



# **COLEGIO DE POSTGRADUADOS**

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

**CAMPUS MONTECILLO**

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA

**ECONOMÍA**

**USO DE LA DISTRIBUCIÓN PARETO CON ESTIMADORES BAYESIANOS  
PARA MODELAR EL INGRESO EN MÉXICO**

**DAMIÁN TORRES HERNÁNDEZ**

**TESIS**

PRESENTADA COMO REQUISITO PARCIAL  
PARA OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS**

MONTECILLO, TEXCOCO, ESTADO DE MÉXICO

2015

La presente tesis titulada: **Uso de la distribución Pareto con estimadores Bayesianos para modelar el ingreso en México**, realizada por el alumno: **Damián Torres Hernández**, 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**

**ECONOMÍA**

**CONSEJO PARTICULAR**

CONSEJERO:



Dr. José Jaime Arana Coronado

ASESOR:



Dr. José de Jesús Espinoza Arellano

ASESOR:



Dr. Enrique de Jesús Arjona Suárez

Montecillo, Texcoco, Estado de México, diciembre de 2015

# Uso de la distribución Pareto con estimadores Bayesianos para modelar el ingreso en México

Damián Torres Hernández

Colegio de postgraduados, 2015

## RESUMEN

Se analiza la distribución del ingreso personal de los mexicanos como un proceso estocástico, modelado a partir de la distribución Pareto de parámetros forma y escala determinados con estimadores Bayesianos. La información utilizada corresponde a los ingresos trimestrales brutos de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), del Instituto Nacional de Estadística y Geografía (INEGI), de los años 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012 y 2014, valiéndose de todos los datos disponibles para realizar la estimación mediante la simulación Monte Carlo vía Cadenas de Markov, obteniendo parámetros estadísticamente significativos que muestran diferentes grados de desigualdad en términos de la distribución del ingreso en México.

**Palabras clave:** México, Economía, Pareto, Ingreso, Bayes, Densidad, Estadística, Monte Carlo, Cadenas de Markov.

# Using the Pareto distribution with Bayesian estimators for modeling the income in México

Damián Torres Hernández

Colegio de postgraduados, 2015

## ABSTRACT

Is analyzed the distribution of personal income of Mexicans like a stochastic process, modeling as from the Pareto distribution of shape and scale parameters determined with Bayesian estimators. The information used corresponds to quarterly gross income from ENIGH, of the years 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012 y 2014, using all data available for the estimation by means of Markov Chain Monte Carlo, obtaining statistically significant parameters showing different degrees of inequality in terms of income distribution in México.

**Keywords:** México, Economy, Pareto, Income, Bayes, Density, Statistics, Monte Carlo, Markov Chains.

## **AGRADECIMIENTOS**

*Agradezco expresamente al Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), especialmente al Doctor Homero Salinas González, por haberme brindado la oportunidad de realizar mis estudios de Postgrado a fin de formarme como investigador y aportar mayores y mejores soluciones científicas en materia de Economía Agrícola.*

*De la misma manera agradezco profundamente al Colegio de Postgraduados, especialmente a los Postgrados de Economía y Estadística por abrirme las puertas de su enseñanza para adquirir los conocimientos científicos que derivan de dichas ciencias.*

*Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo financiero otorgado durante mis estudios de Maestría.*

*Al Doctor Humberto Vaquera Huerta, quien además de ser un extraordinario investigador y catedrático de Estadística, es un gran ser humano que comparte sus conocimientos incondicionalmente siempre mostrándose amable, disponible y amigo de quien quiere aprender.*

*Al Doctor José Jaime Arana Coronado, quien es un Economista muy inteligente, pero sobre todo una gran persona que me brindó su amistad, me oriento y me apoyó totalmente durante mi estancia en el Postgrado.*

*Al Doctor Enrique de Jesús Arjona Suárez, quien es una plena muestra del sentido humanista, dado que siempre está disponible para enseñar a sus alumnos, no conociendo yo a una persona más paciente, tolerante y comprensiva que al Doctor Arjona.*

*Al Doctor José de Jesús Espinoza Arellano, quien para mí es un modelo a seguir dado que es un gran investigador en Socioeconomía y que ha confiado en mí y me ha apoyado incondicionalmente dentro del INIFAP y durante mi Maestría en el Colegio de Postgraduados.*

*Al Doctor José Saturnino Mora Flores, quien además de ser un gran investigador y catedrático entregado a su profesión, es destacado por ser una persona muy amable que comparte sus conocimientos y apoya a quien lo necesite como lo hizo conmigo.*

*Al Doctor Miguel Ángel Martínez Damián, quien es un Economista de amplio reconocimiento y con quien puede uno acercarse para pedirle hasta la más mínima asesoría dada su sencillez y disposición para ayudar a sus alumnos.*

*A todos ellos:*

*MUCHAS GRACIAS*

**DEDICATORIA**

CON TODO MI AMOR

*A mi hija Lluvia Alejandra Torres Rivera*

*quien se ha convertido en mi inspiración y mis ganas de seguir progresando, por ella sería capaz de conquistar el mundo entero*

*A mi esposa Jocabed Zulema Rivera Sánchez*

*quien me ha apoyado y comprendido en momentos de ausencia*

*A mis padres Máximo Guadalupe Torres Cedillo*

*y Julieta Hernández Rodríguez*

*quienes han estado conmigo en todo momento y han confiado en mí; a ellos quienes en su momento me brindaron la oportunidad de superarme permitiéndome estudiar, porque sin ellos y sin su apoyo no estaría escribiendo estas líneas. Los amo y extraño mucho*

*A mi hermana Anahí Torres Hernández*

*quien me ha dado muchos ánimos en momentos de “estancamiento  
anímico”*

*A mis suegros Alejandro Rivera Castillo y*

*María del Carmen Sánchez Ruiz*

*quienes me han brindado su amor y cariño, no olvidando que me han  
apoyado con mi familia en momentos de ausencia*

*MUY AFECTUOSAMENTE*

*A todos mis compañeros de generación con quienes conviví dentro y  
fuera del Colegio, llevándome recuerdos muy gratos e inolvidables*

# Contenido

1. INTRODUCCIÓN .....	1
2. PLANTEAMIENTO Y OBJETIVOS .....	3
2.1.  Objetivo General .....	5
2.2.  Objetivos Específicos .....	5
3. MÉTODOS BAYESIANOS .....	6
3.1.  Modelos Bayesianos .....	6
3.2.  Proceso Estocástico .....	6
3.3.  Teorema de Bayes .....	7
3.4.  Inferencia Bayesiana .....	8
3.4.1. Distribución a priori o inicial .....	9
3.4.2. Distribución a priori informativa .....	10
3.4.3. Distribución a priori no informativa .....	11
3.4.4. Inicial Invariante .....	12
3.4.5. Teorema Bayesiano Central del Límite .....	12
3.4.6. Distribución a posteriori .....	13
4. MONTE CARLO VÍA CADENAS DE MARKOV .....	15

4.1.	Simulación Monte Carlo .....	15
4.2.	Cadenas de Markov .....	16
4.3.	Monte Carlo vía Cadenas de Markov .....	18
4.4.	Algoritmos de la simulación Monte Carlo vía Cadenas de Markov .....	20
4.4.1.	Algoritmo de muestreo Metropolis-Hastings .....	20
5.	DISTRIBUCIÓN PARETO.....	23
5.1.	Antecedentes .....	23
5.2.	Características .....	23
5.3.	Momentos y medidas de posición y dispersión .....	25
5.4.	Aplicaciones de la distribución Pareto.....	26
6.	MATERIALES Y MÉTODOS.....	28
6.1.	Estimadores de Máxima Verosimilitud de la distribución Pareto .....	28
6.2.	Simulación Monte Carlo vía Cadenas de Markov para determinar los parámetros de la distribución Pareto.....	29
6.2.1.	Diagnósticos de convergencia .....	31
6.2.1.1.	Intervalo de credibilidad.....	31
6.2.1.2.	Errores Estándar de Monte Carlo .....	31
6.2.1.3.	Diagnóstico de Geweke.....	31
6.2.1.4.	Diagnóstico de Raftery-Lewis .....	32

6.2.1.5. Diagnóstico de Heidelberger-Welch .....	33
6.2.1.6. Tamaño de muestras efectivo .....	34
6.3. Indicadores comunes para medir la desigualdad del ingreso.....	35
6.3.1. Curva de Lorenz .....	35
6.3.2. Coeficiente de Gini .....	35
6.4. Encuesta Nacional de Ingresos y Gastos de los Hogares.....	37
7. RESULTADOS.....	39
7.1. Validación del modelo .....	39
7.2. Estimación de los parámetros $\theta$ y $\alpha$ .....	39
7.3. Errores estándar de la simulación y diagnósticos de Geweke y Raftery- Lewis.....	42
7.4. Diagnóstico de Heidelberger-Welch .....	45
7.5. Validez de la simulación Monte Carlo vía Cadenas de Markov.....	48
7.6. Análisis del coeficiente de Gini y la curva de Lorenz.....	52
7.7. Análisis de la distribución Pareto .....	59
7.8. Discusión.....	71
8. CONCLUSIONES Y RECOMENDACIONES .....	73
Referencias.....	74
Anexos .....	87

# Lista de Cuadros

Cuadro 1. Sumarios e intervalos posteriores .....	41
Cuadro 2. Errores estándar de Monte Carlo y diagnósticos de Geweke y Raftery-Lewis .....	44
Cuadro 3. Diagnóstico de Heidelberger-Welch .....	46
Cuadro 4. Tamaño efectivo de la muestra y estadísticos muestrales del ingreso .....	49
Cuadro 5. Sintaxis MCMC de Sas 9.4.....	88

# Lista de Gráficas

Gráfica 1. Comportamiento del coeficiente de Gini y los parámetros de la distribución Pareto. ....	54
Gráfica 2. Comportamiento de la curva de Lorenz en los años 1984, 1989, 1992, 1994 y 1996. ....	56
Gráfica 3. Comportamiento de la curva de Lorenz en los años 1998, 2000, 2002, 2004 y 2005. ....	57
Gráfica 4. Comportamiento de la curva de Lorenz en los años 2006, 2008, 2010, 2012 y 2014. ....	58
Gráfica 5. Comportamiento de la función de distribución acumulada en base al valor de $\alpha$ . ....	60
Gráfica 6. Comportamiento de la función de densidad de probabilidad en base al valor de $\alpha$ . ....	62
Gráfica 7. Comportamiento de la función de distribución acumulada de los años 1984, 1989, 1992, 1994 y 1996, en base al valor de $\alpha$ . ....	64
Gráfica 8. Comportamiento de la función de densidad de probabilidad de los años 1984, 1989, 1992, 1994 y 1996, en base al valor de $\alpha$ . ....	65
Gráfica 9. Comportamiento de la función de distribución acumulada de los años 1998, 2000, 2002, 2004 y 2005, en base al valor de $\alpha$ . ....	66
Gráfica 10. Comportamiento de la función de densidad de probabilidad de los años 1998, 2000, 2002, 2004 y 2005, en base al valor de $\alpha$ . ....	67

Gráfica 11. Comportamiento de la función de distribución acumulada de los años 2006, 2008, 2010, 2012 y 2014, en base al valor de $\alpha$ .	68
Gráfica 12. Comportamiento de la función de densidad de probabilidad de los años 2006, 2008, 2010, 2012 y 2014, en base al valor de $\alpha$ .	69

## Lista de Figuras

Figura 1. Procedimiento MCMC. Elaboración propia.	30
Figura 2. Curva de Lorenz y coeficiente de Gini.	36
Figura 3. Diagnóstico visual de convergencia para $\theta$ de 2014.	50
Figura 4. Diagnóstico visual de convergencia para $\alpha$ de 2014.	51
Figura 5. Diagnóstico visual de convergencia para $\theta$ y $\alpha$ de 2008.	90

# 1. Introducción

Todo sistema económico al ser formulado y empleado para el desarrollo de una sociedad posee indicadores que permiten visualizar el progreso de los individuos en el ámbito económico. Al ser partícipes de una economía como mano de obra con necesidades en espera de ser satisfechas, se tiene la expectativa de que la remuneración económica por el empleo de la fuerza de trabajo proporcione bienestar y, por ende, un mejor nivel de vida. Lo anterior se ha comportado como un fenómeno socioeconómico latente a los largo de los años pasando a ser objeto de análisis para muchos economistas en cuanto a la distribución del ingreso, el cual es un indicador del dinamismo y de la eficiencia de un sistema económico (Montes, 2012).

Los temas de bienestar económico y la desigualdad en la distribución del ingreso se han abordado frecuentemente analizando la relación entre ambas, y lo que cada temática afronta. Lo anterior abre paso para realizar investigaciones más a profundidad utilizando técnicas tanto económicas como estadísticas que permitan medir los niveles de desigualdad en cuanto a la distribución de la riqueza.

En la presente investigación se considera analizar la distribución del ingreso per cápita de los mexicanos como un proceso estocástico desde el punto de vista Bayesiano.

Lo anterior se desarrolla en 8 capítulos, en el siguiente desglose: en el capítulo 1 se desarrolla la introducción de la presente tesis; el capítulo 2 refiere al planteamiento de la investigación desde un contexto estadístico, donde se expone el interés por la metodología propuesta para realizar la modelación de los datos del ingreso de los mexicanos; así mismo, se exponen el objetivo general y los objetivos específicos. En los capítulos 3 a 5 se aborda el marco teórico, considerando en el capítulo 3 el análisis de los fundamentos de los modelos Bayesianos. En el capítulo 4 se analiza la estructura, el procedimiento y diagnóstico de la simulación Monte Carlo vía Cadenas de Markov mientras que en el capítulo 5 se analiza la esencia de la distribución Pareto.

El capítulo 6 expone los materiales y métodos, e incluye los métodos de simulación Monte Carlo vía Cadenas de Markov para determinar los estimadores Bayesianos de los parámetros de la distribución Pareto; los indicadores comunes para medir la desigualdad incluyendo la curva de Lorenz y el coeficiente de Gini, y las particularidades de la Encuesta Nacional de Ingresos y Gastos de los Hogares de donde se tomaron los datos para la presente investigación. En el capítulo 7, finalmente se obtienen y analizan los resultados estadísticos del modelado de los datos del ingreso per cápita de los mexicanos dándose la discusión de la información desde el punto de vista socioeconómico. Las conclusiones y recomendaciones se abordan en el capítulo 8, contrastando los objetivos planteados al inicio de la investigación con la información obtenida al final de la misma.

## 2. Planteamiento y objetivos

Varias son las distribuciones probabilísticas que se han vinculado al análisis de la distribución del ingreso, entre ellas, Lusting (1976), menciona la distribución Pareto, la distribución normal y la log-normal, Botargues y Petrecolla (1999) refieren a las distribuciones Pareto, log-normal y la Gamma. Por su parte Montes, Pérez, y Pérez (2012), utilizan y comparan las distribuciones Pareto, log-normal y Dagum.

Por otra parte, los métodos Bayesianos de estimación han sido de gran interés para muchos campos de la investigación aplicada, incluyendo la economía. La esencia de dichos métodos es incorporar información heurística previa a la toma de decisiones lo que posibilita su actualización objetiva dadas nuevas observaciones. Jaramillo (2009); Urbisaia y Brufman (2009); y Zellner (1985, 1988), conjuntan la econometría y la inferencia Bayesiana.

La formulación Bayesiana asume los parámetros como variables aleatorias que tienen distribuciones previas, es decir, incorpora una distribución de probabilidad a priori sobre los parámetros a estimar (antes de observar los datos) para hacer inferencias acerca de sus valores a posteriori (una vez observados los datos). La distribución a priori representa el grado personal de incertidumbre que el investigador tiene respecto a los parámetros, por lo tanto, contiene un componente esencial netamente subjetivo (Press, 2009). Este hecho obliga al investigador a utilizar, tanto su evidencia previa como la empírica proporcionada por el fenómeno de análisis. La metodología Bayesiana hace uso de estas dos fuentes de información para obtener mediante el Teorema de Bayes, la información final sobre la cual se tomara la decisión (Eyzaguirre y Chue, 2001; Geweke, Koop, y Van Dijk, 2011; Salas, 2011; Zellner, 1988).

Por último, el desarrollo de la simulación Monte Carlo vía Cadenas de Markov (MCMC, por sus siglas en inglés) ha permitido la estimación de modelos complejos que hasta entonces habían sido imposibles de resolver por requerir complicados problemas de

integración múltiple. Además, en los casos en que no es posible la simulación directa de muestras independientes puede ser posible simular muestras con algún tipo de dependencia que converjan (bajo ciertas condiciones de regularidad) a la distribución de interés. Un ejemplo de estos problemas es la inclusión de variables latentes (o variables ocultas) que no son directamente observadas por el investigador (Araña, 2004). El procedimiento Monte Carlo vía Cadenas de Markov utiliza algoritmos de simulación para obtener estimaciones posteriores (DeGroot y Schervish, 2002; Gilks, Richardson, y Spiegelhalter, 1996; G. J. Gómez y Palarea, 2003; Lee, 2012; Zellner, 1996).

Por lo anterior planteado, se considera a la conjunción de la distribución Pareto con el método Bayesiano y el procedimiento MCMC, como adecuada para modelar el ingreso personal de los mexicanos. La información estadística objetiva obtenida permitirá inferir sobre el comportamiento de la distribución de la riqueza en México el cual es el objetivo principal de la presente investigación.

## 2.1. Objetivo General

Modelar el ingreso en México para generar información estadística objetiva que permita inferir sobre el comportamiento del ingreso en México.

## 2.2. Objetivos Específicos

- ♣ Analizar las bases teóricas del método Bayesiano, de la distribución Pareto y de la simulación Monte Carlo vía Cadenas de Markov que permitan construir un modelo estadístico apropiado para modelar el ingreso per cápita en México.
- ♣ Obtener la mayor información disponible para ajustar y validar el modelo propuesto.
- ♣ Modelar el ingreso per cápita en México con la distribución Pareto, estimadores Bayesianos y cadenas de Markov.
- ♣ Analizar el comportamiento de la distribución del ingreso en México.

## 3. Métodos Bayesianos

### 3.1. Modelos Bayesianos

Los métodos estadísticos utilizados más frecuentemente se conocen como clásicos (o métodos frecuentistas). Estos métodos suponen que los parámetros desconocidos son constantes o fijos, y definen la probabilidad mediante la limitación de las frecuencias relativas (Gelman, Carlin, Stern, y Rubin, 2014).

Los métodos Bayesianos ofrecen un enfoque alternativo; tratan a los parámetros como variables aleatorias y definen la probabilidad como "grados de creencia", es decir, la probabilidad de un evento es el grado en que se cree que el evento es verdadero. De lo anterior, se deduce entonces que las probabilidades son subjetivas y que se pueden hacer declaraciones de probabilidad acerca de los parámetros. El término "Bayesiano" viene del uso frecuente del Teorema de Bayes, del reverendo Thomas Bayes (1702-1761), un ministro presbiteriano del siglo XVIII. Bayes estaba interesado en la solución de la cuestión de la probabilidad inversa: después de observar una colección de eventos, ¿cuál es la probabilidad de un evento? (Arnold y Press, 1983).

El Teorema de Bayes explica cómo actualizar los conocimientos existentes con nueva información. Se comienza con una "creencia previa" (distribución a priori)  $P(\theta)$ , y después de conocer la información de los datos observados  $X$ , cambia o actualiza la "creencia" en  $\theta$  para obtener la distribución a posteriori  $P(\theta|X)$ .

### 3.2. Proceso Estocástico

Un proceso estocástico es un concepto estadístico que sirve para caracterizar una sucesión de variables aleatorias (estocásticas) que evolucionan en función de otra variable, generalmente el tiempo. Cada una de las variables aleatorias del proceso

tiene su propia función de distribución de probabilidad. Cada variable o conjunto de variables sometidas a influencias o efectos aleatorios constituye un proceso aleatorio o proceso estocástico (Dagum y de Dagum, 1971).

En la mayoría de los procesos estocásticos cada resultado depende de lo que sucedió en etapas anteriores del proceso (Arya y Lardner, 2002). Champernowne (1953), realizó un análisis de los ingresos formulando un modelo como un proceso estocástico.

### 3.3. Teorema de Bayes

Desde el punto de vista Bayesiano, todas las probabilidades son condicionales porque casi siempre existe algún conocimiento o experiencia acerca de los sucesos; así el concepto básico en Estadística Bayesiana es la probabilidad condicional.

Sean  $A_1, A_2, \dots, A_k$ , eventos que forman una partición de un espacio muestral  $M$ . Sea  $B$  un evento en  $M$ . Suponga que  $P(A_1), P(A_2), \dots, P(A_k), P(B|A_1), P(B|A_2), \dots, P(B|A_k)$  son probabilidades desconocidas (Infante y Zàrate, 2012). Entonces:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

donde

- ♠  $P(A)$  corresponde a la distribución *a priori* de  $A$
- ♠  $P(B)$  corresponde a la distribución marginal de  $B$
- ♠  $P(B|A)$  corresponde a la distribución condicional de  $B$ , dado  $A$
- ♠  $P(A|B)$  corresponde a la distribución *a posteriori* de  $A$ , dado  $B$

### 3.4. Inferencia Bayesiana

Representando los datos observados mediante  $X$ , y los parámetros desconocidos de un modelo como  $\theta$ , asumiendo que  $\theta$  es aleatorio, es necesario formular la distribución de probabilidad conjunta  $P(X, \theta)$ . Esta distribución de probabilidad asume que cualquier inferencia sobre  $\theta$  está basada en su distribución final o posterior  $P(X|\theta)$  y tiene dos componentes: la distribución de probabilidad inicial o *a priori*  $P(\theta)$  y la función de verosimilitud  $P(X|\theta)$ . La especificación de ambas constituye el modelo probabilístico completo:

$$P(X, \theta) = P(\theta)P(X|\theta) \quad (2)$$

Una vez observados los datos  $X$ , el Teorema de Bayes permite establecer la distribución de probabilidad de  $\theta$  *a posteriori*, condicionada por  $X$ :

$$P(\theta|X) = \frac{P(\theta)P(X|\theta)}{\int P(\theta)P(X|\theta)d\theta} \quad (3)$$

Donde la distribución *a posteriori* es proporcional a la distribución *a priori* multiplicada por la *función de verosimilitud* (ecuación(2)) (Gilks *et al.*, 1996).

El análisis de esta distribución *a posteriori* de  $\theta$  puede llevarse a cabo desde el punto de vista Bayesiano a partir de los momentos, los cuantiles, las regiones de mayor probabilidad, etc. Todas estas características se pueden expresar en términos de expectativas posteriores de la función de  $\theta$ . Tales expectativas posteriores de una función  $f(\theta)$  son:

$$E\{f(\theta)|X\} = \frac{\int f(\theta)P(\theta)P(X|\theta)d\theta}{\int P(\theta)P(X|\theta)d\theta} \quad (4)$$

La evaluación analítica de esta expresión puede ser prácticamente imposible, incluyéndose como alternativa de solución los procedimientos MCMC (Gilks *et al.*, 1996).

Sea  $X$  un vector de  $k$  variables aleatorias, con distribución de probabilidad  $\pi(\cdot)$  y se trata de evaluar el valor esperado:

$$E[f(X)] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \quad (5)$$

Se sabe por la ley de los grandes números que un buen estimador del valor esperado de una variable aleatoria continua con distribución  $F$  es el valor promedio de una muestra finita de variables aleatorias, independientes con distribución  $F$  (Saavedra y Ibarra, 2013).

Con la integración Monte Carlo, la evaluación de  $E[f(X)]$  sobre la muestra  $\{X_t, t = 1, \dots, n\}$  de distribución  $\pi(\cdot)$  es:

$$E[f(X)] = \frac{1}{n} \sum_{t=1}^n f(X_t) \quad (6)$$

Donde la media poblacional de  $f(X)$  es estimada por la media muestral. Cuando las muestras  $\{X_t\}$  son independientes, la ley de los grandes números asegura que la aproximación puede ser tan precisa como se quiera, aumentando el tamaño de la muestra  $n$ .

En general, la obtención de muestras independientes  $\{X_t\}$  no es factible cuando  $F$  no es una distribución estándar. Sin embargo,  $\{X_t\}$  no tienen que ser necesariamente independientes, pudiendo ser generadas por algún proceso como la cadena de Markov que tenga a  $\pi(\cdot)$  como distribución estacionaria, lo que se logra mediante los procedimientos MCMC (Gilks *et al.*, 1996).

### 3.4.1. Distribución a priori o inicial

En inferencia Bayesiana, una distribución de probabilidad a priori  $P(\theta)$ ;  $\theta \in \Theta$ , de un parámetro  $\theta$  incierto o variable desconocida, es una distribución de

probabilidad que expresa la incertidumbre sobre  $\theta$  antes de tomar en cuenta los datos  $X$ . Los parámetros de una distribución a priori se llaman hiperparámetros, para distinguirlos de los parámetros del modelo.

Al aplicar el Teorema de Bayes, la distribución a priori se multiplica por la función de verosimilitud normalizando luego para estimar la distribución de probabilidad posterior, que es la distribución condicional de  $\theta$  dados los datos. Por otra parte, la distribución previa (a priori) afecta a la distribución posterior.

Si se tiene un conocimiento previo sobre los parámetros, este se traducirá en una distribución a priori, pero como 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 Bayesiano, es el de elegir convenientemente una distribución a priori para los parámetros que se presumen desconocidos (Montes, 2012).

Dentro del paradigma Bayesiano, la probabilidad a priori Bayesiana mide el grado de “creencia” que se tiene sobre el evento aleatorio. Dada esta definición, la probabilidad es muy subjetiva y tiene un impacto considerable sobre la distribución posterior, por lo que han sido grandes los esfuerzos por obtener resultados que sean objetivamente válidos. Esto se puede lograr mediante el uso de distribuciones a priori que son objetivas, es decir que tienen un impacto mínimo sobre la distribución posterior. Estas distribuciones se llaman priores objetivas o no informativas. Sin embargo, mientras las priores no informativas son muy populares en algunas aplicaciones, no siempre son fáciles de construir (DeGroot y Schervish, 2002; Press, 2009).

### **3.4.2. Distribución a priori informativa**

Una distribución a priori informativa es un priori que no está dominado por la probabilidad y que tiene un impacto en la distribución posterior. Si una distribución previa domina la probabilidad, es claramente un prior informativo. Estos tipos de distribuciones deben especificarse con cuidado en la práctica. Por otro lado, el uso

adecuado de distribuciones a priori ilustra el poder del método Bayesiano: la información recopilada del estudio anterior, la experiencia pasada, o la opinión de expertos se puede combinar con la información actual (Kass y Wasserman, 1996).

Las distribuciones Gamma y Gamma-inversa han sido manejadas ampliamente en el análisis Bayesiano. La primera a menudo se ha utilizado como la distribución a priori del parámetro de precisión  $\tau = 1/\sigma^2$  en una distribución normal (Sas, 2013). Esta es una distribución flexible que comúnmente ofrece un buen ajuste como priori de una distribución exponencial (Waller, Johnson, Waterman, y Martz Jr, 1976).

La parametrización de la distribución Gamma suele usarse eligiendo un número mínimo y de mismo valor para ambos hiperparámetros de la misma, situación que ha sido muy usual en el software WinBUGS de análisis Bayesiano (Sas, 2013).

### **3.4.3. Distribución a priori no informativa**

Una distribución a priori es no informativa (o “flat”) si el prior es difuso con respecto a la función de verosimilitud y por lo tanto tiene el mínimo impacto en la distribución posterior de  $\theta$ .

Al utilizar una distribución a priori no informativa, la información resultante en la distribución posterior  $P(\theta|X)$  surge de los datos, y en consecuencia la inferencia resultante será completamente aleatoria (Carlin y Louis, 2000; Gelman, 2009; Press, 2009).

El postulado de Bayes/Laplace establece que cuando nada conocemos acerca de  $\theta$  la función a priori  $P(\theta)$  es una distribución uniforme, es decir, todos los posibles resultados de  $\theta$  tienen la misma probabilidad (Kass y Wasserman, 1996; Stigler, 1982).

### 3.4.4. Inicial Invariante

Teóricamente, el modelo y la distribución inicial deben ser invariantes bajo la acción del mismo grupo empleando la media de Haar derecha como la inicial no informativa; pero una dificultad con este enfoque, es que todos los problemas carecen de una estructura invariante y además, la media de Haar derecha no siempre existe (Dawid, 2004; Hartigan, 1964; Kass y Wasserman, 1996)

### 3.4.5. Teorema Bayesiano Central del Límite

Suponga que  $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} f_i(x_i|\theta)$ , de tal manera que  $L(X|\theta) = \prod_{i=1}^n f_i(x_i|\theta)$ . Suponga además que la densidad a priori  $\pi(\theta)$  y la verosimilitud  $L(X|\theta)$  son positivos y dos veces diferenciables cerca de  $\hat{\theta}^\pi$ ; se asume que existe la moda posterior (o “máxima verosimilitud generalizada”) de  $\theta$ . Entonces bajo condiciones de regularidad, la distribución posterior  $P(\theta|X)$  para  $n$  grande puede aproximarse por una distribución normal con media igual a la moda posterior y matriz de covarianzas igual o menor a la inversa Hessiana (matriz de segundas derivadas) de la log posterior evaluada en la moda. Esta matriz a veces denotada como  $[I^\pi(X)]^{-1}$ , es la información observada por la matriz de Fisher para  $\theta$ . Más específicamente (Berry, Carlin, Lee, y Muller, 2010):

$$I_{ij}^\pi(X) = - \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(X|\theta)\pi(\theta)) \right]_{\theta=\hat{\theta}^\pi} \quad (7)$$

El Teorema Bayesiano Central del Límite parece provenir del hecho de que el teorema muestra la parte posterior de ser aproximadamente normal para grandes tamaños de muestra. Sólo el Teorema Central del Límite “regular” proporciona la normalidad aproximada de pruebas estadísticas frecuentistas en muestras grandes (Berger, 1985; Berry *et al.*, 2010).

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 (Lee, 2012).

### 3.4.6. Distribución a posteriori

Con el Teorema de Bayes se obtiene la expresión general para la densidad posterior de  $\theta$  dada en  $P(\theta|X) = \frac{P(\theta)P(X|\theta)}{\int P(\theta)P(X|\theta)d\theta}$ ; esta expresión permite notar que la densidad posterior involucra una distribución de los datos observados a través de  $L(\theta|X)$ , y una contribución de la información a priori cuantificada a través de  $P(\theta)$ .

Entonces la distribución posterior se calcula mediante el Teorema de Bayes como:

$$P(\theta|X) \propto P(\theta)L(\theta|X) \quad (8)$$

La expresión  $m(X) = \int_{\Omega} L(X|\theta)P(\theta|X)d\theta$ , es la constante de normalización de la densidad  $f(X|\theta)$ , y comúnmente se le llama la distribución marginal de los datos o la distribución predictiva a priori (Gelman, 2009). Se ha expresado que un análisis Bayesiano basa sus inferencias en la densidad posterior, sin embargo, en muchos modelos ocurre que  $m(X)$  no tiene una forma analítica cerrada; este dilema conduce a la pregunta ¿Cómo muestrear la distribución multivariable posterior  $P(\theta|X)$ , cuando ésta no tiene una forma cerrada?; ésta pregunta a la vez conduce a una vasta literatura sobre métodos computacionales para muestrear  $P(\theta|X)$  y métodos para estimar  $m(X)$ . Dentro de los métodos computacionales más conocidos se encuentra el muestreador de Gibbs, el método de aceptación y rechazo, el método de aceptación y rechazo adaptativo, y el algoritmo de Metropolis-Hastings (Montes, 2012).

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

- a) La media de  $\theta$ , si  $h(\theta) = \theta$
- b) La covarianza de  $\theta_i$  y  $\theta_j$ , si  $h(\theta) = [\theta_i - E(\theta_i|X)][\theta_j - E(\theta_j|X)]$ ,

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

- c) La densidad predictiva, si  $h(\theta) = f(z|\theta)$ , donde  $f(z|\theta)$  es la densidad predictiva dado el vector de parámetros.
- d) La probabilidad de un conjunto  $A$ , si  $h(\theta) = I_A(\theta)$ .

De acuerdo con M. H. Chen, Shao, e Ibrahim (2012), otras cantidades que pueden obtenerse con  $E_{p\theta}[h(\theta)]$  son: las densidades marginales de cada uno o de un conjunto de los componentes del vector  $\theta$ , las constantes de normalización, los factores de Bayes, densidades finales, cuantiles e intervalos.

## 4. Monte Carlo vía Cadenas de Markov

### 4.1. Simulación Monte Carlo

Montecarlo es un área administrativa que forma parte del principado de Mónaco, esencialmente famoso por su casino Monte Carlo. El método Monte Carlo se le atribuye al físico italiano Enrico Fermi (1901-1954) quien había inventado durante la década de 1930 dicho método (no llamado así en ese entonces) mientras estudiaba difusión de neutrones en Roma, de lo que no publico nada sobre el tema, pero uso el método para resolver muchos problemas de cálculo. Utilizando el método de Fermi, el nombre Monte Carlo se origina hacia 1942 entre John Von Neumann, Stanislaw Ulam y Nicholas Metropolis como palabra clave en la experimentación de la bomba atómica, por la necesidad de simular en la computadora el proceso de difusión nuclear, para la cual se requiere de la utilización de los números aleatorios similares a los que genera la ruleta que opera en el casino Monte Carlo, así Metropolis sugirió el nombre formal para el método al considerar el interés de Ulam en juegos de azar y sobre todo al escuchar la anécdota del tío de Ulam quien había pedido dinero prestado para apostar en dicho casino (Metropolis, 1987). Como método ya establecido es hasta 1949 cuando Metropolis y Ulam publican el artículo "The Monte Carlo Method" (Dunn y Shultis, 2011; Sobol, 1994).

Simular tiene como objetivo duplicar características y comportamientos de un sistema real. El método Monte Carlo (o simulación Monte Carlo) genera muestras de la distribución de probabilidad de interés y a partir de ellas evalúa los momentos de la distribución de probabilidad a posteriori (Alonso, 2008). El método es útil para obtener soluciones numéricas a problemas que son demasiado complicados de resolver analíticamente realizando repeticiones y reconociendo el comportamiento del sistema. A diferencia de los métodos numéricos que se basan en evaluaciones en  $N$  puntos en un espacio  $M$ -dimensional para producir una solución aproximada, el método Monte

Carlo tiene un error absoluto de la estimación que decrece de acuerdo al Teorema del Límite Central (Geyer, 1992).

## 4.2. Cadenas de Markov

La cadena de Markov debe su nombre al matemático ruso Andrei Andreyevich Markov (1856-1922). Su trabajo en el campo de los procesos de componentes aleatorios (procesos estocásticos) derivaron en un instrumento matemático conocido actualmente como cadena de Markov, la cual se define como una secuencia de valores de una variable aleatoria en las que el valor de la variable en el futuro depende del valor de la variable en el presente, pero es independiente de la historia de dicha variable (propiedad markoviana). La cadena de Markov se considera una herramienta de utilidad en disciplinas como la economía, la ingeniería, la investigación de operaciones y muchas otras (Budnick, 1990).

Intuitivamente, una cadena de Markov representa el movimiento aleatorio de un objeto. Podemos considerar  $(X_n)$  como la posición o el valor de un objeto en el instante  $n$ . En el tal caso existen reglas que permiten conocer las probabilidades del lugar en que se situará el objeto en el siguiente movimiento.

Una cadena de Markov requiere un *espacio de estados*  $S$  que es el conjunto de todos los lugares a los que puede ir el objeto. Por ejemplo  $S = \{1, 2, 3\}$ ,  $S = \{\text{fondo, superficie}\}$ , o  $S$  puede ser el conjunto de todos los enteros positivos, además del *ciclo de Markov*, al cual corresponde el periodo de tiempo que sirve de base para examinar las transiciones entre estados.

La cadena requiere también una *probabilidad de transición*, que da la probabilidad que describe hacia donde se desplazará el objeto en el siguiente salto. Concretamente para  $i, j \in S$ , el valor  $p_{ij}$  es la probabilidad de que, si el objeto está en  $i$ , efectúe el siguiente salto a  $j$ . Por lo tanto el conjunto  $\{p_{ij} : i, j \in S\}$  de probabilidades de transición cumple  $p_{ij} \geq 0$  para todo  $i, j \in S$ , y

$$\sum_{j \in S} p_{ij} = 1$$

Para cada  $i \in S$ .

También es necesario establecer el lugar en que *comienza* la cadena de Markov. Generalmente establecemos sencillamente que  $X_0 = s$  para algún estado determinado  $s \in S$ . De forma más genérica, podríamos tener una *distribución inicial*  $\mu_i = P(X_0 = i)$ . En este caso necesitamos que  $\mu_i \geq 0$  para todo  $i \in S$ , y

$$\sum_{j \in S} \mu_j = 1$$

Resumiendo,  $S$  es el espacio de estados de todos los lugares a los que el objeto puede desplazarse,  $\mu_i$  representa la probabilidad de que el objeto está en el punto  $i$ , se desplace al punto  $j$  en el siguiente movimiento. En términos de la secuencia de valores aleatorios  $X_0, X_1, X_2, \dots$ , tenemos que

$$P(X_{n+1} = j | X_n = i) = p_{ij}$$

Siendo necesario que la probabilidad de desplazamiento no dependa de la historia previa de la cadena. Es decir:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = X_{n-1}, \dots, X_0 = x_0) = p_{ij}$$

Para todo  $n$  y  $i, j, x_0, \dots, x_{n-1} \in S$ .

Es conveniente disponer las probabilidades de transición como la matriz  $P = p_{ij}$  de  $m \times m$ , llamada matriz de transición de la cadena de Markov, matriz estocástica y matriz de probabilidades (Evans y Rosenthal, 2005).

Para considerar un matriz de transición en  $n$  pasos se debe satisfacer:

♠  $p_{ij}^{(n)} \geq 0$  para toda  $i$  y toda  $j$ ,  $n = 0, 1, \dots, m$

- ♠  $\sum_{j=0}^m p_{ij}^{(n)} = 1$  para toda  $i, n = 0, 1, \dots, m$
- ♠ asumiendo las probabilidades iniciales  $P = \{X_0 = 1\}$  para toda  $i$

	<i>Estado</i>	0	1	2	...	<i>m</i>	
$P^{(n)} =$	0	$P_{00}^{(n)}$	$P_{01}^{(n)}$	$P_{02}^{(n)}$	...	$P_{0m}^{(n)}$	
	1	$P_{10}^{(n)}$	$P_{11}^{(n)}$	$P_{12}^{(n)}$	...	$P_{1m}^{(n)}$	
	2	$P_{20}^{(n)}$	$P_{21}^{(n)}$	$P_{22}^{(n)}$	...	$P_{2m}^{(n)}$	
	...	...	...	...	...	...	
	<i>m</i>	$P_{m0}^{(n)}$	$P_{m1}^{(n)}$	$P_{m2}^{(n)}$	...	$P_{mm}^{(n)}$	$= (p_{ij})_{i,j \in S}$

### 4.3. Monte Carlo vía Cadenas de Markov

Las técnicas de Monte Carlo vía Cadenas de Markov (MCMC, por sus siglas en inglés) permiten generar de manera interativa, observaciones de distribuciones multivariadas que difícilmente podrían simularse utilizando métodos directos. La idea básica consiste en construir una cadena de Markov que sea fácil de simular, y cuya distribución simulada corresponda a la distribución final de interés (A. F. M. Smith y Roberts, 1993).

La idea de los métodos Monte Carlo vía Cadenas de Markov consiste en simular valores ligeramente dependientes en una cadena de Markov independiente e idénticamente distribuida (i.i.d.)  $\theta_1, \theta_2, \dots, \theta_n$  cuya distribución estacionaria sea  $\pi(\theta|X)$ . Para aproximarse a esta distribución se requiere generar  $n$  muestras i.i.d., de  $X$  evaluando repetidamente el estadístico (distribución posterior en un análisis Bayesiano o la varianza de un estadístico en un análisis frecuentista) sobre ellas  $t(x)$ , y aproximando  $\pi$  mediante la distribución empírica de valores obtenidos. Cada valor simulado  $\theta_t$ , depende únicamente de su predecesor  $\theta_{t-1}$ . Si el algoritmo se implementa correctamente, la convergencia de la cadena está garantizada independientemente de cuales sean los valores iniciales.

Es necesario simular la cadena para un número elevado de iteraciones para que la distribución estacionaria  $\varphi(\cdot)$  sea precisamente la distribución de interés  $\pi(\cdot)$ . Los primeros valores simulados, iteraciones *burn-in*, se eliminan porque no están en el estado estacionario.

Haciendo referencia a Gilks *et al.* (1996), una cadena de Markov converge a una distribución estacionaria  $\varphi(\cdot)$ , si  $P^t(\cdot | x_0)$  converge eventualmente a  $\varphi(\cdot)$ . Es decir la cadena olvida gradualmente su estado inicial y no depende de  $t$ .

Para que la distribución de  $X_t$  sea convergente a una distribución estacionaria, la cadena debe considerar tres propiedades importantes referidas como condiciones de regularidad de una cadena de Markov:

- ♠ Que la cadena sea *irreducible*, es decir, que para toda  $i, j$  exista  $t \geq 0$ , tal que  $P^t(X_t = j | X_0 = i) > 0$ .
- ♠ La cadena debe ser *no periódica*, es decir, si para cada  $i$  se cumple que el máximo común divisor de  $X_0\{t > 0: P^t(X_t = i | X_0 = i) > 0\}$  es igual a 1.
- ♠ La cadena debe ser *recurrente positiva*, si para toda  $i, j$  se cumple que:

$$\sum_t P^t(X_t = j | X_0 = i) = \infty$$

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

$$\sum_t \varphi(i) P^t(X_t = j | X_0 = i) = \varphi(j), \forall j, t \geq 0$$

Si la cadena de Markov cumple con dichas condiciones, entonces se dice que la cadena es ergódica y el teorema ergódico permite realizar la integración Monte Carlo para simular y así obtener las cantidades de interés ignorando la dependencia en la simulación (Gilks *et al.*, 1996).

## 4.4. Algoritmos de la simulación Monte Carlo vía Cadenas de Markov

Lo que se pretende con los algoritmos MCMC es generar valores de una variable aleatoria  $X$  con distribución de probabilidad  $\pi(x)$ , típicamente multidimensional, y de la que no es posible muestrear directamente. Para ello, se simula una cadena de Markov ergódica que tiene como distribución estacionaria la distribución objetivo  $\pi(x)$ , tras un número suficientemente grande de iteraciones que estarán generando muestras aproximadas de  $\pi(x)$ .

La cuestión clave es cómo se producen las transiciones de un estado a otro de la cadena para que se produzca la convergencia. El algoritmo básico propuesto por Metropolis, Rosenbluth, Rosenbluth, Teller, y Teller (1953), fue generalizado por Hastings (1970), dando lugar al conocido *algoritmo de Metropolis-Hastings*, que constituye la versión más general de la familia de algoritmos MCMC. Más tarde Geman y Geman (1984) elaboran un método de simulación, que también genera una cadena de Markov y que, tras ser distinguido en Gelfand y Smith (1990), pasa a ser conocido en la literatura como *muestreador de Gibbs*. Dicho muestreador puede verse como un caso particular del algoritmo Metropolis-Hastings, sin embargo, como destacan Robert y Casella (1998, 2013), tiene algunas características particulares que le dan entidad propia (G. J. Gómez y Palarea, 2003).

### 4.4.1. Algoritmo de muestreo Metropolis-Hastings

El método en su forma más simple lo publica Metropolis *et al.* (1953). Dada la distribución a posteriori de un parámetro  $\theta$ ,  $\pi(\theta|X)$ , 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  $\pi(\theta|X)$ .

Implementación del algoritmo Metropolis-Hastings.

- 1 Se elige arbitrariamente un valor inicial  $\theta^{(0)}$  de la densidad a posteriori de  $\pi(\theta|X)$  para el cual  $\pi(\theta^{(0)}|X) > 0$ .
- 2 Para  $t = 1, 2, \dots, N$ .
  - a) Se elige un candidato  $\theta^*$  de una distribución de salto  $J_t(\theta^*|\theta^{t-1})$ . 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,

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

- c) Si  $r > 1$ , entonces  $\theta^t = \theta^*$ , ir al paso 2.
- d) Si  $r < 1$ , entonces generar  $u \sim Unif(0,1)$ , si  $u < r$ , entonces  $\theta^t = \theta^*$ .
- e) Ir al paso 2 hasta que  $t = N$ .

Cuando se ejecuta un algoritmo MCMC, es importante examinar si los valores simulados  $\theta_t$ , han convergido aproximadamente a la distribución estacionaria  $\pi(\theta|X)$  (Ausín, 2012).

Es considerable:

- ♠ Examinar el trabajo del algoritmo al estar explorando el espacio de estados
- ♠ Verificar la convergencia de las medias de los valores simulados en MCMC, e.g.  
$$\frac{1}{N} \sum_{t=1}^N \theta_t \rightarrow E(\theta|X)$$
- ♠ Analizar si los valores simulados son aproximadamente una muestra de valores independientes e idénticamente distribuidos (i.i.d.).

Ausín (2012), comenta que existen numerosos procedimientos en la literatura para estudiar la convergencia de la cadena. Una posibilidad es ejecutar al algoritmo varias veces comenzando en distintos valores iniciales y comprobar si el comportamiento de la distribución estacionaria es la misma.

## 5. Distribución Pareto

### 5.1. Antecedentes

La distribución del ingreso se ha analizado desde dos perspectivas científicas que han trascendido en el tiempo. La primera de ellas se debe a David Ricardo, quien en 1817 realizó indagaciones sobre la distribución funcional del ingreso, tema del cual derivó la formación de precios de los factores de la producción y la distribución del producto neto entre los propietarios de dichos factores (Dagum, 1980). El segundo enfoque se debe a Vilfredo Pareto (1848-1923), quien planteó investigaciones sobre la distribución personal del ingreso, especificando su modelo tipo I en (1895) y posteriormente los modelos tipo II en (1896) y III en (1897). En este último año, Pareto publicó en su libro "*Cours d'Économie Politique*" una exposición de la llamada Ley de Pareto, en donde formulaba la primera representación analítica de la curva de ingreso en cuanto a su distribución entre las personas (Bouchaud y Potters, 2000).

De ese análisis dedujo que en todos los países sin importar la época, la distribución de la riqueza sigue un patrón logarítmico que se observa en una imagen como una "cola", en cuyo extremo izquierdo se aloja una pequeña fracción de la población la cual es dueña de la mayoría de la riqueza (Luque, 2007).

### 5.2. Características

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:

$$\log N_x = A - \alpha \log x \quad (9)$$

que es lo mismo que:

$$N_x = e^A x^{-\alpha} \quad (10)$$

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

$$\frac{N_x}{N} = 1 - F(x) = \left(\frac{x}{x_0}\right)^{-\alpha}, x \geq x_0 > 0. \quad (\text{Montes, 2012}).$$

La distribución Pareto tipo I se define como una distribución de probabilidad decreciente y convexa, cuya característica fundamental es que está basada en una Ley de Potencia que implica una forma hiperbólica. La distribución es biparamétrica, tomando a  $\alpha > 0$  como el parámetro de forma, y  $\theta > 0$  como el parámetro de escala (espacio paramétrico). Su modelo probabilístico es:

$$X_i \sim \text{Pareto}(\alpha, \theta) \quad (11)$$

con función de distribución acumulada:

$$F(x|\alpha, \theta) = P(X \leq x) = \int_{\theta}^x \frac{\alpha \theta^{\alpha}}{x^{(\alpha+1)}} dx = 1 - \left(\frac{\theta}{x}\right)^{\alpha} \quad (12)$$

y función de densidad de probabilidad:

$$f(x|\alpha, \theta) = \frac{\alpha \theta^{\alpha}}{x^{(\alpha+1)}} I_{(\theta, \infty)}(x) \quad (13)$$

El parámetro  $\theta$  alude al ingreso mínimo de la población, mientras que  $\alpha$  considera la dispersión, que a mayores valores de  $\alpha$  derivan densidades más concentradas en las proximidades del mínimo ( $\theta$ ). A este último parámetro también se le conoce como el índice de la cola o índice de Pareto (Arnold, 1985, 2015).

Dicha distribución permite modelar valores extremos de cola desconocidos con información conocida (Abdel-Ghaly, Attia, y Aly, 1998), y pertenece a la clase de distribuciones subexponenciales, de “cola derecha pesada o larga”, cuando su parámetro de forma es menor que 1; lo que significa que al dominio de los valores extremos se le asigna un espacio de probabilidad amplio en comparación con el que se le asignaría a una distribución Exponencial (Abdel-Ghaly *et al.*, 1998).

### 5.3. Momentos y medidas de posición y dispersión

Los momentos respecto al origen de la distribución Pareto son:

$$E(X^k) = \frac{\alpha\theta^k}{\alpha - k} \quad (14)$$

donde el k-ésimo momento existe solo si  $k < \alpha$ .

El valor esperado o media (primer momento respecto al origen) es:

$$E(X) = \frac{\alpha}{\alpha - 1}\theta, \quad \alpha > 1 \quad (15)$$

y la varianza (segundo momento respecto a la media) es:

$$Var(X) = \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)}\theta^2, \quad \alpha > 2 \quad (16)$$

y entonces el coeficiente de variación está dado por:

$$CV = \frac{1}{\sqrt{\alpha(\alpha - 2)}}, \quad \alpha > 2 \quad (17)$$

## 5.4. Aplicaciones de la distribución Pareto

La distribución Pareto ha sido muy importante debido a que una gran variedad de datos observados siguen una probabilidad de cola larga o ley de potencia en la cola, es decir, que la distribución de los datos es sesgada hacia los valores extremos, de forma que representa eventos en los que la variable aleatoria alcanza valores altos con poca frecuencia, mientras que los valores medianos o bajos son mucho más comunes (Clauset, Shalizi, y Newman, 2009; Dragulescu y Yakovenko, 2001).

Dicha distribución, es una herramienta utilizada para modelar fenómenos en distintas áreas de estudio como como la física, la hidrología, la geología, la astronomía la biología, las finanzas y la economía (Alzaatreh, Famoye, y Lee, 2012; Forbes, Evans, Hastings, y Peacock, 2011; Zaninetti y Ferraro, 2008). Para añadir flexibilidad a la distribución Pareto, se han derivado varias generalizaciones como la Distribución Generalizada de Pareto (Pickands, 1975) y la distribución Beta-Pareto (Akinsete, Famoye, y Lee, 2008), entre otras no menos importantes (Mahmoudi, 2011).

Debido a lo anterior, la distribución Pareto permite ajustar los datos observados en el modelado de terremotos y áreas de deslizamiento (Pelletier, Malamud, Blodgett, y Turcotte, 1997), fragmentación de las rocas y áreas de incendios forestales (Malamud, Morein, y Turcotte, 1998), volúmenes de hidrocarburos (Barton y Scholz, 1995), valuación de diamantes (Beirlant, de Wet, y Goegebeur, 2006); también se ha utilizado para estudiar la resistencia de las fibras de alfombras de nylon (Grimshaw, 1993). Harris (1968) utilizó la distribución para modelar tiempos de servicio de mantenimiento, Pickands (1975) para estudiar inundaciones extremas, Dargahi-Noubary (1989) la utilizó con datos de la velocidad del viento y los niveles del río Feather (Carolina del Norte, EE.UU.), R. L. Smith (1989), para estudiar los niveles de ozono en la atmosfera. En el ámbito de la salud, se realizaron con ella estudios de tiempos de supervivencia de pacientes dializados (Zeinab, 2007). En lo que respecta a negocios, se ha utilizado para determinar índices de capacidad de procesos (Valdiviezo y Fermín, 2010), para modelar tiempos de vida de los negocios (Nigm y Hamdy, 1987), y tiempos de servicio y sistemas de espera (Harris, 1968).

En el contexto socioeconómico, la distribución Pareto ha permitido realizar un gran número de investigaciones en cuanto al análisis del ingreso y la distribución de la riqueza, como en los estudios de Feenberg y Poterba (1993), y Piketty y Saez (2003), quienes analizaron datos de impuestos de los contribuyentes con ingresos altos en los EE.UU. Así mismo la han utilizado otros investigadores en el análisis de la evolución de la distribución del ingreso en Estados Unidos, entre ellos Bound y Johnson (1992); Castañeda, Díaz, y Ríos (2003); Katz y Murphy (1992); Levy y Michel (1991); Murphy y Welch (1992).

Por su parte Ahmed y Bhattacharya (1972), utilizaron la distribución Pareto para analizar la renta per cápita en la India, Hayakawa (1951) y Champernowne (1952), para estudiar la distribución del ingreso con datos de los japoneses. Similarmente Fujiwara, Souma, Aoyama, Kaizoji, y Aoki (2003), emplearon dicha distribución para inferir sobre las tasas de crecimiento de los ingresos per cápita en Japón, Dragulescu y Yakovenko (2001), la utilizaron para realizar una comparación con respecto al comportamiento de las distribuciones de los ingresos personales tanto en Reino Unido como en los Estados Unidos, Clementi y Gallegati (2005), analizaron con ella la distribución de los datos de los ingresos personales de los italianos, Tartalova (2012), realizó con ella estudios sobre la distribución del ingreso en Eslovaquia.

## 6. Materiales y métodos

### 6.1. Estimadores de Máxima Verosimilitud de la distribución Pareto

La inferencia estadística a partir de datos cuantitativos supone en general el conocimiento de la función de probabilidad (o de distribución) que mejor modela el comportamiento de la variable observada. Lo anterior se concreta con la determinación de los parámetros poblacionales de dicha distribución mediante métodos de estimación a partir de variables aleatorias muestrales (Unalmed, s.f.).

Considerando una distribución Pareto con parámetros  $\alpha$  y  $\theta$ , dada una muestra aleatoria  $x_i = (x_1, x_2, \dots, x_n)$ , la función de verosimilitud  $L$  toma la siguiente forma (Rytgaard, 1990):

$$L(\theta, \alpha | x) = \prod_{i=1}^n \frac{\alpha \theta^\alpha}{x_i^{\alpha+1}}; 0 < \theta \leq \min\{x_i\}, \alpha > 0 \quad (18)$$

Podemos observar entonces que  $L(\theta, \alpha | x)$  se incrementa mayormente que  $\theta$ , considerando al valor más alto de  $\theta$ . Dado lo anterior el valor de  $\theta$  puede ser mayor que el menor valor de  $x_i$ , así en la maximización de  $L$  ajustando  $\theta$  es:

$$\hat{\theta} = \min \{x_i\} \quad (19)$$

Aplicando logaritmo a la función de verosimilitud  $L = (\theta, \alpha | x)$  para encontrar el valor de  $\alpha$ , tenemos:

$$\log L(\theta, \alpha | x) = \sum_{i=1}^n \log \left( \frac{\alpha \theta^\alpha}{x_i^{\alpha+1}} \right)$$

$$= n \log(\alpha) + \alpha n \log(\theta) - (\alpha + 1) \sum_{i=1}^n \log(x_i)$$

$$\frac{\delta \log L}{\delta \alpha} = \frac{n + \alpha n \log(\theta) - \alpha \sum_{i=1}^n \log(x_i)}{\alpha}$$

entonces:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \log\left(\frac{x_i}{\hat{\theta}}\right)} \quad (20)$$

## 6.2. Simulación Monte Carlo vía Cadenas de Markov para determinar los parámetros de la distribución Pareto

La información referente al ingreso trimestral bruto de los mexicanos es procesada bajo el procedimiento “proc MCMC” incluido en el sistema Statistical Analysis Software (SAS/STAT Software) mejor conocido comercialmente como SAS en su versión 9.4. Este procedimiento pertenece a la categoría de Inferencia Bayesiana de SAS 9.4, definiéndolo como un procedimiento de acceso a la modelización e inferencia Bayesiana mediante priores arbitrarios y funciones de verosimilitud (Sas, 2013).

En la siguiente figura se integra el procedimiento MCMC a la presente investigación de la siguiente manera:

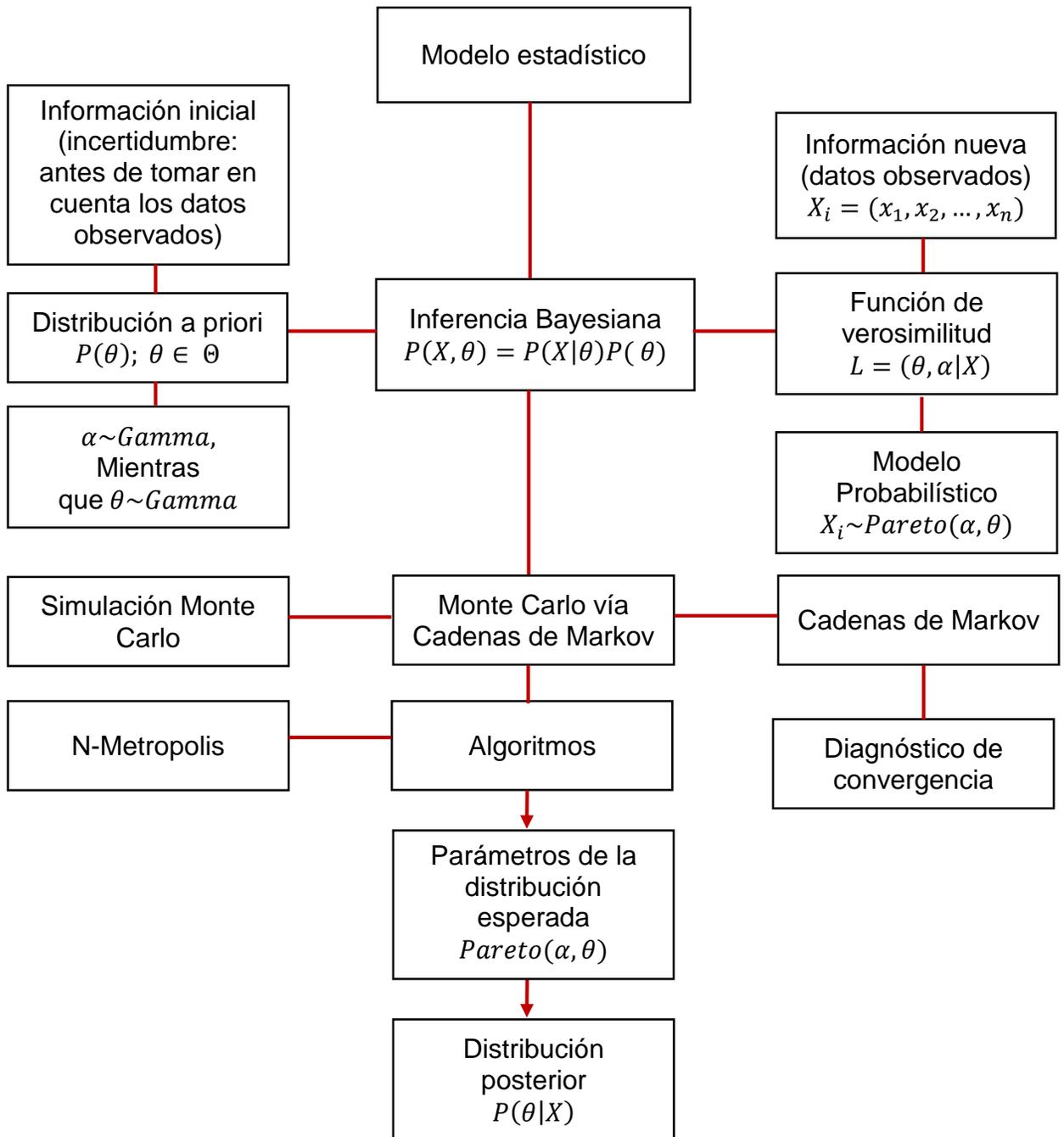


Figura 1. Procedimiento MCMC. Elaboración propia.

## **6.2.1. Diagnósticos de convergencia**

### **6.2.1.1. Intervalo de credibilidad**

La literatura de la estadística Bayesiana refiere a los intervalos de credibilidad como un análogo a los intervalos de confianza de la estadística clásica. Los intervalos Bayesianos corresponden al dominio de la distribución de probabilidad posterior. La elección del intervalo más estrecho en una distribución unimodal implicará la elección de los valores de la densidad de probabilidad más alta (M. H. Chen y Shao, 1999; Eberly y Casella, 2003; Edwards, Lindman, y Savage, 1963; Hartigan, 1966; Wei y Tanner, 1990).

### **6.2.1.2. Errores Estándar de Monte Carlo**

MCSE (por sus siglas en inglés), corresponde al error estándar de la estimación de la media posterior y mide la precisión de la simulación. La desviación estándar posterior describe la incertidumbre en el parámetro y es una función del tamaño de la muestra en el conjunto de datos, mientras que MCSE describe la incertidumbre en la estimación del parámetro como un resultado de la simulación MCMC y es una función del número de iteraciones en la simulación (Sas, 2008).

### **6.2.1.3. Diagnóstico de Geweke**

El diagnóstico de Geweke (1991), hace la comparación de la ubicación del parámetro muestreado en dos espacios diferentes de la cadena, esperando que los valores medios de tales intervalos de tiempo sean iguales o tan cercanos unos de otros, de ser así, entonces se supone que las dos partes de la cadena tienen posiciones similares en el espacio de estados y que las muestras provienen de la misma distribución.

Generalmente, se compara la última mitad de la cadena que se supone ha convergido con algún intervalo más pequeño en el comienzo de la cadena. Dicho diagnóstico utiliza la estimación de densidad espectral para este análisis la cual se utiliza para la detección de comportamientos periódicos en los datos (Cowles y Carlin, 1996; Sahlin, 2011). El diagnóstico de convergencia  $Z$ , es la diferencia entre las dos medias comparadas; divididas por el error estándar asintótico de dicha diferencia. Cuando el largo de la cadena tiende a infinito, la distribución de muestreo de  $Z$  se distribuye  $N(0,1)$ , por lo tanto los valores de  $Z$  que se ubiquen en el extremo de la distribución normal sugiere que la cadena no converge en forma temprana (Alcaíno, 2004).

#### 6.2.1.4. Diagnóstico de Raftery-Lewis

El método de convergencia introducido por Raftery y Lewis (1992, 1996), 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 obtienen  $M$ ,  $N_m$  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  $I = \frac{M+N}{N_m}$ . Valores de  $I$  cuanto más cerca son a 1 menos correlacionadas son las muestras mientras que mayores a 5 indican una fuerte correlación, la cual puede ser resultado de una pobre elección del valor de inicio de la cadena o una alta correlación entre los componentes del vector  $\theta$ , relativa a su distribución final.

El tamaño de muestra estimado 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 si misma una cadena de Markov. Sin embargo,  $Z_t$  puede comportarse como una cadena de Markov si 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 equilibrio de la cadena  $Z_t^k$  (Cowles y Carlin, 1996).

#### **6.2.1.5. Diagnóstico de Heidelberger-Welch**

El diagnóstico de Heidelberger y Welch (1981, 1983), se basa en la suposición de que se tiene un proceso débilmente estacionario cuando la cadena ha alcanzado la convergencia. Un proceso débilmente estacionario tiene las propiedades que si  $X^j$  es definido como la  $j$  – ésima iteración en la secuencia, la función de media  $E[X^j]$  es constante en el tiempo y la covarianza  $Cov(\theta^j, \theta^{j+s})$  no depende de  $j$  sino solo del tamaño de  $s$ . Esto es una suposición razonable ya que la secuencia es generada por una cadena de Markov y, por tanto satisface completa estacionariedad (Sahlin, 2011).

Dicho diagnostico se basa en el estadístico de Cramér-von Mises para evaluar la evidencia de no estacionariedad. Cuando esto último es el resultado de dicha prueba, se desecha el primer 10% de las iteraciones volviéndose a repetir la prueba hasta que la cadena ha pasado la prueba o se han descartado más del 50% de las iteraciones; si se produce esta última situación, basta para concluir que no era suficientemente larga la porción estacionaria de la cadena para evaluar adecuadamente la convergencia. Cada parámetro refleja su estatus si paso o no el estadístico de Cramér-von Mises. Si el parámetro no pasa dicha prueba, es un indicativo de que la cadena necesita una trayectoria de más tiempo antes de alcanzar la convergencia. Posteriormente, el proceso considera una segunda etapa la cual examina las iteraciones que pasaron la prueba de estacionariedad para garantizar la exactitud. Especialmente, si la mitad de la anchura del intervalo de credibilidad de la estimación es menor que una fracción predeterminada de la media, entonces la prueba implica

que la media fue estimada con suficiente precisión. Si un parámetro falla en esta etapa del diagnóstico (indicando poca exactitud en las estimaciones) puede ser necesario más tiempo en la ejecución del muestreador MCMC (Cowles y Carlin, 1996; Kaplan, 2014; Kaplan y Depaoli, 2013).

#### 6.2.1.6. Tamaño de muestras efectivo

ESS (por sus siglas en ingles), realiza el cálculo de los tamaños de muestra efectivos (Kass, Carlin, Gelman, y Neal, 1998), de las muestras posteriores de cada parámetro, adicionalmente calcula el tiempo de correlación y la eficiencia de la cadena para cada parámetro.

ESS se define estadísticamente como:

$$ESS = \frac{n}{\tau} = \frac{n}{1 + 2 \sum_{k=1}^{\infty} \rho_k(\theta)}$$

Donde  $n$  es el tamaño total de la muestra y  $\rho_k$  es la autocorrelación del lag  $k$  para  $\theta$ . La cantidad  $\tau$  es referido al tiempo de autocorrelación. Para estimar este valor, en los procedimientos Bayesianos primero se posiciona en un punto  $k$  después de que las autocorrelaciones son muy cercanas a cero, sumando los valores  $\rho_k$  hasta ese punto. El punto  $k$  es tal que  $\rho_k < 0.05$  o,  $\rho_k < 2sk$  donde  $sk$  es la desviación estándar estimada:

$$sk = 2 \sqrt{\left( \frac{1}{n} \left( 1 + 2 \sum_{j=1}^{k-1} \rho_j^2(\theta) \right) \right)} \quad (\text{Sas, 2008}).$$

## **6.3. Indicadores comunes para medir la desigualdad del ingreso**

### **6.3.1. Curva de Lorenz**

La curva de Lorenz es una herramienta que se utiliza para representar la distribución de ingresos según lo propuesto por Lorenz (1905), que muestra que la proporción del ingreso total está en manos de un determinado porcentaje de la población. Este método es conceptualmente muy similar al método por cuantiles. La curva de Lorenz grafica la fracción acumulada de una variable aleatoria (ingreso) versus la fracción acumulada de población receptora de la variable en estudio (Bellù y Liberati, 2005, 2006).

### **6.3.2. Coeficiente de Gini**

El coeficiente o índice de Gini es una medida de dispersión estadística, para medir cualquier forma de distribución desigual, aunque generalmente se ha utilizado para representar la distribución de los ingresos de los residentes de un país, siendo la medida más utilizada de desigualdad. Dicho coeficiente fue desarrollado por el italiano estadístico y sociólogo Corrado Gini, quien en el año de 1912 publicó su aportación en el documento “*Variabilità e mutabilità*”.

El índice de Gini es el indicador expresado en forma de porcentaje, y es igual al coeficiente de Gini multiplicado por cien.

El índice de Gini mide hasta qué punto la distribución del ingreso entre individuos u hogares dentro de una economía se aleja de una distribución perfectamente equitativa. Un coeficiente de Gini cero expresa perfecta igualdad, donde todos los valores son los mismos, es decir, todos tiene los mismo ingresos. Un coeficiente de Gini de uno o cien por ciento, expresa la desigualdad máxima entre los valores donde solo una persona tiene todos los ingresos o consumo (Bellù y Liberati, 2006).

Partiendo de una curva de Lorenz, el coeficiente de Gini mide la proporción del área entre la diagonal que representa la equidistribución y la curva de Lorenz (A), como proporción del área bajo la diagonal (A+B), es decir

$$Gini = \frac{A}{A + B}$$

Así, entre más cerca de la línea diagonal (distribución del ingreso igualitaria) se encuentre la curva de Lorenz, más cercano a cero estará el coeficiente de Gini.

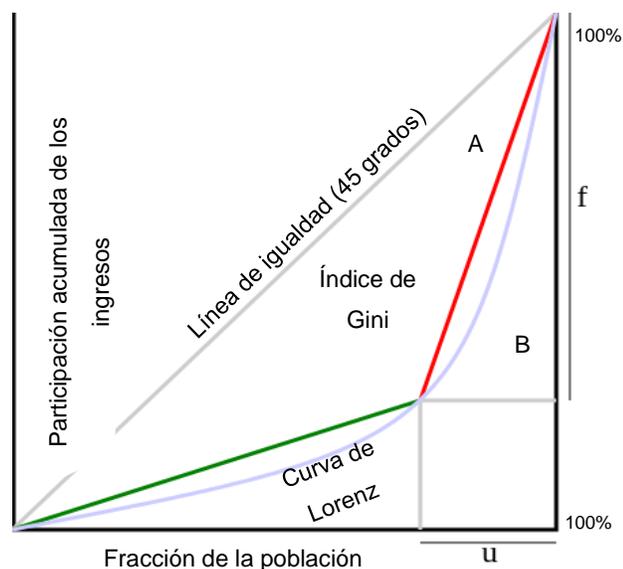


Figura 2. Curva de Lorenz y coeficiente de Gini.

Adaptado de: [https://wikipedia.org/wiki/Gini\\_coefficient#cite\\_ref-2](https://wikipedia.org/wiki/Gini_coefficient#cite_ref-2)

El indicador  $u$  representa el porcentaje de la población con la mayor parte del ingreso porcentual ( $f$ , línea roja), la línea verde indica la riqueza que se comparte con el resto de la población, esto es:  $G = f - u$ . Una distribución uniforme (línea gris) con la misma proporción  $u$  y  $f$  determinará  $G > f - u$ .

## 6.4. Encuesta Nacional de Ingresos y Gastos de los Hogares

La Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), es un instrumento de datos obtenidos a partir de una encuesta, la cual proporciona un panorama estadístico del comportamiento del ingreso y gasto de los hogares en cuanto a su monto, procedencia y distribución, y provee información sobre las características ocupacionales y sociodemográficas de los integrantes del hogar, así como las características de infraestructura de la vivienda y equipamiento del hogar.

Dicho material de análisis apoya en la toma de decisiones en materia de política económica y social, así como para investigadores y organismos internacionales.

Los antecedentes de la ENIGH data desde 1984, cuando en este se tuvo la iniciativa por ampliar los objetivos y homogenizar la metodología, tomando en cuenta las recomendaciones internacionales y los requerimientos de información de los usuarios (Montes, 2012).

La encuesta es levantada formalmente por el Instituto Nacional de Estadística y Geografía (INEGI), con una periodicidad bienal. Tiene una cobertura nacional, con un desglose geográfico que considera áreas urbanas consideradas como localidades de 2,500 y más habitantes, y áreas rurales definidas como localidades menores de 2,500 habitantes (CESOP, 2013).

El diseño muestral de la encuesta es:

- ♣ Probabilístico. Porque todas las unidades de muestreo tienen una probabilidad conocida y distinta de cero.
- ♣ Estratificado. Porque las unidades de muestreo con características socioeconómicas y geográficas se agrupan para formar estratos.
- ♣ Polietápico. Porque la vivienda, que es la última unidad de selección, es seleccionada después de varias etapas.
- ♣ Por conglomerados. Porque previamente se conforman conjuntos de unidades muestrales, de los cuales se obtiene la muestra.

El tamaño de la muestra es calculado con la proporción de preceptores de ingresos por cada vivienda.

En los estratos urbano y rural se fija una fracción de muestra, con las que se construyen las probabilidades de selección de la primera etapa, obteniendo las unidades primarias de muestreo (UPM). Una vez realizado esto, se procede a conformar las probabilidades de selección de la segunda etapa obteniendo las viviendas objeto de entrevista de cada encuesta, en esta selección se incluyen viviendas deshabitadas y de uso temporal que garantiza que los resultados obtenidos a partir de solo una parte de la población (muestra) se pueda generalizar hacia el total.

Dentro de la encuesta está incluido un factor de expansión. El factor de expansión implica que a pesar de que los resultados de la encuesta presenten una muestra reducida, existe un determinado número de individuos (el número de individuos es el equivalente al factor de expansión) que comparten las mismas características.

Las tablas que conforman la base de datos contienen información asociada a tres niveles o grupos, uno asociado a la vivienda como unidad de muestreo, otro asociado al hogar como unidad de observación y el último al integrante del hogar como unidad de análisis.

La ENIGH cuenta con 15 bases de datos (15 años) referentes al ingreso per cápita de los mexicanos de los años 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012 y 2014. De lo anterior, se utilizó toda la información referente a los ingresos trimestrales brutos, catalogados en la encuesta como Microdatos “Ingresos y percepciones financieras y de capital de cada uno de los integrantes del hogar” a través del folio de la vivienda. Este último, consta de seis dígitos donde el primero y segundo indican la clave de la entidad federativa, el tercero corresponde a la decena en la que se realiza el levantamiento, el cuarto, quinto y sexto dígito corresponden al número consecutivo (ENIGH, 2014).

## 7. Resultados

Los resultados de la presente investigación se muestran a partir del procedimiento Bayesiano de la simulación Monte Carlo vía Cadenas de Markov (MCMC) de Sas (2013). Dicho procedimiento se aplicó para cada base de datos de los 15 años. Con esto, se fundamentan los objetivos planteados en la presente investigación.

### 7.1. Validación del modelo

En la siguiente sección, se explica que el proceso MCMC comienza dando lectura al número de datos correspondientes al ingreso per cápita trimestral bruto de los mexicanos para comenzar a procesar la información y dar paso a la salida que corresponde a la estimación de las medias. De esta manera, primeramente se realiza la estimación de los parámetros, en conjunto con el diagnóstico de convergencia del procedimiento MCMC. Al respecto, se llevan a cabo los diagnósticos de Geweke, Raftery-Lewis y Heidelberger-Welch. A continuación se determina la validez de la simulación de la cadena de Markov, para lo cual es importante determinar el tamaño efectivo de la muestra y el nivel de autocorrelación.

Como parte del análisis de los resultados, se incluye el comportamiento de la desigualdad en México medido a través del coeficiente de Gini y la curva de Lorenz.

### 7.2. Estimación de los parámetros $\theta$ y $\alpha$

La distribución a priori utilizada para la estimación posterior de los parámetros corresponde a una distribución gamma, la cual utiliza parámetros mínimos

(shape=0.001; scale=0.001) para expresar ignorancia a priori, no alterando así el resultado de la inferencia (Fink, 1997; Howlader, Hossain, y Makhmin, 2007).

Inicialmente, en el Cuadro 1 se muestran el primer y segundo bloque, los cuales corresponde a la estimación de los parámetros  $\theta$  y  $\alpha$  por separado, con el algoritmo de muestreo Normal-Metropolis (Metropolis basado en una caminata aleatoria con supuesto de normalidad), el cual se ha utilizado para extraer valores al azar de la distribución posterior obteniendo las medias de estos valores, simulando así las medias de la distribución esperada, siendo por tanto la estimación de los parámetros.

El problema de los valores iniciales consiste en que aquel valor por el que se inicia cada cadena de Markov puede influir en los siguientes valores y, por tanto, en los estimadores. Sin embargo, estos valores iniciales son relativamente arbitrarios (Revuelta, 2001). Para reducir su efecto se han elaborado cadenas de entre 50000 y 110000 elementos cada una y se han descartado (burn-in) los primeros 5000 a 10000 elementos en las cadenas (Cuadro 5).

En el Cuadro 1 se muestran los sumarios e intervalos posteriores:

Cuadro 1. Sumarios e intervalos posteriores

Año	No. De observaciones	Parámetro	Método de muestreo N-Metropolis							
			Sumarios posteriores						Intervalos posteriores	
			Valor inicial	Media	Desviación estándar	Percentiles 25, 50 y 75		Intervalo HPD <sub>95%</sub>		
2014	89131	theta	4.032E-7	0.1187	0.000395	0.1184	0.1187	0.119	0.1179	0.1195
		alpha	1.254E-7	0.49	0.000048	0.4899	0.49	0.49	0.4899	0.49
2012	46597	theta	2.07E-11	0.1204	0.000557	0.12	0.1204	0.1207	0.1193	0.1214
		alpha	4.935E-6	0.4799	0.000095	0.4799	0.4799	0.48	0.4797	0.48
2010	126953	theta	0.000015	0.1185	0.000334	0.1183	0.1185	0.1187	0.1179	0.1192
		alpha	1.598E-7	0.5	0.000034	0.5	0.5	0.5	0.4999	0.5
2008	133215	theta	2.07E-11	0.1101	0.000297	0.1099	0.1101	0.1103	0.1095	0.1107
		alpha	4.935E-6	0.25	0.000017	0.25	0.25	0.25	0.25	0.25
2006	75812	theta	2.07E-11	0.1169	0.000421	0.1166	0.1169	0.1171	0.1161	0.1177
		alpha	4.935E-6	0.4999	0.000059	0.4999	0.5	0.5	0.4998	0.5
2005	75598	theta	2.07E-11	0.1331	0.000474	0.1327	0.1331	0.1334	0.1321	0.134
		alpha	4.935E-6	1.4998	0.000179	1.4998	1.4999	1.4999	1.4995	1.5
2004	74290	theta	2.07E-11	0.1462	0.000531	0.1459	0.1462	0.1466	0.1452	0.1473
		alpha	4.935E-6	2.9996	0.000383	2.9995	2.9997	2.9999	2.9988	3
2002	55911	theta	2.07E-11	0.1477	0.000617	0.1473	0.1477	0.1481	0.1465	0.1481
		alpha	4.935E-6	2.4996	0.000436	2.4994	2.4997	2.4999	2.4997	2.499
2000	31629	theta	2.07E-11	0.1490	0.000835	0.1485	0.1490	0.1496	0.1474	0.1506
		alpha	4.935E-6	2.5788	0.00121	2.5783	2.5792	2.5796	2.5764	2.58
1998	33154	theta	4.032E-7	0.1171	0.000648	0.1167	0.1171	0.1176	0.1159	0.1184
		alpha	1.254E-7	0.2999	0.000084	0.2999	0.2999	0.3	0.2997	0.3
1996	38671	theta	4.032E-7	0.1278	0.000655	0.1274	0.1278	0.1283	0.1265	0.1291
		alpha	1.254E-7	0.4999	0.000113	0.4998	0.4999	0.5	0.4997	0.5
1994	34374	theta	4.032E-7	0.1337	0.000726	0.1332	0.1337	0.1342	0.1323	0.1351
		alpha	1.254E-7	0.4999	0.000122	0.4998	0.4999	0.5	0.4996	0.5
1992	36698	theta	4.032E-7	0.0792	0.000412	0.0789	0.0792	0.0795	0.0784	0.08
		alpha	1.254E-7	1.9978	0.0022	1.997	1.9985	1.9994	1.9935	2
1989	27789	theta	2.07E-11	0.0807	0.00050	0.0803	0.0807	0.081	0.0797	0.0817
		alpha	4.935E-6	2.2408	0.0498	2.2075	2.2418	2.2754	2.1436	2.3378
1984	11396	theta	2.07E-11	0.0981	0.000954	0.0975	0.0981	0.0988	0.0962	0.1
		alpha	4.935E-6	1.1192	0.0352	1.0953	1.1185	1.1433	1.0548	1.1914

Fuente: Elaboración propia con datos de la simulación.

Donde las medias estimadas de las simulaciones corresponden al parámetro del ingreso mínimo  $\theta$ , y al parámetro de desigualdad  $\alpha$ , resultando de igual manera las respectivas desviaciones estándar, que al observar detalladamente cada una representa la mínima variación de los datos individuales de la muestra con respecto a la media, permitiendo de esta manera confiar en la representatividad del parámetro; así mismo el percentil 50 indica que el 50% de los datos de la muestra es representado por el valor del parámetro estimado, sin importar valores muy elevados o valores demasiado mínimos.

El siguiente indicador en el Cuadro 1 es el intervalo Bayesiano o de credibilidad de densidad posterior máxima al 95% (Highest Posterior Density at 95%), que asegura que es  $((1 - \alpha)100)$  95% probable que el valor verdadero del parámetro se localice en ese rango estimado, calculado a partir de la distribución posterior, dados los datos observados y la información a priori del mismo.

### **7.3. Errores estándar de la simulación y diagnósticos de Geweke y Raftery-Lewis**

Los errores estándar de la simulación Monte Carlo (MCSE, por sus siglas en inglés), proporcionan información útil sobre la calidad de la estimación y la inferencia posterior. El error estándar es una medida de la variabilidad entre las muestras de una población que representa la cantidad típica de error que se puede esperar de un estimador. Por lo tanto, cuanto menos sea el error estándar, mejor será la estimación del parámetro (Bradley, Christophe, y Denis, 2014). Los errores estándar se pueden utilizar para establecer un intervalo de probabilidad que contenga la verdadera característica de la población (Altman y Bland, 2005; Norman, Streiner, y Tarrés, 1996). Tales intervalos se denominan intervalos de confianza (Norman *et al.*, 1996).

Además de lo anterior, es posible calcular la relación que se da entre el error estándar de la simulación y la desviación estándar de cada una de las medias posteriores,

mostrando valores mínimos en cada dato estimado e indicando mínimas variaciones; fundamentando así la confiabilidad sobre los parámetros obtenidos.

El Cuadro 2 muestra la información de los errores estándar de Monte Carlo, así como los diagnósticos de Geweke y Raftery-Lewis:

Cuadro 2. Errores estándar de Monte Carlo y diagnósticos de Geweke y Raftery-Lewis

Año	Parámetro	Media	Errores estándar de Monte Carlo			Diagnósticos de Geweke		Raftery-Lewis Cuantil q=0.025 Exactitud r=+/- 0.005 Probabilidad s=0.95 Epsilon=0.001			
			MCSE	Desv. Estándar	MCSE/SD	z	Pr >  z  p-valor	Número de muestras			Factor de dependencia
								Bur n-In	Total	Mínimo	
2014	theta	0.1187	3.90E-06	0.000395	0.00989	-0.8734	0.3825	2	3930	3746	1.0491
	alpha	0.49	5.74E-07	0.000048	0.0119	1.1512	0.2496	3	4447	3746	1.1871
2012	theta	0.1204	5.626E-6	0.000557	0.01	1.0237	0.3060	2	3772	3746	1.0069
	alpha	0.4799	1.123E-6	0.000095	0.0118	-0.9172	0.3590	3	4374	3746	1.1676
2010	theta	0.1185	4.87E-06	0.000334	0.0146	-0.2162	0.8288	2	3930	3746	1.0491
	alpha	0.5	5.66E-07	0.000034	0.0169	-1.0584	0.2899	3	4559	3746	1.217
2008	theta	0.1101	4.92E-06	0.000297	0.0166	-1.0316	0.3023	3	4199	3746	1.1209
	alpha	0.25	3.98E-07	0.000017	0.0239	-0.6202	0.5351	5	5873	3746	1.5678
2006	theta	0.1169	3.241E-6	0.000421	0.0077	0.3027	0.7621	3	4249	3746	1.1343
	alpha	0.4999	6.826E-7	0.000059	0.0115	-0.8259	0.4089	7	9716	3746	2.5937
2005	theta	0.1331	4.739E-6	0.000474	0.01	-0.9474	0.3434	2	3803	3746	1.0152
	alpha	1.4998	2.145E-6	0.000179	0.012	-1.6867	0.0917	3	4521	3746	1.2069
2004	theta	0.1462	4.061E-6	0.000531	0.00765	-0.7310	0.4648	3	4249	3746	1.1343
	alpha	2.9996	4.241E-6	0.000383	0.0111	-1.6060	0.1083	6	8993	3746	2.4007
2002	theta	0.1477	4.628E-6	0.000617	0.00750	0.6090	0.5425	3	4308	3746	1.1500
	alpha	2.4996	4.521E-6	0.000436	0.0104	-1.0192	0.3081	5	5987	3746	1.5982
2000	theta	0.1490	6.41E-06	0.000835	0.00769	0.319	0.7490	3	4338	3746	1.1580
	alpha	2.5788	0.000014	0.00121	0.0115	-1.0544	0.2917	7	9140	3746	2.4399
1998	theta	0.1171	6.48E-06	0.000648	0.01	0.084	0.9331	2	3866	3746	1.032
	alpha	0.2999	1.01E-06	0.000084	0.0121	0.2943	0.7685	4	4636	3746	1.2376
1996	theta	0.1278	6.55E-06	0.000655	0.01	1.7463	0.0808	2	3834	3746	1.0235
	alpha	0.4999	1.33E-06	0.000113	0.0119	0.9875	0.3234	4	4792	3746	1.2792
1994	theta	0.1337	7.26E-06	0.000726	0.01	0.9229	0.3561	2	3962	3746	1.0577
	alpha	0.4999	1.45E-06	0.000122	0.0119	0.0884	0.9296	4	4752	3746	1.2686
1992	theta	0.0792	4.18E-06	0.000412	0.0101	-0.4593	0.646	2	3772	3746	1.0069
	alpha	1.9978	0.000026	0.0022	0.0117	0.4381	0.6613	4	4752	3746	1.2686
1989	theta	0.0807	4.24E-06	0.00050	0.00835	-1.7127	0.0868	3	4511	3746	1.2042
	alpha	2.2408	0.000410	0.0498	0.00824	-0.8340	0.4043	3	4428	3746	1.1821
1984	theta	0.0981	0.000011	0.000954	0.0116	-0.2714	0.7861	3	4558	3746	1.2168
	alpha	1.1192	0.000418	0.0352	0.0119	-0.7824	0.4340	3	4374	3746	1.1676

Fuente: Elaboración propia con datos de la simulación.

En la presente investigación, el estadístico de Geweke señala que todos los valores de  $z$  están entre  $-1.7127$  y  $1.7463$ , ubicándose así dentro del intervalo  $\pm 1.96$  (Infante y Zàrate, 2012). Lo anterior determina que no existe diferencia en las medias muestrales de las primeras y últimas partes de las cadenas, aseverando de esta manera que las muestras se han extraído de una distribución posterior estacionaria, y el estadístico de Geweke presenta una distribución asintóticamente normal estándar  $N(0,1)$ .

En lo que refiere al diagnóstico de Raftery y Lewis para la investigación, este muestra el número de iteraciones burn-in al inicio de cada cadena de Markov, resultando valores desde 2 hasta 7 iteraciones descartadas, las cuales son relativamente mínimas si en el modelo se consideraron de 5000 a 10000 iteraciones para el periodo burn-in, y de 50000 a 110000 iteraciones totales de MCMC (Cuadro 5).

Así mismo, el procedimiento indica el tamaño total de la muestra además del valor mínimo necesitado para estimar el cuantil  $q$  de 0.025 que cumple  $P(X \leq x_q) = q$ , indicando que solo el 2.5% de los valores de la muestra es menor o igual que  $x_q$ . El factor de dependencia se mostró dentro de un rango 0.9824 y 2.5937, siendo menor que 5, lo que confirma independencia entre las observaciones de las cadenas simuladas.

#### **7.4. Diagnóstico de Heidelberger-Welch**

La prueba de Heidelberger y Welch (1981, 1983), evalúa desde dos perspectivas: una prueba de parte estacionaria y el cálculo de la mitad de la anchura del intervalo de confianza para la media (Cuadro 3).

Cuadro 3. Diagnóstico de Heidelberger-Welch

Diagnósticos de Heidelberger-Welch										
Año	Parámetro	Media	Test de estacionariedad				Test Half-Width			
			Estadístico de Cramér-Von Mises ( $W^2$ )	Valor p	Resultado del test	iteraciones descartadas	Half-Width	Media	Half-Width relativo	Resultado del test
2014	theta	0.1187	0.1476	0.3971	Superado	0	8.10E-06	0.1187	0.000068	Superado
	alpha	0.49	0.0581	0.8261	Superado	0	0	0.49	0	Superado
2012	theta	0.1204	0.1816	0.3062	Superado	0	0.000010	0.1204	0.000086	Superado
	alpha	0.4799	0.1168	0.5085	Superado	0	1.867E-6	0.4799	3.891E-6	Superado
2010	theta	0.1185	0.2296	0.2167	Superado	0	0.00001	0.1185	0.000087	Superado
	alpha	0.5	0.2926	0.1418	Superado	0	0	0.5	0	Superado
2008	theta	0.1101	0.0479	0.8890	Superado	0	0.000011	0.1101	0.000096	Superado
	alpha	0.25	-	-	Fallido	-	-	-	-	-
2006	theta	0.1169	0.2145	0.2411	Superado	0	6.503E-6	0.1169	0.000056	Superado
	alpha	0.4999	0.7910	0.7910	Superado	0	0	0.4999	0	Superado
2005	theta	0.1331	0.2803	0.1537	Superado	0	8.941E-6	0.1331	0.000067	Superado
	alpha	1.4998	0.2150	0.2401	Superado	0	4.288E-6	1.4998	2.859E-6	Superado
2004	theta	0.1462	0.3927	0.0755	Superado	0	7.429E-6	0.1462	0.000051	Superado
	alpha	2.9996	0.1710	0.3315	Superado	0	7.66E-6	2.9996	2.554E-6	Superado
2002	theta	0.1477	0.0561	0.8387	Superado	0	9.17E-6	0.1477	0.000062	Superado
	alpha	2.4996	0.2309	0.2148	Superado	0	9.105E-6	2.4996	3.643E-6	Superado
2000	theta	0.1490	0.1608	0.3584	Superado	0	0.000012	0.149	0.00008	Superado
	alpha	2.5788	0.0962	0.6036	Superado	0	0.000027	2.5788	0.00001	Superado
1998	theta	0.1171	0.0562	0.8379	Superado	0	0.000013	0.1171	0.00011	Superado
	alpha	0.2999	0.1837	0.3014	Superado	0	2.19E-06	0.2999	7.29E-06	Superado
1996	theta	0.1278	0.4412	0.0564	Superado	1000	0.000012	0.1278	0.000096	Superado
	alpha	0.4999	0.0804	0.6897	Superado	0	3.05E-06	0.4999	6.10E-06	Superado
1994	theta	0.1337	0.0587	0.822	Superado	0	0.000017	0.1337	0.000126	Superado
	alpha	0.4999	0.0603	0.8124	Superado	0	2.64E-06	0.4999	5.28E-06	Superado
1992	theta	0.0792	0.2802	0.1539	Superado	0	9.13E-06	0.0792	0.000115	Superado
	alpha	1.9978	0.0669	0.7704	Superado	0	0.000049	1.9978	0.000025	Superado
1989	theta	0.0807	0.1648	0.3475	Superado	0	7.903E-06	0.0807	0.000098	Superado
	alpha	2.2408	0.1581	0.3657	Superado	0	0.000701	2.2408	0.000313	Superado
1984	theta	0.0981	0.0526	0.8605	Superado	0	0.000021	0.0981	0.000215	Superado
	alpha	1.1192	0.0501	0.8798	Superado	0	0.000764	1.1192	0.000682	Superado

Fuente: Elaboración propia con datos de la simulación.

En la primer prueba del diagnóstico de Heidelberger-Welch basado en el estadístico de Cramér-von Mises, refiere a que los valores muestreados si provienen de un proceso estacionario, al apuntar las pruebas como “superado” en cada parámetro de cada año, a excepción del parámetro  $\alpha$  de 2008, que de forma contraria, dicho test se reportó como “fallido”. Los parámetros que superaron la prueba se fundamentan en los resultados del p-valor los cuales no superan al valor por defecto de  $1 - \alpha = 0.95$  (Cuadro 3). En cuanto al parámetro que no superó la prueba de Cramér-von Mises, según el sistema Sas, no pudo ser calculado porque el error estándar de la autocorrelación ajustada de las muestras posteriores es cero, lo que podría ser causado porque la cadena de Markov es un vector constante; pero al observar las figuras que acompañan las pruebas numéricas (Figura 5), el mezclado de las iteraciones, la gráfica de autocorrelación y la distribución marginal posterior del parámetro, no muestran ninguna anomalía que ponga en duda total la significancia estadística de  $\alpha$  de 2008.

Como una segunda parte del mismo diagnóstico, el estadístico half-width o de medio ancho de banda, utiliza la porción de la cadena que ha superado la prueba de estacionariedad para estimar el error estándar asintótico de la media a través del método de series de tiempo.

Una vez hecho lo anterior el procedimiento obtiene la media muestral de las iteraciones retenidas además del valor half-width, que corresponde a la mitad de la anchura de los intervalos de confianza al 95% para esta media. Si el valor de dicha prueba es menor que  $\varepsilon = 0.1$  (valor de tolerancia) veces la media muestral de las iteraciones retenidas, la prueba se supone superada y la muestra utilizada se considera para calcular la media posterior con una precisión aceptable.

Los valores del half-width relativo han resultado menores al valor de tolerancia  $\varepsilon = 0.1$  concluyendo que los datos de la muestra han sido suficientes para estimar con precisión la media posterior con 95% de confianza bajo la tolerancia de  $\varepsilon$ .

## **7.5. Validez de la simulación Monte Carlo vía Cadenas de Markov**

El Cuadro 4 contiene la información del tamaño efectivo de la muestra (Effective Sample Size), que se relaciona con la función de autocorrelación y los errores estándar de la simulación Monte Carlo.

El tamaño efectivo de la muestra (ESS) y el tiempo de autocorrelación ( $\tau$ ) son inversamente proporcionales; un valor bajo de ESS y alto de  $\tau$ , indican una mala simulación de la cadena de Markov (Sas, 2008).

Cuadro 4. Tamaño efectivo de la muestra y estadísticos muestrales del ingreso

Año	Parm	Media	Tamaño efectivo de la muestra			Estadísticos muestrales del ingreso						Coef. de Gini
			ESS	Tiempo de autocorrr	Eficiencia	Media	Mediana	Moda	Dev std	Mín	Máx	
2014	theta	0.1187	10219.9	0.9785	1.022	7086.9	1956.52	1467.39	22285.37	0.49	4067950.81	0.698
	alpha	0.49	7072.8	1.4139	0.7073							
2012	theta	0.1204	9799.2	1.0205	0.9799	6107.38	1711.95	1467.39	14150.01	0.48	782608.69	0.701
	alpha	0.4799	7151.2	1.3984	0.7151							
2010	theta	0.1185	4693.3	1.0654	0.9387	6778.28	2256.12	1512.3	14905.75	0.5	1504080	0.678
	alpha	0.5	3509.2	1.4248	0.7018							
2008	theta	0.1101	3637.3	1.0997	0.9093	7204.66	2135.37	1512.3	23248.15	0.25	4250521.8	0.699
	alpha	0.25	1749.5	2.2864	0.4374							
2006	theta	0.1169	16866.9	1.1858	0.8433	7489.15	3008.16	544.43	15845.64	0.5	1203264	0.666
	alpha	0.4999	7557.5	2.6464	0.3779							
2005	theta	0.1331	10000	1	1	7837.6	3402.68	6049.2	25369.5	1.5	3998347	0.663
	alpha	1.4998	6964.7	1.4358	0.6965							
2004	theta	0.1462	17097.1	1.1698	0.8549	7863.77	3600	6000	21751.91	3	2400000	0.659
	alpha	2.9996	8172.3	2.4473	0.4086							
2002	theta	0.1477	17781.1	1.2373	0.8082	5837.37	2550	435	11340.69	2.5	1000000	0.646
	alpha	2.4996	9292.0	2.3676	0.4224							
2000	theta	0.1490	16929.5	1.1814	0.8465	5541.8	2550.24	6000	12722.81	2.58	715000	0.651
	alpha	2.5788	7604.2	2.6301	0.3802							
1998	theta	0.1171	10000	1	1	4040.38	1840	3600	10374.02	0.3	955000	0.651
	alpha	0.2999	6841.2	1.4617	0.6841							
1996	theta	0.1278	10000	1	1	2951.08	1500	1800	10036.74	0.5	1380000	0.625
	alpha	0.4999	7110.7	1.4063	0.7111							
1994	theta	0.1337	10000	1	1	2137.18	1100	1800	5036.86	0.5	440700	0.628
	alpha	0.4999	7093.4	1.4098	0.7093							
1992	theta	0.0792	9739.5	1.0267	0.974	1679425	730000	1800000	4732118.94	2	236500000	0.665
	alpha	1.9978	7366.1	1.3576	0.7366							
1989	theta	0.0807	14332.5	1.3954	0.7166	1165818.96	708000	600000	4019236.34	126	535139313	0.589
	alpha	2.2408	14735.8	1.3572	0.7368							
1984	theta	0.0981	9138.7	1.0942	0.9139	60038.55	36000	60000	90411.36	18	2955000	0.569
	alpha	1.1181	9367.6	1.0675	0.9368							

Fuente: Elaboración propia con datos de la simulación.

Los datos del ESS reportados representan el número de las iteraciones independientes que se han utilizado para representar a la cadena simulada, y que de esta manera ha servido satisfactoriamente para calcular las medias posteriores. Así mismo, el tiempo de autocorrelación, el cual se define como el número muestras correlacionadas, se ha mostrado en un rango de entre 0.9785 y 2.6464, indicando una acelerada disminución de las autocorrelaciones en el tiempo (Cuadro 4).

El procedimiento MCMC conjuntamente a la información anterior, muestra un diagnóstico visual como el siguiente a fin de asegurar que el procedimiento ha sido eficaz y eficiente para modelar la información. Las figuras siguientes corresponden a los parámetros  $\theta$  y  $\alpha$  de la simulación de los datos de ingreso del año 2014:

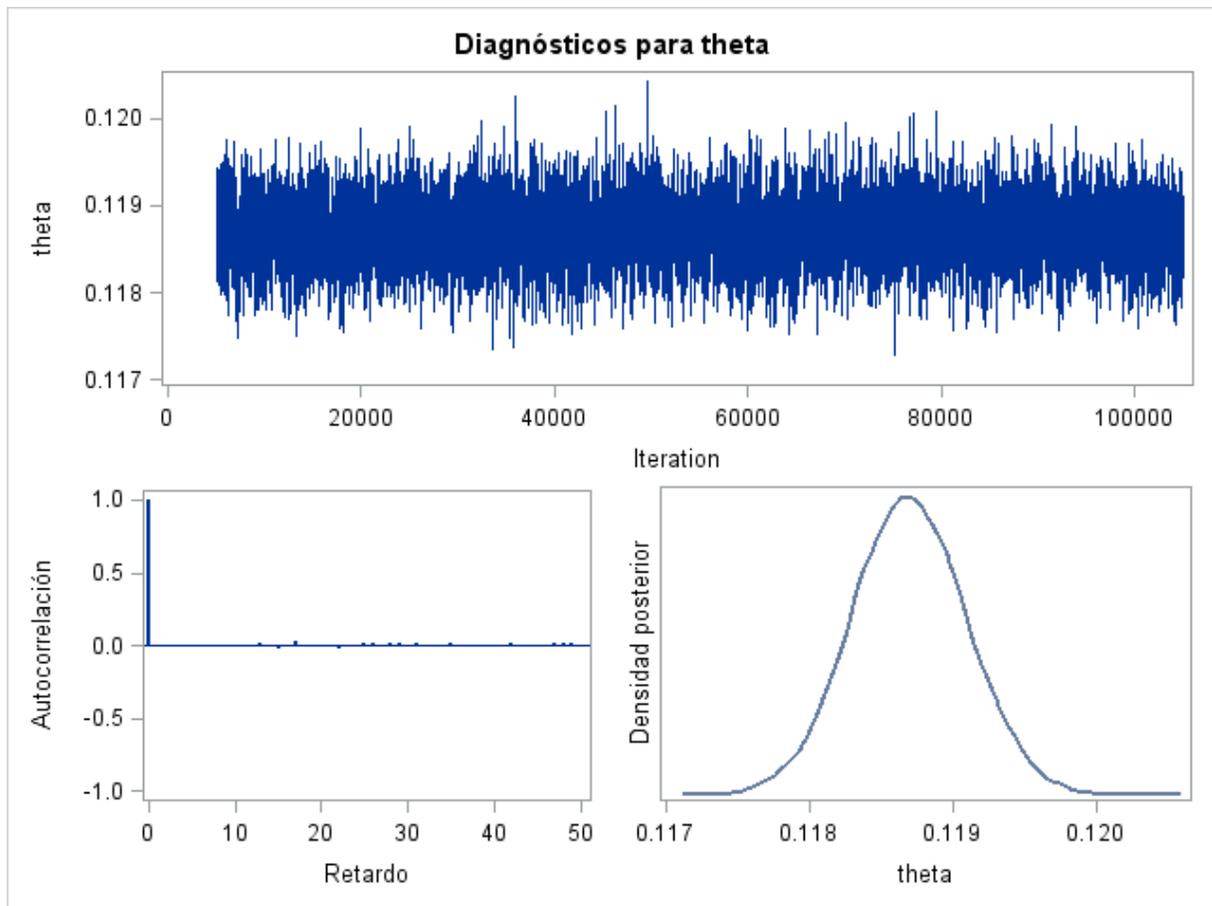


Figura 3. Diagnóstico visual de convergencia para  $\theta$  de 2014.

Dicho procedimiento genera tres tipos de gráfico los cuales son de gran utilidad para observar la convergencia. El primer grafico indica que la cadena de Markov se ha estabilizado y parece constante en el tiempo. Además, la cadena tiene un buen mezclado mostrándose "densa". El segundo gráfico corresponde a la autocorrelación entre las muestras posteriores indicando independencia con muestreo eficiente. Por último la densidad kernel estima la distribución marginal posterior para el parámetro. Lo anterior parece indicar que la cadena ha simulado correctamente alcanzado su distribución estacionaria, concluyendo razonablemente que ha convergido a la distribución posterior.

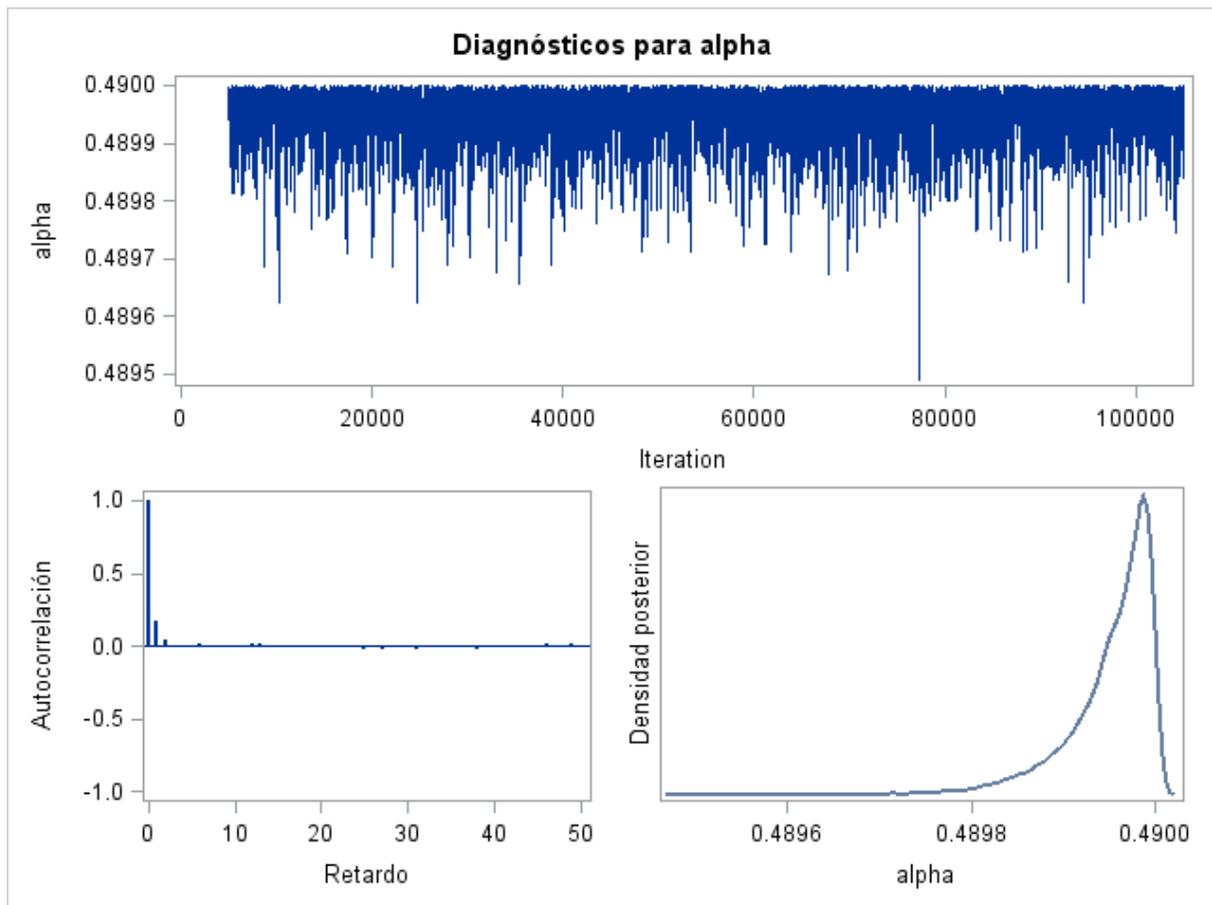


Figura 4. Diagnóstico visual de convergencia para  $\alpha$  de 2014.

El diagnóstico para  $\alpha$  indica que la cadena parece haber llegado a una distribución estacionaria. También tiene un buen mezclado y es densa. La segunda figura indica baja autocorrelación con muestreo eficiente. Finalmente la densidad kernel se muestra suave, con forma unimodal en la distribución marginal posterior para  $\alpha$ .

## **7.6. Análisis del coeficiente de Gini y la curva de Lorenz**

De acuerdo con Tuirán (2005), el índice de Gini es una medida relativa de concentración considerada como un indicador de gran uso a nivel mundial para medir y describir la desigualdad en la distribución del ingreso entre los habitantes de una región o país.

En México, han sido bastantes los estudios que se han realizado para determinar el coeficiente de Gini como una manera de analizar la distribución del ingreso.

Según estimaciones de Tuirán (2005), para el periodo de 1984 a 1994, indica una tendencia hacia la desigualdad al incrementarse dicho indicador pasando de 0.495 a 0.550, respectivamente, pero disminuyendo a 0.528 en 1996. Posteriormente, para el año de 1998 la desigualdad vuelve a crecer al obtener un coeficiente de Gini de 0.548, revirtiéndose el comportamiento en los siguientes dos años ya que en el 2000 fue de 0.540 y en 2002 de 0.516. Para el año 2004 continuó la disminución pasando a 0.512. Los indicadores anteriores, son fundamentados en dicha investigación a partir de los cambios en las participaciones de los deciles de la población mexicana, como consecuencia de la actividad económica del país y del entorno internacional. Por su parte L. E. Hernández (2000), indica para 1984 un coeficiente de Gini de 0.501, mientras que para 1989 un estimado de 0.549. En los siguientes años de análisis calcula 0.519, 0.516 y 0.507 para 1992, 1994 y 1996, respectivamente.

La investigación de Cortés (2003), también muestra un comportamiento de disminuciones e incrementos ya que para 1984 calcula un coeficiente de Gini de 0.477, mientras que para 1989 obtiene 0.518. Para el siguiente año de análisis 1992 calcula

un valor de 0.532 y para 1994 un coeficiente de 0.538; así mismo para los años posteriores, dado que en 1996 señala un valor de 0.521, 0.534 para 1998 y 0.523 para el 2000; contrastando con lo que ha publicado el INEGI con respecto al mismo indicador de medición bajo el siguiente contexto: para 1984 un coeficiente de Gini de 0.456, para 1989 un valor de 0.490, para 1992 un resultado de 0.509, para 1994 un coeficiente de 0.514, consiguientemente 0.489 para 1996, 0.509 para 1998 y 0.503 para el año 2000.

En el análisis sobre la ENIGH 2000-2006 realizado por el CEFP (2008), muestra los siguientes coeficientes de Gini: 0.456, 0.443, 0.466, 0.466, y 0.448 para los años 2000, 2002, 2004, 2005, y 2006, respectivamente. Por su parte CONEVAL (2010), estima los coeficientes de Gini 0.543 para 1992, 0.553 para 2000 y 0.517 para 2006; mientras que para 2008 indica un coeficiente de Gini de 0.530. Flores (2009), en base a datos del Banco mundial, señala que en 2007 México figuró con un coeficiente de Gini de 0.461 a nivel internacional, y L. G. Hernández (2013), que en el 2008 el coeficiente de Gini fue de 0.516.

Montes *et al.* (2012), determinan a partir de los parámetros de la distribución Dagum un coeficiente de Gini de 0.622 para el año de 1998, 0.633 para el año 2002 y un valor de 0.646 para el año 2012.

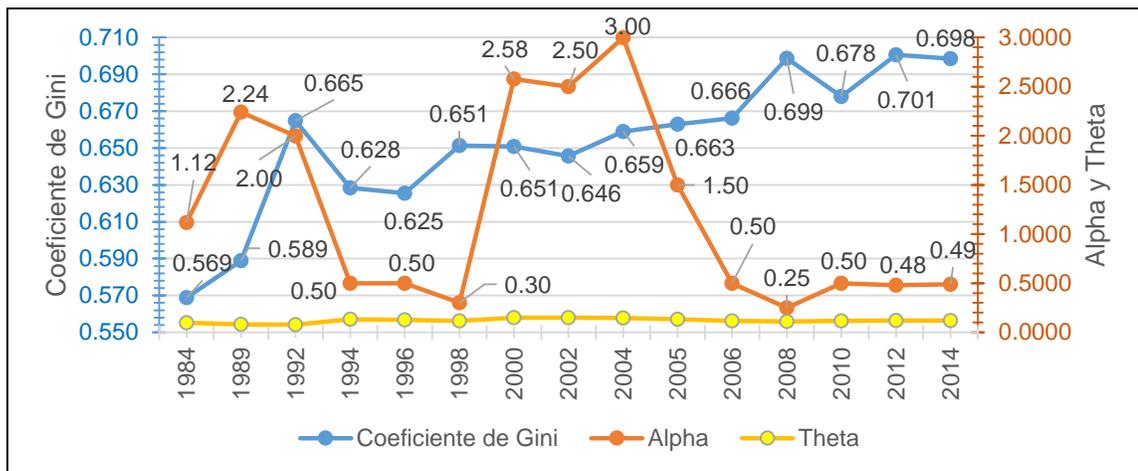
Esquivel (2015), incluye en su análisis (con datos de la Standardized World Income Inequality Database) que entre 2008 y 2012 para México se calculó un coeficiente de Gini de 0.441 en una muestra de 113 países y 0.483 en otra muestra de 132 países.

De estos comparativos, cabe destacar que las variaciones entre las estimaciones derivan de los diferentes criterios empleados por cada uno de los autores para realizar el cálculo de los coeficientes de Gini.

Por consiguiente, a continuación se muestran los resultados de la presente investigación con respecto a dicho indicador, calculados a partir de una metodología que no considera los parámetros de la distribución Pareto.

La Gráfica 1 muestra el comportamiento del coeficiente de Gini así como el comportamiento del parámetro de desigualdad o también conocido como el índice de Pareto ( $\alpha$ ) (Arnold, 2015).

Gráfica 1. Comportamiento del coeficiente de Gini y los parámetros de la distribución Pareto.



Fuente: elaboración propia con datos de la ENIGH.

Basado en el periodo de análisis, el coeficiente de Gini muestra una tendencia hacia el incremento, lo que implica una mayor desigualdad en cuanto a la distribución del ingreso. Dicho indicador ha pasado de 0.569 en 1984 a 0.698 en el año 2014.

El parámetro  $\theta$  de la gráfica tuvo un comportamiento entre 0.079 y 0.149. Dicho indicador se considera como un parámetro que define el espacio paramétrico (Hosking y Wallis, 1987).

En cuanto a los valores del parámetro  $\alpha$ , estos se han comportado de una manera más inestable, habiendo un cambio decisivo cada tres años, desde 1984 hasta el año 2004, a partir de este último los siguientes tres parámetros disminuyeron pasando de 3.0 a

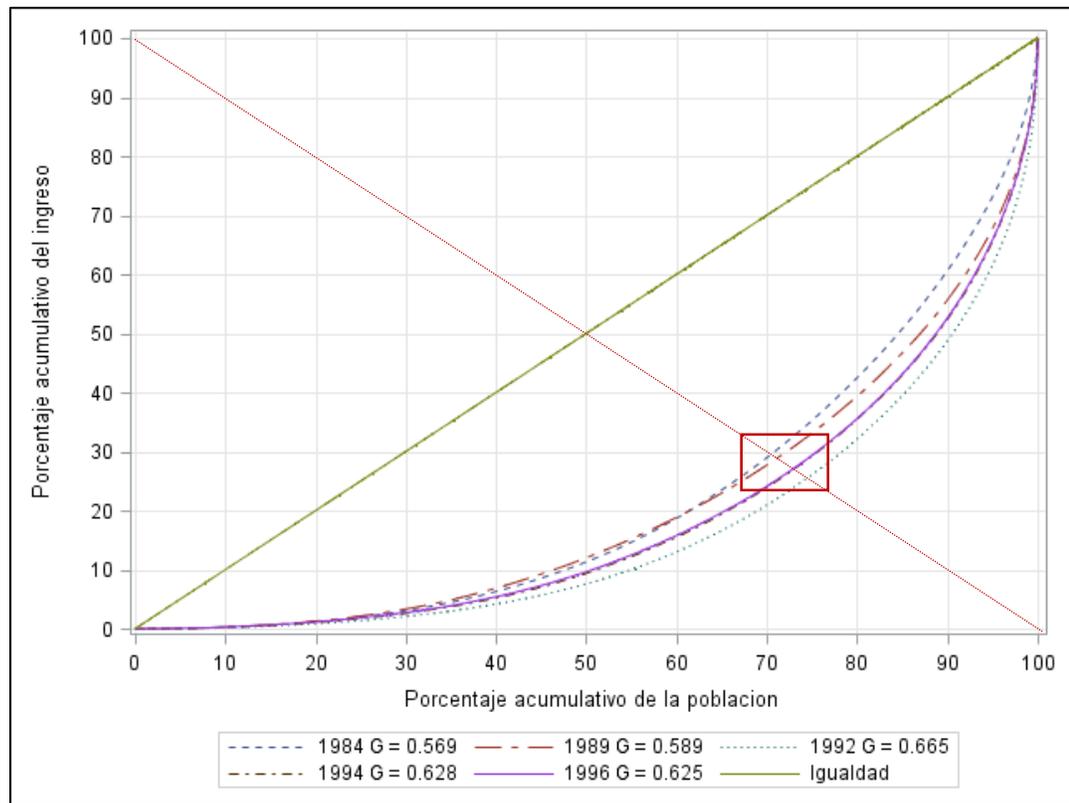
0.25 en 2008. En años más recientes los parámetros muestran un comportamiento más estable manteniéndose en 0.50 para 2010, 0.48 para 2012 y 0.50 para 2014.

El comportamiento del coeficiente de Gini y  $\alpha$  es muy versátil, pero el comportamiento entre ambos indicadores corresponde a una relación inversa.

En términos absolutos, la relación de forma inversa del coeficiente de Gini con el valor de  $\alpha$ , implica que mientras menor sea el parámetro  $\alpha$  mayor será coeficiente de Gini, y el ingreso estará mejor distribuido, mientras mayor sea el parámetro  $\alpha$  la distribución del ingreso será peor, y el coeficiente de Gini mostrará un coeficiente menor (J. Chen, Si, Li, y Zhao, 2014; F. Gómez, 2014).

Una vez examinado lo anterior, se da lugar al análisis de la relación entre el coeficiente de Gini y la curva de Lorenz surgida en la presente investigación.

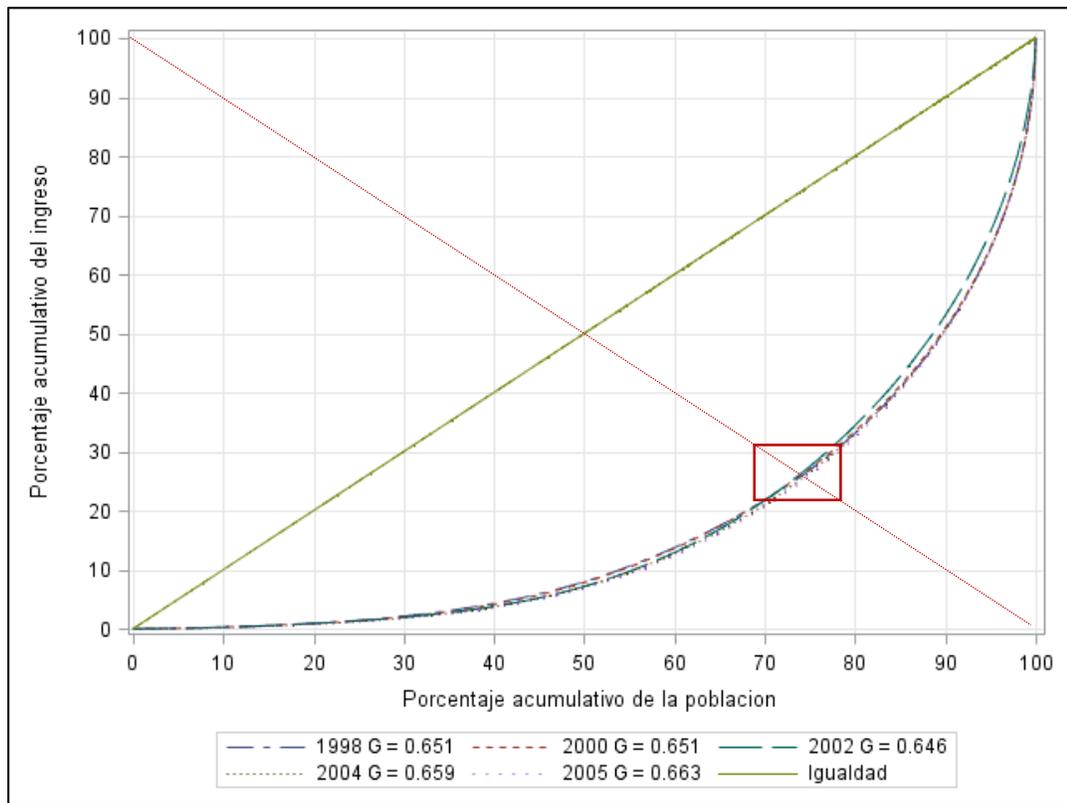
Gráfica 2. Comportamiento de la curva de Lorenz en los años 1984, 1989, 1992, 1994 y 1996.



Fuente: elaboración propia con datos de la ENIGH.

Con respecto a la curva de Lorenz para cada año, en 1984 se acercó más hacia la línea de 45 grados, considerando un coeficiente de Gini de 0.569, mientras que la más alejada a dicha línea es la curva que corresponde al año de 1992, con un coeficiente de Gini de 0.665. De esta manera, se puede decir que la curva de 1984 exhibe menor desigualdad que la curva del año 1992. Como comportamiento general de las curvas se puede interpretar que entre el 70 y 75% de la población percibió entre el 25 y el 30% del ingreso acumulado del país.

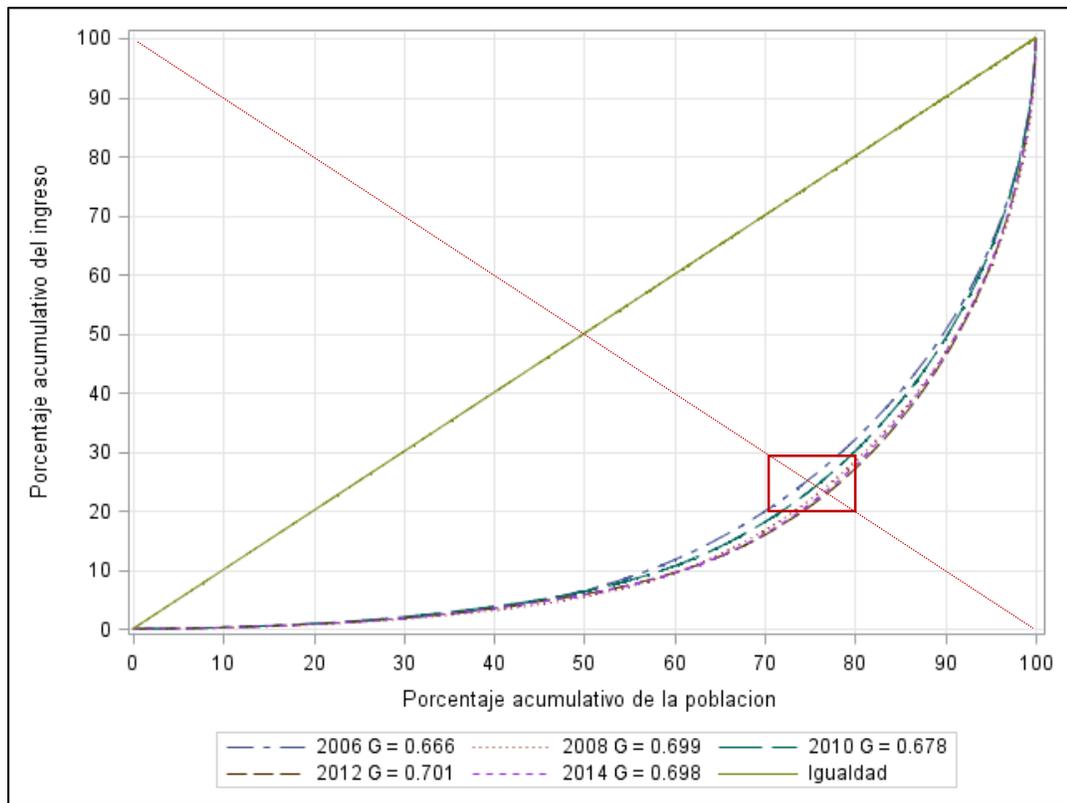
Gráfica 3. Comportamiento de la curva de Lorenz en los años 1998, 2000, 2002, 2004 y 2005.



Fuente: elaboración propia con datos de la ENIGH.

En este grupo de años, el coeficiente de Gini no tuvo una variación tan evidente, ya que en 1998 y 2000 dicho coeficiente se mantuvo en 0.651, el siguiente año disminuyó a 0.646 hasta que en el año 2004 pasó a un valor de 0.659, incrementándose ligeramente para el siguiente año de análisis 2005. Dado lo anterior, las curvas de Lorenz para dichos años tuvieron mucho acercamiento reflejando el comportamiento de los coeficientes de Gini y, por lo tanto, una desigualdad que se mantuvo constante al menos durante ese periodo de tiempo. Lo anterior muestra que menos del 75% de la población percibió poco más del 25% del ingreso acumulado del país

Gráfica 4. Comportamiento de la curva de Lorenz en los años 2006, 2008, 2010, 2012 y 2014.



Fuente: elaboración propia con datos de la ENIGH.

En años recientes, se logra denotar un ligero incremento de la desigualdad, ya que las curvas de Lorenz muestran un desplazamiento hacia la derecha de la gráfica, lo cual se refleja en el cambio del coeficiente de Gini que pasa de 0.666 en 2006 a 0.701 del año 2012. En 2014 el mismo coeficiente disminuyó sutilmente al mostrarse en 0.698. Para el año 2006, el 70% de la población percibió exactamente el 20% del ingreso acumulado mientras que para el 2014 el 70% de la población obtuvo poco más del 15% del ingreso acumulado.

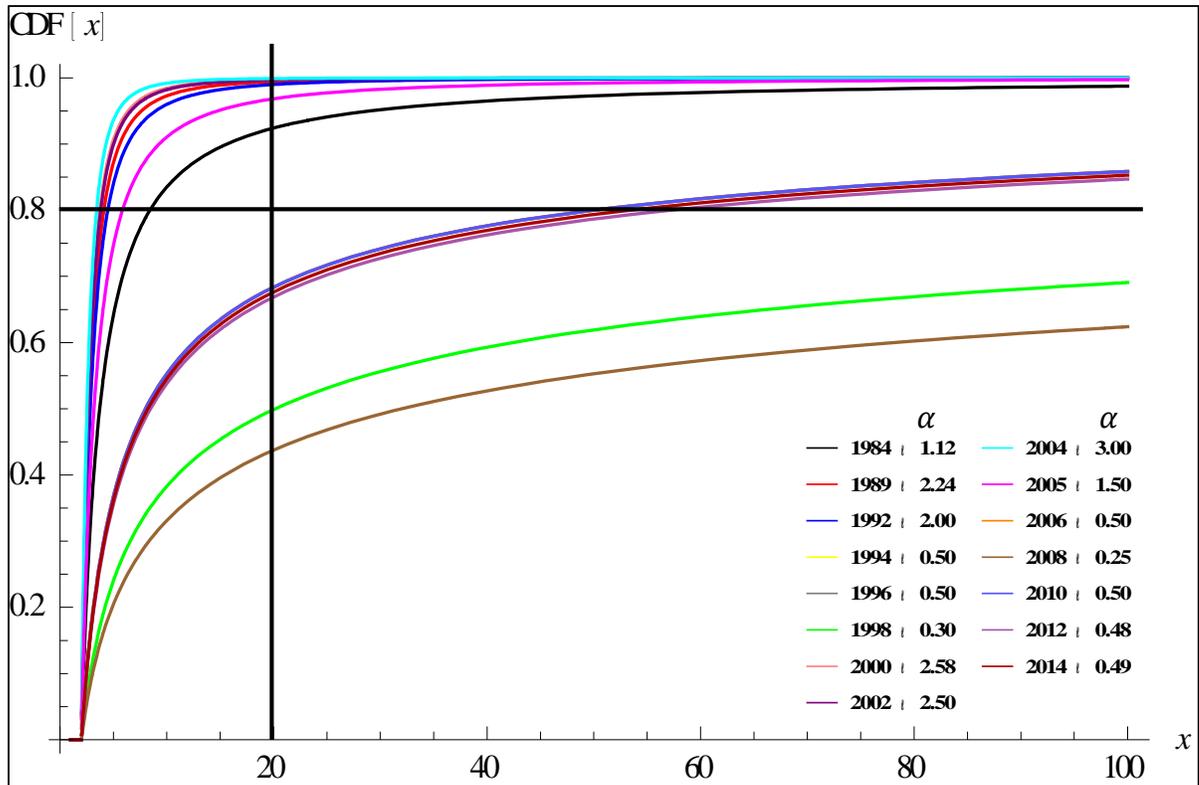
## 7.7. Análisis de la distribución Pareto

El análisis correspondiente al comportamiento de la distribución del ingreso per cápita en México basado en la distribución Pareto, considera los parámetros  $\alpha$  y  $\theta$  de dicha distribución los cuales se sustituyen en la distribución acumulada (ecuación 12) y la función de densidad de probabilidad (ecuación 13), cuyas gráficas (Gráfica 5 y Gráfica 6) permiten observar el comportamiento de la desigualdad en el tiempo.

Dado que los resultados son invariantes respecto al parámetro de escala (Hosking y Wallis, 1987), se fijó el parámetro  $\theta$  igual a 1 (entre 0.079 y 0.149 de la investigación), ya que al ser solo un valor de posición, permite observar mejor el comportamiento del índice de Pareto.

Así resulta lo siguiente:

Gráfica 5. Comportamiento de la función de distribución acumulada en base al valor de  $\alpha$ .



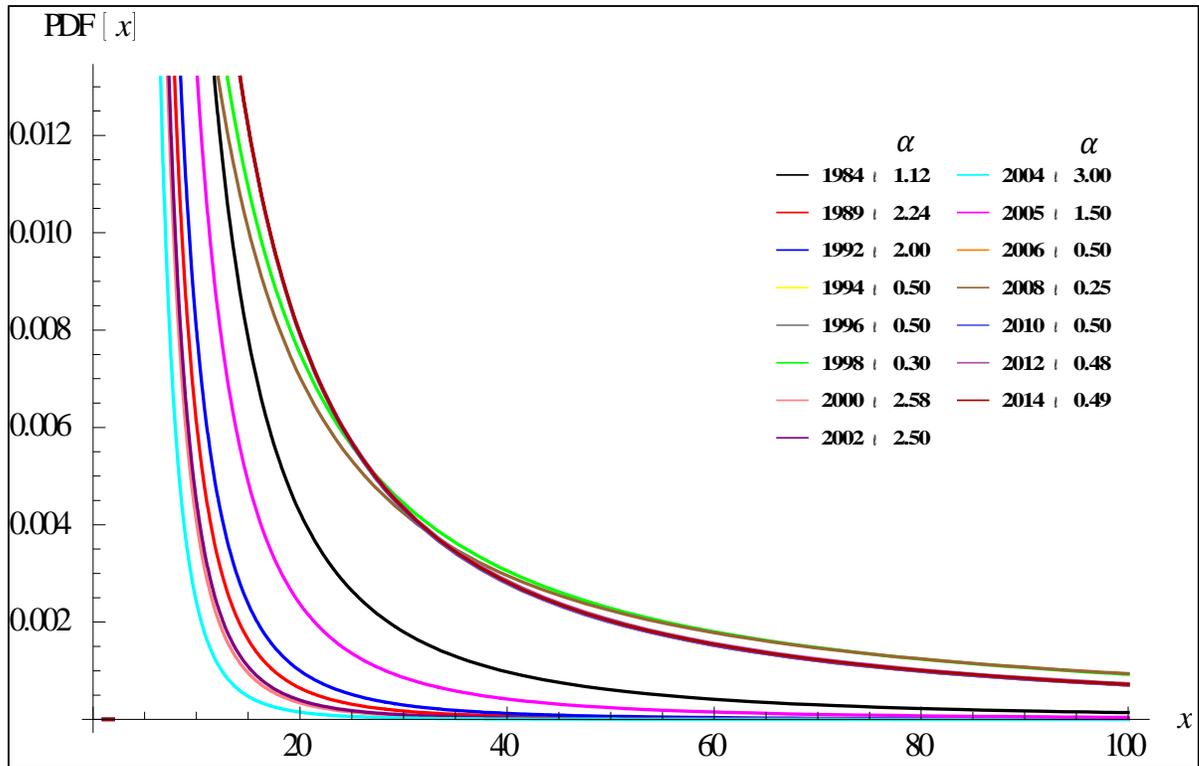
Fuente: elaboración propia con datos de la ENIGH.

La Gráfica 5 corresponde a la función de distribución acumulada en la cual el eje de las abscisas representa la fracción de la población que percibe un ingreso, en tanto el eje de las ordenadas representa la fracción de la riqueza en términos de la distribución del ingreso. En la misma gráfica las líneas que cruzan la figura representa el Principio de Pareto, que básicamente describe el desequilibrio en la distribución de la riqueza (80/20) derivado de la observación hecha por Vilfredo Pareto, que consiste en que el 20 por ciento de los miembros de la sociedad italiana poseían el 80 por ciento de la riqueza acumulada de su país (Arnold, 1985, 2015; Arnold y Press, 1983; Fujiwara et al., 2003; Setiya y Kumar, 2013).

Cada línea basada en el valor del parámetro  $\alpha$  muestra la distribución de la riqueza en México, la que representa la información de dos años, a excepción de 2004, 2005 y 2006.

Durante el periodo de tiempo considerado en la presente investigación, la literatura referente a la desigualdad en el ingreso y distribución de la riqueza en México, crecimiento económico, así como en temas de pobreza, se mencionan una serie de cambios estructurales y crisis económicas que han repercutido en el desarrollo económico de la sociedad mexicana. Bajo este contexto, y con la información de las gráficas, cabe destacar que la distribución del ingreso refiere a distintos comportamientos pero sin mostrar tendencia hacia una distribución más equitativa (Cortés, 2013). Lo anterior, se fundamenta en las distintas investigaciones que han utilizado la información referente al ingreso en México.

Gráfica 6. Comportamiento de la función de densidad de probabilidad en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

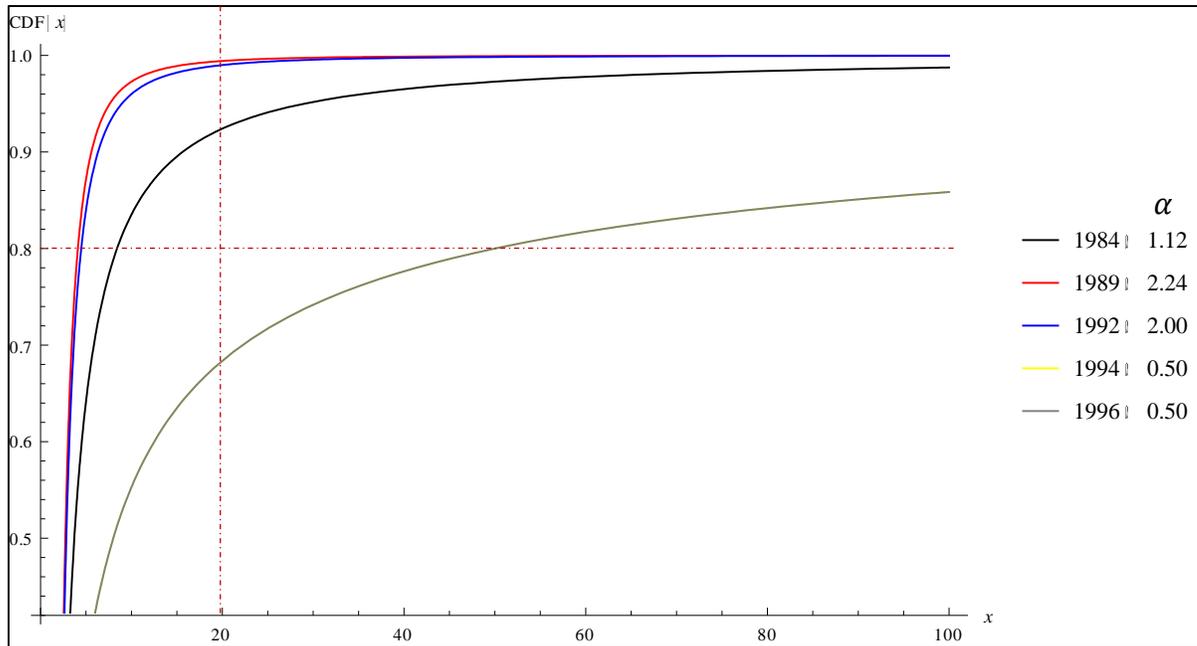
La Gráfica 6 corresponde a la función de densidad de probabilidad, que al igual que en la gráfica anterior está basada en los valores del parámetro  $\alpha$ . Dicha grafica representa el porcentaje del ingreso acumulado a lo largo del eje horizontal y la probabilidad en el eje vertical, lo que caracteriza así el comportamiento probable de los datos del ingreso. La concentración de los datos se carga mayormente hacia la izquierda de la gráfica, lo que significa que es más probable que un dato de menor proporción al ingreso total esté en la región de mayor concentración y una mínima probabilidad de que un valor de mayor proporción se ubique en un porcentaje alto del ingreso.

Cortés (2013), menciona las crisis de gran efecto que tuvieron lugar durante los años 1976, 1982, 1986-1987, 1994-1995 y 2008-2009, así como una retracción económica de 2000 a 2003 como consecuencia del estancamiento de Estado Unidos. De la misma manera, las repercusiones que tuvieron lugar por el cambio del modelo económico sustitutivo de importaciones hacia el modelo de mercado externo a mediados de los ochenta, no olvidando la firma del Tratado de Libre Comercio con Estados Unidos y Canadá a finales de 1993, mientras que hacia finales de los años noventa se inicia una era de programas de transferencias monetarias condicionadas; y el cambio de partido político en el gobierno federal acontecido durante el año 2000.

En años más recientes, dichos temas no han cambiado, siendo persistentes en cuanto a: a) estancamiento económico bajo la disparidad de salarios, siendo que la variable que retrasa al desarrollo socioeconómico del país sigue siendo el ingreso (L. G. Hernández, 2013), b) la desigualdad entre que los “pobres” siguen siendo pobres tendientes hacia la extremidad, y, c) que la fortuna de los que más tienen sigue expandiéndose en base a privilegios y oportunidades exclusivas; además de temas como la concentración de poder y la competencia desleal en los mercados de bienes y servicios, así como la falta de transparencia y corrupción en el sistema político del país (Coutiño, 2015; Esquivel, 2015; L. E. Hernández, 2000).

Para observar mejor la información, se muestran las siguientes graficas divididas en 5 años cada una, así resulta lo siguiente:

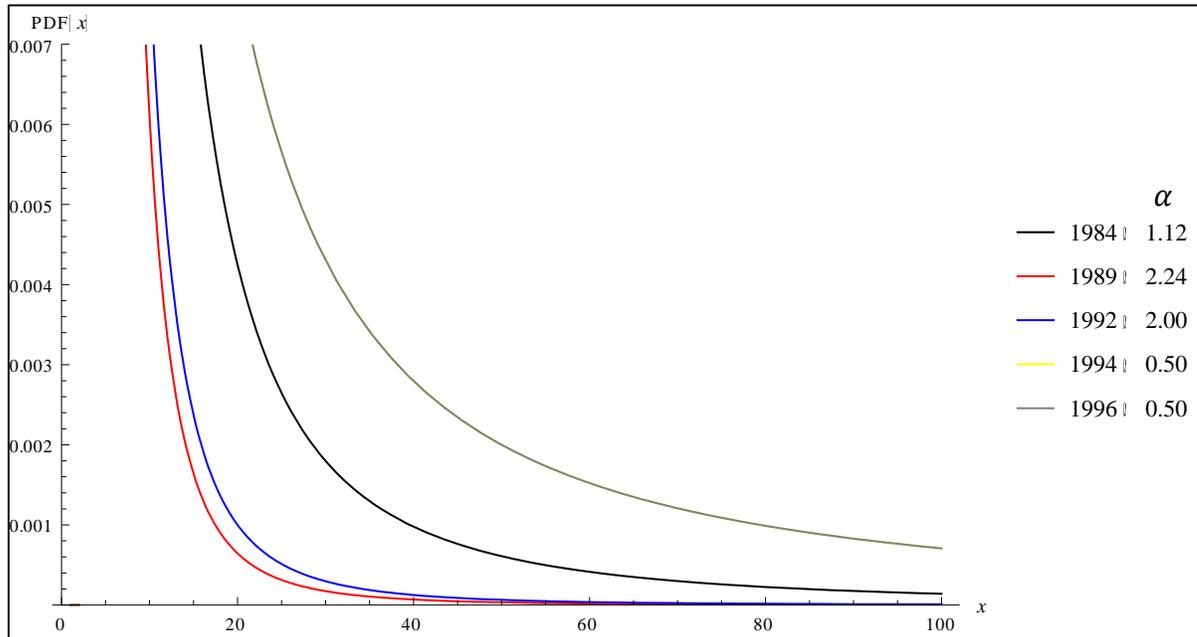
Gráfica 7. Comportamiento de la función de distribución acumulada de los años 1984, 1989, 1992, 1994 y 1996, en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

La Gráfica 7 agrupa los años 1984, 1989, 1992, 1994, y 1996, y muestra que a menores valores de  $\alpha$  se observa una mejor distribución de la riqueza, de tal manera que el parámetro correspondiente al año 1989 ( $\alpha = 2.24$ ) genera una curva más pronunciada hacia la izquierda de la gráfica, lo que indica un acercamiento muy próximo hacia el total de la riqueza ( $y=1$ ) y una pequeñísima fracción de la población ( $x < 10\%$ ).

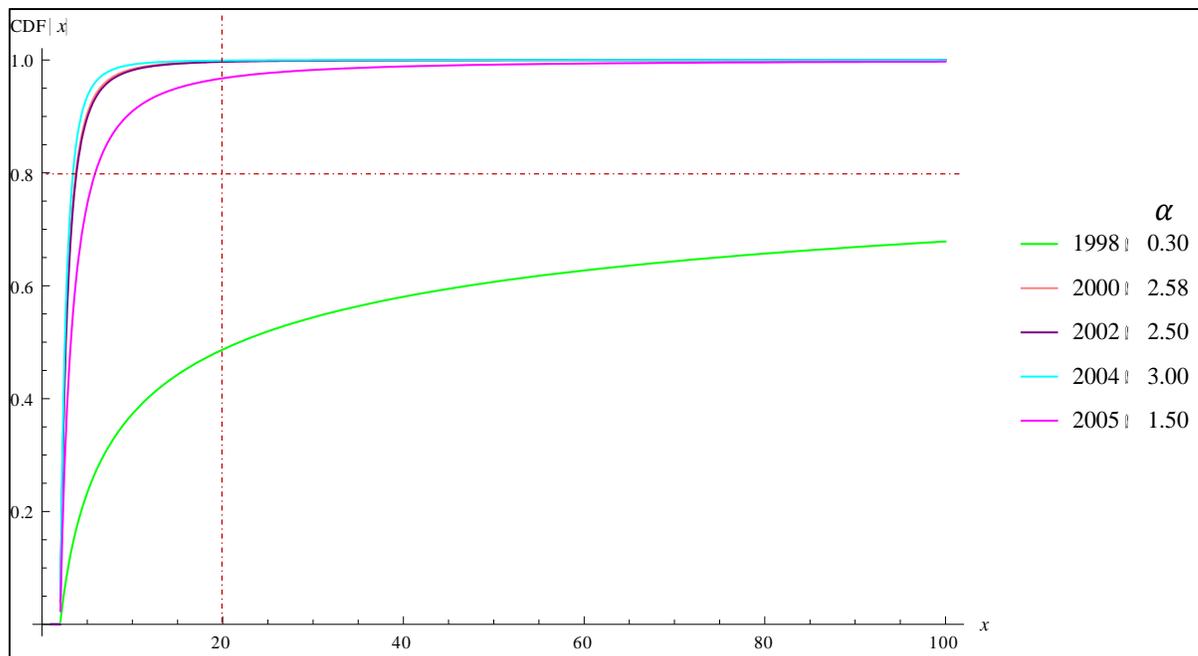
Gráfica 8. Comportamiento de la función de densidad de probabilidad de los años 1984, 1989, 1992, 1994 y 1996, en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

La Gráfica 8 muestra la función de densidad de probabilidad para distintos niveles del parámetro  $\alpha$  de los que cabe resaltar el comportamiento de las curvas cuando toman el valor de parámetro de 0.50 en los años 1994 y 1996, las cuales reflejan una mayor dispersión de las probabilidades en la forma en que la curva de dichos años despega más del eje horizontal a diferencia de las curvas de los años 1989 y 1992 las cuales indican probabilidades casi nulas de que un valor correspondiente a un ingreso alto se localice en esta región.

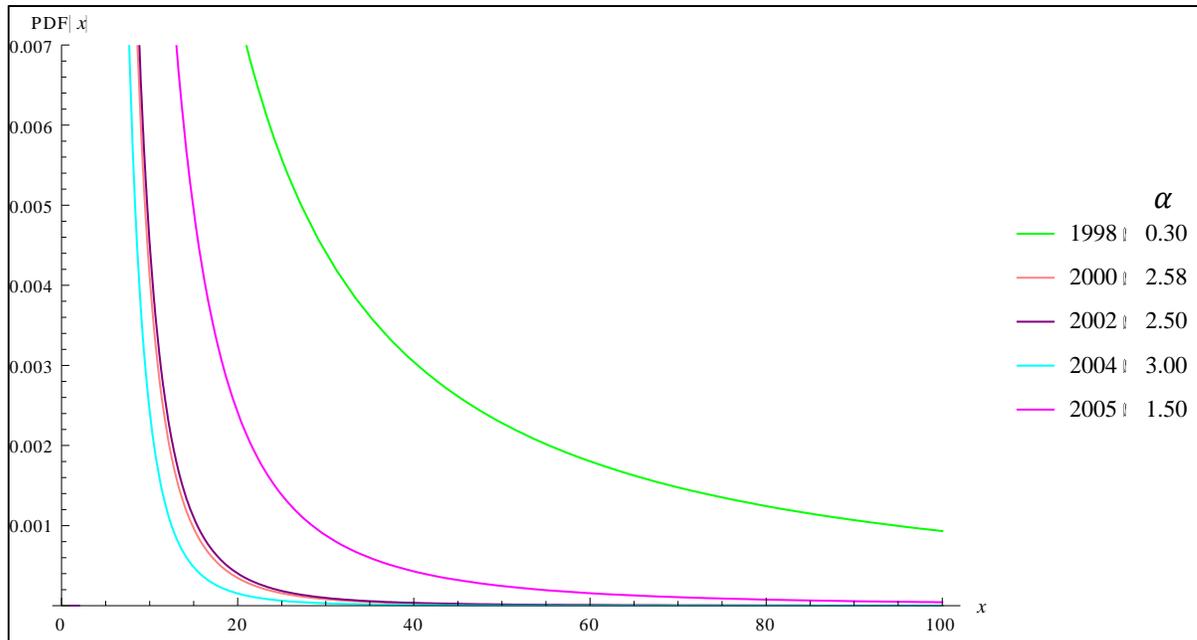
Gráfica 9. Comportamiento de la función de distribución acumulada de los años 1998, 2000, 2002, 2004 y 2005, en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

En la Gráfica 9 se observan 4 años en los que los parámetros  $\alpha$  mayores a 1 se prolongaron hacia la izquierda de la gráfica, lo que marca una diferencia evidente con respecto al parámetro del año 1998, el cual generó una curva por debajo del 80/20 de Pareto.

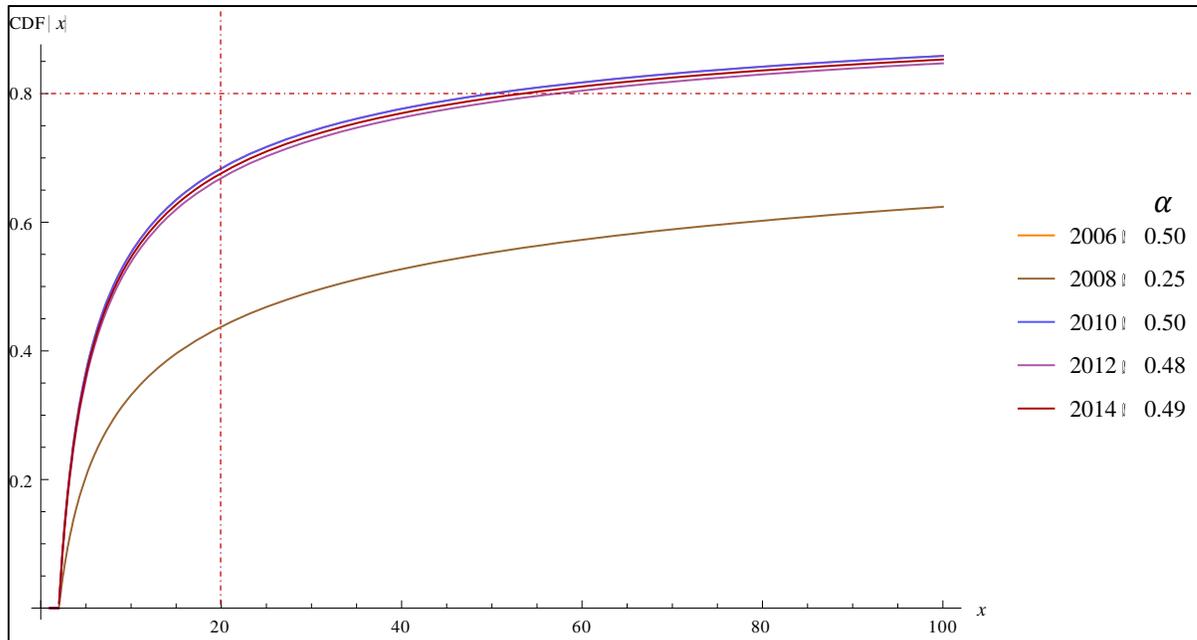
Gráfica 10. Comportamiento de la función de densidad de probabilidad de los años 1998, 2000, 2002, 2004 y 2005, en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

La Gráfica 10 muestra una curva más alejada de ambos ejes, lo que indica una mayor dispersión de las probabilidades del ingreso, a diferencia del año 2004 el cual con un parámetro  $\alpha$  de 3.0 se muestra la línea casi paralela al eje vertical, lo que refleja concentración de las probabilidades para el ingreso de dicho año.

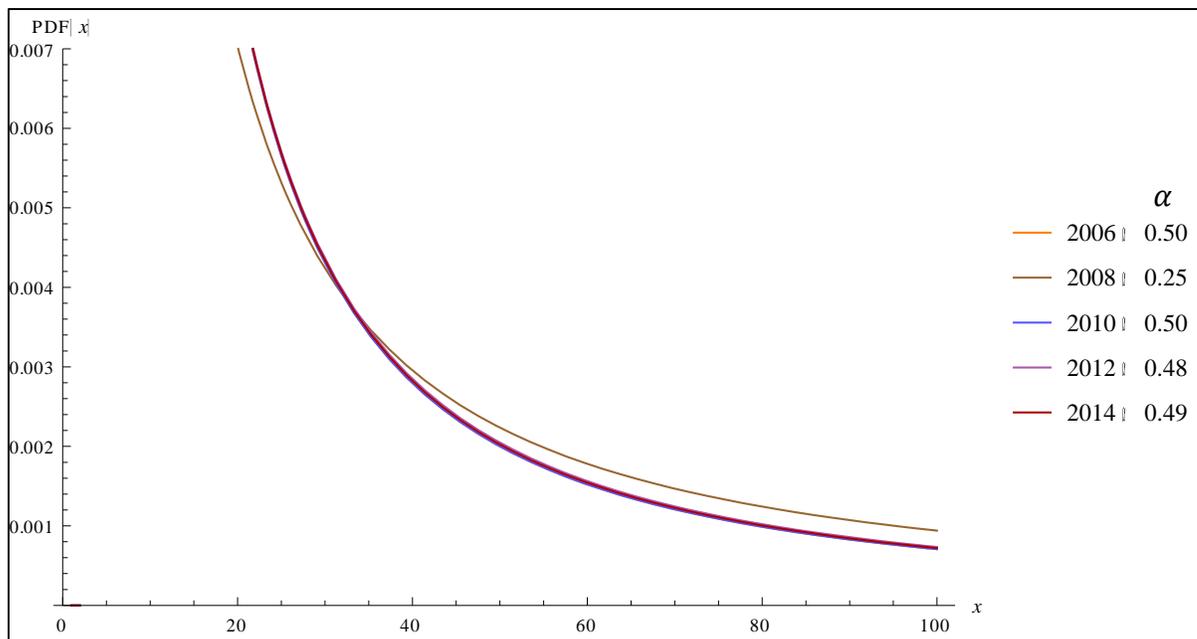
Gráfica 11. Comportamiento de la función de distribución acumulada de los años 2006, 2008, 2010, 2012 y 2014, en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

En lo que corresponde a un periodo de años más recientes, los valores de los parámetros  $\alpha$  originan curvas más alejadas del eje vertical, lo que indica una distribución no tan desigual ya que a excepción del año 2008 la distribución del ingreso de los años 2006, 2010, 2012 y 2014 se mostró en la forma en que el  $\pm 50\%$  de la población concentro el 80% de la riqueza (80/50), es decir, la riqueza se distribuyó entre más personas. Para el año 2008 la curva muestra que el  $\pm 50\%$  de la población concentro más del 50% de la riqueza acumulada del país.

Gráfica 12. Comportamiento de la función de densidad de probabilidad de los años 2006, 2008, 2010, 2012 y 2014, en base al valor de  $\alpha$ .



Fuente: elaboración propia con datos de la ENIGH.

Las probabilidades del ingreso de los 5 años se muestran más dispersas en base a las curvas de la densidad para cada parámetro ya que se encuentran alejadas de los dos ejes, principalmente del eje horizontal. A diferencia de las gráficas de los años anteriores, este grupo de años refleja un comportamiento más estable por la mínima diferencia entre los parámetros  $\alpha$ .

Las gráficas anteriores representan los distintos comportamientos que se han generado en la distribución del ingreso per cápita desde 1984 hasta 2014, y que de alguna manera representan el comportamiento socioeconómico del país.

Los argumentos que se discuten en distintas investigaciones bajo la temática de la distribución del ingreso, refieren a que los grados de desigualdad son ocasionados por las relaciones entre el gobierno y las clases sociales, divididas principalmente en tres grupos, dado que el ingreso acumulado del país está altamente concentrado (Flores,

2009). Esta división de la población señala a la clase de los que menos tienen, la clase media y la clase de ingresos altos. En términos de estadística, la población es dividida en deciles, representando al primero como la población pobre y el décimo como la población de los ricos, o más sutilmente hablando, los que generan pocos recursos y las familias que gana mucho, respectivamente (L. G. Hernández, 2013). Bajo estas circunstancias, Cortés (2003), menciona que en base a las características de baja sensibilidad que representan los ingresos medios con respecto a los extremos, las variaciones en la desigualdad pudieran estar influenciadas por lo que suceda entre el primer y decimo decil de la población como consecuencia de los beneficios que el gobierno promueva para cada clase, particularmente del décimo decil. Dados los resultados de las investigaciones de Cortés (2003, 2013), la desigualdad en la distribución del ingreso monetario crece cuando aumenta la participación del décimo decil y disminuye cuando se reduce dicha participación.

## 7.8. Discusión

La relevancia de la medición de la distribución del ingreso entre individuos u hogares en una economía, se fundamenta en el hecho de que ésta permite conocer que tanto se aleja de una distribución perfectamente equitativa, y así también determinar, a través del tiempo, si dicha medición en la distribución del ingreso tiende a mostrar una mayor o menor desigualdad (CONEVAL, 2012). Al respecto, los elevados niveles de desigualdad en México, medidos mediante el coeficiente de Gini, tienen implicaciones importantes en su desarrollo, ya que limitan el impacto del crecimiento económico en la reducción de la pobreza (De Ferranti, 2004).

Diversos autores han concluido que la desigualdad del ingreso en México se ha incrementado a partir de los años noventa (CEFP, 2008; Montes, 2012). Montes *et al.* (2012) muestran que al comparar los años de 1998 y 2008 dicho ingreso ha tendido a concentrarse, es decir, en México se ha reducido el número de ricos, concentrando mayor riqueza, pero se ha incrementado el número de personas pobres, las cuales disponen de menores recursos. El Centro de Estudios de las Finanzas Públicas (CEFP, 2008) también concluyó que en 2012 con respecto a 2010, la concentración del ingreso aumentó. Por el contrario, Székely (2005) bajo el periodo 1996 -2004 y CONEVAL (2014) para el periodo 1992 – 2012, concluyen que la desigualdad en la distribución del ingreso ha disminuido, y lo cual según Székely (2005) se debe a la expansión de la clase “media” provocada por la reducción de la pobreza.

Dado que no se tiene una posición concluyente, con respecto a la tendencia en la concentración del ingreso en México, en la presente investigación se ha modelado de manera bianual (a excepción de 2004, 2005 y 2006) la información del ingreso per cápita de los años 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012 y 2014 mediante la distribución Pareto con estimadores Bayesianos. Una de las ventajas del método Bayesiano comparado con otros métodos como la estadística clásica, es que permite obtener resultados significativamente más confiables. Reed (2003, 2004), ha indicado que, bajo la estadística clásica en la que

se asume una distribución, por ejemplo la distribución Pareto, se desempeña de manera no adecuada para modelar un amplio rango de observaciones con diferentes niveles de ingreso, y sólo se obtiene un buen comportamiento cuando sólo se consideran observaciones con niveles de ingreso alto.

## 8. Conclusiones

En base a los resultados de la simulación Monte Carlo vía Cadenas de Markov ajustados a la distribución Pareto mediante estimadores Bayesianos, se concluye que la metodología Bayesiana considerada para la presente investigación permitió ajustarse a la distribución Pareto a través de los estimadores Bayesianos de los cuales derivaron parámetros estadísticamente significativos en base a las pruebas de diagnóstico estadístico para la evaluación de la convergencia de las cadenas de Markov; dando soporte de esta manera a la inferencia Bayesiana en el análisis de la distribución de la riqueza en México.

Los parámetros de la distribución Pareto, determinados a partir de la simulación Monte Carlo vía Cadenas de Markov, modelaron la información de 15 años del ingreso per cápita en México como se esperaba en base a la determinación de la función de distribución acumulada y la función de densidad de probabilidad de dicha distribución. Cabe destacar que la cantidad de datos considerados para llevar a cabo la simulación fue toda la información disponible de la ENIGH del INEGI.

De acuerdo con los resultados de la presente investigación, la distribución de la riqueza ha sido desigual en distintos grados. En años más recientes, la distribución del ingreso en México ha logrado estabilizarse bajo un contexto de menor desigualdad.

# Referencias

- Abdel-Ghaly, A. A., Attia, A. F., y Aly, H. M. (1998). Estimation of the parameters of Pareto distribution and the reliability function using accelerated life testing with censoring. *Communications in Statistics-Simulation and Computation*, 27(2), 469-484.
- Ahmed, M., y Bhattacharya, N. (1972). Size distribution of per capita personal income in India: 1955-56, 1960-61 and 1963-64. *Economic and Political Weekly*, 7(31/33), pp. 1581-1588-1588. Retrieved from: <http://www.jstor.org/stable/4361666>
- Akinsete, A., Famoye, F., y Lee, C. (2008). The Beta-Pareto distribution. *Statistics*, 42(6), 547-563. Retrieved from: <http://dx.doi.org/10.1080/02331880801983876><http://www.tandfonline.com/doi/pdf/10.1080/02331880801983876>
- Alcaíno, E. G. I. G. (2004). Modelo Bayesiano semi-paramétrico de respuesta binaria mejorando los métodos de estimación y análisis. (Licenciatura Research Thesis), Universidad Austral de Chile. Retrieved from <http://cybertesis.uach.cl/tesis/uach/2004/bmficia346m/doc/bmficia346m.pdf>
- Alonso, R. A. (2008). El resurgir de Thomas Bayes. *Anuario jurídico y económico escorialense* (41), 327-360. Retrieved from: <http://dialnet.unirioja.es/descarga/articulo/2652087.pdf>
- Altman, D. G., y Bland, J. M. (2005). Standard deviations and standard errors. *BMJ : British Medical Journal*, 331(7521), 903-903. Retrieved from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1255808/>
- Alzaatreh, A., Famoye, F., y Lee, C. (2012). Gamma-Pareto distribution and its applications. *Journal of Modern Applied Statistical Methods*, 11(1), 7-7.
- Araña, J. E. (2004). Métodos Bayesianos en Economía de la Salud. *Cuadernos económicos de ICE*(67), 299-320. Retrieved from: [http://www.revistasice.com/CachePDF/CICE\\_67\\_299-320\\_\\_9735608E60FB38C3894471BDE6B8B7BE.pdf](http://www.revistasice.com/CachePDF/CICE_67_299-320__9735608E60FB38C3894471BDE6B8B7BE.pdf)
- Arnold, B. C. (1985). *Pareto distribution*: John Wiley & Sons, Inc.
- Arnold, B. C. (2015). *Pareto Distributions*: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

- Arnold, B. C., y Press, S. J. (1983). Bayesian inference for Pareto populations. *Journal of Econometrics*, 21(3), 287-306. Retrieved from: <http://www.sciencedirect.com/science/article/pii/0304407683900477>
- Arya, J. C., y Lardner, R. W. (2002). *Matemàtiques Aplicadas a la Administraciòn y a la Economia*: Pearson Educaciòn.
- Ausín, C. (2012). *Métodos de cadenas de Markov Monte Carlo*. Retrieved from [halweb.uc3m.es/esp/Personal/personas/causin/esp/2012.../Tema8.pdf](http://halweb.uc3m.es/esp/Personal/personas/causin/esp/2012.../Tema8.pdf)
- Barton, C. C., y Scholz, C. H. (1995). The fractal size and spatial distribution of hydrocarbon accumulations *Fractals in petroleum geology and earth processes* (pp. 13-34): Springer US.
- Beirlant, J., de Wet, T., y Goegebeur, Y. (2006). A goodness-of-fit statistic for Pareto-type behaviour. *Journal of Computational and Applied Mathematics*, 186(1), 99-116. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0377042705001913>
- Bellù, L. G., y Liberati, P. (2005). *Charting Income Inequality: The Lorenz Curve*. FAO, EASYPol Module, 40. Retrieved from: [http://www.fao.org/docs/up/easypol/302/charting\\_income\\_inequality\\_000en.pdf](http://www.fao.org/docs/up/easypol/302/charting_income_inequality_000en.pdf)
- Bellù, L. G., y Liberati, P. (2006). *Inequality analysis: The Gini index*. FAO, EASYPol Module, 40. Retrieved from: [http://www.fao.org/docs/up/easypol/329/gini\\_index\\_040en.pdf](http://www.fao.org/docs/up/easypol/329/gini_index_040en.pdf)
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*: Springer Science & Business Media.
- Berry, S. M., Carlin, B. P., Lee, J. J., y Muller, P. (2010). *Bayesian adaptive methods for clinical trials*: CRC press.
- Botargues, P., y Petrecolla, D. (1999). *Estimaciones Paramétricas Y No Paramétricas de la distribución del ingreso de los ocupados del Gran Buenos Aires, 1992-1997*. *Económica*. Facultad de Ciencias Económicas Universidad Nacional de La Plata, XLV(1), 13-34.
- Bouchaud, J.-P., y Potters, M. (2000). *Theory of financial risks*: Cambridge University Press, Cambridge.
- Bound, J., y Johnson, G. E. (1992). *Changes in the Structure of Wages during the 1980's: An Evaluation of Alternative Explanations*. Retrieved from <http://www.nber.org/papers/w2983.pdf>

- Bradley, H., Christophe, T., y Denis, C. (2014). Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations. Retrieved from: <http://www.tqmp.org/RegularArticles/vol10-2/p107/p107.pdf>
- Budnick, F. S. (1990). Matemáticas aplicadas para Administración, Economía y Ciencias Sociales (2a ed. ed.): México: Litoarte. .
- Carlin, B. P., y Louis, T. A. (2000). Bayes and empirical Bayes methods for data analysis: New York, Chapman & Hall/CRC.
- Castañeda, A., Díaz, G. J., y Ríos, R. J. V. (2003). Accounting for the US earnings and wealth inequality. *Journal of Political Economy*, 111(4), 818-857. Retrieved from: <http://www.jstor.org/stable/10.1086/375382>
- CEFP. (2008). Distribución del ingreso y desigualdad en México; un análisis sobre la ENIGH 2000 a 2006. Centros de Estudios de las Finanzas Públicas, Cámara de Diputados. Palacio Legislativo de San Lázaro. México. 9,1-18. Retrieved from <http://www.cefp.gob.mx/intr/edocumentos/pdf/cefp/2008/cefp0092008.pdf>.
- CESOP. (2013). Reporte CESOP. Notas acerca de la cruzada contra el hambre. Centro de Estudios Sociales y de Opinión Pública de la Cámara de Diputados, LXII Legislatura. (61). Retrieved from <http://www3.diputados.gob.mx/camara/content/view/full/282829>.
- Clauset, A., Shalizi, C. R., y Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661-703.
- Clementi, F., y Gallegati, M. (2005). Power law tails in the Italian personal income distribution. *Physica A: Statistical Mechanics and its Applications*, 350(24), 427-438. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0378437104014918>
- CONEVAL. (2010). La pobreza por ingresos en México: Consejo Nacional de Evaluación de la Política de Desarrollo Social. México, D.F. Retrieved from: [http://www.coneval.gob.mx/rw/resource/coneval/info\\_public/PDF\\_PUBLICACIONES/POBREZA\\_INGRESOS\\_MEXICO\\_WEB.pdf](http://www.coneval.gob.mx/rw/resource/coneval/info_public/PDF_PUBLICACIONES/POBREZA_INGRESOS_MEXICO_WEB.pdf)
- CONEVAL. (2012). Informe de evaluación de la política de desarrollo social en México 2012: Consejo Nacional de Evaluación de la Política de Desarrollo Social. México, D.F. Retrieved from: [http://www.coneval.gob.mx/Informes/Evaluacion/IEPDS2012/Pages-IEPDSMex2012-12nov-VFinal\\_lowres6.pdf](http://www.coneval.gob.mx/Informes/Evaluacion/IEPDS2012/Pages-IEPDSMex2012-12nov-VFinal_lowres6.pdf)
- CONEVAL. (2014). Pobreza y género en México: hacia un sistema de indicadores. Información 2008-2012: Consejo Nacional de Evaluación de la Política de Desarrollo Social. México, D.F. Retrieved from:

[http://www.coneval.gob.mx/Informes/Coordinacion/Publicaciones%20oficiales/POBREZA\\_Y\\_GENERO\\_WEB.pdf](http://www.coneval.gob.mx/Informes/Coordinacion/Publicaciones%20oficiales/POBREZA_Y_GENERO_WEB.pdf)

Cortés, F. (2003). El ingreso y la desigualdad en su distribución en México. Papeles de población, 9(35). Retrieved from: <http://www.redalyc.org/articulo.oa?id=11203507>

Cortés, F. (2013). Medio siglo de desigualdad en el ingreso en México. Economía-UNAM, 10(29), 12-34. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S1665952X13721935>

Coutiño, A. (2015, Lunes, 07 de septiembre de 2015). México y la desigualdad en la distribución de la riqueza (I). [en línea]. CNNExpansión sección Opinión. Recuperado en noviembre de 2015 de <http://www.cnnexpansion.com/opinion/2015/09/04/por-que-persiste-la-desigualdad-de-ingresos-en-mexico-i>.

Cowles, M. K., y Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association, 91(434), 883-904. Retrieved from: [http://www.math.pitt.edu/~cbsg/Materials/Cowles96\\_MCMCconvergence.pdf](http://www.math.pitt.edu/~cbsg/Materials/Cowles96_MCMCconvergence.pdf)

Champernowne, D. G. (1952). The graduation of income distributions. Econometrica, 20(4), pp. 591-615-615. Retrieved from: <http://www.jstor.org/stable/1907644>

Champernowne, D. G. (1953). A model of income distribution. The Economic Journal, 63(250), pp. 318-351-351. Retrieved from: <http://www.jstor.org/stable/2227127>

Chen, J., Si, Y., Li, F., y Zhao, A. (2014). An analysis of relationship among income inequality, poverty, and income mobility, based on distribution functions. Abstract and Applied Analysis, 2014, 10.

Chen, M. H., y Shao, Q. M. (1999). Monte Carlo estimation of Bayesian Credible and HPD Intervals. Journal of Computational and Graphical Statistics, 8(1), 69-92. Retrieved from: <http://amstat.tandfonline.com/doi/abs/10.1080/10618600.1999.10474802>

Chen, M. H., Shao, Q. M., e Ibrahim, J. G. (2012). Monte Carlo methods in Bayesian computation: Springer Science & Business Media.

Dagum, C. (1980). Sistemas generadores de distribución del ingreso y la Ley de Pareto. El Trimestre Económico, 47(188), 877-917.

Dagum, C., y de Dagum, E. M. B. (1971). Introducción a la Econometría: Siglo XXI.

- Dargahi-Noubary, G. R. (1989). On tail estimation: an improved method. *Mathematical Geology*, 21(8), 829-842. Retrieved from: <http://link.springer.com/article/10.1007/BF00894450#page-1>
- Dawid, A. P. (2004). *Invariant Prior Distributions Encyclopedia of Statistical Sciences*: John Wiley & Sons, Inc.
- De Ferranti, D. M. (2004). *Inequality in Latin America: Breaking with History? : World Bank Publications*. Retrieved from: [http://www.cepal.org/ilpes/noticias/paginas/7/29107/inequality\\_latina\\_america\\_complete.pdf](http://www.cepal.org/ilpes/noticias/paginas/7/29107/inequality_latina_america_complete.pdf).
- DeGroot, M. H., y Schervish, M. J. (2002). *Probability and statistics*: Addison Wesley.
- Dragulescu, A., y Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A: Statistical Mechanics and its Applications*, 299(12), 213-221. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0378437101002989>
- Dunn, W. L., y Shultis, J. K. (2011). *Exploring Monte Carlo Methods*: Elsevier.
- Eberly, L. E., y Casella, G. (2003). Estimating Bayesian Credible Intervals. *Journal of Statistical Planning and Inference*, 112(1-2), 115-132. doi:[http://dx.doi.org/10.1016/S0378-3758\(02\)00327-0](http://dx.doi.org/10.1016/S0378-3758(02)00327-0)
- Edwards, W., Lindman, H., y Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193.
- ENIGH. (2014). *Encuesta Nacional de Ingresos y Gastos de los Hogares*. Instituto Nacional de Estadística y Geografía (INEGI). Retrieved from <http://www.inegi.org.mx/est/contenidos/Proyectos/encuestas/hogares/regulares/enigh/>.
- Esquivel, G. (2015). *Desigualdad Extrema en México: Concentración del Poder Económico y Político*. Oxfam México. Consultado en noviembre de 2015. Retrieved from <http://www.oxfamMexico.org/desigualdad-extrema-en-mexico-concentracion-del-poder-economico-y-politico/#.VkLoBPkvfIU>
- Evans, M. J., y Rosenthal, J. S. (2005). *Probabilidad y Estadística*: Reverte.
- Eyzaguirre, P. R., y Chue, G. J. (2001). *Análisis Bayesiano del Modelo de Regresión Lineal Múltiple*. Paper presented at the *anales científicos*. Universidad Nacional Agraria La Molina.

- Feenberg, D. R., y Poterba, J. M. (1993). Income inequality and the incomes of very high-income taxpayers: evidence from tax returns *Tax Policy and the Economy*, Volume 7 (pp. 145-177): MIT Press.
- Fink, D. (1997). A compendium of conjugate priors. Environmental Statistics Group. Department of Biology. Montana State Univeristy. Bozeman, MT 59717. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.5540&rep=rep1&type=pdf>
- Flores, F. A. (2009). Distribución del ingreso en México 1984-2006: repercusión en las clases sociales. (Mestría Research Thesis), Universidad Autónoma de Nuevo León. Retrieved from <http://eprints.uanl.mx/1919/1/1080187848.pdf>
- Forbes, C., Evans, M., Hastings, N., y Peacock, B. (2011). *Statistical distributions*: John Wiley & Sons.
- Fujiwara, Y., Souma, W., Aoyama, H., Kaizoji, T., y Aoki, M. (2003). Growth and fluctuations of personal income. *Physica A: Statistical Mechanics and its Applications*, 321(3), 598-604. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0378437102016631>
- Gelfand, A. E., y Smith, A. F. M. (1990). Sampling-based approaches to calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410), 398-409. doi:10.2307/2289776
- Gelman, A. (2009). Bayes, Jeffreys, prior distributions and the philosophy of statistics. *Statistical Science*, 176-178. Retrieved from: [http://academiccommons.columbia.edu/download/fedora\\_content/download/ac:125280/CONTENT/jeffreys.pdf](http://academiccommons.columbia.edu/download/fedora_content/download/ac:125280/CONTENT/jeffreys.pdf)
- Gelman, A., Carlin, J. B., Stern, H. S., y Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2): London: Chapman & Hall/CRC.
- Geman, S., y Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 721-741. Retrieved from: [http://www.csee.wvu.edu/~xinl/library/papers/infor/Geman\\_Geman.pdf](http://www.csee.wvu.edu/~xinl/library/papers/infor/Geman_Geman.pdf)
- Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Vol. 196): Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.
- Geweke, J., Koop, G., y Van Dijk, H. (2011). *The Oxford handbook of Bayesian Econometrics*: Oxford University Press.

- Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, 7(4), pp. 473-483-483. Retrieved from: <http://www.jstor.org/stable/2246094>
- Gilks, W. R., Richardson, S., y Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*: London: Chapman and Hall.
- Gini, C. (1912). *Variabilita e mutabilita*, C. Cuppini, Bologna, 156 pages. Reprinted in *Memorie di metodologica statistica* (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955), 1.
- Gómez, F. (2014). Ingreso, desigualdad, pobreza y bienestar social. *Política Económica II*. Escuela de Economía. Universidad Central de Venezuela. Retrieved from <https://2s2014politecon2.files.wordpress.com/2014/12/03-distribucic3b3n-del-ingreso-pobreza-y-bienestar-social.pdf>
- Gómez, G. J., y Palarea, A. J. (2003). Algoritmos Monte Carlo basados en cadenas de Markov aplicados a la imputación múltiple de datos faltantes. Paper presented at the *Actas del 27 Congreso Nacional de Estadística e Investigación Operativa Lleida*, 8–11 de abril de 2003.
- Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, 35(2), 185-191.
- Harris, C. M. (1968). The Pareto distribution as a queue service discipline. *Operations Research*, 16(2), 307-313.
- Hartigan, J. (1964). Invariant Prior Distributions. *The Annals of Mathematical Statistics*, 35(2), pp. 836-845-845. Retrieved from: <http://www.jstor.org/stable/2238537>
- Hartigan, J. (1966). Note on the confidence-prior of Welch and Peers. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55-56.
- Hastings, W. K. (1970). Monte Carlo Sampling methods using Markov Chains and their Applications. *Biometrika*, 57(1), 97-109. Retrieved from: <http://down.cenet.org.cn/upfile/10/20053515710146.pdf>
- Hayakawa, M. (1951). The Application of Pareto's Law of income to Japanese data. *Econometrica*, 19(2), pp. 174-183-183. Retrieved from: <http://www.jstor.org/stable/1905732>
- Heidelberger, P., y Welch, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24(4), 233-245.
- Heidelberger, P., y Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6), 1109-1144.

- Hernández, L. E. (2000). Crecimiento económico, distribución del ingreso y pobreza en México. *Comercio Exterior*, 50(7). Retrieved from: [http://www.cuaed.unam.mx/posgrado/camara\\_diputados/docs/fe/creci\\_econ\\_in\\_g-pob.pdf](http://www.cuaed.unam.mx/posgrado/camara_diputados/docs/fe/creci_econ_in_g-pob.pdf)
- Hernández, L. G. (2013). El desarrollo económico en México. *Estudios-Instituto Tecnológico Autónomo de México*, 11(106), 101-140.
- Hosking, J. R. M., y Wallis, J. R. (1987). Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics*, 29(3), pp. 339-349-349. Retrieved from: <http://www.jstor.org/stable/1269343>
- Howlader, H. A., Hossain, A. M., y Makhmin, O. (2007). Bayesian prediction and estimation from Pareto distribution of the first kind. *Int J Stat Sci*, 6, 137-150. Retrieved from: <http://www.statru.org/ijss/wp-content/uploads/2011/07/P15.Vol6s.pdf>
- Infante, G. S., y Zàrate, d. L. G. P. (2012). *Métodos estadísticos: un enfoque interdisciplinario: La Gaya Ciencia*.
- Jaramillo, G. P. (2009). Estimacion de VAR Bayesianos para la Economía Chilena. *Revista de Análisis Económico*, 24, 101-126. Retrieved from: [http://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-88702009000100005&nrm=iso](http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-88702009000100005&nrm=iso)
- Kaplan, D. (2014). *Bayesian statistics for the Social Sciences (Methodology in the Social Sciences)* (s. Edition Ed.): New York: The Guilford Press.
- Kaplan, D., y Depaoli, S. (2013). Bayesian statistical methods. *Oxford handbook of quantitative methods*, 407-437.
- Kass, R. E., Carlin, B. P., Gelman, A., y Neal, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, 52(2), 93-100.
- Kass, R. E., y Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 1343-1370. Retrieved from: <http://mathfaculty.fullerton.edu/sbehseta/KassWasserman-JASA-1996.pdf>
- Katz, L. F., y Murphy, K. M. (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics*, 107(1), pp. 35-78-78. Retrieved from: <http://www.jstor.org/stable/2118323>
- Lee, P. M. (2012). *Bayesian statistics: an introduction* (4th Edition ed.): John Wiley & Sons.

- Levy, F., y Michel, R. C. (1991). The economic future of american families: Income and wealth trends: The Urban Insitute.
- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. Publications of the American Statistical Association, 9(70), pp. 209-219-219. Retrieved from: <http://www.jstor.org/stable/2276207>
- Luque, G. A. C. (2007). Prueba de bondad de ajuste para la distribución Pareto, basada en la información de Kullback-Leibler. (Maestría Research Thesis), Colegio de Postgraduados. Retrieved from <http://www.biblio.colpos.mx:8080/xmlui/handle/10521/1387>
- Lusting, N. (1976). Algunos aspectos teóricos sobre la distribución del ingreso. Demografía y Economía, 386-398. Retrieved from: [http://codex.colmex.mx:8991/exlibris/aleph/a18\\_1/apache\\_media/HC38AR9APULTTIL6S1DVSJSYRFNYRB.pdf](http://codex.colmex.mx:8991/exlibris/aleph/a18_1/apache_media/HC38AR9APULTTIL6S1DVSJSYRFNYRB.pdf)
- Mahmoudi, E. (2011). The Beta Generalized Pareto distribution with application to lifetime data. Mathematics and Computers in Simulation, 81(11), 2414-2430. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0378475411000851>
- Malamud, B. D., Morein, G., y Turcotte, D. L. (1998). Forest fires: an example of self-organized critical behavior. Science, 281(5384), 1840-1842.
- Metropolis, N. (1987). The beginning of the Monte Carlo method. Los Alamos Science, 15(584), 125-130. Retrieved from: <http://jackman.stanford.edu/mcmc/metropolis1.pdf>
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., y Teller, E. (1953). Equation of state calculations by fast computing machines. The journal of chemical physics, 21(6), 1087-1092. Retrieved from: <http://bayes.wustl.edu/Manual/EquationOfState.pdf>
- Montes, R. F. Y. (2012). Modelación del ingreso en México con un enfoque Bayesiano. (Maestría Research Thesis), Colegio de Postgraduados. Retrieved from <http://colposdigital.colpos.mx:8080/jspui/handle/10521/708>
- Montes, R. F. Y., Pérez, R. P., y Pérez, E. S. (2012). Ajuste del ingreso en México con un enfoque Bayesiano. Estudios Económicos, 27(2 (54)), 273-293. doi:10.2307/41940652
- Murphy, K. M., y Welch, F. (1992). The Structure of Wages. The Quarterly Journal of Economics, 107(1), pp. 285-326-326. Retrieved from: <http://www.jstor.org/stable/2118330>

- Nigm, A. M., y Hamdy, H. I. (1987). Bayesian prediction bounds for the Pareto lifetime model. *Communications in Statistics-Theory and Methods*, 16(6), 1761-1772. Retrieved from: [http://www.tandfonline.com/doi/abs/10.1080/03610928708829470#.VX9Vi\\_I\\_Oko](http://www.tandfonline.com/doi/abs/10.1080/03610928708829470#.VX9Vi_I_Oko)
- Norman, G. R., Streiner, D. L., y Tarrés, J. (1996). *Bioestadística* (M. H. Brace Ed.): Mosby/Doyma Libros.
- Pareto, V. (1895). La legge della domanda. *Giornale degli economisti*, 59-68. Retrieved from: [http://www.jstor.org/stable/23219874?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/23219874?seq=1#page_scan_tab_contents)
- Pareto, V. (1896). *Ecrits sur la courbe de la repartition de la richesse. Oeuvres completes de Vilfredo Pareto publiees sous la direction de Giovanni Busino. Geneve: Librairie Droz, 1965. Busino (1964), Ginebra, Librairie Droz.*
- Pareto, V. (1897). *Cours d'économie politique, Lausana. Rouge: Librairie Droz.*
- Pelletier, J. D., Malamud, B. D., Blodgett, T., y Turcotte, D. L. (1997). Scale-invariance of soil moisture variability and its implications for the frequency-size distribution of landslides. *Engineering Geology*, 48(34), 255-268. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0013795297000410>
- Pickands, I. J. (1975). Statistical inference using extreme order statistics. *The annals of Statistics*, 119-131. Retrieved from: [http://www.jstor.org/stable/2958083?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/2958083?seq=1#page_scan_tab_contents)
- Piketty, T., y Saez, E. (2003). *Income inequality in the United States, 1913-1998 (series updated to 2000 available)*. Retrieved from <http://www.nber.org/papers/w8467.pdf>
- Press, S. J. (2009). *Subjective and objective Bayesian statistics: principles, models, and applications (Second Edition ed.)*: Ney Jersey: John Wiley & Sons.
- Raftery, A. E., y Lewis, S. M. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7, 493-497.
- Raftery, A. E., y Lewis, S. M. (1996). The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms," in W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, eds., *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall, London, UK.
- Reed, W. J. (2003). The Pareto law of incomes an explanation and an extension. *Physica A: Statistical Mechanics and its Applications*, 319(0), 469-486.

- Retrieved from:  
<http://www.sciencedirect.com/science/article/pii/S0378437102015078>
- Reed, W. J. (2004). On Pareto's Law and the determinants of Pareto exponents. *Journal of Income Distribution*, 13, 7-17. Retrieved from:  
<http://www.math.uvic.ca/faculty/reed/JEI.pdf>
- Revuelta, J. (2001). Estimación y evaluación de modelos psicométricos mediante simulaciones posteriores Bayesianas. *Metodología de las Ciencias del Comportamiento*, 3, 1-18. Retrieved from:  
[https://www.uam.es/personal\\_pdi/psicologia/fjabad/cv/articulos/ARTICULOSrmcc/simulacionesbayesianas.pdf](https://www.uam.es/personal_pdi/psicologia/fjabad/cv/articulos/ARTICULOSrmcc/simulacionesbayesianas.pdf)
- Robert, C., y Casella, G. (1998). *Monte Carlo statistical methods*: Springer Science & Business Media.
- Robert, C., y Casella, G. (2013). *Monte Carlo statistical methods*: Springer Science & Business Media.
- Rytgaard, M. (1990). Estimation in the Pareto distribution. *Astin Bulletin*, 20(02), 201-216. Retrieved from:  
<http://journals.cambridge.org/action/displayFulltext?type=1&fid=9000311&jid=ASB&volumeld=20&issuelid=02&aid=9000309&bodyId=&membershipNumber=&societyETOCSession=>
- Saavedra, B. P., y Ibarra, M. V. H. (2013). El método Monte-Carlo y su aplicación a finanzas. Retrieved from  
<http://mat.izt.uam.mx/mat/documentos/notas%20de%20clase/cfenaoe3.pdf>
- Sahlin, K. (2011). Estimating convergence of Markov chain Monte Carlo simulations. (Master Research Thesis), Stockholm University. Retrieved from  
<http://www2.math.su.se/matstat/reports/master/2011/rep2/report.pdf>
- Salas, J. (2011). Estimación Bayesiana de un modelo de pequeña economía abierta con dolarización parcial. *Revista de Estudios Económicos*, 22, 41-62. Retrieved from:  
<https://www.bcrp.gob.pe/docs/Publicaciones/Revista-Estudios-Economicos/22/ree-22-salas.pdf>
- Sas, I. I. (2008). *SAS/STAT® 9.2 User's Guide*: Cary, NC: SAS Institute Inc.
- Sas, I. I. (2013). *SAS/STAT® 13.1 User's guide*: Cary, NC: SAS Institute Inc.
- Setiya, P., y Kumar, V. (2013). Bayesian estimation in Pareto type-I model. *Journal of Reliability and Statistical Studies*, 6(2). Retrieved from:  
<http://jrss.in.net/data/11%20jrss62.pdf>

- Smith, A. F. M., y Roberts, G. O. (1993). Bayesian computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1), pp. 3-23-23. Retrieved from: <http://www.jstor.org/stable/2346063>
- Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 367-377. Retrieved from: [http://www.jstor.org/stable/2245845?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/2245845?seq=1#page_scan_tab_contents)
- Sobol, I. M. (1994). *A primer for the Monte Carlo method*: CRC press.
- Stigler, S. M. (1982). Thomas Bayes's Bayesian Inference. *Journal of the Royal Statistical Society. Series A (General)*, 145(2), pp. 250-258-258. Retrieved from: <http://www.jstor.org/stable/2981538>
- Székely, M. (2005). Pobreza y desigualdad en México entre 1950 Y 2004. *El Trimestre Económico*, 72(288(4)), 913-931. doi:10.2307/20856882
- Tartalova, A. (2012). Modelling income distribution in Slovakia. Prague. The 6th International Days of Statistics and Economics, Prague, September 13-15, 2012., 13, 2012-2015.
- Tuirán, G. A. (2005). *La desigualdad en la distribución del ingreso monetario en México*: CONAPO.
- Unalmed. (s.f.). Estimación de Parametros. Estadística. Universidad Nacional de Colombia. Retrieved from [www.unalmed.edu.co/~estadist/.../Estimacion%20de%20parametros.doc](http://www.unalmed.edu.co/~estadist/.../Estimacion%20de%20parametros.doc)
- Urbisaia, H. L., y Brufman, J. Z. (2009). Frecuencistas versus bayesianos. Implicancias sobre los estudios en economía y administracion. Centro de Investigación en Métodos Cuantitativos Aplicados a la Economía y a la Gestión. Facultad de Ciencias Económicas, Universidad de Buenos Aires. Retrieved from <http://home.econ.uba.ar/economicas/sites/default/files/frecuencistas%20vs%20obayesianos.pdf>
- Valdiviezo, M. M. R., y Fermín, J. S. (2010). Estimación de índices de capacidad de procesos usando la Distribución Generalizada de Pareto. *Ingeniería Industrial*, 9(2), 93-106.
- Waller, R., Johnson, M., Waterman, M., y Martz Jr, H. (1976). Gamma prior distribution selection for Bayesian analysis of failure rate and reliability. Retrieved from <http://www.osti.gov/scitech/servlets/purl/7299878>

- Wei, G. C. G., y Tanner, M. A. (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika*, 77(3), 649-652. doi:10.1093/biomet/77.3.649
- Zaninetti, L., y Ferraro, M. (2008). On the truncated Pareto distribution with applications. *Open Physics*, 6(1), 1-6.
- Zeinab, A. H. (2007). Tests for the validity of the assumption that the underlying distribution of life is Pareto. *Journal of Applied Statistics*, 34(2), 195-201. Retrieved from: <http://www.tandfonline.com/doi/pdf/10.1080/02664760600995098>
- Zellner, A. (1985). Bayesian Econometrics. *Econometrica*, 53(2), pp. 253-269-269. Retrieved from: <http://www.jstor.org/stable/1911235>
- Zellner, A. (1988). Bayesian analysis in econometrics. *Journal of Econometrics*, 37(1), 27-50. Retrieved from: <http://www.sciencedirect.com/science/article/pii/0304407688900723>
- Zellner, A. (1996). Models, prior information, and Bayesian analysis. *Journal of Econometrics*, 75(1), 51-68. Retrieved from: <http://www.sciencedirect.com/science/article/pii/0304407695017682>

# Anexos

## Anexo A: Rutina Sas 9.4 proc MCMC, coeficiente de Gini y curva de Lorenz

```
Proc MCMC
goptions reset=all;
title 'ANALISIS DEL INGRESO AÑO
2014';
ods graphics on;
ods select all;
proc mcmc data=SASUSER.ing2014
outpost=it14out ntu=1000 nmc=100000
seed=1      nbi=5000      thin=10
propcov=quanew init=random
statistics=all diagnostics=all;
/* declare parameters */
parms theta;
parms alpha;
/* prior */
prior theta ~ gamma(shape=0.001,
scale=0.001);
prior alpha ~ gamma(shape=0.001,
scale=0.001);
/* likelihood */
model ing2014 ~ pareto(theta, alpha);
Proc means data=it14out mean median
std;
*proc print;
run;

Coeficiente de Gini y curva de Lorenz*
title 'Income distribution';
ods trace on;
proc freq data=SASUSER.ing2014;
tables ing2014 / noprint out = table;
format ing2014;
run;
ods trace off;
data table;
set table;
retain suminc perpop;
suminc + (ing2014 * count);
perpop + percent;

*proc print;
run;
proc sort data=table;
by descending suminc ;
run;
data table;
set table;
by descending suminc;
if _n_=1 then do;
totalinc=suminc;
end;
retain totalinc;
perinc = (suminc/totalinc) * 100;
run;
proc sort data=table;
by perpop;
run;
data ginidat;
set table;
xlag = lag(perpop);
xlag = xlag / 100;
ylag = lag(perinc);
ylag = ylag / 100;
columna = (perinc/100) * xlag;
columnb = (perpop/100) * ylag;
retain suma sumb;
suma + columna;
sumb + columnb;
gini = suma - sumb;
run;
proc sort data=table; by perpop; run;
data table; set table;
by perpop;
if last.perspop;
keep perpop perinc;
proc print;
run;
title2 'Gini coefficient';
proc print data=ginidat;
```

```

var gini;
where perinc = 100;
run;
title2;
Lorenz curve graph

goptions reset=all;
ods
on/ANTIALIASMAX=8000;

proc sgplot data=table;
  *yaxis label="Density";
  series y=perinc x=perpop /
  legendlabel = "Curva de Lorenz";
  series y=perpop x=perpop /
  legendlabel = "Igualdad";
  discretelegend;
run;
graphics
quit;

```

Cuadro 5. Sintaxis MCMC de Sas 9.4.

<b>AÑO</b>	<b>NTU</b>	<b>NMC</b>	<b>SEED</b>	<b>NBI</b>	<b>THIN</b>	<b>PROPCOV</b>	<b>INIT</b>
<b>1984</b>	1000	50000	123	5000	5	NMSIMP	RANDOM
<b>1989</b>	1000	100000	123	5000	5	NMSIMP	RANDOM
<b>1992</b>	1000	100000	1	5000	10	NMSIMP	RANDOM
<b>1994</b>	1000	100000	1	5000	10	NMSIMP	RANDOM
<b>1996</b>	1000	100000	1	5000	10	NMSIMP	RANDOM
<b>1998</b>	1000	100000	1	5000	10	NMSIMP	RANDOM
<b>2000</b>	1000	100000	123	5000	5	NMSIMP	RANDOM
<b>2002</b>	1000	110000	123	5000	5	NMSIMP	RANDOM
<b>2004</b>	1000	100000	123	5000	5	NMSIMP	RANDOM
<b>2005</b>	1000	100000	123	5000	10	NMSIMP	RANDOM
<b>2006</b>	1000	100000	123	5000	5	NMSIMP	RANDOM
<b>2008</b>	5000	20000	123	10000	5	NMSIMP	RANDOM
<b>2010</b>	1000	50000	10	10000	10	QUANEW	RANDOM
<b>2012</b>	1000	100000	123	5000	10	NMSIMP	RANDOM
<b>2014</b>	1000	100000	1	5000	10	NMSIMP	RANDOM

Fuente: elaboración propia en base al procedimiento MCMC de Sas 9.4.

Según el sistema Sas (2008):

NTU= especifica el número de iteraciones de ajuste para utilizar en cada fase de sintonización de la distribución propuesta. Por defecto, NTU = 500

NMC= especifica el número de iteraciones en la simulación una vez descartados los primeros valores simulados (burn-in). Este es el tamaño de la muestra MCMC si THIN = 1. Por defecto, NMC = 1000.

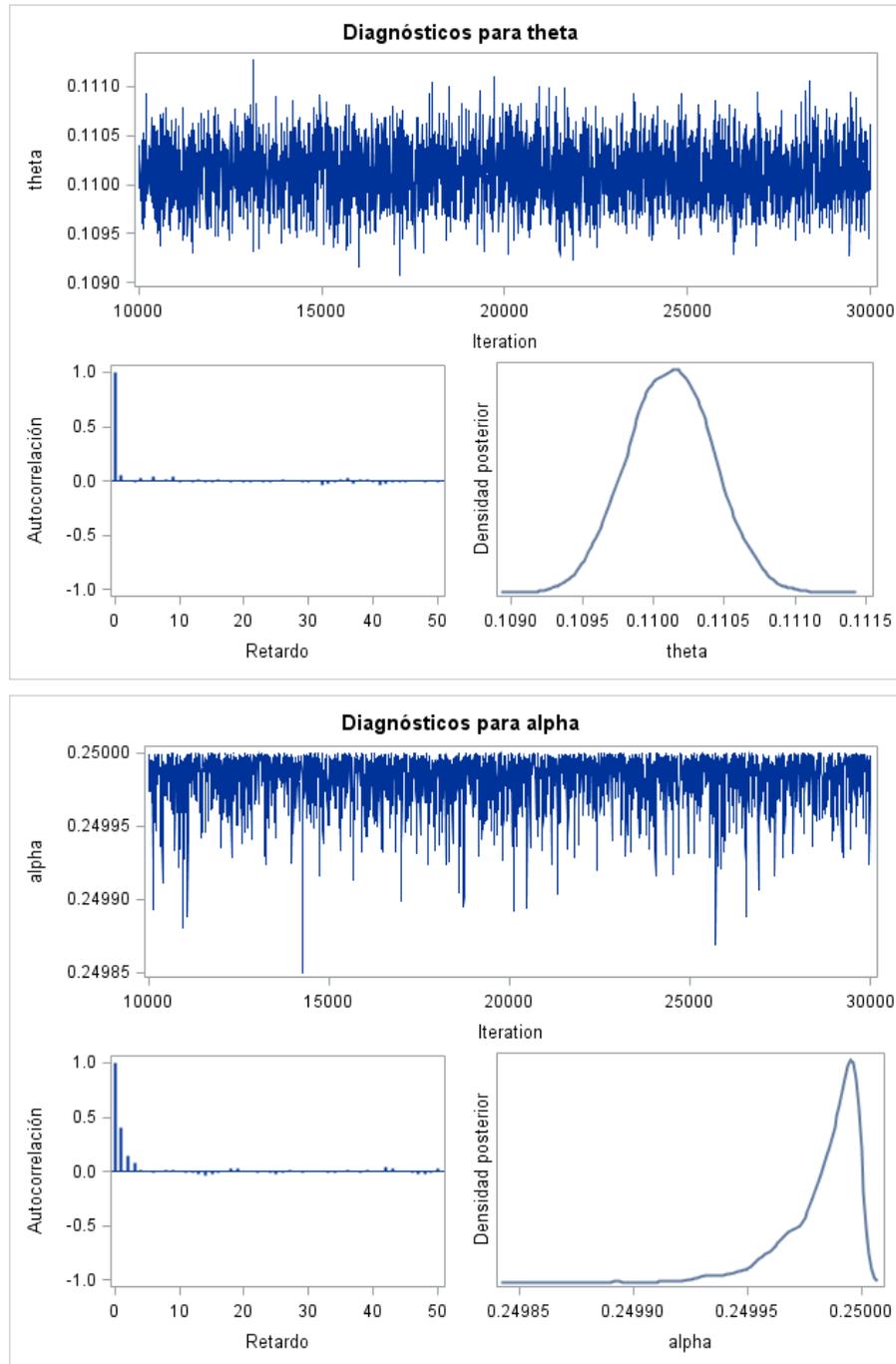
SEED= especifica la semilla de aleatorización para garantizar la reproducibilidad de la cadena de Markov. Por defecto, SEED = 0.

NBI= se refiere al proceso de descarte (burn-in) de una porción inicial de una muestra de la cadena de Markov, a manera de minimizar el efecto de los valores iniciales en la inferencia posterior. Por defecto, NBI = 1,000.

THIN= Con algunos modelos es posible que experimente una mezcla pobre (o convergencia lenta) de la cadena de Markov. Esto puede suceder, por ejemplo, cuando los parámetros están altamente correlacionados entre sí. Una mezcla pobre significa que la cadena de Markov atraviesa lentamente el espacio de parámetros y la cadena tiene una alta dependencia. Una estrategia común es afinar la cadena de Markov con el fin de reducir las autocorrelaciones de la muestra. Con seguridad se puede utilizar una cadena de Markov afinada para la inferencia posterior siempre y cuando converja la cadena. Por defecto, THIN = 1.

PROVCOV= especifica el método utilizado en la construcción de la matriz de covarianza inicial para el algoritmo de Metropolis-Hastings. Los métodos QUANEW y NMSIMP numéricamente encuentran matrices de covarianza aproximadas en el óptimo de la función de densidad posterior con respecto a todos los parámetros continuos. La fase de sintonización se inicia en los valores optimizados; en algunos problemas, esto puede aumentar en gran medida el desempeño de la convergencia. Si la matriz de covarianza aproximada no es definida positiva, entonces en su lugar se utiliza una matriz de identidad. NMSIMP realiza la optimización en base al algoritmo simplex de Nelder-Mead, QUANEW= Realiza una optimización Quasi-Newton.

INIT=RANDOM especifica opciones para generar los valores iniciales para los parámetros de la optimización. RANDOM genera un número aleatorio de la densidad a priori y lo utiliza como valor inicial del parámetro, si no se proporcionó uno.

**Anexo B: Convergencia de los parámetros  $\theta$  y  $\alpha$  del 2008**Figura 5. Diagnóstico visual de convergencia para  $\theta$  y  $\alpha$  de 2008.