



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Modelo de efectos de mezclas en ecuaciones
diferenciales estocásticas

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

PRESENTA:

GABRIELA STEPHANIA REVUELTAS
HERNÁNDEZ



DIRECTOR DE TESIS:
DR. FERNANDO BALTAZAR LARIOS

2014

México, D. F.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos del alumno

Revueltas
Hernández
Gabriela Stephania
56 96 53 42
Universidad Nacional Autónoma de
México
Facultad de Ciencias
Actuaría
306178513

2. Datos del tutor

Dr.
Fernando
Baltazar
Larios

3. Datos del sinodal 1

Dra.
Ruth Selene
Fuentes
García

4. Datos del sinodal 2

Dra.
María Asunción Begoña
Fernández
Fernández

5. Datos del sinodal 3

Dra.
Ana
Meda
Guardiola

6. Datos del sinodal 4

Act.
Harim
García
Lamont

7. Datos del trabajo escrito

Modelo de efectos de mezclas en ecuaciones diferenciales estocásticas
80 p
2015

Agradecimientos

A la Universidad Nacional Autónoma de México, por permitirme formar parte de esta gran institución.

A Fernando, por la paciencia, el tiempo brindado para la realización de este trabajo y en especial por todo el apoyo y la confianza sin la cual no hubiera sido posible la conclusión de esta tesis.

A los miembros del jurado Dra. Ruth Fuentes, Dra. Ana Meda, Dra. Begonña Fernández y Act. Harim García por su dedicación en la revisión del trabajo, así como sus valiosos comentarios y aportaciones.

A mis padres, por darme todas las herramientas necesarias para llegar hasta este punto.

A Einar, por la complicidad de haber crecido juntos.

A mis abuelos, por el amor, el apoyo y el tiempo dedicado todos estos años.

A Kuri, por compartir conmigo cada momento con tanto cariño.

A mis amigos, (que no pierden importancia por no ser nombrados) porque sin ustedes esto no sería igual. A los compañeros que formaron parte del camino por la facultad.

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM "Proyecto PAPIIT IA102214", Estimación de ecuaciones diferenciales estocásticas con efectos aleatorios. Agradezco a la DGAPA-UNAM la beca recibida.

Índice general

Introducción	I
1. Ecuaciones diferenciales estocásticas	1
1.1. Algunos conceptos básicos	1
1.2. Integral de Itô y ecuaciones diferenciales estocásticas	4
1.2.1. Fórmula de Itô	10
1.2.2. Transformación de Lamperti.	14
1.3. Difusiones	15
1.3.1. Propiedades de las difusiones	20
1.3.2. Simulación de difusiones	26
1.3.3. Estimación	29
1.3.4. Puentes de difusiones	37
2. Modelo de efectos de mezclas	39
2.1. Modelos Lineales de Efectos de Mezclas	39
2.1.1. Estimación Máximo Verosímil	41
2.1.2. Generalización del modelo	46
3. Modelo de efectos de mezclas en ecuaciones diferenciales estocásticas	48
3.1. El modelo	49
3.2. Función de verosimilitud	49
3.3. El caso del Movimiento Browniano geométrico	50
3.3.1. Estimación con efectos aleatorios	52
3.3.2. Estimación sin efectos aleatorios	57
4. Simulación y aplicaciones	58
4.1. Simulación del movimiento Browniano geométrico	58
4.2. Simulación con puentes de difusiones	62
4.3. Aplicación	66
A. Descomposición ortogonal-triangular	73
B. Códigos	74

Introducción

El modelo de efectos aleatorios en ecuaciones diferenciales estocásticas como cualquier modelo matemático busca estudiar y analizar el comportamiento de un sistema o experimento para poder comprenderlo y hacer predicciones. La característica primordial de este tipo de modelos es que permiten considerar simultáneamente la aleatoriedad del fenómeno y la variabilidad de la unidad experimental, por tal razón, la principal área de aplicación se concentra en experimentos biomédicos.

Supongamos que tenemos un experimento en el cual se realizan repetidas mediciones en el tiempo a cada individuo de una población, y que el conjunto de observaciones de cada individuo puede modelarse a partir de una ecuación diferencial estocástica. Si además, consideramos que existen diferencias entre los individuos, podemos incorporar un efecto aleatorio al modelo, entonces resulta una ecuación diferencial estocástica donde el coeficiente de deriva está gobernado por una ley de probabilidad, lo que permite tener una mejor descripción de las variaciones del experimento.

En el presente trabajo, se busca encontrar los estimadores de los parámetros del modelo de efectos de mezclas en ecuaciones diferenciales estocásticas a partir de la construcción de la función de verosimilitud considerando ciertos supuestos como normalidad de los efectos aleatorios e independencia de los efectos aleatorios y el movimiento Browniano. Sin embargo, en general no es posible encontrar expresiones explícitas para la verosimilitud o para los estimadores. Por consiguiente, se supone en particular, que la ecuación diferencial estocástica es aquella que tiene como solución al movimiento Browniano geométrico, a partir de la cual sí es posible encontrarlos.

Con el fin de comprender el modelo completo, se expone en el primer capítulo, una breve introducción a la integración estocástica; iniciando con algunos conceptos básicos para seguir con la construcción de la integral de Itô, así como sus características más importantes. Luego, se presentan las ecuaciones diferenciales estocásticas junto con el teorema de existencia y unicidad de la solución de las mismas. Posteriormente, se enuncia la definición de difusión, sus principales propiedades y se describen dos métodos de simulación. Además, se da un resumen de los métodos más utilizados para hacer inferencia sobre los parámetros

de las ecuaciones diferenciales estocásticas. Finalmente, se define un puente de difusiones y se habla de un algoritmo para generarlos.

Por otra parte, en el segundo capítulo, se introduce el modelo lineal de efectos de mezclas, el desarrollo de la función de máxima verosimilitud y sus respectivos estimadores. Estos cálculos plantean la base para hacer la estimación de los parámetros del modelo al incorporarlos a la ecuación diferencial estocástica.

En el tercer capítulo, se integran ambos resultados para la construcción del modelo, se encuentra una expresión para la función de verosimilitud. Luego, se obtienen expresiones cerradas de los estimadores de los parámetros del modelo con la ecuación diferencial estocástica cuya solución es un movimiento Browniano geométrico. Además se encuentran las varianzas asintóticas de cada estimador y las condiciones necesarias para que los estimadores existan.

Finalmente, en el cuarto capítulo, se hace un análisis de los estimadores a partir de simulaciones, es decir, se realizan simulaciones de trayectorias para calcular los estimadores. Luego, se plantea como un modelo de información faltante donde la idea consiste en utilizar puentes de difusiones para completarla con el fin de mejorar los estimadores. Se concluye con la aplicación del modelo a un conjunto de datos que representan la evolución de la disolución de un medicamento.

Capítulo 1

Ecuaciones diferenciales estocásticas

Las ecuaciones diferenciales estocásticas surgen al incorporar un elemento estocástico o “ruido” a una ecuación diferencial, lo que permite en muchas ocasiones obtener un modelo matemático que describa de forma más realista el sistema en estudio.

Para ejemplificarlo, consideremos el modelo simple de crecimiento poblacional

$$\frac{dX_t}{dt} = a(t)X(t), \quad X(0) = X_0 \text{ (constante),}$$

donde $X(t)$ representa el tamaño de la población al tiempo t y $a(t)$ es la tasa de crecimiento. Pero, si $a(t)$ no está completamente determinada ya que depende de elementos aleatorios que provienen del ambiente, es decir, $a(t) = r(t) + \text{“ruido”}$ con $r(t)$ determinista y la parte aleatoria puede describirse mediante una función de distribución, obtenemos una ecuación diferencial estocástica.

1.1. Algunos conceptos básicos

Comenzaremos recordando algunos conceptos necesarios para la construcción de la ecuación diferencial estocástica y sus principales propiedades. Consideremos el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

Definición 1.1.1. *Un proceso estocástico continuo $\{X_t(\omega)\}_{t \in T}$ es una familia de variables aleatorias parametrizadas por un conjunto $T \in \mathbb{R}^+$, usualmente el tiempo.*

Un proceso estocástico $\{X_t(\omega)\}_{t \in T}$ puede pensarse como función de dos variables, donde para cada $t \in T$, $\omega \rightarrow X_t(\omega)$ es una variable aleatoria y para cada $\omega \in \Omega$, $t \rightarrow X_t(\omega)$ es una trayectoria. Entonces se podría interpretar a $X_t(\omega)$ como la posición del individuo ω al tiempo t .

Para simplificar la notación denotaremos a los procesos estocásticos como $\{X_t\}_{t \in T}$.

Uno de los procesos estocásticos más importantes es el movimiento Browniano o proceso de Wiener que fue descubierto por el botánico Robert Brown al observar la trayectoria irregular que siguen las partículas de polen suspendidas en cierta sustancia. Posteriormente, la definición formal y la prueba de existencia fue dada por el matemático Norbert Wiener.

Definición. 1.1.2. *Un movimiento Browniano $\{B_t\}_{t \geq 0}$ es un proceso estocástico continuo tal que:*

1. $B_0 = 0$ con probabilidad 1.
2. $\{B_t\}_{t \geq 0}$ tiene incrementos independientes: si $0 = t_0 < t_1 < \dots < t_n$, entonces $B_{t_i} - B_{t_{i-1}}$, son independientes para $1 \leq i \leq n$.
3. $\{B_t\}_{t \geq 0}$ tiene incrementos estacionarios: $B_{t+s} - B_t$ tiene la misma distribución que B_s para toda s y t . Y además $B_s \sim N(0, s)$.

Antes de referir las principales propiedades del movimiento Browniano, mencionaremos dos conceptos preliminares: procesos de Markov y martingalas.

Primero, recordemos que un proceso de Markov es aquel que cumple la siguiente propiedad: $\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s)$ para $s < t$ (o propiedad de Markov). Luego, recordemos que una *filtración* es una familia de sub- σ -álgebras de \mathcal{F} que cumplen: $\mathcal{F}_s \subset \mathcal{F}_t$, $s < t$.

Se le llama filtración natural o canónica de $\{X_t\}_{t \in T}$ a la σ -álgebra generada por el proceso. Usualmente se piensa a la σ -álgebra \mathcal{F}_t como la información contenida en el proceso hasta el tiempo t , entonces si $\mathcal{F}_s \subset \mathcal{F}_t$, indica que \mathcal{F}_t contiene más información que \mathcal{F}_s .

Definición. 1.1.3. *Una martingala $\{M_t\}_{t \geq 0}$ con respecto a la filtración $\{\mathcal{F}_t\}_{t \geq 0}$ es un proceso estocástico continuo que cumple:*

1. M_t es \mathcal{F}_t -medible.
2. $\mathbb{E}[|M_t|] < \infty$.
3. $\mathbb{E}[M_t | \mathcal{F}_s] = M_s$, $s < t$.

Propiedades del movimiento Browniano:

1. Es un proceso de Markov.
2. Es una martingala.
3. Es no diferenciable en ningún punto.
4. No tiene variación acotada.
5. Tiene variación cuadrática acotada.

Probaremos las propiedades 2 y 5 que se utilizarán más adelante. La demostración del resto de las propiedades pueden encontrarse en [11] y [19].

Proposición 1.1.1. *Sea $\{B_t\}_{t \geq 0}$ un movimiento Browniano, para toda $t \geq 0$ y Δ_n una sucesión de particiones del tiempo $[0, t]$, cuya norma tiende a cero, entonces la sucesión $\sum_{t_i \in \Delta_n} (B_{t_i} - B_{t_{i-1}})^2$ converge a t en \mathcal{L}^2 . Es decir, la variación cuadrática de una trayectoria es la longitud del intervalo.*

Demostración. Basta probar que $\sum_i (B_{t_i} - B_{t_{i-1}})^2 - t$ converge a cero en \mathcal{L}^2 . Denotemos $B_{t_i} - B_{t_{i-1}} = \Delta B_{t_i}$ y $t_i - t_{i-1} = \Delta t_i$

$$\mathbb{E} \left[\left(\sum_i (\Delta B_{t_i})^2 - t \right)^2 \right] = \mathbb{E} \left[\left(\sum_i (\Delta B_{t_i})^2 - (\Delta t_i) \right)^2 \right]$$

desarrollamos el cuadrado

$$\begin{aligned} &= \mathbb{E} \left[\sum_i \sum_j (\Delta B_{t_i})^2 (\Delta B_{t_j})^2 - 2 \sum_i \sum_j (\Delta B_{t_i})^2 (\Delta t_i) \right. \\ &\quad \left. + \sum_i \sum_j \Delta t_i \Delta t_j \right] \\ &= \mathbb{E} \left[\sum_i (\Delta B_{t_i})^2 + \sum_{i \neq j} (\Delta B_{t_i})^2 (\Delta B_{t_j})^2 \right. \\ &\quad \left. - 2 \sum_i \sum_j (\Delta B_{t_i})^2 \Delta t_i + \sum_i \sum_j \Delta t_i \Delta t_j \right] \end{aligned}$$

utilizamos la linealidad de la esperanza y que $\mathbb{E}[(\Delta B_{t_i})^2] = \Delta t_i$ y $\mathbb{E}[(\Delta B_{t_i})^4] = 3(\Delta t_i)^2$

$$\begin{aligned} &= \sum_i 3(\Delta t_i)^2 + \sum_{i \neq j} \Delta t_i \Delta t_j - \sum_i \sum_j \Delta t_i \Delta t_j \\ &= \sum_i 2(\Delta t_i)^2 \\ &= 2\Delta_n t \rightarrow 0. \end{aligned}$$

□

Proposición 1.1.2. *El movimiento Browniano es una martingala continua respecto a la σ -álgebra \mathcal{F}_t generada por $\{B_s\}_{s \leq t}$.*

Demostración. Claramente el proceso es adaptado a la filtración natural. Es integrable ya que $\mathbb{E}[|B_t|] \leq \mathbb{E}[|B_t|^2] = t$.

Y finalmente, probamos la propiedad de martingala,

$$\begin{aligned}\mathbb{E}[B_t | f_s] &= \mathbb{E}[B_t - B_s + B_s | f_s] \\ &= \mathbb{E}[B_t - B_s | f_s] + \mathbb{E}[B_s | f_s] \\ &= \mathbb{E}[B_t - B_s] + B_s = B_s.\end{aligned}$$

□

Con los resultados previos podemos iniciar con el estudio de las ecuaciones diferenciales estocásticas.

1.2. Integral de Itô y ecuaciones diferenciales estocásticas

Consideremos la ecuación diferencial estocástica

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t,$$

donde $\{B_t\}_{t \geq 0}$ es un movimiento Browniano.

O bien, en notación integral

$$X_t = X_0 + \int_0^t b(u, X_u)du + \int_0^t \sigma(u, X_u)dB_u.$$

Como la integral con respecto al movimiento Browniano no puede ser expresada en términos de la integral de Riemman ya que no tiene variación acotada, se define la integral de Itô para funciones $f : [0, \infty) \times \Omega \rightarrow \mathbb{R}$ de la siguiente manera.

Definición 1.2.1. *Se define la integral estocástica (o de Itô) como*

$$I(f)(\omega) = \int_0^T f(t, \omega)dB_t.$$

Consideremos $\mathcal{V} = \mathcal{V}(S, T)$ la clase de funciones $f(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R}$ tales que:

- i. $(t, \omega) \rightarrow f(t, \omega)$ es $(\mathcal{B} \times \mathcal{F})$ -medible.
- ii. $f(t, \omega)$ es \mathcal{F} -adaptado.
- iii. $\mathbb{E}[\int_S^T f(t, \omega)^2 dt] < \infty$.

Nota: $\mathcal{V} \in \mathcal{L}^2_{(d\mathbb{P} \times dt)}$.

Primero se construye la integral de Itô a partir de funciones sencillas llamadas funciones simples denotadas por ϕ y posteriormente se generaliza probando que cada $f \in \mathcal{V}$ puede aproximarse por funciones simples.

Consideremos $\phi \in \mathcal{V}$ tal que

$$\phi = \sum_{j=0}^{n-1} a_j(\omega) \mathbb{I}_{[t_j, t_{j+1})}(t),$$

donde a_j debe ser \mathcal{F}_{t_j} -medible, $\mathbb{E}(a_j^2) < \infty$, $0 = t_1 \leq t_2 \leq \dots \leq t_j \leq \dots \leq t_n = T$.

Definimos la integral de Itô para funciones simples como:

$$I(\phi) = \int_0^T \phi(t, \omega) dB_t(\omega) = \sum_{j=0}^{n-1} a_j(\omega) [B_{t_{j+1}} - B_{t_j}](\omega). \quad (1.1)$$

Con el siguiente lema es posible extender la definición de la integral de Itô para cualquier función en \mathcal{V} .

Lema 1.2.1 (Isometría de Itô). *Para un proceso simple $\phi(t, \omega)$ acotado*

$$\|I(\phi)\|_{\mathcal{L}^2_{(d\mathbb{P})}} = \|\phi\|_{\mathcal{L}^2_{(d\mathbb{P} \times dt)}}$$

es decir,

$$\mathbb{E} \left[\left(\int_0^T \phi(t, \omega) dB_t(\omega) \right)^2 \right] = \mathbb{E} \left[\int_0^T \phi(t, \omega)^2 dt \right].$$

Demostración. Desarrollamos ambos lados de la igualdad para ver que llegamos a lo mismo.

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^T \phi(t, \omega) dB_t(\omega) \right)^2 \right] &= \mathbb{E} \left[\left(\sum_{j=0}^{n-1} a_j(\omega) (B_{t_{j+1}} - B_{t_j})(\omega) \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{j=0}^{n-1} a_j^2(\omega) (B_{t_{j+1}} - B_{t_j})^2 \right. \\ &\quad \left. + 2 \sum_{i < j} a_i a_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j}) \right], \end{aligned}$$

como a_j es independiente a $(B_{t_{j+1}} - B_{t_j})$ y $\mathbb{E}(B_{t_{j+1}} - B_{t_j}) = 0$, los términos cruzados son cero

$$\begin{aligned} \mathbb{E} \left[\sum_{j=0}^{n-1} a_j^2(\omega) (B_{t_{j+1}} - B_{t_j})^2 \right] &= \sum_{j=0}^{n-1} \mathbb{E}[a_j^2(\omega)] \mathbb{E}[(B_{t_{j+1}} - B_{t_j})^2] \\ &= \sum_{j=0}^{n-1} \mathbb{E}[a_j^2(\omega)] (t_{j+1} - t_j). \end{aligned}$$

Por otro lado,

$$\begin{aligned}
\mathbb{E} \left[\int_0^T \phi(t, \omega)^2 dt \right] &= \mathbb{E} \left[\int_0^T \left(\sum_{j=0}^{n-1} a_j(\omega) \mathbb{I}_{[t_j, t_{j+1})}(t) \right)^2 dt \right] \\
&= \mathbb{E} \left[\int_0^T \sum_{j=0}^{n-1} a_j^2(\omega) \mathbb{I}_{[t_j, t_{j+1})}(t) dt \right] \\
&= \mathbb{E} \left[\sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} a_j^2(\omega) dt \right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}[a_j^2(\omega)](t_{j+1} - t_j).
\end{aligned}$$

Por lo tanto, se cumple la igualdad. \square

Ahora, vamos a extender la definición de la integral de Itô para cualquier función en \mathcal{V} .

Primero, consideremos $g \in \mathcal{V}$ acotada y continua para toda ω , entonces existe $\phi_n \in \mathcal{V}$ un proceso simple tal que

$$\mathbb{E} \left[\int_0^T (g - \phi_n)^2 dt \right] \rightarrow 0 \quad \text{si } n \rightarrow \infty.$$

Prueba: definimos $\phi_n = \sum_{j=0}^{n-1} g(t_j, \omega) \mathbb{I}_{[t_j, t_{j+1})}(t)$, con $g \in \mathcal{V}$ y continua para toda ω , ϕ_n es un proceso simple y

$$\int_0^T (g - \phi_n)^2 dt \rightarrow 0 \quad \text{si } n \rightarrow \infty \quad \text{para cada } \omega,$$

entonces, por el teorema de convergencia acotada (véase [9]),

$$\mathbb{E} \left[\int_0^T (g - \phi_n)^2 dt \right] \rightarrow 0 \quad \text{cuando } n \text{ tiende a infinito.}$$

Luego, sea $h \in \mathcal{V}$ acotada entonces existen funciones $g_n(\cdot, \omega) \in \mathcal{V}$ continuas para todo ω y n tales que

$$\mathbb{E} \left[\int_0^T (h - g_n)^2 dt \right] \rightarrow 0 \quad \text{si } n \rightarrow \infty.$$

Prueba: como h acotada $|h(t, \omega)| \leq M$ para todo (t, ω) . Consideremos ψ_n funciones no negativas y continuas en los reales que cumplan:

i) $\psi_n(x) = 0$ para $x \leq -\frac{1}{n}$ y $x \geq 0$.

ii) $\int_{-\infty}^{\infty} \psi_n(x) dx = 1$.

Definimos $g_n(t, \omega) = \int_0^t \psi_n(s - t) h(s, \omega) ds$, entonces $g_n(\cdot, \omega)$ es continua para cada ω y acotada $|g_n(t, \omega)| \leq M$, además $g_n(t, \cdot)$ es \mathcal{F}_t -medible para toda t .

Por el teorema de convergencia uniforme (véase [9]),

$$\lim_{n \rightarrow \infty} \int_0^T (g_n(s, \omega) - h(s, \omega))^2 ds = \int_0^T \lim_{n \rightarrow \infty} (g_n(s, \omega) - h(s, \omega))^2 ds,$$

como $\lim_{n \rightarrow \infty} g_n(t, \omega) = h(t, \omega)$,

$$\lim_{n \rightarrow \infty} \int_0^T (g_n(s, \omega) - h(s, \omega))^2 ds = 0,$$

por lo tanto, $\mathbb{E} \left[\int_0^T (h - g_n)^2 dt \right] \rightarrow 0$ si $n \rightarrow \infty$.

Finalmente, consideremos $f \in \mathcal{V}$ existe una sucesión de funciones $h_n \in \mathcal{V}$ acotada para toda n

$$\mathbb{E} \left[\int_0^T (f - h_n)^2 dt \right] \rightarrow 0 \quad \text{si } n \rightarrow \infty.$$

Prueba: tomamos

$$h_n = \begin{cases} -n & f(t, \omega) < -n \\ f(t, \omega) & -n \leq f(t, \omega) \leq n \\ n & f(t, \omega) > n \end{cases}$$

Como h_n es acotada y converge a f puntualmente podemos usar el teorema de convergencia dominada (véase [9]), por lo tanto, h_n tiende a f en \mathcal{L}^2 .

Con esto termina la parte de aproximación de funciones simples a cualquier función en \mathcal{V} . Ahora ya podemos completar la definición de la integral de Itô.

Definición 1.2.2. Sea $f \in \mathcal{V}$, la integral estocástico de Itô se define:

$$I(f)(\omega) = \int_0^T f(t, \omega) dB_t = \lim_{n \rightarrow \infty} \int_0^T \phi_n(t, \omega) dB_t = \lim_{n \rightarrow \infty} I(\phi_n),$$

donde ϕ_n una sucesión de funciones simples.

Notemos que la sucesión ϕ_n existe como lo demostramos previamente y además notemos que $\int \phi_n dB_t$ forman una sucesión de Cauchy por lo que el límite también existe.

Corolario 1.2.2 (Isometría de Itô). Sea $f \in \mathcal{V}$,

$$\|I(f)\|_{\mathcal{L}^2_{(d\mathbb{P})}} = \|f\|_{\mathcal{L}^2_{(d\mathbb{P} \times dt)}}.$$

Demostración. Sea $\phi_n \in \mathcal{V}$, ya vimos que $\|\phi_n - f\|$ tiende a cero si n crece a infinito, entonces $\|\phi_n\|$ converge a $\|f\|$ en $\mathcal{L}^2_{(d\mathbb{P} \times dt)}$ y por otro lado, $\|I(\phi_n) - I(f)\|$

tiende a cero cuando n crece a infinito, por lo que $\|I(\phi_n)\|$ converge a $\|I(f)\|$ en $\mathcal{L}^2_{(d\mathbb{P})}$ luego, por la isometría de Itô para funciones simples

$$\|I(\phi_n)\|_{\mathcal{L}^2_{(d\mathbb{P})}} = \|\phi_n\|_{\mathcal{L}^2_{(d\mathbb{P} \times dt)}},$$

tomamos el limite

$$\|I(f)\|_{\mathcal{L}^2_{(d\mathbb{P})}} = \lim_{n \rightarrow \infty} \|I(\phi_n)\|_{\mathcal{L}^2_{(d\mathbb{P})}} = \lim_{n \rightarrow \infty} \|\phi_n\|_{\mathcal{L}^2_{(d\mathbb{P} \times dt)}} = \|f\|_{\mathcal{L}^2_{(d\mathbb{P} \times dt)}}.$$

□

Algunas propiedades de la integral de Itô para f y g funciones en $\mathcal{V}(0, T)$ y $0 \leq S < U < T$.

- i) $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$.
- ii) $\int_S^T (cf + g) dB_t = c \int_S^T f dB_t + \int_S^T g dB_t$.
- iii) $\mathbb{E} \left[\int_S^T f dB_t \right] = 0$.
- iv) $Var \left(\int_S^T f dB_t \right) = \int_S^T \mathbb{E}[f^2] dt$.
- v) $\int_S^T f dB_t$ es \mathcal{F}_t -medible.
- vi) El proceso $M(t) = M(0) + \int_0^t f dB_t$ es martingala.

Las demostraciones de las propiedades i) a v) se siguen de la definición de la integral, y la isometría de Itô para la número iv). La propiedad vi) se sigue como corolario del siguiente teorema.

Teorema 1.2.1. *Sea $f \in \mathcal{V}[0, T]$, existe un proceso $\{J_t\}_{t \geq 0}$ continuo en $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $\mathbb{P} \left(J_t = \int_0^t f dB_s \right) = 1$ para toda $t \in [0, T]$.*

Demostración. Sea $\phi_n(t, \omega)$ un proceso simple tal que

$$\mathbb{E} \left[\int_0^T (f - \phi_n)^2 dt \right] \rightarrow 0 \quad \text{cuando } n \rightarrow \infty,$$

hagamos

$$I_n(t, \omega) = \int_0^t \phi_n(s, \omega) dB_s \quad \text{y} \quad I_t = I(t, \omega) = \int_0^t f(s, \omega) dB_s.$$

Notemos que $I_n(\cdot, \omega)$ es continua para toda n . Consideremos $s < t$ para probar que $I_n(t, \omega)$ es \mathcal{F}_t -martingala.

$$\begin{aligned}
\mathbb{E}[I_n(t, \omega) | \mathcal{F}_s] &= \mathbb{E} \left[\int_0^t \phi_n(u, \omega) dB_u \middle| \mathcal{F}_s \right] \\
&= \mathbb{E} \left[\left(\int_0^s \phi_n(u, \omega) dB_u + \int_s^t \phi_n(u, \omega) dB_u \right) \middle| \mathcal{F}_s \right] \\
&= \int_0^s \phi_n(u, \omega) dB_u + \mathbb{E} \left[\int_s^t \phi_n(u, \omega) dB_u \middle| \mathcal{F}_s \right],
\end{aligned}$$

sabemos que la integral de Itô para funciones simples se escribe como

$$\int_s^t \phi_n(u, \omega) dB_u = \sum_{s \leq t_j^{(n)} \leq t_{j+1}^{(n)} \leq t} a_j^{(n)} (B_{t_{j+1}} - B_{t_j}),$$

condicionamos sobre la σ -álgebra \mathcal{F}_{t_j} ,

$$\begin{aligned}
&= \int_0^s \phi_n(u, \omega) dB_u + \sum_{s \leq t_j^{(n)} \leq t_{j+1}^{(n)} \leq t} \mathbb{E} \left[\mathbb{E}[a_j^{(n)} (B_{t_{j+1}} - B_{t_j}) | \mathcal{F}_{t_j}] \middle| \mathcal{F}_s \right] \\
&= \int_0^s \phi_n(u, \omega) dB_u + \sum_{s \leq t_j^{(n)} \leq t_{j+1}^{(n)} \leq t} \mathbb{E} \left[a_j^{(n)} \mathbb{E}[(B_{t_{j+1}} - B_{t_j}) | \mathcal{F}_{t_j}] \middle| \mathcal{F}_s \right],
\end{aligned}$$

notemos que $\mathbb{E}[(B_{t_{j+1}} - B_{t_j}) | \mathcal{F}_{t_j}] = 0$, entonces

$$\mathbb{E}[I_n(t, \omega) | \mathcal{F}_s] = I_n(s, \omega).$$

Por lo tanto, es \mathcal{F}_t -martingala continua.

Entonces, $I_n - I_m$ es también \mathcal{F}_t -martingala y usando la desigualdad de Doob para martingalas continuas (véase [21]),

$$\begin{aligned}
\mathbb{P} \left[\sup_{0 \leq t \leq T} |I_n(t, \omega) - I_m(t, \omega)| > \lambda \right] &\leq \frac{1}{\lambda^2} \mathbb{E} [|I_n(T, \omega) - I_m(T, \omega)|^2] \\
&= \frac{1}{\lambda^2} \mathbb{E} \left[\left| \int_0^T \phi_n(s, \omega) dB_s - \int_0^T \phi_m(s, \omega) dB_s \right|^2 \right] \\
&= \frac{1}{\lambda^2} \mathbb{E} \left[\left| \int_0^T (\phi_n(s, \omega) - \phi_m(s, \omega)) dB_s \right|^2 \right] \\
&\leq \frac{1}{\lambda^2} \mathbb{E} \left[\int_0^T (\phi_n(s, \omega) - \phi_m(s, \omega))^2 dB_s \right] \xrightarrow{m, n \rightarrow \infty} 0.
\end{aligned}$$

Elegimos una sub-sucesión n_k que cuando crece a infinito cumple

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| > 2^{-k} \right] < 2^{-k}.$$

Luego, por el lema de Borel-Cantelli (véase [11]) existe un conjunto Ω_0 con una variable aleatoria, $c(\omega) < \infty$ tal que para toda $\omega \in \omega_0$ y $\sup_{0 \leq t \leq T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| > 2^{-k}$ para $k > c(\omega)$, por lo que,

$$\mathbb{P}[\sup_{0 \leq t \leq T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| > 2^{-k} \text{ para } k \text{ infinito}] = 0,$$

entonces, I_{n_k} es uniformemente convergente para $t \in [0, T]$ y además para cada $\omega \in \Omega_0$ J_t es continua en el tiempo casi seguramente y como $I_{n_k}(t, \omega)$ converge a $I(t, \omega)$ en \mathcal{L}^2 ,

$$I_t = J_t \quad \text{c.s. para toda } t.$$

Con ello se completa la prueba. \square

1.2.1. Fórmula de Itô

La fórmula de Itô es una herramienta importante para el cálculo estocástico porque es útil para resolver algunas ecuaciones diferenciales estocásticas.

Consideremos el proceso $\{Y_t\}_{t \geq 0}$ una función del movimiento Browniano estándar, tal que $Y_t = f(t, B_t)$; la fórmula de Itô para este proceso se presenta en el siguiente teorema.

Teorema 1.2.1.1 (Fórmula de Itô). *Para cualquier función $f(t, x) \in C^{1,2}(\mathbb{R}^+, \mathbb{R})$*

$$Y_t = f(t, B_t) = f(0, 0) + \int_0^t \frac{\partial f}{\partial x}(s, B_s) dB_s + \int_0^t \frac{\partial f}{\partial t}(s, B_s) ds + \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_s) ds.$$

Demostración. Escribimos a $f(t, B_t) - f(0, 0)$ como una suma telescópica y aplicamos el polinomio de Taylor

$$\begin{aligned} f(t, B_t) - f(0, 0) &= \sum_{i=1}^n f(t_i, B_{t_i}) - f(t_{i-1}, B_{t_{i-1}}) \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x}(t_{i-1}, B_{t_{i-1}}) \Delta B_i + \sum_{i=1}^n \frac{\partial f}{\partial t}(t_{i-1}, B_{t_{i-1}}) \Delta t_i \\ &\quad + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}}) (\Delta B_i)^2 + \sum_{i=1}^n r(t_{i-1}, t_i, B_{t_{i-1}}, B_{t_i}). \end{aligned}$$

donde $\Delta t_i = t_i - t_{i-1}$ y $\Delta B_i = B_{t_i} - B_{t_{i-1}}$ Luego, las primeras dos sumas convergen en probabilidad como sigue:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{\partial f}{\partial x}(t_{i-1}, B_{t_{i-1}}) \Delta B_i = \int_0^t \frac{\partial f}{\partial x}(s, B_s) dB_s,$$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{\partial f}{\partial t}(t_{i-1}, B_{t_{i-1}}) \Delta t_i = \int_0^t \frac{\partial f}{\partial t}(s, B_s) ds.$$

Ahora, tomamos la tercer suma,

$$\begin{aligned}
& \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})(\Delta B_i)^2 \\
&= \frac{1}{2} \left[\sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})\Delta t_i - \sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})\Delta t_i \right. \\
&\quad \left. + \sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})(\Delta B_i)^2 \right] \\
&= \frac{1}{2} \left[\sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})\Delta t_i \right. \\
&\quad \left. + \sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})\left((\Delta B_i)^2 - \Delta t_i\right) \right].
\end{aligned}$$

Como queremos hacer el último término igual a cero, lo acotamos utilizando la desigualdad de Markov, entonces

$$\begin{aligned}
& \mathbb{E} \left[\frac{1}{4} \left(\sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})\left((\Delta B_i)^2 - \Delta t_i\right) \right)^2 \right] \\
&= \frac{1}{4} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})^2 \left((\Delta B_i)^2 - \Delta t_i \right)^2 \right] \\
&\quad + \sum_{i \neq j} \mathbb{E} \left[\frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}}) \frac{\partial^2 f}{\partial x^2}(t_{j-1}, B_{t_{j-1}}) \left((\Delta B_i)^2 - \Delta t_i \right) \left((\Delta B_j)^2 - \Delta t_j \right) \right],
\end{aligned}$$

condicionamos la segunda suma sobre la σ -álgebra con el índice menor, $((\Delta B_i)^2 - \Delta t_i)((\Delta B_j)^2 - \Delta t_j)$ son independientes para toda $i \neq j$ y el producto de las esperanzas se hace cero; entonces,

$$\frac{1}{4} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})^2 \mathbb{E} \left[\left((\Delta B_i)^2 - \Delta t_i \right)^2 \right] \right] \leq \frac{t^2}{2n} \left\| \frac{\partial^2 f}{\partial x^2} \right\|_{\infty}^2,$$

puesto que $f \in C^2$ tiene segundas derivadas acotadas. Luego, si tomamos el límite, converge a cero en probabilidad, por lo tanto,

$$\frac{1}{2} \sum_{i=1}^n \frac{\partial^2 f}{\partial x^2}(t_{i-1}, B_{t_{i-1}})(\Delta B_i)^2 = \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_s) ds.$$

Finalmente, si probamos que $\sum_{i=1}^n r(t_{i-1}, t_i, B_{t_{i-1}}, B_{t_i})$ es cero en el límite terminamos la prueba. Sabemos por ser el error en el polinomio de Taylor que

$$\begin{aligned}
|r(t_{i-1}, t_i, B_{t_{i-1}}, B_{t_i})| &\leq (\Delta B_i)h(t_{i-1}, t_i, B_{t_{i-1}}, B_{t_i}) \\
&\quad + (\Delta t_i)k(t_{i-1}, t_i, B_{t_{i-1}}, B_{t_i}),
\end{aligned}$$

donde h y k son funciones acotadas y continuas uniformemente.
Entonces

$$\sum_{i=1}^n |r| \leq \sum_{i=1}^n \Delta B_t^2 h + \Delta t k,$$

por la desigualdad de Cauchy

$$\begin{aligned} \mathbb{E}\left(\left|\sum_{i=1}^n r\right|\right) &\leq \sum_{i=1}^n \mathbb{E}[(\Delta B_t)^4]^{\frac{1}{2}} \mathbb{E}[h^2]^{\frac{1}{2}} + \sum_{i=1}^n \mathbb{E}[(\Delta t)^4]^{\frac{1}{2}} \mathbb{E}[k^2]^{\frac{1}{2}} \\ &= \sum_{i=1}^n \left(\frac{3t^2}{n^2}\right)^{\frac{1}{2}} \mathbb{E}[h^2]^{\frac{1}{2}} + n \left(\frac{t}{n}\right)^4 \mathbb{E}[k^2]^{\frac{1}{2}} \\ &= \sqrt{3t} \mathbb{E}[h^2]^{\frac{1}{2}} + \frac{t^4}{n^3} \mathbb{E}[k^2]^{\frac{1}{2}}. \end{aligned}$$

Por otro lado, utilizando que h y k son acotadas, la continuidad uniforme y posteriormente, la desigualdad de Markov tenemos

$$\begin{aligned} \mathbb{E}(h^2) &\leq \epsilon^2 + \|h\|_{\infty} \mathbb{P}(|\Delta B_t| \geq \delta) \\ &\leq \epsilon^2 + \|h\|_{\infty} \mathbb{E}\left[\frac{|\Delta B_t|^2}{\delta^2}\right] \\ &= \epsilon^2 + \|h\|_{\infty} \frac{t}{\delta^2 n}. \end{aligned}$$

Análogamente

$$\mathbb{E}(k^2) \leq \epsilon^2 + \|k\|_{\infty} \frac{(\Delta t)^2}{\delta^2}.$$

Sustituimos y tomamos el límite cuando n tiende a ∞

$$\mathbb{E}\left(\left|\sum_{i=1}^n r\right|\right) \leq \sqrt{3t} \epsilon^2 + \|h\|_{\infty} \frac{t}{\delta^2 n} + \frac{t^4}{n^3} \epsilon^2 + \|k\|_{\infty} \frac{(\Delta t)^2}{\delta^2} \rightarrow 0,$$

nuevamente por la desigualdad de Markov $\sum_{i=1}^n r$ tiende a cero en probabilidad; con lo cual se concluye la prueba. \square

La regla de la cadena

Supongamos ahora al proceso $\{Y_t\}_{t \geq 0}$ en términos del movimiento Browniano con deriva $X_t = \mu t + \sigma B_t$, es decir $Y_t = f(t, X_t)$. Pero también puede escribirse como función del movimiento Browniano estándar, tal que $Y_t = g(t, B_t)$, donde g está definida como $g(t, x) = f(t, \mu t + \sigma x)$.

Denotemos por simplicidad $\frac{\partial f}{\partial x} = f_x$, $\frac{\partial f}{\partial t} = f_t$ y $\frac{\partial^2 f}{\partial x^2} = f_{xx}$ y consideremos $x = B_t$.

Aplicando la fórmula de Itô para $Y_t = g(t, B_t)$,

$$dY_t = g_t(t, B_t)dt + g_x(t, B_t)dB_t + \frac{1}{2}g_{xx}(t, B_t)dt.$$

Utilizando la regla de la cadena para g

$$\begin{aligned} g_t(t, x) &= f_t(t, \mu t + \sigma x) + \mu f_x(t, \mu t + \sigma x), \\ g_x(t, x) &= f_x(t, \mu t + \sigma x)\sigma, \\ g_{xx} &= f_{xx}(t, \mu t + \sigma x)\sigma^2. \end{aligned}$$

Entonces,

$$\begin{aligned} dY_t &= (f_t(t, \mu t + \sigma x) + \mu f_x(t, \mu t + \sigma x))dt + \sigma f_x(t, \mu t + \sigma x)dB_t \\ &\quad + \frac{1}{2}\sigma^2 f_{xx}(t, \mu t + \sigma x)dt \\ &= f_t(t, X_t)dt + \mu f_x(t, X_t)dt + \sigma f_x(t, X_t)dB_t + \frac{1}{2}\sigma^2 f_{xx}(t, X_t)dt \\ &= f_t(t, X_t)dt + f_x(t, X_t) \underbrace{(\mu dt + \sigma dB_t)}_{dX_t} + \frac{1}{2}\sigma^2 f_{xx}(t, X_t)dt \\ &= f_t(t, X_t)dt + f_x(t, X_t)dX_t + \frac{1}{2}\sigma^2 f_{xx}(t, X_t)dt \end{aligned}$$

Por lo tanto, tenemos la fórmula de Itô para el proceso $\{Y_t\}_{t \geq 0}$ en función del movimiento Browniano con deriva. De manera formal, se enuncia en el siguiente teorema.

Teorema 1.2.1.2. (Fórmula de Itô). Sea $\{X_t\}_{t \geq 0}$ un proceso dado por

$$dX_t = b(t, \omega)dt + \sigma(t, \omega)dB_t$$

y $f(t, x) \in C^{1,2}(\mathbb{R}^+, \mathbb{R})$ tal que el proceso dado por $Y_t = f(t, X_t)$ cumple:

$$Y_t = f(0, 0) + \int_0^t \frac{\partial f}{\partial t}(s, X_s)ds + \int_0^t \frac{\partial f}{\partial x}(s, X_s)dX_s + \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, X_s)\sigma^2(s, \omega)ds. \quad (1.2)$$

La fórmula general también puede escribirse de forma diferencial utilizando la siguiente tabla

\cdot	d_t	dB_t
d_t	0	0
dB_t	0	d_s

y notando que $dX_t = a(s, \omega)dt + b(s, \omega)dB_s$, entonces,

$$\begin{aligned} dY_t &= f_t(t, X_t)dt + f_x(t, X_t)dX_t + \frac{1}{2}f_{xx}(t, X_t)dX_t \cdot dX_t \\ &= f_t(t, X_t)dt + f_x(t, X_t)dX_t + \frac{1}{2}f_{xx}(t, X_t)\sigma^2(s, \omega)ds. \end{aligned}$$

porque

$$\begin{aligned} dX_t \cdot dX_t &= (b(s, \omega)dt + \sigma(s, \omega)dB_s)(b(s, \omega)dt + \sigma(s, \omega)dB_s) \\ &= b^2(s, \omega)dt \cdot dt + 2b(s, \omega)b(s, \omega)dt \cdot dB_s + \sigma^2(s, \omega)dB_s \cdot dB_s \\ &= \sigma^2(s, \omega)ds. \end{aligned}$$

La demostración de la fórmula de Itô es análoga a la del Teorema 1.2.1.

1.2.2. Transformación de Lamperti.

Supongamos que tenemos una ecuación diferencial estocástica de tal forma que el coeficiente de difusión depende solamente del estado del proceso al tiempo t , es decir,

$$dX_t = b(t, X_t)dt + \sigma(X_t)dB_t,$$

la transformación de Lamperti permite convertir tal ecuación en una nueva ecuación diferencial estocástica con coeficiente de difusión unitario, haciendo

$$Y_t = F(X_t) = \int_z^{X_t} \frac{1}{\sigma(u)} du,$$

donde z es un valor arbitrario en el espacio de estados del proceso $\{X_t\}_{t \geq 0}$. Así, Y_t es solución a

$$dY_t = b_Y(t, Y_t)dt + dB_t. \quad (1.3)$$

Esta transformación es muy útil para realizar simulaciones y es empleada en diversos métodos de estimación.

Ahora, aplicamos la fórmula de Itô para encontrar el coeficiente de deriva b_Y en términos de los coeficientes de deriva y difusión del proceso original. Tenemos que

$$f(t, x) = \int_z^x \frac{1}{\sigma(u)} du, \quad f_t(t, x) = 0, \quad f_x(t, x) = \frac{1}{\sigma(x)}, \quad f_{xx}(t, x) = -\frac{\sigma_x(x)}{\sigma^2(x)},$$

luego,

$$\begin{aligned} dY_t &= 0dt + \frac{1}{\sigma(x)}dX_t - \frac{1}{2}\sigma^2(X_t)\frac{\sigma_x(X_t)}{\sigma^2(X_t)}dt \\ &= \frac{1}{\sigma(X_t)}(b(t, X_t)dt + \sigma(X_t)dB_t) - \frac{1}{2}\sigma_x(X_t)dt \\ &= \left(\frac{b(t, X_t)}{\sigma(X_t)} - \frac{1}{2}\sigma_x(X_t) \right) dt + dB_t. \end{aligned}$$

Por lo tanto, igualando los coeficientes,

$$b_Y(t, y) = \frac{b(t, F^{-1}(y))}{\sigma(F^{-1}(y))} - \frac{1}{2}\sigma_x(F^{-1}(y)).$$

1.3. Difusiones

Consideremos una ecuación estocástica de la forma

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t,$$

a la función $b(t, X_t)$ se le conoce como coeficiente de deriva y a $\sigma(t, X_t)$ coeficiente de difusión.

A la solución de la ecuación se le llama *difusión*. El siguiente teorema da las condiciones necesarias para que la solución exista y sea única.

Teorema 1.3.1. *Sea $T > 0$, $b(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\sigma(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{(n,m)}$ son funciones medibles que satisfacen*

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|); \quad x \in \mathbb{R}^n, t \in [0, T],$$

para alguna C constante, y

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|; \quad x, y \in \mathbb{R}^n, t \in [0, T],$$

para D constante. Sea X_0 una v.a. independiente a la σ -álgebra generada por el proceso $\{B_s\}_{s \geq 0}$, tal que $\mathbb{E}[|X_0|^2] < \infty$. Entonces la ecuación diferencial estocástica

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T, \quad (1.4)$$

tiene una única solución X_t continua respecto a t , $\mathbb{E}[\int_0^t |X_t|^2 dt] < \infty$ y además es adaptada a filtración $\mathcal{F}_t^{X_0}$ generada por X_0 y B_s , $s < t$.

Antes de hacer la demostración de existencia y unicidad de las difusiones, veamos otro resultado que necesitaremos para dicha prueba.

Lema 1.3.1 (Desigualdad de Gronwall). *Sea $y(t)$ función no negativa que satisface*

$$y(t) \leq A + B \int_0^t y(s)ds < \infty,$$

para toda $0 < s \leq t \leq T$, A y B son constantes positivas, entonces, $y(t) \leq Ae^{Bt}$, para $t \in [0, T]$.

Demostración.

$$\begin{aligned} y(t) &\leq A + B \int_0^t y(s)ds \\ &\leq A + B \int_0^t A + B \int_0^s y(s_1)ds_1 = A + ABt + B^2 \int_0^t \int_0^s y(s_1)ds_1 ds \\ &\leq A + ABt + \frac{At^2}{2}B^2 + B^3 \int_0^t \int_0^s \int_0^{s_1} y(s_2)ds_2 ds_1 ds \end{aligned}$$

entonces tenemos los primeros k terminos de Ae^{Bt}

$$\sum_{n=1}^k A \frac{(Bt)^n}{n!} + BI_k$$

donde

$$I_k = \int \cdots \int_0^{k-1} y(s_k) ds_k \cdots ds_1 ds$$

por hipótesis $y(t) \leq \sum_{n=1}^k A \frac{(Bt)^n}{n!} + BI_k \leq D$ luego $I_k \leq B^k \frac{t^{k+1}}{k+1!}$ con $k \rightarrow \infty$ converge uniformemente a Ae^{Bt} , por lo tanto, $y(t) \leq Ae^{Bt}$. \square

Seguimos con la demostración del Teorema 1.3.1.

Demostración. Sea $\tilde{X}_t(\omega)$ y $X_t(\omega)$ soluciones a la ecuación (1.4) con valores iniciales X_0 y \tilde{X}_0 respectivamente. Hagamos

$$a(s, \omega) = b(s, X_s) - b(s, \tilde{X}_s),$$

$$\gamma(s, \omega) = \sigma(s, X_s) - \sigma(s, \tilde{X}_s).$$

Comenzamos probando que la solución es única acotando $\mathbb{P}(|X_t - \tilde{X}_t| \geq \epsilon)$ con ayuda de la desigualdad de Markov.

$$\begin{aligned} \mathbb{E}[|X_t - \tilde{X}_t|^2] &= \mathbb{E}\left[|X_0 - \tilde{X}_0 + \int_0^t a ds + \int_0^t \gamma dB_s|^2\right] \\ &\leq 3\mathbb{E}[|X_0 - \tilde{X}_0|^2] + 3\mathbb{E}\left[\left|\int_0^t a ds\right|^2\right] + 3\mathbb{E}\left[\left|\int_0^t \gamma dB_s\right|^2\right]. \end{aligned}$$

Utilizando la isometría de Itô y la desigualdad de Cauchy-Schwarz

$$\begin{aligned} &\leq 3\mathbb{E}[|X_0 - \tilde{X}_0|^2] + 3t\mathbb{E}\left[\int_0^t a^2 ds\right] + 3\mathbb{E}\left[\left|\int_0^t \gamma^2 ds\right|\right] \\ &\leq 3\mathbb{E}[|X_0 - \tilde{X}_0|^2] + 3t\mathbb{E}\left[\int_0^t D^2 |X_s - \tilde{X}_s|^2 ds\right] \\ &\quad + 3\mathbb{E}\left[\int_0^t D^2 |X_s - \tilde{X}_s|^2 ds\right] \\ &= 3\mathbb{E}[|X_0 - \tilde{X}_0|^2] + 3(t+1)D^2\mathbb{E}\left[\int_0^t |X_s - \tilde{X}_s|^2 ds\right]. \end{aligned}$$

Sea $0 \leq t \leq T$, consideremos $\nu(t) = \mathbb{E}[|X_t - \tilde{X}_t|^2]$ tal que satisface

$$\nu(t) \leq F + A \int_0^t \nu(s) ds,$$

con $F = \mathbb{E}[|X_0 - \tilde{X}_0|^2]$ y $A = 3(1 + T)D^2$; por la desigualdad de Gronwall $\nu(t) \leq Fe^{At}$. Asumimos que $X_0 - \tilde{X}_0 = 0$ entonces $\nu(t) = 0$ para cualquier t en el intervalo $[0, T]$. Luego, por la desigualdad de Markov

$$\mathbb{P}(|X_t - \tilde{X}_t| \geq \epsilon) \leq \frac{\mathbb{E}[|X_t - \tilde{X}_t|^2]}{\epsilon^2}, \quad \text{para toda } \epsilon > 0.$$

Por lo tanto, $\mathbb{P}(|X_t - \tilde{X}_t| \geq \epsilon) = 0$, lo cual implica que $\mathbb{P}(|X_t - \tilde{X}_t| < \epsilon) = 1$ para toda ϵ . Si en particular tomamos $t \in \mathbb{Q}$, $\mathbb{P}(|X_t - \tilde{X}_t| = 0, \forall t \in \mathbb{Q} \cap [0, T]) = 1$. Como es continuo en t , $\mathbb{P}(|X_t - \tilde{X}_t| = 0, \forall t \in [0, T]) = 1$ entonces $X_t = \tilde{X}_t$ en probabilidad; por lo tanto, la solución es única.

Probar la existencia se hace de la misma manera que en las ecuaciones diferenciales.

Definimos $Y_t^{(0)} = X_0$ y $Y_t^k = X_t^{(k)}(\omega)$ y la siguiente sucesión de procesos

$$Y_t^{(k+1)} = X_0 + \int_0^t b(s, Y_s^{(k)})ds + \int_0^t \sigma(s, Y_s^{(k)})dB_s,$$

siguiendo el mismo procedimiento que para probar unicidad, obtenemos

$$\mathbb{E}[|Y_t^{(k+1)} - Y_t^{(k)}|^2] \leq (1 - T)D^2 \int_0^t \mathbb{E}[|Y_t^{(k)} - Y_t^{(k-1)}|^2] ds,$$

para $t \leq T$, $k \geq 1$ con

$$\begin{aligned} \mathbb{E}[|Y_t^{(1)} - Y_t^{(0)}|^2] &= \mathbb{E}\left[\left|X_0 - \int_0^t b(s, Y_s^{(0)})ds + \int_0^t \sigma(s, Y_s^{(0)})dB_s - X_0\right|^2\right] \\ &= \mathbb{E}\left[\left|\int_0^t b(s, Y_s^{(0)})ds + \int_0^t \sigma(s, Y_s^{(0)})dB_s\right|^2\right] \\ &\leq 2t\mathbb{E}\left[\left|\int_0^t b^2(s, Y_s^{(0)})ds\right|\right] + 2\mathbb{E}\left[\left|\int_0^t \sigma^2(s, Y_s^{(0)})dB_s\right|\right] \\ &\leq 2t\mathbb{E}\left[\int_0^t C^2(1 + X_0)^2\right] + 2\mathbb{E}\left[\int_0^t C^2(1 + X_0)^2\right] \\ &\leq 2t^2C^2(1 + \mathbb{E}[|X_0|^2]) + 2tC^2(1 + \mathbb{E}[|X_0|^2]) \\ &\leq A_1 t. \end{aligned}$$

Donde A_1 es una constante que depende de T , C y $\mathbb{E}[|X_0|^2]$ entonces

$$\mathbb{E}[|Y_t^{(k+1)} - Y_t^{(k)}|^2] \leq \frac{(tA_2)^{k+1}}{(k+1)!},$$

para $k \geq 0$, $t \in [0, T]$ y A_2 constante que depende de C , D , T y $\mathbb{E}[|X_0|^2]$. Por otro lado, sea $\Delta b^{(k)} = b(s, Y_s^{(k)}) - b(s, Y_s^{(k-1)})$ y

$$\Delta\sigma^{(k)} = \sigma(s, Y_s^{(k)}) - \sigma(s, Y_s^{(k-1)})$$

$$\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| \leq \int_0^T |\Delta b^{(k)}| ds + \sup_{0 \leq t \leq T} \int_0^T |\Delta\sigma^{(k)}|_{B_s}.$$

Luego,

$$\begin{aligned} \mathbb{P}\left[\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| > 2^{-k}\right] &\leq \mathbb{P}\left[\left(\int_0^T |\Delta b^{(k)}| ds\right)^2 > 2^{-2k-2}\right] \\ &\quad + \mathbb{P}\left[\sup_{0 \leq t \leq T} \int_0^T |\Delta\sigma^{(k)}|_{B_s} > 2^{-k-1}\right], \end{aligned}$$

utilizando la desigualdad maximal de Doob, la desigualdad de Cauchy y de Markov, finalmente la isometría de Itô; si además consideramos que $A \geq D^2(T+1)$,

$$\begin{aligned} &\leq 2^{2k+2}T \int_0^t \mathbb{E}[|\Delta b^{(k)}|] ds + 2^{k+2} \int_0^t \mathbb{E}[|\Delta\sigma^{(k)}|^2] ds \\ &\leq 2^{2k+2}D^2(T+1) \int_0^t \frac{A_2^k t^k}{k!} dt \\ &\leq \frac{(4A_2T)^{k+1}}{(k+1)!}. \end{aligned}$$

Por el lema de Borell-Cantelli

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| > 2^{-k} \text{ para una cantidad de } k \text{ infinita}\right] = 0,$$

entonces existe para casi toda ω un $k_0(\omega)$ tal que $\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| \leq 2^{-k}$ con $k \geq k_0$ y $Y_t^{(n)} = Y_y^{(0)} + \sum_{k=0}^{n-1} Y_t^{(k+1)} - Y_t^{(k)}$ es uniformemente convergente. El límite X_t de $Y_t^{(n)}$ es continuo y medible para casi toda ω si $Y_t^{(n)}$ es continua para toda n .

Ahora, para $m > n > 0$

$$\begin{aligned} \mathbb{E}\left[|Y_t^{(m)} - Y_t^{(n)}|^2\right]^{\frac{1}{2}} &= \|Y_t^{(m)} - Y_t^{(n)}\|_{\mathcal{L}_{\mathbb{P}}^2} \\ &= \left\| \sum_{k=n}^{m-1} Y_t^{(k+1)} - Y_t^{(k)} \right\|_{\mathcal{L}_{\mathbb{P}}^2} \\ &\leq \sum_{k=n}^{m-1} \|Y_t^{(k+1)} - Y_t^{(k)}\|_{\mathcal{L}_{\mathbb{P}}^2} \\ &\leq \sum_{k=n}^{\infty} \left[\frac{(A_2t)^{k+1}}{(k+1)!} \right]^{\frac{1}{2}} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Por lo tanto, $Y_t^{(m)}$ converge en \mathcal{L}^2 a Y_t y como ya probamos unicidad, $Y_t = X_t$ casi seguramente.

Solo falta probar que X_t satisface la ecuación (1.4).

Para toda n , $Y_t^{(n)} = X_0 + \int_0^t b(s, Y_s^{(n)})ds + \int_0^t \sigma(s, Y_s^{(n)})dB_s$; X_t es el límite de $Y_t^{(n)}$ cuando n tiende a infinito y además converge uniformemente en $t \in [0, T]$ para casi toda ω .

Por el lema de Fatou (véase [9])

$$\mathbb{E} \left[\int_0^t |X_t - Y_t^{(n)}|^2 dt \right] \leq \limsup_{m \rightarrow \infty} \mathbb{E} \left[\int_0^t |Y_t^{(m)} - Y_t^{(n)}|^2 dt \right] = 0$$

Además $\int_0^t \sigma(s, Y_s^{(n)})dB_s$ converge a $\int_0^t \sigma(s, X_t)dB_s$ y por la desigualdad de Hölder $\int_0^t b(s, Y_s^{(n)})ds$ converge a $\int_0^t b(s, X_s)ds$ en \mathcal{L}^2 .

Tomando el límite cuando n tiende a infinito, recuperamos la ecuación (1.4). \square

A la solución X_t dada a partir del teorema anterior se le llama *solución fuerte* donde X_t es \mathcal{F}_t -adaptado (es la filtración generada por B_t y X_0). Por otra parte, si consideramos el espacio $(\Omega, \mathcal{H}, \mathbb{P})$ y tomamos $(\tilde{X}_t, \tilde{B}_t, \tilde{\mathcal{H}}_t)$ donde \tilde{X}_t cumple la ecuación diferencial estocástica y $\tilde{\mathcal{H}}$ es una filtración, \tilde{X}_t es adaptado y \tilde{B}_t es martingala respecto a $\tilde{\mathcal{H}}_t$ se le llama a \tilde{X}_t *solución débil*.

Una solución fuerte es también débil, pero en general no se cumple lo inverso.

Lema 1.3.2. *Si b y σ cumplen las condiciones del Teorema 1.3.1 entonces la solución (débil o fuerte) es débilmente única*

Es decir, tienen la misma distribución finito-dimensional.

Definición 1.3.1. *Una difusión de Itô es un proceso estocástico $X_t(\omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R}^n$ que satisface la siguiente ecuación*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t,$$

$\{B_t\}_{t \geq 0}$ representa un movimiento Browniano m -dimensional y las funciones $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$ y $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ cumplen las condiciones del Teorema 1.3.1.

Consideremos al proceso $\{X_t\}_{t \geq 0}$ homogéneo respecto al tiempo, es decir,

$$\begin{aligned} X_{s+h}^{s,x} &= x + \int_s^{s+h} b(X_u^{s,x})du + \int_s^{s+h} \sigma(X_u^{s,x})dB_u \\ &= x + \int_0^h b(X_{s+v}^{s,x})dv + \int_0^h \sigma(X_{s+v}^{s,x})d\tilde{B}_v, \end{aligned}$$

donde $u = s + v$ y $\tilde{B}_v = B_{s+v} - B_s$ $v \geq 0$. Por otro lado sabemos que

$$X_h^{0,x} = x + \int_0^h b(X_v^{0,x})dv + \int_0^h \sigma(X_v^{0,x})dB_v,$$

y como \tilde{B}_v y B_v tienen la misma distribución P^0 (ley de probabilidad del movimiento Browniano que inicia en cero) entonces $X_{s+h}^{s,x}$ tiene la misma distribución que $X_h^{0,x}$, por lo tanto, es homogénea en el tiempo. Introducimos Q^x la ley de probabilidad de $\{X_t\}_{t \geq 0}$ para $x \in \mathbb{R}^n$, intuitivamente, nos da la distribución con $X_0 = x$ consideremos μ la σ -álgebra generada por $\omega \rightarrow X_t^x(\omega)$ definimos

$$Q^x(X_{t_1} \in E_1, \dots, X_{t_k} \in E_k) = P^0(X_{t_1}^x \in E_1, \dots, X_{t_k}^x \in E_k)$$

con E_i conjuntos de Borel $1 \leq i \leq k$. Sea $\mathcal{F}_t^{(m)}$ la σ -álgebra generada por $\{B_r\}_{r \leq t}$ y μ_t la σ -álgebra generada por $\{X_r\}_{r \leq t}$. Notemos que si X_t es \mathcal{F}_t -medible, entonces $\mu_t \subset \mathcal{F}_t^{(m)}$.

En la siguiente sección se muestran las propiedades fundamentales de las difusiones.

1.3.1. Propiedades de las difusiones

Comenzaremos demostrando que cumplen la propiedad de Markov, esta característica tiene que ver con la pérdida de memoria de los procesos estocásticos.

Teorema 1.3.1.1 (Propiedad de Markov). *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ borel acotada, entonces para $t, h \geq 0$*

$$\mathbb{E}^x[f(X_{t+h})|\mathcal{F}_t^{(m)}] = \mathbb{E}^{X_t(\omega)}[f(X_h)], \quad (1.5)$$

donde \mathbb{E}^x es la esperanza respecto a Q^x y \mathbb{E} respecto a P^0 , tal que $\mathbb{E}^y[f(X_h)] = \mathbb{E}[f(X_h^y)]$.

Demostración. Para $r \geq t$

$$X_r(\omega) = X_t(\omega) + \int_t^r b(X_u)du + \int_t^r \sigma(X_u)sB_u,$$

por unicidad, $X_r(\omega) = X_r^{t, X_t}(\omega)$. Definimos $F(x, t, r, \omega) = X_r^{t,x}(\omega)$ entonces $X_r(\omega) = F(X_t, t, r, \omega)$. Notemos que para cada ω , F es independiente a $\mathcal{F}_t^{(m)}$. Reescribimos (1.5)

$$\mathbb{E}^x[f(F(X_t, t, t+h, \omega))|\mathcal{F}_t^{(m)}] = \mathbb{E}[f(F(x, 0, h, \omega))]_{x=X_t}.$$

Sea $g(x, \omega) = f \circ F(x, t, t+h, \omega)$ ($x, \omega \rightarrow g(x, \omega)$ es medible.

Podemos aproximar a g de manera puntual y acotada por funciones de la forma $\sum_{k=1}^m \phi_k(x)\psi_k(\omega)$, entonces

$$\begin{aligned}
\mathbb{E}\left[g(x, \omega) | \mathcal{F}_t^{(m)}\right] &= \mathbb{E}\left[\lim_{n \rightarrow \infty} \sum_{k=1}^n \phi_k(x) \psi_k(\omega) | \mathcal{F}_t^{(m)}\right] \\
&= \lim_{n \rightarrow \infty} \sum_{k=1}^n \phi_k(X_t) \mathbb{E}\left[\psi_k(\omega) | \mathcal{F}_t^{(m)}\right] \\
&= \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}\left[\phi_k(y) \psi_k(\omega) | \mathcal{F}_t^{(m)}\right]_{y=X_t} \\
&= \mathbb{E}\left[g(y, \omega) | \mathcal{F}_t^{(m)}\right]_{y=X_t} \\
&= \mathbb{E}[g(y, \omega)]_{y=X_t},
\end{aligned}$$

por ser homogénea en el tiempo

$$\begin{aligned}
\mathbb{E}\left[f(F(X_t, t, t+h, \omega)) | \mathcal{F}_t^{(m)}\right] &= \mathbb{E}\left[f(F(y, t, t+h, \omega)) | \mathcal{F}_t^{(m)}\right]_{y=X_t} \\
&= \mathbb{E}\left[f(F(y, 0, h, \omega))\right]_{y=X_t},
\end{aligned}$$

con lo cual se prueba la propiedad de Markov. □

Ahora, de forma general probaremos la propiedad fuerte de Markov, definimos primero un tiempo de paro.

Definición 1.3.1.1. Sea $\{\mathcal{N}_t\}_{t \geq 0}$ una familia creciente de σ -álgebras. Se le llama tiempo de paro a la función $\tau : \Omega \rightarrow [0, \infty]$ con respecto a $\{\mathcal{N}_t\}_{t \geq 0}$

$$\{\omega \mid \tau(\omega) \leq t\} \in \mathcal{N}_t \quad \text{para toda } t \text{ positiva.}$$

Sea τ un tiempo de paro respecto a $\{\mathcal{N}_t\}_{t \geq 0}$, consideremos \mathcal{N}_∞ la σ -álgebra más pequeña que contiene a \mathcal{N}_t para toda t , la σ -álgebra \mathcal{N}_τ consiste en todos los conjuntos $N \in \mathcal{N}_\infty$ tales que $N \cap \{\tau \leq t\} \in \mathcal{N}_t$ para toda t positiva.

Sea μ_τ es la σ -álgebra generada por $\{X_{s \wedge \tau}\}_{s \geq 0}$ y $\mathcal{F}_\tau^{(m)}$ es la σ -álgebra generada por $\{B_{s \wedge \tau}\}_{s \geq 0}$.

Teorema 1.3.1.2 (Propiedad fuerte de Markov). Sea f una función borel acotada en \mathbb{R}^n , τ un tiempo de paro respecto a $\mathcal{F}_t^{(m)}$, $\tau \leq \infty$ casi seguramente, entonces

$$\mathbb{E}^x[f(X_{\tau+h}) | \mathcal{F}_\tau^{(m)}] = \mathbb{E}^{X_\tau(\omega)}[f(X_h)]. \quad (1.6)$$

Demostración. Para casi toda ω tenemos

$$X_{\tau+h}^{\tau, x} = x + \int_\tau^{\tau+h} b(X_u^{\tau, x}) du + \int_\tau^{\tau+h} \sigma(X_u^{\tau, x}) dB_u.$$

Por la propiedad fuerte de Markov del movimiento Browniano, $\tilde{B}_v = B_{\tau+v} - B_\tau$, $v \geq 0$ es nuevamente Browniano e independiente a $\mathcal{F}_\tau^{(m)}$, entonces

$$X_{\tau+h}^{\tau,x} = x + \int_0^h b(X_{\tau+v}^{\tau,x})dv + \int_0^h \sigma(X_{\tau+v}^{\tau,x})d\tilde{B}_v.$$

Por unicidad fuerte $X_{\tau+h}^{\tau,x}$ coincide casi en todo punto con la solución

$$Y_h = x + \int_0^h b(Y_v)dv + \int_0^h \sigma(Y_v)d\tilde{B}_v,$$

y como Y_h es independiente a \mathcal{F}_τ , $X_{\tau+h}^{\tau,x}$ también lo es y por unicidad débil tienen la misma distribución que $X_h^{0,x}$.

Hagamos $F(x, t, r, \omega) = X_r^{t,x}(\omega)$, $r \geq t$; escribimos (1.6)

$$\mathbb{E}[f(F(x, 0, \tau + h, \omega)) | \mathcal{F}_\tau^{(m)}] = \mathbb{E}[f(F(x, 0, h, \omega))]_{x=X_\tau}.$$

$$\begin{aligned} F(x, 0, \tau + h, \omega) &= X_{\tau+h}(\omega) = x + \int_0^{\tau+h} b(X_s)ds + \int_0^{\tau+h} \sigma(X_s)dB_s \\ &= x + \int_0^\tau b(X_s)ds + \int_0^\tau \sigma(X_s)dB_s \\ &\quad + \int_\tau^{\tau+h} b(X_s)ds + \int_\tau^{\tau+h} \sigma(X_s)dB_s \\ &= X_\tau + \int_\tau^{\tau+h} b(X_s)ds + \int_\tau^{\tau+h} \sigma(X_s)dB_s \\ &= F(X_\tau, \tau, \tau + h, \omega). \end{aligned}$$

Entonces

$$\mathbb{E}[f(F(X_\tau, \tau, \tau + h, \omega)) | \mathcal{F}_\tau^{(m)}] = \mathbb{E}[f(F(x, 0, h, \omega))]_{x=X_\tau}$$

de manera análoga al teorema anterior definimos $g(x, t, r, \omega) = f(F(x, t, r, \omega))$ y suponemos que $g(x, t, r, \omega) = \sum_k \phi_k(x)\psi_k(t, r, \omega)$.

Entonces,

$$\begin{aligned} \mathbb{E}[g(X_\tau, \tau, \tau + h, \omega) | \mathcal{F}_\tau^{(m)}] &= \mathbb{E}\left[\sum_k \phi_k(X_\tau)\psi_k(\tau, \tau + h, \omega) | \mathcal{F}_\tau^{(m)}\right] \\ &= \sum_k \phi_k(X_\tau)\mathbb{E}[\psi_k(\tau, \tau + h, \omega) | \mathcal{F}_\tau^{(m)}] \\ &= \sum_k \mathbb{E}[\phi_k(x)\psi_k(\tau, \tau + h, \omega) | \mathcal{F}_\tau^{(m)}]_{x=X_\tau} \\ &= \mathbb{E}[g(x, \tau, \tau + h, \omega) | \mathcal{F}_\tau^{(m)}]_{x=X_\tau} \\ &= \mathbb{E}[g(x, \tau, \tau + h, \omega)]_{x=X_\tau} \\ &= \mathbb{E}[f(X_{\tau+h}^{\tau,x})]_{x=X_\tau} \\ &= \mathbb{E}[f(X_h^{0,x})]_{x=X_\tau} \\ &= \mathbb{E}[f(F(x, 0, h, \omega))]_{x=X_\tau}. \end{aligned}$$

□

Definición 1.3.1.2. Sea $\{X_t\}_{t \geq 0}$ una difusión homogénea en el tiempo en \mathbb{R}^n . El generador infinitesimal A de X_t se define como:

$$Af(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}, \quad x \in \mathbb{R}^n$$

Se denota $D_A(x)$ al conjunto de funciones tales que el límite existe en x y D_A el conjunto de funciones para las cuáles el límite existe para toda x .

Es posible encontrar relación entre el generador infinitesimal y los coeficiente de deriva y difusión de las ecuaciones diferenciales estocásticas; el siguiente lema es un herramienta para probarlo.

Lema 1.3.1.1. Sea $\{Y_t\}_{t \geq 0} = Y_t^x$ un proceso de Itô de las siguiente forma

$$Y_t^x = x + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s,$$

$\{B_t\}_{t \geq 0}$ un movimiento Browniano m -dimensional, $f \in C^2(\mathbb{R}^n)$ con soporte compacto, τ un tiempo de paro con respecto a $\{\mathcal{F}_t^{(m)}\}$, $\mathbb{E}^x[\tau] < \infty$, $u(s, \omega)$ y $v(s, \omega)$ acotadas tales que $Y(t, \omega)$ pertenece al soporte de f , entonces

$$\mathbb{E}^x[f(Y_\tau)] = f(x) + \mathbb{E}^x \left[\int_0^\tau \left(\sum_i u_i \frac{\partial f}{\partial x_i}(y_s) + \frac{1}{2} \sum_{i,j} (vv^T)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s) \right) ds \right].$$

Demostración. Sea $Z = f(Y)$, aplicamos fórmula de Itô

$$\begin{aligned} dZ &= \sum_i \frac{\partial f}{\partial y_i}(Y) dY_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial y_i \partial y_j}(Y) dY_i dY_j \\ &= \sum_i \frac{\partial f}{\partial y_i}(Y) (u_i dt + v_i dB_t) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial y_i \partial y_j}(Y) (u_i dt + v_i dB_t) (u_j dt + v_j dB_t) \\ &= \sum_i \frac{\partial f}{\partial y_i}(Y) (u_i dt) + \sum_i \frac{\partial f}{\partial y_i}(Y) (v_i dB_t) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial y_i \partial y_j}(Y) (v_i dB_t) (v_j dB_t). \end{aligned}$$

Aparte calculamos

$$\begin{aligned} (v_i dB_t) (v_j dB_t) &= \left(\sum_k v_{i_k} dB_{t_k} \right) \left(\sum_l v_{j_l} dB_{t_l} \right) \\ &= \sum_k v_{i_k} v_{j_k} (dB_{t_k})^2 \\ &= \left(\sum_k v_{i_k} v_{j_k} \right) dt \\ &= (vv^T)_{ij} dt. \end{aligned}$$

Sustituyendo éste último resultado

$$\begin{aligned} dZ_t &= \sum_i \frac{\partial f}{\partial y_i}(Y) u_i dt + \sum_i \frac{\partial f}{\partial y_i}(Y) v_i dB_t + \frac{1}{2} \sum_{i,j} \frac{\partial f}{\partial y_i y_j}(Y) (vv^T)_{ij} dt \\ &= \left(\sum_i \frac{\partial f}{\partial y_i}(Y) u_i + \frac{1}{2} \sum_{i,j} \frac{\partial f}{\partial y_i y_j}(Y) (vv^T)_{ij} \right) dt + \sum_i \frac{\partial f}{\partial y_i}(Y) v_i dB_t, \end{aligned}$$

entonces

$$f(Y_\tau) = f(Y_0) + \int_0^\tau \left(\sum_i u_i \frac{\partial f}{\partial y_i} + \frac{1}{2} \sum_{i,j} (vv^T)_{ij} \frac{\partial f}{\partial y_i y_j} \right) ds + \sum_{i,k} \int_0^\tau v_{ik} \frac{\partial f}{\partial y_i} dB_{t_k}.$$

Aplicando esperanza

$$\begin{aligned} \mathbb{E}^x[f(Y_\tau)] &= f(x) + \mathbb{E}^x \left[\int_0^\tau \left(\sum_i u_i \frac{\partial f}{\partial y_i} + \frac{1}{2} \sum_{i,j} (vv^T)_{ij} \frac{\partial f}{\partial y_i y_j} \right) ds \right] \\ &\quad + \sum_{i,k} \mathbb{E}^x \left[\int_0^\tau v_{ik} \frac{\partial f}{\partial y_i} dB_{t_k} \right] \end{aligned}$$

Por otro lado, si g es una función borel acotada, ($|g| \leq M$), entonces para todo entero k

$$\mathbb{E} \left[\int_0^{\tau \wedge k} g(Y_s) dB_s \right] = \mathbb{E} \left[\int_0^k g(Y_s) \mathbb{I}_{\{s \wedge \tau\}} dB_s \right] = 0,$$

luego,

$$\begin{aligned} \mathbb{E}^x \left[\left(\int_0^\tau g(Y_s) dB_s - \int_0^{\tau \wedge k} g(Y_s) dB_s \right)^2 \right] &= \mathbb{E}^x \left[\int_{\tau \wedge k}^\tau g^2(Y_s) dB_s \right] \\ &\leq M^2 \mathbb{E}^x[\tau - \tau \wedge k], \end{aligned}$$

Si hacemos que k tienda a infinito, la última expresión queda acotada por cero. Utilizando las dos afirmaciones anteriores

$$0 = \lim_{k \rightarrow \infty} \mathbb{E}^x \left[\int_0^{\tau \wedge k} g(Y_s) dB_s \right] = \lim_{k \rightarrow \infty} \mathbb{E}^x \left[\int_0^\tau (Y_s) dB_s \right] = \mathbb{E}^x \left[\int_0^\tau (Y_s) dB_s \right]$$

Por lo tanto,

$$\mathbb{E}^x[f(Y_\tau)] = f(x) + \mathbb{E}^x \left[\int_0^\tau \left(\sum_i u_i \frac{\partial f}{\partial x_i}(y_s) + \frac{1}{2} \sum_{i,j} (vv^T)_{ij} \frac{\partial^2 f}{\partial x_i x_j}(Y_s) \right) ds \right].$$

□

Teorema 1.3.1.3 (Generador Infinitesimal). *Sea $\{X_t\}_{t \geq 0}$ una difusión de Itô tal que $dX_t = b(X_t)dt + \sigma(X_t)dB_t$. Si $f \in C_0^2$ entonces $f \in D_A$ y*

$$Af(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{ij}(x) \frac{\partial^2 f}{\partial x_i x_j}.$$

La demostración se sigue del lema 1.3.1.1 y la definición del generador infinitesimal tomando $t = \tau$.

Los ejemplos más comunes de difusiones son el movimiento Browniano, el movimiento Browniano geométrico, el proceso de Ornstein-Uhlenbeck, el proceso de Bessel y el modelo de Cox-Ingersoll-Ross.

Consideremos la ecuación diferencial estocástica $dX_t = \beta X_t dt + \sigma X_t dB_t$, cuya solución es el movimiento Browniano geométrico. Notemos primero que el coeficiente de deriva y de difusión cumplen las propiedades del teorema 1.3.1 que garantiza la existencia y unicidad de la solución.

Encontramos la solución a partir de la fórmula de Itô igualando los coeficientes, $dX_t = f_t(t, B_t)dt + f_x(t, B_t)dX_t + \frac{1}{2}f_{xx}(t, B_t)dt$, si tomamos a $X_t = f(t, B_t)$ entonces,

$$\beta f(t, x) = f_t(t, x) + \frac{1}{2}f_{xx}(t, x) \quad (1.7)$$

$$\sigma f(t, x) = f_x(t, B_t) \quad (1.8)$$

De la ecuación (1.7) tenemos que $f(t, x) = e^{\sigma x + g(t)}$ donde $g(t)$ es una función arbitraria y por la ecuación (1.8) sabemos que $g'(t) = \beta - \frac{1}{2}\sigma^2$, por lo tanto, la ecuación diferencial estocástica tiene solución $X_t = X_0 e^{(\beta - \frac{1}{2}\sigma^2)t + \sigma B_t}$.

Ahora, consideremos la ecuación diferencial estocástica $dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dB_t$, con $\theta_1, \theta_2 \in \mathbb{R}$ y $\theta_3 \in \mathbb{R}^+$, que tiene como solución al proceso de Ornstein-Uhlenbeck. Primero, veamos que cumple las propiedades del teorema 1.3.1 que garantizan la existencia y unicidad de la solución. $|\theta_1 - \theta_2 x| + |\theta_3| \leq C(1 + |x|)$ si consideramos $C \geq \max\{|\theta_1| + |\theta_3|, |\theta_2|\}$, donde C puede encontrarse porque θ_1, θ_2 , y θ_3 son constantes. Y $|(\theta_1 - \theta_2 x) - (\theta_1 - \theta_2 y)| + |\theta_3 - \theta_3| = |-\theta_2 x + \theta_2 y| = |\theta_2||x - y| \leq D|x - y|$ con D constante tal que $D \geq |\theta_2|$.

Para encontrar la solución elegimos $f(t, x) = x e^{\theta_2 t}$, entonces $f_x = e^{\theta_2 t}$, $f_t = \theta_2 x e^{\theta_2 t}$ y $f_{xx} = 0$, luego

$$\begin{aligned} X_t e^{\theta_2 t} &= f(t, X_t) = f(0, X_0) + \int_0^t \theta_2 X_u e^{\theta_2 u} du + \int_0^t e^{\theta_2 u} dX_u \\ &= x_0 + \int_0^t \theta_2 X_u e^{\theta_2 u} du + \int_0^t e^{\theta_2 u} ((\theta_1 - \theta_2 X_u) du + \theta_3 dB_u) \\ &= x_0 + \frac{\theta_1}{\theta_2} (e^{\theta_2 t} - 1) + \theta_3 \int_0^t e^{\theta_2 u} dB_u. \end{aligned}$$

Entonces, la solución está dada por $X_t = \frac{\theta_1}{\theta_2} + e^{-\theta_2 t} \left(x_0 - \frac{\theta_1}{\theta_2} \right) + \theta_3 \int_0^t e^{-\theta_2(t-u)} dB_u$.

1.3.2. Simulación de difusiones

Debido a que en general no es posible encontrar explícitamente la solución a una ecuación diferencial estocástica se han desarrollado métodos numéricos para simular dicho proceso, cuyo objetivo radica en estudiar la trayectoria misma, o bien, alguna función del proceso que no se conoce analíticamente, por ejemplo, los momentos.

La mayoría de estos métodos están basados en aproximaciones discretas de la solución continua y pueden clasificarse según sus propiedades, en la literatura se usa principalmente un criterio de optimización considerando el orden de convergencia.

Definición 1.3.2.1. Sea X^δ una aproximación a tiempo discreto del proceso continuo $\{X_t\}_{t \geq 0}$, donde δ representa el incremento de tiempo máximo en la discretización. Se dice que es de nivel fuerte de convergencia τ para todo $t > 0$ si

$$\mathbb{E}[X_t^\delta - X_t] \leq C\delta^\tau$$

para todo $\delta < \delta_0$, con $\delta_0 > 0$ y C constante que no depende de δ .

Definición 1.3.2.1. Sea nuevamente X^δ una aproximación a tiempo discreto del proceso continuo $\{X_t\}_{t \geq 0}$. Se dice que es de convergencia débil con nivel τ si

$$|\mathbb{E}[g(X_t^\delta)] - \mathbb{E}[g(X_t)]| \leq C\delta^\tau$$

para todo $t > 0$, $\delta < \delta_0$, con $\delta_0 > 0$ y C constante que no depende de δ y g cualquier función polinómica, creciente, continua y diferenciable de orden $2(\tau + 1)$.

A continuación presentamos los métodos más comunes de simulación de difusiones.

Método de Euler. El método de aproximación de Euler originalmente se utilizaba para generar soluciones a las ecuaciones diferenciales deterministas y posteriormente se extendió al caso estocástico.

Consideremos un proceso de difusión que es solución a la ecuación diferencial estocástica dada por la ecuación (1.4), con $X_0 = x_0$ y discretización del intervalo $[0, T]$, es decir, $0 = t_0 < t_1 < \dots < t_n = T$. La aproximación de Euler de $X = \{X_t\}_{t \geq 0}$ es un proceso estocástico continuo Y que satisface:

$$Y_{t_{i+1}} = Y_{t_i} + b(Y_{t_i}, t_i)(t_{i+1} - t_i) + \sigma(Y_{t_i}, t_i)(B_{t_{i+1}} - B_{t_i})$$

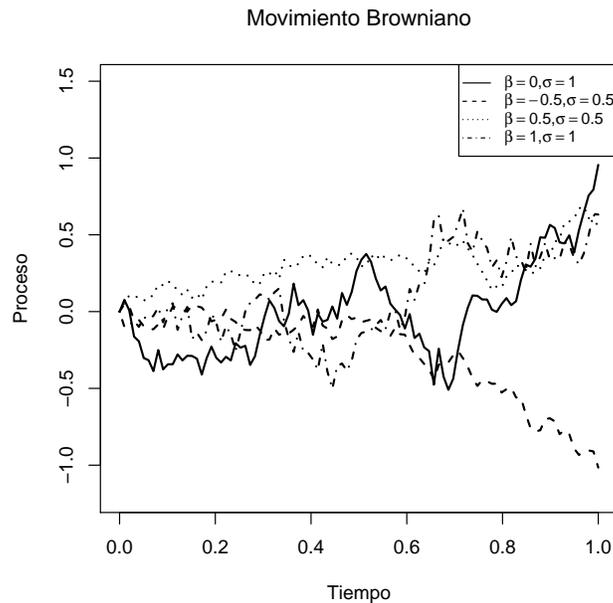
para $i = 0, 1, \dots, n-1$, con $Y_0 = x_0$. Notemos que $B_{t_{i+1}} - B_{t_i}$ tiene distribución normal con media cero y varianza $\Delta_i = t_{i+1} - t_i$. Usualmente, Δ_i se considera constante. El método de Euler tiene orden de convergencia fuerte $\tau = 1/2$.

Método de Milstein. El método de Milstein busca incrementar el orden de convergencia fuerte del método de Euler $\tau = 1/2$ a 1, agregando una corrección al incremento estocástico mediante un término de segundo orden utilizando el lema de Itô y la expansión de Taylor. Entonces, la aproximación es de la siguiente forma:

$$Y_{t_{i+1}} = Y_{t_i} + b(t_i, Y_{t_i})(t_{i+1} - t_i) + \sigma(t_i, Y_{t_i})(B_{t_{i+1}} - B_{t_i}) + \frac{1}{2}\sigma(t_i, Y_{t_i})\sigma_x(t_i, Y_{t_i})((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)).$$

Como ya mencionamos, el método de Milstein tiene orden de convergencia fuerte igual a uno y débil igual a uno. Existe otra versión del método de Milstein que incrementa a segundo orden la convergencia débil pero requiere la primer y segunda derivadas parciales del coeficiente de deriva y difusión.

Existen otros métodos que consisten en aproximar el coeficiente de deriva con una función lineal, la primer aproximación fue propuesta por Ozaki (véase por ejemplo [15]) suponiendo que σ es constante y que el coeficiente de deriva depende solo del proceso; posteriormente, hacen la extensión del método permitiendo que coeficiente de deriva sea función del tiempo y del proceso y al coeficiente de difusión variar. Otro algoritmo interesante fue propuesto en [2], basado en muestreo exacto tomando el algoritmo de aceptación / rechazo para simular trayectorias.



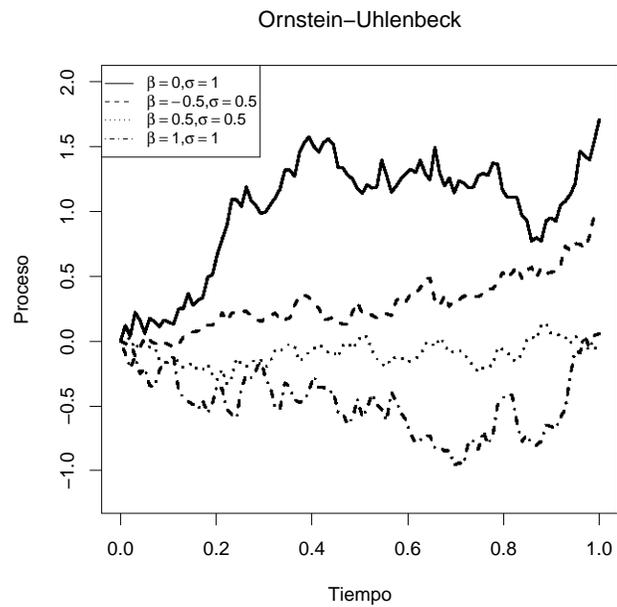
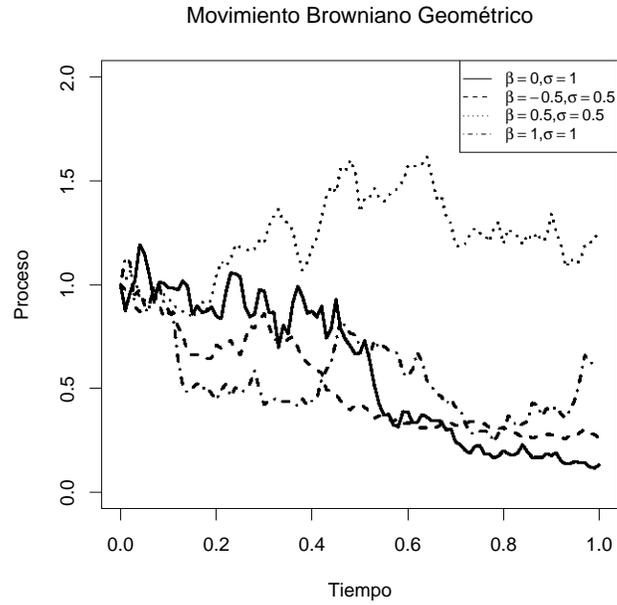


Figura 1.1: Simulación del movimiento Browniano, del movimiento Browniano y del proceso de Ornstein-Uhlenbeck con diferentes parámetros a partir de los métodos antes descritos.

1.3.3. Estimación

En esta sección damos una breve explicación de algunos métodos para la estimación de los parámetros de una difusión. Para profundizar en los detalles véase [10]. Es importante notar que en la práctica las difusiones únicamente se observan de manera discreta, es decir, se tienen observaciones en los tiempos t_0, t_1, \dots, t_n . En general, la inferencia se hace a través de la función de verosimilitud pero en la mayoría de los casos no está disponible de forma explícita.

Ya que las difusiones cumplen la propiedad de Markov, la función de verosimilitud puede escribirse como el producto de las densidades de transición, sin embargo, solo en casos especiales dichas expresiones son conocidas. Como consecuencia, los métodos de estimación están basados en aproximaciones a las trayectorias de la difusión o a la función de verosimilitud misma.

Consideremos la ecuación diferencial estocástica

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dB_t, \quad (1.9)$$

donde θ representa el vector de parámetros, de aquí en adelante, supondremos que $\theta \in \Theta \subset \mathbb{R}^p$ las funciones b y σ cumplen las propiedades del Teorema 1.3.1. Denotemos a p_θ la distribución del proceso $\{X_t\}_{t \geq 0}$ con distribución inicial π_θ y $p_\theta(t, \cdot | x)$ la distribución condicional del proceso dado $X_0 = x$; además, como $\{X_t\}_{t \geq 0}$ es homogéneo en el tiempo, $p_\theta(t, \cdot | x)$ también denota la probabilidad condicional de X_{t+s} dado $X_s = x$ para toda t positiva. Supongamos que el proceso se observa en el tiempo $t_i, i = 1, \dots, n$; denotamos a $\mathcal{F}_n = \sigma\{X_i, i \leq n\}$ la σ -álgebra generada por las primeras n observaciones, con \mathcal{F}_0 la σ -álgebra trivial, $X_i = X_{t_i}$ y $\Delta_i = t_i - t_{i-1}$.

Entonces, la función de verosimilitud queda como sigue:

$$L_n(\theta) = \prod_{i=1}^n p_\theta(\Delta_i, X_i | X_{i-1}) p_\theta(X_0)$$

y la log-verosimilitud

$$\log(L_n(\theta)) = \sum_{i=1}^n \log(p_\theta(\Delta_i, X_i | X_{i-1})) + \log(p_\theta(X_0)).$$

Los métodos que vamos a describir utilizan ciertos supuestos, en primer lugar, asumimos que $p_\theta(X_0) = 1$, luego consideremos las siguientes hipótesis:

1. Crecimiento polinomial. Sea K constante e independiente a θ , $m > 0$ tal que

$$|b(x, \theta)| + |\sigma(x, \theta)| \leq K(1 + |x|^m).$$

Se le llama crecimiento lineal al caso particular cuando $m = 1$.

2. Condición de Lipschitz. Sea K constante e independiente a θ tal que

$$|b(x, \theta) - b(y, \theta)| + |\sigma(x, \theta) - \sigma(y, \theta)| \leq K(|x - y|).$$

3. Coeficiente de difusión positivo.

$$\inf_x \sigma^2(x, \theta) > 0.$$

4. Momentos acotados. Para toda k positiva, todos los momentos de orden k del proceso de difusión existen,

$$\sup_t \mathbb{E}[|X_t|^k] < \infty.$$

5. Los coeficientes b y σ y sus derivadas respecto a θ (al menos de orden 3) deben existir, ser continuas y tener crecimiento polinomial sobre θ .

Notemos que las hipótesis de crecimiento lineal y la condición de Lipschitz son las características de los coeficientes de deriva y difusión que garantizan la existencia y unicidad de la solución a la ecuación diferencial estocástica.

Encontrar los estimadores máximo verosímil con la función de verosimilitud real es posible solo en casos especiales en los cuales la función de densidad condicional es conocida, por ejemplo, el proceso de Ornstein-Uhlenbeck o el proceso de Cox-Ingersoll-Ross y la función de verosimilitud quede expresada de tal forma que pueda maximizarse con técnicas conocidas.

Otra forma de obtener los estimadores deseados se basa en métodos de aproximación, los cuales en lugar de aproximar la función de densidad de transición lo hacen a través de trayectorias del proceso.

Método de Euler. Consideremos una ecuación diferencial estocástica como la ecuación (1.9), si los coeficientes b y σ son constantes en intervalos de tiempo suficientemente pequeños, es decir, $[t, t + \Delta t)$, la aproximación de Euler produce

$$X_{t+\Delta t} - X_t = b(X_t, \theta)\Delta t + \sigma(X_t, \theta)(B_{t+\Delta} - B_t),$$

donde los incrementos $X_{t+\Delta} - X_t$ son variables aleatorias normales e independientes con media $b(X_t, \theta)\Delta t$ y varianza $\sigma(X_t, \theta)\Delta t$, bajo las hipótesis 1,3 y 4, la función de densidad de transición es

$$p_\theta(t, y|x) = \frac{1}{\sqrt{2\pi t\sigma^2(x, \theta)}} \exp \left\{ -\frac{(y - x - b(x, \theta)t)^2}{2t\sigma^2(x, \theta)} \right\}.$$

Y la log-verosimilitud

$$\log(L_n(\theta)) = -\frac{1}{2} \left\{ \frac{\sum_{i=1}^n (X_i - X_{i-1} - b(X_{i-1}, \theta)\Delta t)^2}{2\sigma^2(X_{i-1}, \theta)\Delta t} - n \log(2\pi\sigma^2(x, \theta)\Delta t) \right\}.$$

Cuando σ^2 es constante, la expresión de la log-verosimilitud que debe maximizarse se reduce a una función más sencilla. Si además suponemos que $n\Delta_n^3 \rightarrow 0$

el estimador máximo verosímil es consistente y asintóticamente normal.

Método de Elerian. En este método se propone utilizar la función de densidad de transición obtenida a partir del método de Milstein, es decir,

$$X_{t+dt} = X_t + b(t, X_t)dt + \sigma(t, X_t)(B_{t+dt} - B_t) + \frac{1}{2}\sigma(t, X_t)\sigma_x(t, X_t)((B_{t+dt} - B_t)^2 - dt).$$

donde $\sigma_x(t, x)$ representa la derivada parcial de $\sigma(t, x)$ con respecto a x . Notemos que si $\sigma_x \approx 0$ la expresión se reduce al método de Euler. La verosimilitud se encuentra a partir del siguiente teorema y la demostración del mismo puede encontrarse en [8].

Teorema. 1. *Consideremos la ecuación diferencial estocástica definida en la ecuación (1.9) que satisface las condiciones del Teorema 1.3.1 y $X_0 = x_0$. Definamos*

$$Cf(t, x) = \sigma(t, x) \frac{\partial}{\partial x} f(t, x),$$

y

$$b^*(t, x) = b(t, x) - \frac{1}{2}\sigma(t, x)\sigma_x(t, x).$$

Supongamos que satisfacen las siguientes condiciones para t, s en $[0, T]$, para cada x en los reales y K constante positiva e independiente a la discretización

$$|b^*(t, x)| + |Cb(t, x)| \leq K(1 + |x|), \quad |\sigma(t, x)| + |C\sigma(t, x)| \leq K(1 + |x|),$$

$$|C^2\sigma(t, x)| \leq K(1 + |x|),$$

$$|g(t, x) - g(s, x)| \leq K(1 + |x|)|s - t|^{\frac{1}{2}}.$$

Donde $g(t, x)$ puede ser cualquiera de las siguientes funciones b^* , σ , $C\sigma$. Entonces, la densidad de transición del método de Milstein puede escribirse como:

$$p_\theta(t, y|x) = \frac{z^{-\frac{1}{2}} \cosh(\sqrt{cz})}{|A|\sqrt{2\pi}} e^{-\frac{c+z}{2}},$$

donde

$$A = \frac{\sigma(x)\sigma_x(x)t}{2}, \quad B = \frac{\sigma(x)}{2\sigma_x(x)} + x + b(x)t - A, \\ z = \frac{y-B}{A}, \quad c = \frac{1}{\sigma_x^2(x)t}.$$

Por lo tanto, la aproximación es válida para $\sigma_x \neq 0$ y $z > 0$.

Usualmente, en la implementación se utiliza con $\cosh(x) = \frac{e^x + e^{-x}}{2}$.

Como habíamos mencionado, los métodos anteriores tratan de aproximar la función de densidad de transición a partir de las trayectorias mientras que los siguientes buscan aproximar directamente la función de verosimilitud.

Método de Kessler. Se propone usar la expansión de Itô-Taylor para aproximar la media y la varianza de la función de densidad condicional. La expansión es la siguiente:

$$\mathbb{E}_\theta[\phi(X_i)|\mathcal{F}_{i-1}] = \phi(X_{i-1}) + \Delta_n A_\theta \phi(X_{i-1}) + \frac{1}{2} \Delta_n^2 A_\theta^2 \phi(X_{i-1}) + \dots,$$

donde A_θ representa el generador infinitesimal de la función $\phi(x)$. Además de las hipótesis ya mencionadas, para esta aproximación necesitamos $n \rightarrow \infty$, $\Delta_n \rightarrow 0$, $n\Delta_n \rightarrow \infty$ y $n\Delta_n^2 \rightarrow 0$, y ergodicidad. Por lo tanto, las aproximaciones son:

$$\begin{aligned} \mathbb{E}_\theta &= x + b(t, x)dt + \left(b(t, x)b_x(t, x) + \frac{1}{2}\sigma^2(t, x)b_{xx}(t, x) \right) \frac{(dt)^2}{2}, \\ \text{Var}_\theta &= x^2 + (2b(t, x)x + \sigma^2(t, x))dt + \left[2b(t, x)(b_x(t, x) + b(t, x) \right. \\ &\quad \left. + \sigma(t, x)\sigma_x(t, x)) + \sigma^2(t, x)(b_{xx}(t, x)x + 2b_x(t, x) + \sigma_x^2(t, x) \right. \\ &\quad \left. + \sigma(t, x)\sigma_{xx}(t, x)) \right] \frac{(dt)^2}{2} - \mathbb{E}_\theta^2 \end{aligned}$$

Notemos que para algunas trayectorias, la varianza puede tomar valores negativos para algunos valores de θ y x . Una característica importante de este método, es la posibilidad de obtener intervalos de confianza igualmente aproximados.

Métodos a partir de simulaciones. Sea $p_\theta(\Delta, y|x)$ la función de densidad verdadera de $X_{t+\Delta}$ en el punto y dado $X_t = x$. Si Δ no es suficientemente pequeño, la aproximación de Euler no produce buenas estimaciones de dicha densidad. La idea, es tomar $\delta \ll \Delta$, consideremos por ejemplo, $\delta = \Delta/N$ para N grande y posteriormente utilizar la ecuación de Chapman-Kolmogorov, luego

$$p_\theta(\Delta, y|x) = \int p_\theta(\delta, y|z)p_\theta(\Delta - \delta, z|x)dz = \mathbb{E}_\theta[p_\theta(\delta, y|z)|\Delta - \delta],$$

lo que significa que $p_\theta(\Delta, y|x)$ puede verse como la esperanza del proceso del tiempo $t + (\Delta - \delta)$ al tiempo $t + \Delta$ considerando que el proceso al tiempo t está en x . Como elegimos a δ pequeña empleamos la aproximación de Euler, por lo cual $p_\theta(\Delta - \delta, z|x)$ tiene distribución normal.

Entonces, el método consiste en encontrar los estimadores Monte Carlo. Definamos $\phi_\theta(\delta, y|x)$ la función de densidad de transición encontrada a partir del método de Euler, ahora, se deben simular M trayectorias del proceso del tiempo δ a $\Delta - \delta$ iniciando en $X_0 = x$. Sea z_m el último valor obtenido en la simulación con $m = 1, \dots, M$, el estimador Monte Carlo es el siguiente:

$$\hat{p}_\theta^{(N, M)}(\Delta, X_{t+\Delta}|x) = \frac{1}{M} \sum_{i=1}^M \phi_\theta(\delta, X_{t+\Delta}|z_i),$$

es posible mostrar que $\hat{p}_\theta^{(N, M)}(\Delta, X_{t+\Delta}|x)$ tiende a $p_\theta(\Delta, X_{t+\Delta}|x)$ cuando N y M crecen a infinito. Por consiguiente la aproximación depende fuertemente del

tamaño de N . Notemos que para cada par de puntos se necesitan M trayectorias de tamaño N , entonces el tiempo que lleva simularlas es una desventaja.

Polinomios hermitianos. Supongamos que queremos aproximar la función de densidad de un proceso $\{X_t\}_{t \geq 0}$, en lugar de expandir dicho proceso, se propone hacer una transformación de tal manera que obtengamos algo muy parecido a una normal y entonces hacer la expansión mediante polinomios hermitianos.

Los polinomios hermitianos se definen de dos formas según el área en que se esté trabajando (física o probabilidad). La definición usual que se utiliza en probabilidad es

$$H_j^{proba}(x) = (-1)^j e^{x^2/2} \frac{d^j}{dx^j} e^{-x^2/2},$$

observemos que puede quedar expresada en términos de una función de densidad normal estándar. Sin embargo, la definición que vamos a utilizar es la siguiente:

$$H_j(x) = e^{x^2/2} \frac{d^j}{dx^j} e^{-x^2/2}$$

ya que ayuda a simplificar la notación.

Además de la hipótesis 5 se necesita suponer alguno de los siguientes casos: Sea D_X el dominio del proceso $\{X_t\}_{t \geq 0}$, $D_X = (-\infty, \infty)$ o $D_X = (0, \infty)$.

- Si $D_X = (-\infty, \infty)$, entonces existe $c > 0$ tal que $c < \sigma(x, \theta)$ para toda x en el dominio y θ .
- Si $D_X = (0, \infty)$, entonces para cada $\xi > 0$ existe $c_\xi > 0$ tal que $c_\xi \leq \sigma(x, \theta)$ para $x \in [\xi, \infty]$ y θ . Adicionalmente, si $\lim_{x \rightarrow 0^+} \sigma(x, \theta) = 0$, entonces existen $\xi_0 > 0$, $\omega > 0$ y $\rho > 0$ tales que $\sigma(x, \theta) \geq \omega x^\rho$ para toda $0 < x \leq \xi_0$.

En primer lugar, hacemos $Y_t = F(X_t)$, denotamos al proceso $\{Y_t\}_{t \geq 0} = Y$ donde Y_t debe satisfacer

$$dY_t = \mu_Y(Y_t, \theta) dt + dB_t$$

que es la transformación de Lamperti, por lo tanto, aplicando la fórmula de Itô obtenemos

$$\mu_Y(y, \theta) = \frac{b(F^{-1}(x), \theta)}{\sigma(F^{-1}(x), \theta)} - \frac{1}{2} \sigma_x^2(F^{-1}(x), \theta).$$

Si $D_X = (\underline{x}, \bar{x})$, entonces el dominio del nuevo proceso $D_Y = (\underline{y}, \bar{y})$, donde $\underline{y} = \lim_{x \rightarrow \underline{x}} F(x)$ y $\bar{y} = \lim_{x \rightarrow \bar{x}} F(x)$.

Para controlar el comportamiento de la difusión transformada y de la original incluimos una nueva hipótesis formulada en términos del coeficiente μ_Y , no obstante, tiene repercusiones en los coeficientes b y σ de la difusión original.

Para toda θ , μ_Y y sus derivadas con respecto a y y a θ siguen un crecimiento polinomial y $\lim_{y \rightarrow \underline{y}^+ \text{ o } \bar{y}^-} \lambda_Y(y, \theta) < \infty$ con $\lambda_Y(y, \theta) = -(\mu_y^2(y, \theta) + \partial \mu_Y(y, \theta) / \partial y) / 2$.

-
1. Límite izquierdo. Si $y = 0$, entonces, existe $\epsilon_0, \kappa, \alpha$ tal que para toda $0 < y \leq \epsilon_0$ y θ , $\mu_Y(y, \theta) \geq \kappa y^{-\alpha}$, donde $\alpha > 1$ y $\kappa > 0$ o $\alpha = 1$ y $\kappa \geq 1$. Si $y = -\infty$, entonces existe $E_0 > 0, K > 0$ tal que para toda $y \leq -E_0$ y θ , $\mu_Y(y, \theta) \geq Ky$.
 2. Límite derecho. Si $\bar{y} = \infty$, entonces existe $E_0 > 0, K > 0$ tal que para toda $y \geq E_0$ y θ , $\mu_Y(y, \theta) \leq Ky$. Si $\bar{y} = 0$, entonces, existe $\epsilon_0, \kappa, \alpha$ tal que para toda $-\epsilon_0 \leq y < 0$ y θ , $\mu_Y(y, \theta) \leq -\kappa|y|^{-\alpha}$, donde $\alpha > 1$ y $\kappa > 0$ o $\alpha = 1$ y $\kappa \geq 1/2$.

A partir de los últimos supuestos hechos, se prueba que la ecuación diferencial estocástica de la transformación tiene única solución débil para cualquier distribución inicial y la función de densidad de transición es continua y diferenciable, véase [1].

Ahora, introducimos el incremento pseudo normalizado de Y para hacer a la función de densidad más adecuada para la expansión de los polinomios hermitianos. Sea

$$Z = \Delta^{1/2}(Y - y_0)$$

con $\Delta > 0$ dada, θ y $y \in \mathbb{R}$.

Denotamos a la función de densidad de transición de $Y_{t+\Delta}|Y_t = y_0$ como $p_Y(\Delta, y|y_0, \theta)$ y definimos la función de densidad de Z como:

$$p_Z(\Delta, z|y_0, \theta) = \Delta^{1/2} p_Y(\Delta, \Delta^{1/2}z + y_0|y_0, \theta).$$

Entonces,

$$p_Y(\Delta, y|y_0, \theta) = \Delta^{-1/2} p_Z(\Delta, \Delta^{-1/2}(y - y_0)|y_0, \theta)$$

y

$$p_X(\Delta, x|x_0, \theta) = \sigma(x, \theta)^{-1} p_Y(\Delta, F(x)|F(x_0), \theta).$$

El siguiente paso es aproximar al proceso Z con polinomios hermitianos de $p_Z(\Delta, z|y_0, \theta)$ para Δ, y_0 y θ fijos,

$$p_Z^J(\Delta, z|y_0, \theta) = \phi(z) \sum_{j=0}^J \eta_Z^{(j)}(\Delta, z|y_0, \theta) H_j(z),$$

dado que $H_j(z)/\sqrt{j!}$ es ortonormal en $\mathcal{L}^2(\phi)$, el coeficiente $\eta_Z^{(j)}$ se define

$$\eta_Z^{(j)}(\Delta|y_0, \theta) = \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z|y_0, \theta) dz.$$

También podemos definir

$$p_Y^{(J)}(\Delta, y|y_0, \theta) = \Delta^{-1/2} p_Z^{(J)}(\Delta, \Delta^{-1/2}(y - y_0)|y_0, \theta)$$

y

$$p_X^{(J)}(\Delta, x|x_0, \theta) = \sigma(x, \theta)^{-1} p_Y^{(J)}(\Delta, F(x)|F(x_0), \theta).$$

Igualmente se demuestra que $p_Y^{(J)}(\Delta, y|y_0, \theta)$ converge a $p_X(\Delta, x|x_0, \theta)$ cuando J tiende a infinito.

El coeficiente $\eta_Z^{(j)}$ puede ser expresado en términos de los momentos de la función de densidad condicional del proceso Y , véase [1].

Estimación Bayesiana. El estimador bayesiano de θ se encuentra con el valor esperado de la distribución a posteriori

$$p(\theta|x^{(obs)}) = \frac{L_n(\theta)p(\theta)}{\int L_n(\theta)p(\theta)d\theta},$$

$p(\theta)$ representa la distribución a priori y $x^{(obs)}$ las observaciones de la difusión. Recientemente, se han utilizado algoritmos de simulación basados en Cadenas de Markov Monte Carlo, especialmente el muestreo de Gibbs y el Metropolis-Hastings.

Funciones Estimadoras. Las funciones estimadoras son funciones F_n que tienen como argumentos tanto al vector de parámetros θ como los datos observados X^{obs} . Los estimadores se obtienen como solución a la ecuación $F_n(X^{obs}, \theta) = 0$. Existen muchos tipos de funciones estimadoras ya que se desarrollan según las características propias de cada modelo. Dichas funciones deben ser asintóticamente insesgadas y ser capaces de distinguir entre diferentes valores de θ , es decir, $\mathbb{E}[F_n(X^{obs}, \theta)] = 0$ si y solo si $\theta = \theta_0$ con θ_0 los parámetros reales. A continuación damos ejemplos de las funciones estimadoras más comunes. Las funciones estimadoras simples están relacionadas con la distribución marginal o conjunta del proceso de alguna manera y son de dos tipos principalmente,

$$F_n(\theta) = \sum_{i=1}^n f(X_i, \theta) \quad \text{o} \quad F_n(\theta) = \sum_{i=1}^n f(X_{i-1}, X_i, \theta).$$

La función f puede ser por ejemplo, $f = \partial_\theta \log \pi_\theta$, $f_j(x, \theta) = x^j - \mathbb{E}[X_0^j]$, $j = 1, \dots, p$.

Un caso particular del segundo tipo de función es la función score, si la densidad de transición es conocida, la solución es el estimador máximo verosímil.

Funciones estimadoras Martingala. Como habíamos mencionado, existe una gran diversidad de funciones estimadoras, un tipo muy importante son aquellas funciones estimadoras que además son martingalas respecto a la σ -álgebra, denotadas por $G_n(\theta)$. Se dice que G_n es insesgada si $\mathbb{E}[G(\theta)] = 0$. La forma más general de este tipo de funciones es la siguiente:

$$G_n = \sum_{i=1}^n g(\Delta, X_{i-1}, X_i; \theta)$$

con

$$g(\Delta, x, y; \theta) = \sum_{j=1}^N a_j(\Delta, x; \theta)h_j(\Delta, x, y; \theta),$$

donde $h_j(\Delta, x, y; \theta)$ es una función real que satisface $\int_l^r h_j(\Delta, x, y; \theta) p_\theta(\Delta, y|x) dy = 0$ para toda $\Delta > 0$, $x \in (l, r)$ y θ . Las $a_j(\Delta, x; \theta)$ son funciones arbitrarias llamadas ponderadores. Algunos autores como Godambe and Heyde (1987) han descrito métodos para encontrar las funciones estimadoras óptimas.

Funciones estimadoras martingala polinomiales. Otra forma de funciones estimadoras, son las polinomiales; se caracterizan por depender de los primeros momentos de la función de densidad condicional en lugar de toda la función de densidad de transición. Si dichos momentos no se conocen, pueden aplicarse métodos Monte Carlo para la estimación. Los ejemplos más simples son la lineales y las cuadráticas. Las lineales se utilizan cuando el parámetro se encuentra solamente en el coeficiente de deriva. Y si el parámetro se encuentra en el coeficiente de difusión, el modelo lineal no es suficiente, por lo que se utiliza una función cuadrática.

Funciones estimadoras basadas en funciones propias. Esta forma de construir funciones estimadoras está basado en el generador infinitesimal A_θ y fue introducido en [12]. Una función propia de A_θ es una función dos veces diferenciable que satisface $A_\theta \phi(x; \theta) = -\lambda(\theta) \phi(x; \theta)$, donde $\lambda(\theta)$ es un número real positivo llamado valor propio. Ya que $\mathbb{E}[\phi(X_\Delta; \theta) | X_0 = x] = e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta)$, podemos definir una función estimadora martingala como:

$$h_j(\Delta, x, y; \theta) = \phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta),$$

donde $\phi_j(\cdot, \theta)$ son funciones propias y $\lambda_j(\theta)$ valores propios; $j = 1, \dots, N$.

Método de momentos generalizados. Consiste en igualar los momentos teóricos con los muestrales. Sea $u_i = u(X_i; \theta)$ una función tal que

$$\mathbb{E}[u_i] = \mu, \quad \text{para toda } i$$

y

$$S_j = Cov(u_i, u_{i+j}) = \mathbb{E}[(u_i - \mu)(u_{i+j} - \mu)^T], \quad \text{para toda } i.$$

Asumimos que la esperanza de $u(X_i; \theta)$ es igual a cero solo si $\theta = \theta_0$. Usualmente, $u(\theta)$ es la diferencia entre el k -ésimo momento y X_i^k para algunas k . Sea

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^n u(X_i; \theta)$$

la esperanza muestral, por la condición anterior, se espera que $\mathbb{E}[g_n(\theta_0)] = 0$. Y por la ley fuerte de los grandes números,

$$g_n(\theta) \rightarrow \mathbb{E}[u(X_i; \theta)] \quad \text{c.s.}$$

uniformemente en θ . Entonces, la estimación está dada por el siguiente problema de optimización:

$$\hat{\theta} = \underset{\theta}{\text{mín}} g_n(\theta)^T W g_n(\theta).$$

La matriz de pesos o ponderadores W es positiva definida y está dada por S^{-1} , con $S = \mathbb{E}[uu^T]$ y $\hat{S}_j = \frac{1}{n} \sum_{i=j+1}^n u_i u_{i-j}^T$, $j = 0, 1, \dots, l$, donde l es el máximo valor que podemos elegir a priori. Luego,

$$\hat{S} = \hat{S}_0 + \sum_{j=1}^l w_j (\hat{S}_j + \hat{S}_j^T).$$

Los ponderadores se pueden elegir de muchas formas, comúnmente, se elige $w_j = 1 - \frac{j}{l+1}$, llamados ponderadores de Bartlett.

1.3.4. Puentes de difusiones

Al tratar de hacer inferencia sobre los parámetros que determinan una ecuación diferencial estocástica nos encontramos con que en la práctica se tienen observaciones a tiempo discreto mientras que las difusiones se observan a tiempo continuo, esto puede resolverse planteando el problema como de datos faltantes, donde se simulan trayectorias continuas entre cada par de puntos dados.

Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico dado por la ecuación diferencial estocástica (1.4) donde los coeficientes de difusión y de deriva cumplen las condiciones del Teorema 1.3.1 los cuales aseguran la existencia y unicidad del proceso.

Al proceso $\{X_t\}_{t \geq 0}$ en un intervalo de tiempo $[0, T]$ condicionado a que $X_0 = a$ y $X_T = b$ se le llama puente de difusión, son muy pocos los procesos con los cuales es posible encontrar un algoritmo exacto para la construcción del puente, por ejemplo, el puente-Browniano, o puente del proceso de Ornstein-Uhlenbeck.

Sin embargo, existen métodos para simular puentes basados en algoritmos MCMC, en aproximaciones a la función de verosimilitud o a las funciones de densidad de transición. Para efectos de este trabajo vamos a utilizar el método presentado en [4] que consiste en simular dos procesos de difusión independientes: X^1 con valor inicial a en el tiempo $[0, T]$ y análogamente se simula un proceso X^2 con valor inicial b en el mismo periodo de tiempo pero reversible, es decir, el proceso corre de T a 0 . La idea es realizar una aproximación al puente como una combinación de ambos procesos. Se prueba que el proceso recién construido condicionado al evento de que los dos procesos se intersecten, es una aproximación a la realización del puente de difusión en cuestión. Una ventaja importante de esta construcción a diferencia de la demás, es que funciona bien para intervalos de tiempo grandes.

Denotamos a la función de densidad de transición del proceso $\{X_t\}_{t \geq 0}$ como $p_t(x, y)$, específicamente, la densidad condicional de X_{s+t} dado $X_s = x$ es $y \rightarrow p_t(x, y)$. Definimos a los procesos antes mencionados como solución a

$$dX_t^i = b(X_t^i)dt + \sigma(X_t^i)dB_t^i, \quad i = 1, 2, \quad X_t^1 = a \quad \text{y} \quad X_t^2 = b,$$

con B_t^i movimientos Brownianos estándar independientes

Los procesos $\{X_t^1\}_{t \geq 0}$ y $\{X_t^2\}_{t \geq 0}$ se simulan a partir de los métodos conocidos descritos en la sección anterior. Sean $Y^i = Y_{\delta j}^i$, $j = 0, 1, \dots, N$ simulaciones independientes de los procesos $\{X_t^i\}_{t \geq 0}$ para $i = 1, 2$ en $[0, T]$ con $\delta = T/N$. Entonces, una aproximación del puente es obtenida a partir del siguiente algoritmo de muestreo por rechazo. Simular Y^1 y Y^2 hasta que se crucen, entonces existe alguna j , tal que cumple $Y_{\delta j}^1 \geq Y_{\delta(N-j)}^2$ y $Y_{\delta(j+1)}^1 \leq Y_{\delta(N-(j+1))}^2$ o $Y_{\delta j}^1 \leq Y_{\delta(N-j)}^2$ y $Y_{\delta(j+1)}^1 \geq Y_{\delta(N-(j+1))}^2$. Ya con esto obtenemos una trayectoria del puente definida por:

$$B_{\delta j} = \begin{cases} Y_{\delta j}^1 & j = 0, 1, \dots, \nu - 1 \\ Y_{\delta(N-j)}^2 & j = \nu, \dots, N, \end{cases} \quad (1.10)$$

donde $\nu = \min\{i \in \{1, 2, \dots, N\} | Y_{\delta i}^1 \leq Y_{\delta(N-i)}^2\}$ si $Y_0^1 \geq Y_T^2$ y $\nu = \min\{i \in \{1, 2, \dots, N\} | Y_{\delta i}^1 \geq Y_{\delta(N-i)}^2\}$ si $Y_0^1 \leq Y_T^2$. Entonces el proceso B aproxima al puente de difusión bajo las condiciones del siguiente teorema. Notemos que δ tiene influencia sobre la probabilidad de que las trayectorias se crucen, por lo que es recomendable elegir δ más pequeño de lo usual.

La probabilidad de rechazo, esto es, la probabilidad de que las trayectorias no se crucen depende de los coeficientes de deriva y de difusión, además de los valores a y b y la longitud del intervalo T .

Teorema 1.3.1. *Sea $\{X_t\}_{t \geq 0}$ una difusión con espacio de estados (l, r) , $\tau = \inf\{0 \leq t \leq T | X_t^1 = X_{T-t}^2\}$ ($\inf \emptyset = \infty$) y definamos*

$$Z_t = \begin{cases} X_t^1 & \text{si } 0 \leq t \leq \tau \\ X_{T-t}^2 & \text{si } \tau \leq t \leq T. \end{cases} \quad (1.11)$$

Asumimos que

$$\int_l^r m(x) dx < \infty,$$

con

$$m(x) = \frac{1}{\sigma^2(x)} \exp\left(2 \int_z^x \frac{b(y)}{\sigma^2(y)} dy\right), \quad x \in (l, r)$$

la función de densidad de la medida de aceleración de las difusiones. Entonces la distribución de $\{Z_t\}_{0 \leq t \leq T}$, condicionado al evento $\{\tau \leq T\}$, es igual a la distribución del puente en $[0, T]$ con valor inicial a y valor final b , condicionado al evento de que el puente está dado por una difusión independiente con ecuación diferencial dada por la ecuación (1.4) y distribución inicial $p_T(b, \cdot)$.

La demostración del teorema puede encontrarse en [4] y es muy útil para la construcción de puentes. Además puede probarse que para difusiones ergódicas la probabilidad de que el puente sea una difusión independiente es cercana a 1 si T es grande.

Capítulo 2

Modelo de efectos de mezclas

Supongamos que tenemos una población con características similares, se elige una muestra aleatoria a la cual se le realiza cierto experimento. Las respuestas de los individuos siguen cierto comportamiento común, sin embargo, es razonable pensar que existirá variabilidad debida a las diferencias entre las unidades experimentales.

El modelo de efectos de mezclas incorpora efectos fijos que modelan las características comunes y aleatorios que tratan de modelar dicha variabilidad, es por ello que se utiliza principalmente en análisis de datos longitudinales o curvas de crecimiento, es decir, cuando las observaciones son correlacionadas; así como también en el análisis de grupos donde en algunos casos los efectos aleatorios representan variables no observadas o exógenas.

La creciente popularidad de este modelo radica en la flexibilidad de modelar la variación entre y dentro de los componentes individuales. Han surgido aplicaciones en diferentes áreas tales como la biología, las ciencias sociales, la medicina, la psicología, etc.

2.1. Modelos Lineales de Efectos de Mezclas

En los modelos lineales de efectos de mezclas, los efectos fijos y aleatorios se presentan ambos de manera lineal. Supongamos que se tiene un grupo con M individuos, cada uno con n_i observaciones, $i = 1, \dots, M$.

Sea

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, & i = 1, \dots, M, \\ \mathbf{b}_i &\sim N(\mathbf{0}, \boldsymbol{\Psi}), & \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \end{aligned} \tag{2.1}$$

el modelo lineal de efectos de mezclas. Donde el vector de n_i dimensiones \mathbf{y}_i representa las respuestas del individuo i , $\boldsymbol{\beta}$ es el vector de p dimensiones de efectos fijos y \mathbf{b}_i , el vector de q dimensiones de efectos aleatorios; \mathbf{X}_i y \mathbf{Z}_i matrices de $n_i \times p$ y $n_i \times q$ llamadas matriz regresora de efectos fijos y de efectos aleatorios respectivamente y finalmente $\boldsymbol{\epsilon}_i$, el vector de n_i dimensiones, de errores de cada individuo.

Hacemos supuestos distribucionales de normalidad para los efectos aleatorios \mathbf{b}_i y los errores $\boldsymbol{\epsilon}_i$ además de independencia entre individuos y entre observaciones del mismo individuo. Cabe mencionar que la hipótesis de varianza constante de los errores puede modificarse para distintas estructuras de varianza y covarianza. Respecto a la matriz $\boldsymbol{\Psi}$ debe ser simétrica y definida positiva (es decir, todos sus eigen-valores debe ser estrictamente positivos).

En caso de que las medias de los vectores \mathbf{b}_i sean diferentes de cero, se expresan como parte de los términos de los efectos fijos, en este caso las columnas de \mathbf{Z}_i son un subconjunto de las columnas de \mathbf{X}_i .

Es conveniente expresar la matriz de varianzas y covarianzas como el *factor relativo de precisión* Δ que satisface:

$$\frac{\boldsymbol{\Psi}^{-1}}{\frac{1}{\sigma^2}} = \Delta^T \Delta.$$

Como $\boldsymbol{\Psi}$ es positiva definida, existe Δ aunque no necesariamente es única. Utilizando la factorización de Cholesky para $\sigma^2 \boldsymbol{\Psi}^{-1}$ entonces una posible solución Δ .

Un ejemplo sencillo del modelo de efectos de mezclas estudiado en [17] consiste en medir el tiempo que tarda una onda en viajar a través de ciertas vías férreas como resultado del esfuerzo longitudinal de las mismas. Se eligieron seis vías al azar y se realizaron tres mediciones en cada una, en la Figura 2.1 se muestran los resultados.

Notemos que para este ejemplo, el modelo puede escribirse considerando que tenemos $M=6$, $n_i = 3$, $i = 1, \dots, 6$, $p = q = 1$ y las matrices regresoras

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, i = 1, \dots, 6$$

Los efectos aleatorios b_i se distribuyen normal con media 0 y varianza σ_b^2 que es un escalar, por lo tanto, el factor de precisión es $\Delta = \sqrt{\sigma^2 / \sigma_b^2}$.

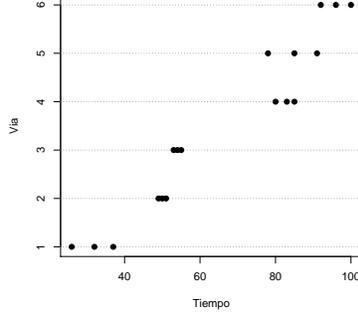


Figura 2.1: Tiempo que tarda en viajar una onda en nanosegundos muestreada en cada una de las seis vías.

2.1.1. Estimación Máximo Verosímil

Consideremos el modelo lineal de efectos de mezclas dado por la ecuación (2.1), notemos que los parámetros a estimar son β, σ^2 y θ donde θ representa los parámetros que definen a Δ . Escribimos la verosimilitud como

$$L(\beta, \theta, \sigma^2 | \mathbf{y}) = p(\mathbf{y} | \beta, \theta, \sigma^2)$$

que es la función de densidad de las respuestas dados los parámetros. Podemos observar que la densidad sólo depende de los efectos fijos pero es necesario incorporar los efectos aleatorios que también son parte del modelo. Entonces la función de verosimilitud queda de la siguiente manera:

$$\begin{aligned} L(\beta, \theta, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^M p(\mathbf{y}_i | \beta, \theta, \sigma^2) \\ &= \prod_{i=1}^M \int p(\mathbf{y}_i | \mathbf{b}_i, \beta, \sigma^2) p(\mathbf{b}_i | \theta, \sigma^2) d\mathbf{b}_i, \end{aligned}$$

donde

$$p(\mathbf{y}_i | \mathbf{b}_i, \beta, \sigma^2) = \frac{\exp(-\|\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{n_i/2}}$$

y

$$p(\mathbf{b}_i | \theta, \sigma^2) = \frac{\exp(-\mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i)}{(2\pi)^{q/2} \sqrt{|\Psi|}} = \frac{\exp(-\|\Delta\mathbf{b}_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{q/2} \text{abs}|\Delta|^{-1}},$$

son normales multivariadas.

Sustituyendo,

$$L(\beta, \theta, \sigma^2 | \mathbf{y}) = \prod_{i=1}^M \int \frac{\exp(-\|\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{n_i/2}} \frac{\exp(-\|\Delta\mathbf{b}_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{q/2} \text{abs}|\Delta|^{-1}} d\mathbf{b}_i$$

$$= \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{\exp\left\{-\frac{\|\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 - \|\Delta\mathbf{b}_i\|^2}{2\sigma^2}\right\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i. \quad (2.2)$$

Ahora, para hacer más sencilla la notación y simplificar el cálculo de los estimadores hagamos

$$\tilde{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{X}_i \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{Z}}_i = \begin{bmatrix} \mathbf{Z}_i \\ \Delta \end{bmatrix},$$

notemos que $\tilde{\mathbf{y}}_i$ es un vector de $(n_i + q)$, las matrices $\tilde{\mathbf{X}}_i$ y $\tilde{\mathbf{Z}}_i$ de dimensión $(n_i + q) \times p$ y $(n_i + q) \times q$ respectivamente. A estas matrices aumentadas se les llama pseudodatos.

Entonces reescribimos la ecuación (2.2) en función de los pseudodatos y obtenemos la siguiente función de verosimilitud

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{\exp\left(-\frac{\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i. \quad (2.3)$$

A partir de la función anterior, es posible encontrar el estimador $\hat{\mathbf{b}}_i$ minimizando la suma de cuadrados de los residuales y luego los estimadores condicionales $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ y $\hat{\sigma}^2(\boldsymbol{\theta})$ como los valores que maximizan $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ para algún $\boldsymbol{\theta}$ dado y posteriormente, encontrar $\boldsymbol{\theta}$ a partir de métodos numéricos como el algoritmo de Newton-Raphson, vía simulación con el algoritmo EM, o bien, una combinación de ambos.

Notemos que $\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2$ es la suma de cuadrados de los residuales, entonces minimizándola encontramos a \mathbf{b}_i .

$$\begin{aligned} \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 &= (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i)^T (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i) \\ &= \tilde{\mathbf{y}}_i^T \tilde{\mathbf{y}}_i - 2(\tilde{\mathbf{X}}_i\boldsymbol{\beta})^T \tilde{\mathbf{y}}_i - 2(\tilde{\mathbf{Z}}_i\mathbf{b}_i)^T \tilde{\mathbf{y}}_i + 2\mathbf{b}_i^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{X}}_i\boldsymbol{\beta} \\ &\quad + \boldsymbol{\beta}^T \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \mathbf{b}_i^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i\mathbf{b}_i, \end{aligned}$$

derivamos e igualamos a cero para encontrar el mínimo,

$$\begin{aligned} -2\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{y}}_i + 2\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{X}}_i\boldsymbol{\beta} + 2\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i\mathbf{b}_i &= 0, \\ \tilde{\mathbf{Z}}_i^T (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta}) &= \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i\mathbf{b}_i, \end{aligned}$$

despejando, obtenemos $\hat{\mathbf{b}}_i = (\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i)^{-1} \tilde{\mathbf{Z}}_i^T (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta})$.

Luego,

$$\begin{aligned} \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 &= \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 + \|\tilde{\mathbf{Z}}_i(\mathbf{b}_i - \hat{\mathbf{b}}_i)\|^2 \\ &= \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 + (\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i). \end{aligned}$$

Sustituimos en la ecuación (2.3),

$$\begin{aligned} & \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \int \frac{\exp(-\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 + (\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)/2\sigma^2)}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i. \\ &= \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{n_i/2}} \exp(-\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2/2\sigma^2) \\ & \quad \int \frac{\exp(-(\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i))}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i, \end{aligned}$$

observemos que podemos completar una distribución normal q variada con media $\hat{\mathbf{b}}_i$ y varianza $\tilde{\mathbf{Z}}_i^2/\sigma^2$

$$= \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{n_i/2}} \frac{\exp(-\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2/2\sigma^2)}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}}.$$

Por lo tanto, la función de verosimilitud queda como sigue:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_{i=1}^M \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2}{2\sigma^2}\right) \prod_{i=1}^M \frac{\text{abs}|\Delta|}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}}, \quad (2.4)$$

con $N = \sum_{i=1}^M n_i$.

De esta expresión, se calculan los estimadores $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ y $\hat{\sigma}^2(\boldsymbol{\theta})$ para $\boldsymbol{\theta}$ dado, de la misma manera que se haría en una regresión lineal.

Notemos que los mínimos cuadrados para estimar $\boldsymbol{\beta}$ dependen de $\hat{\mathbf{b}}_i$ que a su vez depende de $\boldsymbol{\beta}$, por lo que debemos determinar los valores de los mínimos cuadrados conjuntamente, es decir,

$$\left(\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_M^T, \hat{\boldsymbol{\beta}}^T\right)^T = \min_{\mathbf{b}_1, \dots, \mathbf{b}_M, \boldsymbol{\beta}} \|\mathbf{y}_e - \mathbf{X}_e(\mathbf{b}_1, \dots, \mathbf{b}_M, \boldsymbol{\beta})^T\|^2,$$

donde

$$\mathbf{y}_e = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{0} \\ \mathbf{y}_2 \\ \mathbf{0} \\ \vdots \\ \mathbf{y}_M \\ \mathbf{0} \end{bmatrix} \quad \text{y} \quad \mathbf{X}_e = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}_1 \\ \boldsymbol{\Delta} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} & \mathbf{X}_2 \\ \mathbf{0} & \boldsymbol{\Delta} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_M & \mathbf{X}_M \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Delta} & \mathbf{0} \end{bmatrix}.$$

Por lo tanto, escribimos al vector de estimadores dado $\boldsymbol{\theta}$

$$\left(\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_M^T, \hat{\boldsymbol{\beta}}^T\right)^T = \left(\mathbf{X}_e^T \mathbf{X}_e\right)^{-1} \mathbf{X}_e^T \mathbf{y}_e.$$

A partir de la teoría de regresión lineal también encontramos al estimador de la varianza de los errores ,

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{\left\| \mathbf{y}_e - \mathbf{X}_e \left(\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_M^T, \hat{\boldsymbol{\beta}}^T \right)^T \right\|^2}{N}.$$

Finalmente, sustituyendo los estimadores condicionales en la función de verosimilitud dada por la ecuación (2.4), obtenemos que

$$L(\boldsymbol{\theta}) = L(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta})) = \frac{\exp(-N/2)}{[2\pi\hat{\sigma}^2(\boldsymbol{\theta})]^N/2} \prod_{i=1}^M \frac{\text{abs}|\Delta|}{\sqrt{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|}}.$$

Sin embargo, computacionalmente, no es conveniente trabajar con la matriz \mathbf{X}_e ya que puede ser muy grande, por lo cual es preferible utilizar la descomposición QR (*Ortogonal-triangular*) descrita en el Apéndice A para encontrar los estimadores condicionados a $\boldsymbol{\theta}$. Y el vector de parámetros $\boldsymbol{\theta}$ se encuentra a partir de métodos numéricos como ya habíamos mencionado.

Descomposición QR para encontrar los estimadores

Consideremos la descomposición

$$\tilde{\mathbf{Z}}_i = \mathbf{Q}_{(i)} \begin{bmatrix} \mathbf{R}_{11(i)} \\ \mathbf{0} \end{bmatrix},$$

donde $\mathbf{Q}_{(i)}$ es una matriz ortogonal de $(n_i + q) \times (n_i + q)$ y $\mathbf{R}_{11(i)}$ triangular superior de $p \times p$.

Tomamos la parte de la exponencial en la ecuación (2.3), multiplicamos por $\mathbf{Q}_{(i)}^T$ que como es ortogonal preserva la norma y sustituimos a $\tilde{\mathbf{Z}}_i$,

$$\begin{aligned} \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \mathbf{b}_i\|^2 &= \|\mathbf{Q}_{(i)}^T (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \mathbf{b}_i)\|^2 \\ &= \|\mathbf{Q}_{(i)}^T \tilde{\mathbf{y}}_i - \mathbf{Q}_{(i)}^T \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \mathbf{Q}_{(i)}^T \tilde{\mathbf{Z}}_i \mathbf{b}_i\|^2 \\ &= \|\mathbf{Q}_{(i)}^T \tilde{\mathbf{y}}_i - \mathbf{Q}_{(i)}^T \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \mathbf{Q}_{(i)}^T \mathbf{Q}_{(i)} \begin{bmatrix} \mathbf{R}_{11(i)} \\ \mathbf{0} \end{bmatrix} \mathbf{b}_i\|^2. \end{aligned}$$

Hagamos

$$\mathbf{Q}_{(i)}^T \tilde{\mathbf{y}}_i = \begin{bmatrix} \mathbf{C}_{1(i)} \\ \mathbf{C}_{0(i)} \end{bmatrix}, \quad \mathbf{Q}_{(i)}^T \tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{R}_{10(i)} \\ \mathbf{R}_{00(i)} \end{bmatrix},$$

entonces sustituyendo las expresiones anteriores y desarrollando

$$\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \mathbf{b}_i\|^2 = \|\mathbf{C}_{1(i)} - \mathbf{R}_{10(i)} \boldsymbol{\beta} - \mathbf{R}_{11(i)} \mathbf{b}_i\|^2 + \|\mathbf{C}_{0(i)} - \mathbf{R}_{00(i)} \boldsymbol{\beta}\|^2.$$

Regresando a la verosimilitud, podemos escribir la integral de la ecuación (2.3) como

$$\exp \left[\frac{\|\mathbf{C}_{0(i)} - \mathbf{R}_{00(i)} \boldsymbol{\beta}\|^2}{-2\sigma^2} \right] \int \frac{\exp \left(-\frac{\|\mathbf{C}_{1(i)} - \mathbf{R}_{10(i)} \boldsymbol{\beta} - \mathbf{R}_{11(i)} \mathbf{b}_i\|^2}{2\sigma^2} \right)}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i.$$

Para resolver esta integral realizamos un cambio de variable que puede hacerse porque $\mathbf{R}_{11(i)}$ es no singular, luego $\phi_i = (\mathbf{C}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i)/\sigma$ y su diferencial $d\phi_i = \sigma^{-q}\text{abs}|\mathbf{R}_{11(i)}|$ y reescribimos la integral

$$\begin{aligned} & \int \frac{\exp\left(-\|\mathbf{C}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2/2\sigma^2\right)}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= \frac{1}{\text{abs}|\mathbf{R}_{11(i)}|} \int \frac{\exp(-\|\phi_i\|^2/2)}{2\pi^{q/2}} d\phi_i \\ &= \frac{1}{\text{abs}|\mathbf{R}_{11(i)}|}. \end{aligned}$$

Se sigue que la verosimilitud es

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^M \frac{\text{abs}|\Delta|}{\text{abs}|\mathbf{R}_{11(i)}|} \exp\left[\frac{\|\mathbf{C}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2/2\sigma^2}{(2\pi\sigma^2)^{n_i/2}}\right] \\ &= \exp\left[\frac{\sum_{i=1}^M \|\mathbf{C}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2/2\sigma^2}{(2\pi\sigma^2)^{N/2}}\right] \prod_{i=1}^M \frac{\text{abs}|\Delta|}{\text{abs}|\mathbf{R}_{11(i)}|}, \end{aligned}$$

podemos observar que dentro de la exponencial nuevamente tenemos una suma de cuadrados de residuales, por lo cual, utilizamos por segunda vez la descomposición QR para R_{00} con

$$R_{00} = \begin{pmatrix} R_{00(1)} \\ R_{00(2)} \\ \vdots \\ R_{00(M)} \end{pmatrix} \quad \text{y} \quad C_0 = \begin{pmatrix} C_{0(1)} \\ C_{0(2)} \\ \vdots \\ C_{0(M)} \end{pmatrix},$$

luego, $R_{00} = \mathbf{Q}_0 \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{bmatrix}$ y siguiendo el mismo procedimiento multiplicamos por \mathbf{Q}_0^T y sustituimos la descomposición y obtenemos finalmente

$$\|\mathbf{C}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2 = \|\mathbf{C}_{01} - \mathbf{R}_0\boldsymbol{\beta}\|^2 + \|\mathbf{C}_{02}\|^2,$$

donde $\mathbf{Q}_0^T \mathbf{C}_0 = (\mathbf{C}_{01} \mathbf{c}_{02})$.

Por lo tanto, la función de verosimilitud queda en términos de estas últimas matrices, asumiendo \mathbf{R}_{00} no singular

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{N/2} \exp\left(\frac{\|\mathbf{C}_{01} - \mathbf{R}_0\boldsymbol{\beta}\|^2 + \|\mathbf{C}_{02}\|^2}{-2\sigma^2}\right) \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|}\right),$$

de aquí, se ve que los estimadores máximo verosímil dado $\boldsymbol{\theta}$ son los siguientes:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{R}_0^{-1} \mathbf{C}_{01}, \quad \hat{\sigma}^2(\boldsymbol{\theta}) = \frac{\|\mathbf{C}_{02}\|^2}{N} \quad \text{y} \quad \hat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbf{R}_{11(i)}(\mathbf{C}_{1(i)} - \mathbf{R}_{10(i)}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})).$$

Ya que técnicamente los efectos aleatorios \mathbf{b}_i no son parámetros del modelo estadístico, pero también deben estimarse, el estimador condicionado de tales efectos aleatorios al estimador condicional de $\boldsymbol{\beta}$ es un BLUP *Best Linear Unbiased Predictors* (Mejor predictor lineal insesgado) para las \mathbf{b}_i $i = 1, \dots, M$.

Sustituyendo los estimadores para encontrar la función de verosimilitud para $\boldsymbol{\theta}$

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &= L(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta})|\mathbf{y}) \\ &= \left(\frac{N}{2\pi \|\mathbf{C}_{02}\|^2} \right)^{N/2} \exp\left(-\frac{N}{2}\right) \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}}\right), \end{aligned}$$

y la log-verosimilitud

$$\log(L(\boldsymbol{\theta}|\mathbf{y})) = \frac{N}{2}(\log(N) - \log(2\pi) - 1) - N \log \|\mathbf{C}_{02}\|^2 + \sum_{i=1}^M \log \text{abs}\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}}\right).$$

Para concluir, los estimadores quedan al escribir el valor de $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, que se obtiene al maximizar la log-verosimilitud.

2.1.2. Generalización del modelo

Podemos pensar en generalizar el modelo en dos sentidos, extender el modelo lineal a múltiples niveles de efectos aleatorios, en el caso de dos por ejemplo tendríamos lo siguiente:

$$\begin{aligned} \mathbf{y}_{ij} &= \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{b}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, M_i \\ \mathbf{b}_i &\sim N(\mathbf{0}, \boldsymbol{\Psi}_1), \quad \mathbf{b}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Psi}_2), \quad \boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \end{aligned}$$

donde \mathbf{y}_{ij} es el vector de respuestas, M representa el primer nivel de los grupos y M_i el segundo nivel dentro del grupo i . Las matrices de efectos fijos \mathbf{X}_{ij} de $n_{ij} \times p$. Los efectos aleatorios \mathbf{b}_i y \mathbf{b}_{ij} de primer y segundo nivel de tamaño q_1 y q_2 respectivamente; y sus correspondientes matrices $\mathbf{Z}_{i,j}$ de $n_i \times q_1$ y \mathbf{Z}_{ij} de $n_i \times q_2$. La estimación de los modelos lineales con multiniveles se hace de manera análoga a la presentada en la sección anterior aunque se complica un poco la notación.

Por otra parte, se puede generalizar el modelo relajando las hipótesis de varianza de los errores $\boldsymbol{\epsilon}_i$, permitiendo heterocedasticidad y correlaciones dentro de los grupos, y se expresa como

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M, \\ \mathbf{b}_i &\sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i). \end{aligned} \tag{2.5}$$

Donde $\boldsymbol{\Lambda}_i$ son matrices positivo definidas parametrizadas por el conjunto $\boldsymbol{\lambda}$ de parámetros fijos. De la misma manera que en el modelo lineal de efectos de mezclas básico, los errores se asumen independientes para cada i e independientes

a los efectos aleatorios \mathbf{b}_i .

Para la estimación de el modelo lineal generalizado, como Λ_i es positiva definida, admite una raíz cuadrada invertible $\Lambda_i^{-1/2}$ tal que

$$\Lambda_i = (\Lambda_i^{1/2})^T \Lambda_i^{1/2} \quad \text{y} \quad \Lambda_i^{-1} = \Lambda_i^{-1/2} (\Lambda_i^{-1/2})^T.$$

Si hacemos

$$\begin{aligned} \mathbf{y}_i^* &= (\Lambda_i^{-1/2})^T \mathbf{y}_i & \boldsymbol{\epsilon}_i^* &= (\Lambda_i^{-1/2})^T \boldsymbol{\epsilon}_i, \\ \mathbf{X}_i^* &= (\Lambda_i^{-1/2})^T \mathbf{X}_i & \mathbf{Z}_i^* &= (\Lambda_i^{-1/2})^T \mathbf{Z}_i. \end{aligned}$$

Entonces notemos que

$$\boldsymbol{\epsilon}_i^* \sim N\left((\Lambda_i^{-1/2})^T \mathbf{0}, \sigma^2 (\Lambda_i^{-1/2})^T \Lambda_i \Lambda_i^{-1/2}\right) = N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Por lo cual, podemos escribir (2.5) como

$$\begin{aligned} \mathbf{y}_i^* &= \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{Z}_i^* \mathbf{b}_i + \boldsymbol{\epsilon}_i^*, & i &= 1, \dots, M, \\ \mathbf{b}_i &\sim N(\mathbf{0}, \boldsymbol{\Psi}), & \boldsymbol{\epsilon}_i^* &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}). \end{aligned}$$

Esto implica que podemos describir a \mathbf{y}_i^* con un modelo básico lineal de efectos de mezclas.

Como el diferencial de la transformación lineal $\mathbf{y}_i^* = (\Lambda_i^{-1/2})^T \mathbf{y}_i$ es $d\mathbf{y}_i^* = |\Lambda_i^{-1/2}| d\mathbf{y}_i$ la verosimilitud correspondiente al modelo (2.5) se expresa de la siguiente manera

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}) &= \prod_{i=1}^M p(\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) \\ &= \prod_{i=1}^M p(\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}) |\Lambda_i^{-1/2}| = L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}^*) \prod_{i=1}^M |\Lambda_i^{-1/2}|, \end{aligned}$$

donde $p(\cdot)$ denota la función de densidad. Y como $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}^*)$ representa la verosimilitud de un modelo lineal básico, se aplica el mismo desarrollo.

Capítulo 3

Modelo de efectos de mezclas en ecuaciones diferenciales estocásticas

Como habíamos mencionado en los dos capítulos anteriores, tenemos por un lado las ecuaciones diferenciales estocásticas que surgen al incorporar un componente aleatorio a un sistema modelado a partir de una ecuación diferencial y por otro lado, los modelos de efectos de mezclas que asumen que las observaciones de un grupo siguen el mismo modelo, sin embargo existen variaciones entre los individuos.

Si integramos en un mismo modelo ambos conceptos obtenemos el modelo de efectos de mezclas en ecuaciones diferenciales estocásticas, donde el comportamiento de cada individuo se modela a partir de una ecuación diferencial estocástica, esto es, modela la variación en el conjunto de observaciones de cada individuo; mientras que los parámetros de cada ecuación diferencial estocástica son modelados por los efectos aleatorios; es decir, son variables aleatorias que modelan la variabilidad entre los individuos.

Las principales aplicaciones del modelo de efectos de mezclas en ecuaciones diferenciales estocásticas se dan en investigaciones biomédicas, especialmente en las áreas de la farmacología que estudian el tiempo en que cierta sustancia es totalmente eliminada del cuerpo y los efectos bioquímicos y fisiológicos, así como la relación entre la concentración y el efecto que la sustancia produce. (*pharmacokinetic* y *pharmacodynamic*, en inglés).

3.1. El modelo

Sea un proceso estocástico continuo $\{X_t\}_{t \geq 0}$, unidimensional que se desarrolla en M unidades experimentales o individuos elegidos aleatoriamente de una población. El modelo de efectos de mezclas en ecuaciones diferenciales estocásticas se define:

$$dX_t^i = \mu(X_t^i, \boldsymbol{\theta}, \mathbf{b}^i)dt + \sigma(X_t^i, \boldsymbol{\theta}, \mathbf{b}^i)dB_t^i, \quad i = 1, \dots, M,$$

$$\mathbf{b}^i \sim N(0, \boldsymbol{\Sigma}) \quad \text{y} \quad X_0^i = x_0^i,$$

donde X_t^i representa el valor del proceso al tiempo t del individuo i , $\boldsymbol{\theta}$ es un vector de p dimensiones de los parámetros de los efectos fijos, \mathbf{b}^i es el vector de q dimensiones de los parámetros de los efectos aleatorios, $\boldsymbol{\Sigma}$ es la matriz de varianzas y covarianzas de los efectos aleatorios la cual se asume conocida a través del vector de parámetros $\boldsymbol{\Psi}$. Los $\{B_t^i\}_{t \geq 0}$ son movimientos Brownianos estándar independientes a las \mathbf{b}^i y a los valores iniciales X_0^i para toda i .

Los coeficientes de deriva y difusión μ y σ se suponen conocidos a partir de los parámetros y además cumplen las condiciones para obtener la solución única descritas en el Teorema 1.3.1.

Sea $E \subseteq \mathbb{R}$ el espacio de estados de X_t^i y asumimos que la distribución condicional de X_t^i dado \mathbf{b}^i y $X_s = x$ para $s < t$ tiene densidad estrictamente positiva con respecto a la medida de Lebesgue en E , denotada por

$$y \rightarrow p(y, x, t - s | \mathbf{b}^i, \boldsymbol{\theta}) \quad y, x \in E.$$

Consideremos que se tienen las observaciones en los tiempos $(t_0^i, t_1^i, \dots, t_{n_i}^i)$, es decir, se tienen $n_i + 1$ observaciones por individuo. Entonces \mathbf{y}^i denota el vector de $n_i + 1$ dimensiones de respuestas para cada $i = 1, 2, \dots, M$; para simplificar la notación escribimos $y_{t_j^i}^i = y_j^i$ las entradas del vector \mathbf{y}^i , y sea \mathbf{y} el vector de N dimensiones de respuestas, donde $N = \sum_{i=1}^M (n_i + 1)$. Además escribimos a la distancia entre la $j - 1$ y la j observación, como $\Delta_j^i = t_j^i - t_{j-1}^i$ para cada individuo i .

El problema consiste en encontrar los estimadores de los parámetros $\boldsymbol{\theta}$ y $\boldsymbol{\Psi}$, supongamos que contamos con el vector de respuestas \mathbf{y} , a partir del cual escribimos la función de verosimilitud, con la cual queremos hacer la inferencia. Debemos tener en cuenta que en la mayoría de los casos no es posible encontrar una expresión explícita para la función de verosimilitud, sin embargo en casos sencillos es posible hallar expresiones de los estimadores máximo verosimil de $\boldsymbol{\theta}$ y $\boldsymbol{\Psi}$.

3.2. Función de verosimilitud

Ya que la idea es encontrar la función de verosimilitud de los parámetros dado el conjunto datos de todos los individuos e incorporar los efectos aleatorios al

modelo, integramos la densidad condicional de las respuestas dados los efectos aleatorios no observados \mathbf{b}^i con respecto a la densidad marginal de los mismos. Además usando el supuesto de independencia entre \mathbf{b}^i y B_t^i , obtenemos

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) = \prod_{i=1}^M p(\mathbf{y}^i|\boldsymbol{\theta}, \boldsymbol{\Psi}) = \prod_{i=1}^M \int p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta})p(\mathbf{b}^i|\boldsymbol{\Psi})d\mathbf{b}^i \quad (3.1)$$

donde $p(\cdot)$ denota la función de densidad.

Ahora, como X_t^i dado \mathbf{b}^i cumple la propiedad de Markov, podemos escribir la densidad $p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta})$ como probabilidades de transición,

$$p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta}) = \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\mathbf{b}^i, \boldsymbol{\theta}).$$

Luego, por hipótesis \mathbf{b}^i tiene distribución normal,

$$p(\mathbf{b}^i|\boldsymbol{\Psi}) = \frac{1}{\sqrt{|\boldsymbol{\Psi}|}(2\pi)^{q/2}} \exp \left\{ \frac{-(\mathbf{b}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{b}^i}{2} \right\}$$

Sustituyendo ambas expresiones en la ecuación (3.1)

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) = \prod_{i=1}^M \int \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\mathbf{b}^i, \boldsymbol{\theta}) \frac{\exp \left\{ -(\mathbf{b}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{b}^i / 2 \right\}}{\sqrt{|\boldsymbol{\Psi}|}(2\pi)^{q/2}} d\mathbf{b}^i \quad (3.2)$$

Basta con resolver la integral para encontrar los estimadores máximo verosímil de los parámetros, sin embargo, no siempre es posible encontrar una solución explícita ya sea porque se desconoce la densidad de transición o bien, la integral no puede ser resuelta analíticamente.

3.3. El caso del Movimiento Browniano geométrico

Un ejemplo sencillo donde se encuentra la función de densidad de transición y además no resulta difícil resolver la integral con teoría de mínimos cuadrados es el caso del movimiento Browniano geométrico. Consideremos la ecuación diferencial estocástica de la forma

$$dX_t = \beta X_t dt + \sigma X_t dB_t, \quad (3.3)$$

encontramos la solución igualando los coeficientes con los de la fórmula de Itô, $dX_t = f_t(t, B_t)dt + f_x(t, B_t)dX_t + \frac{1}{2}f_{xx}(t, B_t)dt$, si tomamos a $X_t = f(t, B_t)$ entonces,

$$\beta f(t, x) = f_t(t, x) + \frac{1}{2}f_{xx}(t, x) \quad (3.4)$$

$$\sigma f(t, x) = f_x(t, B_t) \quad (3.5)$$

De la ecuación (3.4) tenemos que $f(t, x) = e^{\sigma x + g(t)}$ donde $g(t)$ es una función arbitraria y por la ecuación (3.5) sabemos que $g'(t) = \beta - \frac{1}{2}\sigma^2$, por lo tanto, la ecuación diferencial estocástica tiene solución $X_t = X_0 e^{(\beta - \frac{1}{2}\sigma^2)t + \sigma B_t}$, que es un movimiento Browniano geométrico. Tomamos el proceso $Y_t = \log(X_t)$, luego $Y_t = Y_0 + (\beta - \frac{1}{2}\sigma^2)t + \sigma B_t$.

Asumimos que el experimento que representa la anterior ecuación diferencial estocástica, se realiza a diferentes individuos y además se les realizan distintas mediciones en el tiempo. Como estamos interesados en estimar los parámetros de toda la población incluyendo las características propias de cada individuo incorporamos un efecto aleatorio, lo cual nos lleva a la ecuación diferencial estocástica

$$dX_t^i = \left(\beta + \beta^i - \frac{1}{2}\sigma^2 \right) dt + \sigma dB_t^i, \quad (3.6)$$

cuya solución está dada por el modelo:

$$Y_t^i = Y_0^i + \left(\beta + \beta^i - \frac{1}{2}\sigma^2 \right) t + \sigma B_t^i, \quad (3.7)$$

donde $\beta^i \sim N(0, \sigma_\beta^2)$.

En este ejemplo los parámetros a estimar son $\boldsymbol{\theta} = (\beta, \sigma^2)$ y $\boldsymbol{\Psi} = \sigma_\beta^2$. Se puede observar que por la hipótesis de normalidad sobre las β^i la distribución condicional de Y_t^i dado $Y_0^i = y_0^i; \beta, \beta^i, \sigma^2$ también es normal con media y varianza siguientes:

$$\begin{aligned} \mathbb{E}[Y_t^i | Y_0^i = y_0^i; \beta, \beta^i, \sigma^2] &= y_0^i + \left(\beta + \beta^i - \frac{1}{2}\sigma^2 \right) t + \sigma \mathbb{E}[B_t^i] \\ &= y_0^i + \left(\beta + \beta^i - \frac{1}{2}\sigma^2 \right) t, \end{aligned}$$

$$\begin{aligned} \text{Var}[Y_t^i | Y_0^i = y_0^i; \beta, \beta^i, \sigma^2] &= \text{Var}[y_0^i + \left(\beta + \beta^i - \frac{1}{2}\sigma^2 \right) t + \sigma B_t^i] \\ &= \sigma^2 \text{Var}[B_t^i] \\ &= \sigma^2 t. \end{aligned}$$

Entonces,

$$p(y_j^i, y_{j-1}^i, \boldsymbol{\Delta}_j^i; \beta, \beta^i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2\boldsymbol{\Delta}_j^i}} \exp \left\{ -\frac{(y_j^i - y_{j-1}^i - (\beta + \beta^i - \frac{1}{2}\sigma^2)\boldsymbol{\Delta}_j^i)^2}{2\sigma^2\boldsymbol{\Delta}_j^i} \right\}.$$

Sustituyendo la densidad que acabamos de encontrar en la función de verosimilitud descrita en (3.2),

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) = \prod_{i=1}^M \int \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp \left\{ -\frac{(y_j^i - y_{j-1}^i - (\beta + \beta^i - \frac{1}{2}\sigma^2)\Delta_j^i)^2}{2\sigma^2\Delta_j^i} \right\} \times \frac{1}{\sqrt{(2\pi\sigma_\beta^2)}} \exp \left\{ \frac{-(\beta^i)^2}{2\sigma_\beta^2} \right\} d\beta^i.$$

Para simplificar el modelo, utilizamos el factor de precisión aplicado en el modelo de regresión lineal con efecto aleatorios, definimos $\eta^2 = \sigma^2/\sigma_\beta^2$ para obtener la siguiente verosimilitud:

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) = \prod_{i=1}^M \frac{\sqrt{\eta^2}}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \int \frac{\exp \left\{ -\frac{\sum_{j=1}^{n_i} \frac{1}{\Delta_j^i} (y_j^i - y_{j-1}^i - (\beta + \beta^i - \frac{1}{2}\sigma^2)\Delta_j^i)^2 + (\eta\beta^i)^2}{2\sigma^2} \right\}}{\sqrt{2\pi\sigma^2}} d\beta^i \prod_{j=1}^{n_i} \frac{1}{(\Delta_j^i)^{1/2}}.$$

Si además escribimos $\alpha = \beta - \frac{1}{2}\sigma^2$ y $\Delta^i = (\prod_{j=1}^{n_i} \frac{1}{\Delta_j^i})^{1/n_i}$ la función de verosimilitud queda un poco más reducida,

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) = \prod_{i=1}^M \frac{\sqrt{\eta^2}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}} \int \frac{\exp \left\{ -\frac{\sum_{j=1}^{n_i} \frac{1}{\Delta_j^i} (y_j^i - y_{j-1}^i - (\alpha + \beta^i)\Delta_j^i)^2 + (\eta\beta^i)^2}{2\sigma^2} \right\}}{\sqrt{2\pi\sigma^2}} d\beta^i.$$

3.3.1. Estimación con efectos aleatorios

Al resolver la integral de la función de verosimilitud, obtenemos la densidad marginal de los parámetros independiente a los efectos aleatorios para lo cual definimos los siguientes vectores que nos permitirán escribir la expresión dentro de la exponencial como un problema de mínimos cuadrados.

Sean

$$\begin{aligned} \tilde{\mathbf{y}}^i &= ((\Delta_1^i)^{-1/2}(y_1^i - y_0^i), \dots, (\Delta_{n_i}^i)^{-1/2}(y_{n_i}^i - y_{n_i-1}^i), 0)^T, \\ \tilde{\mathbf{x}}^i &= ((\Delta_1^i)^{1/2}, \dots, (\Delta_{n_i}^i)^{1/2}, 0)^T, \\ \tilde{\mathbf{z}}^i &= ((\Delta_1^i)^{1/2}, \dots, (\Delta_{n_i}^i)^{1/2}, \eta)^T. \end{aligned}$$

Entonces,

$$\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2 = \sum_{j=1}^{n_i} \left((\Delta_j^i)^{-1/2}(y_j^i - y_{j-1}^i) - (\alpha + \beta^i)\Delta_j^i \right)^2 + (\beta^i\eta)^2,$$

desarrollamos el cuadrado y factorizamos $(\Delta_j^i)^{-1}$

$$= \sum_{j=1}^{n_i} (\Delta_j^i)^{-1} \left((y_j^i - y_{j-1}^i) - (\alpha + \beta^i) \Delta_j^i \right)^2 + (\beta^i \eta)^2.$$

Por lo tanto,

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi} | \mathbf{y}) = \prod_{i=1}^M \frac{\sqrt{\eta^2}}{(2\pi\sigma^2 \Delta^i)^{n_i/2}} \int \frac{\exp \left\{ -\frac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i \alpha - \tilde{\mathbf{z}}^i \beta^i\|^2}{2\sigma^2} \right\}}{\sqrt{2\pi\sigma^2}} d\beta^i. \quad (3.8)$$

A partir de la teoría de regresión lineal puede encontrarse el estimador de los efectos aleatorios que minimiza la suma de cuadrados $\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i \alpha - \tilde{\mathbf{z}}^i \beta^i\|^2$ para un α fijo, es decir,

$$\begin{aligned} \hat{\beta}^i &= ((\tilde{\mathbf{z}}^i)^T \tilde{\mathbf{z}}^i)^{-1} (\tilde{\mathbf{z}}^i)^T (\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i \alpha) \\ &= \left(\sum_{j=1}^{n_i} \tilde{z}_j^i \tilde{z}_j^i \right)^{-1} \left(\sum_{j=1}^{n_i} \tilde{z}_j^i \tilde{y}_j^i - \sum_{j=1}^{n_i} \tilde{z}_j^i \tilde{x}_j^i \alpha \right) \\ &= \left(\sum_{j=1}^{n_i} \Delta_j^i + \eta^2 \right)^{-1} \left(\sum_{j=1}^{n_i} y_j - y_{j-1} - \Delta_j^i \alpha \right) \\ &= \frac{y_{n_i} - y_0 - \alpha T^i}{T^i + \eta^2}, \end{aligned} \quad (3.9)$$

con $T^i = \sum_{j=1}^{n_i} \Delta_j^i$, por consiguiente hallamos directamente al estimador de β^i dados los parámetros $\boldsymbol{\theta}$ y $\boldsymbol{\Psi}$.

Por otro lado, la suma de cuadrados que esta dentro de la exponencial puede separarse de tal manera que solo una parte dependa de los efectos aleatorios, entonces reescribimos la integral de la ecuación (3.8)

$$\exp \left\{ -\frac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i \alpha - \tilde{\mathbf{z}}^i \hat{\beta}^i\|^2}{2\sigma^2} \right\} \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{\|\tilde{\mathbf{z}}^i (\beta^i - \hat{\beta}^i)\|^2}{2\sigma^2} \right\} d\beta^i,$$

multiplicamos y dividimos por $((\tilde{\mathbf{z}}^i)^T \tilde{\mathbf{z}}^i)^{1/2}$ la expresión anterior para completar la función de densidad de una normal; por lo tanto, β^i tiene media $\hat{\beta}^i$ y varianza $\sigma^2 / ((\tilde{\mathbf{z}}^i)^T \tilde{\mathbf{z}}^i)$. Recordemos que $(\tilde{\mathbf{z}}^i)^T \tilde{\mathbf{z}}^i = T^i + \eta^2$.

Por consiguiente, la función de verosimilitud queda expresada como:

$$\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) &= \prod_{i=1}^M \frac{\sqrt{\eta^2}}{(2\pi\sigma^2\Delta^i)^{n_i/2}} \frac{1}{\sqrt{T^i + \eta^2}} \exp \left\{ - \frac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\hat{\beta}^i\|^2}{2\sigma^2} \right\} \\
&= \frac{(\eta^2)^{M/2}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^M \frac{1}{(\Delta^i)^{\frac{n_i}{2}} \sqrt{T^i + \eta^2}} \times \\
&\quad \exp \left\{ - \frac{\sum_{i=1}^M \|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\hat{\beta}^i\|^2}{2\sigma^2} \right\} \\
&= \frac{(\eta^2)^{M/2}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^M \frac{1}{(\Delta^i)^{\frac{n_i}{2}} \sqrt{T^i + \eta^2}} \times \\
&\quad \exp \left\{ - \frac{\sum_{i=1}^M \sum_{j=1}^{n_i} (\Delta_j^i)^{-1} \left((y_j^i - y_{j-1}^i) - (\alpha + \hat{\beta}^i)\Delta_j^i \right)^2 + (\hat{\beta}^i\eta)^2}{2\sigma^2} \right\}.
\end{aligned}$$

Tomamos la suma que está dentro de la exponencial en la función de verosimilitud $\sum_{i=1}^M \sum_{j=1}^{n_i} (\Delta_j^i)^{-1} \left((y_j^i - y_{j-1}^i) - \alpha\Delta_j^i - \hat{\beta}^i\Delta_j^i \right)^2$, desarrollamos el cuadrado, distribuimos $(\Delta_j^i)^{-1}$ y sustituimos el estimador $\hat{\beta}^i$ para obtener la verosimilitud

$$\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) &= \frac{(\eta^2)^{M/2}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^M \frac{1}{(\Delta^i)^{\frac{n_i}{2}} \sqrt{T^i + \eta^2}} \times \\
&\quad \exp \left\{ - \frac{1}{2\sigma^2} \left[\sum_{i=1}^M \sum_{j=1}^{n_i} (\Delta_j^i)^{-1} \left(y_j^i - y_{j-1}^i - \alpha\Delta_j^i \right)^2 \right. \right. \\
&\quad \left. \left. - \sum_{i=1}^M (y_{n_i}^i - y_0^i - \alpha T^i)^2 (T^i + \eta^2)^{-1} \right] \right\}. \tag{3.10}
\end{aligned}$$

Y la log-verosimilitud

$$\begin{aligned}
\log(L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y})) &= \frac{M}{2} \log(\eta^2) - \frac{N-M}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^M \log((\Delta^i)^{n_i} (T^i + \eta^2)) \\
&\quad - \frac{1}{2\sigma^2} \left[\sum_{i=1}^M \sum_{j=1}^{n_i} (\Delta_j^i)^{-1} \left(y_j^i - y_{j-1}^i - \alpha\Delta_j^i \right)^2 \right. \\
&\quad \left. - \sum_{i=1}^M (y_{n_i}^i - y_0^i - \alpha T^i)^2 (T^i + \eta^2)^{-1} \right]. \tag{3.11}
\end{aligned}$$

Para hacer el cálculo de los estimadores, asumimos que cada individuo es observado en el mismo intervalo de tiempo, esto quiere decir que $T^i = T$ para

cada $1 \leq i \leq M$, con ello se consigue simplificar las expresiones de los estimadores, luego desarrollamos los cuadrados de las sumas y notamos que quedan expresadas en términos de tres sumas constantes definidas como:

$$\begin{aligned}
S1 &= \sum_{i=1}^M (y_{n_i}^i - y_0^i), \\
S2 &= \sum_{i=1}^M (y_{n_i}^i - y_0^i)^2, \\
S3 &= \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{1}{\Delta_j^i} (y_j^i - y_{j-1}^i)^2.
\end{aligned} \tag{3.12}$$

Por lo tanto,

$$\begin{aligned}
\sum_{i=1}^M \sum_{j=1}^{n_i} \frac{1}{\Delta_j^i} (y_j^i - y_{j-1}^i - \alpha \Delta_j^i)^2 &= \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{1}{\Delta_j^i} (y_j^i - y_{j-1}^i)^2 - 2\alpha \sum_{i=1}^M (y_{n_i}^i - y_0^i) \\
&\quad + \alpha^2 MT \\
&= S3 - 2\alpha S2 + \alpha^2 MT,
\end{aligned}$$

analogamente,

$$\begin{aligned}
\sum_{i=1}^M \frac{(y_{n_i}^i - y_0^i - \alpha T^i)^2}{(T^i + \eta^2)} &= \frac{\sum_{i=1}^M (y_{n_i}^i - y_0^i)^2 - 2\alpha T \sum_{i=1}^M (y_{n_i}^i - y_0^i) + \alpha^2 MT^2}{(T + \eta^2)} \\
&= \frac{S2 - 2\alpha TS1 + \alpha^2 MT^2}{(T + \eta^2)}.
\end{aligned}$$

Reescribimos la log-verosimilitud en términos de los parámetros originales, es decir, sustituyendo $\alpha = \beta - \sigma^2/2$, $\eta^2 = \sigma^2/\sigma_\beta^2$ y los desarrollos de las sumas anteriores

$$\begin{aligned}
\log(L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y})) &= \frac{M}{2} \log\left(\frac{\sigma^2}{\sigma_\beta^2}\right) - \frac{N-M}{2} \log(2\pi\sigma^2) \\
&\quad - \frac{1}{2} \sum_{i=1}^M \log\left((\Delta^i)^{n_i} \left(T + \frac{\sigma^2}{\sigma_\beta^2}\right)\right) - \frac{1}{2\sigma^2} \left[S3 - 2\left(\beta - \frac{\sigma^2}{2}\right) S1 \right. \\
&\quad \left. + \left(\beta - \frac{\sigma^2}{2}\right)^2 MT - \frac{S2 - 2\left(\beta - \frac{\sigma^2}{2}\right) TS1 + \left(\beta - \frac{\sigma^2}{2}\right)^2 MT^2}{\left(T + \frac{\sigma^2}{\sigma_\beta^2}\right)} \right].
\end{aligned} \tag{3.13}$$

Los estimadores máximo verosimiles se encuentran igualando a cero las derivadas de la log-verosimilitud y posteriormente se despeja a los parámetros $\hat{\beta}$, $\hat{\sigma}^2$ y $\hat{\sigma}_\beta^2$.

Derivamos respecto a β ,

$$\begin{aligned}
\frac{d \log(L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}))}{d\beta} &= \frac{1}{\sigma^2} \left[S1 - \left(\beta - \frac{\sigma^2}{2} \right) MT \right. \\
&\quad \left. - \frac{\sigma_\beta^2}{T\sigma_\beta^2 + \sigma^2} \left(TS1 - \left(\beta - \frac{\sigma^2}{2} \right) MT^2 \right) \right] \\
&= \frac{1}{\sigma^2(T\sigma_\beta^2 + \sigma^2)} \left[S1(T\sigma_\beta^2 + \sigma^2 \right. \\
&\quad \left. - T\sigma_\beta^2) - \left(\beta - \frac{\sigma^2}{2} \right) MT(T\sigma_\beta^2 + \sigma^2 - T\sigma_\beta^2) \right] \\
&= \frac{1}{(T\sigma_\beta^2 + \sigma^2)} \left[S1 - \left(\beta - \frac{\sigma^2}{2} \right) MT \right].
\end{aligned}$$

Ahora, la derivada respecto a σ^2 ,

$$\begin{aligned}
\frac{d \log(L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}))}{d\sigma^2} &= -\frac{N-2M}{2\sigma^2} + \frac{M}{2(T\sigma_\beta^2 + \sigma^2)} - \frac{S3}{2\sigma^4} - \frac{\left(\beta - \frac{\sigma^2}{2} \right) S1}{(T\sigma_\beta^2 + \sigma^2)^2} \\
&\quad - \frac{S1}{2(T\sigma_\beta^2 + \sigma^2)} + \frac{\left(\beta - \frac{\sigma^2}{2} \right) MT}{2(T\sigma_\beta^2 + \sigma^2)} + \frac{\left(\beta - \frac{\sigma^2}{2} \right)^2 MT}{2(T\sigma_\beta^2 + \sigma^2)^2} \\
&\quad - \frac{\sigma_\beta^2(T\sigma_\beta^2 + 2\sigma^2)S2}{2\sigma^4(T\sigma_\beta^2 + \sigma^2)^2}.
\end{aligned}$$

Finalmente, derivamos respecto a σ_β^2 ,

$$\begin{aligned}
\frac{d \log(L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}))}{d\sigma_\beta^2} &= -\frac{M}{2\sigma_\beta^2} - \frac{M\sigma^2}{2\sigma_\beta^2(T\sigma_\beta^2 + \sigma^2)} + \frac{S2 - 2\left(\beta - \frac{\sigma^2}{2} \right) TS1}{2(T\sigma_\beta^2 + \sigma^2)^2} \\
&\quad + \frac{\left(\beta - \frac{\sigma^2}{2} \right)^2 MT^2}{2(T\sigma_\beta^2 + \sigma^2)^2}.
\end{aligned}$$

Al resolver el sistema de ecuaciones para encontrar los máximos, los estimadores son los siguientes:

$$\hat{\beta} = \frac{1}{MT} S1 + \frac{\hat{\sigma}^2}{2}. \quad (3.14)$$

$$\hat{\sigma}^2 = \frac{1}{(N-2M)} \left(S3 - \frac{1}{T} S2 \right). \quad (3.15)$$

$$\hat{\sigma}_\beta^2 = \frac{M(N-M)S2 - (N-2M)S1^2 - M^2TS3}{(N-2M)M^2T^2}. \quad (3.16)$$

Notemos que para que los estimadores de σ^2 y σ_β^2 existan necesitamos que las sumas $S1, S2$ y $S3$ cumplan la siguiente desigualdad:

$$\frac{1}{T}S2 < S3 < \frac{N-M}{MT}S2 - \frac{N-2M}{M^2T}S1^2.$$

Calculamos las varianzas asintóticas de los estimadores a partir de la inversa de la información de Fisher evaluadas en el óptimo. Recordemos que la información de Fisher para algún parámetro θ y densidad $f(X; \theta)$ esta dado por

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} \log f(X; \theta)\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(X; \theta)\right].$$

$$\hat{Var}(\hat{\beta}) = \frac{T\hat{\sigma}_\beta^2 + \hat{\sigma}^2}{MT}. \quad (3.17)$$

$$\hat{Var}(\hat{\sigma}^2) = \left[\frac{3(N-2M)}{2\sigma^4} + \frac{2M + MT(T\hat{\sigma}_\beta^2 + \hat{\sigma}^2)}{4(T\hat{\sigma}_\beta^2 + \hat{\sigma}^2)^2} \right]^{-1}. \quad (3.18)$$

$$\hat{Var}(\hat{\sigma}_\beta^2) = \frac{2(T\hat{\sigma}_\beta^2 + \hat{\sigma}^2)^2}{MT^2}. \quad (3.19)$$

3.3.2. Estimación sin efectos aleatorios

Si consideramos que el modelo no tiene efectos aleatorios, la función de log-verosimilitud se simplifica de la siguiente manera:

$$\begin{aligned} \log(L(\beta, \sigma^2 | \mathbf{y})) &= \frac{N-M}{2} \log(2\pi\sigma^2) - \sum_{i=1}^M \frac{n_i}{2} \log(\Delta^i) \\ &\quad - \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{(y_j^i - y_{j-1}^i - (\beta - \frac{1}{2}\sigma^2)\Delta_j^i)^2}{2\sigma^2\Delta_j^i} \end{aligned} \quad (3.20)$$

Cabe mencionar que si escribimos la función de verosimilitud como el producto de las funciones de densidad de transición sin considerar los efectos aleatorios desde el inicio, obtenemos la misma verosimilitud y por tanto la misma log-verosimilitud que al hacer $\beta^i = 0$ en la ecuación (3.11).

Por lo que los estimadores máximo verosímil también se simplifican:

$$\hat{\beta} = \frac{1}{MT}S1 + \frac{\hat{\sigma}^2}{2}. \quad (3.21)$$

$$\hat{\sigma}^2 = \frac{1}{(N-M)} \left(S3 - \frac{1}{MT}S1^2 \right). \quad (3.22)$$

Capítulo 4

Simulación y aplicaciones

4.1. Simulación del movimiento Browniano geométrico

Antes de desarrollar la aplicación, es conveniente verificar que los estimadores que acabamos de encontrar son buenos. Con el fin de estudiar el comportamiento de dichos estimadores para diferentes relaciones de parámetros, en particular de las varianzas, realizamos simulaciones de las observaciones suponiendo que siguen trayectorias de un movimiento Browniano geométrico con el método de Milstein. Utilizamos seis conjuntos de parámetros para dos valores diferentes de β , M y n y analizamos las siguientes relaciones: $\sigma^2 < \sigma_\beta^2$, $\sigma^2 \approx \sigma_\beta^2$ y $\sigma^2 > \sigma_\beta^2$. Notemos que estamos suponiendo que todos los individuos tiene el mismo número de observaciones n . Los valores se muestran en la Tabla 4.1.

	β	σ^2	σ_β^2	M	n
1	-0.02	0.02	0.02	10	50
2	-0.02	0.2	0.02	10	50
3	-0.02	0.02	0.2	10	50
4	-0.02	0.02	0.02	50	10
5	-0.02	0.2	0.02	50	10
6	-0.02	0.02	0.2	50	10
7	-0.2	0.02	0.02	10	50
8	-0.2	0.2	0.02	10	50
9	-0.2	0.02	0.2	10	50
10	-0.2	0.02	0.2	50	10
11	-0.2	0.2	0.02	50	10
12	-0.2	0.02	0.2	50	10

Tabla 4.1: Valores de los parámetros utilizados en las simulaciones.

Para el análisis realizamos 1000 trayectorias para cada individuo por cada conjunto de parámetros, con observaciones equidistantes en el tiempo dependiendo del valor correspondiente de n . Tomamos como valor inicial para cada uno de ellos $\log(100)$ y el intervalo de tiempo en que se realizan las observaciones $T=100$. Los parámetros y las varianzas de cada estimador fueron calculadas a partir de las expresiones de los estimadores encontradas vía máxima verosimilitud expresados en las ecuaciones (3.14), (3.15), (3.16). Consideramos el estimador como la media de los estimadores y los cuantiles empíricos del 2.5 % y 97.5 % como el intervalos de confianza de 95 %; cuyos resultados se muestran en las Tablas 4.2,4.3,4.4 para los parámetros de β , σ^2 y σ_β^2 , respectivamente.

	$\hat{\beta}$		
	$\hat{\beta}$	IC	\hat{Var}
1	-0.0180	(-0.1042;0.0720)	0.0018
2	-0.0185	(-0.1114;0.0668)	0.0020
3	-0.0205	(-0.2937;0.2460)	0.0185
4	-0.0199	(-0.0610;0.0179)	0.0004
5	-0.0195	(-0.0614;0.0253)	0.0004
6	-0.0186	(-0.1356;0.1000)	0.0040
7	-0.2005	(-0.2910;-0.1164)	0.0018
8	-0.2020	(-0.2889;-0.1079)	0.0019
9	-0.2004	(-0.4630;0.0738)	0.0183
10	-0.2011	(-0.2411;-0.1606)	0.0004
11	-0.2007	(-0.2436;-0.1570)	0.0004
12	-0.2014	(-0.3253;-0.0772)	0.0039

Tabla 4.2: Estimadores, intervalos del 95 % de confianza (IC) y varianza.

En general, se obtienen buenos estimadores; aunque, un número considerable de estimadores de β fueron positivos para casi todos los casos, lo cual puede observarse en el intervalo de confianza.

Por otra parte, el estimador de σ^2 además de ser el más exacto, con varianzas asintóticas pequeñas existió en todos los casos, es decir, $\hat{\sigma}^2 > 0$. Por el contrario, aproximadamente el 8 % de los estimadores de la varianza de los efectos aleatorios σ_β^2 fue negativo, en particular cuando $\sigma^2 > \sigma_\beta^2$ y para M grande, lo cual indica la dependencia al tamaño de la muestra; por consiguiente $\hat{\sigma}_\beta^2$ es el más difícil de estimar.

Si tenemos un experimento el cual se modela incluyendo efectos aleatorios y erróneamente lo consideramos sin ellos, es decir, $\beta^i = 0$ para toda i .

Utilizando estos estimadores dados por las ecuaciones (3.21) y (3.22) podemos observar en la Tabla 4.5, que $\hat{\beta}$ y $\hat{\sigma}^2$ son malos. En particular cuando la varianza de los efectos aleatorios y el número de individuos M es grande, ya que el mo-

	$\hat{\sigma}^2$		
	$\hat{\sigma}^2$	IC	$\hat{V}ar$
1	0.01999	(0.01754;0.02253)	2.173e-10
2	0.20000	(0.17610;0.22600)	2.176e-06
3	0.01993	(0.01754;0.02253)	2.147e-10
4	0.02000	(0.01751;0.02275)	2.372e-10
5	0.20000	(0.17640;0.22630)	2.365e-06
6	0.01998	(0.01742;0.02278)	2.362e-10
7	0.02004	(0.01766;0.02252)	2.196e-10
8	0.20040	(0.17650;0.22690)	2.192e-06
9	0.02006	(0.01773;0.02254)	2.203e-10
10	0.02002	(0.01736;0.02280)	2.382e-10
11	0.19930	(0.17550;0.22740)	2.336e-06
12	0.02003	(0.01740;0.02285)	2.387e-10

Tabla 4.3: Estimadores, intervalos del 95 % de confianza (IC) y varianza.

delo está fuertemente influenciado por tales efectos, por consiguiente, suponer que no hay efectos aleatorios produce estimadores deficientes.

Como ya mencionamos, hacer la estimación de los parámetros de difusiones continuas con observaciones a tiempo discreto causa cierto problema y como las principales aplicaciones del modelo de efectos de mezclas en ecuaciones diferenciales estocásticas son en estudios biomédicos o biológicos, usualmente se cuenta con un número pequeño de observaciones. Recientemente, se han estudiado diversos métodos para la estimación máximo verosímil de difusiones a partir de observaciones discretas, ciertos autores proponen completar información entre cada par de observaciones mediante puentes de dicha difusión con el fin de obtener más observaciones y mejores estimadores.

	$\hat{\sigma}_\beta^2$		
	$\hat{\sigma}_\beta^2$	IC	\hat{Var}
1	0.01756	(0.004968;0.03862)	6.306e-05
2	0.01786	(0.003772;0.04053)	7.891e-05
3	0.18460	(0.051750;0.38840)	6.829e-03
4	0.01958	(0.012710;0.02856)	1.565e-05
5	0.01957	(0.011750;0.02897)	1.860e-05
6	0.19750	(0.126800;0.28460)	1.564e-03
7	0.01797	(0.004995;0.03853)	6.604e-05
8	0.01748	(0.003816;0.03801)	7.592e-05
9	0.18330	(0.049300;0.38810)	6.731e-03
10	0.01981	(0.012920;0.02850)	1.601e-05
11	0.01962	(0.012360;0.02848)	1.869e-05
12	0.19590	(0.128200;0.27850)	1.539e-03

Tabla 4.4: Estimadores, intervalos del 95 % de confianza (IC) y varianza.

Estimadores sin efectos aleatorios		
	$\hat{\beta}$	$\hat{\sigma}_\beta^2$
1	-0.00202	0.0570
2	-0.00283	0.2360
3	0.16200	0.3770
4	0.07820	0.2170
5	0.07620	0.3930
6	0.95200	1.9600
7	-0.18200	0.0558
8	-0.18400	0.2350
9	-0.02600	0.3770
10	-0.10200	0.2150
11	-0.10300	0.3940
12	0.78300	1.9900

Tabla 4.5: Estimadores para el modelo sin efectos aleatorios.

4.2. Simulación con puentes de difusiones

En esta sección, tomamos el caso del movimiento browniano geométrico ya estudiado y vemos qué sucede con los estimadores al incluir las observaciones obtenidas mediante la simulación de puentes con el método descrito en el Capítulo 1.

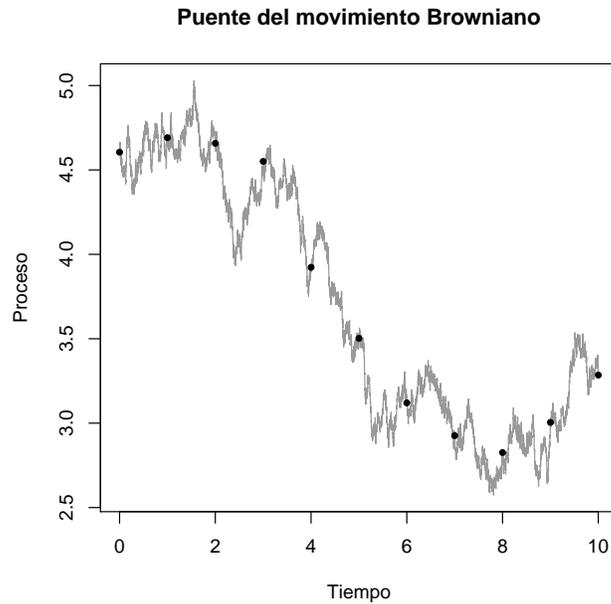


Figura 4.1: Puente del movimiento Browniano.

En primer lugar, realizamos el estudio para una ecuación diferencial estocástica dada por la ecuación (3.3), sin agregar efectos aleatorios. Consideremos que el proceso tiene valor inicial $\log(100)$ y $\Delta = 1$. Tomamos un par de parámetros (los mismos que utilizamos anteriormente) $\beta = -0.02$, $\sigma^2 = 0.2$ y número de individuos $M=10$ y 50 , suponemos que todos los individuos tienen el mismo número de observaciones n .

Calculamos los estimadores dados por las ecuaciones (3.21) y (3.22) con diferentes n para verificar la sensibilidad del estimador al número de observaciones, podemos observar en la Figura 4.2 que mientras n es más grande el estimador converge al valor del parámetro original. Para $M=10$ esto ocurre con n entre 1500 y 2000 aproximadamente; mientras que con $M=50$ se pega al valor original con más rapidez. Sin embargo, para ambos casos, podríamos decir que con muestras no tan grandes el estimador no está tan alejado del valor real.

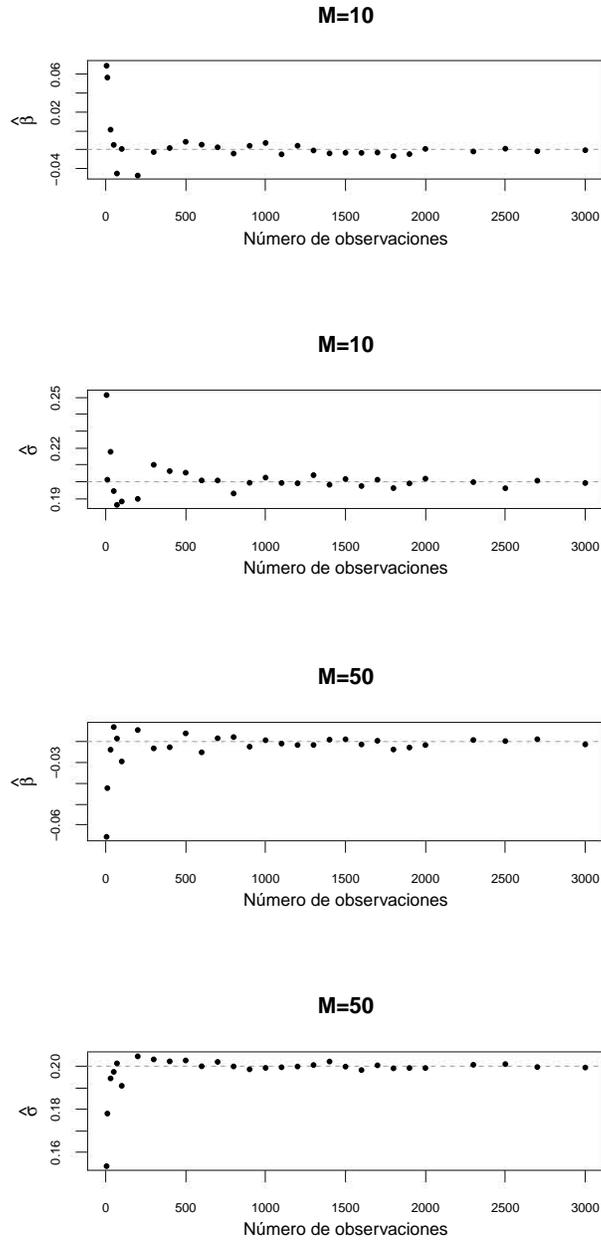


Figura 4.2: Estimadores vs. número de observaciones.

Ahora, consideremos que tenemos un número reducido de observaciones, es decir, $n=10$ a partir de dichos puntos generamos los puentes con el algoritmo

descrito en el Teorema 1.3.1; en la Figura 4.1 podemos observar una realización del puente para un individuo, donde los puntos en negro muestran las observaciones que ya se tenían y el resto (en gris) muestra la trayectoria simulada.

Entonces simulamos observaciones con los parámetros en cuestión, generamos puentes entre cada par de puntos y la idea es verificar que efectivamente recuperamos los valores de β y σ^2 . Para hacer la estimación de esta forma, es importante obtener una buena muestra con las simulaciones ya que la estimación con puentes es muy sensible a estos cambios.

Para notar esto, generamos trayectorias con $M=10$ y $n=10, 100$ y 1000 ; de tal manera que en los tres casos obtengamos $10000, 15000$ y 20000 observaciones, entonces si N número de discretizaciones del puente la Tabla 4.6 muestra los resultados.

Podemos observar que en los tres casos y para las diferentes discretizaciones el valor de $\hat{\sigma}^2$ no varió mucho y está muy cerca de 0.2 como queríamos, incluso, al incluir las observaciones del puente el estimador mejoró. Ya que los puentes están fuertemente influenciados por las observaciones, el estimador de β es un poco más complicado. Si tomamos el modelo sin puentes, con $n=10$, $\hat{\beta}$ no es bueno y podemos observar que los estimadores obtenidos al agregar los puentes se acercan a ese valor de $\hat{\beta}$, por lo que también son malos. Sin embargo, con $n=100$ y $n=1000$ mejora mucho $\hat{\beta}$ y el parámetro adquiere mayor estabilidad, por lo tanto, incluir los puentes parece ser buena idea.

n	N	$\hat{\beta}$	$\hat{\sigma}^2$
10	0	-0.0494	0.1859
10	1000	-0.0427	0.1992
10	1500	-0.0423	0.2000
10	2000	-0.0424	0.1999
100	0	-0.0191	0.2119
100	100	-0.0219	0.1988
100	150	-0.0183	0.2127
100	200	-0.0206	0.2080
1000	0	-0.0194	0.1975
1000	10	-0.0215	0.1932
1000	15	-0.0201	0.1961
1000	20	-0.0194	0.1974

Tabla 4.6: Estimadores del modelo sin efectos aleatorios con puentes.

Ya vimos que el uso de puentes de difusiones se comporta bien para resolver el problema de información faltante para ecuaciones diferenciales estocásticas, sin efectos aleatorios. También ya vimos como se comportan los estimadores en el

modelo con efectos aleatorios, entonces falta hacer el análisis agregando dichos efectos.

Seguimos con los mismos valores de β y σ^2 e incluimos $\sigma_\beta^2 = 0.02$. En la Figura 4.3 se muestra una trayectoria de un puente Browniano con efectos aleatorios con $n=5$ y $N=100$, nuevamente los puntos en negro son las observaciones y la trayectoria en gris es la simulación del puente.

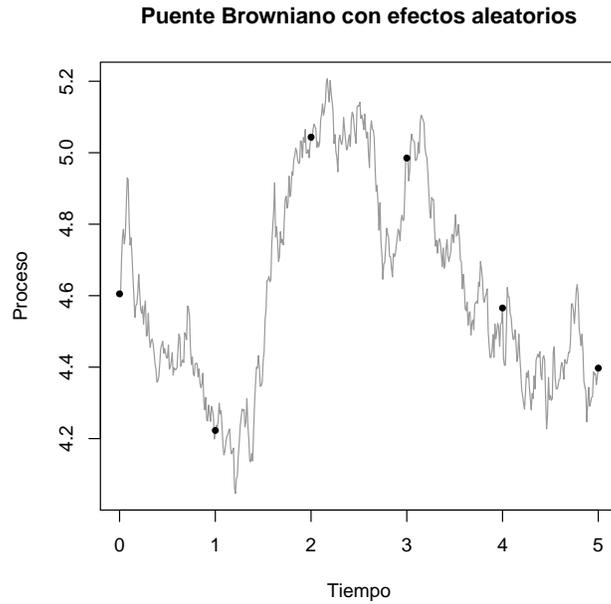


Figura 4.3: Puente del movimiento Browniano con efectos aleatorios.

Supongamos que tenemos $M=10$ y $n=100$ y hacemos 100 simulaciones para cada $N=10, 100$ y 1000 del modelo con efectos aleatorios con los mismo parámetros. En la Tabla 4.7 se muestran los valores de los estimadores y sus respectivas varianzas. Notemos que para este conjunto de parámetros habíamos obtenido en las Tablas 4.2, 4.3 y 4.4 $\hat{\beta} = -0.0185$, $\hat{\sigma}^2 = 0.2000$ y $\hat{\sigma}_\beta^2 = 0.01786$.

N	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_\beta^2$	$Var(\hat{\beta})$	$Var(\hat{\sigma}^2)$	$Var(\hat{\sigma}_\beta^2)$
10	-0.0248	0.1968	0.0214	$1.063e - 03$	$9.515e - 08$	$2.260e - 05$
100	-0.0233	0.1991	0.0213	$1.063e - 03$	$1.050e - 08$	$2.260e - 05$
1000	-0.0220	0.1999	0.0213	$1.754e - 03$	$2.124e - 10$	$6.158e - 05$

Tabla 4.7: Estimadores del modelo con efectos aleatorios con puentes.

Podemos notar que el estimador de β está sobre estimado, sin embargo, cuando incrementamos el número de puntos en el puente la estimación mejora considerablemente, para el caso de σ^2 como ya habíamos mencionado, sigue siendo el valor más exacto incluso con muestras no tan grandes, mientras que para σ_β^2 obtuvimos cambios favorables en el estimador, ya que al considerar los puentes, se acerca más al valor real. Las varianzas permanecen muy pequeñas y similares a las anteriores a pesar de haber agregado las simulaciones.

Con este estudio pudimos observar que generar puentes entre cada par de observaciones resulta un poco más costoso computacionalmente, sin embargo, para obtener los estimadores es de gran ayuda. Pero en este caso ya que solo es de verificación únicamente buscamos recuperar los valores de los parámetros que ya conocíamos.

En la práctica, lo que conocemos realmente son resultados del experimento, entonces, lo primero que tendríamos que suponer es que se distribuyen como un movimiento Browniano geométrico, además de independiencia entre individuos o unidades experimentales. Luego, la estimación se hace mediante algún algoritmo, por ejemplo, el EM (esperanza maximización).

El EM es un algoritmo iterativo que consta de dos pasos, primero se deben dar valores iniciales a los tres parámetros; en el paso E se simulan los puentes con los valores de los parámetros iniciales para completar la información faltante, posteriormente, en el paso M se maximiza la verosimilitud, es decir, se calculan los estimadores con toda la información y se regresa al paso E utilizando los valores recién obtenidos y así hasta la convergencia.

4.3. Aplicación

Ahora, aplicamos el modelo a un conjunto de datos que representan la disolución de Metropolol Tartrate que es un medicamento utilizado para tratar la presión arterial alta (hipertensión); lo cual ayuda a prevenir derrames cerebrales, ataques cardíacos y problemas renales. Este medicamento también se utiliza para tratar el dolor de pecho (angina) y para mejorar la supervivencia después de un ataque al corazón.

Los datos fueron obtenidos de [18], donde se examinan con distintos métodos basados en ANOVA, ratio test, entre otros. Se analizaron cuatro tipos de pastillas de 100 mg, cada experimento se realizó seis veces, con intervalos de tiempo de 5 minutos hasta 45 minutos.

Utilizamos únicamente los datos de lenta disolución (véase Tabla 4.8) para ejemplificar el modelo y encontrar los estimadores correspondientes. Consideremos que el porcentaje aún no disuelto de Metropolol se modela con la ecuación (3.7), donde y_j^i es el logaritmo de la observación de la repetición i al tiempo j . En la Figura 4.4 observamos el porcentaje no disuelto para cada experimento. Notemos

5min	10min	15min	20min	25min	30min	45min
14.6	23.7	34.7	44.2	51.9	58.1	73.4
11.5	21.1	31.3	38.4	45.1	51.0	65.8
12.3	21.4	30.3	38.5	45.7	52.7	67.6
12.9	23.1	33.0	40.4	48.1	53.8	69.6
12.4	21.0	30.9	40.9	47.4	53.2	71.3
13.4	23.4	32.3	39.7	46.1	52.2	66.5

Tabla 4.8: Porcentaje no disuelto de Metoprolol.

que $M = 6$ y $N = 42$.

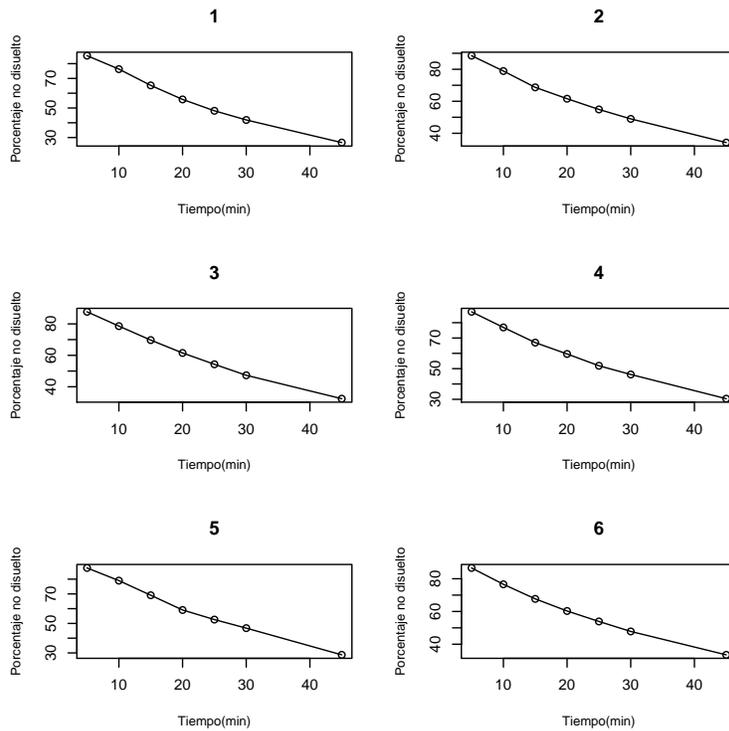


Figura 4.4: Porcentaje de Metoprolol no disuelto.

Utilizando las ecuaciones (3.14), (3.15) y (3.16), encontramos el valor de los estimadores que se muestran en la Tabla 4.9 así como sus respectivas varianzas y los estimadores del modelo si asumimos que no hay efectos aleatorios.

Como la varianza de los efectos aleatorios es muy pequeña y comparada con σ^2 sigue siendo pequeña, suponiendo que no hay efectos aleatorios los estimadores

	Con efectos aleatorios		Sin efectos aleatorios
	Estimador	Desviación estándar	Estimador
$\hat{\beta}$	-0.02593	8.3283e-04	-0.02592
$\hat{\sigma}^2$	5.0936e-05	3.8676e-10	7.0191e-05
$\hat{\sigma}_\beta^2$	2.8882e-06	2.4027e-06	-

Tabla 4.9: Estimadores de Metropool.

no son tan malos como en el caso de las simulaciones.

	$\hat{\beta}^i$
1	-0.00222
2	0.00152
3	0.00074
4	-0.00024
5	-0.00134
6	0.00154

Tabla 4.10: *Estimadores de los efectos aleatorios.*

Además, a partir de la ecuación (3.3.1), obtenemos los estimadores de los efectos aleatorios que se muestran en la Tabla 4.10 para cada repetición i del experimento. Para verificar que dichos estimadores cumplen el supuesto de normalidad, hacemos la gráfica de los cuantiles qqnorm; en la Figura 4.5 se puede observar a pesar de ser pocos datos efectivamente la muestra pertenece a una distribución normal.

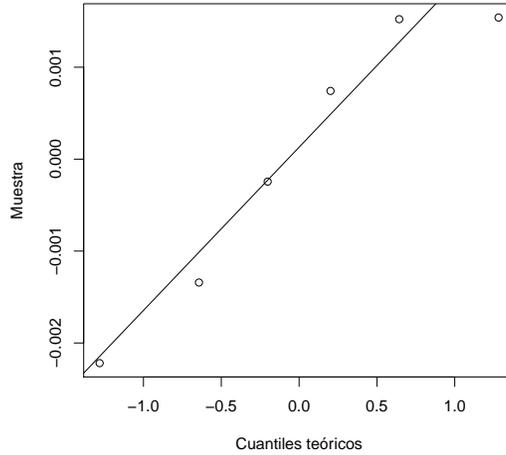


Figura 4.5: Gráfica de los cuantiles de una normal vs. los estimadores de β_i .

Finalmente, en la Figura 4.6 se muestran los datos observados comparados con las trayectorias simuladas a partir de los estimadores de los parámetros y de los efectos aleatorios encontrados para cada repetición del experimento donde podemos observar que se ajustan adecuadamente.

Para modelar la tendencia decreciente de los datos, es decir, que a medida que se incrementa el tiempo el porcentaje no disuelto de Metoprolol disminuya, según la ecuación (3.7) los parámetros deben cumplir la desigualdad $(\hat{\beta} + \hat{\beta}^i - \hat{\sigma}^2/2) < 0$ para toda i .

Notemos que el coeficiente de difusión es muy pequeño comparado con $\hat{\beta}$ y que además los efectos aleatorios son normales con media cero y desviación estándar estimada también muy pequeña, por lo tanto, con $\hat{\beta}$ negativa podemos asegurar que se cumpla dicha desigualdad, por lo tanto el comportamiento del modelo deseado. Recordemos también que los efectos aleatorios $\hat{\beta}^i$ nos dan las variaciones entre las seis veces que se realizó el experimento. Entonces, podemos modelar el porcentaje de Metoprolol no disuelto mediante un movimiento Browniano geométrico con efectos aleatorios.

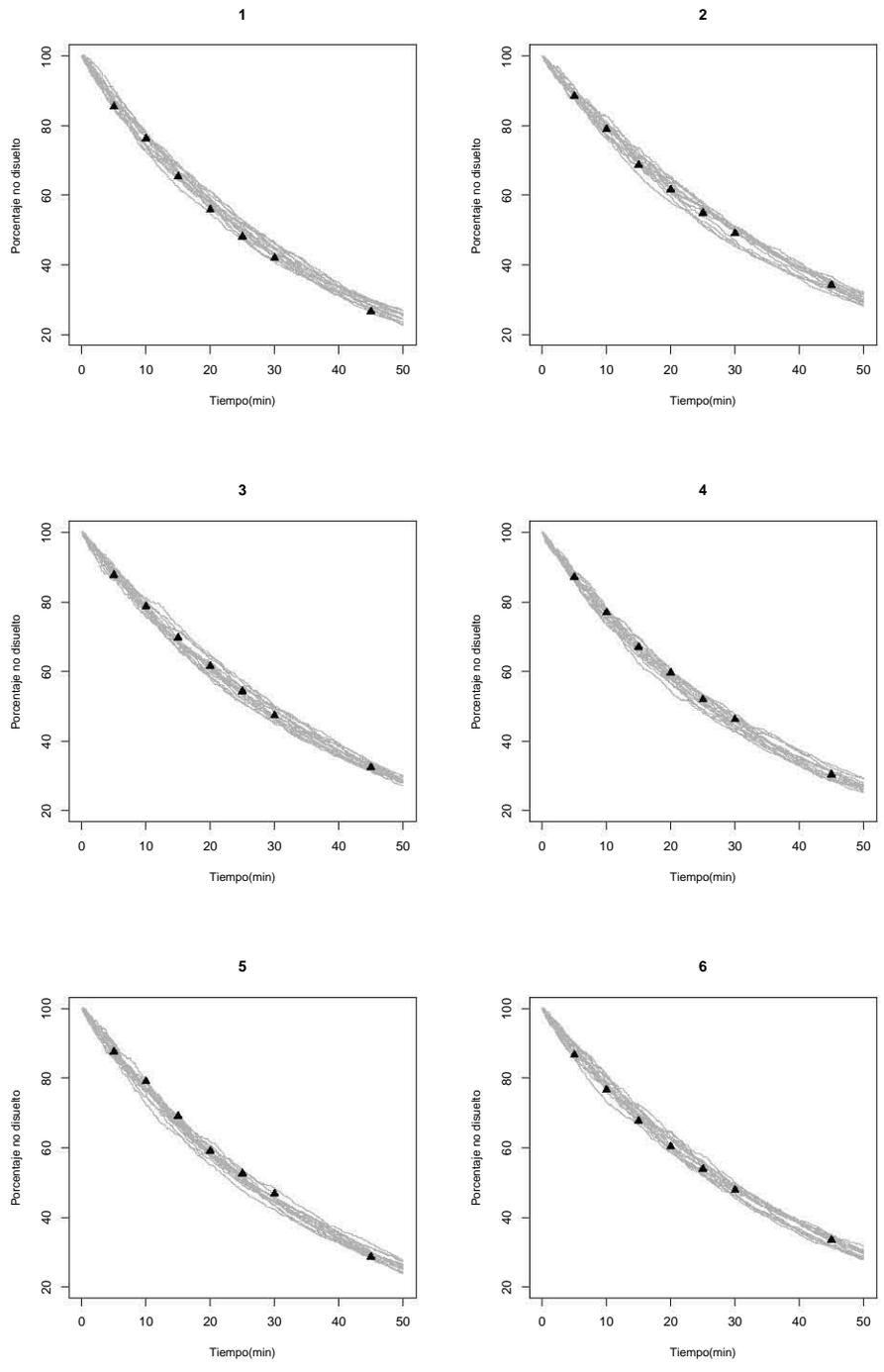


Figura 4.6: Simulación de trayectorias.

Conclusiones

Sin lugar a dudas el modelo de efectos de mezclas en ecuaciones diferenciales estocásticas es muy útil para describir mejor cierto tipo de experimentos, por ejemplo, los estudios biomédicos; ya que modela tanto las variaciones que existen entre los individuos como las variaciones de cada individuo durante el estudio.

Pudimos observar que encontrar los estimadores de manera explícita en el caso del movimiento Browniano geométrico no resulta difícil después de haber realizado ciertos cálculos. También pudimos notar que el desarrollo para encontrar dichos estimadores está basado en la estimación vía verosimilitud tanto de las ecuaciones diferenciales estocásticas como del modelo de efectos de mezclas.

En el estudio de las simulaciones, en general, los valores estimados de los parámetros β , σ^2 y σ_β^2 fueron muy cercanos a los valores reales para los diferentes conjuntos de parámetros. Además, hubo una clara relación de dependencia entre el estimador de la varianza de los efectos aleatorios y el tamaño de la muestra. Asimismo, pudimos observar que si un experimento se modela mediante un modelo de efectos de mezclas en ecuaciones diferenciales estocásticas y al hacer la estimación no se consideran los efectos aleatorios, los valores de β , σ^2 , no representan el comportamiento del experimento.

Desafortunadamente, el modelo es restrictivo en cuanto a las condiciones que se deben cumplir para que existan los estimadores de la varianza de los efectos aleatorios y el coeficiente de difusión. Por otra parte, como ya se ha mencionado la estimación máxima verosímil no funciona para cualquier modelo, ni siquiera en difusiones con funciones de densidad condicional conocidas como por ejemplo, el Ornstein-Uhlenbeck, del cual puede escribirse la función de verosimilitud pero no existen los estimadores explícitos.

Por tal motivo el estudio del modelo es muy amplio, por un lado podría buscarse generalizar el modelo con el movimiento Browniano geométrico donde los efectos aleatorios no sean idénticamente distribuidos y por lo tanto, los individuos tengan varianzas distintas.

Por otro lado, ya algunos autores han desarrollado otros métodos para poder utilizar otras difusiones, considerando aproximaciones a la verosimilitud, a partir

de funciones hermitianas y reglas de cuadratura gaussiana, ([16]) y para modelos no lineales, basados en el algoritmo EM ([6]), y combinando la estimación condicional de primer orden con el filtro de Kalman ([14]).

Apéndice A

Descomposición ortogonal-triangular

La descomposición ortogonal-triangular o QR es un método numérico, usado normalmente para resolver problemas de mínimos cuadrados.

Sea \mathbf{X} una matriz de $n \times m$ con $n \geq m$ de rango m , entonces existe una matriz ortogonal \mathbf{Q} , cuadrada de $n \times n$ y una matriz triangular superior, \mathbf{R} de $p \times p$ únicas tal que:

$$\mathbf{X} = \mathbf{Q} \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_t \mathbf{R},$$

donde \mathbf{Q}_t (\mathbf{Q} - *truncada*) son las primeras p columnas de \mathbf{Q} . Si $m = n$, la matriz \mathbf{Q} es unitaria, si \mathbf{X} es no singular, las entradas de la diagonal de \mathbf{R} pueden elegirse positivas y además si \mathbf{X} es real, \mathbf{Q} y \mathbf{R} también lo son.

Como \mathbf{Q} es ortogonal, entonces $\mathbf{Q}\mathbf{Q}^t = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, por lo cual, $\mathbf{Q}_t^T\mathbf{Q}_t = \mathbf{I}$. Además, recordemos que cualquier matriz ortogonal preserva la norma de los vectores bajo multiplicación, es decir,

$$\|\mathbf{Q}^T \mathbf{y}\|^2 = (\mathbf{Q}^T \mathbf{y})^T \mathbf{Q}^T \mathbf{y} = \mathbf{y}^T \mathbf{Q} \mathbf{Q}^T \mathbf{y} = \mathbf{y}^T \mathbf{y} = \|\mathbf{y}\|^2.$$

El método que más se utiliza para encontrar dichas matrices es el de Gram-Schmidt.

Apéndice B

Códigos

A continuación, presentamos los códigos realizados en R-3.1.0 utilizados para las simulaciones y la aplicación.

Para la simulación del movimiento Browniano geométrico y el puentes de dicha difusión, utilizamos el paquete `sde`, específicamente las funciones `GBM` y `DBridge`.

La función `sim` genera la matriz de respuestas para los M individuos del experimento, cada uno con n observaciones generadas a partir del movimiento Browniano geométrico. Recibe como argumentos M , n , los parámetros del modelo $\beta = \text{beta}$, $\sigma^2 = \text{sigma}$ y $\sigma_\beta^2 = \text{sigmab}$ para el caso donde haya efectos aleatorios y `efectos` que es un argumento lógico; por default genera las trayectorias sin efectos aleatorios. Devuelve una lista que contiene la matriz de respuestas transformadas Y , el vector de tiempos T y en su caso el vector de efectos aleatorios `betai`.

```
sim<-function(M,n,beta,sigma,sigmab,efectos=FALSE){
  T<-numeric
  Y<-matrix(0,M,(n+1))
  i<-0
  betai<-numeric()
  T<-seq(from=0, to=n,length=n+1)

  if(efectos==FALSE){
    while(i<M){
      i<-i+1
      aux<-GBM(x=100,r=beta,sigma=sqrt(sigma),N=n,T=n)
      Y[i,<]<-log(aux)
    }
    resp<-list(Y=Y,T=T)
  }
}
```

```

else{
  betai<-rnorm(M,0,sqrt(sigtab))
  while(i<M){
    i<-i+1
    mu<-beta+betai[i]
    aux<-GBM(x=100,r=mu,sigma=sqrt(sigma),N=n,T=n)
    Y[i,]<-log(aux)
  }
  resp<-list(Y=Y,T=T,betai=betai)
}

return(resp)
}

```

La función `Tray.Comp` genera puentes entre cada par de observaciones de un individuo; nuevamente por default lo hace sin efectos aleatorios. Recibe como argumentos un vector de observaciones `Obs`, de tiempo `Tiempo`, `N` que indica el número puntos a simular entre cada par de observaciones y los parámetros provenientes de la difusión. Regresa una lista con el vector de la trayectoria que incluye las observaciones dadas y los puentes generados `Proceso` y el nuevo vector de tiempos.

```

Tray.Comp<-function(Obs,Tiempo,N,beta,sigma,betai,efectos=FALSE){
  Z<-numeric()
  sd<-sqrt(sigma)
  aux<-numeric()
  auxT<-numeric()
  i<-1
  if(efectos==FALSE){
    mu<-as.expression(beta)
    t<-expression(t)
    d<-eval(substitute(expression(mu*t),
      list(mu=mu[[1]],t=t[[1]])))
    s <- as.expression(sd)
    while(i <length(Obs)){
      aux<-DBridge(x=Obs[i], y=Obs[i+1], t0=0, T=1,
        drift=d,sigma=s,N=N)
      auxT<-seq(from=Tiempo[i],to=Tiempo[i+1],length=N+1)
      if(i==1){
        Z<-aux
        T<-auxT
      }
      else{
        Z<-c(Z,aux[2:(N+1)])
        T<-c(T,auxT[2:(N+1)])
      }
      i<-i+1
    }
  }
}

```

```

    }
    else{
      s <- as.expression(sd)
      while(i <length(Obs)){
        mu<-beta+betai-1/2*sigma
        mu<-as.expression(mu)
        t<-expression(t)
        d <- eval(substitute(expression(mu*t),
          list(mu=mu[[1]],t=t[[1]])))
        aux<-DBridge(x=Obs[i], y=Obs[i+1], t0=0, T=1,
          drift=d, sigma=s,N=N)
        auxT<-seq(from=Tiempo[i],to=Tiempo[i+1],length=N+1)
        if(i==1){
          Z<-aux
          T<-auxT
        }
        else{
          Z<-c(Z,aux[2:(N+1)])
          T<-c(T,auxT[2:(N+1)])
        }
        i<-i+1
      }
    }

    return(list(Proceso=Z,Tiempo=T))
  }
}

```

La función `puentes` utiliza la función anterior para generar las trayectorias con puentes de difusiones de los M individuos. Recibe como argumentos D que es una lista, con una matriz denotada por Y donde cada renglón contiene las observaciones de cada individuo y el vector de tiempos T (puede ser el output de `sim`) y el resto de los argumentos son los ya conocidos.

```

puentes<-function(D,N,beta,sigma,betai,efectos=FALSE){
  M<-dim(D$Y)[1]
  i<-1
  T<-Tray.Comp(D$Y[1,],D$T,N,beta,sigma,betai[i],efectos)$Tiempo
  aux<-matrix(0,nrow=M,ncol=length(T))
  while(i<=M){
    aux[i,]<-Tray.Comp(D$Y[i,], D$T, N, beta, sigma,
      betai[i],efectos)$Proceso
    i<-i+1
  }
  return(list(Y=aux,T=T))
}

```

Ya que presentamos las funciones necesarias para simular las observaciones, nos interesan ahora las funciones para calcular las estimaciones y sus varianzas. El

primer código es auxiliar para calcular los estimadores, tiene como argumento una matriz Y de observaciones y el vector T de tiempos, regresa un vector de tres entradas, cada una con las sumas definidas en las ecuaciones (3.3.1), donde la i-ésima entrada representa la suma Si.

```
Sumas<-function(Y,T){
  aux<-numeric(3)
  tam<-dim(Y)
  for (i in 1:tam[1]){
    aux[1]<-aux[1]+(Y[i,tam[2]]-Y[i,1])
    aux[2]<-aux[2]+(Y[i,tam[2]]-Y[i,1])^2
    delta<-T[2]-T[1]
    for (j in 2:tam[2]){
      rest<-(Y[i,j]-Y[i,j-1])^2
      prod<-(1/delta)*rest
      aux[3]<-aux[3]+prod
    }
  }
  return(aux)
}
```

Las siguientes dos funciones, calculan los estimadores dados en por las ecuaciones (3.21) y (3.22) cuando no hay efectos aleatorios y por las ecuaciones (3.14), (3.15) y (3.16). Ambas reciben como argumento una lista D con la matriz de observaciones Y y el vector de tiempos T. Regresan un vector de dos y tres entradas respectivamente. La primer entrada corresponde a $\hat{\beta}$, seguido de $\hat{\sigma}^2$ y $\hat{\sigma}_{\beta}^2$.

```
estimadores.sinE<-function(D){
  param<-numeric(2)
  M<-dim(D$Y)[1]
  N<-M*dim(D$Y)[2]
  T<-D$T[length(D$T)]-D$T[1]
  S<-Sumas(D$Y,D$T)
  param[2]<-(S[3]-(S[1]^2)/(M*T))/(N-M)
  param[1]<-1/(M*T)*S[1]+param[2]/2
  return(param)
}
```

```
estimadores<-function(D){
  est<-numeric(3)
  M<-dim(D$Y)[1]
  N<-M*dim(D$Y)[2]
  S<-Sumas(D$Y,D$T)
  T<-D$T[length(D$T)]-D$T[1]
  est[2]<-(S[3]*T-S[2])/(T*(N-2*M))
  est[1]<-1/(M*T)*S[1]+est[2]/2
}
```

```

    est[3]<-(S[2]*M*(M-N)+S[3]*M^2*T+S[1]^2*(N-2*M))/(M^2*T^2*(2*M-N))
    return(est)
}

```

Y por último las varianzas de los estimadores con efectos aleatorios, además de tener como argumento a D, recibe los valores de $\hat{\sigma}^2$ y $\hat{\sigma}_\beta^2$ obtenidos con la función `estimadores`.

```

varianza<-function(D,sigma,sigtab){
  M<-dim(D$Y)[1]
  N<-M*dim(D$Y)[2]
  T<-D$T[length(D$T)]-D$T[1]
  vari<-numeric(3)
  vari[1]<-(T*sigtab+sigma)/(M*T)
  vari[2]<-(3*(N-2*M)/(2*sigma^4)
    +(M*(2+T^2*sigtab+T*sigma))/(4*(T*sigtab+sigma)^2))^(-1)
  vari[3]<-(2*(T*sigtab+sigma)^2)/(M*T^2)
  return(vari)
}

```

Los valores obtenidos en el Capítulo 4 fueron obtenidos a partir de la aplicación de las funciones anteriores, ya sea para los diferentes conjuntos de parámetros o para los datos de Metropolit.

Bibliografía

- [1] Ait-Sahalia, Y. (2002) *Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach*. *Econometrica*, 70, 223262.
- [2] Beskos, A., Papaspiliopoulos, O., Roberts, G.O. (2004) *Retrospective exact simulation of diffusion sample paths with applications*, Working paper of Lancaster University.
- [3] Desmond J. Higham. (2001). *An algorithmic introduction to numerical simulation of stochastic differential equations*. *SIAM REVIEW*, Vol. 43, No. 3, 525546.
- [4] Bladt M., Srensen, M. (2014) *Simple simulation of diffusion bridges with application to likelihood inference for diffusions*. *Bernoulli* 20(2), 2014, 645675. MOGENS BLADT1 and MICHAEL SRENSEN2
- [5] Ditlevsen S. y De Gaetano A.(2005) *Mixed effects models in stochastic differential equation models*. *REVSTAT* Vol 3: 137-153.
- [6] Donnet S. y Samson A. (2007). *Parametric inference for mixed models defined by stochastic differential equations*. *ESAIM: Probability and Statistics*.
- [7] Durham G. y Gallant R. (2003). *Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes*. *Journal of Business and Economic Statistics*, Vol.3: 297-316.
- [8] Elerian, O. (1998). *A note on the existence of a closed form conditional density for the Milstein scheme*. Working Paper, Nuel College, Oxford University.
- [9] Grabinsky G. (2013). *Teoría de la medida*. Las prensas de ciencias, Facultad de ciencias, UNAM, primera edición.
- [10] Iacus Stefano M. (2008). *Simulation and inference for stochastic differential equations*. Springer.
- [11] Koralov Leonid B., Sinai Yakov G. (2007). *Theory of probability and random processes*. Springer-Verlag, second edition.

-
- [12] Kessler, M., Sørensen, M. (1999). *Estimating equations based on eigenfunctions for a discretely observed diffusion process*, Bernoulli, 5, 299-314.
- [13] Øksendal B. (1998). *Stochastic differential equations: and introduction with applications*. Springer-Verlag, fifth edition.
- [14] Overgaard, R., Jonsson, N., Tornøe, C., Madsen, H. (2005). *Nonlinear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm*. J. Pharmacokinet. Pharmacodyn. 32, 851-87.
- [15] Ozaki, T. (1993). *A local linearization approach to nonlinear filtering*, Int. J. Control, 57, 759-6.
- [16] Picchini U., De Gaetano A. y Ditlevsen S. (2010). *Stochastic differential mixed-effects models*. Scandinavian Journal of Statistics, Vol. 37: 67-90.
- [17] Pinheiro, J. & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*, Springer-Verlag, New York.
- [18] Polli, J.E.; Singh Rekhi, G.; Augsburger, L.L. y Shah, V.P. (1997). *Methods to compare dissolution profiles and a rationale for wide dissolution specifications for metoprolol tartrate tablets*. J. Pharmaceutical Sciences, 86, 6, 690-700.
- [19] Rincón L. *Introducción a los procesos estocásticos*. Las prensas de ciencias, Facultad de ciencias, UNAM.
- [20] Sørensen. (1997). *Estimating Functions for Discretely Observed Diffusions: A Review*. Lecture Notes-Monograph Series, Vol. 32, Selected Proceedings of the Symposium on Estimating Functions, pp. 305-325.
- [21] Steele J. M. (2001). *Stochastic calculus and financial applications*. Springer-Verlag.
- [22] Uribe Bravo G. *Curso avanzado de probabilidad, 2013-I. Integración estocástica*. Instituto de Matemáticas, UNAM.