

00321

61



UNIVERSIDAD NACIONAL  
AUTONOMA DE MEXICO



FACULTAD DE CIENCIAS

UN MODELO ESTADISTICO  
PARA VALORES EXTREMOS EN  
SERIES DE DATOS  
ACTUARIALES.

TESIS QUE PARA OBTENER EL TITULO DE  
ACTUARIA

PRESENTA

PATRICIA ALEJANDRA MARTINEZ RAMOS

DIRECTOR DE TESIS:  
DR. ALBERTO CONTRERAS CRISTAN



CD. UNIVERSITARIA MEXICO, D.F.



FACULTAD DE CIENCIAS  
SECCION ESCOLAR

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# **PAGINACION DISCONTINUA**



GOBIERNO NACIONAL  
MINISTERIO DE  
EDUCACIÓN

**DRA. MARÍA DE LOURDES ESTEVA PERALTA**

**Jefa de la División de Estudios Profesionales de la**

**Facultad de Ciencias**

**Presente**

Comunicamos a usted que hemos revisado el trabajo escrito: "Un modelo estadístico para valores extremos en series de datos actuariales"

realizado por Patricia Alejandra Martínez Ramos

con número de cuenta 09650320-3, quien cubrió los créditos de la carrera de: Actuaría.

Dicho trabajo cuenta con nuestro voto aprobatorio.

**Atentamente**

Director de Tesis

Propietario

Dr. Alberto Contreras Cristán

Propietario

Dr. Eduardo Arturo Gutiérrez Peña

Propietario

M. en C. Beatriz Eugenia Rodríguez Fernández

Suplente

Mat. Margarita Elvira Chávez Cano

Suplente

Dra. Ma. Asunción Begoña Fernández Fernández

**Consejo Departamental de Matemáticas**

M. en C. José Antonio Flores Díaz

CONSEJO DEPARTAMENTAL

2

A Marco

## AGRADECIMIENTOS

Al Dr. Alberto Contreras Cristán, por sus conocimientos, invaluable apoyo y tiempo para la realización de este trabajo. A la Mtra. Beatriz Rodríguez, Dra. Begoña Fernández, Dr. Eduardo Gutiérrez y a la Mtra. Margarita Chávez por la revisión de esta tesis y sus valiosos comentarios y críticas a la misma.

A mis padres Patricia y Fernando por darme la vida, su amor, apoyo incondicional y sobre todo por ser una inspiración para titularme. Gracias también a mis hermanos Fer, Clan y Gerardo, por dejarme crecer con ellos, ser mis mejores amigos y preguntarme todos los días si ya iba a acabar.

En especial a mi abuela Noemí, por su ejemplo admirable, sus consejos "anticoladera" y constante apoyo.

A Carmen, por compartir conmigo sus filosofías de la vida, ser terapeuta y construir conmigo un bello significado de amistad.

A todo el mar de fueguitos que han iluminado mis caminos, por eso mismo: Candela, Gaby, Yoyontzin, Wendy, Maricarmen, Anaf, Fab, Vaquerito und die kleine Schlange.

A los maestros de la Facultad de Ciencias que fueron parte de mi formación universitaria y al Mtro. José Luis Villarreal y Omar Hernández por su paciencia y apoyo para la impresión de este documento.

A la U.N.A.M. por permitirme ser parte de ella y los diferentes apoyos que me otorgó durante mis estudios.

## INTRODUCCIÓN

En años recientes, se ha incrementado el interés por modelar fenómenos de rara ocurrencia pero que dan lugar a grandes pérdidas de tipo financiero, económico, o que tienen un impacto considerable en el ambiente o en la salud de la población entre otras consecuencias.

Para este tipo de problemas se han desarrollado soluciones a partir de métodos que se basan en la Teoría de Valores Extremos (TVE). El enfoque clásico de esta teoría tuvo origen con el estudio del comportamiento asintótico de la distribución de observaciones "extremas" (máximos o mínimos), que realizaron autores como Fisher, Tippet, Gnedenko, Jenkinson y von Mises. El teorema propuesto por Fisher y Tippet (1928) es el resultado principal de la TVE pues establece que la forma de la distribución límite para máximos centrados y normalizados pertenece a una de tres familias paramétricas conocidas como Distribuciones de Valores Extremos. A Gnedenko (1943) se debe la primera demostración completa y rigurosa de este teorema y von Mises (1936) y Jenkinson (1955) definieron la Distribución Valor Extremo Generalizada (DVEG), que engloba las tres familias paramétricas caracterizadas por Fisher y Tippet.

Un enfoque alternativo de la TVE clásica considera la distribución de los excedentes o valores de la muestra que se encuentran sobre cierto umbral  $u \in \mathbf{R}$ . Este enfoque forma parte de la TVE moderna en el que la idea principal consiste en aproximar la distribución de los excedentes mediante la función de distribución conocida como Distribución Pareto Generalizada (DPG). En las aplicaciones actuales de la TVE, se usa la equivalencia entre la DVEG y la DPG, resultado que se ha demostrado por autores como de Haan (1970) y Pickands (1975).

En el estudio de valores extremos, algunos autores han relacionado el análisis de los excedentes con el uso de Procesos Puntuales, un ejemplo puede encontrarse en Leadbetter et al. (1983). Esta idea se deriva del hecho de que bajo ciertos supuestos, el número de observaciones por encima de un umbral  $u$  se puede aproximar con una variable aleatoria Poisson. En este tipo de aplicaciones es posible considerar métodos de bondad de ajuste para justificar el uso de los modelos Poisson. En Reiss, R.-D. y Thomas M. (1997) se explica la manera en que puede demostrarse el supuesto de que ciertos datos se generan mediante un Proceso Poisson.

En la práctica, al estudiar los valores extremos de cierto fenómeno, no es suficiente con saber que su distribución puede considerarse como una DVEG o una DPG, sino que es necesario conocer el valor de los parámetros que caracterizan a estas distribuciones. Por otra parte, una vez que se estudian los parámetros es deseable realizar predicciones sobre posibles valores futuros de las observaciones extremas. Una forma de estudiar los posibles valores de parámetros que definirán la forma de la distribución del fenómeno bajo estudio, es usando Métodos Bayesianos, en particular con los "Métodos de Simulación vía Cadenas de Markov". Dentro de estos métodos, los algoritmos de Metrópolis Hastings y el Muestreo de Gibbs son usados frecuentemente en problemas prácticos.

En el campo de estudio de la Actuaría, la necesidad de obtener modelos para valores extremos surge, por ejemplo, en las empresas reaseguradoras ya que estas establecen métodos para calcular primas cuyo monto sea suficiente a fin de proteger a otras empresas y a la misma reaseguradora del impacto financiero que se enfrentará en caso de que un riesgo catastrófico ocurra. Mientras más grande es el impacto financiero que se tenga tras la ocurrencia de un siniestro, se tiene un mayor interés en poder modelar de manera exhaustiva y precisa la probabilidad de que esos riesgos se conviertan en siniestros. En la práctica, establecer un modelo que incorpore todas las variables que pudieran intervenir en la ocurrencia de los siniestros, y que además sea adecuado para estimar el tamaño del siniestro es una tarea difícil.

En este trabajo se usa la TVE en un contexto Bayesiano además de métodos de simulación vía Cadenas de Markov con el objeto de analizar una serie de datos de reclamaciones de reaseguro, proponer un modelo con el que sea posible describir procesos de reclamaciones futuras y finalmente estimar probabilidades de pérdidas muy grandes.



El modelo que se propone para ajustar a la serie de datos se conoce como Poisson-DPG y es de tipo completamente Bayesiano (ver Smith (1997) y Smith y Goodman (2000)), es decir, el número de reclamaciones anuales se aproxima como un Proceso de tipo Poisson no homogéneo mientras que los montos de los excesos de cada año se consideran variables independientes e idénticamente distribuidas como Pareto Generalizadas; por otra parte, el carácter "completamente Bayesiano" del modelo consiste en definir distribuciones iniciales para los parámetros de la DPG, a fin de incorporar la información que se tiene sobre ellos en la muestra, vía el Teorema de Bayes.

Los conceptos teóricos que sustentan el modelo se presentan en los Capítulos I y II y los Apéndices B y C de la siguiente manera: en el Capítulo I se definen y ejemplifican las generalidades de la Inferencia estadística desde el punto de vista Bayesiano y los algoritmos de Metrópolis-Hastings y Gibbs como herramientas para simular valores de distribuciones posteriores de los modelos estadísticos Bayesianos. Dada la importancia de estos algoritmos, al final del Capítulo I se muestra un ejemplo en el que se realizan simulaciones vía el Muestreo de Gibbs para resolver un problema práctico de tipo actuarial.

El Capítulo II se dedica a la parte clásica y moderna de la Teoría de Valores Extremos, esta última presenta el enfoque que estudia los elementos de la muestra que exceden el valor de un umbral  $u$  y su distribución asintótica. Se usa el gráfico de la media de los excesos como un método para determinar el valor  $u$  a partir del cual puede considerarse que los datos tienen una distribución que es posible aproximar por medio de la DPG. Para concluir este capítulo, se realiza una breve descripción de los Procesos Puntuales, en particular, de la forma de definir a un proceso de reclamaciones de reaseguro como un Proceso de Poisson Puntual.

Las formalidades teóricas del Capítulo II se definen y en algunos casos, cuando se considera procedente, se bosqueja su demostración en el Apéndice B para lo correspondiente a la Teoría de Valores Extremos y en el Apéndice C para los Procesos Puntuales.

En el tercer Capítulo se lleva a cabo la aplicación de lo que se estudia en los capítulos I y II, con el fin de obtener un modelo que describa a la serie de datos de reclamaciones de reaseguro. Para ello, se realiza en primera instancia el análisis descriptivo de los datos, enseguida se propone el modelo Poisson-DPG estableciendo dos casos respecto a los parámetros de localización y escala: en un primer modelo se consideran efectos aleato-

#### IV

rios anuales en ambos parámetros y en el segundo caso se vuelve a definir el modelo bajo el supuesto de que los efectos aleatorios anuales en la localización no son significativos. Después de definir el modelo para ambos casos, se simularon valores para la distribución posterior de los parámetros usando los algoritmos descritos en el Capítulo I. Con los valores obtenidos se simularon montos de reclamaciones futuras y se estimaron intervalos de confianza para las probabilidades de pérdidas futuras. Para concluir el capítulo, se realiza un análisis de los resultados que se obtuvieron al llevar a cabo las simulaciones a partir del modelo propuesto.

En el Apéndice A se incluyen los códigos de S+ con los que se realizaron las simulaciones necesarias. Las figuras que describen los resultados obtenidos con ambos modelos se pueden apreciar en el Apéndice D para el caso en que se consideraron efectos aleatorios en localización y escala y en el apéndice E para el caso que considera efectos aleatorios sólo en escala.

Para finalizar este trabajo se presentan las conclusiones respecto a los resultados que se obtuvieron en general y entre otras recomendaciones, se propone como investigación posterior, implementar pruebas de análisis de sensibilidad para "calibrar" el modelo propuesto.

# Índice

<b>INTRODUCCIÓN</b>	<b>I</b>
<b>1. INFERENCIA BAYESIANA</b>	<b>1</b>
<b>Inferencia Bayesiana</b>	<b>1</b>
1.1. Inferencia desde el punto de vista Bayesiano . . . . .	1
1.2. Distribución inicial . . . . .	2
1.3. Teorema de Bayes . . . . .	3
1.4. Función de densidad posterior y función de verosimilitud . . . . .	4
1.5. Familias conjugadas . . . . .	5
1.6. Modelos Jerárquicos . . . . .	10
1.7. Algoritmos de Metrópolis-Hastings y Gibbs . . . . .	13
1.7.1. Un ejemplo de Muestreo de Gibbs . . . . .	18
<b>2. TEORIA DE VALORES EXTREMOS</b>	<b>31</b>

<b>Teoría de valores extremos</b>	<b>31</b>
2.1. Introducción	31
2.2. Descripción básica de distribuciones de valores extremos	32
2.2.1. Teoría clásica	36
2.2.2. Teoría moderna	39
2.2.3. Gráfico de la media de los excesos	43
2.3. Procesos puntuales	45
<b>3. APLICACIÓN A DATOS DE REASEGURO</b>	<b>49</b>
<b>Aplicación a datos de reaseguro</b>	<b>49</b>
3.1. Análisis descriptivo	49
3.2. Modelos Poisson-DPG	54
3.2.1. Efectos aleatorios en localización y escala	59
3.2.2. Efectos aleatorios sólo en escala	65
3.3. Simulación de la distribución posterior vía Metrópolis-Hastings	67
3.4. Simulación del valor de reclamaciones futuras y estimación de probabilidades de pérdida	70
3.5. Análisis de resultados	72

	VII
<b>CONCLUSIONES</b>	<b>76</b>
<b>APENDICES</b>	<b>79</b>
<b>A. Programas en S+</b>	<b>81</b>
<b>B. Aspectos formales de la Teoría de Valores Extremos</b>	<b>87</b>
<b>C. Procesos Puntuales</b>	<b>97</b>
<b>D. Gráficos del modelo con efectos aleatorios en localización y escala</b>	<b>107</b>
<b>E. Gráficos del modelo con efectos aleatorios sólo en escala</b>	<b>121</b>
<b>Bibliografía</b>	<b>150</b>

# Capítulo 1

## INFERENCIA BAYESIANA

### 1.1. Inferencia desde el punto de vista Bayesiano

El análisis de datos Bayesiano hace uso de modelos probabilísticos para tratar de cuantificar y hacer inferencias respecto a características del fenómeno bajo estudio. La incertidumbre de las inferencias que se realizan se cuantifica a su vez mediante probabilidades. El análisis Bayesiano sigue generalmente el orden que se describe a continuación:

- (i) Se define un modelo de probabilidad completo: es decir, una distribución de probabilidad conjunta para todas las cantidades consideradas en el problema, observables o no. Este modelo, debe ser consistente con la información que se tenga respecto al problema en estudio y la manera en que se obtienen los datos.
- (ii) Se calcula e interpreta la “distribución posterior”, es decir, la distribución de probabilidad condicional de las cantidades no observadas que son de interés, dados los datos observados.
- (iii) Se evalúa el modelo y las implicaciones que pueden derivarse de la distribución posterior resultante determinando para ello si el modelo describe de manera adecuada a los datos, si las conclusiones son razonables y qué tan sensibles son los resultados finales a las hipótesis realizadas en i).

En caso de que se considere necesario, se altera o amplía el modelo repitiendo los pasos descritos, a fin de mejorar la aproximación obtenida.

De esta manera, en Estadística Bayesiana, la descripción formal del proceso de aprendizaje a partir de valores observados, unido a información previa que se tenga sobre el objeto de estudio, son el punto de partida para realizar inferencias, conclusiones o predicciones acerca de la siguiente observación que se obtendrá de la población.

El caso canónico de un análisis Bayesiano, se tiene al extraer una muestra aleatoria  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ , de una población con función de densidad  $\mathcal{P}(\mathbf{y}|\theta)$ , donde  $\mathbf{y}$  representa el vector de las observaciones  $y_i$ ,  $i = 1, \dots, n$ , independientes dado el valor de  $\theta$ , idénticamente distribuidas con densidad  $\mathcal{P}(y|\theta)$  donde  $\theta$  es el vector de cantidades no observables también conocidos como *parámetros* que caracterizan a la población.

A lo largo de este trabajo se utilizará  $\mathcal{P}(\cdot)$  para denotar densidades de probabilidad con los argumentos determinados por el contexto, al igual que en el caso de  $\mathcal{P}(\cdot)$ , que denota una función de distribución marginal. Los términos "distribución" y "densidad" se emplean indistintamente, salvo en los casos donde se indique explícitamente.

## 1.2. Distribución inicial

Al resolver un problema mediante procedimientos Bayesianos, se especifica primeramente la *distribución a priori*, o *distribución inicial* del parámetro  $\pi(\theta)$ . Tal distribución debe incluir y reflejar toda la información con la que se cuente al principio del estudio, aunque existan casos en que ésta no ha sido observada o comprobada empíricamente. Una vez definida la distribución  $\pi(\theta)$ , es posible mejorar nuestro conocimiento respecto a  $\theta$ , vía el Teorema de Bayes.

La importancia de la distribución inicial del parámetro  $\theta$ , radica en que permite incorporar información acerca de  $\theta$  al modelo. Esta información generalmente tiene un significado muy especial en cada problema específico, dado que generalmente se tiene algún conocimiento previo acerca de su valor.

### 1.3. Teorema de Bayes

Una vez realizado un experimento que produzca observaciones  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , la información que se tiene sobre el valor de  $\theta$  puede describirse a través de la distribución final  $P(\theta|\mathbf{y})$ .

Para poder realizar inferencias acerca de  $\theta$  dado  $\mathbf{y}$ , es necesario contar con un modelo que contenga la distribución de probabilidad conjunta para  $\theta$  y  $\mathbf{y}$ , que se pueda escribir como el producto de dos densidades: la distribución inicial  $\pi(\theta)$  y la distribución de la muestra,  $P(\mathbf{y}|\theta)$ . La distribución de la muestra o verosimilitud contiene información del experimento mismo, ya que se deriva de él y la densidad de probabilidad a priori incluirá en el modelo la información que se tiene de manera independiente al experimento, o que proviene de experimentos anteriores.

El Teorema de Bayes es fundamental en la Estadística Bayesiana y en este caso, para llegar a la densidad  $P(\theta|\mathbf{y})$ , ya que expresa claramente la forma en que las probabilidades cambian cuando se cuenta con cierta evidencia. Dicho en otras palabras, permite incorporar el aprendizaje de tal manera, que se actualiza el conocimiento sobre  $\theta$  con nueva información obtenida del "mundo real".

De manera formal, una vez definidas la distribución inicial  $\pi(\theta)$  y la distribución de la muestra  $P(\mathbf{y}|\theta)$ :

$$P(\theta, \mathbf{y}) = \pi(\theta) \cdot P(\mathbf{y}|\theta) \quad (1.1)$$

Condicionando al valor conocido de  $\mathbf{y}$ , y usando la regla de Bayes, se obtiene la densidad posterior:

$$P(\theta|\mathbf{y}) = \frac{P(\theta, \mathbf{y})}{P(\mathbf{y})} = \frac{\pi(\theta) \cdot P(\mathbf{y}|\theta)}{P(\mathbf{y})} \quad (1.2)$$

donde  $P(\mathbf{y}) = \sum_{\theta} \pi(\theta) \cdot P(\mathbf{y}|\theta)$ , sumando sobre todos los valores de  $\theta$  o bien,  $P(\mathbf{y}) = \int \pi(\theta) \cdot P(\mathbf{y}|\theta) d\theta$ , en el caso continuo. Llamaremos a  $P(\mathbf{y})$  la distribución predictiva.



La expresión (1.2) se conoce como teorema de Bayes.

## 1.4. Función de densidad posterior y función de verosimilitud

Como se mencionó en la sección 1.3, al resolver un problema mediante Estadística Bayesiana, se desea encontrar una función de probabilidad que describa a  $\theta$  dada cierta información, dicha función de probabilidad se conoce como *distribución a posteriori o final* y se denota como  $P(\theta|\mathbf{y})$ .

De acuerdo con el Teorema de Bayes, las funciones de densidad posteriores dependen de las iniciales y los resultados experimentales u observados. Los datos experimentales u observados influyen en la función de probabilidad posterior únicamente a través de la función de densidad que los describe, llamada *verosimilitud* y denotada como  $P(\mathbf{y}|\theta) = L(\mathbf{y}|\theta)$ . Las verosimilitudes dependen solamente de la información con la que se cuenta a través del experimento mientras que las distribuciones a priori dependerán sólo de la evidencia con la que se dispone al principio del análisis.

Una forma equivalente de (1.2) omite la densidad  $P(\mathbf{y})$ , que no depende del valor de  $\theta$ . Así, para  $\mathbf{y}$  fija,  $P(\mathbf{y})$  se puede considerar como una constante, obteniéndose la *densidad posterior no normalizada*,

$$P(\theta|\mathbf{y}) \propto \pi(\theta) \cdot L(\mathbf{y}|\theta) \quad (1.3)$$

Esta relación de proporcionalidad se puede expresar como igualdad siempre que se encuentre y agregue la constante  $k = P(\mathbf{y})$ . Esta constante se puede hallar a partir del hecho de que  $\sum_{\theta} P(\theta|\mathbf{y}) = 1$ , o bien en el caso continuo,  $\int_{\theta} P(\theta|\mathbf{y}) d\theta = 1$ , con  $\mathbf{y}$  fija.

## 1.5. Familias conjugadas

La resolución de cualquier problema al usar Estadística Bayesiana depende en gran parte de la distribución inicial que se proponga, por lo que es muy importante que ésta refleje lo mejor posible el conocimiento inicial que se tenga del parámetro  $\theta$ . Además sería deseable que la forma de  $\pi(\theta)$  sea "tratable" en el sentido de que cuando se calculen esperanzas o se obtenga  $P(\theta|\mathbf{y})$ , los procedimientos no sean complicados.

Una forma de explotar las características conocidas de una familia  $\mathcal{F}$  de densidades para lograr la "tratabilidad" descrita, sería escoger  $\mathcal{F}$  tal que si  $\pi(\theta) \in \mathcal{F}$ , entonces después de usar el teorema de Bayes  $P(\theta|\mathbf{y}) \in \mathcal{F}$ .

### Definición 1.1 Familias conjugadas

Sea  $\mathbf{y}$  una variable aleatoria cuya distribución está dada por

$$P(\mathbf{y}|\theta) \in \mathcal{G} \equiv \{P(\cdot|\theta) : \theta \in \Theta \subseteq \mathbb{R}^n\}.$$

Se dice que una familia de distribuciones  $\mathcal{F} = \{P(\theta) : \theta \in \Theta\}$ , es conjugada para  $\theta$  (o para  $\mathcal{G}$ ), si cuando  $P(\theta) \in \mathcal{F}$  y  $P(\mathbf{y}|\theta) \in \mathcal{G}$  entonces  $P(\theta|\mathbf{y}) \in \mathcal{F}$ .

Si  $\mathcal{F}$  es una familia conjugada para  $P(\mathbf{y}|\theta)$ , entonces  $P(\theta)$  y  $P(\theta|\mathbf{y})$  tienen la misma forma algebraica y difieren de una distribución a otra únicamente por constantes.

Por ejemplo, supóngase que  $P(\mathbf{y}|\theta) \sim \text{Bernoulli}(\theta)$  y  $\pi(\theta) \sim \text{Beta}(\alpha, \beta)$ , esto es:

$$\begin{aligned} P(\mathbf{y}|\theta) &\sim \text{Bin}(1, \theta) \\ \pi(\theta) &\propto \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} \end{aligned} \tag{1.4}$$

entonces, la verosimilitud está dada por

$$L(\mathbf{y}|\theta) = \prod_{i=1}^n P(y_i|\theta) = \theta^{\sum_{i=1}^n y_i} \cdot (1-\theta)^{n-\sum_{i=1}^n y_i}$$

(1.5)

y por tanto se obtiene, que la distribución final para  $\theta$  dados los datos es

$$\begin{aligned} P(\theta|\mathbf{y}) &\propto \theta^{\sum_{i=1}^n y_i} \cdot (1-\theta)^{n-\sum_{i=1}^n y_i} \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} \\ P(\theta|\mathbf{y}) &\propto \theta^{t_n+\alpha-1} \cdot (1-\theta)^{n-t_n+\beta-1} \\ P(\theta|\mathbf{y}) &\propto \text{Beta}(\alpha^*, \beta^*) \end{aligned}$$

donde  $\alpha^* = t_n + \alpha$ ,  $\beta^* = n - t_n + \beta$  y  $t_n = \sum_{i=1}^n y_i$ .

### Familias conjugadas de la distribución exponencial

Se dice que una distribución de probabilidad,  $P(\mathbf{y}|\theta)$  pertenece a la familia exponencial de distribuciones, con un parámetro, si

$$P(\mathbf{y}|\theta) = a(\mathbf{y}) \exp\{\phi(\theta)t(\mathbf{y}) + b(\theta)\} \quad (1.6)$$

La importancia de la familia exponencial se debe al hecho de que un buen número de distribuciones usadas en la práctica pertenecen a ella, tales como:

a) La distribución Normal  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida, donde

$$\phi(\theta) = \frac{\theta}{\sigma^2}, \quad t(\mathbf{y}) = y, \quad b(\theta) = -\frac{\theta^2}{2\sigma^2}, \quad a(y) = \exp\left\{-\frac{y^2}{2\sigma^2}\right\} \quad (1.7)$$

b) La distribución Binomial  $\text{Bin}(n, \theta)$ , donde

$$\phi(\theta) = \log\left(\frac{\theta}{1-\theta}\right), \quad t(\mathbf{y}) = y, \quad b(\theta) = n \log(1-\theta), \quad a(y) = \binom{n}{y}$$

c) Distribución Exponencial  $\text{Exp}(\theta)$

$$\phi(\theta) = -\theta, \quad t(\mathbf{y}) = y, \quad b(\theta) = \log \theta, \quad a(y) = 1$$

d) La distribución Poisson  $Poi(\theta)$

$$\phi(\theta) = \log \frac{\theta}{1-\theta}, \quad t(y) = y, \quad b(\theta) = -\log \theta, \quad a(y) = 1$$

Algunas familias de distribuciones importantes que no pertenecen a la exponencial son la Uniforme y la t-Student.

Las familias exponenciales que consideran distribuciones con un parámetro desconocido se conocen como *familias exponenciales uniparametrales*. Una familia de distribuciones se llama *familia exponencial biparametral* si tiene la forma

$$P(y|\theta, \eta) = a(y)h(\theta, \eta) \exp\{\phi(\theta, \eta)t(y) + u(y)\chi(\theta, \eta)\}$$

o de forma equivalente, si la verosimilitud de las  $n$  observaciones independientes e idénticamente distribuidas,  $y_i$ ,  $i = 1, 2, \dots, n$  puede escribirse como

$$L(\theta, \eta) = h(\theta, \eta)^n \exp\left\{\sum_{i=1}^n t(y_i)\phi(\theta, \eta) + \sum_{i=1}^n u(y_i)\chi(\theta, \eta)\right\}.$$

Es importante notar que el vector  $(\sum_{i=1}^n t(y_i), \sum_{i=1}^n u(y_i))$ , es una estadística suficiente para  $(\theta, \eta)$ . La familia de densidades conjugadas para esta verosimilitud tiene la forma

$$P(\theta, \eta) \propto h(\theta, \eta)^\nu \exp\{\tau\phi(\theta, \eta) + \nu\xi(\theta, \eta)\}.$$

En la práctica existen pocos casos en los que se consideren desconocidos los dos parámetros de una distribución que pertenezca a la familia exponencial, el caso más importante es el de la distribución Normal  $(\mu, \sigma^2)$ .

De acuerdo con Lee (1989), la idea de usar familias exponenciales puede extenderse fácilmente al caso  $k$ -parametral, a partir de la manera en que se definieron aquí las familias exponenciales biparametrales.

### Familia conjugada natural

Una de las razones por las cuales las familias exponenciales son de utilidad, es porque poseen familias conjugadas *naturales* de distribuciones iniciales. Esto significa que si la función de verosimilitud  $L(\mathbf{y}|\theta)$  pertenece a la familia exponencial, se puede proponer que la distribución inicial  $\pi(\theta)$  tenga la misma estructura algebraica, pero sustituyendo estadísticas suficientes y tamaños de muestra por hiperparámetros  $\alpha$ . Esto último quiere decir que en algunos casos  $\pi(\theta)$  puede depender a su vez de una cantidad no observable  $\alpha$ , de forma que  $\pi(\theta) = P(\theta|\alpha)$  donde se asignará a su vez una distribución inicial para  $\alpha$ . Al utilizar el teorema de Bayes, se debe estudiar para qué valores de estos hiperparámetros la distribución final (miembro de la familia exponencial), es propia.

### Ejemplo. Familias conjugadas con distribución Normal

Sea  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  una muestra de observaciones independientes e idénticamente distribuidas de  $y \sim N(\theta, \sigma^2)$  con  $\sigma^2$  conocida y  $\pi(\theta)$  la distribución inicial del parámetro, donde  $\theta \sim N(\mu_0, \tau_0^2)$  con los hiperparámetros  $\mu_0, \tau_0^2$  conocidos. La distribución final será:

$$\begin{aligned}
 P(\theta|\mathbf{y}) &\propto \pi(\theta) \cdot L(\mathbf{y}|\theta) \\
 &= \pi(\theta) \cdot \pi_{i=1}^n P(y_i|\theta) \\
 &\propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right) \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta)^2\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\tau_0^2}(\theta - \mu_0)^2 + \frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \theta)^2\right]\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[\frac{\theta^2 - 2\theta\mu_0 + \mu_0^2}{\tau_0^2} + \frac{\sum_{i=1}^n y_i^2 - 2\theta\sum_{i=1}^n y_i + \sum_{i=1}^n \theta^2}{\sigma^2}\right]\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right]\left\{\theta^2 - 2\theta\frac{\frac{\mu_0}{\tau_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} + \frac{\frac{\mu_0^2}{\tau_0^2} + \frac{\sum_{i=1}^n y_i^2}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}\right\}\right)
 \end{aligned}$$

multiplicando el tercer término en el exponente por

$$\frac{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad (1.8)$$

y desarrollando (A),

$$\begin{aligned} &\propto \exp \left( -\frac{1}{2} \left[ \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right] \left\{ \theta^2 - 2\theta \frac{\left[ \frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2} \right]}{\left[ \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right]} + \frac{\left[ \frac{\mu_0^2}{\tau_0^2} + \frac{n\bar{y}^2}{\sigma^2} \right]}{\left[ \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right]^2} - \left[ \frac{2n\mu_0\bar{y}}{\sigma^2\tau_0^2} + \frac{2n}{(\sigma^2)^2} \sum_{i=1}^{n-1} y_i \sum_{j=i+1}^n y_j \right] \right\} \right) \\ &\propto \exp \left( -\frac{1}{2} \left[ \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right] \left\{ [\theta - \mu_n]^2 - \left[ \frac{2n\mu_0\bar{y}}{\sigma^2\tau_0^2} + \frac{2n}{(\sigma^2)^2} \sum_{i=1}^{n-1} y_i \sum_{j=i+1}^n y_j \right] \right\} \right). \end{aligned}$$

Por lo tanto,

$$P(\theta|\mathbf{y}) = N(\mu_n, \tau_n^2)$$

donde

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad \text{y} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

El parámetro  $\tau$  corresponde a la varianza:  $\tau_0$  y  $\tau_n$  denotan la varianza de la distribución inicial y distribución final, respectivamente. La media de la distribución posterior se interpreta como el promedio ponderado de la media inicial y la media de las observaciones  $\mathbf{y}$ , con pesos proporcionales a las varianza.

## 1.6. Modelos Jerárquicos

Existen modelos estadísticos en los que el o los parámetros de las distribuciones que se involucran pueden su vez ser descritos por nuevos parámetros, llamados hiperparámetros.

Al incluir información sobre los parámetros mediante una distribución que incluye hiperparámetros surge de manera natural un orden en todos los parámetros que se consideran para el modelo general que se convierte entonces en un *modelo jerárquico*.

Los modelos jerárquicos son una herramienta útil cuando se desea reflejar la dependencia inter-parámetros y obtener así resultados más precisos al ajustar distribuciones a conjuntos de datos que exijan el empleo de muchos parámetros.

El siguiente ejemplo es un caso sencillo de modelos jerárquicos y puede encontrarse en Gelman (1995).

**Ejemplo.** Consideremos que se desea estimar el parámetro  $\theta$  y se cuenta con datos provenientes de

- (i) Un experimento reciente y,
- (ii) Series históricas de experimentos anteriores.

Suponemos que los datos de (i) y (ii) provienen de la misma población y que además las condiciones en que se realizaron todos los experimentos no cambian.

Si consideramos el caso particular del estudio sobre el desarrollo de tumores en ratas de laboratorio, nos interesa  $\theta =$  Probabilidad de tumor en una población de ratas que recibe una dosis "cero" de medicamento (grupo de control).

Al realizar el experimento actual se obtiene que 4 de 14 ratas desarrollan el tumor.

En general, el número de tumores  $N$  es modelado como una variable aleatoria Binomial dada la probabilidad del evento, es decir

$$P\{N = n|\theta\} = \binom{M}{n} \theta^n (1 - \theta)^{M-n},$$

donde  $M$  es el número de experimentos.

### Análisis del problema usando familias conjugadas y modelos jerárquicos.

Por conveniencia, elegimos una distribución inicial para  $\theta$  de la familia conjugada

$$\mathcal{F} = \{\pi(\theta|\alpha, \beta) : \pi(\theta|\alpha, \beta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}; (\alpha, \beta) \in \mathbf{R} \times \mathbf{R}\}$$

Entonces si se realiza un análisis bayesiano completo mediante un modelo jerárquico, podríamos considerar la siguiente estructura del mismo:

$$\begin{array}{ccccc} & \nearrow & \theta_1 & \longrightarrow & n_1 \\ (\alpha, \beta) & : & \vdots & & \vdots \\ & \searrow & \theta_k & \longrightarrow & n_k \end{array}$$

Los números  $\alpha$  y  $\beta$  que llamamos hiperparámetros, son parte de las cantidades desconocidas en nuestro modelo, por tanto se deberán considerar cuando pensemos en la distribución inicial

$$P(\theta, \alpha, \beta) = P(\theta_1, \theta_2, \dots, \theta_k, \alpha, \beta). \quad (1.9)$$

La parte "jerárquica" en estos modelos es que como  $(\alpha, \beta)$  es desconocido debe tener su propia distribución inicial  $P(\alpha, \beta)$ , así entonces

$$P(\theta, \alpha, \beta) = P(\alpha, \beta) \cdot P(\theta|\alpha, \beta)$$

es la distribución conjunta inicial y la distribución conjunta posterior de todos los parámetros es



$$\begin{aligned}
 \pi(\theta, \alpha, \beta | \mathfrak{U}) &\approx \pi(\alpha, \beta, \theta) L(\mathfrak{U} | \alpha, \beta, \theta) \\
 &= \pi(\alpha, \beta, \theta) L(\mathfrak{U} | \theta) \\
 &= \pi(\alpha, \beta) P(\theta | \alpha, \beta) L(\mathfrak{U} | \theta)
 \end{aligned}$$

La primera igualdad se cumple pues la distribución de la muestra  $L(\mathfrak{U} | \alpha, \beta, \theta)$  depende sólo de  $\theta$ , es decir que los hiperparámetros  $\alpha, \beta$  afectan a  $\mathfrak{U}$  sólo mediante  $\theta$ .

Entonces

$$\begin{aligned}
 \pi(\theta, \alpha, \beta | \mathfrak{U}) &\propto \pi(\alpha, \beta) \times \prod_{j=1}^k \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1} (1 - \theta_j)^{\beta-1} \\
 &\times \prod_{j=1}^k \theta_j^{n_j} (1 - \theta_j)^{N_j - n_j}
 \end{aligned} \tag{1.10}$$

Los componentes de  $\theta$  tienen distribuciones posteriores independientes dado el valor de  $(\alpha, \beta)$ , que son densidades Beta( $\alpha, \beta$ ) y su distribución conjunta es

$$\begin{aligned}
 P(\theta | \alpha, \beta, \mathfrak{U}) &= \frac{P(\theta, \mathfrak{U}, \alpha, \beta)}{P(\alpha, \beta, \mathfrak{U})} \\
 &= \frac{\pi(\alpha, \beta) P(\theta | \alpha, \beta) L(\mathfrak{U} | \theta, \alpha, \beta)}{P(\alpha, \beta, \mathfrak{U})} \\
 &\propto P(\alpha, \beta) \times L(\mathfrak{U} | \theta, \alpha, \beta) \\
 &\propto P(\alpha, \beta) \times L(\mathfrak{U} | \theta) \\
 &\propto \prod_{j=1}^k \theta_j^{\alpha+n_j-1} (1 - \theta_j)^{\beta+N_j-n_j-1}
 \end{aligned}$$

por lo tanto

$$P(\theta|\alpha, \beta, \mathbf{u}) = \prod_{j=1}^k \frac{\Gamma(\alpha + \beta + N_j)}{\Gamma(\alpha + n_j)\Gamma(\beta + N_j - n_j)} \theta_j^{\alpha+n_j-1} (1 - \theta_j)^{\beta+N_j-n_j-1}.$$

Finalmente, para obtener la distribución marginal posterior de  $(\alpha, \beta)$  notamos que

$$P(\alpha, \beta|\mathbf{u}) = \frac{\pi(\theta, \alpha, \beta|\mathbf{u})}{P(\theta|\alpha, \beta, \mathbf{u})} \quad (1.11)$$

y entonces

$$P(\alpha, \beta|\mathbf{u}) \propto \pi(\alpha, \beta) \times \prod_{j=1}^k \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha + n_j)\Gamma(\beta + N_j - n_j)}{\Gamma(\alpha + \beta + N_j)} \quad (1.12)$$

El producto en (1.12) no se puede especificar analíticamente, pero puede calcularse para valores específicos de  $(\alpha, \beta)$  usando rutinas para calcular la función Gamma.

En Gelman et al. (1995) se presenta una discusión sobre la distribución a priori de los hiperparámetros  $\pi(\alpha, \beta)$ , así como de la "distribución predictiva", es decir, la distribución para valores futuros de  $\mathbf{u}$ .

## 1.7. Algoritmos de Metrópolis-Hastings y Gibbs

### Descripción de Métodos Monte Carlo vía Cadenas de Markov

Los métodos de Monte Carlo vía Cadenas de Markov sirven para obtener muestras de distribuciones que no son fáciles de simular en primera instancia. En muchos casos estas distribuciones pueden ser multivariadas o con forma algebraica complicada, como sucede comúnmente con las distribuciones posteriores en los modelos estadísticos bayesianos. Con las muestras simuladas, es posible estimar algunas cantidades de interés como valores esperados o probabilidades correspondientes al fenómeno bajo estudio.

En Gamerman (1997), Gutiérrez-Peña (1997) y Gelman et al. (1995) se pueden encontrar descripciones detalladas y aplicaciones de estos métodos. Las bases teóricas pueden consultarse en Besag y Green (1993) y Smith y Roberts (1993).

La idea básica es construir una cadena de Markov  $\theta^{(1)}, \theta^{(2)}, \dots$  que sea fácil de simular y cuya distribución de equilibrio  $P(\theta)$  corresponda a la distribución final que nos interesa. La siguiente proposición da la base de los métodos de Monte Carlo vía cadenas de Markov.

**Proposición 1.1** Sea  $\theta^{(1)}, \theta^{(2)}, \dots$  una cadena de Markov homogénea, irreducible, y aperiódica, con espacio de estados  $\Theta \subset \mathbb{R}^d$  y distribución de equilibrio  $p(\theta) = IP(\theta | y)$ . Entonces, cuando  $t \rightarrow \infty$ ,

$$(i) \theta^{(t)} \xrightarrow{d} \theta, \text{ donde } \theta \sim IP(\theta | y);$$

$$(ii) \frac{1}{t} \sum_{k=1}^t g(\theta^{(k)}) \longrightarrow E[g(\theta) | y].$$

Donde  $\xrightarrow{d}$  denota convergencia en distribución de variables aleatorias de acuerdo con la siguiente definición.

**Definición 1.2** *Convergencia en Distribución.*

Se dice que  $(X_n)$  converge en distribución o converge débilmente a la variable aleatoria  $X$  ( $X_n \xrightarrow{d} X$ ) si para todo punto de continuidad  $x$  de la distribución  $F$

$$F_n(x) \longrightarrow F(x) \text{ cuando } n \rightarrow \infty$$

es decir,

$$F_n(x) = IP(X_n \leq x) \longrightarrow F(x) = IP(X \leq x)$$

Entonces, si nuestro objetivo es obtener una muestra  $\theta_1, \theta_2, \dots, \theta_k$  de la distribución  $IP(\theta | y)$ , la proposición 1.1 nos dice que una forma sería construir una cadena de Markov

$\{\theta^{(l)}\}_l$  con las propiedades que se citan, particularmente como ya se mencionó, tal que su distribución de equilibrio sea la misma que aquella de donde se requieren muestras. Entonces si  $l^*$  es suficientemente grande como para que la convergencia en la proposición 1.1 se haya dado, podemos proponer como muestra a

$$\{\theta_1, \theta_2, \dots, \theta_k\} = \{\theta^{l^*+1}, \theta^{l^*+2}, \dots, \theta^{l^*+k}\}$$

### Algoritmo de Metropolis-Hastings

El objetivo de este algoritmo es construir una cadena de Markov apropiada definiendo las probabilidades de transición de la siguiente manera.

Sea  $Q(\theta^*, \theta)$  una distribución de transición (arbitraria) y definamos

$$\alpha(\theta^*, \theta) \approx \min \left\{ \frac{P(\theta^* | y) Q(\theta, \theta^*)}{P(\theta | y) Q(\theta^*, \theta)}, 1 \right\}$$

Dado un valor inicial  $\theta^{(0)}$ , los pasos a repetir en la  $l$ -ésima iteración del algoritmo son:

1. Generar una observación  $\theta^*$  de  $Q(\theta^*, \theta^{(l)})$ .
2. Generar una variable  $u \sim U(0, 1)$
3. Si  $u \leq \alpha(\theta^*, \theta^{(l)})$ , entonces se toma  $\theta^{(l+1)} = \theta^*$ , en caso contrario la cadena no se mueve, es decir  $\theta^{(l+1)} = \theta^{(l)}$ .

Este procedimiento genera una cadena de Markov  $\{\theta^{(l)}\}_l$  con función de transición

$$P(\theta^{(l+1)}, \theta^{(l)}) = \alpha(\theta^{(l+1)}, \theta^{(l)}) Q(\theta^{(l+1)}, \theta^{(l)}).$$

Notemos que la probabilidad de aceptar la nueva muestra  $\alpha(\theta^*, \theta)$  depende de  $P(\theta | y)$  a través de un cociente, por tanto no se requiere conocer la constante de normalización para  $P(\theta | y)$ .

### Muestreo de Gibbs

Al igual que el algoritmo de Metropolis-Hastings, el algoritmo de Gibbs permite simular una cadena de Markov  $\{\theta^{(t)}\}_t$  con distribución de equilibrio  $P(\theta | \mathbf{y})$ . Las densidades

$$\begin{aligned} & P(\theta_1 | \theta_2, \dots, \theta_d, \mathbf{y}) \\ & \vdots \\ & P(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d, \mathbf{y}) \quad (i = 2, \dots, k-1) \\ & \vdots \\ & P(\theta_d | \theta_2, \dots, \theta_{d-1}, \mathbf{y}) \end{aligned}$$

se llaman densidades condicionales completas. En el contexto de la inferencia estadística Bayesiana estas densidades se pueden identificar usando la forma de la distribución final  $P(\theta | \mathbf{y})$  como

$$P(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d, \mathbf{y}) = \frac{P(\theta | \mathbf{y})}{\int_{\mathcal{R}} P(\theta | \mathbf{y}) d\theta_i}; \quad (1.13)$$

Dado un valor inicial  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)})$ , el algoritmo de Gibbs simula una cadena en la cual  $\theta^{(t+1)}$  se obtiene de  $\theta^{(t)}$  siguiendo los pasos:

$$\begin{aligned} \text{generar una observación} & \quad \theta_1^{(t+1)} \quad \text{de} \quad P(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_d^{(t)}, \mathbf{y}), \\ \text{generar una observación} & \quad \theta_2^{(t+1)} \quad \text{de} \quad P(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_d^{(t)}, \mathbf{y}), \\ & \quad \vdots \\ \text{generar una observación} & \quad \theta_d^{(t+1)} \quad \text{de} \quad P(\theta_d | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{d-1}^{(t+1)}, \mathbf{y}) \end{aligned}$$

La sucesión  $\theta^{(1)}, \theta^{(2)}, \dots$  así obtenida es entonces una realización de una cadena de Markov cuya distribución de transición está dada por

$$P(\theta^{(t+1)} | \theta^{(t)}) = \prod_{i=1}^d P(\theta_i^{(t+1)} | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)}, \mathbf{y})$$

### Convergencia

Supongamos que se desea generar una muestra de tamaño  $N$  de la distribución  $\mathcal{P}(\theta | \mathbf{y})$ . Si para cada uno de  $N$  valores iniciales  $\theta_1^{(0)}, \dots, \theta_N^{(0)}$  se aplica alguno de los algoritmos descritos anteriormente, entonces, de acuerdo con la Proposición 1.1 (i), después de cierto número de iteraciones  $T$  suficientemente grande, los valores  $\theta_1^{(T)}, \dots, \theta_N^{(T)}$  pueden considerarse como una muestra de tamaño  $N$  de la distribución final de  $\theta$ . Alternativamente podemos generar una sola cadena y tomar los valores  $\theta^{(T+K)}, \theta^{(T+2K)}, \dots, \theta^{(T+NK)}$  como una muestra de  $\mathcal{P}(\theta | \mathbf{y})$ , donde  $K$  se elige de manera que la correlación entre las observaciones sea pequeña. En general no es fácil determinar en qué momento la(s) cadena(s) ha(n) convergido. Un método empírico comúnmente utilizado, basado en la Proposición 1.1, consiste en graficar los promedios ergódicos de algunas funciones de  $\theta$  contra el número de iteraciones y elegir el valor de  $T$  a partir del cual las gráficas se estabilizan. En este caso es frecuente omitir los primeros valores de la(s) cadena(s) al calcular los promedios ergódicos. La idea de este *periodo de calentamiento* es permitir que la(s) cadena(s) salga(n) de una primera fase de inestabilidad. En el caso particular del muestreo de Gibbs la velocidad de convergencia depende frecuentemente de la correlación de los componentes del vector  $\theta$  bajo la distribución final  $\mathcal{P}(\theta | \mathbf{y})$ : entre más alta la correlación más lenta será la convergencia.

A fin de probar la convergencia del algoritmo se han propuesto métodos informales basados en técnicas gráficas. Uno de ellos es el graficar los promedios ergódicos de la cadena simulada que se definen a continuación

#### Definición 1.3 Promedio Ergódico

Dada una función  $h : \mathcal{R}^d \rightarrow \mathcal{R}$  tal que

$$E \mathcal{P}_{(\theta|\mathbf{y})}[h(\theta)] < \infty,$$

el promedio ergódico  $h$  se define como

$$\bar{h}_n \equiv \frac{1}{n} \sum_{t=1}^n h(\theta^{(t)}).$$

De acuerdo con la proposición 1.1 (ii), el comportamiento asintótico de  $\bar{h}_n$  es tal que

$$\bar{h}_n \rightarrow \mathbf{E} P_{(\theta|y)}[h(\theta)], \quad \text{cuando } n \rightarrow \infty.$$

Entonces, si graficamos  $\bar{h}_n$  contra  $n$  podemos tener una idea del comportamiento asintótico de la cadena despues de cierto número de iteraciones sucesivas y si además se efectúa un análisis empírico de este tipo de gráficos será posible proponer el tiempo  $T$  a partir del cual la trayectoria de los promedios ergódicos se estabiliza.

### 1.7.1. Un ejemplo de Muestreo de Gibbs

El siguiente ejemplo puede encontrarse en Carlin y Gelfand (1991) o en Gamerman (1997).

Una Compañía de seguros tiene una muestra de los números de ocurrencias de cierto tipo de siniestros ocurridos en cada semana de un año

$$y_1, y_2, \dots, y_{52}$$

La compañía tiene la sospecha de que existe un fenómeno ambiental que afecta la ocurrencia de estos siniestros. Se lo ha pedido a los asesores estadísticos de la compañía que verifiquen si existe un cambio en la tasa de ocurrencia de los siniestros a lo largo del año.

#### Formulación de un modelo estocástico

Supóngase que el proceso del número de siniestros ocurridos hasta el tiempo  $t$  se puede modelar con un proceso de Poisson no homogéneo. No es el objeto de este ejemplo describir cómo verificar este supuesto, pero existen metodologías estadísticas con este fin.

Entonces

$$N_t = \text{número siniestros ocurridos en el intervalo de tiempo}(0, t].$$

$N_t$  es una variable aleatoria con distribución Poisson con  $N_0 = 0$  y media

$$E\{N_t\} = \eta t, \quad (1.14)$$

donde la tasa de ocurrencia de los eventos es  $\eta$  y en el caso no homogéneo  $\eta$  puede ser función del tiempo  $\eta = \eta(t)$ .

Si asumimos este modelo entonces

$$y_1 \equiv N_1 - N_0 = N_1$$

y

$$y_t = N_t - N_{t-1}, \quad \text{para } t = 2, \dots, 52.$$

recordemos que para el proceso de Poisson

$$N_t - N_{t-1} \sim \text{Poisson}(\eta).$$

Queremos verificar si existe  $t^* \in \{1, 2, \dots, 52\}$  tal que el valor de  $\eta$  cambia en forma significativa, es decir, verificar si el proceso del número de siniestros es no homogéneo.

### Heterogeneidad en los valores de los parámetros

En el fenómeno de la destrucción de un inmueble, la idea de que la tasa de ocurrencia del siniestro para algunos inmuebles (concreto) no es la misma que para otros (partes de madera), no resulta inverosímil. Tampoco es de esperarse que la tasa sea la misma para diferentes regiones de una ciudad. Así pues  $\eta$  es una cantidad que "varía" y la forma más natural para cuantificar variabilidad e incertidumbre, es asignar distribuciones de probabilidad. En otras palabras, estamos proponiendo modelar a  $\eta$  como una variable aleatoria con una distribución dada.

Cuando se desconoce esta distribución, una forma de modelar a  $\eta$  como variable aleatoria es usar un elemento de una familia paramétrica

$$P(\eta|\alpha, \beta),$$



con valores de hiperparámetros  $\alpha$  y  $\beta$  que expresen nuestra incertidumbre acerca del comportamiento de  $\eta$ .

Por otra parte, si nuestro conocimiento de  $\eta$  es mayor, entonces podemos expresarlo modificando los valores de  $\alpha$  y  $\beta$ .

En cuanto al punto en el tiempo  $t^*$  donde cambia la tasa de ocurrencia  $\eta$ , si no existe evidencia de otro esquema, podemos pensar en que  $t^*$  no tiene por qué ser determinístico y estar fijo año con año, es decir, podemos modelarlo como aleatorio a su vez. Entonces  $t^*$  es una variable aleatoria con distribución en el rango de valores  $\{1, 2, \dots, n\}$ . De nueva cuenta si no tenemos razones para pensar que es más probable que  $t^*$  tome cierto valor, podemos asignarle una distribución uniforme sobre  $\{1, 2, \dots, n\}$ . Si hay evidencia para pensar en otro modelo entonces asignamos otra distribución sobre los enteros  $\{1, 2, \dots, n\}$ .

### Modelo

Las consideraciones expuestas arriba se pueden escribir en el siguiente modelo.

Se tiene una muestra aleatoria  $Y_1, Y_2, \dots, Y_n$ ,  $n = 52$  y una variable aleatoria  $t^*$  con valores sobre los enteros  $1, 2, \dots, n$ . Entonces, dado el valor de  $t^*$ , para  $t = 1, 2, \dots, t^*$ ,  $y_t$  se distribuye Poisson con parámetro de media  $\eta = \lambda$ , pero para  $t = t^* + 1, t^* + 2, \dots, n$ ,  $y_t$  se distribuye Poisson con parámetro de media  $\eta = \phi$ . Podemos suponer que inicialmente  $\lambda, \phi$  y  $t^*$  son independientes y con distribuciones Gamma( $\alpha_1, \beta_1$ ), Gamma( $\alpha_2, \beta_2$ ) y Uniforme en  $1, 2, \dots, n$ , respectivamente. Los valores de  $\alpha_1$  y  $\alpha_2$  se pueden fijar como 1 y los valores de  $\beta_1$  y  $\beta_2$  se pueden fijar como 0,01. Estos valores corresponden a distribuciones iniciales para  $\lambda$  y  $\phi$  con poca concentración de probabilidad en intervalos  $[0, a]$  con  $a$  cercana a 0. De esta forma se expresa poco conocimiento respecto a las distribuciones iniciales de  $\lambda$  y  $\phi$ . Este tipo de distribuciones "vagas" o "difusas" sobre el rango de valores de los parámetros suelen llamarse *no informativas*. Al establecer distribuciones a priori no informativas se permite que los datos jueguen el papel más importante en las conclusiones a posteriori, a través de la verosimilitud, sin que influya demasiado el conocimiento inicial.

## Uso del Gibbs sampler

Podemos entonces usar el esquema de Gibbs y la información de la muestra  $y_1, y_2, \dots, y_n$ , para obtener una muestra de

$$\theta \equiv (\lambda, \phi, t^*) \in (0, \infty) \times (0, \infty) \times \{1, 2, \dots, n\}$$

como sigue.

El teorema de Bayes establece que

$$P(\theta | \mathbf{y}) = \frac{1}{c} P(\mathbf{y} | \theta) P(\theta), \quad (1.15)$$

donde  $c$  es una constante (c.r. a  $\theta$ ) tal que  $P(\theta | \mathbf{y})$  en (1.15) satisface

$$\int_{\theta \in \Theta} P(\theta | \mathbf{y}) d\theta = 1.$$

Para nuestro caso

$$P(\mathbf{y} | \theta) = \prod_{i=1}^{t^*} \frac{e^{-\lambda} \lambda^{y_i}}{(y_i)!} \times \prod_{i=t^*+1}^n \frac{e^{-\phi} \phi^{y_i}}{(y_i)!} \quad (1.16)$$

y

$$P(\theta) = \left\{ \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda^{\alpha_1-1} e^{-\beta_1 \lambda} \right\} \left\{ \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \phi^{\alpha_2-1} e^{-\beta_2 \phi} \right\} \left\{ \frac{1}{n} \right\}. \quad (1.17)$$

Primero, usando la definición (1.13) se obtendrán las densidades condicionales completas

$$P(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d, \mathbf{y}),$$

donde en este caso

$$\theta = (\theta_1, \theta_2, \theta_3) = (\lambda, \phi, t^*).$$

Dadas (1.16) y (1.17), tenemos que la densidad condicional completa para  $\lambda$  es

$$P(\lambda | \phi, t^*, \mathbf{y}) = \frac{\frac{1}{c} P(\mathbf{y} | \theta) \cdot P(\theta)}{\int_0^{\infty} \frac{1}{c} P(\mathbf{y} | \theta) P(\theta) d\lambda} = \frac{P(\mathbf{y} | \theta) \cdot P(\theta)}{\underbrace{\int_0^{\infty} P(\mathbf{y} | \theta) P(\theta) d\lambda}_{(1)}}$$

Desarrollando (1),

$$\begin{aligned} & \int_0^{\infty} \prod_{i=1}^{t^*} \frac{e^{-\lambda} \lambda^{y_i}}{(y_i)!} \cdot \prod_{i=t^*+1}^n \frac{e^{-\phi} \phi^{y_i}}{(y_i)!} \cdot \left\{ \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda^{\alpha_1-1} e^{-\beta_1 \lambda} \right\} \left\{ \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \phi^{\alpha_2-1} e^{-\beta_2 \phi} \right\} \left\{ \frac{1}{n} \right\} d\lambda \\ &= \frac{1}{n} \cdot \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{e^{-\phi(n-t^*+\beta_2)} \phi^{\sum_{i=t^*+1}^n y_i + \alpha_2 - 1}}{\prod_{i=t^*+1}^n (y_i)!} \\ & \times \frac{1}{\prod_{i=1}^{t^*} (y_i)!} \underbrace{\int_0^{\infty} \lambda^{\sum_{i=1}^{t^*} y_i + \alpha_1 - 1} e^{-\lambda(t^*+\beta_1)} d\lambda}_{(2)}. \end{aligned}$$

Pero basta usar la forma de la densidad Gamma para determinar el valor de la integral (2) como

$$\int_0^{\infty} \lambda^{\sum_{i=1}^{t^*} y_i + \alpha_1 - 1} e^{-\lambda(t^*+\beta_1)} d\lambda = \frac{\Gamma(\sum_{i=1}^{t^*} y_i + \alpha_1)}{(t^* + \beta_1)^{\sum_{i=1}^{t^*} y_i + \alpha_1}}.$$

Por lo tanto,

$$\begin{aligned} \int_0^{\infty} P(\mathbf{y}|\theta) P(\theta) d\lambda &= \frac{1}{n} \cdot \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \cdot \frac{\Gamma(\sum_{i=1}^{t^*} y_i + \alpha_1)}{\prod_{i=1}^{t^*} (y_i)! (t^* + \beta_1)^{\sum_{i=1}^{t^*} y_i + \alpha_1}} \\ & \cdot \frac{e^{-\phi(n-t^*+\beta_2)} \phi^{\sum_{i=t^*+1}^n y_i + \alpha_2 - 1}}{\prod_{i=t^*+1}^n (y_i)!} \end{aligned}$$

y

$$P(\lambda|\phi, t^*, \mathbf{y}) = e^{-\lambda(t^*+\beta_1)} \lambda^{\sum_{i=1}^{t^*} y_i + \alpha_1 - 1} \times \frac{(t^* + \beta_1)^{\sum_{i=1}^{t^*} y_i + \alpha_1}}{\Gamma(\sum_{i=1}^{t^*} y_i + \alpha_1)}$$

de lo anterior,

$$\lambda|\phi, t^*, \mathbf{y} \sim \text{Gamma}(\alpha_1 + \sum_{i=1}^{t^*} y_i, \beta_1 + t^*).$$

Con un proceso análogo podemos deducir que la densidad condicional para  $\phi$  es

$$P(\phi|\lambda, t^*, \mathbf{y}) = e^{-\phi(n-t^*+\beta_2)} \phi^{\sum_{i=t^*+1}^n y_i + \alpha_2 - 1} \times \frac{(n-t^*+\beta_2)^{\sum_{i=t^*+1}^n y_i + \alpha_2 - 1}}{\Gamma(\sum_{i=t^*+1}^n y_i + \alpha_2)}$$

es decir

$$\phi|\lambda, t^*, \mathbf{y} \sim \text{Gamma}(\alpha_2 + \sum_{i=t^*+1}^n y_i, \beta_2 + n - t^*).$$

Por último, dado que  $t^*$  toma valores distretos, su densidad condicional completa se calcula como

$$P(t^*|\lambda, \phi, \mathbf{y}) = \frac{\frac{1}{c} P(\mathbf{y}|\theta) \cdot P(\theta)}{\sum_{t^*=1}^n \frac{1}{c} P(\theta|\mathbf{y}) P(\theta)} = \frac{P(\mathbf{y}|\theta) \cdot P(\theta)}{\underbrace{\sum_{t^*=1}^n P(\theta|\mathbf{y}) P(\theta)}_{(3)}}$$

La expresión para (3) es equivalente a

$$\frac{1}{n} \cdot \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \sum_{t^*=1}^n \frac{e^{-\lambda(t^*+\beta_1)} \lambda^{\sum_{i=1}^{t^*} y_i + \alpha_1 - 1} e^{-\phi(n-t^*+\beta_2)} \phi^{\sum_{i=t^*+1}^n y_i + \alpha_2 - 1}}{\prod_{i=1}^n (y_i)!}.$$

Para facilitar la notación usaremos  $l$  en lugar de  $t^*$  para escribir (3) como

$$\frac{1}{n} \cdot \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \sum_{l=1}^n \frac{e^{-\lambda(l+\beta_1)} \lambda^{\sum_{i=1}^l y_i + \alpha_1 - 1} e^{-\phi(n-l+\beta_2)} \phi^{\sum_{i=l+1}^n y_i + \alpha_2 - 1}}{\prod_{i=1}^n (y_i)!}$$

por lo tanto,

$$IP(t^* | \lambda, \phi, \mathbf{y}) = \frac{\lambda^{\alpha_1 + \sum_{t=1}^{t^*} \nu_{t-1}} e^{-(\beta_1 + t^*)\lambda} \phi^{\alpha_2 + \sum_{t=t^*+1}^n \nu_{t-1}} e^{-(\beta_2 + n - t^*)\phi}}{\sum_{t=1}^n \lambda^{\alpha_1 + \sum_{t=1}^t \nu_{t-1}} e^{-(\beta_1 + t)\lambda} \phi^{\alpha_2 + \sum_{t=t+1}^n \nu_{t-1}} e^{-(\beta_2 + n - t)\phi}} ; \quad t^* = 1, 2, \dots, n.$$

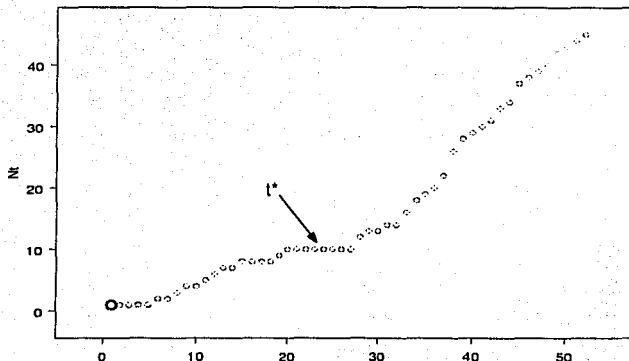
Una vez encontradas las densidades condicionales completas que tienen los parámetros, se debe proponer un valor inicial  $\theta^{(0)}$  adecuado para aplicar el muestreo de Gibbs. Es posible definir este valor a partir de la muestra que se tenga del fenómeno en estudio. A continuación se propone una forma de encontrar  $\theta^{(0)}$  para el problema aquí expuesto.

#### Metodología para encontrar $\theta^{(0)}$

En el gráfico del número de siniestros  $N_t$ , que se muestra enseguida, se observa que la media del proceso parece ser diferente antes y después de algún punto  $t^*$  entre  $t = 20$  y  $t = 30$ , lo que sugiere que un buen candidato a valor inicial de  $t^*$  se encuentra entre esos valores de  $t$ . Este puede ser el punto en el tiempo que divide el proceso en dos poblaciones que tienen la misma distribución con medias distintas.

Para probar la hipótesis anterior de manera formal, se aplicó una prueba Mann-Whitney para igualdad de medias a las observaciones  $y_t$  considerando  $t_0^* = 24$ . Esto es posible ya que si se supone que los siniestros ocurridos hasta el tiempo  $t$ , denotados como  $N_t$ , siguen una distribución Poisson, entonces las observaciones  $y_t = N_t - N_{t-1}$  serán independientes e idénticamente distribuidas como Poisson( $\eta$ ), cumpliéndose los supuestos de la prueba. El estadístico de prueba da suficiente evidencia para afirmar que las medias de las poblaciones son distintas y por lo tanto, inicialmente se supone que los datos provienen de dos procesos: el primero con media  $\eta = \lambda$  para  $t = 1$  a  $t = 24$  y el segundo con media  $\eta = \phi$  de  $t = 25$  a  $t = 52$ .

Ahora bien, para los valores iniciales de  $\lambda^{(0)}$  y  $\phi^{(0)}$ , se obtienen los estimadores de máxima verosimilitud basados en las muestras  $y_1, y_2, \dots, y_{24}$  y  $y_{25}, y_{26}, \dots, y_{52}$  y se calcula su valor:

Figura 1.1: Gráfico de  $N_t$ 

$$\hat{\lambda}^{(0)} = \sum_{t=1}^{24} \frac{N_t}{24} = \frac{5}{12}$$

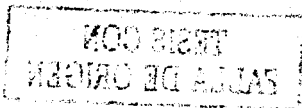
$$\hat{\phi}^{(0)} = \sum_{t=25}^{52} \frac{N_t}{28} = \frac{35}{28}$$

Por lo tanto, se propone usar como valor inicial  $\theta^{(0)} = (24, \frac{5}{12}, \frac{35}{28})$

### Generación de la muestra

Una vez propuesto el valor  $\theta^{(0)}$ , se efectuó un Muestreo de Gibbs usando algoritmos en S-Plus (Apéndice A).

Esta muestra inicial se hizo de tamaño  $n = 2,500$  y se eliminaron las primeras 500 observaciones a fin de trabajar con aquéllas que representarían valores de la cadena cer-



canos a la convergencia. Con los 2,000 valores restantes se realizó el histograma de frecuencias (figura 1.2) y el gráfico de promedios ergódicos (figura 1.3) a fin de comprobar la convergencia de la cadena  $\{\theta^{(t)}\}$  a la distribución límite, al analizar estos resultados puede considerarse que los valores de  $\lambda$  convergen a partir de  $n_\lambda \approx 700$ ,  $\phi$  cuando  $n_\phi \approx 800$  y  $t^*$  desde  $n_{t^*} \approx 900$ , por lo que sería adecuado emplear los valores de  $\{\theta^{(t)}\}$  para  $t \geq n \equiv \max\{n_\lambda, n_\phi, n_{t^*}\} = 900$  a fin de realizar análisis estadísticos usando los valores  $\{\theta^{(t)} : t \geq n\}$  como muestra de la distribución final de  $\theta$ .

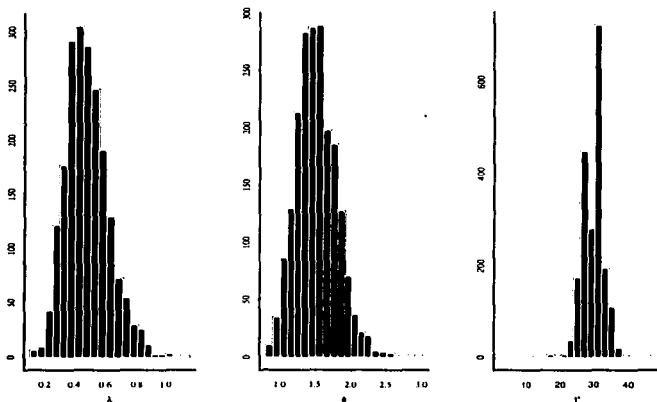
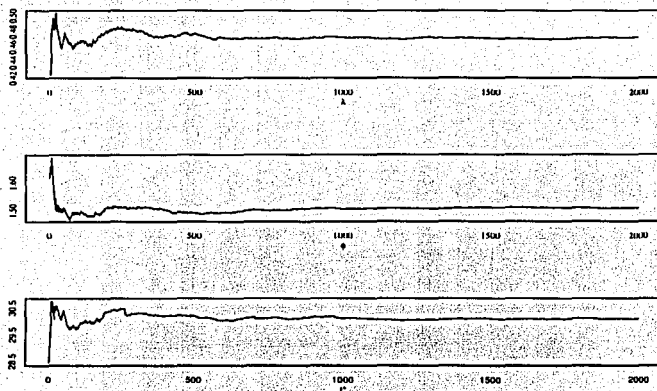


Figura 1.2: Histogramas de frecuencias,  $n = 2000$

Los valores de  $\theta$  que se generan mediante este algoritmo son correlacionados, lo cual puede ser inadecuado para algunos de los estudios que se desee realizar con ellos. Recordemos que en inferencia estadística la hipótesis de independencia de los elementos de una muestra se supone en varios procedimientos y metodologías. Este problema puede disminuirse analizando la autocorrelación de los valores generados para  $\theta$ , en este caso particular, los valores de  $\lambda$ ,  $\phi$  y  $t^*$ ; a partir de este análisis es posible determinar la forma de

TESIS CON  
FALLA DE ORIGEN

Figura 1.3: Promedios ergódicos,  $n = 2000$ 

elegir aquéllos elementos que presenten menor correlación.

En nuestro caso, después de obtener una muestra de tamaño  $n = 7,000$  se eliminaron los primeros 1,400 elementos y se llevó a cabo el análisis de correlación y elección de una muestra final con los últimos 5,600 elementos de la muestra  $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(7000)})$ . La autocorrelación de los parámetros (figura 1.4) sugiere elegir un valor cada 8 elementos (el máximo del mínimo rezago para el cual la autocorrelación de cada uno de los parámetros no es significativa). Con esta metodología se eligieron 700 elementos de la muestra y al repetir el cálculo de la autocorrelación se comprobó que disminuye de manera considerable (figura 1.5).



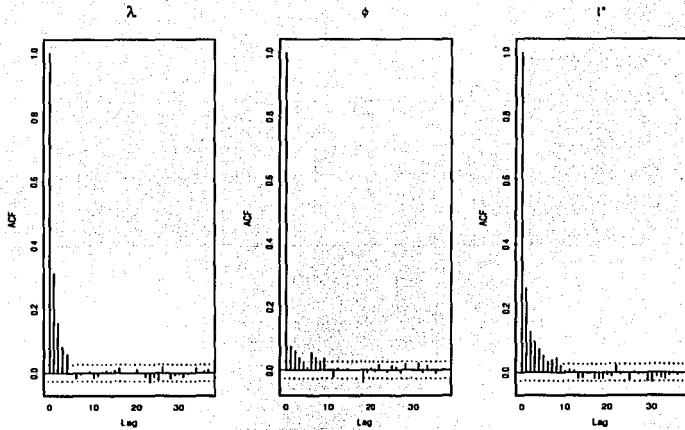


Figura 1.4: Autocorrelogramas de la muestra  $n=5600$

### Conclusiones

La forma no informativa en la que se especificaron las distribuciones iniciales para los parámetros permitió que los datos incorporaran la mayor información al modelo de tal manera que las distribuciones posteriores que se obtienen confirman la intuición a priori respecto al comportamiento de la tasa de ocurrencia del número de reclamaciones. En este sentido se puede decir que el modelo que se planteó es "adecuado".

A partir de los histogramas de frecuencias que se obtuvieron para la distribución final de los parámetros, sería posible construir "intervalos de credibilidad" a los que la mayor parte de las distribuciones posteriores pertenezcan. Este tipo de intervalos se conocen, en Estadística Bayesiana, como "región de densidad (posterior) máxima".

En este modelo por ejemplo, al observar la figura 1.2 podríamos establecer intervalos  $I_\lambda$  e  $I_\phi$  para los parámetros  $\lambda$  y  $\phi$  respectivamente, que sean regiones de densidad máxima

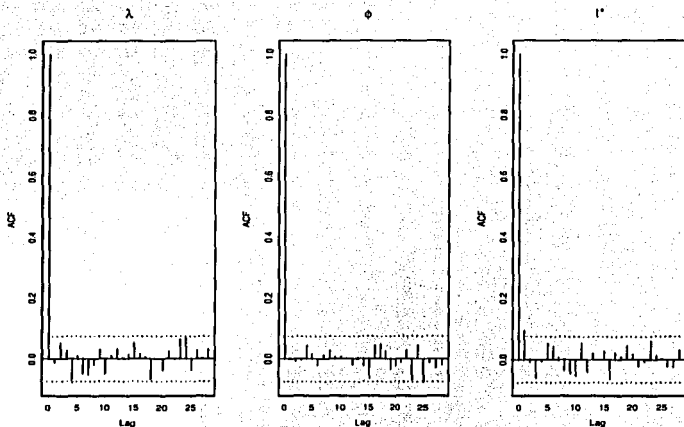
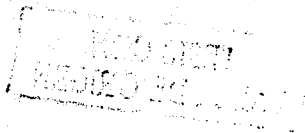


Figura 1.5: Autocorrelogramas con  $n=700$  observaciones elegidas

de 95 %, tales como  $I_\lambda = (0,1, 0,9)$  e  $I_\phi = (0,95, 2,3)$  para  $\phi$ .

Dado que  $I_\lambda \cap I_\phi = \emptyset$ , es decir, son intervalos ajenos, puede afirmarse que el proceso de reclamaciones es un proceso de Poisson no homogéneo, ya que existe  $t^*$  a partir de la cual la media del proceso,  $\eta$ , presenta cambios significativos.

Por lo tanto, sería recomendable que la compañía aseguradora considere cierta la influencia de los efectos ambientales en el comportamiento del número de siniestros anuales para realizar cualquier estudio posterior, como por ejemplo predicciones acerca del valor de los montos esperados de futuros siniestros, a lo largo de cierto número de años.



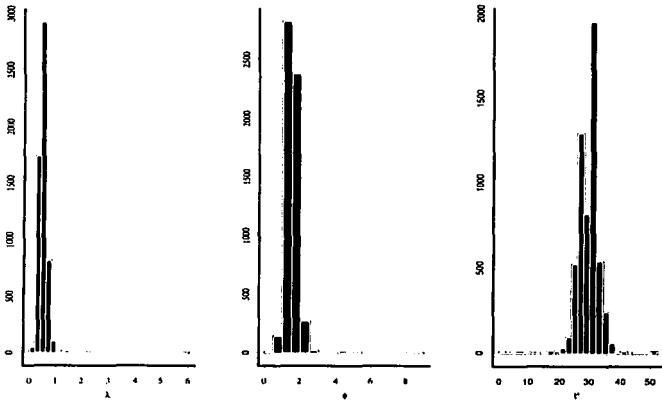


Figura 1.6: Histogramas de frecuencias para  $n=700$  observaciones elegidas

TESIS CON  
FALLA DE ORIGEN

# Capítulo 2

## TEORIA DE VALORES EXTREMOS

### 2.1. Introducción

En este capítulo se hará una breve revisión de algunos elementos de teoría de valores extremos.

En particular es de interés estudiar la distribución del máximo de los elementos de una muestra, cuando el tamaño de la muestra crece. Una parte clásica de esta teoría establece que bajo una renormalización, la distribución asintótica del máximo que se describe puede pertenecer a una de tres familias paramétricas. Estas se pueden englobar a su vez en la familia conocida como distribución de valor extremo generalizada (DVEG).

Por otro lado una parte moderna de la teoría establece la equivalencia de dos condiciones:

- i) La distribución del máximo muestral (renormalizado) se aproxima a la DVEG cuando el tamaño de la muestra crece.
- ii) Dada la muestra  $y_1, y_2, \dots, y_n$  y un valor de umbral  $u \in \mathcal{R}$ , la distribución condicional

$$P\{Y - u \leq y | Y > u\},$$

se puede aproximar utilizando una función de distribución conocida como distribución Pareto generalizada.

En esta tesis, usaremos el segundo de estos enfoques con el fin de analizar posteriormente un conjunto de reclamaciones de siniestros correspondientes a reaseguro.

Usando la teoría de valores extremos podemos estudiar cálculos probabilísticos e inferencias estadísticas referentes a valores máximos (o mínimos) en muestras aleatorias de observaciones y de procesos estocásticos en general.

Existen muchas aplicaciones en los campos de estudio de la ciencia y la ingeniería, que analizan eventos de rara ocurrencia pero de consecuencias significativas tales como fenómenos meteorológicos, resistencia de materiales, hidrología, grandes reclamaciones en seguros y posiciones de alto riesgo en operaciones financieras entre otros.

Llevar a la práctica la teoría de valores extremos no es fácil ya que tiene como objetivo modelar fenómenos que ocurren esporádicamente, de los cuales por consecuencia se tienen pocos datos, lo que hace más difícil la tarea de determinar qué tan pesadas deben ser las colas de la distribución, el punto en el que éstas empiezan y qué método de estimación de parámetros es el mejor.

## 2.2. Descripción básica de distribuciones de valores extremos

La teoría de valores extremos surge formalmente del estudio de distribuciones límite para máximos y mínimos de una sucesión de variables aleatorias. Para el caso específico de máximos, que se usará posteriormente en este trabajo, puede darse la siguiente descripción del problema.

Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias independientes e idénticamente distribuidas con distribución  $F$ , c

$$F(y) = P(Y_j \leq y) \quad , \quad j = 1, 2, \dots, n.$$

La distribución del máximo,  $M_n \equiv \max\{Y_1, Y_2, \dots, Y_n\}$  está dada por la  $n$ -ésima potencia de  $F$ :

$$F_{M_n}(y) \equiv \mathbb{P}(M_n \leq y) = \mathbb{P}(Y_1 \leq y; Y_2 \leq y; \dots; Y_n \leq y) = F^n(y). \quad (2.1)$$

A continuación se mencionan conceptos de convergencia de la Teoría de Probabilidad que se utilizarán a partir de esta sección.

Sea  $(\Omega, \mathcal{F}, \mathbb{P})$  un espacio de Probabilidad.

**Definición 2.1** *Convergencia en Probabilidad.*

Se dice que  $(X_n)$  converge en probabilidad a la variable aleatoria  $X$  ( $X_n \xrightarrow{P} X$ ) si la relación

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0, \quad \text{cuando } n \rightarrow \infty.$$

se cumple para toda  $\varepsilon$  positiva.

La convergencia en probabilidad implica convergencia en distribución, el inverso es cierto si y sólo si  $X = a$  casi seguramente para alguna constante  $a$ .

La relación  $X_n \xrightarrow{P} \infty$  se interpreta como  $\frac{1}{X_n} \xrightarrow{P} 0$ .

**Definición 2.2** *Convergencia casi segura*

Decimos que  $(X_n)$  converge casi seguramente (c.s.) o con probabilidad 1 a la variable aleatoria  $X$  ( $X_n \xrightarrow{c.s.} X$ ) existe  $A \in \mathcal{F}$  tal que para toda  $\omega \in A$

$$X_n(\omega) \rightarrow X(\omega), \quad \text{cuando } n \rightarrow \infty,$$

$$\text{y } \mathbb{P}(A) = 1.$$

Lo anterior quiere decir que

$$\mathbb{P}(X_n \rightarrow X) = \mathbb{P}(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

Nos interesa el comportamiento de la distribución de  $M_n$  cuando  $n$  se hace grande. Se observa que dado un valor de  $y$  en  $\mathbf{R}$ , se tiene que  $0 \leq F(y) \leq 1$  y por tanto,  $\lim_{n \rightarrow \infty} F_{M_n}(y)$  vale 0 ó 1. Entonces, aunque  $F_{M_n}$  se conozca vía la ecuación 2.1, es necesario hacer más para llevar a cabo un estudio asintótico de esta distribución. En particular, será necesario hacer una renormalización o estandarización de  $M_n$  para tal fin.

Se define el punto final a la derecha de la distribución  $F$  como

$$\omega_F \equiv \sup\{x \in \mathbf{R} : F(x) < 1\},$$

notemos que este valor puede ser infinito.

Para la variable aleatoria  $M_n$  tenemos que si  $x < \omega_F$  entonces

$$P(M_n \leq x) = F^n(x) < 1,$$

de donde

$$P(M_n \leq x) \rightarrow 0, \text{ cuando } n \rightarrow \infty.$$

Por otra parte, si  $\omega_F < +\infty$  y  $x \geq \omega_F$  entonces

$$P(M_n \leq x) = F^n(x) = 1,$$

de donde

$$P(M_n \leq x) \rightarrow 1 \text{ cuando } n \rightarrow \infty$$

Dadas estas consideraciones, si  $\omega_F < +\infty$ , entonces para  $\varepsilon > 0$

$$P(-\varepsilon \leq M_n - \omega_F \leq \varepsilon) = P(M_n \leq \varepsilon + \omega_F) - P(M_n < \omega_F - \varepsilon) \rightarrow 1 - 0 = 1,$$

cuando  $n \rightarrow \infty$ .

Por otra parte si  $\omega_F = +\infty$ , entonces para cada  $x \in \mathbf{R}$  cuando  $n \rightarrow \infty$

$$IP(M_n < x) \rightarrow 0.$$

Luego entonces, para cada  $x \in \mathbf{R}$ , cuando  $n \rightarrow \infty$

$$IP(M_n \geq x) = 1 - IP(M_n < x) \rightarrow 1.$$

De ambos casos podemos concluir que  $M_n \xrightarrow{P} \omega_F$ . Pero si por último notamos que  $M_n \leq M_{n+1}$  casi seguramente, podemos concluir que  $M_n \rightarrow \omega_F$ , cuando  $n \rightarrow \infty$  casi seguramente.

Dado que ahora sabemos que cuando  $n$  crece, la sucesión de máximos  $M_n$  es convergente al punto final a la derecha de  $F$ , volvemos a preguntarnos por la distribución de  $M_n$  pero ahora después de aplicarle una transformación.

Para este análisis asintótico, se "estandariza" a  $M_n$  de la siguiente manera.

Sean  $\{a_n\}$  y  $\{b_n\}$  sucesiones de constantes tales que  $a_n > 0$  para cada  $n$ , entonces

$$\begin{aligned} IP \left[ \frac{M_n - b_n}{a_n} \leq y \right] &= IP [M_n \leq a_n y + b_n] \\ &= F^n(a_n y + b_n) \end{aligned}$$

así, el problema se convierte en encontrar límites de la forma

$$H(y) = \lim_{n \rightarrow \infty} F^n(a_n y + b_n).$$

### Definición 2.3 Máximo dominio de atracción

Decimos que la variable aleatoria  $Y$  pertenece al máximo dominio de atracción de la distribución  $H$  ( $Y \in \text{MDA}(H)$ ) si existen sucesiones de constantes  $\{c_n\}$  y  $\{d_n\}$ , tales que para cada  $n$ ,  $c_n > 0$ ,  $d_n \in \mathbf{R}$  y



$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H. \quad (2.2)$$

Notemos que si la distribución  $H$  es continua en toda la recta real  $\mathbf{R}$ , entonces la convergencia en (2.2) es equivalente a tener que para cada  $x$  en  $\mathbf{R}$

$$\lim_{n \rightarrow \infty} P(M_n \leq c_n x + d_n) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x).$$

### 2.2.1. Teoría clásica

La distribución límite  $H(y)$  de los máximos estandarizados  $\frac{M_n - b_n}{a_n}$  ha sido estudiada en forma amplia en la literatura de valores extremos por investigadores como Fischer y Tippett(1928) y Gnedenko (1943). En particular se conoce el siguiente teorema desarrollado por éstos tres autores.

#### Teorema 2.1 (Fischer-Tippet-Gnedenko)

Sea  $\{Y_n\}_{n>0}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas. Supóngase que existen constantes de normalización

$$a_n > 0 \quad \text{y} \quad b_n \in \mathbf{R},$$

tales que

$$P \left[ \frac{M_n - b_n}{a_n} \leq y \right] \rightarrow H(y), \quad \text{cuando} \quad n \rightarrow \infty,$$

donde  $H$  es una función de distribución no degenerada, es decir, que no vale 0 ni 1 para toda  $y$ .

Entonces  $H$  pertenece a una de las tres familias:

$$1. \text{ Fréchet} \quad H(y) = \Phi_{\alpha}(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ \exp\{-y^{-\alpha}\} & \text{si } y > 0 \end{cases}$$

$$2. \text{ Weibull} \quad H(y) = \Psi_{\alpha}(y) = \begin{cases} \exp\{-(-y)^{\alpha}\} & \text{si } y \leq 0 \\ 1 & \text{si } y > 0, \end{cases}$$

$$3. \text{ Gumbel} \quad H(y) = \Lambda(y) = \begin{cases} \exp\{-\exp(-y)\} & \text{si } y \in \mathbf{R} \end{cases}$$

La demostración de este resultado es de naturaleza muy técnica y se avanzará en otra dirección. Un bosquejo de la prueba puede hallarse en Embrechts, et al. (1991).

Los tres casos en que puede clasificarse a  $H$  se han logrado representar mediante una sola distribución conocida como distribución de valores extremos generalizada (DVEG). Esta distribución se obtiene al introducir un parámetro  $\xi$ , tal que cuando

$\xi = \frac{1}{\alpha} > 0$  corresponde a la distribución de tipo Fréchet  $\Phi_{\alpha}$ ,

$\xi = 0$  corresponde a la distribución Gumbel  $\Lambda$  y,

$\xi = -\frac{1}{\alpha} < 0$  corresponde a la distribución Weibull  $\Psi_{\alpha}$ .

A continuación se define la DVEG.

**Definición 2.4** *Distribución de valores extremos generalizada (Jenkinson-von Mises)*

Se define la función de distribución  $H_\xi$  como

$$H_\xi(y) = \begin{cases} \exp\{-(1 + \xi y)^{-\frac{1}{\xi}}\} & \text{si } \xi \neq 0, \\ \exp\{-\exp\{-y\}\} & \text{si } \xi = 0, \end{cases} \quad (2.3)$$

donde  $1 + \xi y > 0$ .

Entonces el rango de  $H_\xi$  corresponde a

$$\begin{aligned} y &> -\xi^{-1} && \text{para } \xi > 0, \\ y &< -\xi^{-1} && \text{para } \xi < 0, \\ y &\in \mathcal{R} && \text{para } \xi = 0. \end{aligned}$$

Es posible obtener una versión análoga de  $H_\xi$  al incluir parámetros de localización y escala y sustituir  $y$  por  $\frac{y-\mu}{\sigma}$  en la expresión para  $H_\xi(y)$  que se establece en la definición 2.4. La distribución que resulta se denota como  $H_{\xi,\mu,\sigma}$  y es referida también como distribución de valores extremos generalizada.

El caso  $H_0$  se interpreta como el límite de  $H_\xi$  cuando  $\xi \rightarrow 0$ , entonces

$$H_\xi(y) = \exp\{-(1 + \xi y)^{-\frac{1}{\xi}}\}, \quad 1 + \xi y > 0,$$

sirve como representación para toda  $\xi \in \mathcal{R}$ .

El teorema 2.1 es un resultado muy importante ya que establece que la distribución asintótica del máximo siempre pertenecerá a alguna de las tres familias, Fréchet, Weibull o Gumbel, sin importar la distribución original de los datos, que generalmente es desconocida, por lo que se pueden hacer estimaciones sin realizar supuestos acerca de ésta última.

Por otra parte, la distribución de valores extremos generalizada resulta especialmente útil para calcular estimadores de máxima verosimilitud ya que en la mayor parte de los

casos no se sabe de antemano a qué familia pertenecerá la distribución límite de la muestra de máximos.

### 2.2.2. Teoría moderna

Un enfoque alternativo en la teoría estadística de valores extremos se basa en seleccionar un valor de umbral  $u$  y estudiar al subconjunto de elementos de la muestra que exceden al valor de  $u$ ; llamaremos a los elementos de tal conjunto excedentes. El problema de escoger un valor adecuado de  $u$  se discutirá más adelante. Por ahora enfocaremos nuestro interés en dos variables:

1. Cuántos excedentes se tienen sobre un periodo de tiempo fijo.
2. La magnitud de los excesos, es decir, por cuánto fue excedido el valor de  $u$  en cada excedente.

Para estas variables, pensemos en la siguiente aproximación. Supóngase que  $F$  denota a la función de distribución común de las observaciones originales  $Y_1, Y_2, \dots, Y_n$ , entonces la distribución de los excesos  $X_i = Y_i - u$ , condicional al evento  $Y_i > u$  es

$$P(Y \leq u + y | Y > u) = \frac{F(u + y) - F(u)}{1 - F(u)} \equiv F_u(y). \quad (2.4)$$

Denotemos por  $\omega_F$  al punto final a la derecha de la distribución  $F$ . Pickands(1975), Embrechts et al. (1991) han justificado que si  $F \in MDA(H_\xi)$ , se puede encontrar un factor de re-escalamiento  $\sigma = \sigma(u) > 0$  tal que

$$\lim_{u \uparrow \omega_F} P\left(\frac{Y - u}{\sigma(u)} \leq y | Y > u\right) = 1 - (1 + \xi y)^{-\frac{1}{\xi}},$$

para  $y \geq 0$ ;  $1 + \xi y > 0$  y  $\xi \neq 0$ .

En el teorema B.3 del apéndice B se presenta una argumentación de este resultado.

Equivalentemente, podemos decir que cuando  $u \uparrow \omega_F$  y  $F \in \text{MDA}(H_\xi)$  para algún  $\xi \neq 0$ , podemos encontrar  $\sigma = \sigma(u) > 0$  tal que

$$F_u(y) \approx 1 - \left(1 + \xi \frac{y}{\sigma(u)}\right)^{-\frac{1}{\xi}} = G(y, \sigma(u), \xi), \quad (2.5)$$

para  $y \geq 0$ ,  $1 + \frac{\xi}{\sigma(u)}y > 0$  y  $\xi \neq 0$ .

La función  $G(y, \sigma(u), \xi)$  en 2.5 recibe el nombre de distribución Pareto Generalizada (DPG). El teorema B.3 en el apéndice B establece que el parámetro de forma  $\xi$  que aparece en la DPG es el mismo que aparece en la DVEG  $H_\xi$  para la cual  $F \in \text{MDA}(H_\xi)$ .

Al igual que para la DVEG, en la DPG pueden identificarse tres casos diferentes dependiendo del signo de  $\xi$ :

En el caso  $\xi > 0$ , la DPG es válida para  $x \in \mathcal{R}^+$  y satisface

$$1 - G(y; \sigma, \xi) \sim cy^{-\xi}, \quad c > 0.$$

Si  $\xi < 0$ , la DPG es similar al tipo Weibull de la DVEG con punto final a la derecha de la distribución  $\omega_F = \frac{\sigma}{|\xi|}$ .

Finalmente, cuando  $\xi = 0$  se tiene la distribución exponencial con media  $\sigma$ ,

$$G(y, \sigma, 0) = 1 - \exp\left(-\frac{y}{\sigma}\right).$$

En Smith y Goodman (2000) y Smith (1997) se establece que el caso  $\xi < 0$  corresponde a distribuciones para las que  $\omega_F < \infty$ .

Como en nuestro caso trabajaremos con datos de reclamaciones, en principio, usar un modelo probabilístico donde se tuviera  $\xi < 0$  implicaría que pensamos en distribuciones que tienen un límite superior finito. Esta situación no es la que se pretende modelar. Por lo anterior, consideraremos que  $\xi \geq 0$ , y que  $\sigma(u)$  es un parámetro de escala.

El caso  $\xi > 0$ , es el más adecuado para nuestra modelación por corresponder a una distribución de cola pesada. De acuerdo al teorema B.3 (del apéndice B, tendríamos que si  $\xi > 0$  entonces  $F$  pertenece al máximo dominio de atracción de la distribución Fréchet  $\Phi_\alpha$ ;  $\alpha = \frac{1}{\xi}$  que es la familia de colas más pesadas en la DVEG.

Por otra parte, la esperanza y la varianza de la DPG existen si  $\xi < 1$  y  $\xi < \frac{1}{2}$ , respectivamente. Para obtenerlas se hace uso del siguiente teorema

**Teorema 2.2** Si  $X$  tiene valor esperado finito y  $X \geq 0$  entonces

$$E[X] = \int_0^{\infty} (1 - F_X(x)) dx$$

Por ejemplo, en el caso de la esperanza para  $Z \sim \text{DPG}(\xi, \sigma)$ , si  $\xi < 1$  entonces

$$F_Z(z) = G(z) = 1 - \left(1 + \frac{\xi z}{\sigma}\right)^{-\frac{1}{\xi}} \quad y,$$

$$\begin{aligned} E(Z) &= \int_0^{\infty} \left(1 + \frac{\xi z}{\sigma}\right)^{-\frac{1}{\xi}} dz \\ &\stackrel{\xi \leq 1}{=} \frac{\sigma}{\xi} \cdot \frac{\left(1 + \frac{\xi z}{\sigma}\right)^{1-\frac{1}{\xi}}}{1 - \frac{1}{\xi}} \Bigg|_0^{\infty} \\ &= \frac{\sigma}{\xi \left(1 - \frac{1}{\xi}\right)} \\ &= \frac{\sigma}{1 - \xi} \end{aligned}$$

por lo tanto,

$$E[Z] = \frac{\sigma}{1 - \xi}. \quad (2.6)$$

En forma análoga la varianza se puede calcular como

$$\text{VAR}[Z] = \frac{\sigma^2}{(1 - \xi)^2(1 - 2\xi)},$$

siempre que  $\xi < \frac{1}{2}$ , usando que  $\text{VAR}[Z] = E[Z^2] - (E[Z])^2$ .

En el teorema B.3 del apéndice B se establece que si la distribución de  $Y-u$  condicional a  $Y > u$  es como en (2.5) para  $u$  suficientemente grande, entonces  $F$  debe pertenecer al máximo dominio de atracción de una distribución como la que aparece en la ecuación (2.3). Como se verá en la sección 3.2, para relacionar los parámetros  $\psi$  y  $\mu$ , de la DVEG en (2.3) con los parámetros  $\xi$  y  $\sigma$  en la DPG en (2.5) será conveniente reparametrizar

$$\sigma(u) = \psi + \xi(u - \mu).$$

Si asumimos que  $\xi < 1$  y haciendo uso del valor esperado en (2.6), obtenemos que para los excesos  $Y_i - u$

$$E[Y_i - u | Y_i > u] = \frac{\psi + \xi(u - \mu)}{1 - \xi}. \quad (2.7)$$

Entonces si graficamos los puntos  $(\omega, E[Y_i - \omega | Y_i > \omega])$  para valores de  $\omega$  en todo un intervalo, deberíamos observar una línea recta con pendiente  $\frac{\xi}{1-\xi}$ .

### 2.2.3. Gráfico de la media de los excesos

Una técnica empleada frecuentemente para encontrar el valor óptimo del umbral  $u$ , a partir del cual los datos siguen una DPG, es la construcción de gráficos de la media de los excesos. Estos gráficos se realizan bajo el supuesto de que los excesos  $Y_i - u$  condicionados a  $Y_i > u$ , siguen una DPG( $\sigma, \xi$ ). Si el modelo es correcto, al graficar la media muestral de dichos excesos contra el valor de  $u$ , para  $u$  en un rango de valores, se podrá identificar un valor de umbral  $u$  a partir del cual la gráfica se asemeja a una línea recta sobre un intervalo  $[u, u_0)$ .

Lo anterior se puede llevar a cabo de manera empírica, calculando la media muestral de los excedentes

$$\hat{\mu}(u) = \frac{\sum_t (Y_t - u) \mathbf{1}_{Y_t > u}}{\sum_t \mathbf{1}_{Y_t > u}} \quad (2.8)$$



y graficándola contra  $u$ , para  $u$  en un rango de valores.

La inspección simple del gráfico que así se obtenga puede darnos una idea de cuál será el umbral óptimo si es posible identificar un valor de  $u$  a partir del cual el gráfico sea aproximadamente una línea recta para todo  $y \geq u$ .

Adicionalmente, se construyen bandas de confianza de la siguiente manera.

Sea  $\mu(y)$ , la media teórica de los excesos definida por (2.7) para  $y \geq u$ .

$$\begin{aligned}\mu(y) &= E[Y - y | Y > y] = \frac{\psi + \xi(y - \mu)}{1 - \xi} \\ &= \frac{\psi + \xi(u - \mu) + \xi(y - u)}{1 - \xi} \\ &= \frac{\sigma + \xi(y - u)}{1 - \xi}, \quad y \geq u.\end{aligned}$$

Sean  $\hat{\mu}(y)$  definida en (2.8) y  $\hat{\xi}, \hat{\sigma}$  los estimadores de máxima verosimilitud de  $\xi$  y  $\sigma$  respectivamente. Smith (1995) sugiere que una forma de probar el supuesto de la distribución Pareto generalizada es usar la estadística

$$\hat{\mu}(y) - \frac{\hat{\sigma} + \hat{\xi}(y - u)}{1 - \hat{\xi}} \quad (2.9)$$

para  $u$  un valor de umbral.

Para desarrollar intervalos de confianza simulamos de la distribución de (2.9) como sigue. Para  $k = 1, 2, \dots, 99$ , se genera una muestra aleatoria a partir de la DPG sobre el umbral  $u$ , del mismo tamaño que la muestra original, tomando como valores de los parámetros a  $\hat{\xi}, \hat{\sigma}$ . Para cada  $k$  se calculan los nuevos estimadores de máxima verosimilitud,  $\hat{\xi}^{(k)}, \hat{\sigma}^{(k)}$  así como la media muestral de los excesos  $\hat{\mu}^{(k)}(y)$ .

Para cada valor de umbral  $y$ , se obtiene

$$\hat{\mu}^{(k)}(y) = \frac{\hat{\sigma}^{(k)} + \hat{\xi}^{(k)}(y - u)}{1 - \hat{\xi}^{(k)}} + \frac{\hat{\sigma} + \hat{\xi}(y - u)}{1 - \hat{\xi}}; \quad k = 1, 2, \dots, 99 \quad (2.10)$$

y se determinan los cuantiles 5 y 95 por ciento de esta muestra, los cuales determinarán los límites superior e inferior de las bandas de confianza para  $\hat{\mu}(y)$  si el modelo es adecuado.

El proceso anterior se repite para varios valores de  $y$  en  $(u, u_0)$  a fin de encontrar el umbral óptimo a proponer.

A continuación se realiza el gráfico de la media de los excesos para una serie de datos correspondiente a reclamaciones de reaseguro en cierta compañía aseguradora.

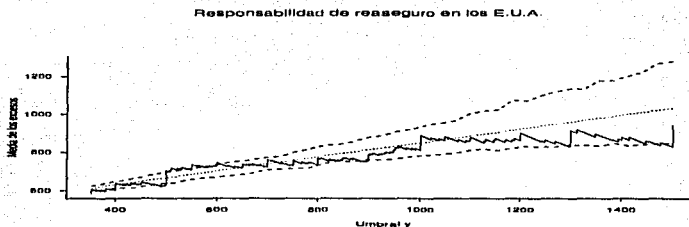


Figura 2.1: Media de los excesos

### 2.3. Procesos puntuales

Esta sección describe el proceso de reclamaciones de reaseguro como un Proceso de Poisson Puntual.

En conexión con las variables 1 y 2 mencionadas al inicio de la sección 2.2.2, una técnica alternativa combina la información de los valores de los excesos con los tiempos en que una observación excede el valor dado de un umbral  $u$ .

**TESIS CON  
FALLA DE ORIGEN**

Esta técnica consiste en ver a los tiempos donde se excede el valor  $u$ ,  $T_1, T_2, \dots, T_k$  y a los valores  $Y_{T_1}, Y_{T_2}, \dots, Y_{T_k}$  como parte de un proceso puntual.

Supóngase que graficamos a las parejas ordenadas  $(T_i, Y_{T_i})$ ;  $i = 1, 2, \dots$  en un periodo de tiempo de  $t_1$  a  $t_2$ .

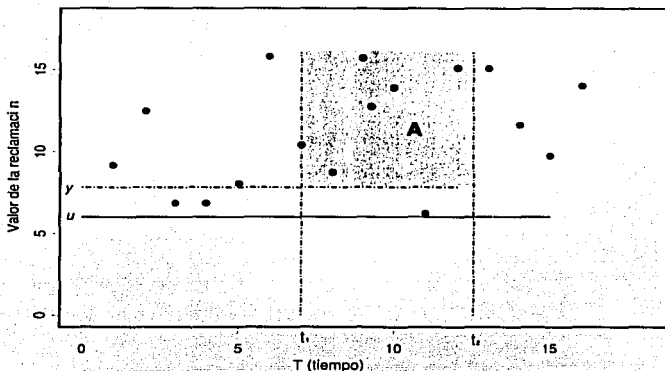


Figura 2.2: Proceso puntual en  $R^2$

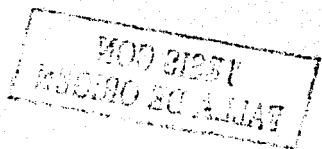
Si  $A$  es un subconjunto de Borel en  $(0, T] \times (u, \infty)$ , por ejemplo

$$A = (t_1, t_2) \times (y, \infty); \quad y > u;$$

entonces

$$\mathcal{N}(A) = \text{número de reclamaciones mayores a } y, \text{ en el intervalo de tiempo } (t_1, t_2).$$

La colección de variables aleatorias  $\{\mathcal{N}(A) : A \in \mathcal{B}((0, T] \times (u, \infty))\}$ , donde  $\mathcal{N}(A)$  es el número de puntos  $(T_i, Y_{T_i})$  que están en  $A$ , es conocida como *proceso puntual* ya que



para cada "índice"  $A$  (un conjunto en  $\mathcal{B}((0, T] \times (u, \infty))$ ), obtenemos un conteo (aleatorio) de puntos correspondiente a este "índice". El apéndice C trata con mayor detalle lo que se describe a continuación.

**Definición 2.5** Para  $E \subset \mathbb{R}^2$ , sean  $\mathcal{B}(E)$  la  $\sigma$ -álgebra de Borel en  $E$  y  $\mathcal{N}$  un proceso puntual con valores en  $E$ .

Decimos que  $\mathcal{N}$  es un proceso de Poisson puntual con medida de intensidad (valor esperado)  $\mu$  si

(i) Para cada  $A \in \mathcal{B}(E)$

$$IP(\mathcal{N}(A) = k) = \begin{cases} \frac{\exp(-\mu(A))(\mu(A))^k}{k!} & \text{si } \mu(A) < \infty \\ 0 & \text{si } \mu(A) = \infty \end{cases}$$

donde  $\mu$  es una medida sobre  $(E, \mathcal{B}(E))$ .

(ii) Si  $A_1, A_2, \dots, A_k$  son elementos de  $\mathcal{B}(E)$  con  $A_i \cap A_j = \emptyset$  para toda  $i \neq j$ , entonces  $\mathcal{N}(A_1), \mathcal{N}(A_2), \dots, \mathcal{N}(A_k)$  son variables aleatorias independientes.

En el apéndice C se discute un resultado de Leadbetter et al. (1983) que establece la siguiente aproximación:

Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias independientes con distribución  $F$ ,  $M_n = \max\{Y_1, \dots, Y_n\}$  y  $\{a_n\}, \{b_n\}$  sucesiones de números reales tales que

$$IP\left(\frac{M_n - b_n}{a_n} \leq y\right) \rightarrow H(y; \xi, \mu, \sigma) = \exp\left\{-\left[1 + \xi\left(\frac{y - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\},$$

cuando  $n \rightarrow \infty$ .

Entonces el proceso  $\{\mathcal{N}(A) : A \in \mathcal{B}((0, T] \times (u, \infty))\}$  descrito al inicio de esta sección es aproximadamente un proceso de Poisson puntual con medida de intensidad  $\Lambda$  tal que

para conjuntos de la forma  $(t_1, t_2] \times (y, \infty)$ , con  $y \geq u$  vale

$$\Lambda((t_1, t_2] \times (y, \infty)) = (t_2 - t_1) \times \left(1 + \xi \left(\frac{y - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}.$$

Si suponemos que el tiempo se mide en unidades de años, el número de reclamaciones que exceden el valor de  $y \geq u$  durante el año  $t$  es  $N_t = \mathcal{N}((t, t + 1] \times (y, \infty))$ .

En la figura 2.2 se ejemplifica de manera gráfica un proceso puntual en  $\mathbf{R}^2$ .

# Capítulo 3

## APLICACIÓN A DATOS DE REASEGURO

### 3.1. Análisis descriptivo

La serie de datos original que se emplea no se da a conocer por razones de confidencialidad. Sin embargo, se realizará un análisis descriptivo breve que permitirá observar las características generales de la misma. Los datos corresponden a responsabilidades de reaseguro para aseguradoras en Estados Unidos.

La compañía reaseguradora ofrece cobertura sólo hasta un límite determinado, lo anterior se refleja en los datos (tamaños de reclamaciones), ya que cuando la reclamación excede el límite establecido, sólo se reporta el valor de este límite. El límite puede cambiar con cada reclamación.

El archivo original contiene registros con la información del año de ocurrencia, límite de cobertura y el valor de la reclamación de cada evento (en caso de no exceder el límite de cobertura), para un total de 11 años.

Así entonces, podemos pensar en los datos como un proceso de reclamaciones

$$Y_1^1, \dots, Y_{N_1}^1, Y_1^2, \dots, Y_{N_2}^2, \dots, Y_1^{11}, \dots, Y_{N_{11}}^{11},$$

con censura por la derecha y  $N_j$  observaciones en el año  $j$ ,  $j = 1, 2, \dots, 11$ .

Año	Número de reclamaciones	Censuradas
1987	71	6
1988	54	3
1989	62	4
1990	63	5
1991	81	2
1992	68	3
1993	77	3
1994	74	1
1995	47	0
1996	28	1
1997	9	1
Total	635	29

Cuadro 3.1: Descripción de la serie datos censurados

En estos datos no hay información acerca de las fechas exactas de los eventos (mes o día).

De las 701 observaciones originales, se trabajó con aquéllas mayores a 45,000 (casi un 10% de los datos están por debajo de este umbral), además se cambió la escala al dividir por 1000.

Una vez hecho lo anterior, se tienen 635 muestras.

Únicamente con el fin de hacer una primera gráfica descriptiva, se supone que las reclamaciones están ordenadas cronológicamente dentro de cada año.

La figura 3.1 muestra las reclamaciones acumuladas en el tiempo. En ella se puede observar que existe cierto patrón anual.

Responsabilidad de reaseguro en E.U.A.

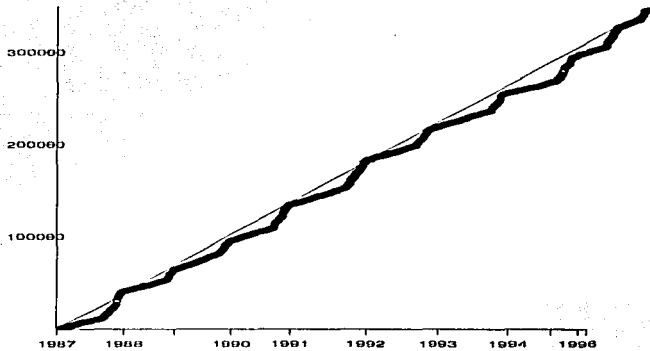


Figura 3.1: Reclamaciones acumuladas

Por otra parte, el histograma en la figura 3.2 corresponde a la frecuencia de las reclamaciones en relación al monto total de las mismas. Los diagramas de caja en la figura 3.3 permiten apreciar la variabilidad que existe entre los grupos anuales de las reclamaciones, la línea horizontal trazada para cada caja corresponde a la mediana  $M_i$  de los datos para el año  $i$ ,  $i = 1, 2, \dots, 11$ .

$M_i$	Intervalo
$M_1, M_2, M_4, M_5, M_6, M_7, M_8, M_9, M_{11}$	[280,450)
$M_3$	[450,500]
$M_{10}$	(500,550]

Cuadro 3.2: Medianas de los datos de reaseguro agrupados anualmente

TESIS CON  
FALLA DE ORIGEN



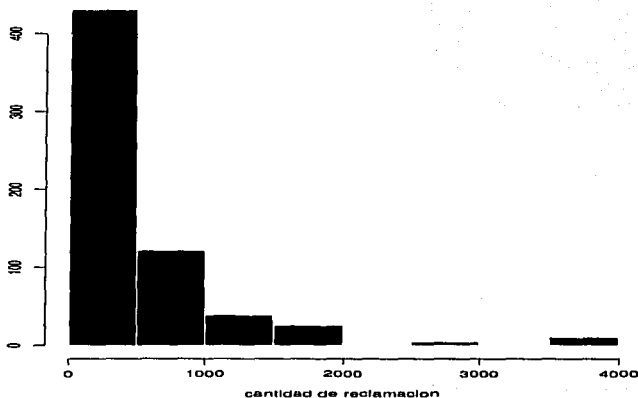


Figura 3.2: Histograma de frecuencias para las 635 reclamaciones de reaseguro

La tabla 3.2 agrupa las medianas por intervalos. En general, la diferencia entre medianas no es tan grande en relación a la diferencia que puede observarse en la dispersión de los datos, año con año.

Los bordes superiores e inferiores de los diagramas de caja corresponden al cuartil superior e inferior, respectivamente; las líneas discontinuas indican la presencia de datos más allá de los cuartiles.

A partir de la figura 3.3 puede afirmarse que en los años 1,4,5,8,9 y 10 la distribución anual de las reclamaciones presenta colas más pesadas; además se observa que en el año ocho el monto de los siniestros fue mayor en comparación a otros.

El gráfico de la media de los excesos para la serie de datos descrita se realizó considerando umbrales en un rango de a)45 a 500, b)45 a 1000, c)45 a 1500 y d)45 a 4000 como se muestra en la figura 3.4.

TESIS CON  
FALLA DE ORIGEN

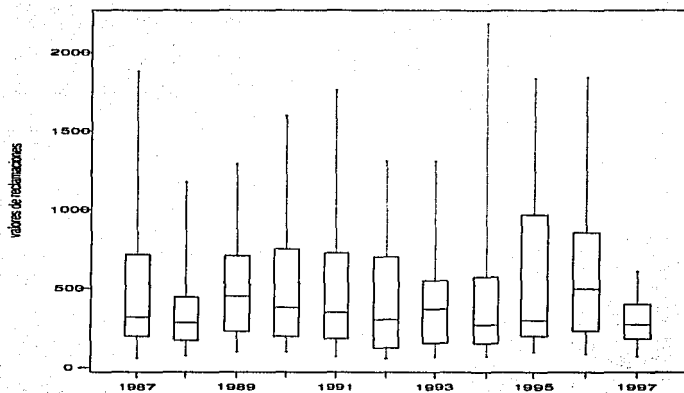


Figura 3.3: Diagramas de caja para los valores de las 635 reclamaciones de reaseguro

Al observar el gráfico de la media de los excesos puede afirmarse que la distribución de los excedentes sigue una DPG para umbrales entre 270 y 460 ya que en ese rango de valores se observa un comportamiento aproximadamente lineal. La figura 2.1 en el capítulo dos muestra la media de los excesos para las observaciones a partir del umbral 350 y hasta 1500, en este gráfico se observa que hay intervalos de umbrales para los cuales la DPG podría ser un buen modelo en la descripción de los excesos.

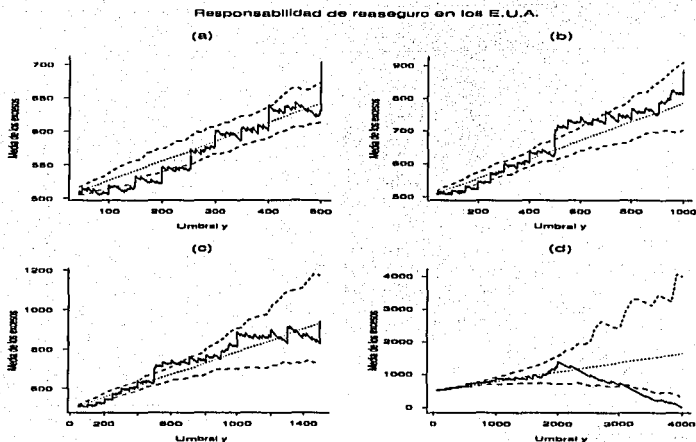


Figura 3.4: Medio de los excesos

### 3.2. Modelos Poisson-DPG

A partir de esta sección se hará uso de la(s) palabra(s) “modelo(s)” en referencia a la verosimilitud, las distribuciones inicial y posterior y a las estructuras jerárquicas que lleguen a considerarse.

A continuación se proponen dos modelos Poisson-DPG para los datos de reaseguro descritos en la primera parte del capítulo; estos modelos contemplan los siguientes casos:

- (i) Efectos aleatorios anuales en los parámetros de localización y escala.
- (ii) Efectos aleatorios anuales únicamente en la escala.

### Descripción general de los modelos Poisson-DPG

Sea  $N_t$  el número de reclamaciones por encima del umbral  $u$  para el año  $t$ , donde  $t = 1, 2, \dots, T^*$ .

Sea  $y_i^t$  el valor de la  $i$ -ésima muestra correspondiente al  $t$ -ésimo año. Denotaremos por  $x_i^t$  a los excecetes  $x_i^t = y_i^t - u$ .

Dado un valor de umbral  $u$ , de acuerdo a Smith (1995 y 1997), suponemos que para cada año  $t$ :

- (a)  $N_t$  se distribuye Poisson( $\lambda_t$ ).
- (b) Condicionalmente en el valor de  $N_t$ , las variables aleatorias  $X_1^t, X_2^t, \dots, X_{N_t}^t$  son independientes y con idéntica distribución dada por la Distribución Pareto Generalizada, DPG:

$$P(X_i^t \leq x_i^t | Y_i^t > u) = 1 - \left(1 + \xi_t \frac{x_i^t}{\sigma_t}\right)_+^{-\frac{1}{\xi_t}}$$

Donde  $x_+ = \max(x, 0)$ ,  $\sigma_t > 0$  es un parámetro de escala y  $\xi_t$  es un parámetro de forma, ambos para el año  $t$ . En Smith (2000) se presenta una discusión respecto al rango de valores de  $\xi_t$ . El caso  $\xi_t < 0$  corresponde a distribuciones cuyo soporte tiene una cota superior finita, entonces si tal parámetro de forma fuese negativo, estaríamos pensando en observar datos (de reclamaciones) que tuvieran tendencia a acumularse cerca de una cota superior, lo cual no tiene mucho sentido en datos de reclamaciones. Por lo anterior consideraremos que  $\xi_t \geq 0$ .

Así entonces, la verosimilitud para el año  $t$  es

$$\begin{aligned} L_t &= P(N_t = n_t) \times f(x_1^t, x_2^t, \dots, x_{n_t}^t | N_t = n_t) \\ &= \frac{\lambda_t^{n_t} e^{-\lambda_t}}{(n_t)!} \times \prod_{i=1}^{n_t} \underbrace{\left( \frac{1}{\sigma_t} \left(1 + \xi_t \frac{x_i^t}{\sigma_t}\right)^{-\left(1 + \frac{1}{\xi_t}\right)} \right)}_{(1)} \end{aligned} \quad (3.1)$$

$$= \frac{\lambda_i^{n_i} e^{-\lambda_i}}{(n_i)!} \times \prod_{i=1}^{n_i} (\sigma_i + \xi x_i^i)^{-\left(1 + \frac{1}{\xi_i}\right)} \sigma_i^{\frac{1}{\xi_i}} \quad (3.2)$$

Smith (1997), establece que el modelo Poisson-DPG es consistente con la Distribución Valor Extremo Generalizada (DVEG) si se consideran las siguientes reparametrizaciones

$$\lambda_i = \left(1 + \xi_i \left(\frac{u - \mu_i}{\psi_i}\right)\right)_+^{-\frac{1}{\xi_i}};$$

$$\sigma_i = \psi_i + \xi_i(u - \mu_i),$$

donde  $\mu_i$  es el parámetro de localización que aparece en la DVEG.

Si además consideramos que los datos fueron censurados, reemplazamos los factores (1) en la ecuación (3.1) correspondientes a observaciones censuradas por

$$\left(1 + \frac{\xi_i}{\sigma_i} \omega_i\right)^{-\frac{1}{\xi_i}} \equiv 1 - GPD(\omega_i, \xi_i, \sigma_i),$$

Donde  $\omega_i = k_i - u$  si  $k_i$  es el valor de censura para la reclamación  $y_i$ .

Obtenemos entonces

$$L_i = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \times \prod_{i \in A_i} \left(\frac{1}{\sigma_i} \left(1 + \xi_i \frac{x_i^i}{\sigma_i}\right)^{-\left(1 + \frac{1}{\xi_i}\right)}\right) \times \prod_{i \in B_i} \left(\left(1 + \xi_i \frac{\omega_i^i}{\sigma_i}\right)_+^{-\left(\frac{1}{\xi_i}\right)}\right).$$

Donde

$$A_i = \{i \in \{1, 2, \dots, n_i\} : x_i^i \text{ no fue censurada}\}$$

$$B_i = \{i \in \{1, 2, \dots, n_i\} : x_i^i \text{ fue censurada}\}$$

$$\text{y } \#(A_i) + \#(B_i) = n_i.$$

De tal manera que sustituyendo los valores de la reparametrización

$$\begin{aligned}
 L_t &= \frac{1}{n_t!} \left( \frac{1}{\psi_t} \right)^{-\frac{n_t}{t}} (\psi_t + \xi_t(u - \mu_t))^{-\frac{n_t}{t}} \\
 &\times e^{-\left(1 + \frac{\xi_t}{\psi_t}(u - \mu_t)\right)_+^{-\frac{1}{t}}} \\
 &\times \prod_{i \in A_t} (\psi_t + \xi_t(u - \mu_t))_+^{\frac{1}{t}} (\psi_t + \xi_t(u - \mu_t) + \xi_t x_i^t)^{-(1 + \frac{1}{t})} \\
 &\times \prod_{i \in B_t} (\psi_t + \xi_t(u - \mu_t))_+^{\frac{1}{t}} (\psi_t + \xi_t(u - \mu_t) + \xi_t \omega_i^t)^{-(\frac{1}{t})} \\
 L_t &= \frac{1}{n_t!} \psi_t^{\frac{n_t}{t}} e^{-\left(1 + \frac{\xi_t}{\psi_t}(u - \mu_t)\right)_+^{-\frac{1}{t}}} \\
 &\times \prod_{i \in A_t} (\psi_t + \xi_t(y_i^t - \mu_t))_+^{-(1 + \frac{1}{t})} \\
 &\times \prod_{i \in B_t} (\psi_t + \xi_t(k_i - \mu_t))_+^{-(\frac{1}{t})}
 \end{aligned}$$

donde  $k_i$  es el valor de censura para  $y_i$ .

Entonces,

$$\begin{aligned}
 L_t &= \frac{1}{n_t!} e^{-\left(1 + \frac{\xi_t}{\psi_t}(u - \mu_t)\right)_+^{-\frac{1}{t}}} \\
 &\times \prod_{i \in A_t} \frac{1}{\psi_t} \left( 1 + \frac{\xi_t}{\psi_t}(y_i^t - \mu_t) \right)_+^{-(1 + \frac{1}{t})} \\
 &\times \prod_{i \in B_t} \left( 1 + \frac{\xi_t}{\psi_t}(k_i - \mu_t) \right)_+^{-(\frac{1}{t})}
 \end{aligned}$$

y por tanto la verosimilitud para los años 1 a 11 se expresará como

$$L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{11} | \Theta) = \prod_{t=1}^{11} L_t(\mathbf{x}_t | \theta)$$

donde  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{11})$  es el vector de excesos tal que

$$\mathbf{x}_1 = (x_{1,1}, x_{2,1}, \dots, x_{1,N_{11}}), \dots, \mathbf{x}_{11} = (x_{1,11}, \dots, x_{11,N_{11}})$$

con  $N_t$ ,  $t = 1, \dots, 11$ , el número de excedentes observados en cada año  $t$ , y

$$\Theta = (\mu_1, \mu_2, \dots, \mu_{11}, \mu, \tau_1^2, \psi_1, \psi_2, \dots, \psi_{11}, \psi, \tau_2^2, \xi) \quad (3.3)$$

y la distribución posterior será

$$P(\Theta | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{11}) \propto L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_{11} | \Theta) \cdot P(\Theta).$$

Ahora reparametrizamos considerando la siguiente transformación y el valor absoluto de su jacobiano

$$\begin{aligned} \mathcal{T}(\Theta) &= \Xi \text{ donde } \Xi = (\mu_1, \mu_2, \dots, \mu_{11}, \mu, \tau_1^2, \phi_1, \phi_2, \dots, \phi_{11}, \phi, \tau_2^2, \xi) \\ \phi_j &= \log(\psi_j); \quad j = 1, 2, \dots, 11 \\ \phi &= \log(\psi) \end{aligned}$$

$$|J_{\mathcal{T}}^{-1}| = \prod_{j=1}^{11} e^{\phi_j} \times e^{\phi}$$

entonces,

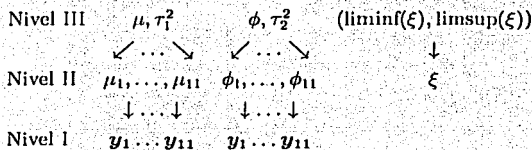
$$\Xi = (\mu_1, \mu_2, \dots, \mu_{11}, \mu, \tau_1^2, \phi_1, \phi_2, \dots, \phi_{11}, \phi, \tau_2^2, \xi)$$

y la distribución posterior se calculará como

$$P(\Xi | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{11}) \propto L(\mathbf{x}_1, \dots, \mathbf{x}_{11} | \mathcal{T}^{-1}(\Xi)) \times |J_{\mathcal{T}^{-1}}|$$

### 3.2.1. Efectos aleatorios en localización y escala

A fin de estudiar efectos aleatorios anuales para  $t = 1, \dots, 11$ , en los parámetros de localización  $(\mu_1, \mu_2, \dots, \mu_{11})$  y escala  $(\phi_1, \phi_2, \dots, \phi_{11})$ , consideraremos un modelo completamente bayesiano, con las siguiente estructura jerárquica:



El límite superior e inferior de  $\xi$  se fijaron como 0.49 y 0.01, respectivamente.

De esta manera se establecen las siguientes distribuciones iniciales para los parámetros.

Para el parámetro de localización, suponemos que  $\mu_1, \mu_2, \dots, \mu_{11}$  son variables aleatorias independientes dado el valor de  $(\mu, \tau_1^2)$

$$\mu_t \sim N(\mu_t | \mu, \tau_1^2) \quad (3.4)$$

y para la escala, suponemos que  $(\phi_1, \phi_2, \dots, \phi_{11})$  son variables aleatorias independientes dado el valor de  $(\phi, \tau_2^2)$  y tales que

$$\phi_t = \log(\psi_t) \sim N(\phi_t | \log(\psi), \tau_2^2) \quad (3.5)$$

donde  $\log(\psi) = \phi$



Por otra parte, la distribuciones conjuntas de  $IP(\mu, \tau_1^2)$ ,  $IP(\phi, \tau_2^2)$  se modelarán de la siguiente forma

$$IP(\mu, \tau_1^2) = IP(\mu|\delta, \tau_1^2) \times IP(\tau_1^2|a, b) \quad (3.6)$$

donde

$$IP(\mu|\delta, \tau_1^2) \sim N(\delta, \tau_1^2) \quad \text{y} \quad IP(\tau_1^2|a, b) \sim GI(a, b)$$

tal que  $GI$  denota una distribución Gamma Inversa y los hiperparámetros  $\delta, a$  y  $b$  se consideran fijos.

por lo tanto

$$\begin{aligned} IP(\mu, \tau_1^2) &= \frac{1}{\sqrt{2\pi}\tau_1} e^{-\frac{1}{2\tau_1^2}(\mu-\delta)^2} \times \frac{b^a}{\Gamma(a)} \left(\frac{1}{\tau_1^2}\right)^{a+1} e^{-\frac{b}{\tau_1^2}} \\ &= c(a, b) \left(\frac{1}{\tau_1^2}\right)^{\frac{1}{2}+a+1} e^{-\frac{1}{\tau_1^2}(b+\frac{1}{2}(\mu-\delta)^2)} \end{aligned} \quad (3.7)$$

Sabemos además que

$$\begin{aligned} IP(\mu_1, \dots, \mu_T | \mu, \tau_1^2) &= \prod_{t=1}^T IP(\mu_t | \mu, \tau_1^2) \\ &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\tau_1} e^{-\frac{1}{2\tau_1^2}(\mu_t - \mu)^2} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\frac{1}{\tau_1^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\tau_1^2} \sum_{t=1}^T (\mu_t - \mu)^2} \end{aligned} \quad (3.8)$$

así, de(3.7) y (3.8) se obtiene que

$$P(\mu_1, \dots, \mu_t; \mu, \tau_1^2) = c(a, b) \left( \frac{1}{\tau_1^2} \right)^{\frac{t}{2} + \frac{1}{2} + a + 1} e^{-\frac{1}{\tau_1^2} [b + \frac{1}{2}(\mu - \delta)^2 + \frac{1}{2} \sum_{i=1}^t (\mu_i - \mu)^2]}. \quad (3.9)$$

De manera análoga, obtenemos las expresiones para el parámetro de escala,  $\phi_t = \log(\psi_t)$ ,

$$\phi_t \sim N(\phi_t | \phi, \tau_2^2)$$

donde  $\phi = \log(\psi)$ , como se mencionó anteriormente.

Para la distribución conjunta,  $P(\phi, \tau_2^2)$  tenemos

$$P(\phi, \tau_2^2) = P(\phi, \tau_2^2) = P(\phi | \eta, \tau_2^2) \times P(\tau_2^2 | \alpha, \beta) \quad (3.10)$$

donde

$$P(\phi | \eta, \tau_2^2) \sim N(\eta, \tau_2^2) \quad \text{y} \quad P(\tau_2^2 | \alpha, \beta) \sim GI(\alpha, \beta)$$

y tales que  $\eta, \alpha$  y  $\beta$  se consideran fijos.

Por lo tanto

$$P(\phi, \tau_2^2) = \frac{1}{\sqrt{2\pi\tau_2}} e^{-\frac{1}{2\tau_2^2}(\phi - \eta)^2} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{1}{\tau_2^2} \right)^{\alpha+1} e^{-\frac{\beta}{\tau_2^2}} \quad (3.11)$$

$$= c(\alpha, \beta) \left( \frac{1}{\tau_2^2} \right)^{\frac{1}{2} + \alpha + 1} e^{-\frac{1}{2}(\beta + \frac{1}{2}(\phi - \eta)^2)}. \quad (3.12)$$

Sabemos además que

$$P(\phi_1, \dots, \phi_T | \phi, \tau_2^2) = \prod_{i=1}^T P(\phi_i | \phi, \tau_2^2) \quad (3.13)$$

$$\begin{aligned} &= \prod_{i=1}^T \frac{1}{\sqrt{2\pi\tau_2^2}} e^{-\frac{1}{2\tau_2^2}(\phi_i - \phi)^2} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\frac{1}{\tau_2^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\tau_2^2} \sum_{i=1}^T (\phi_i - \phi)^2} \end{aligned} \quad (3.14)$$

así, de (3.11) y (3.13) se obtiene que

$$P(\phi_1, \dots, \phi_T, \phi, \tau_2^2) = c(\alpha, \beta) \left(\frac{1}{\tau_2^2}\right)^{\frac{T}{2} + \frac{1}{2} + \alpha + 1} e^{-\frac{1}{2\tau_2^2} [\theta + \frac{1}{2}(\theta - \eta)^2 + \frac{1}{2} \sum_{i=1}^T (\phi_i - \phi)^2]} \quad (3.15)$$

Para el parámetro correspondiente al peso de la cola,  $\xi$ , se considerará una distribución a priori,  $\xi \sim \text{Unif}(0,01, 0,4)$ , ya que fuera de este rango la DPG no tiene momentos finitos de orden 1 y 2.

Además, se supone que las distribuciones a priori de  $(\mu_1, \dots, \mu_T, \mu, \tau_1^2)$ ,  $(\phi_1, \dots, \phi_T, \phi, \tau_2^2)$  y  $\xi$  son independientes. De esta manera se obtiene que la distribución a priori de  $\Xi$  se puede expresar como

$$P(\Xi) = P(\mu_1, \dots, \mu_T; \mu, \tau_1^2) \times P(\phi_1, \dots, \phi_T; \phi, \tau_2^2) \times \frac{1}{0,4 - 0,01} \mathbf{I}_{(0,01,0,4)}(\xi) \times |J_{T-1}|.$$

Con el fin de obtener una versión ligeramente más general de las distribuciones conjuntas a priori,  $P(\mu, \tau_1^2)$  y  $P(\phi, \tau_2^2)$ , se incluirán los hiperparámetros  $\gamma$  y  $\nu$ , respectivamente; de tal manera que

$$P(\mu, \tau_1^2) = NGI(\mu, \tau_1^2 | \delta, \gamma, \alpha, b)$$

y

$$P(\phi, \tau_2^2) = NGI(\phi, \tau_2^2 | \eta, \nu, \alpha, \beta)$$

donde NGI denota una distribución Normal-Gamma Inversa.

Haciendo los cálculos nuevamente, se tiene para el parámetro de localización,

$$\begin{aligned} P(\mu, \tau_1^2) &= N(\mu | \delta, \gamma \tau_1^2) \times GI(\tau_1^2 | a, b) \\ &= c(a, b) \frac{1}{\sqrt{\gamma}} \left( \frac{1}{\tau_1^2} \right)^{\frac{1}{2} + a + 1} e^{-\frac{1}{2\tau_1^2} (b + \frac{1}{\gamma} (\mu - \delta)^2)} \end{aligned}$$

y se obtiene que

$$P(\mu_1, \dots, \mu_t; \mu, \tau_1^2) = c(a, b) \frac{1}{\sqrt{\gamma}} \left( \frac{1}{\tau_1^2} \right)^{\frac{t}{2} + \frac{1}{2} + a + 1} e^{-\frac{1}{2\tau_1^2} (b + \frac{1}{\gamma} (\mu - \delta)^2 + \frac{1}{2} \sum_{i=1}^t (\mu_i - \mu)^2)}. \quad (3.16)$$

De manera análoga, al hacer los cálculos para el parámetro de escala tenemos

$$\begin{aligned} P(\phi, \tau_2^2) &= N(\phi | \eta, \nu \tau_2^2) \times GI(\tau_2^2 | \alpha, \beta) \\ &= c(\alpha, \beta) \frac{1}{\sqrt{\nu}} \left( \frac{1}{\tau_2^2} \right)^{\frac{1}{2} + \alpha + 1} e^{-\frac{1}{2\tau_2^2} (b + \frac{1}{\nu} (\phi - \eta)^2)} \end{aligned}$$

y se obtiene que

$$P(\phi_1, \dots, \phi_t; \phi, \tau_2^2) = c(\alpha, \beta) \frac{1}{\sqrt{\nu}} \left( \frac{1}{\tau_2^2} \right)^{\frac{t}{2} + \frac{1}{2} + \alpha + 1} e^{-\frac{1}{2\tau_2^2} (b + \frac{1}{\nu} (\phi - \eta)^2 + \frac{1}{2} \sum_{i=1}^t (\phi_i - \phi)^2)}. \quad (3.17)$$

Con esta parametrización, para tener una distribución no informativa puede realizarse una de las siguientes opciones

- (i) Hacer  $a$ ,  $b$ ,  $\alpha$  y  $\beta$  muy pequeñas.
- (ii) Hacer  $\nu$  y  $\gamma$  grandes, es decir, se consideraría que las varianzas de  $\mu$  y  $\phi$  tomen valores grandes y las densidades normales por lo tanto tenderán a uniformes sobre toda la recta.

Es importante notar que si se realizaran (i) y (ii) al mismo tiempo se podrían obtener distribuciones posteriores cercanas a no ser propias.

Con las especificaciones que se dieron a lo largo de esta sección y la 3.1, podemos expresar la distribución posterior como

$$P(\Xi | \mathbf{y}_1, \dots, \mathbf{y}_{11}) = \prod_{i=1}^{11} L_i(\mathbf{y}_i | \mathcal{T}^{-1}(\Xi)) \times P(\Xi)$$

De tal manera que

$$\begin{aligned} P(\Xi | \mathbf{y}_1, \dots, \mathbf{y}_{11}) &\propto \prod_{i=1}^{11} \left\{ e^{-(1+\epsilon e^{-\phi_i}(u_i - \mu_i))_+^{-1}} \right. \\ &\quad \times \prod_{i \in A_i} \frac{1}{\phi_i} \left( 1 + \xi e^{-\phi_i}(y_i - \mu_i) \right)_+^{-(1+\xi)} \\ &\quad \times \left. \prod_{i \in B_i} \left( 1 + \xi e^{-\phi_i}(k_i - \mu_i) \right)_+^{-1} \right\} \\ &\quad \times \left( \frac{1}{\tau_1^2} \right)^{\frac{\tau}{2} + \frac{1}{2} + \alpha + 1} e^{-\frac{1}{\tau_1^2} (\alpha + \frac{1}{2}(\mu - \alpha)^2 + \frac{1}{2} \sum_{i=1}^{\tau} (\mu_i - \mu)^2)} \\ &\quad \times \left( \frac{1}{\tau_2^2} \right)^{\frac{\tau}{2} + \frac{1}{2} + \alpha + 1} e^{-\frac{1}{\tau_2^2} (\beta + \frac{1}{2}(\phi - \eta)^2 + \frac{1}{2} \sum_{i=1}^{\tau} (\phi_i - \phi)^2)} \\ &\quad \times \frac{1}{0,4 - 0,01} I_{(0,01,0,4)}(\xi) \times e^{\phi_1} \cdot e^{\phi} \end{aligned}$$

El parámetro  $\Xi$  contiene en este caso un total de 27 parámetros, esto es  $\Xi = (\mu_1, \mu_2, \dots, \mu_{11}, \mu, \tau_1^2, \phi_1, \phi_2, \dots, \phi_{11}, \phi, \tau_2^2, \xi)$ , los cuales será necesario estimar a fin de establecer la distribución posterior de las reclamaciones.

### 3.2.2. Efectos aleatorios sólo en escala

Si estudiamos únicamente los efectos aleatorios en la escala, se omiten los parámetros  $(\mu_1, \dots, \mu_T)$  del modelo jerárquico descrito en la subsección anterior y se sustituyen por un sólo parámetro de escala,  $\mu$ , cuyo valor no dependerá de  $t$ , el número de año. Entonces el modelo jerárquico que se propone en este caso es

Nivel III	$(\delta, \gamma)$	$\phi, \tau_2^2$	$(\text{liminf}(\xi), \text{limsup}(\xi))$
	↓	↙ ... ↘	↓
Nivel II	$\mu$	$\phi_1, \dots, \phi_{11}$	$\xi$
	↓ ... ↓	↓ ... ↓	
Nivel I	$y_1 \dots y_{11}$	$y_1 \dots y_{11}$	

El límite inferior y superior de  $\xi$  se fijaron como 0,01 y 0,49, respectivamente.

Se propone entonces una distribución normal para el parámetro general de localización que no dependerá del tiempo, de la siguiente forma

$$\begin{aligned}
 P(\mu) &= \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{1}{2\gamma^2}(\mu - \delta)^2} \\
 &= N(\mu|\delta, \gamma^2)
 \end{aligned}$$

y los hiperparámetros  $\delta$  y  $\gamma^2$  se fijarán de manera que la distribución inicial de  $\mu$  refleje poca informatividad respecto al valor de la localización y de esta manera los datos incorporen información al modelo.

Por otra parte, se supondrá independencia a priori entre las distribuciones de  $\mu$ ,  $(\phi_1, \dots, \phi_t, \phi, \tau_2^2)$  y  $\xi$ ; de tal forma que

$$P(\Xi) = P(\mu) \times P(\phi_1, \dots, \phi_T, \phi, \tau_2^2) \times \left( \frac{1}{0,4 - 0,01} \right) J_{(0,01,0,4)}(\xi) \times |J_{T-1}|.$$

y de manera análoga a como se obtuvo la distribución posterior en el modelo que contempla efectos aleatorios anuales tanto en localización y escala, podemos expresar la distribución posterior en este caso como sigue

$$\begin{aligned} P(\Xi | y_1, \dots, y_{11}) &\propto \prod_{t=1}^{11} \left\{ e^{-(1+\epsilon^{-\phi_t}(1-\mu))_+^{-1}} \right. \\ &\times \prod_{i \in A_t} \frac{1}{\phi_t} \left( 1 + \xi e^{-\phi_t} (y_i^t - \mu) \right)_+^{-(1+t)} \\ &\times \prod_{i \in B_t} \left( 1 + \xi e^{-\phi_t} (k_i - \mu) \right)_+^{-t} \\ &\times e^{-\frac{1}{2\tau_2^2}(\mu-\phi)^2} \\ &\times \left( \frac{1}{\tau_2^2} \right)^{\frac{T}{2} + \frac{1}{2}n + 1} c^{-\frac{1}{2}(\beta + \frac{1}{\tau_2}(\phi-\eta)^2 + \frac{1}{2} \sum_{t=1}^T (\phi_t - \phi)^2)} \\ &\times \left. \frac{1}{0,4 - 0,01} J_{(0,01,0,4)}(\xi) \times e^{\phi_t} \cdot e^{\phi} \right\}. \end{aligned}$$

En este caso  $\Xi$  se reduce a  $\Xi = (\mu, \phi_1, \phi_2, \dots, \phi_{11}, \phi, \tau_2^2, \xi)$  por lo que se necesitaría conocer el valor de 15 parámetros para establecer la distribución posterior de las reclamaciones.

En la siguiente sección se propone un método para estimar el valor de  $\Xi$  tanto para el modelo con efectos aleatorios anuales en localización y escala como para el modelo con efectos aleatorios anuales únicamente en la escala, usando métodos de simulación Monte Carlo y las distribuciones posteriores descritas para cada caso.

### 3.3. Simulación de la distribución posterior vía Metrópolis-Hastings

Por conveniencia numérica y a fin de simular valores para los parámetros de las distribuciones marginales posteriores en los modelos 3.2.1 y 3.2.2, consideramos el logaritmo negativo de la distribución posterior correspondiente a cada caso, de la siguiente manera.

#### Caso I: Efectos aleatorios anuales en localización y escala

Para el logaritmo negativo de la distribución inicial conjunta de

$$(\mu_1, \dots, \mu_{11}, \mu, \tau_1^2, \phi_1, \dots, \phi_{11}, \phi, \tau_2^2, \xi)$$

se tiene

$$\begin{aligned} -\log[P(\Xi)] &= -\log(c_1) + \left(\frac{T}{2} + a + 1 + \frac{1}{2}\right) \log(\tau_1^2) + \frac{1}{\tau_1^2} \left[ b + \frac{1}{2\gamma}(\mu - \delta)^2 + \frac{1}{2} \sum_{i=1}^T (\mu_i - \mu)^2 \right] \\ &- \log(c_2) + \left(\frac{T}{2} + \alpha + 1 + \frac{1}{2}\right) \log(\tau_2^2) + \frac{1}{\tau_2^2} \left[ \beta + \frac{1}{2\nu}(\phi - \eta)^2 + \frac{1}{2} \sum_{i=1}^T (\phi_i - \phi)^2 \right] \\ &- \log \left[ \frac{1}{0,4 - 0,01} I_{(0,01,0,4)}(\xi) \right] - \phi - \sum_{i=1}^T \phi_i \end{aligned}$$

por lo que el logaritmo negativo de la distribución posterior final puede expresarse como

$$\begin{aligned} -\log[P(\Xi | \mathbf{y}_1, \dots, \mathbf{y}_{11})] &= -\sum_{i=1}^{11} \log[L_i(\mathbf{y}_i)] - \log[P(\Xi)] \\ &\triangleq \sum_{i=1}^{11} (1 + \xi e^{-\phi_i(u - \mu_i)})_+^{-1} \\ &+ \sum_{i=1}^{11} \sum_{i \in A_i} \left\{ \phi_i + \left(1 + \frac{1}{\xi}\right) \cdot \log \left[ (1 + \xi e^{-\phi_i(y_i^t - \mu_i)})_+ \right] \right\} \end{aligned}$$



$$\begin{aligned}
& + \sum_{i=1}^{11} \sum_{i \in B_i} \left\{ \left( \frac{1}{\xi} \right) \cdot \log \left[ (1 + \xi e^{-\phi_i} (k_i - \mu_i))_+ \right] \right\} \\
& + \frac{T}{2} + \alpha + \frac{3}{2} \cdot \log(\tau_1^2) + \frac{1}{\tau_1^2} \left( b + \frac{1}{2\gamma} (\mu - \delta)^2 + \frac{1}{2} \sum_{i=1}^{11} (\mu_i - \mu)^2 \right) \\
& + \frac{T}{2} + \alpha + \frac{3}{2} \cdot \log \left( \frac{1}{\tau_2^2} \right) \frac{1}{2\tau_2^2} \left( \beta + \frac{1}{2\nu} (\phi - \eta)^2 + \frac{1}{2} \sum_{i=1}^{11} (\phi_i - \phi)^2 \right) \\
& - \log \left( \frac{1}{0,4 - 0,01} \mathcal{I}_{(0,01,0,4)}(\xi) \right) - \sum_{i=1}^{11} \phi_i - \phi.
\end{aligned}$$

donde  $\hat{\Xi}$  significa "en relación a una constante aditiva respecto a  $\Xi$ ".

### Caso II: Efectos aleatorios anuales sólo en escala

En este caso el logaritmo negativo de la distribución inicial conjunta  $(\mu, \phi_1, \dots, \phi_{11}, \phi, \tau_2^2, \xi)$  será

$$\begin{aligned}
-\log(\mathcal{I}(\Xi)) &= -\log(c_1) + \frac{1}{2\gamma} (\mu - \delta)^2 \\
&- \log(c_2) + \left( \frac{T}{2} + \alpha + 1 + \frac{1}{2} \right) \log(\tau_2^2) + \frac{1}{\tau_2^2} \left[ \beta + \frac{1}{2\nu} (\phi - \eta)^2 + \frac{1}{2} \sum_{i=1}^T (\phi_i - \phi)^2 \right] \\
&- \log \left[ \frac{1}{0,4 - 0,01} \mathcal{I}_{(0,01,0,4)}(\xi) \right] - \phi - \sum_{i=1}^T \phi_i
\end{aligned}$$

y el logaritmo negativo de la distribución posterior se expresa como

$$\begin{aligned}
-\log[\mathcal{I}(\Xi | \mathbf{y}_1, \dots, \mathbf{y}_{11})] &= - \sum_{i=1}^{11} \log[L_i(\mathbf{y}_i)] - \log[\mathcal{I}(\Xi)] \\
&\hat{=} \sum_{i=1}^{11} \left( 1 + \xi e^{-\phi_i} (u - \mu) \right)_+^{-1}
\end{aligned}$$

*Simulación del valor de reclamaciones futuras y estimación de probabilidades de pérdida* 69

$$\begin{aligned}
 & + \sum_{t=1}^{11} \sum_{i \in A_t} \left\{ \phi_t + \left(1 + \frac{1}{\xi}\right) \cdot \log \left[ (1 + \xi e^{-\phi_t} (v_i^t - \mu))_+ \right] \right\} \\
 & + \sum_{t=1}^{11} \sum_{i \in B_t} \left\{ \left(\frac{1}{\xi}\right) \cdot \log \left[ (1 + \xi e^{-\phi_t} (k_i - \mu))_+ \right] \right\} \\
 & + \frac{1}{2\gamma} (\mu - \delta)^2 \\
 & + \frac{T}{2} + \alpha + \frac{3}{2} \cdot \log \left( \frac{1}{\tau^2} \right) \frac{1}{2\tau^2} \left( \beta + \frac{1}{2\nu} (\phi - \eta)^2 + \frac{1}{2} \sum_{t=1}^{11} (\phi_t - \phi)^2 \right) \\
 & - \log \left( \frac{1}{0,4 - 0,01} \mathbf{I}_{(0,01,0,4)}(\xi) \right) - \sum_{t=1}^{11} \phi_t - \phi.
 \end{aligned}$$

donde  $\triangleq$  significa "en relación a una constante aditiva respecto a  $\Xi$ ".

### Simulación de los parámetros

Para estimar los parámetros en ambos casos, de acuerdo con Smith y Roberts (1993), se llevó a cabo una simulación mediante un algoritmo "híbrido" Gibbs-Metrópolis-Hastings, es decir, se actualizaba un valor de cada parámetro a la vez mediante su distribución posterior vía el muestreo de Gibbs, después se modificaba el valor del parámetro por una cantidad aleatoria pequeña y finalmente se usó el método de Metrópolis-Hastings para decidir si aceptar o rechazar el valor modificado.

De esta forma, el proceso de estimación es una simulación de Monte Carlo vía Cadenas de Markov en la que la actualización de los valores de los parámetros desconocidos se realizó de acuerdo al modelo jerárquico descrito en la sección 3.2.1 para el Caso I y de acuerdo al modelo jerárquico que se especificó para el caso II en la sección 3.2.2.

En ambos casos, se generaron cadenas de tamaño 15,000 para cada parámetro y se trabajó con los últimos 5,000 valores para simular el valor de los excesos y con ellos realizar una estimación de las probabilidades de pérdida futuras de acuerdo a la metodología que se describe en la siguiente sección.

### 3.4. Simulación del valor de reclamaciones futuras y estimación de probabilidades de pérdida

#### Simulación del valor de las reclamaciones

Dado el valor  $\Xi^{(i)} = (\mu_1^{(i)}, \dots, \mu_{11}^{(i)}, \mu^{(i)}, \tau_1^{2(i)}, \phi_1^{(i)}, \dots, \phi_{11}^{(i)}, \phi^{(i)}, \tau_2^{2(i)})$ ,  $i \in N$  en el Caso I y  $\Xi^{(i)} = (\mu^{(i)}, \phi_1^{(i)}, \dots, \phi_{11}^{(i)}, \phi^{(i)}, \tau_2^{2(i)})$ ,  $i \in N$  en el Caso II y para cada año  $t = 1, \dots, 11$ , se simulan  $j$  procesos de reclamaciones,  $j \in N$  en los siguientes pasos

- (i) Se simula una variable aleatoria  $N_t$  del número de excedentes en cada año tal que

$$N_t \sim \text{Poisson}(\lambda_t)$$

donde

$$\lambda_t = \left(1 + \xi e^{(-\phi_t)}(u - \mu_t)\right)^{(-\frac{1}{\xi})}$$

- (ii) Dado el valor de  $N_t$  en (i), se simula el valor de los excesos  $x_{t,1}, x_{t,2}, \dots, x_{t,N_t}$  como variables aleatorias que siguen una Distribución Pareto Generalizada

$$F_{X_{t,j}}(x) = 1 - \left(1 + \frac{\xi}{\sigma_t} x\right)^{-\frac{1}{\xi}}$$

donde  $\sigma_t = e^{\phi_t} + \xi(u - \mu_t)$ .

- (iii) Se calcula  $y_{t,i}$  como  $x_{t,i} + u$  donde  $u$  es el valor del umbral.

El paso (i) puede realizarse directamente. El paso (ii) se puede llevar a cabo vía Inversión, como se describe enseguida

1. Generar  $y \sim \text{Unif}(0, 1)$ .
2. Sustituir el valor calculado en (i) y el valor de los parámetros correspondientes en

$$x = \frac{\sigma \epsilon}{\xi} \left( (1 - y)^{(-\xi)} - 1 \right).$$

Después de obtener los valores de los excesos, el paso (iii) para calcular el valor de las reclamaciones es directo.

De esta forma se obtienen  $j$  procesos de reclamación

$$y_{1,1}^{(j)} \dots y_{1,N_1}^{(j)}; \dots; y_{11,1}^{(j)} \dots y_{11,N_{11}}^{(j)}.$$

### Estimación de las probabilidades de pérdida

Para cada  $j$ -ésimo proceso de reclamación se calcula  $Z^{(j)}$ , el valor los excesos acumulados para cada año  $t$ ,

$$Z^{(j)} = \sum_{t=1}^{11} \sum_{k=1}^{N_t} y_{t,k}^{(j)}.$$

Después usamos el valor de  $Z^{(j)}$  para obtener un estimador empírico de la distribución predictiva

$$p_n \equiv P[Z > z^*] = \frac{\sum_{j=1}^n \mathbb{1}_{\{z^{(j)} - z^* > 0\}}}{n}.$$

Finalmente, si ordenamos los valores estimados  $\hat{p}_n$ , es posible obtener intervalos al  $\alpha \times 100\%$  de confianza con  $\alpha \in (0, 1)$ , para la probabilidad de pérdida, si se toman los cuantiles de orden  $(\frac{\alpha}{2})$  y  $(1 - \frac{\alpha}{2})$  como valores para los límites inferior y superior de los intervalos, respectivamente.

Hiperparámetro	Caso I	Caso II
$\delta$	2200	2500
$\gamma$	100	10
$a$	10	0.5
$b$	10	0.5
$\eta$	5.0	6.4
$\nu$	50.0	1.0
$\alpha$	10.0	0.5
$\beta$	10.0	0.5

Cuadro 3.3: Valores para los hiperparámetros

### 3.5. Análisis de resultados

A continuación se explica la manera en que se implementaron las metodologías descritas en las secciones 3.3 y 3.4, para ambos modelos y valores de umbral  $u = 45$  y  $u = 300$ . Los gráficos o figuras de las que se hace referencia en esta sección se realizaron para los dos valores de  $u$  mencionados.

#### Efectos aleatorios anuales en localización y escala

Se analizaron los resultados para diferentes valores de hiperparámetros, de acuerdo con las ideas explicadas al final de la sección 3.2.1 para tener una distribución inicial del parámetro no informativa.

Una vez elegidos los valores de hiperparámetros se simularon cadenas para  $\Xi$  y se analizó la convergencia de las mismas. Enseguida se hablará de los resultados que se obtuvieron únicamente con dos casos que se consideraron de mayor importancia y cuyos valores de hiperparámetros se presentan en la tabla 3.3.

Adicionalmente, en cada caso se consideraron los valores para el umbral,  $u$  de 45 y 300, por lo que en total se mostrarán los resultados para cuatro casos.

Como se mencionó en la sección 3.3, en todos los casos se generaron cadenas de tamaño

Hiperparámetro	Valor
$\delta$	3500
$\gamma$	1000
$\eta$	6.4
$\nu$	1.0
$\alpha$	0.5
$\beta$	0.5

Cuadro 3.4: Valores para los hiperparámetros

15,000, se consideró un "periodo de calentamiento" de 10,000 y se trabajó con las últimas 5,000 muestras del parámetro.

Con los valores de  $\Xi^{(1)}, \dots, \Xi^{(5000)}$  se verificó la convergencia de las cadenas para cada parámetro considerado en el modelo, mediante los gráficos de promedios ergódicos. En el apéndice D se muestran las figuras de los promedios ergódicos de algunos parámetros seleccionados para los Casos I y II y los valores de umbral  $u = 45$  y  $u = 300$ .

Una vez comprobada la convergencia, se analizaron las autocorrelaciones de las cadenas y se eligió una de cada 100 observaciones; al estudiar la autocorrelación de las nuevas cadenas se comprobó que ésta ya no fuera significativa.

Las figuras en el apéndice D muestran los diagramas de caja para las densidades marginales posteriores de los parámetros de localización y escala, para el caso de las localizaciones puede observarse que el efecto en estos parámetros no es significativo en ninguno de los casos, en el sentido de que la distribución final para los parámetros de localización no cambia significativamente de un año a otro.

#### Efectos aleatorios anuales sólo en escala

En virtud de que aparentemente modelar efectos aleatorios en la localización no nos da información relevante, se planteó estudiar el modelo sin éstos, con valores para los hiperparámetros que se muestran en la tabla 3.4 y valores de umbral  $u = 45$  y  $u = 300$ , por lo que se analizaron un total de 2 casos para este modelo.

Los valores que se fijaron para los hiperparámetros de  $\mu$  se eligieron de tal manera que la distribución inicial en la escala no influyera en el valor final del parámetro y permitiera que los datos aportaran la mayor información posible a la distribución final.

En este caso se usaron también muestras de tamaño 5,000, después de eliminar 10,000 simulaciones del "periodo de calentamiento". Con el objeto de analizar la convergencia de la cadena se calcularon y graficaron los promedios ergódicos para cada uno de los parámetros (ver figuras en apéndice E).

Un análisis previo de autocorrelogramas nos lleva a tomar una de cada 100 observaciones de forma que las autocorrelaciones de cada parámetro no sean significativas como puede verificarse en los autocorrelogramas que se anexan en el apéndice E.

Los diagramas de caja para las densidades marginales posteriores de los parámetros considerados en este modelo se muestran en el apéndice E. En este caso se observa que el efecto en los parámetros es significativo.

Dado que en el modelo que considera efectos aleatorios anuales para la localización y la escala no se observaron cambios significativos en los parámetros, se simularon valores futuros de reclamación de acuerdo con Smith (1996) y la metodología explicada en la sección 3.4, sólo para el modelo que considera que el parámetro de escala cambia con el año de reclamación  $t$ .

Una vez simulados los procesos de reclamación se calcularon las sumas acumuladas  $Z^{(j)}$  para  $j = 1, \dots, 99$  para todas las reclamaciones de los años 1 a 11.

Con los valores  $Z^{(1)}, \dots, Z^{(99)}$  y diferentes valores de  $z^*$ , la máxima cantidad esperada del total de reclamaciones acumuladas, se calcularon estimaciones para las probabilidades de pérdida  $\hat{p}_1, \dots, \hat{p}_{99}$  y dados los valores estimados  $\hat{p}_n$ ,  $n = 1, \dots, 99$ , se tomaron los valores de  $\hat{p}_5$  y  $\hat{p}_{94}$  como el límite inferior y superior, respectivamente, para obtener intervalos al 95% de confianza de la probabilidad de pérdida dados los diferentes valores de  $z^*$ .

En el apéndice E se muestra el histograma del monto de las reclamaciones simuladas para dos procesos de este modelo: uno con  $u = 45$  y otro con  $u = 300$ . El número de reclamaciones por año,  $N_1, N_2, \dots, N_{11}$ , que se simularon para esos mismos procesos se

$t$	$N_t$ con $u = 45$	$N_t$ con $u = 300$
1	65	27
2	44	26
3	52	37
4	62	32
5	89	39
6	83	49
7	64	50
8	74	30
9	43	12
10	26	9
11	9	4

Cuadro 3.5: Número de reclamaciones  $N_t$  simuladas por año

muestran en la siguiente tabla.

Si se comparan

- (i) El histograma y los valores  $N_1, N_2, \dots, N_{11}$  que se obtuvieron para el proceso de simulación de las reclamaciones con  $u = 45$  (Tabla 3.5), contra
- (ii) El histograma y valores para el número de reclamaciones que se mostraron en la figura 3.2 y en la tabla 3.1 del capítulo 3, correspondientes a la serie de datos original,

se puede decir que el modelo con efectos aleatorios sólo en el parámetro de escala, parece incorporar la información de la muestra de forma adecuada y entonces puede pensarse que los valores simulados de la distribución final del parámetro, las reclamaciones futuras y la estimación de las probabilidades de pérdida arrojarán resultados que serán una buena aproximación de lo que pudiera suceder en un nuevo proceso de reclamaciones. Cuando se usó el umbral  $u = 300$  en este modelo, los valores del número de reclamaciones anuales fueron menores a los que se obtuvieron con  $u = 45$  (Tabla 3.5) en la mayoría de los años,



esta situación es intuitivamente congruente con lo que sucedería en la realidad, ya que al elevar el valor de umbral  $u$ , serán menos las reclamaciones que lo sobrepasen en cada año  $t$ .

Los intervalos de confianza al 95 % para la probabilidad de pérdida respecto al monto acumulado de reclamaciones  $z^*$  se anexan al final del apéndice E. Se observa que para valores pequeños de  $z^*$ , estos intervalos son de poca amplitud y contienen valores cercanos a uno; a medida que se eleva  $z^*$ , la amplitud de los intervalos es mayor y sus límites se van haciendo más pequeños, sin embargo, si se consideran valores muy grandes de  $z^*$  la amplitud de los intervalos disminuye y sus límites toman valores que van acercándose a cero, el límite inferior a una velocidad mayor que el superior. Este comportamiento es intuitivamente adecuado pues mientras más grande se considere  $z^*$ , disminuye la probabilidad de observar montos de esa magnitud. La importancia de estos intervalos radica entonces, en que a partir de ellos se puede tener idea de qué tan factible será observar valores de determinada magnitud y a partir de esto, es posible definir por ejemplo, en el caso de una reaseguradora, si se tomarán posiciones de alto riesgo, dependiendo de los montos de pérdida esperados en caso de que ocurrieran siniestros de tipo catastrófico.

En el apéndice A puede consultarse el código en S+ con el que se simularon los procesos de reclamación y el valor estimado para las probabilidades de pérdida.

## CONCLUSIONES

En diversos problemas en que se modelan datos usando métodos estadísticos, es común que exista más de un modelo que proporcione una aproximación adecuada de lo que se desea estimar. Por esta razón, es fundamental hacer una evaluación del modelo que se ha propuesto en el último capítulo de este trabajo.

Dadas las características de los datos y el fenómeno que describen, una buena opción es haber propuesto un modelo a partir de las ideas que se han desarrollado en la Teoría de Valores Extremos, con un entorno Bayesiano. Bajo este esquema, la forma del modelo que se propuso está perfectamente sustentado de manera formal y el problema se convierte en un asunto de elección y aplicación de herramientas estadísticas para definir de manera adecuada los parámetros y la manera mas conveniente de simular sus valores y posteriormente valores predictivos de nuevas reclamaciones.

De acuerdo con lo que se dijo en la sección 3.5, considerar efectos aleatorios anuales en los parámetros de localización no contribuye de forma significativa para explicar los datos. La validez de esta afirmación puede constatarse en las figuras D.8 y D.11 del Apéndice D, en donde se observa que la distribución marginal posterior de los parámetros  $\mu_i$ ,  $i = 1, \dots, 11$ , no cambian considerablemente de un año a otro. Esta observación indujo el análisis de los datos usando el modelo inicial pero suponiendo que el parámetro de localización podía considerarse independiente del año  $t$ , para  $t = 1, \dots, 11$ . El parámetro de escala sí se modeló usando efectos aleatorios anuales. Con éste último modelo se obtuvieron resultados que corroboraron el hecho de que los efectos aleatorios anuales en la escala son significativos (ver Figuras E.7 y E.17).

Por las razones anteriores, el segundo modelo puede considerarse adecuado para describir los datos y simular nuevos procesos de reclamacione. Si se comparan las Tablas 3.5 y

3.1, se observa que con este modelo, los valores simulados para el número de reclamaciones anuales que se ilustran en la Tabla 3.5, para  $u = 45$  son similares a los que se tenían en la Tabla 3.1, como parte de la información inicial de la serie de datos. Por otra parte, el número de reclamaciones por año simuladas cuando el umbral se eleva a 300, es menor en relación a los valores que se obtienen con  $u = 45$ , lo que corresponde a un comportamiento que coincide con lo que se esperaría intuitivamente (ver Tabla 3.5).

Tomando en cuenta que se definieron distribuciones iniciales poco informativas, puede afirmarse que la manera en que se estableció el modelo que considera efectos aleatorios en escala solamente, permite que la muestra aporte suficiente información para obtener distribuciones posteriores de los parámetros, que caracterizan adecuadamente la distribución de las reclamaciones, esto se confirma en las Figuras E.7 y E.17.

Por otra parte, al estudiar los histogramas de frecuencias para valores predictivos de nuevas reclamaciones, que se obtuvieron después de simular valores de los parámetros, se puede afirmar que las simulaciones son congruentes con lo que se esperaría en un nuevo proceso de reclamaciones (ver Figuras E.9 y E.18).

Por otra parte, al analizar los intervalos de confianza obtenidos en el Apéndice E, la reaseguradora podría establecer límites de responsabilidad de manera conservadora o de mediano o alto riesgo, pero con bases sustentadas en la factibilidad de incurrir en posibles pérdidas que pudieran afectar su estabilidad financiera. Así por ejemplo, si se analizan los intervalos de confianza obtenidos para un umbral  $u = 45$ , puede suceder que con una probabilidad mayor a 0,55, la reaseguradora llegara a tener pérdidas de  $z^* = 395,000$ . Lo mismo sucedería al considerar el umbral  $u = 300$  y un valor  $z^*$  de 701,000.

La ventaja de haber propuesto un modelo completamente Bayesiano consiste en que una vez que se cuente con nueva información sobre el proceso de reclamaciones descrito en el Capítulo 3, sería posible incorporarla al modelo ya establecido y obtener nuevas estimaciones y conclusiones, no sin antes llevar a cabo una validación del modelo modificado.

La validación formal y rigurosa del modelo con efectos aleatorios anuales sólo en escala que se propuso, así como de ajustes posteriores que se hicieran al mismo, se puede llevar a cabo como trabajo posterior, siguiendo la línea de investigación de los estudios referentes a la robustez de los modelos Bayesianos y análisis de sensibilidad de las inferencias poste-

riores debido a los supuestos iniciales, que se han llevado a cabo por investigadores como Leamer (1978), Berger (1984), Berger y Berliner (1986), McCulloch (1989) y Wasserman (1992), entre muchos otros. Algunos de estos métodos para verificar el modelo, incluyen pruebas para observaciones aberrantes, gráficos de residuales y gráficos normales y se consideran como pruebas Bayesianas de tipo predictivo-posteriores, en las que se buscan incongruencias entre los resultados obtenidos y los esperados bajo el modelo supuesto. De forma paralela, pueden emplearse técnicas no Bayesianas para checar los modelos y sus modificaciones como se realiza por ejemplo en Atkinson (1985).

**ESTA TESIS NO SALE  
DE LA BIBLIOTECA**



## Apéndice A

### Programas en S+

A continuación se presentan los códigos de S-Plus que se emplearon para realizar el muestreo de Gibbs en el apartado 1.6.1 de este trabajo.

```
gibbs <- function(n,t,Y)
{
  set.seed(100)
  Lam <- {1:n}
  Fi <- {1:n}
  Testrella <- {1:n}
  if (t<52)
  {
    Y1_Y[1:t]
    Y2_Y[(t+1):52]
  }
  if (t==52)
  {
    Y1_Y
    Y2_0
  }
  sum1_sum(Y1)
  sum2_sum(Y2)
```

```

Lam[1] <- rgamma(1,1+sum1,0.01+t)
Fi[1] <- rgamma(1,1+sum2,0.01+52-t)
vec <- FD(Lam[1],Fi[1],Y)
Prob <- vec$p
z <- runif(1)

if (Prob[1]>=z)
  { Testrella[1]_1 }

for(j in 1:51)
  {
    if ((Prob[j]<z)&&(z<=Prob[j+1]))
      {Testrella[1]_j+1}
  }
for(i in 2:n)
  {
    permit <- 0
    if (Testrella[i-1]<52)
      {
        Y1_Y[1:Testrella[i-1]]
        Y2_Y[(Testrella[i-1]+1):52]
      }
    if (Testrella[i-1]==52)
      {
        Y1_Y
        Y2_0
      }
    sum1_sum(Y1)
    sum2_sum(Y2)
    while(permit==0){
      Lam[i] <- rgamma(1,1+sum1,0.01+Testrella[i-1])
      Fi[i] <- rgamma(1,1+sum2,0.01+52-Testrella[i-1])
      vec <- FD(Lam[i],Fi[i],Y)
      permit <- vec$toss
    }
  }

```

```
Prob <- vec$p
z <- runif(1)
if (z<=Prob[1])
  { Testrella[i]_1 }
for(j in 1:51)
  {
    if ((Prob[j]<z)&&(z<=Prob[j+1]))
      { Testrella[i]_j+1 }
  }
print(cbind(i))
}
write(Lam, file="c:/sp2000/users/AMR/lamda.out", ncol=1)
write(Fi, file="c:/sp2000/users/AMR/fi.out", ncol=1)
write(Tes, file="c:/sp2000/users/AMR/Tes.out", ncol=1)
}
```



```
FD <- function(lamda,fi,Y)
{
  resultado <- 0
  denominador_funcion(1,lamda,fi,Y)
  for (l in 2:52)
  {
    denominador_denominador+funcion(l,lamda,fi,Y)
  }
  probab <- {1:52}
  probab[1]_funcion(1,lamda,fi,Y)
  for (i in 2:52)
  {
    probab[i]_funcion(i,lamda,fi,Y) + probab[i-1]
  }
  print(cbind(denominador))
  if(denominador<5.345274e+213)
  {
    probab <- probab/denominador
    resultado <- 1
  }

  funcion <- function(Te,lamda,fi,Y)
  {
    if (Te<52)
    {
      Y1 <- Y[1:Te]
      Y2 <- Y[(Te+1):52]
    }
    if (Te == 52)
    {
      Y1 <- Y
      Y2 <- 0
    }
  }
  sum1_sum(Y1)
```

```

sum2_sum(Y2)
f_exp(log(lamda)*sum1+log(fi)*sum2-Te*(lamda-fi))
return(f)
}

fdvec <- list(p=probab,toss=resultado)

fdvec
}
}

```

El siguiente código corresponde al programa que se empleó para simular los valores de los excesos de acuerdo a la distribución posterior obtenida en la sección 3.4.

```

excess<-function(n,r){
  u_45
  xit_sxis
  lim_380000
  Prob_matrix(0,n,1)
  zfin_0
  for(m in 1:n){
    zn_matrix(0,r,1)
    ind_0
    for(i in 1:r){
      resulta_matrix(0,1,150)
      xt_matrix(0,11,150)
      z_0
      mut_smus[m]
      phit_sfis[m,]
      lamdat_(1+(xit [m]*exp(-phit)*(u-mut)))^(-1/xit [m])
      sigmat_exp(phit)+(xit [m]*(u-mut))
      for(j in 1:11){
        Nt_rpois(1,lamdat [j])
        for(k in 1:Nt){

```

```
xt[j,k]_(sigmat[j]/(xit[m]))*(((1-(runif(1)))^(-xit[m]))-1))
}
yt[j]_sum(u+xt[j,])
z_z+yt[j]
}
zn[i]_z
if(zn[i] > lim){
  ind_ind+1
}
resulta_rbind(resulta,xt)
}
zfin_rbind(zfin,zn)
Prob[m]_ind/r
print(cbind(m))
}
write(t(resulta), file="c:/sp2000/users/AMR/excess.out")
write(t(zfin), file="c:/sp2000/users/AMR/zfin.out")
write(t(Prob), file="c:/sp2000/users/AMR/Prob.out")
}
```

## Apéndice B

# Aspectos formales de la Teoría de Valores Extremos

**Definición B.1** *Función inversa generalizada de una función monótona*

Considérese la inversa generalizada para una función monótona no decreciente  $h$

$$h^{\leftarrow}(t) \equiv \inf\{x \in \mathbf{R} : h(x) \geq t\}$$

En particular, si  $t \in [0, 1]$  y  $F$  es una función de distribución

$$F^{\leftarrow}(t) = \inf\{x \in \mathbf{R} : F(x) \geq t\}$$

Si  $x$  es un número real mayor que 1 entonces

$$\begin{aligned} F^{\leftarrow}\left(1 - \frac{1}{x}\right) &= \inf\left\{x \in \mathbf{R} : F(x) \geq 1 - \frac{1}{x}\right\} \\ &= \inf\left\{x \in \mathbf{R} : F(x) - 1 \geq -\frac{1}{x}\right\} \\ &= \inf\left\{x \in \mathbf{R} : \frac{1}{F(x)} \geq x\right\} \\ &= \left(\frac{1}{F}\right)^{\leftarrow}(x) \end{aligned}$$

**Definición B.2** *Variación regular*

- (i) Una función positiva y Lebesgue-medible  $L$  definida sobre  $(0, \infty)$  se dice que es de variación lenta en  $\infty$  si

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0$$

y se denota  $L \in \mathcal{R}_0$ .

- (ii) Una función positiva, Lebesgue medible  $L$ , definida sobre  $(0, \infty)$  se llama de variación regular en  $+\infty$  con índice  $\alpha \in \mathcal{R}$  si

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha, \quad t > 0$$

y se denota  $h \in \mathcal{R}_\alpha$ . En particular para  $F$  una función de distribución y  $\bar{F}(x) = 1 - F(x)$  la cola de la distribución, diremos que  $\bar{F}$  es de variación regular con índice  $-\alpha$  para la cola de la distribución y algún  $\alpha \geq 0$  (notación  $\bar{F} \in \mathcal{R}_{-\alpha}$ ) si

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\alpha}, \quad t > 0.$$

La demostración de la proposición B.1 y los teoremas B.1 y B.2 que se enuncian a continuación puede hallarse en Embrechts et al. (1991).

**Proposición B.1** *Caracterización del MDA(H)*

La función de distribución  $F$  pertenece al máximo dominio de atracción de la distribución de valor extremo  $H$  con constantes de normalización  $\{c_n\}$  y  $\{d_n\}$  ( $C_n > 0$ ;  $d_n \in \mathcal{R}$ ) si y sólo si

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln H(x) \quad \text{para toda } x \in \mathcal{R}$$

donde el caso  $H(x) = 0$  se interpreta como

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n X + d_n) = \infty.$$

**Teorema B.1** *Máximo dominio de atracción de  $\Phi_\alpha$* 

La función de distribución  $F$  pertenece al máximo dominio de atracción (MDA) de la distribución Fréchet  $\{\Phi_\alpha : \alpha > 0\}$  si y sólo si  $\bar{F}(x) = x^{-\alpha} \cdot L(x)$  donde  $L$  es una función de variación lenta.

**Nota I.** Si la función de distribución  $F$  es tal que  $\bar{F}(x) = x^{-\alpha} \cdot L(x)$ , con  $L(x)$  de variación lenta, entonces

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(x)} &= \lim_{x \rightarrow \infty} \frac{(tx)^{-\alpha} \cdot L(tx)}{x^{-\alpha} \cdot L(x)} \\ &= t^{-\alpha} \lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} \\ &= t^{-\alpha} \end{aligned}$$

Entonces si  $F$  pertenece al MDA de  $\Phi_\alpha$ , tenemos que  $\bar{F} \in \mathcal{R}_{-\alpha}$ .

Por otra parte, se puede probar (Feller, 1968), que si  $\bar{F} \in \mathcal{R}_{-\alpha}$  entonces  $\bar{F}(x) = x^{-\alpha} L(x)$  con  $L(x)$  una función de variación lenta.

De lo anterior, el teorema B.1 se puede reescribir como

$$F \in MDA(\Phi_\alpha) \text{ si y sólo si } \bar{F} \in \mathcal{R}_{-\alpha}.$$

**Nota II.** Si  $F \in MDA(\Phi_\alpha)$  entonces el punto final a la derecha de  $F$ ,  $\omega_F$ , es tal que  $\omega_F = \infty$ . Para verificar lo anterior, sea  $x_1 \in \mathcal{R}$  tal que  $\omega_F < x_1 < +\infty$ , entonces tenemos  $\bar{F}(x_1) = 0$  pero como  $F \in MDA(\Phi_\alpha)$ ,  $\bar{F}(x_1) = x_1^{-\alpha} \cdot L(x_1)$ . Concluimos que si  $\omega_F < \infty$ , entonces para toda  $x > \omega_F$  se tiene  $L(x) = 0$ , lo cual es contradictorio ya que  $L \in \mathcal{R}_0$ .

que es contradictorio.

**Teorema B.2** *Representación para funciones de variación regular*

Si  $h \in \mathcal{R}_\alpha$ , para alguna  $\alpha \in \mathbb{R}$ , entonces

$$h(x) = c(x) \exp \left\{ \int_z^x \frac{\delta(t)}{t} dt \right\}$$

para toda  $x \geq z$  y alguna  $z > 0$ , donde  $c$  y  $\delta$  son funciones medibles tales que

$$c(x) \rightarrow c_0 > 0, \quad \text{cuando } x \rightarrow \infty \quad \text{y}$$

$$\delta(x) \rightarrow \alpha \quad \text{cuando } x \rightarrow \infty.$$

La demostración a este teorema puede encontrarse en Embrechts et.al. (1991).

**Teorema B.3** *Caracterización del Máximo Dominio de Atracción (MDA) para  $H_\xi$ .*

Para  $\xi \in \mathbb{R}$ , las siguientes condiciones son equivalentes

(i)  $F \in MDA(H_\xi)$

(ii) Existe una función positiva y medible  $a(\cdot)$ , tal que si  $1 + \xi x > 0$ ,

$$\lim_{u \uparrow \infty} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0 \end{cases}$$

A continuación se presenta sólo un bosquejo de la demostración para el caso  $\xi > 0$  y asumiendo que  $F$  es continua. En de Haan (1970) se puede encontrar una prueba rigurosa.

*Demostración.* (Bosquejo)

Sean  $\xi > 0$  y  $\alpha \equiv \frac{1}{\xi} > 0$ .

Tenemos que

$$\begin{aligned} H_{\xi}(x) &= \Phi_{\frac{1}{\xi}}(1 + \xi x) \\ &= \exp\left\{-(1 + \xi x)^{-\frac{1}{\xi}}\right\} \\ &= \exp\left\{-\left[\frac{1}{\alpha}(\alpha + x)\right]^{-\alpha}\right\} \\ &= \Phi_{\alpha}\left(\frac{1}{\alpha}(\alpha + x)\right) \end{aligned}$$

donde  $\Phi_{\alpha}(x) = \exp(-x^{-\alpha})$  es la distribución Fréchet. Entonces si  $F$  pertenece al máximo dominio de atracción de  $H_{\xi}$  se tiene que  $F$  pertenece al máximo dominio de atracción de  $\Phi_{\alpha}$ , para  $\alpha = \frac{1}{\xi}$ .

Por el teorema B.1 podemos concluir que  $\bar{F} \in \mathcal{R}_{-\alpha}$

Usando el teorema de representación para funciones de variación regular B.2, existe  $z > 0$  tal que para toda  $x \geq z$

$$\bar{F}(x) = c(x) \exp\left\{\int_z^x \frac{\delta(t)}{t} dt\right\}$$

donde  $c(\cdot)$  y  $\delta(\cdot)$  son funciones medibles tales que

$$c(x) \rightarrow c_0 \in (0, \infty) \text{ cuando } x \rightarrow \infty$$

$$\delta(x) \rightarrow -\alpha \text{ cuando } x \rightarrow \infty.$$

Como  $-\delta(x) \rightarrow \alpha$  cuando  $x \rightarrow \infty$  y  $\alpha > 0$ , entonces  $a(x) \equiv \frac{x}{-\delta(x)}$  es una función medible y positiva para  $x$  suficientemente grande. Además,  $\frac{a(x)}{x} \rightarrow \frac{1}{\alpha}$  cuando  $x \rightarrow \infty$ .



Sea  $u_1 \in \mathbf{R}$  tal que si  $u > u_1$ ,

$$u + xa(u) = \left(1 + x \frac{a(u)}{u}\right) u \approx \left(1 + \frac{x}{\alpha}\right) u$$

Sea  $u_2 \in \mathbf{R}$  tal que si  $u > u_2$ ,  $-\delta(u) \approx \alpha$  y  $u_3 \in \mathbf{R}$  tal que si  $u > u_3$ ,  $c(u) \approx c_0$ .  
Entonces, para  $x > 0$  y  $u > \max(z, u_1, u_2, u_3)$  tenemos que

$$u + xa(u) \approx \left(1 + \frac{x}{\alpha}\right) u > u > z$$

y

$$\begin{aligned} \frac{\bar{F}(u + xa(u))}{F(u)} &= \frac{c(u + xa(u))}{c(u)} \cdot \frac{\exp\left\{-\int_z^{u+xa(u)} \frac{\delta(t)}{t} dt\right\}}{\exp\left\{-\int_z^u \frac{\delta(t)}{t} dt\right\}} \\ &= \frac{c(u + xa(u))}{c(u)} \cdot \exp\left\{-\int_u^{u+xa(u)} \frac{\delta(t)}{t} dt\right\} \\ &= \frac{c(u + xa(u))}{c(u)} \cdot \exp\left\{-\int_1^{1+\frac{x}{\alpha}} \frac{\delta(u\tau)}{\delta(u)} \left(-\frac{\delta(u)}{\tau}\right) d\tau\right\} \\ &\approx \frac{c(u + xa(u))}{c(u)} \cdot \exp\left\{\delta(u) \int_1^{1+\frac{x}{\alpha}} \frac{\delta(u\tau)}{\delta(u)} \frac{1}{\tau} d\tau\right\} \\ &\approx \frac{c_0}{c_0} \cdot \exp\left\{-\alpha \int_1^{1+\frac{x}{\alpha}} \frac{1}{\tau} d\tau\right\}. \end{aligned}$$

En la aproximación anterior usamos que  $\frac{\delta(u\tau)}{\delta(u)} \rightarrow 1$ , cuando  $u \rightarrow \infty$ , uniformemente en  $\tau$  sobre conjuntos compactos de  $(0, \infty)$ . Ver Embrechts et al. (1991), teorema A3.2.

Así entonces, si  $x > 0$  y  $u > \max(z, u_1, u_2, u_3)$ ,

$$\begin{aligned} \frac{\tilde{F}(u + xa(u))}{\tilde{F}(u)} &= \exp \left\{ -\alpha \int_1^{1+\frac{x}{\alpha}} \frac{1}{\tau} d\tau \right\} \\ &= \exp \left\{ -\alpha \left[ \log \left( 1 + \frac{x}{\alpha} \right) - \log(1) \right] \right\} \\ &= \left( 1 + \frac{x}{\alpha} \right)^{-\alpha}. \end{aligned}$$

\* Nota: en Embrechts et.al.(1991) se obtiene

$$\frac{\delta(u\tau)}{\delta(u)} \rightarrow 1 \quad \text{cuando } u \rightarrow \infty$$

uniformemente en  $\tau$  sobre conjuntos compactos de  $(0, \infty)$ .

Es decir

$$\lim_{u \rightarrow \infty} \frac{\tilde{F}(u + xa(u))}{\tilde{F}(u)} = (1 + \xi x)^{-\xi}.$$

Ahora asumimos (ii). Para el conjunto

$$\left\{ x \in \mathbb{R} : F(x) \geq 1 - \frac{1}{n} \right\} = \left\{ x \in \mathbb{R} : \frac{1}{n} \geq \bar{F}(x) \right\}$$

Sea

$$\begin{aligned} d_n &\equiv \inf \left\{ x \in \mathbb{R} : F(x) \geq 1 - \frac{1}{n} \right\} \\ &= F^{\leftarrow} \left( 1 - \frac{1}{n} \right) \\ &= \left( \frac{1}{\bar{F}} \right)^{\leftarrow} (n). \end{aligned}$$

La sucesión  $\{d_n\}_{n \geq 1}$  es tal que

- (i)  $d_{n_1} \leq d_{n_2}$  si  $n_1 \leq n_2$ .
- (ii)  $\lim_{n \rightarrow \infty} d_n = \omega_F$ .

Por continuidad por la derecha de  $F$  tenemos que

$$\frac{1}{n} \geq \bar{F}(d_n)$$

Supongamos ahora que

$$\frac{1}{n} > \bar{F}(d_n)$$

es decir

$$F(d_n) > 1 - \frac{1}{n}$$

Pero por continuidad por la izquierda de  $F$  en  $d_n$ , dado  $\varepsilon = \varepsilon_n = F(d_n) - \left(1 - \frac{1}{n}\right)$  existe  $x_\varepsilon < d_n$  tal que

$$F(d_n) - F(x_\varepsilon) < \varepsilon = F(d_n) - \left(1 - \frac{1}{n}\right),$$

es decir

$$F(x_\varepsilon) > 1 - \frac{1}{n},$$

lo cual es contradictorio por definición de  $d_n$ .

Entonces concluimos que por continuidad de  $F$

$$\frac{1}{n} = \bar{F}(d_n)(1) \quad (\text{B.1})$$

Sea  $\alpha = \frac{1}{\xi} > 0$ , usando la hipótesis (ii) con  $u = d_n$  obtenemos

$$\begin{aligned} \left(1 + \frac{x}{\alpha}\right)^{-n} &= \lim_{n \rightarrow \infty} \frac{\bar{F}(d_n + \alpha a(d_n))}{\bar{F}(d_n)} \\ &= \lim_{n \rightarrow \infty} \frac{n \bar{F}(d_n + \alpha a(d_n))}{n \bar{F}(d_n)} \\ &= \lim_{n \rightarrow \infty} \bar{F}(d_n + \alpha a(d_n)), \end{aligned}$$

donde usamos la ecuación (1) en la última igualdad.

Por último, aplicamos la proposición B.1 para obtener que  $F \in MDA(H_\xi)$  para  $\xi = \frac{1}{\alpha}$ , donde  $H_\xi$  es la DVEG en la ecuación 2.3.



## Apéndice C

### Procesos Puntuales

En esta sección se describen de manera no rigurosa algunos conceptos respecto a procesos estocásticos puntuales.

Sean  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad,  $E$  un subconjunto de  $\mathbb{R}^d$ , para  $d$  un entero mayor o igual a 1 y  $\mathcal{B}(E)$  la  $\sigma$ -álgebra de Borel en  $E$ .

Supóngase que  $\{X_n : n \geq 0\}$  es una colección de variables aleatorias con valores en  $E$ . Para cada  $n \geq 0$  sea  $\delta_{X_n}$  la función

$$\delta_{X_n} = \begin{cases} 1 & \text{si } X_n \in A \\ 0 & \text{si } X_n \notin A \end{cases},$$

donde  $A \in \mathcal{B}(E)$ .

Entonces, el número total de "puntos aleatorios" que caen en  $A$  es

$$\mathcal{N}(A) = \sum_{n \geq 0} \delta_{X_n}(A).$$

Notemos que  $\mathcal{N}$  es una medida de conteo que es a su vez variable aleatoria ya que para cada  $\omega \in \Omega$ , los puntos  $X_n(\omega)$  pueden cambiar de posición y por tanto  $\mathcal{N}$  depende

de  $\omega$ .

$$\mathcal{N}(\omega, A) = \sum_{n \geq 0} \delta_{X_n(\omega)}(A).$$

Para cada  $\omega \in \Omega$  fijo  $\mathcal{N}(\omega, \cdot)$  es una medida y para cada  $A \in \mathcal{B}(E)$  fijo  $\mathcal{N}(\cdot, A)$  es una variable aleatoria

En adelante, la dependencia de  $\mathcal{N}$  en  $\omega$  se hará notar sólo cuando se considere necesario.

i)  $\mathcal{N} : \mathcal{B}(E) \rightarrow [0, \infty)$ .

ii) Si  $A_1, A_2, \dots \in \mathcal{B}(E)$  tales que  $A_i \cap A_j = \emptyset$ , para toda  $i \neq j$  entonces

$$\begin{aligned} \mathcal{N}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{n \geq 0} \delta_{X_n}\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{n \geq 0} \sum_{i=1}^{\infty} \delta_{X_n}(A_i) \\ &= \sum_{i=1}^{\infty} \sum_{n \geq 0} \delta_{X_n}(A_i) \\ &= \sum_{i=1}^{\infty} \mathcal{N}(A_i). \end{aligned}$$

Diremos que  $\{\mathcal{N}(A) : A \in \mathcal{B}(E)\}$  es un proceso puntual si para cada  $A \in \mathcal{B}(E)$  compacto, tenemos que

$$IP(\mathcal{N}(A) < \infty) = 1.$$

Como para cada  $A \in \mathcal{B}(E)$   $\mathcal{N}(A)$  una variable aleatoria, podemos hablar de su valor esperado, ya sea que éste sea finito o infinito.

$$\mu(A) = E(\mathcal{N}(A)).$$

Por las propiedades de valor esperado  $\mu$  es una medida sobre  $\mathcal{B}(E)$  como se demuestra enseguida.

*Demostración.*  $\mu(A) \equiv E(\sum_{i=0}^{\infty} \delta_{X_i}(A))$  es medida si

- (i)  $\mu(A) \geq 0$  y  
 (ii) Para todo  $A \in \mathcal{B}(R^d)$ ,  $A_1, A_2, \dots \in \mathcal{B}(R^d)$  tales que  $A_i \cap A_j = \emptyset$ ,  $\forall i \neq j$ ,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

La demostración de (i) es directa pues  $\sum_{i=0}^{\infty} \delta_{X_i}(A)$  será cero en el caso en que ningún punto  $x_i \in A$  y entonces su valor esperado también será cero; si uno ó más puntos  $x_i \in A$ ,  $\sum_{i=0}^{\infty} \delta_{X_i}(A) > 0$  y por lo tanto  $E(\sum_{i=0}^{\infty} \delta_{X_i}(A)) > 0$  también.

La parte (ii) se demuestra

$$\begin{aligned} \mu\left(\bigcup_{k=1}^{\infty} A_k\right) &= E\left(\sum_{n \geq 0} \delta_{X_n}\left(\bigcup_{k=1}^{\infty} A_k\right)\right) \\ &\stackrel{(1)}{=} E\left(\sum_{n \geq 0} \sum_{k=1}^{\infty} \delta_{X_n}(A_k)\right) \\ &\stackrel{(2)}{=} E\left(\sum_{k=1}^{\infty} \sum_{i \geq 0} \delta_{X_i}(A_k)\right) \\ &= \lim_{n \rightarrow \infty} E\left(\sum_{k=1}^n \sum_{i \geq 0} \delta_{X_i}(A_k)\right) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n E\left(\sum_{i \geq 0} \delta_{X_i}(A_k)\right) \\ &= \sum_{k=1}^{\infty} \mu(A_k) \end{aligned}$$



la igualdad (1) es cierta porque para  $A_1, A_2$  mutuamente excluyentes tenemos

$$\delta_A = \begin{cases} 1 & \text{si } x_i \in A \\ 0 & \text{si } x_i \notin A \end{cases}$$

y entonces para cualesquiera  $j \neq k$

$$\delta_{A_j \cup A_k} = \begin{cases} 1 & \text{si } x_i \in A_j \text{ ó } x_i \in A_k \\ 0 & \text{si } x_i \notin A_j \text{ y } x_i \notin A_k \end{cases}$$

por lo que

$$\delta_{A_j \cup A_k} = \delta_{X_i}(A_j) + \delta_{X_i}(A_k).$$

La igualdad (2) es cierta por el Teorema de Fubini.

*q.e.d.*

Para cada  $A \in \mathcal{B}(E)$ ;  $\mu(A)$  es el "número esperado de puntos en la región  $A$ ".

**Ejemplo 1.** Sean  $d = 1$ ,  $E = [0, \infty)$ ,  $\mathcal{B}(E) = \mathcal{B}[0, \infty)$  y  $T_0, T_1, \dots$  variables aleatorias independientes con distribución exponencial de parámetro  $\alpha > 0$ . Sea  $X_n = \sum_{j=0}^n T_j$  entonces

$$\begin{aligned} N_t &\equiv \mathcal{N}((0, t]) \\ &= \sum_{n \geq 0} \delta_{X_n}((0, t]) \end{aligned}$$

es decir,  $N_t$  es el número de  $X_i$ 's en el intervalo  $(0, t]$ .

$$IP(\mathcal{N}((0, t]) = k) = IP(X_{k-1}, X_k > t).$$

Para hallar tal probabilidad, sea  $G: \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{k+1}$  la transformación

$$\begin{aligned} X_0 &= T_0; \\ X_1 &= T_0 + T_1; \\ &\vdots \\ X_k &= T_0 + T_1 + \cdots + T_k. \end{aligned}$$

Entonces  $G$  es una función uno a uno y su inversa

$$\begin{aligned} T_0 &= X_0; \\ T_1 &= X_1 - X_0; \\ &\vdots \\ T_k &= X_k - X_{k-1}, \end{aligned}$$

tiene Jacobiano con valor absoluto 1.

Así entonces

$$\begin{aligned} f_{X_0, X_1, \dots, X_k}(x_0, x_1, \dots, x_k) &= 1 \times f_{T_0, T_1, \dots, T_k}(x_0, x_1 - x_0, \dots, x_k - x_{k-1}) \\ &= f_{T_0}(x_0) \times \prod_{j=1}^k f_{T_j}(x_j - x_{j-1}) \\ &= \alpha^{k+1} \exp\{-\alpha x_k\}. \end{aligned}$$

Por lo tanto

$$\begin{aligned} IP(\mathcal{N}((0, t])) &= IP(X_{k-1} \leq t; X_k > t) \\ &= \int_0^\infty \int_0^t \int_0^{x_{k-1}} \cdots \int_0^{x_1} \alpha^{k+1} \exp\{-\alpha x_k\} dx_0 dx_1 \cdots dx_{k-2} dx_{k-1} dx_k \\ &= \frac{(\alpha t)^k}{k!} \exp\{-\alpha t\}, \end{aligned}$$

es decir  $\mathcal{N}_t \sim \text{Poisson}(\alpha t)$  de donde

$$\mu((0, t]) = \alpha t = \alpha \cdot \lambda((0, t]),$$

donde  $\lambda$  es la medida de Lebesgue.

Se puede probar que si  $0 < s < t$  entonces  $\mathcal{N}((0, t]) - \mathcal{N}((0, s])$  y  $\mathcal{N}((0, s])$  son variables aleatorias independientes.

**Ejemplo 2.** Sean  $d = 2; u > 0; E = [0, \infty) \times (u, \infty), \mathcal{B}(E) = \mathcal{B}([0, \infty) \times (u, \infty))$ . Consideremos el proceso de reclamaciones de cierto tipo de seguro  $Y_1, Y_2, \dots, Y_n$  en una compañía reaseguradora. Sean  $Y_{T_1}, Y_{T_2}, \dots, Y_{T_{k_n}}$  aquéllas reclamaciones que exceden el valor del umbral  $u$ , donde  $Y_{T_i}$  es la magnitud de la reclamación al tiempo  $T_i$ , es decir las variables aleatorias  $\{T_i : i = 1, 2, \dots, k_n\}$  son los tiempos donde el proceso  $\{Y_i\}$  excede el valor de  $u$ .

Las variables aleatorias  $X_i = (T_i, Y_{T_i}), i = 1, 2, \dots, k_n$  toman valores en  $E$  y para cada  $A \in \mathcal{B}(E)$  podemos definir  $\mathcal{N}_n(A) = \sum_{i=1}^{k_n} \delta_{X_i}(A)$ .

En particular para  $y > u$  y  $A = (t_1, t_2] \times (y, \infty)$ ,  $\mathcal{N}_n(A)$  es el número de reclamaciones mayores que  $y$  entre los tiempos  $t_1$  y  $t_2$ .

### Definición C.1 (Proceso de Poisson Puntual)

Sea  $\mathcal{N}$  un proceso puntual con valores en  $E \subset \mathbb{R}^d$  y  $\mathcal{B}(E)$  la  $\sigma$ -álgebra de Borel en  $E$ .

Decimos que  $\mathcal{N}$  es un proceso de Poisson puntual con medida de intensidad (valor esperado)  $\mu$  si

i) Para cada  $A \in \mathcal{B}(E)$

$$P(\mathcal{N}(A) = k) = \begin{cases} \frac{\exp\{-\mu(A)\}(\mu(A))^k}{k!} & \text{si } \mu(A) < \infty \\ 0 & \text{si } \mu(A) = \infty \end{cases}$$

- ii) Si  $A_1, A_2, \dots, A_k$  son elementos de  $\mathcal{B}(E)$  con  $A_i \cap A_j = \emptyset$  para toda  $i \neq j$ , entonces  $\mathcal{N}(A_1), \mathcal{N}(A_2), \dots, \mathcal{N}(A_k)$  son variables aleatorias independientes.

Es decir,  $\mathcal{N}$  es un proceso de Poisson puntual si el número (aleatorio) de puntos en un conjunto  $A$  tiene distribución Poisson con parámetro  $\mu(A)$  y los números de puntos en regiones ajenas son variables aleatorias independientes.

La propiedad ii) generaliza para  $d > 1$  la propiedad de incrementos independientes del proceso Poisson en  $\mathcal{R}$ .

Cuando la medida de intensidad  $\mu$  está dada por un múltiplo de la medida de Lebesgue (longitud para  $d = 1$ , área para  $d=2$ , volumen para  $d=3$ , etc.), decimos que  $\mathcal{N}$  es homogéneo.

Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias y  $X$  una variable aleatoria, recordemos que si  $F_{X_n}(x) = \mathbb{P}(X_n \leq x)$  y  $F_X(x) = \mathbb{P}(X \leq x)$ , entonces  $X_n \xrightarrow{d} X$  significa que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

para cada  $x \in C_F = \{x \in \mathcal{R} : F_x \text{ es continua en } x\}$ .

Notemos que definir un concepto análogo para una sucesión de procesos puntuales  $\{\mathcal{N}_n\}_{n \geq 1}$  es mucho más complicado.

Para empezar cada proceso puntual  $\mathcal{N}_n$  puede considerarse como un proceso estocástico cuyo conjunto de índices son los conjuntos  $A \in \mathcal{B}(E)$ , es decir, la colección de variables aleatorias  $\{\mathcal{N}_n(A) : A \in \mathcal{B}(E)\}$ .

Otra forma de conceptualizar  $\mathcal{N}_n$  es como una variable aleatoria con valores en un espacio vectorial cuyos elementos son medidas de conteo. No es nuestro objetivo en este trabajo discutir respecto a la distribución de probabilidades  $\mathbb{P}_{\mathcal{N}_n}$  de tales variables aleatorias, así como de la convergencia de las mismas en algún sentido a una distribución límite  $\mathbb{P}_{\mathcal{N}}$ . Una buena discusión de estos temas con referencias para profundizar estos conceptos puede hallarse en Embrechts et al. (1991), capítulo 5.

Aquí nos limitaremos a decir que la sucesión de procesos puntuales  $\{\mathcal{N}_n\}_n$  converge en distribución al proceso puntual  $\mathcal{N}$  cuando  $n \rightarrow \infty$  (en notación  $\mathcal{N}_n \xrightarrow{d} \mathcal{N}$ ), si para cada entero  $m \geq 1$ ; para enteros no negativos  $i_1, i_2, \dots, i_m$  y conjuntos  $A_1, A_2, \dots, A_m \in \mathcal{G}(E)$

$$\mathbb{P}(\mathcal{N}_n(A_1) = i_1; \dots; \mathcal{N}_n(A_m) = i_m) \longrightarrow \mathbb{P}(\mathcal{N}(A_1) = i_1; \dots; \mathcal{N}(A_m) = i_m),$$

cuando  $n \rightarrow \infty$ . La clase de conjuntos  $\mathcal{G}(E)$  está dada por

$$\mathcal{G}(E) = \{A \in \mathcal{B}(E) : \mathbb{P}[\mathcal{N}(\delta A) = 0] = 1\},$$

donde  $\delta A$  denota la frontera de  $A$ .

A continuación enunciamos en un sentido informal un resultado que es de relevancia en nuestro estudio. Su demostración requiere fundamentos matemáticos que van más allá de los objetivos de esta tesis, pero una versión de la prueba puede encontrarse por ejemplo en Leadbetter et al. (1983), p.p. 119.

Consideremos el ejemplo 2 de esta sección, supóngase que las observaciones  $Y_1, Y_2, \dots$  son variables aleatorias independientes y con distribución  $F$ .

Sea  $M_n \equiv \max\{Y_1, Y_2, \dots, Y_n\}$ . Suponemos que existen sucesiones de números reales  $\{a_n\}$  y  $\{b_n\}$  ( $a_n > 0; \forall n$ ), tales que

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G,$$

cuando  $n \rightarrow \infty$ , es decir  $F^n(a_n x + b_n) \rightarrow G(x)$  cuando  $n \rightarrow \infty$  para  $x$  punto de continuidad de  $G$ , donde  $G$  es una función de distribución no degenerada. Entonces bajo algunas hipótesis adicionales, el proceso puntual (renormalizado) correspondiente a los valores que exceden el umbral  $u_n = a_n x + b_n$

$$\mathcal{N}'_n(\cdot) = \sum_{i \geq 1} \delta_{x'_{i,n}}(\cdot);$$

$$X'_{i,n} = \left( \frac{T_i}{n}, \frac{Y_{T_i} - b_n}{a_n} \right),$$

converge en distribución cuando  $n \rightarrow \infty$  a un proceso de Poisson puntual  $\mathcal{N}'$  con medida de intensidad

$$\Lambda(\cdot) = (\lambda \times \mu_\sigma)(\cdot),$$

donde  $\lambda$  es la medida de Lebesgue en  $(0, \infty)$  y  $\mu_\sigma$  es la medida determinada por la función no decreciente  $\log\{G(y)\}$  sobre  $(x_0, \infty)$  con  $x_0 = \inf\{x : G(x) > 0\}$ .

En particular  $\mu_\sigma$  es tal que

$$\mu_\sigma((a, b]) = \log\{G(b)\} - \log\{G(a)\}.$$

En Smith (1989) y Smith y Goodman (2000) se ha hecho uso de resultados de este tipo. En específico, si se supone que para el proceso de reclamaciones  $Y_1, Y_2, \dots$  existen sucesiones de constantes  $\{a_n\}$  y  $\{b_n\}$  tales que

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n}\right) \rightarrow H(y; \xi, \mu, \sigma) = \exp\left\{-\left[1 + \xi\left(\frac{y - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\},$$

cuando  $n \rightarrow \infty$ , entonces podremos considerar que a nivel de aproximación el proceso puntual descrito en el ejemplo 2 se puede modelar como un proceso de Poisson puntual cuya medida de intensidad sobre conjuntos de la forma  $(t_1, t_2] \times (y, +\infty)$ , con  $y > u$  es

$$\begin{aligned} \Lambda((t_1, t_2] \times (y, +\infty)) &= \lambda((t_1, t_2]) \times \mu_\sigma((y, \infty)) \\ &= (t_2 - t_1) \times \log\{H(+\infty; \xi, \mu, \sigma)\} - \log\{H(y; \xi, \mu, \sigma)\} \\ &= (t_2 - t_1) \times \left(1 + \xi\left(\frac{y - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}. \end{aligned}$$

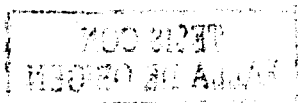


## Apéndice D

### Gráficos del modelo con efectos aleatorios en localización y escala

Figuras correspondientes al modelo con efectos aleatorios en localización y escala, para el Caso I.

Figuras correspondientes al modelo con efectos aleatorios en localización y escala, para el Caso II.





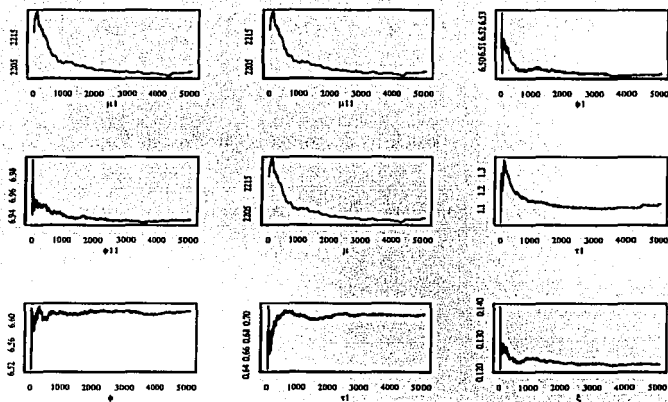


Figura D.1: Gráfico de promedios ergódicos,  $u = 45$

**TESIS CON  
FALLA DE ORIGEN**

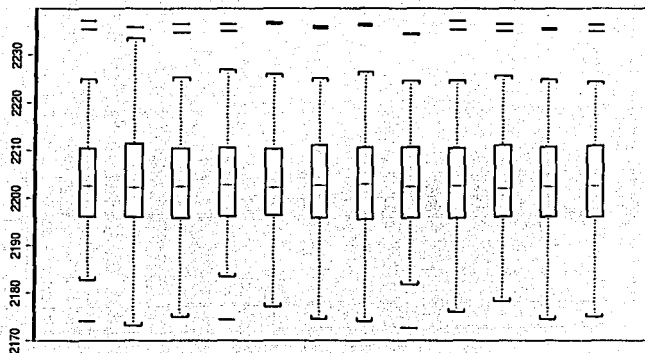


Figura D.2: Gráfico de efectos aleatorios en localizaciones,  $u = 45$

TESIS CON  
FALLA DE ORIGEN

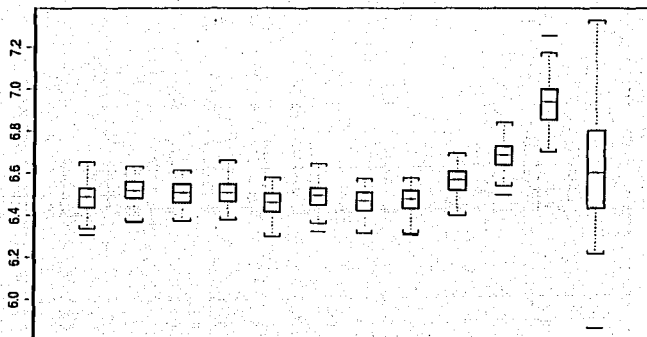


Figura D.3: Gráfico de efectos aleatorios en escalas,  $u = 45$

TESIS CON  
FALLA DE ORIGEN

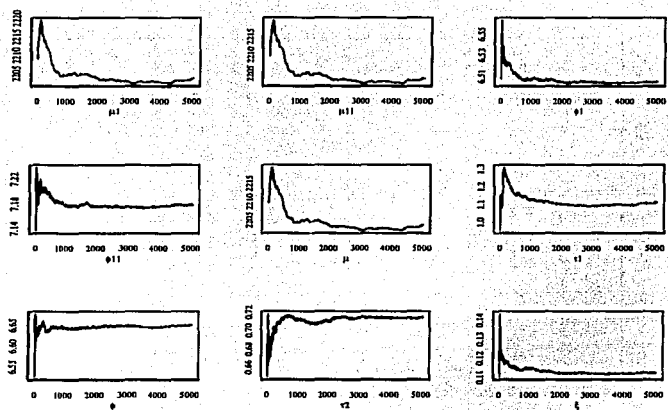


Figura D.4: Gráfico de promedios ergódicos,  $u = 300$

TESIS CON  
FALLA DE ORIGEN

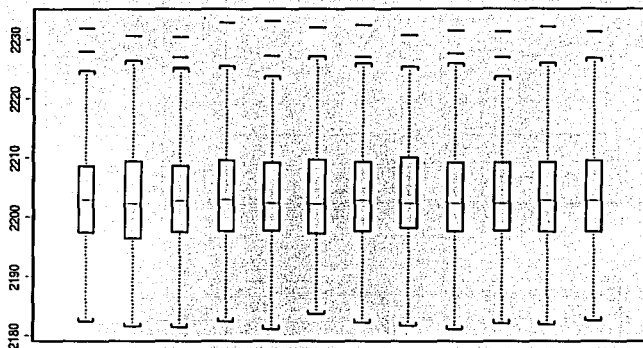


Figura D.5: Gráfico de efectos aleatorios en localizaciones,  $u = 300$

TESIS CON  
FALLA DE ORIGEN

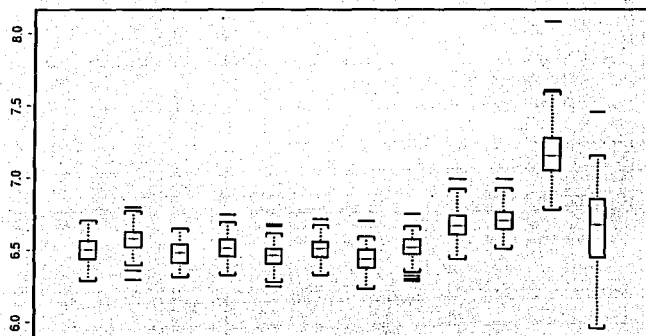


Figura D.6: Gráfico de efectos aleatorios en escalas,  $u = 300$

INSTITUTO VENEZOLANO DE INVESTIGACIONES CIENTÍFICAS  
TESIS CON FALLA DE ORIGEN

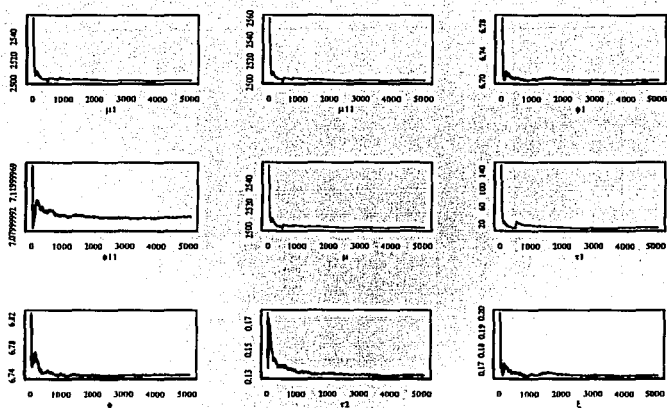


Figura D.7: Gráfico de promedios ergódicos,  $u = 45$

TESIS CON  
FALLA DE ORIGEN

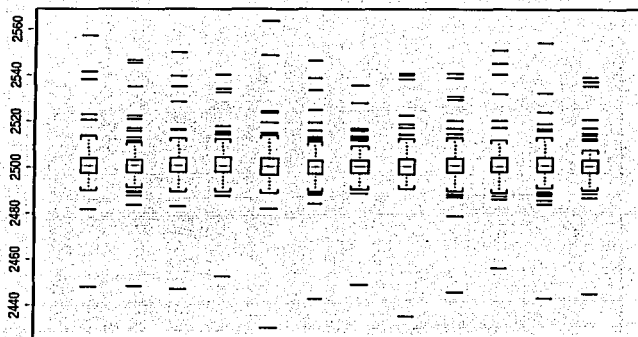


Figura D.8: Gráfico de efectos aleatorios en localizaciones,  $u = 45$

TESIS CON  
FALLA DE ORIGEN



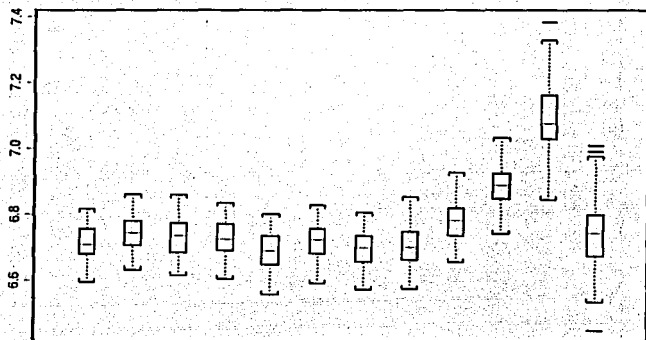


Figura D.9: Gráfico de efectos aleatorios en escalas,  $u = 45$

TESIS CON  
FALLA DE ORIGEN

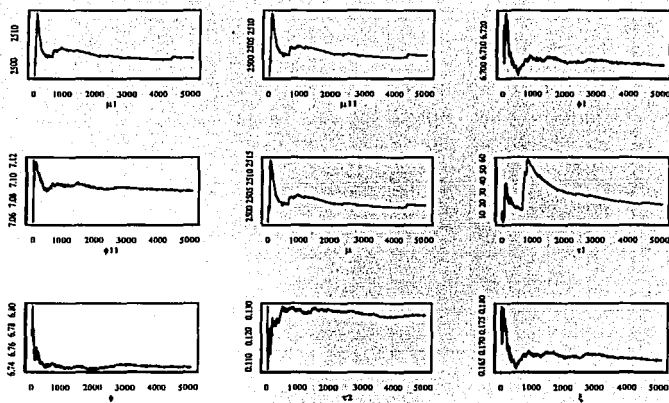


Figura D.10: Gráfico de promedios ergódicos,  $u = 300$

TRASCUN  
FALLA DE ORIGEN

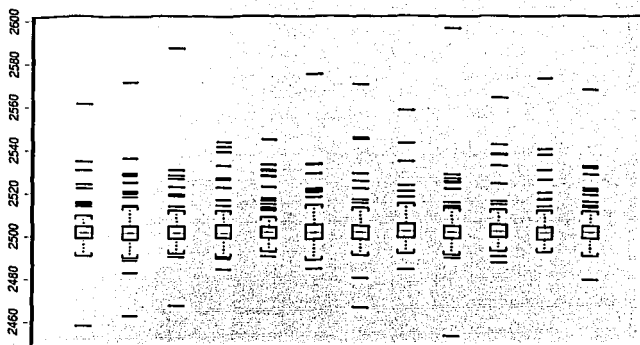


Figura D.11: Gráfico de efectos aleatorios en localizaciones,  $u = 300$

TESIS CON  
FALLA DE ORIGEN

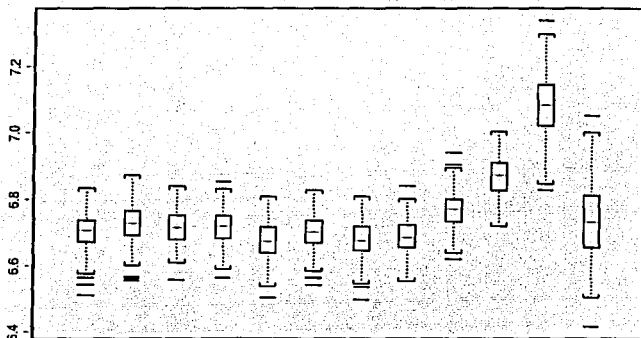


Figura D.12: Gráfico de efectos aleatorios en escalas,  $u = 300$

TESIS CON  
FALLA DE ORIGEN

UNIVERSITY OF TORONTO  
LIBRARY  
130 St. George Street  
Toronto, Ontario M5S 1A5  
416-978-2811

## Apéndice E

### Gráficos del modelo con efectos aleatorios sólo en escala

En este apéndice se incluyen las figuras correspondientes al modelo para efectos aleatorios anuales únicamente en el parámetro de escala primero con  $u = 45$  y después con  $u = 300$ .



# Gráficos de promedios ergódicos con $u=45$

Gráficos del modelo con efectos aleatorios únicamente en la escala, con  $u = 45$ .

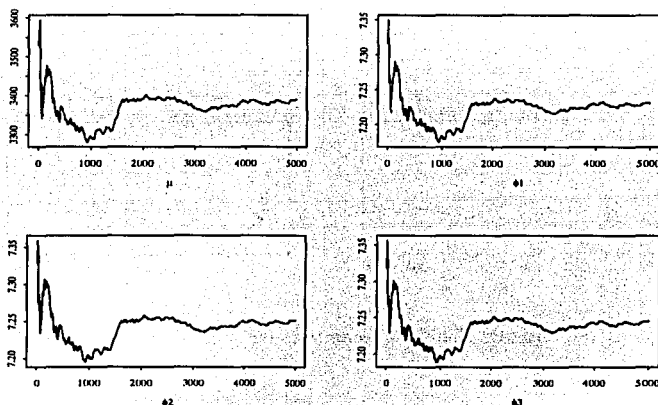


Figura E.1: Gráfico de promedios ergódicos





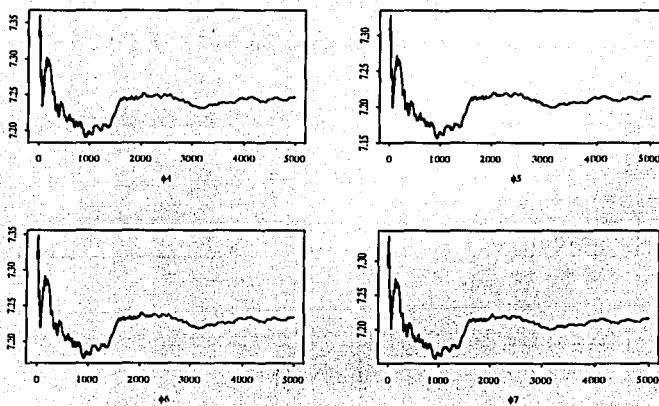


Figura E.2: Gráfico de promedios ergódicos

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
INSTITUTO TECNOLÓGICO Y DE INVESTIGACIONES EN CIENCIAS  
Y TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN  
**TESIS CON FALLA DE ORIGEN**

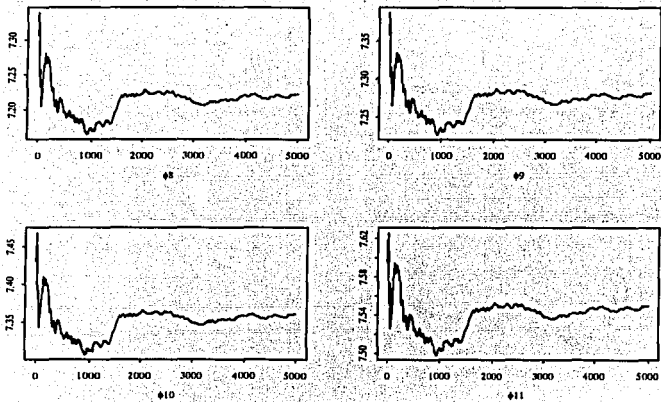


Figura E.3: Gráfico de promedios ergódicos

TESIS CON  
FALLA DE ORIGEN

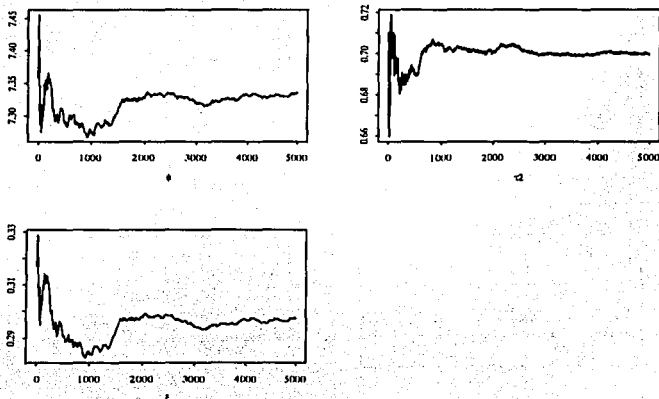


Figura E.4: Gráfico de promedios ergódicos.

TESIS CON  
FALLA DE ORIGEN

# Autocorrelogramas y diagramas de caja con $u=45$

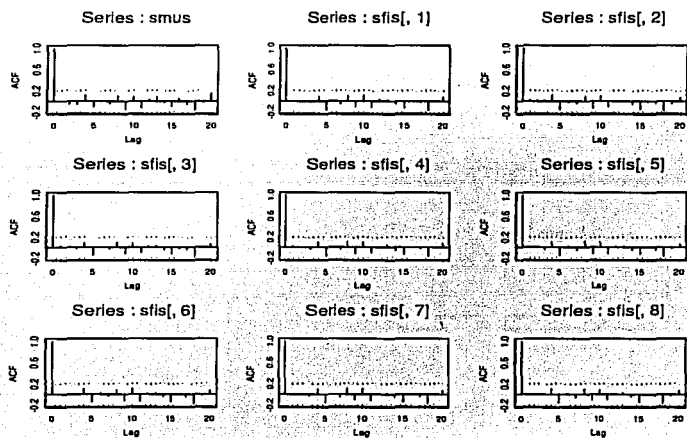


Figura E.5: Autocorrelogramas

TESIS CON  
FALLA DE ORIGEN

FALLA DE ORIGEN

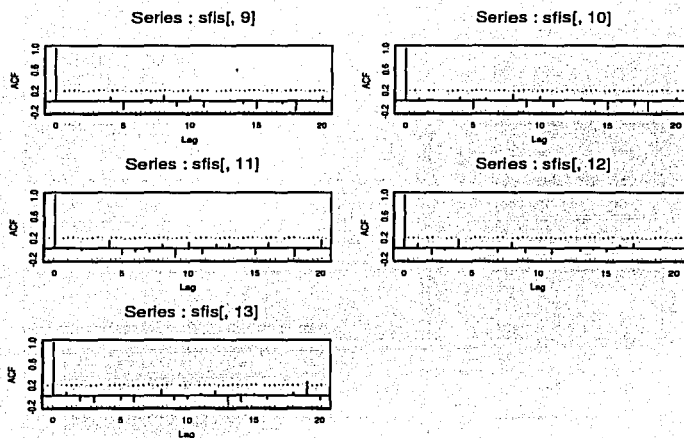


Figura E.6: Autocorrelogramas

CON  
DE ORIGEN

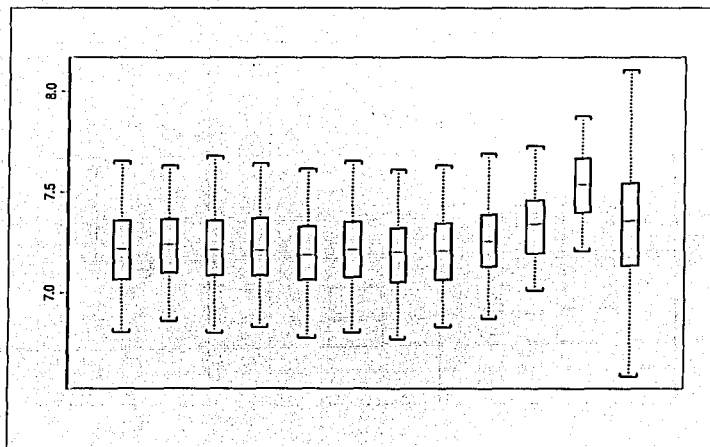


Figura E.7: Gráfico de efectos aleatorios en escala

ES CON  
FALLA DE ORIGEN

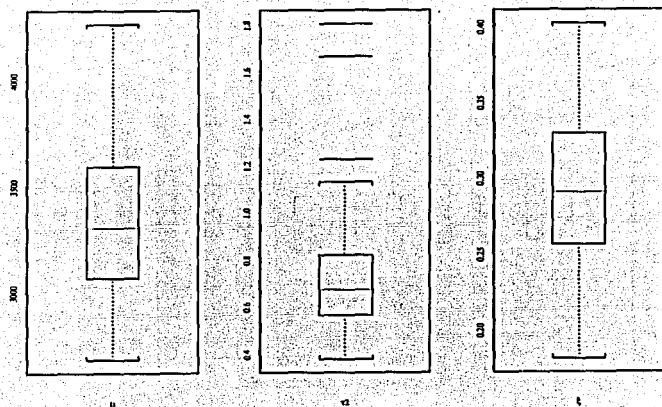


Figura E.8: Gráfico de efectos aleatorios en escala

TESIS CON  
ALA DE ORIGEN

## Histograma de un proceso de reclamaciones con $u=45$

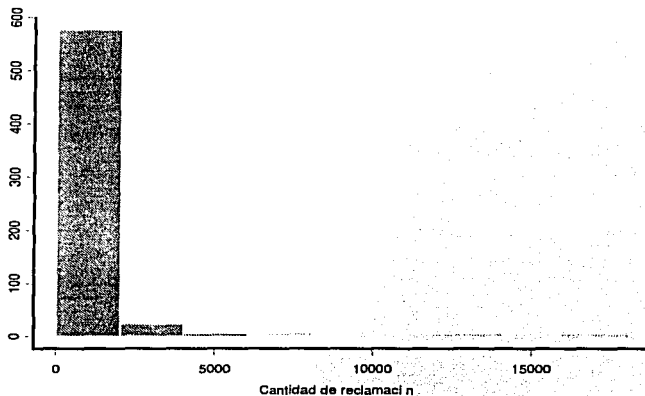


Figura E.9: Histograma del valor de reclamaciones

TESIS CON  
FALLA DE ORIGEN



1970-1971  
1972-1973  
1974-1975  
1976-1977  
1978-1979  
1980-1981  
1982-1983  
1984-1985  
1986-1987  
1988-1989  
1990-1991  
1992-1993  
1994-1995  
1996-1997  
1998-1999  
2000-2001  
2002-2003  
2004-2005  
2006-2007  
2008-2009  
2010-2011  
2012-2013  
2014-2015  
2016-2017  
2018-2019  
2020-2021  
2022-2023  
2024-2025

# Gráficos de promedios ergódicos con $u=300$

Figuras correspondientes al modelo con efectos aleatorios únicamente en la escala con  $u = 300$ .

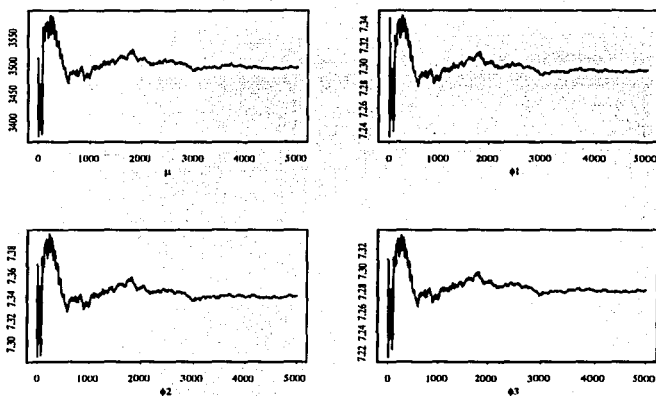


Figura E.10: Gráfico de promedios ergódicos

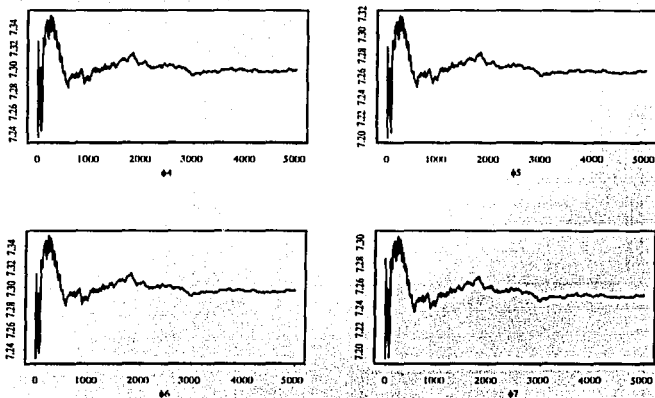


Figura E.11: Gráfico de promedios ergódicos

TESIS CON  
FALLA DE ORIGEN

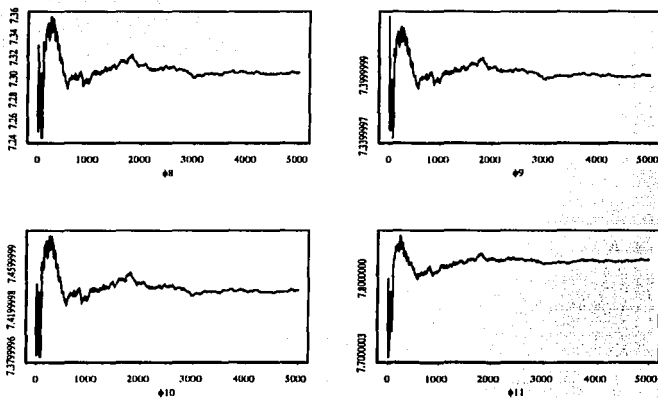


Figura E.12: Gráfico de promedios ergódicos

TESIS CON  
FALLA DE ORIGEN

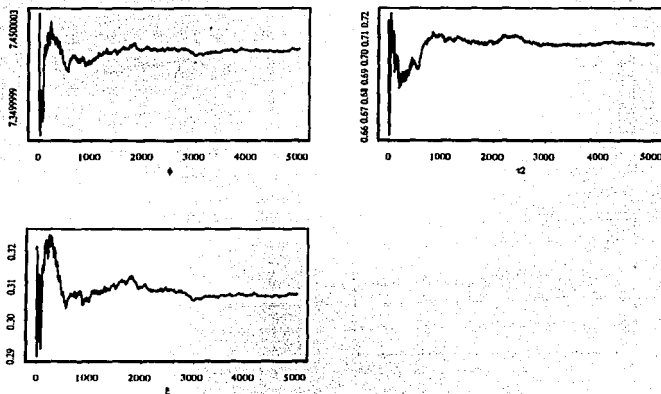


Figura E.13: Gráfico de promedios ergógicos

TESIS CON  
FALLA DE ORIGEN

# Autocorrelogramas y diagramas de caja con $u=300$

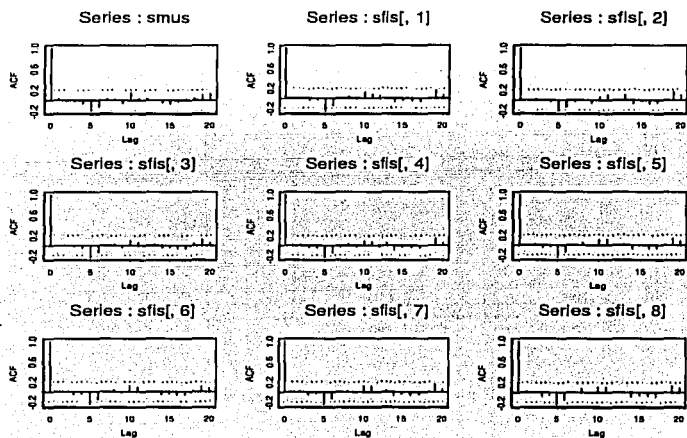


Figura E.14: Autocorrelogramas

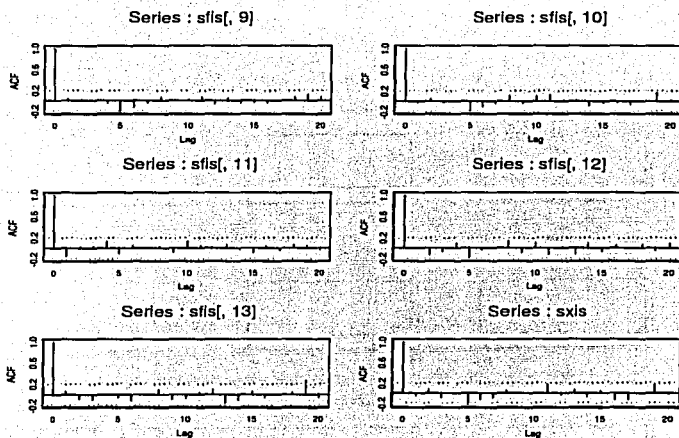


Figura E.15: Autocorrelogramas

TESIS CON  
FALLA DE ORIGEN

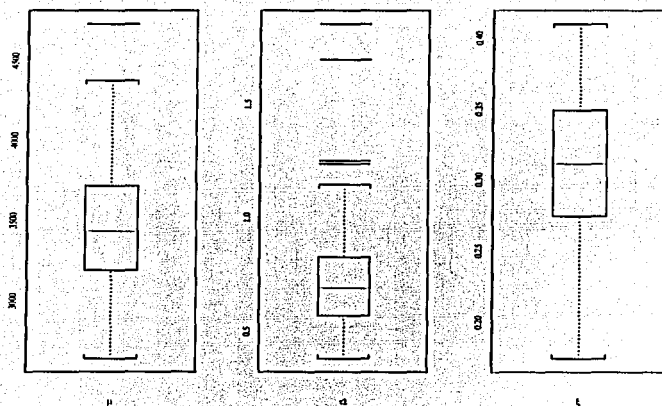


Figura E.16: Gráfico de efectos aleatorios en escala

TESIS CON  
FALLA DE ORIGEN



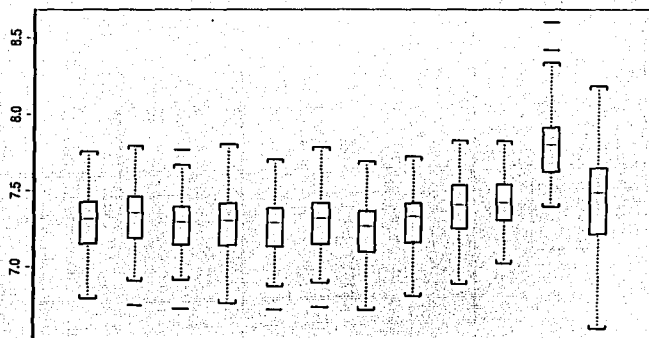


Figura E.17: Gráfico de efectos aleatorios en escala

TESIS CON  
FALLA DE ORIGEN

## Histograma de un proceso de reclamaciones con $u=300$

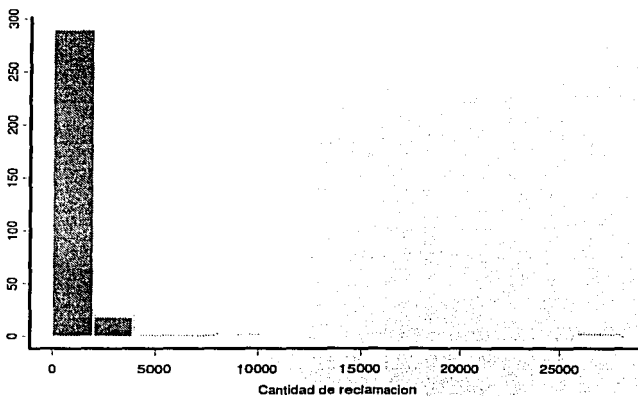


Figura E.18: Histograma del valor de reclamaciones

TESIS CON  
FALLA DE ORIGEN

RECEIVED  
MAY 17 1941

## Intervalos de confianza estimados para la probabilidad de pérdida

A continuación se muestran los intervalos al 95 para la probabilidad de pérdida dado un valor  $z^*$  y los umbrales  $u = 45$  y  $u = 300$ , respectivamente. Los valores que se obtuvieron para  $z^*$  (etiquetada como "zstar") y los cuantiles 5 y 95 (etiquetados como "q5" y "q95", respectivamente) fueron

(i) Para el umbral  $u = 45$ ,

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
zstar	323000.00	326000.00	329000.00	332000.00	335000.00	338000.00
q5	0.99	0.99	0.99	0.99	0.99	0.99
q95	0.99	0.99	0.99	0.99	0.99	0.99
	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
zstar	341000.00	344000.00	347000.00	3.5e+005	353000.00	356000.00
q5	0.98	0.98	0.98	9.7e-001	0.97	0.96
q95	0.99	0.99	0.99	9.9e-001	0.99	0.99
	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
zstar	359000.00	362000.00	365000.00	368000.00	371000.00	374000.00
q5	0.95	0.95	0.93	0.89	0.86	0.83
q95	0.99	0.99	0.99	0.99	0.99	0.99
	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
zstar	377000.00	3.8e+005	383000.00	386000.00	389000.00	392000.00
q5	0.80	7.6e-001	0.73	0.67	0.59	0.56
q95	0.99	9.9e-001	0.99	0.99	0.99	0.99
	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]



	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
zstar	707000.00	7.1e+005	713000.00	716000.00	719000.00	722000.00
q5	0.42	3.6e-001	0.34	0.30	0.25	0.23
q95	0.99	9.9e-001	0.99	0.99	0.99	0.99
	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]
zstar	725000.00	728000.00	731000.00	734000.00	737000.00	7.4e+005
q5	0.17	0.13	0.10	0.08	0.06	5.0e-002
q95	0.99	0.98	0.98	0.97	0.97	9.5e-001
	[,31]	[,32]	[,33]	[,34]	[,35]	[,36]
zstar	743000.00	746000.00	749000.00	752000.00	755000.00	758000.00
q5	0.04	0.03	0.02	0.01	0.01	0.01
q95	0.94	0.94	0.93	0.90	0.89	0.87
	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]
zstar	761000.00	764000.00	767000.00	7.7e+005	773000.00	776000.00
q5	0.01	0.00	0.00	0.0e+000	0.00	0.00
q95	0.85	0.83	0.81	7.6e-001	0.69	0.63
	[,43]	[,44]	[,45]	[,46]	[,47]	[,48]
zstar	779000.0	782000.00	785000.00	788000.00	791000.00	794000.00
q5	0.0	0.00	0.00	0.00	0.00	0.00
q95	0.6	0.57	0.54	0.49	0.48	0.44
	[,49]	[,50]				
zstar	797000.00	8.0e+005				
q5	0.00	0.0e+000				
q95	0.41	3.9e-001				



## Bibliografía

- [1] Atkinson, A. C. (1985). *Plots, Transformations, and Regression*. New York: Cambridge University Press.
- [2] Balkema, A. A. & Haan, L. de (1974). *Residual life time at great age*. Ann. Probab. 2, 792-804.
- [3] Bradley, P.C. & Thomas, A.L. (2000). *Bayes and empirical Bayes methods for data analysis*. 2nd. edition. Chapman & Hall. New York.
- [4] Berliner & Mac Eachern. (1994). *Subsampling the Gibbs Sampler*. The American Statistician. 48.188-190.
- [5] Berger, J.O. *The robust Bayesian viewpoint (with discussion)*. In Robustness in Bayesian Statistics, e. J. Kadane. Amsterdam: North-Holland.
- [6] Berger, J.O. & Berliner, M. (1986). *Robust Bayes and empirical Bayes analysis with epsilon-contaminated priors*. Annals of Statistics. 14, 461-486.
- [7] Bernardo, J.M. y Smith, A.F.M. (1994). *Bayesian Theory*. Wiley & Sons, New York.
- [8] Besag, J. & Green, P. J. (1993). *Spatial statistics and Bayesian computation*. Journal of the Royal Statistical Society. B55, 25-102.
- [9] Carlin, B. P. & Gelfand, A. E. (1991). *An iterative Monte Carlo method for non-conjugate Bayesian analysis*. Statist. Computing. 1, 119-128.
- [10] Contreras-Cristán, A., O'Reilly, F. & Gutiérrez-Peña E. (2002). Monografías. Editada por el I.I.M.A.S., México.



- [11] Davidson, A.C. & Smith R.L. (1990). *Models for exceedances over high thresholds (with discussion)*. J.R. Statist. Soc. **B52**, 393-442.
- [12] de Haan, L. (1970). *On regular variation and its application to weak convergence of sample extremes*. C.W.I. Tract. 32, Amsterdam.
- [13] de Haan, L. (1984). *Slow variation and the characterization of domains of attraction*. En: Tiago de Oliveira, J. (Ed.) *Statistical Extremes and Applications*, pp. 31-48. Reidel, Dordrecht.
- [14] Embrechts, P., Klüpelberg, C. y Mikosch, T. (1991). *Modelling Extremal Events*. Springer-Verlag, Basel.
- [15] Fisher y Tippett, L.H.C. (1928). *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Proc. Camb. Phil. Soc. **24**.
- [16] Feller, W. (1971). *An introduction to Probability Theory and Its Applications I*, 3rd. edition. Wiley, New York.
- [17] Feller, W. (1971). *An introduction to Probability Theory and Its Applications II*. Wiley, New York.
- [18] Gamerman, D. (1997). *Markov Chain Monte Carlo, stochastic simulation for Bayesian inference*. Chapman & Hall, London.
- [19] Gelman, A., Carlin, J. y Stern, H. (1995). *Bayesian Data Analysis*, 1st. edition. Chapman & Hall, London.
- [20] Gelman A. & Rubin, D.B. (1992). *Inference from iterative simulation using multiple sequences*. Statistical Science, 7,457-511.
- [21] Gilli, K. y Küllezi, E. (2000). *Extreme Value Theory for Tailed-related Measures*. Kluwer Academic Publishers, Netherlands.
- [22] Gnedenko, B.V. (1943). *Sur la distribution limitée du terme d'une série aléatoire*. Ann. Math. **44**, 423-453.

- [23] Gutiérrez-Peña, E. (1997). *Métodos computacionales en la Estadística Bayesiana*. Monografías. Vol. 6, No.15. Editada por el I.I.M.A.S., México.
- [24] Jenkinson, A.F. (1955). *The frequency distribution of annual maximum (or minimum) values of meteorological elements*. Quart. J. Roy. Meteorol. Soc. **81**, 158-171.
- [25] Leadbetter, M.R., Lindgren & G. Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.
- [26] Leamer, E. E. (1978). *Regression selection strategies and revealed priors*. Journal of the american statistical association. **73**, 580-587.
- [27] Lee, P.M. (1989). *Bayesian Statistics: An introduction*. 1st. edition. Oxford University Press, New York, pp. 66-72.
- [28] McCulloch, R. E. (1989). *Local model influence*. Journal of the american statistical association. **84**, 473-478.
- [29] Pickands, J. (1975). *Statistical inference using extreme value order statistics*. Ann. Statist. **3**, 119-131.
- [30] Reiss, R.-D. & Thomas M. (1997). *Statistical analysis of extreme values for insurance, finance, hydrology and other fields*. Birkhäuser-Verlag. Basel, Switzerland.
- [31] Smith, R.L. (1989). *Extreme value analysis of environmental time Series: An application to trend detection in ground-level ozone*. Statistical Science, Vol. 4, No. 4 , pp. 367-393.
- [32] Smith, A.F.M. y Roberts, G.O. (1993). *Bayesian computation via the Gibbs sampler and related Markov chain Montecarlo methods*. J.R. Statist. Soc. B55, pp.3-23.
- [33] Smith, R. L. (1995). *A package for threshold-based Extreme Value Analysis*. U.N.C. Chapell-Hill. Unpublished draft.
- [34] Smith, R. L. (1997). *Extreme value statistics in meteorology*. U.N.C. Chapell-Hill. Unpublished draft.

- [35] Smith, R. L. & Goodman, D. J. (2000). Bayesian Risk Analysis. *Extremes and Integrated Risk Management*, Vol. 17, edited by Embrechts, P. Risk Books, London, pp. 235-251.
- [36] von Mises, R. (1936). *La distribution de la plus grande de  $n$  valeurs*. Rev. Math. Union Interbalcanique. 1, 141-160. Reproduced on selected papers of Richard von Mises, Amer. Soc. 2 (1964), 271-294.
- [37] Wasserman, L. (1992). *Recent methodological advances in robust Bayesian inference (with discussion)*. In Bayesian Statistics 4, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 438-502. New York: Oxford University Press.