado permite trasladar varios resultados pertinentes de la familia de distribución Singh-Maddala, que serán útiles en la distribución Dagum.

La Singh-Mandala es también llamada la distribución Burr III, o simplemente la distribución Burr, por lo que no debería sorprender, si a la distribución Dagum se le llama la distribución Burr inversa. Tanto la distribución Singh-Manddala como la Dagum pueden ser consideradas una distribución log-logística generalizada. El caso especial donde  $a = p$  es algunas veces llamado como la distribución paralogística inversa (Klugman y Willmot, 1998). Previo a su uso como una distribución de ingresos, la familia de distribución Da-

## 8.2. Definición

---

gum fue propuesta como una modelo de precipitación en la literatura meteorológica por [Mielke \(1973\)](#), donde esta es llamada distribución capa (de tres parámetros).

[Mielke y Johnson \(1974\)](#) anidaron a ésta dentro de la Beta Generalizada de segundo orden y fue llamada la distribución Beta-k. En un desarrollo paralelo-consiente de [Mielke \(1973\)](#) pero presuntamente inconsciente de [Dagum \(1977\)](#) - [Fattorini y Lemmi \(1979\)](#) propusieron la distribución como una distribución de ingresos, ver también [Lemmi \(1987\)](#). Sin embargo la distribución es usualmente llamada distribución de Dagum en la literatura de distribuciones que modelan ingresos.

[Dagum \(1977\)](#) derivó su modelo de la observación empírica en la elasticidad del ingreso, ya que la f.d.a. del ingreso es una función decreciente y acotada de F. Empezando de la ecuación diferencial,

$$\eta(x, F) = \frac{d \log F(x)}{d \log x} = ap \{1 - [F(x)^{1/p}]\}, x \geq 0, \quad (8.1)$$

sujeito a  $p > 0$  y  $ap > 0$ , uno obtiene la densidad,

$$f(x) = \frac{apx^{ap-1}}{b^{ap} [1 + (x/b)^a]^{p+1}}, x > 0. \quad (8.2)$$

[Fattorini y Lemmi \(1979\)](#) independientemente llegaron a la distribución Dagum como una distribución de equilibrio de un proceso estocástico de tiempo continuo bajo ciertas suposiciones tales como, media y varianza infinitesimal, ver también ([Dagum y Lemmi, 1989](#)).

## 8.2. Definición

La f.d.p de la distribución Dagum está dada por la expresión (8.2), donde  $a$ ,  $b$  y  $p > 0$ .

Por lo que su f.d.a. es está dada por la siguiente expresión:

$$F(x) = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-p}, x > 0.$$

## 8.3. Momentos y otras propiedades básicas

Como en el caso de varias distribuciones, la familia de distribuciones Dagum ha sido considerada en varios ocasiones con diferentes parametrizaciones. [Mielke \(1973\)](#) y más tarde [Fattorini y Lemmi \(1979\)](#) usaron  $(\alpha, \beta, \theta)' = (1/p, bp^{1/a}, ap)'$  mientras que [Dagum \(1977\)](#) empleo  $(\beta, \delta, \lambda)' = (p, a, b^a)'$ .

Así el  $k$ -ésimo momento existe si para  $-ap < k < a$ , y es igual a:

$$E(X^k) = \frac{b^k B(p + k/a, 1 - k/a)}{B(p, 1)} = \frac{b^k \Gamma(p + k/a) \Gamma(1 - k/a)}{\Gamma(p)}.$$

Por lo que la esperanza de una variable con distribución Dagum es:

$$E(X) = \frac{b \Gamma(p + 1/a) \Gamma(1 - 1/a)}{\Gamma(p)},$$

con varianza:

$$var(X) = \frac{b^2 \{ \Gamma(p) \Gamma(p + 2/a) \Gamma(1 - 2/a) - \Gamma^2(p + 1/a) \Gamma^2(1 - 1/a) \}}{\Gamma^2(p)},$$

y coeficiente de variación:

$$CV = \sqrt{\frac{\Gamma(p) \Gamma(p + 2/a) \Gamma(1 - 2/a)}{\Gamma^2(p + 1/a) \Gamma^2(1 - 1/a)} - 1}.$$

## 8.4. Estimadores de máxima verosimilitud de la distribución Dagum

En lo referente a la estimación de parámetros de esta distribución, [Dagum \(1977\)](#) propuso un criterio de mínimos cuadrados no lineales basado en la distancia entre la función de distribución empírica y la función de distribución Dagum,

## 8.4. Estimadores de máxima verosimilitud de la distribución Dagum

---

$$\sum_{i=1}^n \{F_n(x_i) - [1 + (x_i/b)^{-a}]^{-p}\}^2. \quad (8.3)$$

Más tarde, [Stoppa \(1995\)](#) consideró un estimador tipo regresión utilizando la elasticidad en (8.1).

En la actualidad, generalmente se estiman los parámetros de la distribución Dagum por el método de máxima verosimilitud. Cabe señalar que los estimadores de máxima verosimilitud no son muy estables por lo cual se utilizan métodos alternativos, como por ejemplo los basados en elasticidades, y ahora el método Bayesiano.

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño  $n$  de la distribución Dagum, su log-verosimilitud es:

$$\ell = n \log a + n \log p + (ap - 1) \sum_{i=1}^n \log x_i - nap \log b - (p + 1) \sum_{i=1}^n \log \left[ 1 + \left( \frac{x_i}{b} \right)^a \right]. \quad (8.4)$$

Se deben obtener los estimadores  $(\hat{a}, \hat{b}, \hat{p})'$  que maximizan (8.4). Al derivar con respecto a cada parámetro e igualar a cero se obtienen las siguientes ecuaciones:

$$\frac{n}{a} + p \sum_{i=1}^n \log(x_i/b) = (p + 1) \sum_{i=1}^n \frac{\log(x_i/b)}{1 + (b/x_i)^a},$$

$$np = (p + 1) \sum_{i=1}^n \frac{1}{1 + (b/x_i)^a},$$

$$\frac{n}{p} + a \sum_{i=1}^n \log(x_i/b) = \sum_{i=1}^n \log\{1 + (x_i/b)^a\},$$

las cuales deben de resolverse numéricamente.

## 8.5. Estimadores Bayesianos para la distribución Dagum

Supóngase que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria i.i.d con función de distribución Dagum, con parámetros de forma  $b, p$  y parámetro de escala  $a$ , entonces la función de verosimilitud puede ser escrita de la siguiente manera:

$$L(\mathbf{X} \mid a, b, p) = \frac{a^n p^n \left( \prod_{i=1}^n x_i \right)^{ap-1}}{b^{nap} \left[ \prod_{i=1}^n (x_i^a + b^a)^{p+1} / b^{na(p+1)} \right]},$$

donde los parámetros  $a, b, p > 0$  y la variable aleatoria  $\mathbf{X} > 0$ .

Para este caso en particular, debido a que no existen estudios anteriores sobre el ajuste del ingreso en México con la distribución Dagum desde un enfoque bayesiano, se desconoce el comportamiento de los parámetros de esta distribución por lo que se opta por distribuciones a priori no informativas y de la misma familia de distribuciones que la verosimilitud, así nuestras distribuciones a priori para los parámetros son las siguientes:

La variable aleatoria  $a \sim \text{Gamma}(\beta_0, \alpha_0)$ , donde  $\beta_0, \alpha_0 = 0.001$

$$\pi(a) = \frac{\beta_0}{\Gamma(\alpha_0)} b^{-(\alpha_0+1)} e^{-\beta_0 a}; a > 0.$$

La variable aleatoria  $b \sim \text{Gamma}(\beta_1, \alpha_1)$ , donde  $\beta_1, \alpha_1 = 0.001$

$$\pi(b) = \frac{\beta_1}{\Gamma(\alpha_1)} b^{-(\alpha_1+1)} e^{-\beta_1 a}; b > 0.$$

La variable aleatoria  $p \sim \text{Gamma}(\beta_2, \alpha_2)$ , donde  $\beta_2, \alpha_2 = 0.001$

$$\pi(p) = \frac{\beta_2}{\Gamma(\alpha_2)} b^{-(\alpha_2+1)} e^{-\beta_2 a}; p > 0.$$

Como se supone independencia entre las distribuciones a priori, la distribución a priori conjunta es el producto de las distribuciones a priori individuales,

## 8.5. Estimadores Bayesianos para la distribución Dagum

---

$$\pi(a, b, p) = \pi(a) \times \pi(b) \times \pi(p).$$

Así la distribución posterior conjunta queda de la siguiente manera:

$$\pi(a, b, p | \mathbf{X}) \propto L(\mathbf{X} | a, b, p) \times \pi(a, b, p)$$

$$\pi(a, b, p | \mathbf{X}) = \frac{a^n \beta_0^{\alpha_0} \beta_1^{\alpha_1} \beta_2^{\alpha_2} p^{n-(\alpha_2+1)} e^{(-\beta_0 a) - (\beta_1 b) - (\beta_2 p)} \left( \prod_{i=1}^n x_i \right)^{ap-1} a^{-(\alpha_0+1)}}{b^{(anp)+(\alpha_1+1)} \left[ \prod_{i=1}^n (x_i^a + b^a)^{p+1} / b^{na(p+1)} \right] \Gamma(\alpha_0) \Gamma(\alpha_1) \Gamma(\alpha_2)}. \quad (8.5)$$

Dado que de la ecuación (8.5) no se reconoce el kernel, para encontrar las distribuciones posteriores de cada parámetro se utiliza el algoritmo de muestreo “t-walk”, el cual puede ser encontrado en el paquete [Christen \(2010\)](#).

# Capítulo 9

## Metodología

### 9.1. Base de datos

#### 9.1.1. Antecedentes

Los datos utilizados en esta investigación provienen de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH).

La ENIGH, tiene sus antecedentes en varias encuestas realizadas por diferentes dependencias públicas. En 1956 y 1958 la Dirección General de Estadística (DGE), entonces dependiente de la Secretaría de Industria y Comercio (SIC), levantó la encuesta Ingresos y Egresos de la Población en México, y en 1960 las 16 Ciudades de la República Mexicana, Ingresos y Egresos Familiares. Por su parte, para los años 1963 y 1968, el Banco de México llevó a cabo el levantamiento de la encuesta Ingresos y Gastos Familiares.

Durante el periodo 1969-1970, nuevamente la DGE-SIC realizó la encuesta Ingresos y Egresos de la República Mexicana. Posteriormente, en 1975, la Secretaría del Trabajo y Previsión Social (STPS), efectuó Ingresos y Gastos Familiares. En 1977 la DGE, como parte de la Secretaría de Programación y Presupuesto (SPP), desarrolló la Encuesta Nacional de Ingreso-Gasto de los Hogares, trabajo que constituyó el antecedente inmediato de las encuestas que ha llevado a cabo la DGE-INEGI para los periodos: 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006 y 2008.

Es a partir de 1984 cuando se inició una tendencia a ampliar los objetivos y a homogeneizar la metodología, tomando en cuenta las recomendaciones internacionales y los requerimientos de información de los diferentes usuarios, cuidando la comparabilidad histórica. También se incorporaron nuevos productos con el objetivo de adecuarse a los cambios económicos del país y obtener resultados que reflejan la realidad.



## 9.1. Base de datos

---

### 9.1.2. Diseño conceptual de la ENIGH

#### **Población objetivo:**

La constituyen los hogares de nacionales o extranjeros, que residen habitualmente en viviendas particulares dentro del territorio nacional.

#### **Cobertura temática:**

- Ingreso corriente total (monetario y no monetario) de los hogares.
- Gasto corriente monetario de los hogares.
- Percepciones financieras y de capital de los hogares y sus integrantes.
- Erogaciones financieras y de capital de los hogares.
- Características de la vivienda.
- Residentes e identificación de hogares en la vivienda.
- Características sociodemográficas de los residentes de la vivienda.
- Condición de actividad y características ocupacionales de los integrantes del hogar de 12 y más años.
- Equipamiento del hogar, servicios.

#### **Clasificadores utilizados:**

Clasificación Mexicana de Ocupaciones (CMO), 2005. Sistema de Clasificación Industrial de América del Norte (SCIAN) 2004, versión hogares.

#### **Recomendaciones internacionales:**

- Las resoluciones e informes de las 18 Conferencias Internacionales sobre Estadísticas del Trabajo, de la Organización Internacional del Trabajo (OIT).
- El Reporte final y Recomendaciones del Grupo de Canberra, grupo de expertos en las Estadísticas de Ingreso de los Hogares.
- Manual de Encuestas sobre Hogares. Departamento de Asuntos Económicos y Sociales Internacionales, Oficina de Estadística. Naciones Unidas, Nueva York, 1987.
- Articulación con el sistema de cuentas nacionales y con las encuestas de hogares que levanta el INEGI.

## 9.2. Selección del Modelo

---

### 9.1.3. Diseño estadístico de la ENIGH

**Tamaño de la muestra afectiva:** 35 146 viviendas.

**Unidad de observación:** El hogar.

**Unidad de muestreo:** La vivienda.

**Unidad de análisis:** El hogar, la vivienda y los integrantes del hogar.

**Marco muestral:** El marco de muestreo utilizado es el marco de propósitos múltiples del INEGI, está constituido con la información demográfica y cartográfica obtenida a partir del levantamiento del Censo de Población y Vivienda del 2000.

Las viviendas se eligen mediante un minucioso proceso estadístico (incluye las viviendas deshabitadas y de uso temporal) que garantiza que los resultados obtenidos a partir de sólo una parte de la población (muestra), se pueda generalizar hacia el total.

La selección se realiza en dos etapas:

- En la primera etapa, se eligen las unidades primarias de muestreo (UPM).
- En la segunda etapa, se seleccionan las viviendas objeto de entrevista de cada encuesta.

**Esquema de muestreo:** Probabilística, a su vez estratificado, bietápico y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar; en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población

**Ponderador:** Variable “factor” en la tabla HOGARES y “hog” en la tabla de CONCENTRADO; el factor de expansión es el mismo en las dos tablas pero se renombró para que el usuario lo relacionara con el número de hogares que representa.

Para el caso específico de ésta investigación, se trabajó únicamente con los datos proporcionados por el INEGI (1998, 2002, 2008), los cuales fueron procesados en el programa [R Development Core Team \(2011\)](#).

## 9.2. Selección del Modelo

Para seleccionar el modelo a usar se estimaron los parámetros de cada distribución de manera Bayesiana; para la distribución Pareto ésto se realizó mediante el uso del algoritmo del muestreador de *Gibbs*, mientras que en la distribución Lognormal se realizó con un *ars* dentro de un *Gibbs* y por último para la distribución Dagum se utilizó el algoritmo *t-walk*. Ya obtenidos los estimadores de las diferentes distribuciones de interés,

## 9.2. Selección del Modelo

---

éstas se utilizaron para modelar los datos del ingreso de la población mexicana en los tres diferentes años de estudio, mostrando una mejor modelación la distribución Dagum. Lo antes mencionado está descrito de manera explícita en los apartados 6.5, 7.4 y 8.5 de esta misma tesis, además de que se puede consultar el anexo A, el cual contiene las rutinas para la estimación de los diferentes parámetros de manera Bayesiana en cada distribución, para lo cual fue utilizado el paquete “coda” desarrollado por [Plummer \*et al.\* \(2010\)](#), todo lo anterior elaborado en el programa [R Development Core Team \(2011\)](#).

Además de lo antes expuesto era necesario el uso de otro criterio, por lo que se implementó el criterio de akaike (AIC) por sus siglas en Inglés, el cual fue calculado para los tres modelos antes mencionados, además de realizar un análisis gráfico, el cual es mostrado en las figuras [9.1](#), [9.2](#) y [9.3](#).

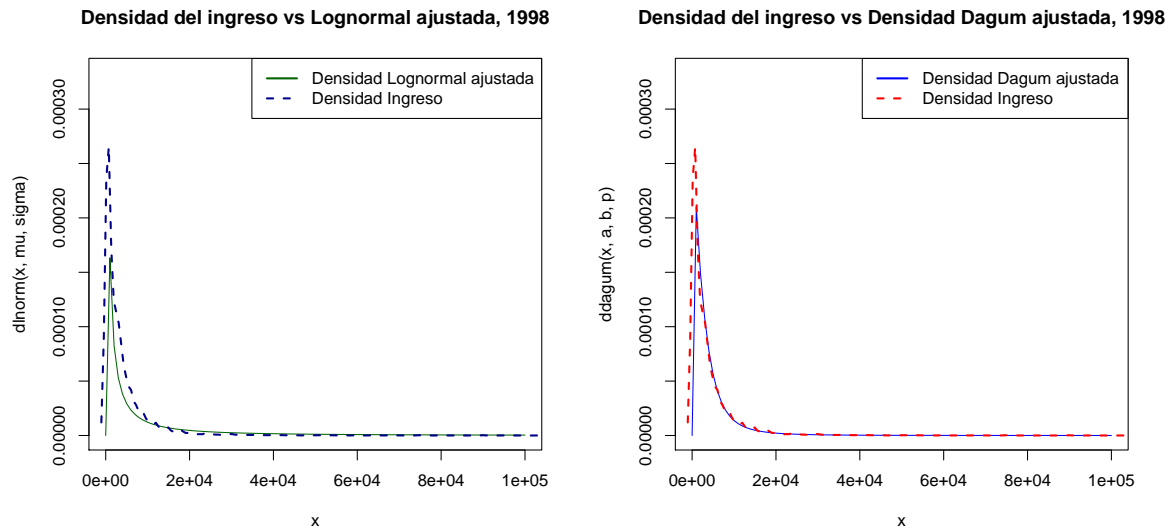
**Tabla 9.1:** Criterio de Akaike.

	1998	2002	2008
Dagum	580985.1	663504.1	675389
Lognormal	517802.2	593367.5	603921.2
Pareto	662455.5	735007.8	764398.4

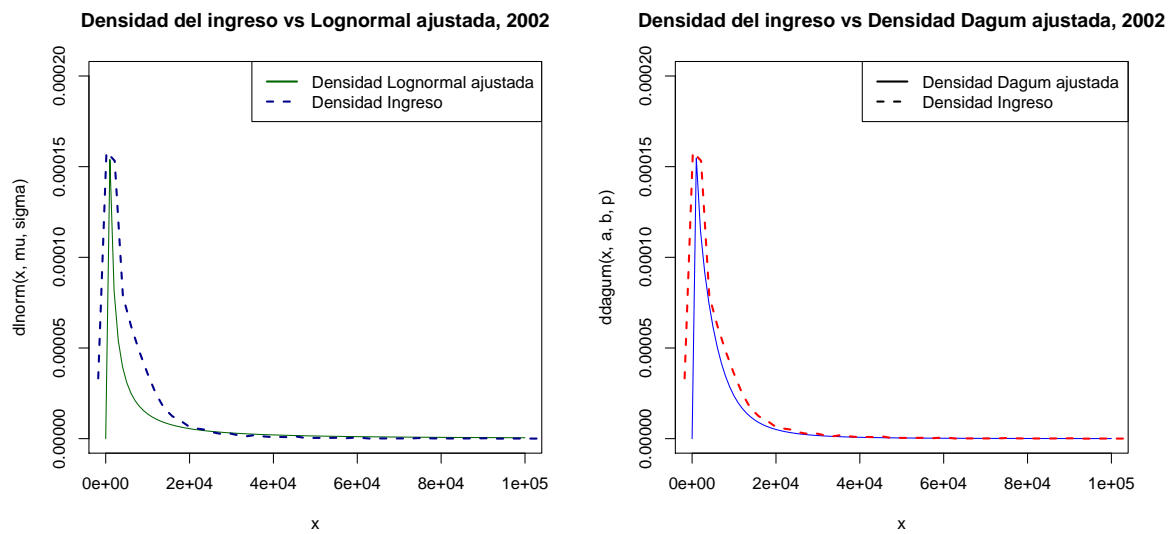
Como lo muestra la tabla [9.1](#), las dos distribuciones que mejor modelan los datos de los ingresos en México en base al AIC son el modelo Lognormal y el Dagum, ya que éstos tienen un menor valor AIC, estos para los tres periodos de estudio.

Al comparar gráficamente estas dos distribuciones, se nota que para los tres años en cuestión la distribución Dagum modela mejor a los datos de ingreso, esto de se puede ver en las figuras [9.1](#), [9.2](#) y [9.3](#).

## 9.2. Selección del Modelo



**Figura 9.1:** Comparación entre las funciones de densidad Lognormal y Dagum, con datos del ingreso de 1998.

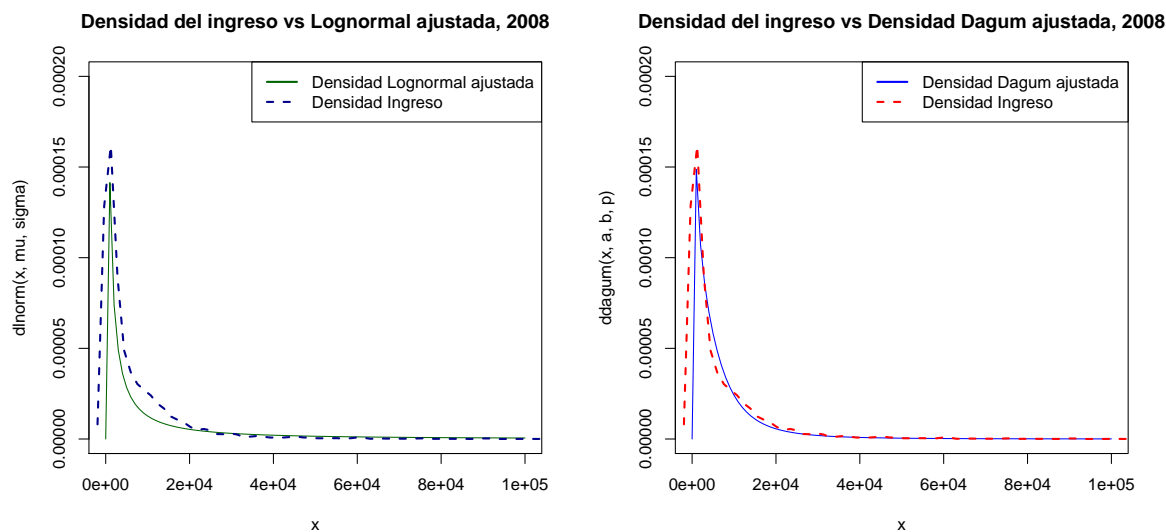


**Figura 9.2:** Comparación entre las funciones de densidad Lognormal y Dagum, con datos del ingreso de 2002.

Al contrastar el gráfico 9.2a con el 9.2b, es visible la diferencia de ajuste de la distribución Lognormal que de la Dagum, al apreciar un mejor modelamiento en la gráfica 9.2b, lo

### 9.3. Análisis del ingreso en México con la Distribución Dagum

cual se mantiene para los otros dos años (1998 y 2008). Es de importancia señalar que el ajuste que presentan los datos con la Distribución Dagum fue realizado de manera Bayesiana, gráfica 9.2b, mientras que para la distribución Lognormal fue realizado por máxima verosimilitud, ésto así para las tres periodos de análisis.



**Figura 9.3:** Comparación entre las funciones de densidad Lognormal y Dagum, con datos del ingreso de 2008.

Lo anterior dio la pauta a este trabajo para optar por el modelo Dagum, ya que resultó tener un AIC menor que el de la distribución Pareto y gráficamente modela mejor el comportamiento de los datos en cuestión.

### 9.3. Análisis del ingreso en México con la Distribución Dagum

El análisis se divide en dos partes principales:

- Modelar los datos del ingreso en México para los años 1998, 2002, 2008 con la distribución Dagum.
- Una vez obtenidas las distribuciones posteriores de los parámetros, realizar un análisis económico sobre la distribución (repartición) del ingreso en México, usando la densidades posteriores.

### 9.3. Análisis del ingreso en México con la Distribución Dagum

---

Para los tres cortes de tiempo de interés, el análisis a la base de datos partió de obtener una muestra representativa, la cual se obtuvo de la siguiente manera:

- Se ordenó en base a los folios de vivienda de cada hogar, la variables ingreso trimestral y factor de expansión.
- Posteriormente, con la variable factor de expansión se construyó un vector de probabilidades.
- Después se obtuvo una nueva muestra tomando en cuenta el vector de probabilidades, es decir se le asignó una probabilidad de ser elegido en la muestra a cada individuo, esto en base a su factor de expansión.
- Finalmente la muestra obtenida anteriormente es una muestra representativa del ingreso de la población en México, la cual permitirá hacer el análisis de manera real.

#### 9.3.1. Prueba de convergencia en la distribución Dagum

Una vez obtenida la muestra, se obtienen los estimadores de la distribución Dagum con el algoritmo “t-walk”, para verificar que se tienen muestras de la distribución estacionaria se realiza la prueba de convergencia de [Gelman y Rubin \(1992\)](#). Los resultados se muestran en las tablas [9.2](#), [9.3](#) y [9.4](#).

**Tabla 9.2:** Prueba de convergencia de Gelman y Rubin para el año de corte 1998.

Parámetro	Estimador puntual	Cuantil al 97.5 %
$a$	1.02	1.02
$b$	1.00	1.00
$p$	1.00	1.00
$\widehat{R}$	1	

**Tabla 9.3:** Prueba de convergencia de Gelman y Rubin para el año de corte 2002.

Parámetro	Estimador puntual	Cuantil al 97.5 %
$a$	1.04	1.12
$b$	1.03	1.12
$p$	1.03	1.09
$\widehat{R}$	1.03	

### 9.3. Análisis del ingreso en México con la Distribución Dagum

---

**Tabla 9.4:** Prueba de convergencia de Gelman y Rubin para el año de corte 2008.

Parámetro	Estimador puntual	Cuantil al 97.5 %
$a$	1.00	1.01
$b$	1.01	1.01
$p$	1.00	1.00
$\widehat{R}$	1	

Las Cadenas de Markov partieron de diferentes puntos (valores aleatorios), por lo que cuando  $\widehat{R}$  (estimador de reducción potencial de escala) es igual a 1 la varianza entre las dos cadenas propuestas no es más grande que la varianza dentro de cada cadena individual, lo cual declara convergencia.

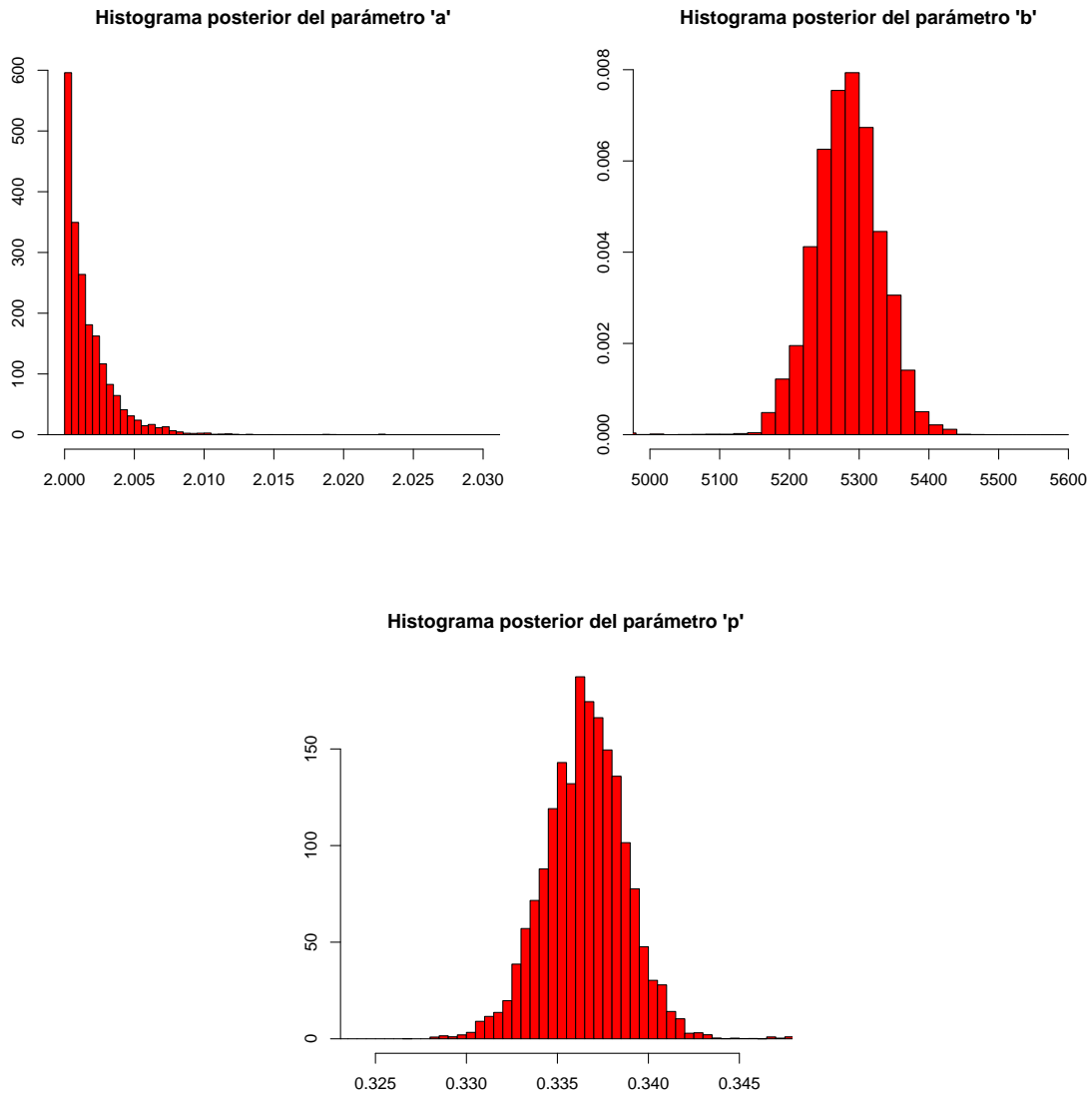
Como lo muestran la tablas 9.2, 9.3 y 9.4 las Cadenas de Markov han convergido, ya que la  $\widehat{R}$  es igual a 1, significando que las Cadenas de Markov están esencialmente sobrepuestas y provienen de la misma distribución, por lo cual éstas son de utilidad para estimar los parámetros de la distribución Dagum. El estadístico que se uso en éste caso fue la moda debido a que se observa que el ajuste proporcionado con este fue mejor para los tres años de estudio, además de estimar el valor de cada parámetro se obtuvo las distribuciones posteriores marginales de los mismos, las cuales pueden verse en las figuras 9.4 - 9.6, para los años 1998, 2002 y 2008 respectivamente.

#### 9.3.2. Distribuciones marginales posteriores

Por ejemplo para el año 1998 se observa en la gráfica 9.4a que la distribución posterior marginal del parámetro  $a$  se asemeja a una exponencial, mientras que para los parámetro  $b$  y  $p$  se puede ver en 9.4b y 9.4c respectivamente, que se asemejan a una distribución normal. Lo anterior puede generalizarse para los otros dos años de estudio, ya que gráficamente las distribuciones posteriores marginales de los parámetros en cuestión coinciden con las de 1998 como lo muestran las figuras 9.5 y 9.6.

### 9.3. Análisis del ingreso en México con la Distribución Dagum

---

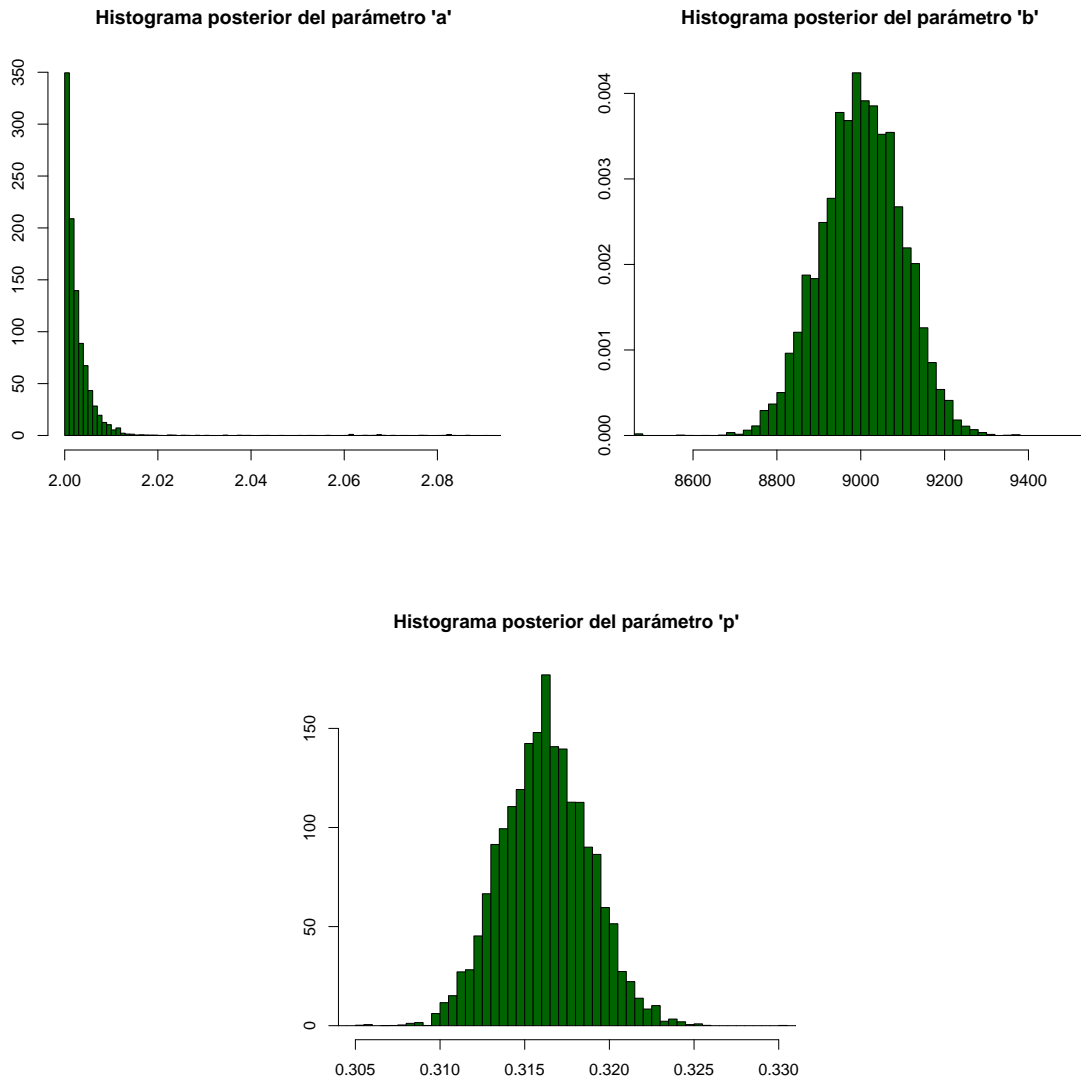


**Figura 9.4:** Distribuciones posteriores de los parámetros en la distribución Dagum, para datos del ingreso de 1998.



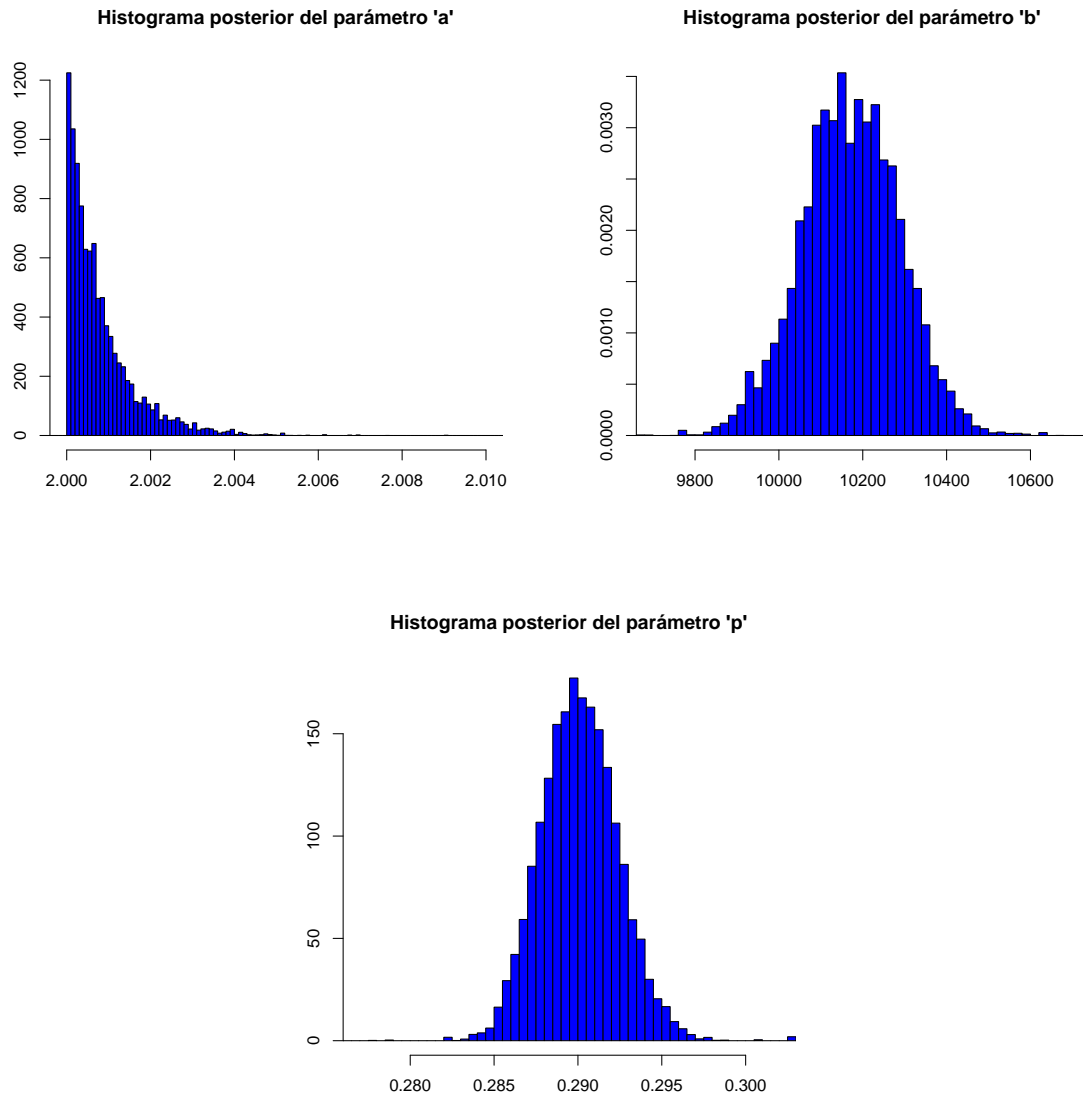
### 9.3. Análisis del ingreso en México con la Distribución Dagum

---



**Figura 9.5:** Distribuciones posteriores de los parámetros en la distribución Dagum, para datos del ingreso de 2002.

### 9.3. Análisis del ingreso en México con la Distribución Dagum



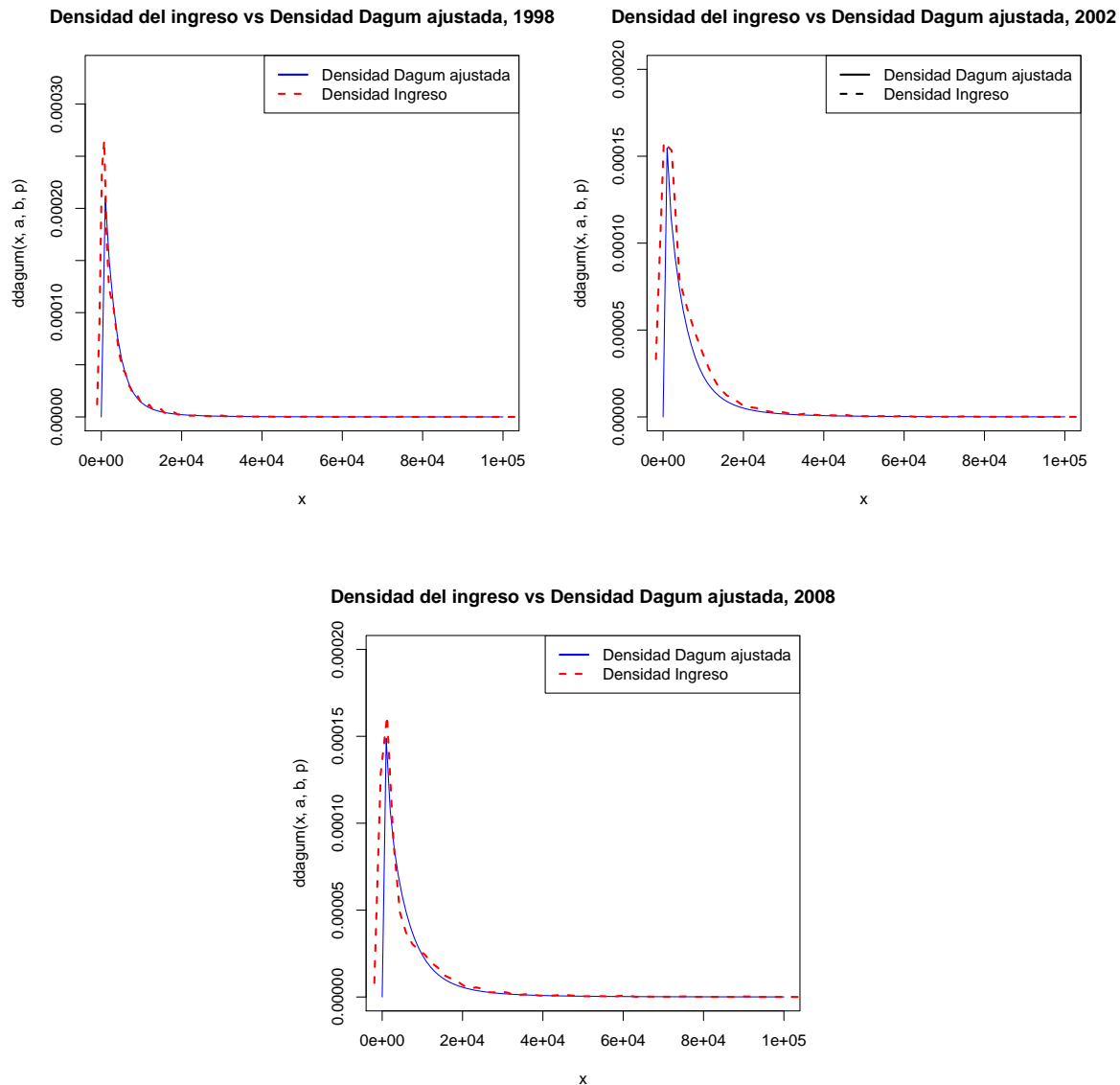
**Figura 9.6:** Distribuciones posteriores de los parámetros en la distribución Dagum, para datos del ingreso de 2008.

#### 9.3.3. Ajuste de la distribución Dagum a los datos del ingreso, para los años 1998, 2002 y 2008

En la figura 9.7 se puede observar el ajuste de la distribución Dagum a los datos de ingresos de la población mexicana para los años de estudio. Para las tres gráficas 9.7a, 9.7b y 9.7c el ingreso esta representado por la línea punteada de color rojo mientras

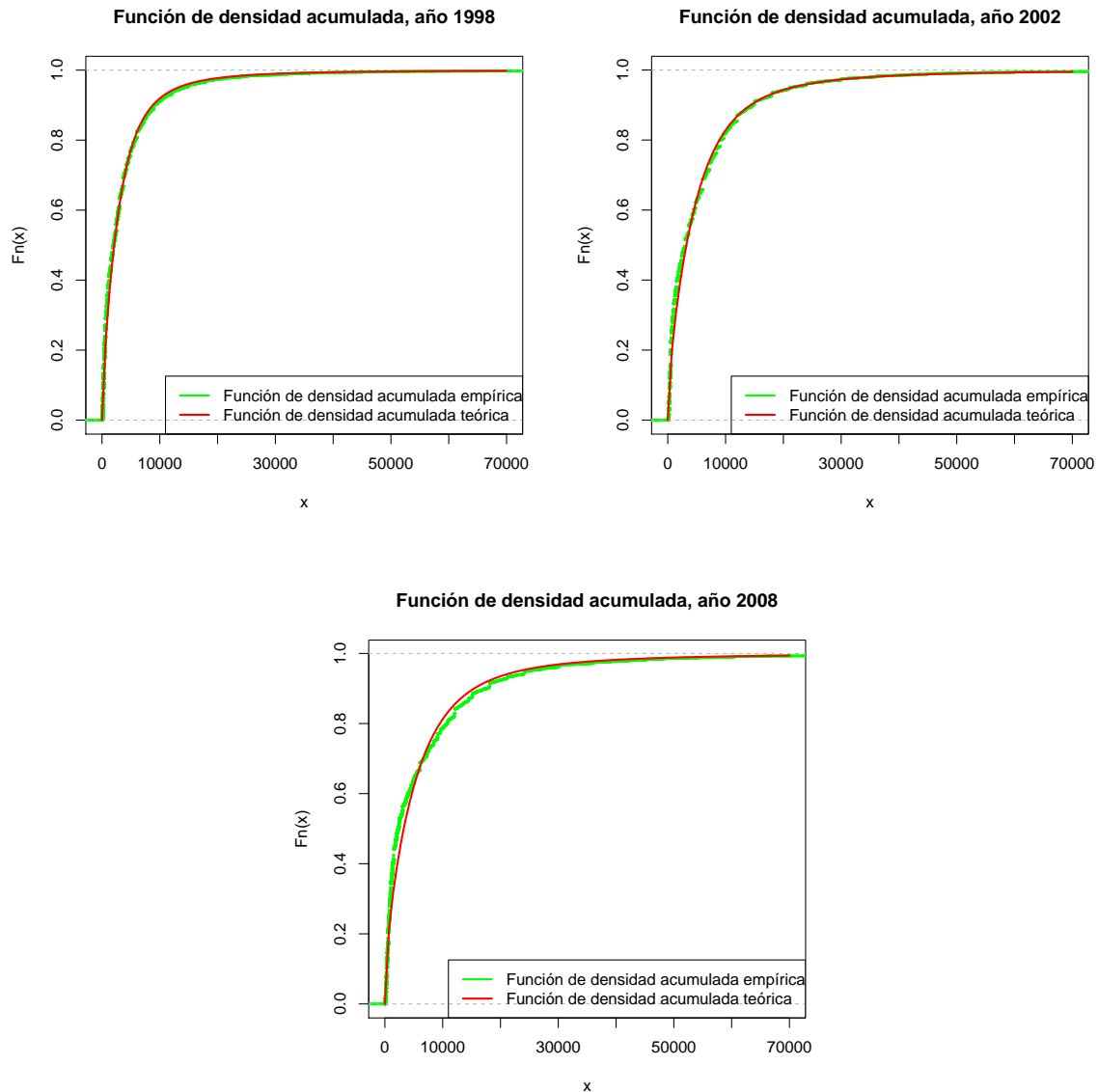
### 9.3. Análisis del ingreso en México con la Distribución Dagum

que la ajustada por estimación bayesiana es la de color azul, claramente se ve un buen ajuste, especialmente para los años 1998 y 2008 en los cuales el acoplo de ambas curvas es mayor, con lo anterior se reafirma que la distribución Dagum estimada de manera bayesiana explica correctamente el comportamiento de los datos de ingreso en México.



**Figura 9.7:** Comparación entre la densidad del ingreso en México y la densidad ajustada con la distribución Dagum.

### 9.3. Análisis del ingreso en México con la Distribución Dagum



**Figura 9.8:** Comparación entre la función de distribución empírica y teórica.

Otra gráfica que permite estudiar la bondad de ajuste de la distribución Dagum a los datos del ingreso, es cuando se grafican de manera conjunta las funciones de densidad acumulada teórica y la empírica. Esto se puede ver en la figura 9.8, en la cual para los tres años de estudio la función de densidad acumulada (fda) teórica está representada por la línea punteada de color verde mientras que la fda empírica es la línea continua de color rojo.

En base a la figura 9.8, el año en el que mejor se ajustan ambas líneas es 1998 (gráfica

### 9.3. Análisis del ingreso en México con la Distribución Dagum

---

9.8a), lo cual suena lógico ya que también en la gráfica 9.7a es la que muestra mejor ajuste. Lo anterior proporciona una mejor pauta para proceder a realizar el análisis económico en estos periodos de tiempo, ya que con lo anterior se corrobora que la distribución Dagum explica de manera correcta el comportamiento de los datos del ingreso en México.

# Capítulo 10

## Resultados

### 10.1. Definición Coeficiente de Gini

El Coeficiente de Gini es una medida de la desigualdad ideada por el estadístico italiano Corrado Gini<sup>1</sup>. Normalmente se utiliza para medir la desigualdad en los ingresos, pero puede utilizarse para medir cualquier forma de distribución desigual. El coeficiente de Gini es un número entre 0 y 1, en donde 0 corresponde a la perfecta igualdad (todos tienen los mismos ingresos) y 1 corresponde a la perfecta desigualdad (una persona tiene todos los ingresos y los demás ninguno). El índice de Gini es el coeficiente de Gini expresado en porcentaje.

### 10.2. Relación entre la curva de Lorenz y el coeficiente de Gini

Aunque el coeficiente de Gini se utiliza sobre todo para medir la desigualdad en los ingresos, también puede utilizarse para medir la desigualdad en la riqueza. Este uso requiere que nadie disponga de una riqueza neta negativa.

El coeficiente de Gini se calcula como una razón de las áreas en el diagrama de la curva de Lorenz. Si el área entre la línea de perfecta igualdad y la curva de Lorenz es  $a$ , y el área por debajo de la curva de Lorenz es  $b$ , entonces el coeficiente de Gini es  $a/(a + b)$ .

Esta razón se expresa como porcentaje o como equivalente numérico de ese porcentaje,

---

<sup>1</sup> Corrado Gini (1884 - 1965), fue un estadístico, demógrafo y sociólogo italiano que propuso el conocido índice de concentración de Gini. Fue también un influyente teórico fascista e ideólogo que escribió las bases científicas del fascismo en 1927.

### 10.3. Coeficiente de Gini para la distribución Dagum

---

que es siempre un número entre 0 y 1. El coeficiente de Gini se calcula a menudo con la Fórmula de Brown, que es más práctica:

$$G = \left| 1 - \sum_{k=1}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k) \right|$$

Donde:

- G: Coeficiente de Gini
- X: Proporción acumulada de la variable población
- Y: Proporción acumulada de la variable ingresos

De forma resumida, la Curva de Lorenz es una gráfica de concentración acumulada de la distribución de la riqueza superpuesta a la curva de la distribución de frecuencias de los individuos que la poseen, y su expresión en porcentajes es el índice de Gini.

#### Propiedades

- Todas las curvas de Lorenz pasan por los puntos (0,0) y (1,1). Si dos curvas de Lorenz no se cortan fuera de esos dos puntos, es posible comparar la desigualdad que representan sin necesidad de calcular el índice de Gini. En el caso general, un mayor índice de Gini significa una mayor desigualdad.
- Para determinar el área entre la curva de Lorenz y la línea de perfecta equidad, lo ideal es calcular una integral definida, pero a veces no se conoce la definición explícita de la curva de Lorenz, por lo que es interesante utilizar otras fórmulas con un número finito de sumandos.
- Las propiedades del índice de Gini son comparables con las del cuadrado del coeficiente de variación.

### 10.3. Coeficiente de Gini para la distribución Dagum

Una vez corroborado el ajuste de la distribución Dagum a los datos del ingreso de la población en México, el cual puede observarse en la figura 9.7, se puede calcular el coeficiente de Gini, ya que esta distribución explica correctamente el compartimiento de nuestros datos.

#### 10.4. Cálculo del Coeficiente de Gini y Curva de Lorenz para México en los años 1998, 2002 y 2008

---

Para el cálculo del coeficiente de Gini se utiliza la expresión propuesta por Dagum (1977), la cual está dada por la siguiente expresión:

$$G = \frac{\Gamma(p)\Gamma(2p + 1/a)}{\Gamma(2p)\Gamma(p + 1/a)} - 1 \quad (10.1)$$

Como se puede observar en la ecuación (10.1), para el cálculo del coeficiente de Gini se necesita tener los valores de los parámetros ( $a$  y  $p$ ) de la distribución Dagum en cuestión. Debido a que en el presente trabajo no solo se obtuvo la estimación puntual de los parámetros, si no que se encontró la distribución posterior de los mismos, esto gracias al enfoque bayesiano que se le dio. Por lo anterior se pudo obtener toda la distribución del coeficiente de Gini, como se puede observa en la figura 10.1, y así realizar una mejor estimación (puntual o por intervalos de credibilidad).

#### 10.4. Cálculo del Coeficiente de Gini y Curva de Lorenz para México en los años 1998, 2002 y 2008

Una vez seleccionado el modelo Dagum, y ya con sus parámetros estimados se procedió a el cálculo de la distribución posterior del coeficiente de Gini para tres años (1998,2002 y 2008) en un intervalo de tiempo de 10 años, con lo cual se pretende realizar una estimación del nivel de desigualdad en la distribución del ingreso en México, y encontrar evidencia estadística de como se ha comportado la concentración de la riqueza en este país. Así con esto se pretende validar o invalidar la afirmación de que cada vez los pobres son más pobres y los ricos más ricos.

**Tabla 10.1:** Coeficiente de Gini.

Año	$\hat{a}$	$\hat{b}$	$\hat{p}$	Coeficiente de Gini
1998	2.000081	5208.986	0.3508603	0.622542
2002	2.000498	9016.87	0.318004	0.633100
2008	2.000156	10216.31	0.290466	0.646406

Como se puede observar en la tabla 10.1 el coeficiente de Gini a aumentado en un 7.87 puntos porcentuales en un lapso de 10 años, es decir que la riqueza en México si se ha concentrado cada vez más en menos personas, o si se quisiera decir en otras palabras, los ricos cada vez son menos pero más ricos, y los pobres cada vez son más y más pobres.

Esta afirmación anterior se puede validar nuevamente al obtener la curva de Lorenz para los diferentes años de estudio, por ejemplo en la figura 10.2 se observa específicamente para las gráficas 10.2c y 10.2b el desplazamiento de la curva del año 2008 representado con



#### 10.4. Cálculo del Coeficiente de Gini y Curva de Lorenz para México en los años 1998, 2002 y 2008

---

la línea verde punteada tiene un ligero desplazamiento hacia abajo (eje de las abscisas), la cual representa y corrobora lo antes concluido con el indicador coeficiente de Gini.

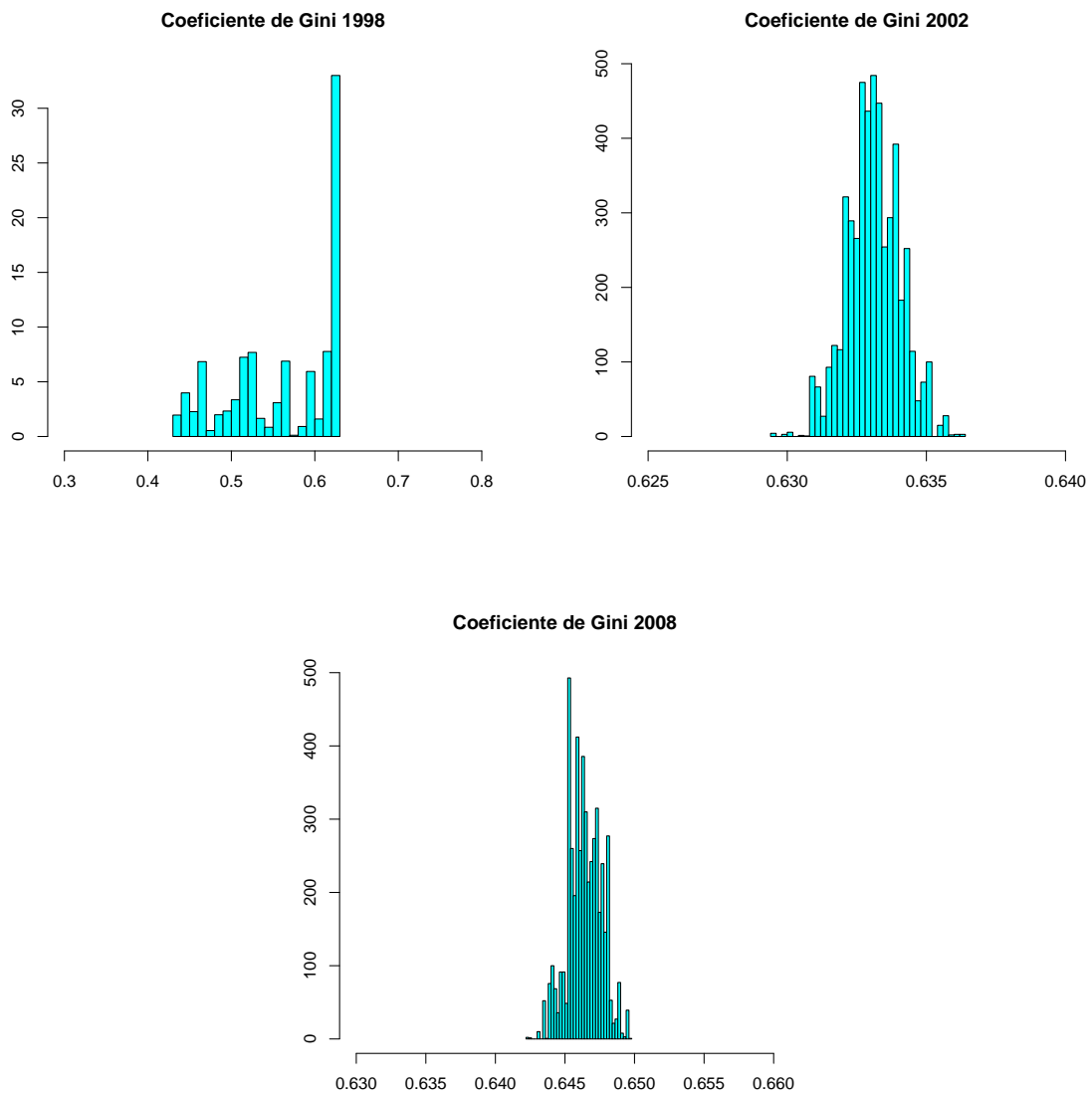
**Tabla 10.2:** Intervalo de máxima densidad al 95 % de probabilidades para el Coeficiente de Gini.

Año	Coeficiente de Gini	Límite inferior	Límite superior	Varianza
1998	0.622542	0.4483455	0.6257984	0.00398220
2002	0.633100	0.6314125	0.6351684	8.93630e-07
2008	0.646406	0.6438346	0.6483595	1.43726e-06

Además se puede sustentar estadísticamente lo antes afirmado, ya que al realizar un intervalo de máxima densidad al 95 % de probabilidades, éste contiene a nuestro valor estimado. La interpretación adecuada sería; se tiene un 95 % de probabilidades de que los intervalo descritos en la tabla 10.2 contengan los valores del parámetro estimado (Coeficiente de Gini) para los tres años de estudio, esto proporciona una herramienta veraz al utilizar el coeficiente de Gini para hacer conclusiones.

## 10.4. Cálculo del Coeficiente de Gini y Curva de Lorenz para México en los años 1998, 2002 y 2008

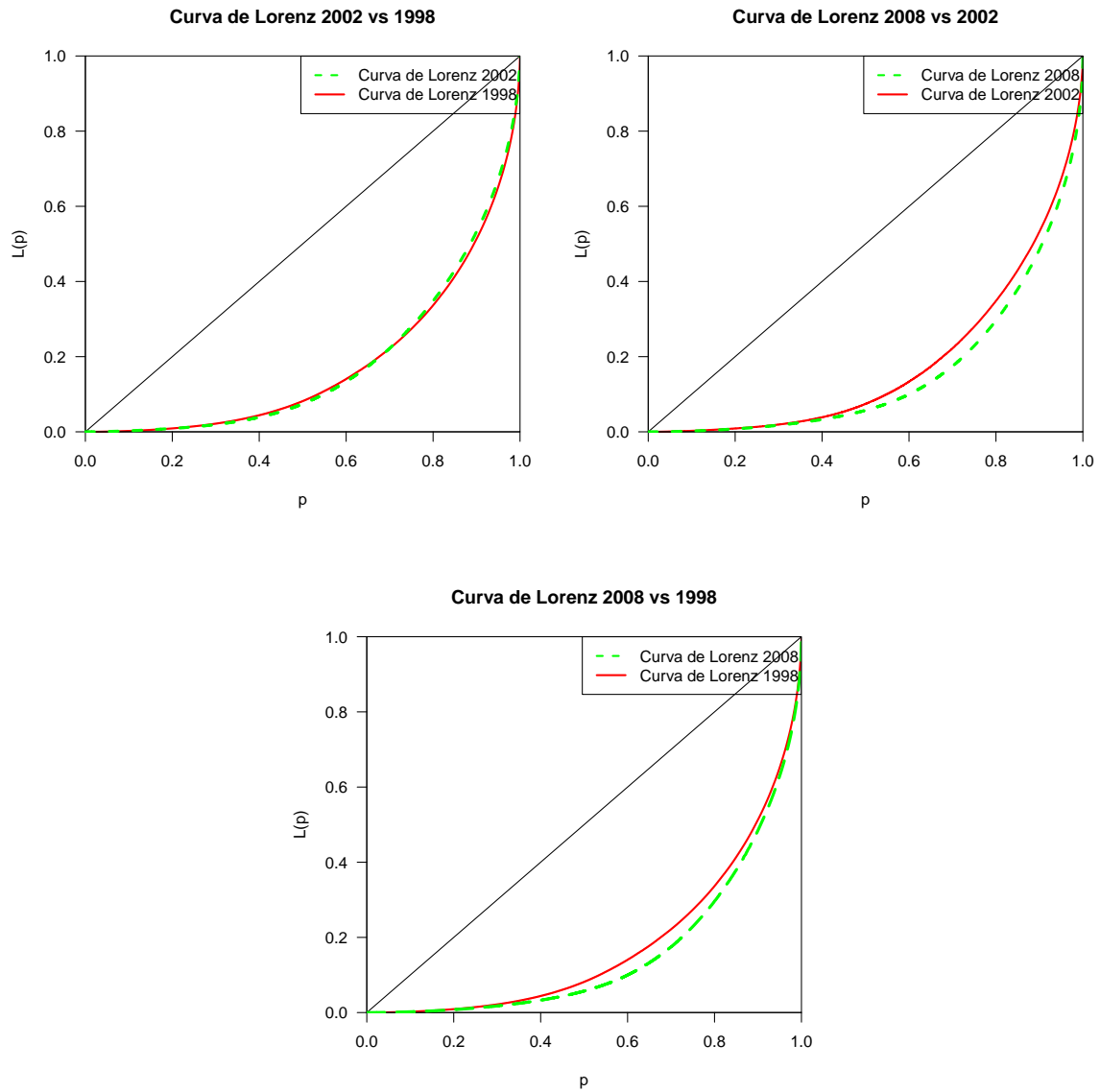
---



**Figura 10.1:** Histogramas del Coeficiente de Gini para los años 1998, 2002 y 2008.

Por ejemplo para la gráfica 10.2c, se puede observar que en el año 1998 aproximadamente un 65 por ciento de la población tenía menos del 20 por ciento de la riqueza total, mientras que para el año 2008 al ocurrir un desplazamiento hacia el eje de las abscisas este porcentaje de personas con menos de 20 por ciento de la riqueza total aumenta, visualmente parece ser que pasa del 65 al 72 por ciento de personas en esta situación.

## 10.4. Cálculo del Coeficiente de Gini y Curva de Lorenz para México en los años 1998, 2002 y 2008



**Figura 10.2:** Curvas de Lorenz para los años 1998, 2002 y 2008.

# Capítulo 11

## Conclusiones

Las conclusiones obtenidas en este trabajo, se pueden dividir en tres partes:

En lo referente a obtener los estimadores bayesianos de la distribución Dagum se concluye lo siguiente:

- Los estimadores bayesianos de la distribución Dagum se pueden obtener de forma eficiente utilizando el algoritmo de muestreo “t-walk”, el cual funciona como un Metropolis Hastings por partes, una vez obtenido los estimadores de esta distribución estos funcionan correctamente, ya que las inferencias realizadas con estos son coherentes con la realidad.

En lo referente a modelar el ingreso con la distribución Dagum:

- Se puede concluir en base a la figura 9.7 que el ajuste de la distribución Dagum a los datos de ingresos de la población mexicana para los años de estudio es buena, especialmente para los años 2002 y 2008 como lo muestran las siguientes gráficas 9.7b y 9.7c respectivamente, en las cuales el empalme de ambas curvas es mejor, con lo anterior se reafirma que la distribución Dagum estimada de manera bayesiana modela correctamente el comportamiento de los datos de ingreso en México.
- Otra figura que permite reafirmar el hecho de que la distribución Dagum modela correctamente los ingresos de la población mexicana es la 9.8, la cual muestra el perfecto empalme entre las funciones de densidad acumulada teórica y la empírica, para los tres años de estudio.

En lo referente al análisis de la distribución del ingreso en México:

## 11. Conclusiones

---

- Como se puede observar en la tabla [10.1](#) el coeficiente de Gini ha aumentado en un 7.87 puntos porcentuales en un lapso de 10 años, es decir que la riqueza en México si se ha concentrado cada vez más en menos personas, o si se quisiera decir en otras palabras, los ricos cada vez son menos pero más ricos, y los pobres cada vez son más y más pobres.
- Otra conclusión que se obtiene, es que el lapso de tiempo en el cual la desigualdad no aumentó considerablemente o aumentó de manera marginal, es el periodo comprendido de 1998 a 2002, ya que tanto numéricamente (tabla [10.1](#)) como gráficamente se puede observar. Visualmente la Figura [10.2a](#) permite apreciar que ambas líneas están casi perfectamente empalmadas, lo cual coincide con sus correspondientes coeficientes de Gini y refuerza la conclusión anterior.
- Mientras que el periodo en el cual la desigualdad aumento en mayor magnitud fue el comprendido entre el año 2002 y 2008, esto se puede apreciar en la gráfica [10.2b](#) o en la tabla [10.1](#).
- En base a las conjeturas antes hechas, se puede concluir que a pesar de que en este periodo de tiempo a aumentado el gasto público y por lo tanto el producto interno bruto, la asignación de estos recursos provenientes del erario público no se ha dado de manera eficiente en términos que mitigue la pobreza y beneficie a los que menos ingresos tienen.

# Referencias

- Arnold, B. C., Castillo, E. y Sarabia, J. M. (1998). Bayesian analysis for classical distributions using conditionally specified priors. *The Indian Journal of Statistics*, 60, 228–245.
- Box, G. y Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company.
- Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Phil. Mag*, 4, 161–173.
- Burr, I. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, 13, 215–232.
- Carlin, B. y Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman & Hall/CRC.
- Casella, G. y George, E. I. (1992). Explaining the Gibbs Sampler. *American Statistical Association*, 46, 167–174.
- CEFP (2008). Distribución del Ingreso y desigualdad en México; un análisis sobre la ENIGH 2000 a 2006. *Cámara de Diputados*, 9, 1–18.
- Chen, M., Shao, Q. y Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer Series in Statistics. New York: Springer-Verlag.
- Christen, J. y Fox, C. (2010). A General Purpose Sampling Algorithm for Continuous Distributions (the t-walk). *Bayesian Analysis*, 5, 263 – 282.
- Christen, J. A. (2010). *Rtwalk: The R implementation of the t-walk*. R package version 1.5.1.
- Dagum, C. (1977). A New Model for Personal Income Distributions: Specification and Estimation. *Economic Appliquée*, 30, 413 – 437.
- Dagum, C. y Lemmi, A. (1989). A contribution to the analysis of income distribution and income inequality, and a case study: Italy. *Research on Economic Inequality*, 1, 123–157.
- Datta, G. y Ghosh, M. (1996). On the Invariance of Noninformative Priors. *The Annals of Statistics*, 24, 141–159.
- Davis, H. T. (1941). *The Analysis of Economic Time Series*. Bloomington, IN: Principia Press.

## Referencias

---

- Devroye, L. (1986). *Non-uniform random variate generation*. New York, Springer. pp, 38, 287.
- Fattorini, L. y Lemmi, A. (1979). Proposta di un modello alternativo per l'analisi della distribuzione personale del reddito. *Atti Giornate di Lavoro AIRO*, 28, 89–117.
- Fernández, J. (2005). La estadística en las ciencias actuariales y finanzas. *Boletín electrónico ITAM.*, 6.
- Gelfand, A. y Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., Carlin, J., Stern, H. y Rubin, D. (2004). *Bayesian data analysis*. 2nd Edn. Chapman and Hall / CRC. USA.
- Gelman, A. y Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- Geman, S. y Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the bayesian reformation of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6, 721–741.
- Gibrat, R. (1931). *Les Inégalités Économiques*.. Paris: Librairie du Recueil Sirey.
- Gilks, W., Richardson, S. y Spiegelhalter. (1996). Markov Chain Monte Carlo in Practice. *Interdisciplinary statistic. London: Chapman & Hall. USA*, 46, 5–7.
- Gilks, W. y Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistical*, 41, 337–348.
- Hart, P. E. (1973). The comparative statics and dynamics of income distributions. *University of Reading Discussion Papers in Economics, Series A*, 49.
- Hartigan, J. (1964). Invariant Prior Distributions. *Annals of Mathematical Statistics*, 35, 836–845.
- INEGI (1998, 2002, 2008). *Encuesta Nacional de Ingresos y Gastos de los Hogares*. Instituto Nacional de Estadística, Geografía e informática, México.
- Kakwani, N. C. (1980b.). *Income Inequality and Poverty. Methods of Estimation and Policy Applications*.. New York: Oxford University Press.
- Kass, R. y Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules. *JASA*, 91, 1343–1370.
- Klugman, P. H. H., S. A. y Willmot, G. E. (1998). *Loss Models*.. New York: John Wiley.
- Lee, P. (1997). *Bayesian Statistics: An Introduction*. Second Edition. John Wiley & Sons.
- Lemmi, A. (1987). *Il modello kappa a tre parametri nell'analisi della distribuzione personale del reddito*. In: M. Zenga (ed.).
- Mandelbrot, B. (1997). *A case against the lognormal distribution*. New York and Berlin: Springer.

## Referencias

---

- Metropolis, N., Rosenbluth, A., Rosenluth, M. y Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of the chemical physics*, 21, 1087–1091.
- Mielke, P. W. (1973). Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology*, 12, 275–280.
- Mielke, P. W. y Johnson, E. S. (1974). Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resources Research*, 10, 223–226.
- Pareto, V. (1895). La legge della domanda. *Giornale degli Economisti*, 10, 59–68.
- Paulino Perez Rodriguez, W. R. G., original C++ code from Arnost Komarek based on `ars.f` written by P. Wild (2009). *ars: Adaptive Rejection Sampling*. R package version 0.4.
- Plummer, M., Best, N., Cowles, K. y Vines, K. (2010). *coda: Output analysis and diagnostics for MCMC*. R package version 0.14-2.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Raftery, A. y Lewis, S. (1992). One long run with diagnostics. Implementations strategies for Markov chain Monte Carlo. *Statistical Science*, 7, 493–497.
- Stoppa, G. (1995). Explicit Estimators for Income Distributions. *Research on Economic Inequality*, 6, 393 – 405.
- Stroustrup, B. (1963). The UNIX System: Data Abstraction in C. *Bell Laboratories Technical Journal*, 8, 1701–1732.
- Upadhyay, S. K., Vasishta, N. y Smith, A. (2000). Bayes Inference in life testing and reliability via Markov Chain Monte Carlo simulation. *Sankhya*, 62(2), 203–222.



# Anexos

## Anexo A:

### Rutinas en R-2.12.2

#### Implementación del muestreador del Gibbs en la distribución Pareto

```
#Carga las librerías necesarias para la rutina
library(VGAM)
library(boa)
library(coda)
library(foreign)

#Carga los datos

#Carga la base de datos de los ingresos por individuos
library(foreign)
setwd("/Users/MARYN/Documents/yair/usb 19-07-11")
data<-read.dbf("ingresos.dbf")
home=data$FOLIOVIV; income=data$ING_TRI
datos=data.frame(home,income)

#Carga la base de datos de los factores de expansión
library(foreign)
setwd("/Users/MARYN/Documents/yair/usb 19-07-11")
data2<-read.dbf("hogares.dbf")
house=data2$FOLIOVIV; factor=data2$FACTOR
factores<-data.frame(house,factor)

#Obtiene la muestra REPRESENTATIVA
```

```
set.seed(123)
complete<- (q=merge(factoros,datos,by.x="house", by.y="home"))
weights<-c(complete[,2]/sum(complete[,2]))
real<-sample(complete[,3],25000,replace=F,prob=weights)
datos<<-real[real>0]
datos<<-datos[datos<4000000]

x<<-datos

#Inicia la función Gibbs
gibbs.pareto=function(vector=x,a1.=-206.55,a2.=42.1,b=0,m11=-43,
m12=.001,c=100,m=104000,burn=4000)
{
  n=length(x)
  ALPHA=numeric(m); TAO=numeric(m)
  tao=1/min(x)
  alpha=n/(n*(log(tao))+sum(log(x)))
  for (i in 2:m)
  {
    alpha.star=(1+a2.+(m12*log(tao)))
    beta.star=-(a1.+(m11*log(tao)))
    ALPHA[i]=rgamma(1,alpha.star,beta.star)

    delta.star=-(1+b+m11*alpha+m12*log(alpha))
    ny.star=c
    TAO[i]=rpareto(1,1/ny.star,delta.star)

    tao=TAO[i]
    alpha=ALPHA[i]
  }
  ALPHA=ALPHA[burn:m]

  TAO=TAO[burn:m]

  ESTIMADORES=cbind(ALPHA,TAO)
  return(ESTIMADORES)
}

#Genera una MCMC
cadena1=gibbs.pareto(x,m=20000,burn=4000)

#Genera otra MCMC
cadena2=gibbs.pareto(x,m=20000,burn=4000)

#Hace una lista de objetos MCMC, para la pruebas de convergencia
cadenas=mcmc.list(list(mcmc(cadena2,1,nrow(cadena2),thin=10),
mcmc(cadena1,1,nrow(cadena1),thin=10)))
```

```
#Aplica la prueba de convergencia de Gelman y Rubin
gelman.diag(cadenas)
gelman.plot(cadenas)

#Aplica la prueba de convergencia de Raftery and Lewis
raftery.diag(cadenas)

#Calcula la moda
moda=function(datos)
{
  dens=density(datos)
  dens$x[dens$y==max(dens$y)]
}

#Calcula los estimadores bayesianos de la distribución Pareto con
#el estadístico moda
cadena=rbind(cadena1,cadena2)
est.b <- c(a=moda(cadena[,1]),b=moda(cadena[,2]))
est.b

#Calcula los estimadores bayesianos de la distribución Pareto con
#el estadístico media
est.b1 <- c(a=mean(cadena[,1]),b=mean(cadena[,2]))
est.b1

#Grafica las densidades posteriores marginales de los parámetros
par(mfrow=c(2,1))
curve(dpareto(x,est.b[1], est.b[2]),0,100000,ylim=c(0,0.001/5),
col="blue",main="estimador moda")
lines(density(x,from=0,to=1000000),col="red")
curve(dpareto(x,est.b1[1],est.b1[2]),0,100000,ylim=c(0,0.001/5),
col="blue",main="estimador media")
lines(density(x,from=0,to=1000000),col="red")

#Calcula el coeficiente de gini con los parámetros obtenidos de
#forma bayesiana
coef.GinP <- 1/(2*est.b[1]-1)
coef.GinP
```

## Implementación del algoritmo ARS en la distribución Lognormal

```
library(ars)
library(coda)

load("datos2008.RData")

f <- function(x)
  {
    suma=0
    for(j in length(datos)) suma=suma+(log(datos[j])-x)^2
    exp(-0.5/sigma*suma)
  }
h <- function(x) log(f(x))

hprima=function(x)
  {
    suma=0
    for(k in length(datos)) suma=suma+log(datos[k])-x
    suma/sigma
  }

#Inicia gibbs sampler

gibbs.lognm=function(burn=1000,m=5000)
{
  MU=numeric(); SIGMA=numeric()
  sigma<<-var(log(datos))
  mu<<-mean(log(datos))
  n1=length(datos)
  for (i in 1:m)
  {

    MU[i]=ars(n=1,f=h,fprima=hprima,x=c(7.5,7.7,8))

    alpha=n1/2
    beta=2/(sum((log(datos)-mu)^2))
    SIGMA[i]=1/rgamma(1,alpha,beta)

    mu<<-MU[i]
    sigma<<-SIGMA[i]
  }
  MU=MU[burn:m]
  SIGMA=SIGMA[burn:m]
  ESTIMADORES=cbind(MU,SIGMA)
  return(ESTIMADORES)
}
```

```
#Genera un objeto MCMC
cadena1=gibbs.lognm(burn=4000,m=10000)

#Genera otro objeto MCMC
cadena2=gibbs.lognm(burn=4000,m=10000)

#Genera una lista de objetos mcmc, para la pruebas de convergencia
cadenas=mcmc.list(list(mcmc(cadena2,1,nrow(cadena2),thin=10),
mcmc(cadena1,1,nrow(cadena1),thin=10)))

#Realiza la prueba de convergencia de Gelman y Rubin
gelman.diag(cadenas)
gelman.plot(cadenas)

#Realiza la prueba de convergencia de Raftery y Lewis
raftery.diag(cadenas)

#Grafica la densidad posterior del parámetro de localización
plot(density(cadena1[,1]))

#Grafica la densidad posterior del parámetro de escala
plot(density(cadena1[,2]))

#Estimadores posteriores
#Calcula la moda
moda=function(datos)
{
  dens=density(datos)
  dens$x[dens$y==max(dens$y)]
}

#Calcula los estimadores posteriores
cadena=rbind(cadena1,cadena2)
est.b <- c(a=moda(cadena[,1]),b=moda(cadena[,2]))
est.b
est.b1 <- c(a=mean(cadena[,1]),b=mean(cadena[,2]))
est.b1

#Grafica la densidad teórica por kernel smoothing vs la densidad
#ajustada
curve(dpareto(x,est.b[1], est.b[2]),0,1000000, ylim=(0,100)col="blue")
lines(density(x,from=0,to=1000000),col="red")
```

## Rutinas para implementar el algoritmo “t-walk” en la distribución Dagum

```
#Carga las librerías necesarias
library(Rtwalk)
library(VGAM)
library(coda)
library(MASS)

#Carga la base de datos de los ingresos por individuos
library(foreign)
setwd("/Users/yair/Documents/TESIS/Bases de datos")
data<-read.dbf("ingresos1998.dbf")
home=data$FOLIO; income=data$ING_TRI
datos=data.frame(home,income)

#Carga la base de datos de los factores de expansión
library(foreign)
setwd("/Users/yair/Documents/TESIS/Bases de datos")
data2<-read.dbf("hogares1998.dbf")
house=data2$FOLIO; factor=data2$FACTOR
factores<-data.frame(house,factor)

#Obtiene la muestra REPRESENTATIVA
set.seed(123)
complete<-(q=merge(factores,datos,by.x="house", by.y="home"))
weights<-c(complete[,2]/sum(complete[,2]))
real<-sample(complete[,3],35000,replace=F,prob=weights)
datos<<-real[real>0]
datos<<-datos[datos<700000]

#Función objetivo
dpos=function(x)
{
  Beta0=0.001; Beta=0.001; Beta1=0.001;
  Alpha=0.001; Alpha1=0.001; Alpha0=0.001
  n=length(datos)
  suma1=(sum(log(datos))*(x[1]*x[3]-1)
  suma2=(sum(log(datos^x[1]+x[2]^x[1])))*(x[3]+1)

  Part1=(n*log(x[1]))+(Alpha0*log(Beta0))+(Alpha*log(Beta))+
  (Alpha1*log(Beta1))+
  ((n-Alpha1-1)*log(x[3]))-Beta0*x[1]-Beta*x[2]-Beta1*x[3]+suma1-
  (Alpha0*log(x[1]))

  Part2=((n*x[1]*x[3]+Alpha+1)*log(x[2]))+(suma2-(n*x[1]*(x[3]+1))
```

```
*log(x[2])))+ lgamma(Alpha)+lgamma(Alpha1)+lgamma(Alpha0)

lvero=-Part1+Part2
return(lvero)
}

#Es la función de soporte
spo=function(x)
{
  #(x[1] > 1) && (x[2]>0) && (-x[1]*x[3]<1) && (x[3]>0)
  (x[1] > 2) && (x[2]>0) && (-x[1]*x[3]<2) && (x[3]>0)
}

#Genera los valores iniciales de forma aleatoria
ini <- function()
{
  x1 <- 2+rgamma(1,0.1,0.1)
  x2 <- rgamma(1,0.1,0.1)
  x3 <- 2/x1+rgamma(1,0.1,0.1)
  return(c(x1,x2,x3))
}

#Número de parámetros a estimar
m<<-3

#Número de iteraciones a ejecutar con el algoritmo t-walk
w<<-4000000

#Inicia el algoritmo t-walk
info1 <- Runtwalk(Tr=w,dim=m, Obj=dpos, Supp=spo, x0=ini(),
xp0=ini(), PlotObj=FALSE);

#Calculala moda
moda=function(datos)
{
  dens=density(datos)
  dens$x[dens$y==max(dens$y)]
}

#Calcula los estimadores de Máxima verosimilitud de la
#distribución Dagum

EMV=vglm(datos~1,family=dagum)
EMV=exp((coef(EMV)))
a1<-EMV[1]
b1<-EMV[2]
```

```
p1<-EMV[3]

#Genera los dos objetos MCMC
cadena1=info1$output
cadena2=info1$outputp

cadenas=mcmc.list(list(mcmc(cadena2,1,nrow(cadena2),thin=10),
                      mcmc(cadena1,1,nrow(cadena1),thin=10)))
#Prueba de convergencia de Gelman y Rubin
gelman.diag(cadenas)
gelman.plot(cadenas)

#Prueba de convergencia de Raftery
raftery.diag(cadenas)

resumen <- summary(cadenas)
resumen$stat[,1]

#Elimina el burn in en el vector de la densidad posterior de los
#parámetros de interés
est <-cadenas[[1]]
est1 <- est[,1]; est1<-est1[3001:10001]
est2 <- est[,2]; est2<-est2[3001:10001]
est3 <- est[,3]; est3<-est3[3001:10001]

#Gráfica de las densidades posteriores marginales de los parámetros
#en cuestión
par(mfrow=c(3,1))
plot(density(info1$output[,1]),col="darkorange",main="Densidad
posterior del parámetro 'a'")
plot(density(info1$output[,2]),col="red",xlim=c(5100,5450),
main="Densidad posterior del parámetro 'b'",ylab="",xlab="")
plot(density(info1$output[,3]),col="darkviolet",main="Densidad
posterior del parámetro 'p'")

#Genera los histogramas de las distribuciones posteriores
#marginales de los parámetros de interés
truehist(info1$output[,1],col="red",breaks=c(seq(0,50,by=.0005)),
xlim=c(2,2.03),main="Histograma posterior del parámetro 'a'",xlab="")
truehist(info1$output[,3],col="red",breaks=c(seq(0,100,
by=0.0005)),xlim=c(.324,.347),main="Histograma posterior
del parámetro 'p'",xlab="")
truehist(info1$output[,2],col="red",breaks=c(seq(0,5600,by=20))
,xlim=c(5000,5600),main="Histograma posterior del parámetro
'b'",xlab="")

#Calcula los estimadores posteriores bayesianos
```



## Anexos

---

```
est.b <- c(modas(est1),modas(est2),modas(est3))
a<-est.b[1]
b<-est.b[2]
p<-est.b[3]

#Gráfica de la densidad del ingreso vs Densidad Dagum ajustada
curve(ddagum(x,a,b,p),0,100000, ylim=c(0,0.001/3),col="blue",main
="Densidad del ingreso vs Densidad Dagum ajustada, 1998",lty=1)
lines(density(datos),col="red",lty=2,lwd=2)
legend("topright",legend=c("Densidad Dagum ajustada ", "Densidad
Ingreso"),lty=c(1,2),col=c("blue", "red"),lwd=c(2,2))

#Calculo del coeficiente de GINI con los Estimadores Bayesianos
ginni.coef <-((gamma(est3)*gamma(2*est3+1/est1))/(gamma(2*est3
*gamma(est3+1/est1)))-1
index=is.na(ginni.coef)
ginni.coef=ginni.coef[!index]
modas(ginni.coef)

#Genera el histograma del coeficiente de Gini
truehist(ginni.coef,xlim=c(0,1),main="Coeficiente de Ginni
1998",xlab="")

#Gráfica de la función de densidad acumulada teórica vs
#la función de densidad acumulada empírica
plot(ecdf(datos),col="green",lty=1,lwd=3,xlim=c(0,7e+04),
main="Función de densidad acumulada, año 1998")
curve(pdagum(x,a,b,p),add=T, col="red",lwd=2)
legend("bottomright",legend=c("Función de densidad acumulada
empírica", "Función de densidad acumulada teórica"),lty=c(1,1),
lwd=c(2,2),col=c("green", "red"))
```