



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE
LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

Comparación de métodos de simulación e implementación del algoritmo de
Esperanza-Maximización en la difusión de Wright-Fisher

T E S I S

QUE PARA OPTAR POR EL GRADO
DE: MAESTRA EN CIENCIAS

PRESENTA:

GABRIELA STEPHANIA REVUELTAS HERNÁNDEZ

DIRECTOR DE TESIS:

DR. FERNANDO BALTAZAR LARIOS

FACULTAD DE CIENCIAS

CIUDAD UNIVERSITARIA, CD. MX., ENERO 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

A la Universidad Nacional Autónoma de México, por permitirme formar parte de esta institución. Así como al Programa de Posgrado en Ciencias Matemáticas y al CONACYT por el apoyo económico otorgado durante la realización mis estudios.

A Fernando, por la confianza y todo el apoyo que me ha dado a pesar de los años que han pasado, el tiempo, los conocimientos y el interés en que termine satisfactoriamente este trabajo.

A cada uno de los sinodales que se tomaron el tiempo de revisar mi trabajo, a Arnaud y Francisco por sus observaciones, a Ramsés por sus aportaciones que suelen ir más allá del tema y finalmente, a Sergio por su interés en que concluya esta etapa.

A todos los compañeros que estuvieron conmigo en algún momento de la maestría con quienes compartí clases, horas de estudio, reflexiones, dudas y conocimientos.

A mi familia por todo el cariño y el apoyo, a mi mamá por enseñarme que todo lo que se incia se tiene que terminar y nunca dudar de mi, a mi papá por siempre confiar en mis decisiones, a mi abue, por su interés en lo que hago aunque no entienda mucho, a mi hermano por siempre estar conmigo y a Kuri por tu forma tan peculiar de hacerme ir hacia adelante y alcanzar mis objetivos,

Índice general

Introducción	IV
1. Procesos de difusión	1
1.1. Algunos conceptos básicos	1
1.2. Ecuaciones diferenciales estocásticas	3
1.2.1. Fórmula de Itô	4
1.2.2. Transformación de Lamperti	5
1.3. Difusiones	6
1.3.1. Difusiones a partir de ecuaciones diferenciales estocásticas	6
1.3.2. Ejemplos de difusiones	7
1.3.3. Difusiones como procesos de Markov	9
1.3.4. Propiedades de las difusiones	9
1.3.5. Generador infinitesimal y algunos funcionales asociados	11
1.3.6. Ecuaciones de Kolmogorov	19
1.3.7. Distribución estacionaria	22
1.3.8. Clasificación de las fronteras de un proceso regular	23
1.3.9. Descomposición espectral de la función de densidad de transición	28
2. Proceso de Wright Fisher	31
2.1. Proceso de Wright Fisher a tiempo discreto	31
2.2. Difusión de Wright Fisher	33
2.2.1. Aproximación a la difusión	34
2.2.2. Ecuación diferencial estocástica	36
2.2.3. Características del proceso	37
2.2.4. Clasificación de las fronteras	41
2.3. Reparametrización	43
2.4. Descomposición espectral de la función de densidad	44
2.4.1. Polinomios de Jacobi	44
2.4.2. Descomposición espectral	45
2.5. Coalescente de Kingman	46
3. Simulación de trayectorias de la difusión de Wright-Fisher	50
3.1. Métodos numéricos	50
3.1.1. Aproximación de Euler	51
3.1.2. Aproximación de Milstein	56
3.1.3. Método BISS	59
3.2. Algoritmo exacto	66
3.2.1. Método de series alternadas aplicado al proceso ancestral del coalescente de Kingman	67
3.2.2. Algoritmo exacto	70
3.3. Construcción vía procesos de Markov	76
3.3.1. Modelo Gamma-Poisson	77
3.3.2. Modelo Beta-Binomial	78
3.4. Comparación gráfica de los algoritmos	85
3.5. Aplicación de las simulaciones	90

3.6. Simulación de puentes de la difusión de Wight-Fisher.	91
3.6.1. Aproximación de puentes de difusiones.	91
4. Estimación	96
4.1. Algoritmo Esperanza-Maximización	96
4.2. Aplicación del algoritmo EM al proceso de Wright-Fisher	98
4.3. Implementación del algoritmo EM	101
Conclusiones	106
Apendice	107
Bibliografía	120

Introducción

Actualmente existe una gran variedad de fenómenos en diversas disciplinas que se estudian desde el punto de vista probabilístico. Los modelos enfocados en procesos estocásticos resultan útiles para describir el comportamiento de un sistema cuyas características lo vuelven complejo. Por ejemplo, en el campo de la biología, la evolución del tamaño de la población a través del tiempo se modela mediante el proceso estocástico llamado proceso de Wright-Fisher, el cual es el tema central de este trabajo.

La evolución del tamaño de una población a través del tiempo está afectada por una gran cantidad de factores o fuerzas evolutivas como la deriva genética, selección, mutación y migración. A lo largo del trabajo vamos a considerar solo el efecto de la mutación. La difusión de Wright-Fisher puede construirse como un proceso límite del proceso de Wright-Fisher que es una cadena de Markov a tiempo discreto.

El objetivo del trabajo es encontrar un mecanismo adecuado para simular este tipo de procesos y un método de estimación preciso. Uno de los problemas que se presenta es que el proceso tiene espacio de estados acotado, específicamente el intervalo $[0,1]$, lo cual complica la generación de trayectorias y por tanto, requiere ajustes en los algoritmos de simulación comunes; aunado al hecho de no tener una expresión explícita para la función de densidad de transición.

En el primer capítulo presentamos brevemente los fundamentos de la teoría de las difusiones que servirá como base para entender la construcción de la difusión de Wright-Fisher.

En el segundo capítulo, introducimos a la cadena de Markov de Wright-Fisher y mostramos el procedimiento para obtener la difusión como proceso límite del observado a tiempo discreto. Posteriormente, encontramos sus principales características como la clasificación de las fronteras, su distribución estacional y la descomposición espectral.

Dado que ya hemos estudiado las principales características de la difusión de Wright-Fisher, en el tercer capítulo, explicamos algunos métodos de simulación como la aproximación de Euler y Milstein que son, como su nombre lo dice, métodos numéricos de aproximación discreta a la solución continua de la ecuación diferencial estocástica. Luego, el algoritmo BISS (Balanced Implicit Split Step) que también es un método de aproximación numérica que, a diferencia de los dos anteriores, preserva las fronteras del proceso. Posteriormente, ilustramos el algoritmo exacto, llamado así por utilizar la expansión en series de la función de densidad de transición y permite recuperar las distribuciones finito dimensionales de la difusión. A continuación, presentamos un método de simulación que utiliza una descomposición latente y el hecho de que la difusión es un proceso de Markov, obteniendo un algoritmo bastante sencillo de simular. Asimismo, en este mismo capítulo incluimos una comparación principalmente gráfica de los cinco algoritmos mencionados anteriormente. Finalmente, introducimos un método para simular puentes de la difusión de Wright-Fisher a partir de la generación de trayectorias.

Para concluir, en el cuarto capítulo presentamos el algoritmo EM (Esperanza - Maximización) el cual consiste en hacer inferencia sobre trayectorias de una difusión observada a tiempo discreto. La idea es considerarlo como un problema de información incompleta y a través de la generación de puentes podemos encontrar estimadores máximo verosímiles. Para evaluar el algoritmo se implementa un ejercicio de estimación de los parámetros de Wright-Fisher que utiliza simulaciones basadas en los mecanismos descritos en el capítulo anterior.

Capítulo 1

Procesos de difusión

Una difusión es un proceso estocástico que modela la evolución de un fenómeno. Formalmente, es un proceso estocástico a tiempo continuo que tiene trayectorias continuas casi seguramente. El estudio de este tipo de procesos es de gran importancia puesto que existe una gran variedad de disciplinas en las que se utilizan, entre ellas, física, química, biología, economía, finanzas, ingeniería y sociología. Algunas aplicaciones particulares son, por ejemplo, el comportamiento de partículas moleculares sujetas a distintas interacciones, fluctuaciones en el precio de activos financieros, sistemas de comunicación que involucran ruido, actividad neurofisiológica, variaciones del crecimiento poblacional, sustituciones genéticas en el desarrollo evolutivo, etc.

En este capítulo presentamos de manera general los principales resultados de la teoría de las difusiones; ya que pueden caracterizarse de distintas maneras, primero se define como solución a una ecuación diferencial estocástica sin profundizar mucho en la teoría y por otro lado, se define como un proceso de Markov. El desarrollo de la difusión partiendo de una ecuación diferencial estocástica está basado principalmente en [16], mientras que el resto del capítulo toma los resultados expuestos en [17].

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 a tiempo 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 elemento $t \in T$, $\omega \rightarrow X_t(\omega)$ 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 de ω al tiempo t .

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

Existen diversos tipos de procesos estocásticos, entre los más comunes y estudiados se encuentran los procesos de Markov, en ellos la evolución del proceso en el futuro depende solamente del valor del proceso en el presente, es decir, no se ve alterado por el comportamiento del pasado.

Definición 1.1.2. *Un proceso estocástico $\{X_t\}$ es llamado proceso de Markov si satisface la propiedad de Markov, es decir, sea $s < t$ y para cualquier $A \subset \mathbb{R}$ boreliano,*

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s),$$

donde $\mathcal{F}_s = \sigma(\{X_r : r \leq s\})$ es la σ -álgebra generada por $\{X_r : r \leq s\}$.

La probabilidad de que el proceso de Markov $\{X_t\}$ se encuentre en $A \subset \mathbb{R}$ boreliano al tiempo t dado que al tiempo s estuvo en el estado x , con $s < t$ está dada por

$$\mathbb{P}(X_t \in A | X_s = x) = \int_A p(x, y; s, t) dy,$$

donde $p(x, y; s, t) = \mathbb{P}(X_t = y | X_s = x)$ es llamada *función de densidad de transición*.

Como consecuencia de la propiedad de Markov, la densidad de transición cumple la ecuación de Chapman-Kolmogorov,

$$p(x, y; s, t) = \int_{-\infty}^{\infty} p(x, z; s, u)p(z, y; u, t)dz,$$

para todo $x, y \in \mathbb{R}$ y tiempos $s < u < t$.

Adicionalmente, si la densidad de transición solo depende del tiempo a través de la longitud del intervalo, $p(x, y; s, t) = p(x, y; 0, t - s)$ se dice que el proceso de Markov es *homogéneo* con respecto al tiempo y la denotaremos por $p(x, y; t - s)$.

El siguiente tipo de procesos surge de la idea de juegos de apuestas justos, donde la ganancia esperada del jugador a través del tiempo es constante. Antes de hacer una definición formal del modelo necesitamos conocer el siguiente concepto.

Definición 1.1.3. Una filtración $\{\mathcal{F}_t\}_{t \geq 0}$ es una familia de sub- σ -álgebras de \mathcal{F} que cumple $\mathcal{F}_s \subseteq \mathcal{F}_t$, $s < t$. Podemos pensar a la σ -álgebra \mathcal{F}_t como la información hasta el tiempo t , de manera que si $\mathcal{F}_s \subseteq \mathcal{F}_t$, indica que \mathcal{F}_t contiene más información que \mathcal{F}_s .

A la σ -álgebra generada por el proceso $\{X_t\}$ se le llama *filtración natural o canónica*. Además, se dice que el proceso $\{X_t\}$ es *adaptado* a la filtración $\{\mathcal{F}_t\}_{t \geq 0}$ si X_t es \mathcal{F}_t medible para todo $t \geq 0$.

Definición 1.1.4. Un proceso estocástico $\{M_t\}_{t \geq 0}$ es una martingala con respecto a la filtración $\{\mathcal{F}_t\}_{t \geq 0}$ si cumple tres propiedades:

- i. M_t es adaptado a la filtración $\{\mathcal{F}_t\}_{t \geq 0}$.
- ii. M_t es integrable, es decir, $\mathbb{E}[|M_t|] < \infty$.
- iii. $\mathbb{E}[M_t | \mathcal{F}_s] = M_s$, $s < t$.

Uno de los ejemplos clásicos de un proceso estocástico es el movimiento browniano o proceso de Wiener, que fue descubierto como un fenómeno físico por el botánico Robert Brown en 1827 al observar la trayectoria irregular que siguen las partículas de polen suspendidas en cierta sustancia. Sin embargo, los inicios de la descripción matemática de este fenómeno fueron a través de las leyes de la física de Einstein en 1905, desde este punto de vista la teoría se desarrolló por autores como Smoluchowski, Fokker, Planck, Ornstein, Uhlenbeck, entre otros. Posteriormente, en 1918, Wiener formuló la definición matemática formal y la prueba de existencia. Para más detalles del proceso puede consultarse [10], [18], [24], [27], y [28].

Definición 1.1.5. Un movimiento browniano estándar $\{B_t\}_{t \geq 0}$ es un proceso estocástico continuo tal que:

- i. $B_0 = 0$ con probabilidad 1.
- ii. $\{B_t\}_{t \geq 0}$ tiene incrementos independientes. Es decir, para los tiempos $0 = t_0 < t_1 < \dots < t_n$, se tiene que las variables aleatorias $B_{t_i} - B_{t_{i-1}}$ son independientes para $1 \leq i \leq n$.
- iii. $\{B_t\}_{t \geq 0}$ tiene incrementos estacionarios. Es decir, la variable aleatoria $B_{t+s} - B_t$ tiene la misma distribución que B_s para cualquier elección de s y t . Más aún, el proceso al tiempo s , B_s , se distribuye como una normal de media cero y varianza s , denotado por $B_s \sim N(0, s)$.

El movimiento browniano estándar cumple las siguientes propiedades:

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.

La demostración de 1, puede encontrarse en [28] y las propiedades 2, 3, 4 y 5 en [10] o [24].

Existen algunas de variaciones del movimiento browniano estándar, entre ellas se considera al movimiento browniano que se construye como

$$X_t = \sigma B_t, \quad t \geq 0,$$

y al movimiento browniano con deriva,

$$X_t = \mu t + \sigma B_t, \quad t \geq 0. \quad (1.1)$$

Ambos procesos heredan las propiedades del movimiento browniano estándar con la diferencia de que los incrementos $X_{t+s} - X_t$ tienen distribución normal de media cero y varianza $\sigma^2 s$ en el primer caso y normal de media μs y varianza $\sigma^2 s$ en el segundo, donde μ y σ son constantes fijas llamadas parámetro de *deriva* y parámetro de *varianza* o *difusión* respectivamente.

1.2. Ecuaciones diferenciales estocásticas

El movimiento browniano puede utilizarse para incorporar un elemento estocástico o “ruido” a una ecuación diferencial ordinaria, lo que permite en muchas ocasiones encontrar un mejor modelo para el sistema en estudio. Por ejemplo, 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. Cuando $a(t)$ no está completamente determinada y depende de elementos aleatorios que provienen del entorno, es decir, $a(t) = r(t) + \text{“ruido”}$ donde $r(t)$ es determinista y el “ruido” es la parte aleatoria que puede describirse mediante una función de distribución, obtenemos una ecuación diferencial estocástica.

Consideremos la ecuación diferencial estocástica

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

donde $\mu(t, x)$ y $\sigma(t, x)$ son dos funciones dadas y $\{B_t\}_{t \geq 0}$ es un movimiento browniano estándar. En notación integral

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

Observamos que tenemos una integral respecto al movimiento browniano, pero como no tiene variación acotada no es posible utilizar la integral de Riemann-Stieltjes. Por esta razón surge la integral de Itô. La construcción de la integral de Itô utiliza, en primer lugar, funciones sencillas llamadas funciones simples y posteriormente se generaliza aproximando cualquier función en \mathcal{V} por funciones simples, donde $\mathcal{V} = \mathcal{V}(S, T)$ es la clase de funciones $f(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R}$ tales que:

1. La función $(t, \omega) \rightarrow f(t, \omega)$ es $(\mathcal{B} \times \mathcal{F})$ -medible, donde \mathcal{B} denota la σ -álgebra de borel del $[0, \infty]$.
2. La función $f(t, \omega)$ es \mathcal{F} -adaptada.
3. $\mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right] < \infty$.

Cabe mencionar que la clase de funciones \mathcal{V} pertenece a $\mathcal{L}^2_{(d\mathbb{P} \times dt)}$. La construcción de esta integral que acabamos de describir rápidamente puede encontrarse en [16].

1.2.1. Fórmula de Itô

La fórmula de Itô es una herramienta importante para el cálculo estocástico porque es útil para resolver la integral de Itô y como consecuencia, encontrar la solución a algunas ecuaciones diferenciales estocásticas.

Hablando de la integral de Riemann rara vez se utiliza la definición para resolver este tipo de integrales, en su lugar se emplea el teorema fundamental del cálculo y la regla de la cadena. En el contexto de la integral de Itô solo se tiene teoría sobre integración, sin embargo, la versión análoga a la regla de la cadena es la fórmula de Itô.

Consideremos al 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, la demostración y algunos ejemplos de su uso pueden hallarse en [16] o [31].

Teorema 1.2.1 (Fórmula de Itô). *Para cualquier función $f(t, x) \in C^{1,2}(\mathbb{R}^+, \mathbb{R})$*

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

Generalizando, supongamos ahora que el proceso $\{Y_t\}_{t \geq 0}$ está 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 $g(t, x) = f(t, \mu t + \sigma x)$. Para más detalle de este enfoque de la fórmula de Itô y el Teorema 1.2.1 se puede consultar [31].

Denotamos 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,

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

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

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

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

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

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

$$\begin{array}{c|cc} \cdot & d_t & dB_t \\ \hline d_t & 0 & 0 \\ dB_t & 0 & ds \end{array} \quad (1.3)$$

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

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

Pues $dX_t \cdot dX_t$ se simplifica

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

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

1.2.2. Transformación de Lamperti.

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

Supongamos que tenemos una ecuación diferencial estocástica de tal forma que la función $\sigma(t, X_t)$ depende solamente del estado del proceso al tiempo t , es decir,

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

la transformación de Lamperti permite convertir tal ecuación en una nueva ecuación diferencial estocástica con $\sigma(X_t) = 1$, 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 la nueva ecuación diferencial

$$dY_t = \mu_Y(t, Y_t)dt + dB_t, \quad (1.4)$$

μ_Y se deriva a continuación.

Aplicamos la fórmula de Itô para $f(X_t)$, como

$$f(t, x) = \int_z^x \frac{1}{\sigma(u)} du,$$

recordando $f_t = \frac{\partial f}{\partial t}$, $f_x = \frac{\partial f}{\partial x}$ y $f_{xx} = \frac{\partial^2 f}{\partial x^2}$, tenemos que

$$f_t(t, x) = 0, \quad f_x(t, x) = \frac{1}{\sigma(x)} \quad \text{y} \quad f_{xx}(t, x) = -\frac{\sigma_x(x)}{\sigma^2(x)}.$$

Luego,

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

Igualando los coeficientes, obtenemos

$$\mu_Y(t, y) = \frac{\mu(t, f^{-1}(y))}{\sigma(f^{-1}(y))} - \frac{1}{2}\sigma_x(f^{-1}(y)).$$

1.3. Difusiones

Las difusiones pueden estudiarse desde distintas perspectivas, en primer lugar vamos a introducirlas como la solución a una ecuación diferencial estocástica y posteriormente como un proceso estocástico que cumple ciertas características. Este tipo de procesos tienen aplicaciones en diversas áreas, por ejemplo, en biología, genética, economía, finanzas, química y sociología.

Viendo a la difusión como solución a una ecuación diferenciales estocástica, podemos interpretar al proceso como un modelo gobernado por una parte determinista que es alterado mediante un ruido aleatorio que corresponde al término que involucra al movimiento browniano.

1.3.1. Difusiones a partir de ecuaciones diferenciales estocásticas

Consideremos la ecuación diferencial estocástica de la forma

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (1.5)$$

A la solución de la ecuación (1.5) se le llama *difusión* y a las funciones $\mu(t, X_t)$ y $\sigma(t, X_t)$ se les llama *coeficiente de deriva* y *coeficiente de difusión* respectivamente.

El teorema que vamos a enunciar ahora da las condiciones necesarias para que tal solución exista y sea única, por la importancia y utilidad que tiene el teorema puede encontrarse la prueba en una gran variedad de textos, por ejemplo en [16] y [31].

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

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

para alguna C constante, y

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

para D constante. Sea X_0 una variable aleatoria 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

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T, \quad (1.6)$$

tiene una única solución fuerte $\{X_t\}$ 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 con $s < t$.

A la solución $\{X_t\}$ dada por el teorema anterior se le llama *solución fuerte* y cumple que X_t es $\mathcal{F}_t^{X_0}$ -adaptado, donde $\{\mathcal{F}_t^{X_0}\}$ es la filtración generada por B_t y X_0 . Por su parte, si tenemos las funciones $\mu(t, x)$ y $\sigma(t, x)$ y un par de procesos $(\{\tilde{X}_t, \tilde{B}_t\}, \mathcal{H}_t)$ en un espacio de probabilidad $(\Omega, \mathcal{H}, \mathbb{P})$ y se cumple la ecuación diferencial estocástica (1.6), entonces la solución $(\tilde{X}_t, \tilde{B}_t)$ es llamada *solución débil*. Donde $\{\mathcal{H}_t\}$ es una familia creciente de σ -álgebras tal que \tilde{X}_t es \mathcal{H}_t -adaptado y el movimiento browniano \tilde{B}_t es martingala respecto a \mathcal{H}_t . Una solución fuerte es también débil, pero en general no se cumple lo inverso.

La unicidad del Teorema 1.3.1 es llamada *unicidad fuerte o de trayectoria*, mientras que la *unicidad débil* hace referencia a que los procesos son idénticos en ley, esto es, que tienen las mismas distribuciones finito dimensionales.

La definición matemática precisa de una difusión es la siguiente.

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

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \text{ con } X_s = x; \quad s \leq t,$$

donde $\{B_t\}_{t \geq 0}$ representa un movimiento browniano estándar y las funciones $\mu : \mathbb{R} \rightarrow \mathbb{R}$ y $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ cumplen las condiciones del Teorema 1.3.1.

Una característica fundamental de las difusiones es la propiedad de Markov, la cual tiene que ver con la pérdida de memoria de los procesos estocásticos. Es decir, el comportamiento futuro del proceso dado lo que ha ocurrido hasta el tiempo t es el mismo que el comportamiento que se obtiene al iniciar el proceso en X_t . La demostración de los dos teoremas relacionados a esta propiedad puede consultarse en [16].

Teorema 1.3.2 (Propiedad de Markov). Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ borel acotada, entonces para $t, h \geq 0$

$$\mathbb{E}_x [f(X_{t+h}) | \mathcal{F}_t^{(m)}] = \mathbb{E}_{X_t(\omega)} [f(X_h)], \quad (1.7)$$

donde \mathbb{E}_x es la esperanza respecto a Q_x y \mathbb{E} respecto a P_0 , tal que $\mathbb{E}_y [f(X_h)] = \mathbb{E} [f(X_h^y)]$.

Ahora, de forma general la propiedad fuerte de Markov nos permite intercambiar el tiempo t por un tiempo de paro τ en la ecuación (1.7), para ello es necesario definir qué es un tiempo de paro.

Definición 1.3.2. Sea $\{\mathcal{F}_t\}_{t \geq 0}$ una familia creciente de σ -álgebras. Se le llama tiempo de paro a la función $\tau : \Omega \rightarrow [0, \infty]$ con respecto a $\{\mathcal{F}_t\}_{t \geq 0}$ si

$$\{\omega \in \Omega | \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \text{para toda } t \text{ positiva.}$$

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

Intuitivamente la definición de tiempo de paro indica que es posible saber si el evento $\tau \leq t$ ha ocurrido o no dado el conocimiento de $\{\mathcal{F}_t\}$. El ejemplo más común de un tiempo de paro que estudiaremos más adelante es la primera vez que un proceso alcanza cierto estado.

Denotamos a $\mathcal{F}_\tau^{(m)}$ la σ -álgebra generada por $\{B_{s \wedge \tau}\}_{s \geq 0}$.

Teorema 1.3.3 (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 para toda $h \geq 0$,

$$\mathbb{E}_x [f(X_{\tau+h}) | \mathcal{F}_\tau^{(m)}] = \mathbb{E}_{X_\tau(\omega)} [f(X_h)]. \quad (1.8)$$

Las difusiones más comunes son el movimiento browniano, el movimiento browniano geométrico, que es una función del movimiento browniano y se utiliza principalmente para modelar precios de acciones mediante la fórmula de Black and Scholes, en [31] se incluye este tipo de procesos. El proceso de Ornstein-Uhlenbeck o Vasicek que puede obtenerse como un proceso límite del modelo de la urna de Ehrenfest; también se utiliza principalmente en finanzas como un modelo de volatilidad. El modelo Cox-Ingersoll-Ross, describe la evolución de las tasas de interés, se puede consultar [14] para más detalle.

1.3.2. Ejemplos de difusiones

A continuación, veamos un par de ejemplos.

1. Consideremos la ecuación diferencial estocástica

$$dX_t = \beta X_t dt + \sigma X_t dB_t,$$

donde $\beta, \sigma \in \mathbb{R}^+$. 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.

Para encontrar la solución igualamos los coeficientes con los de la fórmula de Itô,

$$dX_t = f_t(t, B_t)dt + f_x(t, B_t)dX_t + \frac{1}{2}f_{xx}(t, B_t)dt$$

y consideramos al proceso X_t como función del movimiento browniano, es decir, $X_t = f(t, B_t)$ entonces,

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

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

De la primer ecuación, $f(t, x) = e^{\sigma x + g(t)}$ donde $g(t)$ es una función arbitraria y de la segunda ecuación sabemos que $g_x(t) = \beta - \frac{1}{2}\sigma^2$. De esta manera, la ecuación diferencial estocástica tiene solución

$$X_t = X_0 e^{(\beta - \frac{1}{2}\sigma^2)t + \sigma B_t},$$

a este proceso se le llama movimiento browniano geométrico.

2. Consideremos la ecuación diferencial estocástica

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dB_t,$$

donde $\theta_1, \theta_2 \in \mathbb{R}$ y $\theta_3 \in \mathbb{R}^+$. Esta ecuación cumple las propiedades que garantizan la existencia y unicidad de la solución del del Teorema 1.3.1, pues

$$|\theta_1 - \theta_2 x| + |\theta_3| \leq C(1 + |x|) \quad \text{tomando} \quad \max\{|\theta_1| + |\theta_3|, |-\theta_2|\} \leq C.$$

Y la segunda condición,

$$|(\theta_1 - \theta_2 x) - (\theta_1 - \theta_2 y)| + |\theta_3 - \theta_3| = |-\theta_2 x + \theta_2 y| = |-\theta_2||x - y| \leq D|x - y|$$

con D constante tal que $|-\theta_2| \leq D$.

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

$$X_t = \frac{\theta_1}{\theta_2} + e^{-\theta_2 t} \left(x_0 - \frac{\theta_1}{\theta_2} \right) + \theta_3 \int_0^t e^{-\theta_2(t-u)} dB_u,$$

al proceso encontrado se le llama difusión de Ornstein-Uhlenbeck.

1.3.3. Difusiones como procesos de Markov

Como ya habíamos mencionado, una difusión puede caracterizarse de distintas maneras, ya estudiamos al proceso como solución a una ecuación diferencial estocástica, en esta sección presentamos otras formas de definirlo, el desarrollo de este planteamiento se encuentra en [17].

Definición 1.3.3. Una difusión es un proceso estocástico $\{X_t\}_{t \geq 0}$ a tiempo continuo que posee la propiedad fuerte de Markov cuyas trayectorias están en función del tiempo y son continuas casi seguramente (c.s.).

El siguiente concepto es importante para el desarrollo de la teoría de las difusiones como procesos de Markov.

Definición 1.3.4. Se dice que un proceso estocástico $\{X_t\}_{t \geq 0}$ es regular si iniciando en cualquier punto del espacio de estados E es posible llegar a cualquier otro punto de E con probabilidad positiva. Es decir, consideremos el tiempo de paro $\tau_x = \inf\{t \geq 0 : X_t = x\}$, el proceso $\{X_t\}_{t \geq 0}$ es regular si

$$\mathbb{P}(\tau_x < \infty | X_0 = z) > 0 \text{ para } x, z \in E.$$

El espacio de estados E de una difusión $\{X_t\}_{t \geq 0}$ puede ser un intervalo de la forma (l, r) , $(l, r]$, $[l, r)$ o $[l, r]$ permitiendo que $l = -\infty$ y $r = \infty$.

Para efectos del presente trabajo vamos a considerar difusiones regulares.

1.3.4. Propiedades de las difusiones

A partir de la Definición 1.3.3 podemos enunciar algunas propiedades de las difusiones. Sea $\Delta_h X_t = X_{t+h} - X_t$ el incremento del proceso de tamaño h , luego,

1. Para toda $\varepsilon > 0$,

$$\lim_{h \downarrow 0} \mathbb{P}(|X_{t+h} - x| > \varepsilon | X_t = x) = 0 \quad \forall x \in E.$$

Este hecho asegura que las trayectorias del proceso sean continuas pues indica que para intervalos de tiempo suficientemente pequeños el proceso se desplaza en a lo más ε con probabilidad cercana a cero.

2. $\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[\Delta_h X_t | X_t = x] = \mu(x, t)$.
3. $\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[(\Delta_h X_t)^2 | X_t = x] = \sigma^2(x, t)$.

A las funciones $\mu(x, t)$ y $\sigma^2(x, t)$ se les llama parámetros infinitesimales del proceso, particularmente, a $\mu(x, t)$ se le denomina de media infinitesimal que corresponde al parámetro de deriva y $\sigma^2(x, t)$ se le conoce como varianza infinitesimal, donde $\sigma(x, t)$ es el parámetro de difusión. Si además, la difusión es homogénea en el tiempo, los parámetros infinitesimales no dependen del tiempo, es decir, $\mu(x, t) = \mu(x)$ y $\sigma^2(x, t) = \sigma^2(x)$.

4. $\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[(\Delta_h X_t)^k | X_t = x] = 0 \quad k = 3, 4, \dots$

Las condiciones (2) y (3) describen la media y la varianza infinitesimal del proceso, pues

$$\mathbb{E}[\Delta_h X_t | X_t] = \mu(x)h + \mathcal{O}(h)$$

y

$$\begin{aligned} \text{Var}(\Delta_h X_t | X_t) &= \mathbb{E}[(\Delta_h X_t)^2 | X_t] - \mathbb{E}[\Delta_h X_t | X_t]^2 \\ &= \sigma^2(x)h + \mathcal{O}(h) + (\mu(x)h + \mathcal{O}(h))^2 \\ &= \sigma^2(x)h + \mathcal{O}(h), \end{aligned}$$

donde $\mathcal{O}(h)$ es cualquier función f que cumple $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$.

Cabe mencionar que algunos autores definen a las difusiones a partir de las propiedades antes mencionadas, dicho de otra manera, una difusión es un proceso de Markov que satisface las condiciones (1), (2) y (3).

Es conveniente que tengamos condiciones suficientes con las cuales sea posible verificar si un proceso de Markov es una difusión, para ello introducimos la siguiente definición.

Definición 1.3.5. *Un proceso de Markov $\{X_t\}_{t \geq 0}$ es un proceso estándar si:*

- i. *El proceso $\{X_t\}_{t \geq 0}$ es continuo por la derecha, $\lim_{t \downarrow s} X_t = X_s$.*
- ii. *El proceso $\{X_t\}_{t \geq 0}$ tiene límite por la izquierda, $\lim_{t \uparrow s} X_t$ existe para $s > t$.*
- iii. *El proceso $\{X_t\}_{t \geq 0}$ es continuo por la izquierda casi seguramente en tiempos de paro, a esta propiedad se le llama cuasi continuidad por la izquierda. Sean $\tau_1 \leq \tau_2 \leq \dots$ tiempos de paro que convergen a τ , siendo $\tau \leq \infty$. Entonces $\lim_{n \rightarrow \infty} X_{\tau_n} = X_\tau$ c.s., es decir,*

$$\mathbb{P} \left(\tau < \infty, \lim_{n \rightarrow \infty} X_{\tau_n} \neq X_\tau \right) = 0.$$

La cuasi continuidad no implica que las trayectorias son continuas por la izquierda, sino que los saltos no pueden predecirse. Por ejemplo, el proceso Poisson es un proceso estándar cuyas trayectorias tienen saltos.

Se dice que un proceso es continuo en probabilidad si $\lim_{t \rightarrow s} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0$ para $\varepsilon > 0$ y $s \geq 0$. Todo proceso estocástico $\{X_t\}_{t \geq 0}$ que cumple la propiedad fuerte de Markov y es continuo en probabilidad posee una versión equivalente $\{\tilde{X}_t\}_{t \geq 0}$, que es un proceso estándar. En otras palabras, los procesos $\{X_t\}_{t \geq 0}$ y $\{\tilde{X}_t\}_{t \geq 0}$ tienen las mismas distribuciones finito dimensionales.

Ya que muchos procesos estocásticos cumplen la propiedad de continuidad en probabilidad, podemos suponer que el proceso de Markov que estamos considerando es un proceso estándar.

El siguiente teorema proporciona una condición suficiente para que un proceso estándar sea una difusión. La demostración puede encontrarse en [17].

Teorema 1.3.4. *Sea $\{X_t\}_{t \geq 0}$ un proceso estándar que además cumple la propiedad de Dynkin, sea $\varepsilon > 0$ y para todo $x \in E$,*

$$\frac{1}{h} \mathbb{P}(|X_{t+h} - X_t| > \varepsilon | X_t = x) \xrightarrow{h \downarrow 0} 0. \quad (1.9)$$

Entonces, $\{X_t\}_{t \geq 0}$ es un proceso de difusión.

Así que para saber si un proceso estándar es una difusión basta con probar la condición de Dynkin, lo cual resulta relativamente sencillo bajo el siguiente lema.

Lema 1.3.1. *Si un proceso estándar satisface la condición del momento infinitesimal, esto es, para alguna $p > 2$*

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E} [|\Delta_h X_t|^p | X_t = x] = 0, \quad (1.10)$$

uniformemente para x en cualquier subintervalo compacto del (l, r) , y t en un intervalo finito $[0, N]$, entonces la condición de Dynkin (1.9) se satisface.

Para probar el lema basta utilizar la desigualdad de Chebyshev. El siguiente teorema es de utilidad para encontrar nuevas difusiones como funciones de una difusión conocida. De hecho, la transformación de Lamperti puede verse como un caso particular que utiliza la función $g(X_t) = \int^{X_t} \frac{1}{\sigma(x)} dx$.

Teorema 1.3.5. Sea $\{X_t\}$ un proceso de difusión regular en el intervalo (l, r) con parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$. Sea g una función monótona con segunda derivada continua en el intervalo con fronteras en $g(l)$ y $g(r)$. Entonces $Y_t = g(X_t)$ es una difusión regular en el intervalo dado por los puntos $g(l)$ y $g(r)$ y además, tiene parámetros infinitesimales

$$\mu_Y(y) = \mu(x)g_x(x) + \frac{1}{2}\sigma^2(x)g_{xx}(x)$$

y

$$\sigma_Y^2(y) = \sigma^2(x)g_x^2(x).$$

Demostración. Supongamos que g es creciente y sea g_x y g_{xx} la primer y segunda derivada de g con respecto a x , utilizando la expansión de Taylor,

$$g(x + \Delta x) = g(x) + \Delta x g_x(x) + \frac{1}{2}(\Delta x)^2 g_{xx}(x) + \frac{1}{2}(\Delta x)^2 (g_{xx}(\xi) - g_{xx}(x))$$

con $x \leq \xi \leq x + \Delta x$. Sustituyendo $X_t = x$ y $\Delta X = X_{t+h} - X_t$ tenemos que

$$g(X_{t+h}) = g(X_t) + \Delta X g_x(X_t) + \frac{1}{2}(\Delta X)^2 g_{xx}(X_t) + \frac{1}{2}(\Delta X)^2 (g_{xx}(\xi(\omega)) - g_{xx}(X_t)),$$

$\xi(\omega)$ se encuentra entre X_t y X_{t+h} . Puesto que definimos $Y_t = g(X_t)$ implica que

$$Y_{t+h} - Y_t = \Delta X g_x(X_t) + \frac{1}{2}(\Delta X)^2 g_{xx}(X_t) + \frac{1}{2}(\Delta X)^2 (g_{xx}(\xi(\omega)) - g_{xx}(X_t)).$$

Luego, condicionando de tal manera que $Y_t = g(X_t) = g(x) = y$, calculamos la esperanza, dividimos entre h y tomamos límite

$$\begin{aligned} & \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[Y_{t+h} - Y_t | Y_t = y] \\ &= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E} \left[\Delta X g_x(X_t) + \frac{1}{2}(\Delta X)^2 g_{xx}(X_t) + \frac{1}{2}(\Delta X)^2 (g_{xx}(\xi(\omega)) - g_{xx}(X_t)) \middle| g(X_t) = g(x) \right] \\ &= \mu(x)g_x(x) + \frac{1}{2}\sigma^2(x)g_{xx}(x) + \frac{1}{2} \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E} [(\Delta X)^2 (g_{xx}(\xi(\omega)) - g_{xx}(X_t))], \end{aligned}$$

debido a que $\frac{1}{h} \mathbb{E}[(\Delta X)^2]$ converge y a la continuidad de g se tiene que

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E} [(\Delta X)^2 (g_{xx}(\xi(\omega)) - g_{xx}(X_t))] = 0.$$

Por lo tanto,

$$\mu_Y(y) = \mu(x)g_x(x) + \frac{1}{2}\sigma^2(x)g_{xx}(x).$$

Para encontrar el coeficiente de difusión, notamos que elevando al cuadrado la expresión en series de Taylor para $Y_{t+h} - Y_t$ todos los términos son de orden mayor a $(\Delta X)^3$ salvo el primero, $(\Delta X)^2 (g_{xx}(\xi(\omega)) - g_{xx}(X_t))^2$, entonces aplicamos esperanza, dividimos entre h y posteriormente tomamos límite, así $\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[|\Delta x|^r | X_t = x] = 0$ para $r > 2$ y como consecuencia, obtenemos

$$\sigma_Y^2(y) = \sigma^2(x)g_x^2(x).$$

□

1.3.5. Generador infinitesimal y algunos funcionales asociados

El generador infinitesimal de un proceso contiene mucha información de éste; particularmente, describe cómo el proceso se mueve de un punto x a otro punto del espacio de estados en un tiempo suficientemente pequeño.

Definición 1.3.6. Sea $\{X_t\}_{t \geq 0}$ una difusión, el generador infinitesimal \mathcal{A} de la difusión se define para $x \in \mathbb{R}$ como

$$\mathcal{A}f(x) = \lim_{h \downarrow 0} \frac{\mathbb{E}(f(X_{t+h}) - f(X_t) | X_t = x)}{h} = \lim_{h \downarrow 0} \frac{\mathbb{E}(f(X_{t+h}) | X_t = x) - f(x)}{h}.$$

Al conjunto de funciones $f : \mathbb{R} \rightarrow \mathbb{R}$ para las cuales el límite existe para todo $x \in \mathbb{R}$ se denota por $\mathcal{D}_{\mathcal{A}}$ y al conjunto de funciones donde el límite en x existe se denota por $\mathcal{D}_{\mathcal{A}}(x)$.

Si el proceso es homogéneo en el tiempo, podemos escribir al generador infinitesimal como

$$\mathcal{A}f(x) = \lim_{h \downarrow 0} \frac{\mathbb{E}(f(X_h) - f(X_0) | X_0 = x)}{h}.$$

Existe una relación entre el generador \mathcal{A} y los coeficientes de una ecuación diferencial estocástica, dicha relación está dada en los siguientes dos resultados, los cuales pueden encontrarse en [16].

Lema 1.3.2. Sea $\{X_t\}_{t \geq 0}$ una difusión, escrita de forma integral,

$$X_t = x + \int_0^t \mu(x, s) ds + \int_0^t \sigma^2(x, s) dB_s$$

donde $\{B_s\}_{s \geq 0}$ es un movimiento browniano estándar. Sea $f \in C^2(\mathbb{R})$ con soporte compacto, τ un tiempo de paro con respecto a $\{\mathcal{F}_t\}_{t \geq 0}$ y $\mathbb{E}_x[\tau] < \infty$, los coeficientes μ y σ^2 son funciones acotadas, tal que X_t pertenece al soporte de f . Entonces

$$\mathbb{E}_x[f(X_\tau)] = f(x) + \mathbb{E}_x \left[\int_0^\tau \mu(x, s) \frac{\partial}{\partial x} f(X_s) + \frac{1}{2} \sigma^4(x, s) \frac{\partial^2}{\partial x^2} f(X_s) ds \right], \quad (1.11)$$

con \mathbb{E}_x la esperanza con respecto a la ley de probabilidad del proceso iniciando en x .

Demostración. Para simplificar notación denotaremos los parámetros infinitesimales $\mu(x, t) = \mu$ y $\sigma^2(x, t) = \sigma^2$. Aplicando la fórmula de Itô del Teorema 1.2.1 a $f(X_t)$,

$$\begin{aligned} d(f(X_t)) &= \frac{\partial}{\partial t} f(X_t) dX_t + \frac{\partial}{\partial x} f(X_t) dX_t + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(X_t) (dX_t)^2 \\ &= \frac{\partial}{\partial x} f(X_t) (\mu dt + \sigma^2 dB_t) + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(X_t) (\mu dt + \sigma^2 dB_t) (\mu dt + \sigma^2 dB_t) \\ &= \left(\mu \frac{\partial}{\partial x} f(X_t) + \frac{1}{2} \sigma^4 \frac{\partial^2}{\partial x^2} f(X_t) \right) dt + \sigma^2 \frac{\partial}{\partial x} f(X_t) dB_t. \end{aligned}$$

La última igualdad se cumple debido a que $dt \cdot dt = 0$, $dt \cdot dB_t = 0$ y $dB_t \cdot dB_t = dt$, dado en la Tabla 1.2.1. Entonces,

$$f(X_t) = f(X_0) + \int_0^t \left(\mu \frac{\partial}{\partial x} f(X_s) + \frac{1}{2} \sigma^4 \frac{\partial^2}{\partial x^2} f(X_s) \right) ds + \int_0^t \sigma^2 \frac{\partial}{\partial x} f(X_s) dB_s.$$

Calculando esperanza condicional cuando $X_0 = x$ en el tiempo τ ,

$$\mathbb{E}_x[f(X_t)] = \mathbb{E}[f(x)] + \mathbb{E}_x \left[\int_0^\tau \left(\mu \frac{\partial}{\partial x} f(X_s) + \frac{1}{2} \sigma^4 \frac{\partial^2}{\partial x^2} f(X_s) \right) ds \right] + \mathbb{E}_x \left[\int_0^\tau \sigma^2 \frac{\partial}{\partial x} f(X_t) dB_s \right].$$

Veamos ahora que el último término es cero. Sea g una función de borel acotada, es decir, $|g| \leq M$, entonces para k entero se tiene que

$$\begin{aligned} \mathbb{E}_x \left[\int_0^{k \wedge \tau} g(X_s) dB_s \right] &= \mathbb{E}_x \left[\int_0^k \mathbb{I}_{(s < \tau)} g(X_s) dB_s \right] \\ &\leq \mathbb{E}_x \left[M \int_0^k dB_s \right] \\ &= 0. \end{aligned}$$

Luego, por la isometría de Itô

$$\begin{aligned}\mathbb{E}_x \left[\left(\int_0^\tau g(X_s) dB_s - \int_0^{k \wedge \tau} g(X_s) dB_s \right)^2 \right] &= \mathbb{E}_x \left[\left(\int_{k \wedge \tau}^\tau g(X_s) dB_s \right)^2 \right] \\ &= \mathbb{E}_x \left[\int_{k \wedge \tau}^\tau g(X_s)^2 ds \right] \\ &\leq M^2 \mathbb{E}_x (\tau - \tau \wedge k) \xrightarrow[k \rightarrow \infty]{} 0.\end{aligned}$$

Finalmente, $0 = \lim_{k \rightarrow \infty} \mathbb{E}_x \left[\int_0^{k \wedge \tau} g(X_s) dB_s \right] = \mathbb{E}_x \left[\int_0^\tau g(X_s) dB_s \right]$.

Por lo tanto,

$$\mathbb{E}_x [f(X_\tau)] = f(x) + \mathbb{E}_x \left[\int_0^\tau \left(\mu \frac{\partial}{\partial x} f(X_s) + \frac{1}{2} \sigma^4 \frac{\partial^2}{\partial x^2} f(X_s) \right) ds \right].$$

□

De esta manera, podemos escribir al generador infinitesimal de un proceso de difusión como una ecuación diferencial de segundo orden que depende de los parámetros de deriva y difusión.

Teorema 1.3.6. *Sea $\{X_t\}_{t \geq 0}$ una difusión con coeficiente de deriva $\mu(x)$ y coeficiente de difusión $\sigma^2(x)$. Sea f una función real tal que $f \in C_0^2(\mathbb{R})$, entonces, $f \in \mathcal{D}_A$ y el generador infinitesimal*

$$\mathcal{A}f(x) = \mu(x) \frac{\partial}{\partial x} f(X_s) + \frac{1}{2} \sigma^4(x) \frac{\partial^2}{\partial x^2} f(X_s).$$

Simplificando la notación, $f_x = \frac{\partial f}{\partial x}$ y $f_{xx} = \frac{\partial^2 f}{\partial x^2}$ el generador se escribe como

$$\mathcal{A}f(x) = \mu(x) f_x + \frac{1}{2} \sigma^4(x) f_{xx}. \quad (1.12)$$

Demostración. Por el Lema 1.3.2

$$\begin{aligned}\mathcal{A}f(x) &= \lim_{h \downarrow 0} \frac{\mathbb{E}(f(X_h) - f(X_0) | X_0 = x)}{h} \\ &= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}_x \left[\int_0^h \left(\mu f_x(X_s) + \frac{1}{2} \sigma^4 f_{xx}(X_s) \right) ds \right] \\ &= \lim_{h \downarrow 0} \int_0^h \frac{1}{h} \mathbb{E}_x \left[\left(\mu f_x(X_s) + \frac{1}{2} \sigma^4 f_{xx}(X_s) \right) \right] ds \\ &= \mathbb{E}_x \left[\left(\mu f_x(X_0) + \frac{1}{2} \sigma^4 f_{xx}(X_0) \right) \right] \\ &= \mu f_x(x) + \frac{1}{2} \sigma^4 f_{xx}(x).\end{aligned}$$

□

Existen algunos funcionales relacionados con las ecuaciones diferenciales y con el generador infinitesimal que surgen como problemas básicos en el estudio de algunos procesos estocásticos. El desarrollo de estos funcionales y algunos ejemplos se presentan en [17].

Sea a y b elementos del espacio de estados, $\tau_a = \inf\{t \geq 0 : X_t = a\}$ y $\tau_b = \inf\{t \geq 0 : X_t = b\}$ los tiempos de llegada a a o b respectivamente y otro tiempo de paro $\tau_{a,b} = \min\{\tau_a, \tau_b\}$, es decir, la variable aleatoria que representa el tiempo en que el proceso llega al estado a o al estado b por primera vez.

Sea

$$u(x) = \mathbb{P}(\tau_b < \tau_a | X_0 = x) \quad (1.13)$$

con $a < x < b$ la probabilidad de que el proceso alcance el estado b antes que al estado a .

Bajo los supuestos en que estamos trabajando la función $u(x)$ tiene primera y segunda derivadas acotadas, adicionalmente, satisface la ecuación

$$\mu(x)\frac{du}{dx} + \frac{1}{2}\sigma^2\frac{d^2u}{dx^2} = 0, \quad a < x < b; \quad (1.14)$$

con condiciones $u(a) = 0$, $u(b) = 1$. En primer lugar, notemos que las condiciones iniciales son claras ya que $u(a) = \mathbb{P}(\tau_b < \tau_a | X_0 = a) = 0$ y $u(b) = \mathbb{P}(\tau_b < \tau_a | X_0 = b) = 1$. Para justificar la afirmación (1.14) elegimos un tiempo h suficientemente pequeño de tal manera que la probabilidad de llegar a a o b sea despreciable,

$$\begin{aligned} u(x) &= \mathbb{P}(\tau_b < \tau_a | X_0 = x) \\ &= \int_a^b \mathbb{P}(\tau_b < \tau_a, X_h = y, \tau_{a,b} \leq h | X_0 = x) dy + \int_a^b \mathbb{P}(\tau_b < \tau_a, X_h = y, \tau_{a,b} > h | X_0 = x) dy \\ &= \int_a^b \mathbb{P}(\tau_b < \tau_a | X_h = y, \tau_{a,b} \leq h, X_0 = x) \mathbb{P}(X_h = y, \tau_{a,b} \leq h | X_0 = x) dy \\ &\quad + \int_a^b \mathbb{P}(\tau_b < \tau_a | X_h = y, \tau_{a,b} > h, X_0 = x) \mathbb{P}(X_h = y, \tau_{a,b} > h | X_0 = x) dy \\ &= \int_a^b \mathbb{P}(\tau_b < \tau_a | X_h = y, \tau_{a,b} \leq h, X_0 = x) \mathcal{O}(h) dy \\ &\quad + \int_a^b \mathbb{P}(\tau_b < \tau_a | X_h = y, \tau_{a,b} > h, X_0 = x) \mathbb{P}(X_h = y, | X_0 = x) (1 - \mathcal{O}(h)) dy \\ &= \int_a^b \mathbb{P}(\tau_b < \tau_a | \tau_{a,b} > h, X_h = y) \mathbb{P}(X_h = y, | X_0 = x) dy + \mathcal{O}(h) \\ &= \mathbb{E}[u(X_h) | X_0 = x] + \mathcal{O}(h), \end{aligned}$$

considerando que el término de error $\mathcal{O}(h)$ tiene orden menor a h .

Ahora, utilizando la expansión en series de Taylor para $u(X_h)$ y escribiendo $\Delta X = X_h - x$,

$$\begin{aligned} u(X_h) &= u(x + \Delta X) \\ &= u(x) + \Delta X u_x(x) + \frac{1}{2}(\Delta X)^2 u_{xx}(x) + \dots, \end{aligned}$$

donde el cuarto sumando y los siguientes pueden ser omitidos debido a que son de orden menor que $(\Delta X)^2$.

Además, recordando que $\mathbb{E}[\Delta X | X_0 = x] = \mu(x)h + \mathcal{O}(h)$ y $\mathbb{E}[(\Delta X | X_0 = x)^2] = \sigma^2(x)h + \mathcal{O}(h)$, escribimos a la función u como,

$$\begin{aligned} u(x) &= \mathbb{E}[u(X_h) | X_0 = x] + \mathcal{O}(h) \\ &= \mathbb{E}[u(x + \Delta X) | X_0 = x] + \mathcal{O}(h) \\ &= \mathbb{E} \left[u(x) + \Delta X u_x(x) + \frac{1}{2}(\Delta X)^2 u_{xx}(x) \right] + \mathcal{O}(h) \\ &= u(x) + \mu(x)u_x(x)h + \frac{1}{2}\sigma^2(x)u_{xx}(x)h + \mathcal{O}(h). \end{aligned}$$

Posteriormente, restando $u(x)$ de ambos lados y dividiendo entre h ,

$$\mu(x)u_x(x) + \frac{1}{2}\sigma^2(x)u_{xx}(x) + \frac{\mathcal{O}(h)}{h} = 0,$$

por último al hacer tender h a cero obtenemos el resultado.

Es evidente que la ecuación (1.14) tiene la forma del generador infinitesimal de una difusión.

Las pregunta que surge a continuación es el tiempo medio en que el proceso alcanza alguno de los estados a o b , dado que iniciamos en un punto $x \in (a, b)$, es decir,

$$v(x) = \mathbb{E}[\tau_{a,b} | X_0 = x], \quad a < x < b. \quad (1.15)$$

O bien, de manera más general para una función g continua y acotada,

$$w(x) = \mathbb{E} \left[\int_0^{\tau_{a,b}} g(X_s) ds \middle| X_0 = x \right], \quad a < x < b, \quad (1.16)$$

como las trayectorias del proceso de difusión son continuas la integral está definida. Puede interpretarse a la función $g(x)$ como el costo en que se incurre cuando el proceso está en x , por lo que, la integral representa el costo total hasta el tiempo que se alcanza a o b . Utilizando $g(x) = 1$ para toda x en la función (1.16) obtenemos como caso particular a la función v dada en (1.15).

Dadas las propiedades de las difusiones, las funciones $v(x)$ y $w(x)$ son finitas y poseen primera y segunda derivada acotadas para $a < x < b$, consecuentemente, satisfacen las siguientes ecuaciones respectivamente,

$$\mu(x)v_x(x) + \frac{1}{2}\sigma^2(x)v_{xx}(x) = -1 \quad a < x < b, \quad (1.17)$$

y

$$\mu(x)w_x(x) + \frac{1}{2}\sigma^2(x)w_{xx}(x) = -g(x) \quad a < x < b, \quad (1.18)$$

con condiciones iniciales $v(a) = v(b) = 0$ y $w(a) = w(b) = 0$. Como ya mencionamos, $v(x)$ es un caso particular de $w(x)$ por lo que, basta justificarlo de manera general. Nuevamente las condiciones iniciales son claras y tomamos un tiempo h suficientemente pequeño tal que la probabilidad de alcanzar a o b sea despreciable. Condicionando a X_h ,

$$\begin{aligned} w(x) &= \mathbb{E} \left[\int_0^{\tau_{a,b}} g(X_s) ds \middle| X_0 = x \right] \\ &= \mathbb{E} \left[\int_0^h g(X_s) ds \middle| X_0 = x \right] + \mathbb{E} \left[\int_h^{\tau_{a,b}} g(X_s) ds \middle| X_0 = x \right], \end{aligned} \quad (1.19)$$

como tanto g como las trayectorias del proceso son continuas puede utilizarse una aproximación del primer sumando, es decir,

$$\mathbb{E} \left[\int_0^h g(X_s) ds \middle| X_0 = x \right] = g(x)h + \mathcal{O}(h).$$

Para el segundo, notemos que por la propiedad de Markov y si además el proceso es estacionario

$$\mathbb{E} \left[\int_h^{\tau_{a,b}} g(X_s) ds \middle| X_h = z \right] = \mathbb{E} \left[\int_0^{\tau_{a,b}} g(X_s) ds \middle| X_0 = z \right] = w(z).$$

De tal manera que para $a < x < b$,

$$w(x) = g(x)h + \mathcal{O}(h) + \mathbb{E} [w(X_h) | X_0 = x].$$

Luego, análogo a la función u , utilizamos la expansión de Taylor y la definición de esperanza y varianza infinitesimal de tal manera que

$$\begin{aligned} \mathbb{E}[w(X_h) | X_0 = x] &= \mathbb{E}[w(x + \Delta X) | X_0 = x] \\ &= w(x) + \mu(x)w_x(x)h + \frac{1}{2}\sigma^2(x)w_{xx}(x)h + \mathcal{O}(h). \end{aligned}$$

Como resultado,

$$w(x) = g(x)h + w(x) + \mu(x)w_x(x)h + \frac{1}{2}\sigma^2(x)w_{xx}(x)h + \mathcal{O}(h),$$

restando de ambos lados $w(x)$, dividimos entre h y finalmente, hacemos tender h a cero para obtener el resultado.

Podemos notar que las ecuaciones (1.14), (1.17) y (1.18) involucran al generador infinitesimal $\mathcal{A}f(x)$ donde f es una función dos veces diferenciable en el intervalo (a, b) . Por lo que estamos interesados en encontrar la solución a dicho operador, para ello definiremos las siguientes funciones.

Definición 1.3.7. La función de escala de un proceso con $\sigma^2(x) > 0$ para $l < x < r$ se define como

$$S(x) = \int^x s(y)dy \quad (1.20)$$

donde

$$s(x) = \exp \left\{ - \int^x \frac{2\mu(y)}{\sigma^2(y)} dy \right\}, \quad (1.21)$$

la función de densidad de velocidad,

$$m(x) = \frac{1}{\sigma^2(x)s(x)} \quad (1.22)$$

y la medida de velocidad

$$M(x) = \int^x m(y)dy. \quad (1.23)$$

Podemos observar que las definiciones están dadas como integrales indefinidas, más adelante veremos por qué pueden definirse de esta forma.

La función de escala recibe su nombre pues puede emplearse para reescalar el espacio de estados (l, r) en términos de probabilidades de alcanzar ciertos niveles. Determinamos el origen como un punto fijo x_0 , si definimos el proceso $Y_t = S(X_t)$ en el intervalo $(S(l), S(r))$ la función de escala establece una traslación. Además, como S es monótona y dos veces diferenciable, podemos aplicar el Teorema 1.3.5, considerando $y = S(x)$ obtenemos los parámetros infinitesimales del nuevo proceso

$$\mu_Y(y) = \frac{1}{2}\sigma^2(x)S''(x) + \mu(x)S'(x)$$

y

$$\sigma_Y^2(y) = \sigma^2(x)(S'(x))^2 = \sigma^2(x)s(x)^2.$$

Haciendo los cálculos $S'(x) = s(x)$ y $S''(x) = -s(x)\frac{2\mu(x)}{\sigma^2(x)}$, entonces,

$$\mu_Y(y) = 0$$

y

$$\sigma_Y^2(y) = \sigma^2(x)s(x)^2.$$

Se dice que un proceso $\{Y_t\}$ cuya función de escala es lineal está en *escala natural o canónica* ya que las probabilidades de llegada $\mathbb{P}(\tau_a < \tau_b | Y_0 = y) = \frac{b-y}{b-a}$ son claramente proporcionales a la distancia actual.

El siguiente resultado nos da la relación entre el generador infinitesimal y las medidas de escala y velocidad, lo cual nos permitirá dar solución a las ecuaciones (1.14), (1.17) y (1.18).

Proposición 1.3.1. El generador infinitesimal $\mathcal{A}f$ puede escribirse

$$\mathcal{A}f(x) = \frac{1}{2} \frac{1}{m(x)} \frac{d}{dx} \left(\frac{1}{s(x)} \frac{df(x)}{dx} \right).$$

Demostración. Primero notemos que

$$\begin{aligned} \frac{d}{dx} \left(\frac{1}{s(x)} \frac{df(x)}{dx} \right) &= \frac{1}{s(x)} \frac{d^2f(x)}{dx^2} - \frac{df(x)}{dx} \frac{s'(x)}{s(x)^2} \\ &= \frac{1}{s(x)} \left[\frac{d^2f(x)}{dx^2} + \frac{df(x)}{dx} \left(\frac{2\mu(x)}{\sigma^2(x)} \right) \right], \end{aligned} \quad (1.24)$$

debido a que $s'(x) = -\frac{2\mu(x)}{\sigma^2(x)}s(x)$, luego, $\frac{s'(x)}{s(x)} = -\frac{2\mu(x)}{\sigma^2(x)}$. Entonces,

$$\begin{aligned} \frac{1}{2} \frac{1}{m(x)} \frac{d}{dx} \left(\frac{1}{s(x)} \frac{df(x)}{dx} \right) &= \frac{1}{2} (\sigma^2(x)s(x)) \frac{1}{s(x)} \left[\frac{d^2f(x)}{dx^2} + \frac{df(x)}{dx} \left(\frac{2\mu(x)}{\sigma^2(x)} \right) \right] \\ &= \frac{1}{2} \sigma^2(x) \frac{d^2f(x)}{dx^2} + \mu(x) \frac{df(x)}{dx} \\ &= \mathcal{A}f(x). \end{aligned} \tag{1.25}$$

□

Para simplificar vamos a denotar $dS(x) = s(x)dx$ y $dM = m(x)dx$, así, el generador infinitesimal queda como

$$\mathcal{A}f(x) = \frac{1}{2} \frac{d}{dM} \frac{df(x)}{dS},$$

a esta expresión se le llama *representación canónica del operador infinitesimal diferencial* asociado al proceso de difusión. Esta forma del generador permite resolver la ecuación diferencial (1.12) mediante integraciones sucesivas.

Regresando al problema anterior, buscamos resolver las ecuaciones (1.14), (1.17) y (1.18). En primer lugar, queremos encontrar la probabilidad de que el proceso llegue al estado b antes que al estado a , es decir $u(x)$.

Utilizando la representación canónica, tenemos que

$$\frac{1}{2} \frac{d}{dM} \left(\frac{du(x)}{dS} \right) = 0, \quad a < x < b, \quad u(a) = 0, \quad u(b) = 1.$$

Integrando por primera vez obtenemos $\frac{du(x)}{dS} = k_1$, integrando nuevamente obtenemos $u(x) = k_2 + k_1S(x)$, donde k_1 y k_2 las constantes de integración. Finalmente, según las condiciones iniciales podemos hallar $k_2 = -k_1S(a)$ y $k_1 = \frac{1}{S(b)-S(a)}$. Por lo tanto,

$$u(x) = \frac{S(x) - S(a)}{S(b) - S(a)} \quad a \leq x \leq b. \tag{1.26}$$

Ahora, veamos que $u(x)$ permanece igual si reemplazamos $S(x)$ por $S^*(x) = c + dS(x)$ para c y $d \neq 0$ constantes

$$\frac{S^*(x) - S^*(a)}{S^*(b) - S^*(a)} = \frac{(c + dS(x)) - (c + dS(a))}{(c + dS(b)) - (c + dS(a))} = \frac{d(S(x) - S(a))}{d(S(b) - S(a))} = \frac{S(x) - S(a)}{S(b) - S(a)},$$

por esta razón podemos definir la función de escala a partir de integrales indefinidas pues la función $u(x)$ no depende del límite inferior de integración.

Posteriormente, para encontrar las funciones v y w basta resolver con integraciones sucesivas la ecuación diferencial

$$\frac{1}{2} \frac{d}{dM} \left(\frac{dw(x)}{dS(x)} \right) = -g(x) \quad a < x < b, \quad w(a) = 0, \quad w(b) = 0,$$

pues como ya habíamos mencionado v es un caso particular de w .

Integrando por primera vez,

$$\frac{dw(y)}{dS} = -2 \int_a^y g(z) dM(z) + k_1 = -2 \int_a^y g(z) m(z) dz + k_1.$$

Luego,

$$w(x) = -2 \int_a^x \int_a^y g(z) m(z) dz dS(y) + k_1(S(x) - S(a)) + k_2,$$

sujeto a $w(a) = 0 = w(b)$ implica que $k_2 = 0$ y $k_1 = \frac{2}{S(b)-S(a)} \int_a^b \int_a^y g(z)m(z)dzdS(y)$.

Utilizando el hecho de que $u(x) = \frac{S(x)-S(a)}{S(b)-S(a)}$,

$$\begin{aligned}
w(x) &= 2 \left[u(x) \int_a^b \int_a^y g(z)m(z)dzdS(y) - \int_a^x \int_a^y g(z)m(z)dzdS(y) \right] \\
&= 2 \left[u(x) \int_a^b \int_z^b g(z)m(z)dS(y)dz - \int_a^x \int_z^x g(z)m(z)dS(y)dz \right] \\
&= 2 \left[u(x) \int_a^b (S(b) - S(z))g(z)m(z)dz - \int_a^x (S(x) - S(z))g(z)m(z)dz \right] \\
&= 2 \left[u(x) \int_x^b (S(b) - S(z))g(z)m(z)dz + u(x) \int_a^x (S(b) - S(z))g(z)m(z)dz \right. \\
&\quad \left. - \int_a^x (S(x) - S(z))g(z)m(z)dz \right] \\
&= 2 \left[u(x) \int_x^b (S(b) - S(z))g(z)m(z)dz + \int_a^x \left(u(x)(S(b) - S(z)) - (S(x) - S(z)) \right) g(z)m(z)dz \right] \\
&= 2 \left[u(x) \int_x^b (S(b) - S(z))g(z)m(z)dz + \int_a^x \frac{S(b) - S(x)}{S(b) - S(a)} (S(z) - S(a))g(z)m(z)dz \right] \\
&= 2 \left[u(x) \int_x^b (S(b) - S(z))g(z)m(z)dz + (1 - u(x)) \int_a^x (S(z) - S(a))g(z)m(z)dz \right]. \tag{1.27}
\end{aligned}$$

Con este resultado obtenemos $w(x)$ y como consecuencia con $g(x) = 1$ encontramos $v(x)$, la esperanza del tiempo en que el proceso llega a los estados a o b .

Por otro lado, consideremos un proceso con escala natural o canónica, es decir, con función de escala $S(x) = x$ y como consecuencia $s(x) = 1$. Con la expresión anterior queremos calcular el tiempo medio de salida del intervalo $(x - \varepsilon, x + \varepsilon)$ si inicialmente el proceso se encuentra en x , cabe señalar que

$$u(x) = \frac{S(x) - S(x - \varepsilon)}{S(x + \varepsilon) - S(x - \varepsilon)} = \frac{x - (x - \varepsilon)}{x + \varepsilon - (x - \varepsilon)} = \frac{1}{2},$$

lo cual tiene sentido ya que el intervalo es simétrico, entonces,

$$\mathbb{E}[\tau_{x-\varepsilon, x+\varepsilon} | X_0 = x] = \int_x^{x+\varepsilon} (x + \varepsilon - z)m(z)dz + \int_{x-\varepsilon}^x (z - x + \varepsilon)m(z)dz.$$

Resolviendo ambas integrales por partes,

$$\int_x^{x+\varepsilon} (x + \varepsilon - z)m(z)dz = -\varepsilon M(x) + \int_x^{x+\varepsilon} M(z)dz$$

y

$$\int_{x-\varepsilon}^x (z - x + \varepsilon)m(z)dz = \varepsilon M(x) + \int_{x-\varepsilon}^x M(z)dz,$$

aplicando la aproximación de series de Taylor a $f(\varepsilon) = \int_x^{x+\varepsilon} M(z)dz$ y $f(\varepsilon) = \int_{x-\varepsilon}^x M(z)dz$ alrededor del cero, $f(\varepsilon) = f(0) + f'(0)\varepsilon + \frac{1}{2}f''(0)\varepsilon^2 + \mathcal{O}(\varepsilon^2)$ donde, en ambos casos, $f(0) = 0$, $f'(0) = M(x)$ y $f''(0) = m(x)$, por lo tanto,

$$\int_x^{x+\varepsilon} (x + \varepsilon - z)m(z)dz = \frac{1}{2}m(x)\varepsilon^2 + \mathcal{O}(\varepsilon^2)$$

y

$$\int_{x-\varepsilon}^x (z-x+\varepsilon)m(z)dz = \frac{1}{2}m(x)\varepsilon^2 + \mathcal{O}(\varepsilon^2).$$

Entonces,

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} \mathbb{E}[\tau_{x-\varepsilon, x+\varepsilon} | X_0 = x] &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} \left[\int_x^{x+\varepsilon} (x+\varepsilon-z)m(z)dz + \int_{x-\varepsilon}^x (z-x+\varepsilon)m(z)dz \right] \\ &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} [m(x)\varepsilon^2 + \mathcal{O}(\varepsilon^2)] \\ &= m(x). \end{aligned}$$

Por consiguiente, la función de velocidad $m(x)$ puede construirse como la velocidad a la cual el reloj del proceso corre cuando se encuentra en el estado x . Equivalentemente, si consideramos el reloj del movimiento browniano estándar como estándar y el proceso se encuentra en escala natural, la cantidad $m(x)\varepsilon^2$ es del orden del tiempo esperado que el proceso pasa en el intervalo $(x-\varepsilon, x+\varepsilon)$ dado $X_0 = x$ antes de salir del mismo.

Además del generador infinitesimal existen otras ecuaciones diferenciales asociadas a las difusiones.

1.3.6. Ecuaciones de Kolmogorov

Las ecuaciones de Kolmogorov llamadas ecuaciones hacia atrás y hacia adelante son herramientas muy útiles para analizar procesos de Markov, por ejemplo una sus principales aplicaciones es encontrar distribuciones límite. Si se requiere mayor detalle se puede consultar [17].

En primer lugar, vamos a hacer el desarrollo para encontrar la ecuación hacia atrás. Sea $\{X_t\}$ una difusión regular homogénea en el tiempo en el intervalo (l, r) . Definimos a la distribución de transición de X_t como $P(x, y; t) = \mathbb{P}(X_t \leq y | X_0 = x)$, sujeto a la distribución inicial

$$P(x, y; 0) = \begin{cases} 1 & \text{si } x \leq y \\ 0 & \text{si } x > y, \end{cases} \quad (1.28)$$

es decir, la distribución se concentra en el punto x . Supongamos que $\frac{dP}{dy} = p(x, y; t)$ para $t > 0$ es una densidad continua en (l, r) . Queremos encontrar una ecuación diferencial parcial para la función

$$u(x, t) = \mathbb{E}[g(X_t) | X_0 = x], \quad (1.29)$$

donde $g(x)$ es un función continua y acotada en (l, r) .

Proposición 1.3.2. *Supongamos que $u(x, t)$ es diferenciable en t y con segunda derivada continua en x , entonces $u(x, t)$ satisface*

$$\frac{\partial u}{\partial t} = \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} + \mu(x)\frac{\partial u}{\partial x} \quad (1.30)$$

con condición inicial $\lim_{h \downarrow 0} u(x, h) = g(x)$.

Demostración. Supongamos que u es una función con las propiedades ya mencionadas, notemos que

$$u(x, t) = \mathbb{E}[g(X_{t+h}) | X_h = x] \quad \text{y} \quad u(X_h, t) = \mathbb{E}[g(X_{t+h}) | X_h].$$

Entonces,

$$\begin{aligned} u(x, t+h) &= \mathbb{E}[g(X_{t+h}) | X_0 = x] \\ &= \mathbb{E}\left[\mathbb{E}[g(X_{t+h}) | X_h] \mid X_0 = x\right] \\ &= \mathbb{E}[u(X_h, t) | X_0 = x]. \end{aligned} \quad (1.31)$$

Restando de ambos lados $u(x, t)$, dividimos entre h y tomamos límite cuando h tiende a cero.

$$\frac{\partial}{\partial t} u(x, t) = \lim_{h \rightarrow 0} \frac{u(x, t+h) - u(x, t)}{h} = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E}[u(X_h, t) - u(x, t) | X_0 = x],$$

por tanto, basta encontrar el límite de la derecha.

Bajo la condición $X_0 = x$, denotamos por $\Delta X = X_h - x$ y definimos

$$\Delta X_\varepsilon = \begin{cases} \Delta X & \text{si } |\Delta X| \leq \varepsilon \\ 0 & \text{si } |\Delta X| > \varepsilon. \end{cases}$$

Por otro lado, como g es acotada, digamos $|g(x)| \leq M$ para toda x , y M una constante, se tiene que

$$|u(x, t)| = \left| \mathbb{E}[g(X_t) | X_0 = x] \right| \leq \mathbb{E}[|g(X_t)| | X_0 = x] \leq M.$$

Utilizando la definición de ΔX_ε notemos que

$$u(x + \Delta X, t) - u(x + \Delta X_\varepsilon, t) = \begin{cases} 0 & \text{si } |\Delta X| \leq \varepsilon \\ u(x + \Delta X, t) - u(x, t) & \text{si } |\Delta X| > \varepsilon. \end{cases}$$

Luego,

$$\begin{aligned} \mathbb{E} \left[|u(x + \Delta X, t) - u(x + \Delta X_\varepsilon, t)| \right] &= \mathbb{E} \left[|u(x + \Delta X, t) - u(x + \Delta X_\varepsilon, t)| \mathbb{I}_{(|\Delta X| \leq \varepsilon)} \right] \\ &\quad + \mathbb{E} \left[|u(x + \Delta X, t) - u(x + \Delta X_\varepsilon, t)| \mathbb{I}_{(|\Delta X| > \varepsilon)} \right] \\ &= \mathbb{E} \left[|u(x + \Delta X, t) - u(x + \Delta X_\varepsilon, t)| \mathbb{I}_{(|\Delta X| > \varepsilon)} \right] \\ &\leq 2M \mathbb{P}(|\Delta X| > \varepsilon | X_0 = x) \\ &= \mathcal{O}(h). \end{aligned}$$

Así, para obtener el resultado, incorporamos ΔX_ε mediante $u(x + \Delta X_\varepsilon, t)$ y con las expresiones anteriores,

$$\frac{1}{h} \mathbb{E}[u(X_h, t) - u(x, t) | X_0 = x] = \frac{1}{h} \mathbb{E}[u(x + \Delta X_\varepsilon, t) - u(x, t) | X_0 = x] + \frac{\mathcal{O}(h)}{h}.$$

Para $u(X_h, t)$ utilizamos la expansión en series de Taylor,

$$u(x + \Delta X_\varepsilon, t) = u(x, t) + \Delta X_\varepsilon \frac{\partial}{\partial x} u(x, t) + \frac{1}{2} (\Delta X_\varepsilon)^2 \frac{\partial^2}{\partial x^2} u(x + \xi, t),$$

donde ξ es una variable aleatoria tal que $|\xi| \leq |\Delta X_\varepsilon| \leq \varepsilon$.

Por otro lado, sea $\delta > 0$, como u tiene segunda derivada continua, podemos elegir ε suficientemente pequeña tal que

$$\left| \frac{\partial^2}{\partial x^2} u(x + \xi, t) - \frac{\partial^2}{\partial x^2} u(x, t) \right| < \delta.$$

Entonces, de la expansión en series,

$$\begin{aligned} \frac{1}{h} \mathbb{E}[u(x + \Delta X_\varepsilon, t) - u(x, t) | X_0 = x] &= \frac{1}{h} \mathbb{E} \left[\Delta X_\varepsilon \frac{\partial}{\partial x} u(x, t) + \frac{1}{2} (\Delta X_\varepsilon)^2 \frac{\partial^2}{\partial x^2} u(x + \xi, t) \mid X_0 = x \right] \\ &= \mu(x) \frac{\partial}{\partial x} u(x, t) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} u(x + \xi, t). \end{aligned}$$

Ahora, notemos que sustituyendo la última expresión y la cota anterior tenemos que

$$\lim_{h \downarrow 0} \left| \mathbb{E}[u(x + \Delta X_\varepsilon, t) - u(x, t) | X_0 = x] - \mu(x) \frac{\partial u(x, t)}{\partial x} + \frac{1}{2} \sigma^2(x) \frac{\partial^2 u(x, t)}{\partial x^2} \right| \leq \frac{1}{2} \sigma^2(x) \delta,$$

como consecuencia, por ser δ arbitrario, se tiene la igualdad.

Finalmente,

$$\begin{aligned}
\frac{\partial}{\partial t} u(x, t) &= \lim_{h \downarrow 0} \mathbb{E}[u(X_h, t) - u(x, t) | X_0 = x] \\
&= \lim_{h \downarrow 0} \mathbb{E}[u(x + \Delta X_\varepsilon, t) - u(x, t) | X_0 = x] + \frac{\mathcal{O}(h)}{h} \\
&= \mu(x) \frac{\partial}{\partial x} u(x, t) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} u(x, t).
\end{aligned} \tag{1.32}$$

□

A esta ecuación se le llama *ecuación hacia atrás de Kolmogorov*.

Notemos que la ecuación (1.32) puede escribirse en términos del generador como

$$\frac{\partial}{\partial t} u(x, t) = \mathcal{A}u(x, t)$$

con condición inicial $\lim_{t \downarrow 0} u(x, t) = g(x)$.

Además, si utilizamos $g(z) = 1$ para $z \leq y$ y $g(z) = 0$ en cualquier otro caso, podemos considerar $u(x, t) = P(x, y; t)$, la distribución de transición entonces,

$$\frac{\partial}{\partial t} P(x, y; t) = \mu(x) \frac{\partial}{\partial x} P(x, y; t) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} P(x, y; t), \tag{1.33}$$

es la ecuación hacia a atrás con condición inicial correspondiente a la ecuación (1.28). Asimismo, la densidad de transición satisface la ecuación hacia atrás de Kolmogorov, es decir,

$$\frac{\partial}{\partial t} p(x, y; t) = \mu(x) \frac{\partial}{\partial x} p(x, y; t) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} p(x, y; t). \tag{1.34}$$

Cabe mencionar que las ecuaciones (1.33) y (1.34) no necesariamente tienen una única solución.

A continuación, describiremos otra importante ecuación diferencial parcial que satisface la densidad de transición $p(x, y; t)$. Consideremos una función “suave” $\varphi(y, t)$ que satisfaga

$$\varphi(y, t + s) = \int \varphi(z, t) p(z, y; s) dz,$$

para todo $s, t > 0$. Derivando ambos lados con respecto a s tenemos,

$$\begin{aligned}
\frac{\partial}{\partial t} \varphi(y, t + s) &= \frac{\partial}{\partial s} \varphi(y, t + s) \\
&= \int \varphi(z, t) \frac{\partial}{\partial s} p(z, y; s) dz \\
&= \int \varphi(z, t) \left(\mu(z) \frac{\partial}{\partial z} p(z, y; s) + \frac{1}{2} \sigma^2(z) \frac{\partial^2}{\partial z^2} p(z, y; s) \right) dz;
\end{aligned}$$

la última igualdad se da al sustituir la ecuación hacia atrás para la densidad de transición. Ahora, después de distribuir φ , resolvemos la integral utilizando integración por partes,

$$\int \varphi(z, t) \mu(z) \frac{\partial}{\partial z} p(z, y; s) dz = \varphi(x, t) \mu(z) p(z, x; s) - \int p(z, x; s) \frac{\partial}{\partial z} \mu(z) \varphi(z, t) dz$$

y

$$\begin{aligned}
\int \varphi(z, t) \sigma^2(z) \frac{\partial^2}{\partial z^2} p(z, y; s) dz &= \varphi(z, t) \sigma^2(z) \frac{\partial}{\partial z} p(z, y; s) - p(z, y; s) \frac{\partial}{\partial z} \varphi(z, t) \sigma^2(z) \\
&\quad + \int p(z, y; s) \frac{\partial^2}{\partial z^2} \varphi(z, t) \sigma^2(z) dz.
\end{aligned}$$

Suponiendo que los términos que no están dentro de la integral se cancelan al evaluarlos en los límites, la derivada parcial de u con respecto a t se simplifica,

$$\frac{\partial}{\partial t}\varphi(y, t + s) = \int \left(\frac{1}{2} \frac{\partial^2}{\partial z^2} \varphi(z, t) \sigma^2(z) - \frac{\partial}{\partial z} \varphi(z, t) \mu(z) \right) p(z, y; s) dz.$$

Finalmente, haciendo s tender a cero, $p(z, y; s)$ converge a una medida degenerada concentrada en y ,

$$\frac{\partial}{\partial t}\varphi(y, t + s) = \frac{1}{2} \frac{\partial^2}{\partial y^2} \varphi(y, t) \sigma^2(y) - \frac{\partial}{\partial y} \varphi(y, t) \mu(y). \quad (1.35)$$

A la ecuación (1.35) se le llama *ecuación hacia adelante*. Notemos que la densidad de transición satisface

$$p(z, y; t + s) = \int p(z, x; t) p(x, y; s) dx,$$

pues es la propiedad de Markov para difusiones, por lo tanto, bajo ciertas condiciones cumple la ecuación hacia adelante.

Las ecuaciones de Kolmogorov no necesariamente tienen una única distribución, pues no reflejan el comportamiento de las fronteras del proceso.

1.3.7. Distribución estacionaria

Una distribución estacionaria $\pi(x)$, si existe, satisface

$$\pi(y) = \int \pi(x) p(x, y; t) dx \quad \text{para } t > 0. \quad (1.36)$$

Proposición 1.3.3. *La distribución estacionaria π satisface*

$$\frac{1}{2} \frac{\partial^2 \sigma^2(y) \pi(y)}{\partial y^2} - \frac{\partial \mu(y) \pi(y)}{\partial y} = 0. \quad (1.37)$$

La prueba se hace de manera análoga a la construcción de la ecuación hacia adelante para la función π .

Existe una forma sencilla de encontrar la distribución estacionaria para difusiones mediante la función de escala y de velocidad.

Proposición 1.3.4. *La distribución estacionaria satisface*

$$\pi(x) = m(x)(k_1 S(x) + k_2), \quad (1.38)$$

donde, $m(x)$ es la función de densidad de velocidad dada en la ecuación (1.22) y $S(x)$ la función de escala dada en la ecuación (1.20); k_1 y k_2 son las constantes que garantizan que $\pi(x) > 0$ y $\int_l^r \pi(y) dy = 1$.

Demostración. Por la Proposición 1.3.3 sabemos que $\pi(x)$ satisface la ecuación diferencial parcial

$$\frac{1}{2} \frac{\partial^2}{\partial y^2} \sigma^2(y) \pi(y) - \frac{\partial}{\partial y} \mu(y) \pi(y) = 0,$$

al integrar obtenemos por el teorema fundamental de cálculo

$$\frac{1}{2} \frac{\partial}{\partial y} \sigma^2(y) \pi(y) - \mu(y) \pi(y) = k_0,$$

utilizamos el factor integrante $s(y) = \exp\left\{-\int^y \frac{2\mu(x)}{\sigma^2(x)} dx\right\}$ con derivada $s'(y) = -s(y) \frac{2\mu(y)}{\sigma^2(y)}$,

$$\begin{aligned} k_0 s(y) &= s(y) \left[\frac{1}{2} \frac{\partial}{\partial y} \sigma^2(y) \pi(y) - \mu(y) \pi(y) \right] \\ &= \frac{1}{2} \left[s(y) \frac{\partial}{\partial y} \sigma^2(y) \pi(y) - s(y) \frac{2\mu(y)}{\sigma^2(y)} \pi(y) \sigma^2(y) \right] \\ &= \frac{1}{2} \left[\frac{\partial}{\partial y} s(y) \sigma^2(y) \pi(y) \right], \end{aligned}$$

por lo tanto, considerando $k_1 = 2k_0$, tenemos $k_1 s(x) = \frac{\partial}{\partial y} s(x) \sigma^2(x) \pi(x)$, integrando nuevamente

$$k_1 S(x) = s(x) \sigma^2(x) \pi(x) + k_2.$$

Así, despejamos $\pi(x)$

$$\pi(x) = \frac{k_1 S(x)}{s(x) \sigma^2(x)} - \frac{k_2}{s(x) \sigma^2(x)} = m(x)(k_1 S(x) + k_2).$$

□

Dar condiciones sobre la existencia de la distribución estacionaria nos lleva a estudiar el comportamiento del proceso cerca de sus fronteras. En la literatura, por ejemplo [17], se encuentran dos clasificaciones, la primera dada por Feller y otros probabilistas americanos, que es la que definiremos más adelante y la otra, propuesta por probabilistas rusos.

1.3.8. Clasificación de las fronteras de un proceso regular

Consideremos $\{X_t\}$ una difusión regular con espacio de estados $E = (l, r)$ y coeficientes infinitesimales $\mu(x)$ y $\sigma^2(x) > 0$ para cualquier $x \in (l, r)$. A continuación realizaremos el análisis para la frontera izquierda l , ya que la frontera derecha r se estudia de manera análoga.

En primer lugar, definimos la medida de escala para intervalos cerrados $J = [a, b] \subset (l, r)$

$$S[J] = S[a, b] = S(b) - S(a),$$

donde $S(x)$ es la función de escala. Utilizaremos indistintamente $dS(x) = S[dx]$ para intervalos infinitesimales $[x, x + dx]$, notemos que $S[dx] = S(x + dx) - S(x) = dS(x) = s(x)dx$. Así, considerando la definición de la función de escala,

$$S[a, b] = S(b) - S(a) = \int_a^b s(x)dx - \int_a^a s(x)dx = \int_a^b s(x)dx.$$

Como consecuencia, para $l < a < x < b < r$,

$$0 \leq S[a, b] < \infty \quad \text{y} \quad S[a, b] = S[a, x] - S[x, b].$$

De manera análoga introducimos la medida de velocidad para intervalos cerrados

$$M[J] = M[a, b] = \int_a^b m(x)dx,$$

que también es positiva y finita.

Entonces, podemos escribir (1.13) y (1.15) en términos de estas medidas, como

$$u(x) = u_{a,b}(x) = \frac{S[a, x]}{S[a, b]} \quad l < a < x < b < r \quad (1.39)$$

y

$$v(x) = v_{a,b}(x) = 2 \left[u(x) \int_x^b S[z, b] dM(z) + (1 - u(x)) \int_a^x S[a, z] dM(z) \right]. \quad (1.40)$$

Ya que la medida S es no negativa y monótona en a para un valor b fijo, entonces definimos

$$S(l, b) = \lim_{a \downarrow l} S[a, b] \leq \infty, \quad l < b < r. \quad (1.41)$$

Como consecuencia tenemos que $S(l, b) = \infty$ para alguna $b \in (l, r)$ si y solo si $S(l, b) = \infty$ para toda $b \in (l, r)$. Para ver este hecho consideremos $b \in (l, r)$ tal que $S(l, b) = \infty$ y sabemos que $S[a, b] = S[a, x] + S[x, b]$, entonces

$$\begin{aligned} \lim_{a \downarrow l} S[a, b] &= \lim_{a \downarrow l} S[a, x] + \lim_{a \downarrow l} S[x, b] \\ S(l, b) &= S(l, x) + S[x, b]. \end{aligned}$$

Por lo tanto, como $S[x, b]$ es finito y $S(l, b) = \infty$ entonces $S(l, x) = \infty$ para cualquier $l < x < b < r$, con lo que se sigue el resultado.

Adicionalmente, definimos $\tau_{l+} = \lim_{a \downarrow l} \tau_a$, donde τ_a es el tiempo de llegada al estado a . Observamos que $\tau_{l+} \leq \infty$ y queremos probar que $\tau_{l+} = \tau_l$.

Debido a que para cada trayectoria que inicia en $X_0 = x$, $x \in (a, b)$, τ_a es una función monótona decreciente de a , se tiene que $\tau_a \leq \tau_l$, luego, $\tau_{l+} = \lim_{a \downarrow l} \tau_a \leq \tau_l$. Así, si $\tau_{l+} = \infty$ implica que $\tau_l = \infty$ y se da la igualdad.

Supongamos entonces que $\tau_{l+} < \infty$ y como $X_{\tau_{l+}} = \lim_{a \downarrow l} X_{\tau_a} = \lim_{a \downarrow l} a = l$ y además como τ_l está definido como un ínfimo se tiene que $\tau_{l+} \geq \tau_l = \inf\{t \geq 0 : X_t = l\}$, por lo tanto τ_{l+} y τ_l son iguales.

Notemos que τ_l está definido incluso cuando l no es un elemento del espacio de estados, esto sucede cuando $\tau_l = \infty$.

Lema 1.3.3. (a) Si $S(l, x) < \infty$ para algún $x \in (l, r)$, entonces

$$\mathbb{P}(\tau_{l+} \leq \tau_b | X_0 = x) > 0 \quad \text{para toda } l < x < b < r.$$

(b) Si $S(l, x) = \infty$ para algún $x \in (l, r)$, entonces

$$\mathbb{P}(\tau_{l+} < \tau_b | X_0 = x) = 0 \quad \text{para toda } l < x < b < r.$$

Demostración. (a) Por la ecuación (1.39)

$$\mathbb{P}(\tau_{l+} \leq \tau_b | X_0 = x) = 1 - \mathbb{P}(\tau_{l+} > \tau_b | X_0 = x) = 1 - \lim_{a \downarrow l} \frac{S[a, x]}{S[a, b]} = 1 - \frac{S(l, x)}{S(l, b)}.$$

Por hipótesis $S(l, x) < \infty$, lo que implica que $S(l, b) < \infty$ y además $S(l, x) < S(l, b)$, luego, el cociente $\frac{S(l, x)}{S(l, b)}$ es menor a uno y, por lo tanto, $\mathbb{P}(\tau_{l+} \leq \tau_b | X_0 = x) > 0$.

(b) Nuevamente,

$$\mathbb{P}(\tau_{l+} < \tau_b | X_0 = x) = \lim_{a \downarrow l} \frac{S[x, b]}{S[a, b]} = \frac{S[x, b]}{S(l, b)} = 0.$$

debido a que por hipótesis $S(l, x) = \infty$ para algún x . □

Observemos que si $a \neq b$, $\tau_a = \tau_b$ no puede ocurrir ya que una trayectoria continua no puede estar al mismo tiempo en dos puntos. Pero $\tau_{l+} = \tau_b$ puede ser cierto en el caso en que ambos sean infinito. Para aquellas trayectorias que inician en $x < b$ y nunca alcanzan el estado b , por convención $\tau_b = \infty$. Por su parte, mientras que la deriva del proceso tiende a $l = -\infty$, $\tau_{l+} = \infty$ en un sentido posiblemente diferente.

El Lema 1.3.3 induce la siguiente definición.

Definición 1.3.8. Se dice que la frontera l es atrayente si $S(l, x) < \infty$ para cualquier $x \in (l, r)$.

Ahora estamos interesados en el tiempo promedio en que la difusión alcanza la frontera atrayente antes de un nivel b . Evaluando la expresión (1.40)

$$\lim_{a \downarrow l} \mathbb{E}[\tau_a \wedge \tau_b | X_0 = x] = \lim_{a \downarrow l} \left[\frac{S[a, x]}{S[a, b]} \int_x^b S[z, b] dM(z) + \frac{S[x, b]}{S[a, b]} \int_a^x S[a, z] dM(z) \right].$$

Como l es atrayente, tenemos que tanto $\frac{S[a, x]}{S[a, b]}$ como $\frac{S[x, b]}{S[a, b]}$ son finitas y positivas, por lo que

$$\lim_{a \downarrow l} \mathbb{E}[\tau_a \wedge \tau_b | X_0 = x] < \infty \quad \text{si y solo si} \quad \lim_{a \downarrow l} \int_a^x S[a, z] dM(z) < \infty. \quad (1.42)$$

De esta manera, definimos

$$\Sigma(l) = \lim_{a \downarrow l} \int_a^x S[a, z] dM(z). \quad (1.43)$$

Podemos pensar a $\Sigma(l)$ como la medida del tiempo que toma alcanzar la barrera l o un estado b iniciando en un estado interior $x < b$.

Además, $\Sigma(l)$ puede reescribirse como

$$\begin{aligned} \Sigma(l) &= \lim_{a \downarrow l} \int_a^x S[a, z] dM(z) \\ &= \int_l^x S(l, z] dM(z) \\ &= \int_l^x \int_l^z s(y) dym(z) dz \\ &= \int_l^x \int_y^x m(z) dz s(y) dy \\ &= \int_l^x M[y, x] dS(y). \end{aligned}$$

Definición 1.3.9. Decimos que una frontera l es alcanzable si $\Sigma(l) < \infty$, por el contrario, l no es alcanzable si $\Sigma(l) = \infty$.

No es difícil ver que $S(l, x] < \infty$ si $\Sigma(l) < \infty$, entonces si una frontera es alcanzable es atrayente, en cambio, una frontera no alcanzable puede o no ser atrayente.

El siguiente lema muestra que se puede llegar a una frontera alcanzable en un tiempo finito con probabilidad positiva.

Lema 1.3.4. Sea l una frontera atrayente, supongamos que $l < x < b < r$, entonces las siguientes afirmaciones son equivalentes:

- (a) $\mathbb{P}(\tau_l < \infty | X_0 = x) > 0$.
- (b) $\mathbb{E}[\tau_l \wedge \tau_b | X_0 = x] < \infty$.
- (c) $\Sigma(l) = \int_l^x S(l, z] dM(z) < \infty$.

Demostración. Ya probamos que (b) ocurre si y solo si ocurre (c).

Para ver que (b) implica (a) notemos que si $\mathbb{E}[\tau_l \wedge \tau_b | X_0 = x] < \infty$, entonces, dado que el proceso inicia en x , $\tau_l \wedge \tau_b$ es finito con probabilidad uno. Luego, por hipótesis y la equivalencia (1.42), l es alcanzable, entonces $S(l, x] < \infty$, se sigue del Lema 1.3.3 que $\mathbb{P}(\tau_l \leq \tau_b | X_0 = x) > 0$ y como $\tau_l \neq \tau_b$ tenemos que $\mathbb{P}(\tau_l < \tau_b | X_0 = x) > 0$. Finalmente, $\mathbb{P}(\tau_l < \infty | X_0 = x) \geq \mathbb{P}(\tau_l < \tau_b \leq \infty | X_0 = x) > 0$.

Por último, vamos a probar que (a) implica (c), supongamos que $\mathbb{P}(\tau_l < \infty | X_0 = x) > 0$, entonces existe $t > 0$ para el cual $\mathbb{P}(\tau_l < t | X_0 = x) = \alpha > 0$. Notemos que cada trayectoria que inicia en x y alcanza la frontera l antes del tiempo t visita cada estado $z \in (l, x]$ y posteriormente viaja de z a l . Este hecho lo podemos escribir como $\tau_l = \tau_z + \tau_l^+(z)$ donde $\tau_l^+(z)$ representa la duración de la primera visita de z a la primera visita de l , así que $\tau_l^+(z) \leq \tau_l$. De donde para cualquier $z \in (l, x]$,

$$\begin{aligned} 0 < \alpha &\leq \mathbb{P}(\tau_l^+(z) \leq t) \\ &= \mathbb{P}(\tau_l \leq t | X_0 = z) \\ &\leq \mathbb{P}(\tau_l \wedge \tau_x \leq t | X_0 = z). \end{aligned}$$

De aquí se tiene que

$$\sup_{z \in (l, x]} \mathbb{P}(\tau_l \wedge \tau_x > t | X_0 = z) \leq 1 - \alpha < 1,$$

por inducción, $\sup_{z \in (l, x]} \mathbb{P}(\tau_l \wedge \tau_x > nt | X_0 = z) \leq (1 - \alpha)^n < 1$ para $n \geq 1$, ya que por la propiedad fuerte de Markov,

$$\begin{aligned} \sup_{z \in (l, x]} \mathbb{P}(\tau_l \wedge \tau_x > nt | X_0 = z) &= \sup_{z \in (l, x]} \mathbb{P}(\tau_l \wedge \tau_x > nt | \tau_l \wedge \tau_x > (n-1)t, X_0 = z) \\ &\quad \times \mathbb{P}(\tau_l \wedge \tau_x > (n-1)t | X_0 = z) \\ &\leq \mathbb{P}(\tau_l \wedge \tau_x > t | X_0 = z) (1 - \alpha)^{n-1} \\ &\leq (1 - \alpha)^n. \end{aligned}$$

Entonces, $\mathbb{E}[\tau_l \wedge \tau_x | X_0 = z] \leq \frac{t}{\alpha} < \infty$. Pero, por la equivalencia entre (b) y (c), tenemos que $\int_l^z S(l, y) dM(y) < \infty$.

Finalmente, como $\int_z^x S(l, y) dM(y) < \infty$

$$\int_l^x S(l, y) dM(y) = \int_l^z S(l, y) dM(y) + \int_z^x S(l, y) dM(y) < \infty.$$

□

Para continuar con la clasificación de las fronteras definimos

$$M(l, x] = \lim_{a \downarrow l} M[a, x]$$

y

$$N(l) = \int_l^x S[y, x] dM(y) = \int_l^x M(l, z] dS(z).$$

Entonces, $M(l, x]$ mide la velocidad del proceso cerca de l y $N(l)$ puede interpretarse como la medida del tiempo que le toma a la difusión alcanzar un punto $x \in (l, r)$ empezando en la frontera l .

La clasificación moderna del comportamiento de las fronteras depende del valor (si son finitos o no) de los cuatro funcionales que acabamos de definir, $S(l, x]$, $\Sigma(l)$, $N(l)$ y $M(l, x]$. Recapitulando, a grandes rasgos, $\Sigma(l)$ mide el tiempo que el proceso tarda en alcanzar la frontera l o un punto interior b iniciando en otro punto interior $x < b$ y $N(l)$ el tiempo que toma alcanzar un punto interior desde la frontera l . El siguiente lema permite observar la relación que existe entre estos funcionales.

Lema 1.3.5. *Se tienen las siguientes relaciones:*

- (a) $S(l, x] = \infty$ implica $\Sigma(l) = \infty$.
- (b) $\Sigma(l) < \infty$ implica $S(l, x] < \infty$.
- (c) $M(l, x] = \infty$ implica $N(l) = \infty$.
- (d) $N(l) < \infty$ implica $M(l, x] < \infty$.
- (e) $\Sigma(l) + N(l) = S(l, x]M(l, x]$.

Demostración. Primero notemos que (b) es la contrapositiva de (a) y (d) es la contrapositiva de (c), por lo que bastará probar (a), (c), (e).

Para el primer inciso, como $S(l, x] = \infty$ entonces $S(l, z] = \infty$ para toda $z \in (l, r)$, luego

$$\Sigma(l) = \int_l^x S(l, z] dM(z) = \infty,$$

pues M es una medida estrictamente positiva.

Análogamente para (c) $M(l, x] = \infty$ implica que $M(l, z] = \infty$ para toda $z \in (l, r)$ y

$$N(l) = \int_l^x M(l, z) dS(x) = \infty$$

ya que S es una medida estrictamente positiva.

Finalmente,

$$\begin{aligned}
\Sigma(l) + N(l) &= \int_l^x S(l, z) dM(z) + \int_l^x S[z, x] dM(z) \\
&= \int_l^x (S(l, z] + S[z, x]) dM(z) \\
&= S(l, x] \int_l^x dM(z) \\
&= S(l, x] M(l, x].
\end{aligned}$$

□

De acuerdo al Lema 1.3.5 de las 16 combinaciones que se obtienen respecto a si $S(l, x]$, $M(l, x]$, $\Sigma(l)$, y $N(l)$ son finitas o infinitas, solo son posibles seis de ellas, las cuales se muestran en la Tabla 1.1.

	$S(l, x]$	$M(l, x]$	$\Sigma(l)$	$N(l)$
1	$< \infty$	$< \infty$	$< \infty$	$< \infty$
2	$< \infty$	$= \infty$	$< \infty$	$= \infty$
3	$< \infty$	$= \infty$	$= \infty$	$= \infty$
4	$= \infty$	$< \infty$	$= \infty$	$< \infty$
5	$= \infty$	$< \infty$	$= \infty$	$= \infty$
6	$= \infty$	$= \infty$	$= \infty$	$= \infty$

Tabla 1.1: Combinaciones posibles $S(l, x]$, $M(l, x]$, $\Sigma(l)$, $N(l)$. Tabla tomada de [17].

Para caracterizar completamente a un proceso es necesario conocer el comportamiento de sus fronteras y esto puede lograrse asignando la velocidad $M[\{l\}]$ a la frontera l . El comportamiento puede ir desde la absorción, cuando $M[\{l\}] = \infty$ o la reflexión $M[\{l\}] = 0$. Si la frontera tiene un comportamiento intermedio entre la absorción y la reflexión es una frontera pegajosa y $0 < M[\{l\}] < \infty$.

A partir de las combinaciones de la Tabla 1.1 enunciamos la clasificación de Feller.

- (a) Frontera regular. Una frontera l es regular si el proceso puede entrar y salir de la misma. Basta probar que $S(l, r] < \infty$ y $M(l, x] < \infty$ para determinar si la frontera l es regular.
- (b) Frontera de salida. Una frontera l es de salida si iniciando en l no es posible acceder a un estado interior b (aunque b esté muy cerca de l). Equivalentemente, podemos decir que una vez alcanzada la frontera l no existen trayectorias continuas que salgan de l , es decir, es un punto absorbente. Para determinar si una frontera es de salida es suficiente probar que $\Sigma(l) < \infty$ pero $M(l, x] = \infty$.
- (c) Frontera de entrada. Una frontera l es de entrada si no puede ser alcanzada desde el interior del espacio de estados, sin embargo, el proceso puede iniciar en l . Este proceso saldrá rápidamente de la frontera y no regresará a ella. Para probar que l es de entrada, es necesario verificar que $S(l, x] = \infty$ y $N(l) < \infty$.
- (d) Frontera natural. Una frontera l es natural si no puede ser alcanzada en un tiempo promedio finito ni es posible que la difusión inicie en l . Las fronteras naturales se omiten del espacio de estados, es decir, son de la forma (l, r) o $(l, r]$. Para determinar que l es natural, debe cumplirse que $\Sigma(l) = \infty$ y $N(l) = \infty$.

En la Tabla 1.2 mostramos la clasificación de las fronteras considerando los posibles resultados de los funcionales ya mencionados.

	$S(l, x]$	$M(l, x]$	$\Sigma(l)$	$N(l)$	Feller	Rusa	
1	$< \infty^*$	$< \infty^*$	$< \infty$	$< \infty$	Regular	Regular	Atrayente y alcanzable
2	$< \infty$	$= \infty^*$	$< \infty^*$	$= \infty$	Salida	Salida	Atrayente y alcanzable
3	$< \infty^*$	$= \infty^*$	$= \infty^*$	$= \infty$	Natural		Atrayente y No alcanzable
4	$= \infty^*$	$< \infty$	$= \infty$	$< \infty^*$	Entrada	Natural	No atrayente y No alcanzable
5	$= \infty^*$	$< \infty^*$	$= \infty$	$= \infty$	Natural	Natural	No atrayente y No alcanzable
6	$= \infty^*$	$= \infty^*$	$= \infty$	$= \infty$	Natural	Natural	No atrayente y No alcanzable

Tabla 1.2: Clasificación de las fronteras. El * indica las mínimas condiciones para establecer la clasificación en cada renglón. Tabla tomada de [17].

Para la frontera superior r se siguen criterios análogos definiendo

$$S[x, r) = \lim_{b \uparrow r} S[x, b], \quad M[x, r) = \lim_{b \uparrow r} M[x, b],$$

$$\Sigma(r) = \int_x^r M[x, z] dS(z) = \int_x^r S[y, r) dM(y)$$

y

$$N(r) = \int_x^r S[x, z] dM(z) = \int_x^r M[y, r) dS(y).$$

Para difusiones regulares en una dimensión, donde sus fronteras son de entrada, se espera que el proceso tenga una única distribución estacionaria en el límite. En la ecuación (1.38) tenemos que $\pi(x) = m(x)(k_1 S(x) + k_2)$, y como las fronteras l y r son de entrada, la función $S(x)$ es monótona dentro del intervalo (l, r) , crece a infinito cuando x tiende a r y decrece a menos infinito cuando x tiende a l . Entonces, para que π sea positiva siempre es necesario que $k_1 = 0$ y k_2 debe elegirse de tal manera que $\int_l^r \pi(x) dx = 1$, por lo que, la única distribución estacionaria satisface,

$$\pi(x) = \frac{m(x)}{\int_l^r m(z) dz}. \quad (1.44)$$

1.3.9. Descomposición espectral de la función de densidad de transición.

Ya que las difusiones son procesos de Markov sería de utilidad conocer la función de transición del proceso. Sin embargo, son contadas las difusiones en las que ésta es conocida, ya sea porque no tiene una expresión explícita, o bien, porque está expresada como series infinitas.

Hasta el momento solo sabemos que la densidad de transición cumple las ecuaciones hacia adelante y hacia atrás de Kolmogorov, en esta sección presentamos un método para encontrar a la transición a partir del generador infinitesimal.

Como ya habíamos mencionado, denotamos por $P(x, y; t) = \mathbb{P}(X_t \leq y | X_0 = x)$ a la distribución de transición y $p(x, y; t) = \frac{dP(x, y; t)}{dy}$ la densidad de transición.

Consideremos una difusión regular en el intervalo finito y cerrado $I = [l, r]$ y a la varianza infinitesimal, $\sigma^2(x)$, positiva y continua en el intervalo I , donde sus fronteras pueden ser de salida o reflejantes. Sea $f(x)$ una función continua y acotada en (l, r) , la función

$$u(x, t) = \mathbb{E}[f(X_t) | X_0 = x]$$

satisface la ecuación diferencial hacia atrás,

$$\frac{\partial u}{\partial t} = \frac{1}{2} \sigma^2(x) \frac{\partial^2 u}{\partial x^2} + \mu(x) \frac{\partial u}{\partial x}, \quad (1.45)$$

notemos que la parte derecha es el generador infinitesimal de la difusión, así por la Proposición 1.3.1

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{1}{m(x)} \frac{d}{dx} \left(\frac{1}{s(x)} \frac{du}{dx} \right), \quad (1.46)$$

con condiciones iniciales $u(x, 0) = f(x)$ y condiciones de frontera $u(l, t) = 0$ si l es de salida y $u'(l, t) = 0$ si l es reflejante, análogamente para la frontera r .

Para encontrar la solución a (1.45) vamos a suponer que $u(x, t)$ es de la forma

$$u(x, t) = c(t)\varphi(x)$$

y resolvemos mediante el método de separación de variables, entonces por la descomposición

$$\mathcal{A}u(x, t) = c(t)\mathcal{A}\varphi(x),$$

la ecuación que debemos resolver es

$$\varphi(x) \frac{\partial c(t)}{\partial t} = c(t)\mathcal{A}\varphi(x),$$

reordenando los términos,

$$\frac{c'(t)}{c(t)} = \frac{\mathcal{A}\varphi(x)}{\varphi(x)}.$$

Como el lado izquierdo solo depende de t y el derecho solo de x , la igualdad se da si ambas son iguales a una constante, llamada constante de separación, es decir,

$$\frac{c'(t)}{c(t)} = \frac{\mathcal{A}\varphi(x)}{\varphi(x)} = \lambda.$$

Con ello obtenemos las siguientes ecuaciones diferenciales ordinarias que se resuelven de forma independiente, tomando en cuenta las condiciones de frontera de $\varphi(x)$.

$$\begin{aligned} c'(t) + \lambda c(t) &= 0, \\ \mathcal{A}\varphi(x) + \lambda\varphi(x) &= 0. \end{aligned}$$

La solución a la primer ecuación es $c(t) = ke^{-\lambda t}$, donde k es una constante. Por otro lado, para la segunda ecuación, bajo ciertas condiciones de frontera, particularmente cuando I es acotado, existe una única sucesión, $\{\lambda_n\}$,

$$0 \leq \lambda_0 < \lambda_1 < \dots < \lambda_n < \dots, \quad \lambda_n \uparrow \infty,$$

llamados *valores propios o espectro del operador* \mathcal{A} y sus correspondientes funciones asociadas

$$\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x), \dots,$$

conocidas como *funciones propias*, tal que,

$$\mathcal{A}\varphi_n(x) = -\lambda_n\varphi_n(x).$$

Entonces tenemos soluciones a la primer ecuación de la forma $c_n(t) = k_n e^{-\lambda_n t}$ con k_n constante y por consiguiente, la solución a la ecuación (1.45) es $k_n e^{-\lambda_n t} \varphi_n$. Además, la sucesión $\{\varphi_n(x)\}$ es ortogonal con respecto a la densidad $m(x)$. Para probarlo veamos

$$\begin{aligned} -\lambda_j \int_l^r m(x) \varphi_i(x) \varphi_j(x) dx &= \int_l^r m(x) \varphi_i(x) \mathcal{A}\varphi_j(x) dx \\ &= \frac{1}{2} \int_l^r \varphi_i(x) \left(\frac{d}{dx} \left(\frac{1}{s(x)} \frac{d}{dx} \varphi_j(x) \right) \right) dx \\ &= -\frac{1}{2} \int_l^r \frac{1}{s(x)} \frac{d}{dx} \varphi_i(x) \frac{d}{dx} \varphi_j(x) dx \\ &= \frac{1}{2} \int_l^r \frac{d}{dx} \left(\frac{1}{s(x)} \frac{d}{dx} \varphi_i(x) \right) \varphi_j(x) dx \\ &= \int_l^r m(x) \mathcal{A}\varphi_i(x) \varphi_j(x) dx \\ &= -\lambda_i \int_l^r m(x) \varphi_i(x) \varphi_j(x) dx, \end{aligned}$$

donde la tercer y cuarta igualdad se dan integrando por partes sucesivamente y considerando las condiciones de frontera adecuadas. Por lo tanto, si $i \neq j$ entonces $\lambda_i \neq \lambda_j$, lo que implica

$$\int_l^r m(x)\varphi_i(x)\varphi_j(x)dx = 0,$$

y así se cumple el resultado.

Además, como la ecuación diferencial es lineal, una solución es

$$u(x, t) = \sum_{n=0}^{\infty} k_n e^{-\lambda_n t} \varphi_n(x), \quad (1.47)$$

donde la selección de $\{k_n\}$ garantiza la condición inicial $u(x, 0) = f(x)$. De hecho,

$$k_n = \frac{\int_l^r f(y)\varphi_n(y)m(y)dy}{\int_l^r \varphi_n^2(y)m(y)dy}.$$

Como estamos interesados en encontrar la función de densidad de transición, tomamos a f adecuadamente, es decir, $f(x) = \mathbb{I}_{(a,b)}$, luego,

$$u(x, t) = \int_a^b p(x, y; t)dy = \sum_{n=0}^{\infty} e^{-\lambda_n t} \varphi_n(x) \frac{\int_a^b \varphi_n(y)m(y)dy}{\int_l^r \varphi_n^2(y)m(y)dy},$$

así, dividiendo entre $(b - a)$ y reduciéndolo hasta y obtenemos la representación espectral para la función de transición,

$$p(x, y; t) = m(y) \sum_{n=0}^{\infty} e^{-\lambda_n t} \varphi_n(x)\varphi_n(y) \left(\int_l^r \varphi_n^2(y)m(y)dy \right)^{-1}. \quad (1.48)$$

Las condiciones de frontera adecuadas para el resultado implican que la función $u(x, t)$ sea cero en fronteras de salida, la derivada sea cero en fronteras reflejantes y $d/dS = 0$ para fronteras de entrada.

Por último, observamos que como λ_n tiende a infinito cuando n tiende a infinito, a partir de cierto valor los términos de la exponencial $e^{-\lambda_n t}$ se hacen despreciables.

Capítulo 2

Proceso de Wright Fisher

El modelo central de este trabajo es el proceso de Wright Fisher, su principal aplicación es en la genética de poblaciones y la evolución del tamaño de una población a través del tiempo que está afectada por diversas fuerzas evolutivas como la deriva genética, selección, mutación y migración.

En el presente capítulo desarrollamos la teoría del proceso de Wright Fisher basándonos en [17], en primer lugar, se construye como un proceso límite de una cadena de Markov, se encuentran algunas características del proceso como la clasificación de las fronteras, su distribución estacionaria y la descomposición espectral. Finalmente, presentamos al proceso coalescente, que estudia las líneas de descendencia de una población y está relacionado con el proceso de Wright-Fisher.

2.1. Proceso de Wright Fisher a tiempo discreto

Supongamos que tenemos una población de tamaño N que permanece constante a través del tiempo y está compuesta por dos tipos de individuos, digamos, A y a . Consideremos $\{X\}_{t \geq 0}$ un proceso estocástico que representa el número de individuos de tipo A al tiempo t (o bien, generación t), es decir, si $X_t = i$, $i = 1, \dots, N$ implica que hay i individuos de tipo A y $N - i$ de tipo a en la generación t . Además, suponemos que las generaciones no se traslapan.

En el modelo general, en la siguiente generación se produce un individuo bajo la influencia de diversos factores que pueden ser aleatorios, mutaciones o bien, selección.

En el caso más sencillo del modelo de Wright-Fisher cuando no hay mutación ni selección, se plantea que la siguiente generación se obtiene como una muestra aleatoria con reemplazo de individuos de la generación actual. En la Figura 2.1 mostramos un diagrama ejemplificando este hecho.

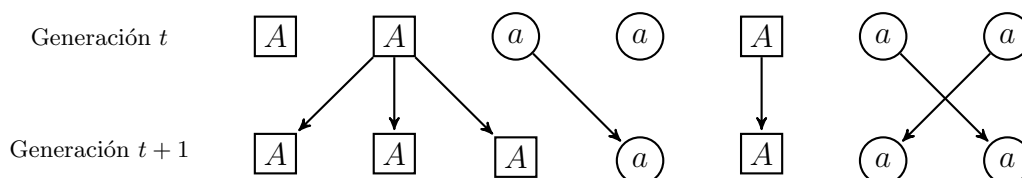


Figura 2.1: Ejemplo de la evolución en la composición de la población entre la generación t y $t+1$.

Supongamos, por ejemplo, que tenemos una urna con N bolas, de las cuales i son de tipo A y $N - i$ son de tipo a entonces para obtener a la generación $t+1$ elegimos al azar N bolas de la urna con reemplazo. Este proceso describe una distribución Binomial donde la probabilidad de producir un individuo tipo A es $p_i = \frac{i}{N}$.

Lo anterior también puede interpretarse como que cada individuo de la generación $t+1$ elige de manera aleatoria y uniforme a su padre (en la generación t), de manera que cada individuo de tipo A de la generación

t puede ser elegido N veces de forma independiente.

Podemos observar que si en la generación t todos los individuos son de tipo A entonces, no será posible obtener individuos de tipo a en las siguientes generaciones y por ende, ocurre la extinción. De igual forma ocurre si en la generación t solo hay individuos de tipo a .

A lo largo del presente trabajo vamos a considerar únicamente el efecto de la mutación. Asumimos entonces que un individuo a nace de uno tipo A con probabilidad γ y por el contrario, un individuo tipo A nace de un tipo a con probabilidad δ .

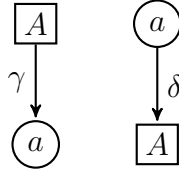


Figura 2.2: Probabilidades de mutación.

Dada una estructura de la población, es decir, si está compuesta de i individuos de tipo A y $N - i$ individuos de tipo a , la fracción esperada de individuos de tipo A después de la mutación es $\frac{i}{N}(1 - \gamma) + \left(\frac{N-i}{N}\right)\delta$, pensando en que no ocurra mutación en los i individuos de tipo A con probabilidad $(1 - \gamma)$ y del $(N - i)$ restante ocurra mutación a A con probabilidad δ . De manera análoga, la fracción esperada de individuos de tipo a es $\frac{i}{N}\gamma + \left(\frac{N-i}{N}\right)(1 - \delta)$, así la probabilidad de producir un individuo de tipo A es

$$p_i = \frac{i}{N}(1 - \gamma) + \left(1 - \frac{i}{N}\right)\delta.$$

Entonces el proceso $\{X_t\}_{t \geq 0}$ que representa el número de individuos de tipo A en la generación t es una cadena de Markov con probabilidades de transición,

$$\mathbb{P}(X_{t+1} = j | X_t = i) = \binom{N}{j} p_i^j (1 - p_i)^{N-j}.$$

La Figura 2.3 representa la evolución de la población entre dos generaciones considerando el efecto de la mutación.

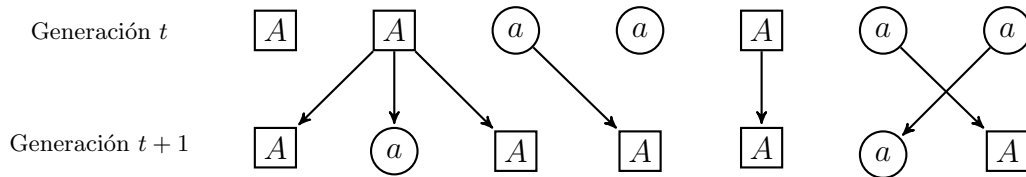


Figura 2.3: Ejemplo de la evolución en la composición de la población entre la generación t y $t+1$ considerando mutación.

A continuación, observamos en la Figura 2.4 la evolución del número de individuos de tipo A , que se modela con un proceso de Markov discreto, en una población de tamaño $N = 10$. Inicialmente hay 5 individuos de tipo A en dicha población y se observa un periodo de 15 generaciones. Podemos notar que cuando alguna de las probabilidades de mutación es cero ($\delta = 0$ o $\gamma = 0$) el proceso alcanza un estado absorbente. Por otro lado, si la probabilidad de mutación de A a a es grande y la tasa de mutación de a a A pequeña, el número de individuos de tipo A será cercano a cero y por el contrario, si la tasa de mutación de A a a chica, el número de individuos de tipo A será cercano a uno.

Al considerar el efecto de la mutación es posible obtener individuos de tipo a a pesar de que en la generación previa solo haya individuos de tipo A o viceversa, por lo que, no ocurre la extinción.

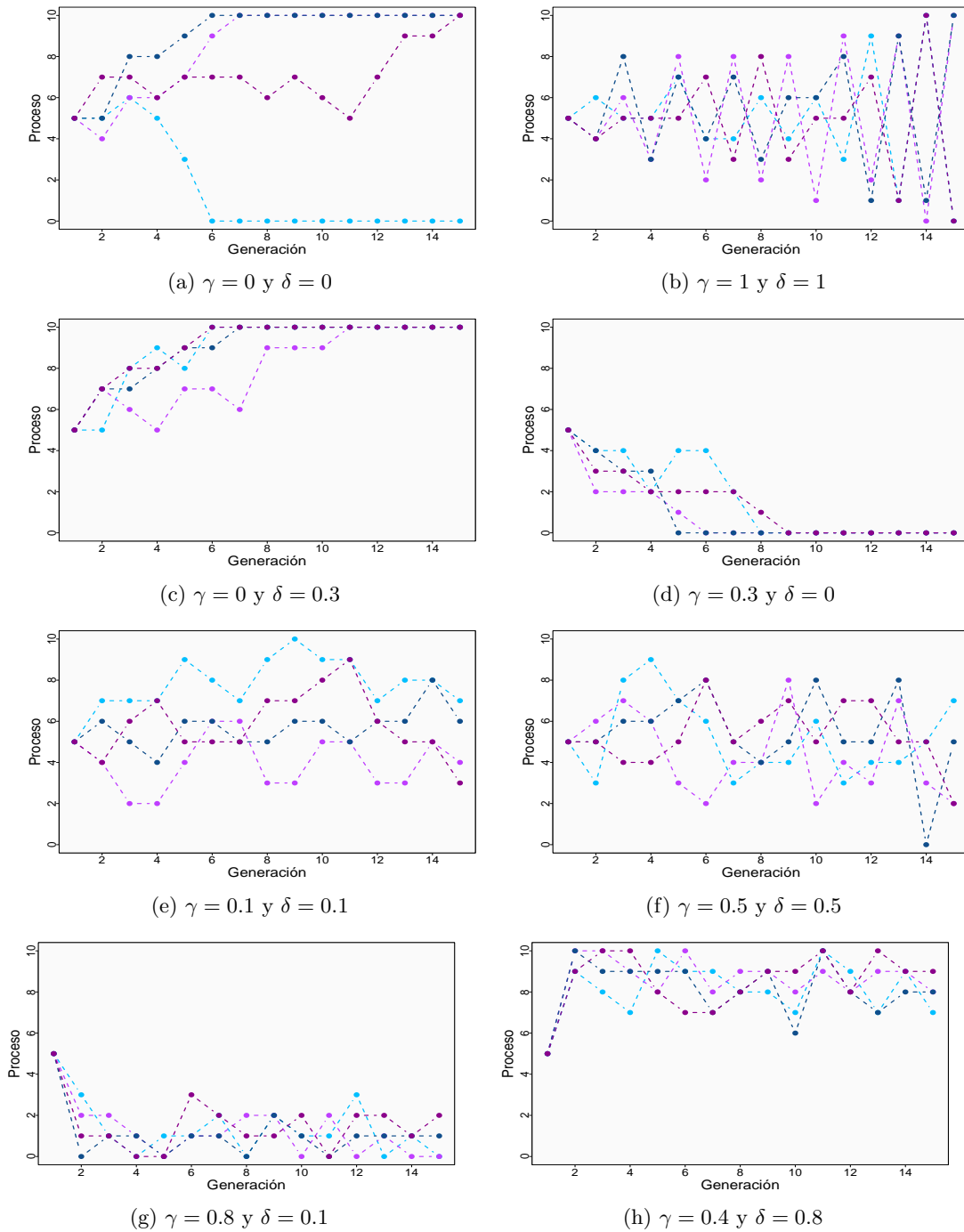


Figura 2.4: Simulación de trayectorias del proceso de Wright-Fisher discreto con distintas probabilidades de mutación. Tamaño de población $N = 10$, $X_0 = 5$ a través de 15 generaciones.

2.2. Difusión de Wright Fisher

A pesar de que el modelo tiene una matriz de transición conocida y con una forma simple no es factible encontrar analíticamente funcionales probabilísticos relevantes del proceso como la esperanza del tiempo de absorción cuando no hay mutación o la esperanza del tiempo que tarda en alcanzar alguna de las fronteras. Por tal razón se busca aproximar la cadena de Markov mediante difusiones cuando N es grande.

2.2.1. Aproximación a la difusión

Para encontrar una difusión a partir del modelo de Wright-Fisher definimos el proceso

$$Y_\tau^N = \frac{X_{[N\tau]}}{N},$$

donde una unidad $\tau = 1$ del proceso Y_τ^N corresponde a N generaciones del proceso original X_t y representa la proporción de individuos de tipo A al tiempo $[N\tau]$.

Sean $\delta = \frac{\alpha}{N}$ y $\gamma = \frac{\beta}{N}$, entonces la intensidad de mutación de A a a es $\beta = N\gamma$ y la intensidad de mutación de a a A es $\alpha = \delta N$. Notemos que al suponer δ y γ de esa manera tenemos una interpretación conveniente ya que las tasas de mutación son constantes por unidad de tiempo del proceso escalado Y_τ^N que equivale a N generaciones del proceso original.

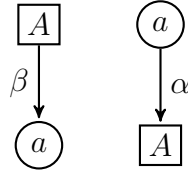


Figura 2.5: Intensidad de mutación.

Con la finalidad de hallar la difusión a la cual converge el proceso Y_τ^N con N suficientemente grande vamos a utilizar las condiciones (2), (3) y (4) de la Sección 1.3.4.

Consideramos $h = 1/N$ y notemos que

$$\Delta_h Y_\tau^N = Y_{\tau+\frac{1}{N}}^N - Y_\tau^N = \frac{X_{[N\tau+1]}}{N} - \frac{X_{[N\tau]}}{N}.$$

Además, sabemos que la distribución condicional de $X_{[N\tau+1]}$ dado $X_{[N\tau]} = i$ tiene distribución Binomial de parámetros N y p_i , luego,

$$\begin{aligned} N\mathbb{E}\left[\Delta_h Y_\tau^N \mid Y_\tau^N = \frac{i}{N}\right] &= N\mathbb{E}\left[Y_{\tau+\frac{1}{N}}^N - Y_\tau^N \mid Y_\tau^N = \frac{i}{N}\right] \\ &= N\mathbb{E}\left[\frac{X_{[N\tau+1]}}{N} - \frac{X_{[N\tau]}}{N} \mid \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] \\ &= N\left(\mathbb{E}\left[\frac{X_{[N\tau+1]}}{N} \mid \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] - \frac{i}{N}\right) \\ &= N\left(p_i - \frac{i}{N}\right). \end{aligned}$$

Al tomar el límite, asumimos que $y = \lim_{N \rightarrow \infty} \frac{i}{N}$, obtenemos

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E}\left[\Delta_h Y_\tau^N \mid Y_\tau^N = \frac{i}{N}\right] &= \lim_{N \rightarrow \infty} N\mathbb{E}\left[Y_{\tau+\frac{1}{N}}^N - Y_\tau^N \mid Y_\tau^N = \frac{i}{N}\right] \\ &= \lim_{N \rightarrow \infty} N\left(p_i - \frac{i}{N}\right) \\ &= \lim_{N \rightarrow \infty} N\left(\frac{i}{N}(1-\gamma) + \left(1 - \frac{i}{N}\right)\delta - \frac{i}{N}\right) \\ &= \lim_{N \rightarrow \infty} \left(1 - \frac{i}{N}\right)(N\delta) - \left(\frac{i}{N}\right)(N\gamma) \\ &= \alpha\left(1 - \frac{i}{N}\right) - \beta\left(\frac{i}{N}\right) \\ &= \alpha(1-y) - \beta y. \end{aligned} \tag{2.1}$$

Entonces, la media infinitesimal está dada por

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[\Delta_h Y_\tau^N \mid Y_\tau^N = \frac{i}{N} \right] = \alpha(1-y) - \beta y,$$

donde la convergencia es uniforme si $0 \leq y \leq 1$.

Ahora, utilizando el segundo momento de una variable aleatoria Binomial, evaluamos

$$\begin{aligned} N \mathbb{E} \left[(\Delta_h Y_\tau^N)^2 \mid Y_\tau^N = \frac{i}{N} \right] &= N \mathbb{E} \left[\left(Y_{\tau + \frac{1}{N}}^N - Y_\tau^N \right)^2 \mid Y_\tau^N = \frac{i}{N} \right] \\ &= N \mathbb{E} \left[\left(\frac{X_{[N\tau+1]}}{N} - \frac{X_{[N\tau]}}{N} \right)^2 \mid \frac{X_{[N\tau]}}{N} = \frac{i}{N} \right] \\ &= N \mathbb{E} \left[\frac{X_{[N\tau+1]}^2}{N^2} \mid \frac{X_{[N\tau]}}{N} = \frac{i}{N} \right] - 2 \left(\frac{i}{N} \right) \mathbb{E} \left[\frac{X_{[N\tau+1]}}{N} \mid \frac{X_{[N\tau]}}{N} = \frac{i}{N} \right] + \left(\frac{i}{N} \right)^2 \\ &= N \left[\frac{1}{N^2} (N(N-1)p_i^2 + Np_i) - 2 \left(\frac{i}{N} \right) \left(\frac{1}{N} \right) Np_i + \left(\frac{i}{N} \right)^2 \right] \\ &= N \left[\frac{1}{N^2} \left(N^2 \left(p_i - \frac{i}{N} \right)^2 - Np_i(1-p_i) \right) \right] \\ &= \frac{1}{N} \left(N^2 \left(p_i - \frac{i}{N} \right)^2 - Np_i(1-p_i) \right). \end{aligned}$$

Notamos de la ecuación (2.1), que $N \left(p_i - \left(\frac{i}{N} \right) \right)$ converge cuando N tiende a infinito, así, $\frac{1}{N} \left(N^2 \left(p_i - \frac{i}{N} \right)^2 \right)$ es una función de la forma $\mathcal{O} \left(\frac{1}{N} \right)$, donde $\mathcal{O} \left(\frac{1}{N} \right)$ es de orden $\frac{1}{N}$.

Luego, desarrollando el segundo término y sustituyendo $\delta = \frac{\alpha}{N}$ y $\gamma = \frac{\beta}{N}$,

$$\begin{aligned} p_i(1-p_i) &= \left(\left(\frac{i}{N} \right) (1-\gamma) + \left(1 - \frac{i}{N} \right) \delta \right) \left(\left(1 - \frac{i}{N} \right) (1-\delta) + \left(\frac{i}{N} \right) \gamma \right) \\ &= \left(\frac{i}{N} \right)^2 (1-\gamma)\gamma + \left(\frac{i}{N} \right) \left(1 - \frac{i}{N} \right) \left((1-\delta)(1-\gamma) + \delta\gamma \right) + \left(1 - \frac{i}{N} \right)^2 (1-\delta)\delta \\ &= \left(\frac{i}{N} \right) \left(1 - \frac{i}{N} \right) + \mathcal{O} \left(\frac{1}{N} \right). \end{aligned}$$

Por lo tanto, considerando $y = \lim_{N \rightarrow \infty} \frac{i}{N}$,

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[(\Delta_h Y_\tau^N)^2 \mid Y_\tau^N = \frac{i}{N} \right] &= \lim_{N \rightarrow \infty} N \mathbb{E} \left[(\Delta_h Y_\tau^N)^2 \mid Y_\tau^N = \frac{i}{N} \right] \\ &= \lim_{N \rightarrow \infty} \left(\frac{i}{N} \right) \left(1 - \frac{i}{N} \right) + \mathcal{O} \left(\frac{1}{N} \right) \\ &= y(1-y), \end{aligned} \tag{2.2}$$

la convergencia es uniforme para $0 \leq y \leq 1$.

Finalmente, de acuerdo al Lema 1.3.1 debemos verificar que $\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left[(\Delta_h Y_\tau^N)^4 \mid Y_\tau^N = \frac{i}{N} \right] = 0$ con el fin de justificar que se cumple la propiedad de Dynkin y como consecuencia, obtener la aproximación deseada. Utilizando los momentos de una distribución Binomial.

$$\begin{aligned}
N\mathbb{E}\left[\left(\Delta_h Y_\tau^N\right)^4 \middle| Y_\tau^N = \frac{i}{N}\right] &= N\mathbb{E}\left[\left(Y_{\tau+\frac{1}{N}}^N - Y_\tau^N\right)^4 \middle| Y_\tau^N = \frac{i}{N}\right] \\
&= N\mathbb{E}\left[\left(\frac{X_{[N\tau+1]}}{N} - \frac{X_{[N\tau]}}{N}\right)^4 \middle| \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] \\
&= N\mathbb{E}\left[\frac{X_{[N\tau+1]}^4}{N^4} \middle| \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] - 4\left(\frac{i}{N}\right)\mathbb{E}\left[\frac{X_{[N\tau+1]}^3}{N^3} \middle| \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] \\
&\quad + 6\left(\frac{i}{N}\right)^2\mathbb{E}\left[\frac{X_{[N\tau+1]}^2}{N^2} \middle| \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] - 4\left(\frac{i}{N}\right)^3\mathbb{E}\left[\frac{X_{[N\tau+1]}}{N} \middle| \frac{X_{[N\tau]}}{N} = \frac{i}{N}\right] + \left(\frac{i}{N}\right)^4 \\
&= N\left[\frac{1}{N^4}[N(N-1)(N-2)(N-3)p_i^4 + 6N(N-1)(N-2)p_i^3 + 7N(N-1)p_i^2 + Np_i] \right. \\
&\quad \left. - 4\left(\frac{1}{N^3}\right)\left(\frac{i}{N}\right)[N(N-1)(N-2)p_i^3 + 3N(N-1)p_i^2 + Np_i] \right. \\
&\quad \left. + 6\left(\frac{1}{N^2}\right)\left(\frac{i}{N}\right)^2[N(N-1)p_i^2 + Np_i] - 4\left(\frac{1}{N^3}\right)\left(\frac{i}{N}\right)^3 Np_i + \left(\frac{i}{N}\right)^4\right] \\
&= \frac{1}{N^3}\left[N^4\left(p_i - \left(\frac{i}{N}\right)\right)^4 + 6N^3p_i(1-p_i)\left(p_i - \left(\frac{i}{N}\right)\right)^2 + \mathcal{O}\left(\frac{1}{N^3}\right)\right].
\end{aligned}$$

Finalmente, al tomar el límite,

$$\begin{aligned}
\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E}\left[\left(\Delta_h Y_\tau^N\right)^4 \middle| Y_\tau^N = \frac{i}{N}\right] &= \lim_{N \rightarrow \infty} N\mathbb{E}\left[\left(Y_{\tau+\frac{1}{N}}^N - Y_\tau^N\right)^4 \middle| Y_\tau^N = \frac{i}{N}\right] \\
&= \left(\frac{1}{N^3}\right)\left[N^4\left(p_i - \left(\frac{i}{N}\right)\right)^4 + 6N^3p_i(1-p_i)\left(p_i - \left(\frac{i}{N}\right)\right)^2 + \mathcal{O}\left(\frac{1}{N^3}\right)\right] \\
&= 0.
\end{aligned}$$

Por lo que, el proceso $Y_\tau^N = \frac{X_{[N\tau]}}{N}$ converge, cuando N tiende a infinito, a una difusión Y_τ con espacio de estados en el intervalo $[0,1]$, coeficiente de deriva $\mu(y) = \alpha(1-y) - \beta y$ y difusión $\sigma = \sqrt{y(1-y)}$.

En la Figura 2.6 podemos ver el comportamiento de la aproximación con $N=50, 100$ y 500 .

2.2.2. Ecuación diferencial estocástica

A partir del coeficiente de deriva y difusión podemos escribir el generador infinitesimal del proceso,

$$\mathcal{A}f(x) = \left(\alpha(1-x) - \beta x\right)f'(x) + \frac{1}{2}\left(x(1-x)\right)f''(x). \quad (2.3)$$

Por lo tanto, a la solución de la ecuación diferencial estocástica

$$dX_t = \left(\alpha(1-X_t) - \beta X_t\right)dt + \sqrt{X_t(1-X_t)}dB_t, \quad (2.4)$$

se le llama difusión de Wright-Fisher.

Ya que hemos definido nuestro proceso, estamos interesados en conocer algunas características del mismo.

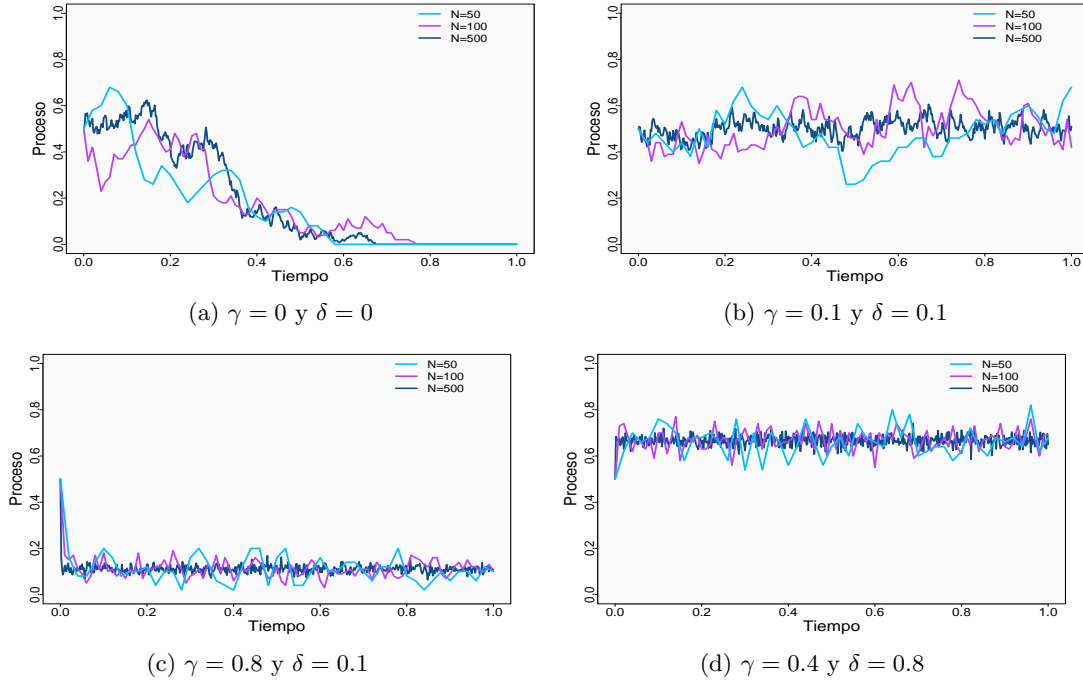


Figura 2.6: Simulación de trayectorias de la aproximación del proceso de Wright-Fisher con distintas probabilidades de mutación y $N = 50, 100$ y 500 generaciones y estado inicial $X_0 = 0.5$.

2.2.3. Características del proceso

Sea $\tau_y = \inf\{t \geq 0 : X_t = y\}$ el tiempo de paro que representa el tiempo en el que el proceso alcanza el estado $y \in (0, 1)$. Consideramos los estados b y c fijos, definimos a la variable aleatoria $\tau_{b,c} = \min\{\tau_b, \tau_c\}$, la primera vez que el proceso llega al estado c o al estado b .

De manera natural surge la pregunta de conocer la probabilidad de alcanzar el estado b antes que el estado c , esta probabilidad está dada por la ecuación (1.13), es decir,

$$u(x) = \mathbb{P}(\tau_b < \tau_c | X_0 = x), \quad c < x < b.$$

Como vimos en el capítulo anterior, $u(x)$ satisface la ecuación (1.14), al sustituir los parámetros infinitesimales del proceso que estamos trabajando obtenemos,

$$(\alpha(1-x) - \beta x)u' + \frac{1}{2}x(1-x)u'' = 0, \quad c < x < b \quad (2.5)$$

y condiciones iniciales $u(c) = 0$ y $u(b) = 1$.

La solución a la ecuación (2.5) que representa la probabilidad de que el proceso llegue al estado b antes que a c está dada por la ecuación (1.26), es decir,

$$u(x) = \frac{S(x) - S(c)}{S(b) - S(c)}. \quad (2.6)$$

Debemos encontrar a la función de escala de acuerdo a la Definición 1.3.7, para $x^* \in [0, 1]$ arbitrario,

$$\begin{aligned}
s(x) &= \exp\left\{-\int_{x^*}^x \frac{2\mu(y)}{\sigma^2(y)} dy\right\} \\
&= \exp\left\{-\int_{x^*}^x \frac{2(\alpha(1-y) - \beta y)}{y(1-y)} dy\right\} \\
&= \exp\left\{-\left(\int_{x^*}^x \frac{2\alpha}{y} dy - \int_{x^*}^x \frac{2\beta}{1-y} dy\right)\right\} \\
&= \exp\left\{-\left(2\alpha \ln(x) - 2\alpha \ln(x^*) + 2\beta \ln(1-x) - 2\beta \ln(1-x^*)\right)\right\} \\
&= \exp\left\{-\left(2\alpha \ln(x) + 2\beta \ln(1-x) + k_1\right)\right\} \\
&= k_2 x^{-2\alpha} (1-x)^{-2\beta},
\end{aligned}$$

donde k_1 y k_2 son constantes.

Entonces, la función de escala está dada por

$$\begin{aligned}
S(x) &= \int_{x_0}^x s(y) dy \\
&= k_2 \int_{x_0}^x y^{-2\alpha} (1-y)^{-2\beta} dy,
\end{aligned} \tag{2.7}$$

para algún x_0 arbitrario en el espacio de estados. Recordemos que la elección de x^* y x_0 es irrelevante.

Bajo las condiciones $\alpha < \frac{1}{2}$, $\beta < \frac{1}{2}$, y $x_0 = 0$ la función de escala resulta la función Beta truncada en x con parámetros $1 - 2\alpha$ y $1 - 2\beta$. Por otro lado, si $\alpha \geq \frac{1}{2}$, $S(0) = -\infty$ y contrariamente si $\beta \geq \frac{1}{2}$, $S(1) = \infty$.

Analizando de manera general la ecuación (2.6) y la función de escala del proceso de Wright-Fisher encontrada en la ecuación (2.7); si $\beta \geq \frac{1}{2}$ y b tiende a uno, entonces $\mathbb{P}(\tau_b < \tau_c) = \frac{S(x) - S(c)}{S(b) - S(c)}$ converge a cero, es decir, ya que con probabilidad uno, el proceso alcanza el estado c antes que el estado uno, la frecuencia del número de individuos de tipo A no alcanza el valor uno. Por su parte, si $\alpha \geq \frac{1}{2}$ y c tiende a cero, se tiene que $\mathbb{P}(\tau_c < \tau_b) = \frac{S(b) - S(x)}{S(b) - S(c)}$ converge a cero, de manera que la frecuencia del número de individuos no alcanza el valor cero. En otras palabras, las tasas de mutación no son lo suficientemente fuertes para que la frecuencia de individuos de tipo A llegue a las fronteras, más adelante estudiaremos el comportamiento de las fronteras con mayor detalle.

Otra cuestión que también podría interesarnos es encontrar el tiempo medio en que el proceso llega a alguno de los estados b o c ,

$$v(x) = \mathbb{E}[\tau_{b,c} | X_0 = x], \tag{2.8}$$

de acuerdo la ecuación (1.17) satisface

$$(\alpha(1-x) - \beta x)v' + \frac{1}{2}x(1-x)v'' = -1, \quad c < x < b, \tag{2.9}$$

cuya solución está dada por la ecuación (1.27) donde $g(z) = 1$,

$$v(x) = 2 \left[u(x) \int_x^b (S(b) - S(z))g(z)m(z)dz + (1 - u(x)) \int_c^x (S(z) - S(c))g(z)m(z)dz \right].$$

De la misma manera que la función de escala encontramos la función de velocidad a partir de la Definición 1.3.7,

$$\begin{aligned}
m(x) &= \frac{1}{\sigma^2(x)s(x)} \\
&= \frac{x^{2\alpha}(1-x)^{2\beta}}{x(1-x)} \\
&= x^{2\alpha-1}(1-x)^{2\beta-1}.
\end{aligned} \tag{2.10}$$

Sustituimos la función de velocidad y la solución a la ecuación (2.8) queda de la siguiente manera,

$$\begin{aligned}
v(x) &= 2 \left[u(x) \int_x^b (S(b) - S(z)) z^{2\alpha-1} (1-z)^{2\beta-1} dz \right. \\
&\quad \left. + (1-u(x)) \int_c^x (S(z) - S(c)) z^{2\alpha-1} (1-z)^{2\beta-1} dz \right].
\end{aligned} \tag{2.11}$$

A continuación veremos un par de ejemplos donde es posible encontrar una expresión explícita para la función de escala.

Ejemplo 2.2.1. *Supongamos que no hay mutación, esto es, $\alpha = 0$ y $\beta = 0$. En este caso los estados frontera 0 y 1 son puntos absorbentes, lo cual significa que a partir de cierto momento todos los individuos de la población serán del mismo tipo.*

La función de escala se simplifica, $s(x) = 1$, para $x \in (0, 1)$ pues el coeficiente de deriva $\mu(x) = 0$, y por consiguiente, $S(x) = x$. Entonces, la probabilidad de alcanzar el estado b antes que el estado c dada en la ecuación (2.6) es

$$u(x) = \frac{x-c}{b-c}, \quad 0 < c \leq x \leq b < 1. \tag{2.12}$$

Ya que la fórmula se sigue cumpliendo si tomamos el límite cuando c tiende a cero y b tiende a uno, asumiendo que el proceso inicia en x , $X_0 = x$, la probabilidad de fijación en 1 donde todos los individuos son de tipo A es x , mientras que la fijación en cero donde todos los individuos son de tipo a ocurre con probabilidad $1-x$.

Por su parte, la función de velocidad en la ecuación (2.10) se reduce a

$$m(x) = \frac{1}{x(1-x)},$$

de manera que el tiempo promedio en que el proceso llega a alguno de los estados b o c obtenido en la ecuación (2.11) es

$$\begin{aligned}
v(x) &= 2 \left[\frac{x-c}{b-c} \int_x^b \frac{b-z}{z(1-z)} dz + \frac{b-x}{b-c} \int_c^x \frac{z-c}{z(1-z)} dz \right] \\
&= 2 \left[\frac{x-c}{b-c} (b \ln(b) - b \ln(x) + (1-b) \ln(1-b) - (1-b) \ln(1-x)) \right. \\
&\quad \left. + \frac{b-x}{b-c} (c \ln(c) - c \ln(x) + (1-c) \ln(1-c) - (1-c) \ln(1-x)) \right].
\end{aligned}$$

En este ejemplo obtener el tiempo medio de fijación es válido si calculamos $v(x)$ cuando c tiende a cero y b a uno, es decir,

$$v(x) = -2(x \ln(x) + (1-x) \ln(1-x)).$$

Resulta sencillo verificar que la función $v(x)$ se maximiza cuando $x = \frac{1}{2}$ y el tiempo medio de fijación es $2 \ln(2)$.

Ejemplo 2.2.2. Supongamos que solo hay mutación de A a a , es decir, $\alpha = 0$ y $\beta > 0$.

La función de escala en el intervalo $(0, 1)$ se reduce a

$$\begin{aligned} S(x) &= \int_0^x (1-y)^{-2\beta} dy \\ &= \frac{1 - (1-x)^{-2\beta+1}}{1-2\beta}, \quad 2\beta \neq 1, \end{aligned}$$

que es una función creciente.

Entonces, la probabilidad de llegar al estado b antes que al estado c dado que iniciamos en x es la siguiente,

$$u(x) = \frac{(1-c)^{-2\beta+1} - (1-x)^{-2\beta+1}}{(1-c)^{-2\beta+1} - (1-b)^{-2\beta+1}}.$$

Es claro que si b tiende a uno con la condición $\beta > \frac{1}{2}$ se tiene que $u(x)$ tiende a cero, es decir, con probabilidad uno el proceso alcanza el estado c antes de llegar al estado 1. Por el contrario, $\beta < \frac{1}{2}$ implica que $u(x) > 0$. En este caso podríamos decir que para tasas de mutación β suficientemente grandes, el proceso no llega al estado 1.

Además, esperaríamos que haya fijación en el estado 0 en algún momento del proceso, o sea, que a partir de cierto tiempo los individuos de tipo A se extingan. De acuerdo a la ecuación (2.8) el tiempo esperado de fijación es τ_0 . Como ya hemos visto, si $\beta > \frac{1}{2}$, el uno es inalcanzable, entonces, $\tau_0 = \tau_{0,1}$.

$$\begin{aligned} v(x) &= 2 \left[\frac{(1-c)^{-2\beta+1} - (1-x)^{-2\beta+1}}{(1-c)^{-2\beta+1} - (1-b)^{-2\beta+1}} \int_x^b \left(\frac{(1-z)^{-2\beta+1} - (1-b)^{-2\beta+1}}{1-2\beta} \right) \frac{(1-z)^{2\beta-1}}{z} dz \right. \\ &\quad \left. + \frac{(1-x)^{-2\beta+1} - (1-b)^{-2\beta+1}}{(1-c)^{-2\beta+1} - (1-b)^{-2\beta+1}} \int_c^x \left(\frac{(1-c)^{-2\beta+1} - (1-z)^{-2\beta+1}}{1-2\beta} \right) \frac{(1-z)^{2\beta-1}}{z} dz \right]. \end{aligned}$$

Ahora, tomando límite cuando c tiende a cero y b tiende a uno,

$$v(x) = 2 \left[1 - \frac{(1-x)^{-2\beta+1}}{1-2\beta} \int_x^1 \frac{(1-z)^{2\beta-1}}{z} dz + \int_c^x \left(\frac{(1-z)^{-2\beta+1} - 1}{2\beta-1} \right) \frac{(1-z)^{2\beta-1}}{z} dz \right].$$

Por último, para encontrar el tiempo promedio en que el proceso llega a la fijación debemos resolver la expresión anterior. Es importante notar que si $\beta < \frac{1}{2}$, $\tau_0 \neq \tau_{0,1}$ por lo que se tendría que encontrar el resultado de otra manera.

Finalmente, estamos interesados en encontrar la distribución estacionaria del proceso, la Proposición 1.38 asegura que el conjunto de posibles distribuciones estacionarias está determinado por,

$$\pi(x) = x^{2\alpha-1}(1-x)^{2\beta-1} \left(k_1 \int_{x^*}^x y^{-2\alpha}(1-y)^{-2\beta} dy + k_2 \right), \quad (2.13)$$

donde k_1 y k_2 son constantes que garantizan que sea distribución.

Como consecuencia, la existencia de la distribución estacionaria dependerá de la magnitud de α y β . Observamos que si $\alpha \geq \frac{1}{2}$ o $\beta \geq \frac{1}{2}$ la integral no converge. Luego, la condición $\pi(x) \geq 0$ requiere que x^* sea igual a 0 o 1 y la condición de integrabilidad de $\pi(x)$ obliga a $k_1 = 0$ y que k_2 sea la constante de normalización de una distribución Beta de parámetros 2α y 2β .

$$\pi(x) = \frac{1}{B(2\alpha, 2\beta)} x^{2\alpha-1}(1-x)^{2\beta-1}. \quad (2.14)$$

Además sabemos que una difusión tiene una única distribución estacionaria cuando sus fronteras son de entrada, por consiguiente, es necesario clasificar las fronteras del proceso según los valores que toman las tasas de mutación.

2.2.4. Clasificación de las fronteras

En esta sección vamos a estudiar el comportamiento de las fronteras del proceso de Wright-Fisher con el fin de proporcionar condiciones bajo las cuales dichas fronteras son regulares, de salida o de entrada, considerando la clasificación de Feller de la Tabla 1.2.

En primer lugar, realizamos los cálculos para averiguar qué sucede con el estado 0 recordando que $s(x) = x^{-2\alpha}(1-x)^{-2\beta}$ y $m(x) = x^{2\alpha-1}(1-x)^{2\beta-1}$.

Como ya hemos mencionado si $\alpha \geq \frac{1}{2}$, $S(0) = -\infty$ y por consiguiente, $S(0, x] = \lim_{a \downarrow 0} S[a, x] = \infty$, implica $\Sigma(0) = \infty$.

Por el contrario, cuando $\alpha < \frac{1}{2}$, notando que para x lo suficientemente pequeño, podemos aproximar $s(x) \approx x^{-2\alpha}$ y $m(x) \approx x^{2\alpha-1}$ entonces,

$$S(0, x] = \lim_{a \downarrow 0} S[a, x] \approx k_1 x^{-2\alpha+1} < \infty,$$

con k_1 constante. Adicionalmente, tenemos

$$\Sigma(0) = \int_0^x M[y, x] dS(y) = \int_0^x \int_y^x z^{2\alpha-1} (1-z)^{2\beta-1} dz dS(y) \approx \int_0^x (k_2 y^{2\alpha} + k_3) y^{-2\alpha} dy < \infty,$$

donde k_2 y k_3 son constantes.

Por otro lado, $M(0, x] = \lim_{a \downarrow 0} M[a, x] \approx k_4 x^{2\alpha} < \infty$ siempre y cuando $\alpha \neq 0$ y k_4 constante, luego,

$$N(0) = \int_0^x S[y, x] dM(y) = \int_0^x \int_y^x z^{-2\alpha} (1-z)^{-2\beta} dz dM(y) \approx \int_0^x (k_5 y^{-2\alpha+1} + k_6) y^{2\alpha-1} dy < \infty,$$

con k_5 y k_6 constantes. Finalmente, si $\alpha = 0$ resulta que $M(0, x]$ no converge y $N(0)$ tampoco.

Análogamente para la frontera 1, si $\beta \geq \frac{1}{2}$ ya vimos que $S(1) = \infty$, entonces $S[x, 1) = \infty$ y $\Sigma(1) = \infty$.

Al considerar $\beta < \frac{1}{2}$ podemos aproximar para x lo suficientemente cercano a 1, $s(x) \approx (1-x)^{-2\beta}$ y $m(x) \approx (1-x)^{2\beta-1}$, así $S[x, 1) \approx k_1 (1-x)^{-2\beta+1} < \infty$, donde k_1 es constante y

$$\Sigma(1) \approx \int_x^1 (k_2 (1-y)^{2\beta} + k_3) (1-y)^{-2\beta} dy < \infty,$$

con k_2, k_3 constantes.

Luego, si $\beta \neq 0$, $M[x, 1) = \lim_{b \uparrow 1} M[x, b] \approx k_4 (1-x)^{2\beta} < \infty$ con k_4 constante y como consecuencia,

$$N(1) \approx (k_5 (1-y)^{-2\beta+1} + k_6) (1-y)^{2\beta-1} dy < \infty.$$

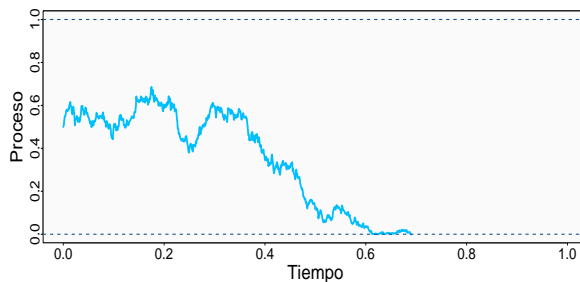
Caso contrario, si $\beta = 0$, ambas medidas divergen.

En la Tabla 2.1 se muestra la clasificación de las fronteras de acuerdo a los distintos valores que pueden tomar α y β .

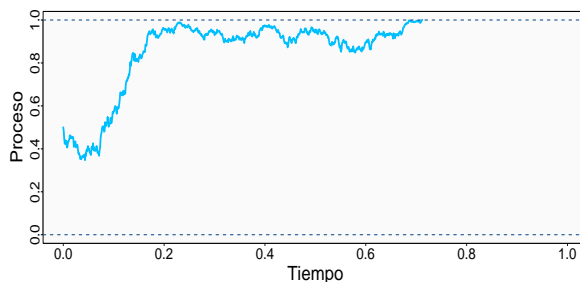
Estos resultados nos indican que si $\alpha = 0$ y $\beta = 0$ las fronteras 0 y 1 son de salida, en efecto, como el proceso representa la proporción de individuos de tipo A y no hay mutación, si la proporción es cero no es posible que en las siguientes generaciones se produzcan individuos de tipo A , de la misma forma sucede si el proceso llega al estado 1, y por consiguiente estos son estados absorbentes. En la Figura 2.7 podemos observar el comportamiento del proceso cuando alguna de las fronteras es de salida.

Frontera	Parámetro	Σ	N	Tipo
0	$\alpha = 0$	$< \infty$	∞	Salida
	$0 < \alpha < \frac{1}{2}$	$< \infty$	$< \infty$	Regular
	$\alpha \geq \frac{1}{2}$	∞	$< \infty$	Entrada
1	$\beta = 0$	$< \infty$	∞	Salida
	$0 < \beta < \frac{1}{2}$	$< \infty$	$< \infty$	Regular
	$\beta \geq \frac{1}{2}$	∞	$< \infty$	Entrada

Tabla 2.1: Tipos de fronteras del proceso de Wright-Fisher. Tabla tomada de [17].



(a) Frontera 0 de salida
 $\alpha = 0$ y $\beta = 0.5$

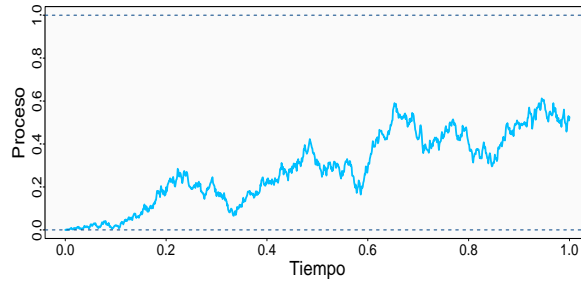


(b) Frontera 1 de salida
 $\alpha = 0.5$ y $\beta = 0$

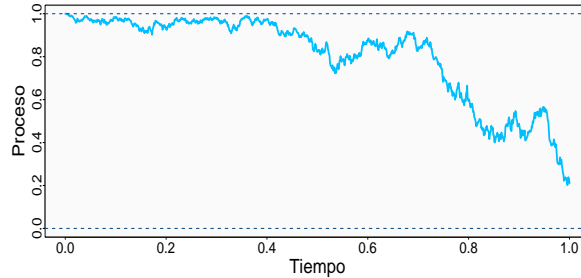
Figura 2.7: Simulación de trayectorias del proceso de Wright-Fisher con fronteras de salida.

Por otro lado, si $0 < \alpha < \frac{1}{2}$ y $0 < \beta < \frac{1}{2}$ tenemos fronteras regulares, entonces el proceso puede entrar y salir de ellas. Cabe mencionar que en la mayoría de las aplicaciones en genética se utilizan barreras reflejantes.

Por último si $\alpha \geq \frac{1}{2}$ y $\beta \geq \frac{1}{2}$ las fronteras son de entrada, esto quiere decir que el proceso no puede alcanzar los estados 0 y 1 del interior del espacio de estados. Sin embargo, el proceso puede iniciar en las fronteras pero no regresar a ellas. En otras palabras, es posible tener una población donde todos los individuos son de tipo A y que se produzcan nacimientos de individuos de tipo a pero la población no regresa a tener puros individuos tipo A , esto debido a que las tasas de mutación son lo suficientemente grandes. Podemos notar que el resultado es consistente con los cálculos obtenidos como solución al funcional $u(x)$. En la Figura 2.8 podemos observar la evolución del proceso cuando alguna de sus fronteras es de entrada.



(a) *Frontera 0 de entrada*
 $\alpha = 0.5$ y $\beta = 0.125$



(b) *Frontera 1 de entrada*
 $\alpha = 0.125$ y $\beta = 0.5$

Figura 2.8: Simulación de trayectorias del proceso de Wright-Fisher con fronteras de entrada.

Obviamente, falta considerar todas las combinaciones posibles de parámetros, por ejemplo, si $\alpha = 0$ y $\beta > \frac{1}{2}$ donde el estado 0 es una frontera de salida y el estado 1 es una frontera de entrada, es decir, si todos los individuos son de tipo a por la falta de mutación de a a A el proceso se fija en el estado cero, por el contrario, al inicio, la población puede estar compuesta solo de individuos tipo A pero esta configuración no se tendrá posteriormente.

En el presente trabajo estaremos interesados principalmente cuando ambas fronteras son de entrada, es decir, si $\alpha \geq \frac{1}{2}$ y $\beta \geq \frac{1}{2}$ puesto que tenemos una única distribución estacionaria. Como ya habíamos mencionado, si se cumplen estas condiciones la distribución estacionaria es una Beta de parámetros 2α y 2β .

2.3. Reparametrización

El proceso de Wright-Fisher también puede interpretarse desde el punto de vista genético. Este enfoque se encuentra en [8]. Antes de explicar el modelo daremos algunos conceptos. El término *genotipo* se refiere a la composición genética de un organismo, el genotipo está especificado por el genoma que codifica la información genética que pasa de padres a hijos en la reproducción sexual, es posible que dicha información pase modificada. El *genoma* está constituido por un conjunto de cromosomas, donde el *cromosoma* es una molécula de ADN que contiene muchos genes, elementos regulatorios y secuencias de nucleótidos. Una posición dada del cromosoma es llamada *locus* (*loci* en plural) y puede estar ocupada por uno o más genes. A las diferentes variaciones de un gen en un locus particular se les conoce como *alelos*, es decir, son las diferentes versiones de información genética codificada en un locus. Un *mapa genético* es la lista ordenada de loci en una genoma particular.

Los individuos *diploides* tienen dos copias homólogas de cada cromosoma, usualmente en el caso de la reproducción sexual, un cromosoma es heredado por el padre y el otro por la madre. Los cromosomas homólogos contienen los mismos genes en el mismo locus pero posiblemente diferentes alelos en esos genes.

Supongamos una población diploide de tamaño constante N donde las generaciones no se traslapan y ex-

perimentan un apareamiento aleatorio, además, en un locus tenemos dos alelos A y a , que tienen la misma habilidad tanto para sobrevivir como para reproducirse. Ya que la población es diploide regularmente se trata a los N individuos como las $2N$ copias del locus. En este sentido se desarrolla la teoría a partir de $2N$ individuos. Regresando al ejemplo de la urna, esta vez tenemos $2N$ bolas, i son tipo A y $2N - i$ son tipo a y elegimos $2N$ bolas con reemplazo. Nuevamente el proceso en t representa el número de individuos tipo A el tiempo t y tiene una distribución Binomial de parámetros $2N$ y $p_i = \frac{i}{2N}(1 - \gamma) + \frac{2N-i}{2N}\delta$.

De esta manera obtenemos los parámetros infinitesimales,

$$\mu(x) = \frac{1}{2}(\alpha(1-x) - \beta x) \quad \text{y} \quad \sigma^2(x) = x(1-x).$$

Por lo tanto, el generador infinitesimal

$$\mathcal{A}f(x) = \frac{1}{2}(\alpha(1-x) - \beta x) \frac{df}{dx} + \frac{1}{2}(x(1-x)) \frac{d^2f}{dx^2}, \quad (2.15)$$

y la ecuación diferencial estocástica

$$dX_t = \frac{1}{2}(\alpha(1-X_t) - \beta X_t)dt + \sqrt{X_t(1-X_t)}dB_t. \quad (2.16)$$

Podemos observar que es una reparametrización de las ecuaciones (2.3) y (2.4), más aún, los parámetros son los mismos salvo por una constante multiplicativa. De hecho, podríamos interpretarlo como que la tasa de mutación de A a a es $\frac{\beta}{2}$ y la tasa de a a A es $\frac{\alpha}{2}$, por lo que es indistinto utilizar una u otra. En este caso, la distribución estacionaria es una distribución Beta de parámetros α y β .

Por simplicidad, a partir de ahora, trabajaremos con la segunda parametrización.

2.4. Descomposición espectral de la función de densidad

Finalmente, vamos a encontrar la descomposición espectral del proceso de Wright-Fisher, es decir, expresar su función de densidad como una serie. Para ello supongamos que las fronteras son de salida o reflejantes.

2.4.1. Polinomios de Jacobi

En primer lugar, daremos una breve introducción a los polinomios de Jacobi, lo cuales nos ayudarán a encontrar la descomposición espectral de la función de transición de la difusión de Wright-Fisher, para más detalles [1], [13], [30]. Los polinomios de Jacobi han sido utilizados en campos del análisis matemático, así como en aplicaciones prácticas. Algunos casos especiales son los polinomios de Gegenbauer, Zernike, Legendre y Chebyshev, éstos últimos juegan un papel importante en los métodos espectrales de las ecuaciones diferenciales parciales.

Para $z \in [-1, 1]$ y $n \in \mathbb{N}$ los polinomios de Jacobi, $P_n^{(\alpha, \beta)}(z)$, satisfacen la ecuación diferencial

$$(1-z^2) \frac{d^2 f(z)}{dz^2} + (\beta - \alpha - (\alpha + \beta + 2)z) \frac{df(z)}{dz} + n(n + \alpha + \beta + 1)f(z) = 0,$$

para $\alpha, \beta > -1$, $P_0^{(\alpha, \beta)} = 1$ y además, $\{P_n^{(\alpha, \beta)}(z)\}$ forman un sistema ortogonal con respecto a la función de pesos $(1-z)^\alpha(1+z)^\beta$ en el intervalo $[-1, 1]$.

Para nuestros propósitos, el dominio y los parámetros de la función no son lo adecuados, sin embargo, definimos la siguiente modificación a los polinomios de Jacobi, para $x \in [0, 1]$ y $\alpha, \beta > 0$

$$R_n^{(\alpha, \beta)}(x) = c_n P_n^{(\beta-1, \alpha-1)}(2x-1),$$

la constante c_n está dada por

$$c_n = \sqrt{\frac{(2n + \alpha + \beta - 1)(\alpha + \beta)_{(n-1)} n!}{\alpha_{(n)} \beta_{(n)}}}, \quad (2.17)$$

donde $x_{(n)} = x(x+1)\dots(x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)}$ y $x_{(0)} = 1$.

Notemos que para α y β fijos, $\{R_n^{(\alpha,\beta)}(x)\}$ es ortogonal a la función $x^{\alpha-1}(1-x)^{\beta-1}$ en $[0, 1]$,

$$\int_0^1 R_n^{(\alpha,\beta)}(x)R_m^{(\alpha,\beta)}(x)x^{\alpha-1}(1-x)^{\beta-1}dx = \delta_{mn} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

δ_{mn} denota la delta de Kronecker, lo cual implica que

$$\int_0^1 \left(R_n^{(\alpha,\beta)}(x)\right)^2 x^{\alpha}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

y satisface la ecuación diferencial

$$x(1-x)\frac{d^2f(x)}{dx^2} + (\alpha(1-x) + \beta x)\frac{df(x)}{dx} + \frac{1}{2}n(n+\alpha+\beta-1)f(x) = 0.$$

Por lo tanto, los polinomios de Jacobi son las funciones propias del generador infinitesimal de la difusión de Wright-Fisher con valores propios $n(n+\alpha+\beta-1)$.

2.4.2. Descomposición espectral

De acuerdo a la modificación de los polinomios de Jacobi y la ecuación (1.48) la función de densidad de transición

$$\begin{aligned} p(x, y; t) &= x^{\alpha-1}(1-x)^{\beta-1} \sum_{n=0}^{\infty} e^{-n(n+\alpha+\beta-1)t} R_n^{(\alpha,\beta)}(x)R_n^{(\alpha,\beta)}(y) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \mathcal{D}_{\alpha,\beta}(y) \left(1 + \sum_{n=0}^{\infty} e^{-\frac{1}{2}n(n+\alpha+\beta-1)t} R_n^{(\alpha,\beta)}(x)R_n^{(\alpha,\beta)}(y) \right), \end{aligned}$$

con $\mathcal{D}_{\alpha,\beta}(y)$ es la función de densidad Beta de parámetros α y β evaluada en y .

El siguiente resultado nos ayuda a escribir de otra manera la densidad de transición del proceso de Wright-Fisher; es un corolario de Ethier y Griffiths [9] quienes encontraron una expresión para la densidad de transición del proceso del Fleming Voigt, el cual es una generalización al proceso de Wright-Fisher. Está dado para K tipos de individuos en la población.

En nuestro proceso, $K = 2$ y denotamos $\Delta = \{x = (x_1, x_2) : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 = 0\}$.

Supongamos que el proceso tiene generador

$$\mathcal{A} = \frac{1}{2} \sum_{i,j=1}^2 x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{j=1}^2 \sum_{i=1}^2 q_{ij} x_i \frac{\partial}{\partial x_j}$$

donde δ_{ij} es la delta de Kronecker y (q_{ij}) es la matriz infinitesimal que describe la estructura de las mutaciones.

Dada $\theta \geq 0$, sea $\{D_t\}_{t \geq 0}$ un proceso de muerte puro en los enteros positivos incluyendo infinito. Si el proceso inicia en ∞ con tasas de muerte

$$\lambda_n = \frac{1}{2}n(n+\theta-1),$$

para cualquier $n \geq 0$, definimos $q_m^\theta(t) = \mathbb{P}(D_t = m)$, $m \geq 0$ y $t > 0$, tenemos que

$$q_0^\theta(t) = 1 - \sum_{k=1}^{\infty} \frac{\theta + 2k - 1}{k!} (-1)^{k-1} \theta_{(m-1)} e^{-\lambda_k t}$$

$$\begin{aligned}
q_m^\theta(t) &= \sum_{k=m}^{\infty} \frac{\theta + 2k - 1}{k!} (-1)^{k-m} \binom{k}{m} (m + \theta)_{(k-1)} e^{-\lambda_k t} \\
&= \sum_{k=m}^{\infty} \frac{\theta + 2k - 1}{m!(k-m)!} (-1)^{k-m} (m + \theta)_{(k-1)} e^{-\lambda_k t}.
\end{aligned} \tag{2.18}$$

Teorema 2.4.1. Sean $\theta_1 \geq 0$ y $\theta_2 \geq 0$, hagamos $\theta = \theta_1 + \theta_2$ y $\Theta = (\theta_1, \theta_2)$. El proceso de difusión en Δ y generador \mathcal{A} , donde $q_{ij} = \frac{1}{2}\theta_j$ para toda $i, j \in \{1, \dots, K\}$ con $i \neq j$ tiene función de densidad de transición,

$$p((x_1, x_2), (y_1, y_2); t) = \sum_{n=0}^{\infty} q_n^\theta(t) \sum_{\xi \in (\mathbb{Z}_+)^K: |\xi|=n} \binom{n}{\xi} x_1^{\xi_1} x_2^{\xi_2} \frac{\Gamma(\theta + \xi_1 + \xi_2)}{\Gamma(\xi_1 + \theta_1)\Gamma(\xi_2 + \theta_2)} y_1^{\xi_1 + \theta_1} y_2^{\xi_2 + \theta_2}. \tag{2.19}$$

Notemos que podemos traducir el proceso del Teorema 2.4.1 dado en [9] a uno en una dimensión con $x = x_1$ y $(1-x) = x_2$. Reescribiendo la ecuación (2.19), haciendo $l = \xi_1$, implica $m-l = \xi_2$ y el coeficiente multinomial $\binom{n}{\xi} = \binom{n}{\xi_1 \xi_2} = \frac{n!}{l!(m-l)!} = \binom{m}{l}$, adicionalmente sustituimos las tasas de mutación $\theta_1 = \alpha$ y $\theta_2 = \beta$ la función de transición

$$\begin{aligned}
p(x, y; t) &= \sum_{n=0}^{\infty} q_n^\theta(t) \sum_{l=0}^m \binom{m}{l} x^l (1-x)^{m-l} \frac{\Gamma(\theta + m)}{\Gamma(\alpha + l)\Gamma(\beta + m - l)} y^{\alpha+l} (1-y)^{\beta+m-l} \\
&= \sum_{n=0}^{\infty} q_n^\theta(t) \sum_{l=0}^m \mathcal{B}_{m,x}(l) \mathcal{D}_{\alpha+l, \beta+m-l}(y),
\end{aligned} \tag{2.20}$$

donde $\mathcal{B}_{m,x}(l) = \binom{m}{l} x^l (1-x)^{m-l}$ la función de densidad Binomial evaluada en l de parámetros m y x y $\mathcal{D}_{\alpha+l, \beta+m-l}(y) = \frac{\Gamma(\theta+m)}{\Gamma(\alpha+l)\Gamma(\beta+m-l)} y^{\alpha+l} (1-y)^{\beta+m-l}$ la función de densidad Beta evaluada en y .

2.5. Coalescente de Kingman

Asociado a los procesos de genética de poblaciones como el modelo de Wright-Fisher se encuentran los procesos de coalescencia, uno de los primeros autores en estudiar este tipo de procesos fue Kingman [19]. El proceso coalescente describe el comportamiento genealógico de ciertos individuos y sus ancestros, representado por árbol coalescente, esto es, un árbol aleatorio binario que identifica a los ancestros de una muestra de genes o individuos. A continuación, vamos a dar una breve explicación basada principalmente en [12], [26] y [32] donde encontramos un resumen.

Consideremos el proceso de Wright-Fisher a tiempo discreto presentado en la Sección 2.1 con población constante y donde no hay mutación. La Figura 2.9a representa una posible realización del proceso que describe la evolución de una población de tamaño 7 a través de 9 generaciones bajo las hipótesis del modelo. Observamos que una línea de descendencia se divide o ramifica cuando un individuo tiene dos o más descendientes y termina cuando un individuo no los tiene. Por ejemplo, la línea de descendencia del primer individuo termina con él ya que no tiene descendientes mientras que la línea de descendencia del segundo presenta una ramificación hacia tres individuos.

Lo que nos interesa ahora es rastrear la ascendencia de un subconjunto de individuos de la población, por tanto, observaremos hacia atrás en el tiempo la genealogía de cada uno de ellos. En la primer generación donde al menos dos individuos tienen un ancestro común se dice que ocurre la primer coalescencia, de manera sucesiva cuando se vuelve a encontrar un ancestro común ocurre la siguiente coalescencia y así sucesivamente hasta llegar al primer ancestro en común. En el ejemplo de la Figura 2.9b seleccionamos un subconjunto de la población y notamos que la primer coalescencia se da en la generación 8, la segunda coalescencia tiene lugar en la generación 6, la siguiente en la generación 3 y el primer ancestro en común se encuentra en la

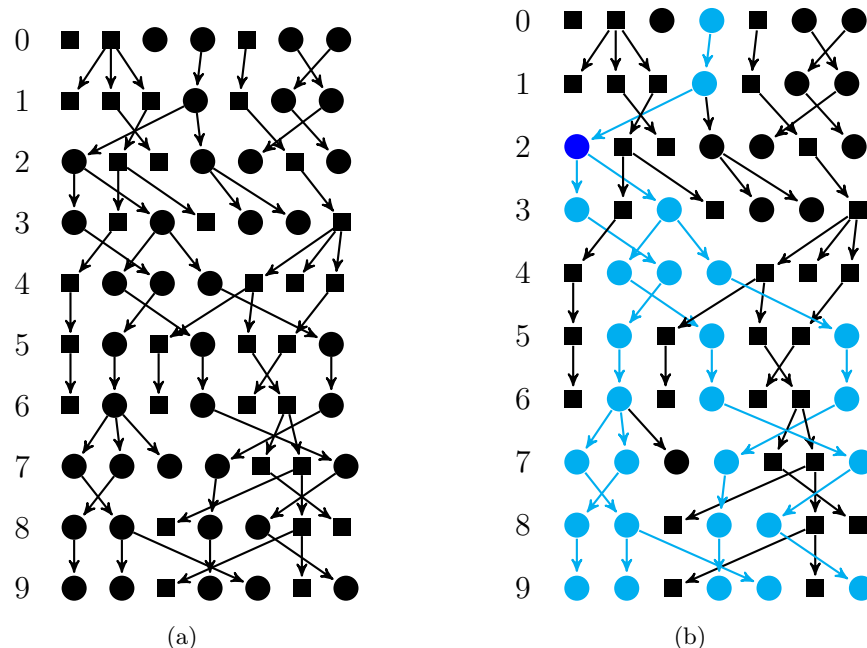


Figura 2.9: Ejemplo de la evolución de una población de tamaño 7 a través de 9 generaciones. En la Figura 2.9b los individuos en azul claro representan el subconjunto de la población con ancestro común marcado en azul oscuro.

generación 2 (el individuo marcado con azul oscuro).

Es claro que al identificar las líneas de descendencia de un grupo de individuos hacia atrás en el tiempo el número de líneas decrece hasta alcanzar uno cuando se encuentra al primer ancestro común como ya lo vimos en el ejemplo. De acuerdo con Nordborg [26] existen dos ideas fundamentales para estudiar la genealogía de una población, la primera consiste en que es posible separar el proceso de mutación neutral del proceso genealógico y el segundo la viabilidad de modelar la genealogía de un grupo de individuos sin importar el resto de la población; en suma, podemos modelar conjuntamente la deriva genética y la mutación neutral utilizando la genealogía como ya hemos mencionado con el tiempo hacia atrás y posteriormente superponiendo el efecto de la mutación con el tiempo hacia adelante.

Así, al proceso que describe la genealogía, es decir, el comportamiento de las líneas de descendencia de los individuos de una población se le llama proceso genealógico o proceso ancestral [32] y puede modelarse mediante una cadena de Markov $\{A_n\}_{n \in \mathbb{N}}$. Kingman [19] estudió el proceso que describe el comportamiento del árbol genealógico de la población, demostró que el proceso coalescente converge a una cadena de Markov a tiempo continuo, por esta razón, usualmente se le denomina el coalescente de Kingman.

Antes de describir la formulación del modelo que propuso Kingman observamos que la Figura 2.9b puede representarse de forma más sencilla como un árbol donde nos va a interesar la estructura, es decir, los individuos con ancestro común y la longitud de las ramas, Figura 2.10.

Formalmente, para un tiempo n fijo y una población de tamaño N seleccionamos sin reemplazo una muestra de individuos que etiquetamos I_1, I_2, \dots, I_i . Kingman demostró que la familia de árboles de estos individuos y sus ancestros puede modelarse mediante una cadena de Markov a tiempo discreto $\{B_n^N\}_{n \in \mathbb{N}}$ donde el espacio de estados \mathcal{E} son las relaciones de equivalencia del conjunto $\{1, 2, \dots, i\}$, de tal manera que en el estado al tiempo r del proceso, l y m están en la misma clase de equivalencia si tienen el mismo ancestro en la generación $n - r$. Por consiguiente, las distintas clases de equivalencia en B_r^N , representan a los ancestros de la muestra en la generación $n - r$. Al tiempo cero, la cadena se encuentra en el estado donde cada individuo es una clase de equivalencia y bajo la hipótesis de que todos tienen un mismo ancestro la cadena alcanza un

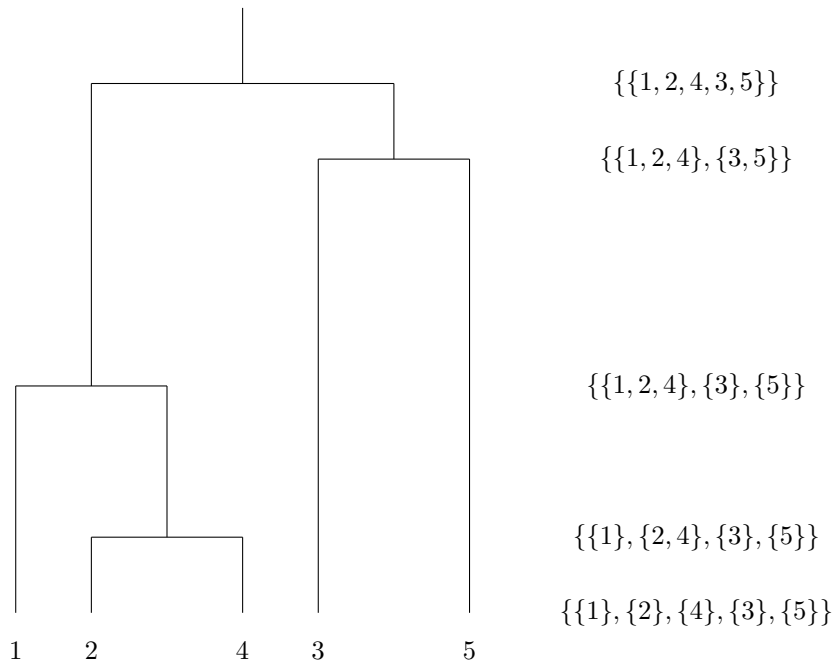


Figura 2.10: Representación mediante un árbol binario de las líneas de descendencia o la genealogía de un subconjunto de la población (Sin pérdida de generalidad, se reordenan los individuos).

estado absorbente.

Ahora, para $\alpha \in \mathcal{E}$, $|\alpha|$ representa el número de clases de equivalencia de α , así para $n \in \mathbb{N}$, $|B_n^N|$ es equivalente a A_n , el proceso ancestral definido previamente. Finalmente, al hacer converger N a infinito, el proceso converge a un proceso de saltos de Markov con espacio de estados \mathcal{E} que representa al conjunto de relaciones de equivalencia de $\{1, 2, \dots, n\}$ y generador infinitesimal $Q = (q_{\xi\eta})_{\xi, \eta \in \mathcal{E}}$ dado por

$$q_{\xi\eta} = \begin{cases} -\frac{k(k-1)}{2} & \text{si } \xi = \eta \\ 1 & \text{si } \xi \prec \eta \\ 0 & \text{otro caso,} \end{cases}$$

donde $k = |\xi|$ es el número de clases de equivalencia de ξ y $\xi \prec \eta$ si η se obtiene de ξ como la unión de dos de sus clases de equivalencia.

O bien, en términos del proceso ancestral, $\{A_n\}$ converge a una cadena de Markov a tiempo continuo con espacio de estados $\{1, 2, \dots, i\}$ con generador infinitesimal $Q = (q_{lj})_{l, j \in \{1, \dots, i\}}$ tal que

$$q_{lj} = \begin{cases} -\frac{k(k-1)}{2} & \text{si } l = j \\ \frac{k(k-1)}{2} & \text{si } k = l - 1 \\ 0 & \text{otro caso.} \end{cases}$$

Esto es, es un proceso de muerte que inicia en el estado i y termina en el estado 1.

Por su parte, Griffiths [11] estudió desde la perspectiva de las difusiones las líneas de descendencia del proceso de Wright-Fisher con mutación partiendo de un solo individuo al tiempo 0 y las ramificaciones ocurren en cada generación cuando un miembro de la línea de descendencia se reproduce. Bajo el supuesto de que las mutaciones no se incluyen en las líneas de descendencia del padre pero sí se genera una nueva línea, presentó, entre otras cosas, la función de distribución conjunta para el número de líneas de descendencia sobrevivientes del tiempo 0 al tiempo t cuando la tasa de mutación es la misma para los distintos tipos de individuos.

Representamos por $A_n^\theta(t)$ al número de líneas de descendencia que no han tenido mutaciones hacia atrás en el tiempo en un árbol coalescente de n individuos y θ representa la tasa de mutación. En [11] y [32] estudiaron la distribución de este proceso de tal manera que

$$\mathbb{P}(A_n^\theta(t) = j) = \sum_{k=j}^n e^{-k(k+\theta-1)t/2} \frac{(-1)^{k-j}(\theta + 2k - 1)(\theta + j)_{(k-1)} n_{[k]}}{j!(k-j)!(n+\theta)_{(k)}},$$

para $j = 0, 1, \dots, n$, $a_{(j)} = a(a+1) \cdots (a+j-1)$ que ya habíamos definido antes y $b_{[j]} = b(b-1) \cdots (b-j+1)$. Considerando que $\frac{n_{[k]}}{n+\theta_{(k)}} = 1$ cuando n es infinito, se sigue cumpliendo que

$$\mathbb{P}(A_\infty^\theta(t) = j) = \sum_{k=j}^n e^{-k(k+\theta-1)t/2} \frac{(-1)^{k-j}(\theta + 2k - 1)(\theta + j)_{(k-1)}}{j!(k-j)!}. \quad (2.21)$$

El proceso $\{A_n^\theta(t)\}$ es un proceso de muerte que representa el número de individuos que hay al tiempo t después de que ha ocurrido coalescencia o mutación con tasa $\frac{k(k+\theta-1)}{2}$ para $k = n, n-1, \dots, 1$. Podemos notar que la ecuación (2.5) es igual a la ecuación (2.4.2), por lo tanto, el proceso de muerte $\{A_\infty^\theta\}$ corresponde al proceso de muerte $\{D_t\}$ que se utiliza para encontrar la descomposición espectral del proceso de Wright-Fisher.

Capítulo 3

Simulación de trayectorias de la difusión de Wright-Fisher

Ya que hemos estudiado el proceso de Wright-Fisher desde el punto de vista teórico, ahora estamos interesados en simular sus trayectorias. Dos de los principales objetivos de la simulación de trayectorias de procesos obtenidos como solución de una ecuación diferencial estocástica es observar el comportamiento de la trayectoria y por otro lado, encontrar el valor esperado de algunos funcionales que no es posible obtener una expresión analítica explícita. En general, si se conoce la función de densidad de transición es sencillo simular trayectorias. Sin embargo, en la mayoría de los procesos de difusión no es conocida por lo que uno de los principales métodos es la aproximación discreta a la solución continua de la ecuación diferencial estocástica, por ejemplo, la aproximación de Euler y Milstein.

En el presente capítulo estudiaremos algunos métodos de simulación de trayectorias del proceso de Wright-Fisher. Consideremos $\{X_t\}$ la difusión de Wright-Fisher, es decir, la solución a la ecuación diferencial estocástica.

$$dX_t = \frac{1}{2}(\alpha(1 - X_t) - \beta X_t)dt + \sqrt{X_t(1 - X_t)}dB_t, \quad (3.1)$$

donde $\alpha, \beta > 0$ son los parámetros y $\{B_t\}$ un movimiento browniano estándar.

Debido a que no se conoce de forma explícita la función de densidad de transición de la solución a la ecuación (3.1) para simular trayectorias se emplean métodos de aproximación. El problema de utilizar estos métodos en la difusión de Wright-Fisher es que las trayectorias salen del espacio de estados $[0,1]$, como consecuencia, es necesario desarrollar métodos alternativos.

En primer lugar se introducen las aproximaciones de Euler y Milstein, posteriormente, otro mecanismo de aproximación que preserva las fronteras del espacio de estados, nombrado BISS por sus siglas en inglés (Balanced Implicit Split Step), seguido de un algoritmo llamado exacto por utilizar la expansión en series de la función de transición, luego, un método que se basa en una descomposición latente y la propiedad de Markov. También comparamos dichos algoritmos gráficamente y finalmente presentamos un método para simular puentes que solo necesita poder simular trayectorias.

3.1. Métodos numéricos.

En primer lugar, consideramos algoritmos que generan trayectorias mediante técnicas numéricas de aproximación a la ecuación diferencial estocástica en (3.1). Los métodos más sencillos de implementar son las aproximaciones de Euler y Milstein, ambos métodos pueden encontrarse en [14]. Sin embargo, para la difusión de Wright-Fisher estos mecanismos no son muy buenos, ya que no siempre se preservan las fronteras del proceso debido a que solo toma valores entre 0 y 1. Posteriormente, enunciamos un método novedoso que

genera trayectorias dentro del espacio de estados del proceso.

Supongamos que queremos simular trayectorias del proceso $\{X_t\}$ en el intervalo de tiempo $[0, T]$ y consideramos la discretización del intervalo $0 = t_0 < t_1 < \dots < t_n = T$.

Sea $X_0 = x_0$ el valor inicial del proceso, adicionalmente, definimos $\Delta_i = t_{i+1} - t_i$ el incremento entre los tiempos t_i y t_{i+1} y $\Delta B_i = B_{t_{i+1}} - B_{t_i}$, tiene distribución normal de media cero y varianza Δ_i . Usualmente se consideran incrementos de tiempo constantes, o sea, $\Delta_i = \Delta_j$ para todo i, j . A lo largo de este trabajo haremos uso de este supuesto, en este caso, lo denotaremos simplemente por Δ .

3.1.1. Aproximación de Euler

La aproximación de Euler al proceso de Wright-Fisher que inicia en x_0 es el proceso continuo $\{Y_t\}$ dado por

$$Y_{t_{i+1}} = Y_{t_i} + \frac{1}{2} \left(\alpha(1 - Y_{t_i}) - \beta Y_{t_i} \right) \Delta_i + \sqrt{Y_{t_i}(1 - Y_{t_i})} \Delta B_i, \quad (3.2)$$

para $i = 0, 1, \dots, n - 1$ y $Y_0 = x_0$.

En la Figura 3.1 observamos una trayectoria simulada con la ecuación (3.2) para cinco conjuntos de parámetros, $(\alpha = 0.7, \beta = 0.7)$, $(\alpha = 1, \beta = 1)$, $(\alpha = 2, \beta = 2)$, $(\alpha = 3, \beta = 1)$ y $(\alpha = 10, \beta = 20)$. Las trayectorias se generan sobre el intervalo de tiempo $[0, 5]$ e inician en el estado $X_0 = 0.5$ con incremento de tiempo constante $\Delta = 0.01$.

Recordando que la distribución estacionaria del proceso de Wright-Fisher existe siempre y cuando $\alpha \geq 1$ y $\beta \geq 1$. Además, sigue una distribución Beta(α, β), donde si $Y \sim \text{Beta}(\alpha, \beta)$, tiene esperanza y varianza

$$\mathbb{E}(Y) = \frac{\alpha}{\alpha + \beta},$$

$$\text{Var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

De manera que cuando los parámetros son iguales la trayectoria estará alrededor del punto medio del espacio de estados, esto tiene sentido pues como las tasas de mutación son iguales la proporción de individuos de tipo A y a , en promedio, será la misma.

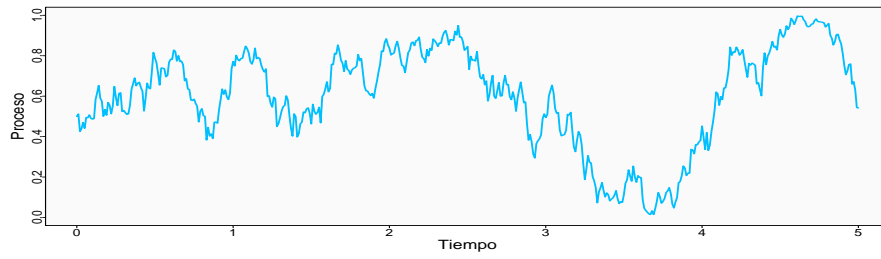
Por el contrario, si la tasa de mutación $\alpha > \beta$ la proporción de individuos de tipo A tenderá a ser mayor y si la tasa de mutación $\beta > \alpha$ la proporción de individuos de tipo A tenderá a ser menor teniendo en cuenta que como las fronteras son de entrada cuando $\alpha \geq 1$ y $\beta \geq 1$ el proceso no alcanzará los estados cero o uno.

Por otro lado, otra cuestión importante por señalar es respecto a la varianza, podemos concluir que mientras más grandes los parámetros, la varianza es menor y, por tanto, la trayectoria será menos variable.

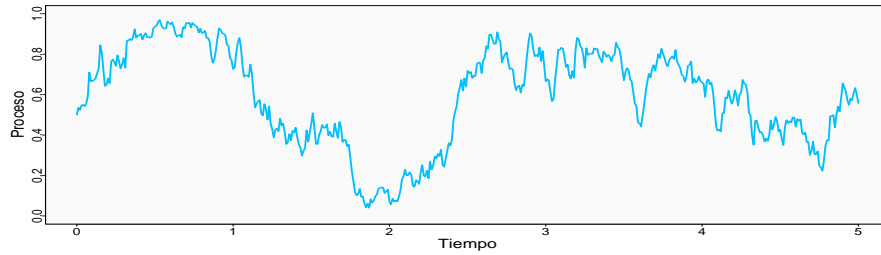
Ahora, simulamos 2000 trayectorias en el intervalo de tiempo $[0, 5]$ que inician en $X_0 = 0.5$ con los parámetros antes mencionados cuya distribución estacionaria existe y comparamos con la distribución estacionaria del proceso de Wright-Fisher.

Para realizar la comparación tomamos distintos puntos del tiempo, $t = 0.5$, $t = 1$, el punto medio del intervalo $t = 2.5$ y el punto final $t = 5$ con distintas discretizaciones $\Delta = 0.01, 0.002, 0.001$ y 0.0005 .

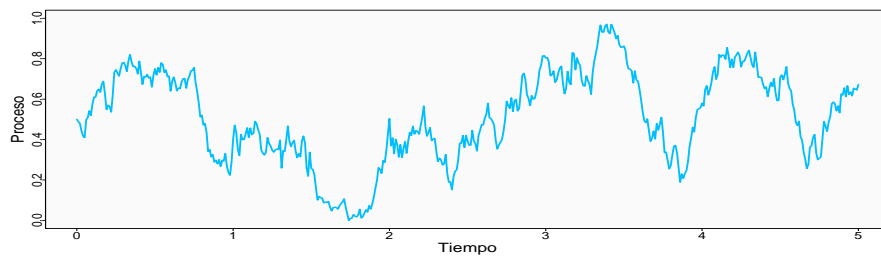
En la Tabla 3.1 presentamos los *p-values* de la prueba de Kolmogorov-Smirnov de bondad de ajuste para una muestra, la cual compara a la muestra con una distribución dada. La hipótesis nula corresponde a que la muestra sigue una distribución Beta de parámetros α y β y en contraste, la hipótesis alternativa donde la muestra no proviene de una distribución Beta. Posteriormente, presentamos algunas comparaciones gráficas.



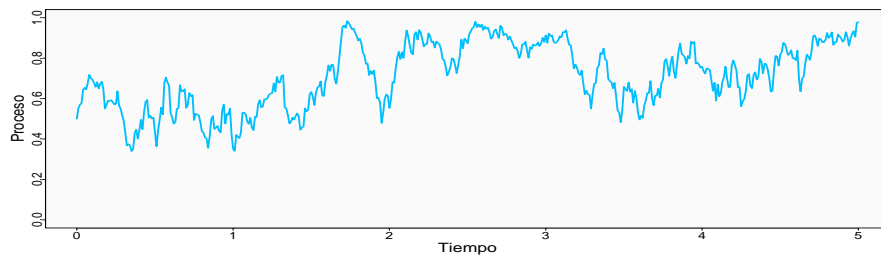
(a) $\alpha = 0.7$ y $\beta = 0.7$



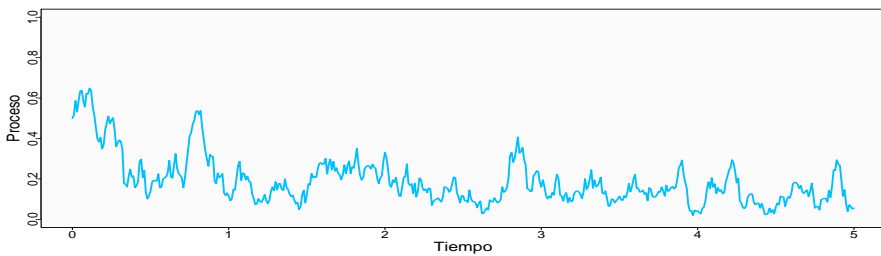
(b) $\alpha = 1$ y $\beta = 1$



(c) $\alpha = 2$ y $\beta = 2$



(d) $\alpha = 3$ y $\beta = 1$



(e) $\alpha = 10$ y $\beta = 20$

Figura 3.1: Trayectorias simuladas con el método de Euler en el intervalo de tiempo $[0,5]$ con $\Delta = 0.01$ y $X_0 = 0.5$.

Parámetros	Tiempo	$\Delta = 0.01$	$\Delta = 0.002$	$\Delta = 0.001$	$\Delta = 0.0005$
$\alpha = 1$	0.5	6.4e-13	9.2e-12	6.5e-11	3.9e-11
	1	4.3e-09	5.46e-06	1.7e-04	1.8e-04
$\beta = 1$	2.5	3.8e-06	3.0e-04	1.0e-04	0.001
	5	0.004	0.051	0.016	0.125
$\alpha = 2$	0.5	0.123	0.537	0.501	0.076
	1	0.435	0.382	0.286	0.366
$\beta = 2$	2.5	0.496	0.720	0.243	0.609
	5	0.787	0.303	0.781	0.614
$\alpha = 3$	0.5	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	1	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
$\beta = 1$	2.5	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	5	1.2e-13	9.2e-12	1.1e-09	3.1e-05
$\alpha = 10$	0.5	0	0.667	0.480	0.076
	1	0	0.321	0.379	0.900
$\beta = 20$	2.5	0	0.500	0.601	0.447
	5	0	0.490	0.799	0.013

Tabla 3.1: Resultados de la pruebas de bondad de ajuste para simulaciones con el método de Euler contra la distribución estacionaria.

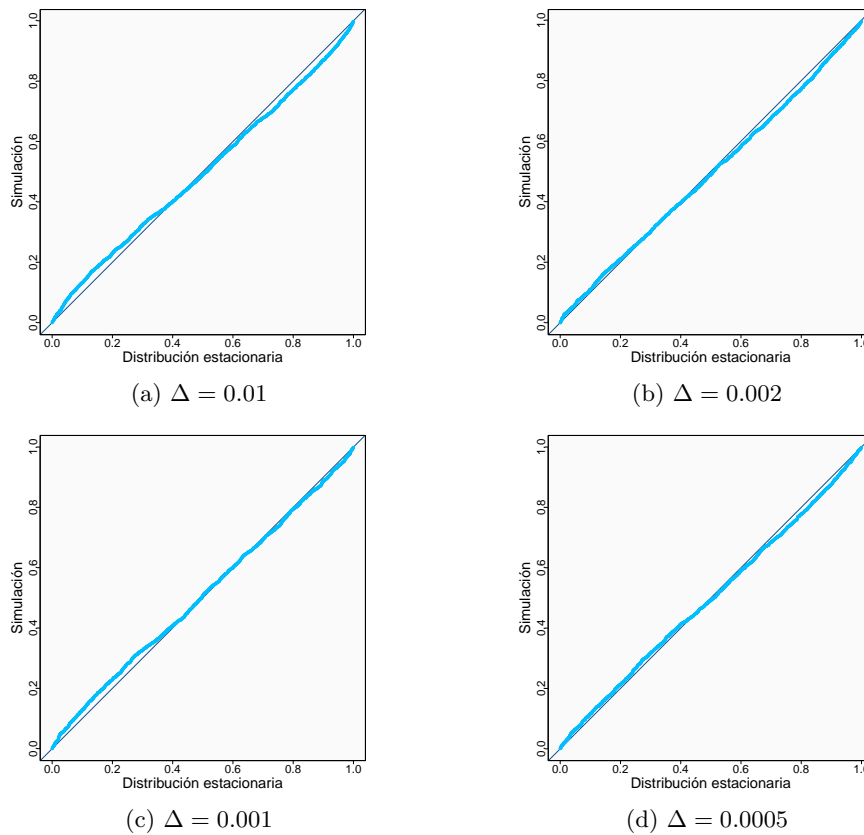


Figura 3.2: QQplot comparando la muestra de 2000 trayectorias al tiempo $t = 5$ simuladas con el método de Euler con parámetros ($\alpha = 1$, $\beta = 1$) contra la distribución estacionaria.

En primer lugar, de acuerdo con los resultados de las pruebas para ($\alpha = 1$, $\beta = 1$) pareciera que el proceso

alcanza la distribución estacionaria en $t = 5$ excepto quizá con $\Delta = 0.01$. Veamos en la Figura 3.2 las gráficas qqplot que contrastan una variable aleatoria Beta en el eje “x” contra la muestra obtenida de las simulaciones al tiempo final, $t = 5$, en el eje “y”. Por ejemplo, se observan ligeras diferencias respecto a la recta identidad alrededor del cuantil 0.2 y 0.8. La conclusión cuando $\Delta = 0.001$ dependerá de la exigencia que requiera la aplicación.

Por otro lado, cuando $(\alpha = 2, \beta = 2)$, a diferencia del ejemplo anterior, observamos muy buenos resultados cuando comparamos con la distribución estacionaria desde tiempos pequeños. En este caso graficamos solamente la función de densidad de la distribución estacionaria del proceso respecto a la curva ajustada con las simulaciones al tiempo $t = 5$, Figura 3.3.

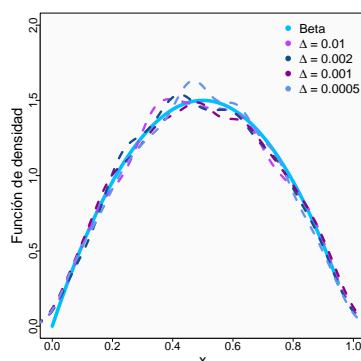


Figura 3.3: Comparación de la función de densidad de una variable aleatoria $Beta(2,2)$ y la muestra de 2000 trayectorias al tiempo $t = 5$ obtenidas con el método de Euler.

En los siguientes dos casos comparamos cuando los parámetros no son iguales, considerando $(\alpha = 3, \beta = 1)$, según los resultados de las pruebas de bondad de ajuste es claro que la muestra obtenida por medio de simulaciones no corresponden a una distribución Beta, a pesar de que el ajuste mejora al incrementar el tiempo y hacer Δ pequeño, no se obtienen resultados favorables. Para evidenciar estas diferencias mostramos en la Figura 3.4 el qqplot con las distintas discretizaciones, la función de distribución empírica y la función de densidad al tiempo $t = 5$, en cualquier gráfica que observemos se nota el sesgo que existe respecto a la distribución estacionaria.

Finalmente, al hacer más grandes los parámetros con $(\alpha = 10, \beta = 20)$ es suficiente con $\Delta = 0.002$ para que no se rechace la hipótesis nula, incluso en $t = 0.5$. Esto nos da indicios de que el proceso alcanza rápidamente la distribución estacionaria.

A continuación, evaluamos la cantidad de simulaciones al comparar nuevamente con la distribución estacionaria. Considerando los mismos parámetros y tomamos la discretización $\Delta = 0.002$. A diferencia de los ejercicios anteriores, dejamos que el valor inicial sea un número aleatorio en el intervalo $[0,1]$.

En la Tabla 3.2 mostramos los p -values de las pruebas. Es curioso que con $(\alpha = 1, \beta = 1)$ se obtienen buenos resultados si el número de simulaciones es chico, de entre 100 y 500, mientras que con una cantidad considerable de simulaciones se rechaza la hipótesis de igualdad en distribución. Por su parte, para $(\alpha = 2, \beta = 2)$ y $(\alpha = 10, \beta = 20)$, en general, se obtienen muestras donde no se rechaza que sigan una distribución Beta con cualquiera de las cantidades estudiadas. Respecto a los parámetros $(\alpha = 3, \beta = 1)$ los resultados son similares a los obtenidos en la Tabla 3.1.

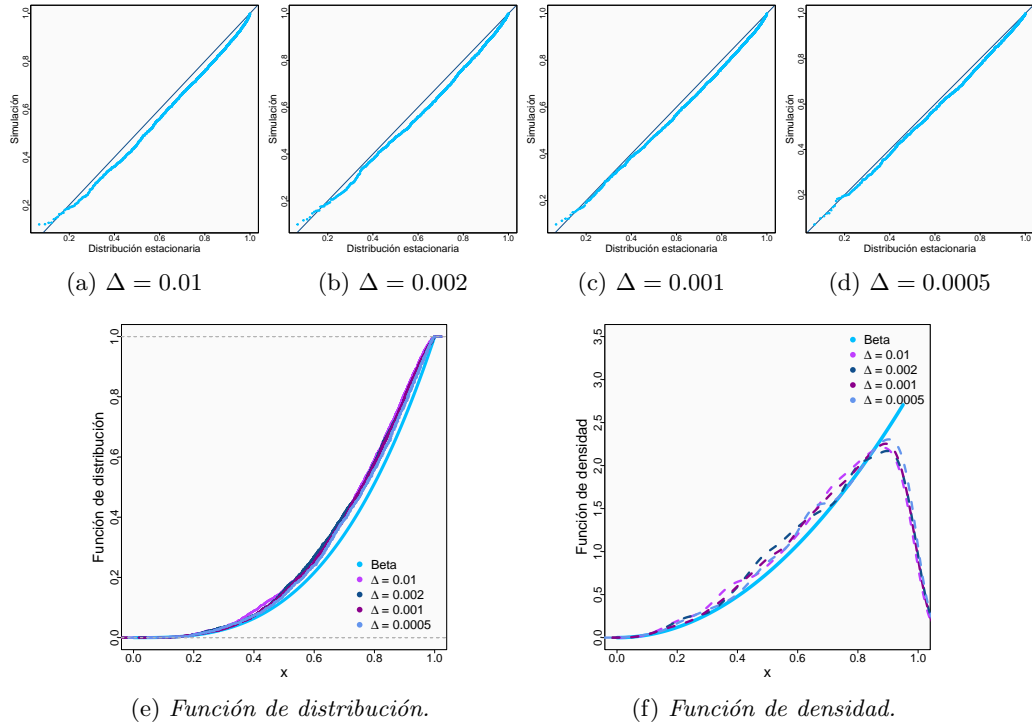


Figura 3.4: Comparaciones gráficas de la muestra de 2000 trayectorias al tiempo las $t = 5$ simuladas con el método de Euler de parámetros ($\alpha = 3$, $\beta = 1$) contra la distribución estacionaria.

Parámetros	Tiempo	100	500	1000	1500	2000	5000	10000
$\alpha = 1$	0.5	0.321	0.548	0.086	0.002	2.2e-04	4.1e-08	4.7e-14
	1	0.079	0.042	0.005	6.6e-04	6.5e-06	1.4e-10	< 2.2e-16
	2.5	0.437	0.040	0.007	0.001	8.8e-05	1.4e-10	< 2.2e-16
	5	0.877	0.133	0.115	0.024	0.023	3.3e-04	1.0e-06
$\alpha = 2$	0.5	0.288	0.082	0.268	0.030	0.058	0.015	0.004
	1	0.770	0.669	0.634	0.246	0.151	0.825	0.951
	2.5	0.442	0.443	0.367	0.382	0.335	0.812	0.189
	5	0.555	0.283	0.396	0.605	0.816	0.350	0.039
$\alpha = 3$	0.5	1.3e-07	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
	1	0.007	4.6-13	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
	2.5	0.008	2.1e-10	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
	5	0.190	1.7e-04	4.0e-08	2.8e-10	2.0e-14	0	0
$\alpha = 10$	0.5	0.932	0.799	0.277	0.276	0.476	0.832	0.466
	1	0.385	0.545	0.550	0.914	0.840	0.966	0.998
	2.5	0.814	0.544	0.167	0.031	0.037	0.021	0.064
	5	0.484	0.515	0.887	0.708	0.645	0.438	0.255

Tabla 3.2: Resultados de la pruebas de bondad de ajuste para simulaciones con el método de Euler contra la distribución estacionaria con distinto número de simulaciones.

3.1.2. Aproximación de Milstein

La aproximación de Milstein es más precisa que la aproximación de Euler pues agrega una corrección al incremento estocástico mediante un término de segundo orden.

Para obtener este término se utiliza el lema de Itô y la expansión de Taylor, por lo que la aproximación queda de la siguiente forma,

$$Y_{t_{i+1}} = Y_{t_i} + \frac{1}{2} \left(\alpha(1 - Y_{t_i}) - \beta Y_{t_i} \right) \Delta_i + \sqrt{Y_{t_i}(1 - Y_{t_i})} \Delta B_i + (1 - 2Y_{t_i})(\Delta B_i^2 - \Delta_i).$$

con $i = 0, 1, \dots, n - 1$ y $Y_0 = x_0$. En este caso, análogo a la aproximación de Euler el método será bueno si la discretización converge a cero.

Nuevamente realizamos simulaciones para cinco conjuntos de parámetros, $(\alpha = 0.7, \beta = 0.7)$, $(\alpha = 1, \beta = 1)$, $(\alpha = 2, \beta = 2)$, $(\alpha = 3, \beta = 1)$ y $(\alpha = 10, \beta = 20)$. Estudiamos trayectorias en el intervalo de tiempo $[0,5]$ que inician en el estado $X_0 = 0.5$ con incrementos de tiempo $\Delta = 0.01, 0.002, 0.001$ y 0.0005 . En la Figura 3.5 mostramos una trayectoria con cada uno de los parámetros y $\Delta = 0.01$.

Para evaluar el comportamiento de las simulaciones realizamos pruebas de bondad de ajuste en los tiempos $t = 0.5, 1, 2.5$ y 5 . Los *p-values* que obtuvimos se encuentran en la Tabla 3.3, a pesar de que el método de Milstein tiene un término más en la aproximación los resultados son bastante similares a la aproximación Euler.

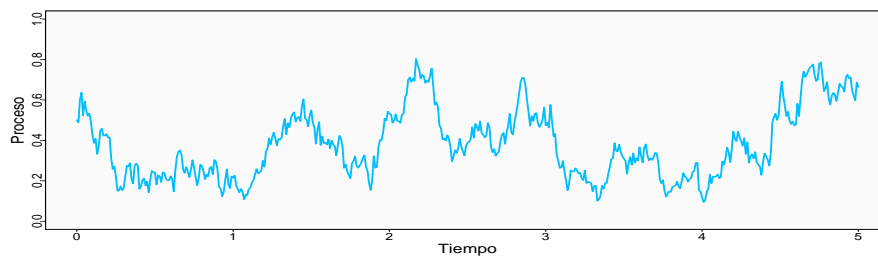
Parámetros	Tiempo	$\Delta = 0.01$	$\Delta = 0.002$	$\Delta = 0.001$	$\Delta = 0.0005$
$\alpha = 1$	0.5	1.0e-09	5.7e-09	3.5e-11	1.1e-11
	1	1.1e-05	1.8e-06	5.6e-07	2.5e-04
$\beta = 1$	2.5	7.6e-08	0.002	7.9e-04	3.9e-04
	5	3.3e-04	0.016	0.067	0.009
$\alpha = 2$	0.5	0.007	0.065	0.087	0.426
	1	0.687	0.982	0.671	0.437
$\beta = 2$	2.5	0.161	0.413	0.494	0.022
	5	0.284	0.259	0.281	0.753
$\alpha = 3$	0.5	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	1	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
$\beta = 1$	2.5	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	5	5.6e-15	2.0e-07	1.2e-05	7.1e-13
$\alpha = 10$	0.5	0.208	0.063	0.142	0.954
	1	0.053	0.209	0.223	0.771
$\beta = 20$	2.5	0.569	0.186	0.877	0.413
	5	0.050	0.221	0.930	0.630

Tabla 3.3: Resultados de la pruebas de bondad de ajuste para simulaciones con el método de Milstein contra la distribución estacionaria.

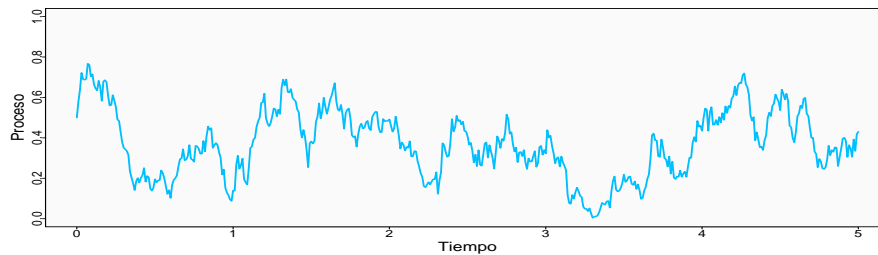
Como ejemplo ponemos en la Figura 3.6 las gráficas con los parámetros $(\alpha = 10, \beta = 20)$ al tiempo final, $t = 5$, con las distintas discretizaciones, donde tenemos que las simulaciones siguen la distribución deseada.

En la Tabla 3.4 reportamos los *p-values* de la prueba Kolmogorov-Smornov para los parámetros que hemos estado trabajando, $\Delta = 0.002$ y X_0 un punto aleatorio entre cero y uno considerando distinto número de simulaciones (100, 500, 1000, 1500, 2000, 5000 y 10000). Nuevamente los resultados son similares a los obtenidos con la aproximación Euler. La muestra simulada con $(\alpha = 1, \beta = 1)$ se comportan mejor cuando la cantidad de simulaciones es menor. El desempeño del algoritmo con $(\alpha = 3, \beta = 1)$ es deficiente mientras que con $(\alpha = 2, \beta = 2)$ y $(\alpha = 10, \beta = 20)$ se comporta bastante bien tanto con muestras pequeñas como

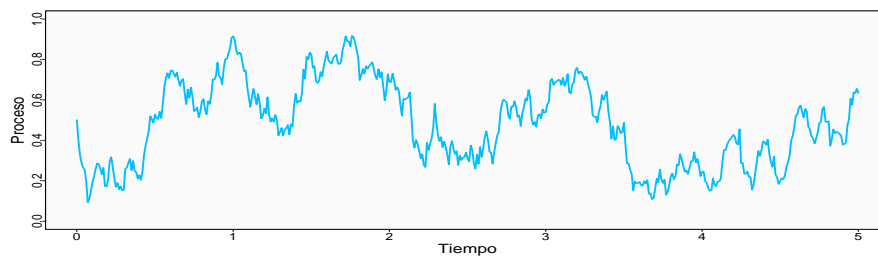
con muestras grandes. Ya que se obtienen rendimientos similares a los de la Tabla 3.3 podemos concluir que no afecta desde donde inicia la trayectoria.



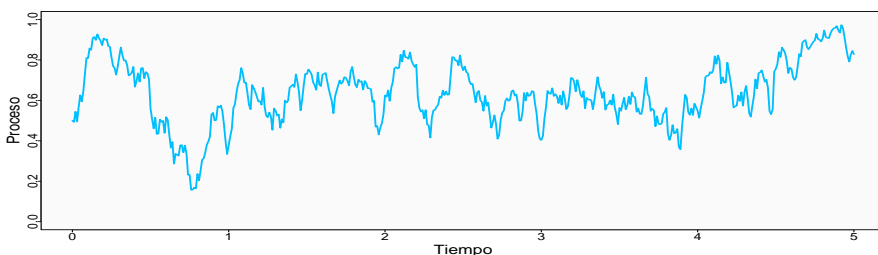
(a) $\alpha = 0.7$ y $\beta = 0.7$



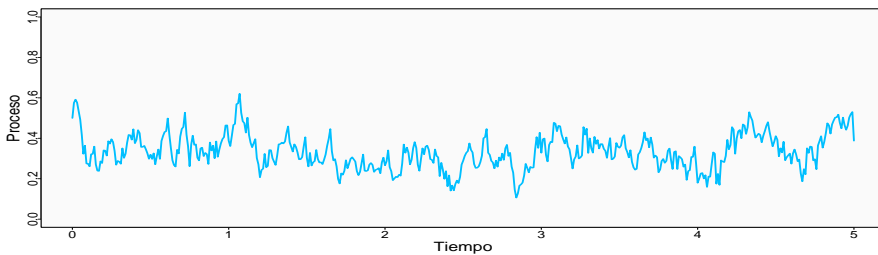
(b) $\alpha = 1$ y $\beta = 1$



(c) $\alpha = 2$ y $\beta = 2$



(d) $\alpha = 3$ y $\beta = 1$



(e) $\alpha = 10$ y $\beta = 20$

Figura 3.5: Trayectorias simuladas con el método de Milstein en el intervalo de tiempo $[0,5]$ con $\Delta = 0.01$ y $X_0 = 0.5$.

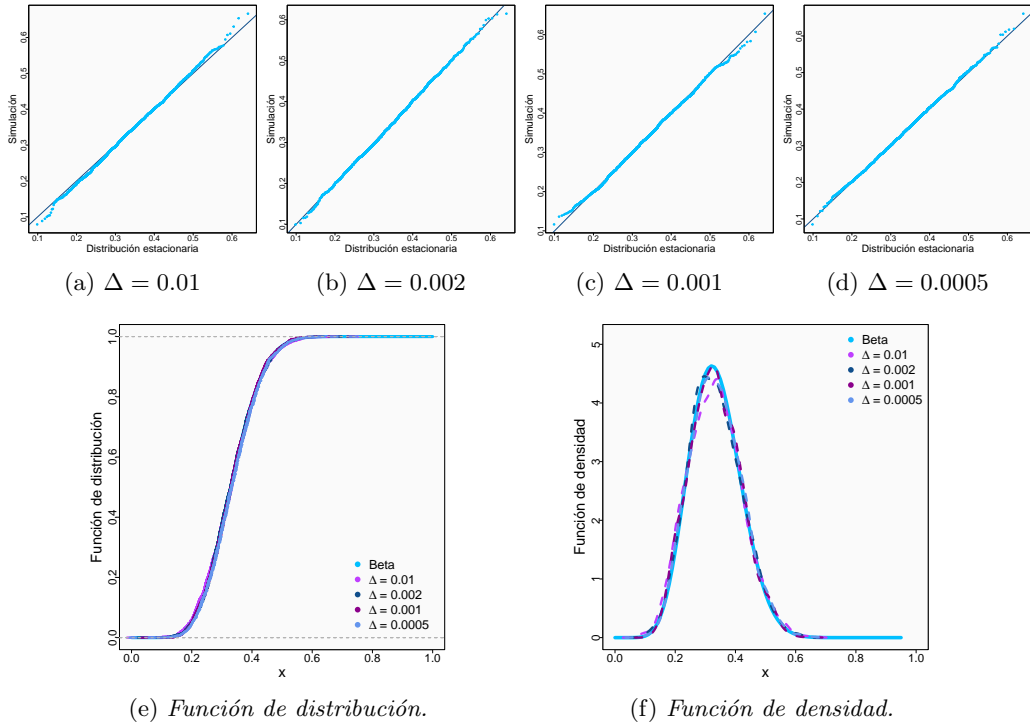


Figura 3.6: Comparaciones gráficas de la muestra de 2000 trayectorias al tiempo las $t = 5$ simuladas con el método de Milstein de parámetros ($\alpha = 10$, $\beta = 20$) contra la distribución estacionaria.

Parámetros	Tiempo	100	500	1000	1500	2000	5000	10000
$\alpha = 1$	0.5	0.182	0.403	0.060	0.007	0.013	2.1e-06	5.6e-15
	1	0.647	0.350	0.064	0.005	2.1e-04	8.0e-12	< 2.2e-16
$\beta = 1$	2.5	0.342	0.040	0.001	4.4e-05	3.0e-06	2.0e-12	< 2.2e-16
	5	0.527	0.541	0.019	0.020	0.010	1.2e-06	6.2e-08
$\alpha = 2$	0.5	0.711	0.724	0.246	0.081	0.024	0.005	0.004
	1	0.691	0.479	0.791	0.880	0.978	0.817	0.900
$\beta = 2$	2.5	0.947	0.612	0.596	0.404	0.218	0.532	0.271
	5	0.541	0.491	0.012	0.002	0.004	0.031	0.016
$\alpha = 3$	0.5	1.6e-07	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
	1	3.8e-05	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
$\beta = 1$	2.5	0.001	3.9e-11	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
	5	0.044	0.003	4.4e-07	6.3e-09	8.4e-10	0	0
$\alpha = 10$	0.5	0.422	0.162	0.780	0.704	0.260	0.276	0.470
	1	0.385	0.545	0.550	0.914	0.840	0.966	0.998
$\beta = 20$	2.5	0.664	0.524	0.925	0.502	0.568	0.301	0.829
	5	0.016	0.154	0.851	0.846	0.335	0.847	0.864

Tabla 3.4: Resultados de la pruebas de bondad de ajuste para simulaciones con el método de Milstein contra la distribución estacionaria con distinto número de simulaciones.

Ya que los algoritmos de Euler y Milstein basados en aproximaciones a la ecuación diferencial estocástica no siempre generan trayectorias que permanecen en el intervalo $[0,1]$ debido a que los incrementos ΔB_i cerca de la frontera podrían ser lo suficientemente grandes como para salir de ella se utilizan algunos ajustes.

Por este motivo, para evitar sesgo en las simulaciones de los métodos de Euler y Milstein generamos trayectorias hasta obtener una que cae completamente en el intervalo $[0,1]$, lo cual implica un costo computacional elevado y el tiempo de ejecución para generar trayectorias aceptables sea mayor. Por ejemplo, para generar una trayectoria aceptable con el método de Euler y parámetros $(\alpha = 1, \beta = 1)$ en promedio, se tuvieron que generar aproximadamente tres trayectorias.

Otro tipo de modificaciones que comúnmente se utilizan es que cuando el proceso sale de las fronteras, se generan tantos valores del incremento browniano, ΔB_i , como sean necesarios hasta que la solución queda dentro del intervalo. Alternativamente, si el proceso se vuelve negativo o mayor a uno, se asigna en ese tiempo el valor de cero o uno, según corresponda. Sin embargo, estos ajustes pueden generar sesgo en la solución numérica.

3.1.3. Método BISS

El siguiente algoritmo propuesto por Dangerfield, Kay, MacNamara y Burrage [4] conocido como método BISS (Balanced Implicit Split Step) es una aproximación numérica que preserva las fronteras del proceso. Este algoritmo es una combinación de dos algoritmos existentes, BIM (Balanced Implicit Method) introducido en [21] y el método de Moro and Schurz dado en [23].

El algoritmo desarrollado en [4] se basa en la ecuación diferencial estocástica

$$dY_t = (A - (A + B)Y_t)dt + C\sqrt{Y_t(1 - Y_t)}dB_t, \quad (3.3)$$

donde A , B y C son constantes positivas. Esta ecuación es utilizada para modelar la dinámica de los canales de iones.

Notemos que la ecuación (3.3) es una reparametrización de la ecuación diferencial estocástica de Wright-Fisher que estamos trabajando, por lo que si $A = \frac{\alpha}{2}$, $B = \frac{\beta}{2}$ y $C = 1$ recuperamos la ecuación (3.1) que nos interesa simular.

A continuación presentamos dos definiciones conceptualizadas por Schurz en [29] importantes para el desarrollo del BIM y, como consecuencia, para la aproximación BISS.

Definición 3.1.1. *Sea Y_{t_i} una aproximación a la solución de la ecuación (3.3) al tiempo t_i calculada mediante algún método numérico estocástico. Se dice que Y_{t_i} posee vida eterna si*

$$\mathbb{P}(Y_{t_i} \in [0, 1] | Y_{t_0} \in [0, 1]) = 1 \quad \text{para toda } i \geq 0.$$

Si la solución numérica no posee vida eterna, se dice que tiene vida finita y es cuando la aproximación puede salir del espacio de estados $[0,1]$ con probabilidad positiva. La siguiente definición relaja este supuesto y asegura que la aproximación permanezca en el intervalo $[0,1]$ en un paso.

Definición 3.1.2. *Sea Y_{t_i} una aproximación a la solución de la ecuación (3.3) al tiempo t_i calculada mediante algún método numérico estocástico. Se dice que Y_{t_i} posee ϵ -vida si*

$$\mathbb{P}(Y_{t_i} \in [0, 1] | Y_{t_{i-1}} \in [\epsilon, 1 - \epsilon]) = 1 \quad \text{para alguna } \epsilon \geq 0.$$

Bajo la definición de ϵ -vida observamos que con probabilidad positiva, Y_{t_i} puede caer en el conjunto $[0, \epsilon) \cup (1 - \epsilon, 1]$ lo cual no garantiza que al tiempo t_{i+1} la aproximación permanezca en la región adecuada, de manera que no se garantiza que el método preserve las fronteras en todo el tiempo.

BIM

El BIM por sus siglas en inglés (Balanced Implicit Method) fue introducido por Milstein, Platen y Schurz en [21], quienes desarrollaron un método implícito para encontrar la solución numérica a una ecuación diferencial estocástica en d dimensiones el cual nombraron método Balanceado. Lo interpretan como una familia de métodos específicos que proporcionan una especie de equilibrio entre la aproximación de los términos estocásticos en los métodos numéricos.

Describimos brevemente la idea general de la aproximación para ecuaciones diferenciales estocásticas en una sola dimensión, por su parte, las condiciones necesarias y el estudio de la convergencia del método puede encontrarse en [21].

Sea $\{X_t\}$ un proceso estocástico con $\mathbb{E}[(X_0)^2] < \infty$ que satisface la ecuación diferencial estocástica

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

μ y σ son funciones Lipschitz que cumplen la condición de crecimiento lineal y $\{B_t\}$ un movimiento browniano estándar.

La familia de métodos puede escribirse como:

$$Y_{t_{i+1}} = Y_{t_i} + \mu(t_i, Y_{t_i})\Delta_i + \sigma(t_i, Y_{t_i})\Delta B_i + D_i(Y_{t_i} - Y_{t_{i+1}}), \quad (3.4)$$

donde

$$D_i = d_0(t_i, Y_{t_i})\Delta_i + d_1(t_i, Y_{t_i})|\Delta B_i|$$

con d_j , $j = 0, 1$ funciones. La elección de d_0 y d_1 podría afectar la estabilidad de la aproximación por lo que debe elegirse cuidadosamente.

Despejando la ecuación (3.4) obtenemos la solución obtenida con el BIM.

$$Y_{t_{i+1}} = Y_{t_i} + (1 + D_i)^{-1}(\mu(t_i, Y_{t_i})\Delta_i + \sigma(t_i, Y_{t_i})\Delta B_i).$$

Luego, utilizando el BIM para dar una aproximación a la solución de la ecuación (3.3),

$$Y_{t_{i+1}} = Y_{t_i} + (A - (A + B)Y_{t_i})\Delta_i + C\sqrt{Y_{t_i}(1 - Y_{t_i})}\Delta B_i + D(Y_{t_i})(Y_{t_i} - Y_{t_{i+1}}), \quad (3.5)$$

con $D(y) = d_0(y)\Delta_i + d_1(y)|\Delta B_i|$, a $D(y)$ se le llama sistema de funciones de control.

Schurz demostró en [29] que con una elección adecuada de las funciones de control para el proceso de Wright Fisher sin deriva ($A = 0$ y $B = 0$) el BIM posee ϵ -vida. Es decir, si tomamos $d_0(y) = 0$ pues el proceso no tiene deriva y

$$d_1(y) = \begin{cases} \sqrt{\frac{1-\epsilon}{\epsilon}} & \text{si } y < \epsilon \\ \sqrt{\frac{1-y}{y}} & \text{si } \epsilon \leq y < \frac{1}{2} \\ \sqrt{\frac{y}{1-y}} & \text{si } \frac{1}{2} \leq y \leq 1 - \epsilon \\ \sqrt{\frac{1-\epsilon}{\epsilon}} & \text{si } y > 1 - \epsilon, \end{cases} \quad (3.6)$$

la aproximación BIM posee ϵ -vida. De manera que podemos reescribir la solución (3.5) como

$$Y_{t_{i+1}} = Y_{t_i} + C\sqrt{Y_{t_i}(1 - Y_{t_i})}\Delta B_i + d_1(Y_{t_i})|\Delta B_i|(Y_{t_i} - Y_{t_{i+1}}),$$

lo cual implica que,

$$Y_{t_{i+1}} = Y_{t_i} + \frac{C\sqrt{Y_{t_i}(1 - Y_{t_i})}}{1 - d_1(Y_{t_i})} \frac{\Delta B_i}{|\Delta B_i|}.$$

Método por pasos divididos.

Moro y Schurz proponen en [23] un mecanismo basado en los métodos por pasos divididos (*Split step*) para encontrar la solución a una ecuación diferencial estocástica que está sujeta a condiciones de frontera naturales. De tal manera que se garantice que las solución numérica permanezca en el espacio de estados y se preserve el comportamiento de las fronteras. Su enfoque consiste en explotar la estructura de la ecuación diferencial estocástica de tal forma que permita una división natural de la misma.

Supongamos que tenemos una ecuación diferencial de la forma

$$dX_t = (\mu_1(t, X_t) + \mu_2(t, X_t))dt + \sigma(t, X_t)dB_t. \quad (3.7)$$

La estructura general del algoritmo es descomponer la ecuación diferencial estocástica (3.7) de la siguiente forma

$$dX_t^1 = \mu_2(t, X_t^1)dt + \sigma(t, X_t^1)dB_t \quad (3.8)$$

$$dX_t^2 = \mu_1(t, X_t^2)dt, \quad (3.9)$$

tal que se conoce la solución fuerte exacta de X_t^1 o la probabilidad condicional $\mathbb{P}(X_t^1|X_0^1)$.

Notemos que la primer ecuación es una ecuación diferencial estocástica y la segunda es una ecuación diferencial ordinaria. Entonces, podemos aproximar la solución a la ecuación original por el proceso estocástico $\{Y_t\}$ en intervalos de tiempo $[t, t + \Delta_t]$ usando el siguiente algoritmo.

Algoritmo 1 Método pasos divididos.

- 1: Sabiendo el valor de Y_t se obtiene un valor intermedio \bar{Y}_t como solución exacta a la ecuación (3.8) $\bar{Y}_t = X_{t+\Delta_t}^1$ y con condición inicial $X_t^1 = Y_t$.
 - 2: \bar{Y}_t se usa como condición inicial para la ecuación (3.9) que se resuelve utilizando cualquier método numérico determinista para obtener un \bar{X}_t^2 . Entonces, $Y_{t+\Delta_t} = \bar{X}_{t+\Delta_t}^2$.
-

Es decir, la solución exacta de (3.8) se utiliza como condición inicial para la ecuación diferencial ordinaria (3.9) en ese tiempo. Algunas propiedades sobre la convergencia del método y ejemplos pueden encontrarse en [23].

Aplicando este método a la ecuación (3.3), de acuerdo a [3] puede separarse como sigue

$$dX_t^1 = \frac{C^2(1 - 2X_t^1)}{4}dt + C\sqrt{X_t^1(1 - X_t^1)}dB_t \quad (3.10)$$

$$dX_t^2 = \left[A - \frac{C^2}{4} - \left(A + B - \frac{C^2}{2} \right) X_t^2 \right] dt \quad (3.11)$$

Bajo este esquema, la ecuación (3.10) es equivalente a la ecuación diferencial estocástica de Stratonovich $dX_t^1 = C\sqrt{X_t^1(1 - X_t^1)} \circ dB_t$ que tiene solución

$$X_{t_{i+1}}^1 = \left(\text{sen} \left(\frac{C}{2} \Delta B_i + \text{sen}^{-1} \sqrt{X_{t_i}^1} \right) \right)^2, \quad (3.12)$$

que claramente está en el intervalo $[0,1]$, posteriormente, utilizando (3.12) como condición inicial de (3.11) tenemos que

$$X_t^2 = \frac{A - C^2/4}{C^2/2 - A - B} \left(e^{-(A+B-C^2/2)\Delta_i} - 1 \right) + e^{-(A+B-C^2/2)\Delta_i} \left(\text{sen} \left(\frac{C}{2} \Delta B_i + \text{sen}^{-1} \sqrt{X_{t_i}^1} \right) \right)^2. \quad (3.13)$$

Método BISS

Combinando los dos enfoques anteriores Dangerfield, Kay, MacNamara y Burrage [4] proponen una descomposición de la ecuación (3.3) similar al método de separación, la principal diferencia se encuentra en la manera de encontrar la solución a la primer ecuación, describimos a continuación el algoritmo.

La primera ecuación de la descomposición comprende únicamente el término de la difusión, es decir,

$$dX_t^1 = C\sqrt{X_t^1(1-X_t^1)}dB_t \quad (3.14)$$

y la segunda, es la ecuación diferencial ordinaria que corresponde a la deriva

$$dX_t^2 = \left(A - (A+B)X_t^2 \right) dt. \quad (3.15)$$

Posteriormente, la solución numérica de la primer ecuación (3.14) se utiliza como condición inicial de la segunda (3.15). Finalmente, la solución a (3.15) representa una aproximación a la ecuación diferencial estocástica original. La propuesta es utilizar el BIM para resolver (3.14) en cada tiempo utilizando la función de control (3.6) y como consecuencia el proceso tendrá ϵ -vida. Siguiendo con la ecuación diferencial ordinaria utilizamos la aproximación de Euler, lo que nos lleva a la siguiente aproximación

$$Y_{t_{i+1}} = Y_{t_i} + \left(A - (A+B)Y_{t_i} \right) \Delta_i + \frac{C\sqrt{Y_{t_i}(1-Y_{t_i})}}{1+d_1(Y_{t_i})} \frac{\Delta B_i}{|\Delta B_i|} (1 - (A+B)\Delta_i).$$

En el método BISS se determina ϵ de manera que asegure que para un incremento particular Δ la solución a la ecuación permanezca dentro de la barrera ϵ , es decir, dentro del intervalo $[\epsilon, 1 - \epsilon]$. Para ello tomamos

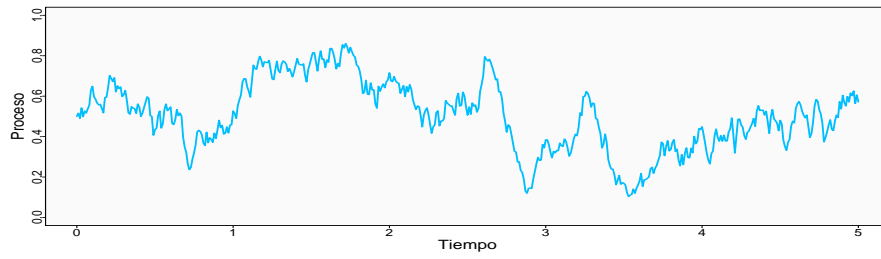
$$\epsilon = \min\{A\Delta, B\Delta, 1 - A\Delta, 1 - B\Delta\} > 0,$$

cuando Δ es suficientemente pequeño. Bajo estas condiciones, Dangerfield, Kay, MacNamara y Burrage prueban en [4] que el proceso posee ϵ -vida, además de la convergencia del método.

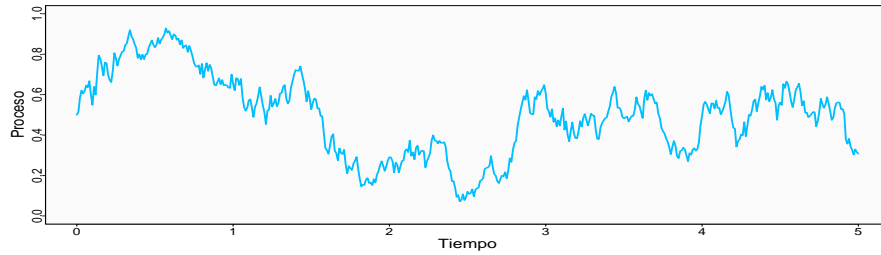
Ya que hemos descrito brevemente el método, realizamos el mismo estudio de las trayectorias tomando en cuenta la parametrización adecuada del proceso. Consideramos los cinco conjuntos de parámetros, $(\alpha = 0.7, \beta = 0.7)$, $(\alpha = 1, \beta = 1)$, $(\alpha = 2, \beta = 2)$, $(\alpha = 3, \beta = 1)$ y $(\alpha = 10, \beta = 20)$ y generamos 2000 trayectorias en el intervalo de tiempo $[0,5]$ con valor inicial $X_0 = 0.5$ e incrementos de tiempo $\Delta = 0.01, 0.002, 0.001$ y 0.0005 . En la Figura 3.7 observamos una trayectoria con cada uno de los parámetros y $\Delta = 0.01$.

También realizamos las pruebas de Kolmogorov-Smirnov en los tiempos $t = 0.5, 1, 2.5$ y 5 para cada uno de los parámetros que aceptan una distribución estacionaria con el fin de hacer la comparación. Los p -values se presentan en la Tabla 3.5. Es claro que cuando $\Delta = 0.01$, independientemente de los parámetros, las simulaciones no siguen una distribución Beta, sin embargo, observamos mejoras sustanciales a medida que Δ se hace pequeño. Así mismo, para complementar, presentamos algunas comparaciones gráficas analizando el comportamiento de las simulaciones respecto a la distribución estacionaria con distintos valores de Δ .

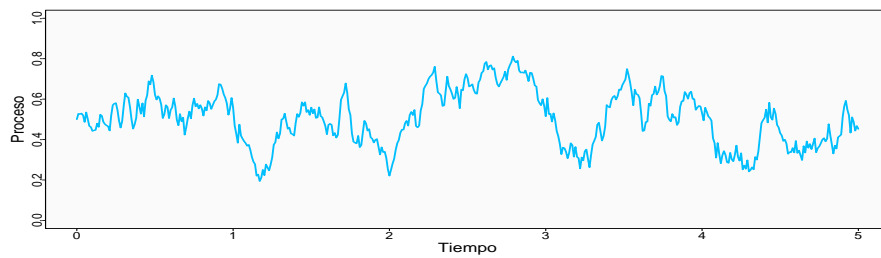
En primer lugar, observamos en la Figura 3.8 el qqplot de la muestra obtenida al tiempo $t = 5$ a partir de 2000 simulaciones de las trayectorias con parámetros $(\alpha = 1, \beta = 1)$ comparada con la distribución estacionaria, una variable aleatoria Beta con los mismos parámetros. Con los distintos valores de Δ observamos un comportamiento similar donde para valores menores al punto medio la gráfica queda por arriba de la recta identidad y para valores mayores, queda por debajo.



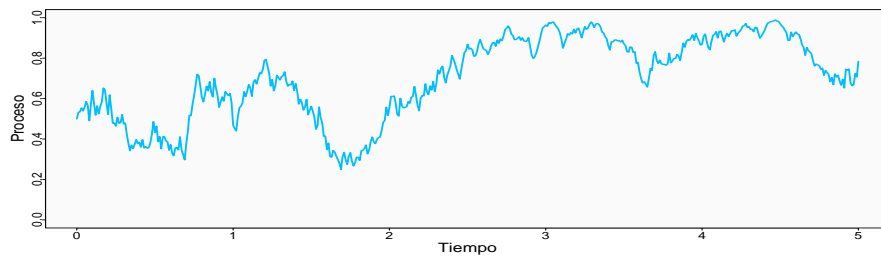
(a) $\alpha = 0.7$ y $\beta = 0.7$



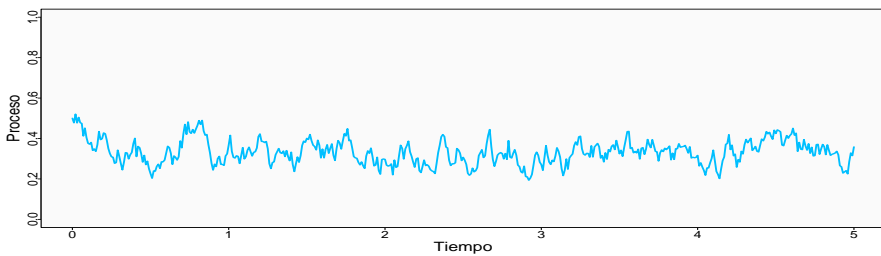
(b) $\alpha = 1$ y $\beta = 1$



(c) $\alpha = 2$ y $\beta = 2$



(d) $\alpha = 3$ y $\beta = 1$



(e) $\alpha = 10$ y $\beta = 20$

Figura 3.7: Trayectorias simuladas con el método BISS en el intervalo de tiempo $[0,5]$ con $\Delta = 0.01$ y $X_0 = 0.5$.

Parámetros	Tiempo	$\Delta = 0.01$	$\Delta = 0.002$	$\Delta = 0.001$	$\Delta = 0.0005$
$\alpha = 1$	0.5	$< 2.2e-16$	$3.9e-13$	$1.4e-11$	$1.4e-09$
	1	$5.9e-12$	$7.2e-05$	$4.2e-04$	0.037
$\beta = 1$	2.5	$6.0e-08$	$1.6e-04$	0.066	0.006
	5	$2.4e-08$	0.015	0.015	0.004
$\alpha = 2$	0.5	$6.8e-08$	$2.0e-06$	0.024	0.024
	1	$4.1e-08$	0.014	0.003	0.046
$\beta = 2$	2.5	$5.4e-07$	0.024	0.065	0.476
	5	$7.5e-07$	$4.7e-05$	0.093	0.190
$\alpha = 3$	0.5	0	$< 2.2e-16$	$< 2.2e-16$	$< 2.2e-16$
	1	0	$1.2e-14$	$6.3e-14$	$3.6e-09$
$\beta = 1$	2.5	0	$3.3e-07$	$3.8e-04$	0.012
	5	0	$1.6e-06$	$8.2e-04$	0.016
$\alpha = 10$	0.5	$2.3e-14$	0.002	0.105	0.020
	1	$1.9e-10$	0.005	0.005	0.028
$\beta = 20$	2.5	$2.2e-16$	$7.1e-04$	0.128	0.308
	5	$4.6e-12$	0.018	0.164	0.118

Tabla 3.5: Resultados de la pruebas de bondad de ajuste para simulaciones con el método BISS contra la distribución estacionaria.

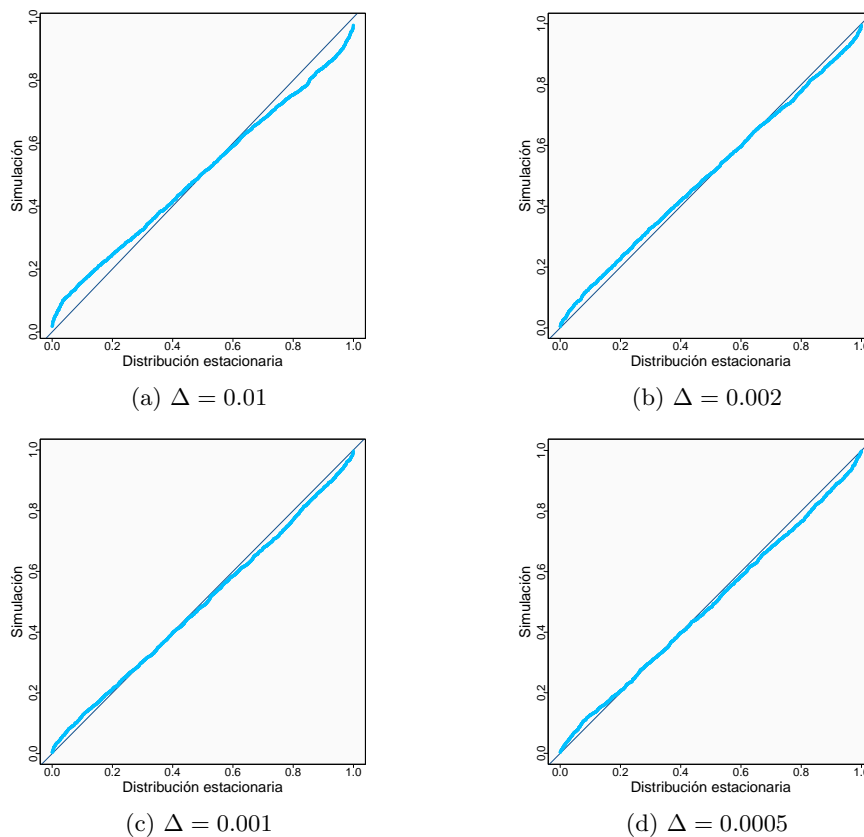


Figura 3.8: QQplot comparando la muestra de 2000 trayectorias al tiempo $t = 5$ simuladas con el método BISS con parámetros $(\alpha = 1, \beta = 1)$ contra la distribución estacionaria.

Seguido del caso cuando $(\alpha = 2, \beta = 2)$ mostramos la función de distribución y la función de densidad, en la Figura 3.9, donde es claro que mientras Δ se hace menor, obtenemos mejores resultados. A diferencia de los

dos métodos anteriores donde las trayectorias simuladas alcanzan la distribución estacionaria rápidamente, el método BISS requiere mayor atención.

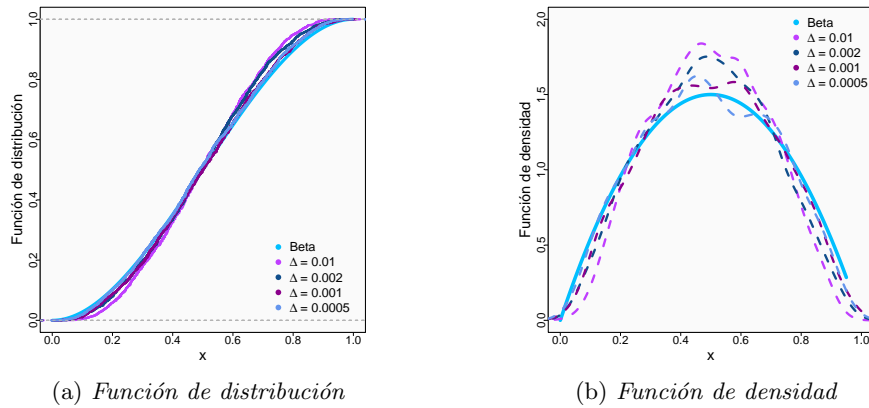


Figura 3.9: Función de distribución y densidad comparando la muestra de 2000 trayectorias al tiempo las $t = 5$ obtenidas con el método BISS de parámetros $(\alpha = 2, \beta = 2)$ contra la distribución estacionaria.

Luego, cuando $(\alpha = 3, \beta = 1)$ donde las aproximaciones de Euler y Milstein fallaron para las simulaciones en el intervalo de tiempo $[0,5]$ observadas en $t = 0.5, 1, 2.5$ y 5 el método BISS presenta mejores resultados con $\Delta = 0.0005$ los cuales podrían llegar a ser aceptables si no se es tan exigente. En la Figura 3.10 son evidentes estas diferencias. Por ejemplo, en el qqplot vemos que los valores cercanos a cero son los más alejados de la recta identidad mientras que no se presentan problema con valores cercanos a uno.

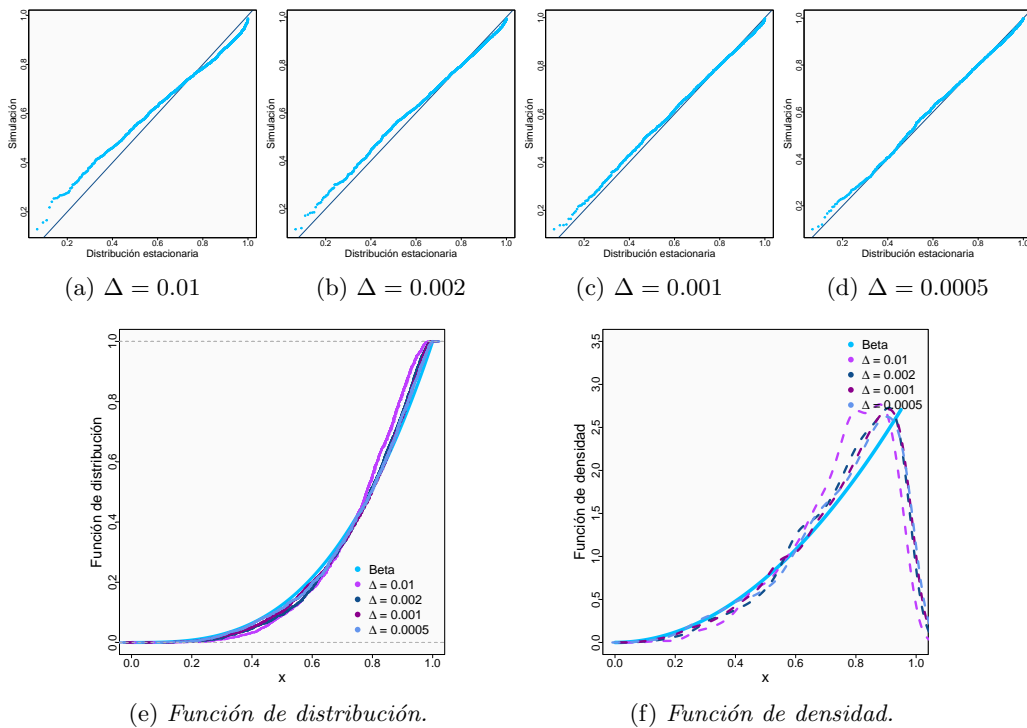


Figura 3.10: Comparaciones gráficas de la muestra de 2000 trayectorias al tiempo las $t = 5$ simuladas con el método BISS de parámetros $(\alpha = 3, \beta = 1)$ contra la distribución estacionaria.

Finalmente, vale la pena mostrar también algún gráfico para $(\alpha = 10, \beta = 30)$ en la Figura 3.11 la evolución de la función de densidad a medida que Δ decrece.

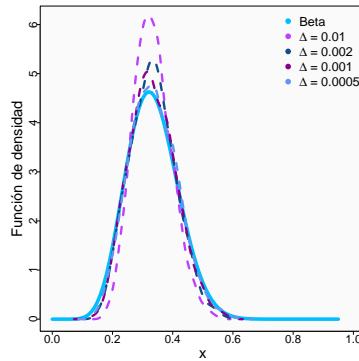


Figura 3.11: Comparación de la función de densidad de una variable aleatoria $Beta(10, 30)$ y la muestra de 2000 trayectorias al tiempo $t = 5$ obtenidas con el método BISS.

A continuación, realizamos las pruebas Kolmogorov-Smirnov con $\Delta = 0.002$, los parámetros ya estudiados pero tomando como valor inicial un número aleatorio entre cero y uno e incrementando el número de simulaciones. Los p -values obtenidos se encuentran en la Tabla 3.6, donde podemos notar que, en general, tamaños de muestra pequeños generan un mejor rendimiento.

Parámetros	Tiempo	100	500	1000	1500	2000	5000	10000
$\alpha = 1$	0.5	0.736	0.360	0.571	0.079	0.015	1.5e-04	1.2e-06
	1	0.420	0.276	0.016	0.006	6.9e-04	3.4e-06	7.0e-10
$\beta = 1$	2.5	0.253	0.031	0.040	0.057	0.016	4.0e-05	4.4e-09
	5	0.880	0.174	0.041	0.036	0.028	4.0e-05	4.4e-09
$\alpha = 2$	0.5	0.112	0.121	0.155	0.787	0.684	0.131	0.005
	1	0.987	0.226	0.040	0.030	0.021	4.5e-04	1.6e-06
$\beta = 2$	2.5	0.492	0.157	0.278	0.044	0.010	2.0e-05	5.3e-09
	5	0.093	0.009	0.028	0.020	0.003	0.002	1.9e-06
$\alpha = 3$	0.5	0.035	9.9e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	0	0
	1	0.127	1.0e-05	7.7e-10	1.0e-12	2.2e-16	< 2.2e-16	0
$\beta = 1$	2.5	0.155	0.001	0.001	4.8e-05	4.0e-08	< 2.2e-16	0
	5	0.438	0.005	6.6e-05	3.4e-06	8.3e-07	< 2.2e-16	0
$\alpha = 10$	0.5	0.466	0.324	0.095	0.037	0.037	4.3e-06	9.5e-10
	1	0.114	0.090	0.020	0.045	0.002	2.3e-05	5.5e-11
$\beta = 20$	2.5	0.387	0.027	0.004	0.005	0.002	4.7e-06	1.5e-10
	5	0.528	0.644	0.258	0.028	0.007	9.5e-06	1.6e-12

Tabla 3.6: Resultados de la pruebas de bondad de ajuste para simulaciones con el método de BISS contra la distribución estacionaria con distinto número de simulaciones.

3.2. Algoritmo exacto.

En esta sección presentamos un algoritmo propuesto por Jenkins y Spanò [15] para simular trayectorias de la difusión de Wright-Fisher dada por la ecuación (3.1). La idea principal del algoritmo consiste en utilizar la representación en series infinitas de la función de transición (3.16) en combinación con una variante del método de series alternadas. Se dice que este método es exacto en el sentido de que es posible recuperar, sin error de aproximación (salvo por el error de precisión de la máquina), muestras de las distribuciones finito dimensionales de la difusión en cuestión.

Previamente vimos la descomposición en series de la función de transición $f(x, \cdot; t)$ de la difusión de Wright-Fisher a partir de sus funciones propias. De acuerdo a la ecuación (2.20), la densidad de transición está dada por

$$f(x, y; t) = \sum_{m=0}^{\infty} q_m^\theta(t) \sum_{l=0}^m \mathcal{B}_{m,x}(l) \mathcal{D}_{\alpha+l, \beta+m-l}(y), \quad (3.16)$$

donde

$$\mathcal{B}_{m,x}(l) = \binom{m}{l} x^l (1-x)^{m-l}$$

es la función de densidad de una variable aleatoria Binomial de parámetros m y x y

$$\mathcal{D}_{\theta_1, \theta_2}(y) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\theta_1-1} (1-y)^{\theta_2-1},$$

la función de densidad de una variable aleatoria Beta de parámetros $\theta_1 = \alpha + l$ y $\theta_2 = \beta + m - l$.

Por último, como ya vimos $\{q_m^\theta(t)\}$ es la función de transición de un proceso de muerte $A_\infty^\theta(t)$ con $\theta = \alpha + \beta$ y frontera de entrada infinito. Formalmente, si $\{A_n^\theta(t)\}_{t \geq 0}$ es un proceso de muerte sobre los naturales de tal forma que $A_n^\theta(0) = n$ casi seguramente y cuyas transiciones son solo de m a $m-1$ con tasa $m(m+\theta+1)/2$ para $m = 1, 2, \dots, n$. Entonces $q_m^\theta(t) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n^\theta(t) = m)$.

Además, la ecuación (3.16) tiene una interpretación en términos del coalescente de Kingman que vimos en la Sección 2.5. El proceso ancestral $A_\infty^\theta(t)$ representa el número de ascendencia sobreviviente al tiempo t en un árbol de coalescencia infinito, cuando la ascendencia se pierde por mutación o por coalescencia.

El algoritmo para obtener trayectorias del procesos de Wright-Fisher es sencillo si se tiene un mecanismo para simular el proceso ancestral.

3.2.1. Método de series alternadas aplicado al proceso ancestral del coalescente de Kingman.

En primer lugar, vamos a describir el método de series alternadas para variables aleatorias, así como las modificaciones pertinentes para generar muestras del proceso ancestral del coalescente de Kingman.

Método de series alternadas.

El método de series alternadas es una modificación al algoritmo de la transformada inversa y el método de series dado por Devroye en 1986/1980 para simular variables aleatorias, puede pensarse como un método de aceptación/rechazo con una infinidad de aceptaciones y rechazos, comúnmente se utiliza cuando la función de densidad se escribe en términos de una serie. Para mayor detalle de los Algoritmos 2 y 3 puede consultarse [6].

Para utilizar el algoritmo de series, consideremos X una variable aleatoria con densidad f que puede ser aproximada por sucesiones de funciones fáciles de evaluar, por abajo con f_n y por arriba con g_n , donde f_n no necesariamente debe ser positiva y g_n puede no ser integrable, de tal manera que se cumple:

- i. $\lim_{n \rightarrow \infty} f_n = f$.
- ii. $\lim_{n \rightarrow \infty} g_n = f$.
- iii. $f_n \leq f \leq g_n$.
- iv. $f \leq ch$ para alguna constante $c \geq 1$ y h una función de densidad fácil de simular.

Bajo estos supuestos, el algoritmo de series queda como se describe a continuación.

Algoritmo 2 Método de series para simular variables aleatorias.

- 1: Generar un valor X con densidad h .
 - 2: Generar una variable aleatoria uniforme, $U \sim \text{Unif}(0, 1)$.
 - 3: Hacer $W = Uch(X)$, $n = 0$.
 - 4: Incrementar $n = n + 1$.
 - 5: Si $W \leq f_n(X)$. Detener y devolver X .
 - 6: Mientras $W \leq g_n(X)$ regresar al paso (4). Si $W > g_n(X)$, regresar a (1).
-

Notemos que tenemos dos ciclos anidados, el ciclo exterior es un ciclo infinito, lo cual no es de importancia pues con probabilidad 1 el algoritmo para en el ciclo interior pues $f_n \rightarrow f$ y $g_n \rightarrow f$.

Ahora, supongamos que X es una variable aleatoria discreta que toma valores en los enteros no negativos. Recordando el algoritmo de la función inversa para este tipo de variables, si $U \sim \text{Unif}(0, 1)$, el valor dado por $\inf\{M \in \mathbb{N} : \sum_{m=0}^M f(m) > U\}$ tiene distribución con función de densidad f .

Además supongamos que su función de densidad se escribe como un serie de la forma

$$f(x) = ch(x) (1 - a_1(x) + a_2(x) - a_3(x) + \dots)$$

donde $\{a_n\}_{n \in \mathbb{N}}$ es una sucesión de funciones que decrece a cero cuando n crece a infinito, c una constante y h una función de densidad fácil de simular.

Entonces, el algoritmo de series alternadas se escribe de la siguiente forma.

Algoritmo 3 Método de series alternadas para simular variables aleatorias.

- 1: Generar un valor X con densidad h .
 - 2: Generar una variable aleatoria uniforme, $U \sim \text{Unif}(0, c)$.
 - 3: Hacer $n = 0$ y $W = 0$.
 - 4: Incrementar $n = n + 1$.
 - 5: Calcular $W = W + a_n(X)$.
 - 6: Si $U \geq W$. Detener y devolver X .
 - 7: Incrementar $n = n + 1$.
 - 8: Calcular $W = W - a_n(X)$.
 - 9: Mientras $W < U$ regresar a (4). Si $U < W$, regresar a (1).
-

Este algoritmo es válido pues la función de densidad f está acotada por arriba y por abajo por dos series convergentes,

$$1 + \sum_{j=1}^k (-1)^j a_j(x) \leq \frac{f(x)}{ch(x)} \leq 1 + \sum_{j=1}^{k+1} (-1)^j a_j(x), \quad k \text{ impar.}$$

La validez de las desigualdades es gracias a la monotonía tomada por pares de la sucesión. Observamos que igual que en el primer método de series la densidad f nunca se calcula completamente, adicionalmente, tampoco se evalúa la densidad h .

Simulación del proceso ancestral del coalescente de Kingman.

Para poder utilizar el método de series alteranadas para simular trayectorias de la difusión de Wright-Fisher Jenkins y Spanò en [15] consideran algunas variaciones al método anterior. Resumiendo, supongamos que tenemos una función de densidad de la forma

$$f(x) = \sum_{i=0}^{\infty} (-1)^i a_i(x)$$

tal que $a_n \downarrow 0$ si $n \rightarrow \infty$.

Entonces, para toda $M, K \in \mathbb{N}$, se tiene que

$$T_k^-(M) = \sum_{m=0}^M \sum_{k=0}^{2K+1} (-1)^k a_k(m) \leq \sum_{m=0}^M f(m) \leq \sum_{m=0}^M \sum_{k=0}^{2K} (-1)^k a_k(m) = T_k^+(M).$$

Y como $\sum_{m=0}^M f(m) = \mathbb{P}(X \leq M)$ acabamos de encontrar dos funciones que acotan a la función de distribución de tal manera que

$$T_k^-(M) \uparrow \mathbb{P}(X \leq M) \quad \text{y} \quad T_k^+(M) \downarrow \mathbb{P}(X \leq M) \quad \text{cuando} \quad K \rightarrow \infty.$$

Finalmente, definimos para $U \sim \text{Unif}(0, 1)$,

$$K_0(M) = \inf\{K \in \mathbb{N} : T_k^-(M) > U \quad \text{o} \quad T_k^+(M) < U\}$$

y posteriormente, siguiendo un procedimiento análogo al algoritmo de la inversa,

$$\inf\{M \in \mathbb{N} : T_{K_0(M)}^-(M) > U\}$$

puede ser calculado con una cantidad finita de términos y tiene distribución con densidad f .

Consideremos la función de transición $f(x, \cdot; t)$ del proceso de Wright-Fisher dada por la ecuación (3.16). Jenkins y Spanò [15] probaron que bajo ciertas modificaciones puede implementarse el método anterior para simular valores del proceso ancestral del coalescente de Kingman $A_\infty^\theta(t)$, es decir, simular valores de la función de densidad $\{q_m^\theta(t)\}_{m=0}^\infty$, donde $\theta = \alpha + \beta$.

Como $q_m^\theta(t)$ es conocida y tiene una forma explícita, el método de inversión permite simular valores de dicha distribución, ya que para $U \sim \text{Unif}(0, 1)$ $\inf\{M \in \mathbb{N} : \sum_{m=1}^M q_m^\theta(t) > U\}$ tiene distribución $q_m^\theta(t)$. Sin embargo, $q_m^\theta(t)$ sólo es conocida a partir de series, ([32])

$$q_m^\theta(t) = \sum_{k=m}^{\infty} (-1)^{k-m} a_{km}^\theta e^{-k(k+\theta-1)t/2} \quad (3.17)$$

y $a_{km}^\theta = \frac{(\theta+2k-1)(\theta+m)(k-1)}{m!(k-m)!}$ donde $a_{(x)} := \frac{\Gamma(a+x)}{\Gamma(a)} = a(a+1) \cdots (a+x-1)$, para $a > 0$ y $x \geq -1$, el factorial ascendente o símbolo de Pochhammer.

Definimos $b_k^{(t,\theta)}(m) = a_{km}^\theta e^{-k(k+\theta-1)t/2}$, la siguiente proposición muestra que a partir de cierto valor los coeficientes $b_k^{(t,\theta)}(m)$ de la función de densidad $\{q_m^\theta(t)\}_{m=0}^\infty$ cumplen con las condiciones para el algoritmo de series alternadas. La demostración puede encontrarse en [15].

Proposición 3.2.1. *Sea $b_k^{(t,\theta)}(m) = a_{km}^\theta e^{-k(k+\theta-1)t/2}$ y sea $C_m^{(t,\theta)} = \inf\{i \geq 0 : b_{i+m+1}^{(t,\theta)}(m) < b_{i+m}^{(t,\theta)}(m)\}$. Entonces,*

i) $C_m^{(t,\theta)} < \infty$ para toda m .

ii. $b_k^{(t,\theta)}(m) \downarrow 0$, $k \rightarrow \infty$ si $k \geq m + C_m^{(t,\theta)}$.

iii. $C_m^{(t,\theta)} = 0$ para toda $m > D_0^{(t,\theta)}$, donde para $\epsilon \in [0, 1)$ se define

$$D_\epsilon^{(t,\theta)} := \inf \left\{ k \geq \max \left\{ \left(\frac{1}{t} - \frac{\theta+1}{2} \right), 0 \right\} : (\theta+2k+1)e^{\frac{2k+\theta}{2}t} < 1 - \epsilon \right\}. \quad (3.18)$$

El ajuste que se hace al método de series alternadas consiste en calcular los términos $q_m^\theta(t)$ mientras que $m \leq D_0^{(t,\theta)}$ hasta encontrar la primera vez que los coeficientes en (3.17) empiecen a decrecer, éste hecho ocurre gracias a la Proposición 3.2.1. Adicionalmente, para que los coeficientes dependan de m , introducimos el vector $\mathbf{k} = (k_1, \dots, k_m)$ y definimos,

$$S_{\mathbf{k}}^-(M) = \sum_{m=0}^M \sum_{i=0}^{2k_m+1} (-1)^i b_{m+i}^{(t,\theta)}(m) \quad \text{y} \quad S_{\mathbf{k}}^+(M) = \sum_{m=0}^M \sum_{i=0}^{2k_m} (-1)^i b_{m+i}^{(t,\theta)}(m).$$

De modo que la Proposición 3.2.1 asegura que el algoritmo pare en un tiempo finito. El algoritmo se presenta a continuación, donde $\lceil a \rceil$ representa la función techo de a .

Algoritmo 4 Método de series alternadas para simular el coalescente de Kingman $A_\infty(t)$.

- 1: Hacer $m = 0$, $k_0 = 0$ y $\mathbf{k} = (k_0)$.
 - 2: Generar $U \sim \text{Unif}(0, 1)$.
 - 3: Sea $k_m = \lceil C_m^{(t,\theta)}/2 \rceil$.
 - 4: Mientras $S_{\mathbf{k}}^-(M) < U < S_{\mathbf{k}}^+(M)$, hacer $\mathbf{k} = \mathbf{k} + (1, 1, \dots, 1)$.
 - 5: Si $S_{\mathbf{k}}^-(M) > U$. Detener y devolver el valor m .
 - 6: Si $S_{\mathbf{k}}^+(M) < U$, $\mathbf{k} = (k_0, \dots, k_m, 0)$ e incrementar $m = m + 1$.
 - 7: Regresar a (3).
-

3.2.2. Algoritmo exacto.

Ya que hemos simulado un valor del coalescente de Kingman, es sencillo generar trayectorias de la difusión de Wright-Fisher a partir de la función de transición mediante el siguiente algoritmo.

Algoritmo 5 Método exacto para simular el proceso de Wright-Fisher.

- 1: Simular $A_\infty^\theta(t)$. (Mediante el Algoritmo 4).
 - 2: Dado $A_\infty^\theta(t) = m$, generar $L \sim \text{mathsf{Bin}}(m, x)$.
 - 3: Dado $L = l$, generar $Y \sim \text{Beta}(\alpha + l, \beta + m - l)$
-

Respecto a la eficiencia del algoritmo propuesto por Jenkins y Spanò en [15], es importante observar el número de coeficientes $b_k^{(t,\theta)}(m)$ que se calculan en el Algoritmo 4, el cual depende, de acuerdo a la Proposición 3.2.1, de los valores de $D_0^{(t,\theta)}$ y $C_m^{(t,\theta)}$ que normalmente son pequeños, así que el número de coeficientes no es muy grande. La excepción ocurre cuando t es cercana a cero y el número de coeficientes crece rápidamente. La siguiente proposición, que puede encontrarse en [15], confirma tal afirmación.

Proposición 3.2.2. *Siguiendo con la notación de la Proposición 3.2.1. Cuando t tiende a cero,*

- i) $C_m^{(t,\theta)} = \mathcal{O}(t^{-1})$.
- ii) $\max_m C_m^{(t,\theta)} = \mathcal{O}(t^{-1} \ln(t^{-1}))$.
- iii) $D_0^{(t,\theta)} = \mathcal{O}(t^{-(1+\kappa)})$, para cualquier $\kappa > 0$.
- iv) Sea $N^{(t,\theta)}$ el número de coeficientes que se calculan al implementar el Algoritmo 4.
Entonces $\mathbb{E}[N^{(t,\theta)}] < \infty$ y en particular, $\mathbb{E}[N^{(t,\theta)}] = \mathcal{O}(t^{-(1+\kappa)})$, para cualquier $\kappa > 0$.

Así que la deficiencia en el algoritmo para t cercanos a cero está dada por la inestabilidad numérica de la ecuación (3.17) cuando t tiende a cero. Es preciso señalar que éste hecho afecta a cualquier método que utilice expansiones con base en polinomios ortogonales. Por tanto, en la implementación del algoritmo se deben utilizar aproximaciones cuando la separación entre dos puntos es muy pequeña. Jenkins y Spanò ([15]) encontraron que el Algoritmo 5 falla para $t < 0.05$ aproximadamente. Una posible solución para la implementación es utilizar el algoritmo de Euler o alguna aproximación normal. La aproximación normal que ellos proponen es la siguiente.

Teorema 3.2.1. Supogamos que $\kappa = \frac{1}{2}(\theta - 1)t$ y sea $\mu^{(t,\theta)} = \frac{2\eta}{t}$ y

$$\left(\sigma^{(t,\theta)}\right)^2 = \begin{cases} \frac{2\eta}{t}(\eta + \kappa)^2 \left(1 - \frac{\eta}{\eta + \kappa} - 2\eta\right) \kappa^{-2} & \text{si } \kappa \neq 0 \\ \frac{2}{3t} & \text{si } \kappa = 0, \end{cases}$$

donde

$$\eta = \begin{cases} \frac{\kappa}{e^{\kappa} - 1} & \text{si } \kappa \neq 0 \\ 1 & \text{si } \kappa = 0. \end{cases}$$

Entonces, $\mathbb{P}\left(\frac{A_{\infty}^{\theta}(t) - \mu^{(t,\theta)}}{\sigma^{(t,\theta)}} \leq x\right) \rightarrow \Phi(x)$ cuando $t \rightarrow 0$, con $\Phi(\cdot)$ la función de distribución de una variable aleatoria normal estándar.

Generamos trayectorias con parámetros, $(\alpha = 0.7, \beta = 0.7)$, $(\alpha = 1, \beta = 1)$, $(\alpha = 2, \beta = 2)$, $(\alpha = 3, \beta = 1)$ y $(\alpha = 10, \beta = 20)$ en el intervalo de tiempo $[0,5]$ y con valor inicial $X_0 = 0.5$. De acuerdo con los hallazgos de Jenkins y Spanò, si $t < 0.05$ utilizamos la aproximación normal para simular el coalescente de Kingman y en caso contrario el Algoritmo 5.

Cabe mencionar que para generar trayectorias nos apoyamos en la propiedad de Markov del proceso, de manera que para simular una observación al tiempo t_i para $i = 1, 2, \dots, n$ utilizamos la observación al tiempo anterior, t_{i-1} , como valor inicial. En la Figura 3.12 presentamos una trayectoria simulada con cada uno de los parámetros con 100 puntos por unidad de tiempo.

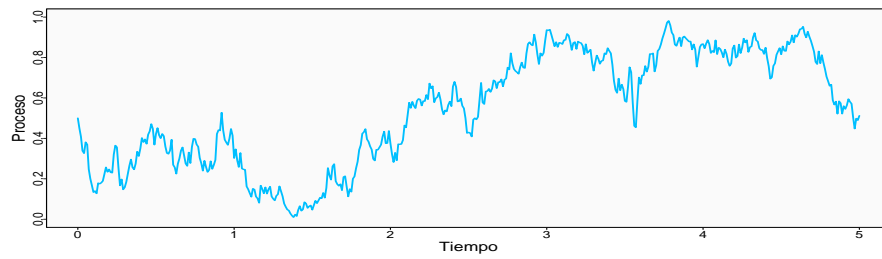
Luego, utilizamos la prueba de Kolmogorov-Smirnov para comparar con la distribución estacionaria con aquellos parámetros donde es posible. En este caso podemos hacer la comparación en dos sentidos. En primer lugar, ya que el algoritmo permite simular directamente de la función de densidad de transición, $f(x, y; t)$, generamos con cada conjunto de parámetros 2000 observaciones con valor inicial $x = 0.5$ en los tiempos $t = 0.5, 1, 2.5$ y 5 . Los p -values se encuentran en la Tabla 3.7, donde observamos muy buenos resultados pues para todos los conjuntos de parámetros no se rechaza la hipótesis nula de igualdad en distribución en $t = 2.5$ e incluso en tiempos menores con algunos parámetros. Lo cual da indicios de un buen comportamiento del algoritmo para simular en tiempos grandes.

Parámetros	Tiempo	p-value
$\alpha = 1$	0.5	1.0e-08
	1	0.092
	$\beta = 1$	2.5
	5	0.652
$\alpha = 2$	0.5	0.201
	1	0.672
	$\beta = 2$	2.5
	5	0.214
$\alpha = 3$	0.5	< 2.2e-16
	1	2.0e-08
	$\beta = 1$	2.5
	5	0.785
$\alpha = 10$	0.5	0.781
	1	0.519
	$\beta = 20$	2.5
	5	0.312

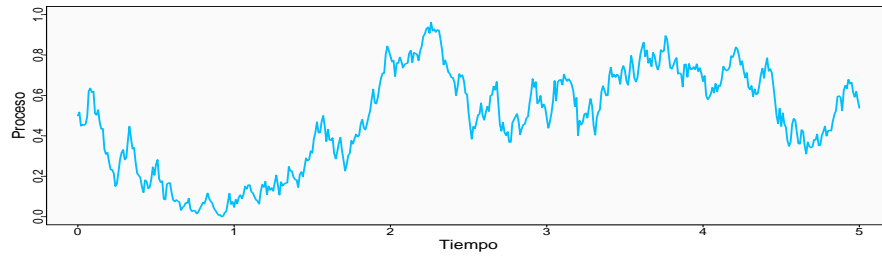
Tabla 3.7: Resultados de la pruebas de bondad de ajuste para simulaciones de observaciones con el método exacto contra la distribución estacionaria.

La otra comparación es generar trayectorias con los parámetros que estamos trabajando en el intervalo de tiempo $[0,5]$ y evaluar su comportamiento en los tiempos $t = 0.5, 1, 2.5$ y 5 con los distintos valores de Δ .

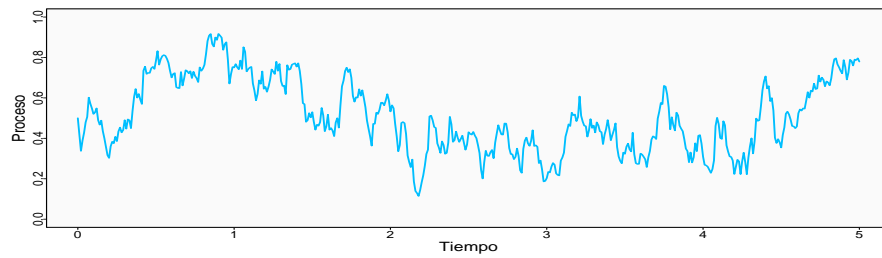
Los resultados de la prueba de hipótesis se muestran en la Tabla 3.8, donde notamos que existe congruencia entre ambas formas.



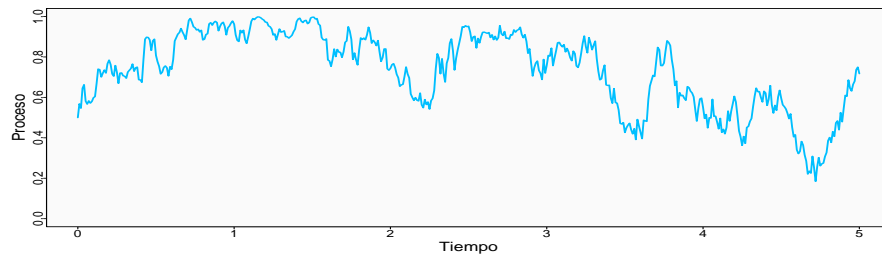
(a) $\alpha = 0.7$ y $\beta = 0.7$



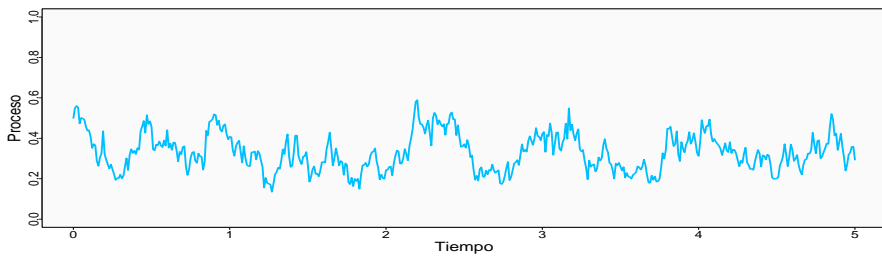
(b) $\alpha = 1$ y $\beta = 1$



(c) $\alpha = 2$ y $\beta = 2$



(d) $\alpha = 3$ y $\beta = 1$

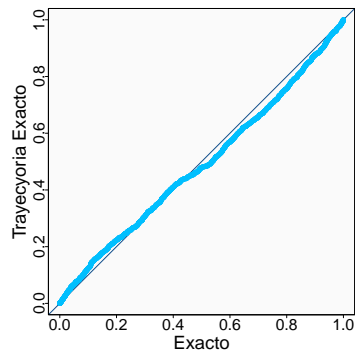


(e) $\alpha = 10$ y $\beta = 20$

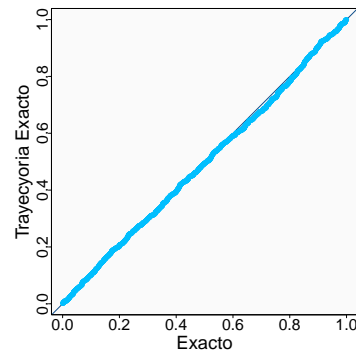
Figura 3.12: Trayectorias simuladas con el método exacto en el intervalo de tiempo $[0,5]$ con 100 puntos simulados por unidad de tiempo y $X_0 = 0.5$.

Parámetros	Tiempo	$\Delta = 0.01$	$\Delta = 0.002$	$\Delta = 0.001$	$\Delta = 0.0005$
$\alpha = 1$	0.5	9.3-07	5.6e-06	4.7e-10	7.90e-07
	1	0.42	0.7	0.76	0.75
	$\beta = 1$	2.5	0.2	0.86	0.51
	5	0.13	0.16	0.72	0.63
$\alpha = 2$	0.5	0.28	0.33	0.06	0.17
	1	0.68	0.81	0.82	0.59
	$\beta = 2$	2.5	0.91	0.58	0.97
	5	0.55	0.99	0.95	0.36
$\alpha = 3$	0.5	0	0	0	0
	1	3.5e-10	8e-11	1.4e-09	5.3 e-05
	$\beta = 1$	2.5	0.52	0.77	0.24
	5	0.42	0.16	0.82	0.31
$\alpha = 10$	0.5	0.84	0.57	0.90	0.46
	1	0.01	0.90	0.01	0.33
	$\beta = 20$	2.5	0.66	0.68	0.81
	5	0.22	0.22	0.87	0.88

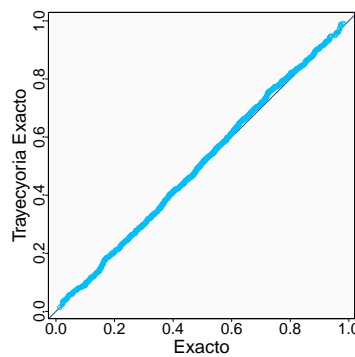
Tabla 3.8: Resultados de la pruebas de bondad de ajuste para las simulaciones de trayectorias con el método exacto contra la distribución estacionaria.



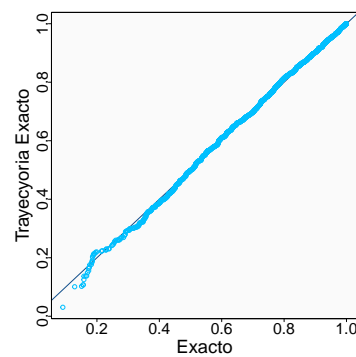
(a) ($\alpha = 0.7, \beta = 0.7$)



(b) ($\alpha = 1, \beta = 1$)



(c) ($\alpha = 2, \beta = 2$)



(d) ($\alpha = 3, \beta = 1$)

Figura 3.13: QQplot comparando la simulación de 2000 observaciones simuladas de 0 a $t = 1$ contra una muestra de 2000 trayectorias al tiempo $t = 1$ con $\Delta = 0.01$, ambas generadas con algoritmo exacto.

Veamos una comparación gráfica del algoritmo exacto, por un lado, obtenemos una muestra de tamaño 2000

generando observaciones del tiempo 0 al tiempo $t = 1$ y por otro lado, simulamos 2000 trayectorias en el intervalo de tiempo $[0,5]$ con $\Delta = 0.01$ y tomamos la muestra al tiempo uno. Mostramos en la Figura 3.13 un qqplot para algunos de los parámetros de estudio. Podemos observar que el comportamiento de ambas muestras es ligeramente diferente cuando $(\alpha = 0.7, \beta = 0.7)$ pues para valores menores a 0.5 los puntos caen por arriba de la diagonal mientras que para valores mayores quedan por debajo.

Resulta interesante notar que a pesar de que al considerar $(\alpha = 3, \beta = 1)$ en $t = 1$ las muestras no siguen la distribución estacionaria, sin embargo, gráficamente el comportamiento es muy similar, salvo diferencias en la parte izquierda de la distribución. Para el resto de los parámetros, como ya esperábamos, el comportamiento es similar entre ellos.

Ya que en general nos va a interesar obtener trayectorias del proceso y no solo observaciones en tiempo fijos, de aquí en adelante vamos a considerar, para cualquier análisis, dichas simulaciones.

A continuación, mostramos algunas comparaciones gráficas de las muestras obtenidas contra la distribución estacionaria. En particular, veamos aquellos parámetros con los cuales no obtuvimos resultados tan favorables en los métodos numéricos. Por ejemplo, con $(\alpha = 1, \beta = 1)$ y longitud del intervalo entre puntos de $\Delta = 0.01$ en los diferentes tiempos de estudio, Figura 3.14, verificamos los resultados obtenidos con la prueba de hipótesis.

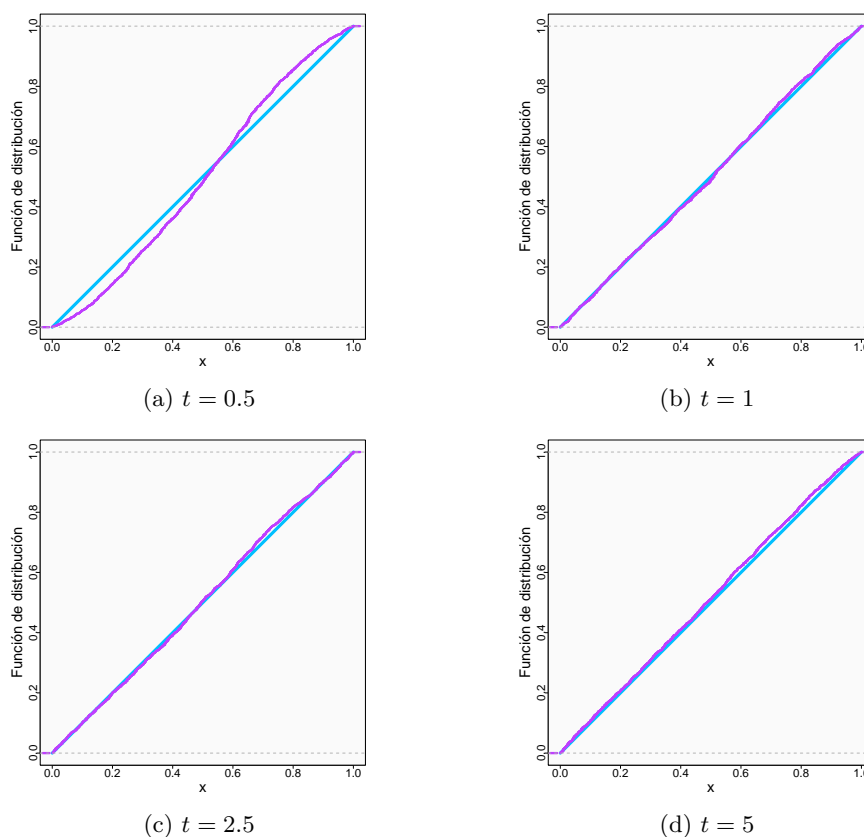


Figura 3.14: Función de distribución de la muestra de 2000 trayectorias de parámetros $(\alpha = 1, \beta = 1)$ en distintos tiempos obtenidas con el método de exacto contra la función de distribución de una variable aleatoria $Beta(1, 1)$.

También mostramos las comparaciones cuando $(\alpha = 3, \beta = 1)$ y $\Delta = 0.01$ en los distintos tiempos de estudio, Figura 3.15, como se observa gráficamente y de acuerdo a los resultados de las pruebas de hipótesis no se rechaza la hipótesis nula pareciera que el proceso nuevamente tarda en alcanzar la distribución estacionaria

en los tiempos $t = 0.5$ y $t = 1$ pero a diferencia de los métodos numéricos sí obtenemos una muestra con distribución Beta cuando el tiempo se incrementa.

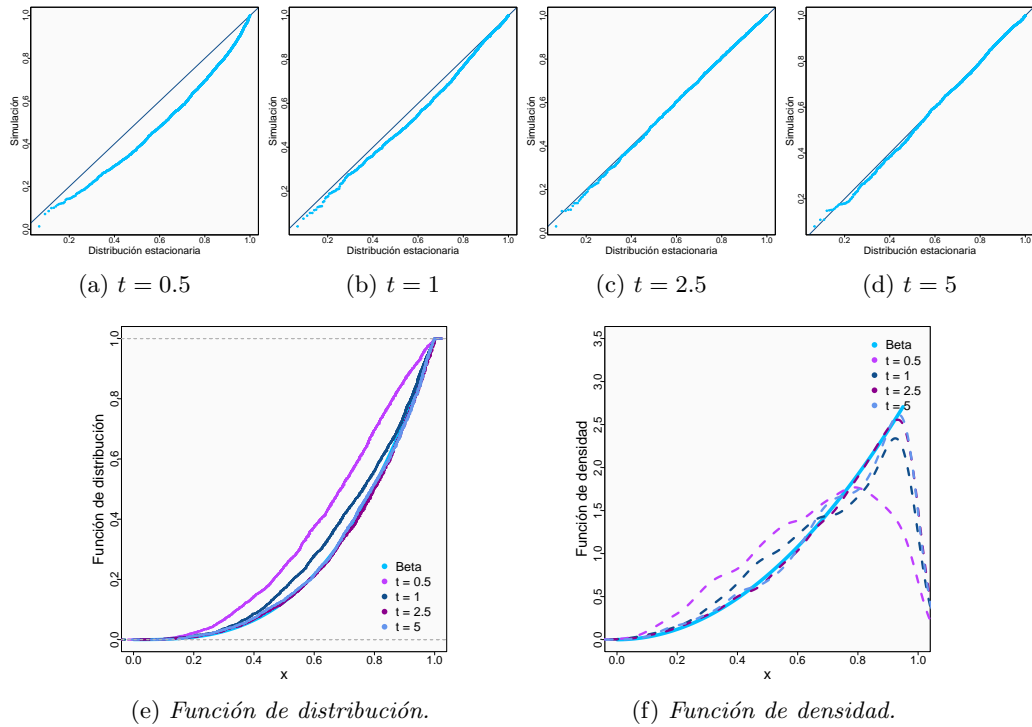


Figura 3.15: Comparaciones gráficas de la muestra de 2000 trayectorias de parámetros ($\alpha = 3$, $\beta = 1$) en distintos tiempos simuladas con el método de exacto contra la distribución estacionaria.

Parámetros	Tiempo	100	500	1000	1500	2000	5000	10000	
$\alpha = 1$	0.05	0.507	0.976	0.827	0.825	0.730	0.441	0.549	
	1	0.413	0.308	0.696	0.736	0.876	0.344	0.196	
	$\beta = 1$	2.5	0.317	0.279	0.892	0.837	0.655	0.979	0.556
		5	0.945	0.616	0.623	0.674	0.877	0.827	0.656
$\alpha = 2$	0.05	0.249	0.942	0.607	0.571	0.366	0.508	0.543	
	1	0.7	0.885	0.732	0.915	0.902	0.991	0.652	
	$\beta = 2$	2.5	0.552	0.927	0.865	0.955	0.99	0.74	0.604
		5	0.431	0.428	0.191	0.280	0.192	0.1	0.236
$\alpha = 3$	0.5	1.3e-07	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	
	1	< 0.001	3.9e-04	4.2e-06	< 1.1e-07	3.0e-16	< 2.2e-16	< 2.2e-16	
	$\beta = 1$	2.5	0.391	0.231	0.577	0.931	0.51	0.299	0.425
		5	0.968	0.793	0.778	0.29	0.679	0.454	0.44
$\alpha = 10$	0.5	0.275	0.262	0.486	0.505	0.642	0.105	0.131	
	1	0.451	0.33	0.396	0.799	0.941	0.859	0.408	
	$\beta = 20$	2.5	0.996	0.431	0.262	0.167	0.084	0.474	0.776
		5	0.458	0.467	0.857	0.625	0.187	0.196	0.031

Tabla 3.9: Resultados de la pruebas de bondad de ajuste para simulaciones con el método de exacto contra la distribución estacionaria con distinto número de simulaciones.

Finalmente, veamos el comportamiento del método exacto respecto al número de simulaciones con 100, 500,

1000, 1500, 2000, 5000 y 10000. Además, para hacerlo un poco más general consideramos el estado del proceso al tiempo cero como un número aleatorio en el espacio de estados. En la Tabla 3.9 registramos los *p-values* que se obtienen al contrastar con la distribución estacionaria utilizando la prueba de Kolmogorov-Smirnov. Los resultados son consistentes con los obtenidos anteriormente y además en todos los caso se tienen las mismas conclusiones.

3.3. Construcción vía procesos de Markov

El siguiente método simula trayectorias de procesos de Markov estacionarios en el tiempo, fue propuesto por Mena y Walker [20], quienes tomaron el estudio de Pitt, Chatfield y Walker (2002) que consiste en la construcción de series de tiempo estacionarias con distribuciones marginales dadas. La idea es construir un proceso markoviano con distribución estacionaria π_X conocida. Para ello, hay que suponer que π_X pertenece a una familia paramétrica y encontrar las probabilidades de transición que conserven dicha distribución estacionaria incorporando una variable latente Y cuya relación de dependencia está dada por la densidad condicional $f_{Y|X}(y|x)$.

Escribimos la función de transición de x_{n-1} a x_n como:

$$p(x_{n-1}, x_n) = \int f_{X|Y}(x_n|y)f_{Y|X}(y|x_{n-1})\eta_1(dy) \quad (3.19)$$

donde, η_1 es una medida de referencia, en la práctica puede ser la medida de Lebesgue o la medida de conteo. Desde el punto de vista bayesiano $f_{X|Y}$ representa la distribución posterior de la verosimilitud $f_{Y|X}(y|x)$ considerando una única observación x y π_X su distribución inicial, es decir,

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\pi_X(x)}{\int f_{Y|X}(y|x)\pi_X(x)} = \frac{f_{Y|X}(y|x)\pi_X(x)}{\pi_Y(y)} \propto f_{Y|X}(y|x)\pi_X(x). \quad (3.20)$$

Bajo estos supuestos, π_X constituye una medida invariante para la densidad de transición (3.19) ya que,

$$\begin{aligned} \int p(x_{n-1}, x_n)\pi_X(x_{n-1})\eta_2(dx_{n-1}) &= \int \int f_{X|Y}(x_n|y)f_{Y|X}(y|x_{n-1})\eta_1(dy)\eta_2(dx_{n-1}) \\ &= \int \int \frac{f_{Y|X}(y|x_n)\pi_X(x_n)}{\pi_Y(y)}f_{Y|X}(y|x_{n-1})\eta_1(dy)\eta_2(dx_{n-1}) \\ &= \int \frac{f_{Y|X}(y|x_n)\pi_X(x_n)}{\pi_Y(y)} \int f_{Y|X}(y|x_{n-1})\eta_2(dx_{n-1})\eta_1(dy) \\ &= \int f_{Y|X}(y|x_n)\pi_X(x_n)\eta_1(dy) \\ &= \pi_X(x_n), \end{aligned} \quad (3.21)$$

nuevamente η_2 es una medida de referencia, usualmente la medida de Lebesgue o la medida de conteo.

Puesto que las funciones de transición de un proceso de Markov se caracterizan, entre otras cosas, por cumplir las ecuaciones de Chapman-Kolmogorov, debe elegirse la relación de dependencia $f_{Y|X}(y|x)$ de tal manera que se cumpla esta condición. Esto es,

$$p(x_0, x_{t+s}) = \int p(x_s, x_{t+s})p(x_0, x_s)\eta_2(dx_s). \quad (3.22)$$

No parece haber una forma general para encontrar $f_{Y|X}$ de tal manera que se satisfaga la ecuación (3.22), pero se encontraron algunos resultados interesantes cuando dicha función pertenece a ciertas familias paramétricas. El enfoque de Mena y Walker [20] consiste en hacer depender del tiempo uno de los parámetros subyacentes y dar condiciones para que se satisfagan las ecuaciones de Chapman-Kolmogorov.

Los procesos de Markov que se construyen de esta manera heredan algunas características de una cadena de Markov generada por el algoritmo de muestreo de Gibbs. Además, esta construcción permite obtener

procesos de Markov reversibles en el tiempo. Asimismo, esta representación de la transición permite tratar la ley del proceso lo cual puede ser útil para diversos propósitos.

Regresando a la construcción de la difusión de Wright-Fisher, sabemos que la distribución estacionaria del proceso es una distribución Beta. Mena y Walker proponen en [20] hacer la construcción de un modelo Gamma-Poisson, el cual desarrollamos a continuación, y posteriormente realizar las transformaciones pertinentes para obtener el proceso deseado.

3.3.1. Modelo Gamma-Poisson

Como primer ejemplo, podemos obtener el Proceso de Cox-Ingersoll-Ross (CIR) mediante un modelo sencillo, sea $X_0 \sim \text{Gamma}(a, b)$ con función de densidad f_X , la propuesta para la función de dependencia $f_{Y|X}$ es una distribución Poisson de parámetro ϕX_0 con $\phi > 0$, es decir, $Y_1|X_0 \sim \text{Pois}(\phi X_0)$ y como son familias conjugadas se sigue que $X_1|Y_1 \sim \text{Gamma}(a + Y_1, b + \phi)$, claramente la densidad marginal de X_1 sigue siendo $\text{Gamma}(a, b)$. Siguiendo el mismo razonamiento $Y_2|X_1 \sim \text{Pois}(\phi X_1)$ y $X_2|Y_2 \sim \text{Gamma}(a + Y_2, b + \phi)$ y así sucesivamente.

La distribución conjunta de X y Y es $f(x, y) = f_{Y|X}(y; \phi x) f_X(x; a, b)$ donde, podemos observar que el parámetro ϕ controla la correlación del proceso a tiempo discreto $\{X_n\}$. Por lo tanto, la función de transición está dada por:

$$\begin{aligned} p(x_{n-1}, x_n) &= \sum_{y=0}^{\infty} f_{X_n|Y_n}(x_n; a + y, \phi + b) f_{Y|X_{n-1}}(y; \phi x_{n-1}) \\ &= \sum_{y=0}^{\infty} \frac{(\phi + b)^{a+y}}{\Gamma(a + y)} e^{-(\phi+b)x_n} x_n^{a+y-1} \cdot \frac{e^{-\phi x_{n-1}} (\phi x_{n-1})^y}{y!} \\ &= \frac{e^{-(\phi(x_n+x_{n-1})+bx_n)}}{(\phi + b)^{-(a+1)/2} \phi^{(a-1)/2}} \left(\frac{x_n}{x_{n-1}} \right)^{\frac{a-1}{2}} I_{a-1}(2\sqrt{x_n x_{n-1} \phi(\phi + b)}), \end{aligned} \quad (3.23)$$

donde $I_\nu(\cdot)$ es la función de Bessel modificada de primer tipo.

Para que el proceso resultante sea continuo, es necesario hacer que el parámetro de dependencia ϕ sea una función determinista del tiempo ϕ_t y si además, es posible encontrar una función que satisfaga las ecuaciones de Chapman-Kolmogorov tendremos entonces un proceso de Markov homogéneo a tiempo continuo. La demostración de la Proposición 3.3.1

Proposición 3.3.1. *El proceso estocástico definido por la función de transición (3.23) con distribución estacionaria Gamma satisface las ecuaciones de Chapman-Kolmogorov si*

$$\phi_t := \frac{b}{e^{ct} - 1}, \quad c > 0.$$

Finalmente, se prueba utilizando la desigualdad de Chebyshev que

$$\lim_{t \downarrow 0} \frac{1}{t} \int_{|x_t - x_0| > \epsilon} p_t(x_0, x_t) dx_t = 0,$$

el coeficiente de deriva

$$\mu(x) := \lim_{t \downarrow 0} \frac{1}{t} \mathbb{E}(X_t - X_0 | X_0) = c \left(\frac{a}{b} - x \right)$$

y el coeficiente de difusión,

$$\sigma^2(x) := \lim_{t \downarrow 0} \frac{1}{t} \mathbb{E}(|X_t - X_0|^2 | X_0) = \sqrt{\frac{2c}{b}} x,$$

además

$$\mathbb{E}(|X_t - X_0|^4 | X_0) = 0.$$

Por lo tanto, el proceso de Markov que acabamos de definir es una difusión, lo cual nos lleva a la ecuación diferencial estocástica

$$dX_t = c\left(\frac{a}{b} - X_t\right)dt + \sqrt{\frac{2c}{b}}X_t dB_t,$$

donde B_t es un movimiento browniano estándar. La demostración de la Proposición 3.3.1 y más detalles del modelo pueden encontrarse en [20].

3.3.2. Modelo Beta-Binomial

La construcción de la difusión de Wright-Fisher dada en [20] está dada por el modelo Beta-Binomial el cual se basa en el modelo Gamma-Poisson. Sean $X_1 \sim \text{Gamma}(\alpha, 1)$ y $X_2 \sim \text{Gamma}(\beta, 1)$ independientes entonces $P = \frac{X_1}{X_1 + X_2} \sim \text{Beta}(\alpha, \beta)$ que además es independiente a $X = X_1 + X_2 \sim \text{Gamma}(\alpha + \beta, 1)$.

También sabemos que si $Y_i \sim \text{Pois}(\phi X_i)$ con $i = 1, 2$ independientes, luego $Y_1 | Y_1 + Y_2 \sim \text{Bin}\left(Y_1 + Y_2, \frac{X_1}{X_1 + X_2}\right)$.

La idea consiste en construir dos procesos independientes Gamma-Poisson, es decir,

$$X_{1t} | Y_{1t} \sim \text{Gamma}(\alpha + Y_{1t}, 1 + \phi_t), \quad X_{2t} | Y_{2t} \sim \text{Gamma}(\beta + Y_{2t}, 1 + \phi_t)$$

y

$$Y_{it} | X_{it} \sim \text{Pois}(\phi X_{it}) \text{ para } i = 1, 2,$$

luego definimos al proceso $\{P_t\}$ como

$$P_t = \frac{X_{1t}}{X_{1t} + X_{2t}}.$$

Por consiguiente, $P_t | Y_{1t}, Y_{2t} \sim \text{Beta}(\alpha + Y_{1t}, \theta - \alpha + Y_{2t} - Y_{1t})$ donde $Y_t = Y_{1t} + Y_{2t}$ y $\theta = \alpha + \beta$. Como ya habíamos visto en la Proposición 3.3.1 $\phi_t = (e^{ct} - 1)^{-1}$.

Condicionando la función de densidad de transición, al evento $Y_t = m$,

$$\begin{aligned} p(p_t | p_0, Y_t = m) &= \sum_{k=0}^m \frac{\Gamma(\theta + m)}{\Gamma(\alpha + k)\Gamma(\beta + m - k)} p_t^{\alpha+k-1} (1-p_t)^{\beta+m-k-1} \binom{m}{k} p_0^k (1-p_0)^{m-k} \\ &= \Gamma(\theta + m) \left((1-p_t)(1-p_0) \right)^m p_t^{\alpha-1} (1-p_t)^{\beta-1} \sum_{k=0}^m \frac{\binom{m}{k}}{\Gamma(\alpha + k)\Gamma(\beta + m - k)} \\ &\quad \times \left(\frac{p_t p_0}{(1-p_t)(1-p_0)} \right)^k \\ &= \frac{\Gamma(\theta + m)}{\Gamma(\alpha)\Gamma(\beta + m)} \left((1-p_t)(1-p_0) \right)^m p_t^{\alpha-1} (1-p_t)^{\beta-1} \sum_{k=0}^m \frac{\Gamma(m+1)}{\Gamma(m+1-k)} \frac{\Gamma(\beta+m)}{\Gamma(\beta+m-k)} \frac{\rho^k}{k!} \\ &= \frac{\Gamma(\theta + m)}{\Gamma(\alpha)\Gamma(\beta + m)} \left((1-p_t)(1-p_0) \right)^m p_t^{\alpha-1} (1-p_t)^{\beta-1} {}_2F_1(-m, -\beta - m + 1; \alpha; \rho), \end{aligned}$$

donde $\rho = \frac{p_t p_0}{(1-p_t)(1-p_0)}$, ${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{a(n)b(n)}{c(n)} \frac{z^n}{n!}$ que representa la función hiperbólica de Gauss y teniendo en cuenta que el factorial ascendente satisface $(-a)_{(x)} = (a - x + 1)_{(x)}$.

Por otro lado, $\mathbb{P}(Y_t = m) = \mathbb{E}\left(e^{-\phi_t X_0} \frac{(\phi_t X_0)^m}{m!} | X_0\right) = \frac{\phi_t^m}{m!} \frac{\theta(m)}{(1+\phi_t)^{\theta+m}}$ con $X_0 \sim \text{Gamma}(\theta, 1)$.

Finalmente, como

$$p(p_t | p_0) = \sum_{m=0}^{\infty} p(p_t | p_0, Y_t = m) \mathbb{P}(Y_t = m),$$

y sustituyendo ϕ_t obtenemos la función de transición

$$p(p_t | p_0) = \frac{p_t^{\alpha-1} (1-p_t)^{\beta-1} (1-e^{-ct})^\theta}{\Gamma(\theta)\Gamma(\alpha)} \sum_{m=0}^{\infty} \frac{\Gamma(\theta + m)^2}{\Gamma(\beta + m)m!} \left(e^{-ct} (1-p_t)(1-p_0) \right)^m {}_2F_1(-m, -\beta - m + 1; \alpha; \rho).$$

Bajo este enfoque hemos encontrado un proceso estacional $\{p_t\}$ con distribuciones marginales $Beta(\alpha, \beta)$.

Posteriormente, es posible asociar un proceso de difusión al modelo anterior. Primero, recordamos que si $p \sim Beta(\alpha, \beta)$ tiene momentos de la forma $\mathbb{E}[p^j] = \frac{a^{(j)}}{(a+b)^{(j)}}$, y $P_t|Y_{1t}, Y_t \sim Beta(\alpha + Y_{1t}, \theta - \alpha + Y_t - Y_{1t})$.

$$\begin{aligned}\mathbb{E}\left[P_t^j|P_0 = p_0, Y_t = m\right] &= \mathbb{E}\left[\frac{(\alpha + Y_{1t})^{(j)}}{(\theta + m)^{(j)}}|P_0 = p_0, Y_t = m\right] \\ &= \frac{1}{(\theta + m)^{(j)}} \sum_{k=0}^m \binom{m}{k} p_0^k (1-p_0)^{m-k} (\alpha + k)^{(j)}.\end{aligned}$$

Luego, sustituyendo $\phi_t = (e^{ct} - 1)^{-1}$ en la densidad de Y_t ,

$$\mathbb{P}(Y_t = m) = \frac{\theta^{(m)}}{m!} (1 - e^{-ct})^\theta e^{-mct},$$

los momentos no condicionales están dados por

$$\sum_{m=0}^{\infty} \mathbb{E}\left[P_t^j|P_0 = p_0, Y_t = m\right] \mathbb{P}(Y_t = m).$$

En particular,

$$\begin{aligned}\mathbb{E}[(P_t - P_0) | P_0 = p_0] &= \sum_{m=0}^{\infty} \frac{1}{\theta + m} \left(\sum_{k=0}^m \binom{m}{k} p_0^k (1-p_0)^{m-k} (\alpha + k) - p_0 \right) \frac{\theta^{(m)}}{m!} (1 - e^{-ct})^\theta e^{-mct} \\ &= (1 - e^{-ct})^\theta \sum_{m=0}^{\infty} \frac{1}{\theta + m} (\alpha + mp_0) - p_0 \frac{\theta^{(m)} e^{-mct}}{m!} \\ &= (1 - e^{-ct})^\theta \left[\frac{\alpha}{\theta} \sum_{m=0}^{\infty} \frac{\theta}{\theta + m} \frac{\theta^{(m)} e^{-mct}}{m!} - p_0 \sum_{m=0}^{\infty} \frac{\theta}{\theta + m} \frac{\theta^{(m)} e^{-mct}}{m!} \right] \\ &= \frac{(1 - e^{-ct})^\theta (\alpha - \theta p_0) {}_2F_1(\theta, \theta; \theta + 1; e^{-mct})}{\theta}.\end{aligned}$$

De acuerdo a la media infinitesimal, varianza infinitesimal y los momentos no condicionales, suponiendo que $\theta > 1$, se puede encontrar,

$$\mu(p) = \frac{c}{\theta - 1} (\alpha - \theta p),$$

análogamente,

$$\sigma(p) = \frac{2c}{\theta - 1} p(1 - p).$$

Así, la difusión asociada puede verse como la solución a la ecuación

$$dP_t = \frac{c}{\theta - 1} (\alpha - \theta p) dt + \sqrt{\frac{2c}{\theta - 1} p(1 - p)} dB_t.$$

Haciendo $c = \frac{\theta - 1}{2}$ obtenemos

$$dP_t = \frac{1}{2} [\alpha - \theta P_t] dt + \sqrt{P_t(1 - P_t)} dB_t,$$

que es la ecuación diferencial estocástica de Wright-Fisher de la ecuación (3.1).

Utilizando esta construcción podemos escribir un algoritmo para la simulación de trayectorias de la difusión de Wright-Fisher:

Algoritmo 6 Simulación del proceso de Wright-Fisher a partir de la construcción vía procesos de Markov

- 1: Generar x_1 con densidad $\text{Gamma}(\alpha, 1)$ y x_2 con densidad $\text{Gamma}(\beta, 1)$.
 - 2: Hacer $c = \frac{\theta-1}{2}$ y $\phi_t = (e^{ct} - 1)^{-1}$.
 - 3: Generar y_1 con densidad $Y|X_1 \sim \text{Pois}(\phi_t x_1)$.
 - 4: Generar x_1 con densidad $X_1|Y \sim \text{Gamma}(\alpha + y_1, 1 + \phi_t)$.
 - 5: Generar y_2 con densidad $Y|X_2 \sim \text{Pois}(\phi_t x_2)$.
 - 6: Generar x_2 con densidad $X_2|Y \sim \text{Gamma}(\beta + y_2, 1 + \phi_t)$.
 - 7: Hacer $X = \frac{x_1}{x_1 + x_2}$.
-

Es importante señalar que esta construcción es muy sencilla de simular pues sólo depende de funciones de distribución conocidas.

Realizamos el mismo estudio que con los enfoques anteriores simulando en el intervalo de tiempo $[0,5]$ con valor inicial $X_0 = 0.5$, y los mismos parámetros $(\alpha = 0.7, \beta = 0.7)$, $(\alpha = 1, \beta = 1)$, $(\alpha = 2, \beta = 2)$, $(\alpha = 3, \beta = 1)$ y $(\alpha = 10, \beta = 20)$. En la Figura 3.16 presentamos una trayectoria con cada uno de los parámetros y 100 puntos simulados por unidad de tiempo.

Es importante resaltar que las trayectorias generadas con éste método visualmente son más variables respecto a los algoritmos previos. Más adelante veremos una comparación un tanto más detallada entre las distintas formas de simular que presentamos.

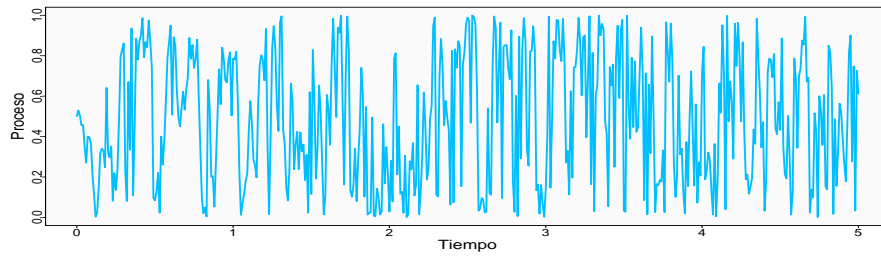
Ya que esta construcción tampoco tiene alguna restricción respecto al intervalo de tiempo en que se pueden simular observaciones, es posible generar valores directamente del tiempo 0 al tiempo t , con $t = 0.5, 1, 2.5, 5$, o bien, obtener trayectorias con Δ pequeños. A continuación, analizamos ambos casos.

Realizamos las pruebas de Kolmogorov-Smirnov para comparar con la distribución estacionaria, los p -values se muestran en la Tabla 3.10 con 2000 simulaciones del proceso generadas del tiempo cero al tiempo t donde vemos que al tiempo $t = 2.5$ con todos los parámetros no se rechaza la hipótesis de igualdad en distribución.

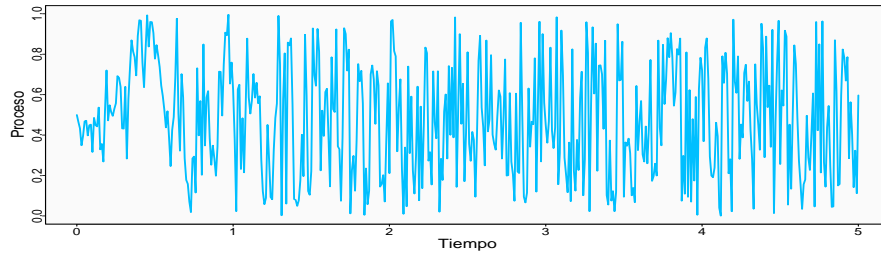
Parámetros	Tiempo	p-value
$\alpha = 1$	0.5	< 2.2e-16
	1	4.0e-04
	$\beta = 1$	0.230
$\beta = 1$	5	0.479
	$\alpha = 2$	0.013
	1	0.030
$\beta = 2$	2.5	0.609
	5	0.325
	$\alpha = 3$	0.5
$\beta = 1$	1	< 2.2e-16
	2.5	0.541
	5	0.317
$\alpha = 10$	0.5	0.312
	1	0.173
	$\beta = 20$	2.5
	5	0.109

Tabla 3.10: Resultados de la pruebas de bondad de ajuste para simulaciones con la construcción vía procesos de Markov contra la distribución estacionaria.

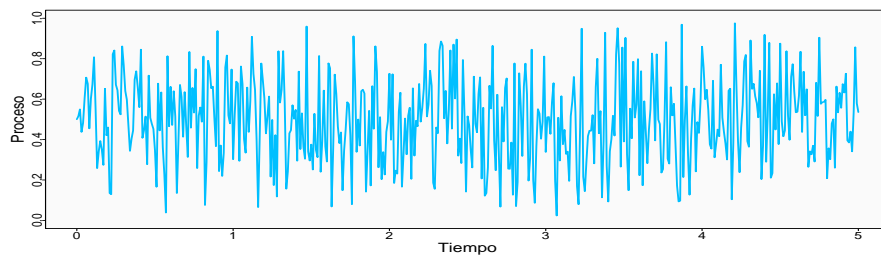
Ahora, realizamos las mismas pruebas para muestras obtenidas con simulaciones para distintos valores Δ y los parámetros de estudio. Los resultados p -values se muestran en la Tabla 3.11 donde se observa que a diferencia de las pruebas anteriores desde el tiempo $t = 0.5$ no se rechaza la hipótesis de igualdad de distribución.



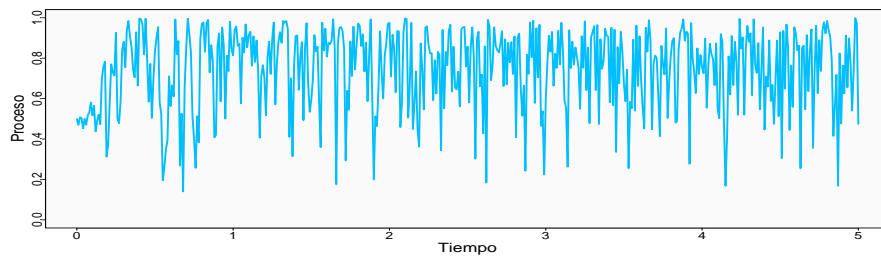
(a) $\alpha = 0.7$ y $\beta = 0.7$



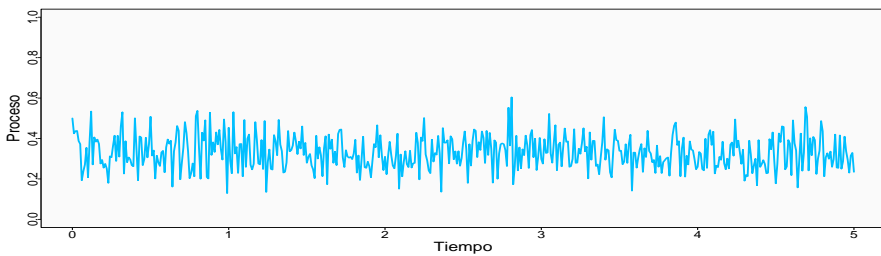
(b) $\alpha = 1$ y $\beta = 1$



(c) $\alpha = 2$ y $\beta = 2$



(d) $\alpha = 3$ y $\beta = 1$

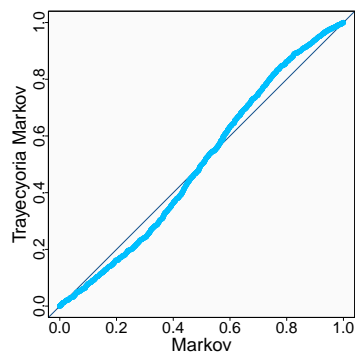


(e) $\alpha = 10$ y $\beta = 20$

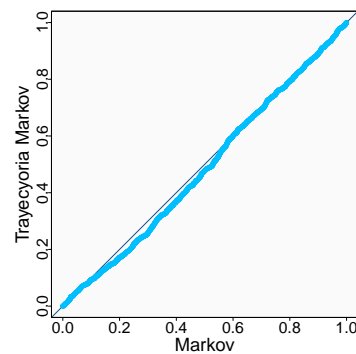
Figura 3.16: Trayectorias simuladas con la construcción vía procesos de Markov en el intervalo de tiempo $[0,5]$ con 100 puntos simulados por unidad de tiempo y $X_0 = 0.5$.

Parámetros	Tiempo	$\Delta = 0.01$	$\Delta = 0.002$	$\Delta = 0.001$	$\Delta = 0.0005$
$\alpha = 1$	0.5	0.485	0.128	0.549	0.061
	1	0.375	0.564	0.979	0.422
$\beta = 1$	2.5	0.175	0.290	0.178	0.691
	5	0.495	0.395	0.154	0.191
$\alpha = 2$	0.5	0.165	0.04	0.623	0.103
	1	0.324	0.859	0.547	0.649
$\beta = 2$	2.5	0.628	0.39	0.633	0.449
	5	0.064	0.908	0.106	0.52
$\alpha = 3$	0.5	0.172	0.773	0.226	0.312
	1	0.582	0.762	0.368	0.319
$\beta = 1$	2.5	0.654	0.907	0.769	0.944
	5	0.303	0.157	0.199	0.965
$\alpha = 10$	0.5	0.651	0.581	0.730	0.652
	1	0.983	0.123	0.854	0.082
$\beta = 20$	2.5	0.782	0.673	0.549	0.358
	5	0.057	0.051	0.78	0.088

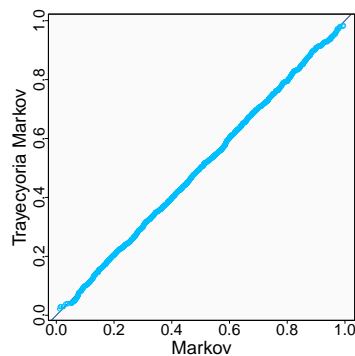
Tabla 3.11: Resultados de la pruebas de bondad de ajuste para simulaciones de trayectorias con la construcción vía cadenas de Markov contra la distribución estacionaria.



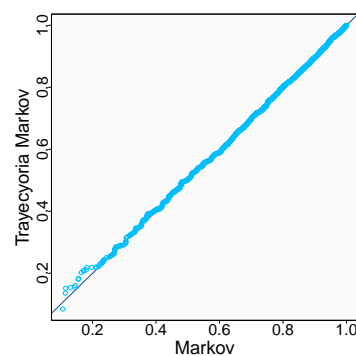
(a) ($\alpha = 0.7, \beta = 0.7$)



(b) ($\alpha = 1, \beta = 1$)



(c) ($\alpha = 2, \beta = 2$)



(d) ($\alpha = 3, \beta = 1$)

Figura 3.17: QQplot comparando la simulación de 2000 observaciones simuladas de 0 a $t = 1$ contra una muestra de 2000 trayectorias al tiempo $t = 1$ con $\Delta = 0.01$, ambas generadas con la construcción vía cadenas de Markov.

Adicionalmente, comparamos mediante un gráfico las muestras generadas de las dos maneras, es decir, de las 2000 simulaciones tomamos una muestra al tiempo medio $t = 2.5$ y $\Delta = 0.01$ generamos un qqplot para cada conjunto de parámetros. En la Figura 3.17 vemos que con $(\alpha = 0.7, \beta = 0.7)$ ambas muestras provienen de distribuciones diferentes, observamos ligeras diferencias con $(\alpha = 1, \beta = 1)$, sin embargo, en ambos casos siguen una distribución Beta. En la gráfica del resto de los parámetros los puntos caen prácticamente sobre la recta identidad, lo cual sugiere que tiene la misma distribución.

Veamos gráficamente la evolución del comportamiento de las trayectorias simuladas en distintos tiempos. Tomando $(\alpha = 1, \beta = 1)$, en la Figura 3.18 se muestran los qqplot donde es claro que en todos los tiempos las muestras provienen de una distribución Beta con los parámetros respectivos.

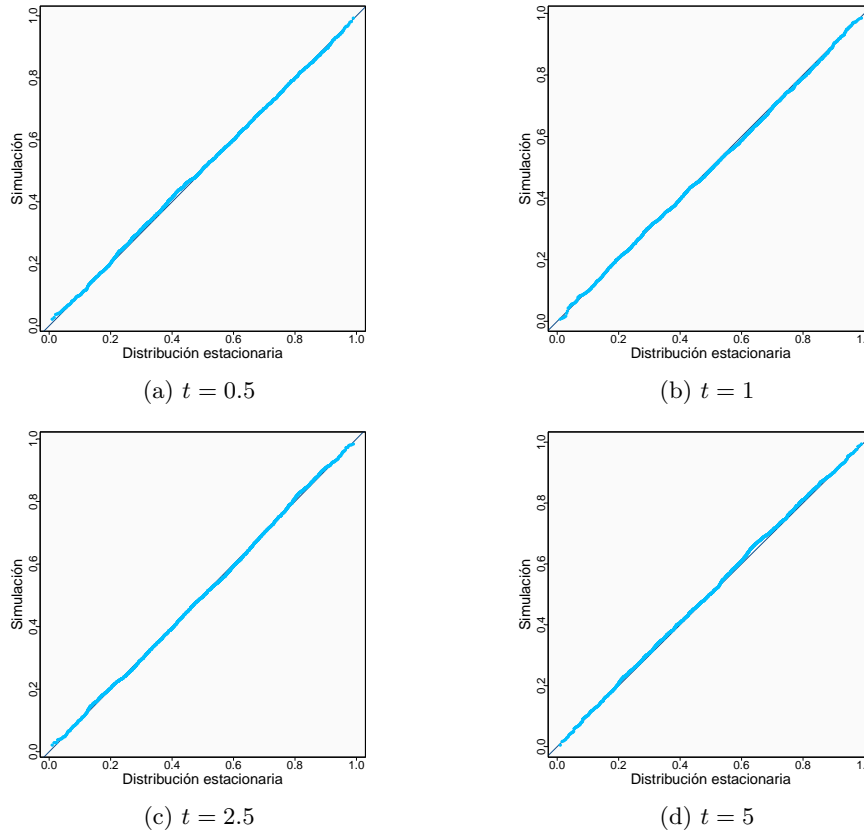


Figura 3.18: QQplot comparando la muestra de 2000 trayectorias en distintos tiempos simuladas con la construcción vía cadenas de Markov con parámetros $(\alpha = 1, \beta = 1)$ contra la distribución estacionaria.

Las conclusiones cuando los parámetros son diferentes son análogas al caso descrito anteriormente. Es decir, con $(\alpha = 3, \beta = 1)$, las muestras se comportan como la distribución estacionaria desde tiempos pequeños. Mientras que al hacer más grandes los parámetros, y consecuentemente disminuir la variabilidad del proceso, las muestras se comportan como la distribución estacionaria.

Finalmente, para evaluar el comportamiento respecto al número de simulaciones, probamos generando 100, 500, 1000, 1500, 2000, 5000 y 10000 puntos de la trayectoria con valor inicial aleatorio, los parámetros y los tiempos que hemos trabajado a lo largo del capítulo.

En la Tabla 3.12 presentamos los p -values de la prueba de Kolmogorov-Smirnov para contrastar con la distribución estacionaria. Con base en los resultados de la prueba observamos que el algoritmo tiene un rendimiento similar cuando el valor inicial es un número aleatorio, como se indica en el Algoritmo 6, que cuando se simulan trayectorias con un punto inicial fijo.

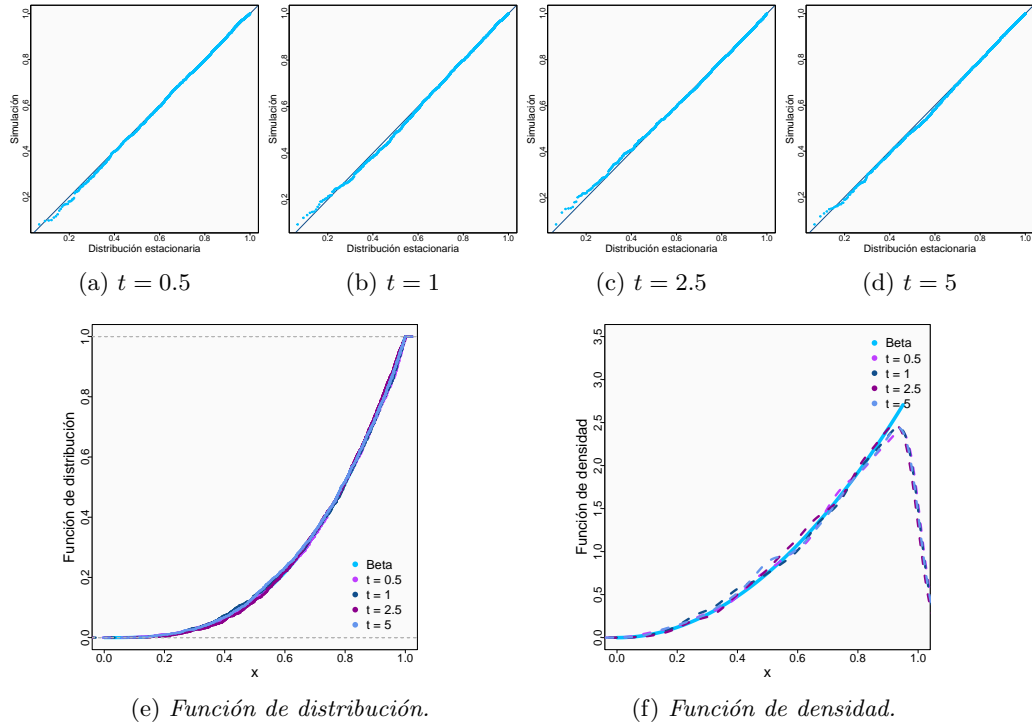


Figura 3.19: Comparaciones gráficas de la muestra de 2000 trayectorias en distintos tiempos simuladas con la construcción vía cadenas de Markov de parámetros ($\alpha = 3$, $\beta = 1$) contra la distribución estacionaria.

Parámetros	Tiempo	100	500	1000	1500	2000	5000	10000	
$\alpha = 1$	0.05	0.009	0.009	0.332	0.618	0.636	0.292	0.595	
	1	0.296	0.911	0.432	0.306	0.22	0.317	0.597	
	$\beta = 1$	2.5	0.517	0.597	0.936	0.366	0.48	0.149	0.416
		5	0.176	0.167	0.366	0.873	0.679	0.863	0.851
$\alpha = 2$	0.5	0.045	0.17	0.24	0.5	0.451	0.303	0.044	
	1	0.345	0.185	0.205	0.384	0.778	0.685	0.675	
	$\beta = 2$	2.5	0.483	0.826	0.858	0.571	0.551	0.851	0.792
		5	0.085	0.437	0.335	0.206	0.159	0.975	0.869
$\alpha = 3$	0.5	0.826	0.072	0.004	0.087	0.15	0.731	0.861	
	1	0.256	0.417	0.25	0.31	0.273	0.679	0.349	
	$\beta = 1$	2.5	0.515	0.342	0.803	0.867	0.359	0.436	0.552
		5	0.806	0.978	0.999	0.888	0.624	0.857	0.704
$\alpha = 10$	0.5	0.703	0.923	0.904	0.515	0.320	0.141	0.586	
	1	0.031	0.116	0.064	0.04	0.46	0.121	0.48	
	$\beta = 20$	2.5	0.232	0.924	0.708	0.705	0.308	0.001	0.413
		5	0.088	0.786	0.68	0.787	0.664	0.656	0.536

Tabla 3.12: Resultados de la pruebas de bondad de ajuste para simulaciones con la construcción vía procesos de Markov contra la distribución estacionaria con distinto número de simulaciones.

3.4. Comparación gráfica de los algoritmos.

Ya que hemos estudiado diversos métodos de simulación y observamos algunas diferencias entre ellos, en esta sección realizamos pruebas gráficas para comparar los algoritmos utilizando distintos parámetros con el fin de investigar cuáles son los más adecuados. Elegimos escenarios considerando los parámetros con los que ya hemos trabajado.

Entre las principales diferencias se encuentran las gráficas de las trayectorias. Las trayectorias obtenidas mediante los tres métodos numéricos y el método exacto son muy similares entre ellas (Figuras 3.1, 3.5, 3.7 y 3.12) pero muy diferentes a las obtenidas mediante la construcción vía procesos de Markov (Figura 3.16) debido a la magnitud de los cambios entre un tiempo y otro y la variabilidad del proceso.

Respecto a la comparación con la distribución estacionaria, obtuvimos resultados análogos con las aproximaciones de Euler y Milstein los cuales dependen principalmente de los parámetros, es decir, solo bajo ciertos parámetros es posible obtener muestras de la distribución estacionaria. Por otro lado, notamos en nuestras simulaciones que el algoritmo BISS no solo depende de los parámetros sino de qué tan pequeño es el incremento Δ . Finalmente, el algoritmo exacto y la construcción vía procesos de Markov parecen ser mejores opciones pues es posible obtener muestras de la distribución estacionaria con cada uno de los parámetros.

Primero realizamos un gráfico qqplot, Figura 3.20, para evaluar el comportamiento entre cada uno de los métodos, para este fin realizamos 2000 simulaciones de trayectorias con $(\alpha = 0.7, \beta = 0.7)$, valor inicial $X_0 = 0.5$ en el intervalo de tiempo $[0,5]$ y tomamos la muestra de la trayectoria al tiempo final, además tomamos $\Delta = 0.001$. Recordando que la distribución estacionaria no existe si $\alpha < 1$ o $\beta < 1$ no fue posible contrastar el comportamiento de estos parámetros anteriormente.

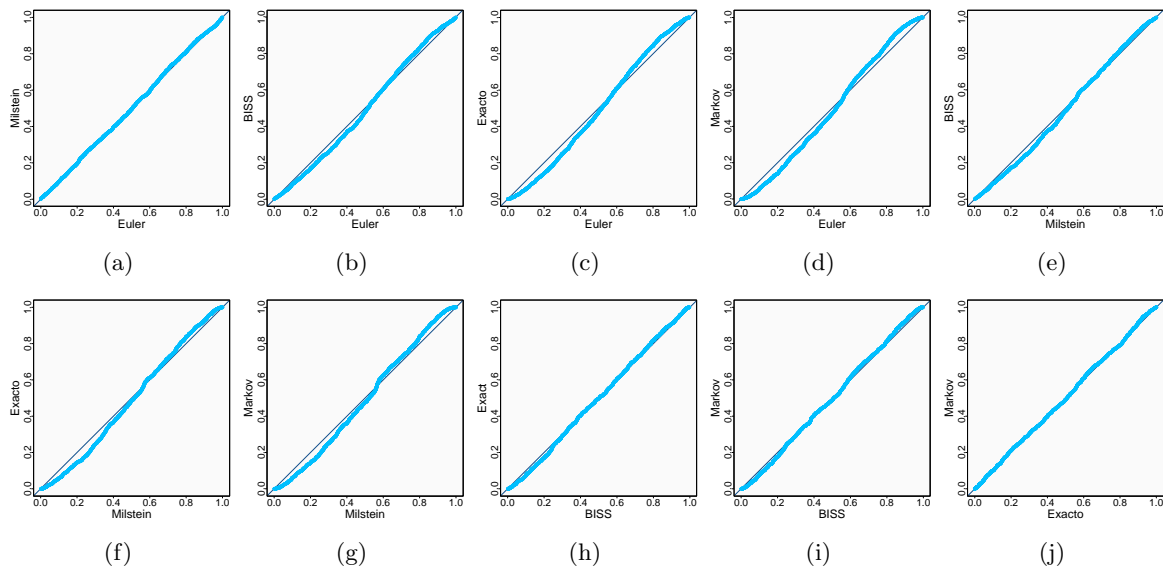


Figura 3.20: QQplot comparando los resultados de distintos algoritmos con $(\alpha = 0.7, \beta = 0.7)$, simulando 2000 trayectorias con valor inicial $X_0 = 0.5$ en el intervalo de tiempo $[0,5]$ y la muestra se toma en $t = 5$.

De acuerdo a la Figura 3.20a las aproximaciones de Euler y Milstein tienen un comportamiento similar, el cual evidentemente contrasta con el resto de los enfoques (Figuras 3.20b-3.20g). Por otro lado, en la Figura 3.20j se observa que los puntos caen sobre la recta identidad por lo que el algoritmo exacto y la construcción a partir de procesos de Markov sugieren la misma distribución. Por su parte, el algoritmo tiene un comportamiento similar con ligeras diferencias del lado izquierdo de la distribución respecto a la construcción por procesos de Markov al algoritmo exacto, Figuras 3.20i y 3.20h respectivamente; aún existiendo esas diferencias con una prueba de hipótesis no existe evidencia para rechazar la hipótesis de que siguen la misma distribución.

El siguiente ejemplo considera los parámetros $(\alpha = 1, \beta = 1)$, valor inicial $X_0 = 0.5$ en un intervalo de tiempo $[0,5]$ y $\Delta = 0.001$ para los métodos numéricos. Se generan 2000 simulaciones y tomamos la muestra al tiempo $t = 2.5$.

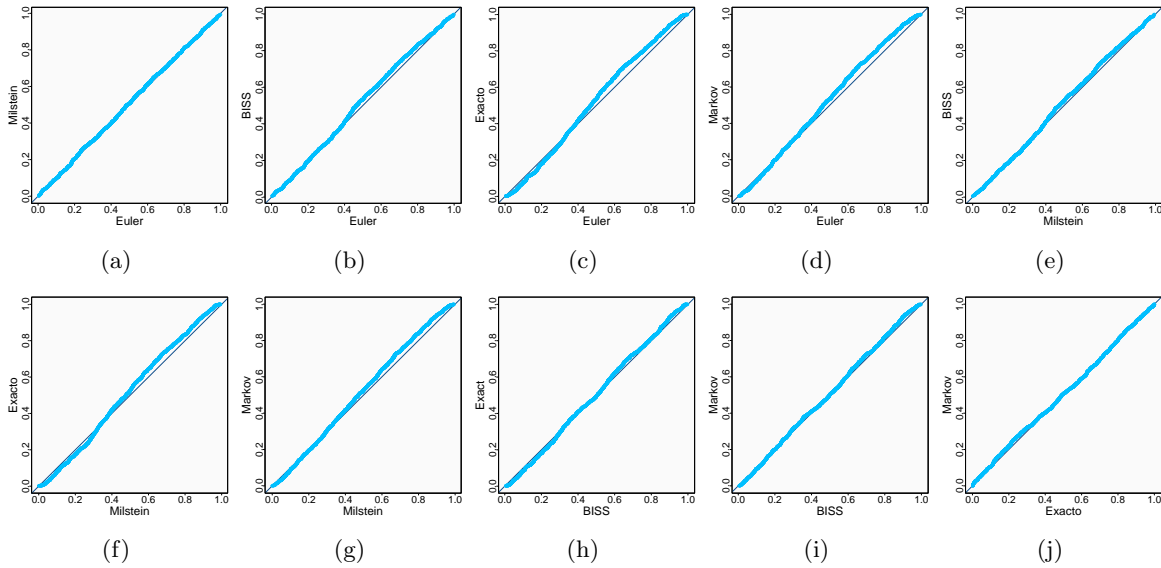


Figura 3.21: QQplot comparando los resultados de distintos algoritmos con $(\alpha = 1, \beta = 1)$, simulando 2000 trayectorias con valor inicial $X_0 = 0.5$ en el intervalo de tiempo $[0,5]$ y la muestra se toma en $t = 2.5$.

Las principales diferencias se observan en la Figura 3.21c, es decir, en la comparación de la aproximación Euler y el algoritmo exacto, sin embargo, también se observan diferencias con el método BISS y la construcción vía procesos de Markov, Figuras 3.21b y 3.21b. Un sesgo similar se tiene con la aproximación Milstein y el algoritmo exacto, Figura 3.21f. El resto de las combinaciones posibles se ajustan bastante bien. Estos resultados tienen sentido pues considerando el análisis de los métodos individuales al tiempo $t = 2.5$ y parámetros $(\alpha = 1, \beta = 1)$ la aproximación de Euler y Milstein no habían alcanzado la distribución estacionaria a diferencia de los otros tres.

Ahora, utilizamos $(\alpha = 2, \beta = 2)$, $\Delta = 0.001$ ya que en las simulaciones con estos parámetros se producen resultados satisfactorios a tiempos chicos, tomamos una muestra de tamaño 2000 al tiempo $t = 1$. En general, observamos comportamientos parecidos en los distintos algoritmos excepto el BISS del lado derecho de la distribución. En las Figuras 3.22b y 3.22b para valores mayores a 0.7 aproximadamente, la gráfica produce una ligera curva por debajo de la recta de pendiente uno y notamos que si intercambiamos los ejes en la Figura 3.22h ocurre el mismo comportamiento.

El siguiente conjunto de parámetros como ya hemos visto es el que genera mayores problemas al contrastar con las distribución estacionaria, lo que sugiere que se presentarán diferencias considerables entre los mecanismos de simulación que hemos estudiado. Sea $(\alpha = 3, \beta = 1)$, nuevamente consideramos la discretización $\Delta = 0.001$ y una muestra de 2000 observaciones obtenidas al final del intervalo de tiempo, es decir, $t = 5$.

En la Figura 3.23 se hace evidente que las aproximaciones de Euler y Milstein tienen un comportamiento similar debido a su construcción, este comportamiento difiere en mayor o menor medida de los otros tres enfoques. En la Figura 3.24j la gráfica queda sobre la recta identidad, por lo que, nuevamente el algoritmo exacto y la construcción a partir de procesos de Markov siguen la misma distribución. No siendo tan estrictos podríamos decir que las aproximaciones de Milstein y la construcción por cadenas de Markov, Figura 3.24g siguen el mismo comportamiento lo cual también puede comprobarse con la prueba de Kolmogorov-Smirnov.

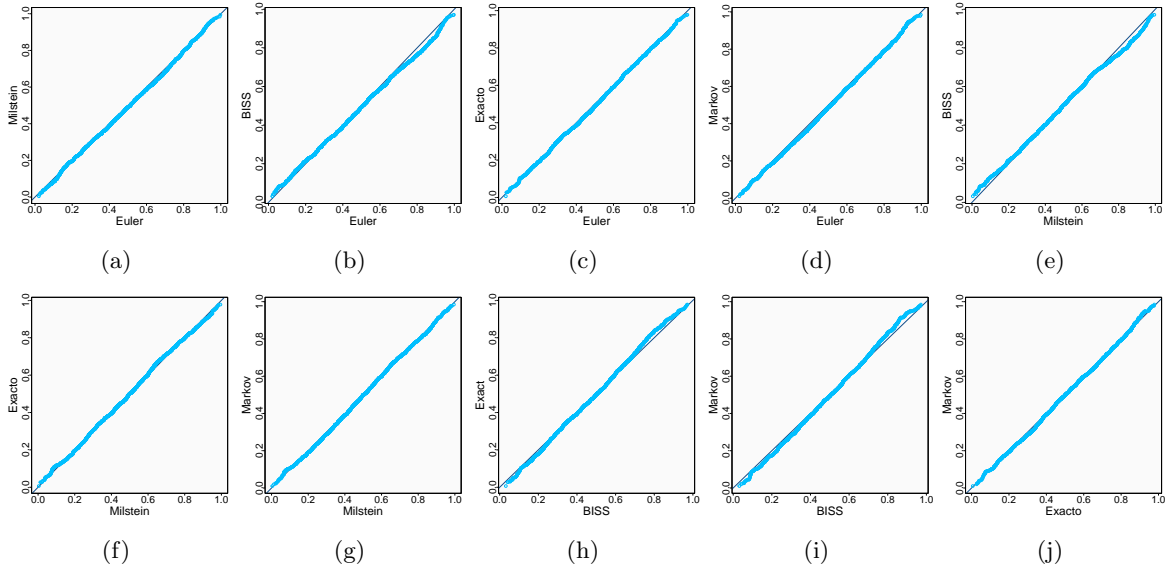


Figura 3.22: QQplot comparando los resultados de distintos algoritmos con $(\alpha = 2, \beta = 2)$, simulando 2000 trayectorias con valor inicial $X_0 = 0.5$ en el intervalo de tiempo $[0,5]$ y la muestra se toma en $t = 1$.

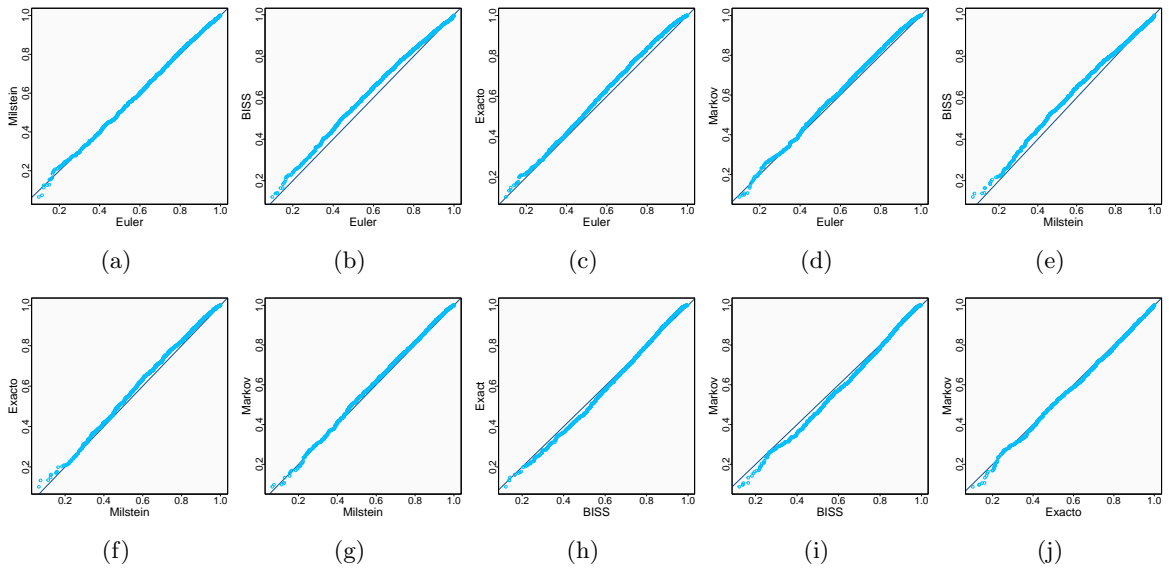


Figura 3.23: QQplot comparando los resultados de distintos algoritmos con $(\alpha = 3, \beta = 1)$, simulando 2000 trayectorias con valor inicial $X_0 = 0.5$ en el intervalo de tiempo $[0,5]$ y la muestra se toma en $t = 5$.

Por último, comparamos muestras de 2000 simulaciones al tiempo $t = 1$ con los parámetros $(\alpha = 10, \beta = 20)$ y $\Delta = 0.001$ para los algoritmos que corresponde. No hay mucho que agregar respecto a estos parámetros pues al comparar con la distribución estacionaria se obtienen buenos resultados, que también se ven reflejados en las comparaciones entre algoritmos, Figura 3.24.

Con la evidencia que hemos recabado hasta ahora, podemos concluir que los algoritmos más precisos para la generación de trayectorias de la difusión de Wright-Fisher son el exacto y el generado a partir de procesos de Markov, pues se comportan bien tanto para parámetros grandes como para parámetros pequeños. Así mismo, al compararlos entre ellos mediante un qqplot o pruebas de hipótesis para muestras en distintos tiempos resulta que tienen un comportamiento similar. Por otro lado, el algoritmo BISS resulta una opción

hasta cierto punto aceptable para algunos parámetros y con una buena elección de Δ .

En cuanto a las aproximaciones numéricas de Euler y Milstein encontramos que se comportan de manera similar entre ellos, pero no siempre similar al resto de los algoritmos, además de los resultados desfavorables con algunos parámetros al comparar con la distribución estacionaria y el problema que se genera al tener espacio de estados acotado.

El siguiente contraste es el tiempo de ejecución, para ello seleccionamos solamente el algoritmo exacto y la construcción por procesos de Markov, registramos el tiempo en segundos que tarda cada uno de ellos en generar 10000 observaciones del tiempo cero a un cierto tiempo t con los parámetros que hemos utilizado a lo largo del estudio y un valor inicial aleatorio X_0 .

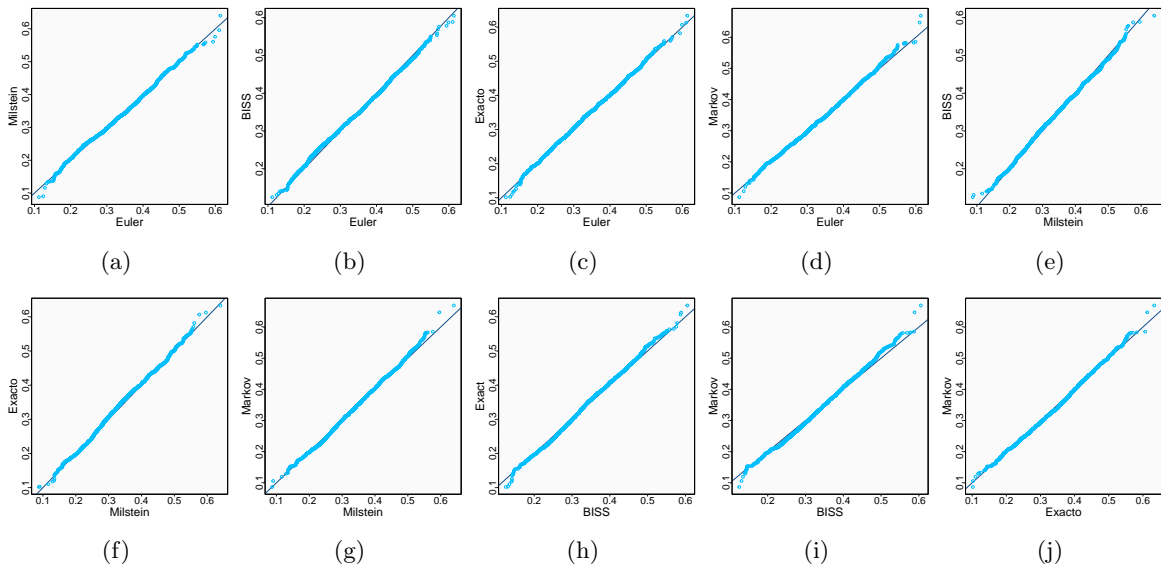


Figura 3.24: *QQplot* comparando los resultados de distintos algoritmos con $(\alpha = 10, \beta = 20)$, simulando 2000 trayectorias con valor inicial $X_0 = 0.5$ en el intervalo de tiempo $[0,5]$ y la muestra se toma en $t = 1$.

Los tiempos se muestran en la Tabla 3.13. Notamos que hay un incremento considerable entre el tiempo 0.01 y 0.05, esto se debe a que el método exacto se basa en la expansión en series de la función de densidad de transición, en la ecuación (3.16) que a su vez depende del coalescente de Kingman (3.17) el cual es inestable numéricamente cuando t tiende a cero y por tanto el Algoritmo 4 hereda dicha característica. Por lo que, de acuerdo a la observación de Jenkins y Spanò en [15], si $t < 0.05$ utilizamos la aproximación dada en la Proposición 3.2.2. Posteriormente, a medida que el tiempo crece, los tiempos de ejecución disminuyen considerablemente aún con la complejidad del Algoritmo 4.

En comparación, la construcción vía procesos de Markov, es más sencilla de implementar y conlleva menor costo computacional pues solo depende de distribuciones conocidas; esto permite que los tiempos de ejecución sean constantes con los distintos parámetros y los distintos tiempos. En ambos casos no observamos diferencias sustanciales al cambiar el valor del proceso al tiempo inicial.

Ya que el interés principal es obtener trayectorias, también registramos los tiempos de ejecución para generar 10000 trayectorias en el intervalo de tiempo $[0,1]$ con cada uno de los parámetros, y diferentes valores de Δ y valor inicial $X_0 = 0.5$. La implementación se hizo en la versión 4.2.0 de R en una laptop DELL Inspiron 3521 con Procesador Intel(R) Core(TM) i3-3227U CPU @ 1.90GHz, 1901 Mhz, 2 procesadores principales, 4 procesadores lógicos y 6GB de memoria RAM.

Parámetros	Tiempo	Exacto		Proceso de Markov	
		$X_0 = 0.01$	$X_0 = 0.5$	$X_0 = 0.01$	$X_0 = 0.5$
$\alpha = 0.7$	0.01	2.32	2.61	3.53	3.59
	0.05	1457.57	1887.78	3.22	3.03
$\beta = 0.7$	0.1	707.02	702.87	3.36	3.02
	1	13.52	13.62	3.38	3.06
$\alpha = 1$	0.01	3.01	3.03	3.05	3.40
	0.05	3332.78	3351.50	3.11	3.01
$\beta = 1$	0.1	692.71	691.58	3.43	3.05
	1	10.54	12.05	3.46	3.17
$\alpha = 2$	0.01	2.98	3.03	3.02	3.41
	0.05	3186.41	3296.92	3.00	3.41
$\beta = 2$	0.1	643.08	659.90	3.08	3.05
	1	9.49	9.19	3.42	3.03
$\alpha = 3$	0.01	3.13	3.11	3.39	3.09
	0.05	3362.11	3340.21	3.11	3.41
$\beta = 1$	0.1	644.78	651.31	3.02	3.39
	1	8.89	9.31	3.03	3.27
$\alpha = 10$	0.01	3.07	2.95	3.01	3.03
	0.05	1649.87	1622.06	3.42	2.98
$\beta = 20$	0.1	200.96	197.38	3.44	3.05
	1	5.67	6.66	3.39	3.03

Tabla 3.13: Comparación del tiempo en segundos que se necesitan para generar 1000 observaciones con el método exacto y la construcción vía procesos de Markov en distintos tiempos.

Parámetros	Δ	Exacto	Proceso de Markov
		$X_0 = 0.5$	$X_0 = 0.5$
$\alpha = 0.7$	0.01	74.7	60.25
	0.05	295.84	259.47
$\beta = 0.7$	0.1	552.53	507.19
	1	1069.26	991.5
$\alpha = 1$	0.01	56.25	64.75
	0.05	523.21	270.59
$\beta = 1$	0.1	534.42	518.45
	1	1054.86	991.59
$\alpha = 2$	0.01	62.86	53.86
	0.05	274.06	256.62
$\beta = 2$	0.1	523.21	505.80
	1	1043.08	1007.71
$\alpha = 3$	0.01	67.90	61.47
	0.05	280.28	266.35
$\beta = 1$	0.1	534.00	503.63
	1	1047.31	985.67
$\alpha = 10$	0.01	79.25	64.50
	0.05	269.31	258.50
$\beta = 20$	0.1	540.00	520.85
	1	1047.44	1061.77

Tabla 3.14: Comparación del tiempo en segundos que se necesitan para generar 1000 trayectorias con el método exacto y la construcción vía procesos de Markov.

Los tiempos de ejecución se registran en la Tabla 3.14, es claro que cuando las trayectorias del algoritmo

exacto se generan mediante la aproximación al coalescente de Kingman, es decir, cuando $\Delta = 0.01$, los tiempos son comparables. De manera similar ocurre cuando $\Delta = 1$.

3.5. Aplicación de las simulaciones

Como ya hemos mencionado, las simulaciones son de gran utilidad para el estudio de las difusiones. A continuación las vamos a utilizar para encontrar la solución a ciertos funcionales, que como vimos en la Subsección 2.2.3 solo tienen solución explícita bajo ciertas condiciones.

Ejemplo 3.5.1. *Supongamos que solo hay mutación de A a a , es decir, $\alpha = 0$ y $\beta > 0$.*

Tenemos que la probabilidad de llegar al estado b antes que al estado c dado que iniciamos en x con la parametrización que estamos utilizando es la siguiente,

$$u(x) = \frac{(1-c)^{-\beta+1} - (1-x)^{-\beta+1}}{(1-c)^{-\beta+1} - (1-b)^{-\beta+1}}.$$

Sea $\beta = 2$, $c = 0.1$ y $b = 0.6$ y vamos a considerar dos escenarios dependiendo del valor inicial $X_0 = 0.2$ y $X_0 = 0.5$. Entonces $\mathbb{P}_{0.2}(\tau_{0.6} < \tau_{0.1}) = u(0.2) = 0.1$ y $\mathbb{P}_{0.5}(\tau_{0.6} < \tau_{0.1}) = u(0.5) = 0.64$. Para el cálculo de dichas probabilidades, generamos 2000 trayectorias con $\Delta = 0.0001$ y calculamos el número de veces que el proceso alcanzó el estado 0.6 antes que el estado 0.1. De los cinco métodos que hemos estudiado solo podemos utilizar Euler y Milstein debido a que en ellos no se les pide que las tasas de mutación sean estrictamente positivas.

X_0	Probabilidad	Euler	Milstein
0.2	0.1	0.1085	0.1083
0.5	0.64	0.621	0.6255

Tabla 3.15: Cálculo de probabilidades de llegada con los distintos métodos de simulación.

Como podemos notar, las probabilidades estimadas mediante simulaciones son cercanas al valor calculado analíticamente.

Ejemplo 3.5.2. *Supongamos que las tasas de mutación son positivas, por ejemplo, $\alpha = 2$ y $\beta = 2$.*

Recordamos de la ecuación (2.6) que la probabilidad de llegar al estado b antes que al estado c dado que iniciamos en x es,

$$u(x) = \frac{S(x) - S(c)}{S(b) - S(c)}, \tag{3.24}$$

donde, de acuerdo a la ecuación (2.7), para un x_0 arbitrario y la reparametrización adecuada,

$$\begin{aligned} S(x) &= k_2 \int_{x_0}^x y^{-\alpha} (1-y)^{-\beta} dy \\ &= k_2 \int_{x_0}^x y^2 - 2y^3 + y^4 dy \end{aligned} \tag{3.25}$$

Esa integral es sencilla de calcular y para simplificar tomamos $x_0 = 0$.

Sea $c = 0.1$ y $b = 0.6$ y probamos con valor inicial $X_0 = 0.2$ y $X_0 = 0.5$. Entonces $\mathbb{P}_{0.2}(\tau_{0.6} < \tau_{0.1}) = u(0.2) = 0.45$ y $\mathbb{P}_{0.5}(\tau_{0.6} < \tau_{0.1}) = u(0.5) = 0.88$.

En la Tabla 3.16 mostramos la comparación de las probabilidades teóricas contra las obtenidas mediante simulaciones con los métodos estudiados. Adicionalmente, agregamos una comparación con $\alpha = 3$ y $\beta = 1$.

Parámetros	X_0	Probabilidad	Euler	Milstein	BISS	Exacto	Markov
$\alpha = 2, \beta = 2$	0.2	0.4529	0.4645	0.465	0.4515	0.472	0.5015
	0.5	0.8898	0.898	0.893	0.9005	0.8875	0.895
$\alpha = 3, \beta = 1$	0.2	0.7273	0.745	0.730	0.8345	0.736	0.815
	0.5	0.9773	0.9815	0.978	0.987	0.98	0.986

Tabla 3.16: Cálculo de probabilidades de llegada con los distintos métodos de simulación.

3.6. Simulación de puentes de la difusión de Wight-Fisher.

En algunas ocasiones se tienen observaciones a tiempo discreto, esto es, en los tiempos $t_0, t_1 \dots, t_n$ mientras que las difusiones evolucionan a tiempo continuo, por lo que nos interesa poder simular trayectorias que nos ayuden a hacer inferencia sobre los parámetros que describen el comportamiento del proceso. Un puente de difusión es un proceso $\{X_t\}$ en el intervalo de tiempo $[0, T]$ condicionado a $X_0 = a$ y $X_T = b$, son limitados los procesos en los cuales es posible encontrar la función de densidad de transición del proceso condicionado y por consiguiente un algoritmo exacto para la construcción del puente. Por esta razón se han desarrollado algoritmos para poder simular este tipo de trayectorias.

3.6.1. Aproximación de puentes de difusiones.

El siguiente método de simulación de puentes fue propuesto por Bladt y Sørensen [2] se basa principalmente en el proceso a tiempo reverso. 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 que corre hacia atrás en el tiempo, es decir, de T a 0. La idea es considerar una trayectoria de cada proceso y si éstas se cruzan, se combinan para generar una nueva trayectoria. 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. En la Figura (3.25) mostramos la construcción descrita de un puente.

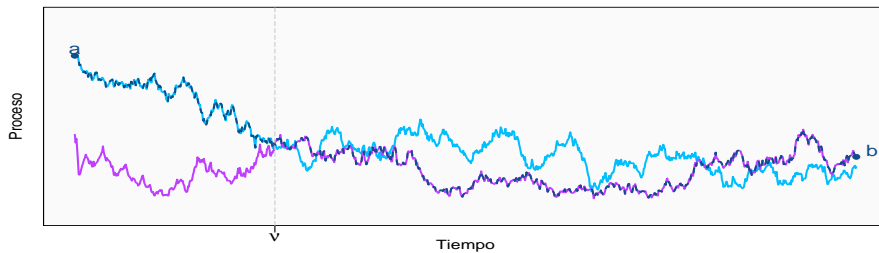


Figura 3.25: Construcción de un puente de difusión. La primera vez que se cruza el proceso X^1 en (azul claro) con valor inicial a y X^2 en (morado) con valor inicial b es en el tiempo v . El puente resultante se muestra en azul oscuro con línea punteada.

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 $\{X_t^1\}_{t \geq 0}$ y $\{X_t^2\}_{t \geq 0}$ como solución a la ecuación diferencial estocástica

$$dX_t^i = \mu(X_t^i)dt + \sigma(X_t^i)dB_t^i, \quad i = 1, 2, \quad X_0^1 = a \quad \text{y} \quad X_0^2 = b, \quad (3.26)$$

con $\{B_t^i\}$ movimientos brownianos estándar independientes. Y al proceso reverso en el tiempo $\tilde{X}_t^2 = X_{T-t}^2$ para $t \in [0, T]$.

Los procesos $\{X_t^1\}$ y $\{X_t^2\}$ se simulan con algún método conocido, por ejemplo, alguno de los descritos en la sección anterior. Sean $Y^1 = Y_{\Delta j}^1, j = 0, 1, \dots, N$ y $Y^2 = Y_{\Delta j}^2, j = 0, 1, \dots, N$ simulaciones independientes de los procesos $\{X_t^1\}$ y $\{X_t^2\}$ respectivamente en el intervalo de tiempo $[0, T]$ 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, es decir que existe alguna j , tal que cumple

$$Y_{\Delta j}^1 \geq Y_{\Delta(N-j)}^2 \text{ y } Y_{\Delta(j+1)}^1 \leq Y_{\Delta(N-(j+1))}^2 \quad \text{o} \quad Y_{\Delta j}^1 \leq Y_{\Delta(N-j)}^2 \text{ y } Y_{\Delta(j+1)}^1 \geq Y_{\Delta(N-(j+1))}^2.$$

Así obtenemos una trayectoria del puente definida de la siguiente manera,

$$W_{\Delta j} = \begin{cases} Y_{\Delta j}^1 & j = 0, 1, \dots, \nu - 1 \\ Y_{\Delta(N-j)}^2 & j = \nu, \dots, N, \end{cases} \quad (3.27)$$

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 $\{W\}$ definido en (3.27) aproxima al puente de difusión bajo las condiciones del Teorema 3.6.1. La demostración del teorema puede encontrarse en [2] y es muy útil para la construcción de puentes.

Antes de enunciar dicho teorema, cabe mencionar que el valor de Δ tiene influencia sobre la probabilidad de que las trayectorias se crucen, por lo que es recomendable elegir Δ más pequeño de lo usual y por su parte, la probabilidad de rechazo, es decir, la probabilidad de que las trayectorias no se crucen dependerá de los coeficientes de deriva y de difusión, además de los valores a y b y la longitud del intervalo T .

El siguiente Lema referente a la distribución del proceso reverso también puede encontrarse en [2].

Lema 3.6.1. *El proceso reverso, $\{\tilde{X}_t\}$ definido por $\tilde{X}_t = X_{T-t}$, y el proceso $\{X_t\}$ condicionado a $X_T = b$ tienen las mismas densidades de transición,*

$$q(x, s, y, t) = \frac{p_{t-s}(x, y)p_{T-t}(y, b)}{p_{T-s}(x, b)} = \frac{p_{t-s}(y, x)p_{T-t}(b, y)}{p_{T-s}(b, x)}, \quad s < t < T. \quad (3.28)$$

Además, si se cumple

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

con $m(x)$ la función de velocidad definida en la ecuación (1.22). Entonces, la distribución de $\{\tilde{X}_t\}$ es igual a la distribución condicional del proceso $\{X_t\}$ con $X_0 \sim \pi$ dado que $X_T = b$, donde π es la distribución invariante o estacionaria.

Teorema 3.6.1. *Sea $\{X_t\}_{t \geq 0}$ una difusión con espacio de estados (l, r) ,*

$$\tau = \inf\{0 \leq t \leq T | X_t^1 = X_{T-t}^2\}$$

(por convención, $\inf\{\emptyset\} = \infty$) y definamos

$$Z_t = \begin{cases} X_t^1 & \text{si } 0 \leq t \leq \tau \\ X_{T-t}^2 & \text{si } \tau \leq t \leq T. \end{cases} \quad (3.29)$$

Asumimos que

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

con

$$m(x)$$

la función de velocidad definida en la ecuación (1.22). Esto es,

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

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 el intervalo de tiempo $[0, T]$ con valor inicial a y valor final b , condicionado al evento de que el puente se cruza con una difusión independiente con ecuación diferencial dada por la ecuación (3.26) y distribución inicial $p_T(b, \cdot)$.

Además puede probarse que para difusiones ergódicas la probabilidad de que el puente sea una difusión independiente es cercana a uno si T es grande.

Ya que este es un método genérico que se basa en que sea posible simular trayectorias de la difusión, cualquiera de los algoritmos ya descritos será de utilidad para construir el puente de Wright-Fisher. Dados los resultados que obtuvimos al comparar las simulaciones de cada uno de los enfoques decidimos utilizar el algoritmo exacto y la construcción que utiliza el hecho de ser proceso de Markov.

A continuación calculamos las probabilidades de aceptación, es decir, la probabilidad de que la trayectoria y la trayectoria reversible se crucen en el intervalo $[0, T]$ para formar el puente en 10000 simulaciones. Evaluamos los cinco conjuntos de parámetros que ya hemos estudiado variando el intervalo de tiempo, la distancia entre puntos simulados, Δ y dos distintos valores para $X_0 = a$ y $X_T = b$.

Parámetros	Exacto			Markov		
	$\Delta = 0.01$	$\Delta = 0.001$	$\Delta = 0.0005$	$\Delta = 0.01$	$\Delta = 0.001$	$\Delta = 0.0005$
	$a = 0.5, b = 0.5, T = 0.1$					
$\alpha = 0.7, \beta = 0.7$	0.713	0.790	0.793	0.809	0.945	0.986
$\alpha = 1, \beta = 1$	0.719	0.785	0.802	0.815	0.988	0.999
$\alpha = 2, \beta = 2$	0.725	0.796	0.803	0.854	0.999	1
$\alpha = 3, \beta = 1$	0.744	0.811	0.827	0.876	1	1
$\alpha = 10, \beta = 20$	0.949	0.976	0.977	0.999	1	1
	$a = 0.4, b = 0.6, T = 0.1$					
$\alpha = 0.7, \beta = 0.7$	0.320	0.386	0.407	0.423	0.918	0.981
$\alpha = 1, \beta = 1$	0.331	0.392	0.406	0.592	0.982	0.999
$\alpha = 2, \beta = 2$	0.339	0.410	0.421	0.697	0.999	1
$\alpha = 3, \beta = 1$	0.352	0.435	0.439	0.656	0.999	1
$\alpha = 10, \beta = 20$	0.719	0.797	0.808	0.992	1	1
	$a = 0.5, b = 0.5, T = 0.5$					
$\alpha = 0.7, \beta = 0.7$	0.777	0.813	0.825	0.986	1	1
$\alpha = 1, \beta = 1$	0.802	0.836	0.839	0.999	1	1
$\alpha = 2, \beta = 2$	0.848	0.883	0.889	1	1	1
$\alpha = 3, \beta = 1$	0.894	0.910	0.914	1	1	1
$\alpha = 10, \beta = 20$	0.999	0.999	1	1	1	1
	$a = 0.2, b = 0.9, T = 0.5$					
$\alpha = 0.7, \beta = 0.7$	0.09	0.109	0.111	0.896	1	1
$\alpha = 1, \beta = 1$	0.104	0.134	0.132	0.987	1	1
$\alpha = 2, \beta = 2$	0.193	0.228	0.224	0.999	1	1
$\alpha = 3, \beta = 1$	0.202	0.225	0.237	0.999	1	1
$\alpha = 10, \beta = 20$	0.973	0.988	0.991	1	1	1

Tabla 3.17: Probabilidad de cruce en 10000 iteraciones del algoritmo para la aproximación del puente con el método exacto y la construcción vía cadenas de Markov en el intervalo de tiempo $[0, T]$ para $T = 0.1$ y $T = 0.5$ con distintos parámetros, valores de Δ y diferentes valores iniciales y finales.

En general, podemos observar en la Tabla 3.17 que las probabilidades de aceptación son más altas cuando generamos el puente a partir de la construcción vía cadenas Markov que con el algoritmo exacto, lo cual tiene sentido pues las trayectorias aparentan ser más volátiles. En ambos casos obtenemos probabilidades altas cuando el punto inicial y el punto final son igual a 0.5, especialmente cuando Δ es menor, como ya se había anticipado. Otro factor importante es el intervalo de tiempo, notamos que cuando $T = 0.1$ dichas probabilidades son menores al compararlas con las obtenidas con $T = 0.5$.

Por otro lado, al contrastar las probabilidades con los distintos puntos a y b las probabilidades disminuyen considerablemente mejorando cuando Δ es lo suficientemente pequeño, este comportamiento se observa principalmente en el algoritmo exacto. Finalmente, como las trayectorias son más volátiles cuando los parámetros

son pequeños la probabilidad de aceptación se ve afectada negativamente pues es más complicado que las trayectorias se crucen.

Asimismo, registramos en la Tabla 3.18 el tiempo (en segundos) necesario para simular 10000 puentes, los cuales son consistentes con las probabilidades de aceptación, mientras menor sean dichas probabilidades se requiere una mayor cantidad de trayectorias simuladas y por consiguiente mayor tiempo. Como habíamos visto antes que los tiempos para simular trayectorias son comparables entre ambos métodos los resultados sugieren que las diferencias grandes se deben al porcentaje de veces que las trayectorias se cruzan.

Parámetros	Exacto			Markov		
	$\Delta = 0.01$	$\Delta = 0.001$	$\Delta = 0.0005$	$\Delta = 0.01$	$\Delta = 0.001$	$\Delta = 0.0005$
	$a = 0.5, b = 0.5, T = 0.1$					
$\alpha = 0.7, \beta = 0.7$	15.50	88.91	176.05	12.55	75.86	139.87
$\alpha = 1, \beta = 1$	13.83	89.30	169.08	12.38	69.51	137.11
$\alpha = 2, \beta = 2$	14.83	88.20	170.88	12.28	70.95	132.83
$\alpha = 3, \beta = 1$	13.44	85.94	167.25	11.99	73.28	137.33
$\alpha = 10, \beta = 20$	10.60	72.06	138.69	10.31	69.02	134.09
	$a = 0.4, b = 0.6, T = 0.1$					
$\alpha = 0.7, \beta = 0.7$	30.55	176.61	339.37	22.91	75.49	136.62
$\alpha = 1, \beta = 1$	32.64	174.89	335.09	17.00	69.82	133.98
$\alpha = 2, \beta = 2$	29.45	167.28	327.58	14.28	68.74	133.70
$\alpha = 3, \beta = 1$	27.52	161.74	308.37	15.25	69.13	133.11
$\alpha = 10, \beta = 20$	13.97	88.03	166.11	10.20	68.77	132.63
	$a = 0.5, b = 0.5, T = 0.5$					
$\alpha = 0.7, \beta = 0.7$	55.00	416.33	808.36	38.63	322.42	634.79
$\alpha = 1, \beta = 1$	45.92	400.91	819.43	38.69	321.97	634.56
$\alpha = 2, \beta = 2$	42.94	377.46	755.98	36.70	323.86	638.20
$\alpha = 3, \beta = 1$	41.74	365.11	720.19	36.64	321.85	638.22
$\alpha = 10, \beta = 20$	36.93	330.69	662.13	37.87	320.27	627.51
	$a = 0.2, b = 0.9, T = 0.5$					
$\alpha = 0.7, \beta = 0.7$	445.67	3169.76	5872.00	42.01	345.71	674.34
$\alpha = 1, \beta = 1$	336.00	2961.03	5488.22	38.64	329.97	654.46
$\alpha = 2, \beta = 2$	194.11	1487.12	2899.24	38.19	331.02	713.75
$\alpha = 3, \beta = 1$	183.34	1404.58	2758.45	43.38	344.60	680.62
$\alpha = 10, \beta = 20$	38.16	15784.15	669.41	38.22	358.03	660.41

Tabla 3.18: Tiempo de ejecución en segundos para generar 10000 aproximaciones efectivas del puente con el método exacto y la construcción vía cadenas de Markov en el intervalo de tiempo $[0, T]$ para $T = 0.1$ y $T = 0.5$ con distintos parámetros, valores de Δ y diferentes valores iniciales y finales.

Ahora, presentamos las probabilidades de cruce haciendo las distintas combinaciones entre parámetros y dejando fijo el resto de los valores, tomamos $a = 0.5, b = 0.5$, el intervalo de tiempo de 0 a $T = 0.1$ y $T = 0.5$ con $\Delta = 0.001$ con el fin de determinar de manera clara el efecto en las probabilidades de cruce al cambiar los parámetros.

En la Tabla 3.19 otra vez observamos que las probabilidades de cruce son menores en el algoritmo exacto y que mientras mayor sea el periodo de tiempo mayores serán las probabilidades de aceptación. En este ejemplo dejamos α fijo e incrementamos el valor de β y comprobamos que mientras uno de los parámetros se incremente las probabilidades de cruce también lo harán.

Parámetros	Exacto		Markov	
	$T = 0.1$	$T = 0.5$	$T = 0.1$	$T = 0.5$
$\alpha = 0.7, \beta = 0.7$	0.391	0.785	0.917	0.945
$\alpha = 0.7, \beta = 1$	0.400	0.793	0.969	0.980
$\alpha = 0.7, \beta = 2$	0.410	0.800	0.998	0.999
$\alpha = 0.7, \beta = 20$	0.973	0.998	1	1
$\alpha = 1, \beta = 0.7$	0.393	0.784	0.955	0.974
$\alpha = 1, \beta = 1$	0.398	0.790	0.982	0.988
$\alpha = 1, \beta = 2$	0.407	0.802	0.999	0.999
$\alpha = 1, \beta = 20$	0.971	0.998	1	1
$\alpha = 2, \beta = 0.7$	0.410	0.800	0.996	0.998
$\alpha = 2, \beta = 1$	0.409	0.804	0.996	0.999
$\alpha = 2, \beta = 2$	0.412	0.792	0.999	0.999
$\alpha = 2, \beta = 20$	0.962	0.997	1	1
$\alpha = 3, \beta = 0.7$	0.429	0.818	0.999	1
$\alpha = 3, \beta = 1$	0.446	0.812	0.999	1
$\alpha = 3, \beta = 2$	0.428	0.807	1	1
$\alpha = 3, \beta = 20$	0.945	0.997	1	1
$\alpha = 10, \beta = 0.7$	0.723	0.951	1	1
$\alpha = 10, \beta = 1$	0.707	0.950	1	1
$\alpha = 10, \beta = 2$	0.681	0.939	1	1
$\alpha = 10, \beta = 20$	0.798	0.973	1	1

Tabla 3.19: Probabilidades de cruce del puente de Wright-Fisher con el método exacto y la construcción vía cadenas de Markov en el intervalo de tiempo $[0, T]$ para $T = 0.1$ y $T = 0.5$ con distintos parámetros, $\Delta = 0.001$ y valor inicial y final de 0.5.

Capítulo 4

Estimación

La inferencia estadística de los procesos de difusión es un tema en constante desarrollo debido a la gran cantidad de aplicaciones y de complejidad de los modelos. Además, independientemente del fenómeno que describa el proceso de difusión, solo se observa información a tiempo discreto. Existen pocos modelos donde las estimaciones son sencillas de calcular ya que se conoce de forma explícita la solución a la ecuación diferencial estocástica.

En el presente capítulo desarrollamos el algoritmo de Esperanza-Maximización, EM, para encontrar estimadores máximo verosímil de difusiones observadas a tiempo discreto aplicado a la difusión de Wright-Fisher. Dicho algoritmo fue presentado por Bladt y Sørensen [2] y se basa principalmente, en la simulación de puentes de difusión.

4.1. Algoritmo Esperanza-Maximización

El algoritmo EM originalmente fue presentado en [5] para encontrar los estimadores máximo verosímil donde las observaciones pueden considerarse como información incompleta. Este algoritmo recibe su nombre pues cada iteración consta de dos pasos, el paso de esperanza (Paso E) y el paso de maximización (Paso M). En el paso E se estiman los datos faltantes mientras que en el Paso M se encuentran los nuevos parámetros como solución al problema de maximización y se repite hasta alcanzar la convergencia.

Puesto que conocemos una trayectoria de la difusión solo a través de los tiempos t_1, t_2, \dots, t_n , es natural ver al problema de estimación como un problema de datos incompletos, por lo que, es posible utilizar el algoritmo EM para hacer inferencia sobre sus parámetros.

Supongamos que tenemos una ecuación diferencial estocástica de la forma,

$$dX_t = \mu_\eta(X_t)dt + \sigma_\theta dB_t, \quad (4.1)$$

donde η y θ son los parámetros desconocidos que deseamos estimar y B_t es un movimiento browniano estándar.

En primer lugar, utilizamos la transformación de Lamperti descrita en la Sección 1.2.2 con la finalidad de obtener un nuevo proceso con coeficiente de difusión unitario. Esto es, consideremos la función

$$h_\theta(x) = \int_{x^*}^x \frac{1}{\sigma(y)} dy$$

donde x^* es un punto arbitrario en el espacio de estados. Como ya hemos mencionado el nuevo proceso $Y_t = h_\theta(X_t)$ tiene coeficiente de difusión unitario y coeficiente de deriva $\tilde{\mu}_{\eta,\theta}(y)$, dada por:

$$\tilde{\mu}_{\eta,\theta}(y) = \frac{\mu_\eta(h_\theta^{-1}(y))}{\sigma_\theta(h_\theta^{-1}(y))} - \frac{1}{2}\sigma'_\theta(h_\theta^{-1}(y)).$$

Ya que el coeficiente de difusión no depende de ningún parámetro puede encontrarse la verosimilitud, para ello se define la función

$$g_{\eta,\theta}(x) = s_{\eta,\theta}(x) - \frac{1}{2} \ln(\sigma_\theta(x)), \quad \text{donde} \quad s_{\eta,\theta}(x) = \int_{x^*}^x \frac{\mu_\eta(z)}{\sigma_\theta^2(z)} dz$$

y x^* un punto arbitrario. La función $s_{\eta,\theta}(x)$ está relacionada con la distribución estacionaria del proceso, de manera que $s_{\eta,\theta}(x)$ es igual a $\frac{1}{2} \ln(\sigma_\theta^2(x)\pi(x))$ salvo una constante aditiva. Además se cumple que $\int_{y^*}^y \tilde{\mu}_{\eta,\theta}(z) dz = g_{\eta,\theta}(h_\theta^{-1}(y)) - g_{\eta,\theta}(h_\theta^{-1}(y^*))$.

Para continuar con el algoritmo EM propuesto en [2] y hacer la estimación, definamos

$$Y_t^*(\theta, \theta_0) = Z_t^{(i,\eta_0,\theta_0)} + \frac{(t_i - t)(h_\theta(x_{i-1}) - h_{\theta_0}(x_{i-1})) + (t - t_{i-1})(h_\theta(x_i) - h_{\theta_0}(x_i))}{t_i - t_{i-1}}, \quad (4.2)$$

donde $Z_t^{(i,\eta_0,\theta_0)}$ denota el puente $(t_{i-1}, h(x_{i-1}), t_i, h(x_i))$ con parámetros θ_0 y además $Z_t^{(i,\eta,\theta)}$ con $i = 2, \dots, n$ son independientes. El segundo sumando se agrega para mantener fijos los datos originales al aplicar el algoritmo después de la transformación, debido a la dependencia de θ .

El algoritmo EM dado por [2] para la difusión (4.1) es el siguiente.

Algoritmo 7 EM.

1: **Paso E.** Calcular

$$q(\eta, \theta) = g_{\eta,\theta}(x_n) - g_{\eta,\theta}(x_1) - \frac{1}{2} \sum_{i=2}^n \frac{(h_\theta(x_i) - h_\theta(x_{i-1}))^2}{t_i - t_{i-1}} - \sum_{i=2}^n \ln(\sigma_\theta(x_i)) - \frac{1}{2} \sum_{i=2}^n \mathbb{E}_{Z^{(i,\eta_0,\theta_0)}} \left[\int_{t_{i-1}}^{t_i} \left(\tilde{\mu}'_{\eta,\theta}(Y_t^*) + \tilde{\mu}_{\eta,\theta}(Y_t^*)^2 \right) dt \right]. \quad (4.3)$$

2: **Paso M.** $(\eta_0, \theta_0) = \operatorname{argmax}_{\eta,\theta} q(\eta, \theta)$.

3: Regresar al Paso E.

Notemos que en la esperanza del paso E la parte aleatoria está dada solamente por los puentes y además, se calcula con respecto a la distribución de los mismos, por lo que, puede ser aproximada utilizando Monte Carlo, es decir, simulando puentes y promediándolos.

No obstante, si el coeficiente de deriva tiene forma lineal, es decir, $\mu(x) = \eta_1 a_1(x) + \eta_2 a_2(x) \cdots + \eta_n a_n(x)$, donde $a_1(x), \dots, a_n(x)$ son funciones conocidas y el coeficiente de difusión fijo, digamos β , la optimización se reduce considerablemente ya que la función (4.3) del paso E puede reescribirse de la siguiente forma,

$$q(\eta, \theta) = \sum_{i=1}^n \eta_i H_{i,\theta} - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \eta_i \eta_j B_{i,j,\theta} + G_\theta$$

donde,

$$H_{i,\theta} = s_{i,\theta}(x_n) - s_{i,\theta}(x_1) + \sum_{j=2}^n \mathbb{E}_{Z^{(k,\eta_0,\theta_0)}} \left[\int_{t_{j-1}}^{t_j} \left(a_i(h_\theta^{-1}(Y_t^*(\theta, \theta_0))) \left(\ln(\sigma_\theta) \right)' (h_\theta^{-1}(Y_t^*(\theta, \theta_0))) - \frac{1}{2} a_i'(h_\theta^{-1}(Y_t^*(\theta, \theta_0))) \right) dt \right],$$

con $s_{i,\theta}(x) = \int_{x^*}^x \frac{a_i(y)}{\sigma_\theta^2(y)} dy$,

$$B_{i,j,\theta} = \sum_{k=2}^n \mathbb{E}_{Z^{(k,\eta_0,\theta_0)}} \left[\int_{t_{k-1}}^{t_k} \frac{a_i(h_\theta^{-1}(Y_t^*(\theta, \theta_0))) a_j(h_\theta^{-1}(Y_t^*(\theta, \theta_0)))}{\sigma_\theta^2(h_\theta^{-1}(Y_t^*(\theta, \theta_0)))} dt \right]$$

y

$$G_\theta = -\frac{1}{2} \ln \left(\frac{\sigma_\theta(x_n)}{\sigma_\theta(x_1)} \right) - \frac{1}{2} \sum_{i=2}^n \frac{(h_\theta(x_i) - h_\theta(x_{i-1}))^2}{t_i - t_{i-1}} - \sum_{i=2}^n \ln(\sigma_\theta(x_i)) \\ + \frac{1}{4} \sum_{i=2}^n \mathbb{E}_{Z^{(i, \eta_0, \theta_0)}} \left[\int_{t_{i-1}}^{t_i} \left(\sigma_\theta''(h_\theta^{-1}(Y_t^*(\theta, \theta_0))) \sigma_\theta(h_\theta^{-1}(Y_t^*(\theta, \theta_0))) - \frac{1}{2} \left(\sigma_\theta'(h_\theta^{-1}(Y_t^*)) \right)^2 \right) dt \right].$$

Para un valor fijo de θ , la función $\eta \mapsto q(\eta, \theta)$ se maximiza cuando,

$$\hat{\eta}(\theta) = \mathbf{B}_\theta^{-1} \mathbf{H}_\theta, \quad (4.4)$$

donde $\hat{\eta} = (\eta_1, \dots, \eta_k)^T$, $\mathbf{H}_\theta = (H_{1,\theta}, \dots, H_{k,\theta})^T$ y $\mathbf{B}_\theta = \{B_{i,j,\theta}\}$. Teniendo en cuenta que \mathbf{B}_θ es invertible cuando las funciones a_i , $i = 1, \dots, k$ son linealmente independientes.

Entonces, $q(\eta, \theta)$ alcanza su máximo valor en $(\hat{\eta}(\hat{\theta}), \hat{\theta})$, donde $\hat{\theta}$ se maximiza,

$$\theta \mapsto q(\hat{\eta}(\theta), \theta) = \frac{1}{2} \mathbf{H}_\theta^T \mathbf{B}_\theta^{-1} \mathbf{H}_\theta + \mathbf{G}_\theta.$$

4.2. Aplicación del algoritmo EM al proceso de Wright-Fisher

En esta sección vamos a implementar el Algoritmo 7 a la ecuación diferencial estocástica de Wright-Fisher (3.1). Tenemos que $\mu_\eta(x) = \frac{1}{2}(\alpha(1-x) - \beta x)$ y $\sigma_\theta(x) = \sqrt{x(1-x)}$, por lo tanto, el parámetro a estimar es $\eta = (\alpha, \beta)$. Como en la ecuación de Wright-Fisher el coeficiente de difusión no depende de algún parámetro, por simplicidad omitiremos los subíndices.

En primer lugar, hacemos los cálculos de la transformación de Lamperti,

$$\int \frac{1}{\sqrt{y(1-y)}} dy = \int \frac{1}{\sqrt{\frac{1}{4} - (y - \frac{1}{2})^2}} dy \\ = \int \frac{1}{\sqrt{1 - (2y - 1)^2}} dy \\ = \text{sen}^{-1}(2y - 1).$$

Cabe mencionar que si utilizamos el cambio de variable $y = \text{sen}^2(u)$ para resolver la integral $h(x)$ en lugar de completar el cuadrado obtenemos que $h(x) = 2 \text{sen}^{-1}(\sqrt{x})$. Sin embargo, ambas expresiones son iguales salvo por la constante $\frac{\pi}{2}$, lo cual verificamos notando que $\cos(\text{sen}^{-1}(2x-1) + \pi/2) = -\text{sen}(\text{sen}^{-1}(2x-1)) = 1-2x$ y por otro lado, $\cos(2 \text{sen}^{-1}(\sqrt{x})) = 1 - 2 \text{sen}^2(\text{sen}^{-1}(\sqrt{x})) = 1 - 2x$.

Por consiguiente, $h(x) = \text{sen}^{-1}(2x-1) - \text{sen}^{-1}(2x^*-1)$ y observamos que si elegimos $x^* = \frac{1}{2}$ la función se simplifica $h(x) = \text{sen}^{-1}(2x-1)$ y además como $x \in (0, 1)$ el rango de la función h es el intervalo $(-\frac{\pi}{2}, \frac{\pi}{2})$.

Luego,

$$h^{-1}(y) = \frac{1}{2} (\text{sen}(y) + 1).$$

Posteriormente, necesitamos encontrar la primer y segunda derivada del coeficiente de difusión,

$$\sigma(x)' = \frac{1-2x}{2\sqrt{x(1-x)}}$$

y

$$\sigma(x)'' = -\frac{1}{4(x(1-x))^{3/2}}.$$

Así el coeficiente de deriva del proceso transformado,

$$\begin{aligned}\tilde{\mu}(y) &= \frac{\frac{1}{2} \left[\alpha \left(\frac{1 - \text{sen}(y)}{2} \right) - \beta \left(\frac{\text{sen}(y) + 1}{2} \right) \right]}{\sqrt{\left(\frac{\text{sen}(y) + 1}{2} \right) \left(\frac{1 - \text{sen}(y)}{2} \right)}} - \frac{1}{2} \left(\frac{1 - 2 \left(\frac{\text{sen}(y) + 1}{2} \right)}{2 \sqrt{\left(\frac{\text{sen}(y) + 1}{2} \right) \left(\frac{1 - \text{sen}(y)}{2} \right)}} \right) \\ &= \frac{1}{2} \left[\alpha \left(\frac{1 - \text{sen}(y)}{\cos(y)} \right) - \beta \left(\frac{\text{sen}(y) + 1}{\cos(y)} \right) + \frac{\text{sen}(y)}{\cos(y)} \right].\end{aligned}$$

Por lo tanto, la nueva ecuación diferencial estocástica está dada por:

$$dY_t = \frac{1}{2} \left[\alpha \left(\frac{1 - \text{sen}(y)}{\cos(y)} \right) - \beta \left(\frac{\text{sen}(y) + 1}{\cos(y)} \right) + \frac{\text{sen}(y)}{\cos(y)} \right] dt + dB_t. \quad (4.5)$$

Donde el espacio de estados del nuevo proceso, $\{Y_t\}$, asociado a la ecuación diferencial estocástica (4.5) es el intervalo $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Por su parte,

$$\begin{aligned}s(x) &= \int_{\frac{1}{2}}^x \frac{\frac{1}{2}(\alpha(1-z) - \beta z)}{z(1-x)} dz \\ &= \frac{1}{2} \left(\alpha \ln(x) + \beta \ln(1-x) - (\alpha + \beta) \ln\left(\frac{1}{2}\right) \right) \\ &= \frac{1}{2} \ln\left(x^\alpha(1-x)^\beta\right) + \frac{\alpha + \beta}{2} \ln(2).\end{aligned}$$

Como la distribución estacionaria es una $Beta(\alpha, \beta)$, es decir,

$$\pi(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

es fácil verificar que $s(x)$ es igual a $\frac{1}{2} \ln(\sigma^2(x)\pi(x))$ salvo una constante aditiva.

Entonces,

$$\begin{aligned}g(x) &= \frac{1}{2} \ln\left(x^\alpha(1-x)^\beta\right) + \frac{\alpha + \beta}{2} \ln(2) - \frac{1}{2} \ln(\sqrt{x(1-x)}) \\ &= \frac{1}{2} \left[\ln\left(x^\alpha(1-x)^\beta\right) - \ln(\sqrt{x(1-x)}) \right] + \frac{\alpha + \beta}{2} \ln(2) \\ &= \frac{1}{2} \left[\ln\left(x^{\alpha-\frac{1}{2}}(1-x)^{\beta-\frac{1}{2}}\right) \right] + \frac{\alpha + \beta}{2} \ln(2) \\ &= \frac{1}{2} \left[\left(\alpha - \frac{1}{2}\right) \ln(x) + \left(\beta - \frac{1}{2}\right) \ln(1-x) \right] + \frac{\alpha + \beta}{2} \ln(2).\end{aligned}$$

Además, para comprobar que $\int_{y^*}^y \tilde{\mu}(z) dz = g(h^{-1}(y)) - g(h^{-1}(y^*))$ basta revisar que la derivada de $g(h^{-1}(y))$ respecto a y es igual a $\tilde{\mu}(y)$. Notando que $h^{-1}(y)' = \sigma(h^{-1}(y))$,

$$\begin{aligned}\sigma(h^{-1}(y)) &= \sqrt{h^{-1}(y)(1-h^{-1}(y))} \\ &= \frac{1}{2} \sqrt{(\text{sen}(y) + 1)(1 - \text{sen}(y))} \\ &= \frac{1}{2} \cos(y).\end{aligned} \quad (4.6)$$

Tenemos entonces,

$$\begin{aligned} g(h^{-1}(y)) &= \frac{1}{2} \left[\left(\alpha - \frac{1}{2} \right) \ln \left(\frac{\text{sen}(y) + 1}{2} \right) + \left(\beta - \frac{1}{2} \right) \ln \left(\frac{1 - \text{sen}(y)}{2} \right) \right] + \frac{\alpha + \beta}{2} \ln(2) \\ &= \frac{1}{2} \left[\left(\alpha - \frac{1}{2} \right) \ln(\text{sen}(y) + 1) + \left(\beta - \frac{1}{2} \right) \ln(1 - \text{sen}(y)) \right] + \frac{1}{2} \ln(2), \end{aligned}$$

usando la regla de la cadena para derivar $g(h^{-1}(y))$ respecto a y

$$\begin{aligned} \frac{dg(h^{-1}(y))}{dy} &= \frac{1}{2} \left[\left(\alpha - \frac{1}{2} \right) \frac{\cos(y)}{\text{sen}(y) + 1} - \left(\beta - \frac{1}{2} \right) \frac{\cos(y)}{1 - \text{sen}(y)} \right] \\ &= \frac{1}{2} \left(\frac{\cos(y)}{\cos(y)} \right) \left[\left(\alpha - \frac{1}{2} \right) \frac{\cos(y)}{\text{sen}(y) + 1} - \left(\beta - \frac{1}{2} \right) \frac{\cos(y)}{1 - \text{sen}(y)} \right] \\ &= \frac{1}{2} \left[\left(\alpha - \frac{1}{2} \right) \frac{\cos^2(y)}{(\text{sen}(y) + 1) \cos(y)} - \left(\beta - \frac{1}{2} \right) \frac{\cos^2(y)}{(1 - \text{sen}(y)) \cos(y)} \right] \\ &= \frac{1}{2} \left[\left(\alpha - \frac{1}{2} \right) \left(\frac{1 - \text{sen}(y)}{\cos(y)} \right) - \left(\beta - \frac{1}{2} \right) \left(\frac{\text{sen}(y) + 1}{\cos(y)} \right) \right] \\ &= \frac{1}{2} \left[\alpha \left(\frac{1 - \text{sen}(y)}{\cos(y)} \right) - \beta \left(\frac{\text{sen}(y) + 1}{\cos(y)} \right) + \frac{\text{sen}(y)}{\cos(y)} \right], \end{aligned}$$

que efectivamente es la deriva $\tilde{\mu}$ del proceso transformado.

Ahora, como el coeficiente de difusión de (3.1) solo depende del proceso, los términos de la forma $h_\theta(x) - h_{\theta_0}(x)$ se cancelan, de manera que la ecuación (4.2) se reduce a

$$Y_t^* = Y_t^*(\theta, \theta_0) = Z_t^{(i, \eta_0, \theta_0)},$$

donde $Z_t^{(i, \eta_0, \theta_0)}$ representa el puente del proceso transformado $\{Y_t\}$ entre los tiempos t_{i-1} y t_i , con estado inicial $h(x_{i-1})$ y estado final $h(x_{i-1})$.

Como el coeficiente de deriva del proceso de Wright-Fisher tiene forma lineal, consideramos $\eta_1 = \alpha$ y $\eta_2 = \beta$ y $a_1(x) = \frac{1}{2}(1-x)$, $a_2 = -\frac{1}{2}(x)$. De tal manera que la maximización en el Algoritmo 7 se simplifica utilizando la ecuación (4.4). Obtenemos que $s_1 = \frac{1}{2} \int_{x^*}^x \frac{1}{y} dy = \frac{1}{2} (\ln(x) - \ln(x^*))$ y $s_2 = -\frac{1}{2} \int_{x^*}^x \frac{1}{1-y} dy = \frac{1}{2} (\ln(1-x) - \ln(1-x^*))$.

La verosimilitud del paso E

$$q(\eta) = \alpha H_1 + \beta H_2 - \frac{1}{2} \left[\alpha^2 B_{1,1} + \alpha \beta (B_{1,2} + B_{2,1}) + \beta^2 B_{2,2} \right] + G, \quad (4.7)$$

donde,

$$H_1 = \frac{1}{2} (\ln(x_n) - \ln(x_1)) + \sum_{j=2}^n \mathbb{E}_{Z^{(j, \eta_0)}} \left[\int_{t_{j-1}}^{t_j} -\frac{\text{sen}(Y_t^*)}{2(1 + \text{sen}(Y_t^*))} + \frac{1}{4} dt \right]$$

y

$$H_2 = \ln(1-x_n) - \ln(1-x_1) + \sum_{j=2}^n \mathbb{E}_{Z^{(j, \eta_0)}} \left[\int_{t_{j-1}}^{t_j} \frac{\text{sen}(Y_t^*)}{2(1 - \text{sen}(Y_t^*))} + \frac{1}{4} dt \right]$$

$$B_{1,2} = -\frac{1}{4}(t_n - t_1)$$

$$\begin{aligned}
B_{1,1} &= -\frac{1}{4} \sum_{j=2}^n \mathbb{E}_{Z^{(j,\eta_0)}} \left[\int_{t_{j-1}}^{t_j} \frac{1 - \text{sen}(Y_t^*)}{1 + \text{sen}(Y_t^*)} dt \right] \\
B_{2,2} &= \frac{1}{4} \sum_{j=2}^n \mathbb{E}_{Z^{(j,\eta_0)}} \left[\int_{t_{j-1}}^{t_j} \frac{1 + \text{sen}(Y_t^*)}{1 - \text{sen}(Y_t^*)} dt \right] \\
G &= -\frac{1}{2} \ln \left(\frac{\sqrt{x_n(1-x_n)}}{\sqrt{x_1(1-x_1)}} \right) - \frac{1}{2} \sum_{i=1}^n \left[\frac{\text{sen}^{-1}(2x_i - 1) - \text{sen}^{-1}(2x_{i-1} - 1)}{t_i - t_{i-1}} \right] - \sum_{i=1}^n \ln \left(\sqrt{x_i(1-x_i)} \right) \\
&\quad - \frac{1}{4} \sum_{i=1}^n \mathbb{E}_{Z^{(j,\eta_0)}} \left[\int_{t_{j-1}}^{t_j} \frac{2 + \text{sen}^2(Y_t^*)}{\cos^2(Y_t^*)} dt \right].
\end{aligned}$$

Notamos que la verosimilitud de la ecuación (4.7) puede escribirse matricialmente como

$$q(\eta) = \eta \mathbf{H} - \frac{1}{2} (\eta \mathbf{B} \eta) + G, \text{ con } \mathbf{B} = \{B_{i,j}\}_{i,j=1,2} \text{ y } \mathbf{H} = (H_1, H_2).$$

Además, no es necesario conocer G pues no tenemos dependencia del parámetro θ . Entonces, la función $\eta = (\alpha, \beta) \rightarrow q(\eta)$ se maximiza en $\hat{\eta} = (\hat{\alpha}, \hat{\beta}) = \mathbf{B}^{-1} \mathbf{H}$.

Resumiendo, el Algoritmo 7 implementado a la difusión de Wright-Fisher se describe a continuación.

Algoritmo 8 Wright-Fisher EM

1: **Paso E.** Calcular

$$q(\eta) = \alpha H_1 + \beta H_2 - \frac{1}{2} \left[\alpha^2 B_{1,1} + \alpha \beta (B_{1,2} + B_{2,1}) + \beta^2 B_{2,2} \right] + G, \quad (4.8)$$

2: **Paso M.** Encontrar $\hat{\eta} = \mathbf{B}^{-1} \mathbf{H}$.

3: Regresar al Paso E.

4.3. Implementación del algoritmo EM

En la sección anterior presentamos el algoritmo EM para hacer estimaciones de los parámetros de la difusión de Wright-Fisher, lo siguiente es probar el algoritmo. Ya que no tenemos datos reales, lo haremos a través de simulaciones. La idea es generar una trayectoria y eliminar algunas observaciones con la finalidad de obtener un problema de datos incompletos.

Al repetir muchas veces el algoritmo EM es posible obtener una muestra de estimadores y con ella intervalos de confianza empíricos. En particular, vamos a generar una muestra de tamaño 1000, a partir de la cual tomamos al estimador como el promedio y un intervalo formado con los cuantiles al 2.5 % y 97.5 %.

En primer lugar, elegimos los parámetros ($\alpha = 2$, $\beta = 2$), simulamos una trayectoria en el intervalo de tiempo $[0,100]$ con 1000 observaciones, es decir, se tienen observaciones en intervalos de tiempo $\Delta = 0.1$. Se utiliza el paso E y el paso M (sin generar puentes) para hacer estimación sobre la trayectoria completa. En la Tabla 4.1 mostramos los resultados para distintos tamaños de muestra.

Posteriormente, se elimina el 90 % de las observaciones y volvemos a hacer estimación con la trayectoria incompleta, es decir, con los 100 puntos restantes ubicados en cada unidad de tiempo. Nuevamente aplicamos el paso E y el paso M sin generar puentes y obtenemos una muestra de estimadores. En la Tabla 4.2 se presentan los resultados, como era de esperarse, las estimaciones no son tan buenas como las anteriores y la amplitud del intervalo es mayor.

Finalmente, utilizamos el Algoritmo 8 y la construcción de puentes de difusión para hacer estimaciones observando de manera parcial la trayectoria, esto es, observando solo el 10 % y completando con puentes simulados. Cabe mencionar que para generar la trayectoria original, utilizamos el algoritmo exacto, Algoritmo

N	α			β		
	Promedio	Inferior	Superior	Promedio	Inferior	Superior
50	2.027	1.705	2.333	2.062	1.761	2.444
100	2.026	1.692	2.297	2.05	1.729	2.441
200	2.048	1.694	2.386	2.046	1.682	2.452
300	2.048	1.709	2.380	2.047	1.723	2.451
400	2.052	1.718	2.4	2.052	1.732	2.429
500	2.048	1.709	2.4	2.052	1.730	2.403
600	2.043	1.707	2.401	2.052	1.729	2.429
700	2.044	1.709	2.404	2.053	1.728	2.427
800	2.042	1.710	2.390	2.049	1.727	2.425
900	2.044	1.714	2.404	2.047	1.728	2.443
1000	2.045	1.710	2.421	2.051	1.729	2.444

Tabla 4.1: Estimaciones e intervalos de confianza para $(\alpha = 2, \beta = 2)$ con la trayectoria completa (1000 observaciones) de acuerdo con el número de implementaciones del algoritmo EM.

N	α			β		
	Promedio	Inferior	Superior	Promedio	Inferior	Superior
50	2.112	1.552	2.767	2.159	1.674	2.662
100	2.091	1.540	2.740	2.113	1.599	2.67
200	2.148	1.605	2.872	2.129	1.569	2.773
300	2.131	1.568	2.814	2.113	1.558	2.795
400	2.126	1.588	2.790	2.122	1.569	2.795
500	2.123	1.551	2.793	2.123	1.563	2.81
600	2.117	1.557	2.789	2.119	1.55	2.791
700	2.127	1.593	2.789	2.129	1.553	2.791
800	2.123	1.580	2.78	2.125	1.55	2.79
900	2.128	1.584	2.785	2.126	1.547	2.791
1000	2.129	1.557	2.790	2.133	1.557	2.795

Tabla 4.2: Estimaciones e intervalos de confianza para $(\alpha = 2, \beta = 2)$ con la trayectoria incompleta (100 observaciones) de acuerdo con el número de implementaciones del algoritmo EM.

5, y para la simulación de los puentes la aproximación presentada en la Sección 3.6.1 utilizando trayectorias construidas a partir de la propiedad de Markov, Algoritmo 6, debido a la rapidez de dicho método.

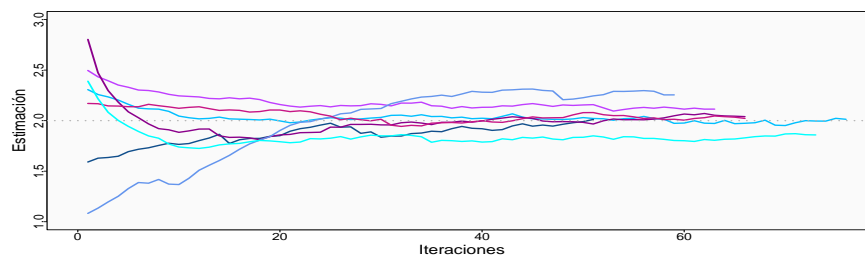
Los parámetros iniciales para cada repetición del EM fueron números aleatorios entre 1 y 3. En cada iteración del EM, las integrales para el cálculo de la verosimilitud del paso E en la ecuación (4.7) se aproximan mediante el método del trapecio generando puentes con 9 observaciones entre cada par de puntos, de tal manera que recuperamos trayectorias del tamaño original. Por su parte, las esperanzas se calculan mediante Monte Carlo con tamaño de muestra 100. En la Tabla 4.3 mostramos los valores de α y β .

Cabe mencionar que al ejecutar el algoritmo EM consideramos dos criterios de paro, el primero que el valor absoluto de la diferencia de la verosimilitud en la iteración i respecto a la iteración $i + 1$ sea menor a $\varepsilon = 0.005$ o bien, que se realizaran a lo sumo 10000 iteraciones, en todos los casos paró bajo el primer criterio.

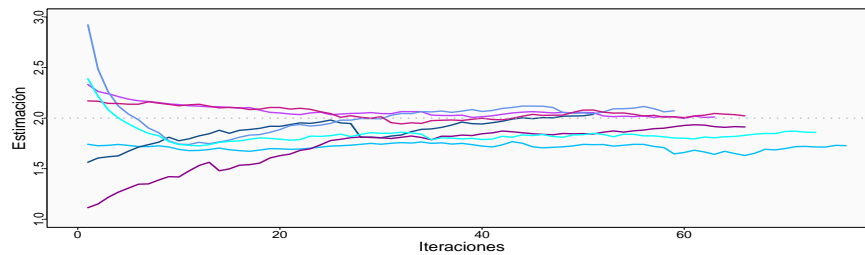
En la Figura 4.1 mostramos algunos ejemplos de la evolución de los parámetros en cada iteración del algoritmo EM hasta la convergencia bajo el primer criterio. Se puede notar que aproximadamente en la iteración 25, los valores ya se encuentran fluctuando cerca del valor real.

N	α			β		
	Promedio	Inferior	Superior	Promedio	Inferior	Superior
50	2.110	1.7	2.731	2.126	1.728	2.981
100	2.085	1.689	2.733	2.101	1.668	2.816
200	2.073	1.655	2.707	2.097	1.671	2.768
300	2.077	1.577	2.7	2.085	1.643	2.739
400	2.087	1.566	2.760	2.088	1.636	2.727
500	2.080	1.563	2.743	2.082	1.625	2.733
600	2.072	1.566	2.741	2.076	1.624	2.718
700	2.073	1.570	2.718	2.077	1.605	2.711
800	2.074	1.574	2.707	2.079	1.607	2.718
900	2.071	1.566	2.692	2.076	1.613	2.712
1000	2.076	1.574	2.728	2.079	1.618	2.712

Tabla 4.3: Estimaciones e intervalos de confianza para ($\alpha = 2 \beta = 2$) con el algoritmo EM y puentes de difusión de acuerdo con el número de implementaciones del algoritmo EM.



(a) Estimación de α



(b) Estimación de β

Figura 4.1: Gráfica de la evolución de los parámetros en cada iteración en 7 implementaciones del algoritmo EM simulando puentes de difusión. El valor inicial para cada implementación es un número aleatorio en el intervalo (1,3).

Además, ya que obtuvimos una muestra de 1000 estimadores, encontramos que el promedio de iteraciones para la convergencia es de aproximadamente 60, el mínimo de 3 y el máximo de 2173. La Tabla 4.4 contiene el promedio de iteraciones de acuerdo al tamaño de muestra de las estimaciones.

Ahora, con el objetivo de que sea más sencillo contrastar las muestras obtenidas en cada ejercicio: datos completos (1000 observaciones), datos incompletos (100 observaciones) y completando la muestra con puentes de difusión (100 observaciones más 900 observaciones simuladas), presentamos la función de densidad ajustada en cada uno de ellos. En la Figura 4.2 notamos que las tres curvas están centrada cerca del valor real y notamos que la muestra de la trayectoria completa está más concentrada alrededor 2 con una desviación estándar de 0.17 tanto para α como para β , le sigue la muestra de estimadores que utilizan los puentes con una desviación estándar de 0.28 y 0.27 respectivamente. Finalmente, con una desviación estándar de 0.31 para los dos parámetros, la gráfica obtenida con la trayectoria incompleta es la que tiene mayor varianza.

N	Promedio iteraciones
50	48.73
100	57.78
200	59.32
300	57.38
400	59.26
500	59.78
600	58.60
700	58.93
800	60.83
900	62.69
1000	62.13

Tabla 4.4: Promedio de iteraciones para la convergencia en el algoritmo EM para $(\alpha = 2, \beta = 2)$ de acuerdo con el número de implementaciones del algoritmo.

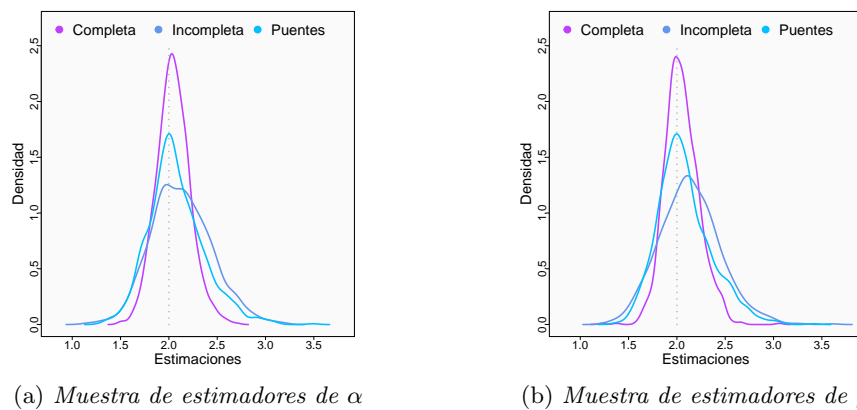
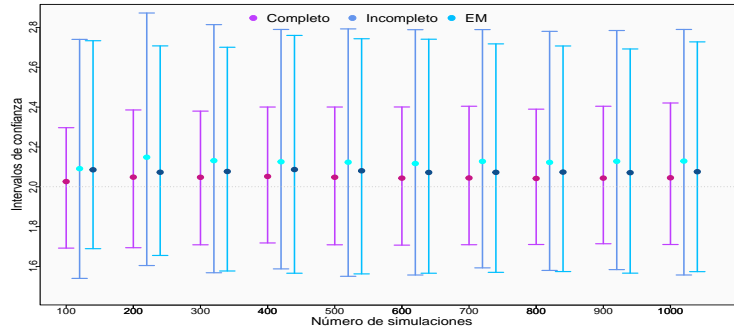


Figura 4.2: Función de densidad ajustada a la muestra de estimadores obtenida en 1000 implementaciones del algoritmo EM con la muestra completa, la muestra incompleta y utilizando puentes de difusión.

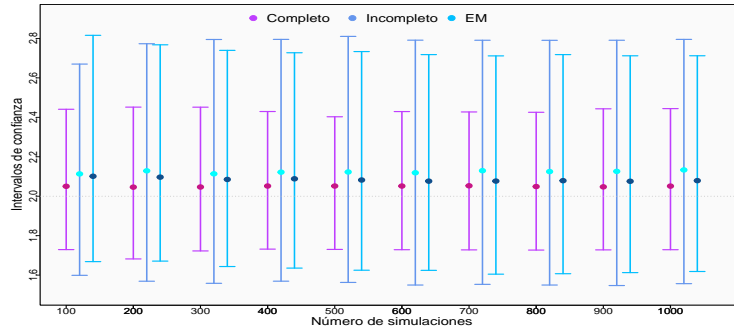
En conclusión, podemos percibir que no es necesario ejecutar tantas veces el algoritmo EM, pues desde los primeros 100 se tienen resultados consistentes y no se observan mejoras considerables mientras se incrementa el número de veces que se replica el algoritmo. Ahora, respecto al promedio de las estimaciones están muy cercanas al valor 2, el valor más alto se encuentra en la estimación con información incompleta siendo de aproximadamente 2.1. Por otro lado, como era de esperarse debido a la variabilidad que se tiene con cada estimación, los intervalos de confianza son más amplios con datos incompletos, le sigue las estimaciones con el EM y finalmente con la trayectoria completa. En la Figura 4.3 mostramos gráficamente los intervalos resultantes en cada ejercicio.

También, probamos el algoritmo con $(\alpha = 10, \beta = 20)$, previamente vimos que las trayectorias generadas con este conjunto de parámetros son menos volátiles pero no están centradas alrededor del punto medio. Utilizamos el mismo procedimiento que el ejercicio anterior, es decir, generar trayectorias en el intervalo de tiempo de $[0,100]$ y se hacen estimaciones con el algoritmo EM con la trayectoria completa, con la trayectoria incompleta, esto es, quitando el 90% de las observaciones y posteriormente, completando la trayectoria con puentes.

Con base en los resultados del ejercicio anterior donde sugiere que no es necesario realizar tantas réplicas del algoritmo EM para obtener buenas estimaciones, los resultados se muestran para un tamaño de muestra igual a 100. Cabe mencionar que el valor inicial para α y β en el algoritmo fueron números aleatorios distribuidos uniformemente entre 5 y 25 y el resto de las consideraciones continúan igual.



(a) Intervalos de confianza para α



(b) Intervalos de confianza para β

Figura 4.3: Intervalos de confianza obtenidos con la muestra de estimadores al implementar 1000 veces el algoritmo EM con la trayectoria completa, la trayectoria incompleta y utilizando puentes de difusión para ($\alpha = 2$, $\beta = 2$).

Respecto al número de iteraciones en cada realización del algoritmo EM en promedio se necesitaron 439.25 iteraciones para alcanzar la convergencia de acuerdo a los criterios establecidos, un mínimo de 16 y un máximo de 2789.

Parámetro	Estimación	Información completa	Información incompleta	EM
α	Promedio	10.038	9.912	9.948
	Inferior	9.163	7.469	7.477
	Superior	11.176	12.29	12.039
β	Promedio	20.149	19.925	20.009
	Inferior	18.425	14.988	15.032
	Superior	22.428	25.068	25.302

Tabla 4.5: Estimaciones e intervalos de confianza obtenidos con 100 implementaciones del algoritmo EM con la trayectoria completa, la trayectoria incompleta y utilizando puentes de difusión para ($\alpha = 10$, $\beta = 20$).

Conclusiones

En el presente trabajo desarrollamos la teoría del proceso de Wright-Fisher que modela la dinámica de poblaciones afectada solamente por el efecto de mutación. Se parte de individuos con dos tipos de genes A y a que pueden mutar de uno a otro con tasas α y β , respectivamente, el proceso al tiempo t representa la proporción de individuos de tipo A en la población a tiempo t . Ya que la difusión de Wright-Fisher modela una proporción, el espacio de estados correspondiente es el intervalo $[0,1]$. Cabe resaltar que este tipo de difusiones podrían utilizarse en un contexto más allá de biológico, donde se busque estudiar el comportamiento de la proporción de cierta característica en algún fenómeno a través del tiempo.

A lo largo de este trabajo también realizamos una compilación de algunos métodos de simulación de la difusión de Wright-Fisher permitiéndonos observar algunas semejanzas y diferencias entre ellos. Sin duda con los métodos de aproximación más conocidos para simular difusiones se tiene el problema del espacio de estados acotado, por lo que probamos algunos otros métodos.

La primer evaluación consistió en comparar una muestra de observaciones obtenidas de trayectorias simuladas en distintos tiempos con la distribución estacionaria y posteriormente, efectuamos comparaciones gráficas entre cada uno de los métodos. Obtuvimos que las aproximaciones numéricas de Euler y Milstein son similares entre sí, sin embargo, tienen ciertas deficiencias al contrastar con la distribución estacionaria, particularmente con parámetros pequeños aunque en algunos casos el comportamiento mejora ligeramente al hacer Δ suficientemente pequeño. Por su parte el algoritmo BISS depende tanto de los parámetros como del incremento Δ , lo cual resulta problemático si α y β son pequeños. Dados los resultados obtenidos, consideramos que los métodos que generan trayectorias más confiables son el algoritmo exacto y el algoritmo basado en la construcción vía cadenas de Markov, entre ellos tienen un comportamiento similar. La principal ventaja del segundo algoritmo respecto al primero es la simplicidad y la rapidez para generar una gran cantidad de trayectorias, lo cual lleva a un menor costo computacional.

Finalmente, probamos el algoritmo EM con parámetros ($\alpha = 2$, $\beta = 2$) para hacer estimación sobre las tasas de mutación que resultan ser los únicos parámetros de la difusión de Wright-Fisher. El paso M de maximización se simplifica considerablemente pues la deriva del proceso puede escribirse como una combinación lineal, esto implica que sea más sencillo de implementar.

Replicamos el algoritmo EM con la finalidad de obtener una muestra de estimadores, así como un intervalo de confianza empírico del 95% a partir de los cuantiles del 2.5% y 97.5%, observamos que con una muestra de tamaño 100 los resultados son consistentes a muestras de mayor tamaño. Además, el valor de la estimación obtenido como el promedio de los elementos de la muestra es bastante cercano al valor real del parámetro, de hecho, la densidad ajustada a la muestra de estimadores se encuentra centrada alrededor del valor real.

En [15] Jenkins y Spanò también propusieron un método “exacto” para simular trayectorias de puentes para la difusión de Wright-Fisher, una posible extensión a este trabajo es la comparación de este método con la aproximación propuesta por Bladt y Sørensen [2] que aquí utilizamos.

Apendice

Simulación de trayectorias

Métodos numéricos: Euler y Milstein

Algoritmo de Euler y Milstein para simular trayectorias de la difusión de de Wright- Fisher presentados en las Secciones 3.1.1 y 3.1.2.

```
# Coeficiente de deriva, difusión y la derivada del coeficiente de difusión

# Función para evaluar la difusión:
# Args:
# x: punto en el espacio de estados [0,1]
# alpha: parámetro de la difusión de Wright-Fisher (alpha > 0)
# beta: parámetro de la difusión de Wright-Fisher (beta > 0)
# Returns:
# Data frame con los coeficientes evaluados en el punto x, mu, sigma ^2
# y sigma^2 * derivada de sigma respecto a x

WrightFisherCoef <- function(x, alpha, beta) {
  if (alpha < 0 | beta < 0) {
    stop("Error: Los parámetros no son válidos")
  }
  if (x < 0 | x > 1) {
    stop("Error: Fuera del espacio de estados")
  }

  mu <- 1/2 * (alpha * (1 - x) - beta * x)
  sigma2 <- sqrt(x * (1 - x))
  dsigma <- (1 - 2 * x) / 2
  coef <- data.frame(mu=mu,sigma2=sigma2,dsigma=dsigma)
  return(coef)
}

# Función para simular trayectorias con el método de Euler y Milstein
# con Delta constante
# Args:
# param: vector de parámetros (alpha,beta), alpha > 0 y beta > 0
# X0: estado del proceso al tiempo cero
# M: número de puntos por simular
# T0: tiempo inicial (default T0 = 0)
# T: tiempo final (default T = 1)
# método: indica el método, Euler "E" (default) o Milstein "M"
# Returns:
# Matriz de dimensión (M+1)x2 con la simulacion de una trayectoria
# la primer columna representa el tiempo y la segunda el proceso

SimEDE <- function(param, X0, M, T0=0, T=1, método="E") {

  if (param[1] < 0 | param[2] < 0) {
    stop("Error: Los parámetros no son válidos")
  }
  if (X0 < 0 | X0 > 1) {
    stop("Error: X0 fuera del espacio de estados")
  }
}
```

```

Delta <- (T - T0) / M
t <- seq(from=T0, to=T, by=Delta)
X <- numeric(M)
X[1] <- X0
for (i in 2:(M+1)) {
  wf <- WrightFisherCoef(X[i - 1], param[1], param[2])
  Bt <- sqrt(Delta) * rnorm(1)
  X[i] <- X[i-1] + wf$mu * Delta + wf$sigma2 * Bt
  if (método=="M") {
    X[i] <- X[i] + 1/2 * wf$sigma2 * wf$dsigma * (Bt^2 - Delta)
  }
  if (X[i] > 1 | X[i] < 0) {
    return("Error: Fuera del espacio de estados")
  }
}
return(cbind(t,X))
}

```

Métodos numéricos: BISS

Algoritmo para simular trayectorias de Wright-Fisher con la aproximación BISS, desarrollada en la Sección 3.1.3. La primer función es auxiliar y corresponde a al ecuación (3.6).

```

# Función de control auxiliar para el Algoritmo BISS

# Función para evaluar la difusión:
# Args:
#   y: punto en el espacio de estados [0,1]
#   eps: número positivo
# Returns:
#   Evaluacion de la Función

controlf <- function(y, eps) {
  if (y < eps){
    x <- sqrt((1 - eps) / eps)
  }
  else if (y >= eps && y < 1/2){
    x <- sqrt((1 - y) / y)
  }
  else if ( y >= 1/2 && y <= 1 - eps){
    x <- sqrt( y / (1 - y))
  }
  else{
    x <- sqrt((1 - eps) / eps)
  }
  return(x)
}

# Función para simular trayectorias con el Algoritmo BISS
# Args:
#   param: vector de parámetros (alpha,beta), alpha > 0 y beta > 0
#   X0: estado del proceso al tiempo cero
#   M: número de puntos por simular
#   T0: tiempo inicial (default T0 = 0)
#   T: tiempo final (default T = 1)
# Returns:
#   Matriz de dimensión (M+1)x2 con la simulación de una trayectoria
#   la primer columna representa el tiempo y la segunda el proceso

SimBISS <- function(param, X0, M, T0=0, T=1) {
  if (param[1] < 0 | param[2] < 0) {
    stop("Error: Los parámetros no son válidos")
  }
}

```

```

if (X0 < 0 | X0 > 1) {
  stop("Error: X0 fuera del espacio de estados")
}

X <- numeric()
A <- param[1] / 2
B <- param[2] / 2
Delta <- (T - T0) / M
t <- seq(from=T0, to=T, by=Delta)
X[1] <- X0
epsilon <- min(A * Delta, B * Delta, 1 - A * Delta, 1 - B * Delta)
if (epsilon <= 0) {
  return("Error: valor de epsilon no es válido")
}
else{
  for(i in 1:M){
    deltaW <- rnorm(1,0,sd=sqrt(Delta))
    X[i+1] <- X[i] + (A - (A + B) * X[i]) * Delta +
      (1 - (A + B) * Delta) * sqrt(X[i] * (1 - X[i])) * deltaW /
      (1 + controlf(X[i], epsilon) * abs(deltaW))
  }
  return(cbind(t,X))
}
}

```

Algoritmo exacto

El algoritmo exacto por su complejidad se divide en varias funciones, la mayoría de ellas son auxiliares para simular el coalescente de Kingman.

```

# Función auxiliar para simular el proceso ancestral del coalescente de Kingman
# Calcula de los terminos de la serie alternante
# Args:
#   t: tiempo (t>0)
#   theta: suma de los parámetros de la difusión de Wright-Fisher
#   k: número entero positivo ( k >= m)
#   m: punto en el espacio de estados del proceso ancestral del coalescente de
#       Kingman (m = 0,1,2,...)
# Returns:
#   Evaluacion de la Función

bm <- function(t, theta, k, m) {

  if (k < m | theta == 0){ # Revisa condiciones sobre los parámetros
    stop("Error: parámetros no válidos")
  }

  if ( m==0 & k == 0 & theta > 0) {
    b <- 1
  }

  else {
    lafact <- sign(theta + m + k - 1) * lgamma(theta + m + k - 1) - lgamma(theta + m) -
      (lgamma(m+1) + lgamma(k - m + 1))
    a <- (theta + 2 * k - 1) * exp(lafact)
    b <- a * exp(-k * (k + theta - 1) * t / 2)
    if (is.infinite(a) | is.nan(a)) {
      b<-"Error: bm"
    }
  }
  return(b)
}

```

```

# Función auxiliar para simular el proceso ancestral del coalescente de Kingman
# Encuentra el valor de m para el cual la Función Cm se hace cero decrece
# Args:
#   t: tiempo (t>0)
#   theta: suma de los parámetros de la difusión de Wright-Fisher
#   m: punto en el espacio de estados del proceso ancestral del coalescente de
#       Kingman (m = 0,1,2,...)
# Returns:
#   Entero positivo con el cual bm decrece por primera vez

```

```

Deps <- function(t, theta, eps) {
  k <- max(1 / t - (theta + 1) / 2, 0)
  repeat {
    if ((theta + 2 * k + 1) * exp(-(2 * k + theta) * t / 2) < 1 - eps) {
      break
    }
    else {
      k <- k + 1
    }
  }
  return(k)
}

```

```

# Función auxiliar para simular el proceso ancestral del coalescente de Kingman
# encuentra el valor de k para el cual la Función bm decrece
# Args:
#   t: tiempo (t>0)
#   theta: suma de los parámetros de la difusión de Wright-Fisher
#   m: punto en el espacio de estados del proceso ancestral del coalescente de
#       Kingman (m = 0,1,2,...)
# Returns:
#   Entero positivo con el cual bm decrece por primera vez

```

```

Cm <- function(t, theta, m) {
  i<-0
  D <- Deps(t, theta, 0)
  if (m <= D) {
    repeat {
      bi <- bm(t, theta, i + m, m)
      bimas1 <- bm(t, theta, i + m + 1, m)
      if (bimas1 < bi & is.numeric(bi) & is.numeric(bimas1)) {
        break
      }
      else if (!is.numeric(bimas1) | !is.numeric(bi)) {
        i <- "Error: Cm"
        break
      }
      else {
        i<-i+1
      }
    }
  }
  return(i)
}

```

```

# Función auxiliar para simular el proceso ancestral del coalescente de Kingman
# calcula las cotas para la Función de distribución del proceso evaluado en M
# El algoritmo toma las sumas del paso anterior y agrega los sumandos
# correspondientes
# Args:
#   t: tiempo (t>0)
#   theta: suma de los parámetros de la difusión de Wright-Fisher
#   k: vector de enteros positivos de dimensión M+1
#   M: entero positivo
#   S: vector de dos dimensiones
#   step: "M" Indica que se actualiza M (en una unidad) y se deja fijo k

```

```

#           "K" Indica que se actualiza K (en una unidad) y se deja fijo M
# Returns:
#   Vector de dimensión 2 con la cota inferior y superior

cotaS <- function(t, theta, k, M = 0, S, step = "M") {
  Smas <- 0
  Smenos <- 0
  cotas <- S
  if (step == "M" & M == 0 ) {
    S <- c(0,0)
    i <- 0:(2 * k[1] + 1)
    b <- sapply(i, FUN = bm, t = t, theta = theta, m = 0)
    if (any(!is.finite(b))) {
      cotas <- "Error: M=0"
    }
    else {
      bsigno <- (-1) ^ i * b
      Smenos <- sum(bsigno)
      Smas <- sum(bsigno[-length(b)])
      cotas <- cotas + c(Smenos, Smas)
    }
  }

  else if (step == "M" & M > 0 & length(k) == (M + 1)) {
    i <- 0:(2 * k[M+1] + 1)
    b <- sapply(M + i, FUN = bm, t = t, theta = theta, m = M)
    if (any(!is.finite(b))) {
      cotas <- "Error: M > 0"
    }
    else {
      bsigno <- (-1) ^ i * b
      Smenos <- sum(bsigno)
      Smas <- sum(bsigno[-length(b)])
      cotas <- cotas + c(Smenos, Smas)
    }
  }

  else if (step == "K" & length(k) == (M + 1)){
    for (m in M:0) {
      b <- c(bm(t, theta, m + 2 * k[m+1], m),
            bm(t, theta, m + 2 * k[m+1] + 1, m),
            bm(t, theta, m + 2 * k[m+1] - 1, m))
      if (any(!is.finite(b))) {
        cotas <- "Error: K"
        break
      }
      else if (sum(b) == 0) {
        break
      }
      else {
        bsigno <- c((-1)^(2*k[m+1]), (-1)^(2*k[m+1] + 1), (-1)^(2*k[m+1] - 1)) * b
        Smenos <- Smenos + bsigno[1] + bsigno[2]
        Smas <- Smas + bsigno[3] + bsigno[1]
      }
    }
    if (all(is.finite(cotas))) {
      cotas <- cotas + c(Smenos, Smas)
    }
  }

  else {
    stop("Error S: parámetros")
  }
  return(cotas)
}

```

Simulación del coalescente de Kingman, Algoritmo 4, utiliza las funciones auxiliares previas.

```

# Simular el proceso ancestral del coalescente de Kingman
# Args:
#   t: tiempo (t>0)
#   theta: suma de los parámetros de la difusión de Wright-Fisher
# Returns:
#   Entero positivo resultado de simular un valor del proceso

```

```

coales <- function(t, theta) {
  m <- 0
  k <- numeric(1)
  u <- runif(1)
  S <- c(0,0)
  repeat {
    C <- Cm(t, theta, m)
    if (!is.numeric(C)) {
      m <- "Error 1"
      break
    }
    k[m+1] <- ceiling(C / 2)
    S <- cotaS(t, theta, k, m, S, step = "M")
    if (!is.numeric(S)) {
      m <- "Error 2"
      break
    }
    while (u < S[2] & u > S[1] & is.numeric(m)) {
      k <- k+1
      S <- cotaS(t, theta, k, m, S, step = "K")
      if (!is.numeric(S)) {
        m <- "Error 3"
        break
      }
    }
    if (u > S[2] & is.numeric(m)) {
      k <- c(k,0)
      m <- m + 1
    }
    else {
      break
    }
  }
  return (m)
}

```

De acuerdo al Teorema 3.2.2, debido a problemas numéricos es recomendable utilizar una aproximación cuando t es cercano a cero. La siguiente función genera dicha aproximación.

```

# Aproximación al proceso ancestral del coalescente de Kingman con t chica
# Args:
#   t: tiempo (t>0)
#   theta: suma de los parámetros de la difusión de Wright-Fisher
# Returns:
#   Aproximación normal

```

```

aproxCoales <- function(t, theta) {
  beta <- (1 / 2) * (theta - 1) * t
  eta <- 1
  if (beta != 0) {
    eta <- beta / (exp(beta) - 1)
    sigma2 <- (2 * eta / t) * (eta + beta) ^ 2 *
      (1 + eta / (eta + beta) - 2 * eta) * beta ^ (-2)
  }
  if (beta == 0) {
    sigma2 <- 2 / (3 * t)
  }
  mu <- 2 * eta / t
  aprox <- rnorm(1, mu, sqrt(sigma2))
}

```

```

if (aprox <= 0) {
  A <- 0
}
else {
  A <- round(aprox)
}
return(A)
}

```

De acuerdo al Algoritmo 5 escribimos una función que genera un valor simulado en un intervalo de tiempo t dado el punto inicial, es decir una observación x_t .

```

# Simular una observacion del proceso de Wright- Fisher
# Args:
#   t: tiempo (t>0)
#   param: vector de parámetros (alpha,beta), alpha > 0 y beta > 0
#   método: E, exacto (default) o A, aproximacion normal
# Returns:
#   Observacion del proceso al tiempo t

exactWF <- function(t, param, X0, método="E"){
  if (param[1] < 0 | param[2] < 0) {
    return("Error: Los parámetros no son válidos")
  }

  if (X0 < 0 | X0 > 1) {
    stop("Error: X0 fuera del espacio de estados")
  }

  theta <- param[1] + param[2]
  if (método == "E"){
    A <- coales(t, theta)
  }
  else if (método == "A"){
    A <- aproxCoales(t, theta)
  }
  if (is.numeric(A)) {
    L <- rbinom(1, A, X0)
    x <- rbeta(1, param[1]+L, param[2] + A - L)
  }
  else {
    x <- "Error: Kingman"
  }
  return(x)
}

```

Finalmente generalizamos la función anterior para simular una trayectoria completa, con intervalos de tiempo constantes.

```

# Simular una trayectoria del proceso de Wright - Fisher
# Args:
#   param: vector de parámetros (alpha,beta), alpha > 0 y beta > 0
#   X0: estado del proceso al tiempo cero
#   N: número de puntos por simular
#   T0: tiempo inicial (default T0 = 0)
#   T: tiempo final (default T = 1)
# Returns:
#   Matriz de dimensión (M+1)x2 con la simulacion de una trayectoria
#   la primer columna representa el tiempo y la segunda el proceso

exactWFt <- function(param, X0, N, T0=0, T=1, taprox=0.05){
  if (param[1] < 0 | param[2] < 0) {
    return("Error: Los parámetros no son válidos")
  }

  if (X0 < 0 | X0 > 1) {
    stop("Error: X0 fuera del espacio de estados")
  }

```

```

}

X <- numeric(N + 1)
X[1] <- X0
Delta <- (T - T0) / M
t <- seq(from=T0, to=T, by=Delta)
for (i in 2:(N + 1)) {
  if(t[i] - t[i-1] < taprox) {
    X[i] <- exactWF(t[i] - t[i-1], param, X[i-1], "A")
  }
  else {
    X[i] <- exactWF(t[i] - t[i-1], param, X[i-1], "E")
    if (!is.numeric(X[i])) {
      break
    }
  }
}
return(cbind(t,X))
}

```

Construcción vía cadenas de Markov

A continuación, el código correspondiente al Algoritmo 6.

```

# Función para simular trayectorias del modelo Beta-Binomial (Wright-Fisher)
# Args:
# a1, a2: parámetros de  $X1 \sim \text{Gamma}(a1,1)$  y  $X2 \sim \text{Gamma}(a2,1)$  ( $a1, a2 > 0$ )
# X0: estado del proceso al tiempo cero
# M: número de puntos por simular
# T0: tiempo inicial (default T0 = 0)
# T: tiempo final (default T = 1)
# xcero: valor lógico, FALSE el estado al tiempo cero está dado,
# TRUE el estado al tiempo cero es un aleatorio.
# Returns:
# Matriz de dimensión (M+1)x2 con la simulación de una trayectoria
# la primer columna representa el tiempo y la segunda el proceso

```

```

MarkovWF <- function(a1, a2, X0 = 0.5, M, T0 = 0, T = 1, xcero = FALSE) {
  if (a1 < 0 | a2 < 0) {
    stop("Error: Los parámetros no son válidos")
  }

  if (X0 < 0 | X0 > 1) {
    stop("Error: X0 fuera del espacio de estados")
  }

  Delta <- (T - T0) / M
  t <- seq(from=T0, to=T, by=Delta)
  XY1 <- numeric()
  YX1 <- numeric()
  XY2 <- numeric()
  YX2 <- numeric()
  P <- numeric()

  if (xcero) {
    XY1[1] <- rgamma(1, a1, 1)
    XY2[1] <- rgamma(1, a2, 1)
  }
  else {
    XY1[1] <- a1
    XY2[1] <- a1 * (1 - X0) / X0
  }

  c <- (a1 + a2 - 1) / 2
  for(i in 1:M) {
    phi <- (exp(c * t[i+1]) - 1) ^ (-1)

```

```

    YX1[i] <- rpois(1,phi * XY1[i])
    XY1[i+1] <- rgamma(1,a1 + YX1[i], 1 + phi)
    XY2[i] <- rpois(1,phi * XY2[i])
    XY2[i+1] <- rgamma(1,a2 + YX2[i],1 + phi)
  }
  P <- XY1 / (XY1 + XY2)
  return(cbind(t,P))
}

```

Simulación de trayectorias de Wright-Fisher

Ahora, ya que tenemos distintos métodos para obtener trayectorias, se resumen en una función, de tal manera que siempre se obtenga una trayectoria consistente con el espacio de estados.

```

# Función para simular trayectorias
# Args:
#   param: vector de parámetros (alpha,beta), (alpha, beta > 0)
#   X0: estado del proceso al tiempo cero
#   M: número de puntos por simular (Devuelve M+1)
#   T0: tiempo inicial (default T0 = 0)
#   T: tiempo final (default T = 1)
#   método: indica el método, Euler "E", Milstein "M", BISS "BISS",
#           Exacto "Exact", Construcción vía cadenas de Markov "Markov"
# Returns:
#   Matriz de dimensión (M+1)x2 con la simulación de una trayectoria
#   la primer columna representa el tiempo y la segunda el proceso

SimulacionWF <- function(param, X0, M, T0=0, T=1, método="E", xcero=FALSE) {

  if (param[1] < 0 | param[2] < 0) {
    stop("Error: Los parámetros no son válidos")
  }
  if (X0 < 0 | X0 > 1) {
    stop("Error: X0 fuera del espacio de estados")
  }

  if (método=="E" | método=="M") {
    repeat {
      X <- SimEDE(param, X0, M, T0, T, método)
      if (is.numeric(X)) {
        break
      }
    }
  }
  else if (método == "Exact") {
    repeat{
      X <- exactWFt(param, X0, M, T0, T)
      if (is.numeric(X) & all(!is.na(X)) ) {
        break
      }
    }
  }
  else if (método == "Markov") {
    repeat {
      X <- MarkovWF(param[1], param[2], X0, M, T0, T, xcero)
      if (is.numeric(X)) {
        break
      }
    }
  }
  else if (método == "BISS") {
    repeat {
      X <- SimBISS(param, X0, M, T0, T)
      if (is.numeric(X)) {
        break
      }
    }
  }
}

```

```

    }
  }
  return(X)
}

```

Aproximación puentes de Wright-Fisher

Los siguientes códigos corresponden al Algoritmo de la Sección 3.6.1.

```

# Función para indicar si dos trayectorias se cruzan
# Args:
#   Y1: trayectoria
#   Y2: trayectoria
# Returns:
#   El tiempo en que se cruzan por primera vez o -1 si no lo hacen

cruzan <- function(Y1, Y2) {
  inter <- -1
  if(Y1[1] >= Y2[1]) {
    if(!all(Y1 > Y2)) {
      inter <- min(which(Y1 <= Y2)) - 1
    }
  }
  else {
    if(!all(Y1 < Y2)) {
      inter <- min(which(Y1 >= Y2)) - 1
    }
  }
  if(inter == 0 || inter == length(Y1)) {
    inter <- -1
  }
  return(inter)
}

# Función para generar un puente entre dos puntos
# Args:
#   param: parámetros de la difusión
#   a: valor inicial
#   b: valor final
#   t0: tiempo inicial
#   tn: tiempo final
#   M: número de puntos por simular entre los tiempos tn y t0
# método: método utilizado para generar las trayectorias
# Returns:
#   Puente entre dos tiempos

Puente <- function(param, a, b, M, t0, tn, método) {
  c <- -1
  while(c == -1) {
    Y1 <- SimulacionWF(param, a, M + 1, T0=t0, T=tn, método, xcero=FALSE)[,2]
    Y2 <- rev(SimulacionWF(param, b, M + 1, T0=t0, T=tn, método, xcero=FALSE)[,2])
    c <- cruzan(Y1, Y2)
  }
  p <- c(Y1[1:c], Y2[(c + 1):length(Y1)])
  return(p)
}

# Función para generar un puente con puntos dados
# Args:
#   tray: puntos de la trayectoria
#   param: parámetros de la difusión

```

```

# t0: tiempo inicial
# tn: tiempo final
# M: número de puntos por simular entre puntos
# método: método utilizado para generar las trayectorias
# Returns:
# Puente entre dos tiempos

PuentesDif <- function(param, tray, método, M) {
  X <- numeric()
  X[1] <- tray[1,2]
  auxX <- numeric()
  for (i in 2:dim(tray)[1]) {
    auxX <- Puente(param, tray[i-1,2], tray[i,2], M, tray[i-1,1], tray[i,1], método)
    X <- c(X, auxX[-1])
  }
  t <- seq(tray[1,1], tray[dim(tray)[1],1], (tray[i,1]-tray[i-1,1]) / M)
  return(cbind(t,X))
}

```

Algoritmo EM

Implementación del Algoritmo EM.

Las primeras dos funciones son auxiliares para hallar la función de verosimilitud del Paso E.

```

# Función auxiliar para aproximar las integrales
# utilizando el método del trapecio
# Args:
# Y: matriz de dos columnas donde la primera representa el tiempo
# y la segunda una trayectoria.
# Returns:
# Vector de dimensión 6 con la aproximación de las integrales.

integral <- function(param, x1, x2, M, t1, t2, metodo = 'Exact') {
  delta <- (t2 - t1) / (M + 1)
  integ <- c(0, 0, 0, 0, 0, 0)
  puente <- Puente(param, x1, x2, M, t1, t2, metodo)
  puente <- asin(2 * puente - 1)
  cte <- c(1, rep(2, length(puente)-2), 1)
  fun1 <- - sin(puente) / (2 * (1 + sin(puente))) + (1/4) # H1
  fun2 <- sin(puente) / (2 * (1 - sin(puente))) + (1/4) # H2
  fun3 <- (1/4) * (1 - sin(puente)) / (1 + sin(puente)) # B11
  fun4 <- (1/4) * (1 + sin(puente)) / (1 - sin(puente)) # B22
  fun5 <- -1/4 # B12
  fun6 <- -(1 / cos(puente)^2) - (1/2) * (sin(puente)^2) / (cos(puente)^2) # G
  integ[1] <- (delta / 2) * sum(fun1 * cte)
  integ[2] <- (delta / 2) * sum(fun2 * cte)
  integ[3] <- (delta / 2) * sum(fun3 * cte)
  integ[4] <- (delta / 2) * sum(fun4 * cte)
  integ[5] <- (delta / 2) * sum(fun5 * cte)
  integ[6] <- (delta / 2) * sum(fun6 * cte)
  return(integ)
}

# Función auxiliar para calcular la esperanza a partir de los puentes
# Args:
# N_esp: número de iteraciones para el cálculo de la esperanza
# param: parámetros para la simulación de puentes
# x1: punto inicial
# x2: punto final
# M_pto: número de puntos para generar el puente
# t1: tiempo inicial
# t2: tiempo final

```



```

# metodo: indica el método para generar los puentes, Exacto "Exacto,"
# construcción vía cadenas de Markov, "Markov"
# Returns:
# Vector de dimensión 6 con la esperanza

esperanza <- function(N_esp, param, x1, x2, M_pto, t1, t2, metodo='Exact') {
  i <- 1
  iter <- matrix(ncol= 6, nrow= N_esp)
  repeat{
    iter[i,] <- integral(param, x1, x2, M_pto, t1, t2, metodo)
    if (i == N_esp){
      break
    }
    i <- i+1
  }
  esp <- apply(iter, 2, mean)
  return(esp)
}

```

La idea de la siguiente función es agregar todos los elementos para hallar la verosimilitud de la ecuación 4.7.

```

# Paso E
# Args:
# X: trayectoria, matriz de dimensión dos, la primer columna
# representa el tiempo y la segunda las observaciones
# N_esp: número de iteraciones para el cálculo de la esperanza
# param: parámetros para la simulación de puentes
# M_pto: número de puntos para generar el puente
# metodo: indica el método para generar los puentes, Exacto "Exacto,"
# construcción vía cadenas de Markov, "Markov"
# Returns:
# Lista H vector de dimensión dos, B matriz de 2x2 y V la verosimilitud

PasoE <- function (X, N_esp, param, M_pto, metodo = 'Exact') {
  h <- numeric(2)
  b <- matrix(ncol=2, nrow=2)
  n <- dim(X)[1]
  esp <- c(0, 0, 0, 0, 0, 0)
  for (i in 2:n){
    esp <- esp + esperanza(N_esp, param, X[i-1,2], X[i,2], M_pto,
                          X[i-1,1], X[i,1], metodo)
  }
  h[1] <- (1/2) * (log(X[n,2]) - log(X[1,2])) + esp[1]
  h[2] <- (1/2) * (log(1 - X[n,2]) - log(1 - X[1,2])) + esp[2]
  b[1,1] <- esp[3]
  b[2,2] <- esp[4]
  b[1,2] <- esp[5]
  b[2,1] <- esp[5]
  G <- - (1/4) * log((X[n,2] * (1 - X[n,2])) / (X[1,2] * (1 - X[1,2])))
  - (1/2) * sum(diff(X[,2])^2 / diff(X[,1]))
  - sum(log(sqrt(X[,2] * (1 - X[,2]))))
  + (1/4) * esp[6]
  verosimil <- param %*% h - (1/2) * param %*% b %*% param + G
  return(list(H=h, B=b, V=verosimil))
}

```

Ahora, resulta sencillo implementar el Algoritmo 8.

```

# Algoritmo EM
# Args:

```

```

# X: trayectoria, matriz de dimensión dos, la primer columna
#     representa el tiempo y la segunda las observaciones
# N_esp: número de iteraciones para el cálculo de la esperanza
# param: parámetros iniciales
# M_pto: número de puntos para generar el puente
# metodo: indica el método para generar los puentes, Exacto "Exacto,"
#           construcción vía cadenas de Markov, "Markov"
# eps: valor para condición de paro respecto a la diferencia de
#       verosimilitud
# max_iter: número máximo de iteraciones
# Returns:
# Lista: V la verosimilitud, estimacion

EM <- function(X, N_esp, M_pto, param, eps, max_iter, metodo = 'Exact') {
  pasoM <- param
  aux <- matrix(0, ncol=3, nrow=max_iter + 1)
  difV <- Inf
  iter <- 1
  aux[1,1:2] <- pasoM
  while(iter <= max_iter & difV > eps) {
    pasoE <- PasoE(X, N_esp, pasoM, M_pto, metodo)
    aux[iter,3] <- pasoE$V
    pasoM <- as.numeric(solve(pasoE$B) %*% pasoE$H)
    aux[iter + 1, c(1,2)] <- pasoM
    if (iter == 1) {
      difV <- Inf
    } else {
      difV <- abs(aux[iter - 1,3] - aux[iter,3])
    }
    iter <- iter + 1
  }
  return(list(V=aux[1:iter,], estimacion=pasoM))
}

```

Bibliografía

- [1] ABRAMOWITZ, M., & STEGUN, I. A. (Eds.). (1965). *Handbook of mathematical functions*. Dover Publications.
- [2] BLADT, M., & SØRENSEN, M. (2014). *Simple simulation of diffusion bridges with application to likelihood inference for diffusions*. Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability, 20(2), 645–675.
- [3] DANGERFIELD, C. E., KAY, D., & BURRAGE, K. (2010). *Stochastic models and simulation of ion channel dynamics*. Procedia Computer Science, 1(1), 1581–1590.
- [4] DANGERFIELD, C. E., KAY, D., MACNAMARA, S., & BURRAGE, K. (2012). *A boundary preserving numerical algorithm for the Wright-Fisher model with mutation*. BIT Numerical Mathematics, 52(2), 283–304.
- [5] DEMPSTER, A. P., LAIRD, N. M., & RUBIN, D. B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), 39(1), 1–38.
- [6] DEVROYE, L. (1986). *Non-uniform random variate generation*. Springer.
- [7] DUDLEY, R. M.. (2002) *Cambridge studies in advanced mathematics: Real analysis and probability series number 74*. Cambridge University Press.
- [8] DURRETT, R. (2013). *Probability models for DNA sequence evolution (2002a ed.)*. Springer.
- [9] ETHIER, S. N., & GRIFFITHS, R. C. (1993). *The transition function of a Fleming-Viot process*. Annals of probability, 21(3), 1571–1590.
- [10] GALL, L. (2016). *Brownian Motion, Martingales, and Stochastic Calculus (Graduate Texts in Mathematics)*. Springer International Publishing Switzerland.
- [11] GRIFFITHS, R. C. (1980). *Lines of descent in the diffusion approximation of neutral Wright-Fisher models*. Theoretical Population Biology, 17(1), 37–50.
- [12] GRIFFITHS, R. C. (2006). *Coalescent lineage distributions*. Advances in Applied Probability, 38(2), 405–429.
- [13] GRIFFITHS, R. C., & SPANÓ, D. (2010). *Diffusion processes and coalescent trees*. En N. H. Bingham & C. M. Goldie (Eds.), Probability and Mathematical Genetics (pp. 358–379). Cambridge University Press.
- [14] IACUS, S. M. (2010). *Simulation and inference for stochastic differential equations: With R examples*. Springer.
- [15] JENKINS, P. A., & SPANÒ, D. (2017). *Exact simulation of the Wright-Fisher diffusion*. The annals of applied probability: an official journal of the Institute of Mathematical Statistics, 27(3), 1478–1509.
- [16] ØKSENDAL, B. K. (1998). *Stochastic differential equations: An introduction with applications*. Springer.
- [17] KARLIN, S., & TAYLOR, H. E. (1981). *A Second Course in Stochastic Processes*. Academic Press.

-
- [18] KARLIN, S., & TAYLOR, H. M. (1975). *A first course in stochastic processes*. Academic Press.
- [19] KINGMAN, J. F. C. (1982). *The coalescent*. *Stochastic Processes and Their Applications*, 13(3), 235–248.
- [20] MENA, R.H. & WALKER, S.G. (2009). *On a construction of Markov models in continuous time*. *Metron*, 67, 303–323.
- [21] MILSTEIN, G. N., PLATEN, E., & SCHURZ, H. (1998). *Balanced implicit methods for stiff stochastic systems*. *SIAM journal on numerical analysis*, 35(3), 1010–1019.
- [22] MODE, C. J., & SLEEMAN, C. K. (2012). *Stochastic processes in genetics and evolution: Computer experiments in the quantification of mutation and selection: Computer experiments in the quantification of mutation and selection*. World Scientific Publishing.
- [23] MORO, E., & SCHURZ, H. (2007). *Boundary preserving semianalytic numerical algorithms for stochastic differential equations*. *SIAM journal on scientific computing: a publication of the Society for Industrial and Applied Mathematics*, 29(4), 1525–1549.
- [24] MÖRTERS, P., & PERES, Y. (2010). *Brownian Motion (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.
- [25] NAOUARA, NOUREDDINE & TRABELSI, FAOUZI. (2016). A short review on boundary behavior of linear diffusion processes. 1. 138-149.
- [26] NORDBORG, M. (2004). *Coalescent Theory. En Handbook of Statistical Genetics*. John Wiley & Sons, Ltd.
- [27] REVUZ, D., & YOR, M. (2005). *Continuous Martingales and Brownian Motion*. Springer.
- [28] ROSS, S. M. (1996). *Stochastic Processes* (2a ed.). John Wiley & Sons.
- [29] SCHURZ, H. (1995). *Numerical Regularization for SDEs: Construction of Nonnegative Solutions*.
- [30] SONG, Y. S., & STEINRÜCKEN, M. (2012). A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*, 190(3), 1117–1129. <https://doi.org/10.1534/genetics.111.136929>
- [31] STEELE, J. M. (2001). *Stochastic calculus and financial applications*. Springer.
- [32] TAVARÉ, S. (1984). *Line-of-descent and genealogical processes, and their applications in population genetics models*. *Theoretical Population Biology*, 26(2), 119–164.