

Estimación Bayesiana para la Volatilidad de la Tasa de Intercambio del Nuevo Sol versus Dólar Americano

por

Luis Alberto Navarro Huamaní

Disertación sometida al cuerpo docente de la Escuela Profesional de Ingeniería Estadística de la Universidad Nacional de Ingeniería como parte de los requisitos necesarios para obtener el título profesional de Licenciado en Estadística

Area de concentración: Estadística

Universidad Nacional de Ingeniería
Escuela Profesional de Ingeniería Estadística

Agradecimientos

- a Víctor Valdivieso Benavides asesor de esta tesis por el apoyo, paciencia y dedicación en la ejecución de este trabajo;
- a los amigos Miguel Melchor, Daniel Huamani, Hector Altamirano y colegas de la EPIES-FIECS de la Universidad Nacional de Ingeniería por el apoyo y amistad;
- al Núcleo de Estadística Computacional (NEC, <http://nec.uni.edu.pe>), EPIE - FIECS por haber facilitado la ejecución de este trabajo;
- al comite del IX Congreso Latinoamericano de Probabilidad y Estadística Matemática (CLAPEM), Punta del Este - Uruguay (22 al 26 de marzo del 2004), por la aceptación del presente trabajo de tesis a participar de dicho congreso y por las sugerencias.
- A Dios por haber iluminado mi alma y por la fuerza en los momentos mas difíciles.

RESUMEN

Un algoritmo de simulación estocástica iterativa Metropolis-Hasting por componentes es desarrollada para propósitos de inferencia estadística Bayesiana sobre el modelo GARCH(1,1) t-Student univariada. Esta estrategia representa una alternativa diferente a las propuestas anteriores de MCMC. Los resultados derivados de un estudio de simulación demuestran el buen desempeño de la propuesta de solución. La volatilidad de la tasa de intercambio del Nuevo Sol versus el Dólar Americano descrita por un proceso GARCH(1,1) t-Student es estimada a través de éste algoritmo MCMC.

Índice General

1	Introducción	1
1.1	Series Financieras	2
1.2	Distribución de los Retornos	4
1.3	Volatilidad	6
1.4	Perfil de la Tesis	6
2	Procesos GARCH	7
2.1	Proceso ARCH univariado	10
2.2	Proceso GARCH univariado	13
3	Estadística Bayesiana	14
3.1	Naturaleza de la Inferencia Bayesiana	15
3.1.1	Teorema de Bayes	16
3.1.2	Determinación a priori	19
3.1.3	Inferencia Estadística Bayesiana	26
3.1.4	Inferencia Bayesiana a través del simulador estocástico MCMC - (Solución via Remuestreo)	27
3.1.5	Gibbs Sampling	30
3.1.6	El algoritmo de Metropolis-Hastings	31
4	Volatilidad de Tasa de intercambio del Nuevo Sol versus Dólar Americano	37
4.1	Aproximación Bayesiana para GARCH(1,1) t-Student	38
4.1.1	Metropolis-Hasting por componentes para GARCH(1,1) t-Student	39
4.1.2	Diagnóstico de Convergencia	40

5	Conclusiones, Recomendaciones y Sugerencias	51
A	Algunas Funciones de Densidades	53
A.1	Normal Univariada	53
A.2	Normal Multivariada	53
A.3	t-Student Multivariada	53
A.4	Gamma - 1era parametrización	54
A.5	Gamma - 2da parametrización	54
A.6	Gamma Inversa	54
B		55
C	Ley de esperanzas iteradas	57

Índice de Figuras

1.1	<i>Evolución de tasas de intercambio del Nuevo sol versus Dólar americano . .</i>	2
1.2	<i>Evolución de los retornos de las tasas de intercambio del Nuevo sol versus Dólar americano</i>	3
3.1	<i>función de verosimilitud para diferentes valores de y</i>	16
3.2	<i>inferencia via Muestreo</i>	28
3.3	<i>Inferencia via muestreo por componentes</i>	33
4.1	<i>ploteos CUMSUM para estimativa de la media w (datos simulados)</i>	41
4.2	<i>ploteos CUMSUM para estimativa de la media α (datos simulados)</i>	42
4.3	<i>ploteos CUMSUM para estimativa de la media β (datos simulados)</i>	42
4.4	<i>ploteos CUMSUM para estimativa de la media ν (datos simulados)</i>	43
4.5	<i>Histograma de w (datos simulados)</i>	43
4.6	<i>Histograma de α (datos simulados)</i>	44
4.7	<i>Histograma de β (datos simulados)</i>	44
4.8	<i>Histograma de ν (datos simulados)</i>	45
4.9	<i>ploteos CUMSUM para estimativa de la media w (datos reales)</i>	46
4.10	<i>ploteos CUMSUM para estimativa de la media α (datos reales)</i>	46
4.11	<i>ploteos CUMSUM para estimativa de la media β (datos reales)</i>	47
4.12	<i>ploteos CUMSUM para estimativa de la media ν (datos reales)</i>	47
4.13	<i>Histograma de w (datos reales)</i>	48
4.14	<i>Histograma de α (datos reales)</i>	48
4.15	<i>Histograma de β (datos reales)</i>	49
4.16	<i>Histograma de ν (datos reales)</i>	49

Índice de Tablas

4.1	<i>Descripción de la serie de retornos de la tasa de intercambio del Nuevo Sol versus el Dólar americano</i>	38
4.2	<i>Estimación clásica y Bayesiana para un modelo GARCH(1,1) t-Student (datos artificiales)</i>	45
4.3	<i>Estimación clásica y Bayesiana para un modelo GARCH(1,1) t-Student (datos reales)</i>	50

Capítulo 1

Introducción

Toda actividad económica está sujeta a riesgo, sea de un negocio propio o de factores externos. Con la globalización de la economía los inversionistas pueden, por ejemplo, comprar dólares en el mercado alemán de monedas y vender en el mercado americano, exigiendo de ellos un conocimiento de la situación económica, política y social de esos países para que sean concientes de los posibles riesgos. Debido a la complejidad del mercado, las instituciones están sujetas no sólo a nuevos riesgos por actuar en nuevos mercados, sino también al aumento de riesgos a los cuales ellos se exponen, en la medida en que haya una significativa elevación de las volatilidades de esos mercados. Volatilidad es uno de los conceptos mas importantes hoy en día en finanzas a nivel mundial y el mercado Bursátil limeño no puede ignorar estas nuevas tecnologías y seguir la tendencia mundial hacia una administración eficiente de riesgos.

El presente trabajo es una propuesta metodológica, ver Navarro [14], de inferencia estadística Bayesiana en el modelamiento de volatilidad para un único instrumento financiero. Durante la ejecución del proceso de inferencia estadística Bayesiana sobre el modelo GARCH(1,1) t-Student, usado para describir volatilidad, el algoritmo de simulación estocástica de Metropolis-Hastings por componentes es propuesto para resolver el problema computacional de inferencia. Los resultados alcanzados usando datos artificiales generados por un modelo GARCH(1,1) t-Student nos proporciona la consistencia del algoritmo propuesto para después aplicarlo a un conjunto de datos reales de tasa de intercambio del Nuevo Sol versus el Dólar americano en un periodo determinado de análisis.

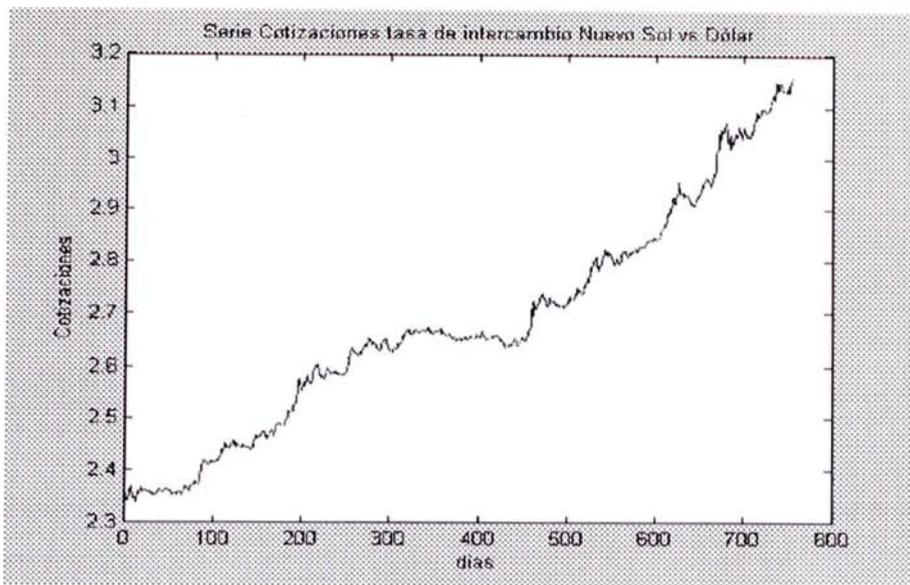


Figura 1.1: *Evolución de tasas de intercambio del Nuevo sol versus Dólar americano*

1.1 Series Financieras

Activos, tasas de intercambio, tasa de interés, etc. son negociados en los diversos mercados financieros mundiales (bolsas de valores), y la secuencia de estos valores negociados de cada uno de estos "instrumentos financieros" conforman una *Serie financiera*. Si se intenta modelar las series financieras con las metodologías tradicionales de procesos lineales, veremos que no son capaces de reproducir características peculiares que estas series de tiempo presentan. Estas características son conocidas como *Hechos Estilizados (Stylized Facts)*. En las últimas tres décadas ha habido un cambio significativo de esta situación, con el desenvolvimiento de modelos probabilísticos más complejos que capturen las propiedades de estas series financieras, así como de esquemas de simulación estocásticas numéricas para efectuar un análisis estadístico completo de estos modelos.

El objetivo principal en el análisis de estas series financieras es la interpretación y pronósticos a través de un proceso de inferencia estadística. Existe una preocupación en entender la evolución de este tipo de series de tiempo. Es decir, entender el comportamiento de los precios de los diversos instrumentos financieros. Uno de los aspectos de mayor interés para los inversionistas se refiere a pronosticar alguna característica en particular de una serie financiera o conocer algunas características particulares de la distribución del estimador bajo análisis. Por lo tanto, los modelos propuestos deben reproducir los *hechos estilizados* y hacer buenos pronósticos.

Los precios de los productos negociados en los mercados financieros son determinados por las condiciones del mercado (oferta y demanda). Los días en que el mercado permanece

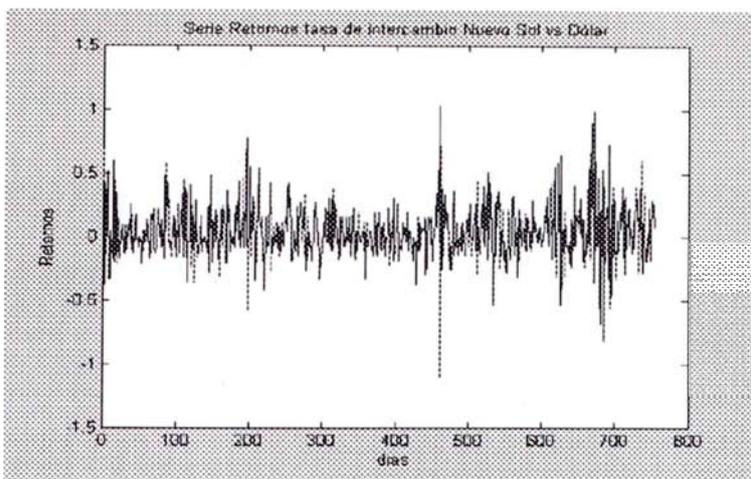


Figura 1.2: *Evolución de los retornos de las tasas de intercambio del Nuevo sol versus Dólar americano*

abierto son llamados de *Trading Days* y los restantes *non Trading Days*. Al mes tenemos usualmente 22 días de funcionamiento del mercado y al año 252 días. Generalmente, los modelos teóricos de series de tiempo al analizar la evolución diaria de las series financieras, consideran como indexador del tiempo los días en que existe movimiento en el mercado y como observaciones perdidas (*missing*) los días sin movimiento .

Una característica común en las series financieras es el de *volatilidad*. Informalmente podemos definir la volatilidad como el grado de variabilidad que presentan los cambios en los precios de algún instrumento financiero en particular o cartera de instrumentos financieros, y que es interpretada como una *medida de riesgo* asociada a cada producto o cartera. Por ejemplo, en la figuras 1.1 y 1.2 tenemos la evolución de las tasas de intercambio del Nuevo sol versus el Dólar americano y los retornos correspondientes, ambos variando en el tiempo y es posible identificar períodos para los cuales las varianzas varían en el tiempo (volatilidad)

Los modelos que se ocupan de cuantificar la volatilidad son conocidos como *Modelos de volatilidad* y entre ellos destacan dos modelos estadísticos: Modelos ARCH/GARCH¹ y sus extensiones, y los modelos de Volatilidad Estocástica. Los modelos de Volatilidad Estocástica no son actualmente muy usados por la comunidad financiera ya que se encuentran en una etapa de investigación prometedora como una opción a los modelos ARCH/GARCH. Dentro de la familia ARCH univariado el mas utilizado es el proceso GARCH(1,1) univariado.

¹ *Autoregressive Conditional Heteroscedasticity* y *Generalized Autoregressive Conditional Heteroscedasticity* respectivamente

1.2 Distribución de los Retornos

En principio debemos establecer el tipo de serie de tiempo que va a ser usada para nuestro análisis y lo que queremos de ella. Es decir, nuestro análisis se basará en un tipo de serie financiera denominada de *Retornos* y el interés recaerá en el concepto de Volatilidad. Los retornos serán obtenidos a partir de los precios de los instrumentos financieros (o cotizaciones), de la siguiente manera:

Retorno.- es la variación del precio (por ejemplo, precio de un activo, tasas de intercambio, etc.) ocurrida en un determinado periodo de tiempo. Tal variación puede ser calculada en términos absolutos, relativos, porcentuales o logarítmicas. Calculado en escala logarítmica, el retorno de un instrumento financiero R_t es dado por

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (1.1)$$

donde P_t es el precio en t

A partir de ahora, los retornos serán calculados usando datos diarios, a menos que se especifique lo contrario. La ventaja de la especificación de retorno dada por la ecuación (1.1), es que los retornos logarítmicos se aproximan satisfactoriamente a la variación relativa de precios, para grandes cantidades de datos, que es el caso al momento de usar datos diarios.

El argumento para usar R_t en la forma dada por la ecuación (1.1), se basa en el siguiente análisis: se sabe que $\ln(1+x) \approx x$, cuando $x \rightarrow 0$. Para datos diarios es razonable suponer que $(P_t/P_{t-1}) \rightarrow 1$. Entonces si consideráramos $x = (P_t/P_{t-1}) - 1$, tenemos:

$$\ln\left(\frac{P_t}{P_{t-1}}\right) \approx \left(\frac{P_t}{P_{t-1}}\right) - 1 = \frac{(P_t - P_{t-1})}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}} \quad (1.2)$$

donde $\frac{\Delta P_t}{P_{t-1}}$ es la variación relativa. Luego, la especificación logarítmica aproxima satisfactoriamente la variación relativa de precios, justificando su utilización. Usaremos esta formulación para el análisis de nuestra serie de tiempo financiera.

La derivación de la ecuación (1.2) fue hecha partiendo de que R_t fue aproximada por una serie de Taylor. Si $r_t = \left(\frac{P_t}{P_{t-1}}\right) - 1$, entonces $R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(1+r_t)$. Luego, $R_t = r_t - \frac{1}{2}r_t^2 + \frac{2}{3}r_t^3 - \dots$. Como r_t es pequeño, podemos concluir que $R_t \simeq r_t$ y así obtener finalmente la ecuación (1.2). Esto no necesariamente es cierto para mercados altamente volátiles.

Una ventaja al usar la ecuación (1.1) es que el logaritmo natural es mas realista, pues si los retornos R_t siguen una distribución Normal, entonces no se obtienen precios negativos. Esto pasa si hacemos que la cola izquierda de la distribución Normal de $\ln(\frac{P_t}{P_{t-1}})$ tienda a $-\infty$, lo que llevaría a que $\frac{P_t}{P_{t-1}} \rightarrow 0$ ó $P_t \rightarrow 0$. Por otra parte, si hacemos que la cola izquierda de la distribución Normal de $R_t = \frac{(P_t - P_{t-1})}{P_{t-1}}$ tienda a $-\infty$, conllevaría a que $(\frac{P_t}{P_{t-1}} - 1) < -1$ ó $P_t < 0$. Economicamente éste resultado no tiene sentido.

Otra ventaja de usar el retorno definido por la ecuación (1.1) es que permite una descomposición en multiperiodos debido a las propiedades matemáticas de la función logaritmo. Así, es posible calcular el retorno mensual a partir de datos diarios simplemente sumando los retornos diarios. Por ejemplo, considere R_{t-2} el retorno acumulado de los dos últimos días. El retorno puede ser descompuesto en

$$R_{t-2} = \ln\left(\frac{P_t}{P_{t-2}}\right) = \ln\left(\frac{P_t}{P_{t-1}}\right) + \ln\left(\frac{P_{t-1}}{P_{t-2}}\right) = R_t + R_{t-1} \quad (1.3)$$

El problema de la distribución de probabilidad a ser adoptado para los retornos es objeto de investigación en los últimos años. El modelo paramétrico a ser adoptado debe partir de un presupuesto o suposición acerca de la distribución de los retornos y la verdad no existen resultados concluyentes respecto a este tema.

Un modelo *sensato* a ser usado, para explicar las series financieras, debe ser uno que incorpore lo mas aproximadamente posible características propias o hechos estilizados de las series financieras.

Por ejemplo, un hecho estilizado es el exceso de curtosis de las series financieras, tema que será visto en detalle en el capítulo 4. El hecho que las series sean bastantes volátiles, conlleva a que los retornos asuman valores extremos con razonable frecuencia. Traduciendo este hecho en términos probabilísticos, la distribución de probabilidad empírica tendrá colas anchas ² ocasionando exceso de curtosis o leptocurtosis. La distribución Normal, generalmente usada para modelar retornos, no tiene exceso de curtosis. La verdad su curtosis es 3, que es el valor de referencia y llamada de mesocúrtica. Cualquier distribución con curtosis mayor que 3 será llamada de leptocúrtica (exceso de curtosis). Por otro lado, cualquier distribución con curtosis menor que 3, será llamada de platicúrtica.

²*fat tails*

1.3 Volatilidad

Existen algunos aspectos mas o menos genéricos de las series de retornos financieros que permite determinar algunas características inherentes a la distribución a ser adoptada. Se sabe que las series de retornos financieras tienen considerable dispersión con respecto a la media. Además, se observa que la magnitud de tal dispersión asume un comportamiento variante en el tiempo.

Ahora, si consideramos a la varianza como la medida de esta dispersión con respecto a la media, lo que es bastante razonable, podemos decir que las series financieras presentan varianza variante en el tiempo. Esta propiedad es designada como volatilidad, y caracteriza un hecho estilizado de las series financieras que explicaremos a través de la familia ARCH/GARCH en particular. Debemos recalcar que existen otras formas alternativas de modelar volatilidad que no serán tratadas en éste trabajo.

1.4 Perfil de la Tesis

La tesis esta estructurada en 5 capítulos. En el capítulo 2, presentamos a los procesos GARCH(1,1) univariados que serán usados en ésta tesis para explicar el comportamiento de volatilidad de la serie de retornos de la tasa de intercambio del Nuevo Sol versus el Dólar americano en un determinado periodo de análisis. En el capítulo 3, describiremos el paradigma Bayesiano así como su uso en un proceso de inferencia estadística. Algunos problemas presentados en la ejecución de un proceso de inferencia estadística Bayesiana son resueltos a través de ciertos esquemas de simulación iterativos. En el capítulo 4, una estrategia de solución de Metropolis-Hasting por componentes será desarrollada para resolver el problema de inferencia estadística Bayesiana usando un proceso GARCH(1,1) t-Student. Finalmente en el capítulo 5 recomendaciones, conclusiones y sugerencias resumen el trabajo de la presente tesis.

Capítulo 2

Procesos GARCH

Un análisis de series de tiempo para modelar movimientos estocásticos de varianzas ocupa un lugar importante en la toma de decisiones en diversas instituciones financieras en la necesidad de cuantificar los riesgos de diversos instrumentos financieros. Dos clases de modelos destacan en la interpretación de las características de estas series financieras: los modelos de la familia ARCH/GARCH y los modelos de volatilidad estocástica. Estos últimos son en general más robustos que los de la familia ARCH/GARCH. Sin embargo, los modelos ARCH/GARCH tienen ventajas computacionales que los convierte en herramientas de mayor uso actualmente por la comunidad financiera.

El interés central en este capítulo es la familia GARCH. Un hecho interesante de esta estructura es el modelamiento del segundo momento condicional. Engle [7] sugiere modelar los primeros y segundos momentos de manera conjunta. Existen diversas formulaciones que extienden la propuesta original. Sin embargo, como es citado por Engle [7] las varianzas condicionales deben depender de los elementos del conjunto de información disponible de manera autorregresiva.

Modelos lineales

Antes de examinar los modelos ARCH/GARCH no lineales¹, definiremos brevemente lo que significa un modelo lineal. Un requerimiento necesario es que el estimador cuadrado medio mínimo (ECMM)² de observaciones futuras deben ser combinaciones lineales de observa-

¹la no linealidad es reflejada en momentos de mayor orden tal como la varianza

²el ECMM es definido como

$$\hat{y}_{T+l|T} = E[y_{T+l}|Y_T]$$

ciones muestrales. Sin embargo, linealidad en los pronósticos de la media no es suficiente para que muchos modelos exhiban efectos no lineales.

Si se asume una estructura univariada por simplicidad, una definición débil de linealidad es el requerimiento de que el modelo debe ser escrito como

$$y_t = \sum_{j=0}^{t-1} \psi_{jt} \epsilon_{t-j} + \lambda_{0t} \quad t = 1, 2, \dots, T \quad (2.1)$$

donde ψ_{jt} 's son pesos no estocásticos, λ_{0t} denota condiciones iniciales observables, que pueden ser fijos o aleatorios, y los ϵ_t 's es una secuencia de variables aleatorias con media cero.

Ley de esperanzas iteradas

Un resultado importante para analizar modelos no lineales es la *Ley de esperanzas iteradas*, la base estadística esta en el apéndice C. En el presente contexto esto significa que la esperanza de la observación actual, o una función de ella $g(y_t)$, con respecto a la información en el tiempo $t - J$ puede ser hallado primero tomando la esperanza condicional sobre la información en el periodo anterior $t - 1$, luego tomando la esperanza condicional sobre la información en el periodo anterior y así sucesivamente hasta $t - J$. Es decir,

$$E[g(y_t)|Y_{t-J}] = E_{t-J}[g(y_t)] = E_{t-J} \dots E_{t-1}[g(y_t)]$$

la esperanza incondicional es hallada haciendo que J tienda a ∞ , en cuyo caso el subíndice será dejado de lado. Si el proceso es considerado a iniciarse en $t = 0$, la esperanza dado la condición inicial es dado por $J = t$.

Ruido blanco

El ruido blanco (RB) es definido como una secuencia de variables aleatorias no correlacionadas con media y varianza constante. Contrario a lo que uno espera, es posible hacer pronósticos no triviales de series que tiene la propiedad RB. Únicamente si las variables son independientes es imposible hacer uso de valores actuales y pasados para pronosticar cualquier hecho del futuro. El ruido blanco *estricto* (RBE) es una secuencia de variable aleatorias independientes e idénticamente distribuidos (IID). Observe que si un RB es gaussiano, debe ser también RBE.

La relevancia de la distinción entre RB y RBE es que las observaciones generadas por un modelo no lineal posiblemente tengan la propiedad de RB. Aunque posiblemente concluyamos, de manera correcta, que no hay forma de ensayar un modelo lineal, existe la posibilidad de ensayar con un modelo no lineal para hacer pronósticos.

Diferencia martingale

Si se asume que $E|y_t| < \infty$, un proceso con la propiedad *Diferencia martingale* es tal que

$$E_{t-1}[y_t] = E(y_t|Y_{t-1}) = 0$$

donde Y_{t-1} denota toda la información hasta el tiempo $t - 1$. Se sigue a partir de la ley de esperanza iteradas que la esperanza incondicional de y_t es cero,

$$E[y_t] = E[E(y_t|Y_{t-1})] = 0$$

además y_t es no correlacionada con cualquier función de observaciones pasadas $f(Y_{t-1})$. Esto es porque

$$E[y_t f(Y_{t-1})|Y_{t-1}] = f(Y_{t-1})E[y_t|Y_{t-1}] = 0$$

se sigue de esto que la esperanza incondicional de $y_t f(Y_{t-1})$ es también cero

Modelando volatilidad

Muchas series financieras tales como retornos de stocks y tasas de intercambio exhiben cambios en la varianza a través del tiempo. Estos cambios tienden a ser serialmente correlacionados, con grupos de observaciones de alta volatilidad ocurriendo en el tiempo. Existen varias formas de modelar cambios en la varianza. La configuración básica usada en esta tesis es considerar a la serie como generada por una secuencia de variables aleatorias independientes e idénticamente distribuido ϵ_t , con varianza unitaria, multiplicado por un factor $\sqrt{h_t}$ ³, el desvío estándar.

$$y_t = \epsilon_t h_t^{1/2} \quad \epsilon_t \sim IID(0, 1)$$

la aproximación a ser desarrollada por la presente tesis es que $\sqrt{h_t}$ sea modelada por un proceso GARCH(1,1) en la cual la varianza es modelada en términos de observaciones pasadas y desvíos padrones pasados.

³Aquí se asume que $\sigma_t = \sqrt{h_t}$

2.1 Proceso ARCH univariado

En esta parte explicaremos brevemente el origen de la postulación de los modelos ARCH para luego extenderlos para un modelo GARCH. Si una variable aleatoria y_t es generada a partir de la densidad condicional $f(y_t|y_{t-1})$ el pronóstico, dado la información pasada, es dado por $E(y_t|y_{t-1})$. La varianza de este pronóstico es dado por $V(y_t|y_{t-1})$. Esta última expresión establece que la varianza condicional depende de la información pasada y tal vez se comporte como una variable aleatoria. Sin embargo, en los modelos econométricos tradicionales, la varianza condicional no depende de y_{t-1} . Una nueva clase de modelos fue propuesta por Engle [7], donde la varianza condicional depende de la información pasada y que será discutida en breve.

Considere un proceso autorregresivo de primer orden:

$$y_t = \gamma y_{t-1} + \epsilon_t$$

donde ϵ_t es ruido blanco con $V(\epsilon_t) = \sigma^2$. La media condicional de y_t es γy_{t-1} y la media incondicional es cero. Una cierta mejora en términos de pronósticos aparece si se usa esta media condicional. La varianza condicional de y_t es σ^2 , y la varianza incondicional es $\frac{\sigma^2}{1-\gamma^2}$. Para procesos reales es interesante tener intervalos de los pronósticos que dependan de la información pasada, dado que ella puede afectar la varianza del pronóstico. Así lo mejor sería tener una clase de modelos mas general

Una aproximación clásica de heteroscedasticidad es introducir una variable exógena x_t que ayude a pronosticar la varianza. Con una media a priori de cero el modelo tendrá la siguiente forma:

$$y_t = \epsilon_t x_t$$

donde $V(\epsilon_t) = \sigma^2$. La varianza de y_t es dado por $\sigma^2 x_{t-1}^2$ y el intervalo de pronóstico dependerá de la evolución de la variable exógena. Es imposible encontrar una solución a este problema de variación en la varianza, dado que esto implica encontrar las causas de esta variación. Una aproximación alternativa es reconocer que las medias y varianzas condicionales evolucionan conjuntamente en el tiempo.

Un modelo que permite a la varianza condicional depender sobre la información pasada de la serie de tiempo es el modelo Bilineal propuesto por Granger y Andersen. Un caso simple es:

$$y_t = \epsilon_t y_{t-1}$$

La varianza condicional ahora es dada por $\sigma^2 y_{t-1}^2$. Sin embargo, la varianza incondicional es cero o infinito, lo que hace que esta formulación no tenga atractivo a pesar que este problema puede ser controlable a través de algunas modificaciones.

Un modelo que contorna los problemas anteriores fue sugerido por Engle [7], en donde la parametrización de la varianza condicional permite modelar los segundos momentos no observables. Es decir, permite que la varianza condicional dependa de elementos del conjunto de información disponible de una manera autorregresiva. Sea ψ_{t-1} el conjunto de información disponible dado por los valores pasados de y_t . El modelo ARCH univariado puede ser escrito de la siguiente forma:

$$\begin{aligned} y_t &= \epsilon_t h_t^{1/2} \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 \end{aligned}$$

con $V(\epsilon_t) = 1$. Este modelo será llamado de Autoregressive Conditional Heteroscedasticity —ARCH— de orden uno o ARCH(1). Este modelo no es exactamente un modelo Bilineal, pero es mucho más próximo de él. Si se aumenta la suposición de Normalidad para ϵ_t , el modelo se convierte en:

$$\begin{aligned} y_t | \psi_{t-1} &\sim N(0, h_t) \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 ; \alpha_0 > 0 , \alpha_1 \geq 0 \end{aligned} \quad (2.2)$$

restricciones en los parámetros son necesarios para garantizar que h_t sea positivo. Una varianza condicional alta en el periodo de tiempo anterior puede fácilmente generar una observación cerca de cero, con el resultado que la varianza condicional actual puede ser relativamente pequeña. En general uno espera que la varianza condicional actual cambie levemente. Esto indica la necesidad a dispersar la memoria del proceso sobre un número de observaciones pasadas. Es decir, cuanto más observaciones pasadas sean consideradas mejor. Una expresión más general es

$$\begin{aligned} y_t | \psi_{t-1} &\sim N(0, h_t) \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_p y_{t-p}^2 \end{aligned}$$

este modelo es llamado de ARCH de orden p , o ARCH(p).

Un modelo ARCH regresivo es obtenido suponiendo que la media de y_t es dado por $x_t \beta$, una combinación lineal de variables endógenas y exógenas incluidas en x_t con β un vector

de parámetros desconocidos. Es decir,

$$\begin{aligned} y_t | \psi_{t-1} &\sim N(x_t \beta, h_t) \\ h_t &= h(\epsilon_{t-1}, \dots, \epsilon_{t-p}, \alpha) \\ \epsilon_t &= (y_t - x_t \beta) \end{aligned}$$

Propiedades del proceso ARCH(1)

El modelo ARCH es una DM y su media incondicional es cero y es serialmente no correlacionada. La varianza incondicional puede ser hallada usando la ley de esperanzas iteradas.

$$E_{t-2} E_{t-1}(y_t^2) = E_{t-2}[\alpha_0 + \alpha_1 y_{t-1}^2] = \alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 y_{t-2}^2$$

repetiendo esta operación hasta que la última expresión es tomada en $t - J$ da

$$E_{t-J} \dots E_{t-1}(y_t^2) = \alpha_0 + \alpha_0 \alpha_1 + \alpha_0 \alpha_1^2 + \dots + \alpha_0 \alpha_1^{J-1} + \alpha_1^J y_{t-J}^2$$

si $\alpha_0 < 1$, haciendo $J \rightarrow \infty$ y sumando como una progresión geométrica tenemos que

$$Var(y_t) = E(y_t^2) = \alpha_0 / (1 - \alpha_1)$$

el proceso ARCH(1) es por lo tanto un RB, pero no un RBE. Además, si se supone que la distribución condicional es Normal, incondicionalmente no es distribución Normal. Ella no podría tener una distribución incondicional Normal pues se trataría de un modelo lineal. La distribución incondicional obtenida a partir de este modelo ARCH(1) no tiene una forma conocida, pero es posible probar que es simétrica, ver Engle [7]. Es posible también probar que si $3\alpha_1^2 < 1$, la curtosis es dada por $3(1 - \alpha_1^2)/(1 - 3\alpha_1^2)$. Esto es mayor que 3 para α_1 positivo, y así el modelo ARCH(1) produce observaciones con colas más pesadas que las de una distribución Normal. Esto es una propiedad deseada pues muchas series financieras son observadas a tener colas pesadas.

Cuando consideramos pronósticos, la esperanza condicional de cualquier observación futura es cero, pero aplicando la ley de esperanzas iteradas es posible probar que el error cuadrático medio (ECM) del pronóstico ⁴, el cual es justamente la varianza condicional de

⁴El ECM del pronóstico $\hat{y}_{T+l|T}$ es dado por

$$ECM(\hat{y}_{T+l|T}) = Var(y_{T+l}|Y_T) + [\hat{y}_{T+l|T} - E[y_{T+l}|Y_T]]^2$$

la observación futura es

$$ECM(\hat{y}_{T+l|T}) = Var(y_{T+l}|Y_T) = \alpha_0(1 + \alpha_1 + \alpha_1^2 + \dots + \alpha_1^{l-1}) + \alpha_1^l y_T^2$$

Si la serie fuese tratada como si fuesen RB, el pronóstico es dado por la varianza incondicional $Var(y_t) = \alpha_0/(1 - \alpha_1)$. La última expresión de la varianza condicional tiende a la expresión de la varianza condicional cuando $l \rightarrow \infty$, pero para un horizonte pequeño esta puede ser muy diferente.

2.2 Processo GARCH univariado

Una mejor aproximación que la propuesta de insertar observaciones pasadas a la ecuación de la varianza condicional actual, es introducir valores rezagados de h_t . Bollerslev [4] generaliza el proceso ARCH permitiendo que las varianzas condicionales pasadas aparezcan en la ecuación de la varianza condicional. Consecuentemente la ecuación de la varianza condicional se convierte en:

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_p y_{t-p}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \dots + \beta_q h_{t-q}$$

este proceso es llamado de *generalized* ARCH de orden (p, q) , o GARCH(p, q). El modelo particular GARCH(1,1) ofrece una descripción parcimoniosa de los datos, satisfactoria para el modelamiento de muchas series reales (ver Bollerslev [4]).

Propiedades del proceso GARCH(1,1)

Todos los modelos GARCH son DM. Si la suma de los α 's y los β 's es menor que uno, el modelo tiene varianza finita y es RB. En particular para el modelo GARCH(1,1) podemos ver que

$$\begin{aligned} E_{t-2} E_{t-1}(y_t^2) &= E_{t-2} [\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}^2] = \alpha_0 + (\alpha_1 + \beta_1) h_{t-1}^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) [\alpha_0 + \beta_1 y_{t-2}^2 + \beta_1 h_{t-2}^2] \end{aligned}$$

repetiendo este proceso hacia el infinito, podemos ver que si $\alpha_1 + \beta_1 < 1$, tenemos que

$$Var(y_t) = \alpha_0 / (1 - \alpha_1 - \beta_1)$$

El proceso de inferencia estadística Bayesiana para este modelo será tratada en el capítulo 4 de la presente tesis.

Capítulo 3

Estadística Bayesiana

El propósito fundamental de un análisis estadístico es hacer inferencias y pronósticos. Datos son coleccionados y necesitamos hacer inferencias sobre uno o mas factores desconocidos relacionados al sistema generador de estos datos, para así pronosticar sobre eventos futuros. Aunque la teoría frecuentista proporcione resultados simples en casos en donde suposiciones especiales puedan ser hechas (tales como Normalidad e independencia de los errores), en otros casos, y particularmente cuando ninguna estadística suficiente exista, las soluciones no son satisfactorias. A pesar que estas suposiciones especiales involucren un número de situaciones de interés científico, es ideal pretender que un conjunto de problemas estadísticos cuya solución fue colocada como necesaria por el investigador coincida con el conjunto de problemas sujetos a un tratamiento padrón. La metodología Bayesiana dá mayor énfasis al interés científico que conveniencia matemática.

Una distribución a priori, que representa toda la información sobre los parámetros desconocidos *antes* que los datos sean disponibles, asume un papel esencial en la aproximación Bayesiana. Tal distribución puede ser usada para representar el conocimiento a priori o ignorancia relativa. En problemas de inferencia científica deberíamos, siempre que fuera posible, llevar a cabo un análisis como si un estado de ignorancia relativa existiese *a priori*. El temor por parte de algunos estadísticos acerca del uso de distribuciones a priori esta asociado con el hecho que la distribución a priori domine y distorcionen la información que los datos traen con respecto a los parámetros de interés.

Una ventaja del paradigma Bayesiano es que él puede ser usado para explorar las consecuencias de cualquier tipo de modelo probabilístico, sin restricción para aquellos que tiene formas matemáticas especiales o convenientes. Para mayor detalle sobre estadística

Bayesiana, ver Box y Tiao [5], O'Hagan [17], Berger [2], etc.

3.1 Naturaleza de la Inferencia Bayesiana

Es de interés presentar a la metodología Bayesiana como una alternativa a la estadística clásica o frecuentista a través del uso del principio de Verosimilitud, principio comunmente adoptado por la comunidad estadística clásica, para luego enlazar de manera natural estas ideas en el teorema de Bayes. El principio de Verosimilitud hace explícita la idea condicional natural que *unicamente* el valor observado \mathbf{y} debe ser relevante a conclusiones o evidencias acerca de θ (se asume que un modelo paramétrico es el proceso generador de ésta observación \mathbf{y} y que además que depende de θ). El concepto clave en el principio de Verosimilitud es el de Verosimilitud.

Definición 3.1.1 Dada la observación \mathbf{y} , la función $l(\theta; \mathbf{y}) = p(\mathbf{y}|\theta)$, considerada como una función de θ , es llamada de función de Verosimilitud

La razón intuitiva para el nombre de función de Verosimilitud se establece a partir de la siguiente idea: un valor de θ para el cual $p(\mathbf{y}|\theta)$ le corresponde el mayor valor con respecto al resto de valores de θ , será el θ mas *plausible* (dada la información \mathbf{y}). Para ilustrar esto consideremos el siguiente experimento: Cierta individuo lanza de manera independiente una moneda dos veces. La probabilidad de que salga cara es θ y la probabilidad de que salga sello es $(1 - \theta)$. Un modelo probabilístico que explique éste fenómeno es dada por un distribución Binomial. Es decir, si se define una variable aleatoria \mathbf{Y} : número de caras en los dos lanzamientos independientes, lo que tenemos es que $\mathbf{Y} \sim \text{Binomial}(2, \theta)$. Entonces la f.d.p es dada por

$$p(\mathbf{y}|\theta) = l(\theta; \mathbf{y}) = \binom{2}{\mathbf{y}} \theta^{\mathbf{y}} (1 - \theta)^{2-\mathbf{y}}, \quad \mathbf{y} = 0, 1, 2. \quad \theta \in \Theta = \langle 0, 1 \rangle$$

note que :

- a.-) si $\mathbf{y} = 0$ entonces $l(\theta; \mathbf{y} = 0) = (1 - \theta)^2$, y el valor mas *plausible* (o verosímil) de θ es 0
- b.-) si $\mathbf{y} = 1$ entonces $l(\theta; \mathbf{y} = 1) = 2\theta(1 - \theta)$, y el valor mas *plausible* de θ es $\frac{1}{2}$
- c.-) si $\mathbf{y} = 2$ entonces $l(\theta; \mathbf{y} = 2) = \theta^2$, y el valor mas *plausible* de θ es 1

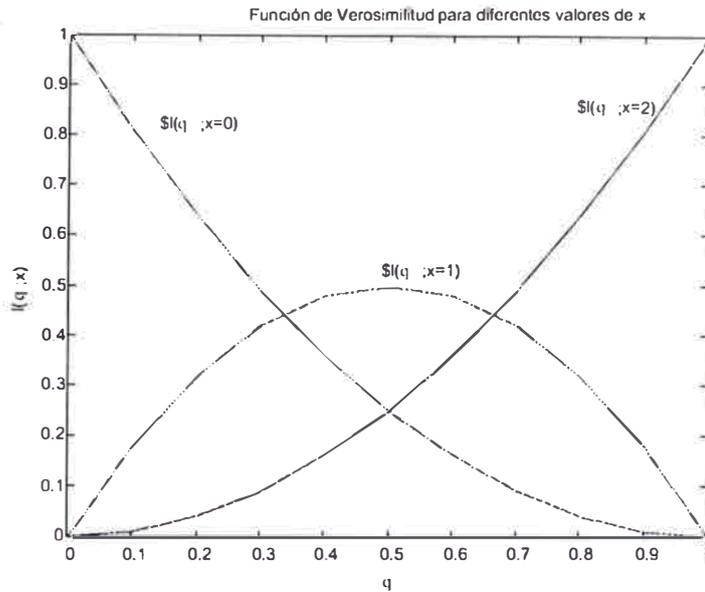


Figura 3.1: función de verosimilitud para diferentes valores de y

esas verosimilitudes son graficadas en la figura (3.1). Observe por ejemplo que la curva de función de verosimilitud para $y=2$ sugiere que el valor más plausible es uno. Esto es, toda información proporcionada por y es absorbida por $l(\theta; y)$. Ideas similares pueden ser obtenidas para los otros casos ($y = 1$ e $y = 0$)

3.1.1 Teorema de Bayes

Suponga que $\underline{y} = (y_1, \dots, y_n)$ es un vector de n observaciones generadas por $\underline{Y} = (Y_1, \dots, Y_n)$ y que $Y = u(\underline{Y})$ es una variable aleatoria cuya distribución de probabilidades $p(\mathbf{y}|\theta)$ depende de un vector de parámetros de dimensión k , digamos $\theta' = (\theta_1, \dots, \theta_k)$. Suponga que θ tiene una distribución de probabilidades cuya función de densidad de probabilidad (f.d.p.) es dada por $p(\theta)$. Entonces, la f.d.p. conjunta de θ e \mathbf{y} es:

$$p(\mathbf{y}|\theta)p(\theta) = p(\mathbf{y}, \theta) = p(\theta|\mathbf{y})p(\mathbf{y}) \tag{3.1}$$

Condicionando a los datos observados \mathbf{y} , la f.d.p. de θ es dada por:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} \tag{3.2}$$

note además que la f.d.p. marginal de los datos es:

$$p(\mathbf{y}) = E[p(\mathbf{y}|\theta)] = c^{-1} = \begin{cases} \int p(\mathbf{y}|\theta)p(\theta)d\theta & \theta \text{ continuo} \\ \sum p(\mathbf{y}|\theta)p(\theta) & \theta \text{ discreto} \end{cases} \tag{3.3}$$

donde la utilización de los operadores suma o integral dependerá de las características del dominio de θ , y $E[f(\theta)]$ es el operador esperanza matemática de $f(\theta)$ con respecto a la

distribución de θ . Así (3.2) puede ser escrita como:

$$p(\theta|\mathbf{y}) = c p(\mathbf{y}|\theta)p(\theta) \quad (3.4)$$

la ecuación (3.2) o equivalentemente (3.4), es comumente conocido como el *teorema de Bayes*, pues θ es una variable aleatoria no observable. En esta expresión $p(\theta)$ representa toda la información acerca de θ sin conocimiento de los datos. $p(\theta)$ es llamada de f.d.p. *a priori* de θ . Por otra parte, $p(\theta|\mathbf{y})$ informa todo sobre θ una vez conocidos los datos. $p(\theta|\mathbf{y})$ es llamada de f.d.p. *a posteriori* de θ dado \mathbf{y} . La cantidad c es una constante de normalización necesaria para garantizar que la distribución a posteriori $p(\theta|\mathbf{y})$ integre (ou sume) 1.

Debemos mencionar en este instante que la *identificación* de la f.d.p a posteriori $p(\theta|\mathbf{y})$ vía el teorema de Bayes según la ecuación (3.2) o (3.4), es a veces complicado y/o imposible determinarlo analíticamente. Métodos de simulación determinística y/o estocásticos deben ser usados para obtener resúmenes de interés de esta f.d.p a posteriori $p(\theta|\mathbf{y})$ de interés. En próximas secciones trataremos de un método particular denominado de Metroplis-Hastings para obtener los resúmenes de nuestro interés.

Teorema de Bayes y la Función de Verosimilitud

Si usamos la *función de verosimilitud* $l(\theta|\mathbf{y})$ dada la información \mathbf{y} , dada por la ecuación (3.4) en vez de $p(\mathbf{y}|\theta)$ en el teorema de Bayes, tenemos que:

$$\begin{aligned} p(\theta|\mathbf{y}) &= c l(\theta|\mathbf{y})p(\theta) \\ &\propto l(\theta|\mathbf{y})p(\theta) \end{aligned} \quad (3.5)$$

Es decir, el teorema de Bayes establece que la distribución de probabilidades de θ posterior a los datos \mathbf{y} , es proporcional al producto de la distribución a priori de θ y de la verosimilitud de θ dado \mathbf{y} . Esto es,

distribución a posteriori \propto verosimilitud \times distribución a priori.

La función de verosimilitud $l(\theta|\mathbf{y})$ destaca en la fórmula de Bayes, a través de ella los datos (denotados por \mathbf{y}) modifican el conocimiento a priori de θ ; ella representa por lo tanto toda la información acerca de θ proporcionado por los datos .

La función de verosimilitud es definida a no ser por una constante multiplicativa, esto es, la multiplicación por una constante produce una verosimilitud invariante en términos

de información. Esto concuerda con el papel que ella asume en la fórmula de Bayes, pues multiplicar la función de verosimilitud por una constante arbitraria no afecta la distribución a posteriori de θ . Apenas el valor relativo de la verosimilitud es importante como el que proporciona la información a la distribución a posteriori de θ . La verdad, nuestro interés es el "núcleo" de la distribución a posteriori $l(\theta|\mathbf{y})p(\theta)$ en la ecuación (3.5) como la que proporciona la información en la fórmula de Bayes.

Naturaleza secuencial del Teorema de Bayes

La ecuación (3.5) nos da una formulación matemática de como la información anterior puede ser combinada con nuevas informaciones. Igualmente, el teorema permite una actualización continua de información acerca de los parámetros θ a medida que nuevas informaciones aparezcan.

Suponga que tenemos una muestra inicial de observaciones $\underline{\mathbf{y}}_1 = (y_1, \dots, y_{n_1})$ y definimos a $\mathbf{y}_1 = u(\underline{\mathbf{y}}_1)$. Entonces, por el teorema de Bayes se sigue que:

$$p(\theta|\mathbf{y}_1) \propto p(\theta)l(\theta|\mathbf{y}_1). \quad (3.6)$$

Suponga ahora que tenemos una segunda muestra de observaciones $\underline{\mathbf{y}}_2 = (y_{n_1+1}, \dots, y_{n_1+n_2})$ obtenida independientemente de la primera muestra (dado θ) y que $\mathbf{y}_2 = u(\underline{\mathbf{y}}_2)$, entonces

$$\begin{aligned} p(\theta|\mathbf{y}_1, \mathbf{y}_2) &\propto p(\theta)l(\theta|\mathbf{y}_1)l(\theta|\mathbf{y}_2) \\ &\propto p(\theta|\mathbf{y}_1)l(\theta|\mathbf{y}_2) \end{aligned} \quad (3.7)$$

la expresión (3.7) tiene la misma forma que (3.6) donde $p(\theta|\mathbf{y}_1)$ (distribución posteriori de θ dado \mathbf{y}_1) asume el papel de la distribución a priori para la segunda muestra. Obviamente este proceso puede ser repetido muchas veces. Para extender estas ideas, imagine que tenemos n observaciones independientes, la distribución a posteriori puede ser re-calculada después de cada nueva observación, así en el m -ésimo estado la verosimilitud asociada a la m -ésima observación es combinada con la distribución posteriori de θ después de $m-1$ observaciones dadas, para obtener la nueva distribución a posteriori:

$$p(\theta|\mathbf{y}_1, \dots, \mathbf{y}_m) \propto p(\theta|\mathbf{y}_1, \dots, \mathbf{y}_{m-1})l(\theta|\mathbf{y}_m) \quad (3.8)$$

donde

$$p(\theta|\mathbf{y}_1) \propto p(\theta)l(\theta|\mathbf{y}_1)$$

Esto es, el teorema de Bayes describe el proceso de aprendizaje a través de la experiencia, además muestra como el conocimiento actual del estado de la naturaleza cambia continuamente a medida que nuevas informaciones son disponibles.

Este proceso de aprendizaje no esta sujeto a la suposición de observaciones independientes. Es decir, podemos establecer un mecanismo de aprendizaje para el caso general. Ver detalles en O'Hagan [17]

3.1.2 Determinación a priori

Este es el único elemento nuevo en el análisis Bayesiano con respecto al análisis clásico o frecuentista. La determinación es en general subjetiva, sin embargo nada impide que datos de experimentos pasados sean utilizados. El único compromiso es que la distribución represente el conocimiento sobre θ antes de ejecutar experimentos. Dos formas alternativas de distribuciones a priori serán presentados en esta sección, que son la distribución a priori no informativa y a priori conjugada. Sin embargo, el lector interesado en otras formas alternativas de distribuciones a priori pueden revisar los libros de Berger [2], O'Hagan [17].

Priori no Informativas

En esta parte formularemos la distribución a priori no informativa e ilustraremos con algunos ejemplos la determinación de distribuciones a posteriori $p(\theta|y)$ a partir de esta distribución a priori $p(\theta)$ no informativa. Este tipo de a priori fue desarrollada por Jeffreys (1961), como una respuesta a los usuarios e investigadores de la estadística clásica que manifiestan que las distribuciones a prioris son arbitrarias y altera conclusiones y que por lo tanto no puede ser aceptada dentro de un procedimiento científico. Este concepto de distribución a priori no informativa parte del hecho de realizar un análisis a partir de un mínimo de información subjetiva a priori. El concepto de mínimo es un concepto relativo a lo que esta contenido en la muestra.

Suponga por un momento que usamos distribuciones uniformes como representantes de situaciones donde no se dispone de información inicial o no se desea usarla; es decir, $p(\theta) \propto cte$ (θ variando en toda la recta), lo que implica no favorecer ninguna valor en particular de θ . Algunas dificultades que origina la formulación anterior son:

- a) La f.d.p. $p(\theta)$ es impropia. Es decir, $\int p(\theta)d\theta \rightarrow \infty$;

b) $\phi = \phi(\theta)$ es una transformación 1 a 1 de θ y si θ tiene distribución uniforme, entonces por transformación de variables aleatorias la f.d.p. de ϕ es

$$p(\phi) = p(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|$$

el raciocinio que conlleva a que $p(\theta) \propto cte$ debería también ser mantenido para que $p(\phi) \propto cte$, lo que no es verdad por b). Lo ideal sería construir una regla en donde $p(\phi)$ no fuese impropia e invariante bajo transformación.

La verdad lo que queremos es que la distribución a posteriori sea propia aún cuando la a priori no lo sea, no dando así eventualmente importancia a esta característica indeseable de la distribución a priori. La clase de distribuciones a priori no informativas propuesta por Jeffreys es invariante bajo transformación y eventualmente impropia. Antes de definir exactamente este tipo de a priori no informativa, necesitamos de la siguiente definición:

Definición 3.1.2 Sea \mathbf{X} una variable aleatoria observacional con f.d.p. $p(x|\theta)$. La medida de información esperada de Fisher de θ a través de \mathbf{X} es definida por

$$I(\theta) = E_{X|\theta} \left[- \frac{\partial^2 \log p(X|\theta)}{\partial \theta^2} \right]$$

Si $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ fuese un vector de parámetros, entonces la matriz de de información esperada de Fisher de \mathbf{X} es dada por

$$I(\theta) = E_{X|\theta} \left[- \frac{\partial^2 \log p(X|\theta)}{\partial \theta \partial \theta'} \right]$$

con elementos $I_{ij}(\theta)$ dados por

$$I_{ij}(\theta) = E_{X|\theta} \left[- \frac{\partial^2 \log p(X|\theta)}{\partial \theta_i \partial \theta_j} \right] \quad i, j = 1, 2, \dots, p$$

La información así definida es una especie de valor medio de la curvatura de la verosimilitud. A medida que es mayor la curvatura mas precisa es la información contenida en la verosimilitud y por lo tanto mayor sera $I(\theta)$. Como la curvatura es negativa, se toma el valor cambiado de ella. La media es tomada con respecto a los valores posibles de las observaciones con pesos dados por la distribución muestral. Cuando no se toma la media lo que tenemos es la información observada de Fisher $J(\theta)$ dado por

$$\left[- \frac{\partial^2 \log p(X|\theta)}{\partial \theta^2} \right]$$

que es una medida local mientras que $I(\theta)$ es una medida global.

Definición 3.1.3 Sea X una variable aleatoria observacional con f.d.p. $p(x|\theta)$. La f.d.p a priori de Jeffreys es definida por

$$p(\theta) \propto [I(\theta)]^{1/2}$$

Para el caso multivariado tenemos que $p(\theta) \propto |\det I(\theta)|^{1/2}$.

Lema 3.1.1 La f.d.p. a priori $p(\theta) \propto [I(\theta)]^{1/2}$ es invariante; es decir, si $\phi = \phi(\theta)$ es una transformación 1 a 1 de θ , entonces $p(\phi) \propto [I(\phi)]^{1/2}$

Este resultado también es válido para el caso multivariado. Los detalles de la propuesta de Jeffreys, así como la demostración del lema anterior pueden ser encontrados en Box y Tiao [5].

A continuación daremos dos definiciones que nos permitirá lidiar con tipos especiales de a prioris no informativas de parámetros de distribuciones observacionales, que son los modelos de localización y de escala respectivamente.

Modelos de Localización X tiene un modelo de localización si existen funciones f y una cantidad θ tal que la distribución de X dado θ satisface que $p(x|\theta) = f(x - \theta)$. Si tenemos este caso, θ es llamado de parámetro de localización

Ejemplo 3.1.1 Sea $(X|\theta) \sim N(\theta, \sigma^2)$ con σ^2 conocido. La f.d.p. de $(X|\theta)$ es dado por

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-0.5 \frac{(x - \theta)^2}{\sigma^2}\right\}$$

que es una función de $(x - \theta)$.

Una estadística importante involucrando la función de verosimilitud es la función Score:

Definición 3.1.4 La función Score de X , denotada por $U(X; \theta)$, es dada por

$$U(X; \theta) = \frac{\partial \log p(X|\theta)}{\partial \theta}$$

En el caso multivariado de un vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_p)$, la función Score es un vector $\mathbf{U}(X; \theta)$ con componentes dados por $U_i(X; \theta) = \frac{\partial \log p(X|\theta)}{\partial \theta_i}$ $i = 1, 2, \dots, p$.

El siguiente lema muestra una forma alternativa de cálculo de la información esperada de Fisher basada en la función Score. La condición de regularidad exigida es básicamente que los operadores de integración y derivada pueden ser aplicados en orden diferentes para una determinada expresión.

Lema 3.1.2 bajo condiciones de regularidad se tiene que

$$I(\theta) = E_{X|\theta} [U^2(X; \theta)]$$

en el caso de un vector de parámetros se tiene que

$$\mathbf{I}(\theta) = E_{X|\theta} [\mathbf{U}(X; \theta)\mathbf{U}'(X; \theta)]$$

Ahora bien si X tiene modelo de localización entonces,

$$\frac{\partial \log p(x|\theta)}{\partial \theta} = \frac{\partial \log f(x - \theta)}{\partial \theta} = \frac{f'(x - \theta)}{f(x - \theta)}$$

donde $f' = \frac{\partial f}{\partial \theta}$.

Por el lema 3.1.2 anterior, $I(\theta) = E_{X|\theta} \left[\left(\frac{f'(x-\theta)}{f(x-\theta)} \right)^2 \right]$. Ahora bien, si hacemos la transformación $s = x - \theta$ se obtiene que

$$I(\theta) = E_S \left(\frac{f'(S)}{f(S)} \right)^2$$

que no depende de θ . Por lo tanto, $I(\theta) = cte$ y la f.d.p. a priori $p(\theta) \propto cte$. Este resultado también es válido cuando θ es un vector.

Modelos de Escala X tiene un modelo de Escala si existen funciones f y una cantidad σ tal que la distribución de X dado σ satisface que $p(x|\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$. Si tenemos este caso, σ es llamada de parámetro de Escala

Ejemplo 3.1.2 Sea $(X|\sigma^2) \sim N(\theta, \sigma^2)$ con θ conocido. La f.d.p. de $(X|\sigma^2)$ ó $(X|\sigma)$ es dado por

$$p(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-0.5 \frac{(x - \theta)^2}{\sigma^2}\right\}$$

que es de la forma $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$.

Ahora bien si X tiene un modelo de escala, entonces:

$$\begin{aligned} \frac{\partial \log p(x|\sigma)}{\partial \sigma} &= \frac{\partial \log [\sigma^{-1} f(x|\sigma)]}{\partial \sigma} \\ &= \frac{\partial [-\log \sigma + \log f\left(\frac{x}{\sigma}\right)]}{\partial \sigma} \\ &= -\frac{1}{\sigma} + \frac{\partial \log f(x/\sigma)}{\partial \sigma} \left(-\frac{x}{\sigma^2}\right) \\ &= -\frac{1}{\sigma} \left[1 + \frac{x f'(x/\sigma)}{f(x/\sigma)} \right] \end{aligned}$$

donde $f' = \frac{\partial f}{\partial \sigma}$.

Luego la medida de información esperada de Fisher es dada por

$$\begin{aligned} I(\sigma) &= \frac{1}{\sigma^2} E_{X|\sigma} \left[\left(1 + \frac{x}{\sigma} \frac{f'(x/\sigma)}{f(x/\sigma)} \right)^2 \middle| \sigma \right] \\ &= \frac{1}{\sigma^2} E \left(1 + U \frac{f'(U)}{f(U)} \right)^2 \end{aligned}$$

la última expresión se obtiene después de hacer uso de la transformación $u = x/\sigma$. En esta expresión la distribución de U no depende σ y de aquí $I(\sigma) = cte * \sigma^{-2}$. La f.d.p a priori no informativa por lo tanto será $p(\sigma) \propto \sigma^{-1}$

Si hacemos uso del lema 2.1.1, tenemos también que $p(\sigma^2) \propto \sigma^{-2}$, pues tenemos una transformación 1 a 1 con $\sigma > 0$

Modelos de Localización-Escala X tiene un modelo de Localización-Escala si existen funciones f y cantidades θ y σ tal que la distribución de X dado (θ, σ) satisface que $p(x|\theta, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$. Si tenemos este caso, θ es llamado de parámetro de localización y σ es llamado de parámetro de Escala.

Ejemplo 3.1.3 Sea $(X|\theta, \sigma^2) \sim N(\theta, \sigma^2)$. La f.d.p. de $(X|\theta, \sigma^2)$ ó $(X|\theta, \sigma)$ es dado por

$$p(x|\theta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-0.5 \frac{(x-\theta)^2}{\sigma^2}\right\}$$

que es de la forma $\frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$.

Una vez que la forma de $p(x|\theta, \sigma)$ es de la forma $\frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$ es fácil probar que $p(\theta, \sigma) \propto \frac{1}{\sigma}$.

Priori Conjugadas

Aquí también definiremos la distribución a priori conjugada e ilustraremos con algunos ejemplos la determinación de distribuciones a posteriori $p(\theta|y)$ a partir de esta distribución a priori $p(\theta)$ conjugada. En general la distribución marginal $p(y)$ y distribución a posteriori $p(\theta|y)$ no son fáciles de calcular. Si por ejemplo, $y \sim N(\theta, \sigma^2)$ y $\theta \sim Cauchy(\mu, \beta)$ entonces la f.d.p. a posteriori $p(\theta|y)$ puede ser evaluado únicamente de manera numérica. Una gran parte de la literatura Bayesiana se dedica a hallar distribuciones a posteriori para el cual $p(\theta|y)$ pueda ser fácilmente calculado. Uno de estas formas de a priori son las llamadas priori conjugada.

Definición 3.1.5 Sea $F = \{f(y|\theta); \theta \in \Theta\}$ la clase de f.d.p observacionales. Se dice que una clase P de f.d.p. a priori es familia conjugada a F , si $p(\theta|y)$ pertenece siempre a P para todo $f \in F$ y $p \in P$

Ejemplo 3.1.4 Sean $(Y|\theta) \sim N(\theta, \sigma^2)$ con σ^2 conocido, y $\theta \sim N(\mu, \tau^2)$ con μ y τ^2 conocidos. Estamos interesados en probar que $P = \{N(\mu, \tau^2); \mu \in \mathfrak{R}, \tau > 0\}$, es conjugada a $F = \{N(\theta, \sigma^2); \theta \in \mathfrak{R}\}$. Para ello necesitamos probar que $p(\theta|y) \in P$. Por el teorema de Bayes tenemos que:

$$\begin{aligned} p(\theta|y) &\propto l(\theta; y)p(\theta) \\ &\propto \exp \left[-\frac{1}{2\sigma^2}(y - \theta)^2 - \frac{1}{2\tau^2}(\theta - \mu)^2 \right] \\ &\propto \exp \left[-\frac{\theta^2}{2\sigma^2} - \frac{\theta^2}{2\tau^2} + \frac{y\theta}{\sigma^2} + \frac{\mu\theta}{\tau^2} \right] \\ &\propto \exp \left[-\frac{\theta^2}{2}(\sigma^{-2} + \tau^{-2}) + \theta(\sigma^{-2}y + \tau^{-2}\mu) \right] \end{aligned}$$

si definimos a

$$\mu_1 = \frac{\tau^{-2}\mu + \sigma^{-2}y}{\tau^{-2} + \sigma^{-2}} \quad y \quad \tau_1^{-2} = \tau^{-2} + \sigma^{-2}$$

tenemos que

$$\begin{aligned} p(\theta|y) &\propto \exp \left[-\frac{\theta^2}{2\tau_1^2} + \frac{\theta\mu_1}{\tau_1^2} \right] \\ &\propto \frac{1}{\sqrt{2\pi\tau_1^2}} \exp \left[-\frac{1}{2\tau_1^2}(\theta - \mu_1)^2 \right] \end{aligned}$$

y ésta última expresión corresponde a la f.d.p. de una distribución Normal. Es decir, si $(\theta|y) \sim N(\mu_1, \tau_1^2)$ entonces $p(\theta|y) \in P$.

En éste ejemplo vimos que el análisis fue hecho usando una única observación. Para cuando usemos una muestra de tamaño n , digamos $\mathbf{y} = (y_1, y_2, \dots, y_n)$, es fácil probar que la verosimilitud es dada por

$$l(\theta; \mathbf{y}) \propto \exp \left\{ -\frac{n}{2\sigma^2}(\bar{y} - \theta)^2 \right\}$$

Luego la verosimilitud es la misma que la obtenida con una única observación con la sustitución de y por \bar{y} y de σ^2 por $\frac{\sigma^2}{n}$. Así los resultado del ejemplo anterior son válidas con las sustituciones mencionadas anteriormente, y la f.d.p. a posteriori de θ sería dada por $N(\mu_1, \tau_1^2)$, con

$$\mu_1 = \frac{\tau^{-2}\mu + n\sigma^{-2}\bar{y}}{\tau^{-2} + n\sigma^{-2}} \quad y \quad \tau_1^{-2} = \tau^{-2} + n\sigma^{-2}$$

Una estrategia comunmente usada para la obtención de la f.d.p. a priori conjugada es ver el núcleo de la verosimilitud como función del parámetro de interés, e *imitar* la forma de ella.

Ejemplo 3.1.5 Sea $\mathbf{y} = (y_1, y_2, \dots, y_n)$ una muestra aleatoria de $N(\theta, \sigma^2)$ con θ conocido y definamos a $\phi = \sigma^{-2}$. De esta manera la verosimilitud de ϕ dada la muestra es dada por

$$\begin{aligned} l(\phi; \mathbf{y}) = p(\mathbf{y}|\theta, \phi) &= \prod_{i=1}^n p(y_i|\phi) \\ &= \prod_{i=1}^n \frac{1}{(2\pi\phi^{-1})^{1/2}} \exp\left\{-\frac{1}{2} \frac{(y_i - \theta)^2}{\phi^{-1}}\right\} \\ &= \frac{1}{(2\pi\phi^{-1})^{n/2}} \exp\left\{-\frac{1}{2\phi^{-1}} \sum_{i=1}^n (y_i - \theta)^2\right\} \\ &\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} n s_0^2\right\}; \quad s_0^2 = \frac{1}{n} \sum (y_i - \theta)^2 \end{aligned}$$

según esta verosimilitud, la a priori conjugada debe tener un núcleo como $l(\phi; \mathbf{y})$ que tiene la forma de una f.d.p Gamma - 1era parametrización (ver apéndice A). Es decir, podemos postular una distribución a priori $\phi \sim Ga(\frac{n_0}{2}, \frac{n_0}{2}\sigma_0^2)$ con f.d.p. a priori dada por $p(\phi) \propto \phi^{n_0/2-1} \exp\{-\phi n_0 \frac{\sigma_0^2}{2}\}$. La f.d.p. a posteriori de ϕ es obtenida usando el teorema de Bayes:

$$\begin{aligned} p(\phi|\mathbf{y}) &\propto l(\phi; \mathbf{y})p(\phi) \\ &\propto \phi^{n/2} \exp\left\{-n\phi \frac{s_0^2}{2}\right\} \phi^{n_0/2-1} \exp\left\{-n_0\phi \frac{\sigma_0^2}{2}\right\} \\ &\propto \phi^{[(n+n_0)/2]-1} \exp\left\{-(n_0\sigma_0^2 + n s_0^2) \frac{\phi}{2}\right\} \end{aligned}$$

cuyo núcleo corresponde a la f.d.p de una Gamma. Así la familia $P = \{Ga(\frac{n_0}{2}, \frac{n_0}{2}\sigma_0^2)\}$, con parámetros $\frac{n_0}{2}$ y $\frac{n_0}{2}\sigma_0^2$ conocidos, es conjugada a $F = \{N(\theta, \phi^{-1})\}$ con θ conocida.

Si queremos postular una f.d.p. a priori para la varianza σ^2 observamos la verosimilitud $l(\phi; \mathbf{y})$ como función de σ^2 . Es decir

$$l(\phi(\sigma^2); \mathbf{y}) = l(\sigma^2; \mathbf{y}) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} n s_0^2\right\}; \quad s_0^2 = \frac{1}{n} \sum (y_i - \theta)^2$$

y según este núcleo de la verosimilitud postulamos una f.d.p. a priori $GI(\frac{n_0}{2}, \frac{n_0}{2}\sigma_0^2)$ Gamma Inversa (ver apéndice A).

3.1.3 Inferencia Estadística Bayesiana

Hacer uso del paradigma Bayesiano no implica finalmente en la determinación de la distribución a posteriori $p(\theta|y)$, sino en hacer inferencia estadística a partir de esta distribución a posteriori. En estadística clásica o frecuentista el proceso de inferencia estadística significa estimación puntual, intervalos de confianza y Pruebas de Hipótesis para el parámetro o vector de parámetros de interés de un modelo probabilístico en particular, digamos $f(y|\theta)$. De igual forma, Un proceso de inferencia estadística Bayesiana involucra Estimación Puntual Bayesiana, Regiones de Credibilidad (intervalos de confianza Bayesianos) y Pruebas de Hipótesis Bayesianas. Para mayores detalles de este tema ver O'Hagan [17] y Berger [2]

En años recientes, la aproximación Bayesiana para Inferencia Estadística a recepcionado gran atención entre los investigadores en estadística aplicada y teórica. Esta alternativa es percibida como un paradigma natural para resolver problemas aplicados. Sin embargo, parece que la difusión de la metodología Bayesiana ha sido restringida por las dificultades de implementación. La estadística Bayesiana puede ser difícil de ser aplicada debido a la dificultad en derivar la distribución a posteriori analíticamente, en casos en donde no se trabaje con problemas conjugados comunmente.

En los últimos 20 años, una gran cantidad de progreso ha sido alcanzado en el área de computación Bayesiana. Un buen número de métodos de integración numérica han sido desarrollados para calcular las integrales que son comunes en Inferencia Bayesiana. (ver por ejemplo Naylor y Smith [15], Ripley [18], etc.). Una clase mas general de algoritmos de computación usa técnicas de muestreo para simular valores de distribuciones a posteriori. Un método de simulación en particular, Markov Chain Monte Carlo (MCMC), ha tenido notable éxito en simular valores de distribuciones a posteriori para un vector de parámetros de gran dimensión (ver por ejemplo Chib and Greenberg [6], Tierney [21], etc.).

A pesar de que existe una gran cantidad de buenos algoritmos para ejecutar cálculos Bayesianos, estos métodos no son ampliamente usados. En particular, estos métodos computacionales no son considerados en cursos introductorios sobre metodología Bayesiana. Dos razones pueden ser sugeridas para la escases en el uso de estas técnicas. Primero, muchos de los métodos computacionales a ser aplicado por el usuario, exige que esté familiarizado con los algoritmos y entender cuando los métodos convergen o producido una respuesta suficientemente acurística. Segundo, y probablemente la razón mas importante es que existe pocos software disponibles. Muchos de los software comercialmente disponibles

son diseñados para un problema de inferencia en particular, tal como modelos lineales, series de tiempo, etc. Existe pocos software disponibles que implemente el Teorema de Bayes para especificaciones a priori y de verosimilitud arbitrarias.

3.1.4 Inferencia Bayesiana a través del simulador estocástico MCMC - (Solución via Remuestreo)

Este trabajo de tesis presentará un caso en donde no es posible identificar la distribución a posteriori de un vector de parametros arbitrario θ , mediante el teorema de Bayes según la ecuación (3.5). Entre los muchos ejemplos en donde no es posible identificar la distribución a posteriori de interés, es importante destacar los modelos para explicar volatilidades de diversos instrumentos financieros en el mercado financiero. Nos ocuparemos particularmente del modelo GARCH(1,1) univariado, como el proceso que explica la volatilidad para nuestra serie de retornos de interés. Si éste modelo fuese sometido a un proceso de inferencia estadística Bayesiana, el núcleo de la distribución a posteriori del vector θ de interés, no presenta una estructura conocida como función de θ para identificarla, lo que imposibilita hacer resúmenes estadísticos de forma analítica. Para éste modelo GARCH(1,1) univariado, una estrategia particular de Markov Chain Monte Carlo (MCMC) denominada Metropolis-Hasting con propuesta independiente, será utilizado para obtener una muestra de la distribución a posteriori del vector de parametros de interés y así derivar algunos resúmenes particulares.

Vamos a presentar el simulador estocástico MCMC en forma descriptiva sin profundizar en detalles técnicos. Existe un sin número de referencias sobre este tema, ver por ejemplo O'Hagan [17], Roberts e Smith [19], Smith y Roberts [20], entre otros. La intención en esta subsección es destacar el mecanismo de funcionamiento del simulador de manera informal. Cadenas de Markov son procesos estocásticos en los cuales el comportamiento futuro del proceso depende del estado actual del proceso, y las probabilidades de los estados futuros no son afectados por la historia pasada del proceso, dado el estado actual. En muchos casos, estos procesos tienden al equilibrio y las cantidades límites siguen una distribución invariante. La Técnica MCMC proporciona un esquema de simulación donde la distribución de interés es colocada como la distribución de equilibrio. La introducción de cadenas de Markov en el esquema de simulación permite considerar distribuciones muy complicadas, principalmente aquellas que corresponden a las de gran dimensión. El hecho interesante

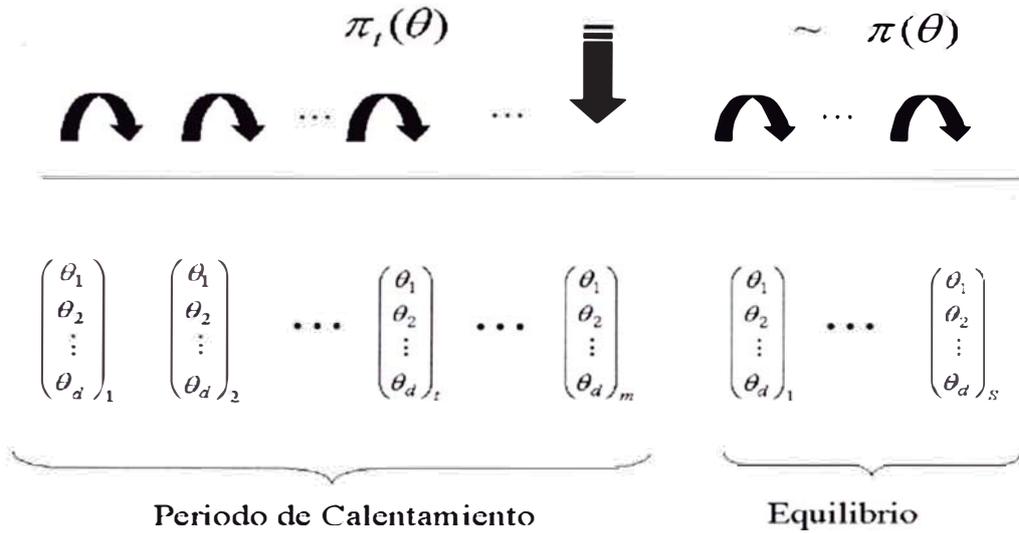


Figura 3.2: *inferencia via Muestreo*

de esta aproximación es que la introducción de esta estructura adicional (la cadena de Markov) en un problema complejo acaba resolviendo el problema original. Sin embargo, la introducción de esta estructura adicional genera un trabajo extra, pues es necesario obtener una secuencia completa de valores de esta cadena hasta alcanzar el equilibrio, y apenas los valores en el equilibrio son considerados como valores generados de la distribución límite (ver figura (3.2)). Aún mas, existen resultados análogos a la ley de los grandes números y al teorema Central del Limite para cadenas de Markov, que garantizan que los valores simulados por la cadena, en el límite, puedan ser usados para obtener información sobre la distribución de interés.

La idea que fundamenta el simulador estocástico MCMC es simple. Suponga que deseamos generar indirectamente una muestra de $\pi(\theta)$ onde $\theta \in \Theta \subseteq \mathbb{R}^n$. Suponga que es posible construir una cadena de Markov con espacio de estados Θ cuya distribución de equilibrio sea $\pi(\theta)$. Si iteramos la cadena por un periodo largo de tiempo, en el limite los valores generados de la cadena pueden ser usados como base para inferir sobre $\pi(\theta)$. Para implementar esta estrategia, necesitamos de algoritmos para construir cadenas con distribuciones de equilibrio específicas.

Para una discusión matemática rigurosa de estas ideas, necesitamos colocar la aproximación dentro de una estructura teórica apropiada. Detalles de esto pueden ser revisados en Smith y Roberts [20], Roberts y Smith [19], Tierney [21], etc., por ejemplo.

Bajo ciertas condiciones de regularidad, resultados asintóticos establecen condiciones para que una muestra resultante de esta cadena con distribución de equilibrio $\pi(\theta)$, pueda

ser usada para *imitar* una muestra generada por $\pi(\theta)$ o para estimar el valor esperado, con respecto a $\pi(\theta)$, de cualquier función de θ de interés.

Si $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(t)}, \dots$ fuese una realización de una cadena apropiada, resultados asintóticos garantizan que

$$\theta^{(t)} \xrightarrow{D} \theta \sim \pi(\theta);$$

y para cualquier $f(\cdot)$

$$\frac{1}{t} \sum_{i=1}^t f(\theta^{(i)}) \xrightarrow{a.s.} E_{\pi}\{f(\theta)\} \quad (3.9)$$

conocido como el Teorema Ergódico. La verdad, el Teorema Ergódico es una versión de la Ley de los Grandes Números para observaciones no necesariamente independientes. Ciertamente en el límite, sucesivos $\theta^{(t)}$ estarán correlacionados debido a que el esquema de simulación usa una cadena de Markov, es decir, si los primeros resultados asintóticos son usados para imitar una muestra de $\pi(\theta)$, ciertos espacios entre realizaciones deberían ser considerados para garantizar una muestra aproximadamente independiente, o de manera alternativa, experimentos independientes en paralelo pueden ser considerados. El segundo resultado asintótico implica que la media ergódica de una función de interés, usando las realizaciones de un único experimento, proporcionan un estimador consistente del valor esperado teórico.

Existe la cuestión de como construir una cadena de Markov cuya distribución limite sea exactamente la distribución de interés. Existen varias clases de esquemas para responder esto, siendo una de ellas el Gibbs Samplig (Muestrador de Gibbs). Este esquema es basado en una cadena de Markov cuya dependencia entre estados es gobernada por distribuciones condicionales obtenidas a partir del modelo. Muchos modelos tienen distribución conjunta compleja, sin embargo las distribuciones condicionales completas son relativamente fáciles de obtener.

Otro esquema de simulación es dado por el algoritmo de Metropolis-Hasting. Este esquema esta basado en una cadena de Markov cuya dependencia sobre el antecesor es dividida en dos partes: una propuesta y una probabilidad de aceptación de esta propuesta. la propuesta sugiere un próximo paso arbitrario a seguir en la trayectoria de la cadena, y la probabilidad de aceptación garantiza la dirección limite apropiada, rechazando movimientos no deseables de la cadena. Esta estrategia representa una alternativa de solución cuando las distribuciones condicionales de interés son complejas.

3.1.5 Gibbs Sampling

El Gibbs Sampling aplicado a la f.d.p de $(\theta|y)$ dado por $\varphi(\theta|y)$ produce, a través de simulaciones, una cadena de Markov $\{\theta^n\}_{n=1}^N$ con distribución de equilibrio $\varphi(\theta|y)$. Los valores pueden ser usados para estimar consistentemente características de interés de la distribución a posteriori, tales como los momentos de primer y segundo orden, densidades marginales, etc.

Suponga que particionamos θ en θ_1 y θ_2 de modo que las densidades a posteriori condicionales $\varphi(\theta_1|\theta_2, y)$ e $\varphi(\theta_2|\theta_1, y)$ tengan formas analíticas conocidas y sea posible generar muestras facilmente a partir de ellas. Esta característica de muestreo de distribuciones condicionales es llamada de *conjugacidad*. El próximo valor θ^n de la densidad a posteriori conjunta será generada usando el valor anterior θ^{n-1} de acuerdo al siguiente formato cíclico

$$\begin{aligned}\theta_1^n &\sim \varphi(\theta_1|\theta_2^{n-1}, y) \\ \theta_2^n &\sim \varphi(\theta_2|\theta_1^n, y)\end{aligned}\tag{3.10}$$

($x^n \sim \varphi(x)$ significa que x^n es un valor generado de la distribución $\varphi(x)$). El procedimiento es inicializado usando algún valor θ_2^0 perteneciente al dominio de la distribución a posteriori. La influencia de la condición inicial desaparece despues de un cierto número de valores generados. Si el algoritmo converge, después de n valores generados, lo que tenemos es una muestra de la densidad a posteriori de θ . Los valores generados no son independientes debido a la estructura de la cadena de Markov.

El Gibbs Sampling usa al máximo las propiedades analíticas de la densidad a posteriori. El Gibbs Sampling es simple cuando el vector de parametros θ puede ser particionado en θ_1 y θ_2 y las expresiones analíticas de ambas densidades condicionales a posteriori son conocidas. A través de un Muestreo Secuencial en cada densidad a posteriori condicional obtenemos una cadena de Markov que, en la convergencia, produce una muestra de la densidad a posteriori conjunta, según el esquema de la figura (3.2). En algunos casos la densidad condicional a posteriori $(\theta_1|\theta_2)$ es conocida, y la densidad condicional a posteriori $(\theta_2|\theta_1)$ no lo es. En estos caso un muestreo directo en $\varphi(\theta_1|\theta_2, y)$ es fácil, pero un muestreo de $\varphi(\theta_2|\theta_1, y)$ es posible de un modo indirecto solamente.

3.1.6 El algoritmo de Metropolis-Hastings

Para construir una cadena de Markov $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(t)}, \dots$ con espacio de estados Θ y distribución de equilibrio $\pi(\theta)$, el algoritmo del Metropolis-Hasting construye una probabilidad de transición (kernel de transición) a partir de un estado $\theta^{(t)} = \phi$ para un tiempo t arbitrario, hasta el próximo estado $\theta^{(t+1)}$. Sea $q(\phi, \phi')$ una probabilidad de transición arbitraria (kernel de transición), de tal forma que, si $\theta^{(t)} = \phi$, ϕ' generado a partir de $q(\phi, \phi')$ será considerado como un valor posible propuesto para $\theta^{(t+1)} = \phi'$. En este momento un proceso de aleatorización debe acompañar el algoritmo. Con una probabilidad $\alpha(\phi, \phi')$, $\theta^{(t+1)} = \phi'$ será aceptado, caso contrario rechazamos el valor generado a partir de $q(\phi, \phi')$ y usamos el valor del estado anterior, o sea $\theta^{(t+1)} = \phi$. Esta construcción define una cadena de Markov con probabilidad de transición dada por:

$$p(\phi, \phi') = \begin{cases} q(\phi, \phi')\alpha(\phi, \phi') & \text{se } \phi' \neq \phi \\ 1 - \sum_{\rho} q(\phi, \rho)\alpha(\phi, \rho) & \text{se } \phi' = \phi. \end{cases} \quad (3.11)$$

con

$$\alpha(\phi, \phi') = \begin{cases} \min \left\{ \frac{\pi(\phi')q(\phi', \phi)}{\pi(\phi)q(\phi, \phi')}, 1 \right\} & \text{se } \pi(\phi)q(\phi, \phi') > 0, \\ 1 & \text{se } \pi(\phi)q(\phi, \phi') = 0. \end{cases} \quad (3.12)$$

es posible comprobar que $\pi(\phi)p(\phi, \phi') = \pi(\phi')p(\phi', \phi)$, es decir la cadena de Markov construida es reversible. Un hecho interesante aparece en esta definición de la cadena: $q(\phi, \phi')$ es arbitraria y puede ser seleccionada de tal forma que la cadena resultante sea irreducible y ergódica sobre el espacio de estado de interés, condición suficiente para que $\pi(\theta)$ sea la distribución de equilibrio de la cadena de Markov construida. Ver Roberts y Smith [19], y Hoel *et al* [10] para mayores detalles.

Es importante destacar que la distribución de interés (equilibrio), $\pi(\theta)$, aparece en $p(\phi, \phi')$ unicamente a través de $\alpha(\phi, \phi')$, como una proporción $\pi(\phi')/\pi(\phi)$. En un contexto Bayesiano, donde $\pi(\theta)$ es una distribución a posteriori, esto es un hecho trascendental pues no es necesario el conocimiento completo de esta distribución, pues el núcleo de la distribución (dado por el producto de la verosimilitud y la apriori) es suficiente para implementar el algoritmo del Metropolis-Hasting.

Para simular valores generados a partir de $\pi(\cdot)$ usando una cadena de Markov definida por la probabilidad de transición (3.11), procedemos de la siguiente forma:

1. Inicie el contador de la iteración $t=1$ y proporcione un valor inicial arbitrario para $\theta^{(0)}$.

2. Mueva la cadena para un nuevo valor ϕ generado a partir de la densidad $q(\theta^{(t-1)}, \cdot)$.
3. Evalúe la probabilidad de aceptación del movimiento $\alpha(\theta^{(t-1)}, \phi)$ dada por (3.12). Si el movimiento fuese aceptado, $\theta^{(t)} = \phi$. Si el movimiento fuese rechazado, $\theta^{(t)} = \theta^{(t-1)}$ y la cadena permanece en el estado anterior.
4. Cambie el contador de t para $t+1$ y retorne a (2) hasta que la convergencia sea alcanzada.

el paso 3. es ejecutado después de la generación independiente de un valor u a partir de una distribución Uniforme. Si $u \leq \alpha$, el movimiento es aceptado y si $u > \alpha$ el movimiento no es permitido. La probabilidad de transición q define únicamente un movimiento posible a ser confirmado según el valor de α . Esta etapa del algoritmo es basada en las ideas de Remuestreo, particularmente el método de Aceptación-Rechazo (ver Ripley [18] para detalles de éste método). Sin embargo, el método de Aceptación-Rechazo no son iterativos.

Como fue mencionado anteriormente, existe una flexibilidad total en cuanto a la selección de la probabilidad de transición q . Diferentes selecciones de q conducen a diferentes algoritmos específicos. Tierney [21] proporciona diferentes tipos de selección disponibles. Por ejemplo, si $q(\phi, \phi') = q(\phi', \phi)$, tenemos que $\alpha(\phi, \phi') = \min\{\pi(\phi')/\pi(\phi), 1\}$, y éste algoritmo es conocido como algoritmo de Metropolis [12]. Si $q(\phi, \phi') = q(\phi' - \phi)$, la cadena es gobernada por un proceso camino aleatorio, donde la distribución Normal multivariada y/o t -Student multivariada son candidatas posibles para $q(\phi' - \phi)$. Si $q(\phi, \phi') = q(\phi')$, tenemos que $\alpha(\phi, \phi') = \min\{\omega(\phi')/\omega(\phi), 1\}$, donde $\omega(\phi) = \pi(\phi)/q(\phi)$ define el peso importancia. Ideas sobre metodologías de Remuestreo por Importancia sugieren que distribuciones t -Student con pequeños grados de libertad son buenas candidatas.

Finalmente, estrategias híbridas pueden ser creadas combinando diferentes cadenas en diversas formas, como por ejemplo esquemas de simulación por componentes cíclico y mixtura) o una utilización del Metropolis-Hasting dentro del condicionamiento del Muestrador de Gibbs.

Esquema de Muestreo por Componentes

En todos los casos presentados en el algoritmo de Metropolis-Hasting, la cantidad de interés para muestrear $\theta = (\theta_1, \dots, \theta_d)$ fue actualizada en un único bloque. En esta parte, una nueva forma de transición es presentada, cuando los componentes $\theta_1, \theta_2, \dots, \theta_d$ son usados

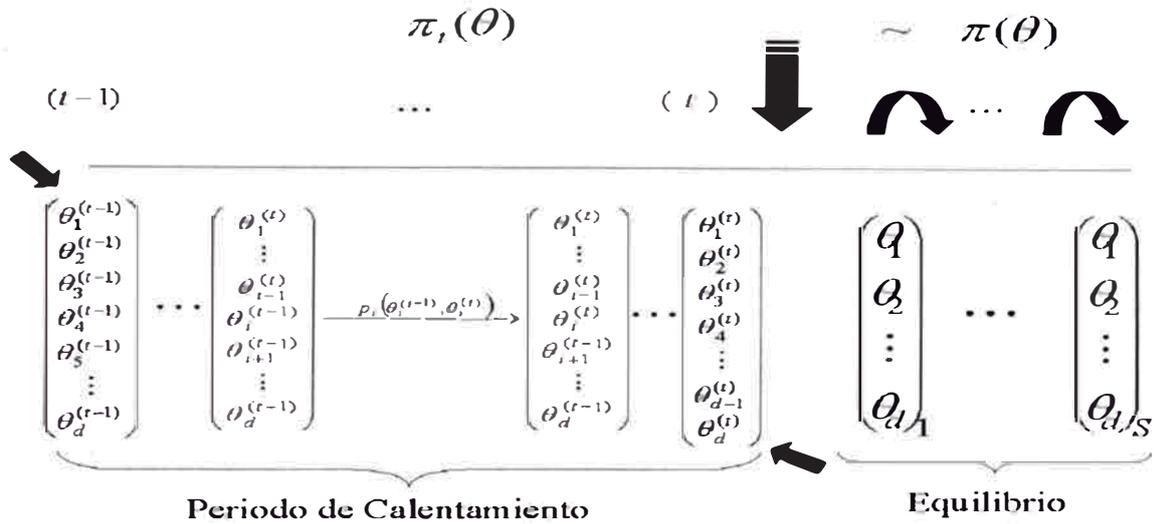


Figura 3.3: Inferencia via muestreo por componentes

separadamente. Tierney [21] discute esta posibilidad. En particular, las componentes de θ pueden ser actualizadas o cambiadas de la siguiente forma:

- a) En cada iteración, una única componente es actualizada y la selección de la componente es hecha a través de una selección aleatoria entre las componentes.
- b) En cada iteración, una única componente es actualizada y la selección de la componente es hecha en un orden predeterminado (fija y pre-especificada) entre los n componentes. Por ejemplo, $1 \rightarrow 2 \rightarrow \dots \rightarrow d$.

Las formas anteriores son ejemplos de transiciones mixturas (caso a) y transiciones cíclicas (caso b), ver figura (3.3). En el primer caso, se define la transición p_m como una distribución de equilibrio común π y pesos ω_m , $m = 1, \dots, r$ con $\omega_m \geq 0$ e $\sum_{m=1}^r \omega_m = 1$. Un kernel de transición mixtura p es definida por $p = \sum_{m=1}^r \omega_m p_m$. El item anterior (a) es un caso especial de una mixtura con $r = d$, $\omega_m = 1/d$ y cada transición p_m mueve unicamente la m -ésima componente de θ , $m = 1, \dots, d$.

Propiedades de las transiciones p_m son pasadas a la transición mixtura p . Ante todo el kernel mixtura p define un kernel de transición de una cadena de Markov con distribución de equilibrio π . Por otra parte, si uno de los kernel de transición componente es irreducible y aperiódico entonces el kernel mixtura es irreducible y aperiódico.

Para transiciones cíclicas con transición componente p_c , $c = 1, \dots, r$, una iteración de la nueva cadena será ejecutada después que todas las componentes cíclicas de transición fuesen ejecutadas bajo p_c , en un ciclo determinado.

Para ir desde ϕ hasta ϕ' en una única iteración de una cadena cíclica, movimientos en estados intermedios ψ_c con probabilidades de transición p_c , $c = 1, \dots, r-1$ conduce finalmente para $\psi_r = \phi'$ a través de p_r . Sea $\psi_0 = \phi$, entonces:

$$p(\phi, \phi') = \int \dots \int \prod_{c=1}^r p_c(\psi_{c-1}, \psi_c) d\psi_1 \dots d\psi_{r-1} \quad (3.13)$$

El item anterior (b) es un caso especial de un ciclo con $r = d$, y en cada iteración p_c mueve unicamente la c -ésima componente de ϕ , $c = 1, \dots, d$.

Muchas de las propiedades de las transiciones p_c son pasadas para la transición cíclica p . La transición cíclica p define la transición de una cadena de Markov con distribución de equilibrio π . Al contrario de transiciones mixturas, irreducibilidad y ergodicidad de una de las transiciones componentes no son suficientes para irreducibilidad y ergodicidad de la transición cíclica. Si todas las transiciones componentes son irreducibles y ergódicas, entonces la transición cíclica es irreducible y ergódica; ver Tierney [21] para mayores detalles.

Estas formas pueden ser integradas al algoritmo de Metropolis-Hasting. Cada Transición p_i puede ser asociada a una transición propuesta q_i y a una probabilidad de aceptación α_i .

En una aproximación Bayesiana, si $\theta = (\theta_1, \dots, \theta_d)$ tiene densidad a posteriori conjunta $\pi(\theta_1, \dots, \theta_d)$, la densidad marginal a posteriori de θ_i , $i = 1, \dots, d$ es dada por:

$$\pi(\theta_i) = \int \pi(\theta_1, \dots, \theta_d) d\theta_{-i}$$

donde $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$ es el vector θ con el i -ésimo componente removido. Para cada θ_i , las distribuciones condicionales posibles son:

$$\pi(\theta_i | \theta_j, j \in C) = \pi(\theta_i, \theta_j, j \in C) / \pi(\theta_j, j \in C)$$

para todo $C \subset \{1, \dots, i-1, i+1, \dots, d\}$. La distribución condicional de $\theta_i | \theta_{-i}$ sera llamada distribución condicional completa de θ_i y a la densidad por $\pi_i(\theta_i)$, osea :

$$\pi_i(\theta_i) = \pi(\theta_i | \theta_{-i})$$

En este contexto, considere el esquema cíclico con transición por componentes. Es decir, la componente de transición $q_i(\phi, \phi')$ propone un movimiento de la i -ésima componente de ϕ , $i = 1, \dots, d$. Además, tenemos que $\pi(\phi) = \pi_i(\phi)\pi(\phi_{-i})$ y el movimiento determinado por q_i cambia apenas ϕ_i , luego $\phi_{-i} = \phi'_{-i}$ y $\pi(\phi') | \pi(\phi) = \pi_i(\phi'_i) | \pi_i(\phi_i)$.

Una vez que p_i es de nuestro interés, las otras componentes de ϕ permanecen fijas y no son afectadas. Es decir, tenemos una cadena de Markov reducible. La transición propuesta

q_i puede ser escrita como $q_i(\phi_i, \phi'_i)$, a pesar que ella puede depender de ϕ_{-i} . Por lo tanto, la probabilidad de aceptación es:

$$\alpha_i(\phi_i, \phi'_i) = \min \left\{ \frac{\pi_i(\phi'_i)q_i(\phi'_i, \phi_i)}{\pi_i(\phi_i)q_i(\phi_i, \phi'_i)}, 1 \right\} \quad (3.14)$$

Note que cada componente de transición anterior define una cadena reversible con distribución de equilibrio π_i , $i = 1, \dots, d$; además de que $\pi(\phi')|\pi(\phi) = \pi_i(\phi'_i)|\pi_i(\phi_i)$. Es decir, cada componente de transición satisface la siguiente ecuación :

$$\pi(\phi'_i|\phi_{-i}) = \int \pi(\phi_i|\phi_{-i})p_i(\phi_i, \phi'_i)d\phi_i \quad (3.15)$$

a continuación se probará, para el caso particular de 2 componentes que la transición cíclica $p(\phi, \phi')$ de la ecuación (3.13) define una cadena de Markov con distribución de equilibrio π . Un movimiento a partir de $\phi = (\phi_1, \phi_2)$ hasta $\phi' = (\phi'_1, \phi'_2)$ es realizado sólo si las dos componentes fuesen ejecutadas de acuerdo con sus respectivas transiciones cíclica, la siguiente ecuación debe ser satisfecha:

$$\pi(\phi') = \int \int \pi(\phi)p_1(\phi_1, \phi'_1)p_2(\phi_2, \phi'_2)d\phi \quad (3.16)$$

Para probar la validez de la ecuación anterior desarrollemos el término del lado derecho de la ecuación (3.16).

$$\begin{aligned} &= \int \int \pi(\phi_1|\phi_2)\pi(\phi_2)p_1(\phi_1, \phi'_1)p_2(\phi_2, \phi'_2)d\phi_1 d\phi_2 \\ &= \int \pi(\phi_2) \left[\int \pi(\phi_1|\phi_2)p_1(\phi_1, \phi'_1)d\phi_1 \right] p_2(\phi_2, \phi'_2)d\phi_2 \\ &= \int \pi(\phi_2)\pi(\phi'_1|\phi_2)p_2(\phi_2, \phi'_2)d\phi_2 \\ &= \int \pi(\phi'_1)\pi(\phi_2|\phi'_1)p_2(\phi_2, \phi'_2)d\phi_2 \\ &= \pi(\phi'_1) \int \pi(\phi_2|\phi'_1)p_2(\phi_2, \phi'_2)d\phi_2 \\ &= \pi(\phi'_1)\pi(\phi'_2|\phi'_1) \\ &= \pi(\phi') \end{aligned}$$

Lo que comprueba la validez de la ecuación (3.16). Los resultados pueden ser extendidos para cualquier número de componentes usando inducción sobre los mismos argumentos. La demostración anterior usa transiciones continuas por una cuestión de simplicidad notacional. Es posible probar a pesar de la reductibilidad de las transiciones componentes, que la transición cíclica es irreductible y aperiódica, ver Tierney [21]. Así, la distribución límite es única.

El algoritmo de Metropolis-Hasting por componentes, que define un movimiento de $\theta^{(t)}$ para $\theta^{(t+1)} = \psi$, puede ser esquematizada como:

1. Inicie el contador de la iteración $t=1$ y, defina un valor inicial arbitrario para $\theta^{(0)}$;
2. Inicie el contador de componente $i=1$.
3. Mueva la i -ésima componente del vector de estados de la cadena para un nuevo valor ψ_i generado a partir de la densidad $q_i(\theta_i^{(t-1)}, \cdot)$.
4. Evalúe la probabilidad de aceptación del movimiento $\alpha_i(\theta_i^{(t-1)}, \psi_i)$ dada por (3.14). Si el movimiento es aceptado, $\theta_i^t = \psi_i$. Si el movimiento es rechazado, $\theta_i^t = \theta_i^{(t-1)}$ y la cadena permanece en ese estado anterior.
5. Cambie el contador de i para $i+1$ y retorne a (3) hasta $i=d$. Cuando $i = d$, ir al paso (6).
6. Cambie el contador de t para $t+1$ y retorne a (2) hasta que la convergencia sea alcanzada.

Capítulo 4

Volatilidad de Tasa de intercambio del Nuevo Sol versus Dólar Americano

Mientras que los modelos econométricos y de series de tiempo operan bajo la suposición de varianza constante, los procesos ARCII/GARCII introducidos por Engle [7] y Bollerslev [4] permiten que la varianza condicional cambien en el tiempo como una función de los errores pasados y varianzas condicionales pasadas, dejando que la varianza incondicional sea constante.

El proceso de inferencia estadística puede ser comunmente hecha desde una perspectiva clásica o Bayesiana. Desde una perspectiva clásica, la estimación de los parametros de los modelos univariados ARCII/GARCII pueden ser ejecutados bajo el principio de máxima verosimilitud, métodos de mínimos cuadrados, etc. Existen diversas propuestas para la resolución numérica de estos problemas de optimización de parámetros para la familia ARCII/GARCH entre los que se encuentra por ejemplo la propuesta de BIIIII (Berndt *all* [3]).

Inferencia Bayesiana sobre los modelos ARCII/GARCII fueron implementados bajo diversas estrategias de solución numérica. Geweke [8] y Kleibergen-van Dijk [11] usaron métodos de remuestreo para estimación de parámetros de un modelo ARCII bajo una perspectiva Bayesiana. Geweke [9] desarrolló una estrategia MCMC de Metropolis-Hastings para inferencia Bayesiana de un modelo GARCII(1,1) Normal. Bauwens y Lubrano [1] propusi-

erón otra estrategia MCMC de Gibbs Sampling para un modelo GARCH(1,1) t-Student.

El presente trabajo de tesis tiene por objetivo desarrollar una estrategia MCMC de Metropolis-Hastings por componentes para un modelo GARCH(1,1) t-Student como una alternativa de solución para propósitos de inferencia estadística Bayesiana. La propuesta será evaluada con datos artificiales de manera similar al trabajo de Bauwens [1] y enseguida implementada para explicar la volatilidad de la tasa de intercambio del Nuevo Sol versus el Dólar americano. El argumento para usar este modelo se deriva de las características propias que esta serie de retornos presenta cuyas propiedades de hechos estilizados es confirmada en el trabajo monográfico de Navarro [13]

La selección del modelo GARCH(1,1) t-Student recae en el hecho de que se trata de un modelo padrón usado por muchos usuarios de finanzas. La implementación inicial en una serie artificial nos permitirá describir algunas posibles complicaciones que se pueden presentar en un estudio de caso real y concluir adecuadamente los resultados.

En la tabla (4.1) damos una breve descripción de la serie de retornos que vamos a utilizar (ver apéndice B). Durante este periodo de tiempo se registra días en donde no hay negociación como son los días feriados, sábados y domingos y que no son considerados en la recopilación de datos.

Serie	Tipo de serie	Mercado	Inicio	Fin	Retornos diarios
SOL/US\$	tasa de intercambio	Nacional	02/01/1996	31/12/1998	754

Tabla 4.1: Descripción de la serie de retornos de la tasa de intercambio del Nuevo Sol versus el Dólar americano

La gráfica de la serie de retornos de tasa de intercambio del Nuevo Sol versus el Dólar norteamericano fue dada en la figura 1.2 en donde podemos observar inicialmente que la serie de retornos presenta volatilidades por conglomerados. Los hechos estilizados de esta serie de retornos son descritos en detalle en Navarro [13]

4.1 Aproximación Bayesiana para GARCH(1,1) t-Student

A partir de los indicios encontrados de que la serie de retornos de la tasa de intercambio del Nuevo Sol versus el Dólar americano presenta características propias de los *hechos estilizados*, el siguiente paso es seleccionar el modelo que explique estas características propias. El

modelo GARCH(1,1) t-Student es un modelo estocástico común en los usuarios que tienen que lidiar con la explicación de volatilidades de diversos instrumentos financieros y será el que usaremos para explicar la volatilidad de la tasa de intercambio del Nuevo Sol versus el Dólar americano. A continuación desarrollaremos los pasos para ejecutar el proceso de inferencia estadística Bayesiana sobre los parámetros de interés del modelo GARCH(1,1) t-Student propiamente dicho. El presente trabajo pretende alcanzar los resultados o resúmenes de interés haciendo uso de técnicas de simulación estocástica iterativas dado que es imposible llevar obtener una solución analítica de algunos resúmenes de interés de la f.d.p a posteriori del vector de parámetros involucrado en el modelo GARCH(1,1) t-Student.

4.1.1 Metropolis-Hasting por componentes para GARCH(1,1) t-Student

Sea θ el vector de parámetros (w, α, β, ν) . Para una muestra de T observaciones $\mathbf{y} = (y_1, \dots, y_T)$, la f.d.p. es

$$p(\theta|\mathbf{y}) \propto p(\theta)l(\theta; \mathbf{y}) \quad (4.1)$$

donde la función de verosimilitud es dado por

$$l(\theta|\mathbf{y}) = \prod_{t=2}^T p(y_t|y_{t-1}, \theta)p(y_1|y_0\theta) \quad (4.2)$$

$$\propto \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (\nu h_t)^{-\frac{1}{2}} \left[1 + \frac{y_t^2}{\nu h_t} \right]^{-\frac{\nu+1}{2}} \quad (4.3)$$

La f.d.p. a priori $p(\theta)$ debe respetar al menos las restricciones de parámetros positivos y que $\beta < 1$ (condición de estacionariedad, ver Nelson [16]). Siguiendo el argumento dado por Bauwens [1], una insinuación importante en inferencia estadística Bayesiana es la integrabilidad de la densidad a posteriori de θ . Según este análisis es suficiente considerar prioris Uniformes para los parámetros w , α y β , mientras que para el parámetro ν la a priori es dado por $p(\nu) \propto (1 + \nu^2)^{-1}$ para $\nu > 0$.

Como se puede observar no es posible identificar la densidad a posteriori de θ y recurriremos a un algoritmo iterativo de simulación estocástica MCMC de Metropolis-Hastings por componentes para obtener resúmenes de interés de la densidad a posteriori de θ . El esquema del algoritmo es el de cadenas independientes enunciado en el capítulo anterior y todo lo que necesitamos es definir el tipo de kernel de transición propuesta q_i en cada componente. Dada las características de nuestro problema usaremos un kernel de transición propuesta del tipo $q_i(\theta'_i, \theta_i) = f_i(\theta_i)$. La selección del tipo de distribución para $f_i(\theta_i)$ es

una distribución t -Student centrado en la moda $\hat{\theta}_i$ del logaritmo de la función de verosimilitud condicional $l(\theta_i|\theta_{-i}, \mathbf{y}) \propto l(\theta|\mathbf{y})$. La matriz escala para cada distribución t -Student es dado por $c_i \times V_i$ con c_i una constante de afinamiento que se determinará de manera experimental y V_i es el i -ésimo componente de la matriz de escala $V = D \left(\frac{\alpha^* + p}{\alpha^*} \right)$ de una distribución multivariada t -Student de dimensión $p = 4$ y D matriz de dispersión modal (ver O'Hagan [17]). La obtención de esta matriz de escala V suele hacerse de manera aproximada y la forma en que la obtenemos es a través de una aproximación de la matriz Hessiana del logaritmo de $l(\theta|\mathbf{y})$. Sea $Hess$ la aproximación a la matriz Hessiana y hagamos que $Hess = -V^{-1} \left(\frac{\alpha^* + p}{\alpha^*} \right)$. La matriz de dispersión modal D para una distribución t -Student multivariada de dimensión p es dada por $D = -Hess^{-1} = V \left(\frac{\alpha^*}{\alpha^* + p} \right)$. La matriz de escala se obtiene de la anterior relación como $V = D \left(\frac{\alpha^* + p}{\alpha^*} \right)$. El valor de α^* es dado por el estimador máximo verosímil $\hat{\theta}_1 = \hat{\nu}$.

Para evaluar el desempeño del algoritmo de simulación estocástico MCMC de Metropolis-Hastings por componentes propuesta por esta tesis, generaremos una serie de datos artificiales de tamaño 150. El modelo usado es un proceso GARCH(1,1) t -Student univariado con parámetros $w = 0.1$, $\alpha = 0.4$, $\beta = 0.4$ y $\nu = 5$.

4.1.2 Diagnóstico de Convergencia

Una vez que se ha determinado los elementos necesarios para la implementación del algoritmo MCMC de Metropolis-Hastings por componentes procedemos a desarrollarlo bajo un entorno MatLab. Un criterio de convergencia debe acompañar toda corrida de un programa de implementación MCMC y el usado en esta tesis es el criterio CUMSUM propuesto por Bauwens [1]. Este criterio se basa en una única corrida de la cadena y define una estadística CUMSUM definida por

$$CS_t = \left(\frac{1}{t} \sum_{n=1}^t \theta^{(n)} - \mu_0 \right) / \sigma_0 \quad t = 50, 100, 150, \dots, T \quad (4.4)$$

donde μ_0 y σ_0 son las medias y desvios empíricas para T observaciones. Si el MCMC converge el gráfico CS_t vs. t converge levemente a cero. Caso contrario, valores distantes de cero son una indicación de la ausencia de convergencia, ver detalles en Bauwens [1].

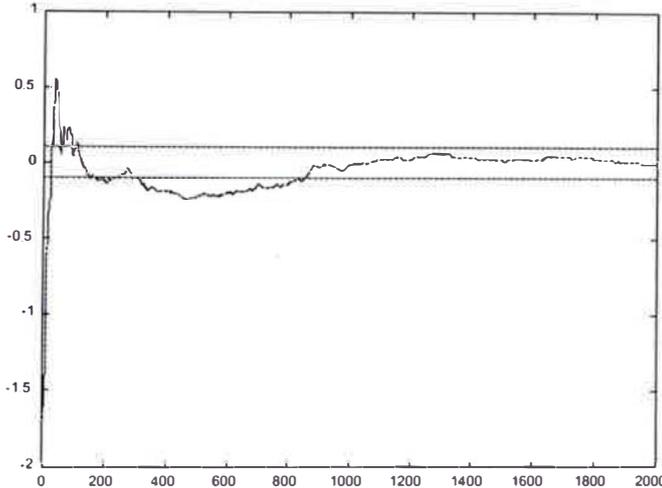


Figura 4.1: *ploteos CUMSUM para estimativa de la media w (datos simulados)*

Datos Artificiales

Los gráficos (4.1), (4.2), (4.3) y (4.4) generados por una corrida de 2000 iteraciones nos indican que aproximadamente el CUMSUM para la estimativa de w converge aproximadamente a cero a partir de la iteración 800, el CUMSUM para α converge aproximadamente a cero a partir de la iteración 400, el CUMSUM para β converge aproximadamente a cero a partir de la iteración 800, y el CUMSUM para ν converge aproximadamente a cero a partir de 300. De los resultados anteriores ejecutamos 2000 iteraciones para hacer uso de las últimas 1000 iteraciones y así obtener resúmenes de interés de la densidad a posteriori de θ .

Los Histogramas (4.5), (4.6), (4.7) y (4.8) obtenidos usando los últimos 1000 valores generados por el algoritmo MCMC Metropolis-Hastings por componentes presenta formas de las densidades a posterioris de cada componente de θ similares a las halladas por Bauwens [1] sin problemas en las colas.

La tabla (4.2) muestra las estimativas [desvío padrón asintótico] de cada parámetro obtenidos por máxima verosimilitud (inferencia clásica), así como las estimativas Bayesianas [desvío padrón] obtenidos por el algoritmo Metropolis-Hastings por componentes (inferencia Bayesiana). Debemos mencionar que utilizamos $c_i = (1.2)^2$ para calibrar los parámetros de escala de las distribuciones t-Student univariados considerados como los kernel propuestos para cada componente del vector θ . Los grados de libertad para estos kernel propuestos t-Student son iguales a 5. Las tasas de aceptación de cada componente fueron 41.7%, 40.7%,

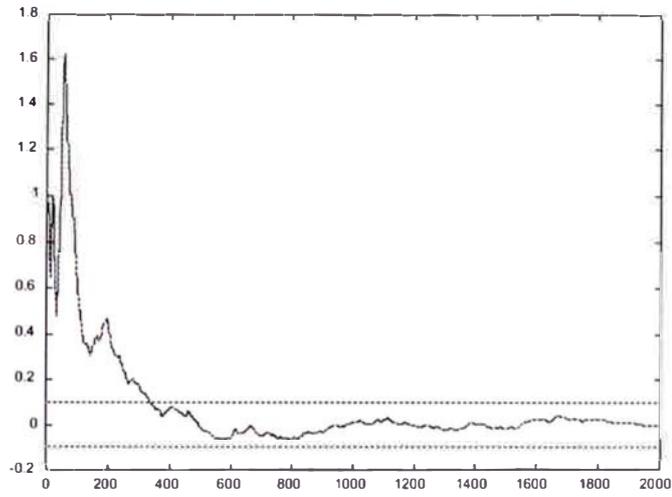


Figura 4.2: *ploteos CUMSUM para estimativa de la media α (datos simulados)*

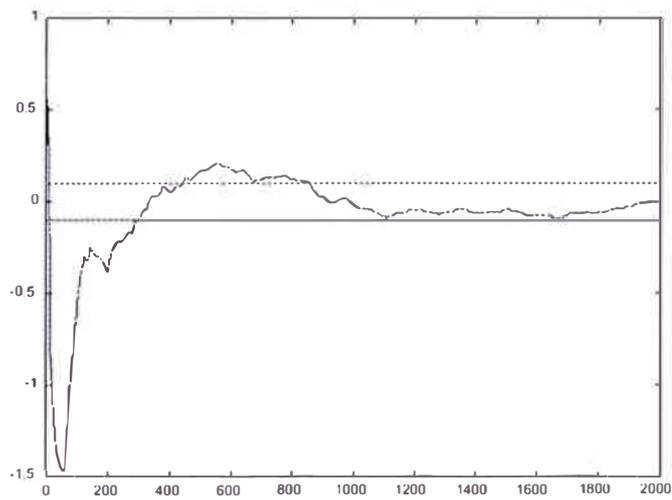


Figura 4.3: *ploteos CUMSUM para estimativa de la media β (datos simulados)*

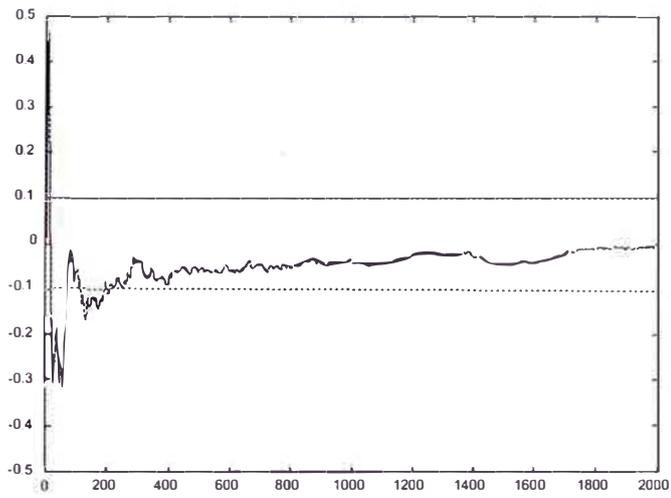


Figura 4.4: *ploteos CUMSUM para estimativa de la media ν (datos simulados)*

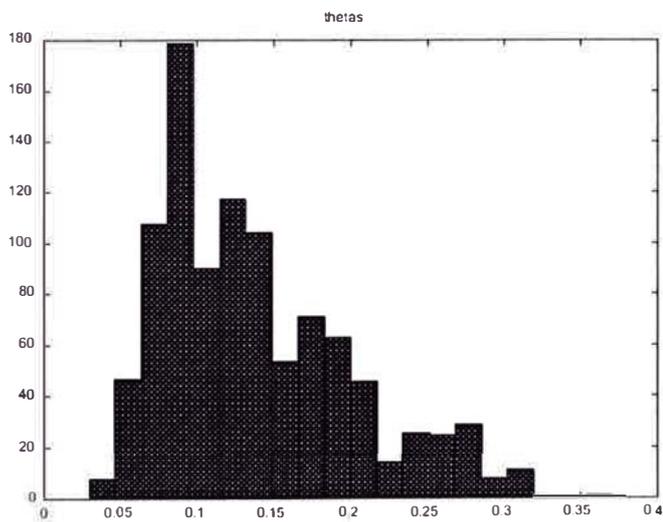


Figura 4.5: *Histograma de w (datos simulados)*

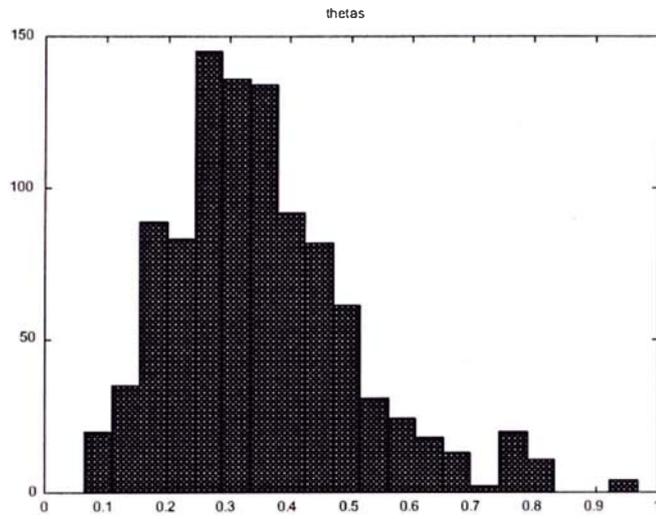


Figura 4.6: *Histograma de α (datos simulados)*

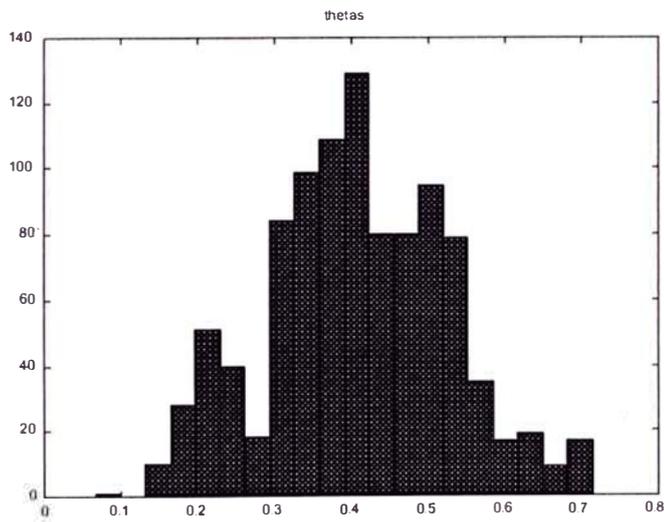


Figura 4.7: *Histograma de β (datos simulados)*

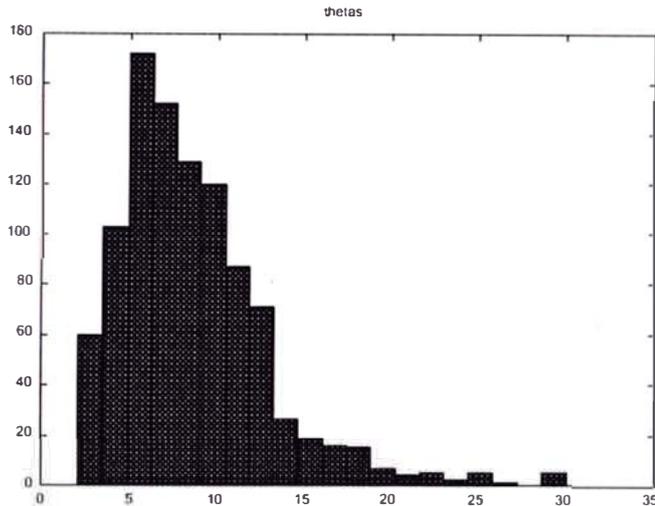


Figura 4.8: *Histograma de ν (datos simulados)*

29.6%, y 77.05% respectivamente.

Parámetros	Máxima Verosimilitud	Metropolis-Hastings por componentes
$w = 0.1$	0.0977 (0.0343)	0.1349 (0.0545)
$\alpha = 0.4$	0.2893 (0.1155)	0.3424 (0.1421)
$\beta = 0.4$	0.4839 (0.1287)	0.4140 (0.1078)
$\nu = 5$	6.6793 (3.0158)	8.449 (4.1864)

Tabla 4.2: *Estimación clásica y Bayesiana para un modelo GARCH(1,1) t-Student (datos artificiales)*

Datos retornos Sol versus Dólar

Los datos de retornos fueron inicialmente ajustados por su media muestral para luego modelarlos por un proceso GARCH(1,1) t-Student. Los gráficos (4.9), (4.10), (4.11) y (4.12) para una rodada de 3000 iteraciones nos indican que aproximadamente el CUMSUM para la estimativa de w converge aproximadamente a cero a partir de la iteración 1400, el CUMSUM para α converge aproximadamente a cero a partir de la iteración 1200, el CUMSUM para β converge aproximadamente a cero a partir de la iteración 2200, y el CUMSUM para ν converge aproximadamente a cero a partir de 300. De los resultados anteriores ejecutamos

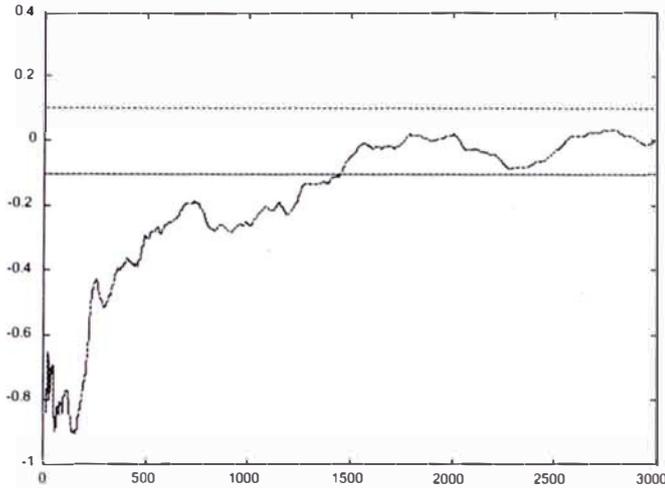


Figura 4.9: *ploteos CUMSUM para estimativa de la media w (datos reales)*

4000 iteraciones para hacer uso de las últimas 1500 iteraciones y así obtener resúmenes de interés de la densidad a posteriori de θ . El algoritmo converge lentamente debido a que existe una correlación alta entre w y α , así como w y β .

Los histogramas (4.13), (4.14), (4.15) y (4.16) obtenidos usando los últimos 1500 valores generados por el algoritmo MCMC Metropolis-Hastings por componentes presenta formas de las densidades a posterioris de cada componente de θ similares a las halladas para los datos artificiales, y esto se debe a la forma de la a priori para ν que garantiza densidad a posteriori de θ integrable.

La tabla (4.3) muestra las estimativas [desvío padrón asintótico] de cada parámetro obtenidos por máxima verosimilitud (inferencia clásica), así como las estimativas Bayesianas [desvío padrón] obtenidos por el algoritmo Metropolis-Hastings por componentes (inferencia Bayesiana). De manera similar utilizamos $c_i = (1.2)^2$ para calibrar los parámetros de escala de las distribuciones t-Student univariados considerados como los kernel propuestos para cada componente del vector θ . Los grados de libertad para estos kernel propuestos t-Student son iguales a 10. Las tasas de aceptación de cada componente fueron 31.76%, 33.20%, 19.70%, y 83.86% respectivamente.

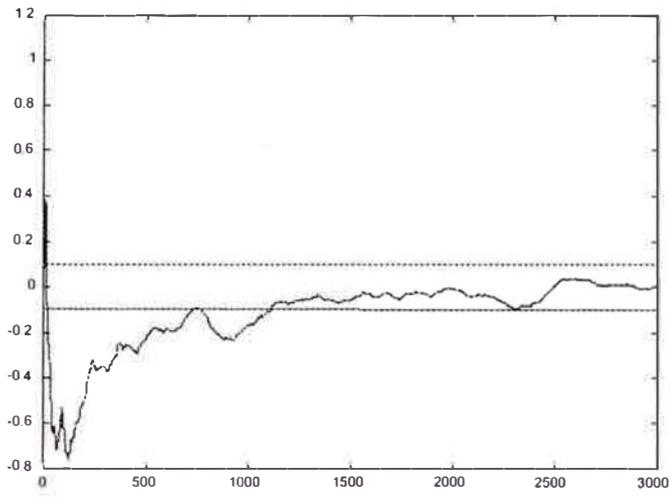


Figura 4.10: *ploteos CUMSUM para estimativa de la media α (datos reales)*

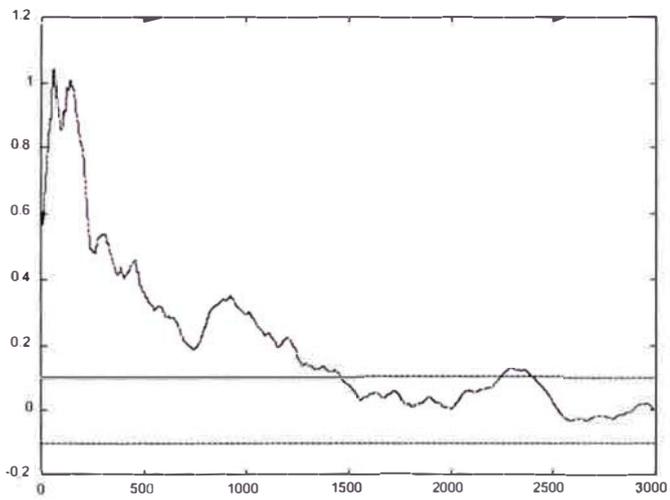


Figura 4.11: *ploteos CUMSUM para estimativa de la media β (datos reales)*

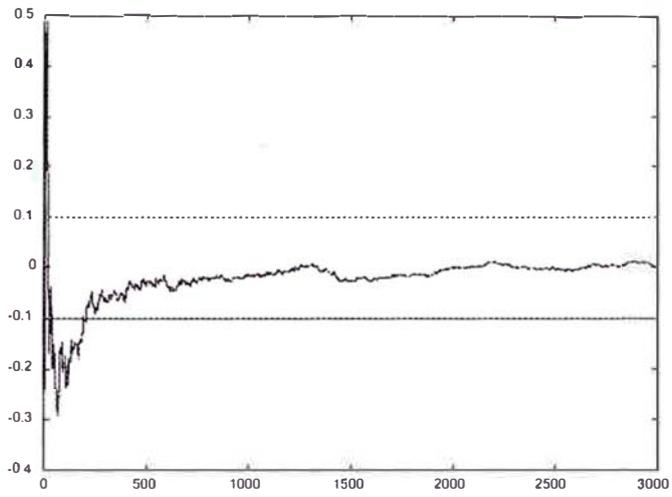


Figura 4.12: *ploteos CUSUM para estimativa de la media ν (datos reales)*

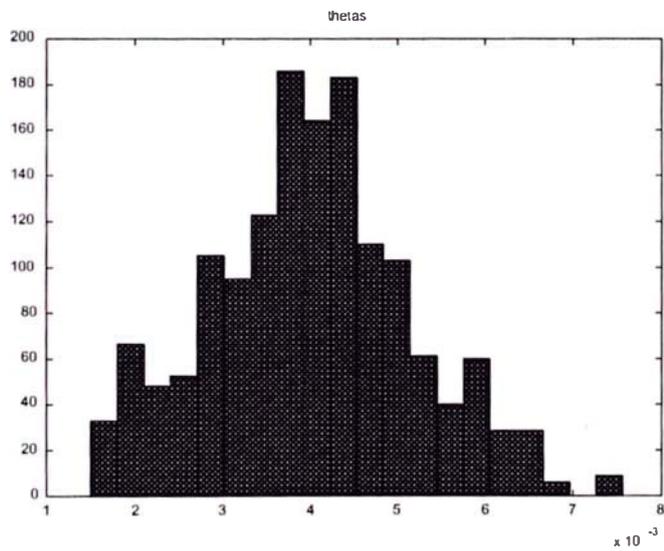


Figura 4.13: *Histograma de w (datos reales)*

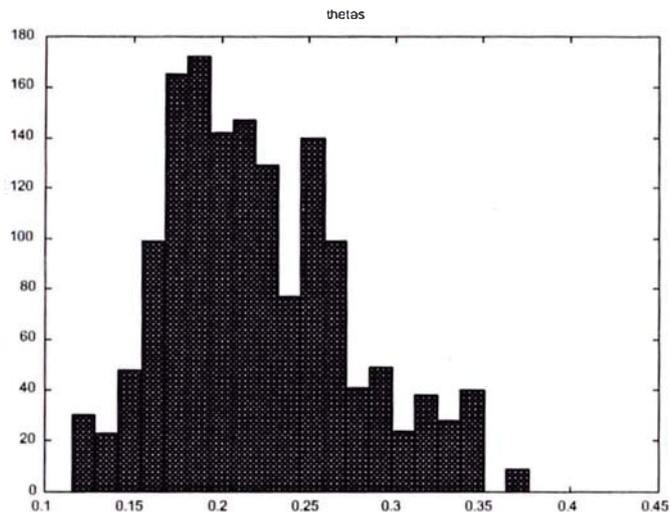


Figura 4.14: *Histograma de α (datos reales)*

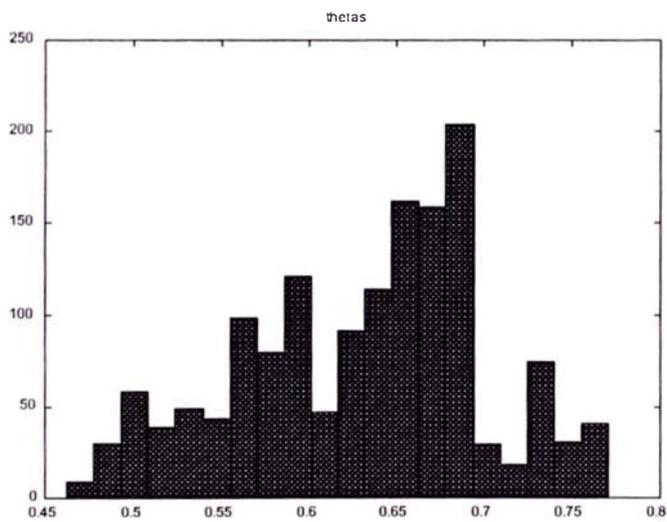


Figura 4.15: *Histograma de β (datos reales)*

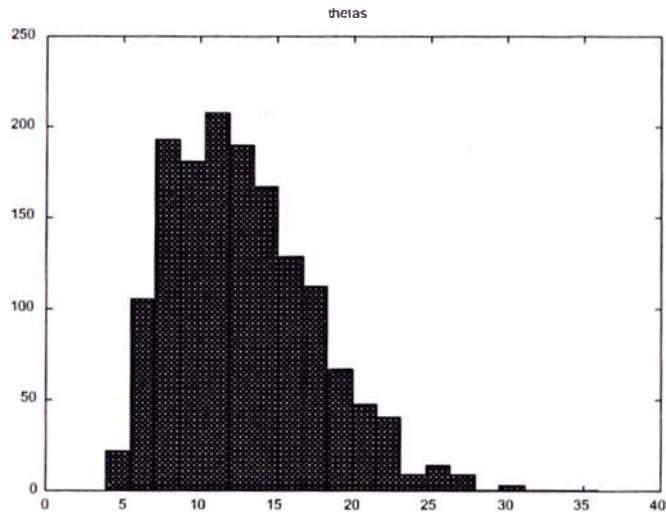


Figura 4.16: *Histograma de ν (datos reales)*

Parámetros	Máxima Verosimilitud	Metropolis-Hastings por componentes
w	0.0035 (0.0012)	0.0040 (0.0008)
α	0.2024 (0.0486)	0.2208 (0.0384)
β	0.6629 (0.0696)	0.6299 (0.0380)
ν	11.5848 (3.8789)	12.8839 (4.7057)

Tabla 4.3: *Estimación clásica y Bayesiana para un modelo GARCH(1,1) t-Student (datos reales)*

Capítulo 5

Conclusiones, Recomendaciones y Sugerencias

- 1) El presente trabajo de tesis propone un nuevo algoritmo MCMC Metropolis-Hastings por componentes para propósitos de inferencia estadística Bayesiana sobre el modelo GARCH(1,1) t-Student univariado. Los resultados encontrados para el estudio de datos artificiales son satisfactorios, validando el esquema MCMC de Metropolis-Hastings por componentes propuesto en esta tesis y luego ser usada en datos de retornos del Nuevo Sol versus el Dólar americano.
- 2) Los parámetros utilizados en la determinación de los kernel propuestos para cada componente del vector θ son obtenidos por experimentación. El gráfico de diagnóstico CUMSUM para la determinación del número de iteraciones necesarias para que la cadena converja nos indica a partir de que iteración debemos tomar los valores de θ generados por el algoritmo MCMC Metropolis-Hastings por componentes y luego usarla en las estimaciones de los resúmenes de interés de la densidad a posteriori de θ . Para los datos artificiales, los gráficos CUMSUM nos indican que aproximadamente a partir de la iteración 800 podemos usar los valores generados por el algoritmo MCMC Metropolis-Hastings por componentes. Para los datos de retornos del Nuevo Sol versus el Dólar americano, los gráficos CUMSUM nos indican que aproximadamente a partir de la iteración 2500 podemos usar los valores generados por el algoritmo MCMC Metropolis-Hastings por componentes. Esta último se deriva pues existe una correlación alta entre los componentes w y α , y entre w y β .

- 3) Las tasas de aceptación obtenidas por el uso del algoritmo MCMC Metropolis-Hastings por componentes son razonables tanto para el caso de datos artificiales y datos reales, para el conjunto de parámetros de calibración obtenidos por experimentación. Se debe mencionar que el hecho de diseñar un esquema MCMC por componentes del vector θ , nos permite una mejor calibración dado que se tiene un control por separado de las tasas de aceptación.
- 4) Las estimaciones del vector de parámetros θ obtenidas bajo una aproximación Bayesiana a través del algoritmo MCMC Metropolis-Hastings por componentes son mostrados en conjunto con las estimaciones máximo verosímiles obtenidas bajo una aproximación clásica para indicar la coherencia en el proceso de inferencia estadística. Debemos advertir aquí que ambas aproximaciones son planteadas bajo principios diferentes y no cabe comparaciones. Ambas aproximaciones son formas alternativas de hacer inferencia estadística.
- 5) Se recomienda a los usuarios interesados en la utilización del algoritmo MCMC Metropolis-Hastings por componentes que se debe tener cuidado en una calibración adecuada de los parámetros de los kernel propuestos para obtener tasas de aceptación razonables. El tiempo de convergencia para los datos artificiales es relativamente rápido y depende en buena parte del tipo de programación utilizado y del procesador utilizado. El tiempo de convergencia para los datos de retornos del Nuevo Sol versus el Dólar americana es mas lento debido a que el número de observaciones es mayor al utilizado por los datos artificiales.
- 6) Finalmente, un trabajo a futuro sería extender los alcances del algoritmo MCMC de Metropolis-Hastings por componentes para modelos GARCH(1,1) t-Student regresivos univariados y para modelos GARCH(1,1) multivariados.

Apéndice A

Algunas Funciones de Densidades

A.1 Normal Univariada

Si $\Theta \sim N(\mu, \tau^2)$, entonces la f.d.p. es dada por

$$p(\theta|\mu, \tau^2) = \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{1}{2} \frac{(\theta - \mu)^2}{\tau^2}\right\}$$

donde $\theta \in \mathfrak{R}$, $\mu \in \mathfrak{R}$ y $\tau^2 > 0$. La media y varianza son dados por $E[\theta] = \mu$ y $Var[\theta] = \tau^2$ respectivamente.

A.2 Normal Multivariada

Si $\Theta \sim N_p(\mu, \Sigma)$ con $\mu = (\mu_1, \mu_2, \dots, \mu_p)' \in \mathfrak{R}^p$, $\Sigma = \Sigma_{p \times p}$ es una matriz definida positiva. Entonces la f.d.p. es dada por

$$p(\theta|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2} (\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2} (\theta - \mu)' \Sigma^{-1} (\theta - \mu)\right\}$$

$\theta \in \mathfrak{R}^p$, la media y matriz de varianza-covarianza son dados respectivamente por $E[\theta] = \mu$ y Σ respectivamente

A.3 t-Student Multivariada

Si $\Theta \sim t_p(\alpha, \mu, \Sigma)$ con α grados de libertad, vector de parámetros de localización $\mu \in \mathfrak{R}^p$, matriz de escala $\Sigma = \Sigma_{p \times p}$. Entonces la f.d.p. es dada por

$$p(\theta|\alpha, \mu, \Sigma) = \frac{\Gamma[(\alpha + p)/2]}{\Gamma[\alpha/2] (\det \Sigma)^{1/2} (\alpha\pi)^{p/2}} \left[1 + \frac{1}{\alpha} (\theta - \mu)' \Sigma^{-1} (\theta - \mu)\right]^{-(\alpha+p)/2}$$

$\theta \in \mathbb{R}^p$, la media y matriz de varianza-covarianza son dados respectivamente por $E[\theta] = \mu$ si $\alpha > 1$, y $\alpha\Sigma/(\alpha - 2)$ si $\alpha > 2$ respectivamente

A.4 Gamma - 1era parametrización

Si $\Theta \sim Ga(\alpha, \beta)$, entonces la f.d.p. es dada por

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta}$$

donde $\theta \in \langle 0, \infty \rangle$, $\alpha > 0$ y $\beta > 0$. La media y varianza son dados por $E[\theta] = \frac{\alpha}{\beta}$ y $Var[\theta] = \frac{\alpha}{\beta^2}$ respectivamente.

A.5 Gamma - 2da parametrización

Si $\Theta \sim Ga(\alpha, \beta)$, entonces la f.d.p. es dada por

$$p(\theta|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \theta^{\alpha-1} e^{-\theta/\beta}$$

donde $\theta \in \langle 0, \infty \rangle$, $\alpha > 0$ y $\beta > 0$. La media y varianza son dados por $E[\theta] = \alpha\beta$ y $Var[\theta] = \alpha\beta^2$ respectivamente.

A.6 Gamma Inversa

Si $\Theta \sim GI(\alpha, \beta)$, entonces la f.d.p. es dada por

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)\theta^{\alpha+1}} e^{-\frac{\beta}{\theta}}$$

donde $\theta \in \langle 0, \infty \rangle$, $\alpha > 0$ y $\beta > 0$. La media y varianza son dados por $E[\theta] = \frac{\beta}{(\alpha-1)}$ si $\alpha > 1$, y $Var[\theta] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ si $\alpha > 2$, respectivamente.

Apéndice B

1o	2o	3o	4o	5o	6o	7o	8o	9o	10o	11o	12o	13o	14o	15o
2.3	2.3	2.	2.4	2.5	2.58	2.62	2.66	2.65	2.65	2.	2.81	2.84	2.93	3.061
2.34	2.3	2.41	2.46	2.56	2.5	2.62	2.66	2.65	2.	2.72	2.80	2.84	2.93	3.057
2.3	2.3	2.41	2.4	2.5	2.5	2.62	2.66	2.65	2.64	2.	2.80	2.84	2.93	3.054
2.34	2.3	2.41	2.46	2.5	2.60	2.62	2.66	2.65	2.	2.	2.7	2.84	2.	3.044
2.34	2.35	2.	2.4	2.5	2.61	2.63	2.66	2.66	2.	2.73	2.7	2.84	2.	3.048
2.3	2.35	2.	2.4	2.5	2.	2.63	2.66	2.6	2.	2.	2.80	2.84	2.	3.044
2.3	2.35	2.41	2.4	2.5	2.62	2.63	2.66	2.65	2.65	2.72	2.80	2.85	2.	3.041
2.3	2.35	2.42	2.47	2.5	2.63	2.63	2.66	2.65	2.66	2.72	2.80	2.	2.95	3.047
2.3	2.35	2.42	2.4	2.57	2.63	2.63	2.6	2.65	2.66	2.73	2.79	2.	2.	3.049
2.35	2.35	2.42	2.4	2.56	2.63	2.64	2.6	2.	2.67	2.73	2.7	2.	2.95	3.051
2.34	2.35	2.4	2.46	2.56	2.62	2.64	2.6	2.64	2.6	2.	2.7	2.	2.	3.059
2.34	2.35	2.43	2.4	2.56	2.62	2.	2.6	2.	2.72	2.73	2.80	2.86	2.	3.065
2.34	2.3	2.43	2.46	2.5	2.62	2.64	2.6	2.	2.6	2.	2.81	2.87	2.94	3.077
2.3	2.3	2.44	2.46	2.58	2.62	2.64	2.6	2.65	2.71	2.	2.81	2.8	2.94	3.072
2.35	2.35	2.45	2.46	2.5	2.62	2.	2.6	2.65	2.71	2.74	2.	2.8	2.	3.08
2.3	2.35	2.45	2.4	2.5	2.	2.6	2.6	2.	2.70	2.	2.81	2.8	2.96	3.09
2.35	2.3	2.44	2.47	2.5	2.62	2.65	2.6	2.65	2.70	2.	2.81	2.8	2.	3.081
2.35	2.35	2.44	2.47	2.59	2.	2.65	2.6	2.6	2.71	2.73	2.81	2.8	2.96	3.084
2.3	2.3	2.44	2.4	2.60	2.62	2.66	2.6	2.65	2.71	2.	2.	2.90	2.97	3.082
2.36	2.3	2.44	2.46	2.59	2.63	2.6	2.65	2.65	2.72	2.74	2.80	2.92	3.00	3.088
2.3	2.3	2.4	2.47	2.5	2.63	2.6	2.65	2.65	2.72	2.	2.81	2.92	3.01	3.095
2.3	2.36	2.44	2.4	2.58	2.63	2.66	2.	2.65	2.73	2.	2.81	2.91	3.0	3.093
2.3	2.3	2.4	2.4	2.57	2.	2.66	2.	2.65	2.73	2.	2.81	2.91	3.0	3.095

16o	17o	18o	19o	20o	21o	22o	23o	24o	25o	26o	27o	28o	29o	30o
2.35	2.3	2.4	2.4	2.5	2.63	2.6	2.65	2.65	2.73	2.	2.81	2.91	3.03	3.096
2.3	2.3	2.45	2.4	2.57	2.	2.66	2.65	2.65	2.73	2.	2.81	2.93	3.05	3.096
2.35	2.36	2.44	2.4	2.57	2.64	2.6	2.	2.	2.72	2.76	2.81	2.93	3.0	3.092
2.3	2.3	2.44	2.48	2.5	2.65	2.6	2.	2.	2.72	2.7	2.82	2.	3.0	3.09
2.3	2.3	2.4	2.48	2.57	2.	2.6	2.	2.64	2.71	2.7	2.	2.93	3.05	3.094
2.35	2.3	2.4	2.48	2.5	2.	2.6	2.65	2.64	2.71	2.7	2.82	2.	3.05	3.094
2.35	2.3	2.4	2.48	2.5	2.64	2.66	2.65	2.64	2.71	2.7	2.	2.93	3.06	3.095
2.35	2.3	2.4	2.4	2.5	2.	2.66	2.	2.63	2.	2.7	2.	2.93	3.06	3.107
2.35	2.37	2.44	2.48	2.5	2.64	2.6	2.	2.	2.	2.80	2.82	2.93	3.0	3.114
2.3	2.37	2.44	2.4	2.5	2.	2.6	2.	2.	2.72	2.80	2.82	2.92	3.03	3.118
2.3	2.37	2.44	2.4	2.5	2.	2.6	2.	2.64	2.72	2.	2.83	2.93	3.0	3.123
2.35	2.3	2.4	2.	2.5	2.	2.66	2.	2.	2.	2.79	2.	2.93	3.04	3.114
2.35	2.3	2.4	2.50	2.58	2.63	2.66	2.65	2.63	2.72	2.7	2.83	2.92	3.0	3.126
2.35	2.3	2.44	2.51	2.5	2.63	2.66	2.65	2.	2.72	2.7	2.82	2.92	3.03	3.13
2.35	2.3	2.44	2.51	2.5	2.63	2.66	2.	2.	2.71	2.7	2.82	2.92	3.02	3.149
2.3	2.40	2.44	2.	2.58	2.63	2.6	2.	2.64	2.71	2.7	2.82	2.92	3.0	3.147
2.3	2.41	2.44	2.51	2.5	2.	2.66	2.	2.64	2.71	2.7	2.83	2.91	3.03	3.138
2.3	2.42	2.44	2.51	2.58	2.63	2.6	2.	2.64	2.71	2.80	2.83	2.91	3.0	3.134
2.3	2.	2.44	2.	2.5	2.	2.66	2.65	2.	2.71	2.80	2.	2.91	3.04	3.145
2.3	2.41	2.44	2.53	2.5	2.64	2.66	2.65	2.	2.71	2.81	2.	2.90	3.0	3.146
2.35	2.41	2.4	2.53	2.5	2.64	2.6	2.	2.64	2.71	2.	2.	2.90	3.0	3.14
2.3	2.41	2.44	2.	2.5	2.	2.6	2.	2.	2.71	2.81	2.84	2.91	3.0	3.134
2.3	2.	2.4	2.53	2.5	2.	2.66	2.	2.	2.71	2.81	2.84	2.	3.04	3.133
2.3	2.41	2.44	2.5	2.5	2.64	2.6	2.65	2.	2.71	2.	2.	2.92	3.0	3.134
2.3	2.41	2.45	2.5	2.5	2.63	2.6	2.65	2.64	2.	2.81	2.	2.	3.03	3.13
2.3	2.41	2.46	2.5	2.5	2.63	2.6	2.6	2.	2.71	2.81	2.	2.92	3.0	3.13
	2.41	2.4	2.56	2.5	2.62	2.6	2.65	2.	2.71	2.81	2.84	2.	3.05	3.137

Nota.- Las observaciones estan ordenadas por columnas

Apéndice C

Ley de esperanzas iteradas

Muchas veces es más fácil evaluar la esperanza de una variable aleatoria Y , hallando primero su esperanza condicional sobre otra variable aleatoria X , y luego tomar la esperanza de la expresión resultante con respecto a la distribución de X . Es decir, la operación es

$$E[E(Y|X = x)] = \int \left[\int yp(y|x)dy \right] p(x)dx$$

arreglando y sustituyendo la densidad conjunta para el producto de la condicional y densidad marginal, tenemos que

$$\int \int yp(y, x)dydx = E[Y]$$

la cual es la esperanza incondicional de Y .

El resultado se generaliza para cualquier función de Y , es decir

$$E[E(g(Y)|X = x)] = E[g(Y)]$$

Bibliografía

- [1] Luc Bauwens and Michel Lubrano. Bayesian Inference on GARCH models using the gibbs sample. *Econometrics Journal*, 1:C23–C46, 1998.
- [2] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Second Edition, Springer-Verlag, 1985.
- [3] E. K. Berndt, B. H. Hall, R. E. Hall, and J.A. Hausman. Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, 3:653–665, 1974.
- [4] Tim Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal Econometrics*, 31:307–27, 1986.
- [5] George E. P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company, 1973.
- [6] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49:327–35, 1994.
- [7] Robert Engle. Autoregressive Conditional Heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 1-50:987–1008, 1982.
- [8] J. Geweke. Exact predictive densities in linear models with ARCH disturbances. *Journal of Econometrics*, 40:63–86, 1989.
- [9] J. Geweke. Bayesian comparison of econometrics models. Research Department, Federal Reserve Bank of Minneapolis, 1994.
- [10] P.G. Hoel, S. C. Port, and C.J. Stone. *Introduction to Stochastic Processes*. Houghton Mifflin, 1972.

- [11] F. Kleibergen and H. K. van Dijk. Non-stationarity in GARCH models. *Journal Applied Econometrics*, 8:41–61, 1993.
- [12] N. Metropolis, A. W. Rosenbluth, and M.N. Teller. Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21:1087–1091, 1953.
- [13] L. A. Navarro Huamani. Estimación de la Volatilidad de la tasa de intercambio del Nuevo Sol versus el Dólar americano. Núcleo de Estadística Computacional - NEC, UNI, 2002.
- [14] L. A. Navarro Huamani. Metropolis - hastings por componentes para modelos garch. IX Congreso Latinoamericano de Probabilidad y Estadística Matemática (CLAPEM), Punta del Este - Uruguay, 2004.
- [15] J.C. Naylor and A.F.M Smith. Econometrics Illustrations of Novel Numerical Integration Strategies for Bayesian Statistics. *Journal of Econometrics*, 38:103–125, 1988.
- [16] D.B. Nelson. Stationarity and Persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990.
- [17] A. O'Hagan. *Bayesian Inference, Volume 2B of Kendall's Advanced Theory of Statistics*. Edward Arnold, London, 1994.
- [18] B. D. Ripley. *Stochastic Simulation*. New York, 1987.
- [19] G. O. Roberts and A. F. Smith. Some convergence theory for Markov Chain Monte Carlo. *Stochastic Processes and their Applications*, 49:207–16, 1993.
- [20] A. F. Smith and G. O. Roberts. Bayesian Computation via the Gibbs Sampler and related Markov Chain Monte Carlo Methods. *Journal Royal Statistical Society*, 55:3–23, 1993.
- [21] L. Tierney. Markov Chain for exploring posterior distributions (with discussion). *Annals of Statistics*, 22:528–550, 1994.