

Estimación de un modelo de difusión con saltos con distribución de error generalizada asimétrica usando algoritmos evolutivos

JUAN DAVID OSPINA ARANGO

Director:

MsC. Norman Diego Giraldo Gómez

Profesor Asociado

Escuela de Ingeniería de la Organización

Universidad Nacional de Colombia, Sede Medellín

Trabajo presentado como requisito para optar
al título de Magister en Estadística

Universidad Nacional de Colombia

Sede Medellín

Facultad de Ciencias

Escuela de Estadística

2009



Agradecimientos

Quiero expresar mi más profunda gratitud a mi director en este trabajo, el profesor Norman Diego Giraldo Gómez, y a los profesores Jesús Antonio Hernández Riveros y Juan Carlos Correa Morales, quienes me orientaron acertadamente cada vez que busqué su consejo. También quiero hacer llegar mis sinceros agradecimientos a los jurados de este trabajo, los profesores Elkin Argermiro Castaño Vélez y Víctor Ignacio López Ríos, porque sus observaciones se constituyeron en aportes valiosos para el logro del mismo y para mi formación.

A mi universidad, la Universidad Nacional de Colombia, mi cariño y gratitud porque me dio todo para lanzarme al mundo de la investigación y de la academia.

A la Escuela de Estadística de la Universidad Nacional de Colombia - Sede Medellín, le agradezco el haberme acogido y permitirme acceder a la formación de altísima calidad que imparte.

Resumen

En este trabajo se considera el problema de estimar los parámetros de un proceso de difusión con saltos (*jump-diffusion*) donde la magnitud de los saltos se asume dada por una variable aleatoria que se distribuye bajo la distribución de error generalizada asimétrica (SGED). Se aproxima la densidad de los rendimientos del proceso utilizando la aproximación Bernoulli-Poisson y recurriendo a la transformada rápida de Fourier (FFT). Con la densidad aproximada se encuentra, vía máxima verosimilitud, los parámetros del proceso usando varias técnicas de optimización para encontrar el óptimo de la log-verosimilitud, entre las cuales el mejor desempeño lo brinda la Evolución Diferencial (DE).

Palabras clave: Procesos de difusión con saltos, distribución de error generalizada asimétrica, transformada rápida de Fourier, máxima verosimilitud, evolución diferencial.

Abstract

In this work is considered the parameters estimation problem of a jump-diffusion process with jumps magnitudes given by a random variable distributed under the asymmetric generalized error distribution (SGED). The returns density is approached using the Bernoulli-Poisson approximation and the fast Fourier transform (FFT). With the approached density, using maximum likelihood, the parameters of the models are found through the application of several optimization techniques, among which the Differential Evolution (DE) gives the better performance.

Keywords: Jump-diffusion process, asymmetric generalized error distribution, fast Fourier transform, maximum likelihood, differential evolution.

Índice general

Índice general	I
Índice de tablas	III
Índice de figuras	V
Índice de algoritmos y programas	VII
Introducción	1
1. El Modelo de Difusión con Saltos	3
1.1. El proceso de Wiener o de Difusión	4
1.1.1. Momentos del proceso de Wiener	6
1.2. El proceso Poisson	8
1.2.1. Momentos del proceso Poisson	12
1.2.2. El proceso Poisson compuesto	13
1.3. Distribución de error generalizada asimétrica (SGED)	15
1.4. Solución del modelo de difusión con saltos	19
2. Revisión de los Métodos de Estimación de Modelos de Difusión con Saltos	25
2.1. Estimación por máxima verosimilitud	25
2.2. Estimación usando la función característica	27
2.3. Estimación por GMM, SMM y EMM	28
3. Evolución Artificial Aplicada a la Optimización	31
3.1. Introducción	31
3.2. Métodos heurísticos de optimización	32
3.3. ¿Por qué evolución artificial?	33
3.4. Algoritmos evolutivos	34
3.5. Evolución diferencial	35

4. Estimación por Máxima Verosimilitud	41
4.1. Introducción	41
4.2. Estimación por máxima verosimilitud asumiendo la aproximación Bernoulli para el proceso Poisson	44
4.2.1. Análisis aplicado del problema de estimación	50
4.2.2. Aplicaciones	61
4.3. Estimación por máxima verosimilitud asumiendo que el proceso de saltos es Poisson	68
Conclusiones	71
A. Operaciones genéticas y mecanismos de diferenciación	73
A.1. Operaciones genéticas	73
A.1.1. La aptitud	74
A.1.2. La selección	74
A.2. Estrategias de diferenciación	75
Bibliografía	77

Índice de tablas

4.1. Resultados de la estimación con <i>optim</i> usando BFGS con condición inicial θ_{01} . . .	52
4.2. Resultados de la estimación con <i>optim</i> usando BFGS con condición inicial θ_{02} . . .	52
4.3. Resultados de la estimación con <i>optim</i> usando L-BFGS-B con condición inicial θ_{01} . . .	53
4.4. Resultados de la estimación con <i>optim</i> usando L-BFGS-B con condición inicial θ_{02} . . .	53
4.5. Resultados de la estimación con <i>optim</i> usando L-BFGS-B con condición inicial θ_{02} y restricciones más amplias.	54
4.6. Resultados de la estimación con <i>optim</i> usando temple simulado con condición ini- cial θ_{02}	55
4.7. Resultados de la estimación con <i>optim</i> usando evolución diferencial, los límites θ_L y θ_U y los límites $\theta_{L'}$ y $\theta_{U'}$	55
4.8. Estadísticos descriptivos de la muestra de estimadores.	63
4.9. Resultados para la acción de IBM.	64
4.10. Resultados para la acción de IBM obtenidos en (Ramezani y Zeng, 1998).	65
4.11. Resultados para la acción de Laboratorios Abbott.	68

Índice de figuras

1.1. Trayectorias simuladas de un proceso de Wiener de un solo estado con diferentes valores para Δt	7
1.2. Trayectorias simuladas de un proceso Poisson homogéneo con $\lambda = 8$ y diferentes valores para Δt	11
1.3. Trayectorias simuladas de un proceso Poisson compuesto con $\lambda = 8$, proceso subyacente iid SGED y diferentes valores para Δt y h la función identidad.	16
1.4. Función de densidad SGED para diferentes valores de los parámetros ν y ξ	19
1.5. Trayectorias simuladas de un proceso de difusión con saltos $\mu = 0.25$, $\sigma = 0.5$, $\lambda = 4$, proceso subyacente a los saltos $Q_k \sim SGED(0, 0.3, 1.5, 2)$ y diferentes valores para Δt	24
3.1. Un esquema general de métodos de optimización.	38
3.2. Superficie definida por la función de Rastrigin.	39
3.3. Evolución de la población en la optimización de la función de Rastrigin.	40
4.1. Densidad aproximada del proceso.	47
4.2. Trayectoria, rendimientos e histograma de una simulación de un proceso de difusión con saltos asumiendo la aproximación Bernoulli de la Poisson.	49
4.3. Evolución del valor de -log-verosimilitud y de los estimadores de los parámetros λ , μ , y σ	56
4.4. Evolución de los estimadores de los parámetros μ_q , σ_q , ν y ξ	57
4.5. -log-verosimilitud restringida evaluada para los parámetros λ , μ , σ y μ_q	58
4.6. -log-verosimilitud restringida evaluada para los parámetros σ_q , ν , y ξ	59
4.7. -log-verosimilitud restringida evaluada para los parámetros (λ, μ) , (λ, σ) y (μ, σ)	60
4.8. Histogramas de los estimadores de los parámetros λ , μ , σ y μ_q	61
4.9. Histogramas de los estimadores de los parámetros σ_q , ν , y ξ	62
4.10. Series del precio y rendimientos e histograma de los rendimientos de la acción de IBM.	64
4.11. Funciones de autocorrelación y autocorrelación parcial de los rendimientos de la acción de IBM.	65

4.12. Ajuste del modelo estimado a los rendimientos de IBM.	66
4.13. Series del precio y rendimientos e histograma de los rendimientos de la acción de Laboratorios Abbott.	67
4.14. Funciones de autocorrelación y autocorrelación parcial de los rendimientos de la acción de IBM.	68
4.15. Ajuste del modelo estimado a los rendimientos de Laboratorios Abbot.	69

Índice de algoritmos y programas

1.	Simulación de una trayectoria del proceso de Wiener.	5
2.	Código en R para simular el proceso de Wiener.	6
3.	Código en R para simular un proceso de Wiener de tres estados independientes entre sí.	6
4.	Simulación de una trayectoria de un proceso Poisson homogéneo.	10
5.	Código en R para simular el proceso de Poisson homogéneo.	10
6.	Simulación de una trayectoria de un proceso Poisson homogéneo con la aproximación binomial.	12
7.	Código en R para simular el proceso Poisson con la aproximación binomial.	12
8.	Simulación de una trayectoria de un proceso Poisson compuesto.	15
9.	Simulación de una trayectoria del proceso de difusión con saltos de las ecuaciones (1.53) y (1.58).	23
10.	Algoritmo Evolutivo General	35
11.	Algoritmo de Evolución Diferencial	36
12.	Código en R para aproximar la densidad del proceso asumiendo la aproximación Bernoulli del proceso Poisson.	48
13.	Código en R para aproximar la densidad del proceso usando interpolación lineal.	50
14.	Código en R para evaluar la verosimilitud usando interpolación lineal.	51

Introducción

Desde su introducción, el proceso de difusión o movimiento browniano ha ganado gran aceptación en muchos campos para modelar fenómenos que se asumen estocásticos. A partir del desarrollo del cálculo estocástico y de la teoría de valoración de opciones se propone el proceso de difusión para modelar los precios en mercados perfectos. Esta teoría se amplía para modelar precios que presentan discontinuidades por saltos y surge de esta manera el proceso de difusión con saltos (*jump-diffusion*). Desde que fue propuesto, este modelo ha sido utilizado en econometría intensivamente, así como también en otros campos. El uso de procesos estocásticos como modelos de procesos reales, constantemente conlleva a estimar los parámetros de dichos modelos para entender, a través de estos, aspectos de la realidad que se quiere capturar con el modelo.

Son muchas las propuestas para estimar modelos de difusión con saltos, todas ellas dependientes del tipo de distribución que se asume para los saltos y de la complejidad del modelo. En este trabajo se desarrolla una metodología para la estimación de procesos de difusión con saltos, donde los saltos se asumen dados de una Distribución de Error Generalizada Asimétrica (SGED). Esta elección se justifica a partir de una revisión de la literatura donde se expone que la mayoría de los rendimientos financieros exhiben altas curtosis y asimetrías, características que posee la SGED. Esto, sumado a la presencia de fuertes incrementos que aparecen en forma esporádica, conlleva a pensar que la SGED podría aportar conocimiento del mercado que se modele con este proceso. El modelo que se considera corresponde a un proceso de difusión con saltos unidimensional y de parámetros constantes.

Este trabajo hace aportes en tres aspectos: 1) Explora el comportamiento del proceso de difusión con saltos utilizando una distribución SGED para los saltos; 2) Propone una metodología para aproximar numéricamente la densidad del proceso utilizando la transformada rápida de Fourier (FFT) y 3) Explora la capacidad y ventajas de los algoritmos de evolución artificial para encontrar el óptimo de la log-verosimilitud aproximada del proceso.

En el primer capítulo se introduce el modelo de difusión con saltos y la SGED. Posteriormente se presenta una revisión de la literatura sobre estimación de procesos de difusión con saltos. Seguidamente se hace una corta exposición de los métodos de evolución artificial para justificar su uso en este trabajo. Luego, en el capítulo 4 se propone una metodología para aproximar numéricamente la densidad del proceso cuando se asume la aproximación Bernoulli-Poisson y se lleva a cabo un análisis de su desempeño con datos simulados para mostrar después la aplicación con datos reales.

En este mismo capítulo se discute la dificultad de implementar una metodología similar para el caso en que la aproximación Bernoulli-Poisson no es realista. En la parte final se discute los resultados y se plantea trabajos futuros.

Capítulo 1

El Modelo de Difusión con Saltos

En esta sección se presenta formalmente el modelo de difusión con saltos (*jump-diffusion*), cuya estimación es el objetivo principal de este trabajo. El modelo de difusión con saltos es un modelo de proceso estocástico en tiempo continuo cuyo antecedente es el modelo propuesto en (Black y Scholes, 1973) por Fischer Black y Myron Scholes para estudiar la dinámica del precio de activos financieros y de las opciones europeas del tipo *call*. El documento donde se proponía el modelo de Black y Scholes fue rechazada por las revistas *Journal of Political Economy* y *Review of Economics and Statistics*, siendo la primera de estas la que finalmente publicaría el trabajo corregido por Robert Merton. Este último, en (Merton, 1973), muestra que el modelo de Black y Scholes se puede plantear a partir de una ecuación diferencial estocástica (EDE) como la que se muestra en la ecuación (1.1), donde $S(t)$ es el precio del activo en el tiempo t , μ es el valor esperado instantáneo del precio, σ^2 es la varianza instantánea del precio y $W(t)$ es un proceso de Wiener estándar.

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t). \quad (1.1)$$

La solución de dicha EDE se obtiene a partir de la teoría de Cálculo Estocástico propuesta por Itô en (Itô, 1951). El modelo de difusión con saltos es planteado por Merton en (Merton, 1976) adicionando el término $v dP_t$ al lado derecho de la ecuación (1.1), donde P_t es un proceso Poisson homogéneo y v es la magnitud de los saltos. El análisis de Merton busca modelar las discontinuidades o saltos que presentan los precios de algunos activos. En su planteamiento inicial, se asume que la magnitud de los saltos v es constante en el tiempo, dando como resultado la ecuación 1.2

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t) + v dP(t). \quad (1.2)$$

En (Hanson, 2006) se muestra que este modelo univariado puede extenderse a uno multivariado de n dimensiones y de coeficientes variables en el tiempo, dónde el término de los saltos es una

variable aleatoria $\mathbf{J}(t) = [J_i(t)]$, sobre cuyas componentes se impone la condición $J_i(t) > -1$ con probabilidad 1. Este modelo se presenta en la ecuación (1.3)

$$d\mathbf{X}(t) = \mathbf{X}(t) (\mathbf{a}(t) dt + \mathbf{b}(t) d\mathbf{W}(t) + \mathbf{J}(t) d\mathbf{P}(t)), \quad (1.3)$$

donde $\mathbf{X}(t) = [X_i(t)]$ es el vector del proceso, $\mathbf{a}(t) = [a_i(t)]$ y $\mathbf{b}(t) = [b_i(t)]$ son funciones vectoriales del tiempo y $\mathbf{W}(t) = [W_i(t)]$ y $\mathbf{P}(t) = [P_i(t)]$ son procesos de Wiener y Poisson multivariados.

1.1. El proceso de Wiener o de Difusión

El proceso de Wiener, de difusión o movimiento browniano, es un proceso estocástico en tiempo continuo ($t \geq 0$). A lo largo de este trabajo se utiliza \mathbf{W}_t para referirse a dicho proceso. El proceso de Wiener, al igual que el proceso Poisson, es un proceso markoviano en tiempo continuo, lo que equivale a satisfacer la siguiente definición:

Definición 1 **[Proceso de Markov o proceso markoviano] Un proceso estocástico en tiempo continuo $\mathbf{X}(t)$ es un proceso de Markov si*

$$Prob(\mathbf{X}(t + \Delta t) = \mathbf{x} | \mathbf{X}(s), s \leq t) = Prob(\mathbf{X}(t + \Delta t) = \mathbf{x} | \mathbf{X}(t)), \quad (1.4)$$

para cualesquiera $t \geq 0, \Delta t \geq 0$ y \mathbf{x} en el espacio de estados del proceso.

El proceso de Wiener puede caracterizarse por las siguientes propiedades. Es un proceso de Markov que satisface:

1. $\mathbf{W}(0) = 0$, con probabilidad 1.
2. Es continuo.
3. Sus incrementos son independientes e idénticamente distribuidos normal.

La propiedad 1 es una propiedad de estandarización. Si el proceso no comienza en cero sino, dígame, en t_0 , puede escribirse $\mathbf{W}(t_0) = 0$, con probabilidad 1. La propiedad 2 puede expresarse formalmente como en la ecuación (1.5):

$$\mathbf{W}(t^-) = \lim_{t_0 \rightarrow t^-} \mathbf{W}(t_0) = \mathbf{W}(t) = \lim_{t_0 \rightarrow t^+} \mathbf{W}(t_0) = \mathbf{W}(t^+), \text{ casi en todas partes.} \quad (1.5)$$

La propiedad 3, que entre otras cosas implica el carácter estacionario de los incrementos del proceso de Wiener, resulta clave para simular trayectorias de este proceso. Esta propiedad establece que $\mathbf{W}(t) - \mathbf{W}(s) \sim N(\mathbf{0}, \mathbf{R}(t - s))$, para $0 \leq s < t$, donde $\mathbf{R}(t - s)$ es la matriz de covarianzas y depende exclusivamente de $t - s$.

En el caso univariado, si $t_1 < t_2 < t_3 < t_4$, entonces $W(t_2 - t_1)$ es independiente de $W(t_4 - t_3)$ y ambos incrementos se distribuyen normal con media cero y varianza $t_2 - t_1$ y $t_4 - t_3$, respectivamente. En general, se puede mostrar que $W(t) - W(s) \sim N(0, |t - s|)$, ver (Hanson, 2006), página 82.

Como se mencionó anteriormente, estas propiedades del proceso de Wiener resultan de gran utilidad para simularlo. Si se define el incremento del proceso de Wiener univariado en el tiempo t como en la ecuación (1.6) para $t \geq 0$ y $\Delta t > 0$

$$\Delta W(t) = W(t + \Delta t) - W(t), \quad (1.6)$$

entonces estos satisfacen

$$\Delta W(t) \sim N(0, \Delta t). \quad (1.7)$$

Ahora, si el proceso está definido en el intervalo $[0, T]$ y se considera una partición de N subintervalos de igual longitud, se puede definir $\Delta t = \frac{T}{N}$, $t_0 = 0$, $t_{N+1} = T$ y $t_i = i \times \Delta t$. Adicionalmente $\Delta \mathbf{W}(t_i) = \mathbf{W}(t_i + \Delta t) - \mathbf{W}(t_i)$. De esta manera se tiene

$$\mathbf{W}(t_i) = \sum_{j=0}^i \Delta \mathbf{W}(t_j). \quad (1.8)$$

Utilizando la ecuación (1.8) puede simularse una trayectoria del proceso de Wiener como se muestra en el algoritmo 1. El algoritmo 2 muestra el código en R para simular una trayectoria del proceso de Wiener univariada y el algoritmo 3 muestra el código en R para simular un proceso de Wiener multivariado de tres estados independientes entre sí. En la figura 1.1 se muestra distintas trayectorias simuladas de un proceso de Wiener para diferentes valores de Δt . Por supuesto, mientras más pequeño sea Δt más realista será la simulación de las trayectorias del proceso de Wiener. Así, las gráficas (c) y (d) de la figura 1.1 se aproximan más a las realizaciones del proceso de Wiener..

Algoritmo 1 Simulación de una trayectoria del proceso de Wiener.

- 1: $\Delta t := \frac{T}{N}, t_i := i \times \Delta t$.
 - 2: Obtenga N muestras $\Delta \mathbf{W}(t_i)$ de una $N(\mathbf{0}, \mathbf{R}(\Delta t))$
 - 3: Haga $\mathbf{W}(t_i) = \sum_{j=0}^i \Delta \mathbf{W}(t_j)$
-

Como el objeto de este trabajo es la estimación de un proceso de difusión con saltos univariado, de este punto en adelante sólo se mencionará las propiedades y particularidades de los modelos univariados del proceso de Wiener y de Poisson. El proceso de Wiener, además de ser un proceso markoviano continuo y de incrementos *iid* normales, es un proceso con incrementos estacionarios. Esto se desprende de que la varianza de los incrementos no depende del tiempo, sino solamente de

Algoritmo 2 Código en R para simular el proceso de Wiener.

```
t=seq(0,T,Dt) #T: tiempo final, Dt=T/N, N: número de subintervalos.  
DWt=rnorm((length(t)-1),0,sqrt(Dt))  
Wt=cumsum(c(0,DWt))
```

Algoritmo 3 Código en R para simular un proceso de Wiener de tres estados independientes entre sí.

```
library(MASS)  
t=seq(0,T,Dt) #T: tiempo final, Dt=T/N, N: número de subintervalos.  
R=Dt*matrix(c(1,0,0,0,1,0,0,0,1),ncol=3)  
DWt=mvrnorm((length(t)-1),c(0,0,0),R)  
Wt=apply(rbind(c(0,0,0),DWt),2,cumsum)
```

la distancia entre ellos y además de que la covarianza entre $W(s)$ y $W(t)$ es igual al mín (s, t) . Para ver esto último, suponga sin pérdida de generalidad que $s < t$. Así, se tiene la ecuación (1.9)

$$\begin{aligned} Cov [W(t), W(s)] &= E [W(t) W(s)] \\ &= E [W(t) W(s) + W^2(s) - W^2(s)] \\ &= E [W^2(s) + W(s) (W(t) - W(s))] \\ &= E [W^2(s)] + E [(W(s) - W(0)) (W(t) - W(s))] \\ &= Var [W(s)] = s, \end{aligned} \tag{1.9}$$

donde se ha utilizado el hecho de que $W(s) - W(0)$ es independiente de $W(t) - W(s)$.

1.1.1. Momentos del proceso de Wiener

Los métodos de estimación basados en momentos, mencionados en el capítulo 2, para estimar el proceso de difusión con saltos dependen del cálculo de los momentos (o sus estimadores) de los procesos de Wiener y de Poisson. Debido a que la función de densidad de probabilidad de los incrementos del proceso de Wiener es par - de hecho es una densidad normal con media cero - se tiene que $E [(\Delta W(t))^{2k+1}] = 0$, para $k = 1, 2, \dots$. Es decir que la esperanza de las potencias impares de $\Delta W(t)$ es cero. En (Hanson, 2006), página 83, se muestra que los momentos de orden $m = 2k$, con $k = 0, 1, 2, \dots$ se obtienen como en la ecuación 1.10.

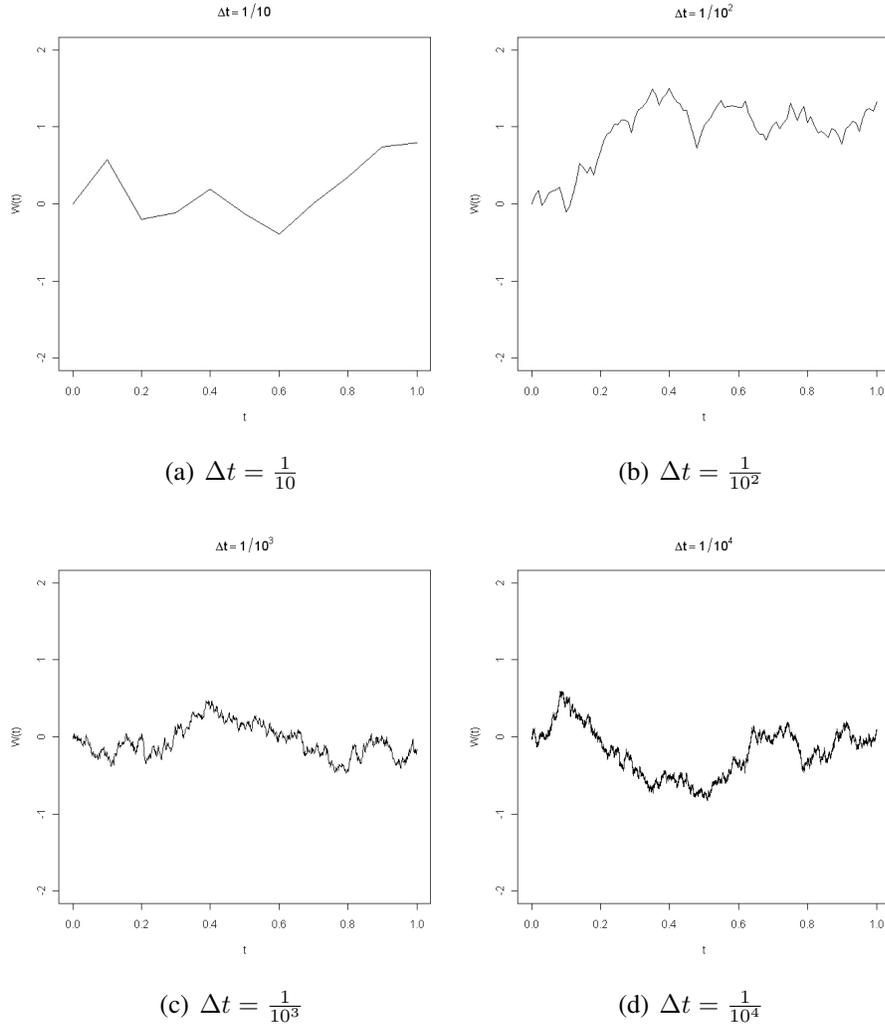


Figura 1.1: Trayectorias simuladas de un proceso de Wiener de un solo estado con diferentes valores para Δt .

$$\begin{aligned}
 E \left[(\Delta W(t))^{2k} \right] &= \int_{-\infty}^{+\infty} \frac{w^{2k}}{\sqrt{2\pi\Delta t}} e^{-\frac{w^2}{2\Delta t}} dw \\
 &= \frac{2}{\sqrt{2\pi\Delta t}} \int_0^{+\infty} w^{2k} e^{-\frac{w^2}{2\Delta t}} dw \\
 &= \frac{2(2\Delta t)^k}{\sqrt{2\pi\Delta t}} \int_0^{+\infty} \left(\frac{w^2}{2\Delta t} \right)^k e^{-\frac{w^2}{2\Delta t}} dw =
 \end{aligned} \tag{1.10}$$

$$\begin{aligned}
&= \frac{(2\Delta t)^k}{\sqrt{\pi}} \int_0^{+\infty} u^{k-\frac{1}{2}} e^{-u} du \\
&= \frac{(2\Delta t)^k \Gamma\left(k + \frac{1}{2}\right)}{\sqrt{\pi}}.
\end{aligned}$$

Así mismo, si se reemplaza $\Delta W(t)$ por $|\Delta W(t)|$ en la ecuación (1.10), ésta es válida ahora para cualquier entero m , lo que resulta en la ecuación (1.11).

$$E[|\Delta W(t)|^m] = \frac{(2\Delta t)^{\frac{m}{2}} \Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi}}. \quad (1.11)$$

En forma más simple, los momentos del valor absoluto de los incrementos del proceso de Wiener se pueden definir para $m = 2k$ y $m = 2k + 1$, $k = 0, 1, \dots$, como en las ecuaciones (1.12) y (1.13), respectivamente.

$$E\left[|\Delta W(t)|^{2k}\right] = (2k - 1)!! (\Delta t)^k, \quad (1.12)$$

$$E\left[|\Delta W(t)|^{2k+1}\right] = (2k)!! (\Delta t)^k \sqrt{\frac{2\Delta t}{\pi}}, \quad (1.13)$$

donde la función doble factorial $(.)!!$ se define para argumentos enteros impares como en la ecuación (1.14) y para argumentos pares como en la ecuación 1.15:

$$(2k - 1)!! = (2k - 1)(2k - 3) \dots 1, \quad (1.14)$$

$$(2k)!! = 2^k k! \quad (1.15)$$

Debido a que las técnicas de estimación empleadas en este trabajo se basan exclusivamente en los incrementos de los procesos de Wiener ($\Delta W(t)$) y de Poisson ($\Delta P(t)$) y no en sus diferenciales ($dW(t)$ y $dP(t)$), no se enfatizará en los momentos de estos últimos, aunque se obtienen al cambiar Δt por dt y aproximando todas las potencias de orden mayor que uno de dt a cero, es decir, para $k = 2, 3, \dots$, se asume $(dt)^k \approx 0$.

1.2. El proceso Poisson

En este apartado se menciona algunas de las características más importantes del proceso Poisson homogéneo $P(t)$ y del proceso Poisson compuesto $\Pi(t)$ univariados. Al igual que el proceso de Wiener, el proceso de Poisson es un proceso markoviano en tiempo continuo, es decir, satisface

la definición 1. El papel de este proceso dentro del modelo de difusión con saltos, es suministrar la característica de los saltos que, dependiendo del contexto, se asumen como desastres, eventos inesperados, bonanzas atípicas, etc. Este proceso se puede caracterizar por las siguientes propiedades:

1. $P(0) = 0$, con probabilidad 1.

2. El proceso Poisson tiene saltos unitarios: si T_k es el tiempo donde se da el k -ésimo salto del proceso, entonces

$$P(T_k^+) = P(T_k^-) + 1, \quad (1.16)$$

con $P(T_k^+)$ y $P(T_k^-)$ los límites por derecha e izquierda de $P(T_k)$, respectivamente.

3. Es discontinuo a derecha: si $t > 0$, entonces

$$P(t^+) = P(t) \geq P(t^-). \quad (1.17)$$

4. Sus incrementos son independientes y se distribuyen Poisson.

5. $P(t) \sim Poisson(\lambda t)$, con $\lambda > 0$ y $t \geq 0$.

En el caso del proceso Poisson, la propiedad 1, aunque al igual que en el proceso de Wiener se trata de estandarización, se deriva de la propiedad 5, ya que cuando $\lambda t \rightarrow 0^+$, $Prob(P(t) = 0) \rightarrow 1$. De las propiedades 2 y 4 se deriva una metodología para simular trayectorias del proceso Poisson. Específicamente, la propiedad 4 aún se debe definir formalmente. Ésta establece que para todo $0 \leq t_1 < t_2 < t_3 < t_4$, entonces $P(t_2) - P(t_1) \sim Poisson(\lambda(t_2 - t_1))$, $P(t_4) - P(t_3) \sim Poisson(\lambda(t_4 - t_3))$ y $P(t_2) - P(t_1)$ es independiente de $P(t_4) - P(t_3)$. Esta notación se usa para indicar que la tasa promedio de los incrementos es una función de la diferencia en que se miden los tiempos.

Como se hizo para el proceso de Wiener, se define el incremento del proceso Poisson en el tiempo t como en la ecuación (1.18), para $\Delta t > 0$

$$\Delta P(t) = P(t + \Delta t) - P(t), \quad (1.18)$$

estos satisfacen

$$\Delta P(t) \sim Poisson(\lambda \Delta t). \quad (1.19)$$

De esta manera, si el proceso está definido en el intervalo $[0, T]$ y se considera una partición de N subintervalos de igual longitud $\Delta t = \frac{T}{N}$ y se define $t_0 = 0$, $t_i = i\Delta t$, $i = 1, 2, \dots, N$, entonces se puede escribir $P(t_i)$ como

$$P(t_i) = \sum_{j=0}^i \Delta P(t_j). \quad (1.20)$$

Con la ecuación 1.20 puede construirse un algoritmo para simular un proceso Poisson homogéneo como el que se muestra en el algoritmo 4. El código en R de este procedimiento se presenta en el algoritmo 5. La figura 1.2 muestra diferentes trayectorias de un proceso Poisson homogéneo con $\lambda = 4$. Estrictamente hablando, estas trayectorias no corresponden a un proceso Poisson sino a mediciones de un proceso Poisson. Esto quiere decir que lo que se ha hecho es simular las trayectorias que podrían observarse midiendo el estado del proceso a determinada frecuencia de muestreo.

Algoritmo 4 Simulación de una trayectoria de un proceso Poisson homogéneo.

- 1: $\Delta t := \frac{T}{N}, t_i := i \times \Delta t.$
 - 2: Obtenga N muestras $\Delta P(t_i)$ de una *Poisson* ($\lambda \Delta t$)
 - 3: Haga $P(t_i) = \sum_{j=0}^i \Delta P(t_j)$
-

Algoritmo 5 Código en R para simular el proceso de Poisson homogéneo.

```
t=seq(0,T,Dt) #T: tiempo final, Dt=T/N, N: número de subintervalos.
DPt=rpois(length(t)-1,lambda*Dt)
Pt=cumsum(c(0,DPt))
```

Otra característica del proceso Poisson es que los tiempos entre saltos siguen una distribución exponencial. Es decir, si T_k y T_{k+1} son los tiempos en los que se dan el k -ésimo y el $(k + 1)$ -ésimo saltos, respectivamente, y $\Delta T_k = T_{k+1} - T_k$, entonces $\Delta T_k \sim Exp(\frac{1}{\lambda})$. De esta característica del proceso Poisson se deriva una técnica usada frecuentemente para simularlo. Dicha técnica consiste en simular los tiempos donde se dan los saltos y luego reconstruir el proceso a partir de la propiedad 2 del proceso Poisson. Las trayectorias obtenidas de esta manera son más realistas que las que se presentan en la figura 1.2.

Si se asume que para el proceso Poisson no se dispone de los tiempos en que ocurren los saltos, sino que con determinada frecuencia se mide el estado del proceso, entonces la estrategia basada en simular los tiempos en que ocurren los saltos ya no es práctica, pues estos tiempos no necesariamente coinciden con los tiempos de medición. En este caso se emplea otra característica del proceso Poisson que lleva a una técnica simple para simularlo, conocida como la aproximación Poisson a la distribución binomial. En otras palabras, esta aproximación asume que en un lapso de tiempo suficientemente pequeño puede ocurrir a lo más un salto. Esto permite simplificar el modelo de saltos a uno binomial. Para verlo, se considera la expansión en series de Taylor de $Prob(\Delta P(t) = 0) = e^{-\lambda \Delta t}$, $Prob(\Delta P(t) = 1) = e^{-\lambda \Delta t} \lambda \Delta t$ y $Prob(\Delta P(t) > 1)$ alrededor de cero,

$$\begin{aligned}
\text{Prob}(\Delta P(t) = 0) &= e^{-\lambda\Delta t} = 1 - \lambda\Delta t + \frac{1}{2}(\lambda\Delta t)^2 + O^3(\lambda\Delta t), \\
\text{Prob}(\Delta P(t) = 1) &= e^{-\lambda\Delta t}\lambda\Delta t = \lambda\Delta t - (\lambda\Delta t)^2 + O^3(\lambda\Delta t), \\
\text{Prob}(\Delta P(t) > 1) &= 1 - \text{Prob}(\Delta P(t) \leq 1) = \frac{1}{2}(\lambda\Delta t)^2 + O^3(\lambda\Delta t).
\end{aligned}
\tag{1.21}$$

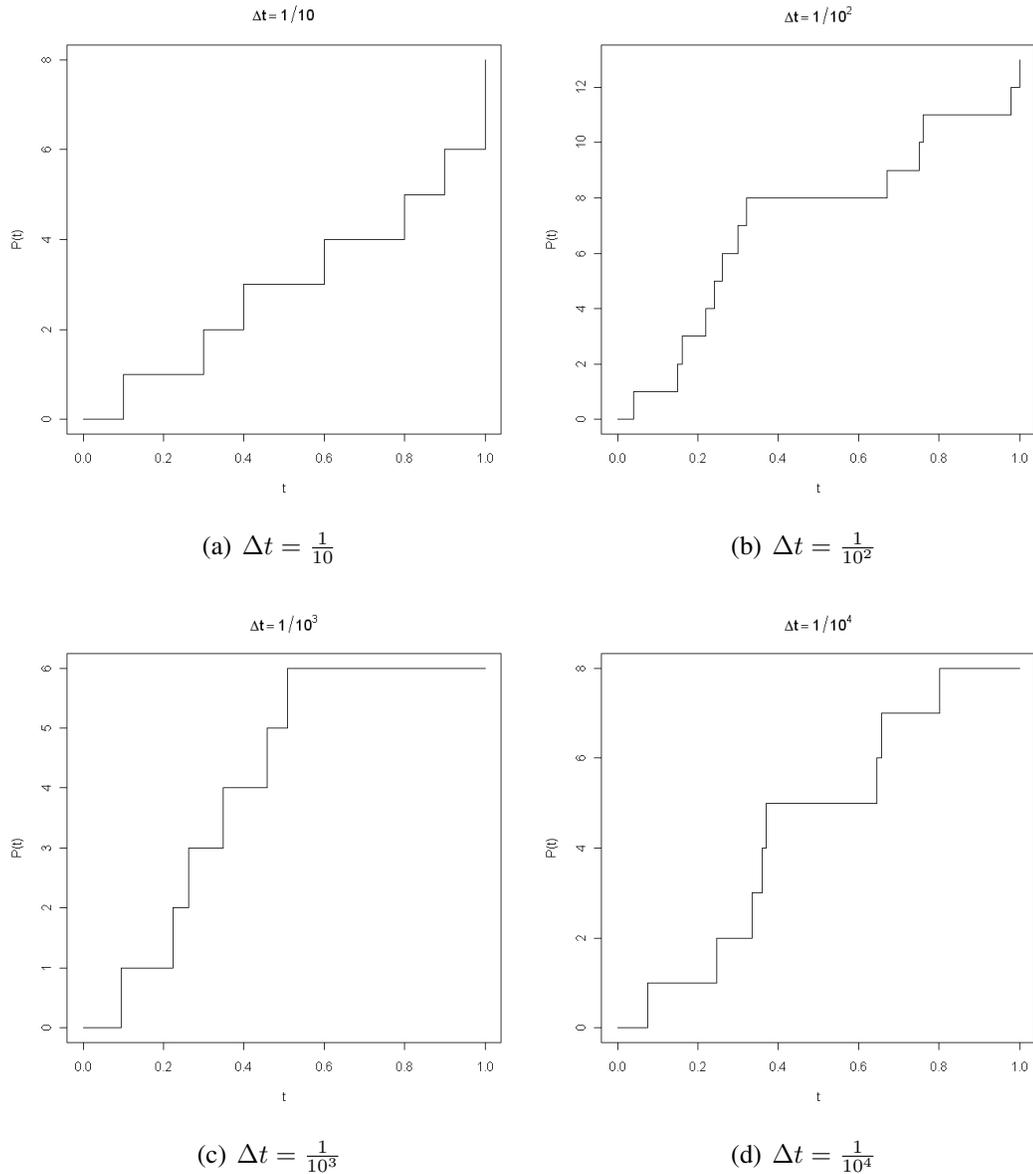


Figura 1.2: Trayectorias simuladas de un proceso Poisson homogéneo con $\lambda = 8$ y diferentes valores para Δt .

Luego, si $\lambda\Delta t$ es suficientemente pequeño, los términos $(\lambda\Delta t)^2$ y $O^3(\lambda\Delta t)$ (el residuo de Taylor) son despreciables, de tal forma que se tiene las siguientes aproximaciones:

$$\begin{aligned} \text{Prob}(\Delta P(t) = 0) &\approx 1 - \lambda\Delta t, \\ \text{Prob}(\Delta P(t) = 1) &\approx \lambda\Delta t, \\ \text{Prob}(\Delta P(t) > 1) &\approx 0. \end{aligned} \tag{1.22}$$

Así, la distribución de $\Delta P(t)$ se puede aproximar con una distribución binomial de parámetros 1 y $\lambda\Delta t$, es decir $\Delta P(t) \sim \text{Bin}(1, \lambda\Delta t)$ ó $\Delta P(t) \sim \text{Bernoulli}(\lambda\Delta t)$. Esto permite simular el proceso Poisson homogéneo utilizando variables binomiales. En la literatura, cuando se utiliza esta aproximación en el modelo de difusión con saltos se habla de un proceso de difusión con saltos binomial (*binomial jump-diffusion*) ó de difusión con saltos Bernoulli (*bernoulli jump-diffusion*) ((Ball y Torous, 1983), (Ball y Torous, 1985) y (Honoré, 1988)). Bajo esta aproximación es posible simular el proceso Poisson homogéneo como se muestra en el algoritmo 6 e implementar dicha simulación en R como en el algoritmo 7. La idea bajo este esquema de simulación es que el proceso Poisson es muestreado con un periodo de muestreo Δt y constituye una aproximación razonable para simular el proceso Poisson si Δt es suficientemente pequeño.

Algoritmo 6 Simulación de una trayectoria de un proceso Poisson homogéneo con la aproximación binomial.

- 1: $\Delta t := \frac{T}{N}, t_i := i \times \Delta t.$
 - 2: Obtenga N muestras $\Delta P(t_i)$ de una $\text{Bin}(1, \lambda\Delta t)$
 - 3: Haga $P(t_i) = \sum_{j=0}^i \Delta P(t_j)$
-

Algoritmo 7 Código en R para simular el proceso Poisson con la aproximación binomial.

```
t=seq(0,T,Dt) #T: tiempo final, Dt=T/N, N: número de subintervalos.
DPt=rbinom(length(t)-1,1,lambda*Dt)
Pt=cumsum(c(0,DPt))
```

1.2.1. Momentos del proceso Poisson

A partir del hecho de que el proceso Poisson es un proceso markoviano en (Hanson, 2006), página 94, se muestra que

$$\text{Cov}[P(s), P(t)] = \lambda \text{mín}(s, t). \tag{1.23}$$

En esta última referencia se muestra también que, en general, los momentos de los incrementos del proceso Poisson homogéneos vienen dados por las ecuaciones (1.24) y (1.25):

$$E [(\Delta P(t))^m] = e^{-\lambda\Delta t} \sum_{k=0}^{+\infty} \frac{(\lambda\Delta t)^k k^m}{k!}, \quad (1.24)$$

$$E [(\Delta P(t))^m] = \left[e^{-u} \left\{ \left(u \frac{d}{du} \right)^m e^u \right\} \right]_{|u=\lambda\Delta t}. \quad (1.25)$$

De esta manera se puede calcular el valor esperado de cualquier potencia entera de $\Delta P(t)$. Utilizando estas fórmulas se puede verificar que la varianza del proceso Poisson homogéneo es $Var [\Delta P(t)] = \lambda\Delta t$.

1.2.2. El proceso Poisson compuesto

Al proceso Poisson se le llama *proceso Poisson compuesto* o *proceso Poisson marcado* si la amplitud de sus saltos está determinada por una variable aleatoria. El proceso Poisson compuesto es una suma aleatoria de variables aleatorias y tiene aplicaciones en áreas como la actuaría. A la variable aleatoria subyacente, que determina la amplitud de los saltos, se le llama *marca* y usualmente se asume que estas marcas son un proceso $Q(t)$ de variables independientes e idénticamente distribuidas. De esta manera, en el tiempo T_k donde se da el k -ésimo salto del proceso Poisson homogéneo $P(t)$, el valor la magnitud de dicho salto está determinada por la realización de la variable aleatoria Q_k ($Q_k := Q(T_k)$) o una función h de ésta, de manera se puede escribir $\Pi(t, Q(t))$ ó $\Pi(t, h(Q(t)))$ para denotar el proceso Poisson compuesto. Sin embargo, para no recargar la notación, el proceso Poisson compuesto se denota simplemente por $\Pi(t)$, especificando cuando corresponda el proceso de marcas $Q(t)$ y la función h .

De esta manera, el proceso Poisson compuesto y sus incrementos se pueden definir en términos de un proceso Poisson homogéneo, $P(t)$, y el proceso de marcas subyacente, $Q(t)$, como en las ecuaciones (1.26) y (1.27), respectivamente,

$$\Pi(t) = \sum_{k=1}^{P(t)} h(Q_k), \quad (1.26)$$

$$\Delta\Pi(t) = \sum_{k=1}^{\Delta P(t)} h(Q_k), \quad (1.27)$$

con la convención de la sumatoria inversa $\sum_{j=1}^0 a_j = 0$. Esto, equivale precisamente a que $\Pi(t) = 0$ si y sólo si $P(t) = 0$.

En el proceso de difusión con saltos interviene un proceso Poisson compuesto en el cual, la magnitud de los saltos depende posiblemente, además del proceso de marcas $Q(t)$, del tiempo y del estado del proceso de difusión con saltos, dígame $X(t)$. Para expresar esta dependencia en

forma funcional se cambia $h(Q(t))$ por $h(X(t), t, Q(t))$, entendiendo que al tener un número de argumentos distinto se trata de funciones diferentes.

El proceso de difusión con saltos se puede definir a partir de una ecuación diferencial estocástica en la que interviene el diferencial del proceso Poisson compuesto. Si se asume que la magnitud de las marcas depende del estado del proceso de difusión con saltos, $X(t)$, entonces se puede utilizar la siguiente convención para denotar el diferencial del proceso Poisson compuesto $d\Pi(X(t), t)$:

$$h(X(t), t, Q) dP(t) := d\Pi(X(t), t). \quad (1.28)$$

Si en cambio, este supuesto no es necesario, se puede escribir

$$h(t, Q) dP(t) := d\Pi(t). \quad (1.29)$$

La notación de la ecuación (1.28) y la ecuación (1.29), utilizada en (Hanson, 2006), ayuda a ver que un cambio diferencial en el proceso Poisson compuesto está dado por un cambio diferencial en un proceso Poisson homogéneo, ponderado por la función que determina la magnitud de los saltos. El diferencial del proceso Poisson homogéneo es una variable aleatoria que toma los valores cero ó uno, de manera que el cambio diferencial en el proceso Poisson homogéneo es el valor de la función h si $dP(t) = 1$ o cero en otro caso. En forma integral, por la regla de Íto, se tiene:

$$\int_0^t d\Pi(X(t), t) = \int_0^t h(X(t), t, Q) dP(t) = \sum_{k=0}^{P(t)} h(X(T_k), T_k, Q_k). \quad (1.30)$$

Los momentos del proceso Poisson compuesto son de interés en actuaría y se usan en la estimación de modelos de difusión con saltos a través de metodologías GMM (Método de los Momentos Generalizado), EMM (Método de los Momentos Eficiente) y SMM (Método de los Momentos Simulados), que se comentan en el capítulo 2. En este trabajo es importante la distribución del proceso definido por la ecuación (1.30), pero esto se abordará cuando se deduzca la verosimilitud del modelo en el capítulo 4.

Si el proceso se define en el intervalo $[0, T]$ y se considera una partición de N subintervalos de igual longitud $\Delta t = \frac{T}{N}$ y se define $t_0 = 0, t_i = i\Delta t, i = 1, 2, \dots, N$, se puede definir el incremento del proceso Poisson compuesto $\Pi(t; Q)$ como

$$\Delta\Pi(t) = \Pi(t + \Delta t) - \Pi(t) = \sum_{k=1}^{P(t+\Delta t)} h(Q_k) - \sum_{k=1}^{P(t)} h(Q_k) = \sum_{k=P(t)}^{P(t+\Delta t)} h(Q_k), \quad (1.31)$$

ó también

$$\Delta\Pi(t) = \sum_{k=1}^{\Delta P(t)} h(Q_k). \quad (1.32)$$

Cabe anotar que las fórmulas anteriores se pueden modificar para incluir la dependencia del proceso Poisson compuesto con el estado del proceso $X(t)$ y el tiempo incluyendo a estos como argumento de la función h . Como en el proceso de Wiener y el proceso Poisson homogéneo, se puede mostrar que los incrementos del proceso Poisson compuesto son independientes entre sí (si los intervalos de tiempo no se traslapan). De esta manera se puede escribir el valor del proceso $\Pi(t_i)$ como una suma de variables aleatorias independientes

$$\Pi(t_i) = \sum_{j=0}^i \Delta\Pi(t_j) = \sum_{k=1}^{P(t_i)} h(Q_k). \quad (1.33)$$

Esta propiedad, como en los casos anteriores, se utiliza para simular trayectorias del proceso Poisson compuesto. El algoritmo 8 presenta el pseudocódigo para simular una trayectoria de un proceso Poisson compuesto. La figura 1.3 muestra diferentes trayectorias de un proceso Poisson compuesto con $Q_k \sim SGED(0, 1.2, 0, 0)$ (definida en la siguiente sección), es decir que el proceso de marcaje tiene una distribución de error generalizada asimétrica para cada t , $h(Q_k) = Q_k$ y diferentes valores de Δt . Es importante aclarar que esta técnica para simular el proceso Poisson compuesto es válida si Δt es pequeño. En caso contrario, las trayectorias corresponden a simulaciones de un proceso Bernoulli compuesto. Una manera de verificar que la aproximación funciona razonablemente bien es, como en el proceso Poisson, verificar que los tiempos entre saltos sigan una distribución aproximadamente exponencial con media $\beta = 1/\lambda$, donde λ es la tasa de ocurrencia promedio de los saltos del proceso Poisson compuesto.

Algoritmo 8 Simulación de una trayectoria de un proceso Poisson compuesto.

- 1: $\Delta t := \frac{T}{n}, t_i := i \times \Delta t.$
 - 2: Obtenga n muestras $\Delta P(t_i)$ de una $Poisson(\lambda\Delta t)$.
 - 3: **para** $i = 1 : N$ **haga**
 - 4: Simule una muestra de tamaño $\Delta P(t_i)$ de la distribución de la variable subyacente Q .
 - 5: Haga $\Delta\Pi(t_i) = \sum_{k=1}^{\Delta P(t_i)} h(Q_k)$
 - 6: **fin para**
 - 7: Haga $\Pi(t_i) = \sum_{j=0}^i \Delta\Pi(t_j)$
-

1.3. Distribución de error generalizada asimétrica (SGED)

La distribución de error generalizada asimétrica (*Skewed Generalized Error Distribution*) SGED tiene diferentes versiones. Una de ellas es la propuesta por Theodossiou, en (Theodossiou, 2000), inspirada en la necesidad de modelar las características de los rendimientos financieros, específicamente asimetría y curtosis. Sin embargo, en este trabajo se utilizará una versión de la SGED

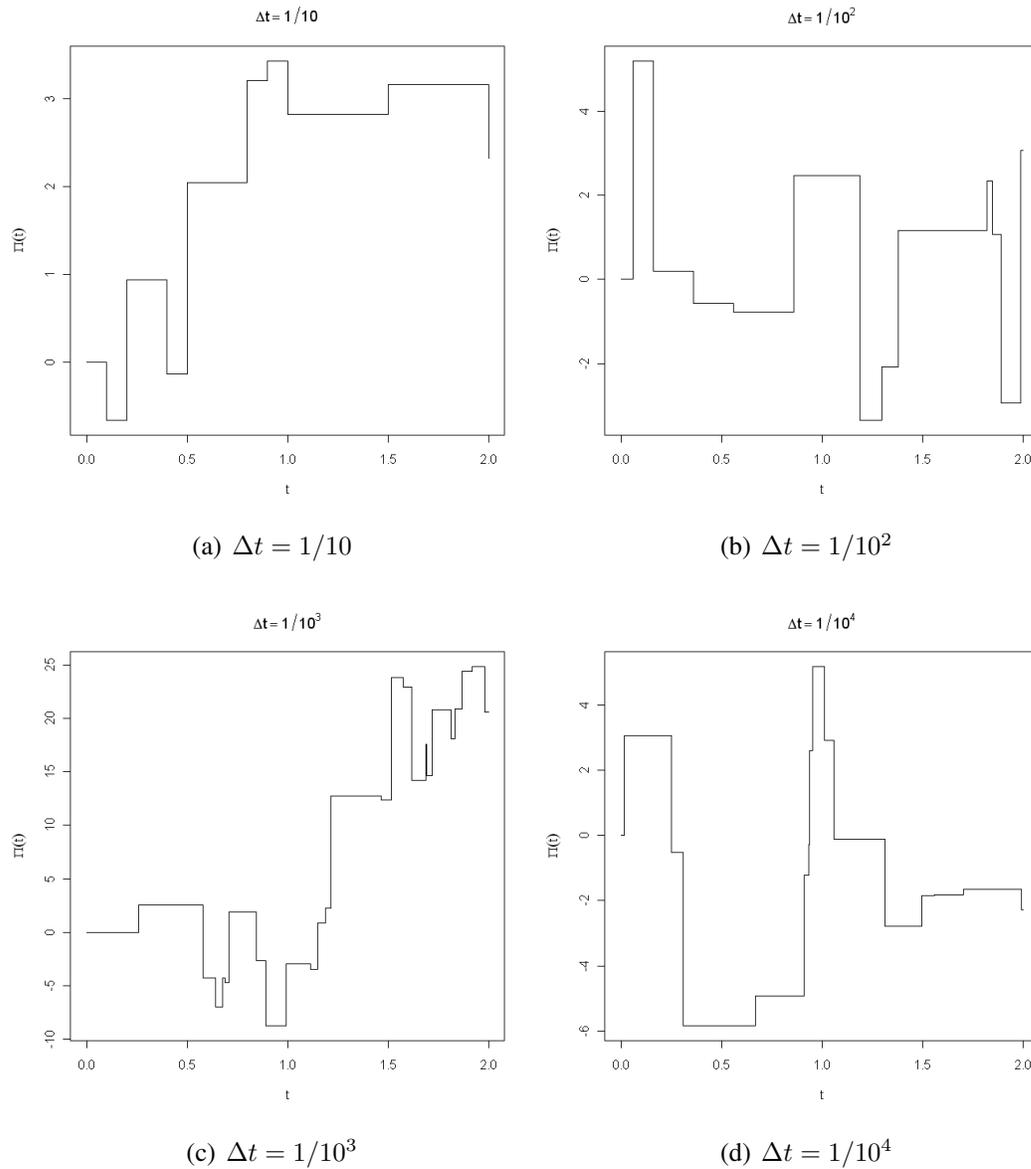


Figura 1.3: Trayectorias simuladas de un proceso Poisson compuesto con $\lambda = 8$, proceso subyacente iid SGED y diferentes valores para Δt y h la función identidad.

incluida dentro del paquete fGarch del software R, (Wuertz y Chalabi, 2008). Dicha versión se construye aplicando la técnica de Fernández y Steel, propuesta en (Fernandez y Steel, 1996), para generar densidades asimétricas a partir de densidades simétricas, usando transformaciones de escala. A continuación se ilustra la construcción de la versión de la SGED a la que se hará referencia posteriormente.

Considere la densidad de la distribución de error generalizada (GED) con parámetro ν , $GED(\nu)$, dada en la ecuación (1.34)

$$f(z|\nu) = \frac{\nu \exp\left[-\frac{1}{2} \left|\frac{z}{\lambda}\right|^\nu\right]}{\lambda 2^{(1+1/\nu)} \Gamma\left(\frac{1}{\nu}\right)}, \quad (1.34)$$

donde $z \in \mathbb{R}$, $\nu > 0$ y $\lambda = [2^{(-2/\nu)} \Gamma(1/\nu) / \Gamma(3/\nu)]^{1/2}$. El parámetro ν controla el peso de las colas. Se puede comprobar fácilmente que cuando $\nu = 2$ entonces la GED coincide con la distribución normal estándar.

Para cualquier valor del parámetro ν la densidad de la GED, $f(\cdot|\nu)$, es una función par y por lo tanto su esperanza es cero. También es sencillo ver que la varianza de una variable aleatoria GED es 1. De esta manera se tiene que si $Z \sim GED(\nu)$, entonces $E[Z|\nu] = 0$ y $Var[Z|\nu] = E[Z^2|\nu] = 1$. Así, la variable aleatoria Z es, por definición, una variable aleatoria estandarizada.

El primer paso para obtener una versión asimétrica de la GED, de acuerdo con Fernández y Steel, página 3, es el reescalamiento de la densidad GED introduciendo un parámetro de asimetría $\xi > 0$. De esta forma se obtiene la densidad asimétrica de la ecuación (1.35),

$$f(z|\nu, \xi) = \frac{2}{\xi + 1/\xi} f(\xi^{-\text{sign}(z)} z|\nu), \quad (1.35)$$

que satisface $f(z|\nu, 1) \equiv f(z|\nu)$. Además, $P(Z \geq 0, \xi) / P(Z < 0, \xi) = \xi^2$. Esta función de densidad se muestra en la ecuación 1.36,

$$f(z|\nu, \xi) = \frac{2}{\xi + \frac{1}{\xi}} \frac{\nu \exp\left[-\frac{1}{2} \left|\frac{\xi^{-\text{sign}(z)} z}{\lambda}\right|^\nu\right]}{\lambda 2^{(1+1/\nu)} \Gamma\left(\frac{1}{\nu}\right)}. \quad (1.36)$$

Los momentos de esta distribución se pueden obtener, de acuerdo con Fernández y Steel, página 4, con la ecuación (1.37),

$$E(Z^k|\nu, \xi) = M_k \times \left(\frac{\xi^{k+1} + (-1)^k / \xi^{k+1}}{\xi + 1/\xi} \right), \quad k = 1, 2, \dots, \quad (1.37)$$

donde $M_k = 2 \int_0^\infty s^k f(s) ds$. Adicionalmente, con la identidad de la ecuación (1.38),

$$\int_0^\infty x^r e^{-a^{-k} x^k} dx = \frac{1}{k} \Gamma\left(\frac{r+1}{k}\right) a^{r+1}, \quad (1.38)$$

se puede comprobar que

$$M_k = \Gamma((k+1)/\nu) \lambda^k 2^{k/\nu} / \Gamma(1/\nu). \quad (1.39)$$

Así, siguiendo a Fernández y Steel, se tiene que

$$E(Z|\nu, \xi) = (\xi - 1/\xi) M_1 = \mu_\xi$$

y

$$Var(Z|\nu, \xi) = (\xi^2 + 1/\xi^2) (M_2 - M_1^2) + 2M_1^2 - M_2 = \sigma_\xi^2.$$

Con lo anterior es posible ver que $E(Z^2|\nu, 1) = M_2 = 1$. El siguiente paso en la técnica de Fernández y Steel es estandarizar la densidad asimétrica. Para esto se utiliza una transformación que escala y traslada la densidad asimétrica. Dicha transformación se muestra en la ecuación (1.40),

$$g(z|\nu, \xi) = \sigma_\xi f(z_\xi|\nu, \xi), \quad (1.40)$$

donde $z_\xi = \sigma_\xi z + \mu_\xi$, y $\mu_\xi = E(Z|\nu, \xi)$ es la media con la densidad de la ecuación (1.35) y $\sigma_\xi^2 = Var(Z|\xi)$ es la varianza con esa densidad. Es decir que,

$$g(z|\nu, \xi) = \frac{2\sigma_\xi}{\xi + \frac{1}{\xi}} \frac{\nu \exp \left[-\frac{1}{2} \left| \frac{\xi^{-\text{sign}(\sigma_\xi z + \mu_\xi)} (\sigma_\xi z + \mu_\xi)}{\lambda} \right|^\nu \right]}{\lambda^{2(1+1/\nu)} \Gamma\left(\frac{1}{\nu}\right)}, \quad (1.41)$$

con

$$\mu_\xi = \left(\xi - \frac{1}{\xi} \right) \Gamma\left(\frac{2}{\nu}\right) \lambda^{2^{1/\nu}} / \Gamma\left(\frac{1}{\nu}\right)$$

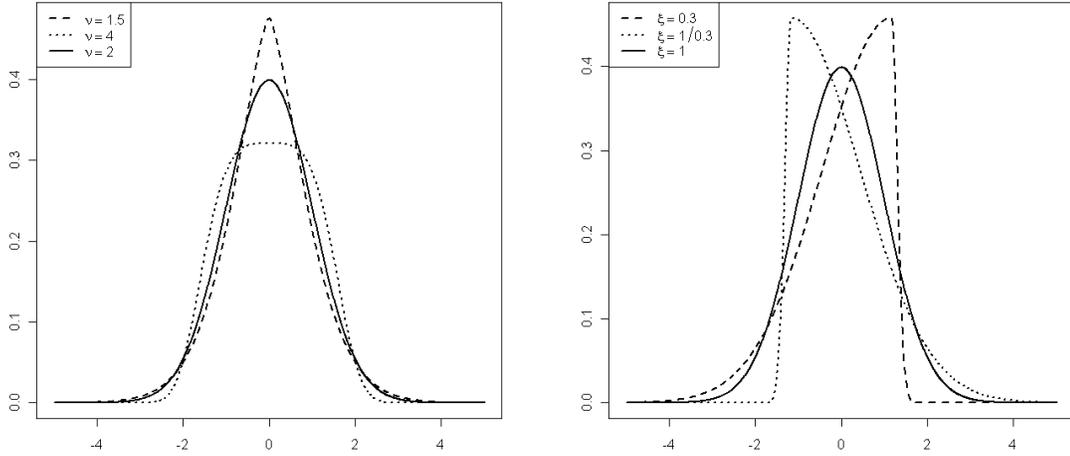
y

$$\sigma_\xi = \left(\xi^2 + \frac{1}{\xi^2} \right) \left(\frac{\Gamma\left(\frac{3}{\nu}\right) \lambda^{2^{2/\nu}}}{\Gamma\left(\frac{1}{\nu}\right)} - \left(\frac{\Gamma\left(\frac{2}{\nu}\right) \lambda^{2^{1/\nu}}}{\Gamma\left(\frac{1}{\nu}\right)} \right)^2 \right) + 2 \left(\frac{\Gamma\left(\frac{2}{\nu}\right) \lambda^{2^{1/\nu}}}{\Gamma\left(\frac{1}{\nu}\right)} \right)^2 - \frac{\Gamma\left(\frac{3}{\nu}\right) \lambda^{2^{2/\nu}}}{\Gamma\left(\frac{1}{\nu}\right)}.$$

Finalmente, se define la SGED para variables aleatorias con media y varianza diferentes de cero y uno, respectivamente. Se dice que $X \sim SGED(\mu, \sigma, \nu, \xi)$ si la densidad de $Z = \frac{X-\mu}{\sigma}$ está dada por la ecuación 1.41.

Por la forma en que se llega a la densidad de la ecuación 1.41 se tiene que $\mu = E[X]$ y $\sigma^2 = Var[X]$. De esta manera se completa la definición de la SGED asimétrica. Cabe anotar que cuando $\nu = 2$ y $\xi = 1$ la SGED se convierte en una densidad normal y cuando $\nu = 1$ y $\xi = 1$ la SGED se convierte en la distribución de Laplace.

La figura 1.4 presenta algunas gráficas de la función de densidad SGED para distintos valores de sus parámetros. El propósito de dichas gráficas es identificar el efecto de los parámetros ν y ξ , en particular como influyen estos en la desviación del comportamiento normal de la distribución SGED. En la gráfica de la izquierda se presenta la densidad GED (aquí $\xi = 1$) y se puede apreciar que valores de ν inferiores a 2 aumentan la curtosis de la densidad mientras que valores mayores que 2 tornan la densidad platocúrtica. En la gráfica de la derecha se aprecia el efecto del parámetro de asimetría ξ . Una propiedad de las distribuciones asimétricas obtenidas a partir de la técnica de Fernández y Steel es que $f(z|\nu, \xi) = f(-z|\nu, 1/\xi)$, como se observa en la gráfica.



(a) Efecto del parámetro ν . $\xi = 1$.

(b) Efecto del parámetro ξ . $\nu = 2$.

Figura 1.4: Función de densidad SGED para diferentes valores de los parámetros ν y ξ .

1.4. Solución del modelo de difusión con saltos

En esta sección se presenta la solución del modelo de difusión con saltos. Dicha solución se obtiene a partir del cálculo estocástico de Itô, (Itô, 1951). En este trabajo no se abordan todos los aspectos técnicos formales, sino que se sigue un desarrollo intuitivo que permite ilustrar cómo se llega a la solución del modelo. Antes de esto, conviene resaltar la aplicación del modelo de difusión con saltos en diferentes contextos. Uno, en el que ha venido ganando gran aceptación, es el modelamiento de precios de energía en mercados no regulados. Prueba de esto son los trabajos de (Huisman y Ronald, 2001), (Weron, Bierbrauer, y Trück, 2004), (Borovkova y Permana, 2006), (Cartea y Figueroa, 2005) y (Weron, 2006), entre otros. También se ha utilizado para modelar el crecimiento económico, como se muestra en (Posch, 2007). En (Fernandes, 2006) utilizan un modelo de difusión con saltos para pronosticar crisis financieras y en (Srivastava, Grenander, Jensen, y Miller, 2002) se usa un modelo de difusión con saltos para el reconocimiento de objetos en imágenes en sistemas de procesamiento digital de imágenes. Como se puede observar, el modelo de difusión con saltos tiene una gama de aplicaciones diversa y, por tanto, los esfuerzos por diseñar metodologías que permitan la estimación de sus parámetros pueden repercutir en avances en otras técnicas y disciplinas.

En (Hanson, 2006) se explica claramente el cálculo de derivadas e integrales que involucran los procesos de difusión y de Poisson y se muestra el desarrollo matemático para llegar a la solución del modelo de difusión con saltos. Antes de presentar directamente la solución del modelo se establecerá el contexto en el que ésta se obtiene.

Suponga que se tiene el proceso estocástico $X(t)$ y que f es una función tal que

$$f : \mathbb{R} \times [0, +\infty] \rightarrow \mathbb{R}.$$

Además, considere la partición del intervalo $[0, t]$ en $n + 1$ subintervalos de longitud $\Delta t = \frac{t}{n}$ y $0 = t_0 < t_1 < \dots < t_{n+1} = t$, con $t_i = i \times \Delta t$. Así, $\Delta t_i = t_{i+1} - t_i = \Delta t$. Esta es una partición regular y su norma es $\delta t_n = \max_{0 \leq i \leq n} (\Delta t_i) = \Delta t$. Adicionalmente, $\Delta X(t_i) = X(t_{i+1}) - X(t_i)$. Así, si se aproxima la integral de f utilizando la regla de Euler (o rectangular izquierda) se obtiene

$$\int_0^t f(X(s), s) dX(s) = \lim_{n \rightarrow +\infty}^{ms} \sum_{k=0}^n f(X(t_k), t_k) \Delta X(t_k). \quad (1.42)$$

Aquí, $\lim_{n \rightarrow +\infty}^{ms}$ representa el límite en media cuadrática y se entiende de la siguiente manera: si $\{Z_n\}$ es una sucesión de variables aleatorias, esta sucesión converge en media cuadrática a la variable aleatoria Z , y se escribe $\lim_{n \rightarrow +\infty}^{ms} Z_n = Z$, si se cumple que

$$\lim_{n \rightarrow +\infty} E[(Z_n - Z)^2] = 0. \quad (1.43)$$

Lo anterior tiene el propósito exclusivo de establecer el contexto en el que la solución de la ecuación (1.3) se obtiene. En (Hanson, 2006) también se presenta la solución general del modelo de difusión con saltos de un solo estado, cuyos parámetros y su estimación son el objeto principal de este trabajo. El modelo a considerar de ahora en adelante es el que se muestra en la ecuación 1.44,

$$\frac{dX(t)}{X(t)} = \mu dt + \sigma dW(t) + h(t, Q(t)) dP(t), \text{ con } X(0) = x_0 \text{ con probabilidad 1.} \quad (1.44)$$

Aquí, $\mu, \sigma \in \mathbb{R}$ y $Q(t)$ es proceso estocástico tal que $\ln(1 + Q(t)) > 1$. Esta última condición se impone para garantizar que el proceso sólo tome valores positivos. En el modelo de la ecuación (1.44) los procesos de difusión y de Poisson no dependen del estado del proceso $X(t)$ y son, además, independientes entre sí.

Si para el proceso en tiempo continuo dado en la ecuación (1.44) se define la transformación $Y(t) = F(X(t), t) = \ln X(t)$, entonces, usando el lema de Íto, se llega a una expresión para el diferencial $dY(t)$ como la que se muestra en la ecuación (1.45)

$$\begin{aligned} dY(t) = & \left(F_t + \mu X(t) F_x + \frac{1}{2} (\sigma X(t))^2 F_{xx} \right) dt + \sigma X(t) F_x dW(t) \\ & + \int [F(X(t) + X(t)h(t, Q(t)), t) - F(X(t), t)] dP(t), \end{aligned} \quad (1.45)$$

donde, para el caso particular, $F_t = \frac{\partial F}{\partial t} = 0$, $F_x = \frac{\partial F}{\partial x} = \frac{1}{x}$ y $F_{xx} = \frac{\partial^2 F}{\partial x^2} = -\frac{1}{x^2}$. Además, el integrando de la parte que corresponde a los saltos se puede escribir como en la ecuación (1.46)

$$\begin{aligned} F(X(t) + X(t)h(t, Q(t)), t) - F(X(t), t) &= \ln(X(t) + X(t)h(t, Q(t))) - \ln(X(t)) \\ &= \ln(1 + h(t, Q(t))). \end{aligned} \quad (1.46)$$

De esta manera, la ecuación (1.45) puede reescribirse como en la ecuación (1.47),

$$dY(t) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW(t) + \int \ln(1 + h(t, Q)) dP(t). \quad (1.47)$$

Así, la solución de la ecuación diferencial estocástica (1.46) se obtiene como en la ecuación (1.48)

$$Y(t) = Y(0) + \int_0^t \left(\left(\mu - \frac{\sigma^2}{2}\right) ds + \sigma dW(s) + \ln(1 + h(s, Q(s))) dP(s) \right), \quad (1.48)$$

que se puede expresar como en la ecuación (1.49)

$$Y(t) = \ln(X(0)) + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W(t) + \sum_{k=1}^{P(t)} \ln(1 + h(t, Q_k)). \quad (1.49)$$

Ahora, como $X(t) = \exp(Y(t))$ la solución general del modelo de la ecuación (1.44) está dada por la ecuación (1.50)

$$X(t) = x_0 \exp \left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W(t) + \sum_{k=1}^{P(t)} \ln(1 + h(t, Q(T_k))) \right), \quad (1.50)$$

o también

$$X(t) = x_0 \exp \left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W(t) \right) \prod_{k=1}^{P(t)} (1 + h(t, Q(T_k))). \quad (1.51)$$

Las ecuaciones (1.50) y (1.51) corresponden a la solución de la ecuación 1.44. Queda por especificar la función h . Ésta se define como en la ecuación (1.52)

$$h(t, Q(t)) = e^{Q(t)} - 1. \quad (1.52)$$

Esta definición de h tiene la ventaja de hacer que el proceso $X(t)$ dependa explícitamente del proceso $Q(t)$. De esta manera se puede escribir la solución de la ecuación (1.44) como en la ecuación (1.53)

$$X(t) = x_0 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t) + \sum_{k=1}^{P(t)} Q_k \right). \quad (1.53)$$

En este trabajo se considera que $Q(t)$ es una sucesión iid de variables que siguen una distribución de error generalizada asimétrica (SGED) con parámetros μ_Q, σ_Q, ν_Q y ξ_Q . Note que el exponente del lado derecho de la ecuación (1.53) es una suma de variables aleatorias cuyos incrementos son independientes (las variables en si mismas también lo son). Por esta razón resulta útil proceder y operar sobre el exponente. Así, si se define el proceso $Y(t)$ como en la ecuación (1.54),

$$Y(t) = \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t) + \sum_{k=1}^{P(t)} Q_k, \quad (1.54)$$

entonces

$$X(t) = x_0 e^{Y(t)}. \quad (1.55)$$

Esta forma de escribir el proceso de difusión con saltos presenta la ventaja de verlo como una función de un proceso con incrementos iid, ya que el proceso $Y(t)$ es tal. Para ver que los incrementos del proceso $Y(t)$ son independientes considere $\Delta Y(t) = Y(t + \Delta t) - Y(t)$. Este satisface la ecuación (1.56),

$$\Delta Y(t) = \left(\mu - \frac{\sigma^2}{2} \right) \Delta t + \sigma \Delta W(t) + \sum_{k=1}^{\Delta P(t)} Q_k. \quad (1.56)$$

El proceso $\Delta Y(t)$ es el proceso de rendimientos logarítmicos de $X(t)$ en el periodo Δt . Es decir,

$$\Delta Y(t) = \ln X(t + \Delta t) - \ln X(t) = \ln \left(\frac{X(t + \Delta t)}{X(t)} \right). \quad (1.57)$$

Una presentación alternativa del proceso $\Delta Y(t)$, como la que se hace en (Craine, Lochstoer, y Syrtveit, 2000) y (Ranjan Das y Sundaram, 1999), se presenta en la ecuación (1.58), donde $Z \sim N \left(\left(\mu - \frac{\sigma^2}{2} \right) \Delta t, \Delta t \sigma^2 \right)$ y las Q_i son las variables aleatorias que definen la magnitud de los saltos,

$$\Delta Y(t) = \begin{cases} Z & \text{si } \Delta P(t) = 0 \\ Z + Q_1 + \dots + Q_{\Delta P(t)} & \text{en otro caso.} \end{cases} \quad (1.58)$$

De aquí surge una estrategia para simular trayectorias del proceso de difusión con saltos definido por la ecuación (1.53), la cual consiste en simular el exponente, es decir el proceso $Y(t)$ definido por la ecuación (1.54) y luego encontrar la trayectoria del proceso como se deduce de la ecuación (1.55). De la misma forma en que se hizo con los procesos de Wiener y de Poisson, si se asume que

el proceso de la ecuación (1.54) está definido en el intervalo $[0, T]$ y se considera una partición en N subintervalos de igual longitud $\Delta t = \frac{T}{N}$ y se define $t_0 = 0, t_i = i\Delta t, i = 1, 2, \dots, N$, entonces se puede escribir $Y(t_i)$ como en la ecuación (1.59),

$$Y(t_i) = \sum_{j=0}^i \Delta Y(t_j). \quad (1.59)$$

El algoritmo 9 muestra el esquema general para la simulación del proceso definido por las ecuación (1.53) y (1.58). Con este esquema se pueden construir realizaciones simuladas del proceso como las que se muestran en la figura 1.5. Estas trayectorias simuladas corresponden a un proceso de difusión con saltos en la medida en que $\Delta t \rightarrow 0$, es decir, mientras más fina sea la partición del intervalo. Si Δt no es lo suficientemente pequeño, entonces es posible que para ciertos valores de λ (la tasa de ocurrencia de saltos del proceso Poisson) la aproximación Poisson-Bernoulli utilizada para simular trayectorias del proceso Poisson no sea adecuada, y en lugar de tener un proceso Poisson compuesto se tiene un proceso Bernoulli compuesto. Lo anterior quiere decir que mientras más pequeño sea Δt , más realistas son las trayectorias simuladas con el algoritmo 9. En la figura 1.5, por ejemplo, las trayectorias (a) y (b) no son muy realistas. Sin embargo, puede pensarse de dichas trayectorias que corresponden a un proceso de difusión con saltos muestreado con periodo Δt . Esto es lo que se hace en la práctica. Se asume que el precio de cierto activo financiero en el mercado bursátil se puede modelar con un proceso de difusión con saltos. Sin embargo, no se dispone de la evolución del precio en forma continua sino de mediciones en tiempo discreto con diferentes tipos de resolución (horaria, diaria, mensual, etc.).

Algoritmo 9 Simulación de una trayectoria del proceso de difusión con saltos de las ecuaciones (1.53) y (1.58).

- 1: $\Delta t := \frac{T}{N}, t_i := i \times \Delta t.$
 - 2: Obtenga N muestras $\Delta D(t_i)$ de una $N \left(\left(\mu - \frac{\sigma^2}{2} \right) \Delta t, \Delta t \sigma^2 \right).$
 - 3: Obtenga N muestras $\Delta P(t_i)$ de una *Poisson* $(\lambda \Delta t)$
 - 4: **para** $i = 1 : n$ **haga**
 - 5: Simule una muestra de tamaño $\Delta P(t_i)$ de la distribución subyacente, f_Q , del proceso de marcas $Q(t)$.
 - 6: Haga $\Delta J(t_i) = \sum_{k=1}^{\Delta P(t_i)} Q_k$
 - 7: **fin para**
 - 8: Haga $Y(t_i) = \sum_{j=0}^i \Delta D(t_j) + \Delta J(t_j)$
-

De esta manera queda cubierta la teoría básica acerca del proceso de difusión con saltos. Se ha presentado una expresión para los rendimientos del proceso de difusión con saltos y algunas de sus características, que se usarán en el capítulo 4 para construir la metodología de estimación.

En (Hanson, 2006) se explica con más detalle los aspectos técnicos acerca de la integración de los procesos de Wiener y de Poisson, se considera una gama más amplia de modelos de difusión con saltos y se abordan aplicaciones para ciencias e ingenierías del modelo.

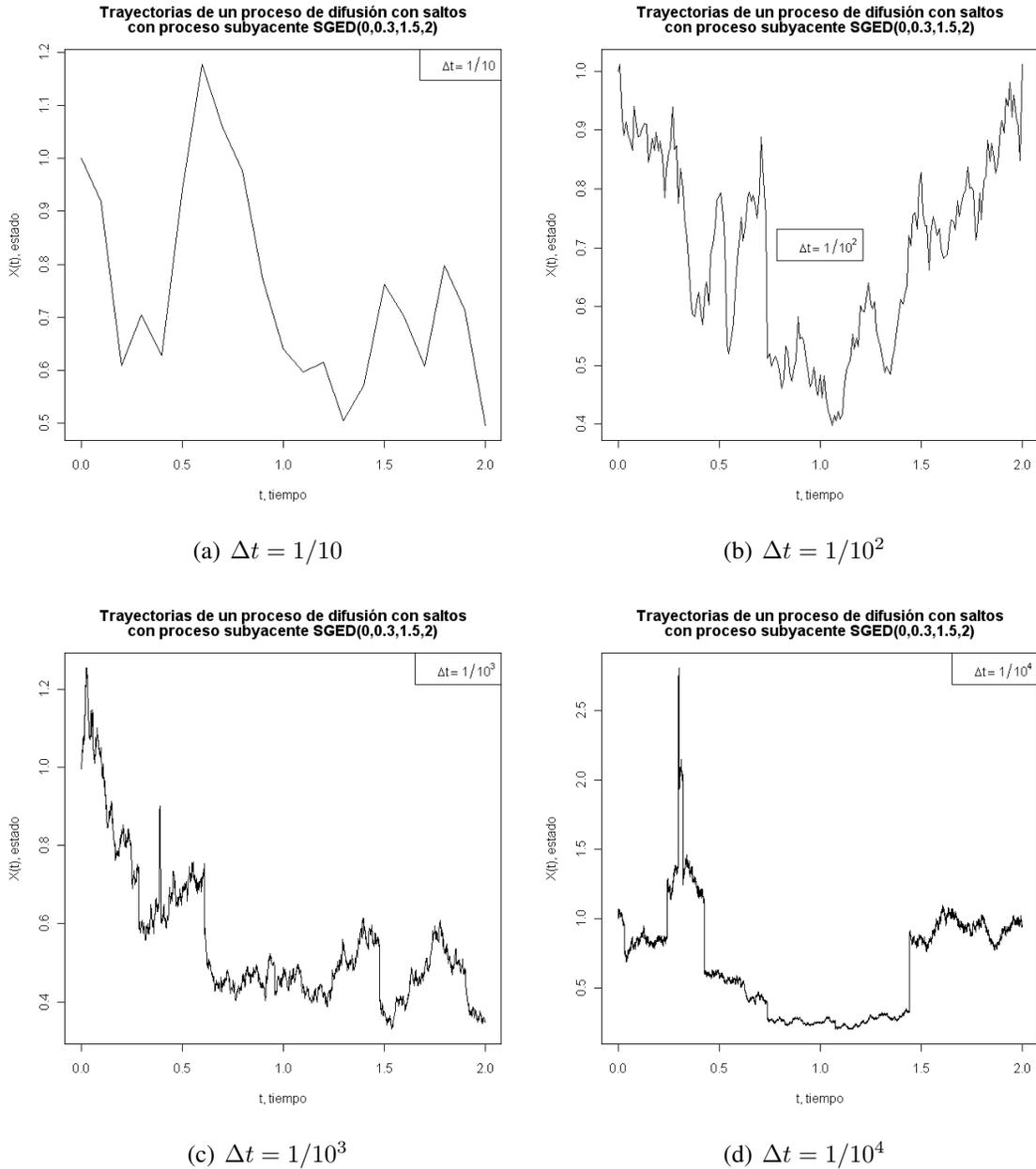


Figura 1.5: Trayectorias simuladas de un proceso de difusión con saltos $\mu = 0.25$, $\sigma = 0.5$, $\lambda = 4$, proceso subyacente a los saltos $Q_k \sim SGED(0, 0.3, 1.5, 2)$ y diferentes valores para Δt .

Capítulo 2

Revisión de los Métodos de Estimación de Modelos de Difusión con Saltos

En este capítulo se presenta una revisión de la literatura relacionada con la estimación de procesos de difusión con saltos. En la primera parte se presenta los métodos basados en método de máxima verosimilitud. Posteriormente se menciona los métodos que hacen uso de la función característica y por último los métodos basados en momentos.

2.1. Estimación por máxima verosimilitud

El modelo de difusión con saltos, aunque ha sido aplicado fuera de la econometría, es en esta área principalmente donde surge y se desarrolla. Son muchas las adaptaciones que se han hecho sobre el modelo inicial propuesto por Merton (Merton, 1976) para tratar de modelar varios de las características propias de las series financieras. El enfoque de estimación por máxima verosimilitud fue el primero en aplicarse a la estimación de procesos de difusión con saltos. Para ello se concentró el trabajo en derivar la función de densidad incondicional del proceso dado por el logaritmo del precio de los activos, es decir la densidad incondicional de los rendimientos logarítmicos. Un trabajo pionero en esto es (Beckers, 1981). En este trabajo se presentó la densidad del proceso como una mezcla Poisson de densidades Gaussianas. Una metodología, explicada con cierto grado de detalle, con la cual se puede derivar la función de densidad de probabilidad de los rendimientos del proceso se presenta en (Craine, Lochstoer, y Syrtveit, 2000) y en (Planchet y Therond, 2005). El problema que surge en estas situaciones es que la densidad del proceso de rendimientos logarítmicos se convierte en una suma infinita de densidades Gaussianas ponderadas por los términos de una distribución Poisson. En (Ball y Torous, 1983), Ball y Torous, presentan una versión simplificada del modelo asumiendo que si la tasa λ , del proceso Poisson, es pequeña entonces en un periodo suficientemente corto de tiempo puede darse a lo más un salto. A esta simplificación del modelo se le conoce como el modelo de difusión con saltos Bernoulli (*Bernoulli jump-diffusion*) y permi-

tió truncar la densidad del proceso en el segundo término de la serie (el primero corresponde a la probabilidad de que no haya salto y el segundo a la probabilidad de que sí ocurra el salto). Los mismos autores, en (Ball y Torous, 1985), presentan una serie de argumentos que permiten truncar la mezcla Poisson de densidades Gaussianas en el décimo término. Para ello calculan el error de truncamiento de una serie y encuentran que, en general, para valores realistas de los parámetros (es decir, aquellos que pueden asimilarse a un proceso observado en la realidad) truncar en el décimo término no representa un problema. Sobre los tres trabajos mencionados anteriormente se construye la metodología de estimación empleada en este trabajo, con dos diferencias importantes: se asume que los saltos siguen una distribución SGED y se usa la transformada rápida de Fourier (FFT).

En (Posch, 2007) se utiliza la función de verosimilitud del proceso y se aplica simulación de Monte Carlo para demostrar que los parámetros del modelo pueden recuperarse de las trayectorias muestrales. En dicho trabajo, Posch, en un contexto completamente econométrico en el que se estudia las tasas de crecimiento, utiliza tanto métodos exactos como aproximados basados en imponer restricciones a los parámetros.

Los retos al aplicar el método de estimación de máxima verosimilitud surgen cuando se considera modificaciones del modelo de difusión con saltos. Estas modificaciones pueden agruparse en tres clases:

1. Procesos de saltos con distribuciones más complejas que la lognormal (esta densidad es la que permite llegar a la densidad incondicional del proceso como una mezcla Poisson de densidades Gaussianas), con el fin de introducir asimetría en el modelo de difusión con saltos.
2. Estructuras del proceso de difusión con saltos más complejas que la propuesta por Merton.
3. La combinación de las dos anteriores: estructuras complejas del proceso y distribuciones no Gaussianas (y posiblemente que no pertenecen a la familia exponencial) para el proceso Poisson compuesto que suministra la característica de los saltos.

En el primer grupo se tiene por ejemplo el trabajo de Ramezani, donde en (Ramezani y Zeng, 1998) se introdujo una variación al modelo inicial de (Merton, 1976), conocida como el modelo de difusión con saltos Pareto-Beta (*Pareto-Beta jump-diffusion PBJD*) introduciendo dos procesos Poisson independientes y asociando una distribución Pareto a los saltos positivos y una distribución Beta a los saltos negativos, con el fin de reproducir el fenómeno de asimetría de los rendimientos de los precios de acciones. El autor deriva la densidad incondicional del proceso obteniendo una suma de mezclas Poisson de densidades Gaussianas, Pareto y Beta, y estima el proceso usando dos métodos diferentes: el método de los cumulantes (*cumulants*) y el método de máxima verosimilitud, aunque advierte que se puede utilizar otros métodos. Sobre los el uso de los cumulantes, señala el autor, muchas veces pueden no existir o tener el signo equivocado. Respecto al método de máxima verosimilitud, el autor indica que de acuerdo con (Kiefer, 1978) no existe un óptimo global en la log-verosimilitud para una mezcla de densidades. La solución a este problema, de acuerdo con

la última referencia, es acotar adecuadamente el espacio de búsqueda de los parámetros y se ha notado que los métodos de descenso por gradiente del tipo Newton-Rhaphson son efectivos, siempre y cuando se suministre las condiciones iniciales cercanas a los valores reales. Este problema es quizá el más importante que se presenta al utilizar máxima verosimilitud y establece dos condiciones que el método de optimización que se aplica a la log-verosimilitud debe cumplir: necesita incorporar restricciones y proporcionar valores iniciales razonables.

En la misma dirección Ramezani y Zeng en (Ramezani y Zeng, 2007) desarrollan la metodología para estimar vía máxima verosimilitud el modelo de difusión con saltos y distribución de saltos doble exponencial (Double Exponential Jump-Diffusion DEJD), propuesto por Kou en (Kou, 2002). Este modelo propone para los saltos una distribución del tipo doble exponencial. Esto puede pensarse como una distribución exponencial de ciertos parámetros para los saltos positivos y otra con parámetros diferentes para los negativos.

En cuanto a la estimación de versiones multivariadas del modelo de difusión con saltos, Shimizu y Yoshida, en (Shimizu y Yoshida, 2006) presentan un método para estimar por máxima verosimilitud el proceso, para lo cual separan las observaciones de los rendimientos logarítmicos en dos clases mediante un filtro: observaciones correspondientes a un movimiento browniano y observaciones debidas a los saltos. Esta separación da pie a una función de contraste que al ser minimizada permite hallar unos estimadores de los parámetros que son asintóticamente consistentes. Otro trabajo interesante en esta dirección se presenta en (Yu, 2007), donde Yu da una expresión cerrada para aproximar la verosimilitud del proceso de difusión con saltos multivariado y así estimar los parámetros del modelo.

2.2. Estimación usando la función característica

Cuando se considera versiones del proceso de difusión con saltos alternativas a la propuesta por Merton se hace necesario emplear otros métodos de estimación. Un ejemplo de esto es la metodología de estimación presentada en (Moreno, Serrano, y Stute, 2008) para estimar el modelo propuesto por Altman et. al en (Altmann, Schmidt, y Stute, 2004). Esta metodología considera un modelo denominado de impacto del ruido (*shot-noise jump-diffusion SNJD*), que busca caracterizar el impacto de un salto en el largo plazo. La metodología de estimación que se propuso para estimar este modelo se basa en la obtención de la función característica del proceso de rendimientos logarítmicos y su densidad espectral. Los parámetros se obtienen al minimizar alguna distancia espectral entre la densidad espectral de los datos observados y la del modelo. Otra técnica de estimación para el modelo SVJD y su extensión, el modelo SVJJD (*stochastic volatility with jumps and jump-diffusion*) que utiliza la función característica, se presenta en (Rockinger y Semenova, 2005). En este trabajo, Rockinger et. al. indican que los procedimientos basados en la función característica pueden dividirse en tres clases:

1. De máxima verosimilitud: utilizando la transformada inversa de Fourier de la función característica del proceso se llega a la función de densidad. Luego mediante un procedimiento de optimización se obtiene los estimadores. El problema que reportan los autores para aplicar este método es el alto costo computacional.
2. GMM (*Generalized Method of Moments*): diferenciando la función característica e igualando sus derivadas a cero se obtiene los momentos del proceso en función de los parámetros, que posteriormente se pueden igualar con los momentos observados para encontrar los estimadores. Sin embargo, si el vector de parámetros es de un orden alto, el método GMM presenta problemas de inestabilidad numérica. En (Chacko y Viceira, 2003) se discute el diseño de estimadores GMM espectrales a partir de la función característica empírica.
3. La función característica empírica: se deriva la función característica del proceso en función de los parámetros del modelo y se estima los parámetros de tal manera que se minimice una distancia entre la función característica empírica y la función característica del proceso. Este método tiene el mismo problema del método GMM, inestabilidad numérica, no obstante puede aplicarse para variables que no son iid. Por esta razón los autores lo escogen sobre los demás y lo aplican en su trabajo concluyendo que aporta un buen compromiso entre eficiencia y costo computacional.

2.3. Estimación por GMM, SMM y EMM

Ho et. al. en (Ho, Perraudin, y Sorensen, 1996), consideran una versión del proceso de difusión con saltos cuyo término de la volatilidad sigue a su vez un proceso estocástico autorregresivo. Este modelo se conoce en la literatura como SVJD por sus siglas en Inglés (*Stochastic Volatility Jump-Diffusion*) y para estimarlo el autor propone usar el método de momentos generalizado (GMM). Posteriormente, Craine et. al. en (Craine, Lochstoer, y Syrtveit, 2000) proponen otro modelo de la clase SVJD y desarrollan una metodología para estimarlo utilizando el método de momentos simulados (SMM, por sus siglas en inglés) de McFadden y Pakes y Pollard propuesto en (McFadden, 1989) y en (Pakes y Pollard, 1989). Craine et. al. también aplican el método de momentos eficiente (EMM) descrito en (Gallant y Tauchen, 1996) y (Gourieroux y Montfort, 1996). Una conclusión interesante de este trabajo, obtenida a través de un estudio de Montecarlo, es que los métodos basados en momentos no son muy efectivos cuando se dispone de 2000 observaciones o menos, para el proceso SVJD.

En (Bollerslev y Zhou, 2002) se plantea el uso del método GMM explotando la información de alta frecuencia (i.e. series de medición diaria), proponiendo para ello la solución analítica de los dos primeros momentos condicionales para la volatilidad, que está planteada como una variable latente en este modelo, que es del tipo SVJD. Algo similar a lo anterior se presenta en (Jiang y Oomen, 2007), donde se estima un modelo del tipo de difusión con saltos afín (*affine jump diffusion*) en

el cual hay una dependencia de la deriva del proceso y de la volatilidad con el estado actual del proceso. Es notable en este trabajo el desarrollo de estimadores insesgados para los parámetros del modelo. También con el método GMM se presenta en (Arnold, 2006) la estimación de un modelo del tipo SVJD. En esta referencia se explota toda la información que brinda el mercado utilizando la información de todos los precios de las opciones de dicho mercado.

Capítulo 3

Evolución Artificial Aplicada a la Optimización

3.1. Introducción

En ciencias e ingeniería es común encontrarse con el problema de hallar el estado óptimo de una variable o parámetros de un sistema¹. Para ello, la práctica más común es definir una función de costo que refleja que tan adecuado es un conjunto de parámetros o estados. Asumiendo que la función de costo f es de valor real y que los parámetros o estados óptimos están contenidos en una región $D \subseteq \mathbb{R}^k$ denominada el espacio de búsqueda, el problema se define formalmente como encontrar el vector $\hat{\mathbf{x}} = (x_1, \dots, x_k)'$ que satisface la ecuación (3.1), donde por $'$ se denota la transpuesta de una matriz,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in D} f(\mathbf{x}). \quad (3.1)$$

En estadística, este problema es bastante común, pues una gran cantidad de estimadores están planteados a partir de la optimización de una función objetivo. Ejemplos de esto son el método de estimación de máxima verosimilitud y los métodos basados en momentos. En algunos casos, como por ejemplo el Método de Momentos Eficiente (EMM, por sus siglas en Inglés), la función objetivo no tiene una expresión analítica cerrada y en cambio debe evaluarse a través de iteraciones de un algoritmo. Cuando no es fácil obtener el gradiente, o el conocimiento *a priori* de la solución es pobre, o más aún, cuando la función f no tiene una expresión matemática, se debe emplear métodos heurísticos para obtener soluciones. En este trabajo se hace uso de los algoritmos evolutivos para optimizar la log-verosimilitud y encontrar los parámetros del modelo de difusión con saltos que se considera. Sobre esta log-verosimilitud ya se ha reportado en la literatura el problema de multimodalidad, (Craine, Lochstoer, y Syrtveit, 2000), (Fernandes, 2006) y (Ramezani y Zeng, 2007),

¹Parte de esta sección se ha tomado textualmente de (Hernández y Ospina, 2009), texto del cual el autor del presente trabajo es también coautor.

entre otros. También, para ciertos valores de los rendimientos en (Ball y Torous, 1983, 1985) se reporta que la log-verosimilitud asume valores infinitos, tal y como se advierte en (Kiefer, 1978). Estos problemas justifican la exploración de otras técnicas de optimización para la estimación del proceso de difusión con saltos, que es uno de los objetivos principales de este trabajo.

3.2. Métodos heurísticos de optimización

Según Gilli y Winker, (Gilli y Winker, 2007), los métodos de optimización heurística se pueden clasificar en dos clases: los métodos constructivos y los métodos de búsqueda local. Estos últimos son los de interés en este trabajo. Son similares a los métodos de optimización para funciones continuas denominados de descenso por gradiente (*gradient descent*), en el cual la solución en la iteración n , x_n satisface una ecuación recursiva $x_{n+1} = x_n + \alpha_n \nabla g(x_n)$ donde $g(x)$ es la función de valor real a minimizar y $\nabla g(x)$ es su gradiente. Los métodos de descenso por gradiente definen α_n de tal forma que siguen la dirección de descenso más rápido hacia un mínimo global. Entre los algoritmos de descenso por gradiente se incluyen: BFGS (por los apellidos de sus creadores: Broyden, Fletcher, Goldfarb y Shanno), DFP (debido a Davidon, Fletcher y Powell), Newton, Newton-Gauss, Steepest Descent, PRCG (*Polak-Ribiere-type conjugate gradient*), y BHHH (Berndt-Hall-Hall-Hausman), ver Nocedal y Wright, (Nocedal y Wright, 1999). Cada uno define α_n de manera diferente, pero con base en la información sobre el gradiente y la matriz hessiana. Son algoritmos deterministas, sensibles a la inicialización y pueden quedar atrapados en mínimos locales ó aún, terminar sin converger cuando llegan a una región tipo meseta (*plateau*). En contraste con esto, los métodos heurísticos de búsqueda local definen una vecindad de x_n , $B_\delta(x_n)$ y escogen x_{n+1} en esa vecindad, usualmente de manera aleatoria según un procedimiento algorítmico. Además, definen criterios para aceptar el punto escogido como posible mejor solución y cuándo terminar el proceso de iteración.

Los métodos heurísticos de búsqueda local se dividen a su vez en métodos de trayectorias y métodos multipoblacionales, entre los que se cuenta los denominados métodos de evolución artificial, en los cuales no se genera un solo candidato x_{n+1} sino todo un vector (población) de posibles candidatos, $(x_n^{(1)}, \dots, x_n^{(k)})$, sobre los cuales se busca simultáneamente la nueva solución. La ventaja de estos métodos multipoblacionales es que se afectan en menor grado por topologías adversas de la función objetivo tales como: múltiple mínimos locales, rugosidades, mesetas planas (*plateau*) y no convexidad. Una característica particular de los métodos multipoblacionales es que estos pueden generar penalizaciones a tales características indeseables asignando valores muy grandes a la función objetivo $g(x)$ en un punto x que pertenezca a una región no deseable.

Sin embargo, usar este tipo de algoritmos tiene un precio. Los algoritmos convencionales son determinísticos, y dadas condiciones sobre el espacio parametral y la función objetivo, convergen al óptimo global. Los algoritmos heurísticos son, en cambio, probabilísticos y su convergencia es en probabilidad o en otro modo de convergencia. No siempre convergerán al óptimo global, pero

en varias repeticiones se espera que el promedio se acerque al óptimo global. Este es, junto con el alto costo computacional, el precio que se paga al usar algoritmos heurísticos.

Entre los de la primera clase están el temple simulado (*simulated annealing*) y aceptación con umbral (*threshold accepting*). Entre los de la segunda clase están los algoritmos genéticos, evolución diferencial (*differential evolution*) y colonias de hormigas (*ants colonies*), ver (Gilli y Winker, 2007) y (Spall, 2003), para una descripción y referencias adicionales.

Una estrategia utilizada en varios programas consiste en algoritmos híbridos, por ejemplo, BFGS con Nelder-Mead, utilizado en R-Metrics para estimación de modelos tipo GARCH, (Wuertz, Chalabi, y Luksan, 2004), o algoritmos genéticos con restricciones y el método SQP (Sequential Quadratic Programming), utilizado en Matlab. En el paquete *tsm* se usa BFGS con la posibilidad de generar los valores iniciales usando algoritmos evolutivos. La figura 3.1 presenta un esquema general de los métodos de optimización. Construir esta clase de esquemas representa siempre un problema debido a que hay una gran cantidad de criterios para agrupar las distintas técnicas. Un posible criterio podría ser la base formal que sustenta las técnicas. Otro podría ser la capacidad del método para encontrar el óptimo en una región específica del espacio de búsqueda (i.e. incorporar restricciones en la búsqueda), otro es la cantidad de soluciones que se evalúa simultáneamente. En el esquema de la figura 3.1 se ha querido separar los métodos entre tradicionales y heurísticos, entendiendo por los primeros aquellos que tienen una base matemática que soporta sus resultados, da cuenta de las condiciones bajo las cuales los métodos convergen y se puede estimar el error de su solución. Los segundos, son muchas veces separados en heurísticos y metaheurísticos, pero el presente trabajo no se ocupa de esas diferencias. La idea es ubicar al lector en el campo donde se encuentra la evolución artificial y las técnicas que abarca, aunque algunas como la programación evolutiva y las estrategias de evolución no se tratan aquí. De la misma manera, en el esquema se ve que la evolución artificial no es pues la única heurística y que en esta misma área se encuentran técnicas ingeniosas, como las basadas en inteligencia colectiva (*swarm intelligence*), colonias de hormigas, inteligencia de partículas y búsqueda por difusión estocástica (*stochastic diffusion search*). Tampoco se quiere decir, como podría pensarse de este esquema, que los métodos heurísticos no permiten incorporar restricciones, sino que esta característica es un diferenciador importante en las metodologías tradicionales. Finalmente, se quiere anotar que los métodos multipoblacionales requieren que se especifique una región de búsqueda, a diferencia de los métodos tradicionales donde se indica una condición inicial, y que esto disminuye en gran medida la cantidad de información *a priori* sobre la solución que se necesita para poder usarlos.

3.3. ¿Por qué evolución artificial?

Los métodos de optimización heurísticos abarcan una amplia variedad de algoritmos entre los cuales uno de los más conocidos son los algoritmos evolutivos. En esta sección se revisan algunas aplicaciones de la optimización heurística a problemas de estimación en estadística y econometría,

con el fin de mostrar antecedentes de la investigación presente.

Un argumento que puede darse a favor de utilizar optimización heurística es que existe actualmente una gama de métodos de estimación con base en simulación que utilizan algoritmos relativamente complejos, como el método de momentos eficiente (EMM). En los casos en los cuales la función objetivo se describe mediante algoritmos Spall, (Spall, 2003), afirma que los algoritmos genéticos son mucho más apropiados para la optimización que algoritmos convencionales. Estos métodos se han desarrollado al tiempo que han aparecido modelos de series de tiempo continuos y discretos para diversas aplicaciones, como los modelos de volatilidad estocástica para series de precios y tasas de interés. En Zivot y Wang, (Zivot y Wang, 2006), se desarrolla más la aplicación de estos métodos a estas clases de modelos.

Un trabajo reciente es el de Goffe y Rogers, (Goffe y Rogers, 1994). Allí, comparan el desempeño de un tipo de algoritmo heurístico, el temple simulado, en cuatro problemas econométricos y lo comparan con tres algoritmos convencionales. Uno de los problemas es una red neuronal con 35 parámetros, usando un algoritmo de gradiente conjugado con derivada numérica, en supercomputadores Cray, programados en Fortran 77. Encuentran que el temple simulado tiene mayor capacidad de encontrar el óptimo local y tiene propiedades de robustez. También pueden mencionarse los análisis de Mayer, (Meyer, 2003) sobre aplicaciones estadísticas de un nuevo método tipo de algoritmo heurístico multipoblacional.

Pero tal vez el argumento que ha dado fama a los algoritmos heurísticos es su capacidad de encontrar el óptimo global en presencia de varios óptimos locales. Este es un problema que se presenta en la estimación estadística cuando existe multimodalidad de la función objetivo, por ejemplo, la función log-verosimilitud. La multimodalidad puede deberse a mala especificación del modelo, por ejemplo, pretender estimar un modelo ARMA(p,q) con órdenes bajos de p y q cuando el modelo real tiene órdenes altos. Puede deberse también a la no identificabilidad del modelo, por ejemplo en modelos MA(q) no invertibles. O a dificultades propias del modelo como el caso bien conocido de la estimación en mezclas finitas de distribuciones. Se presenta también en el caso de procesos tipo GARCH con variables dicótomas en la volatilidad, ver Doornik y Ooms, (Doornik y Ooms, 2003) y en los modelos de regresión aparentemente no relacionadas, ó modelo SUR de Zellner, para el cual Burton y Richardson, (Burton y Richardson, 2001), reportan que la distribución de probabilidad del modelo presenta hasta cinco puntos estacionarios. La multimodalidad está presente en muchos casos de estimación y la aplicación de algoritmos evolutivos es una solución recomendable para el proceso de optimización involucrado. Lo anterior busca motivar el uso de las técnicas de evolución artificial en la optimización de funciones que provienen de problemas de estimación de parámetros.

3.4. Algoritmos evolutivos

La Evolución Artificial (EA) es un campo de las ciencias computacionales que se ha venido desarrollando a gran velocidad. Como concepto, la EA, busca emular el paradigma neodarwiniano

de la evolución de las especies. Algunas de sus técnicas más representativas son los Algoritmos Evolutivos (AE), la Programación Evolutiva (PE), las Estrategias de Evolución (EE) y la Evolución Diferencial (ED). Todas estas técnicas tienen una estructura similar y se aplican siguiendo casi los mismos pasos. En esta sección se describe los AE y la ED, por ser métodos enmarcados dentro de la evolución artificial y que guardan ciertas diferencias desde su base conceptual. Los algoritmos de ED se emplean en este trabajo para optimizar la log-verosimilitud de los rendimientos del proceso de difusión con saltos luego de que al compararlos con los métodos de AE, los primeros tuvieran un mejor desempeño.

Los AE son un método de búsqueda y optimización que emula el mecanismo de la selección natural y la herencia genética a partir del paradigma neodarwiniano de la evolución biológica (Holland, 1992). Un AE es un proceso iterativo que finaliza cuando se verifica una condición. Esta condición se satisface normalmente cuando se llega a un resultado aceptable o se alcanza un número determinado de iteraciones. Los AE se han aplicado en casi todas las disciplinas (Fleming y Purshouse, 2002; Spall, 2003). Algunos autores consideran que los algoritmos genéticos (AG) son un caso especial de los algoritmos evolutivos (Whitley, 2001, 2002), puesto que usan una codificación binaria en lugar de un alfabeto real para las posibles soluciones. La eficiencia computacional de los AE depende en gran medida de la programación y de la codificación escogida. Un AE se caracteriza por las llamadas operaciones genéticas, una Función de Desempeño y un Mecanismo de Selección. El esquema general de un AE se muestra en el algoritmo 10. En el apéndice A se presenta una breve descripción de las operaciones genéticas, la función de aptitud y el mecanismo de selección.

Algoritmo 10 Algoritmo Evolutivo General

- 1: Generación de la Población Inicial $M(t)$ dentro del espacio de búsqueda $D \subseteq \mathbb{R}^k$.
 - 2: Evalúe $M(t)$.
 - 3: **repetir**
 - 4: $t := t + 1$
 - 5: Selección de individuos de $M(t)$.
 - 6: Creación de la nueva población, $M'(t)$, a partir de las operaciones genéticas.
 - 7: $M(t) = M'(t)$
 - 8: Evalúe $M(t)$
 - 9: **hasta que** se satisfaga algún criterio de finalización.
-

3.5. Evolución diferencial

En esta sección se presenta la ED, una heurística que se agrupa dentro de los algoritmos evolutivos y que tiene su origen en el temple genético (*genetic annealing*), propuesto por Price en (Price, 1994) y que más tarde sería modificado por Price y Storn al adaptar el método para resolver el problema de ajuste de un polinomio de Tchebychev. Estos dos autores presentan el método en (Price y Storn, 1997) y desde entonces ha recibido toda la atención de la comunidad de los algoritmos

evolutivos y la evolución artificial.

La evolución diferencial es también un método multipoblacional. Para cada individuo de la población, hasta satisfacer un criterio de parada, se genera un individuo de prueba que se compara con éste utilizando la función objetivo. El mejor de los dos se conserva en la nueva población y el otro se descarta. La creación de los individuos de prueba es el punto en el que la evolución diferencial es radicalmente distinto a los algoritmos genéticos, por ejemplo. Supóngase que se está creando un individuo de prueba x' para el j -ésimo individuo x^j de la generación t . Para ello se selecciona aleatoriamente otros tres individuos, distintos entre sí y también diferentes a x^j , dígase x^{j1} , x^{j2} y x^{j3} . De manera aleatoria se determina los parámetros del individuo x^j que serán variados para crear el individuo de prueba. Esto se hace siempre de manera que al menos un parámetro cambie. Si se determina que el parámetro m -ésimo ha de cambiarse, este cambio se lleva a cabo de acuerdo a la siguiente regla probabilística

$$x'_m = \begin{cases} x^{j1}_m + F(x^{j2}_m - x^{j3}_m) & \text{si } r < Cr \\ x^j_m & \text{en otro caso} \end{cases},$$

donde r es un número aleatorio, Cr es la constante de cruzamiento y F es la constante de diferenciación. Si el parámetro cambiado resulta caer por fuera de la región factible, entonces se genera aleatoriamente de manera que cumpla con las restricciones. De esta manera el individuo de prueba x' se forma a partir del individuo x^j y las diferencias de los individuos x^{j2} y x^{j3} ponderadas y agregadas al individuo x^{j1} . Ésta, por supuesto, no es la única estrategia en la evolución diferencial para crear los nuevos individuos. En el apéndice A se muestra las cinco estrategias más usuales y que están implementadas en la función *DEoptim* de R, (Ardia, 2008). El pseudocódigo de un algoritmo típico de evolución diferencial se presenta en el algoritmo 11.

Algoritmo 11 Algoritmo de Evolución Diferencial

- 1: NP : tamaño de la población, F : constante de diferenciación, Cr : constante de cruzamiento.
 - 2: Generación de la Población Inicial $M(t)$ dentro del espacio de búsqueda $D \subseteq \mathbb{R}^k$.
 - 3: Evalúe $M(t)$.
 - 4: **repetir**
 - 5: $t := t + 1$
 - 6: **para** Cada individuo **haga**
 - 7: Seleccione tres individuos diferentes entre sí de $M(t)$.
 - 8: A partir de estos individuos cree un nuevo individuo de prueba x' .
 - 9: Haga que x' cumpla las restricciones.
 - 10: Evalúe x' en la función objetivo y en caso de ser mejor reemplace el individuo actual por x' , de lo contrario conserve el individuo actual.
 - 11: **fin para**
 - 12: Creación de la nueva población, $M'(t)$ a partir de la población antigua y los nuevos individuos que resultaron mejores.
 - 13: $M(t) = M'(t)$
 - 14: **hasta que** se satisfaga algún criterio de finalización.
-

Ejemplo: optimización de la función de Rastrigin. Para ilustrar como trabaja un algoritmo de evolución diferencial se presenta la optimización de la función de Rastrigin. Esta función se define

como en la ecuación (3.2) y, para dos dimensiones, su mínimo global es $f(0, 0) = 0$. La superficie definida por esta función para dos dimensiones se muestra en la figura 3.2. En esta gráfica se puede apreciar que la función de Rastrigin tiene una gran cantidad de mínimos y máximos locales que pueden dificultar la aplicación de los métodos de descenso por gradiente, por ejemplo.

$$f(\mathbf{x}) = 10 \times k + \sum_{i=1}^k (x_i^2 - 10 \cos(2\pi x_i)) \quad (3.2)$$

$$-5.12 \leq x_i \leq 5.12, i = 1, 2, \dots, k.$$

Con el paquete *DEoptim* de R, (Ardia, 2008), se lleva a cabo la optimización con una población de 50 individuos, una constante de diferenciación $F = 0.8$, 200 generaciones y una estrategia rand/1. La figura 3.3 muestra el proceso de evolución. Observe que en la generación 1, las posibles soluciones o población inicial está distribuida de manera uniforme sobre el espacio de búsqueda. Ya en la generación 30, la población se ha concentrado un poco más y se puede observar que hay individuos que están muy cerca del óptimo global (ubicado en $\mathbf{x} = (0, 0)'$). En la generación 70 se observa ya una concentración de la población en el óptimo global. Este ejemplo, en el cual la función objetivo tiene múltiples óptimos (mínimos) locales sirve para ilustrar la capacidad de los algoritmos de evolución artificial, y en particular de la evolución diferencial, para llegar al óptimo global.

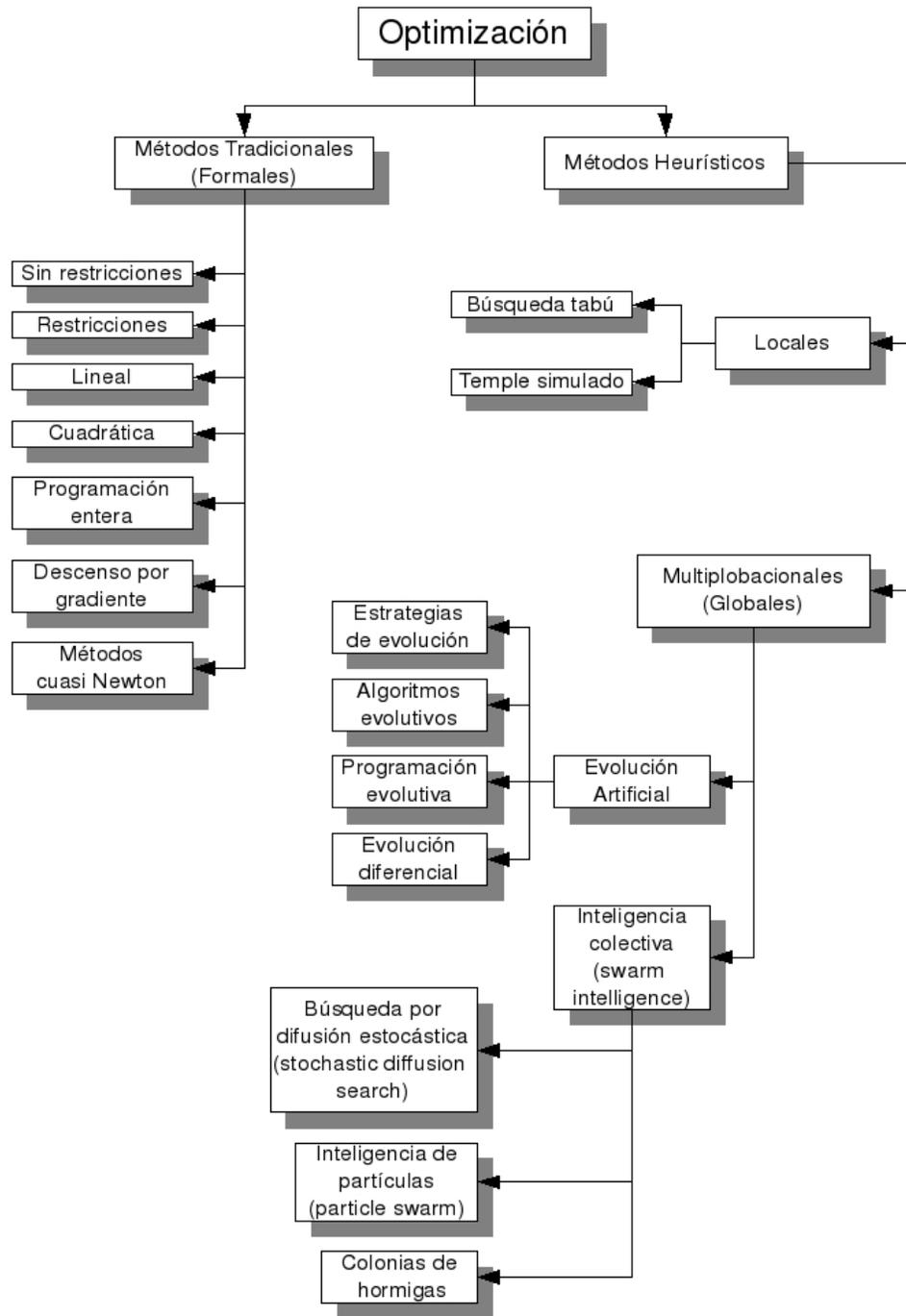


Figura 3.1: Un esquema general de métodos de optimización.

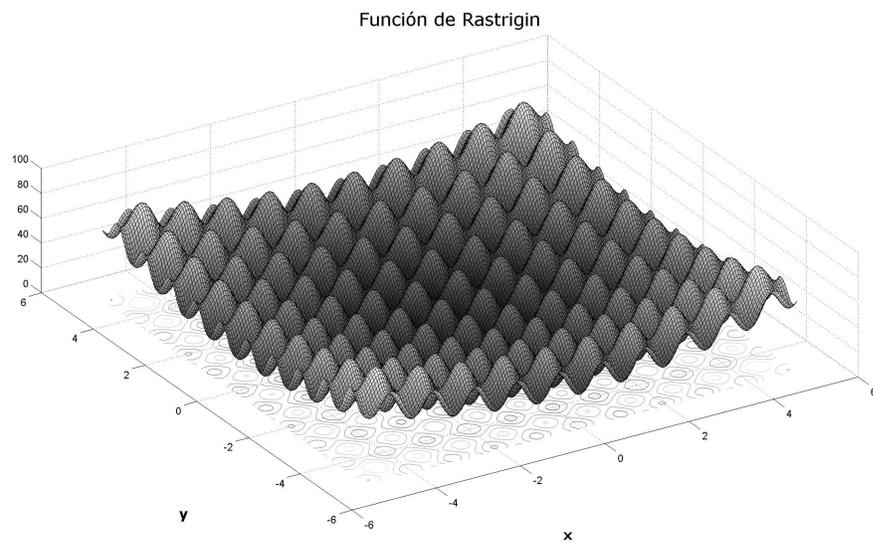
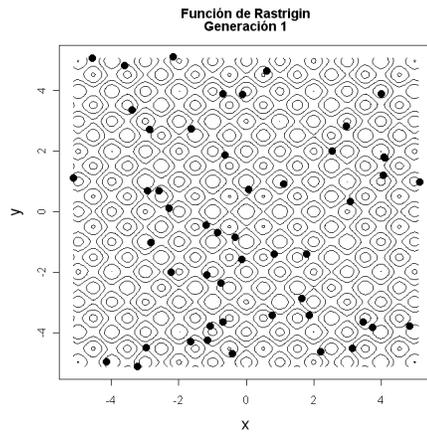
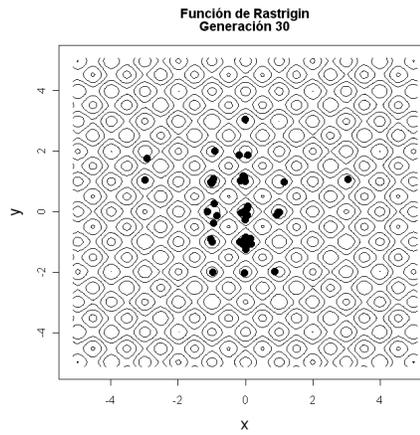


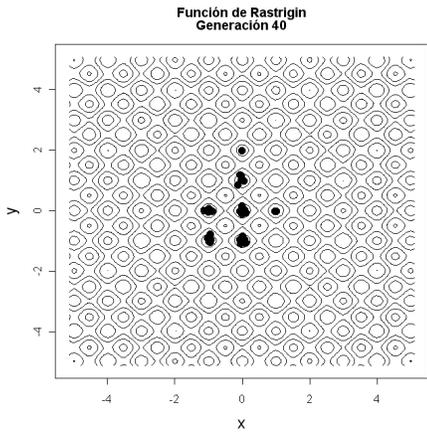
Figura 3.2: Superficie definida por la función de Rastrigin.



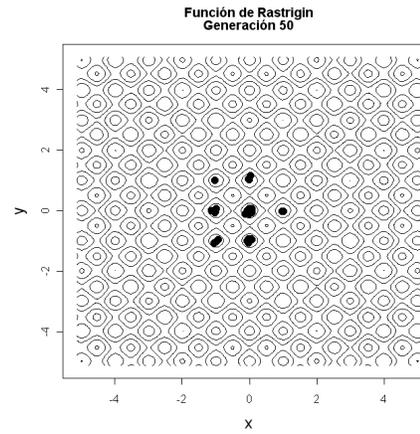
(a) Generación 1



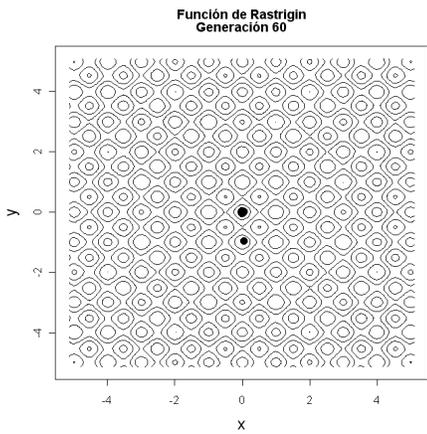
(b) Generación 30



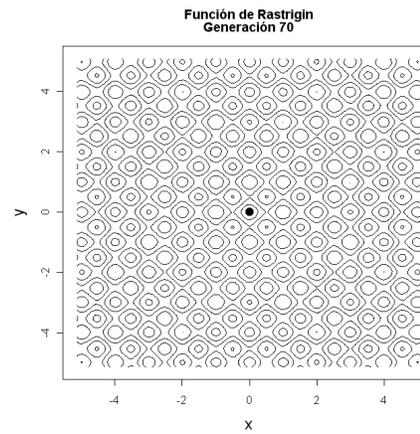
(c) Generación 40



(d) Generación 50



(e) Generación 60



(f) Generación 70

Figura 3.3: Evolución de la población en la optimización de la función de Rastrigin.

Capítulo 4

Estimación por Máxima Verosimilitud

4.1. Introducción

En este capítulo se introduce el método de máxima verosimilitud para estimar el modelo de difusión con saltos de parámetros constantes y donde el término de los saltos sigue una distribución SGED ($Q_k \sim SGED(\mu, \sigma, k, \lambda)$). Se considera el caso en el que se dispone de observaciones o mediciones del proceso en instantes de tiempo discretos y equiespaciados en el intervalo $[0, T]$. La aplicación del método de máxima verosimilitud, según el enfoque asumido en este trabajo, tiene dos maneras posibles de acuerdo a si se asume o no la aproximación del proceso Poisson a un proceso Bernoulli, y requiere el uso de técnicas de convolución para obtener la densidad de probabilidad de una suma de variables aleatorias independientes. En este trabajo se asume la aproximación Bernoulli-Poisson y la estimación de los parámetros del modelo se hace al maximizar la función de verosimilitud que se obtiene como la productoria de la densidad incondicional de los datos. Esta densidad, a su vez, se obtiene numéricamente discretizando las densidades de las variables aleatorias involucradas en los rendimientos logarítmicos del proceso de difusión con saltos y aplicando la transformada rápida de Fourier (*Fast Fourier Transform* - FFT) para obtener, vía convolución, la densidad incondicional de los rendimientos. En la primera parte del capítulo se presenta el problema de aproximación numérica de la log-verosimilitud y la solución propuesta en este trabajo cuando se asume la aproximación Poisson-Bernoulli. Posteriormente se muestra el estudio del problema de optimización de la log-verosimilitud aproximada mediante diferentes técnicas de optimización, entre las que se destaca como más efectiva la evolución diferencial (DE). Posteriormente se muestra un análisis aplicado del problema de estimación, seguido por la aplicación del modelo a dos series: el precio de cierre de la acción de IBM y de Laboratorios Abbot. Al final del capítulo se menciona el problema de aproximación de la log-verosimilitud cuando no se asume la aproximación Bernoulli-Poisson y se bosqueja posibles soluciones que se podrían probar en trabajos futuros.

El procedimiento que aquí se sigue es análogo hasta cierto punto al que siguen en (Planchet y

Therond, 2005), donde muestran en forma bastante clara como se deduce la función de densidad de los datos cuando los saltos siguen una distribución lognormal. Considere la ecuación (4.1),

$$\Delta Y(t) = \left(\mu - \frac{\sigma^2}{2} \right) \Delta t + \sigma \Delta W(t) + \sum_{k=1}^{\Delta P(t)} Q_k. \quad (4.1)$$

Observe que si se define los procesos $D(t)$ y $J(t)$ como en las ecuaciones (4.2) y (4.3),

$$\Delta D(t) = \left(\mu - \frac{\sigma^2}{2} \right) \Delta t + \sigma \Delta W(t), \quad (4.2)$$

$$\Delta J(t) = \sum_{k=1}^{\Delta P(t)} Q_k, \quad (4.3)$$

se puede escribir el proceso $Y(t)$ como en la ecuación (4.4),

$$\Delta Y(t) = \Delta D(t) + \Delta J(t). \quad (4.4)$$

Por las propiedades de carencia de memoria, independencia de los incrementos de los procesos markovianos e independencia de los procesos Wiener y de Poisson, para cada $t \in [0, T]$ se tiene que $\Delta D(t)$ y $\Delta J(t)$ son variables aleatorias independientes y además $\Delta D(t) \sim N\left(\left(\mu - \frac{\sigma^2}{2}\right) \Delta t, \sigma^2 \Delta t\right)$. De esta manera, por la independencia entre $\Delta D(t)$ y $\Delta J(t)$, conociendo la función de densidad de probabilidad de $\Delta J(t)$ se puede obtener, vía convolución, la función de densidad de $\Delta Y(t)$ y así aplicar el método de máxima verosimilitud. Lo anterior se apoya en el hecho de que si X y Z son dos variables aleatorias independientes con rango en los reales, con densidades f_X y f_Z , respectivamente, la densidad de $W = X + Z$, se obtiene como $f_W = f_X * f_Z$, donde $*$ denota el operador convolución.

La convolución de las densidades f_X y f_Z se define como en la ecuación (4.5),

$$(f_X * f_Z)(t) = \int_{-\infty}^{+\infty} f_X(\tau) f_Z(\tau - t) d\tau. \quad (4.5)$$

Evaluar esta expresión depende de que tan sencillas o complejas sean f_X y f_Z . Una alternativa para obtener la convolución de dos funciones es utilizar la transformada de Fourier. Esta transformada es usada comúnmente en ingeniería para hacer análisis de frecuencia y en series de tiempo para modelar ciclos periódicos. La transformada de Fourier de una función se define como en la ecuación (4.6) y la transformada de Fourier inversa como en la ecuación (4.7),

$$\mathfrak{F}(f)(\omega) = F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-i\omega x} dx, \quad (4.6)$$

$$\mathfrak{F}^{-1}(F)(x) = f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega x} d\omega, \quad (4.7)$$

donde $i = \sqrt{-1}$.

Una propiedad importante de la transformada de Fourier es la relacionada con la convolución entre dos funciones. Si F_X y F_Z denotan las transformadas de Fourier de f_X y f_Z , respectivamente, entonces, la convolución entre f_X y f_Z es la transformada de Fourier inversa del producto entre F_X y F_Z , como se muestra en la ecuación (4.8),

$$(f_X * f_Z)(t) = \mathfrak{S}^{-1}(F_X \times F_Z)(t). \quad (4.8)$$

Esta propiedad, si bien es útil, solo puede aplicarse cuando se conoce explícitamente F_X y F_Z , para lo cual es necesario llevar a cabo la integral de la ecuación (4.6) para f_X y f_Z . Como se mencionó anteriormente, la transformada de Fourier se aplica en campos donde lo usual es disponer de secuencias de observaciones (o señales). Esto motivó el desarrollo de un algoritmo para aplicar una versión discreta de la transformada de Fourier, conocida como la Transformada Discreta de Fourier (*Discrete Fourier Transform*), DFT por sus siglas en inglés. Si $x_k \in \{x_0, x_1, \dots, x_{N-1}\}$ es una sucesión de números complejos, la DFT de esta sucesión es la sucesión $X_k \in \{X_0, X_1, \dots, X_{N-1}\}$ que se define como en la ecuación (4.9),

$$X_j = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi i}{N}kj}, \quad j = 1, 2, \dots, N-1, \quad (4.9)$$

y la inversa de esta transformación, denotada por IDFT, se muestra en la ecuación (4.10),

$$x_j = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N}kj}, \quad j = 1, 2, \dots, N-1. \quad (4.10)$$

Análogo a la convolución entre dos funciones, la convolución entre dos sucesiones g_k y h_k se define para cada k como en la ecuación (4.11)

$$(g * h)_k = \sum_n g(n) h(k-n). \quad (4.11)$$

La propiedad de convolución de la DFT permite tener una versión discreta de la ecuación (4.8). Así, la convolución entre dos sucesiones g_k y h_k se puede obtener como en la ecuación (4.12)

$$(g * h)_k = IDFT(DFT(g) \times DFT(h))_k. \quad (4.12)$$

Con el fin de disminuir el costo computacional de la aplicación de la DFT se diseñaron los algoritmos de Transformada Rápida de Fourier (*Fast Fourier Transform*), FFT. Éstos, son métodos que subdividen el problema de encontrar la DFT de una sucesión de números en problemas más simples que, al resolverse e integrarse nuevamente en un nivel superior, permiten llegar a la DFT más rápido. Debido a que la mayoría de programas tienen implementado algoritmos FFT en lugar de la ecuación (4.9), se hará referencia a la FFT en lugar de la DFT y a la IFFT en vez de la IDFT.

Finalmente, para mostrar como se conecta la FFT con lo que se hace en este trabajo, considere que las densidades de probabilidad de las variables aleatorias X y Z , f_X y f_Z , se discretizan con un tamaño de paso Δt en un intervalo $(-h, h)$ en el cuál está contenida la mayor parte de su masa. Es decir que $\int_{-h}^h f_X(t) dt = \int_{-h}^h f_Z(t) dt \approx 1$. La densidad de $W = X + Z$, f_W , discretizada en el intervalo $(-h, h)$ con un tamaño de paso Δt , se puede aproximar como se muestra en la ecuación (4.13)

$$f_W = (f_X * f_Z) = IFFT(FFT(f_X) \times FFT(f_Z)), \quad (4.13)$$

donde f_X , f_Z y f_W representan las sucesiones que resultan al discretizar las funciones de densidad de probabilidad de las variables aleatorias X , Z y W , respectivamente. Este abuso notacional pretende no complicar la escritura introduciendo nuevos nombres para las versiones discretas de las densidades y no debe prestarse a confusiones, pues cuando se mencione la FFT o la IFFT siempre se estará haciendo referencia a las densidades discretizadas. Para ver aplicaciones de la FFT en la obtención de densidades aproximadas se recomienda el trabajo (Embrechts, Grübel, y Pitts, 1993).

4.2. Estimación por máxima verosimilitud asumiendo la aproximación Bernoulli para el proceso Poisson

En este enfoque se asume que $\Delta P(t) \sim Bernoulli(\lambda \Delta t)$, es decir que se asume la aproximación del proceso Poisson a un proceso Bernoulli de acuerdo con (Ball y Torous, 1983). Esto simplifica, en gran medida, el trabajo necesario para obtener la densidad de $\Delta J(t)$, $f_{\Delta J}(x)$. Si se denota por $f_{\Delta J, \Delta P}(x, m)$ la densidad conjunta de los procesos $\Delta J(t)$ y $\Delta P(t)$ y por $f_{\Delta P}(m)$ la densidad de los incrementos del proceso de saltos entonces, por la definición de funciones de densidad condicional, se tiene la ecuación (4.14),

$$f_{\Delta J, \Delta P}(x, m) = f_{\Delta J}(x | \Delta P) f_{\Delta P}(m). \quad (4.14)$$

De esta manera, usando el Teorema de Probabilidad Total, puede obtenerse la función de densidad de probabilidad de $\Delta J(t)$ como en la ecuación (4.15),

$$f_{\Delta J}(x) = \sum_{m=0}^{+\infty} f_{\Delta J, \Delta P}(x, m) = \sum_{m=0}^{+\infty} f_{\Delta J}(x | \Delta P = m) f_{\Delta P}(m). \quad (4.15)$$

Esta densidad se conoce como la densidad de una variable aleatoria Poisson compuesta y su tratamiento se abordará a nivel exploratorio al final de este capítulo.

Ahora, bajo el supuesto de que el proceso de saltos es aproximadamente Bernoulli, ΔP solo toma los valores 0 ó 1. Si $\Delta P = 0$ se tiene que $\sum_{k=1}^{\Delta P=0} Q_k = 0$ y por consiguiente la función de densidad condicional se convierte en la función delta de Kronecker $\delta_{0,x}$, la cual es igual a uno si

$x = 0$ y cero en otro caso. Cuando $\Delta P = 1$, entonces $\sum_{k=1}^{\Delta P=1} Q_k = Q_1 = Q$ y la función de densidad condicional es la de una variable aleatoria, en este caso, distribuida SGED cuya definición se da en la ecuación (1.41) y que aquí se denota por f_Q . De esta manera se encuentra que la densidad de $\Delta J(t)$ se puede escribir como en la ecuación (4.16),

$$f_{\Delta J}(x) = (1 - \lambda\Delta t) \delta_{0,x} + (\lambda\Delta t) f_Q(x). \quad (4.16)$$

En este punto se tiene las dos densidades de las variables que componen a $\Delta Y(t)$, $f_{\Delta D}$ y $f_{\Delta J}$. La primera, $f_{\Delta D}$, corresponde a una densidad normal con media $\left(\mu - \frac{\sigma^2}{2}\right) \Delta t$ y varianza $\sigma^2 \Delta t$ y la segunda, $f_{\Delta J}$, es la que se presenta en la ecuación (4.16). El incremento $\Delta Y(t)$ es la suma de los dos incrementos $\Delta D(t)$ y $\Delta J(t)$ que a su vez son independientes entre sí. Luego, la densidad de $\Delta Y(t)$, $f_{\Delta Y}$, se puede obtener como la convolución entre las funciones de densidad $f_{\Delta D}$ y $f_{\Delta J}$, así

$$f_{\Delta Y} = f_{\Delta D} * f_{\Delta J}. \quad (4.17)$$

donde por $*$ se denota la operación convolución. La convolución funciona como un operador lineal, así que la densidad de $\Delta Y(t)$ se puede escribir como en la ecuación (4.18),

$$f_{\Delta Y}(x) = (1 - \lambda\Delta t) (f_{\Delta D} * \delta_{0,x})(x) + (\lambda\Delta t) (f_{\Delta D} * f_Q)(x). \quad (4.18)$$

Finalmente, como al convolucionar una función con el delta de Kronecker se obtiene la misma función, la densidad $f_{\Delta Y}(x)$ se puede plantear como en la ecuación (4.19),

$$f_{\Delta Y}(x) = (1 - \lambda\Delta t) f_{\Delta D}(x) + (\lambda\Delta t) (f_{\Delta D} * f_Q)(x). \quad (4.19)$$

De esta manera, la densidad de $\Delta Y(t)$ se obtiene como una suma ponderada de dos densidades, una densidad normal y la otra resultante de la convolución entre una densidad normal y la densidad de la variable subyacente al proceso de saltos, que en este caso corresponde a la SGED. Es importante tener presente que la función de densidad de los incrementos, $f_{\Delta Y}$, está determinada por un conjunto de parámetros $\theta = (\mu, \sigma, \lambda, \mu_Q, \sigma_Q, \nu, \xi)'$, correspondientes a la tendencia y varianza del proceso de difusión, la intensidad con que se dan los saltos ($\lambda\Delta t$) y los parámetros de la distribución SGED, cuya media y varianza se distinguen de la media y la varianza del proceso normal con el subíndice Q . De esta manera, si se dispone de una realización de rendimientos logarítmicos del proceso $\{\Delta y_t\}_{t=0}^T$, la verosimilitud de los datos se puede escribir como en la ecuación (4.20),

$$L(\theta | \{\Delta y_t\}_{t=0}^T) = \prod_{t=0}^T f_{\Delta Y}(\Delta y_t | \theta). \quad (4.20)$$

La propuesta de este trabajo para aproximar numéricamente la densidad $f_{\Delta Y}$ consiste en:

1. Discretizar las densidades $f_{\Delta D}$ y f_Q en un intervalo suficientemente amplio y con un tamaño de paso adecuado.
2. Obtener la convolución $f_{\Delta D} * f_Q$ como $IFFT(FFT(f_{\Delta D}) \times FFT(f_Q))$, donde FFT y $IFFT$ denotan la transformada rápida de Fourier y la inversa de la transformada rápida de Fourier, respectivamente.
3. Obtener $f_{\Delta Y}$ llevando a cabo la suma ponderada que se muestra en 4.19.

Con el propósito de no complicar la notación, no se hará diferencia en la forma de denotar las densidades y sus correspondientes discretizaciones.

La discretización se lleva a cabo siguiendo el trabajo de (Embrechts, Grübel, y Pitts, 1993). Allí, los autores consideran aproximar numéricamente la densidad de un proceso Poisson compuesto con severidades no negativas. Proponen evaluar la densidad de las severidades, en M puntos, donde $M = 2^j$. Esto debido a que la FFT es más eficiente si se aplica a secuencias cuya longitud es una potencia de dos. La variable j , determina entonces el número de puntos en que se evalúa la densidad de las severidades. Así, se evalúa la densidad de las severidades en los puntos $x = kh$, con $k = 0, 1, \dots, M-1$ y h el tamaño de paso. En el presente trabajo las severidades se distribuyen SGED con sus respectivos parámetros. La propuesta es evaluar la densidad en los puntos $x = kh$, con $k = -M + 0.5, -M + 1.5, \dots, -1.5, -0.5, 0.5, 1.5, \dots, M - 1.5, M - 0.5$ y h el tamaño de paso. De esta manera se obtendría la densidad discretizada en $2 \times 2^j = 2^{j+1}$ puntos alrededor de cero. Es razonable asumir que es alrededor de cero que debe discretizarse las densidades, pues la media de la mayoría de los rendimientos financieros es cercana a cero. Una consecuencia de esta metodología es que el rango de discretización queda automáticamente establecido en el intervalo $[(-M + 0.5)h, (M - 0.5)h]$.

El tamaño de paso adecuado depende de la dispersión de la densidad. Densidades con altas curtosis (muy concentradas alrededor de la media) requieren valores pequeños de h , mientras que densidades con dispersiones altas no requieren tamaños de paso tan pequeños. Una serie de pruebas, basadas en la integración numérica de las densidades normal y SGED sobre los rangos de discretización, mostró que valores adecuados del tamaño de paso se obtienen al tomar $h = \frac{\sigma}{200}$, donde σ es la desviación estándar de la densidad. Este procedimiento heurístico probó ser efectivo tanto para la densidad normal como para la densidad SGED. De esta manera, no es necesario especificar el tamaño de paso, sino que este parámetro se ajusta dentro del algoritmo cada vez que se evalúa un vector candidato de parámetros del modelo.

A manera de ejemplo, suponga que se tiene el proceso de difusión con saltos con parámetros $\mu = 0.250$, $\sigma = 0.5$, $\lambda = 10$ y $Q_k \sim SGED(0, 0.2, 1.5, 2)$, muestreado 260 veces al año (i.e. $\Delta t = 1/260$) y observado durante cinco años completos. Es decir que la tasa de saltos $\lambda = 10$ corresponde a tener un promedio de diez saltos por año. Otra alternativa, sería asumir que se tiene un proceso muestreado en cada unidad de tiempo, por ejemplo diaria ($\Delta t = 1$) y que la intensidad con

la que ocurren los saltos es $\lambda = 10/260$. Ambas escalas tienen el mismo resultado para $\lambda\Delta t$, que es en últimas lo que define si la aproximación Bernoulli-Poisson es o no factible. El algoritmo 12 muestra el código en R utilizado para aproximar numéricamente la densidad empleando el procedimiento antes descrito. En dicho programa *mun* y *sigman* son la media y la varianza de la componente normal del proceso que se muestra en la ecuación (4.2). Con este tamaño de paso para los incrementos del proceso, $\Delta t = 1/260$, un tamaño de paso $h = \left[\min\left(\frac{0.5}{\sqrt{260}}, 0.2\right) \right] / 200 \approx 1.5 \times 10^{-4}$ y un rango que va desde -5.08 hasta 5.08 para discretizar las densidades, se obtiene la función de densidad aproximada que se presenta en la figura 4.1. Allí puede apreciarse el efecto que tiene el componente de los saltos en las colas de la densidad aproximada de los rendimientos y cómo este componente causa una desviación del comportamiento normal.

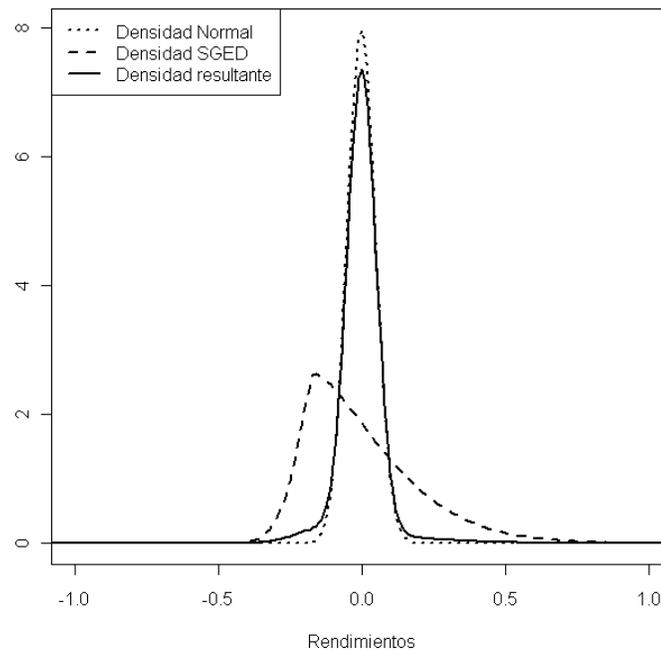


Figura 4.1: Densidad aproximada del proceso.

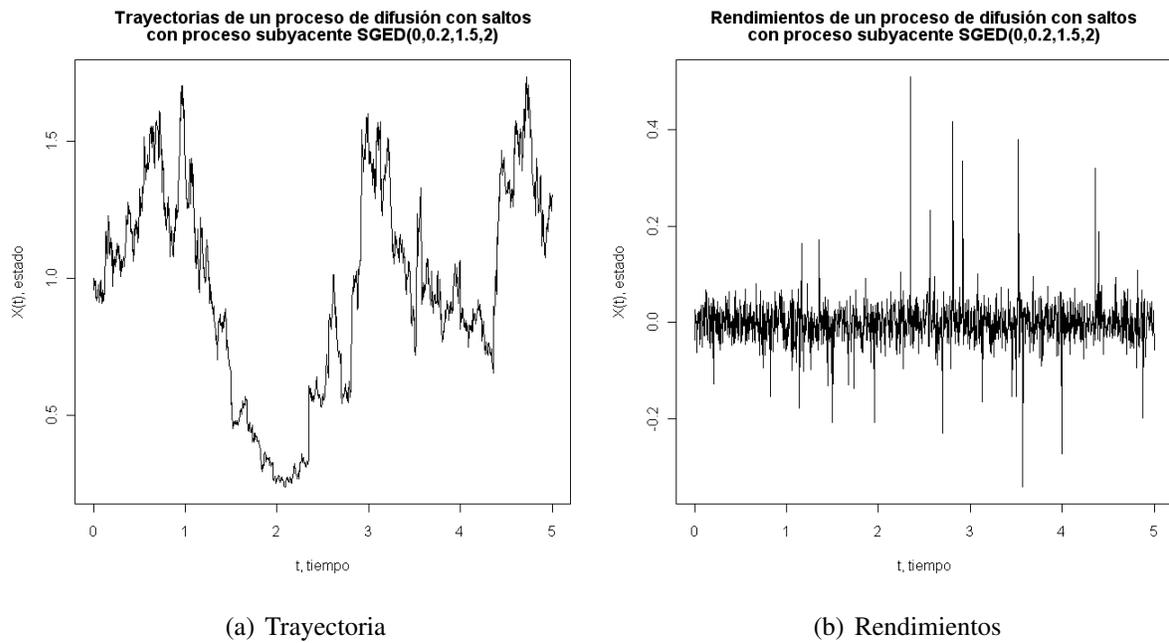
La figura 4.2 presenta una trayectoria de este proceso con sus correspondientes rendimientos e histograma. Sobre este último se ha superpuesto la densidad aproximada de los rendimientos vía convolución y se observa un buen ajuste.

Cabe anotar que con los parámetros seleccionados el proceso es la suma de una tendencia determinística, un movimiento browniano y una variable aleatoria Poisson compuesta. Por supuesto, la desviación del comportamiento normal en los rendimientos no es muy grande, ya que $\lambda\Delta t \approx 0.038$ y prevalece este comportamiento sobre el que induce el proceso de saltos, como puede observarse de la ecuación (4.19).

Algoritmo 12 Código en R para aproximar la densidad del proceso asumiendo la aproximación Bernoulli del proceso Poisson.

```
f.rec=function(Dt, lambda, mun, sdn, muq, sdq, nuq, xiq) {  
# Este programa aproxima numéricamente la densidad de una suma  
# de una normal de media mun y desv. est. sdn con una Poisson  
# compuesta con severidades SGED de parámetros mean=muq, sd=sdq,  
# nu=nuq y xi=xiq. El cálculo se hace vía convolución utilizando  
# la transformada rápida de Fourier y la aproximación Bernoulli  
# para el proceso Poisson.  
  j=15  
  M=2^j #Número de puntos a cada lado de cero.  
  k=as.array(seq((-M+0.5), (M-0.5))) #Vector de longitud 2^(j+1).  
  h=min(abs(c(sdn, sdq))/200) #Tamaño del paso para la discreti-  
  #zación.  
  x=k*h #Puntos en los que se aproxima la función.  
  n=2*M  
  #Se calcula las densidades normal y SGED en el rango discreto:  
  fn=dnorm(x, mean=mun, sd=sdn)  
  fs=dsged(x, mean=muq, sd=sdq, nu=nuq, xi=xiq)  
  #Se calcula la transformada de Fourier de las densidades  
  #discretizadas:  
  Fn=fft(fn)  
  Fs=fft(fs)  
  #Cálculo de la convolución de la densidad normal con la SGED:  
  F=abs(Re(fft(Fs*Fn, inverse=T))) *h/n  
  F=c(F[(M+1):(2*M)], F[1:M])  
  #F=F/sum(F*h)  
  #Densidad resultante:  
  f.rec=(1-lambda*Dt)*fn+lambda*Dt*F  
  return(f.rec=list(x=x, f=f.rec)) }
```

El algoritmo 12 muestra el programa para obtener la densidad aproximada del proceso. Dicho programa devuelve un objeto de la clase lista con un dominio discreto y los correspondientes valores aproximados de la densidad en cada punto. En el proceso de estimación de los parámetros del modelo se requiere evaluar la verosimilitud de la ecuación (4.20), para conocer que tan adecuado es un conjunto de parámetros θ . Dado un vector de parámetros candidato, se debe aproximar la densidad de los rendimientos con estos parámetros y luego evaluar la densidad en cada uno de los datos observados con el fin de obtener el valor de la función de verosimilitud para este vector de parámetros. Como se dispone de un dominio discreto y los valores de la densidad en cada uno de



Histograma de los rendimientos de un proceso de difusión con saltos con proceso subyacente SGED(0,0.2,1.5,2)

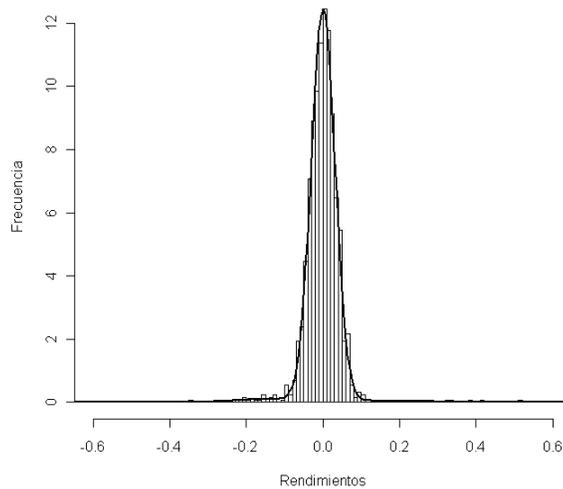


Figura 4.2: Trayectoria, rendimientos e histograma de una simulación de un proceso de difusión con saltos asumiendo la aproximación Bernoulli de la Poisson.

sus puntos, es necesario escoger una de las posibles maneras para aproximar el valor de la densidad en uno de los puntos intermedios del dominio discretizado. Si se ha aproximado la densidad con un tamaño de paso suficientemente pequeño es de esperar que cualquier método de interpolación funcione razonablemente bien. En este trabajo se ha escogido la interpolación lineal para aproximar

la función de densidad incondicional de los rendimientos del proceso en los puntos intermedios del dominio discreto en que está definida. El algoritmo 13 presenta el código en R de la función de argumentos vectoriales con la que se evalúa los datos en la densidad.

Algoritmo 13 Código en R para aproximar la densidad del proceso usando interpolación lineal.

```
f.rec.evalvec=function(t, x, y) {
# x es un vector e y son los valores de la
# densidad correspondientes a cada punto
# de x. También es un vector. t es el vector
# en el que se desea evaluar la densidad.
p1=rep(0, length(t))
  for (i in 1:length(t)) {
    p1[i]=which.min(abs(x-t[i]))
  }
d=ifelse(x[p1]==t, y[p1],
  ifelse(x[p1]>t, y[p1-1]+
  (y[p1-1]-y[p1]) / (x[p1-1]-x[p1]) * (t-x[p1-1]),
  y[p1+1]+ (y[p1+1]-y[p1]) / (x[p1+1]-x[p1]) * (t-x[p1+1])))
return(d)
}
```

Para llevar a cabo la estimación del modelo solo resta definir la función con la que se evalúa el logaritmo natural de la verosimilitud, log-verosimilitud, $-\ln L(\theta)$. El código en R para evaluar la log-verosimilitud se presenta en el algoritmo 14. Algo destacable en este código es que se ha definido que los valores de los parámetros que producen resultados no numéricos (NaN) tienen un valor de la log-verosimilitud de 10^7 (una constante positiva grande). El efecto de esta modificación es que los algoritmos evolutivos descartarán este tipo de soluciones. Los resultados no numéricos provienen de conjuntos de parámetros no factibles, para los cuales la evaluación de la densidad, ya sea normal o SGED, no es numéricamente posible con las funciones implementadas en R.

4.2.1. Análisis aplicado del problema de estimación

En este apartado se prueba la metodología desarrollada, utilizando un conjunto de datos simulados, con los algoritmos propuestos. La idea es analizar como el tipo de algoritmo de optimización empleado para encontrar el óptimo de la log-verosimilitud incide en los estimadores que se obtienen y también como el conocimiento *a priori* acerca de la solución afecta el resultado final. Como aplicación se presenta la estimación del modelo con los mismos parámetros de la serie simulada anteriormente. Es decir,

$$(\lambda, \mu, \sigma, \mu_q, \sigma_q, \nu, \xi) = (10, 0.25, 0.5, 0, 0.2, 1.5, 2).$$

Algoritmo 14 Código en R para evaluar la verosimilitud usando interpolación lineal.

```
logver=function(parametros,Dt,datos){
# Log-verosimilitud utilizando la aproximación Bernoulli-Poisson:
  lambda=parametros[1]
  # mun es la media del componente normal del proceso:
  mun=(parametros[2]-(parametros[3]^2)/2)*Dt
  # sdn es la varianza del componente normal del proceso:
  sdn=parametros[3]*sqrt(Dt)
  # Parámetros de la SGED:
  muq=parametros[4]
  sdq=parametros[5]
  nuq=parametros[6]
  xiq=parametros[7]
  # Aproximación de la densidad con los parámetros dados:
  f=f.rec(Dt,lambda,mun,sdn,muq,sdq,nuq,xiq)
  if((min(f$x)>min(datos)) | (max(f$x)<max(datos))){
  return(10^7)}else{
  # Evaluación de los datos en la densidad:
  d=f.rec.evalvec(datos,f$x,f$f)
  # Expresión de la log-verosimilitud:
  ld=log(d)
  logver=sum(ld)
  if(is.nan(logver) | is.na(logver)){
    logver=-10^7}
    resultado=-logver
  return(-logver)
  }
}
```

La intención de mostrar esto, es ilustrar el funcionamiento de la metodología de estimación utilizando la función de verosimilitud aproximada propuesta.

Optimización de la log-verosimilitud con BFGS. El primer algoritmo de optimización seleccionado es el BFGS, descrito en (Fletcher y Reeves, 1964), que es un método quasi-Newton implementado en la función *optim* de R, (Team, 2008). Dicho algoritmo es ideal porque no requiere una expresión analítica para el gradiente de la log-verosimilitud y en lugar de ésta utiliza una aproximación numérica. Una ventaja adicional de este algoritmo es que permite aproximar la matriz Hessiana (matriz de segundas derivadas parciales) de la log-verosimilitud, que se utiliza al final para calcular los errores estándar reportados que, vale recordar, sólo son válidos cuando se alcanza el

óptimo global de la log-verosimilitud. Este algoritmo requiere la especificación de un valor inicial y es importante analizar que tan sensible es a la especificación de éste. En la tabla 4.1 se presentan los resultados obtenidos especificando el vector de parámetros inicial como

$$\theta_{01} = (9.8, 0.3, 0.52, 0.001, 0.19, 1.4, 2.1).$$

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.250	0.500	0.000	0.200	1.500	2.000
Valor estimado	9.621	0.286	0.507	-0.012	0.184	1.192	1.790
Error estándar	2.144	0.241	0.012	0.027	0.034	0.464	0.356

Tabla 4.1: Resultados de la estimación con *optim* usando BFGS con condición inicial θ_{01} .

Los resultados presentados en la tabla 4.1 confirman de alguna manera la capacidad de la metodología empleada para aproximar la verosimilitud. El haber llegado a unos parámetros cercanos a aquellos con los que se ha simulado el conjunto de datos, es un indicio del éxito de la metodología para aproximar la verosimilitud. No obstante, un segundo experimento muestra la deficiencia del algoritmo BFGS para resolver este problema. Si se optimiza de nuevo la función objetivo con el vector de parámetros inicial

$$\theta_{02} = (15, -0.5, 1.2, 0.2, 0.8, 0.8, 1.1),$$

entonces se obtienen los resultados que se muestran en la tabla 4.2.

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.250	0.500	0.000	0.200	1.500	2.000
Valor estimado	10.395	0.428	0.499	0.124	1.086	0.215	5.300
Error estándar	2.005	0.234	0.011	1.2e-4	0.005	0.002	0.116

Tabla 4.2: Resultados de la estimación con *optim* usando BFGS con condición inicial θ_{02} .

En la tabla 4.2 se aprecia que los resultados obtenidos difieren bastante de los valores reales. Adicionalmente, si se define θ_1 y θ_2 como la soluciones obtenidas en el primer caso y en el segundo, respectivamente, se tiene que $-\ln L(\theta_1) = -2458.468$ y $-\ln L(\theta_2) = -2445.44$. Luego, la solución θ_2 no es el óptimo de la log-verosimilitud y por tanto los errores estándar calculados con la matriz Hessiana aproximada en ese punto no tienen validez. Con base en estos resultados se puede afirmar que el algoritmo de optimización a emplear debe permitir incorporar restricciones, de manera que se pueda obtener soluciones consistentes con el problema. En este segundo experimento se ha puesto en evidencia que:

1. El problema de optimización a resolver para encontrar los parámetros del modelo de difusión con saltos en cuestión debe tratarse con un algoritmo que permita incorporar restricciones.
2. El problema de optimización a resolver es sensible a la especificación de un vector de parámetros iniciales y por tanto requiere algún conocimiento a priori de la solución para emplear algoritmos de descenso por gradiente o métodos quasi-Newton en general.

Optimización de la log-verosimilitud con L-BFGS-B En R se cuenta con el algoritmo L-BFGS-B, descrito en (Byrd, Lu, Nocedal, y Zhu, 1995), implementado también dentro de la función *optim*. Éste, es una modificación del algoritmo BFGS, utilizado anteriormente, que permite incorporar restricciones al dominio en que se busca los parámetros. Se considera ahora la aplicación del algoritmo L-BFGS-B usando como límite inferior y superior para los parámetros

$$\theta_L = (8, 0, 0.3, 0.001, 0.1, 1.4, 1.9)$$

y

$$\theta_U = (20, 0.5, 0.7, 0.2, 0.4, 2, 3),$$

respectivamente. En un primer experimento se utiliza como punto inicial θ_{01} y se obtiene los resultados que aparecen en la tabla 4.3. Seguidamente, se cambia la condición inicial por θ_{02} para estimar nuevamente el modelo y se obtiene los resultados de la tabla 4.4. Estos últimos resultados son evidencia de que, si se tiene suficientemente acotada la región de búsqueda de la solución, la sensibilidad del algoritmo de búsqueda frente a la condición inicial disminuye, lo cual es un hecho conocido en optimización.

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.250	0.500	0.000	0.200	1.500	2.000
Valor estimado	9.708	0.288	0.507	-0.012	0.183	1.191	1.786
Error estándar	2.265	0.242	0.012	0.029	0.036	0.466	0.454

Tabla 4.3: Resultados de la estimación con *optim* usando L-BFGS-B con condición inicial θ_{01} .

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.250	0.500	0.000	0.200	1.500	2.000
Valor estimado	9.430	0.266	0.507	-0.012	0.185	1.213	1.877
Error estándar	2.162	0.241	0.012	0.028	0.036	0.480	0.491

Tabla 4.4: Resultados de la estimación con *optim* usando L-BFGS-B con condición inicial θ_{02} .

Si se establece como límites superior e inferior

$$\theta_{L'} = (0, -1.5, 0, -1.5, 0, 0, 0)$$

y

$$\theta_{U'} = (20, 5, 10, 5, 4, 10, 10),$$

respectivamente y se lleva a cabo nuevamente la optimización con el algoritmo L-BFGS-B usando la condición inicial θ_{02} , se obtiene los resultados de la tabla 4.5. Estos resultados son bastante pobres y los errores estándar reportados carecen de validez, pues el valor estimado no es el óptimo global de la log-verosimilitud.

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.250	0.500	0.000	0.200	1.500	2.000
Valor estimado	9.690	-1.500	0.000	5.000	0.000	0.000	5.745
Error estándar	2.162	0.241	0.012	0.028	0.036	0.480	0.491

Tabla 4.5: Resultados de la estimación con *optim* usando L-BFGS-B con condición inicial θ_{02} y restricciones más amplias.

Optimización de la log-verosimilitud con temple simulado Una alternativa a los métodos de optimización de descenso por gradiente y quasi-Newton son los métodos heurísticos, tratados en el capítulo 3. Entre estos se encuentra el temple simulado descrito en (Spall, 2003; Belisle, 1992) e implementado en la función *optim* de R. Este algoritmo requiere también un punto inicial a partir del cual se comienza a generar soluciones candidatas utilizando cadenas de Markov. Usando este algoritmo con el punto de partida θ_{02} se obtiene los resultados de la tabla 4.6. Los errores estándar que no aparecen en dicha tabla son los de variables asociados a valores negativos de la diagonal de la matriz Hessiana inversa evaluada en la solución encontrada. Hay que anotar que para el parámetro ν del componente SGED se obtuvo un valor negativo y el valor de la log-verosimilitud fue -2356.11 para esta solución, que es inferior al obtenido en el primer caso por el algoritmo BFGS. Desafortunadamente, el temple simulado de la función *optim* no permite incorporar restricciones. Esta limitación es propia de esta función pues en Matlab, por ejemplo, el algoritmo de temple simulado si permite especificar restricciones.

Optimización de la log-verosimilitud con Evolución Diferencial (DE) Otra alternativa heurística, enmarcada dentro de los algoritmos evolutivos, es la Evolución Diferencial descrita en (Nocedal y Wright, 1999; Spall, 2003) y explicada brevemente en el capítulo 3. En R se cuenta con el paquete *DEoptim*, (Ardia, 2008), que permite llevar a cabo la optimización de una función de valor real usando evolución diferencial. Dicho algoritmo fue escogido sobre los algoritmos evolutivos

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.125	0.500	0.000	0.200	1.500	2.000
Valor estimado	7.835	0.308	0.325	-0.207	4.942	-1.208	0.499
Error estándar	1.429	0.093	0.007	x	x	x	0.107

Tabla 4.6: Resultados de la estimación con *optim* usando temple simulado con condición inicial θ_{02} .

tradicionales disponibles en el paquete *rgenoud* de R, (Mebane, Jr., y Singh Sekhon, 2007), luego de una serie de pruebas en las que se determinó que la evolución diferencial tenía un mejor desempeño. Este algoritmo de evolución diferencial permite la especificación de unas restricciones en la forma de límites superior e inferior para los parámetros, del mismo tipo que requiere el algoritmo L-BFGS-B. Usando como límites superior e inferior θ_L y θ_U se obtuvo los resultados de la tabla 4.7, después de 400 generaciones del algoritmo con una población de 70 individuos y una estrategia de selección rand/1. Usando los límites $\theta_{L'}$ y $\theta_{U'}$ como restricciones de los parámetros se obtuvo los mismos resultados de la tabla 4.7, utilizando la evolución diferencial con las mismas condiciones anteriores. Para reportar el error estándar se evaluó numéricamente la matriz Hessiana en la solución encontrada con ayuda de la función *fdHess* del paquete *nlme* de R, (Pinheiro, Bates, DebRoy, Sarkar, y the R Core team, 2008). Lo interesante de este resultado es que usar las restricciones $\theta_{L'}$ y $\theta_{U'}$ equivale a asumir un menor conocimiento de la solución. Sin embargo, el algoritmo de optimización con evolución diferencial llegó a la misma respuesta.

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.000	0.250	0.500	0.000	0.200	1.500	2.000
Valor estimado	9.621	0.286	0.507	-0.012	0.184	1.192	1.790
Error estándar	1.288	0.153	0.004	0.231	0.996	0.137	0.085

Tabla 4.7: Resultados de la estimación con *optim* usando evolución diferencial, los límites θ_L y θ_U y los límites $\theta_{L'}$ y $\theta_{U'}$.

Análisis del desempeño de la evolución diferencial Resulta interesante ver el proceso de evolución de los parámetros y de la log-verosimilitud en el proceso de optimización. En las figuras 4.3 y 4.4 se muestra estos procesos. En estas gráficas se puede observar que cerca de las 200 generaciones el algoritmo ya había llegado a un valor óptimo, lo que significa que el proceso de optimización podría hacerse en un tiempo más corto si en lugar de un número máximo de iteraciones se especifica como condición de finalización un criterio de convergencia (i.e., no cambiar el valor de la función objetivo significativamente durante un número predeterminado de generaciones).

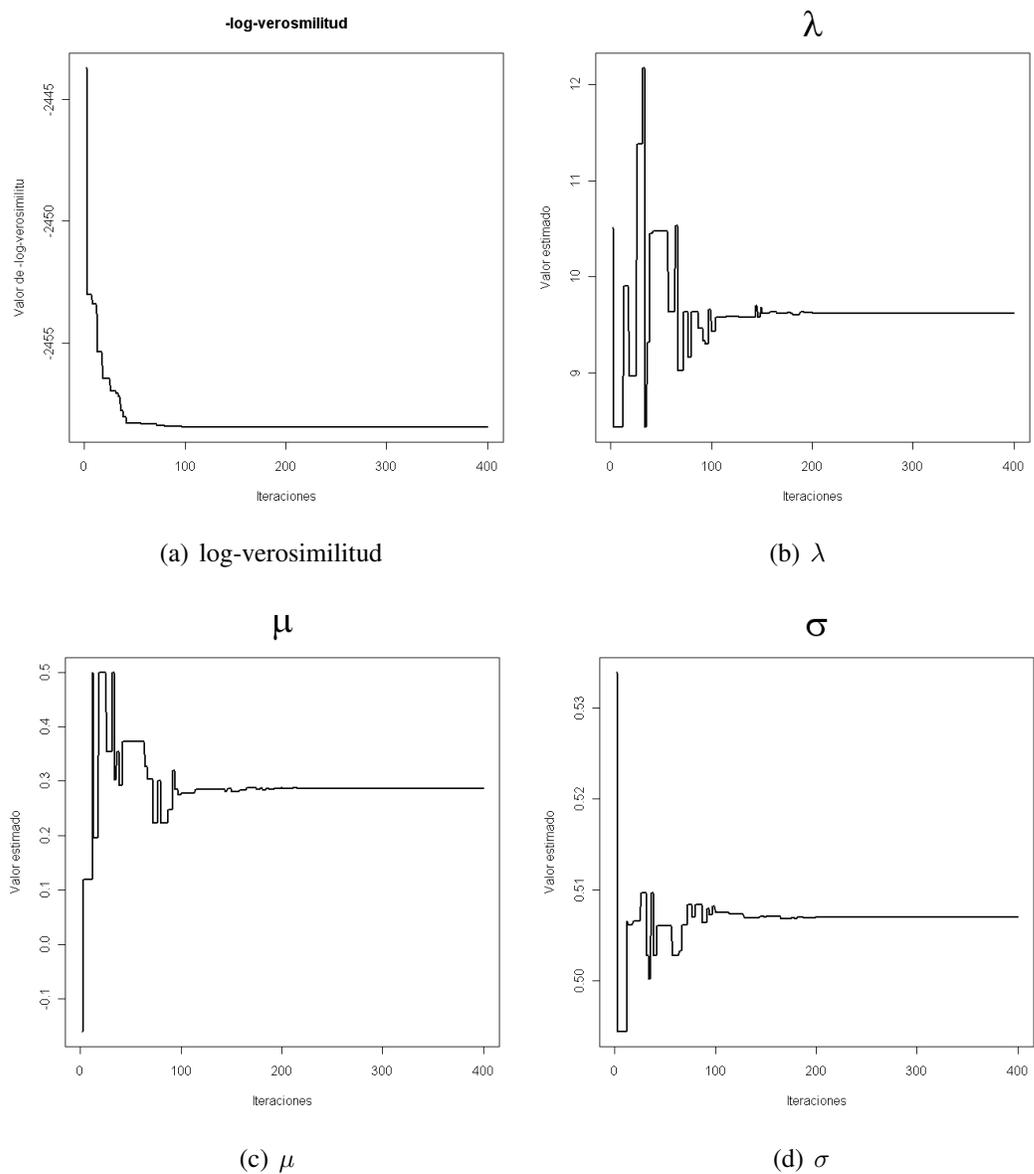


Figura 4.3: Evolución del valor de $-\log$ -verosimilitud y de los estimadores de los parámetros λ , μ , y σ .

Un análisis interesante para comprobar que la solución hallada es un óptimo resulta de observar las gráficas de $-\log$ -verosimilitud restringida por grupos de parámetros. Esto significa que se escoge un punto dentro del espacio parametral y se varía un solo parámetro para evaluar el negativo de la log-verosimilitud. En este caso se escoge como punto de referencia el óptimo hallado por el algoritmo de evolución diferencial que se muestra en la tabla 4.7 y las gráficas se presentan en las figuras 4.5 y 4.6. Note que estas gráficas de $-\log$ -verosimilitud ($-\ln(L)$) son todas unimodales, por lo que se sospecha que el punto en el que están restringidas es un óptimo. Esto, por supuesto, no es

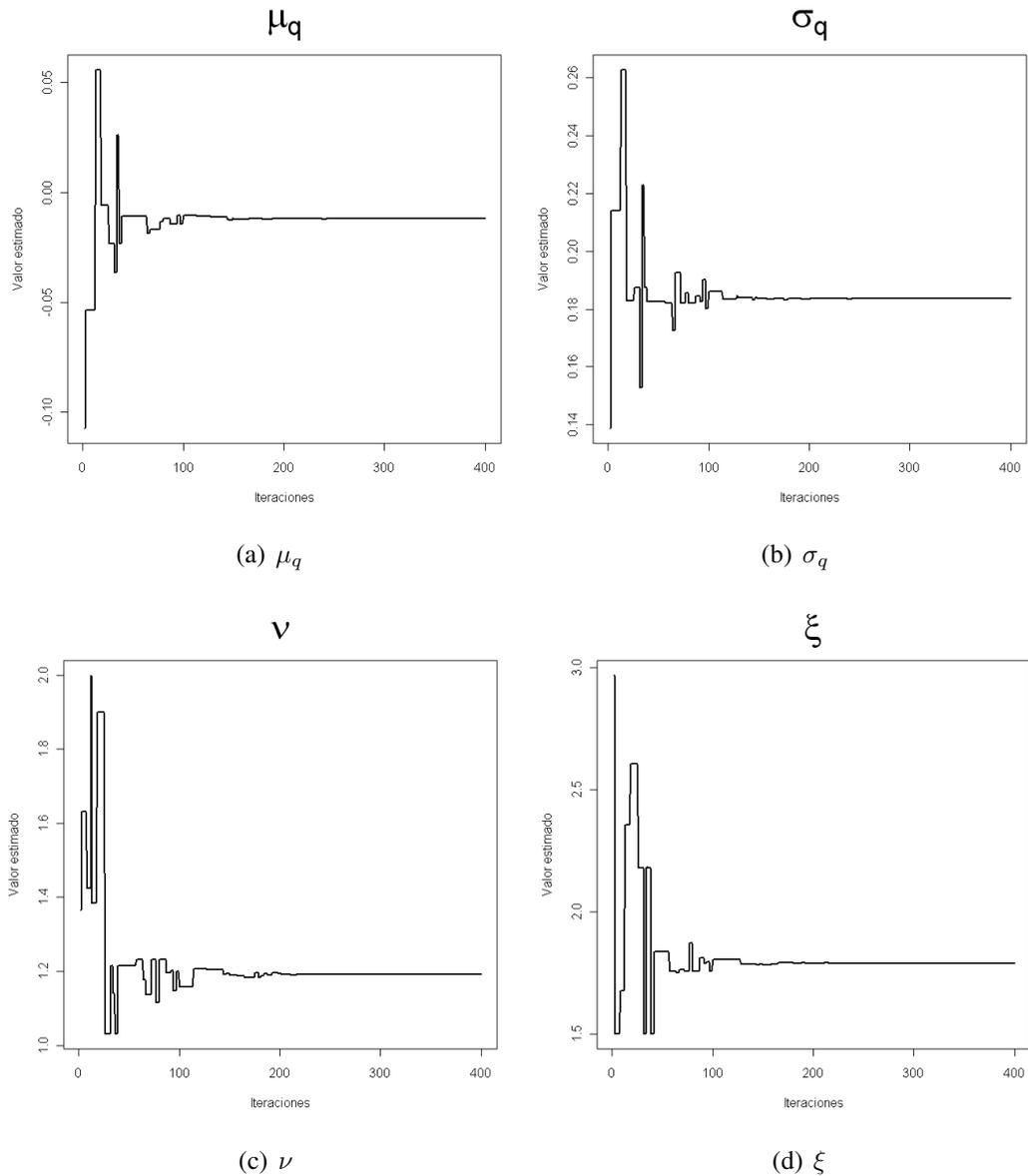


Figura 4.4: Evolución de los estimadores de los parámetros μ_q , σ_q , ν y ξ .

una prueba formal.

Otras gráficas que pueden ser de interés son las de la log-verosimilitud restringida, pero esta vez variando dos parámetros simultáneamente. La figura 4.7 presenta las gráficas de la log-verosimilitud restringida variando simultáneamente los parámetros λ y μ (figura 4.7-(a)), variando simultáneamente los parámetros λ y σ (figura 4.7-(b)) y los parámetros μ y σ (figura 4.7-(c)). Estas gráficas, computacionalmente costosas de obtener, permiten hacerse a una idea de qué tan difícil es encontrar los estimadores de los parámetros variables cuando los parámetros restantes son conocidos. Es claro que el parámetro λ se relaciona con el parámetro μ en una manera diferente a como

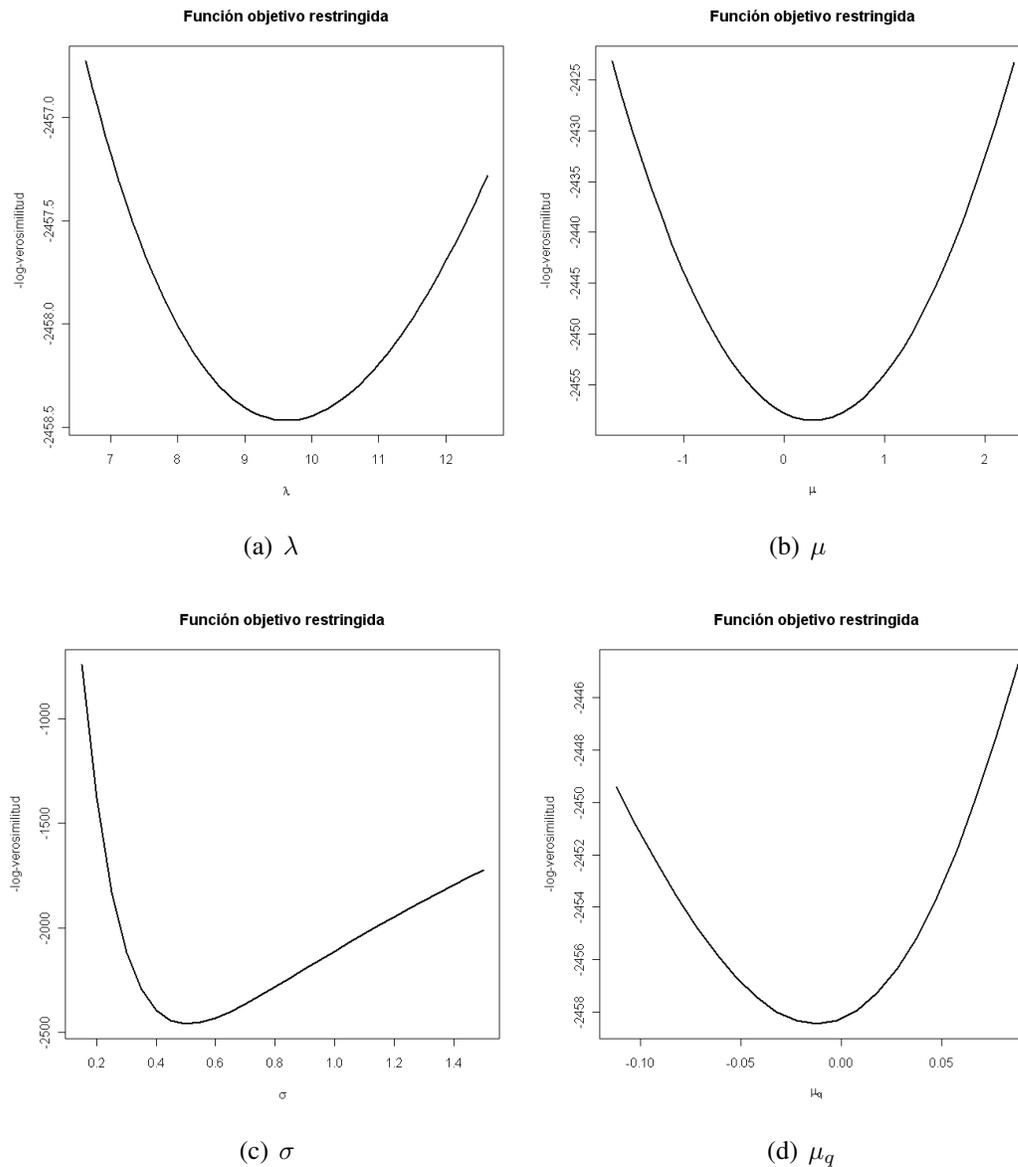


Figura 4.5: -log-verosimilitud restringida evaluada para los parámetros λ , μ , σ y μ_q .

lo hace con el parámetro σ . En cambio el parámetro σ parece relacionarse de manera similar con λ y con μ . Esto se debe a que la función objetivo (ver algoritmo 14) se evalúa de tal manera que si un conjunto de parámetros produce una densidad que no es factible, entonces se le asigna el valor de 10^7 . Esta clase de relaciones, diferentes y complejas y restricciones sobre los parámetros y la función objetivo es la que dificulta en gran parte que un algoritmo de descenso por gradiente, por ejemplo, llegue al óptimo global.

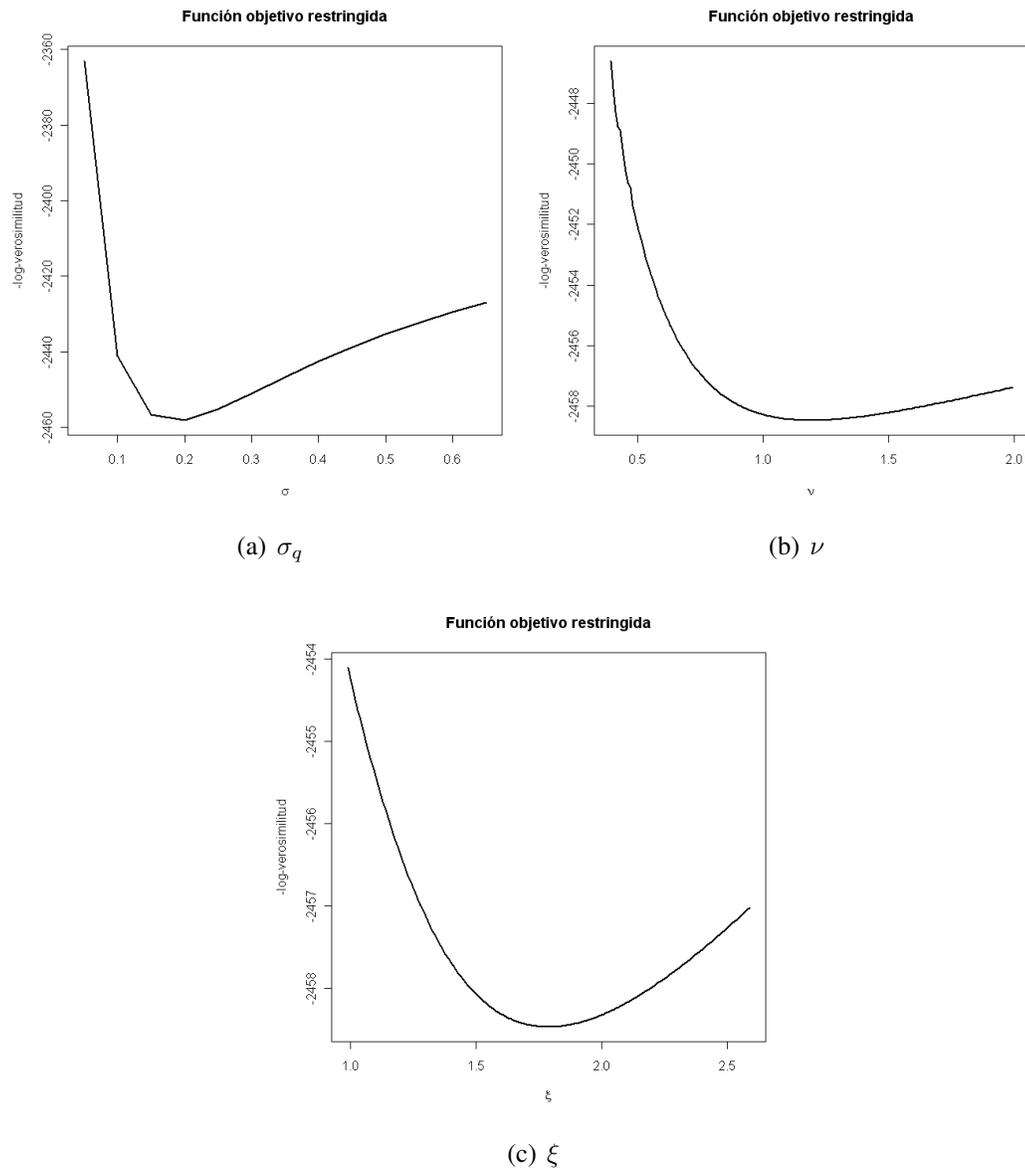
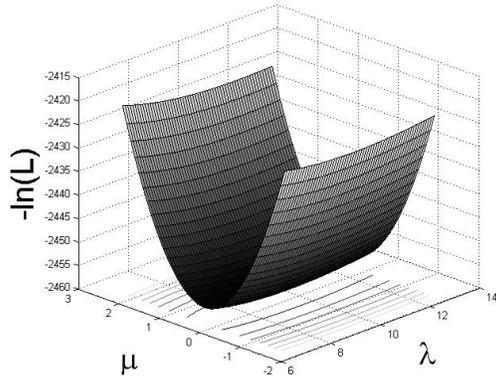
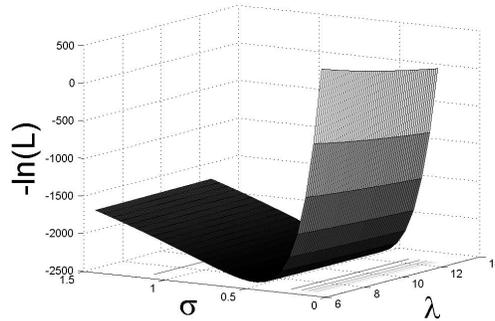


Figura 4.6: $-\log$ -verosimilitud restringida evaluada para los parámetros σ_q , ν , y ξ .

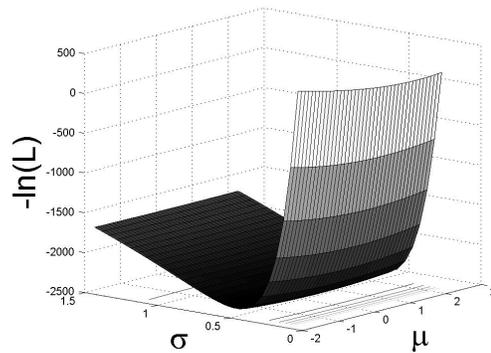
Análisis mediante simulación Con el fin de analizar el desempeño de la evolución diferencial se ha simulado 200 trayectorias del proceso de difusión con saltos utilizado para las estimaciones de este apartado. Para cada serie se ha estimado los parámetros del modelo de manera que se tiene 200 muestras de estimadores de los parámetros $(\lambda, \mu, \sigma, \mu_q, \sigma_q, \nu, \xi)$. El número de simulaciones hechas obedece a una restricción de tiempo. Cada proceso de estimación toma aproximadamente dos horas y media en un *Servidor Dell 2950*, con dos procesadores Procesadores Intel Xeon Quad Core (en total ocho procesadores de 1595.924 MHz cada uno) y 8 Gb de memoria RAM. Para obtener esta muestra de estimadores se necesitó aproximadamente 20 días.



(a) Superficie $(\lambda, \mu, -\ln(L))$.



(b) Superficie $(\lambda, \sigma, -\ln(L))$.



(c) Superficie $(\mu, \sigma, -\ln(L))$.

Figura 4.7: $-\log$ -verosimilitud restringida evaluada para los parámetros (λ, μ) , (λ, σ) y (μ, σ) .

Los histogramas de esta muestra se presentan en las figuras 4.8 y 4.9. Sobre cada uno de los histogramas se ha sobrepuesto una curva normal con la media y la desviación estándar que se muestra en la tabla 4.2.1. En estas gráficas se puede apreciar que la estimación de los parámetros λ , μ y σ tienen una distribución mucho más aproximada a la normal que los parámetros de la distribución SGED, μ_q , σ_q , ν y ξ . Esto, sin embargo, no es concluyente con un número pequeño de simulaciones, aunque es de esperar que la estimación de dichos parámetros represente una dificultad, sobre todo en procesos con tasas de saltos pequeñas.

El resumen de los estadísticos descriptivos de la muestra de estimadores de los parámetros se presenta en la tabla 4.2.1. Observe que para los siete parámetros que se estima, los valores de la media son muy cercanos al valor real. Esto confirma que tanto la metodología para aproximar la verosimilitud como el método para optimizarla funcionan. El parámetro de mayor error estándar es λ . Esto puede explicarse si se piensa que se está ajustando una tasa de saltos anual con cinco años observados. Adicionalmente, los saltos están mezclados con otro proceso. Las consideraciones

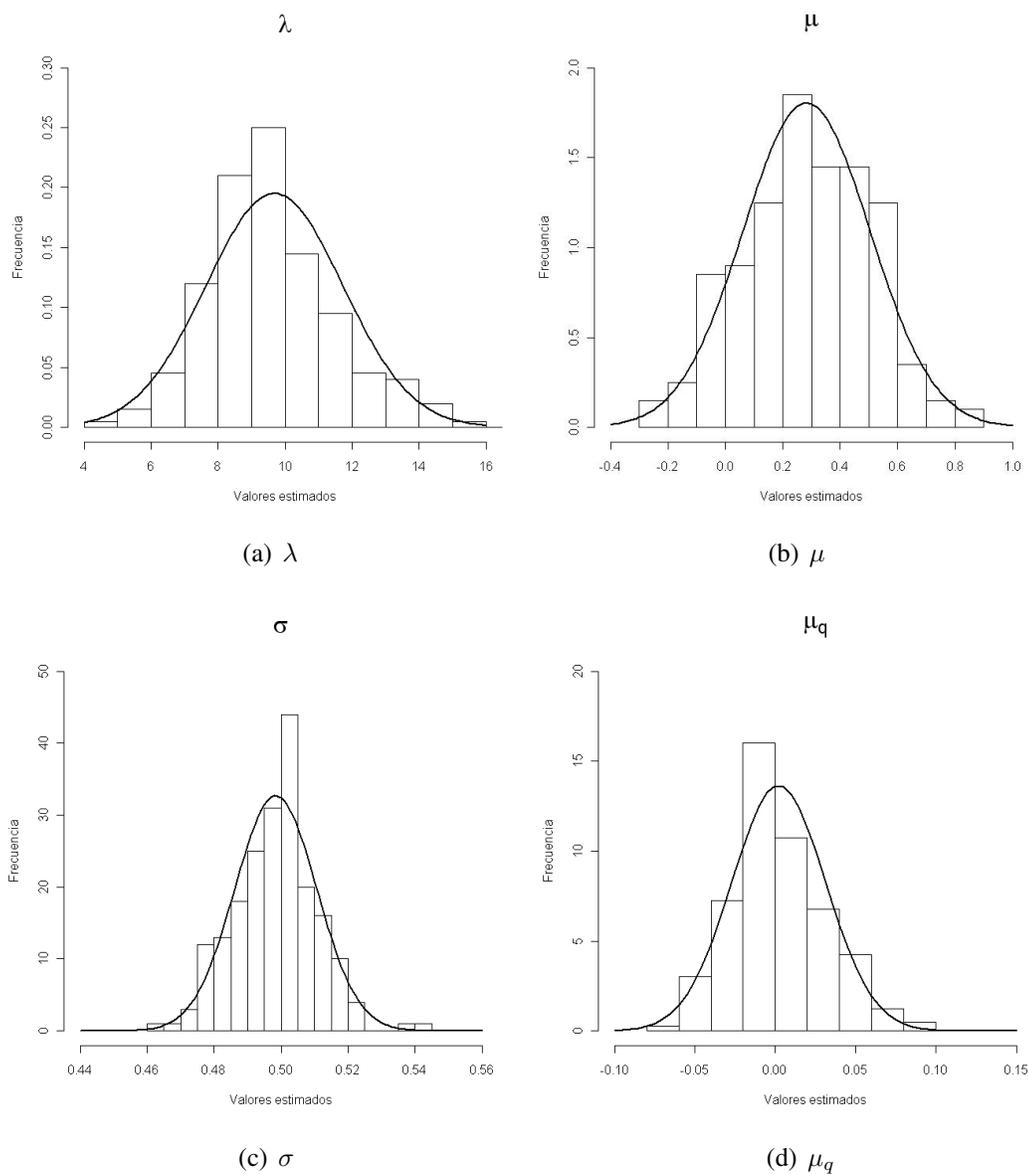


Figura 4.8: Histogramas de los estimadores de los parámetros λ , μ , σ y μ_q .

anteriores llevan a concluir que los parámetros que más dificultad presentan en su estimación son los asociados al proceso Poisson compuesto.

4.2.2. Aplicaciones

Luego de haber desarrollado la metodología de estimación por máxima verosimilitud mediante el uso de la aproximación Bernoulli-Poisson, resulta interesante ajustar el modelo con datos reales para ver su desempeño. A continuación se presenta dos aplicaciones: la serie de precios de cierre

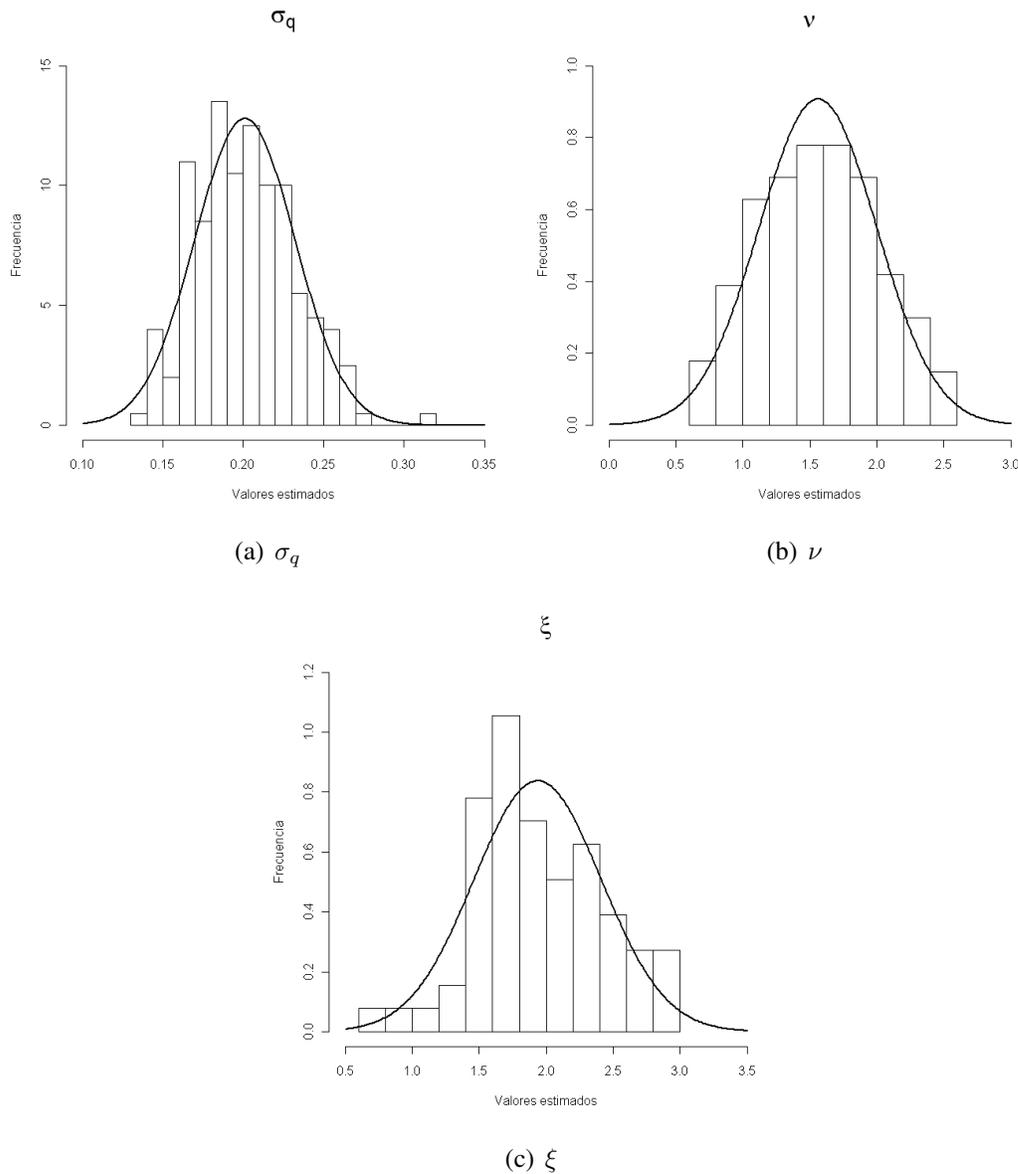


Figura 4.9: Histogramas de los estimadores de los parámetros σ_q , ν , y ξ .

de IBM y la serie de precios de cierre de la acción de Laboratorios Abbott. Ambas series de precios se han ajustado por pago de dividendos y salidas al mercado de nuevos paquetes de acciones (*stock split*). En los dos casos la medida de bondad de ajuste es el negativo de la log-verosimilitud. Los límites superior e inferior, respectivamente, para el algoritmo de evolución diferencial fueron

$$\theta_L = (0, -1, 0, -1, 0, 0, 0)$$

y

$$\theta_U = (60, 1, 1, 1, 1, 3, 3).$$

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor real	10.0000	0.2500	0.5000	0.0000	0.2000	1.5000	2.0000
Media	9.6015	0.2829	0.4984	0.0021	0.2010	1.5607	1.9368
Mediana	9.5574	0.2850	0.4995	-0.0021	0.2000	1.5542	1.8845
Desv. est.	1.8713	0.2212	0.0122	0.0292	0.0312	0.4396	0.4762
EMC	3.6430	0.0498	0.0002	0.0009	0.0010	0.3222	0.2290

Tabla 4.8: Estadísticos descriptivos de la muestra de estimadores.

Precio de la acción de IBM En (Ramezani y Zeng, 1998) se considera, entre varias series, la serie del precio en la bolsa de Nueva York de la acción de IBM. En dicho trabajo los autores ajustan varios modelos de difusión con saltos a esta serie. Los datos corresponden al periodo comprendido entre el primero de enero de 1991 y el 31 de diciembre de 1992. La figura 4.10 presenta las series de precio y rendimientos, así como también el histograma de los rendimientos de la acción de IBM en el periodo que se considera. Para propósitos de comparación, sobre el histograma de los datos se ha superpuesto una curva normal ajustada a los datos. Puede apreciarse que los datos tienen una curtosis mayor que la de la distribución normal.

Para determinar si es adecuado ajustar un modelo de difusión con saltos es importante verificar que los rendimientos sean un proceso iid y estacionario. Esta tarea, no se lleva a cabo en la mayoría de los trabajos que aplican modelos de difusión con saltos a series financieras. La figura 4.11 muestra las funciones de autocorrelación y autocorrelación parcial muestrales de los rendimientos de la acción de IBM. En esta gráfica se puede apreciar que no hay correlaciones significativas. Sin embargo, la aplicación de la prueba BDS (ver (Broock, Scheinkman, Dechert, y LeBaron, 1996)) rechaza la hipótesis de que los rendimientos son una sucesión iid. Esto, si bien no descarta que un modelo de difusión con saltos se pueda ajustar de manera adecuada a los datos, sirve para estar alerta frente al desempeño del modelo. En cuanto a la estacionaridad, la prueba KPSS (ver (Kwiatkowski, Phillips, Schmidt, y Shin, 1992)) no rechaza, con un nivel de significancia $\alpha = 0.05$, la hipótesis nula de que la serie es estacionaria en la media. El valor-p de esta prueba fue 0.09. Ambas pruebas están implementadas en el paquete *tseries* de R, (Hornik, 2009).

En la tabla 4.2.2 se presenta el ajuste del modelo considerado en este trabajo y en la tabla 4.2.2 los resultados obtenidos en (Ramezani y Zeng, 1998), de manera que el lector pueda comparar. Se debe tener en cuenta que la escala que se considera aquí es anual, a diferencia de Ramenazi y Zeng que asumen los días como las unidades de tiempo, en lugar de los años. De esta manera, el valor de λ estimado aquí equivale al valor de $\lambda/260$ estimado por Ramenazi y Zeng.

En (Ramezani y Zeng, 1998) se propone un modelo de difusión con saltos con dos procesos Poisson compuestos: uno, con tasa λ_u y severidades *Pareto* (r_u), para las buenas noticias (saltos positivos) y otro, con tasa λ_r y severidades *Beta* ($r_d, 1$), para las malas noticias. La tabla 4.2.2 muestra los resultados de la estimación. Los parámetros μ y σ equivalen a los coeficientes de deriva

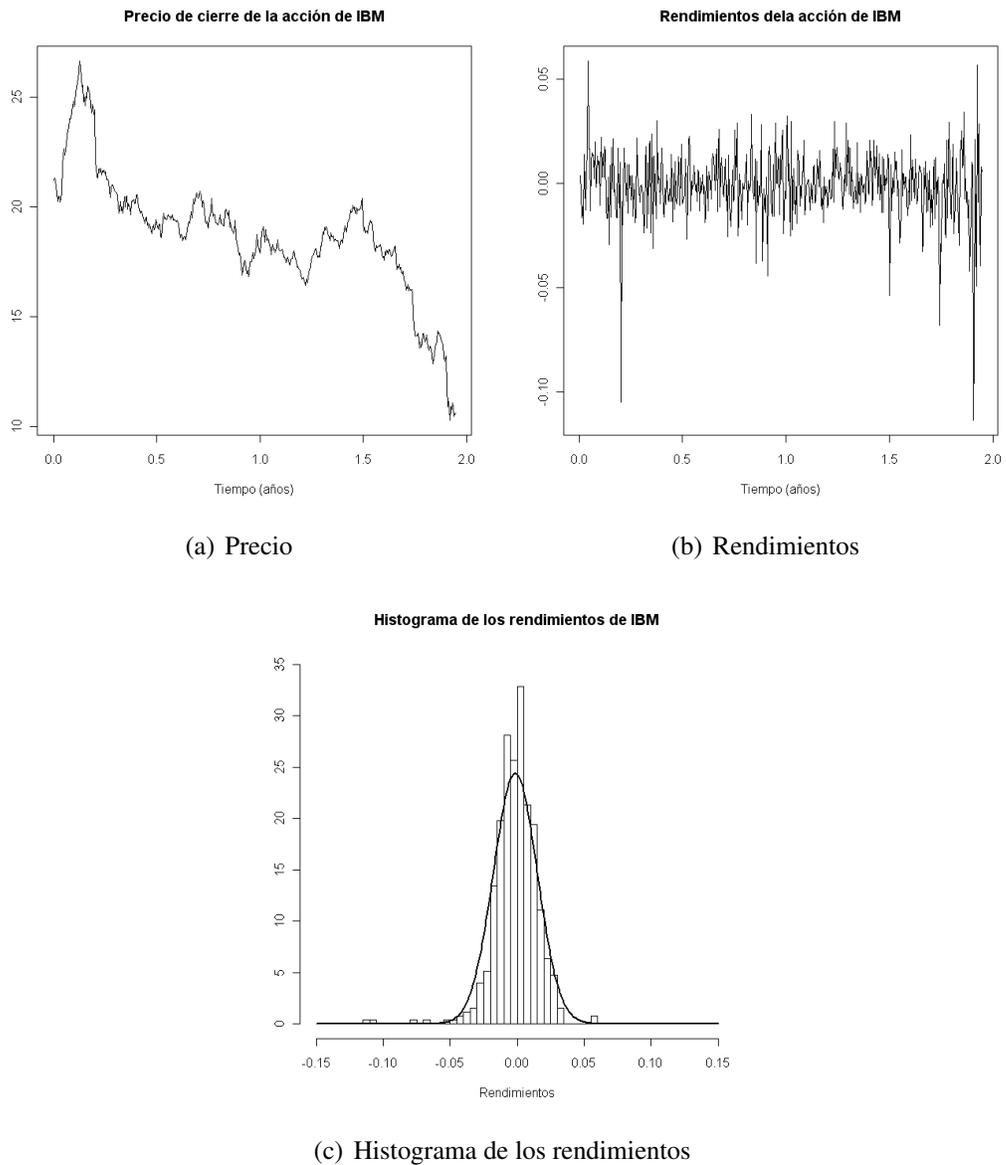
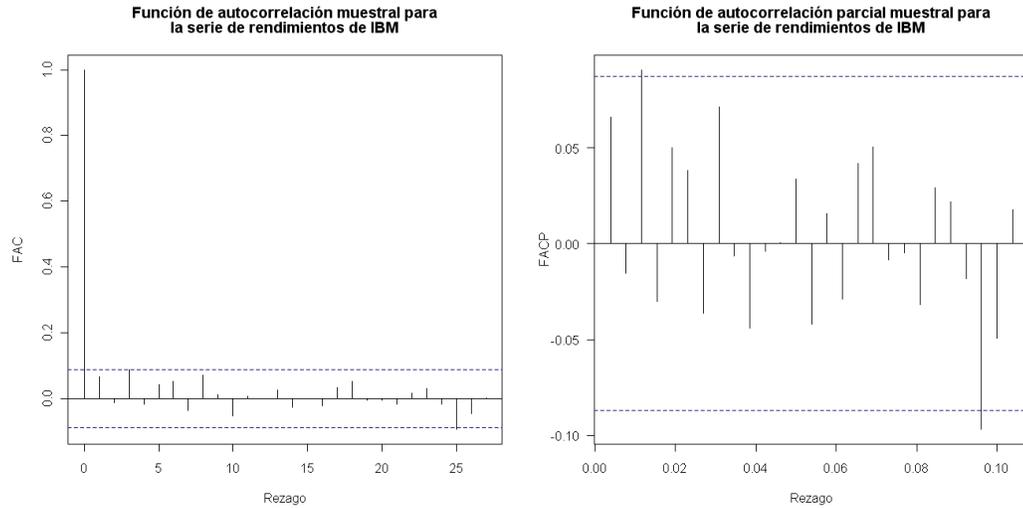


Figura 4.10: Series del precio y rendimientos e histograma de los rendimientos de la acción de IBM.

	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor estimado	11.8865	-0.1030	0.2062	-0.0195	0.0434	3.0000	0.7247
Error estándar	8.9839	0.1693	0.0098	0.0133	0.0145	5.6839	0.5898

Tabla 4.9: Resultados para la acción de IBM.

y volatilidad del modelo que se considera en el presente trabajo, por lo tanto es posible establecer



(a) Función de autocorrelación muestral para los rendimientos de IBM. (b) Función de autocorrelación parcial muestral para los rendimientos de IBM.

Figura 4.11: Funciones de autocorrelación y autocorrelación parcial de los rendimientos de la acción de IBM.

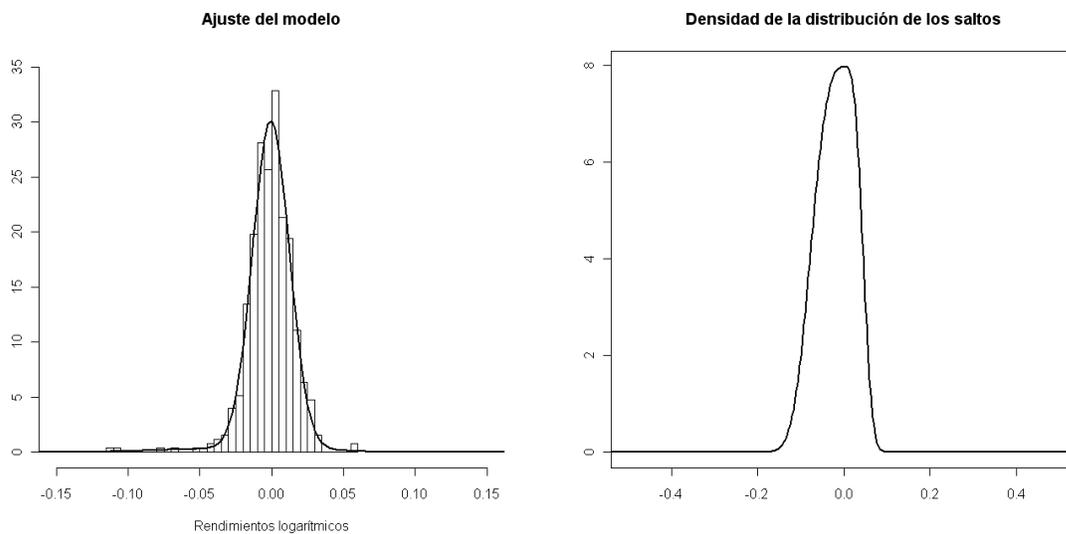
	λ_u	λ_d	r_u	r_d	μ	σ
Valor estimado	0.0884	0.0631	83.5300	37.7800	-5.6690E-04	1.2100E-02
Error estándar	0.6434	0.5721	5.8740	2.3440	1.5530E-04	4.8500E-06

Tabla 4.10: Resultados para la acción de IBM obtenidos en (Ramezani y Zeng, 1998).

una comparación entre los dos modelos. Observe que $\lambda' = \lambda_u + \lambda_r$, es el número esperado de saltos en la unidad de tiempo. De acuerdo con la tabla 4.2.2, $\lambda' = 0.0884 + 0.0631 = 0.15147$. Si multiplicamos este número por 260 (el número de observaciones por año), obtenemos una tasa de saltos anual de 39.38. En (Ramezani y Zeng, 1998) obtuvieron $-\ln(L) = -1422.19$ y en el presente trabajo se obtuvo $-\ln(L) = -1419.886$. Si bien el negativo del logaritmo de la verosimilitud es ligeramente mayor en el presente trabajo, el modelo estimado ajusta razonablemente bien los datos, como se muestra en la figura 4.12-(a). Este gráfico puede compararse con el que se presenta en la figura 4.10-(c), donde se ajusta una distribución normal a los rendimientos. En la figura 4.12-(b) se puede apreciar la gráfica de la densidad estimada del proceso subyacente a los saltos, la cual presenta asimetría, que es una de las características que se busca abarcar con el modelo que se considera en este trabajo.

Los errores estándar obtenidos en ambos trabajos son grandes respecto a los valores de los parámetros, en particular los asociados a los parámetros de los procesos Poisson compuestos. Esto se debe a que sólo se dispone de 506 observaciones, lo que representa estimar tasas anuales (de

saltos) con menos de dos años.



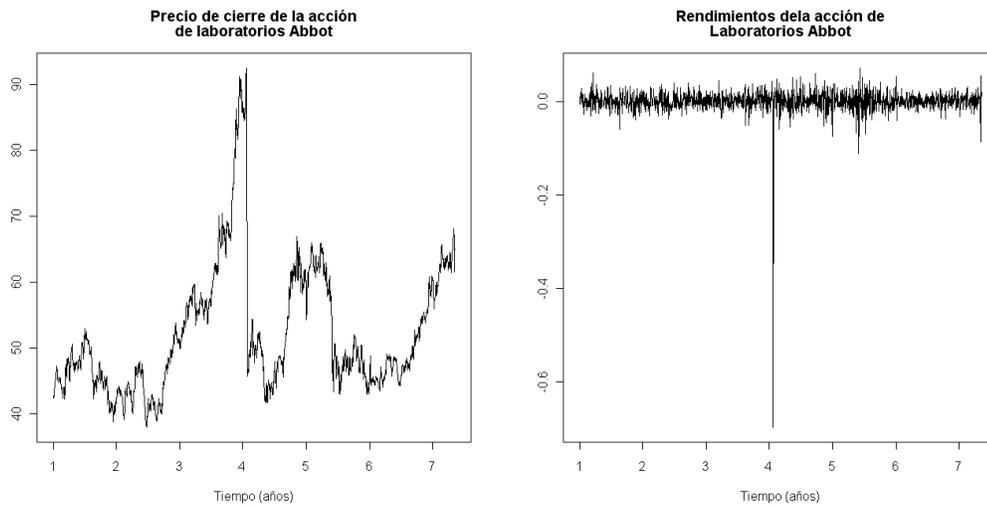
(a) Ajuste del modelo a los rendimientos (b) Densidad $SGED(-0.0195, 0.0434, 3.0000, 0.7247)$.

Figura 4.12: Ajuste del modelo estimado a los rendimientos de IBM.

Precio de la acción de Laboratorios Abbot Laboratorios Abbot es una compañía de Illinois (Estados Unidos) dedicada al desarrollo y comercialización de productos farmacéuticos entre los que se destaca una amplia gama de medicamentos para el tratamiento del VIH. Las gráficas del precio, sus rendimientos y el histograma de los rendimientos se presentan en la figura 4.13. Sobre los rendimientos se ha superpuesto la densidad normal ajustada a los datos, que indica que la curtosis de los datos es mayor que la de una distribución normal. Note que en la figura 4.13-(b) se observa un salto cerca del comienzo del cuarto año. Dicho salto no es un dato atípico, sino que hace parte de una secuencia de saltos que la serie del precio de Laboratorios Abbot muestra desde su salida a la bolsa hasta la fecha. La información utilizada para la estimación del modelo corresponde al periodo comprendido entre el primero de enero de 2001 y el 30 de agosto de 2007, para un total de 1653 observaciones.

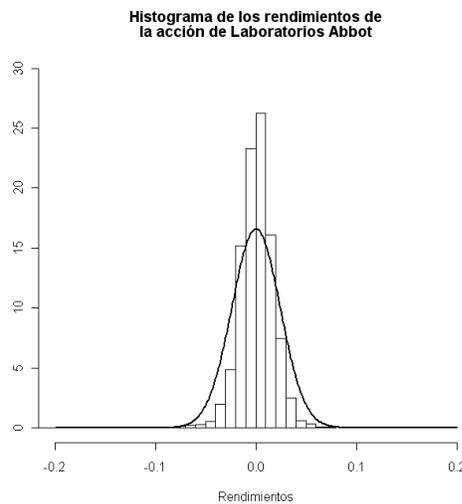
Como puede observarse en la figura 4.14 la serie no muestra correlación serial significativa. Sin embargo, como en el caso de la acción de IBM, la prueba BDS rechaza la hipótesis nula de que la serie es una sucesión iid. La prueba de estacionaridad KPSS, en cambio, no rechaza la hipótesis nula de que la serie es estacionaria en la media. Para esta prueba el valor-p fue superior a 0.1.

Los resultados de la estimación para la acción de Laboratorios Abbot se presentan en la tabla 4.2.2. Aunque los errores estándar son grandes, en algunos casos del orden de magnitud de los estimadores, todos los parámetros son estadísticamente significativos al 99 %. La figura 4.15-(a) muestra el ajuste del modelo a los datos y la figura 4.15-(b) muestra la densidad estimada de las



(a) Precio

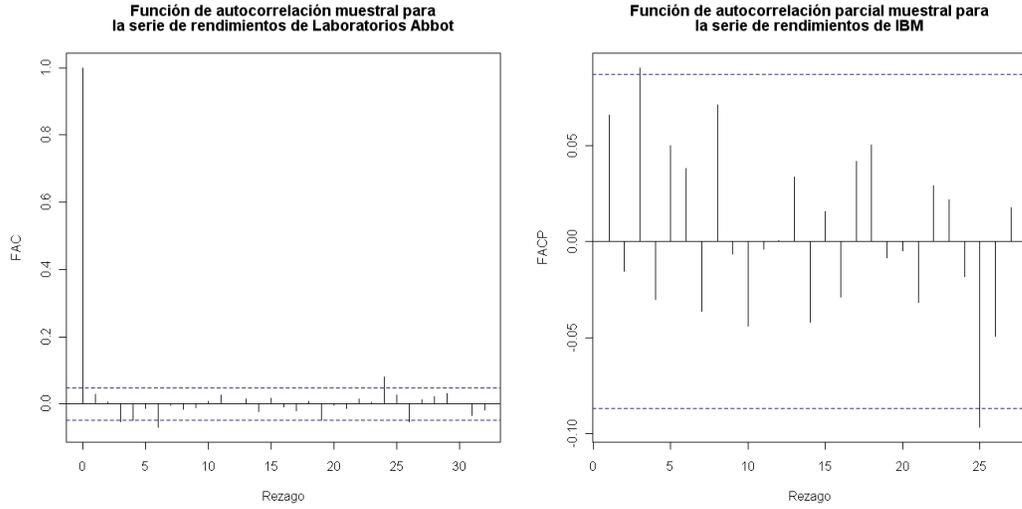
(b) Rendimientos



(c) Histograma de los rendimientos

Figura 4.13: Series del precio y rendimientos e histograma de los rendimientos de la acción de Laboratorios Abbott.

severidades del proceso subyacente a los saltos. Puede apreciarse que en este caso la media de la distribución SGED subyacente a los saltos no es cero. Además, esta distribución tiene la mayor parte de su masa concentrada en una región a la derecha de la media, aunque tiene una parte importante de su cola a la izquierda del cero. Esto es parte de lo que se puede apreciar en la figura 4.13-(b). El negativo de la log-verosimilitud para la estimación de este modelo es $-\ln(L) = -4447.318$.



(a) Función de autocorrelación muestral para los rendimientos de Laboratorios Abbott. (b) Función de autocorrelación parcial muestral para los rendimientos de Laboratorios Abbott.

Figura 4.14: Funciones de autocorrelación y autocorrelación parcial de los rendimientos de la acción de IBM.

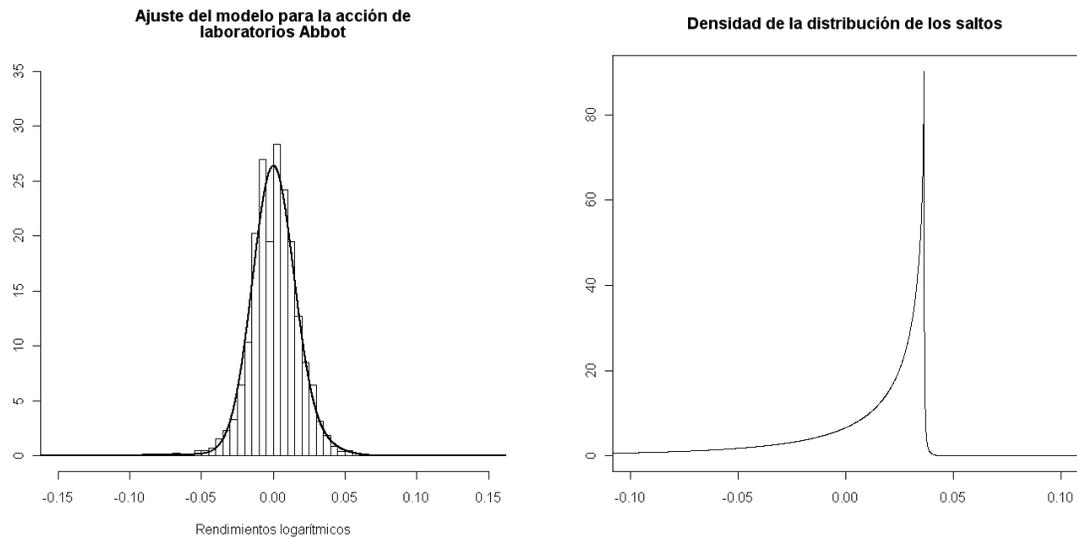
	λ	μ	σ	μ_q	σ_q	ν	ξ
Valor estimado	17.7277	0.0459	0.2322	0.0011	0.0587	0.4581	0.1486
Error estándar	4.2880	0.0229	0.0068	0.0133	0.0205	0.1736	0.5971

Tabla 4.11: Resultados para la acción de Laboratorios Abbott.

4.3. Estimación por máxima verosimilitud asumiendo que el proceso de saltos es Poisson

Cuando $\lambda\Delta t$ no es pequeño la aproximación Bernoulli-Poisson deja de ser adecuada. Esta sección pretende examinar una posible metodología para aproximar la densidad del proceso de difusión con saltos SGED en este caso. Resulta conveniente pensar en el proceso como se expresa en la ecuación (1.58). Al igual que en el caso anterior, la densidad del proceso se escribe utilizando las propiedades de las densidades condicionales, como una suma ponderada por los términos (valores de probabilidad) de una distribución Poisson de parámetro $\lambda\Delta t$. La densidad del proceso $\Delta Y(t)$ se puede escribir como en la ecuación (4.21), donde se define $f_{Q_0} := \delta_{0,x}$ y $f_{Q_i} = f_{Q_i}$, la densidad SGED, para $i = 1, 2, \dots$

$$f_{\Delta Y}(x) = \left[f_{\Delta D} * \sum_{p=0}^{+\infty} \frac{e^{-\lambda\Delta t} (\lambda\Delta t)^p}{p!} (f_{Q_0} * f_{Q_1} * \dots * f_{Q_p}) \right] (x). \quad (4.21)$$



(a) Ajuste del modelo a los rendimientos (b) Densidad $SGED(-0.0195, 0.0434, 3.0000, 0.7247)$.

Figura 4.15: Ajuste del modelo estimado a los rendimientos de Laboratorios Abbot.

Evaluar esta función de densidad requiere, en la práctica, truncar la serie. En (Ball y Torous, 1985) se presenta argumentos que muestran que, en general, se puede truncar la serie en $p = 10$ y para valores pequeños de $\lambda\Delta t$ mucho antes, sin sacrificar demasiado la precisión. Con esta función de densidad se puede escribir la verosimilitud de los datos de la misma manera que en la ecuación (4.20). En la ecuación 4.21 aparece una Poisson compuesta. Este tipo de densidades aparecen frecuentemente en actuaría y se utilizan para modelar riesgos colectivos y de esta manera calcular pagos de primas de seguros. En este contexto, la variable aleatoria Poisson compuesta, que es una suma aleatoria de variables aleatorias llamadas severidades, considera que las variables que se suman sólo toman valores positivos. Bajo esta consideración se ha desarrollado métodos importantes para aproximar la densidad de la Poisson compuesta, como el algoritmo de Panjer presentado en (Panjer, 1981). Ejemplos sobre como se aborda el problema en actuaría se presentan en (Daykin, Pentikäinen, y Pesonen, 1994) y (Klugman, Panjer, y Willmot, 2004). En esta última referencia se toca el problema de severidades que pueden tomar valores negativos y se aconseja recurrir a métodos numéricos y de transformación (del tipo de Fourier, por ejemplo) para obtener la densidad vía convolución, como se ilustra en el trabajo pionero de Embrechts et. al., (Embrechts, Grübel, y Pitts, 1993).

Esta solución no ha sido muy explotada en la literatura y para severidades SGED no hay referencia alguna. Así pues, una continuación de este trabajo podría pasar por el diseño de una metodología para aproximar numéricamente la densidad de una Poisson compuesta con severidades SGED. Esta labor no es trivial, requiere estudiar cuidadosamente el problema y es en sí misma un tema interesante de investigación para continuar este trabajo.

Conclusiones

En este trabajo se ha presentado un modelo de difusión con saltos que siguen una distribución SGED, que puede servir como herramienta para modelar la asimetría y curtosis de los precios de activos financieros.

También se ha desarrollado una metodología para aproximar numéricamente la densidad y se ha probado esta metodología con datos simulados observando buenos resultados. Como producto de este trabajo quedan varios programas en R para simular modelos de difusión con saltos y aproximar la verosimilitud del proceso asumiendo la aproximación Bernoulli-Poisson. La dirección en que podría continuar este trabajo es en la de extender la metodología propuesta cuando no se asume dicha aproximación.

En cuanto al problema de optimización de la log-verosimilitud, se ha presentado evidencia de que no es un problema trivial y que se requiere de técnicas de optimización avanzadas, cuando la información acerca de los valores de los parámetros del modelo que se desea estimar es escasa. Para destacar, están los resultados obtenidos con la evolución diferencial. Estos algoritmos, aunque computacionalmente son costosos por la cantidad de evaluaciones de la función objetivo que requieren, tienen un desempeño superior respecto a los métodos convencionales, aunque al contrario de estos sus bases teóricas son escasas y no hay garantía alguna de que la solución a la que lleguen sea el óptimo global.

El aspecto más importante a destacar del uso de la Evolución Diferencial es que, para llegar a resultados buenos, no requiere especificar una condición inicial cercana a la verdadera solución. Esto representa una ventaja importante, sobre todo cuando el conocimiento *a priori* que se tiene de la solución es pobre.

El estudio de simulación mostró que los parámetros se pueden relacionar entre ellos mismos en formas muy complejas y diferentes. Esta situación representa una dificultad para los algoritmos de búsqueda basados en descenso por gradiente (métodos quasi Newton), que la Evolución Diferencial pudo superar en este problema específico.

Acerca del costo computacional de la metodología propuesta, hay que advertir que sin la capacidad de cómputo con la que se cuenta hoy en día, sería impensable plantear un método que, para cada evaluación de la función objetivo, requiera utilizar la FFT varias veces. No obstante, es necesario desarrollar metodologías que hagan uso de la capacidad de cómputo disponible y aún más presionen el avance de ésta, tal y como se ha hecho en este trabajo.

Las aplicaciones a datos reales muestran que el modelo propuesto tiene la capacidad de aproximar ciertas series financieras y por tanto se puede convertir en una herramienta para econometría. En particular, el modelo propuesto probó aproximar bien la alta curtosis y la asimetría. Otra ventaja que puede aprovecharse para el trabajo econométrico, es que el método permite tener la densidad aproximada de los rendimientos. Esto facilitaría el cálculo de cuantiles y percentiles, como por ejemplo el valor en riesgo (*Value At Risk* - VAR).

Así pues, con lo planteado y desarrollado aquí queda disponible una nueva metodología para estimar procesos de difusión con saltos que se constituye en un paso intermedio hacia otros trabajos de mayor complejidad.

Una de estas posibilidades es la extensión de la metodología cuando no se asume la aproximación Bernoulli-Poisson. Otra, es extender la metodología para procesos con un grado de complejidad mayor, como por ejemplo aquellos cuyos coeficientes son funciones y no constantes. También, la metodología en sí misma puede mejorarse explorando métodos de interpolación y discretización diferentes, que establezcan criterios para el tamaño de paso y el número de puntos a utilizar distintos a los que se ha propuesto aquí. El problema de optimización sigue siendo un campo de trabajo interesante y se requiere de métodos para determinar la región en la que los algoritmos de optimización deben buscar los estimadores.

Apéndice A

Operaciones genéticas y mecanismos de diferenciación

A.1. Operaciones genéticas

A continuación se describe los componentes de un algoritmo evolutivo. Esta clase de algoritmos son conducidos por operaciones, denominadas genéticas, un criterio de aptitud y un mecanismo de selección. Posteriormente se presentan algunas de las estrategias más comunes en evolución diferencial.

Replicación La replicación o clonación lleva a cabo una copia idéntica de los individuos más aptos, aunque no necesariamente estos son los únicos que se reproducen (Whitley, 1993), pues algunos investigadores en el área señalan como conveniente mantener la diversidad de la población en cada generación. Por ejemplo, si la solución de un problema se puede codificar como una cadena binaria de cinco bits, la replicación de los individuos 11010 y 10001 actúa como sigue: $R(I_i) = (I_i)$, es decir, que el individuo I_i presente en la generación k se copia a la generación $k + 1$ sin sufrir cambios. Así que $R(11010) = (11010)$ y $R(10001) = (10001)$.

Cruzamiento El cruzamiento es una operación que combina dos individuos de la generación k para producir dos individuos de la generación $k + 1$. En su forma más general (cruce simple), la operación de cruzamiento incluye la selección de un único punto de la cadena en el cual ambos individuos se separarán. Por ejemplo, el cruzamiento entre los individuos 11010 y 10001 a partir del tercer gen puede representarse como $C(11010, 10001, 3) = (11001, 10010)$. En esta operación es importante tener cuidado con la selección del gen que determina el punto de cruce, pues dentro de un esquema de codificación se debe respetar la agrupación de características, es decir, que si los primeros tres genes de la cadena representan en forma compacta una característica de la solución, separar la cadena en el segundo gen no tiene que tener necesariamente un efecto positivo en la

creación de mejores individuos.

La razón de cruzamiento es el porcentaje de la población a la cual se le aplicará el cruzamiento. Esta razón está relacionada con el tamaño del espacio de búsqueda, dado que si aumenta, el espacio de búsqueda también lo hace, pero si es muy alta se puede perder gran cantidad de tiempo de cómputo explorando regiones no promisorias dentro y hasta por fuera del espacio de búsqueda.

Mutación La mutación es una operación que requiere de mucho cuidado pues en ella se realizan cambios generalmente aleatorios que pueden introducir ruido en la siguiente generación. Por otro lado, es una operación que contribuye en gran medida al sostenimiento de la diversidad en las futuras generaciones. En esta operación se escoge uno o varios puntos de la cadena para mutar y, en el caso de los algoritmos genéticos (AG), la mutación consiste en la inversión del bit. Por ejemplo, la operación de mutación del individuo 11010 en la cuarta posición se escribe como $M(11010) = (11000)$.

La mutación sirve para explorar genes que no estaban incluidos en la población inicial o para recuperar genes perdidos en el proceso evolutivo con el fin de explorarlos en un nuevo contexto.

A.1.1. La aptitud

La forma en que cada individuo responde al problema se mide con la aptitud (*fitness*). La aptitud se deriva de la función objetivo de tal manera que se eviten problemas debidos a la escala de ésta. Esto implica que la función de aptitud no es necesariamente la función objetivo (o problema objetivo), sino que puede ser un resultado a partir de ésta para propósitos de la selección. Los dos métodos más utilizados en la literatura para determinar la aptitud son el escalamiento y el ordenamiento (*scaling* y *ranking*, en inglés, respectivamente). En el escalamiento se usa una función que mapea de alguna forma el valor de la función objetivo para cada individuo. Su principal intención es mitigar la disminución de la varianza de los valores de la función objetivo para los individuos de cada generación a medida que el proceso evolutivo transcurre.

El ordenamiento, en cambio, hace caso omiso de la disminución de la varianza antes mencionada y organiza los individuos de cada generación de acuerdo a algún criterio, como puede ser organizar los individuos de mayor a menor valor de la función objetivo. Generalmente no asigna un valor cuantitativo de aptitud a cada individuo.

A.1.2. La selección

La selección es el proceso que dirige la búsqueda estocástica y puede llevarse a cabo de muchas maneras. La forma en que se lleva a cabo este proceso requiere un gran cuidado puesto que una política de selección que favorezca la permanencia de los individuos más aptos y parecidos entre sí puede llevar a una pérdida de la diversidad y por consiguiente a la convergencia prematura a una

solución no deseada, que es uno de los problemas más comunes en los algoritmos de optimización, en especial de los AE.

Los mecanismos de selección más utilizados son: la ruleta y el torneo. El primero consiste en asignarle a cada individuo una probabilidad de supervivencia, la cual se calcula como el valor de la aptitud del individuo dividido entre la sumatoria de los valores de la aptitud de todos los individuos presentes en la población. Luego de esto se genera un número aleatorio r entre 0 y 1 y se comienza a elegir individuos aleatoriamente hasta que la suma de sus probabilidades de supervivencia sea mayor al número r , el último individuo elegido es el seleccionado. Este proceso se repite hasta completar la población necesaria para aplicar las operaciones genéticas.

Por otra parte, en el torneo se busca realizar competencias entre grupos de individuos (usualmente parejas) para elegir al individuo con mayor aptitud de cada grupo. En general, estos grupos se eligen de manera aleatoria, dándole así posibilidad de ser seleccionados a los individuos con bajos valores de aptitud. El elitismo es una política utilizada para garantizar que el individuo con mayor aptitud (superindividuo), o un grupo de ellos, de cada generación sea copiado en la siguiente generación, en (Spall, 2003) se presenta este concepto como una condición para garantizar la convergencia de un AG.

A.2. Estrategias de diferenciación

La operación principal en los algoritmos de evolución artificial es la diferenciación. Esta es la operación que define la creación de los nuevos individuos. Dentro de los algoritmos evolutivos tradicionales la diferenciación equivale a un cruzamiento entre varios individuos simultáneamente y un proceso de mutación. Las estrategias de diferenciación están enfocadas a crear nuevos individuos ω como $\omega = \beta + F\delta$, donde β es el individuo base, F es la constante de diferenciación y δ es un vector de diferencias. Esto es similar en apariencia a la manera en que se genera una nueva solución en los métodos de descenso por gradiente. En estos métodos δ representa un avance en la dirección en que se llega al óptimo. El individuo base, β , se puede tomar aleatoriamente o puede ser el mejor individuo de la generación actual. Para cada individuo x^j de la población se crea un nuevo individuo ω , con el que se recombina x^j para generar un nuevo individuo x^{jn} . A la siguiente generación pasa el mejor entre x^j y x^{jn} , de acuerdo con la función objetivo. A continuación se describen las cinco estrategias implementadas en el paquete *DEoptim* de R, (Ardia, 2008). Estas y otras estrategias más recientes pueden consultarse en (Feoktistov, 2006), así como también la interpretación y el desempeño de cada una de ellas.

Rand/1 En esta estrategia, para cada individuo x^j , se selecciona aleatoriamente tres individuos x^{j1} , x^{j2} y x^{j3} , mutuamente diferentes entre sí y distintos de x^j . El individuo base es $\beta = x^{j1}$ y el vector de diferenciación es $\delta = x^{j2} - x^{j3}$ y así el nuevo individuo es $\omega = x^{j1} + F(x^{j2} - x^{j3})$.

Rand/2 En esta estrategia, para cada individuo x^j se selecciona aleatoriamente cinco individuos x^{j_1}, \dots, x^{j_5} , mutuamente diferentes entre sí y distintos de x^j . El nuevo individuo se genera como $\omega = x^{j_5} + F(x^{j_1} + x^{j_2} - x^{j_3} - x^{j_4})$.

Best/1 En la estrategia best/1, el individuo base es el mejor individuo de la generación actual x^b . Para cada individuo x^j de la población en la generación actual, se selecciona aleatoriamente dos individuos x^{j_1} y x^{j_2} , mutuamente distintos entre sí y diferentes de x^j . El nuevo individuo se construye como $\omega = x^b + F(x^{j_1} - x^{j_2})$.

Best/1 Análogo a la estrategia Rand/2 pero con x^b , el mejor individuo de la generación actual, como individuo base. En esta estrategia se crea el nuevo individuo como $\omega = x^b + F(x^{j_1} + x^{j_2} - x^{j_3} - x^{j_4})$. Aquí también x^{j_1}, \dots, x^{j_4} son mutuamente diferentes entre sí y distintos de x^j .

Rand-to Best/1 Para cada individuo de la población se selecciona aleatoriamente tres individuos x^{j_1}, x^{j_2} y x^{j_3} , escogidos como en las otras estrategias. El individuo base es $\beta = x^b$, el mejor individuo de la generación actual. El individuo de prueba se construye como $\omega = x^j + \lambda(x^b - x^{j_1}) + F(x^{j_2} - x^{j_3})$.

Bibliografía

- ALTMANN, T., T. SCHMIDT, Y W. STUTE (2004): “Jump Diffusion Processes with Shot Noise Effects and their Applications to Finance,” Documento de Trabajo, University of Giessen, Germany.
- ARDIA, D. (2008): *DEoptim package: Differential Evolution Optimization in R.R* package version 1.3-0.
- ARNOLD, T. (2006): “Using GMM to flatten the option volatility smile,” *Research in International Business and Finance*, 20, 1–21.
- BALL, C. A., Y W.Ñ. TOROUS (1983): “A Simplified Jump Process for Common Stock Returns,” *The Journal of Financial and Quantitative Analysis*, 18(1), 53–65.
- (1985): “On Jumps in Common Stock Prices and Their Impact on Call Option Pricing,” *The Journal of Finance*, 40(1), 155–173.
- BECKERS, S. (1981): “A note on estimating the parameters of the jump-diffusion model of stock returns,” *Journal of financial and quantitative analysis*, 16, 127–140.
- BELISLE, C. J. P. (1992): “Convergence theorems for a class of simulated annealing algorithms,” *Rd. Journal of Applied Probability*, 29, 885–895.
- BLACK, F., Y M. SCHOLES (1973): “The Pricing of Options and Corporate Liabilities,” *The Journal of Political Economy*, 81(3), 637–654.
- BOLLERSLEV, T., Y H. ZHOU (2002): “Estimating stochastic volatility diffusion using conditional moments of integrated volatility,” *Journal of Econometrics*, 109, 35–65.
- BOROVKOVA, S., Y F. J. PERMANA (2006): *Stochastic Finance. Cap. Modelling electricity prices by the potential jump-diffusion*, pp. 239–263. Springer US.
- BROOCK, W. A., J. A. SCHEINKMAN, W. D. DECHERT, Y B. LEBARON (1996): “A test for independence based on the correlation dimension,” *Econometric Reviews*, 15(3), 197–235.

- BURTON, M., Y T. S. RICHARDSON (2001): “Multimodality of the likelihood in the bivariate seemingly unrelated regressions model,” Documento de Trabajo, Department of Statistics. Washington University.
- BYRD, R. H., P. LU, J. ÑOCEDAL, Y C. ZHU (1995): “A limited memory algorithm for bound constrained optimization,” *SIAM Journal of Scientific Computing*, 16, 1190–1208.
- CARTEA, Á., Y M. G. FIGUEROA (2005): “Pricing in electricity markets: a mean reverting jump diffusion model with seasonality,” dans *Birkbeck working papers in economics and finance*, . Cap. BWPEF 0507. University of London.
- CHACKO, G., Y L. M. VICEIRA (2003): “Spectral GMM estimation of continuous-time processes,” *Journal of Econometrics*, 116, 259–292.
- CRAINE, R., L. A. LOCHSTOER, Y K. SYRTVEIT (2000): “Estimation of a stochastic-volatility Jump-Diffusion Model,” *Revista de análisis económico*, 15(1), 61–87.
- DAYKIN, C., T. PENTIKÄINEN, Y M. PESONEN (1994): *Practical Risk Theory for Actuaries*. Chapman & Hall.
- DOORNIK, J. A., Y M. OOMS (2003): “Multimodality in the GARCH Regression Model,” Documento de Trabajo, Nuffield College. University of Oxford.
- EMBRECHTS, P., R. GRÜBEL, Y S. PITTS (1993): “Some applications of the fast Fourier transform algorithm in insurance mathematics,” *Statistica Neerlandica*, 47(1), 59–75.
- FEOKTISTOV, V. (2006): *Differential evolution, in search of solutions*, vol. 5 of *Optimization and its applications*. Springer Verlag, New York.
- FERNANDES, M. (2006): “Financial crashes as endogenous jumps: estimation, testing and forecasting,” *Journal of Economic Dynamics & Control*, 30, 111–141.
- FERNANDEZ, C., Y M. STEEL (1996): “On Bayesian modelling of fat tails and skewness,” Discussion Paper 58, Tilburg University, Center for Economic Research.
- FLEMING, P., Y R. PURSHOUSE (2002): “Evolutionary algorithms in control systems engineering: a survey,” *Control Engineering Practice*, 10, 1223–1241.
- FLETCHER, R., Y C. M. REEVES (1964): “Function minimization by conjugate gradients,” *Computer Journal*, 7, 148–154.
- GALLANT, A., Y G. TAUCHEN (1996): “Which moments to match?,” *Econometrics Theory*, 12, 657–681.

- GILLI, M., Y P. WINKER (2007): *Handbook of Computational Econometrics*. Cap. Heuristic Optimization Methods in Econometrics. Elsevier, Amsterdam.
- GOFFE, W. L., F. G. D., Y J. ROGERS (1994): “Global optimization of statistical functions with simulated annealing,” *Journal of Econometrics*, 60(1-2), 66–99.
- GOURIEROUX, C., Y A. MONTFORT (1996): *Simulation-Based Econometric Methods*. Oxford University Press. Oxford. UK.
- HANSON, F. B. (2006): *Applied stochastic processes and control for jump-diffusions: modeling, analysis and computation*. Society for industrial and applied mathematics, University of Illinois. Chicago, Illinois, USA.
- HERNÁNDEZ, J. A., Y J. D. OSPINA (2009): *Estimación de parámetros en modelos de volatilidad estocástica usando algoritmos evolutivos*. Centro de Publicaciones, Universidad Nacional de Colombia - Sede Medellín, Medellín, 1 edn.
- HO, M. S., W. R. M. PERRAUDIN, Y B. E. SORENSEN (1996): “A Continuous-Time Arbitrage Pricing Model with Stochastic Volatility and Jumps,” *Journal of Business & Economic Statistics*, 14(1), 31–43.
- HOLLAND, J. (1992): *Adaptation In Natural and Artificial Systems*. The University of Michigan Press, 2° edn.
- HONORÉ, P. (1988): “Pitfalls in Estimating Jump-Diffusion Models,” Documento de Trabajo, Centre for Analytical Finance, University of Aarhus Aarhus School of Business.
- HORNIK, K. (2009): *Time series analysis and computational finance*. R package version 0.10-17.
- HUISMAN, R., Y M. RONALD (2001): “Regime jumps in electricity prices,” Erim report series, Erasmus Research Institute of Management (ERIM), Rotterdam School of Management / Faculteit Bedrijfskunde, Erasmus Universiteit Rotterdam, P.O. Box 1738.
- ITÔ, K. (1951): “On stochastic differential equations,” *Memories of American Mathematical Society*, 4, 1–51.
- JIANG, G. J., Y R. C. OOMEN (2007): “Estimating Latent Variables and Jump Diffusion Models Using High-Frequency Data,” *Journal of Financial Econometrics*, 5(1), 1–30.
- KIEFER, N. (1978): “Discrete parameter variation: Efficient estimation in diffusion process,” *Econometrica*, 46(2), 427–434.
- KLUGMAN, S. A., H. H. PANJER, Y G. E. WILLMOT (2004): *Loss Models, From Data to Decisions*. Wiley, 2 edn.

- KOU, S. (2002): “A jump-diffusion model for option pricing,” *Management Science*, 48(8), 1086–1101.
- KWIATKOWSKI, D., P. C. B. PHILLIPS, P. SCHMIDT, Y Y. SHIN (1992): “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root,” *Journal of Econometrics*, 54, 159–178.
- MCFADDEN, D. (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models with Numerical Integration,” *Econometrica*, 57, 995–1026.
- MEBANE, W. R., JR., Y J. SINGH SEKHON (2007): *rgenoud: R version of GENetic Optimization Using Derivatives* R package version 5.4-7.
- MERTON, R. C. (1973): “Theory of Rational Option Pricing,” *The Bell Journal of Economics and Management Science*, 4(1), 141–183.
- (1976): “Option pricing when underlying stock returns are discontinuous,” *Journal of Financial Economics*, 3, 125–144.
- MEYER, M. C. (2003): “An Evolutionary Algorithm With Applications to Statistics,” *Journal of Computational and Graphical Statistics*, 12(2), 1–17.
- MORENO, M., P. J. SERRANO, Y W. STUTE (2008): “Statistical Properties and Economic Implications of Jump-Diffusion Processes with Shot-Noise Effects,” Business Economics Working Papers wb084912, Universidad Carlos III, Departamento de Economía de la Empresa.
- NOCEDAL, J., Y S. WRIGHT (1999): *Numerical Optimization*. Springer Verlag, New York.
- PAKES, A., Y D. POLLARD (1989): “Simulation and the Aysmptotics of Optimization Estimators,” *Econometrica*, 57, 1027–57.
- PANJER, H. (1981): “Recursive evaluation of a family of compound distributions,” *Astin Bulletin*, 12(1), 22–26.
- PINHEIRO, J., D. BATES, S. DEBROY, D. SARKAR, Y THE R CORE TEAM (2008): *nlme: Linear and Nonlinear Mixed Effects Models* R package version 3.1-90.
- PLANCHET, F., Y P.-E. THEROND (2005): “Límpact de la prise en compte des sauts boursiers dans les problématiques d’assurance,” dans *15th AFIR Colloquium*.
- POSCH, O. (2007): “Structural estimation of jump-diffusion processes in macroeconomics,” CREATES Research Papers 2007-23, School of Economics and Management, University of Aarhus.

- PRICE, K. V. (1994): “Genetic annealing,” *Dr. Dobb’s Journal*, pp. 127–132.
- PRICE, K. V., Y R. STORN (1997): “Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces,” *Journal of Global Optimization*, 11(4), 341–359.
- RAMEZANI, C., Y Y. ZENG (1998): “Maximum likelihood Estimation of Asymmetric Jump-Diffusion Processes: Application to Security Prices,” *Tech Report*.
- (2007): “Maximum likelihood estimation of the double exponential jump-diffusion process,” *Annals of Finance*, 3(4), 487–507.
- RANJAN DAS, S., Y R. K. SUNDARAM (1999): “Of Smiles and Smirks: A Term Structure Perspective,” *The Journal of Financial and Quantitative Analysis*, 34(2), 211–239.
- ROCKINGER, M., Y M. SEMENOVA (2005): “Estimation of Jump-Diffusion Process vis Empirical Characteristic Function,” FAME Research Paper Series rp150, International Center for Financial Asset Management and Engineering.
- SHIMIZU, Y., Y N. YOSHIDA (2006): “Estimation of parameters for diffusion processes with jumps from discrete observations,” *Statistical Inference for Stochastic Processes*, 9, 227–277.
- SPALL, J. C. (2003): *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley and Sons, New York.
- SRIVASTAVA, A., U. GRENANDER, G. R. JENSEN, Y M. I. MILLER (2002): “Jump-diffusion Markov processes on orthogonal groups for,” *Journal of Statistical Planning and Inference*, 103, 15–37.
- TEAM, R. D. C. (2008): *R A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- THEODOSSIOU, P. (2000): “Skewed Generalized Error Distribution of Financial Assets and Option Pricing,” Available at SSRN: <http://ssrn.com/abstract=219679> or DOI: 10.2139/ssrn.219679.
- WERON, R. (2006): *Modelling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley Finance Series. John Wiley & Sons, Ltd., England.
- WERON, R., M. BIERBRAUER, Y S. TRÜCK (2004): “Modeling electricity prices: jump diffusion and regime switching,” *Physica A*, 336, 39–48.
- WHITLEY, D. (1993): “A Genetic Algorithm Tutoria,” Documento de Trabajo, Computer Science Department, Colorado State University.

- (2001): “An overview of evolutionary algorithms: practical issues and common pitfalls,” Publicado como notas de trabajo en Computer Science Department, Colorado State University.
- (2002): “Genetic algorithms and evolutionary computing,” Publicado como notas de trabajo en Computer Science Department, Colorado State University.
- WUERTZ, D., Y Y. CHALABI (2008): *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*R package version 280.75.
- WUERTZ, D., Y. CHALABI, Y L. LUKSAN (2004): “Parameter Estimation of ARMA Models with GARCH/APARCH Errors. An R and SPlus Software Implementation,” *Journal of Statistical Software*, 10.
- YU, J. (2007): “Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese Yuan,” *Journal of Econometrics*, 141(2), 1245–1280.
- ZIVOT, E., Y J. WANG (2006): *Modeling Financial Time Series with S-Plus*. Springer Verlag, New York.