



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

¿CÓMO EVOLUCIONA LA LONGEVIDAD?

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

P R E S E N T A :

Fernanda López Eslava

TUTOR

ASESOR: VERÓNICA MIRÓ PINA

CO-ASESOR: ADRIÁN GONZÁLEZ CASANOVA SOBERÓN



CIUDAD DE MÉXICO

7 DICIEMBRE 2021



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Índice general

Agradecimientos	2
Notación	3
Introducción	3
I. Algunos modelos en genética de poblaciones y herramientas para estudiarlos.	6
1. El modelo de Wright-Fisher.	6
1.1 El modelo de Wright-Fisher neutro.	6
1.2 El modelo de Wright-Fisher con selección.	8
1.3 El modelo de Wright-Fisher con mutaciones.	10
2. El modelo de Moran.	12
2.1 Cadena de Markov a tiempo continuo (espacio de estados discreto).	12
2.2 El modelo de Moran neutro.	13
2.3 El modelo de Moran con selección.	15
2.4 El modelo de Moran con mutaciones.	16
3. Generadores.	17
3.1 Generador de un proceso de Markov.	17
3.2 Convergencia de generadores y movimiento Browniano estándar.	20
4. Difusiones.	21
4.1 Difusiones.	21
4.2 Función de escala y medida de velocidad.	26
4.3 Probabilidad de fijación	27
4.4 Fronteras (bordes) accesibles	28
5. Dualidad.	30
5.1 Dualidad de la difusión de Wright-Fisher con el proceso de conteo de bloques.	33
5.2 Dualidad de muestreo.	34
II. Modelo de Wright-Fisher con eficiencia.	35
6.1 El modelo.	36
6.2 La difusión asociada al modelo de Wright-Fisher con eficiencia bajo la regla ($M1$)	37
6.3 El proceso genealógico asociado.	38
6.4 El generador del modelo de Wright-Fisher con eficiencia.	40
6.5 Probabilidades de fijación.	41
III. El modelo de Moran con selección y eficiencia.	44
7.1 Modelo de Moran neutro con eficiencia	45
7.2 Modelo de Moran con selección y con eficiencia	47
IV. Modelo de Wright-Fisher con longevidad.	51
8.1 Caso simple: $d_0 = d_1 = d$	52
8.2 Probabilidades de fijación	58
Bibliografía	67

Agradecimientos

Agradezco a mis asesores, Verónica y Adrián, el apoyo que he recibido para desarrollar este trabajo, agradezco sus correcciones y reconozco el interés que tienen para impulsar y guiar a sus alumnos a lo largo de nuestra vida académica.

Agradezco a mi familia por apoyarme siempre, a mis padres, Ignacio y Claudia, a mis abuelos Jorge y Sofia y mis tíos Julio, Guillermina, Alma y Rocío.

Notación

$$\mathbb{N} = \{1, 2, \dots\}.$$

$$\mathbb{Z}^+ = \{0, 1, 2, \dots\}.$$

$$\mathbb{R} = (-\infty, \infty).$$

$$[N] = \{1, 2, \dots, N\}.$$

\mathbb{E} esperanza.

Var varianza.

$\mathbb{1}_A$ función indicadora de un conjunto A .

$(B_t, t \geq 0)$ un movimiento Browniano estándar.

$(K_t, t \geq 0)$ el coalescente de Kingman.

$(Z_t, t \geq 0)$ proceso de conteo de bloques del coalescente de Kingman.

Introducción

El objetivo de la presente tesis es estudiar el modelo de Wright-Fisher con longevidad, el cual es una extensión del modelo clásico de Wright-Fisher (Fisher, 1930, Wright, 1931), en donde una población conformada por dos tipos de individuos que tienen distintas tasas de envejecimiento, en el sentido de pasar de un estado fértil a un estado infértil en el cual no se puede reproducir más. El objetivo es entender cómo estas estrategias (ser más o menos longevo que el resto de la población) modelan la evolución de una población. Las herramientas que usaremos incluyen generadores, procesos de difusión y dualidad. Este trabajo busca desarrollar las herramientas necesarias para estudiar los modelos que aquí se presentan sin ser exhaustivos en la teoría que hay detrás de las técnicas necesarias para hacerlo.

En el capítulo I se presentan los modelos fundamentales en genética de poblaciones, el modelo de Wright-Fisher (Fisher (1930) [4], Wright (1931) [19]) y el modelo de Moran (Moran (1958) [16]). En el modelo de Wright-Fisher los individuos de una población finita eligen a su padre de manera uniforme e independiente de los demás, la población evoluciona en tiempo discreto a través de generaciones y suponemos que cada individuo tiene un solo padre (población haploide). En el modelo de Moran (Moran (1958) [16]) cada individuo en una población finita se reproduce a tasa 1, eligiendo a otro individuo, uno de ellos muere y el otro se divide en dos, en este modelo el tiempo es continuo. Estos modelos se consideran nulos cuando cada individuo tiene la misma probabilidad de reproducción, pero se pueden agregar estrategias como selección y mutación.

Las secciones 3-5 del capítulo I presentan la teoría necesaria para estudiar los modelos estocásticos que se desarrollarán en los capítulos II, III y IV, de temas tales como generador de un proceso de Markov, convergencia de generadores, difusiones y dualidad, referencias de estos temas son [1] y [2]. Las funciones de escala, la función de medida de velocidad y el concepto de ecuaciones diferenciales estocásticas se presentan de manera heurística, ya que no se pretende profundizar en estos temas, también se exponen ejemplos en cada una de estas secciones.

En el capítulo II, basado en el artículo [6], se define el modelo de Wright-Fisher con eficiencia (M1), en el cual, a diferencia de los demás modelos, el tamaño de la población no está fijo ya que depende de cómo esté conformada la población. Los individuos de la población se dividen en eficientes, aquellos que consumen menos recursos, e ineficientes, aquellos que consumen más recursos. La cantidad de recursos en cada generación está fija, y el número de individuos en cada generación depende de cuántos individuos eficientes e ineficientes conformen a la población en cada generación. Las distintas estrategias de consumo de recursos entre los individuos resultan en un beneficio para los individuos ineficientes cuando el parámetro de selección es alto. En el capítulo III se aborda el problema abierto planteado en [6], sobre encontrar un proceso a tiempo continuo que converja a la difusión de Wright-Fisher con eficiencia y que también su proceso genealógico (de ancestría) converja al proceso genealógico asociado a la difusión de Wright-Fisher. Un experimento a largo plazo de la evolución de la bacteria *Escherichia coli* dirigido por Richard Lenski (Lenski y Travisano (1994)) [12], que se realiza en la Universidad Estatal de Michigan sugiere que consumir más recursos puede ser una estrategia benéfica. En este experimento, realizado desde febrero de 1988, se han registrado las generaciones de 12 poblaciones de esta bacteria, las cuales permanecen en un medio de cultivo con glucosa y cada día

un porcentaje de la población es transferido a un medio de cultivo nuevo, uno de los resultados interesantes es que después de 10, 000 generaciones se vió que las bacterias se han hecho más grandes. Otras referencias sobre el estudio del consumo de recursos en una población y sobre la eficiencia en bacterias son: Miller et al (2005) [14], MacArthur and Wilson (1967)[13] y Molenaar et al. (1970) [17].

El capítulo IV trata el modelo de Wright-Fisher con longevidad en su caso más simple, cuando los individuos tienen distintas tasas de envejecimiento pero igual tasa de mortalidad. Estudiamos el modelo Wright-Fisher con longevidad simple utilizando técnicas desarrolladas en el capítulo I, tales como convergencia de generadores y probabilidades de fijación, además se explica la equivalencia entre este modelo y el modelo Wright-Fisher con eficiencia.

Algunos modelos en genética de poblaciones y herramientas para estudiarlos.

1. El modelo de Wright-Fisher.

1.1 El modelo de Wright-Fisher neutro.

Definición 1.1.1 (El modelo de Wright-Fisher neutro). Sean $g \in \mathbb{Z}^+$ y $N \in \mathbb{N}$ fijo. En una población de tamaño N formada por individuos de tipo 1 y de tipo 0, cada uno con la misma oportunidad de reproducción. Los individuos de la generación g se reproducen al ser elegidos de manera independiente, uniformemente al azar y con reemplazo, para ser el único padre de uno o más individuos de la generación $g+1$. Los individuos de la generación $g+1$ copian el tipo de su padre.

Definición 1.1.2 (La gráfica de Wright-Fisher). Sea $N \in \mathbb{N}$ fijo. Definimos el conjunto de vértices $V = \mathbb{Z} \times [N]$ y sean $\{U_w\}_{w \in V}$ una colección de variables aleatorias independientes e idénticamente distribuidas (iid) con distribución Uniforme en $[N]$. Sea el conjunto de aristas $E = \{ \{(g, i), (g-1, U_{(g,i)})\} \}$, $v, w \in V$ están conectados si $\{v, w\} \in E$. Definimos (V, E) como la gráfica de Wright-Fisher con N individuos por generación.

A los subconjuntos de V que coinciden en la primera coordenada les llamamos generaciones y los denotamos por $Gen(g)$, es decir, para cada $g \in \mathbb{Z}$, $Gen(g) = \{(g, i) \in V | i \in [N]\} \subset V$. En la gráfica de Wright-Fisher se definen dos cadenas de Markov, el proceso de ancestría y el proceso de frecuencia, que definimos a continuación.

Definición 1.1.3 (Proceso de ancestría de Wright-Fisher neutro). Sea $g_0 \in \mathbb{Z}$, $N \in \mathbb{N}$ y (V, E) la gráfica de Wright-Fisher con N individuos por generación. Consideramos una muestra de $n \in [N]$ individuos en la generación g_0 a la cual denotamos por S . Definimos el proceso de ancestría de $S \subset Gen(g_0) \subset V$ para cada $g \in \mathbb{Z}^+$ como

$$A_g^{(N, S)} = |\{v \in Gen(g_0 - g) | \exists u \in S \text{ tal que } u \text{ conecta con } v\}|.$$

Denotamos al proceso de ancestría como $(A_g^{(N, S)}, g \in \mathbb{Z}^+)$.

Observaciones:

- Para cada $g \in \mathbb{Z}^+$, $A_g^{(N, S)} \in [S]$, donde $[S] = \{1, 2, \dots, |S|\}$.
- El proceso de ancestría es un proceso decreciente.
- El proceso de ancestría es un proceso de Markov.

Supongamos que tenemos una población de tamaño $N \in \mathbb{N}$ en la generación g_0 y una muestra S de tamaño $n \in [N]$ y quisiéramos calcular cuál es la probabilidad de que los n individuos no tengan el mismo padre, es decir, que cada uno tenga un ancestro diferente en la generación $g_0 - 1$,

$$\begin{aligned}
\mathbb{P}[A_1^{(N, S)} = n | A_0^{(N, S)} = n] &= \prod_{i=1}^n \left(\frac{N - (i - 1)}{N} \right) \\
&= \prod_{i=0}^{n-1} \left(1 - \frac{i}{N} \right) \\
&= 1 - \sum_{i=0}^{n-1} \frac{i}{N} + o\left(\frac{1}{N^2}\right) \\
&= 1 - \frac{1}{N} \cdot \left(\frac{(n-1)n}{2} \right) + o\left(\frac{1}{N^2}\right) \\
&= 1 - \frac{1}{N} \cdot \binom{n}{2} + o\left(\frac{1}{N^2}\right).
\end{aligned}$$

Ahora calculamos el caso en el que todos los individuos en S eligen el mismo padre. Una vez que un individuo elige un padre de entre los N posibles, cada uno de los demás $n - 1$ individuos elige al mismo padre de forma independiente, con probabilidad $\frac{1}{N}$, entonces la probabilidad de que todos los individuos en S elijan un padre distinto es

$$\mathbb{P}[A_1^{(N, S)} = 1 | A_0^{(N, S)} = n] = \left(\frac{1}{N} \right)^{n-1}.$$

Entonces, la probabilidad de que dos individuos de S elijan el mismo padre es $\frac{1}{N}$ y el número de parejas en S que pueden elegir el mismo padre es $\binom{n}{2}$, por lo tanto

$$\begin{aligned}
\mathbb{P}[A_1^{(N, S)} = n - 1 | A_0^{(N, S)} = n] &= \binom{n}{2} \cdot \frac{1}{N} \left(1 - \frac{1}{N} \cdot \binom{n-1}{2} + o\left(\frac{1}{N^2}\right) \right) \\
&= \binom{n}{2} \cdot \frac{1}{N} + o\left(\frac{1}{N^2}\right).
\end{aligned}$$

es la probabilidad de que todos los individuos en S elijan un padre distinto excepto una pareja.

Así notamos que para $j \geq 3$ individuos del conjunto S , la probabilidad de que elijan el mismo padre será $\frac{1}{N^{j-1}} = o\left(\frac{1}{N^{j-2}}\right)$, por lo tanto para $j \geq 3$ tenemos que $\mathbb{P}[A_1^{(N, S)} = n - j | A_0^{(N, S)} = n] = o\left(\frac{1}{N}\right)$.

Decimos que $i, j \in [N]$ son g -equivalentes si (g_0, i) y (g_0, j) tienen un ancestro común en la generación $g_0 - g$, es decir, si existe v en la generación $g_0 - g$ y un camino tal que (g_0, i) conecta con v y otro camino (g_0, j) conecta con v .

El N-S coalescente de Wright-Fisher es la colección de relaciones de equivalencia inducida por la gráfica de Wright-Fisher. El proceso de ancestría $(A_g^{(N, S)}, g \in \mathbb{Z}^+)$ se le conoce como el proceso de conteo en cada bloque o partición.

Definición 1.1.4 (Proceso de frecuencia de Wright-Fisher neutro). *Sea $N \in \mathbb{N}$ fijo, (V, E) la gráfica de Wright-Fisher, $x \in [0, 1]$ y $\zeta \subseteq \text{Gen}(g_0) \subseteq V$ tal que $|\zeta| = x \cdot N$. Definimos el proceso de frecuencia de Wright-Fisher neutro para cada $g \in \mathbb{Z}^+$ como*

$$X_g^{(N, \zeta)} = \frac{|\{v \in \text{Gen}(g_0 + g) \mid \exists u \in \zeta \text{ tal que } v \text{ conecta con } u\}|}{N}.$$

Dado que en la generación cero $X_0^{(N)} = |\zeta|$, denotamos $X_g^{(N)} := X_g^{(N, \zeta)}$ y al proceso de frecuencia como $(X_g^{(N)}, g \in \mathbb{Z}^+)$.

Observaciones:

- Para cada $g \in \mathbb{Z}^+$, $X_g^{(N)} \in \frac{[N]}{N} \subset [0, 1]$, es una cadena de Markov aperiódica.
- Los estados $\{0\}$ y $\{1\}$ son absorbentes.
- La probabilidad de transición es

$$\begin{aligned} \mathbb{P}[X_1^{(N)} = y | X_0^{(N)} = x] &= \mathbb{P}_x[X_1^{(N)} = y] = \mathbb{P}[\text{Bin}(N, x) = y \cdot N] \\ &= \binom{N}{Ny} x^{Ny} (1-x)^{N-Ny}. \end{aligned}$$

- El proceso de frecuencia de Wright-Fisher es una martingala. De la observación anterior sabemos que $\mathbb{P}[X_1^{(N)} = y | X_0^{(N)} = x] = \mathbb{P}[\frac{\text{Bin}(N, x)}{N} = y]$, es decir, $X_{g+1}^{(N)} | X_g^{(N)} \sim \frac{1}{N} \text{Bin}(N, X_g^{(N)})$ y así

$$\mathbb{E}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N}(Nx) = x.$$

- La varianza del proceso de frecuencia cumple

$$\text{Var}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} Nx(1-x) = \frac{1}{N} x(1-x).$$

1.2 El modelo de Wright-Fisher con selección.

En este modelo se hace diferencia entre los individuos que conforman a la población. Al ser elegidos como padres de la generación siguiente, algunos de los individuos a elegir tienen ventaja selectiva, lo cual sucede habitualmente en poblaciones de seres vivos con ventajas como forma, tamaño, edad, etcétera. En este caso tenemos una población de tamaño N , algunos de los cuales tienen una desventaja selectiva, es decir, que su reproducción ocurre con probabilidad menor que la de los demás individuos. En la siguiente definición consideraremos que los individuos de tipo 0 serán los que tienen desventaja selectiva y los individuos de tipo 1 los que no tienen desventaja selectiva, o equivalentemente, los individuos de tipo 1 serán los que tienen una ventaja selectiva y los individuos de tipo 0 los que no la tienen.

Definición 1.2.5 (El modelo de Wright-Fisher con selección). Sean $s \in (0, 1)$, $g \in \mathbb{Z}^+$, $N \in \mathbb{N}$, N fijo y $k \in \{1, 2, \dots, N-1\}$. En una población de tamaño N formada por k individuos de tipo 0 y $N-k$ de tipo 1, los individuos de la generación $g+1$ eligen padre de manera independiente a los demás y con reemplazo, como sigue, con probabilidad

$$p_0 = \frac{(1-s)k}{(1-s)k + (N-k)}$$

el padre es de tipo 0 y con probabilidad

$$p_1 = \frac{(N-k)}{(1-s)k + (N-k)}$$

el padre es de tipo 1. Los individuos de la generación $g+1$ copian el tipo de su padre.

Podemos extender la definición anterior, permitiendo que el parámetro de selección $s \in (-1, 1)$. Si $s = 0$ se trata de un modelo de Wright-Fisher neutro, pero si $s \in (-1, 0)$, entonces el parámetro de selección de los individuos de tipo 0 sería $-s > 0$, en tal caso se dice que la selección es benéfica o equivalentemente que los individuos de tipo 1 tienen desventaja selectiva.

Para definir el proceso de frecuencia de individuos de tipo 0, los individuos con parámetro de selección $-s < 0$, podemos pensarlo de dos maneras. La primera es que tenemos N individuos en la generación g de los cuales sabemos su tipo (1 o 0), para que los individuos de la generación $g + 1$ elijan a su padre en la generación g y copien su tipo, cada individuo en la generación $g + 1$ elige a su padre con probabilidad

$$\mathbb{P}[\{j \text{ es el padre } i\}] = \frac{1 - s \mathbb{1}_{\{j \text{ es tipo } 0\}}}{N - s \sum_{i \in [N]} \mathbb{1}_{\{i \text{ es tipo } 0\}}}, \quad \forall i, j \in [N].$$

Es decir, que si un individuo tiene distintos posibles padres de tipo 1 y de tipo 0, es más probable que elija un padre de tipo 1, y cada padre de tipo 1 será elegido de manera uniforme.

La segunda forma de pensarlo es la siguiente, elegimos una muestra de tamaño x de individuos de tipo 0 en la generación g , cada individuo i_{g+1} en la generación $g + 1$ elige una cantidad $k_{i_{g+1}}$ de *posibles padres*, cada individuo i_{g+2} en la generación $g + 2$ elige una cantidad $k_{i_{g+2}}$ de *posibles padres*, etcétera. El número de individuos de tipo 0 en la generación $g + r$ será el número de individuos que no tienen algún ancestro de tipo 1. En este caso se puede avanzar en las generaciones conectándolas una a una mediante los *posibles padres*, aquí no es necesario saber el tipo de los individuos en la generación g para formar la generación $g + 1$, y para saber la frecuencia de individuos de tipo 0 en la generación $g + r$ basta con elegir una muestra de individuos de tipo 0 en la generación g e ir asignando el tipo a los individuos en la generación $g + 1, g + 2, \dots, g + r$ con probabilidad:

$$\mathbb{P} \left[\begin{array}{l} \{j \text{ es el padre de entre los } k \\ \text{posibles padres del individuo} \} \end{array} \right] = \frac{1 - s \mathbb{1}_{\{j \text{ es tipo } 0\}}}{N - s \sum_{i \in [N]} \mathbb{1}_{\{i \text{ es tipo } 0\}}}, \quad \forall i, j \in [N].$$

Definición 1.2.6 (Proceso de frecuencia de Wright-Fisher con selección). *Sea $V = \mathbb{Z} \times [N]$, dado que conocemos los tipos de los individuos $(g, i) \in V \forall i \in [N]$ definimos*

$$X_g^{(N)} = \frac{\sum_{i \in [N]} \mathbb{1}_{\{\tau(g, i) = 0\}}}{N}$$

la frecuencia de individuos de tipo 0 en la generación g en un modelo de Wright-Fisher con selección, con $\tau : V \rightarrow \{1, 0\}$ la función de tipo definida como $\tau(v) := \mathbb{1}_{\{v \text{ es tipo } 1\}}$. Llamamos al proceso $(X_g^{(N)}, g \in \mathbb{Z}^+)$ como el proceso de frecuencia de Wright-Fisher con selección.

Observaciones:

- Los estados $\{0\}$ y $\{1\}$ son absorbentes.
- Las probabilidades de transición son

$$\mathbb{P}[X_1^{(N)} = y | X_0^{(N)} = x] = \mathbb{P}_x[X_1^{(N)} = y] = \mathbb{P} \left[\text{Bin} \left(N, \frac{(1-s)x}{1-sx} \right) = y \cdot N \right],$$

es decir, $X_{g+1}^{(N)} | X_g^{(N)} \sim \frac{1}{N} \text{Bin}(N, \frac{(1-s)X_g^{(N)}}{1-sX_g^{(N)}})$.

- Considerando el inciso anterior y que $\frac{1}{(1-x)} = 1 + x + O(x^2)$, la esperanza de $X_1^{(N)}$ dado que $X_0^{(N)} = x$ es

$$\begin{aligned} \mathbb{E}[X_1^{(N)} - X_0^{(N)} | X_0^{(N)} = x] &= \frac{1}{N} N \frac{(1-s)x}{1-sx} - x \\ &= \frac{(1-s)x}{1-sx} - x \\ &= (1-s)x(1+sx + O(s^2)) - x \\ &= x - sx(1-x) + O(s^2) - x \\ &= -sx(1-x) + O(s^2). \end{aligned}$$

- Llamamos μ_x a la esperanza de X_1 dado $X_0^{(N)} = x$

$$\mu_x = \mathbb{E}[X_1^{(N)} | X_0^{(N)} = x], \quad (1.2.1)$$

que por el inciso anterior es $\mu_x = \frac{(1-s)x}{1-sx}$.

La varianza de $X_1^{(N)}$ dado que $X_0^{(N)} = x$ cumple

$$\begin{aligned} \text{Var}[X_1^{(N)} | X_0^{(N)} = x] &= \frac{1}{N^2} \left[N \frac{(1-s)x}{1-sx} \left(1 - \frac{(1-s)x}{1-sx} \right) \right] \\ &= \frac{1}{N} \left[\frac{(1-s)x}{1-sx} \left(1 - \frac{(1-s)x}{1-sx} \right) \right] \\ &= \frac{1}{N} [\mu_x(1-\mu_x)]. \end{aligned}$$

1.3 El modelo de Wright-Fisher con mutaciones.

En esta sección agregamos mutación al proceso de Wright-Fisher. Una mutación se define como “un cambio al azar en en la secuencia del ADN”, en el modelo de Wright-Fisher con mutaciones, consideraremos que ocurre una mutación cuando un individuo cambia de tipo, es decir, siendo su padre de cierto tipo (0 o 1), el individuo cambia de tipo. Los individuos de tipo 0 cambian a tipo 1 con probabilidad $\mu_{0,1}$ y los individuos de tipo 1 cambian a tipo 0 con probabilidad $\mu_{1,0}$.

Definición 1.3.7 (El modelo de Wright-Fisher con mutaciones). *Sean $\mu_{0,1}, \mu_{1,0} \in (0, 1)$, $g \in \mathbb{Z}^+$, $N \in \mathbb{N}$, N fijo y $k \in \{1, 2, \dots, N-1\}$. En una población de tamaño N formada por k individuos de tipo 0 y $N-k$ de tipo 1, los individuos de la generación g se reproducen al ser elegidos de manera independiente, uniformemente al azar y con reemplazo. Los individuos de la generación $g+1$ copian el tipo de su padre. Además con probabilidad $\mu_{(0,1)}$ los individuos de tipo 0 sufren una mutación y cambia a tipo 1 y con probabilidad $\mu_{(1,0)}$ los individuos de tipo 1 cambian de tipo 1 a tipo 0.*

Definición 1.3.8 (Proceso de frecuencia de Wright-Fisher con mutaciones). Sea $V = \mathbb{Z} \times [N]$, dado que conocemos los tipos de los individuos $(g, i) \in V \forall i \in [N]$ definimos

$$X_g^{(N)} = \frac{\sum_{i \in [N]} \mathbb{1}_{\{\tau(g,i)=0\}}}{N}$$

la frecuencia de individuos de tipo 0 en la generación g en un modelo de Wright-Fisher con mutación, con $\tau : V \rightarrow \{1, 0\}$ la función de tipo que a cada $(g, i) \in V$ le asigna su tipo (0 o 1). Llamamos al proceso $(X_g^{(N)}, g \in \mathbb{Z}^+)$ como el proceso de frecuencia de Wright-Fisher con mutaciones.

Observaciones:

- Para cada $g \in \mathbb{Z}^+$, $X_g^{(N)} \in \frac{[N]}{N} \subset [0, 1]$, es una cadena de Markov irreducible si y solo si $\mu_{(0,1)}$ y $\mu_{(1,0)}$ son positivos y ya que el espacio de estados es finito, la cadena tiene una única distribución estacionaria.
- Debido a las probabilidades positivas de cambiar de tipo 0 a tipo 1 a través de una mutación, no hay estados absorbentes.
- La probabilidad de transición es

$$\mathbb{P}[X_1^{(N)} = y | X_0^{(N)} = x] = \mathbb{P}_x[X_1^{(N)} = y] = \mathbb{P}[\text{Bin}(N, x(1 - \mu_{(0,1)}) + (1 - x)\mu_{(1,0)}) = y \cdot N],$$

es decir, $X_{g+1}^{(N)} | X_g^{(N)} \sim \frac{1}{N} \text{Bin}(N, X_g^{(N)}(1 - \mu_{(0,1)}) + (1 - X_g^{(N)})\mu_{(1,0)})$.

- Considerando el inciso anterior la esperanza de $X_{g+1}^{(N)}$ dado que $X_g^{(N)} = x$ es:

$$\mathbb{E}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N} [N(x(1 - \mu_{(0,1)}) + (1 - x)\mu_{(1,0)})] = x(1 - \mu_{(0,1)}) + (1 - x)\mu_{(1,0)}.$$

- Si llamamos $\mu_x = x(1 - \mu_{(0,1)}) + (1 - x)\mu_{(1,0)}$, entonces la varianza de $X_{g+1}^{(N)}$ dado que $X_g^{(N)} = x$ es

$$\text{Var}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} N [\mu_x(1 - \mu_x)] = \frac{1}{N} [\mu_x(1 - \mu_x)].$$

- Si $\mu_{0,1} = \frac{\beta_1}{N}$ y $\mu_{1,0} = \frac{\beta_2}{N}$, entonces $\mathbb{E}[X_{g+1}^{(N)} | X_g^{(N)} = x] = x + o(1)$, y así

$$\text{Var}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} N [x(1 - x) + o(1)] = \frac{1}{N} x(1 - x) + o\left(\frac{1}{N}\right).$$

2. El modelo de Moran.

En el modelo de Wright-Fisher se considera la evolución de una población de tamaño constante a través del tiempo medido en generaciones $g \in \mathbb{Z}$, en las cuales la población se renueva por completo, es decir, no hay interacción entre individuos de distintas generaciones. Presentamos ahora un modelo que a diferencia del modelo de Wright-Fisher es a tiempo continuo, el Modelo de Moran (Moran (1958) [16]), en el cual los individuos se reproducen a tasa 1, como sigue: un individuo al azar muere y es reemplazado por una copia del individuo que se reprodujo, es decir, en un evento de reproducción participan dos individuos, uno muere y otro se reproduce dando lugar a un hijo que reemplaza al individuo que muere. También podemos pensar que a tasa 1 un individuo de la población muere y elige a un individuo de la población el cual se divide en dos.

Ya que el modelo de Moran es una cadena de Markov a tiempo continuo, necesitaremos recordar algunos aspectos de este tema, basados principalmente en [5].

2.1 Cadena de Markov a tiempo continuo (espacio de estados discreto).

Una cadena de Markov a tiempo continuo con espacio de estados discreto es un proceso estocástico $(X_t, t \geq 0)$ que cambia del estado i al estado j después de un tiempo aleatorio U_i que se distribuye exponencial de parámetro ν_i y que solo depende del estado en el que esta la cadena al tiempo t , $X_t = i$, y la sucesión de distintos estados forman una cadena de Markov a tiempo discreto $(X_n, n \in \mathbb{Z}^+)$ que es llamada la cadena de Markov incrustada o esqueleto.

Definición 2.1.1 (Cadena de Markov a tiempo discreto). *Una cadena de Markov a tiempo discreto es un proceso estocástico $(X_n, n \in \mathbb{Z}^+)$ en donde cada variable aleatoria X_n toma valores en un espacio contable S y el valor de X_n para $n \geq 1$ solo depende de X_{n-1} , es decir, se cumple que para cada $n \geq 1$ y estados $i, j, k, \dots, l \in S$,*

$$\mathbb{P}[X_n = j | X_{n-1} = i, X_{n-2} = x_{n-2}, \dots, X_0 = x_0] = \mathbb{P}[X_n = j | X_{n-1} = i],$$

llamada la propiedad de Markov, si además esta propiedad no depende de n se dice que $(X_n, n \geq 0)$ es homogénea. En este caso, llamamos $\mathbb{P}[X_n = j | X_{n-1} = i]$ a la probabilidad de transición del estado i al estado j y se le denota como $\mathbb{P}[X_n = j | X_{n-1} = i] = p_{ij}$.

Definición 2.1.2 (Cadena de Markov a tiempo continuo (espacio de estados discreto)). *Sea $(X_t, t \geq 0)$ un proceso estocástico tal que para cada $t \in \mathbb{R}^+$ la variable aleatoria X_t toma valores como sigue,*

$$X_t = X_n \quad \text{para } S_n \leq t < S_{n+1}; \quad S_0 = 0; \quad S_n = \sum_{m=1}^n U_m \quad \text{para } n \geq 1,$$

con $(X_n, n \in \mathbb{Z}^+)$ una cadena de Markov con espacio de estados S finito o infinito, llamada cadena incrustada o esqueleto, y cada U_n dado $X_{n-1} = i$, es una variable aleatoria exponencial de parámetro $\nu_i > 0$ y es condicionalmente independiente de todas las demás U_m y X_m .

Proposición 2.1.3. *Sea $(X_t, t \geq 0)$ una cadena de Markov a tiempo continuo entonces, para todo $t_0 \leq t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n \in \mathbb{R}^+$ y estados $i, j, k, \dots, l \in S$ tales que $\mathbb{P}[X_{t_{n-1}} = i, X_{t_{n-2}} = k, \dots, X_{t_0} = l] > 0$, se cumple que*

$$\mathbb{P}[X_{t_n} = j | X_{t_{n-1}} = i, X_{t_{n-2}} = k, \dots, X_{t_0} = l] = \mathbb{P}[X_{t_n} = j | X_{t_{n-1}} = i].$$

Decimos que es homogénea si para todo $s, t \geq 0$

$$\mathbb{P}[X_{t+s} = j | X_s = i] = \mathbb{P}[X_t = j | X_0 = i].$$

Otra forma de plantear la transición de una cadena de Markov a tiempo continuo (espacio de estados discreto), de un estado i a un estado j , es asociando un tiempo de espera exponencial de parámetro ν_i al estado i y cuando sea momento de cambiar de estado, después de dicho tiempo de espera, elegiremos ir al estado j con probabilidad $p_{i,j}$, la probabilidad de transición del estado i al estado j de la cadena incrustada o esqueleto, y así la tasa de transición de un estado i a cualquier estado j del proceso es

$$q_{ij} = \nu_i p_{ij}.$$

De esta manera una cadena de Markov a tiempo continuo (espacio de estados discreto) está determinado por sus tasas de transición.

2.2 El modelo de Moran neutro.

Definición 2.2.1 (Modelo de Moran). *Sea $N \in \mathbb{N}$ fijo. En el modelo de Moran con tamaño de población constante N cada individuo se reproduce a tasa 1 y su hijo reemplaza a otro individuo de la población escogiéndolo uniformemente al azar.*

En cada evento de reproducción participan dos individuos, uno vive y se reproduce y el otro muere con la misma probabilidad. Si la población esta formada por dos tipos de individuos k individuos de tipo 0 y $(N - k)$ individuos de tipo 1, se elige a un individuo de tipo 1 con probabilidad $(N - k)/N$, y a tasa 1 hay k individuos de tipo 0 que lo pueden elegir para reproducirse, aumentando a $k + 1$ el número de individuos de tipo 0, de manera similar el número de individuos de tipo 1 aumenta a tasa $k/N(N - k)$. El número de individuos de tipo 0, equivalentemente de tipo 1, en la población cambia si los individuos elegidos en cada reproducción son de distinto tipo y si se eligen dos individuos del mismo tipo no hay ningún cambio.

Las tasas de transición del modelo de Moran neutro que cuenta el número de individuos de tipo 0 son

$$k \text{ a } \begin{cases} k + 1 & \text{a tasa } \frac{(N-k)k}{N}, \\ k - 1 & \text{a tasa } \frac{k(N-k)}{N}. \end{cases}$$

La Figura 2.1 ilustra el modelo de Moran con una población de 5 individuos, las flechas que unen a dos individuos representan los eventos de reproducción, que aparecen a tasa 1 para cada individuo. Cada línea horizontal representa la vida de un individuo, cuando aparece una flecha, el individuo que está al final de la flecha (la punta de la flecha) muere y es reemplazado por el individuo que está en el comienzo de la flecha (la raíz de la flecha) el cual se reproduce o divide en dos.

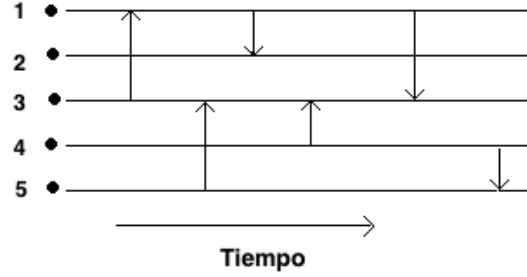


Figura 2.1: Representación gráfica del modelo de Moran con cinco individuos, las flechas representan los eventos de reproducción, $i \rightarrow j$ indica que el individuo i muere y el individuo j se reproduce.

Definición 2.2.2 (Proceso de frecuencia del modelo de Moran). *En el modelo de Moran con tamaño de población N y formada por dos tipos de individuos (tipo 1 y tipo 0), definimos el proceso de frecuencia de individuos de tipo 0 del modelo de Moran neutro como:*

$$X_t^{(N)} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{\text{individuo } i \text{ es tipo } 0\}}.$$

Observaciones:

- $(X_t^{(N)}, t \geq 0)$ es una cadena de Markov a tiempo continuo en $\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$.
- Las tasas de transición del proceso de frecuencia del modelo de Moran son

$$\frac{i}{N} a \begin{cases} \frac{i+1}{N} & \text{a tasa } \frac{(N-i)i}{N}, \\ \frac{i-1}{N} & \text{a tasa } \frac{i(N-i)}{N}. \end{cases}$$

Al igual que en el modelo de Wright-Fisher, podemos construir un proceso de ancestría para el modelo de Moran, en la Figura 2.2 comenzamos con 5 individuos y recuperamos a sus ancestros, avanzando ahora hacia atrás en el tiempo en dirección contraria a como avanzamos en la Figura 2.1, como sigue: si $i \rightarrow j$, entonces j está al final de la flecha (la punta de la flecha) y esto indica que j coalesce con i , si los individuos están en el comienzo de la flecha (la raíz de la flecha) el linaje del individuo no es afectado. Este proceso de coalescencia se representa con el árbol genealógico ilustrado de lado derecho en la Figura 2.2.

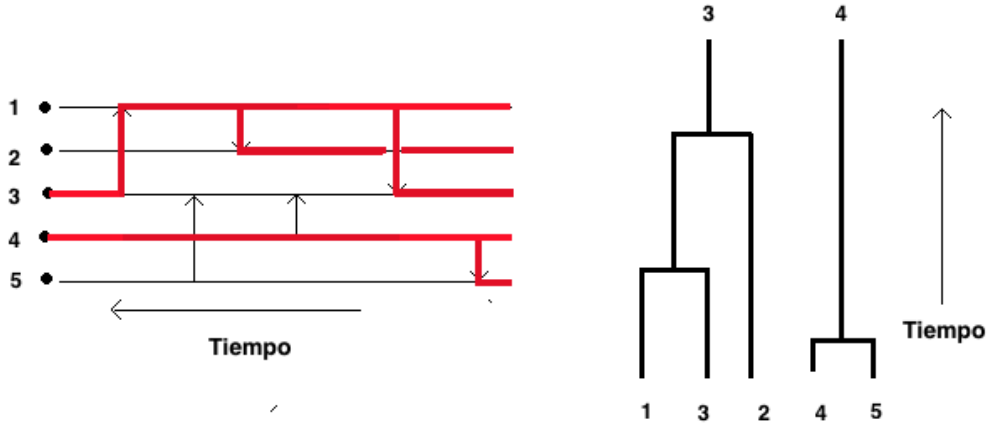


Figura 2.2: Representación gráfica del proceso de ancestría del modelo de Moran, el cual avanza hacia atrás en el tiempo. Las flechas $i \rightarrow j$ indica que el individuo j coalesce con el individuo i .

2.3 El modelo de Moran con selección.

Al igual que en el modelo de Moran neutro, cada individuo se reproduce a tasa 1, eligiendo a un individuo al azar el cual muere y es reemplazado por el individuo nuevo. En el modelo de Moran con parámetro de selección $s \in [0, 1]$, diremos que los individuos de tipo 1 tienen ventaja selectiva si cada individuo además de reproducirse a tasa 1, se reproduce también a tasa $s > 0$, es decir, existe una reproducción exclusiva de su tipo, en la que cada individuo se reproduce a tasa $s > 0$, si los individuos elegidos para un evento de reproducción son del mismo tipo no sucede nada, si el par de individuos elegidos son de distinto tipo probabilidad 1 el individuo de tipo 1 vive y se reproduce (se divide en dos) mientras que el individuo de tipo 0 es reemplazado (muere).

En cada evento de reproducción que llamaremos *común* se eligen dos individuos, uno vive y se reproduce y el otro muere lo cual sucede con la misma probabilidad, al igual que en el modelo de Moran neutro, la tasa de transición del número de individuos de tipo 0 o tipo 1, si hay k individuos de tipo 0 y $(N-k)$ individuos de tipo 1, siguiendo esta forma de reproducción es $(N-k)k/N$ cuando los individuos elegidos son de distinto tipo. Además de la reproducción *común* los individuos se reproducen a tasa $s > 0$, en la cual, si los individuos son de distinto tipo, los individuos de tipo 1 se reproducen y reemplazan al individuos de tipo 0 con probabilidad 1, la cual llamamos reproducción *especial*, así tenemos que la transición de los individuos de tipo 1 va de $(N-k)$ a $(N-K+1)$ a tasa $k(N-k)/N + k(N-k)s$.

Definición 2.3.1 (Modelo de Moran con selección). *Sea $N \in \mathbb{N}$ fijo y $s \in [0, 1]$. En el modelo de Moran con selección y tamaño de población constante N , consideramos dos tipos de individuos, tipo 0 y tipo 1. Cada individuo, ya sea de tipo 1 o tipo 0 se reproduce a tasa 1, escogiendo a un individuo de la población uniformemente al azar, un individuo muere y el otro vive y se reproduce, ambos individuos tienen la misma probabilidad de vivir y reproducirse o morir. Además cada individuo, ya sea de tipo 1 o tipo 0 se reproduce a tasa s , escogiendo a un individuo de la población uniformemente al azar, si los individuos son del mismo tipo no sucede nada, de lo contrario el individuo de tipo 1 se reproduce y reemplaza al individuo de tipo 0 con probabilidad 1.*

Las tasas de transición de los individuos de tipo 1 modelo de Moran con selección son

$$k \text{ a } \begin{cases} k + 1 & \text{a tasa } (N - k) \frac{k}{N} + s(N - k) \frac{k}{N}, \\ k - 1 & \text{a tasa } k \frac{(N-k)}{N}. \end{cases}$$

Las tasas de transición de los individuos de tipo 0 en el modelo de Moran con selección son

$$k \text{ a } \begin{cases} k + 1 & \text{a tasa } (N - k) \frac{k}{N}, \\ k - 1 & \text{a tasa } k \frac{(N-k)}{N} + s(N - k) \frac{k}{N}. \end{cases}$$

Definición 2.3.2 (Proceso de frecuencia del modelo de Moran con selección). *En el modelo de Moran con selección y tamaño de población N formada por dos tipos de individuos (tipo 1 y tipo 0), definimos el proceso de frecuencia de individuos de tipo 0 del modelo de Moran con selección como*

$$X_t^{(N)} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{\text{individuo } i \text{ es tipo } 0\}}.$$

Observaciones:

- $(X_t^{(N)}, t \geq 0)$ es una cadena de Markov a tiempo continuo en $\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$.
- Las tasas de transición del proceso de frecuencia del modelo de Moran con selección son

$$\frac{i}{N} \text{ a } \begin{cases} \frac{i+1}{N} & \text{a tasa } (N - i) \frac{i}{N}, \\ \frac{i-1}{N} & \text{a tasa } \frac{i(N-i)}{N} + s(N - i) \frac{i}{N}. \end{cases}$$

2.4 El modelo de Moran con mutaciones.

A igual que en el modelo de Wright-Fisher, agregamos mutaciones en el modelo de Moran en una población conformada por dos tipos de individuos. En el modelo de Moran con mutaciones cada individuo se reproduce a tasa 1 y además incluimos mutaciones, cambios de un tipo a otro, incluyendo un proceso de mutación (que puede ser Poisson u otro) independiente de las reproducciones de cada individuo, es decir, que entre cada reproducción un individuo puede cambiar de tipo.

Definición 2.4.1 (Modelo de Moran con mutaciones). *Sea $N \in \mathbb{N}$ fijo y $\mu_{0,1}, \mu_{1,0} \in [0, 1]$. En el modelo de Moran con mutaciones y tamaño de población N , consideramos dos tipos de individuos, tipo 0 y tipo 1. Cada individuo, sin importar su tipo, se reproduce a tasa 1, escogiendo a un individuo de la población uniformemente al azar, un individuo se reproduce y el otro muere*

y es reemplazado por el nuevo individuo, ambos individuos tienen la misma probabilidad de vivir y reproducirse o morir. Además cada individuo tiene una mutación y cambia de tipo, a tasa $\mu_{0,1}$ los individuos de tipo 0 tienen una mutación a tipo 1 y a tasa $\mu_{1,0}$ los individuos de tipo 1 tienen una mutación a tipo 0.

Definición 2.4.2 (Proceso de frecuencia del modelo de Moran con mutaciones). *En el modelo de Moran con mutaciones y tamaño de población N formada por dos tipos de individuos (tipo 1 y tipo 0), definimos el proceso de frecuencia de individuos de tipo 0 del modelo de Moran con mutaciones como:*

$$X_t^{(N)} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{\text{individuo } i \text{ es tipo } 0\}}.$$

Observaciones:

- $(X_t^{(N)}, t \geq 0)$ es una cadena de Markov a tiempo continuo en $\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$.
- Las tasas de transición del proceso de frecuencia del modelo de Moran con mutaciones son

$$\frac{i}{N} a \begin{cases} \frac{i+1}{N} & \text{a tasa } (N-i)\frac{i}{N} + \mu_{1,0}(N-i)\frac{i}{N}, \\ \frac{i-1}{N} & \text{a tasa } \frac{i(N-i)}{N} + \mu_{0,1}(N-i)\frac{i}{N}. \end{cases}$$

3. Generadores.

El tema de generadores se aborda de manera heurística ya que no se pretende un estudio exhaustivo del tema.

3.1 Generador de un proceso de Markov.

Definición 3.1.1. [5] *Sea X una v.a. no negativa, y algún valor de una v.a. discreta Y con $\mathbb{P}[Y = y] > 0$ y*

$$F_{X|Y}(x|y) = \frac{\mathbb{P}[X \leq x, Y = y]}{\mathbb{P}[Y = y]} \quad (3.1.2)$$

la función de distribución condicional de X dado que $Y = y$, definimos la esperanza condicional de X dado que $Y = y$ como

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} 1 - F_{X|Y}(x|y) dx. \quad (3.1.3)$$

Denotamos por $\mathbb{E}_y[X]$ a la esperanza condicional de X dado $Y = y$.

Definición 3.1.2 (Generador de un proceso de Markov a tiempo continuo). [2] Sea $(X_t, t \geq 0)$ un proceso de Markov a tiempo continuo, homogéneo y evaluado en los números reales. Para una función $f : \mathbb{R} \rightarrow \mathbb{R}$ se define

$$Af(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}_x[f(X_t) - f(x)]}{t}, \quad (3.1.4)$$

si el límite existe. Y denotamos por $\mathcal{D}(A)$ al conjunto de funciones para las cuales el límite existe, al cual llamamos dominio de A , y al operador A el generador infinitesimal de $(X_t, t \geq 0)$.

Definición 3.1.3 (Generador de una cadena de Markov). Sea $(X_N, N \in \mathbb{Z}^+)$ con valores en S un conjunto numerable y p_{ij} la probabilidad de transición del estado i al estado j , para una función $f : \mathbb{R} \rightarrow \mathbb{R}$ definimos el generador discreto como

$$A^{(N)}f(x) = \sum_{y \in S} [f(y) - f(x)]p_{xy}. \quad (3.1.5)$$

Ahora presentamos algunos ejemplos de generadores.

Ejemplo 3.1.4 (Proceso de Poisson). Sea $(N_t, t \geq 0)$ un proceso Poisson de intensidad λ . Para cada $f : \mathbb{N} \rightarrow \mathbb{R}$ acotada calculamos el generador como

$$\begin{aligned} Af(n) &= \lim_{h \rightarrow 0} \frac{\mathbb{E}_n[f(N_h) - f(n)]}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sum_{i=1}^{\infty} \frac{(\lambda h)^i e^{-\lambda h}}{i!} [f(n+i) - f(n)]}{h} \\ &= \lambda[f(n+1) - f(n)] + \lim_{h \rightarrow 0} \frac{\sum_{i=2}^{\infty} \frac{(\lambda)^i h^i e^{-\lambda h}}{i!} [f(n+i) - f(n)]}{h} \\ &= \lambda[f(n+1) - f(n)]. \end{aligned}$$

Ejemplo 3.1.5 (Caminata aleatoria). Sea $(X_N, N \in \mathbb{Z}^+)$ una cadena de Markov tal que $(X_N, N \in \mathbb{Z}^+)$ tiene espacio de estados $S = \mathbb{Z}$ y p_{ij} probabilidad de transición del estado i al estado j igual a $\frac{1}{2}$, con $j = i + 1$ ó $j = i - 1$. El generador es

$$A^{(N)}f(x) = \frac{1}{2}[f(x+1) - f(x)] + \frac{1}{2}[f(x-1) - f(x)].$$

Ejemplo 3.1.6 (Ecuación diferencial). Sea R_t la solución de la ecuación diferencial $\frac{dR_t}{dt} = g(R_t)$,

con $R_0 = x$ y f una función C_2 acotada. El generador es

$$\begin{aligned}
Af(x) &= \lim_{h \rightarrow 0} \frac{\mathbb{E}_x[f(R_h) - f(x)]}{h} \\
&= \lim_{h \rightarrow 0} \frac{\mathbb{E}_x[f(R_h) - f(R_0)]}{h} \\
&= \lim_{h \rightarrow 0} \frac{[f(R_h) - f(R_0)]}{h} \\
&= \lim_{h \rightarrow 0} \frac{\int_0^h [f(R_s)]' ds}{h} \\
&= \lim_{h \rightarrow 0} \frac{\int_0^h f'(R_s) \cdot R'_s ds}{h} \\
&= \lim_{h \rightarrow 0} \frac{hf'(R_0) \cdot R'_0 + o(h)}{h} \\
&= f'(R_0) \cdot R'_0 \\
&= f'(R_0) \cdot g(R_0) \\
&= f'(x) \cdot g(x).
\end{aligned}$$

Ejemplo 3.1.7 (Proceso de Yule). *Es un proceso de nacimiento puro en el cual cada individuo produce uno nuevo a tasa exponencial de parámetro λ y de manera independiente a los demás, $(X_t, t \geq 0)$ representa el tamaño de la población al tiempo t . El proceso de Yule pasa del estado n al estado $n + 1$ a tasa $\lambda \cdot n$, entonces el generador del proceso de Yule es*

$$A^{(n)}f(n) = (\lambda \cdot n)[f(n + 1) - f(n)].$$

Si $f(n) = x^n$, entonces

$$\begin{aligned}
A^{(n)}f(x) &= (\lambda \cdot n)[x^{n+1} - x^n] \\
&= (\lambda \cdot n)x(x - 1)x^{n-1} \\
&= (\lambda)x(x - 1)f'(x) \\
&= -(\lambda)x(1 - x)f'(x),
\end{aligned}$$

en las últimas igualdades tomando $f(x) = x^n$.

Ejemplo 3.1.8 (Modelo de Moran). *Consideramos el proceso de frecuencia del modelo de Moran neutro de la Definición 2.2.1, si $x \in [0, 1]$, reescalando el tiempo por $\frac{N}{2}$, tenemos que el proceso $(X_{Nt/2}^{(N)}, t \geq 0)$, tiene transiciones de x a $x + \frac{1}{N}$ a tasa $N(1-x)x$ y de x a $x - \frac{1}{N}$ a tasa $Nx(1-x)$, entonces su generador es*

$$A^{(N)}f(x) = \frac{N}{2}N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right]. \quad (3.1.10)$$

3.2 Convergencia de generadores y movimiento Browniano estándar.

Definición 3.2.1 (Convergencia en el sentido de los generadores). [10] Sea $(X_t, t \geq 0)$ un proceso estocástico con generador A y dominio $\mathcal{D}(A)$ y $(X_t^{(n)}, n \geq 0)$ una sucesión de procesos de Markov a tiempo continuo con generadores $(A^{(n)}, n \geq 0)$. Si existe $C \in \mathcal{D}(A)$ tal que para toda función $f \in C$ existe una sucesión de funciones $(f_n, n \geq 0)$, $f_n \in \mathcal{D}(A^{(n)})$, tales que f_n converge uniformemente a f y $A^{(n)}(f_n)$ converge uniformemente a $A(f)$, entonces decimos que $(X_t^{(n)}, n \geq 0)$ converge a $(X_t, t \geq 0)$ en el sentido de los generadores.

Si $(X_t^{(n)}, n \geq 0)$ es una sucesión de cadenas de Markov en tiempo discreto decimos que $(X_{\lambda_n t}^{(n)}, n \geq 0)$ converge a $(X_t, t \geq 0)$ en el sentido de los generadores si $(X_{M_t^{(n)}}^{(n)}, n \geq 0)$ converge en el sentido de los generadores a $(X_t, t \geq 0)$, en donde $(M_t^{(n)}, n \geq 0)$ es una familia de procesos Poisson independientes con intensidad λ_n .

El generador de $(X_{M_t^{(n)}}^{(n)}, n \geq 0)$ es llamado el generador discreto de $(X_{\lambda_n t}^{(n)}, n \geq 0)$ el cual es

$$A^{(n)} f(x) = \lambda_n \mathbb{E}[f(X_1^{(n)}) - f(x)].$$

Definición 3.2.2 (Movimiento Browniano estándar). Un movimiento Browniano estándar es un proceso estocástico $(B_t, t \geq 0)$ que satisface $B_0 = 0$, la función $t \rightarrow B_t$ es continua c.s y para $0 = t_0 < t_1 < \dots < t_n$, $B_{t_1} - B_{t_0}$ se tiene que $B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ son independientes con $B_{t_i} - B_{t_{i-1}}$ con distribución normal de media 0 y varianza $\sigma^2(t_i - t_{i-1})$.

Ejemplo 3.2.3 (Movimiento Browniano). Sea $(S_t^{(N)}, t \geq 0)$ una caminata aleatoria simple y simétrica, se sabe que $(\frac{1}{\sqrt{N}} S_{Nt}^{(N)}, t \geq 0)$ converge en distribución a $(B_t, t \geq 0)$ un movimiento Browniano. El generador de $(\frac{1}{\sqrt{N}} S_{Nt}^{(N)}, t \geq 0)$ es

$$A^{(N)} f(x) = \frac{N}{2} \left[f\left(x + \frac{1}{\sqrt{N}}\right) - f(x) \right] + \frac{N}{2} \left[f\left(x - \frac{1}{\sqrt{N}}\right) - f(x) \right].$$

Tomando el límite cuando $N \rightarrow \infty$ y usando la expansión de Taylor tenemos que

$$\begin{aligned} A^{(N)} f(x) &= \frac{N}{2} \left[f\left(x + \frac{1}{\sqrt{N}}\right) - f(x) \right] + \frac{N}{2} \left[f\left(x - \frac{1}{\sqrt{N}}\right) - f(x) \right] \\ &= \frac{N}{2} \left[f(x) - f'(x) \frac{1}{\sqrt{N}} + f''(x) \frac{1}{2N} + f(x) + f'(x) \frac{1}{\sqrt{N}} \right. \\ &\quad \left. + f''(x) \frac{1}{2N} - 2f(x) + o\left(\frac{1}{N}\right) \right] \\ &\xrightarrow{N \rightarrow \infty} \frac{1}{2} f''(x). \end{aligned}$$

4. Difusiones.

Este capítulo, al igual que el anterior, tiene como objetivo presentar las definiciones de tal manera que nos permitan entender los resultados de los siguientes capítulos, ya que un estudio riguroso del tema requiere teoría avanzada, como la de ecuaciones diferenciales estocásticas, que se escapa del propósito del presente trabajo.

4.1 Difusiones.

Como sucede muchas veces con modelos u objetos matemáticos en general, nos interesa saber que sucede en el límite. En nuestro caso con los modelos de poblaciones, cuando el tamaño de la población $N \in \mathbb{N}$ y el tiempo al reescalarlo en términos de tamaño de la población son muy grandes, es decir, cuando N tiende a infinito, para responder preguntas sobre el comportamiento de la población, como qué tipo de individuo se fija en la población, con qué probabilidad sucede la fijación de alguno de los tipos de individuos, o cómo influye la ventaja o desventaja selectiva en la frecuencia de individuos en la población.

En este capítulo presentamos el concepto de difusiones con el cual se trabajará en los siguientes capítulos y que nos es útil para realizar cálculos. Una manera de pensar a una difusión, que nos parece intuitiva es la siguiente, según A. Etheridge (2011) [2] una difusión es: “Un proceso de Markov fuerte en \mathbb{R} , el cual describe un camino continuo al transcurrir el tiempo”.

Definición 4.1.1 (Difusión unidimensional). *Una difusión unidimensional $(X_t, t \geq 0)$ es un proceso de Markov con generador infinitesimal*

$$Af(x) = \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} + \mu(x)\frac{df}{dx}. \quad (4.1.1)$$

Los coeficientes $\mu(x)$ y $\sigma^2(x)$ son llamados esperanza infinitesimal o deriva y varianza infinitesimal, respectivamente.

Para que el generador de una difusión este bien definido la función f debe estar en el dominio $\mathcal{D}(A)$ y la función f debe ser dos veces diferenciable en $(a, b) \subset \mathbb{R}$ el rango de la difusión (unidimensional) y otras condiciones de frontera, explicadas en el Teorema 3.17 en [2], que dependen del comportamiento de la difusión en las fronteras de su dominio, lo cual da lugar a la clasificación de los puntos frontera del rango de una difusión según Feller [3]. Si escribimos $\Delta_h X(t) = X_{t+h} - X_t$ y tomando a la función $f_1(x) = x$ y usando la propiedad de Markov tenemos que

$$Af_1(X_t) = \lim_{h \rightarrow 0} \frac{\mathbb{E}[f_1(X_{t+h}) - f_1(X_t)|X_t]}{h} = \mu(X_t)f_1'(X_t) + \frac{1}{2}\sigma^2(X_t)f_1''(X_t) = \mu(X_t).$$

Ahora escribimos

$$(X_{t+h} - X_t)^2 = X_{t+h}^2 - X_t^2 - 2X_t(X_{t+h} - X_t)$$

y tomamos $f_2(x) = x^2$

$$\begin{aligned} Af_2(X_t) - 2X_t Af_1(X_t) &= Af_2(X_t) \\ &= \lim_{h \rightarrow 0} \frac{\mathbb{E}[(X_{t+h} - X_t)^2 | X_t]}{h} \\ &= \mu(X_t)2X_t + \frac{1}{2}\sigma^2(X_t)2 \\ &= \sigma^2(X_t) + 2Af_1(X_t). \end{aligned}$$

y así $Af_2(X_t) = \sigma^2(X_t)$, lo cual motiva la denominación de $\mu(x)$ y $\sigma^2(x)$ como *esperanza infinitesimal* o *deriva y varianza infinitesimal*.

Ejemplo 4.1.2. Sea $(B_t, t \geq 0)$ un movimiento Browniano estandar, usando la expansión de Taylor se tiene que

$$f(B_t) - f(B_0) \approx f'(B_0)(B_t - B_0) + \frac{1}{2}f''(B_0)(B_t - B_0)^2,$$

tomando esperanza condicional

$$\mathbb{E}[f(B_t) - f(x) | B_0 = x] \approx \frac{1}{2}f''(x)\sigma^2 t$$

y así

$$Af(x) = \frac{\sigma^2}{2}f''(x)$$

y así $\mu(x) = 0$ y $\sigma^2(x) = \sigma^2$.

De manera heurística una difusión tiene un comportamiento para intervalos de tiempo pequeños como

$$X_t - X_0 = \mu(x)t + \sigma(x)(B_t - B_0).$$

Se puede demostrar que si μ y σ son Lipschitz continuas, es decir, $|\mu(x) - \mu(y)| \leq K|x - y|$ y $|\sigma(x) - \sigma(y)| \leq K|x - y|$, entonces la ecuación

$$X_t - X_0 = \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s \quad (4.1.2)$$

tiene solución única, en donde el segundo sumando es el límite de la sucesión de sumas $\sum_i \sigma(X(s_{i-1}))(B(s_i) - B(s_{i-1}))$. Lo anterior corresponde a temas de ecuaciones diferenciales estocásticas, pero podemos rescatar la siguiente idea intuitiva: una difusión es una ecuación diferencial, más fluctuaciones estocásticas dirigidas por un movimiento Browniano con una varianza dependiente del estado de la difusión.

Definición 4.1.3 (Ecuación diferencial estocástica). Una ecuación diferencial estocástica (EDE) es una ecuación de la forma

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \quad (4.1.3)$$

definida para valores de $t \in [0, T]$ y con condición inicial X_0 , los coeficientes de la ecuación $\mu(t, x)$ y $\sigma(t, x)$ son funciones reales definidas sobre $[0, T] \times \mathbb{R}$. Los coeficientes $\mu(t, x)$ y $\sigma^2(t, x)$ son la esperanza y la varianza infinitesimal, respectivamente.

En nuestro caso trabajaremos con la idea heurística de que una difusión es la solución de una EDE que determina su comportamiento, para profundizar en este tema puede ver por ejemplo Etheridge A. (2011) [2].

Ejemplo 4.1.4 (Difusión de Wright-Fisher). Sea $(X_g, g \in \mathbb{Z}^+)$ el proceso de frecuencia de Wright-Fisher con tamaño de población $N \in \mathbb{N}$ y $X_0 = x$ para algún $x \in [0, 1]$, sabemos que $X_{g+1}^{(N)} | X_g^{(N)} \sim \frac{1}{N} \text{Bin}(N, X_g^{(N)})$,

$$\mathbb{E}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N}(Nx) = x,$$

y la varianza del proceso de frecuencia es

$$\text{Var}[X_{g+1}^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2}Nx(1-x) = \frac{1}{N}x(1-x).$$

Por lo tanto

$$\mathbb{E}[X_{g+1}^{(N)} - X_g^{(N)} | X_g^{(N)} = x] = 0,$$

y

$$\text{Var}[X_{g+1}^{(N)} - X_g^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} [Nx(1-x) - x] = \frac{1}{N}x(1-x) + o\left(\frac{1}{N}\right).$$

Entonces su generador infinitesimal reescalando el tiempo por N es

$$A^{(N)}f(x) = \frac{1}{2} [x(1-x) + o(1)] \frac{d^2f}{dx^2}, \quad (4.1.4)$$

en donde el término $o(1)$ depende de x pero converge a cero uniformemente en x y así cuando la población tiende a infinito tenemos que el generador del proceso de Wright-Fisher reescalado el tiempo por N converge uniformemente en x a el siguiente límite

$$\lim_{N \rightarrow \infty} A^{(N)}f(x) = \frac{1}{2} [x(1-x)] \frac{d^2f}{dx^2} = Af(x), \quad (4.1.5)$$

Al proceso $(X_t, t > 0)$ cuyo generador es (4.1.5) se le conoce como difusión de Wright-Fisher $(X_t, t > 0)$ y es la única solución de la EDE

$$dX_t = \sqrt{X_t(1-X_t)}dB_t. \quad (4.1.6)$$

Ejemplo 4.1.5 (Difusión de Wright-Fisher con selección). En el modelo de Wright-Fisher con parámetro de selección $s \in (0, 1)$ y tamaño de población $N \in \mathbb{N}$, dado que la frecuencia inicial de individuos de tipo 0 es x , la frecuencia esperada en la generación siguiente es $\mu_x = x - sx(1-x) + O(s^2)$, según se vió en la ecuación (1.2.1) así la frecuencia para la siguiente generación será una v.a $X_1^{(N)} | X_0^{(N)} \sim \frac{1}{N} \text{Bin}(N, \mu_x)$,

Por lo tanto

$$\mathbb{E}[X_1^{(N)} - X_0^{(N)} | X_0^{(N)} = x] = -sx(1-x) + O(s^2),$$

y

$$\text{Var}[X_{g+1}^{(N)} - X_g^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} [N\mu_x(1-\mu_x)].$$

Si consideramos que el parámetro de selección es de orden N , fijamos $\alpha > 0$ y consideramos $s_N \in (0, 1)$ como $s_N = \alpha/N$, entonces $\mu_x = x - s_Nx(1-x) + O((s_N)^2) = x + o(1)$, y así

$$\text{Var}[X_{g+1}^{(N)} - X_g^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} [Nx(1-x) + o(1)].$$

Entonces el generador infinitesimal del proceso de frecuencia del modelo Wright-Fisher con selección reescalando el tiempo por N es

$$A^{(N)}f(x) = \frac{1}{2} [x(1-x) + o(1)] \frac{d^2f}{dx^2} + [-\alpha x(1-x) + o(1)] \frac{df}{dx}, \quad (4.1.7)$$

donde el término $o(1)$ depende de x pero converge a cero uniformemente en x , y al tomar el límite del generador tenemos el siguiente límite

$$\lim_{N \rightarrow \infty} A^{(N)}f(x) = \frac{1}{2} [x(1-x)] \frac{d^2f}{dx^2} + [-\alpha x(1-x)] \frac{df}{dx} = Af(x). \quad (4.1.8)$$

El proceso $(X_t, t \geq 0)$ cuyo generador es (4.1.8), es conocido como la difusión de Wright-Fisher con selección y es la única solución de la EDE

$$dX_t = -\alpha X_t(1-X_t)dt + \sqrt{X_t(1-X_t)}dB_t. \quad (4.1.9)$$

Ejemplo 4.1.6 (Difusión de Wright-Fisher con mutaciones). En el modelo de Wright-Fisher con mutación de parámetros $\mu_{0,1}$ y $\mu_{1,0} \in (0, 1)$, sabemos que $X_{g+1}^{(N)} | X_g^{(N)} \sim \frac{1}{N} \text{Bin}(N, X_g^{(N)}(1 - \mu_{0,1}) + (1 - X_g^{(N)})\mu_{1,0})$.

Por lo tanto

$$\mathbb{E}[X_1^{(N)} - X_0^{(N)} | X_0^{(N)} = x] = (1-x)\mu_{1,0} - x\mu_{0,1},$$

y si fijamos $\beta_1 > 0$ y $\beta_2 > 0$ $\mu_{0,1} = \beta_1/N$ y $\mu_{1,0} = \beta_2/N$ se tiene que

$$\text{Var}[X_{g+1}^{(N)} - X_g^{(N)} | X_g^{(N)} = x] = \frac{1}{N^2} N [x(1-x) + o(1)] = \frac{1}{N} x(1-x) + o\left(\frac{1}{N}\right).$$

Entonces su generador infinitesimal reescalando el tiempo por N es

$$A^{(N)}f(x) = \frac{1}{2} [x(1-x) + o(1)] \frac{d^2f}{dx^2} + [(1-x)\beta_2 - x\beta_1] \frac{df}{dx}. \quad (4.1.10)$$

La difusión de Wright-Fisher con mutación $(X_t, t > 0)$ tiene generador

$$\lim_{N \rightarrow \infty} A^{(N)}f(x) = \frac{1}{2} [x(1-x)] \frac{d^2f}{dx^2} + [(1-x)\beta_2 - x\beta_1] \frac{df}{dx}. \quad (4.1.11)$$

El proceso $(X_t, t \geq 0)$ cuyo generador es (4.1.11), es conocido como la difusión de Wright-Fisher con mutaciones y es la única solución de la EDE

$$dX_t = [(1-X_t)\beta_2 - X_t\beta_1] dt + \sqrt{X_t(1-X_t)}dB_t. \quad (4.1.12)$$

Ejemplo 4.1.7 (Convergencia del proceso de frecuencia del modelo de Moran a la difusión de Wright-Fisher). Consideramos el proceso de frecuencia del modelo de Moran neutro de la Definición 2.2.1, $(X_t^{(N)}, t \geq 0)$ es una cadena de Markov a tiempo continuo, la cual al reescalar el tiempo por $\frac{N}{2}$ tiene generador (3.1.10). Este proceso de frecuencia, al igual que el proceso de frecuencia de Wright-Fisher neutro converge en el sentido de los generadores a $(X_t, t \geq 0)$ la difusión de Wright-Fisher, la única solución de la EDE (4.1.6). Tenemos que

$$A^{(N)}f(x) = \frac{N}{2} N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2} Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right].$$

Tomando el limite cuando $N \rightarrow \infty$ y usando la expansión de Taylor tenemos que

$$\begin{aligned}
A^{(N)}f(x) &= \frac{N}{2}N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\
&= \frac{N}{2}Nx(1-x) \left[f(x) + f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} + f(x) - f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} \right. \\
&\quad \left. - 2f(x) + o\left(\frac{1}{N^2}\right) \right] \\
&= \frac{N^2}{2}x(1-x) \left[f''(x)\frac{1}{N^2} + o\left(\frac{1}{N^2}\right) \right] \\
&\xrightarrow{N \rightarrow \infty} x(1-x)\frac{f''(x)}{2}.
\end{aligned}$$

Ejemplo 4.1.8 (Convergencia del proceso de frecuencia del modelo de Moran con selección). *El proceso de frecuencia del modelo de Moran con selección de la definición 2.3.1, $(X_t^{(N)}, t \geq 0)$, es un proceso de Markov a tiempo continuo, el cual al reescalar el tiempo por $\frac{N}{2}$, converge en el sentido de los generadores a $(X_t, t \geq 0)$ la difusión de Wright-Fisher con selección, la única solución de la EDE (4.1.9).*

Tenemos que

$$\begin{aligned}
A^{(N)}f(x) &= \frac{N}{2}N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\
&\quad + \frac{N}{2}Ns(1-x)x \left[f\left(x - \frac{1}{N}\right) - f(x) \right].
\end{aligned}$$

Tomando el limite cuando $N \rightarrow \infty$ y usando la expansión de Taylor tenemos que

$$\begin{aligned}
A^{(N)}f(x) &= \frac{N}{2}N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\
&\quad + \frac{N}{2}Ns(1-x)x \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\
&= \frac{N}{2}Nx(1-x) \left[f''(x)\frac{1}{N^2} + o\left(\frac{1}{N^2}\right) \right] + \frac{N}{2}Ns(1-x)x \left[-f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} + o\left(\frac{1}{N^2}\right) \right] \\
&= \frac{N}{2}N\left(1 + \frac{s}{2}\right)x(1-x) \left[f''(x)\frac{1}{N^2} + o\left(\frac{1}{N^2}\right) \right] - \frac{N}{2}Ns(1-x)x \left[f'(x)\frac{1}{N} \right] \\
&= \frac{1}{2}\left(1 + \frac{s}{2}\right)x(1-x) \left[f''(x) + o(1) \right] - \frac{N}{2}s(1-x)x f'(x) \\
&= \left(1 + \frac{s}{2}\right)x(1-x)\frac{f''(x)}{2} - Ns(1-x)x\frac{f'(x)}{2} + o(1) \\
&\xrightarrow{N \rightarrow \infty} x(1-x)\frac{f''(x)}{2} - \alpha(1-x)x\frac{f'(x)}{2}.
\end{aligned}$$

Ejemplo 4.1.9 (Convergencia del proceso de frecuencia del Modelo de Moran con mutaciones). *El proceso de frecuencia del modelo de Moran con mutaciones de la Definición 2.4.1 es un proceso de Markov a tiempo continuo, el cual al reescalar el tiempo por $\frac{N}{2}$, en el sentido de los generadores a $(X_t, t \geq 0)$ la difusión de Wright-Fisher con mutaciones, la única solución de la*

EDE (4.1.12). *Tenemos que*

$$\begin{aligned} A^{(N)}f(x) &= \frac{N}{2}N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\ &\quad + \frac{N}{2}N\mu_{0,1}(1-x)x \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\ &\quad + \frac{N}{2}N\mu_{1,0}(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right]. \end{aligned}$$

Tomando el limite cuando $N \rightarrow \infty$ y usando la expansión de Taylor tenemos que

$$\begin{aligned} A^{(N)}f(x) &= \frac{N}{2}N(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}Nx(1-x) \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\ &\quad + \frac{N}{2}N\mu_{0,1}(1-x)x \left[f\left(x - \frac{1}{N}\right) - f(x) \right] + \frac{N}{2}N\mu_{1,0}(1-x)x \left[f\left(x + \frac{1}{N}\right) - f(x) \right] \\ &= \frac{N}{2}Nx(1-x) \left[f''(x) \frac{1}{N^2} + o\left(\frac{1}{N^2}\right) \right] + \frac{N}{2}Ns(1-x)x \left[-f'(x) \frac{1}{N} + f''(x) \frac{1}{2N^2} + o\left(\frac{1}{N^2}\right) \right] \\ &= \frac{N}{2}N \left(1 + \frac{s}{2}\right) x(1-x) \left[f''(x) \frac{1}{N^2} + o\left(\frac{1}{N^2}\right) \right] - \frac{N}{2}Ns(1-x)x \left[f'(x) \frac{1}{N} \right] \\ &= \frac{1}{2} \left(1 + \frac{s}{2}\right) x(1-x) \left[f''(x) + o(1) \right] - \frac{N}{2}s(1-x)x f'(x) \\ &= \left(1 + \frac{s}{2}\right) x(1-x) \frac{f''(x)}{2} - Ns(1-x)x \frac{f'(x)}{2} + o(1) \\ &\xrightarrow{N \rightarrow \infty} x(1-x) \frac{f''(x)}{2} - \alpha(1-x)x \frac{f'(x)}{2}. \end{aligned}$$

4.2 Función de escala y medida de velocidad.

En esta sección vemos que pasa con una difusión cuando cambiamos la escala del tiempo. Supongamos que tenemos una difusión $(X_t, t \geq 0)$ con generador A_X , deriva $\mu_X(x)$ y varianza infinitesimal $\sigma_X^2(X)$, es decir $(X_t, t \geq 0)$ es un proceso de Markov a tiempo continuo con generador (4.1.1). Consideramos un proceso nuevo $(Y_t, t \geq 0)$, con $Y_t = X_{\tau(t)}$ y

$$\tau(t) = \int_0^t \beta(Y_s) ds,$$

para alguna función $\beta(x)$ acotada, continua y estrictamente positiva. En el caso más sencillo, $\beta(x)$ es una constante, siendo así $(Y_t, t \geq 0)$ un cambio de tiempo homogéneo con el espacio. Si $Y_0 = X_0$ y si Y_t se mueve en un intervalo pequeño de tiempo $(0, dt)$, X_t se mueve de la misma manera pero en el intervalo $(0, d\tau(t)) = (0, \beta(Y_0)dt)$, entonces el generador de $(Y_t, t \geq 0)$ puede ser calculado como

$$A_Y f(x) = \beta(x) A_X f(x).$$

Sea $S(x)$ una función estrictamente creciente en el intervalo (a, b) en donde se cumplen las condiciones para que $f \in \mathcal{D}(A_X)$, y consideramos el nuevo proceso $Z_t = S(X_t)$, entonces el

generador A_Z puede ser calculado como

$$\begin{aligned}
A_Z f(x) &= \lim_{h \rightarrow 0} \frac{\mathbb{E}[f(Z_h) - f(x) | Z_0 = x]}{h} \\
&= \lim_{h \rightarrow 0} \frac{\mathbb{E}[f(S(X_h)) - f(x) | S(X_0) = x]}{h} \\
&= A_X(f \circ S)(S^{-1}(x)) \\
&= \frac{1}{2} \sigma^2(S^{-1}(x)) \frac{d^2(f \circ S)}{dx^2}(S^{-1}(x)) + \mu(S^{-1}(x)) \frac{d(f \circ S)}{dx}(S^{-1}(x)) \\
&= \frac{1}{2} \sigma^2(S^{-1}(x)) \frac{d^2(f \circ S)}{dx^2}(S^{-1}(x)) + \mu(S^{-1}(x)) \frac{d(f \circ S)}{dx}(S^{-1}(x)) \\
&= \frac{1}{2} \sigma^2(S^{-1}(x)) [S'(S^{-1}(x)) \frac{d^2 f}{dx^2}(x) + S''(S^{-1}(x)) \frac{df}{dx}(x)] + \mu(S^{-1}(x)) S'(S^{-1}(x)) \frac{df}{dx}(x) \\
&= \frac{1}{2} \sigma^2(S^{-1}(x)) S'(S^{-1}(x)) \frac{d^2 f}{dx^2}(x) + \frac{df}{dx}(x) AS(S^{-1}(x)).
\end{aligned}$$

Si encontramos una función creciente $S(x)$ tal que $AS \equiv 0$, el segundo sumando de la última igualdad es igual cero y así, con la explicación de la ecuación (4.1.2), el proceso $(Z_t, t \geq 0)$ sigue la ley de un movimiento Browniano en el intervalo $(S(a), S(b))$. Dicha función S es llamada función de escala.

Definición 4.2.1 (Función de escala). *Para una difusión $(X_t, t \geq 0)$ en (a, b) con deriva μ y varianza σ^2 , la función de escala esta definida por*

$$S(x) = \int_{x_0}^x \exp \left\{ - \int_{\nu}^y \frac{2\mu(z)}{\sigma^2(z)} dz \right\} dy, \quad (4.2.1)$$

con x_0 y ν puntos fijos en (a, b) .

Con el Ejemplo 4.1.2 y la intuición de la ecuación (4.1.2), una difusión unidimensional puede ser transformada en un movimiento Browniano a través de un cambio de la variable espacio y un cambio en la variable del tiempo, con el uso de la función de escala y la medida de velocidad respectivamente, en este trabajo no hace falta profundizar en la medida de velocidad, solo daremos su definición, ya que nos será útil para clasificar las fronteras del rango de una difusión, dos buenas referencias sobre este tema son la sección 3.3 de [2] y 2.2 de [8].

Definición 4.2.2 (Función de medida de velocidad). *Para una difusión $(X_t, t \geq 0)$ en (a, b) con deriva μ y varianza σ^2 , la función $m(\xi) = \frac{1}{\sigma^2(\xi)S'(\xi)}$ es la densidad de la medida de velocidad, o solo la densidad de velocidad, del $(X_t, t \geq 0)$ escribimos*

$$M(x) = \int_{x_0}^x m(\xi) d\xi. \quad (4.2.2)$$

4.3 Probabilidad de fijación

En el modelo de Wright-Fisher neutro o con selección, en donde los parámetros de mutación $\mu_{0,1} = \mu_{1,0} = 0$, una vez que todos los individuos en una generación son de tipo 0, en las siguientes generaciones seguirán siendo todos de tipo 0, pues los individuos de la generación $g+1$ solo podrán elegir padres de tipo 0 y por lo tanto decimos que los individuos tipo 0 se han fijado en la población.

Definición 4.3.1 (Tiempo de fijación). Sea $(X_t, t \geq 0)$ un proceso de difusión en $[a, b]$ definimos el tiempo de fijación como

$$T_{a,b} = \inf\{t \geq 0 \mid X_t = a \text{ ó } X_t = b\}.$$

El siguiente lema nos permite calcular la probabilidad de fijación de una difusión, ver Lema 5.7 en [2].

Lema 1 (Probabilidades de fijación). Sea $(X_t, t > 0)$ una difusión unidimensional en (a, b) con $\mu(x)$ su esperanza infinitesimal y $\sigma^2(x)$ su varianza infinitesimal. Si $a < a_0 < b_0 < b$ y T_y denota el tiempo en que $X_t = y$ por primera vez entonces,

$$\mathbb{P}[T_{a_0} < T_{b_0} \mid X_0 = x] = \frac{S(b_0) - S(x)}{S(b_0) - S(a_0)},$$

en donde S es la función de escala de la difusión.

Ejemplo 4.3.2 (Probabilidad de fijación del proceso de frecuencia Wright-Fisher). [1] Sean $N \in \mathbb{N}$, $(X_g^{(N)}, g \in \mathbb{Z}^+)$ el proceso de frecuencia de Wright-Fisher y $X_0 = i/N$, recordando que $X_{g+1}^{(N)} \mid X_g^{(N)} \sim \frac{1}{N} \text{Bin}(N, X_g^{(N)})$ con lo que

$$\mathbb{E}[X_{g+1}^{(N)} \mid X_g = i/N] = \frac{i}{N},$$

es decir, el valor esperado del proceso de frecuencia de Wright-Fisher permanece constante en el tiempo, pues $\mathbb{E}[X_g^{(N)}] = \frac{i}{N}$. Entonces la probabilidad de fijación es

$$\begin{aligned} \mathbb{E}[X_{T_{0,1}}^{(N)} = 1 \mid X_0 = i/N] &= 1 \cdot \mathbb{P}[X_{T_{0,1}}^{(N)} = 1 \mid X_0 = i/N] \\ &= \frac{i}{N}. \end{aligned}$$

4.4 Fronteras (bordes) accesibles

Los puntos frontera de las difusiones unidimensionales son importantes ya que si es posible alcanzarlos nos brinda información sobre el comportamiento de la población en el futuro. En el caso de nuestros modelos, los puntos 0 y 1 son los puntos frontera que nos interesa estudiar, ya que si una difusión, como la difusión del modelo de Wright-Fisher con selección por ejemplo, alcanza el 1 entonces todos los individuos son de tipo 0 con desventaja selectiva, es decir, el tipo 0 se ha fijado en la población y los individuos de tipo 1 dejan de existir a partir de ese momento. Si una difusión en $[a, b]$ alcanza los puntos $\{a, b\}$ decimos que los puntos $\{a, b\}$ son *accesibles*, de lo contrario son *inaccesibles*.

Proposición 4.4.1 (Clasificación de frontera de Feller). [3] Para una difusión unidimensional en el intervalo con puntos extremos a, b (con $a < b$), definimos

$$u(x) = \int_{x_0}^x M(y) dS(y) \text{ y } v(x) = \int_{x_0}^x S(y) dM(y),$$

en donde S es la función de escala (ver Definición 4.2.1) y M la medida de velocidad (ver Definición 4.2.2) de una difusión 4.1.1 y $x_0 \in (a, b)$. La frontera b es

frontera regular si $u(b) < \infty$ y $v(b) < \infty$

frontera de salida $u(b) < \infty$ y $v(b) = \infty$

frontera de entrada $u(b) = \infty$ y $v(b) < \infty$

frontera natural $u(b) = \infty$ y $v(b) = \infty$

las condiciones son las mismas para el punto a . Las fronteras regulares y de salida son accesibles y las fronteras de entrada y naturales son llamadas inaccesibles.

Cada una de las distintas clases de fronteras se comportan de la siguiente forma, según [8], en el caso de ser b (o a) un punto frontera *regular*, un recorrido de la difusión $(X_t, t > 0)$ puede alcanzarlo desde el interior de (a, b) y volver a (a, b) en tiempo finito, en el caso de ser b (o a) *regular* se puede modificar la difusión para que el punto b (o a) sea absorbente cuando $dX_t = 0$ en b (o a) o reflectante cuando $dX_t = -X_t$ en b (o a). En la difusión del modelo de Wright-Fisher con mutación 0 y 1 son fronteras *regulares* si $\mu_{0,1} < 1/2$ y $\mu_{1,0} < 1/2$ y son fronteras de *entrada* si $\mu_{0,1} > 1/2$ y $\mu_{1,0} > 1/2$.

En caso de ser b (o a) frontera de *salida*, un recorrido de la difusión $(X_t, t > 0)$ puede alcanzarlo desde el interior de (a, b) en tiempo finito, pero ya no puede volver a (a, b) , entonces b (o a) es un punto absorbente. Si $\{a, b\}$ son puntos frontera absorbentes, un recorrido de la difusión $(x_t, t > 0)$ se para en el tiempo $\tau = \min\{T_a, T_b\}$, si solo hay un punto frontera de *salida*, al considerar un recorrido $(x_t, t > 0)$ de la difusión $(X_t, t > 0)$, este es transitorio y el recorrido para en ese punto de salida con probabilidad 1. Y si ninguno de los puntos frontera es de *salida*, el tiempo en que en el proceso se detiene es $\tau = \infty$. En las difusiones de los modelo de Wright-Fisher neutro y con selección los puntos frontera $\{0, 1\}$ son fronteras de *salida*.

Si b (o a) es frontera de *entrada* un recorrido $(x_t, t > 0)$ de la difusión $(X_t, t > 0)$ puede entrar al intervalo (a, b) desde el punto b (o a) pero no puede volver a salir desde (a, b) . Finalmente b (o a) es un punto frontera *natural* en todos los demás casos, en este caso los recorrido $(x_t, t > 0)$ de la difusión no pueden entrar ni salir de (a, b) . Los puntos $\{a, b\}$ son *inaccesibles* si son frontera de *entrada* o *natural*.

Ejemplo 4.4.2 (Probabilidad de fijación de los individuos tipo 0 en la difusión de Wright-Fisher). *Calculamos la probabilidad de fijación de los individuos tipo 0 en la difusión de Wright-Fisher (neutro) si $\alpha = 0$ y con selección vista en el Ejemplo 4.1.5 si $\alpha > 0$. Si $x \in (0, 1)$ es la frecuencia inicial de individuos de tipo 0, entonces*

$$\mathbb{P}_x[\{fix\}] = \begin{cases} \frac{1 - \exp\{-2\alpha x\}}{1 - \exp\{-2\alpha\}} & \text{si } \alpha \neq 0 \\ x & \text{si } \alpha = 0, \end{cases}$$

Si $\alpha \neq 0$ entonces $\mu(x) = -\alpha x(1-x)$ y $\sigma^2(x) = x(1-x)$ usando el Lema 1, tenemos que

$$\mathbb{P}_x[\{fix\}] = \frac{S(0) - S(x)}{S(0) - S(1)}.$$

Sean x_0 y ν dos puntos fijos en $(0, 1)$, la función de escala de la difusión de Wright-Fisher con selección es

$$S(x) = \int_{x_0}^x \exp\left\{\int_{\nu}^y 2\alpha\right\} dy = \int_{x_0}^x \exp(2\alpha(y-\nu)) dy = \frac{\exp(2\alpha(x-\nu))}{2\alpha} - \frac{\exp(2\alpha(x_0-\nu))}{2\alpha},$$

llamamos $c_1 = \exp(-2\alpha\nu)$ y $c_2 = \frac{\exp(2\alpha(x_0-\nu))}{2\alpha}$, con lo cual $S(x) = \frac{c_1 \exp(2\alpha x)}{2\alpha} - c_2$ entonces,

$$\begin{aligned}\mathbb{P}_x[\{fix\}] &= \frac{S(0) - S(x)}{S(0) - S(1)} \\ &= \frac{(\frac{c_1}{2\alpha} - c_2) - (\frac{c_1 \exp(2\alpha x)}{2\alpha} - c_2)}{(\frac{c_1}{2\alpha} - c_2) - (\frac{c_1 \exp(2\alpha)}{2\alpha} - c_2)} \\ &= \frac{\frac{c_1}{2\alpha} - \frac{c_1 \exp(2\alpha x)}{2\alpha}}{\frac{c_1}{2\alpha} - \frac{c_1 \exp(2\alpha)}{2\alpha}} \\ &= \frac{1 - \exp\{2\alpha x\}}{1 - \exp\{2\alpha\}}.\end{aligned}$$

Si $\alpha = 0$ entonces $\mu(x) = 0$ y $\sigma^2(x) = x(1-x)$, entonces la función de escala de la difusión de Wright-Fisher es

$$S(x) = \int_{x_0}^x dy = x - x_0,$$

entonces

$$\begin{aligned}\mathbb{P}_x[\{fix\}] &= \frac{S(0) - S(x)}{S(0) - S(1)} \\ &= x.\end{aligned}$$

5. Dualidad.

En esta sección presentamos una técnica para estudiar procesos desde otra perspectiva, en nuestro caso nos interesa estudiar los procesos de frecuencia definidos en las secciones anteriores. Usaremos la *dualidad de momentos y dualidad de muestreo* como un puente para conectar dos procesos definidos en diferentes espacios, mediante una propiedad que los relaciona, para ir de un proceso a otro y estudiarlos en donde resulte más sencillo.

Hemos hablado sobre procesos de frecuencia ($X_t^{(N)}$, $t > 0$) de individuos en una población de tamaño constante N , formada por dos tipos de individuos, que para cada $t > 0$ nos dicen la frecuencia de individuos del tipo que es interesante para algún estudio, de tal manera que si $t_0 = x_0$, entonces al tiempo t_0 hay Nx_0 individuos del tipo que es de interés. Si ahora queremos ir hacia atrás en el tiempo, ya sea para buscar a los ancestros comunes de una población, o para estudiar el proceso que trajo al presente alguna mutación o enfermedad, podemos estudiar procesos como el de ancestría de Wright-Fisher (ver Definición 1.1.3) o los modelos de coalescencia que se enfocan en la genealogía de una población.

El proceso de ancestría del modelo de Wright-Fisher neutro (ver Definición 1.1.3), avanza hacia atrás en el tiempo, en una población de tamaño constante N , y como vimos en la sección

1, si comenzamos con una muestra de n individuos en la generación $g_0 \in \mathbb{Z}$, la probabilidad de que la muestra de individuos tenga $n - j$ ancestros es $o\left(\frac{1}{N}\right)$ para $j \geq 3$, es decir, que la probabilidad de que 3 o más individuos de la muestra coalescan eligiendo el mismo padre en la generación anterior es muy pequeña, con lo cuál las coalescencias suceden por parejas con probabilidad $1/N$ y los individuos que colasen se eligen al azar de forma $\binom{n}{2}$. Si avanzamos en el tiempo en unidades del tamaño de la población N y al mismo tiempo dejamos tender N a infinito obtenemos un proceso de Markov a tiempo continuo conocido como el *n-coalescente de Kingman*.

Un proceso de coalescencia en $[N]$ definido para cada $t > 0$ es un proceso de Markov a tiempo continuo $(\pi_t, t > 0)$, que toma valores en las clases de equivalencia de $\{1, 2, \dots, n\}$, que para cada $t > 0$ π_t corresponde al ancestro vivo al tiempo $t_0 - t$, con t_0 el tiempo en el que tenemos n individuos o linajes, que consiste en las etiquetas de los individuos que son descendientes de dicho ancestro. Pensando de manera general, estos procesos pueden tener distintas suposiciones para que los individuos coalescan. Notemos que aquí estamos suponiendo que la coalescencia sucede entre dos individuos elegidos de manera uniformemente al azar de una muestra de n individuos con probabilidad $1/N$ de elegir el mismo padre.

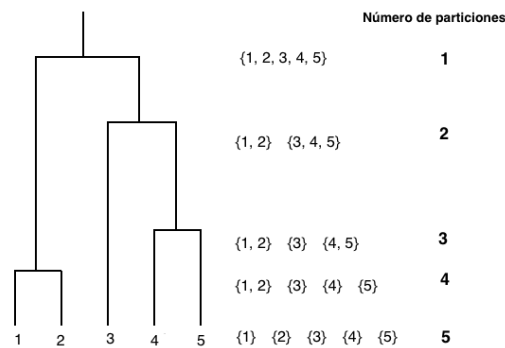


Figura 5.3: Una realización del proceso de coalescencia con una muestra de 5 individuos.

Como podemos ver en la figura 5.3, las coalescencias de linajes hace que el número de particiones disminuya hacia atrás en el tiempo. A los procesos que cuentan el número de particiones de un proceso de coalescencia les llamaremos procesos de conteo de bloques. El proceso de ancestría del modelo de Wright-Fisher neutro cuenta las particiones de un proceso de coalescencia cuyas tasas de transición son las mencionadas en la sección 1.

Definición 5.1.1 (Coalescente de Kingman). *El coalescente de Kingman es una cadena de Markov a tiempo continuo, $(K_t, t > 0)$, en $S_{[k]}$, el espacio de las relaciones de equivalencia en $[k]$, con tasas de transición $q_{\xi, \eta}$, con ξ y $\eta \in S_{[k]}$ dadas por*

$$q_{\xi, \eta} = \begin{cases} 1 & \text{si } \eta \text{ se obtiene al coalescer dos clases de equivalencia de } \xi, \\ 0 & \text{en otro caso.} \end{cases} \quad (5.1.1)$$

El *coalescente de Kingman* en \mathbb{N} es un proceso de equivalencia con la propiedad de que al tomar una muestra de tamaño $n \in \mathbb{N}$, el proceso de coalescencia en $S_{[n]}$ vuelve a ser un proceso de coalescencia de Kingman.

Definición 5.1.2 (Proceso de conteo de bloques). *Definimos el proceso de conteo de bloques como el proceso estocástico $(N_t, t > 0)$, que cuenta cuántos clases de equivalencia hay al tiempo $t > 0$ en un proceso de coalescencia.*

Denotaremos $(\tilde{B}_t, t > 0)$ proceso de conteo de bloques del coalescente de Kingman.

Proposición 5.1.3. *El proceso de conteo de bloques $(\tilde{B}_t, t > 0)$ en $[n]$ es un proceso de Markov a tiempo continuo que va de k a $k - 1$ a tasa $\frac{k(k-1)}{2}$.*

Proposición 5.1.4. *El proceso de conteo de bloques $(\tilde{B}_t, t > 0)$ con $\tilde{B}_0 = n$ tiene generador*

$$A^{(n)}f(k) = \binom{k}{2}[f(k-1) - f(k)].$$

Proposición 5.1.5. *El proceso de ancestría de Wright-Fisher de la definición 1.1.3, con una muestra inicial de tamaño n , al reescalar el tiempo por N , converge al coalescente de Kingman cuando el tamaño de la población tiende a infinito.*

Demostración.

$$\begin{aligned} A^{(N)}f(k) &= N \left[\left(1 - \frac{\binom{k}{2}}{N} + o\left(\frac{1}{N}\right) \right) (f(k) - f(k)) + \left(\frac{\binom{k}{2}}{N} + o\left(\frac{1}{N}\right) \right) (f(k-1) - f(k)) \right. \\ &\quad \left. + \sum_{j>1}^{k-1} o\left(\frac{1}{N}\right) (f(k-j) - f(k)) \right] \\ &= N \left[\left(\frac{\binom{k}{2}}{N} + o\left(\frac{1}{N}\right) \right) (f(k-1) - f(k)) + o\left(\frac{1}{N}\right) \right] \\ &= N \left[\left(\frac{\binom{k}{2}}{N} \right) (f(k-1) - f(k)) + o\left(\frac{1}{N}\right) \right] \\ &= \binom{k}{2} (f(k-1) - f(k)) + o(1) \\ &\xrightarrow{N \rightarrow \infty} \binom{k}{2} (f(k-1) - f(k)). \end{aligned}$$

Con lo cual tenemos que

$$\lim_{N \rightarrow \infty} A^{(N)}f(k) = \binom{k}{2} (f(k-1) - f(k)) = A^{(n)}f(k).$$

en donde n es el número de bloques (linajes) en el tiempo $t = 0$ en el proceso de ancestría de Wright-Fisher y en el coalescente de Kingman. \square

Definición 5.1.6 (Procesos H-duales). *Sean $(X_t, t > 0)$ y $(Z_t, t > 0)$ dos procesos estocásticos con valores en S_1 y S_2 y con generadores A_1 y A_2 respectivamente y $H : S_1 \times S_2 \rightarrow \mathbb{R}$ una función continua y acotada. Decimos que X_t y Z_t son H-duales si para cada $x \in S_1$, $z \in S_2$ y $t \geq 0$, se cumple que*

$$\mathbb{E}_x[H(X_t, z)] = \mathbb{E}_z[H(x, Z_t)]. \quad (5.1.2)$$

Para la prueba de la siguiente proposición ver proposición 1.2 de Jansen and Kurt (2014) [9].

Proposición 5.1.7 (Procesos H-duales en el sentido de los generadores). Sean $(X_t, t > 0)$ y $(Z_t, t > 0)$ dos procesos de Markov con generadores con valores en S_1 y S_2 y con generadores A_1 y A_2 respectivamente y $H : S_1 \times S_2 \rightarrow \mathbb{R}$ continua y acotada. Si $H(x, \cdot), \mathbb{E}_x[H(X_t, \cdot)] \in \mathcal{D}(A_2)$ para toda $x \in S_1, t \geq 0$ y $H(\cdot, z), \mathbb{E}_z[H(\cdot, Z_t)] \in \mathcal{D}(A_1)$ para toda $z \in S_2, t \geq 0$, y si

$$A_1(H(\cdot, z))|_x = A_2(H(x, \cdot))|_z \quad \forall x \in S_1, z \in S_2, \quad (5.1.3)$$

entonces $(X_t, t > 0)$ y $(Z_t, t > 0)$ son H-duales.

Diremos que dos procesos estocásticos son *momento duales* si se cumple 5.1.2 con la función $H(x, n) = x^n$.

5.1 Dualidad de la difusión de Wright–Fisher con el proceso de conteo de bloques.

Sea $(X_t, t > 0)$ la difusión de Wright–Fisher con generador 4.1.5 y consideramos la función $H : [0, 1] \times \mathbb{N} \rightarrow [0, 1]$, $H(x, n) = x^n$, notamos que es continua y acotada, recordemos que la difusión de Wright–Fisher se obtienen con el proceso de frecuencia de Wright–Fisher, al dejar crecer el tamaño de la población $N \rightarrow \infty$ y reescalando el tiempo por N , y el coalescente de Kingman 5.1.1 se obtiene en el límite, cuando dejamos tender el tamaño de la población a infinito y medimos el tiempo hacia atrás, de la muestra a sus ancestros, en unidades de tamaño N para poder observar el ancestro común más cercano. Veamos ahora que estos procesos son momento duales en el sentido de los generadores.

Proposición 5.1.8. Sean $(X_t, t \geq 0)$ la difusión de Wright–Fisher y $(Z_t, t > 0)$ el proceso de conteo de bloques del coalescente de Kingman, estos procesos son momento duales.

Demostración. Sea A_1 el generador infinitesimal de $(X_t, t \geq 0)$ y $H : [0, 1] \times \mathbb{N} \rightarrow [0, 1]$, definida como $H(x, n) = x^n$. Como $H(x, n)$ y $\mathbb{E}[x^{Z_t}]$ son polinomios en x y $H(X, n)$ y $\mathbb{E}[Z_t^x]$ son acotadas en n , por la Proposición 5.1.7, basta probar que para toda $x \in [0, 1]$ y para toda $n \in \mathbb{N}$, se cumple que

$$A_1 H(x, n) = A_2 H(x, n).$$

$$\begin{aligned} A_1 H(x, n) &= \frac{x(1-x)}{2} \frac{dH(x, n)}{dx^2} \\ &= x(1-x) \frac{n \cdot (n-1)}{2} x^{n-2} \\ &= \binom{n}{2} (x^{n-1} - x^n) \\ &= \binom{n}{2} (H(x, n-1) - H(x, n)) \\ &= A_2 H(x, n). \end{aligned}$$

□

5.2 Dualidad de muestreo.

En esta sección presentamos una función que nos permite relacionar procesos, al igual que $H(x, n) = x^n$ nos permitió ver que la difusión de Wright–Fisher con el proceso de conteo de bloques son *momento duales*. El proceso de ancestría y el proceso de frecuencia de un modelo de Wright–Fisher neutro, son ambos procesos de conteo, el primero hacia atrás en el tiempo y el segundo hacia adelante, de manera que parece bueno relacionarlos.

Si conocemos la frecuencia inicial $X_0 = x$ de individuos en un población y nos interesa seguir su proceso de evolución, pensamos hacia adelante en el tiempo, y nos interesan todos sus descendientes. Pero si buscamos una respuesta hacia atrás en el tiempo, es decir, estamos interesados en saber cómo llegamos a tener una frecuencia x de individuos en el presente, queremos saber el número de ancestros que tuvieron en el pasado.

Consideramos $N, n \in \mathbb{N}, x \in [0, 1]$ y $(X_g^{(N)}, g \in \mathbb{Z})$ y $(A_g^{(NS)}, g \in \mathbb{Z})$ los procesos de frecuencia y ancestría respectivamente, de la definición 1.1.3 y 1.1.4, con una muestra con reemplazo \bar{S} de tamaño n en la generación $g = 1$ y con una frecuencia $X_0 = x$, en la generación $g = 0$ y la función $S : \frac{[N]}{N} \times [N] \rightarrow [0, 1]$ definida como

$$S(x, n) = \mathbb{P} \left[\left\{ \begin{array}{l} \text{todos los individuos de una muestra con reemplazo} \\ \text{de tamaño } n \text{ en la generación } g = 1 \text{ son tipo 0,} \\ \text{y la frecuencia de individuos de tipo 0 en la generación } g = 0 \text{ es } x \end{array} \right\} \right],$$

La probabilidad de que un individuo en la generación $g = 1$, de los n que hay en la muestra, elija un padre de tipo 0 es la frecuencia de individuos de tipo 0 en la generación anterior $g = 0$, que es x , y la probabilidad de que dos individuos en la generación $g = 1$ elijan padre de tipo 0 es x^2 , ya que cada uno elige de manera independiente y con reemplazo, entonces $S(x, n) = x^n$, lo cual se ilustra en la figura 5.4.

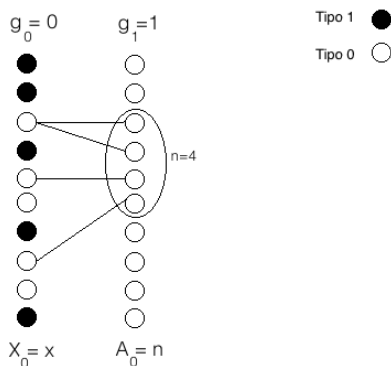


Figura 5.4: Representación del proceso de ancestría en una generación en la gráfica de Wright–Fisher con tamaño de población 10 y muestra inicial $n = 4$.

En general para $g \in \mathbb{Z}^+$, para que en una muestra de n individuos hayan solo individuos de tipo 0, los ancestros de dicha muestra en g generaciones anteriores deben ser todos tipo 0, lo cual sucede con probabilidad $(x)^{A_0^{(N)}}$, con x la frecuencia inicial de individuos de tipo 0, y si vamos hacia adelante en el tiempo con el proceso de frecuencia, entonces con probabilidad $(X_0)^n$, el proceso de ancestría en la generación $g + 1$ esta en el estado n .

Proposición 5.2.1 (Dualidad de muestreo de Möhle). [15] Sean $(X_g^{(N)}, g \in \mathbb{Z})$ y $(A_g^{(NS)}, g \in \mathbb{Z}^+)$ el proceso de frecuencia y el proceso de ancestría, y la función de de muestreo $S : \frac{[N]}{N} \times [N] \times \mathbb{Z} \rightarrow [0, 1]$, definida como

$$S(x, n, g) = \mathbb{P} \left[\left\{ \begin{array}{l} n \text{ individuos en la generación } g + 1 \text{ son tipo } 0, \\ \text{y la frecuencia inicial de individuos de tipo } 0 \text{ es } x \end{array} \right\} \right]$$

y $S(x, n, 0) = S(x, n) = x^n$, entonces $(X_g^{(N)}, g \in \mathbb{Z})$ y $(A_g^{(NS)}, g \in \mathbb{Z})$ son S-duales.

Demostración. Tomamos una muestra \bar{S} de n individuos en la generación $g + 1$, entonces los ancestros de los n individuos en g generaciones deben ser de tipo 0 para que todos los individuos de la muestra sean de tipo 0, por lo tanto

$$\begin{aligned} S(x, n, g) &= \sum_{i=1}^n \mathbb{P}[A_g^{(N\bar{S})} = i] \cdot S(x, i) \\ &= \mathbb{E}_n[S(x, A_g^{(N\bar{S})})]. \end{aligned}$$

Si la frecuencia inicial de individuos de tipo 0 es $X_0 = x$, en las siguientes generaciones tendremos una muestra de n individuos en la generación $g + 1$ si los n individuos eligen un padre de tipo 0 en la generación g con probabilidad X_g , entonces tenemos que

$$\begin{aligned} S(x, n, g) &= \sum_{y \in \frac{[N]}{N}} \mathbb{P}[X_g^{(N\bar{S})} = y] \cdot S(y, n) \\ &= \mathbb{E}_x[S(X_g^{(N\bar{S})}, n)], \end{aligned}$$

por lo tanto

$$\mathbb{E}_n[S(x, A_g^{(N\bar{S})})] = \mathbb{E}_x[S(X_g^{(N\bar{S})}, n)] \quad (5.2.1)$$

lo cual muestra que los procesos de frecuencia y ancestría $(X_g^{(N)}, g \in \mathbb{Z})$ y $(A_g^{(NS)}, g \in \mathbb{Z}^+)$ son S-duales y momento duales pues $S(x, n) = x^n$. \square

Diremos que dos procesos son *muestreo duales* si son S-duales.

Modelo de Wright-Fisher con eficiencia.

En los modelos estudiados en las secciones anteriores hemos supuesto que el tamaño de la población permanece siempre constante a través de las generaciones, la cuál es una suposición que no muchas veces modela situaciones reales en las poblaciones. En esta sección presentamos un modelo propuesto en González Casanova A. et al. [6], en el cuál el tamaño de la población no está fijo, además el mecanismo de reproducción de los individuos se modifica por la eficiencia de los individuos.

El desarrollo de una población está relacionado al entorno en el que se encuentra, en particular en los recursos disponibles en dicho entorno, con los cuales la población vive y se reproduce. Si los recursos son escasos o hay competencia para conseguirlos, el tamaño de la población puede verse afectada.

El modelo presenta dos tipos de individuos con diferentes estrategias de consumo de recursos para ser progenitores. Consideramos individuos con un costo alto de recursos para poder reproducirse, a los cuales llamaremos *ineficientes*, e individuos que tienen un costo bajo de recursos para reproducirse a los cuales llamaremos *eficientes*, la cantidad de recursos a lo largo de las generaciones es fija. La cantidad de individuos en la población depende de los recursos consumidos por la misma, que a su vez depende de la frecuencia de individuos *eficientes*, pues ellos necesitan menos para reproducirse.

6.1 El modelo.

En el modelo, cada individuo en la generación presente elige a un padre de la generación anterior independientemente de la elección de los demás individuos, de igual forma que en el modelo de Wright-Fisher neutro, pero la cantidad de individuos que eligen a un padre no está determinada, solo fijamos la cantidad de recursos disponibles para formar la generación. Tendremos más individuos en la población si se eligen más padres *eficientes*, pues el costo de recursos para producirlos es menor.

Fijamos $N \in \mathbb{N}$ y $\kappa, s, x \in [0, 1]$, en donde N es la cantidad de recursos, κ el coeficiente de eficiencia, $-s$ el coeficiente de selección de los individuos *eficientes* y x la frecuencia inicial de dichos individuos. Los individuos de tipo 0 son: *eficientes*, necesitan $(1 - \kappa)$ unidades de recursos para reproducirse y tienen coeficiente de selección $-s$, los individuos de tipo 1 son: *ineficientes*, necesitan 1 unidad de recursos para reproducirse y tienen coeficiente de selección 0.

Cada generación se construye considerando que la cantidad de recursos disponible es N fijo, el costo para crear el primer individuo puede ser 1 o $(1 - \kappa)$, así al momento de crear el segundo individuo tenemos $N - 1$ o $N - (1 - \kappa)$ recursos disponibles, el proceso continúa hasta que los recursos se agotan. En [6], se consideran dos posibles reglas para detener el proceso de reproducción, el modelo cambia al elegir una u otra regla. En la primera regla de paro, que llamamos **(M1)**, los recursos disminuyen cada vez que se crea un nuevo individuo (según el tipo de individuo creado se consumirá $(1 - \kappa)$ unidades de recursos o 1 unidad de recursos), se detiene el proceso de reproducción la primera vez que la cantidad de recursos es igual o mayor a N , es decir, el primer individuo que al ser creado (consumiendo $(1 - \kappa)$ unidades de recursos o 1 unidad de recursos) exceda o iguale por primera vez los recursos disponibles será el último individuo creado. En la segunda regla de paro, que llamamos **(M2)**, nuevos individuos son creados disminuyendo los recursos en cada reproducción, el proceso de reproducción se detiene la primera vez que la cantidad de recursos es igual o mayor a N , si el último individuo al ser creado termina por consumir exactamente N recursos entonces es creado, pero si el último individuo al ser creado excede los N recursos disponibles es descartado, es decir que no se crea, con lo cual tenemos dos posibles modelos determinados por las reglas de paro **(M2)** o **(M1)**.

Definición 6.1.1 (El modelo de Wright-Fisher con eficiencia de parámetros N, κ, s y x). *Cada generación es una colección de individuos $\{1, 2, \dots, M_n\}$, con M_n el tamaño de la población en la generación n . Sea $\tau(n, i) \in \{1, 0\}$ el tipo del i -ésimo individuo en la generación n y*

$$X_n^{(N)} = \frac{1}{M_n} \sum_{j=1}^{M_n} \mathbb{1}_{\{\tau(n,j)=0\}} \text{ y } C_{(n+1,i)} = i - \kappa \sum_{j=1}^{M_n} \mathbb{1}_{\{\tau(n,j)=0\}},$$

la frecuencia de individuos eficientes y el costo de producir los primeros i individuos en la generación $n + 1$ respectivamente, con las condiciones iniciales $M_0 = N_x = (1 - \kappa x)^{-1}N$, que es la solución de

$$N = (1 - \kappa)xN_x + (1 - x)N_x,$$

$$\text{y } X_0 = \lfloor xN_x \rfloor / N_x.$$

Los individuos de la generación $n + 1$ son creados recursivamente por una de las siguientes reglas. Si $C_{(n+1,i)} < N$ el i -ésimo individuo es creado y elige a su padre de la siguiente forma:

- Si $s = 0$, elige a un padre de la generación anterior de forma uniformemente al azar.
- Si $s > 0$, elige a un padre de tipo 0 con probabilidad

$$\frac{(1 - s)X_n^{(N)}}{1 - sX_n^{(N)}},$$

de otra forma elige un padre de tipo 1. En ambos casos el individuo copia el tipo de su padre.

Si $C_{(n+1,i)} \geq N$ entonces,

- **(M1)** $M_{n+1} = i$, es decir, no se crean más individuos.
- **(M2)** Si $C_{(n+1,i)} = N$, $M_{n+1} = i$, es decir, no se crean más individuos. Si $C_{(n+1,i)} > N$, $M_{n+1} = i - 1$ y el individuo $(n + 1, i)$ es descartado pues los recursos para producirlo son insuficientes.

La frecuencia de individuos eficientes y el tamaño total de la población están altamente relacionadas, cuando $X_n = 1$ el tamaño total de la población es $M_n = \lfloor N/(1 - \kappa) \rfloor + 1$ con la regla de paro **(M1)** y $M_n = \lfloor N/(1 - \kappa) \rfloor$ con la regla de paro **(M2)**, y cuando $X_n = 0$, entonces $M_n = N$. Las diferencias principales entre las dos reglas es que en **(M1)** siempre podemos crear individuos eficientes e ineficientes y en **(M2)** si los recursos disponibles son < 1 unidades solo podemos crear individuos eficientes, es decir, cuando la cantidad de recursos consumidos está en $(N - 1, N]$, crear individuos ineficientes en este caso no es posible bajo la regla de paro **(M2)**. En lo que resta de esta sección estudiaremos al modelo de Wright-Fisher con eficiencia como el modelo de la definición 6.1.1 bajo la regla **(M1)**, ya que estamos interesados en la relación de éste modelo y el modelo de Wright-Fisher con longevidad de la sección 8, ya que a las difusiones a las que converge cada uno de ellos son equivalentes. Resultados interesantes para el modelo de Wright-Fisher con eficiencia bajo la regla de paro **(M2)** pueden ser consultados en [6].

6.2 La difusión asociada al modelo de Wright-Fisher con eficiencia bajo la regla (M1)

Al igual que el modelo de Wright-Fisher neutro, con selección y con mutaciones, se obtiene un proceso límite cuando la población crece hacia infinito y cuando reescalamos el tiempo por N .

Teorema 6.2.1 (Teorema 1. en [6]). Sean $\kappa \in [0, 1]$ y $\alpha \geq 0$ fijos. Consideramos una sucesión de procesos $\{(X_{\lfloor Nt \rfloor}^{(N)}, t > 0), N \geq 1\}$, como en la definición anterior bajo la regla **(M1)** y neutro

(si $s_N = \alpha/N = 0$), o con selección si $s_N = \alpha/N > 0$, y tal que $X_0^{(N)}$ converge en distribución a $x \in (0, 1)$. Entonces la sucesión $\{(X_{[Nt]}^{(N)}, t > 0), N \geq 1\}$ converge en el sentido de los generadores a la solución de la siguiente EDE

$$dX_t = -\alpha X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t)(1 - \kappa X_t)}dBt, \quad (6.2.1)$$

donde B denota un movimiento Browniano estándar y con condición inicial $X_0 = x$.

A la solución de la EDE (6.2.1), que denotamos por $(X_t, t > 0)$, le llamaremos la *difusión de Wright-Fisher con eficiencia*. La EDE (6.2.1) es similar a la EDE de Wright-Fisher con selección (4.1.9), con el factor $(1 - \kappa X_t)$ en la varianza infinitesimal, al igual que antes, el tiempo fue reescalado por N y supusimos que $s_N = \alpha/N$, pero el coeficiente κ no fue necesario reescalarlo.

6.3 El proceso genealógico asociado.

Definamos ahora la gráfica que describe la genealogía asociada al modelo de Wright-Fisher con eficiencia bajo la regla de paro (M1).

Definición 6.3.1. Sea $n \in \mathbb{N}$, $\kappa \in [0, 1]$ y $\alpha \geq 0$ la gráfica aleatoria dirigida G_T^x de parámetros $T > 0$ y $x \in [0, 1]$ que llamamos la *Gráfica de Eficiencia/ Ancestría y Selección (Ancestral Selection/Efficiency Graph ASEG)* es un proceso de Markov a tiempo continuo que se construye de la siguiente forma. Sea Z_t el número de vértices activos al tiempo t y supongamos que $Z_0 = n$. Si $Z_t = j$ entonces,

- (Coalescencia) a tasa $j(j - 1)/2$ dos vértices activos escogidos uniformemente al azar se vuelven inactivos y producen un nuevo vértice activo el cual está conectado con ellos.
- (Ramificación) a tasa αj un vértice activo escogido uniformemente al azar se vuelve inactivo y produce dos nuevos vértices activos conectados a él.
- (Ramificación en pareja) a tasa $\kappa j(j - 1)/2$ un vértice activo escogido uniformemente al azar se vuelve inactivo y produce dos nuevos vértices activos conectados a él.
- (Coloreando los vértices extremos) al tiempo $T > 0$, el procedimiento se detiene y cada vértice activo obtiene un tipo, con probabilidad x es tipo 0 (eficiente) y con probabilidad $1 - x$ es tipo 1 (ineficiente).
- (Coloreando los vértices internos) cada vértice inactivo es de tipo 0 si y solo si no hay un camino dirigido que lo conecte con algún vértice de tipo 1.

El proceso de conteo de vértices activos de ASEG $Z = (Z_t, t > 0)$ es un proceso de nacimiento y muerte con transiciones

$$j \text{ a } \begin{cases} j + 1 & \text{a tasa } \alpha j + \kappa \frac{j(j-1)}{2}, \\ j - 1 & \text{a tasa } \frac{j(j-1)}{2}. \end{cases} \quad (6.3.2)$$

Lema 2. *La difusión de Wright-Fisher con eficiencia X definida como la única solución de 6.2.1 con parámetros $\alpha \geq 0$ y $\kappa \in [0, 1]$ y el proceso de conteo de vértices de ASEG Z definido anteriormente son momento duales, i.e. para toda $x \in [0, 1]$, $n \in \mathbb{N}$ y $t > 0$,*

$$E_x[X_t^n] = E_n[x^{Z_t}],$$

donde \mathbb{E}_x y \mathbb{E}_n son las esperanzas asociadas a los procesos X y Z empezando en x y n respectivamente.

El lema anterior nos muestra de forma analítica que el proceso ASEG es momento dual de la *difusión de Wright-Fisher con eficiencia*, recordamos que de manera similar la *difusión de Wright-Fisher* es momento dual con el *proceso de conteo de bloques del proceso de ancestría de Wright-Fisher*, eso nos hace pensar en un proceso que vaya hacia adelante en el tiempo y que converja a la *difusión de Wright-Fisher con eficiencia*, así como tenemos que el *modelo de Moran* reescalando el tiempo por $N/2$ converge a la *difusión de Wright-Fisher*, esta discusión la retomaremos en el capítulo siguiente.

Demostración. Denotamos por Q al generador del proceso de conteo de vértices de ASEG Z , el cual satisface, para una función f acotada con dominio en \mathbb{N} , que

$$\begin{aligned} Qf(x) &= \left(\alpha n + \frac{n(n-1)}{2}\kappa\right)(f(n+1) - f(n)) \\ &\quad + \frac{n(n-1)}{2}(f(n-1) - f(n)), \end{aligned}$$

y denotamos por A al generador infinitesimal de la difusión de Wright-Fisher con eficiencia, la única solución de la EDE 6.2.1, para cualquier función $C^2([0, 1])$. Consideramos la función $h(x, n) = x^n$. Ya que $h(x, n)$ y $\mathbb{E}[x^{Z_t}]$ son polinomios en x , y $h(x, n)$ y $E[X_t^n]$ son funciones acotadas de n , usando la Proposición 5.1.7 para probar la dualidad entre dichos procesos demostraremos que

$$Ah(x, n) = Qh(x, n).$$

Para cada $x \in [0, 1]$ y $n \in \mathbb{N}$, tenemos

$$\begin{aligned} Ah(x, n) &= -\alpha x(1-x)\frac{dh(x, n)}{dx} + x(1-x)(1-\kappa x)\frac{d^2h(x, n)}{dx^2} \\ &= -\alpha x(1-x)nx^{n-1} + x(1-x)(1-\kappa x)\frac{n(n-1)}{2}x^{n-2} \\ &= -\alpha nx^n + \alpha nx^{n+1} + \frac{n(n-1)}{2}x^{n-1}(1-\kappa x - x + \kappa x^2) \\ &= \alpha n(x^{n+1} - x^n) + \frac{n(n-1)}{2}(x^{n-1} - \kappa x^n - x^n + \kappa x^{n+1}) \\ &= \alpha n(x^{n+1} - x^n) + \frac{n(n-1)}{2}(x^{n-1} - x^n) + \frac{n(n-1)}{2}(\kappa x^{n+1} - \kappa x^n) \\ &= \left(\alpha n + \frac{n(n-1)}{2}\kappa\right)(x^{n+1} - x^n) + \frac{n(n-1)}{2}(x^{n-1} - x^n) \\ &= Qh(x, n). \end{aligned}$$

□

6.4 El generador del modelo de Wright-Fisher con eficiencia.

Lema 3. *En el modelo de Wright-Fisher con eficiencia con parámetros $s = 0$, $\kappa \in [0, 1]$ y regla de paro (M1), dado $X_n^{(N)} = x \in [0, 1]$ para cada $a \in (-1/2, 0)$, tenemos que*

$$\lim_{N \rightarrow \infty} \inf_{x \in (0, 1)} \mathbb{P} \left(\frac{M_{n+1}}{N_x} \in [1 - N^a, 1 + N^a] \right) = 1.$$

Demostración. Sea $\{B_i, i \geq 1\}$ una sucesión de variables aleatorias Bernoulli de parámetro x . Para cada $a \in (-1/2, 0)$, usando la definición de N_x se tiene que

$$\begin{aligned} \mathbb{P} \left(\frac{M_{n+1}^{(N)}}{N_x} \leq 1 - N^a \right) &= \mathbb{P} (C_{(n+1, \lfloor N_x(1-N^a) \rfloor})} \geq N) \\ &= \mathbb{P} \left(\lfloor N_x(1 - N^a) \rfloor - \kappa \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} B_i \geq N \right) \\ &\leq \mathbb{P} \left(N_x(\kappa x - N^a) \geq \kappa \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} B_i \right). \end{aligned}$$

Sumando y restando $\kappa x N_x N^a$, usando que $N = (1 - \kappa)x N_x + (1 - x)N_x$ y después sumando $\kappa x \lfloor N_x(1 - N^a) \rfloor$ entonces,

$$\begin{aligned} \mathbb{P} \left(\frac{M_{n+1}^{(N)}}{N_x} \leq 1 - N^a \right) &\leq \mathbb{P} \left(\kappa x N_x N^a - N^{1+a} \geq \kappa \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} B_i \right) \\ &= \mathbb{P} \left(-N^{1+a} + \kappa x (N_x(1 - N^a) - \lfloor N_x(1 - N^a) \rfloor) \right. \\ &\quad \left. \geq \kappa \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} B_i - \kappa x \lfloor N_x(1 - N^a) \rfloor B_i \right) \\ &\leq \mathbb{P} \left(-N^{1+a} + 1 \geq \kappa \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} (B_i - x) \right) \\ &\leq \mathbb{P} \left(\kappa \left| \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} (B_i - x) \right| > N^{1+a} + 2 \right). \end{aligned}$$

Ya que $B_i - x$ es una variable aleatoria centrada con varianza $\mathbb{V}ar[B_i] = x(1 - x)$, tenemos que

$$\begin{aligned} \mathbb{V}ar \left[\sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} (B_i - x) \right] &\leq \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} \mathbb{V}ar[B_i - x] \\ &= \frac{N}{1 - \kappa x} (1 + N^a) x(1 - x) \\ &\leq N(1 + N^a). \end{aligned}$$

En ambos casos, por la desigualdad de Chebychev obtenemos

$$\begin{aligned}
\lim_{N \rightarrow \infty} \sup_{x \in (0,1)} \mathbb{P} \left(\frac{M_{n+1}^{(N)}}{N_x} \notin [1 - N^a, 1 + N^a] \right) &\leq \lim_{N \rightarrow \infty} \sup_{x \in (0,1)} \mathbb{P} \left(\kappa \left| \sum_{i=1}^{\lfloor N_x(1-N^a) \rfloor} (B_i - x) \right| > N^{1+a} + 1 \right) \\
&\leq \lim_{N \rightarrow \infty} \frac{N(1 + N^a)}{(N^{1+a} + 1)^2} = \lim_{N \rightarrow \infty} N^{1-2(1+a)} = 0.
\end{aligned}$$

pues $1 - 2(a + 1) < 0$ ya que $a \in (-1/2, 0)$. □

Demostración del Teorema 6.2.1. Resultados clásicos para EDEs prueban que (6.2.1) tiene solución única, tema que se escapa de los resultados vistos en este trabajo, ver por ejemplo el Teorema 2 en González Casanova et al. (2017) [7]. Denotamos por A al generador infinitesimal de la difusión de Wright-Fisher con eficiencia. Su dominio contiene a las funciones continuas y dos veces diferenciables en $[0, 1]$, denotadas por $C^2([0, 1])$, y para cada $f \in C^2([0, 1])$ y para cada $x \in [0, 1]$ tenemos que

$$Af(x) = -\alpha x(1-x)f'(x) + x(1-x)(1-\kappa x)\frac{f''(x)}{2}. \quad (6.4.1)$$

Recordemos que, dado $X_0^{(N)} = x$ y $M_1 = m$, $mX_1^{(N)}$ sigue una distribución binomial de parámetros m y $(1-s_N)x/(1-s_Nx)$. usando la expansión de Taylor, como en el caso de modelo de Wright-Fisher con selección ejemplo 4.1.5, tenemos que

$$\begin{aligned}
\mathbb{N}\mathbb{E}[X_1^{(N)} - X_0^{(N)} | X_0^{(N)} = x, M_1 = m] &= -\alpha x(1-x) - o(1), \\
\mathbb{N}\mathbb{E}[(X_1^{(N)} - X_0^{(N)})^2 | X_0^{(N)} = x, M_1 = m] &= \frac{N}{m}x(1-x) + o(1),
\end{aligned}$$

con $N/m \geq 1$. Si tomamos $f \in C^2([0, 1])$, $x \in [0, 1]$ y $a \in (0, 1/2)$ y usando la proposición 3 el generador discreto de $X_{\lfloor Nt \rfloor}^{(N)}$ satisface

$$\begin{aligned}
A^N f(x) &= \mathbb{N}\mathbb{E} \left[f(X_1^{(N)}) - f(x) | X_0^{(N)} = x \right] \\
&= \mathbb{N}\mathbb{E} \left[X_1^{(N)} - x | X_0^{(N)} = x \right] f'(x) + \mathbb{N}\mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x \right] \frac{f''(x)}{2} + o(1) \\
&= \mathbb{N}\mathbb{E} \left[X_1^{(N)} - x | X_0^{(N)} = x \right] f'(x) + \mathbb{N}\mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x, M_1^N \in [N_x(1 - N^a), \right. \\
&\quad \left. N_x(1 + N^a)] \right] \frac{f''(x)}{2} + o(1) \\
&= -\alpha x(1-x)f'(x) + \frac{N}{\lfloor N_x \rfloor} x(1-x)f''(x) + o(1) \\
&= -\alpha x(1-x)f'(x) + x(1-x)(1-\kappa x)f''(x) + o(1) \\
&\xrightarrow{N \rightarrow \infty} Af(x).
\end{aligned}$$

donde el término $o(1)$ depende de x pero converge a 0 uniformemente en x . □

6.5 Probabilidades de fijación.

Lema 4. *Sea $(X_t, t \geq 0)$ la única solución de (6.2.1). Se cumple lo siguiente:*

- i) Si $\alpha \geq 0$ y $\kappa < 1$ los puntos 0 y 1 son accesibles.

ii) Si $\alpha \geq 0$ y $\kappa = 1$ el punto 1 no es accesible y el punto 0 es accesible.

Demostración. Recordamos que la difusión de Wright-Fisher con parámetro de selección $c \in \mathbb{R}$, $(\hat{Y}_t^{(c)}, t \geq 0)$ es la única solución de la EDE:

$$d\hat{Y}_t^{(c)} = c\hat{Y}_t^{(c)}(1 - \hat{Y}_t^{(c)})dt + \sqrt{\hat{Y}_t^{(c)}(1 - \hat{Y}_t^{(c)})}dB_t,$$

en donde B denota un movimiento Browniano estándar. En el caso $\alpha \geq 0$ y $\kappa \in (0, 1)$ utilizamos un argumento de dominancia estocástica. Consideramos la difusión $\bar{Y}^{(c)} = (\bar{Y}_t^{(c)}, t \geq 0)$, que se obtiene de $\hat{Y}^{(c)}$ al realizar un cambio estocástico de tiempo, es decir, para el reloj:

$$A_t = \int_0^t \frac{ds}{1 - \kappa\hat{Y}_s^{(c)}}, \text{ para } t \geq 0,$$

definimos $\bar{Y}_t^{(c)} = \hat{Y}_{\theta_t}^{(c)}$, para $t \geq 0$ en donde $\theta_t = \inf\{u : A_u > t\}$ es el inverso continuo por la derecha del reloj A . Observamos que satisface la siguiente EDE:

$$d\bar{Y}_t^{(c)} = c\bar{Y}_t^{(c)}(1 - \bar{Y}_t^{(c)})(1 - \kappa\bar{Y}_t^{(c)}) + \sqrt{\bar{Y}_t^{(c)}(1 - \bar{Y}_t^{(c)})(1 - \kappa\bar{Y}_t^{(c)})}d\beta_t$$

en donde $(\beta_t, t \geq 0)$ es un movimiento Browniano estándar. Para cada $t \geq 0$ tenemos que:

$$\theta_t = \int_0^t (1 - \kappa\bar{Y}_s^{(c)})ds \in [(1 - \kappa)t, t],$$

y ya que $\hat{Y}^{(c)}$ se fija en tiempo finito c.s. (ecuación (3.6) en Ewens(1963)) concluimos que $\bar{Y}_t^{(c)}$ se fija en tiempo finito. Es decir ambos estados absorbentes son accesibles para $\bar{Y}_t^{(c)}$ para cada $c \in \mathbb{R}$.

Finalmente, para $\alpha > 0$, tenemos que c.s

$$\bar{Y}_t^{(-\alpha)} \geq X_t \geq \bar{Y}_t^{(\alpha/(\kappa-1))},$$

y para $\alpha = 0$ tenemos $X \equiv \bar{Y}^{(0)}$, concluimos que los estados $\{0,1\}$ son accesibles para X . \square

Ahora denotamos $(Y_t, t > 0)$ al proceso de frecuencia de individuos ineficientes, es decir, $Y = 1 - X$. Considerando la definición de probabilidad de fijación, tenemos el siguiente resultado.

Proposición 6.3.1 (Probabilidad de fijación). *La probabilidad de fijación de los individuos ineficientes en la difusión de Wright-Fisher con eficiencia parametrizada por $\kappa \in (0, 1)$ y $\alpha \geq 0$ esta dada por*

$$\mathbb{P}_y[\{fix\}] = \begin{cases} 1 - C_{\alpha,\kappa}(1 - (1 - \kappa(1 - y))^{1 - \frac{2\alpha}{\kappa}}), & \text{si } 2\alpha \neq \kappa \\ 1 - C_{\alpha,\kappa} \ln\left(\frac{1}{1 - \kappa(1 - y)}\right) & \text{si } 2\alpha = \kappa, \end{cases}$$

en donde $C_{\alpha,\kappa}$ esta dada por

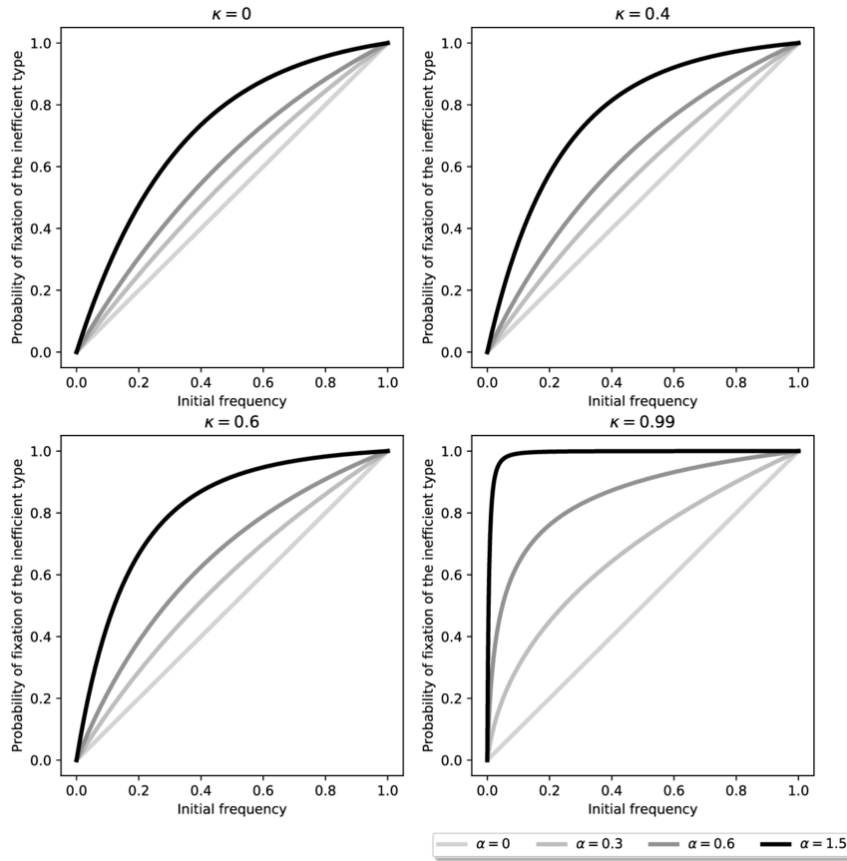


Figura 6.5: Probabilidades de fijación de individuos ineficientes, cuando $\kappa = 0$ es la gráfica correspondiente a un modelo de Wright-Fisher con selección y sin eficiencia. Adaptado de The Wright-Fisher model with efficiency (p. 8), por González Casanova A., Miró Pina V., Pardo J.C., 2020, Theor. Popul. Biol., pp. 33-46.

$$C_{\alpha, \kappa} = \begin{cases} 1 - (1 - \kappa)^{1 - \frac{2\alpha}{\kappa}} & \text{si } 2\alpha \neq \kappa, \\ \ln\left(\frac{1}{1 - \kappa}\right) & \text{si } 2\alpha = \kappa. \end{cases}$$

En la Figura 6.5 podemos observar que para un valor fijo de α (línea negra por ejemplo), la probabilidad de fijación de individuos ineficientes incrementa cuando κ incrementa. Es decir, que ser individuo ineficiente tiene una ventaja selectiva, ya que la probabilidad de fijación de los individuos ineficientes incrementa con κ , comparado con el modelo de Wright-Fisher con selección, en cuyo caso $\kappa = 0$. En el modelo de Wright-Fisher con eficiencia, mientras más alto sea el parámetro de selección α la probabilidad de fijación de los individuos ineficientes incrementa en comparación con el modelo de Wright-Fisher con selección, es decir, la eficiencia hace una diferencia entre los individuos de una población cuando el parámetro de selección es alto.

Demostración de la Proposición 6.3.1. Teniendo en cuenta el Lema 4, la probabilidad de fijación

de los individuos eficientes con desventaja selectiva está dada por

$$\begin{aligned}\mathbb{P}_x[\{fix\}] &= \mathbb{P}[X_{T_{0,1}} = 1 | X_0 = x] \\ &= \frac{S(0) - S(x)}{S(0) - S(1)},\end{aligned}$$

en donde S es la función de escala de la difusión de Wright-Fisher con eficiencia, la cual para $x \in [0, 1]$ esta dada por:

$$S(x) = \int_0^x \exp \left\{ - \int_\eta^v \frac{-2\alpha}{1 - \kappa u} du \right\} dv = \begin{cases} \hat{K} [1 - (1 - \kappa x)^{1 - \frac{2\alpha}{\kappa}}] & si \quad 2\alpha \neq \kappa \\ \hat{K} \ln \left(\frac{1}{1 - \kappa x} \right) & si \quad 2\alpha = \kappa, \end{cases}$$

con η cualquier número positivo y \hat{K} una constante que depende de (α, κ, η) . Así tenemos que la probabilidad de fijación de los individuos eficientes es

$$\mathbb{P}_x[\{fix\}] = \begin{cases} C_{\kappa, \alpha} [1 - (1 - \kappa x)^{1 - \frac{2\alpha}{\kappa}}] & si \quad 2\alpha \neq \kappa \\ C_{\kappa, \alpha} \ln \left(\frac{1}{1 - \kappa x} \right) & si \quad 2\alpha = \kappa. \end{cases}$$

$$C_{\kappa, \alpha} = \begin{cases} 1 - (1 - \kappa)^{1 - \frac{2\alpha}{\kappa}} & si \quad 2\alpha \neq \kappa \\ \ln \left(\frac{1}{1 - \kappa} \right) & si \quad 2\alpha = \kappa, \end{cases}$$

y así concluimos que la probabilidad de fijación de individuos ineficientes con desventaja selectiva es

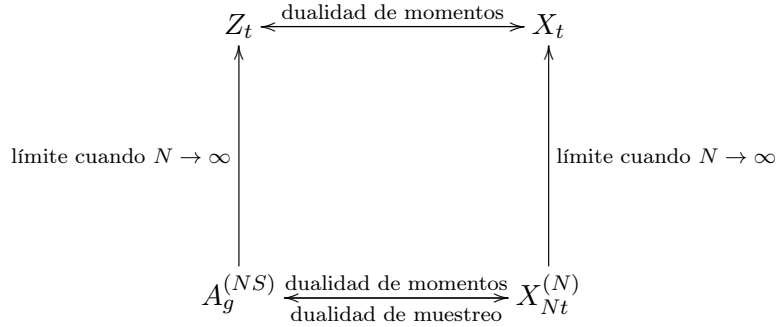
$$\begin{aligned}\mathbb{P}_y[\{fix\}] &= \mathbb{P}[Y_{T_{0,1}} = 1 | Y_0 = y] \\ &= \mathbb{P}[X_{T_{0,1}} = 0 | X_0 = 1 - y] \\ &= 1 - \mathbb{P}_x[\{fix\}] \\ &= 1 - \frac{S(0) - S(x)}{S(0) - S(1)} \\ &= 1 - \frac{S(0) - S(1 - y)}{S(0) - S(1)}.\end{aligned}$$

□

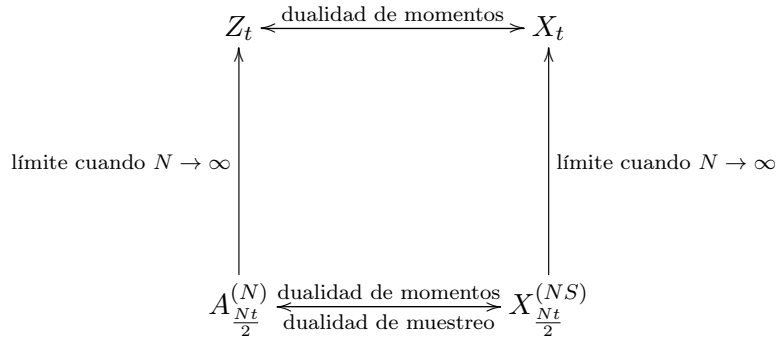
El modelo de Moran con selección y eficiencia.

En la gráfica de Wright-Fisher con una población constante de tamaño $N \in \mathbb{N}$, se definen dos proceso de Markov, $(A_g^{(NS)}, g \in \mathbb{Z}^+)$ el proceso de ancestría y el proceso de frecuencia $(X_g^{(N)})$,

$g \in \mathbb{Z}^+$), sabemos que ambos procesos son momento duales y muestreo duales (ver Proposición 5.2.1), además $(X_g^{(N)}, g \in \mathbb{Z}^+)$ convergen uniformemente a $(X_t, t > 0)$ la difusión de Wright-Fisher al reescalar el tiempo por N (ver ecuación 4.1.5) y $(A_g^{(NS)}, g \in \mathbb{Z}^+)$ converge al proceso de conteo de bloques del coalescente de Kingman $(Z_t, t > 0)$ (ver Proposición 5.1.5), finalmente $(X_t, t > 0)$ la difusión de Wright-Fisher y $(Z_t, t > 0)$ son momento duales (ver Proposición 5.1.8), lo cuál se representa en el siguiente diagrama:



El modelo de Moran, que es un modelo a tiempo continuo, también tiene asociado un proceso de frecuencia y uno proceso de ancestría, el primero converge a la difusión de Wright-Fisher (ver Ejemplo 4.1.7) y el segundo converge al coalescente de Kingman $(K_t, t \geq 0)$, por lo tanto también podemos construir el siguiente diagrama:



Consideramos ahora el modelo de Wright-Fisher con eficiencia y regla de paro **(M1)** de la Definición 6.1.1, el proceso de frecuencia converge en el sentido de los generadores a X , la única solución de la *EDE* (6.2.1), también tenemos que X y el proceso de conteo de vértices de *ASEG* Z son momento duales (Lema 2).

En esta sección definimos un modelo que describe la reproducción de una población con individuos eficientes e ineficientes a tiempo continuo y cuyo proceso de frecuencia converge a la difusión de Wright-Fisher con eficiencia, con el objetivo de completar un diagrama similar al anterior.

7.1 Modelo de Moran neutro con eficiencia

Comenzamos definiendo un modelo neutro, sin selección, en el cual solo consideramos eficiencia. Este modelo es una contribución nueva que resulta de intentar resolver el problema abierto que se plantea en [6] explicado anteriormente.

Definición 7.1.1 (Modelo de Moran neutro con eficiencia). Sean $N \in \mathbb{N}$ fijo y κ y $x \in [0, 1]$, en donde N es el tamaño de la población, x es la frecuencia de inicial de individuos eficientes al tiempo t_0 y κ el parámetro de eficiencia. Consideramos dos tipos de individuos, de tipo 0 (eficientes) y tipo 1 (ineficientes). Cada individuo se reproduce a tasa 1, y elige con quien reproducirse de manera aleatoria. El individuo elegido es reemplazado por un nuevo individuo, el cual copia el tipo del individuo que lo eligió con probabilidad $1 - \kappa x$ y con probabilidad κx conserva su tipo.

Para cada $t > 0$ definimos el proceso de frecuencia de individuos de tipo 0 del modelo Moran neutro con eficiencia $(X_t^{(N)}, t \geq 0)$ como sigue:

$$X_t^{(N)} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{\text{individuo } i \text{ es tipo } 0\}}.$$

- $(X_t^{(N)}, t \geq 0)$ es una cadena de Markov a tiempo continuo en $\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$.
- Las tasas de transición son:

$$\frac{i}{N} a \begin{cases} \frac{i+1}{N} & \text{a tasa } (N-i)\frac{i}{N}(1-\kappa\frac{i}{N}), \\ \frac{i-1}{N} & \text{a tasa } i\frac{(N-i)}{N}(1-\kappa\frac{i}{N}). \end{cases}$$

Proposición 7.1.2. Sean $N \in \mathbb{N}$ fijo y κ y $x \in [0, 1]$. Consideramos una sucesión de procesos de frecuencia del modelo Moran neutro con eficiencia $\{(X_{\lfloor Nt/2 \rfloor}^{(N)}, N \geq 0)\}$, como en la definición anterior, y tal que $X_0^{(N)}$ converge en distribución a $x \in (0, 1)$. Entonces la sucesión $\{(X_{\lfloor Nt/2 \rfloor}^{(N)}, N \geq 1)\}$ converge en el sentido de los generadores a la solución de la siguiente EDE

$$dX_t = \sqrt{Xt(1-X_t)(1-\kappa X_t)} dB_t. \quad (7.1.1)$$

Demostración. Sea $\lambda_x = \frac{N}{2}Nx(1-x)(1-\kappa x)$. El generador del proceso de frecuencia de individuos eficientes es

$$\begin{aligned} A^N f(x) &= \lambda_x \left[(f(x + \frac{1}{N}) - f(x)) + (f(x - \frac{1}{N}) - f(x)) \right] \\ &= \lambda_x \left[\left(f(x) + f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} - f(x) + o(\frac{1}{N^2}) \right) + \right. \\ &\quad \left. \left(f(x) - f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} - f(x) + o(\frac{1}{N^2}) \right) \right] \\ &= \lambda_x \left[f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} - f'(x)\frac{1}{N} + f''(x)\frac{1}{2N^2} + o(\frac{1}{N^2}) \right] \\ &= \lambda_x \left[f''(x)\frac{1}{N^2} + o(\frac{1}{N^2}) \right] \\ &= \frac{N^2}{2}x(1-x)(1-\kappa x) \left[f''(x)\frac{1}{N^2} + o(\frac{1}{N^2}) \right] \\ &= x(1-x)(1-\kappa x) \left[f''(x)\frac{1}{2} + o(1) \right], \end{aligned}$$

y así $\{(X_{\lfloor Nt/2 \rfloor}^{(N)}, t > 0), N \geq 1\}$ converge en el sentido de los generadores a la única solución de la EDE (7.1.1). \square

7.2 Modelo de Moran con selección y con eficiencia

Ahora incluimos selección en el modelo, sin pérdida de generalidad, consideramos que los individuos ineficientes tiene ventaja selectiva. Como se vio en la sección 6, la probabilidad de fijación de los individuos ineficientes en la difusión de Wright-Fisher con eficiencia **(M1)** incrementa cuando el parámetro de eficiencia κ incrementa, lo cual les aporta una ventaja sobre los individuos eficientes. En el modelo de Moran con selección y eficiencia al igual que en el modelo de Moran con selección, los eventos de reproducción suceden a tasa 1 para ambos tipos de individuos y a tasa $s > 0$ para los individuos de tipo 1. En un evento de reproducción a tasa 1, se eligen dos individuos para reproducirse, uno muere y el otro vive y se reproduce, el nuevo individuo producto de la reproducción reemplaza al individuo que muere, y copia el tipo del individuo que vive con probabilidad $(1 - \kappa x)$ y con probabilidad κx copia el tipo del individuo que muere. En un evento de reproducción a tasa s , se eligen dos individuos para reproducirse, si son de distinto tipo, el individuo de tipo 1 vive y se reproduce con probabilidad 1, el individuo nuevo es siempre de tipo 1. Si en un evento de reproducción ambos individuos son del mismo tipo, podemos decir que no pasa nada, ya que si importar que individuo muera, será reemplazado por uno de su mismo tipo, con lo que el número de individuos de ese tipo en la población no cambia. La forma de reproducción que sucede a tasa 1 la llamaremos reproducción *común* y reproducción que sucede a tasa $s > 0$ la llamaremos reproducción *especial*.

Fijamos $N \in \mathbb{N}$ y κ, s y $x \in [0, 1]$, como en el modelo de Wright-Fisher con eficiencia con dichos parámetros, en este caso N es el tamaño de la población, ya que para el modelo de Moran, la población es constante, $-s < 0$ el coeficiente de selección de los individuos *eficientes*, κ el parámetro de eficiencia y x la frecuencia inicial de individuos eficientes. Queremos definir el proceso de frecuencia de individuos eficientes $(X_t, t > 0)$ del modelo de Moran con eficiencia tal que sus tasas de transición sean las siguientes

$$\frac{i}{N} a \begin{cases} \frac{i+1}{N} & \text{a tasa } (N-i)\frac{i}{N}(1-\kappa\frac{i}{N}), \\ \frac{i-1}{N} & \text{a tasa } \frac{N-i}{N}i(1-\kappa\frac{i}{N}+s_N). \end{cases}$$

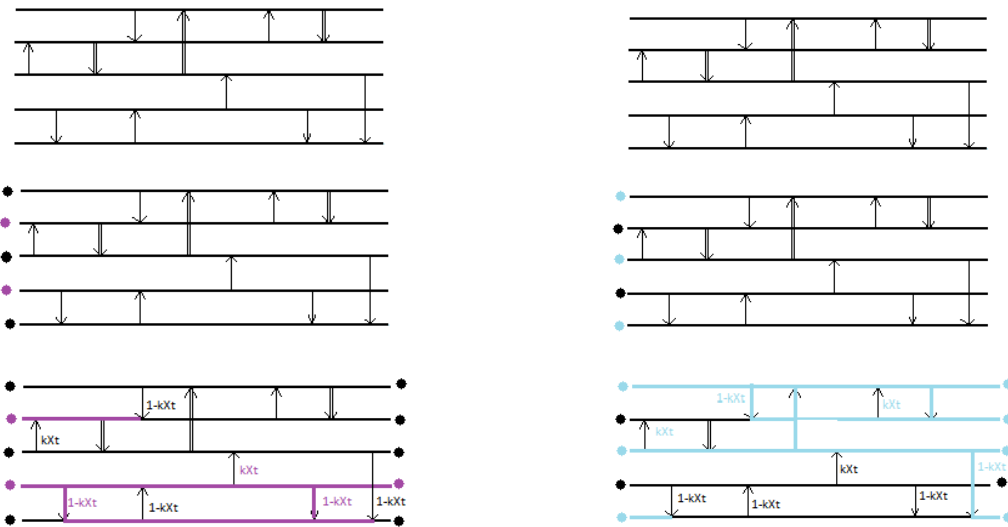
Definición 7.2.1 (Modelo de Moran con selección y eficiencia). *Sean $N \in \mathbb{N}$, $t_0 > 0$ y κ y $s \in [0, 1]$, en el modelo de Moran con selección y eficiencia con tamaño de población N , parámetro de selección s y parámetro de eficiencia κ , al tiempo t_0 elegimos una muestra de tamaño x de individuos eficientes. Los individuos en la población se reproducen de la siguiente forma:*

- i) Consideramos dos formas de reproducción, las comunes (\uparrow) que aparecen según un proceso Poisson de parámetro 1 y las reproducciones especiales ($\uparrow\uparrow$) que aparecen según un proceso Poisson de parámetro s . Los individuos eficientes solo cuentan con la forma de reproducción común e ignoran la reproducción especial, mientras que los individuos ineficientes pueden reproducirse de ambas formas.*
- ii) En un evento de reproducción común un par de individuos es elegido de manera aleatoria y con reemplazo, uno vive y se reproduce (se divide en dos) y otro muere y es reemplazado, con probabilidad $(1 - \kappa x)$ el individuo nuevo copia el tipo del individuo que vive, de lo contrario copia el tipo del individuo que muere.*
- iii) En un evento de reproducción especial un par de individuos es elegido de manera aleatoria*

y con reemplazo, el individuo ineficiente vive con probabilidad 1 y se reproduce, el nuevo individuo es de tipo 1.

A continuación explicamos los eventos de reproducción siguiendo la frecuencia de ambos tipos de individuos.

- i) En una población de tamaño constante N conformada por dos tipos de individuos, tipo 1 y tipo 0, consideramos dos formas de reproducción, reproducción *común* y reproducción *especial*, representadas por flechas, las simples (\uparrow) que aparecen según un proceso Poisson de parámetro 1 y que representan la reproducción *común*, y las dobles (\Uparrow) que aparecen según un proceso Poisson de parámetro $s > 0$ y que representan la reproducción *especial*.
- ii) Al tiempo $t_0 > 0$ coloreamos una muestra de tamaño $x \in [0, 1]$ de individuos eficientes con desventaja selectiva (tipo 0), las flechas dobles (\Uparrow) siempre son ignoradas, y las flechas simples (\uparrow) se toman en cuenta con probabilidad $(1 - \kappa x)$ y con probabilidad (κx) son ignoradas.
- ii') Al tiempo $t_0 > 0$ coloreamos una muestra de tamaño $1 - x \in [0, 1]$ de individuos ineficientes con ventaja selectiva (tipo 1), las flechas dobles (\Uparrow) siempre se toman en cuenta, y las flechas simples (\uparrow) se toman en cuenta con probabilidad $(1 - \kappa x)$ y con probabilidad (κx) son ignoradas.
- iii) Si el individuo es de tipo 1 (ineficiente con ventaja selectiva), se reproducen a tasa 1 con probabilidad $1 - \kappa(1 - x)$ y a tasa $s > 0$ con probabilidad 1. Si el individuo es de tipo 0 (eficiente con desventaja selectiva), se reproduce a tasa 1 con probabilidad $1 - \kappa x$ y a tasa s con probabilidad 0.



(a) ii) Proceso de reproducción de individuos tipo 0 (eficientes)

(b) ii') Proceso de reproducción de individuos tipo 1 (ineficientes)

Figura 7.6

En la Figura 7.6 se ilustran los pasos i), ii) para el proceso de reproducción de individuos eficientes, ii') para el proceso de reproducción de individuos de ineficientes, y iii) en el modelo de Moran con eficiencia y selección de la Definición 7.2.1. El paso iii) está representado

en la Figura 7.6 al colorear a los individuos de la población, los individuos eficientes de color morado solo pueden tomar en cuenta las flechas simples (\uparrow) con probabilidad $1 - \kappa X_t$ y siempre ignoran las flechas dobles ($\uparrow\uparrow$), a diferencia de los individuos ineficientes (en color azul) quienes además de tomar las flechas simples (\uparrow) con probabilidad $1 - \kappa X_t$, siempre toman las flechas dobles ($\uparrow\uparrow$) para reproducirse.

En esta Figura 7.6 también podemos observar que al intentar reconstruir la genealogía de los individuos, en ii) o en ii'), como lo hicimos en el modelo de Moran, recorriendo las flechas desde la punta a la raíz, necesitamos saber la frecuencia de individuos de tipo 0 para saber si la flecha se tomó o no en cuenta, lo cual nos impide definir un proceso de ancestría para el modelo de Moran con eficiencia y selección como en la Definición 7.2.1.

Para cada $t > 0$ definimos el proceso de frecuencia de individuos eficientes del modelo de Moran con eficiencia y selección como

$$X_t^{(N)} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{\text{individuo } i \text{ es tipo } 0\}}.$$

Proposición 7.2.2. Sean $N \in \mathbb{N}$ fijo y $\kappa, \alpha \vee x \in [0, 1]$. Consideramos una sucesión de procesos de frecuencia del modelo Moran con eficiencia y con parámetro de selección $s_N = \alpha/N > 0$, $\{(X_{\lfloor Nt/2 \rfloor}^{(N)}, t > 0), N \geq 0\}$, como en la definición anterior, y tal que $X_0^{(N)}$ converge en distribución a $x \in (0, 1)$. Entonces la sucesión $\{(X_{\lfloor Nt/2 \rfloor}^{(N)}, t > 0), N \geq 1\}$ converge en el sentido de los generadores a la solución de la EDE (6.2.1).

Demostración. Sean

$$\begin{aligned} \lambda_x^+ &= \frac{N}{2} N x (1-x) (1 - \kappa x) \\ \lambda_x^- &= \frac{N}{2} N x (1-x) (1 - \kappa x + s_N), \end{aligned}$$

las tasas de transición del modelo de Moran con selección y eficiencia reescalando el tiempo por $\frac{N}{2}$, entonces

$$\begin{aligned} \lambda_x^+ + \lambda_x^- &= \frac{N^2}{2} x (1-x) [2(1 - \kappa x) + s_N] \\ \lambda_x^+ - \lambda_x^- &= -s_N \frac{N^2}{2} x (1-x). \end{aligned}$$

Usando la expansión de Taylor tenemos que el generador del proceso de frecuencia del modelo de Moran con selección esta dado por

$$\begin{aligned}
A^N f(x) &= \lambda_x^+ \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + \lambda_x^- \left[f\left(x - \frac{1}{N}\right) - f(x) \right] \\
&= \lambda_x^+ \left[f(x) + f'(x) \frac{1}{N} + f''(y) \frac{1}{2N^2} - f(x) + o\left(\frac{1}{N^2}\right) \right] \\
&\quad + \lambda_x^- \left[f(x) - f'(x) \frac{1}{N} + f''(x) \frac{1}{2N^2} - f(x) + o\left(\frac{1}{N^2}\right) \right] \\
&= \frac{f'(x)}{N} (\lambda_x^+ - \lambda_x^-) + \frac{f''(x)}{2N^2} (\lambda_x^+ + \lambda_x^-) + o\left(\frac{1}{N^2}\right) (\lambda_x^+ + \lambda_x^-) \\
&= -\frac{N}{2} s_N x(1-x) f'(x) + \frac{f''(x)}{2} x(1-x) \left[(1 - \kappa x) + \frac{s_N}{2} \right] + o(1) \\
&= -\alpha x(1-x) f'(x) + \frac{f''(x)}{2} x(1-x)(1 - \kappa x) + o(1).
\end{aligned}$$

Por lo tanto $\{(X_{\lfloor Nt/2 \rfloor}^{(N)}, t > 0), N \geq 1\}$ converge en el sentido de los generadores a la solución de la *EDE* (6.2.1). \square

En conclusión el proceso de frecuencia del modelo Moran con selección y eficiencia de la Definición 7.1.1 converge a X , la difusión de Wright-Fisher con eficiencia, la única solución de la *EDE* (6.2.1), que es dual con el proceso de conteo de vértices de *ASEG* $Z = (Z_t, t > 0)$ de la Definición 6.3.1. Sin embargo, al movernos hacia atrás en el tiempo, este modelo no nos permite definir un proceso genealógico.

Modelo de Wright-Fisher con longevidad.

En una población de tamaño constante formada por individuos fértiles, aquellos que se pueden reproducir con parámetro de selección $s > 0$, e individuos infértiles aquellos que no se pueden reproducir, nos preguntamos si tener una tasa de envejecimiento (tasa de pasar a estado infértil) mayor a la de los demás individuos en la población representa una ventaja. En el modelo de Wright-Fisher con eficiencia, los individuos ineficientes, que consumen más recursos con parámetro de selección $s > 0$, tienen una ventaja a nivel de población, ya que la probabilidad de fijarse aumenta conforme aumenta el parámetro κ de eficiencia, por lo que consumir más recursos, en comparación con los individuos eficientes, les aporta una ventaja. En el modelo de Wright-Fisher con longevidad se puede ver un fenómeno similar, en el sentido que, los individuos que se vuelven infértiles más rápido, ocupan un lugar en la población que puede pensarse como ineficiente ya que estos individuos consumen recursos pero no contribuyen a la población fértil.

Definición 8.1.1 (La gráfica aleatoria de Wright-Fisher con longevidad). Sean $N \in \mathbb{N}$ fijo y $s, a_0, a_1, d_0, d_1 \in [0, 1]$. Consideramos un modelo de Wright-Fisher con población constante N y con dos tipos de individuos (tipo 0 y tipo 1), cada uno de los individuos está en dos posibles estados: fértil (f) o infértil (i). Los individuos fértiles de tipo j pasan al estado infértil con probabilidad $a_j \in [0, 1]$ (tasa de envejecimiento) y los individuos infértiles de tipo j mueren con probabilidad $d_j \in [0, 1]$ (tasa de mortalidad). Además, los individuos de tipo 0 tienen parámetro de selección $-s \in [0, 1]$. Definimos la gráfica aleatoria de Wright-Fisher con longevidad (GWFL) con conjunto de vértices $V = \mathbb{N} \times [N]$, en donde $[N] = \{1, 2, \dots, N\}$ y consideramos una generación como el subconjunto de vértices $V_g = [N] \times \{g\} \subset V$, el conjunto de aristas y la función de tipo $\tau : V \mapsto \{0, 1\} \times \{f, i\}$ son construidos recursivamente usando los siguientes pasos:

1. Se define la condición inicial: en la generación cero se asigna de manera aleatoria el tipo a cada individuo, i.e. fijamos el valor de $\tau(v)$ para cada $v \in V_0$.
2. Dado que $\tau(v)$ es conocido para todo $v \in V_{g-1}$, se divide V_{g-1} (aleatoriamente) en tres conjuntos, $D_g \subset V_{g-1}$ individuos muertos, $I_g \subset V_{g-1}$ individuos infértiles y $F_g \subset V_{g-1}$ individuos fértiles.
 - Si $v \in V_{g-1}$ es tal que $\tau(v) = (j, i)$, entonces $\mathbb{P}(v \in D_g) = 1 - \mathbb{P}(v \in I_g) = d_j$.
 - Si $v \in V_{g-1}$ es tal que $\tau(v) = (j, f)$, entonces $\mathbb{P}(v \in I_g) = 1 - \mathbb{P}(v \in F_g) = a_j$.
3. Si $(l, g-1) \in I_g$, entonces $(l, g-1)$ es el padre de (l, g) y (l, g) copia el tipo de su padre y permanece infértil. En otras palabras, si $(l, g-1) \in I_g$ ya sea que $\tau((l, g-1)) = (j, i)$ o $\tau((l, g-1)) = (j, f)$, entonces $\tau((l, g)) = (j, i)$.
4. Si $(l, g-1) \in D_g \cup F_g$, con $|D_g \cup F_g| = NF_g^{(N)}$, entonces (l, g) es un individuo fértil y escoge a su padre de F_g los individuos fértiles de la generación g y copia su tipo. Con probabilidad

$$\frac{(1-s)X_g^{(N)}}{1-sX_g^{(N)}}$$

el individuo elige un padre de tipo 0 y con probabilidad $1 - \frac{(1-s)X_g^{(N)}}{1-sX_g^{(N)}}$ el individuo elige

un padre de tipo 1, en donde

$$X_g^{(N)} = \frac{1}{NF_g^{(N)}} \sum_{k=1}^{NF_g^{(N)}} \mathbb{1}_{\{\tau((k,g))=(0,f)\}}$$

es la frecuencia de individuos de tipo 0 entre los individuos fértiles en la generación g .

8.1 Caso simple: $d_0 = d_1 = d$

Comenzamos con el caso en el que $d_0 = d_1 = d$ y $s = \alpha/N$, $\alpha \geq 0$. Observamos que si $d = 1$, todos los individuos infértiles de la generación g mueren antes de la formación de la generación $g + 1$, es decir, una vez que un individuo pasa al estado infértil, éste solo sobrevive una generación.

En este caso, consideramos dos variables aleatorias: $F_g^{(N)}$, la proporción de individuos fértiles en la generación g (de cualquier tipo), y $X_g^{(N)}$ la proporción de individuos fértiles en la generación g de tipo 0. El proceso $(X_g^{(N)}, F_g^{(N)}, g \geq 0)$ es markoviano cuyas tasas de transición pueden ser caracterizadas de la siguiente forma:

Dado que $X_g^{(N)} = x$ y $F_g^{(N)} = f$,

- $NF_{g+1}^{(N)} = N - B_g^{(0)} - B_g^{(1)} - B_g^{(d)}$, en donde $B_g^{(0)}$, $B_g^{(1)}$ y $B_g^{(d)}$ son variables aleatorias que cuentan el número de individuos fértiles en la generación g que se vuelven infértiles en la generación $g + 1$ y que tienen distribución binomial de parámetros (a_0, Nfx) , $(a_1, Nf(1-x))$ y $(1-d, N(1-f))$ respectivamente.
- $NF_{g+1}^{(N)} X_{g+1}^{(N)}$ tiene distribución binomial de parámetros $(\frac{(1-s_N)x}{1-s_N x}, NF_{g+1}^{(N)})$.

Para $x \in [0, 1]$ definimos

$$N_x := \frac{d}{1 + a_0 x + a_1(1-x) - (1-d)}$$

que es la solución de $N_x = 1 - a_0 x N_x - a_1(1-x)N_x - (1-d)(1-N_x)$. En donde N_x es la proporción de individuos fértiles dado que comenzamos con una proporción de x individuos fértiles de tipo 0.

Proposición 8.1.1. Sea $b \in (-1, 0)$, dado $X_g^{(N)} = x \in [0, 1]$, y suponiendo que $F_g^{(N)} = f \in [N_x(1 - N^b), N_x(1 + N^b)]$ tenemos que:

$$\lim_{N \rightarrow \infty} \inf_{x \in (0,1)} \mathbb{P} \left(\frac{F_{g+1}^{(N)}}{N_x} \in [1 - N^b, 1 + N^b] \right) = 1.$$

Demostración. Suponiendo lo contrario, sustituyendo el valor de N_x y usando que $NF_{g+1}^{(N)} = N - B_g^{(0)} - B_g^{(1)} - B_g^{(d)}$, llamamos $B = B_g^{(0)} + B_g^{(1)} + B_g^{(d)}$, entonces $NF_{g+1}^{(N)} = N - B$ y se tiene que

$$\begin{aligned}
\mathbb{P}\left(\frac{F_{g+1}^{(N)}}{N_x} \geq 1 + N^b\right) &= \mathbb{P}\left(NF_{g+1}^{(N)} \geq N(1 + N^b)N_x\right) \\
&= \mathbb{P}\left(NF_{g+1}^{(N)} \geq N(1 + N^b)[1 - a_0xN_x - a_1(1 - x)N_x - (1 - d)(1 - N_x)]\right) \\
&= \mathbb{P}\left(-dN^{b+1} \geq B - N(1 + N^b)N_x(a_0x + a_1(1 - x))\right. \\
&\quad \left. - N(1 - d)(1 - N_x(1 + N^b))\right),
\end{aligned}$$

aplicando valor absoluto obtenemos

$$\begin{aligned}
\mathbb{P}\left(NF_{g+1}^{(N)} \geq N(1 + N^b)N_x\right) &\leq \mathbb{P}\left(-dN^{b+1} \geq B - a_0xN(1 + N^b)N_x - a_1(1 - x)N(1 + N^b)N_x\right. \\
&\quad \left. - (1 - d)N(1 - (1 + N^b)N_x)\right) \\
&\leq \mathbb{P}\left(|B - a_0xN(1 + N^b)N_x - a_1(1 - x)N(1 + N^b)N_x\right. \\
&\quad \left. - (1 - d)N(1 - (1 + N^b)N_x)| > dN^{b+1}\right)
\end{aligned}$$

y finalmente por la desigualdad de Hoeffding concluimos que

$$\mathbb{P}\left(\frac{F_{g+1}^{(N)}}{N_x} \geq 1 + N^b\right) \leq 2 \exp\left(-\frac{2(dN^{b+1})^2}{3}\right).$$

De manera similar tenemos que:

$$\begin{aligned}
\mathbb{P}\left(\frac{F_{g+1}^{(N)}}{N_x} \leq 1 - N^b\right) &= \mathbb{P}\left(NF_{g+1}^{(N)} \leq NN_x(1 - N^b)\right) \\
&\leq \mathbb{P}\left(dN^{b+1} \leq B - a_0xN(1 - N^b)N_x - a_1(1 - x)N(1 - N^b)N_x\right. \\
&\quad \left. - (1 - d)N(1 - (1 - N^b)N_x)\right)
\end{aligned}$$

aplicando valor absoluto y usando la desigualdad de Hoeffding

$$\begin{aligned}
\mathbb{P}\left(\frac{F_{g+1}^{(N)}}{N_x} \leq 1 - N^b\right) &\leq \mathbb{P}\left(|B - a_0xN(1 - N^b)N_x - a_1(1 - x)N(1 - N^b)N_x\right. \\
&\quad \left. - (1 - d)N(1 - (1 - N^b)N_x)| > dN^{b+1}\right) \\
&\leq 2 \exp\left(-\frac{2(dN^{b+1})^2}{3}\right).
\end{aligned}$$

Por lo tanto tenemos que

$$\begin{aligned}
\lim_{N \rightarrow \infty} \inf_{x \in (0,1)} \mathbb{P} \left(\frac{F_{g+1}^{(N)}}{N_x} \in [1 - N^b, 1 + N^b] \right) &\leq \lim_{N \rightarrow \infty} \inf_{x \in (0,1)} \left(\exp \left(-\frac{2(dN^{b+1})^2}{3} \right) \right. \\
&\quad \left. + \exp \left(-\frac{2(dN^{b+1})^2}{3} \right) \right) \\
&\leq \lim_{N \rightarrow \infty} \exp \left(-\frac{2(dN^{b+1})^2}{3} \right) \\
&\quad + \exp \left(-\frac{2(dN^{b+1})^2}{3} \right) \\
&= 0,
\end{aligned}$$

pues $b \in (-1, 0)$. □

La proposición anterior muestra que $F_{g+1}^{(N)}$ fluctúa poco alrededor de N_x , (tiene fluctuaciones de orden N^b con $b \in (-1, 0)$), y solo sale de $[N_x(1 - N^b), N_x(1 + N^b)]$ con probabilidad exponencialmente pequeña, por lo que estudiaremos un modelo simplificado, es decir, en lugar de considerar el proceso $(X_g^{(N)}, F_g^{(N)}, g \geq 0)$ estudiaremos el proceso unidimensional $(X_g^{(N)}, g \geq 0)$ y dado que $X_0 = x$ suponemos que $F_{g+1}^{(N)} = N_x$, así tenemos que la cantidad de fértiles es determinista si conocemos $X_g^{(N)}$.

Definición 8.1.2 (El modelo de Wright-Fisher con longevidad). Sean $N \in \mathbb{N}$ fijo y $s, x \in [0, 1]$. En una población de tamaño N con dos tipos de individuos, tipo 1 y tipo 0, cada individuo esta en dos posibles estados, fértil o infértil, los individuos fértiles son capaces de reproducirse y al envejecer pasan al estado infértil. Definimos el proceso de frecuencia de individuos fértiles de tipo 0 como

$$X_g^{(N)} = \frac{1}{NF_g} \sum_{k=1}^{NF_g} \mathbb{1}_{\{\tau((k,g))=0\}}$$

con $\tau(n, j) \in \{0, 1\}$ el tipo del j -ésimo individuo en la generación n .

Las generaciones se construye de la siguiente manera

1. Se define la condición inicial: en la generación cero se asigna de manera aleatoria un tipo a cada individuo, con lo cual fijamos la frecuencia de individuos de tipo 0 $X_0^{(N)} = x$ y la frecuencia de individuos fértiles de tipo 0 es $F_1^{(N)} = N_x$.
2. Dado que $X_g^{(N)} = x', F_{g+1}^{(N)} = N_{x'}$.
3. Los $N_{x'}$ individuos fértiles de la generación $g + 1$ eligen al azar a un padre de entre los individuos fértiles de la siguiente forma:
 - si $s = 0$, cada individuo elige a su padre uniformemente al azar e independiente de los demás individuos.
 - si $s > 0$, cada individuo elige a su padre independientemente de los demás individuos, con probabilidad

$$\frac{(1-s)X_g^{(N)}}{1-sX_g^{(N)}}$$

el individuo elige un padre de tipo 0 y con probabilidad $1 - \frac{(1-s)X_g^{(N)}}{1-sX_g^{(N)}}$ el individuo elige un padre de tipo 1.

En la definición anterior la única variable aleatoria de interés es $X_g^{(N)}$ el número de individuos fértiles de tipo 0, ya que la cantidad de fértiles es determinista. Estudiaremos este modelo unidimensional, en lugar de el modelo de la definición 8.1.1, ya que podemos pensar que son muy parecidos pues la proporción de individuos fértiles permanece muy cerca de N_x según vemos en la proposición 8.1.1.

Al reescalar el tiempo por N , es decir, tomando N generaciones consecutivas como una sola, considerando el parámetro de selección $-s_N$ de los individuos de tipo 0 de orden $\frac{1}{N}$ y haciendo tender el tamaño de la población $N \in \mathbb{N}$ a infinito tenemos convergencia en el sentido de los generadores a la difusión de Wright-Fisher con eficiencia con parámetros $\hat{\alpha}$ y $\hat{\kappa}$ que dependen de a_0, a_1, d .

Teorema 8.1.3. Sean $a_0, a_1 \in (0, 1)$ fijos. Consideramos la sucesión de procesos $\{(X_{[Nt]}^{(N)}, t \geq 0)\}_{N \geq 1}$, definidos como arriba. Supongamos que $X_0^{(N)}$ converge en distribución a x y que $F_0^{(N)} = N_x$. Entonces la sucesión $\{(X_{[Nt]}^{(N)}, t \geq 0)\}_{N \geq 1}$ converge en el sentido de los generadores a la solución de la siguiente ecuación diferencial estocástica (EDE)

$$dX_t = -\alpha X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t) \frac{a_0 X_t + a_1(1 - X_t) + d}{d}} dB_t \quad (8.1.1)$$

en donde B denota un movimiento Browniano estándar y con condición inicial $X_0 = x$.

La única solución de la EDE (8.1.1) la llamamos la *difusión de Wright-Fisher con longevidad* y la denotamos como $(X_t, t \geq 0)$.

Prueba del Teorema 8.1.3. Recordemos que dado $X_0^{(N)} = x$ y $F_0^{(N)} = N_x$, los $NF_1^{(N)}$ individuos fértiles que hay en la generación 1 eligen a un padre de tipo 0 con probabilidad $\frac{(1-s_N)x}{1-s_Nx}$, con $0 < s_N = \alpha/N$, por lo tanto tenemos que

$$\begin{aligned} N\mathbb{E} \left[X_1^{(N)} - x \mid X_0^{(N)} = x, F_0^{(N)} = N_x \right] &= N \left[\frac{(1 - s_N)x}{1 - s_Nx} - x \right] \\ &= N[(1 - s_N)x(1 + s_Nx + o(s_N^2)) - x] \\ &= N[-s_Nx + s_Nx^2 - s_N^2x^2 + o(s_N^2)] \\ &= N[-s_Nx(1 - x) + o(s_N^2)] \\ &= -\alpha x(1 - x) + o(s_N^2), \end{aligned}$$

$$\begin{aligned}
N\mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x, F_0^{(N)} = N_x \right] &= N \left[\frac{1}{(NF_1^{(N)})^2} [NF_1^{(N)} \frac{(1-s_N)x}{1-s_Nx} (1 - \frac{(1-s_N)x}{1-s_Nx})] + o(s_N^2) \right] \\
&= \frac{1}{F_1^{(N)}} \left[\frac{(1-s_N)x}{1-s_Nx} (1 - \frac{(1-s_N)x}{1-s_Nx}) \right] + o(s_N^2) \\
&= \frac{1}{F_1^{(N)}} \left[\frac{(1-s_N)x(1-x)}{(1-s_Nx)^2} \right] + o(s_N^2)
\end{aligned}$$

usando que $\frac{1}{(1-x)^2} = 1 + 2x + O(x^2)$ tenemos que

$$\begin{aligned}
N\mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x, F_0^{(N)} = N_x \right] &= \frac{1}{F_1^{(N)}} \left[(1-s_N)x(1-x)(1+2s_Nx+o(s_N^2)) \right] \\
&\quad + o(s_N^2) \\
&= \frac{1}{F_1^{(N)}} \left[x(1-x) - s_Nx(1-x) + 2s_Nx \right. \\
&\quad \left. [x(1-x)(1-s_N)] + o(s_N^2) \right] \\
&= \frac{1}{F_1^{(N)}} \left[x(1-x) + o(s_N) + o(s_N^2) \right].
\end{aligned}$$

Tomamos $f \in C^2([0, 1])$, $x \in [0, 1]$, $b \in (-1, 0)$ y usando la expansión de Taylor y la Proposición 8.1.1, tenemos que

$$\begin{aligned}
\mathbb{E} \left[f(X_1^{(N)}) - f(x) | X_0^{(N)} = x \right] &= \mathbb{E} \left[X_1^{(N)} - x | X_0^{(N)} = x \right] f'(x) + \mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x, \right. \\
&\quad \left. F_1^N \in N_x(1-N^b), N_x(1+N^b) \right] f''(x) + o(1),
\end{aligned}$$

ya que

$$\mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x, F_1^N \notin [N_x(1-N^b), N_x(1+N^b)] \right] = 0$$

por la Proposición 8.1.1, por lo tanto el generador infinitesimal de $X_g^{(N)}$ es

$$\begin{aligned}
A^N f(x) &= N\mathbb{E} \left[f(X_1^{(N)}) - f(x) | X_0^{(N)} = x \right] \\
&= N\mathbb{E} \left[X_1^{(N)} - x | X_0^{(N)} = x \right] f'(x) + N\mathbb{E} \left[(X_1^{(N)} - x)^2 | X_0^{(N)} = x, F_1^N \in [N_x \right. \\
&\quad \left. (1-N^b), N_x(1+N^b)] \right] f''(x) + o(1) \\
&= -\alpha x(1-x) f'(x) + \left(\frac{1}{N_x} [x(1-x) + o(s_N) + o(s_N^2)] \right) f''(x) + o(1) \\
&= -\alpha x(1-x) f'(x) + \left(\frac{1+a_0x+a_1(1-x)-(1-d)}{d} \right) [x(1-x) \\
&\quad + o(s_N) + o(s_N^2)] f''(x) + o(1) \xrightarrow{N \rightarrow \infty} \mathcal{A}f(x),
\end{aligned}$$

en donde el término $o(1)$ depende de x pero cuando N tiende a infinito converge a 0 uniformemente en x , con lo cual el proceso de frecuencia del modelo de Wright-Fisher con

longevidad converge en el sentido de los generadores a la solución de la EDE 8.1.1, cuyo generador es \mathcal{A} . \square

La difusión de Wright-Fisher con longevidad, con $d_0 = d_1 = d$, resulta ser equivalente a la difusión de Wright-Fisher con selección.

Lema 5. Sean d , a_1 y $a_0 \in (0, 1)$ fijos, y sea $(X_t, t \geq 0)$ la difusión de Wright-Fisher con longevidad, solución de la EDE 8.1.1 y consideramos $Y_t = X_{Ct}$, entonces $(Y_t, t \geq 0)$ es solución de la EDE 6.2.1 con $\hat{\kappa} = \frac{a_1 - a_0}{a_1 + d}$, $\hat{\alpha} = \frac{\alpha}{C}$ y $C = \frac{a_1 + d}{d}$.

Demostración. Sea $C = \frac{a_1 + d}{d}$, $\hat{\kappa} = \frac{a_1 - a_0}{a_1 + d}$ y $\hat{\alpha} = \frac{\alpha}{C}$, en la ecuación 8.1.1 factorizamos C de la parte del de la varianza infinitesimal y multiplicamos y dividimos por C en la parte de la esperanza (*drift*) infinitesimal de la difusión.

$$\begin{aligned} dX_t &= -\alpha X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t) \frac{1 + a_0 X_t + a_1(1 - X_t) - (1 - d)}{d}} dB_t \\ &= -\alpha X_t(1 - X_t)dt + \sqrt{\frac{a_1 + d}{d} X_t(1 - X_t) \left(1 - \frac{a_1 - a_0}{a_1 + d} X_t\right)} dB_t \\ &= -\frac{\alpha}{C} C X_t(1 - X_t)dt + \sqrt{C X_t(1 - X_t)(1 - \hat{\kappa} X_t)} dB_t \\ &= -\hat{\alpha} C X_t(1 - X_t)dt + \sqrt{C X_t(1 - X_t)(1 - \hat{\kappa} X_t)} dB_t. \end{aligned}$$

La EDE anterior es la misma que 6.2.1 con parámetros $\hat{\alpha}$ y $\hat{\kappa}$, y multiplicada por C , entonces si avanzamos el tiempo en C unidades en el recorrido de X_t veremos el mismo recorrido que la difusión Y_t sigue en su tiempo natural, así tenemos que $X_{Ct} = Y_t$ es solución de la EDE:

$$dY_t = -\hat{\alpha} Y_t(1 - Y_t)dt + \sqrt{Y_t(1 - Y_t)(1 - \hat{\kappa} Y_t)} dB_t. \quad (8.1.2)$$

\square

Con el resultado anterior tenemos que la difusión de Wright-Fisher con eficiencia y la difusión de Wright-Fisher con longevidad son equivalentes. La primera, que es solución de la EDE 6.2.1, depende de α y κ y la segunda depende de $\hat{\alpha}$ y $\hat{\kappa}$, estos últimos parámetros están determinados por a_0 , a_1 y $d \in [0, 1]$, notamos que con las condiciones $\hat{\kappa} > 0$ y $a_0, a_1 \in [0, 1]$ se tiene que $a_1 > a_0$, es decir, para que haya un parámetro de eficiencia $\hat{\kappa}$, los individuos tipo 1 tienen que envejecer a una tasa mayor que la de los individuos tipo 0, y esto les aportará un ventaja al aumentar su probabilidad de fijación (comparado con el modelo clásico de Wright-Fisher con selección s/C). Además $\frac{\alpha d}{d + a_1} = \hat{\alpha} \leq \alpha$, por lo que esperamos que el efecto de selección que se obtenía para individuos de tipo ineficientes en el modelo de Wright-Fisher con eficiencia de parámetros α y κ sea menor, pues como se vió en la sección 6 en la figura 6.5, el parámetro de selección marca una diferencia en las probabilidades de fijación de los individuos ineficientes cuando el parámetro de selección α es alto. En la siguiente sección calculamos la probabilidad de fijación de los individuos fértiles de tipo 1 en el modelo de Wright-Fisher con longevidad.

8.2 Probabilidades de fijación

Una vez que todos los individuos fértiles en una generación son tipo 0 con desventaja selectiva, en las siguientes generaciones seguirán siendo todos de tipo 0, pues los individuos de la generación $g+1$ solo podrán elegir padres de tipo 0. En este caso la difusión de Wright-Fisher con longevidad ($X_t, t \geq 0$) se absorbe en 1, en caso contrario, cuando solo queden individuos fértiles de tipo 1, la difusión se absorbe en el estado 0. En esta sección calculamos la probabilidad de fijación de los individuos fértiles de tipo 1, la equivalencia entre el modelo de Wright-Fisher con eficiencia y el modelo de Wright-Fisher con longevidad del Lema 5, nos sugiere esperar un efecto de selección cuando $\hat{\kappa} > 0$, similar al que estudiamos en la Sección 6.5 para el modelo de Wright-Fisher con eficiencia, lo cual indicaría que envejecer más rápido podría significar un aumento en la probabilidad de fijación de los individuos tipo 1, sin embargo esto dependerá de la relación que tienen los parámetros $\hat{\kappa}$ y $\hat{\alpha}$ con a_1 , a_0 y d .

Lema 6. *Sea $(X_t, t \geq 0)$ la única solución de la EDE 8.1.1 se cumple lo siguiente:*

- i) *Si $\hat{\alpha} \geq 0$ y $\hat{\kappa} < 1$ los puntos 0 y 1 son accesibles.*
- ii) *Si $\hat{\alpha} \geq 0$ y $\hat{\kappa} = 1$ el punto 1 no es accesible y el punto 0 es accesible.*

Demostración. Recordamos que la difusión de Wright-Fisher con parámetro de selección $s \in \mathbb{R}$, $(Y_t^{(s)}, t \geq 0)$ es la única solución de la EDE:

$$dY_t^{(s)} = sY_t^{(s)}(1 - Y_t^{(s)})dt + \sqrt{Y_t^{(s)}(1 - Y_t^{(s)})}dB_t,$$

en donde B denota un movimiento Browniano estándar. En el caso $\hat{\alpha} \geq 0$ y $\hat{\kappa} \in (0, 1/2)$ utilizamos un argumento de dominancia estocástica. Consideramos la difusión $\bar{Y}^{(s)} = (\bar{Y}_t^{(s)}, t \geq 0)$, que se obtiene de $Y^{(s)}$ al realizar un cambio estocástico de tiempo, es decir, para el reloj:

$$A_t = \int_0^t \frac{ds}{(1 - \hat{\kappa})C}, \text{ para } t \geq 0,$$

definimos $\bar{Y}_t^{(s)} = Y_{\theta_t}^{(s)}$, para $t \geq 0$ en donde $\theta_t = \inf\{u : A_u > t\}$ es el inverso continuo por la derecha del reloj A . Observamos que satisface la siguiente EDE:

$$d\bar{Y}_t^{(s)} = s\bar{Y}_t^{(s)}(1 - \bar{Y}_t^{(s)}) + \sqrt{C\bar{Y}_t^{(s)}(1 - \bar{Y}_t^{(s)})(1 - \hat{\kappa}\bar{Y}_t^{(s)})}d\beta_t$$

en donde $(\beta_t, t \geq 0)$ es un movimiento Browniano estándar. Para cada $t \geq 0$ tenemos que:

$$\theta_t = \int_0^t C(1 - \hat{\kappa}\bar{Y}_s^{(s)})ds \in [(1 - \hat{\kappa})t, t],$$

y ya que $\hat{Y}^{(s)}$ se fija en tiempo finito c.s. (ecuación (3.6) en Ewens(1963)) concluimos que $\bar{Y}_t^{(s)}$ se fija en tiempo finito. Es decir ambos estados absorbentes son accesibles para $\bar{Y}_t^{(s)}$ para cada $S \in \mathbb{R}$. Finalmente, para $\hat{\alpha} > 0$, tenemos que casi seguramente

$$\bar{Y}_t^{(-\hat{\alpha})} \geq X_t \geq \bar{Y}_t^{(-\hat{\alpha} \wedge (\hat{\kappa}-1))},$$

y para $\hat{\alpha} = 0$ tenemos $X \equiv \bar{Y}^{(0)}$, concluimos que los estados $\{0,1\}$ son accesibles para X . \square

Proposición 8.2.1. *La probabilidad de fijación de los individuos fértiles de tipo 0 con desventaja selectiva está dada por:*

$$\mathbb{P}_x[\{fix\}] = \begin{cases} C_{\hat{\kappa}, \hat{\alpha}} [1 - (1 - \hat{\kappa}x)^{1 - \frac{2\hat{\alpha}}{\hat{\kappa}}}] & \text{si } 2\hat{\alpha} \neq \hat{\kappa} \\ C_{\hat{\kappa}, \hat{\alpha}} \ln\left(\frac{1}{1 - \hat{\kappa}x}\right) & \text{si } 2\hat{\alpha} = \hat{\kappa}, \end{cases}$$

en donde la constante $C_{\hat{\kappa}, \hat{\alpha}} = S(1)^{-1}$ y S la función de escala de la difusión de Wright-Fisher con longevidad.

Demostración.

$$\begin{aligned} \mathbb{P}_x[\{fix\}] &= \mathbb{P}[X_{T_{0,1}} = 1 | X_0 = x] \\ &= \frac{S(x) - S(0)}{S(1) - S(0)}, \end{aligned}$$

en donde S es la función de escala de la difusión de Wright-Fisher con longevidad, la cual para $x \in [0, 1]$ esta dada por:

$$\begin{aligned} S(x) &= \int_0^x \exp\left\{-\int_\eta^v \frac{-2\alpha(1+d)}{1+a_0u+a_1(1-u)+d} du\right\} dv \\ &= \begin{cases} K[1 - (1 - \hat{\kappa}x)^{1 - \frac{2\hat{\alpha}}{\hat{\kappa}}}] & \text{si } 2\hat{\alpha} \neq \hat{\kappa} \\ K \ln\left(\frac{1}{1 - \hat{\kappa}x}\right) & \text{si } 2\hat{\alpha} = \hat{\kappa}, \end{cases} \end{aligned}$$

con η cualquier número positivo y K una constante que depende de $(\hat{\alpha}, \hat{\kappa}, \eta)$. Así concluimos que:

$$\mathbb{P}_x[\{fix\}] = \begin{cases} C_{\hat{\kappa}, \hat{\alpha}} [1 - (1 - \hat{\kappa}x)^{1 - \frac{2\hat{\alpha}}{\hat{\kappa}}}] & \text{si } 2\hat{\alpha} \neq \hat{\kappa} \\ C_{\hat{\kappa}, \hat{\alpha}} \ln\left(\frac{1}{1 - \hat{\kappa}x}\right) & \text{si } 2\hat{\alpha} = \hat{\kappa}. \end{cases}$$

$$C_{\hat{\kappa}, \hat{\alpha}} = \begin{cases} 1 - (1 - \hat{\kappa})^{1 - \frac{2\hat{\alpha}}{\hat{\kappa}}} & \text{si } 2\hat{\alpha} \neq \hat{\kappa} \\ \ln\left(\frac{1}{1 - \hat{\kappa}}\right) & \text{si } 2\hat{\alpha} = \hat{\kappa}. \end{cases}$$

\square

Si denotamos $(Y_t, t \geq 0)$ como el proceso de frecuencia de los individuos fértiles de tipo 1, entonces $Y = 1 - X$ y así la probabilidad de fijación de este proceso está dada por

$$\mathbb{P}_y[\{fix\}] = \begin{cases} 1 - C_{\hat{\kappa}, \hat{\alpha}} [1 - (1 - \hat{\kappa}x)^{1 - \frac{2\hat{\alpha}}{\hat{\kappa}}}] & si \quad 2\hat{\alpha} \neq \hat{\kappa} \\ 1 - C_{\hat{\kappa}, \hat{\alpha}} \ln\left(\frac{1}{1 - \hat{\kappa}x}\right) & si \quad 2\hat{\alpha} = \hat{\kappa}. \end{cases}$$

En las figuras 8.7 y 8.8 observamos que la probabilidad de fijación de los individuos fértiles de tipo 1 con ventaja selectiva en el modelo de Wright-Fisher con longevidad, al igual que en el modelo de Wright-Fisher con eficiencia, es mayor en comparación al modelo de Wright-Fisher con longevidad cuando $\hat{\kappa} = 0$, en la Figura 8.7 para un valor fijo de $\hat{\alpha}$ la probabilidad de fijación de individuos fértiles de tipo 1 aumenta al aumentar $\hat{\kappa}$, y en la Figura 8.8 corroboramos que esta diferencia en las probabilidades de fijación es mayor cuando $\hat{\alpha}$ es mayor.

Por lo tanto, cuando fijamos $\hat{\kappa} = \frac{a_1 - a_0}{d + a_1}$ y $\hat{\alpha} = \frac{\alpha d}{d + a_1}$, la probabilidad de fijación de los individuos fértiles de tipo 1 con ventaja selectiva incrementa en comparación con el modelo clásico de Wright-Fisher con el mismo parámetro de selección. Para observar este efecto en una población finita de tamaño N , recordamos que el parámetro de selección del modelo debe ser pequeño ($s_N = \hat{\alpha}/N$) en comparación con el parámetro $\hat{\kappa}$.

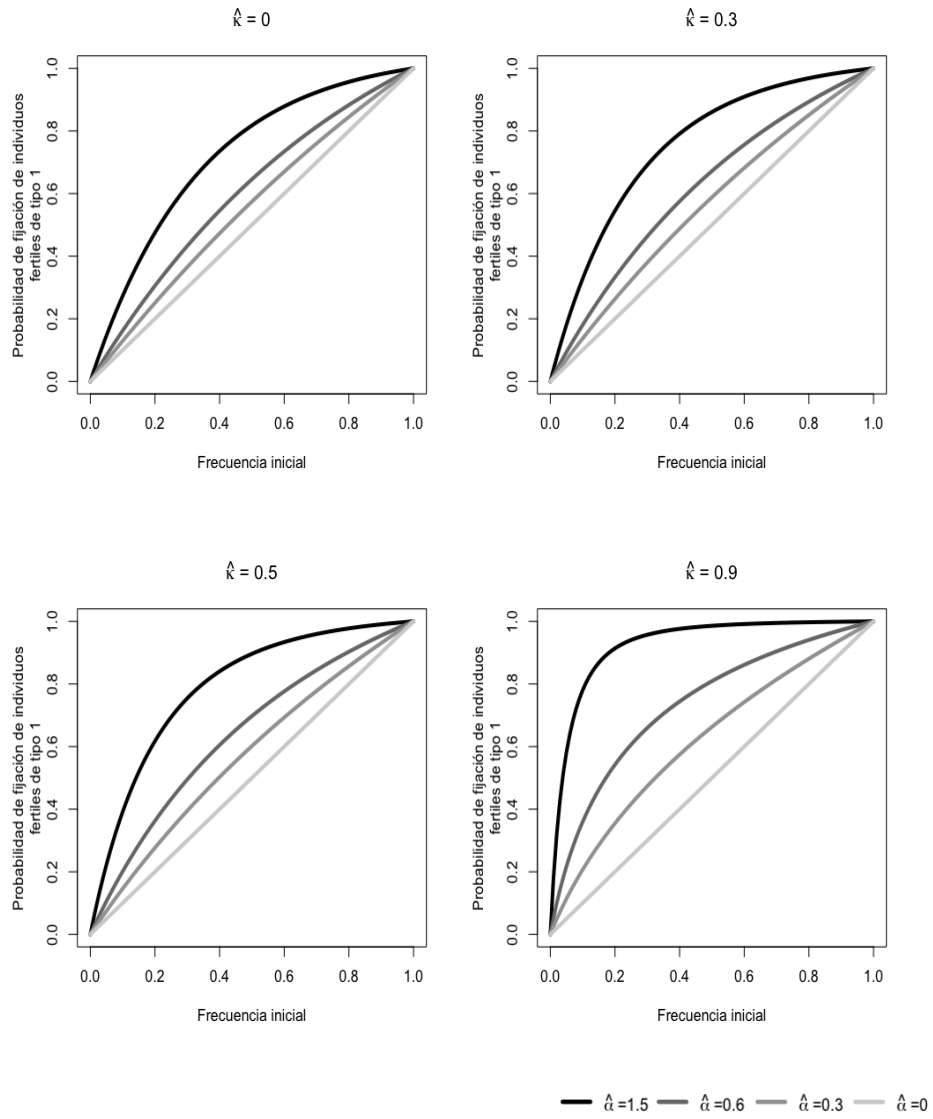


Figura 8.7: Probabilidad de fijación de individuos fértiles tipo 1 en el modelo de Wright-Fisher con longevidad fijando $\hat{\kappa}$.

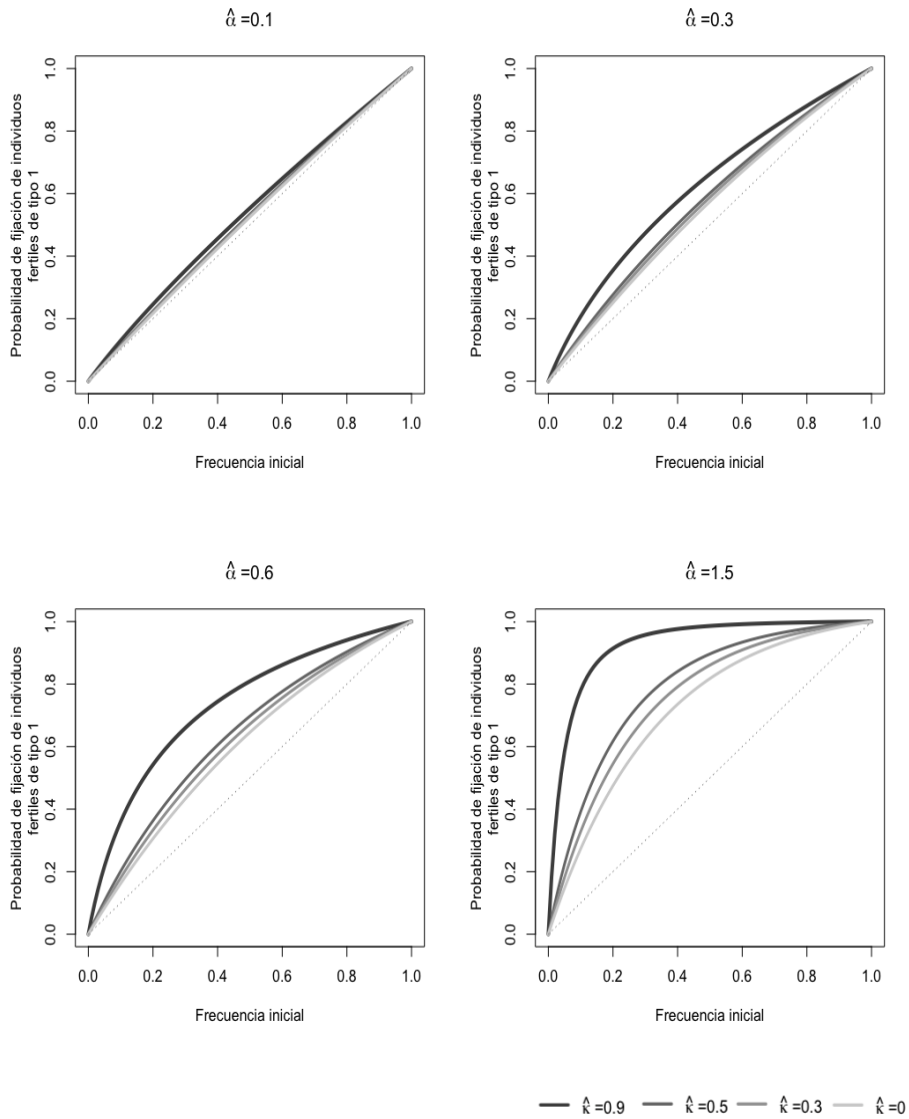


Figura 8.8: Probabilidad de fijación de individuos fértiles tipo 1 en el modelo de Wright-Fisher con longevidad fijando $\hat{\alpha}$.

Los parámetros $\hat{\kappa}$ y $\hat{\alpha}$ están inversamente relacionados, en las figuras 8.9 y 8.10 observamos la relación entre $\hat{\kappa}$ y $\hat{\alpha}$ con los parámetros α , a_0 , a_1 y d , al aumentar los parámetros a_1 y d , $\hat{\kappa}$ incrementa mientras que $\hat{\alpha}$ disminuye, además necesitamos que $a_1 > a_0$ para que el parámetro $\hat{\kappa}$ sea positivo. Por lo tanto, debemos elegir ciertos valores de a_0 , a_1 y d para incrementar $\hat{\alpha}$ y $\hat{\kappa}$ y así obtener un efecto de selección para los individuos de tipo 1 al incrementar su probabilidad de fijación en comparación con un modelo clásico de Wright-Fisher con selección de parámetro s/C .

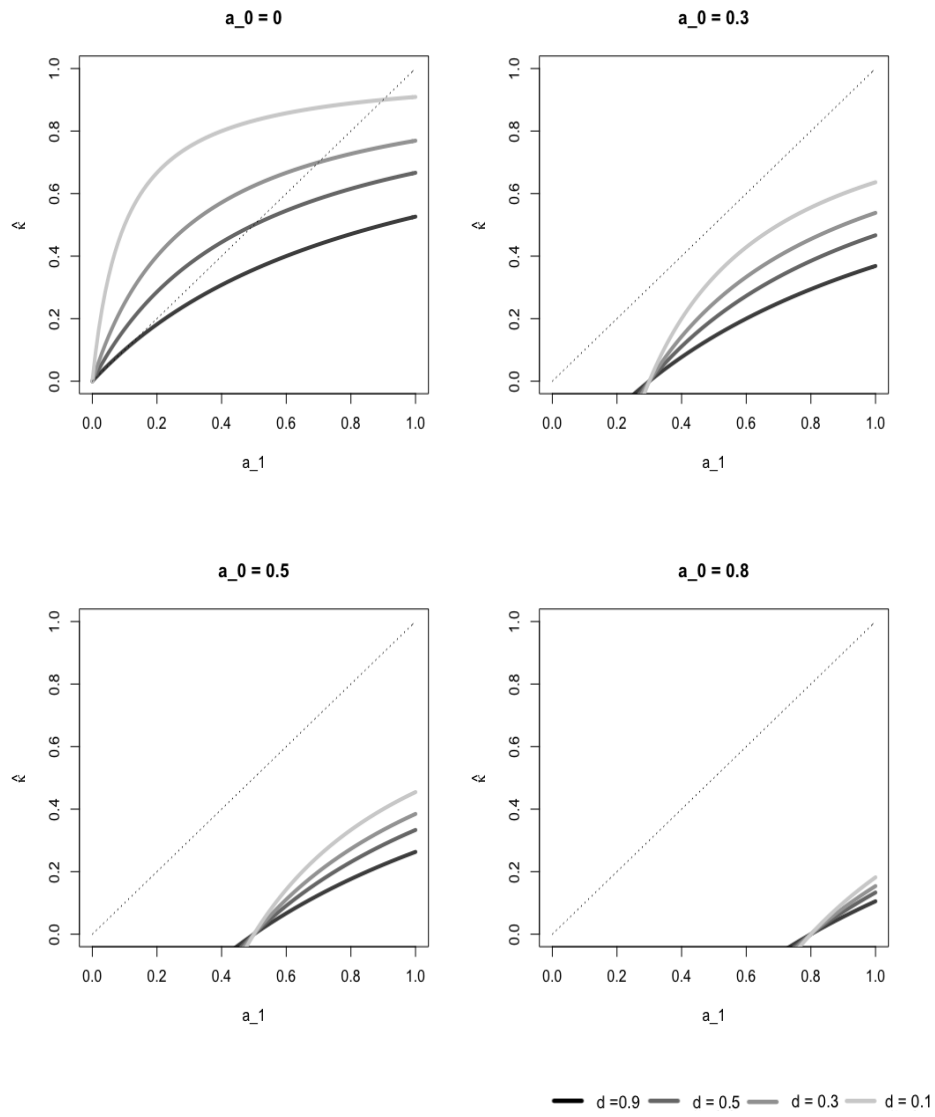


Figura 8.9: Gráficas del parámetro $\hat{\kappa}$ dependiendo de la tasa de envejecimiento a_1 fijando a_0 .

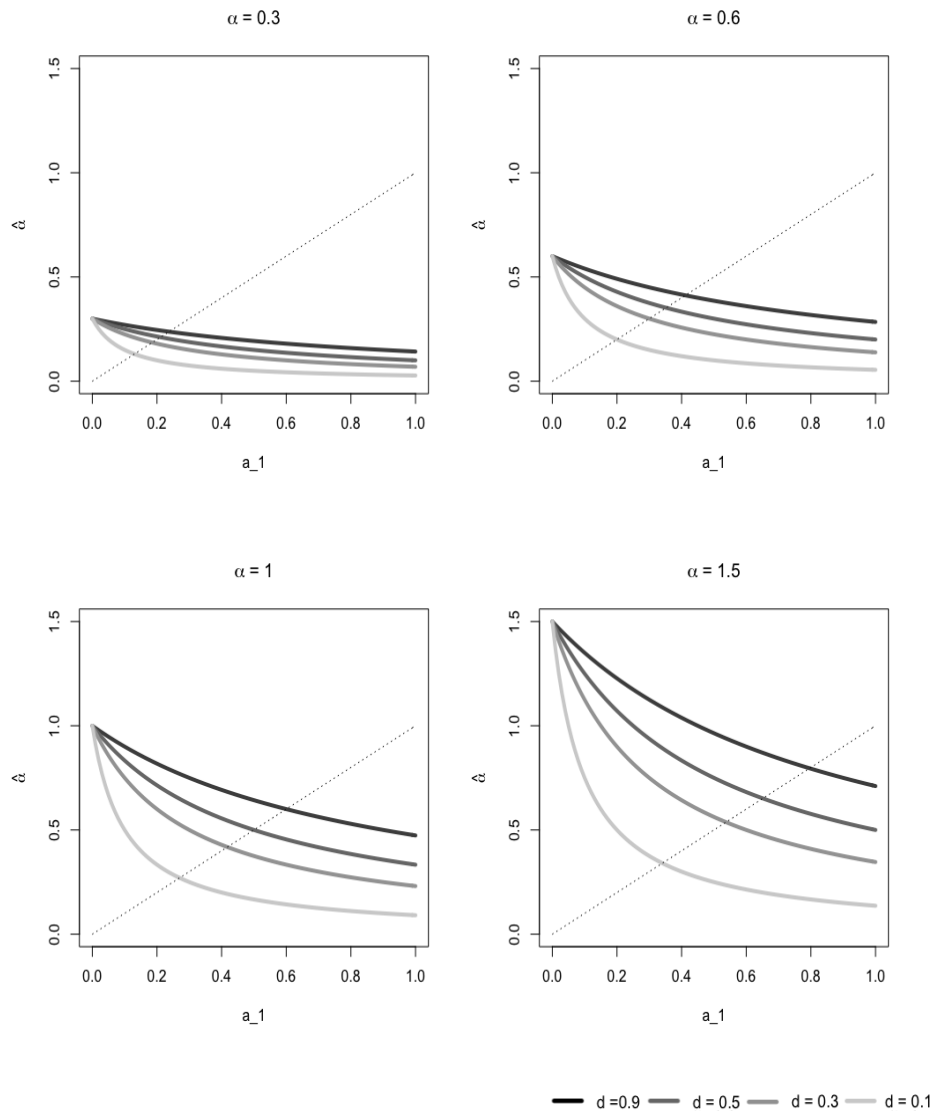


Figura 8.10: Gráficas del parámetro $\hat{\alpha}$ dependiendo de la tasa de envejecimiento a_1 fijando α .

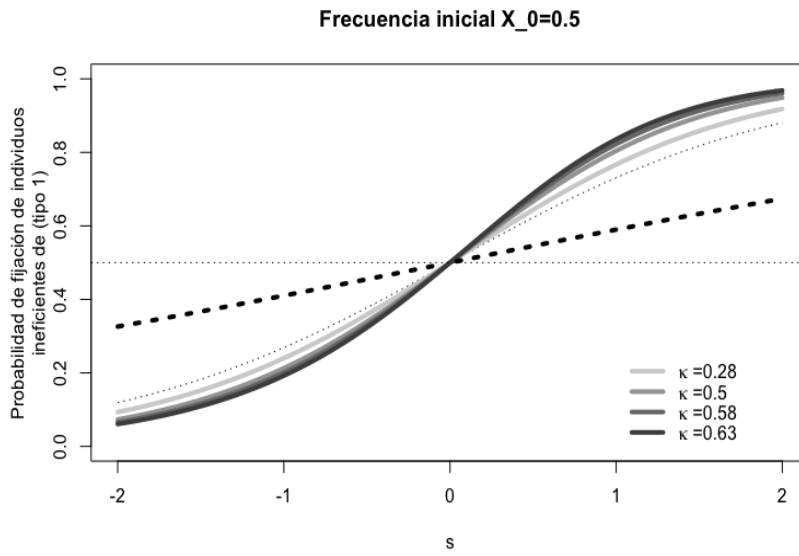


Figura 8.11: Probabilidades de fijación de individuos ineficientes en el modelo de Wright-Fisher con eficiencia con parámetros s y κ al fijar una frecuencia inicial de 0.5.

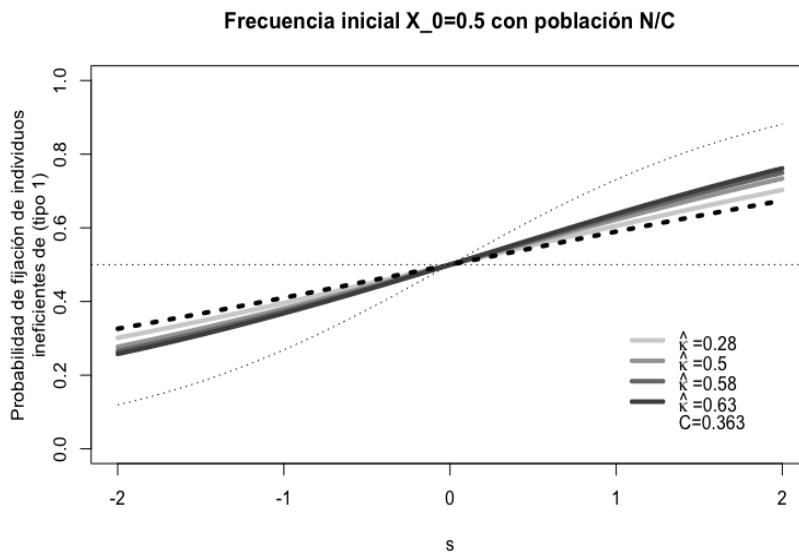


Figura 8.12: Probabilidades de fijación de individuos ineficientes en el modelo de Wright-Fisher con eficiencia con parámetros s/C y κ al fijar una frecuencia inicial de 0.5.

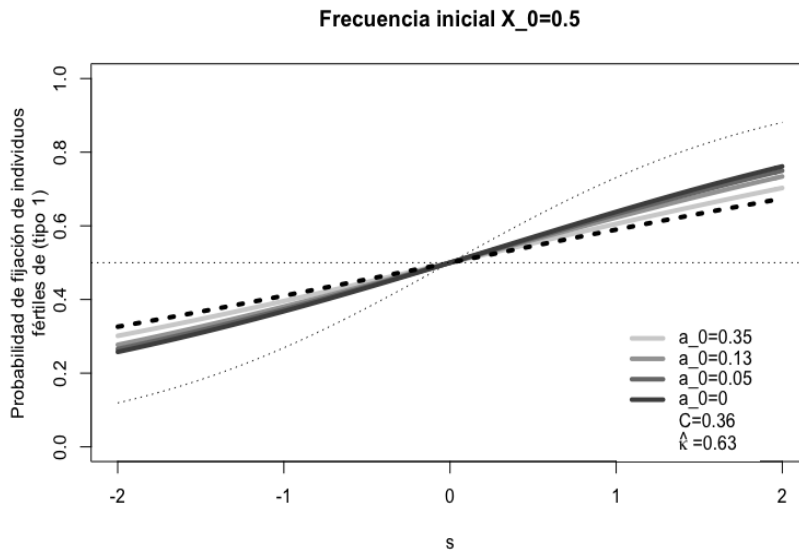


Figura 8.13: Probabilidades de fijación de individuos fértiles de tipo 1 en el modelo de Wright-Fisher con longevidad con parámetro de selección s al fijar C (a_1 y d) y aumentar $\hat{\kappa}$ (disminuir a_0) y al fijar una frecuencia inicial de 0.5.

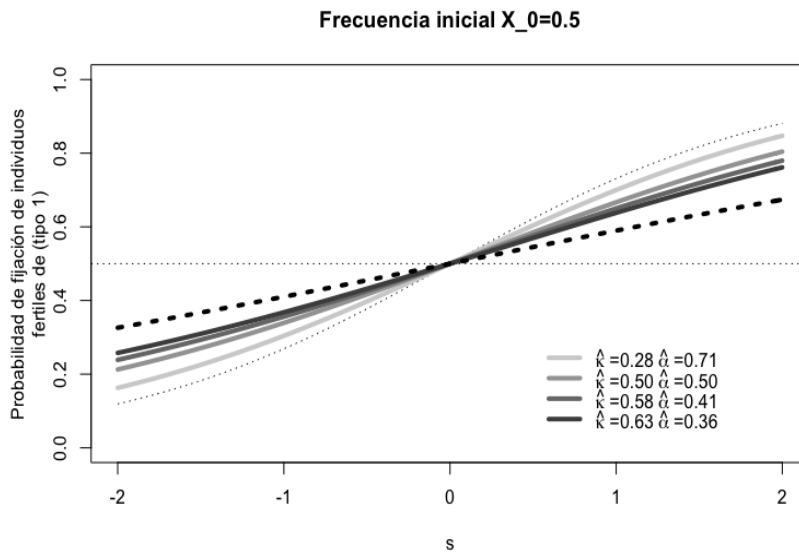


Figura 8.14: Probabilidades de fijación de individuos fértiles de tipo 1 en el modelo de Wright-Fisher con longevidad con parámetro de selección s al fijar una frecuencia inicial de 0.5

Finalmente comparamos las probabilidades de fijación entre el modelo clásico de Wright-Fisher con selección s , el modelo de Wright-Fisher con longevidad de parámetros s/C y $\hat{\kappa} = 0$ (modelo clásico de Wright-Fisher con selección s/C) y el modelo de Wright-Fisher con longevidad de parámetros a_0 , a_1 y $d \in [0, 1]$ (modelo de Wright-Fisher con eficiencia de parámetros s/C y $\hat{\kappa} > 0$). El modelo clásico de Wright-Fisher con selección s/C corresponde a el modelo clásico de Wright-Fisher con selección s con una población N/C en lugar de

N .

En el modelo de Wright-Fisher con longevidad de parámetros s y κ las probabilidades de fijación disminuyen, en comparación con el modelo clásico de Wright-Fisher con selección s , cuando fijamos s/C y $\hat{\kappa}$ decrece (ver Figura 8.13), equivalentemente en el modelo de Wright-Fisher con longevidad de parámetros s y κ las probabilidades de fijación aumentan, en comparación con el modelo clásico de Wright-Fisher con selección s/C , cuando fijamos s/C y $\hat{\kappa}$ crece, como sucedería en el modelo de Wright-Fisher con eficiencia de parámetros s/C y $\hat{\kappa}$ (ver Figura 8.11).

Por otro lado, cuando ambos parámetros ($\hat{\kappa}$ y s/C) varían a consecuencia de su relación con d y a_1 , tenemos que las probabilidades de fijación disminuyen en comparación con el modelo clásico de Wright-Fisher con selección s , cuando $\hat{\kappa}$ crece (y al mismo tiempo s/C decrece) (ver Figura 8.14), equivalentemente en el modelo de Wright-Fisher con eficiencia de parámetros s/C y κ las probabilidades de fijación aumentan, en comparación con el modelo clásico de Wright-Fisher con selección s/C , cuando $\hat{\kappa}$ decrece (y al mismo tiempo s/C crece), esto debido a la relación inversa que tienen al depender de los parámetros d y a_1 .

Es decir, aun cuando el modelo de Wright-Fisher con longevidad de parámetros a_0 , a_1 y $d \in [0, 1]$ puede verse como el modelo de Wright-Fisher con eficiencia con parámetros $\hat{\alpha} = s_N/C$ (cuando el parámetro de selección es de orden $s_N = \alpha/N$) y $\hat{\kappa}$, la diferencia está en que los parámetros $\hat{\alpha}$ y $\hat{\kappa}$ dependen inversamente de d y a_1 , por lo que la probabilidad de fijación de individuos de tipo 1 (fértiles con ventaja selectiva) disminuye cuando $\hat{\kappa}$ crece y $\hat{\alpha}$ decrece al mismo tiempo (ver Figura 8.14), en comparación con el modelo clásico de Wright-Fisher con parámetro de selección disminuido s/C .

Pero si fijamos s/C y $\hat{\kappa}$ crece (ver Figura 8.13) (al fijar s , a_1 y d y disminuir a_0), la probabilidad de fijación de individuos de tipo 1 (fértiles con ventaja selectiva) aumenta en comparación con el modelo clásico de Wright-Fisher con parámetro de selección s/C , tal como un modelo de Wright-Fisher con eficiencia de parámetro $\hat{\alpha} = s_N/C$ y $\hat{\kappa}$.

En conclusión la longevidad tiene dos efectos opuestos, por un lado hace que el efecto de selección disminuya ($\hat{\alpha} < \alpha$) y por otro lado gracias al parámetro $\hat{\kappa}$, el modelo tiene un efecto benéfico para los individuos de tipo 1 al aumentar sus probabilidades de fijación, el segundo efecto es similar al del modelo de Wright-Fisher con eficiencia.

Bibliografía

- [1] Durrett Richard., 2008 Probability Models for DNA Sequence Evolution, Springer.
- [2] Etheridge A., 2011. Some mathematical models from population genetics. École d'Été de Probabilités de Saint-Flour XXXIX-2009, Lecture Notes in Mathematics, Springer.
- [3] Feller, W., 1951. Diffusion processes in genetics. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. 277-246.
- [4] Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Clarendon Press. Dover, New York.
- [5] Gallager Robert G., 2013. Stochastic Processes theory for applications. Cambridge University Press. 527-533.
- [6] González Casanova A., Miró Pina V., Pardo J.C., 2020. The Wright-Fisher model with efficiency. Theor. Popul. Biol., pp. 33-46.
- [7] González Casanova A., Pardo J.C., Pérez J. L., 2017. Branching processes with interactions: the subcritical cooperative regime, Arxiv. 1704.04203
- [8] Huillet Thierry., 2007. On Wright-Fisher diffusion and its relatives. Journal of Statistical Mechanics: Theory and Experiment, IOP Publishing, vol 11.
- [9] Jansen, S., Kurt, N., 2014. On the notion(s) of duality of Markov Processes. Probab. Surv. 62.
- [10] Kallenberg, O., 1997. Foundations of modern probability. Probability and its Applications. Springer-Verlag, New York.
- [11] Kingman, J. F C., 1982. The coalescent. Stochastic Processes and their Applications, 13(3): 235-248.
- [12] Lenski, R.E., E. coli long-term experiment experimental evolution project site. Disponible en: <http://lenski.mmg.msu.edu/ecoli/>
- [13] McArthur, R., Wilson. E.O., 1967. The Theory of Island Biogeography. Princeton University Press.
- [14] Miller, T.E., Burns, J.H., Munguia, P., Walters, E. L., Kneitel, J.M., Richards, P.M., Mouquet, N., Buckley, H.L., 2005. A critical review of twenty years use of the resource-ratio theory. Am. Nat. 165 (4), 165-175.
- [15] Möhle, M., 1999. The Concept of Duality and Applications to Markov Processes Arising in Neutral Population Genetics Models. Bernoulli, 5(5), 761-777.
- [16] Moran P. A. P., 1958. Random processes in genetics. Mathematical Proceedings of the Cambridge Philosophical Society. 54(1): 60-67.
- [17] Molenaar, D., Van Berlo, R., De Ridder, D., Teusink B., 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol. Syst. Biol.

- [18] Rincón L., 2005. Introducción a las ecuaciones diferenciales estocásticas. Memorias del Congreso Regional de Probabilidad y Estadística de la UAA. Disponible en: <http://lya.ciencias.unam.mx/lars/pub/ecuaciones.pdf>
- [19] Wright, S., 1931. Evolution in Mendelian Population Genetics. *Genetics*, 16, 97-159.