



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Modelos genéticos de poblaciones con eventos
de recombinación

TESIS

que para obtener el título de:

ACTUARIA

presenta:

JOSÉ BENITO DÍAZ HERNÁNDEZ

Director de tesis: DRA. ELIANE REGINA RODRÍGUES



2010



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central

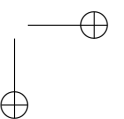
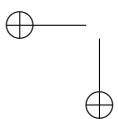
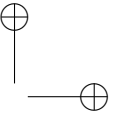
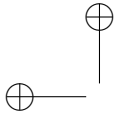


UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



AGRADECIMIENTOS

A la DRA. ELIANE REGINA ROGRÍGUES por sus consejos y su gran apoyo, y por la dirección y revisión de esta tesis. A los miembros del jurado:

DR. JUAN GONZÁLEZ HERNÁNDEZ

DR. LUIS ANTONIO RINCÓN SOLÍS

DR. RAMSÉS HUMBERTO MENA CHÁVEZ

M. en A.P. MARÍA DEL PILAR ALONSO REYES

Por su revisión y sus comentarios.

A mi padre JOSÉ BENITO ANTONIO DÍAZ ÁVILA por su guía, por compartir conmigo sus experiencias y haberme apoyado siempre.

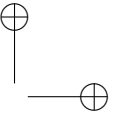
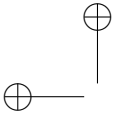
A mi madre GABINA HERNÁNDEZ MÉNDEZ por todas sus bendiciones, sus enseñanzas y sobre todo su amor.

A mi hermano ÁDAN por su compañía, por todas las buenas experiencias que hemos vivido a lo largo de nuestras vidas y por guiarme en muchos momentos.

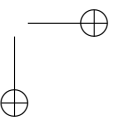
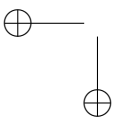
A mi hermana MARIBEL por su cariño y por los consejos que me ha dado para salir adelante.

A JAVIER por ser un buen amigo y estar siempre conmigo y mi familia.

A NAYE por estar a mi lado en todo momento y dejarme pasar momentos inolvidables a su lado. Gracias por apoyarme y cuidarme, siempre estaré a tu lado mi cachetona.



A mis amigos JONATHAN, FERNANDO, ADRIAN, CHRISTIAN y DON ALLHAN por haber compartido mucho momentos dentro y fuera de la facultad y por haberme brindado su amistad.



CONTENIDO

Introducción	iii
1 Procesos de Difusión	1
1.1 Proceso de Difusión	1
1.1.1 Filtración y tiempo de paro	1
1.1.2 Proceso de difusión	2
1.1.3 Procesos de difusión multivariados	5
1.2 Ecuaciones diferenciales de Kolmogorov	6
1.2.1 Ecuación de Kolmogorov hacia adelante	7
1.2.2 Ecuación de Kolmogorov hacia atrás	8
1.3 Distribución estacionaria	8
1.3.1 Cálculo de la distribución estacionaria	9
1.4 Clasificación de las fronteras para un proceso de difusión regular	11
1.5 Convergencia a una difusión	13
2 Modelo de Wright-Fisher y modelo de Moran	15
2.1 Modelo de Wright-Fisher	16
2.1.1 Caso sin mutación	16
2.1.2 Caso con presencia de mutación	17
2.2 Modelo de Moran	18
2.2.1 Caso sin mutación	19
2.2.2 Caso con presencia de mutación	20

CONTENIDO	ii
2.2.3 Modelo de Moran a tiempo continuo	20
2.3 Aproximación por difusión	27
2.3.1 Aproximación por difusión para el modelo de Wright-Fisher	27
2.3.2 Aproximación por difusión para el modelo de Moran	34
3 Modelo genético con recombinación	39
3.1 Caso univariado	39
3.2 Caso multivariado	48
4 Dos-locus ancestrales	55
4.1 El modelo dos-locus	55
4.2 Estructura del gráfico de la recombinación ancestral	57
Conclusión	61
Apéndice	63
Bibliografía	69

INTRODUCCIÓN

La recombinación genética es un proceso que lleva a la obtención de un genotipo a través del intercambio de material genético entre secuencias de DNA en cromosomas homólogos. La recombinación genética es una forma de aumentar la variabilidad del material genético de una población, debido a que la información genética de dos genotipos se puede unir en un nuevo genotipo.

Debido a lo anterior, se han hecho investigaciones que han conducido a estudios en el campo de la modelación estocástica relacionados con la genética de poblaciones, donde surgen familias discretas o cadenas de Markov a tiempo continuo.

Es por ello que en la presente tesis se estudia el modelo de Wright-Fisher y el modelo de Moran donde se considera la presencia de recombinación. Estos dos modelos matemáticos han sido desarrollados y aplicados al estudio de poblaciones en genética, dicho estudio es uno de los temas de mayor interés en la evolución de poblaciones. Para los dos modelos antes mencionados se realizará una aproximación por medio de difusión. En el caso del modelo de Moran se desarrollará el modelo con presencia de recombinación. Por último se presentará el modelo dos-Locus donde la recombinación proviene de un solo ancestro.

En el primer capítulo se presentarán algunos resultados básicos sobre los procesos de difusión, la teoría para calcular la distribución estacionaria de un proceso estocástico y la convergencia a una difusión que permitirán la adecuada comprensión de los modelos que se describirán en los siguientes capítulos.

El Capítulo 2 se conformará de la descripción y análisis de los modelos de Wright-Fisher y Moran. Se presentarán dos casos para ambos, en el primero se observará uno en donde hay presencia de mutación, atributo que

no ocurrirá en el otro caso. Adicionalmente, en este capítulo se realizará la aproximación por difusión para ambos modelos, así como el cálculo de su distribución estacionaria. Se observará que los dos modelos tienen los mismos parámetros infinitesimales y por lo tanto tendrán la misma distribución estacionaria.

En cuanto al Capítulo 3 se refiere, se analizará el modelo de Moran con recombinación para los casos univariado y multivariado, realizando en ambos casos una aproximación por difusión y el cálculo sus respectivos parámetros infinitesimales.

Finalmente en el Capítulo 4 se describirá el modelo dos-Locus ancestral cuya evolución estocástica sigue el modelo de Moran .

CAPÍTULO 1

PROCESOS DE DIFUSIÓN

En este capítulo se presentan algunos resultados básicos sobre los procesos de difusión. Antes de desarrollar el capítulo, se mencionan algunos conceptos que nos ayudarán a entender la definición de un proceso de difusión.

Las definiciones que se mencionan se encuentran ampliamente desarrolladas en Tudor (2002) y Karlin y Taylor (1981).

1.1 PROCESO DE DIFUSIÓN

1.1.1 FILTRACIÓN Y TIEMPO DE PARO

Definición 1.1 Una **filtración** sobre un espacio medible (Ω, \mathcal{F}) es una familia $\{\mathcal{F}_t\}_{t \in T \subset \mathbb{R}}$ de *sub- σ -álgebra* de \mathcal{F} , tales que $\mathcal{F}_s \subset \mathcal{F}_t$ si $s < t$. La **filtración canónica** asociada a un proceso estocástico $\{X_t\}_{t \in T \subset \mathbb{R}}$ es por definición $\mathcal{F}_t^X = \sigma(X_s : s \leq t), t \in T$.

Observación: Si se considera el caso en donde $T = \{0, 1, \dots\}$, generalmente se denota

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{s \in T} \mathcal{F}_s\right), \text{ para } t \geq 0, \quad (1.1)$$

Definición 1.2

- (a) Una filtración $\{\mathcal{F}_t\}_{t \geq 0}$ se llama **continua por la derecha** si $\mathcal{F}_t = \mathcal{F}_{t+}$ para cada $t \geq 0$.

- (b) Una filtración $\{\mathcal{F}_t\}_{t \geq 0}$ que sea continua por la derecha de tal manera que \mathcal{F}_0 contenga a los conjuntos insignificantes de \mathcal{F} , es decir, los conjuntos de medida cero se le llama **filtración estándar**.
- (c) Sea $\{\mathcal{F}_t\}_{t \geq 0}$ una filtración. Al sistema $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t \in T})$ se le llama **espacio de probabilidad filtrado**.

Definición 1.3 Sea $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t \in T})$ un espacio de probabilidad filtrado. Una aplicación $\tau : \Omega \rightarrow T \cup \{\infty\}$ se llama **\mathcal{F}_t -tiempo de paro** si para cada $t \in T$, $\{\tau \leq t\} \in \mathcal{F}_t$. A cada tiempo de paro τ se le asocia la σ -álgebra \mathcal{F}_τ definida por

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{F}_t\}, \quad \text{para } t \in T.$$

Definición 1.4 Sea $X = \{X_t : t \geq 0\}$ un proceso de Markov con espacio de estados S y probabilidad de transición $p_{ij}(t)$, $i, j \in S$, $t \geq 0$. El proceso $X = \{X_t : t \geq 0\}$ es un **proceso fuerte de Markov** si para cada tiempo de paro τ con respecto a X , se tiene que para cada $t \geq 0$ y $A \subset S$

$$P\{X_{\tau+t} \in A | X_s = x_s, s \leq \tau\} = P\{X_{\tau+t} \in A | X_\tau = x_\tau\}$$

para todo $x_s \in S$.

1.1.2 PROCESO DE DIFUSIÓN

Definición 1.5 Sea $X = \{X_t : t \geq 0\}$ un proceso fuerte de Markov con espacio de estados $S = I \subseteq \mathbb{R}$, entonces el proceso X es llamado **proceso de difusión**, si X es un proceso de Markov con trayectorias continuas casi seguramente.

Definición 1.6 Un proceso de difusión $X = \{X_t : t \geq 0\}$ con espacio de estados $S = (l, r)$, $l < r$, se llama **regular** si

$$P\{T_z < \infty | X_0 = x\} > 0$$

con $l < x$ y $z < r$, donde T_z es el tiempo que tarda X en tomar por primera vez el valor z .

Observación: Nótese que T_z puede tomar el valor infinito si para $t \geq 0$, $X_t \neq z$.

Definición 1.7 Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión con espacio de estados S . Sea $\Delta_h X_t = X_{t+h} - X_t$ el incremento de X en el intervalo

$[t, t + h]$, a este proceso de difusión se le puede asociar lo que llamamos los **parámetros infinitesimales** de X que están definidos como:

$$\mu(x, t) = \lim_{h \rightarrow 0} \frac{1}{h} E[\Delta_h X_t | X_t = x] \quad (1.2)$$

$$\sigma^2(x, t) = \lim_{h \rightarrow 0} \frac{1}{h} E[\{\Delta_h X_t\}^2 | X_t = x] \quad (1.3)$$

para $x \in S$ y $t \geq 0$.

Observaciones:

1. El parámetro $\mu(x, t)$ es conocido como media infinitesimal o parámetro de deriva; mientras que al parámetro $\sigma^2(x, t)$ se le conoce como varianza infinitesimal o parámetro de difusión.
2. Por definición se tiene que

$$E[\Delta_h X_t | X_t = x] = E(X_{t+h} - X_t | X_t = x) = \mu(x, t)h + \mathcal{O}(h)$$

$$E[\{\Delta_h X_t\}^2 | X_t = x] = E\{(X_{t+h} - X_t)^2 | X_t = x\} = \sigma(x, t)h + \mathcal{O}(h)$$

donde $\mu(x, t)$ y $\sigma(x, t)$ son llamados parámetros infinitesimales del proceso X y $\mathcal{O}(h)$ es tal que $\lim_{h \rightarrow 0} \frac{\mathcal{O}(h)}{h} = 0$.

3. Si se satisface la condición de Dynkin, es decir, dada $\epsilon > 0$

$$\lim_{h \rightarrow 0} P\{|X_{t+h} - X_t| > \epsilon | X_t = x\} = 0$$

uniformemente respecto a $t \in [0, N]$ para todo x en un subintervalo compacto de S , entonces satisface (1.2) y (1.3) y además tenemos que:

$$\lim_{h \rightarrow 0} \frac{E[|\Delta_h X_t|^r | X_t = x]}{h} = 0, r > 2.$$

Definición 1.8 Cuando el proceso de difusión X es homogéneo en el tiempo (independiente de t), se tienen los parámetros infinitesimales los cuales se denotan como $\mu(x, t) = \mu(x)$ y $\sigma^2(x, t) = \sigma^2(x)$.

Definición 1.9 Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados S y sea ξ un tiempo posiblemente infinito. Se define $X_\xi = \{X_t, 0 \leq t < \xi\}$ una **difusión con muerte** si se comporta como un

proceso de difusión regular hasta un posible valor ξ infinito, en donde el proceso “muere”, es decir, donde termina.

Definición 1.10 Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión con espacio de estados S , entonces se dice que el proceso es **conservativo** si

$$P \{X_t \in S | X_0 = x\} = P \{\xi > t | X_0 = x\} = 1, \quad t \geq 0, x \in S.$$

Observación: Para un proceso de difusión con muerte para cada punto x , existe la probabilidad $k(x) dt + \mathcal{O}(h)$ de que en el intervalo de tiempo $(t, t + dt)$ ocurra el suceso (muerte) y la probabilidad $1 - k(x) dt + \mathcal{O}(h)$ de que no ocurra muerte. Por lo tanto se define la tasa de muerte como:

$$\lim_{h \downarrow 0} \frac{1}{h} P \{t < \xi < t + h | X_t = x\} = k(x, t) \quad (1.3)$$

Teorema 1.1.1 Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados en $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Sea g una función estrictamente monótona en S con segunda derivada $g''(x)$ continua para $l < x < r$. Entonces $Y_t = g(X_t)$ define un proceso de difusión regular en el intervalo S' con extremos $g(l)$ y $g(r)$ y Y_t tiene los siguientes parámetros infinitesimales:

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

$$\sigma_Y^2(y) = \sigma^2(x) [g'(x)]^2$$

donde $y = g(x)$.

Demostración. Sea g estrictamente creciente. La demostración es similar para g estrictamente decreciente. Como g es dos veces continuamente diferenciable entonces tiene la siguiente expansión de Taylor,

$$g(x + \Delta x) = g(x) + \Delta x g'(x) + \frac{1}{2} (\Delta x)^2 g''(x) + \frac{1}{2} (\Delta x)^2 (g''(y) - g''(x)),$$

con $x \leq y \leq x + \Delta x$. Tome $X_t = x$ y $\Delta X_t = X_{t+h} - X_t$, entonces

$$\begin{aligned} g(X_{t+h}) &= g(X_t) + \Delta X_t g'(X_t) + \frac{1}{2} (\Delta X_t)^2 g''(X_t) \\ &\quad + \frac{1}{2} (\Delta X_t)^2 (g''(y(w)) - g''(X_t)) \end{aligned} \quad (1.4)$$

donde $y(w)$ es una cantidad aleatoria entre X_t y X_{t+h} .

Dado que $X_t = x$ se tiene que $g(x) = g(X_t) = Y_t = y$, por tanto (1.5) es de la forma

$$Y_{t+h} - Y_t = \Delta X_t g'(X_t) + \frac{1}{2} (\Delta X_t)^2 g''(X_t) + \frac{1}{2} (\Delta X_t)^2 (g''(y(w)) - g''(X_t)), \quad (1.5)$$

tomando esperanza, dividiendo a ambos lados y haciendo $h \downarrow 0$ se obtiene:

$$\begin{aligned} \lim_{h \downarrow 0} \frac{1}{h} E [Y_{t+h} - Y_t | Y_t = y] &= \mu(x) g'(x) + \frac{1}{2} \sigma^2(x) g''(x) \\ &+ \frac{1}{2} \lim_{h \downarrow 0} \frac{1}{h} E \left[(\Delta X_t)^2 (g''(y(w)) - g''(X_t)) \right]. \end{aligned}$$

Observación: Nótese que la continuidad de g'' implica la convergencia de $g''(y(x))$ a $g''(X_t)$ y dado que $h^{-1} E [(\Delta X)^2]$ converge, se tiene que

$$\lim_{h \downarrow 0} \frac{1}{h} E \left[(\Delta X)^2 (g''(y(w)) - g''(X_t)) \right] = 0.$$

Por lo tanto,

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

La varianza infinitesimal de Y_t se obtiene que manera similar. Tomando (1.6) y elevando al cuadrado se tiene:

$$[Y_{t+h} - Y_t]^2 = (\Delta X)^2 [g'(X_t)]^2 + R_h,$$

donde R_h contiene $(\Delta X)^3$ y términos con orden mayor a tres. Sabemos que $\lim_{h \downarrow 0} h^{-1} E [|\Delta_h X_t|^r | X_t = x] = 0$ para $r \geq 3$, por lo tanto se obtiene:

$$\begin{aligned} \sigma_Y^2(y) &= \lim_{h \downarrow 0} \frac{1}{h} E \left[\{Y_{t+h} - Y_t\}^2 | Y_t = y \right] \\ &= \sigma^2(x) [g'(x)]^2. \end{aligned}$$

1.1.3 PROCESOS DE DIFUSIÓN MULTIVARIADOS

Hasta ahora se han considerado procesos de difusión en una dimensión, es importante mencionar que existen vectores de procesos de difusión sobre un espacio $I = E^n$ (espacio n -Euclidiano) o sobre una región abierta de E^n .

Definición 1.11 Sea $\mathbb{X}_t = (X_t^1, X_t^2, \dots, X_t^n)$, $t \geq 0$ un **proceso de difusión n-variado**, entonces los parámetros infinitesimales de \mathbb{X}_t son:

$$\lim_{h \downarrow 0} \frac{1}{h} E [X_{t+h}^i - X_t^i | \mathbb{X}_t = \mathbf{x} = (x^1, x^2, \dots, x^n)] = \mu^i(\mathbf{x}, t), \text{ para } i = 1, 2, \dots, n$$

y

$$\lim_{h \downarrow 0} \frac{1}{h} E \left[\{X_{t+h}^i - X_t^i\} \{X_{t+h}^j - X_t^j\} | \mathbb{X}_t = \mathbf{x} \right] = \sigma^{ij}(\mathbf{x}, t), \text{ para } i, j = 1, 2, \dots, n.$$

La matriz $\|\sigma^{ij}(\mathbf{x}, t)\|_1^n$ debe ser definida positiva, es decir, $\sum_{i,j=1}^n \sigma^{ij}(\mathbf{x}, t) a^i a^j > 0$ para toda (a_1, \dots, a_n) n -tupla real no trivial y $\mathbf{x} \in I$ y además satisfacer

$$\sum_{i,j=1}^n a^i a^j \{X_{t+h}^i - X_t^i\} \{X_{t+h}^j - X_t^j\} = \sum_{\nu=1}^n a^\nu \{X_{t+h}^\nu - X_t^\nu\}^2.$$

Observación: Los momentos de orden superior son insignificantes en comparación con h , es decir:

$$\lim_{h \downarrow 0} \frac{1}{h} E \left[\{X_{t+h}^i - X_t^i\}^k | \mathbb{X}_t = \mathbf{x} \right] = 0, k \geq 4 \quad \text{para } i = 1, 2, \dots, n.$$

1.2 ECUACIONES DIFERENCIALES DE KOLMOGOROV

Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular y homogéneo en el tiempo con espacio de estados $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Se indica por $P_{xy}(t) = P\{X_t \leq y | X_0 = x\}$ la función de distribución de X_t , sujeta a la distribución inicial

$$P_{xy}(0) = \begin{cases} 1 & \text{si } x \leq y \\ 0 & \text{si } x > y, \end{cases} \quad (1.6)$$

es decir, una distribución puntual concentrada en x . Por tanto $p_{xy}(t)$ la densidad de transición de X_t , es igual a

$$p_{xy}(t) = \frac{dP_{xy}(t)}{dy}, \quad t \geq 0. \quad (1.7)$$

1.2.1 ECUACIÓN DE KOLMOGOROV HACIA ADELANTE

La ecuación de Kolmogorov hacia adelante, también es conocida con el nombre de ecuación de Fokker-Planck o la segunda ecuación de Kolmogorov. Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular y homogéneo en el tiempo con espacio de estados $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Defina la función $u(x, t)$ por

$$u(x, t) = E[g(X_t) | X_0 = x], \quad t \geq 0, \quad (1.8)$$

donde g es un función acotada y continua en S .

Teorema 1.2.1 *Sea $u(x, t)$ continua y diferenciable respecto a t , tiene segunda derivada con respecto a x y es continua en cero por la derecha, entonces $u(x, t)$ satisface ecuación diferencial parcial*

$$\frac{\partial u(x, t)}{\partial t} = \frac{1}{2}\sigma^2(x) \frac{\partial^2 u(x, t)}{\partial x^2} + \mu(x) \frac{\partial u(x, t)}{\partial x} \quad (1.9)$$

con la condición inicial $\lim_{t \rightarrow 0^+} u(x, t) = g(x)$. En particular si $g(\eta)$ es de la forma

$$g(\eta) = \begin{cases} 1 & \text{si } \eta \leq y \\ 0 & \text{si } \eta > y, \end{cases} \quad (1.10)$$

entonces $u(x, t) = P_{xy}(t)$, de donde la ecuación (1.10) se transforma en la ecuación de Kolmogorov hacia adelante, y es de la forma

$$\frac{\partial P_{xy}(t)}{\partial t} = \frac{1}{2}\sigma^2(x) \frac{\partial^2 P_{xy}(t)}{\partial x^2} + \mu(x) \frac{\partial P_{xy}(t)}{\partial x} \quad (1.11)$$

con $t > 0$ y $l < x, y < r$ y con condiciones iniciales

$$P_{xy}(0+) = \begin{cases} 1 & \text{si } x \leq y \\ 0 & \text{si } x > y, \end{cases} \quad (1.12)$$

La densidad de transición p_{xy} también satisface la ecuación de Kolmogorov hacia adelante:

$$\frac{\partial p_{xy}(t)}{\partial t} = \frac{1}{2}\sigma^2(x) \frac{\partial^2 p_{xy}(t)}{\partial x^2} + \mu(x) \frac{\partial p_{xy}(t)}{\partial x} \quad (1.13)$$

para $t > 0$ y $x, y \in S$.

Demostracion. Ver Karlin y Taylor (1981).

1.2.2 ECUACIÓN DE KOLMOGOROV HACIA ATRÁS

La ecuación de Kolmogorov hacia atrás es también conocida como la primera ecuación de Kolmogorov o ecuación progresiva de Kolmogorov.

Teorema 1.2.2 *Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular y homogéneo en el tiempo con espacio de estados $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Y sea p_{xy} la densidad de transición de X_t , entonces $p_{xy}(t)$ satisface la ecuación diferencial parcial*

$$\frac{\partial p_{xy}(t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial y^2} [\sigma^2(y) p_{xy}(t)] - \frac{\partial}{\partial y} [\mu(x) p_{xy}(t)], \quad (1.14)$$

para $t > 0$ y $x, y \in S$.

Demostración. Ver Karlin y Taylor (1981).

1.3 DISTRIBUCIÓN ESTACIONARIA

Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular y homogéneo en el tiempo con espacio de estados $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Si existe una distribución estacionaria $\psi(y)$, entonces la densidad estacionaria $\psi(x)$ necesariamente satisface

$$\psi(y) = \int \psi(x) p_{xy}(t) dx, \quad \forall t > 0. \quad (1.15)$$

y por lo tanto $\psi(y)$ satisface la ecuación diferencial parcial:

$$0 = \frac{1}{2} \frac{\partial^2}{\partial y^2} [\sigma^2(y) \psi(y)] - \frac{\partial}{\partial y} [\mu(y) \psi(y)], \quad y \in S \quad (1.16)$$

Las demostraciones de las afirmaciones anteriores fueron omitidas, se pueden encontrar en Karlin y Taylor (1981).

1.3.1 CÁLCULO DE LA DISTRIBUCIÓN ESTACIONARIA

Definición 1.12 Sea $\mu(x)$ y $\sigma^2(x)$, $x \in S$ los parámetros infinitesimales del proceso de difusión X , entonces se define $s(y)$ como:

$$s(y) = \exp \left\{ - \int^y \left[\frac{2\mu(\xi)}{\sigma^2(\xi)} \right] d\xi \right\}$$

- Defínase $S(x)$ como la **función escala** cuya expresión está dada por:

$$S(x) = \int^x s(y) dy. \tag{1.17}$$

- Sea $J = [c, d] \subset (l, r)$, entonces definamos $S[J]$ como la **medida escala** del intervalo J y se denota como

$$S[J] = S[c, d] = S(d) - S(c).$$

Se usará libremente la medida escala $dS(x) = S[dx]$ de un intervalo infinitesimal $[x, x + dx]$ con $S[dx] = S(x + dx) - S(x)$ donde

$$0 < S[c, d] < \infty, \quad l < c < x < d < r$$

y

$$S[c, d] = S[c, x] + S[x, d], \quad l < c < x < d < r. \tag{1.18}$$

- Sea $m(x)$ la **densidad de velocidad** definida por,

$$m(x) = \frac{1}{s(x)\sigma^2(x)}. \tag{1.19}$$

- Se denota $M[J]$ como la **medida de velocidad** inducida por la densidad de velocidad $m(x)$, cuya expresión es

$$M[J] = M[c, d] = \int_c^d m(x) dx$$

donde $0 < M[J] < \infty$ para $J = [c, d] \subset (l, r)$; además se tiene que $dM[x] = m(x) dx$.

- Sea $M(l, x)$ la **medida de rapidez** de un proceso cerca de l , denotada por:

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

Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular y homogéneo en el tiempo con espacio de estados $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Sea $\psi(y)$ la distribución estacionaria de X , entonces (1.17) es válida y si se integra con respecto a y se obtiene:

$$\frac{1}{2}C_1 = \frac{d}{dy} \left[\frac{\sigma^2(y)}{2} \psi(y) \right] - \mu(y) \psi(y), \quad (1.20)$$

donde C_1 es una constante. Multiplicando (1.21) de ambos lados por el factor

$$s(y) = \exp \left\{ - \int^y \left[\frac{2\mu(\xi)}{\sigma^2(\xi)} \right] d\xi \right\}$$

se obtiene

$$C_1 \exp \left\{ - \int^y \left[\frac{2\mu(\xi)}{\sigma^2(\xi)} \right] d\xi \right\} = \frac{d}{dy} \left[\exp \left\{ - \int^y \left[\frac{2\mu(\xi)}{\sigma^2(\xi)} \right] d\xi \right\} \sigma^2(y) \psi(y) \right],$$

es decir,

$$C_1 s(y) = \frac{d}{dy} [s(y) \sigma^2(y) \psi(y)]. \quad (1.21)$$

Si se integra (1.22) otra vez con respecto a y , se obtiene

$$C_1 S(x) + C_2 = s(x) \sigma^2(x) \psi(x),$$

donde C_2 es constante. Por lo tanto la distribución estacionaria cuando existe es de la forma

$$\begin{aligned} \psi(x) &= C_1 \frac{S(x)}{s(x) \sigma^2(x)} + C_2 \frac{1}{s(x) \sigma^2(x)} \\ &= m(x) [C_1 S(x) + C_2]. \end{aligned}$$

Observación: La distribución estacionaria $\psi(y)$, $y \in S$ existe, si se pueden encontrar constantes C_1 y C_2 tales que $\psi(x) \geq 0$ en S y $\int_r^l \psi(y) dy = 1$, de lo contrario no se puede afirmar su existencia. Ver Karlin y Taylor (1975).

1.4 CLASIFICACIÓN DE LAS FRONTERAS PARA UN PROCESO DE DIFUSIÓN REGULAR

Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados $S = (l, r)$ y parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$, $x \in S$. Ahora se analizará el comportamiento de las fronteras l y r .

Puesto que $S[a, b]$ es una medida no negativa y por (1.19) se tiene que $S[a, b]$ es monótona en a y fija en b (ver Karlin y Taylor (1981)), por lo tanto se define $S(l, b) \leq \infty$ por

$$S(l, b) = \lim_{a \downarrow l} S[a, b] \leq \infty, \quad l < b < r. \quad (1.22)$$

Observación: Si $S[a, b] \subset (l, r)$, entonces $0 \leq S[a, b] < \infty$; y además si se cumple (1.19) se tiene que

$$S(l, b) = \infty, \quad \text{para algún } b \in (l, r)$$

si y sólo si

$$S(l, b) = \infty, \quad \text{para todo } b \in (l, r).$$

Definición 1.13 Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados $S = (l, r)$, entonces se dice que la frontera l es **atractora** si $S(l, x) < \infty$, independientemente de $x \in S$.

Definición 1.14 Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados $S = (l, r)$. Se denota por $\sum(l)$ a la siguiente cantidad:

$$\begin{aligned} \sum(l) &= \lim_{a \downarrow l} \int_a^x S[a, \xi] dM(\xi) = \int_l^x S(l, \xi) dM(\xi) \\ &= \int_l^x \left\{ \int_l^x s(\eta) d\eta \right\} m(\xi) d\xi = \int_l^x \left\{ \int_\eta^x m(\eta) d\eta \right\} s(\eta) d(\eta) \\ &= \int_l^x M[\eta, x] dS(\eta). \end{aligned} \quad (1.23)$$

Definición 1.15 La frontera l es llamada **alcanzable** si $\sum(l) < \infty$ y es llamada **inalcanzable** si $\sum(l) = \infty$.

Lema 1.4.1 *Sea l una frontera atractora y supóngase $l < x < b < r$, entonces:*

$$\sum(l) = \int_l^x S(l, \eta] dM(\eta) < \infty.$$

Demostración. Ver Karlin y Taylor (1981).

Definición 1.16 *Sea $N(l)$ la medida del tiempo en que tarda en llegar a un punto interior x en (l, r) empezando en la frontera l . Entonces, se define*

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

Lema 1.4.2 *Las relaciones entre $S(l, x]$, $\sum(l)$, $N(l)$ y $M(l, x]$ son:*

- i) $S(l, x] = \infty$ implica $\sum(l) = \infty$,
- ii) $\sum(l) < \infty$ implica $S(l, x] < \infty$,
- iii) $M(l, x] = \infty$ implica $N(l) = \infty$,
- iv) $N(l) < \infty$ implica $M(l, x] < \infty$,
- iii) $\sum(l) + N(l) = S(l, x] M(l, x]$.

Demostración. Ver Karlin y Taylor (1981).

Las definiciones y lemas anteriores se pueden obtener de igual manera para la frontera r .

Definición 1.17 *Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados $S = (l, r)$, entonces l puede ser una frontera:*

- i) **Regular**, si $S(l, x] < \infty$ y $M(l, x] < \infty, x \in S$.
- ii) **De salida**, si $\sum(l) < \infty$ pero $M(l, x] = \infty, x \in S$.
- iii) **De entrada**, si $S(l, x] = \infty$ mientras que $N(l) < \infty$.
- iv) **Natural** (Feller), si $\sum(l) = \infty$ y $N(l) = \infty$.

Teorema 1.4.1 *Sea $X = \{X_t : t \geq 0\}$ un proceso de difusión regular con espacio de estados $S = (l, r)$ y sean l y r fronteras de entrada, entonces la distribución estacionaria del proceso X_t es:*

$$\psi(x) = \frac{m(x)}{\int_l^r m(\xi) d\xi} = \frac{1}{\sigma^2(x) s(x) \int_l^r [\sigma^2(\xi) m(\xi)]^{-1} d\xi}. \quad (1.24)$$

Demostración. Ver Karlin y Taylor (1981).

1.5 CONVERGENCIA A UNA DIFUSIÓN

Al realizar la modelación estocástica de problemas en genética de poblaciones, pueden surgir familias o cadenas de Markov a tiempo continuo o a tiempo discreto $Z^K = \{Z_t^K : t \geq 0\}$ indexada por $K = 1, 2, \dots$. Se define $X^K = \{X_t^K : t \geq 0\}$ como

$$X_t^K = a(K) \left[Z_{b(K)t}^K - c(K) \right],$$

donde $c(K)$ es una constante que centra el proceso, $a(K)$ cambia la escala al estado de la variable y $b(K)$ cambia la escala de tiempo. En consecuencia, una unidad de tiempo del proceso X_t^K corresponde a un periodo de $b(K)$ unidades de tiempo del proceso original Z_t^K .

Sea $X^K = \{X_n^K : n = 0, 1, 2, \dots\}$ la secuencia de procesos estocásticos en tiempo discreto y con espacio de estados $S \in \mathfrak{R}$; con S cerrado, y tal que X^K acotado y adaptado¹ a una secuencia creciente $\{\mathcal{F}_n, n \geq 0\}$ de σ -álgebras. Se suele usar $\mathcal{F}_n = \sigma(X_m^K, m \leq n)$.

Sea $\Delta X_n^K = X_{n+1}^K - X_n^K$ tal que

$$E[\Delta X_n^K | \mathcal{F}_n^K] = h_K \mu(X_n^K) + \epsilon_{1,n}^K, \quad (1.25)$$

$$E[(\Delta X_n^K)^2 | \mathcal{F}_n^K] = h_K \sigma^2(X_n^K) + \epsilon_{2,n}^K, \quad (1.26)$$

y

$$E[(\Delta X_n^K)^4 | \mathcal{F}_n^K] = \epsilon_{3,n}^K, \quad (1.27)$$

¹ X_t^K es adaptado si X_t^K es medible con respecto a $\mathcal{F}_t, t \geq 0$.

donde $h_K > 0$ y $h_K \rightarrow \infty$ cuando $N \rightarrow \infty$. Debido a que los términos de error son suficientemente pequeños y se pueden extender que para $n \geq 0$ se tiene

$$\sum_{n > [n/h_K]} E [|\epsilon_{i,n}^K|] \rightarrow 0,$$

para $i = 1, 2, 3$ cuando $K \rightarrow \infty$ donde $[x]$ es el mayor entero menor que x .

Cuando los procesos $\{X_n^K\}$ son cadenas de Markov, entonces en (1.26), (1.27) y (1.28) se utiliza X_n^K en lugar de \mathcal{F}_n^K . Defínase el proceso a tiempo continuo:

$$X_t^K = X_{[t/h_K]}^K.$$

Por lo tanto bajo ciertos requerimientos de suavidad (ver Tudor (2002) o Karlin y Taylor (1981)) se tiene que

$$X_t^K \longrightarrow X_t, \quad t \geq 0 \tag{1.28}$$

convergencia en distribución, cuando $K \rightarrow \infty$, y además para algún conjunto de puntos $0 \leq t_1 < t_2 < \dots < t_r$ la distribución finito dimensional de $\{X_{t_1}^K, X_{t_2}^K, \dots, X_{t_r}^K\}$ converge para la distribución finito dimensional de $\{X_{t_1}, X_{t_2}, \dots, X_{t_r}\}$, cuando $K \rightarrow \infty$, donde X_t es el proceso de difusión con parámetros infinitesimales $\mu(x)$ y $\sigma^2(x)$.

CAPÍTULO 2

MODELO DE WRIGHT-FISHER Y MODELO DE MORAN

En este capítulo se describen dos modelos matemáticos aplicados al estudio de poblaciones en genética. Dicho estudio es uno de los temas de mayor interés en la evolución de poblaciones genéticas. El primer modelo que se presenta en este capítulo fue introducido por S. Wright y R. Fisher mientras que el segundo por Moran, los cuales son conocidos como modelo Wright-Fisher y modelo de Moran, respectivamente. Ambos modelos tienen el supuesto de que la población es haploide, es decir, cada individuo posee un único progenitor. Se considera un cromosoma con un único locus¹, que por hipótesis tiene dos alelos² posibles denotados por **a** y **A**. Otra hipótesis es que se permite mutación entre ambos tipos de alelos, además de que los modelos son neutrales, es decir, el tipo de alelo de un individuo no influye en su capacidad de reproducción. Primero se presenta la versión en tiempo discreto de ambos modelos y posteriormente en tiempo continuo el modelo de Moran.

Para ambos modelos se considera que el tamaño de la población N se mantiene constante. Se denota por α_1 a la probabilidad de que el alelo **a** mute al alelo **A**, y por α_2 a la probabilidad de que el alelo **A** mute al alelo **a**, con $0 \leq \alpha_1, \alpha_2 \leq 1$. Nótese que $\alpha_1 = \alpha_2 = 0$ corresponde al caso en donde no se presenta mutación.

En el caso discreto en ambos modelos se denota por X_n a la variable aleatoria que indica el número de individuos del tipo **a** después del n -ésimo evento de reproducción $n=0,1,2,\dots$. En el caso continuo la variable que

¹Posición de un gen en un cromosoma.

²Forma alternativa de un gen.

indica el número de individuos del tipo \mathbf{a} al tiempo t es denotada por X_t , $t \geq 0$.

A medida de que el tiempo transcurre la evolución de la población será descrita por el proceso $X = \{X_n: n = 0, 1, 2, \dots\}$ y $X = \{X_t: t \geq 0\}$, dependiendo de cada caso. En ambos casos X es una cadena de Markov con espacio de estados $S = \{0, 1, 2, \dots, N\}$.

2.1 MODELO DE WRIGHT-FISHER

El modelo de Wright-Fisher es un modelo de evolución de una población donde no existe intersección de generaciones pues cada evento de reproducción da origen a una generación compuesta en su totalidad por nuevos individuos.

La demografía con la cual el modelo evoluciona es de la siguiente forma: dada un generación, se seleccionan con reemplazo los N individuos presentes, y dicha selección se realiza de forma independiente y uniforme. Cada vez que un individuo es seleccionado produce un descendiente hasta que N descendientes sean producidos. Por lo tanto cuando el proceso de reproducción termina se tiene que la siguiente generación está formada por los descendientes de los individuos seleccionados.

El proceso de mutación en la población está dado por: un individuo con alelo del tipo \mathbf{a} es seleccionado y produce un individuo con alelo del tipo \mathbf{A} con probabilidad α_1 , o produce uno de su mismo tipo con probabilidad $1 - \alpha_1$. Cuando se trata de una selección de un individuo con alelo del tipo \mathbf{A} , se tiene que produce uno con alelo del tipo \mathbf{a} con probabilidad α_2 y produce uno de su mismo tipo con probabilidad $1 - \alpha_2$.

Se tiene que $X = \{X_n: n = 0, 1, 2, \dots\}$ es una cadena de Markov que cuenta el número de individuos del tipo \mathbf{a} y se nota que tiene espacios de estados $S = \{0, 1, 2, \dots, N\}$.

2.1.1 CASO SIN MUTACIÓN

El primer caso que se presentará es cuando no se considera presencia de eventos de mutación, es decir, cuando tenemos $\alpha_1 = \alpha_2 = 0$. Supóngase que

el estado actual de la cadena X al tiempo n es el estado $j \in S$, es decir, existen j individuos del tipo \mathbf{a} al tiempo n . Denote por p_j a la probabilidad de que se produzca un individuo tipo \mathbf{a} en la siguiente generación, mientras que q_j es la probabilidad de que se produzca un individuo tipo \mathbf{A} en la siguiente generación, con $j \in S$. Por lo tanto, dado que los individuos son seleccionados uniforme e independientemente para reproducir se tiene que,

$$p_j = \frac{j}{N} \quad (2.1)$$

$$q_j = \frac{N-j}{N} = 1 - p_j. \quad (2.2)$$

Ahora si se supone que al tiempo n se tienen j individuos del tipo \mathbf{a} , con $j = 0, 1, 2, \dots, N$ individuos en la generación progenitora, entonces el evento se puede ver como N ensayos Bernoulli con probabilidad de éxito p_j y q_j la probabilidad de fracaso. Entonces la probabilidad de que haya k individuos del tipo \mathbf{a} en la generación $n + 1$, dado que hay j en la generación n , es:

$$\begin{aligned} P_{jk} &= P(X_{n+1} = k | X_n = j) \\ &= \binom{N}{k} p_j^k q_j^{N-k} = \binom{N}{k} \left(\frac{j}{N}\right)^k \left(\frac{N-j}{N}\right)^{N-k}, \quad j, k \in S. \end{aligned} \quad (2.3)$$

2.1.2 CASO CON PRESENCIA DE MUTACIÓN

El segundo caso es cuando se presenta mutación, es decir, cuando se tiene $\alpha_1 > 0$ y $\alpha_2 > 0$. Supóngase que el estado actual de la cadena X es $j \in S$, es decir, existen j individuos del tipo \mathbf{a} en la población actual. Entonces se denotará p_j a la probabilidad de que se produzca para la siguiente generación un descendiente tipo \mathbf{a} y q_j a la probabilidad de que se produzca para la siguiente generación un descendiente tipo \mathbf{A} , con $j \in S$, por lo tanto se tiene:

$$p_j = \frac{j}{N} (1 - \alpha_1) + \frac{N-j}{N} \alpha_2. \quad (2.4)$$

$$q_j = \frac{j}{N} \alpha_1 + \frac{N-j}{N} (1 - \alpha_2). \quad (2.5)$$

Para el caso p_j , $j \in S$, la expresión obtenida se debe a que existen dos eventos mutuamente excluyentes de que el descendiente producido sea del tipo **a**. El primer evento consiste en seleccionar un individuo del tipo **a** para reproducirse y que el descendiente no sea un mutante, ocurre con probabilidad $(j/N)(1 - \alpha_1)$. El segundo evento consiste en seleccionar un individuo del tipo **A** para que reproduzca un descendiente mutante, esto ocurre con probabilidad $[(N - j)/N]\alpha_2$. Para el caso de q_j , $j \in S$, se realiza un razonamiento análogo.

Ahora si se considera como un evento exitoso a la producción de un individuo del tipo **a**, con $j = 0, 1, 2, \dots, N$ individuos en la generación progenitora, entonces el evento se puede ver como N ensayos Bernoulli con probabilidad de éxito p_j y q_j la probabilidad de fracaso. Entonces la probabilidad de que haya k individuos del tipo **a** en la generación $n + 1$, dado que hay j en la generación n , es:

$$P_{jk} = P(X_{n+1} = k | X_n = j) = \binom{N}{k} p_j^k q_j^{N-k}, \quad j, k \in S. \quad (2.6)$$

En este modelo se considera una población haploide, sin embargo es posible considerar una población diploide³, en donde el número de individuos será $2N$ (el tamaño total de la población).

2.2 MODELO DE MORAN

En esta sección desarrollamos el modelo de Moran, en un principio será a tiempo discreto y posteriormente se desarrolla el modelo a tiempo continuo.

La demografía de población con la cual el modelo evoluciona es de la siguiente forma: se selecciona un individuo en cada unidad de tiempo, de manera uniforme e independiente y con reemplazo para producir un descendiente; además se selecciona un individuo de manera uniforme e independiente para morir. El proceso de mutación en la población está dado de igual forma como se menciona en el modelo de Wright-Fisher.

³Cada individuo tiene dos progenitores.

2.2.1 CASO SIN MUTACIÓN

En primer lugar se presentará el modelo a tiempo discreto, para el caso cuando no existe mutación ($\alpha_1 = \alpha_2 = 0$) y cuando existe mutación ($\alpha_1 > 0$ y $\alpha_2 > 0$).

Se tiene que $X = \{X_n: n = 0, 1, 2, \dots\}$ es una cadena de Markov que cuenta el número de individuos del tipo **a** después de n eventos de reproducción y notesé que tiene espacios de estados $S = \{0, 1, 2, \dots, N\}$.

El primer caso es cuando no se presenta mutación, es decir, cuando se tiene $\alpha_1 = \alpha_2 = 0$. Tomando en cuenta que la probabilidad de que el individuo **a** sea seleccionado para reproducirse es $p_j = \frac{j}{N}$ y la probabilidad de que el individuo **A** sea seleccionado para morir es $q_j = \frac{N-j}{N}$, se tiene que las probabilidades de transición del proceso se expresan de la siguiente manera:

$$P_{jj-1} = \frac{j}{N} q_j. \tag{2.7}$$

$$P_{jj} = \frac{j}{N} p_j + \left(1 - \frac{j}{N}\right) q_j. \tag{2.8}$$

$$P_{jj+1} = \left(1 - \frac{j}{N}\right) p_j. \tag{2.9}$$

$$P_{jk} = 0 \quad |k - j| > 1, j, k \in S. \tag{2.10}$$

Se afirma que (2.7), (2.8), (2.9) y (2.10) son válidas, por ejemplo para (2.9) basta notar que la cadena pasa del estado j al $j + 1$ si y sólo si un individuo del tipo **a** es seleccionado para reproducirse y un individuo del tipo **A** es seleccionado para morir.

Mientras que en (2.8), se tiene el caso donde no hay cambios en los tipos de individuos. Por lo tanto se puede expresar como la unión de dos eventos mutuamente excluyentes, dichos eventos son:

- i) Un individuo del tipo **a** es seleccionado para reproducirse y un individuo del tipo **a** es seleccionado para morir.
- ii) Un individuo del tipo **A** es seleccionado para reproducirse y un individuo del tipo **A** es seleccionado para morir.

Para los otros casos el razonamiento es semejante.

2.2.2 CASO CON PRESENCIA DE MUTACIÓN

El segundo caso es cuando se presenta mutación, es decir, cuando se tiene $\alpha_1 > 0$ y $\alpha_2 > 0$. Tomando en cuenta que la probabilidad de que el individuo **a** sea seleccionado para reproducirse es $p_j = \frac{j}{N}(1 - \alpha_1) + \frac{N-j}{N}\alpha_2$ y la probabilidad de que el individuo **A** sea seleccionado para morir es $q_j = \frac{j}{N}\alpha_1 + \frac{N-j}{N}(1 - \alpha_2)$, tenemos que las probabilidades de transición del proceso se expresan de la siguiente manera:

$$P_{jj-1} = \frac{j(N-j)}{N^2}(1 - \alpha_2) + \left(\frac{j}{N}\right)^2 \alpha_1, \quad j = 1, \dots, N. \quad (2.11)$$

$$P_{jj} = \left(\frac{N-j}{N}\right)^2(1 - \alpha_2) + \left(\frac{j}{N}\right)^2(1 - \alpha_1) + \frac{j(N-j)}{N^2}(\alpha_1 + \alpha_2) \quad (2.12)$$

$$P_{jj+1} = \frac{j(N-j)}{N^2}(1 - \alpha_1) + \left(\frac{N-j}{N}\right)^2 \alpha_2, \quad j = 0, 1, \dots, N-1. \quad (2.13)$$

$$P_{jk} = 0, \quad \text{si } |k - j| > 1, j, k \in S. \quad (2.14)$$

Se afirma que (2.11), (2.12), (2.13) y (2.14) son válidas, por ejemplo (2.13) esta determinada por la probabilidad de que haya un incremento en los individuos del tipo **a**, esto puede ocurrir de dos formas: la primera es que se elija un individuo del tipo **A** para morir y otro del tipo **a** para reproducirse y que no mute; la segunda es que se elija un individuo del tipo **A** para que se reproduzca y produzca un mutante y que se elija otro individuo del tipo **A** para que muera, puede ocurrir que sea el mismo individuo.

Para los otros casos el razonamiento es semejante.

2.2.3 MODELO DE MORAN A TIEMPO CONTINUO

Ahora se presenta el modelo a tiempo continuo. Como la elección de un individuo no depende del tiempo, se tiene que el proceso que describe la demografía y el proceso de mutación son los mismos que en el caso del modelo a tiempo discreto. Sin embargo, los eventos de nacimiento y muerte ocurren en intervalos de tiempo exponencial con parámetro $\lambda_i > 0$, con $i \in S$.

Denote τ_i al tiempo que el proceso se queda en el estado i antes de hacer una transición a un estado diferente, entonces:

$$P[\tau_i > s + t | \tau_i > s] = P[\tau_i > t], \quad s, t \geq 0,$$

es decir, τ_i es Markoviana lo cual implica que τ_i es exponencial. Por lo tanto la cadena de Markov en tiempo continuo que cuenta el número de individuos del tipo **a** es $X = \{X_t : t \geq 0\}$, cumple que $P(\tau_i > t) = \exp(-\lambda_i t)$ en un tiempo $\tau_i, t \geq 0$.

Observación: Recuerde que un proceso estocástico tiene la propiedad Markoviana si las probabilidades de transición en un paso sólo dependen del estado del sistema en el período de paro más reciente.

De esta forma se tiene que las probabilidades de transición del proceso se expresan de la siguiente manera:

$$P(X_{t+h} = j - 1 | X_t = j) = \lambda \left(\frac{j(N-j)}{N^2} (1 - \alpha_2) + \left(\frac{j}{N} \right)^2 \alpha_1 \right) h + \mathcal{O}(h). \quad (2.15)$$

$$P(X_{t+h} = j | X_t = j) = \lambda \left[\left(\frac{N-j}{N} \right)^2 (1 - \alpha_2) + \left(\frac{j}{N} \right)^2 (1 - \alpha_1) + \frac{j(N-j)}{N^2} (\alpha_1 + \alpha_2) \right] h + \mathcal{O}(h). \quad (2.16)$$

$$P(X_{t+h} = j + 1 | X_t = j) = \lambda \left[\frac{j(N-j)}{N^2} (1 - \alpha_1) + \left(\frac{N-j}{N} \right)^2 \alpha_2 \right] h + \mathcal{O}(h). \quad (2.17)$$

Se afirma que (2.15), (2.16) y (2.17) son válidas. La justificación es similar que en el caso discreto.

Se tiene que para el caso con mutación la tasa de nacimiento λ_i y muerte μ_i de la cadena X están dadas por:

$$\lambda_i = \lambda \left(\frac{i(N-i)(1 - \alpha_1) + (N-i)^2 \alpha_2}{N^2} \right), \quad i = 0, 1, 2, \dots, N-1, \quad (2.18)$$

$$\mu_i = \lambda \left(\frac{i(N-i)(1 - \alpha_2) + i^2 \alpha_1}{N^2} \right), \quad i = 1, 2, \dots, N. \quad (2.19)$$

De esta forma las probabilidades de transición de la cadena de saltos asociada a X están dadas por:

$$\begin{aligned}
 P_{ii+1} &= \frac{\lambda_i}{\lambda_i + \mu_i} \\
 &= \frac{i(N-1)(1-\alpha_2) + i^2\alpha_1}{i(N-i)(2-\alpha_1-\alpha_2) + (N-i)^2\alpha_2 + i^2\alpha_1}, \quad i = 0, 1, 2, \dots, N-1 \\
 P_{ii-1} &= \frac{\mu_i}{\lambda_i + \mu_i} \\
 &= \frac{i(N-1)(1-\alpha_1) + (N-i)^2\alpha_2}{i(N-i)(2-\alpha_1-\alpha_2) + (N-i)^2\alpha_2 + i^2\alpha_1}, \quad i = 1, 2, \dots, N \\
 P_{ij} &= 0, \quad \text{en otro caso.}
 \end{aligned}$$

Dado que X representa el proceso de nacimiento y muerte irreducible con espacio de estados $S = \{0, 1, \dots, N\}$ finito y la cadena de saltos $X^s = \{X_n^s : n = 0, 1, \dots\}$ asociada a X es recurrente positiva, entonces X tiene una única distribución estacionaria y coincide con su distribución límite.

Teorema 2.2.1 *Sea $X = \{X_t : t \geq 0\}$ la cadena de Markov con espacio de estados $S = \{0, 1, \dots, N\}$, que describe la evolución de una población de acuerdo al modelo de Moran con mutación. Entonces la distribución estacionaria de X , que es igual a la distribución límite está dada por:*

$$P(k) = \frac{\Gamma(N+1)\Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right)\Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2} + k\right)\Gamma\left(\frac{(1-\alpha_1)}{1-\alpha_1-\alpha_2} - k\right)}{\Gamma(k+1)\Gamma(N-k+1)\Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2}\right)\Gamma\left(\frac{N}{1-\alpha_1-\alpha_2}\right)\Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right)} \quad (2.20)$$

para todo $k \in S$, donde

$$\Gamma(\alpha) = \begin{cases} (\alpha-1)! & \text{si } \alpha \text{ es un entero} \\ \int_0^\infty y^{\alpha-1} \exp^{-y} dy & \text{en otro caso.} \end{cases}$$

Demostración. Dado que X es un proceso de nacimiento y muerte entonces su distribución límite está dada por:

$$P(k) = P(0) \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_n}, \quad k = 1, 2, \dots, N. \quad (2.21)$$

donde

$$P(0) = \frac{1}{1 + \sum_{n=1}^N \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}$$

y como S el espacio de estados es finito se tiene que:

$$\sum_{n=1}^N \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty.$$

Sólo basta probar que las tasas de nacimiento y muerte producen la expresión (2.20).

Utilizando las tasa de nacimiento y muerte vistas como:

$$\lambda_k = \lambda \left(\frac{N-k}{N} \right) p_k$$

$$\mu_{k+1} = \lambda \left(\frac{k+1}{N} \right) q_{k+1}$$

Se tiene que:

$$\begin{aligned} P(k) &= P(0) \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \\ &= P(0) \left(\prod_{j=0}^{k-1} \frac{\lambda_j}{\mu_{j+1}} \right) \\ &= P(0) \left(\prod_{j=0}^{k-1} \frac{N-j}{j+1} \frac{p_j}{q_{j+1}} \right) \\ &= P(0) \frac{N!}{k! (N-k)!} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \\ &= P(0) \binom{N}{k} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \end{aligned}$$

por otro lado se tiene que,

$$\begin{aligned}
 P(0) &= \left[1 + \sum_{k=1}^N \binom{N}{k} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \right]^{-1} \\
 &= \left[\frac{1}{q_1 q_2 \cdots q_N} \left(q_1 q_2 \cdots q_N + \sum_{k=1}^N \binom{N}{k} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} q_1 q_2 \cdots q_N \right) \right]^{-1} \\
 &= \left[\frac{1}{q_1 q_2 \cdots q_N} \left(q_1 q_2 \cdots q_N + \sum_{k=1}^N \binom{N}{k} p_0 p_1 \cdots p_{k-1} q_{k+1} q_{k+2} \cdots q_N \right) \right]^{-1}.
 \end{aligned}$$

Se denota como $S_N(x, y)$ al término entre paréntesis en la ecuación anterior y por el **Lema A.1** se tiene:

$$S_N(x, y) = \prod_{k=1}^N \left(1 - \frac{k}{N} (1 - \alpha_1 - \alpha_2) \right).$$

Por otro lado, de (2.5) se obtiene:

$$q_1 q_2 \cdots q_N = \prod_{k=1}^N \left(1 - \alpha_2 - \frac{k}{N} (1 - \alpha_1 - \alpha_2) \right)$$

entonces:

$$\begin{aligned}
 P(0) &= \frac{\prod_{k=1}^N (1 - \alpha_2 - \frac{k}{N} (1 - \alpha_1 - \alpha_2))}{\prod_{k=1}^N (1 - \frac{k}{N} (1 - \alpha_1 - \alpha_2))} \\
 &= \frac{\prod_{k=1}^N \left(\frac{N(1-\alpha_2)}{N} - \frac{k}{N} (1 - \alpha_1 - \alpha_2) \right)}{\prod_{k=1}^N \left(\frac{N}{N} - \frac{k}{N} (1 - \alpha_1 - \alpha_2) \right)} \\
 &= \frac{\frac{1-\alpha_1-\alpha_2}{N} \prod_{k=1}^N \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k \right)}{\frac{1-\alpha_1-\alpha_2}{N} \prod_{k=1}^N \left(\frac{N}{1-\alpha_1-\alpha_2} - k \right)} \\
 &= \frac{\prod_{k=1}^N \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k \right)}{\prod_{k=1}^N \left(\frac{N}{1-\alpha_1-\alpha_2} - k \right)}
 \end{aligned}$$

y por el **Lema A.3** se tiene que:

$$\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right) = \prod_{k=1}^N \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right) \Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right)$$

y

$$\Gamma\left(\frac{N}{1-\alpha_1-\alpha_2}\right) = \prod_{k=1}^N \left(\frac{N}{1-\alpha_1-\alpha_2} - k\right) \Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right)$$

entonces:

$$\begin{aligned} P(0) &= \frac{\Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right) \prod_{k=1}^N \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right) \Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right) \prod_{k=1}^N \left(\frac{N}{1-\alpha_1-\alpha_2} - k\right) \Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right)} \\ &= \frac{\Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N}{1-\alpha_1-\alpha_2}\right)} \end{aligned} \quad (2.22)$$

Ahora como,

$$\begin{aligned} q_1 q_2 \cdots q_k &= \prod_{j=1}^k \left(1 - \alpha_2 - \frac{j}{N} (1 - \alpha_1 - \alpha_2)\right) \\ &= \frac{1 - \alpha_1 - \alpha_2}{N} \prod_{j=1}^k \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - j\right) \\ &= \frac{1 - \alpha_1 - \alpha_2}{N} \frac{\prod_{j=1}^k \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - j\right) \Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right)}{\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right)} \end{aligned}$$

y dado que,

$$\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right) = \prod_{j=1}^k \left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - j\right) \Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2} - k\right)$$

entonces:

$$q_1 q_2 \cdots q_k = \frac{1 - \alpha_1 - \alpha_2}{N} \frac{\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right)}. \quad (2.23)$$

Por (2.4) se tiene que,

$$\begin{aligned}
 p_0 p_1 \cdots p_{k-1} &= \prod_{j=0}^{k-1} \left(\alpha_2 + \frac{j}{N} (1 - \alpha_1 - \alpha_2) \right) \\
 &= \frac{(1 - \alpha_1 - \alpha_2)}{N} \prod_{j=0}^{k-1} \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + j \right) \\
 &= \frac{(1 - \alpha_1 - \alpha_2)}{N} \prod_{j=1}^k \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k - j \right) \\
 &= \frac{\frac{(1 - \alpha_1 - \alpha_2)}{N} \prod_{j=1}^k \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k - j \right) \Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} \right)}{\Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} \right)}
 \end{aligned}$$

y por el **Lema A.3**, se cumple:

$$\Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k \right) = \prod_{j=1}^k \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k - j \right) \Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} \right)$$

entonces:

$$p_0 p_1 \cdots p_{k-1} = \frac{\frac{1 - \alpha_1 - \alpha_2}{N} \Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k \right)}{\Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} \right)}. \quad (2.24)$$

Por lo tanto de (2.23) y (2.24) se sigue,

$$\frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} = \frac{\frac{1 - \alpha_1 - \alpha_2}{N} \Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k \right)}{\Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} \right)} \frac{\Gamma \left(\frac{N(1 - \alpha_2)}{1 - \alpha_1 - \alpha_2} \right)}{\frac{1 - \alpha_1 - \alpha_2}{N} \frac{\Gamma \left(\frac{N(1 - \alpha_2)}{1 - \alpha_1 - \alpha_2} - k \right)}{\Gamma \left(\frac{N(1 - \alpha_2)}{1 - \alpha_1 - \alpha_2} - k \right)}}$$

y entonces

$$\frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} = \frac{\Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} + k \right) / \Gamma \left(\frac{N\alpha_2}{1 - \alpha_1 - \alpha_2} \right)}{\Gamma \left(\frac{N(1 - \alpha_2)}{1 - \alpha_1 - \alpha_2} \right) / \Gamma \left(\frac{N(1 - \alpha_2)}{1 - \alpha_1 - \alpha_2} - k \right)}. \quad (2.25)$$

Por lo tanto,

$$\begin{aligned}
 P(k) &= P(0) \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \\
 &= P(0) \frac{N!}{k! (N-k)!} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \\
 &= P(0) \frac{\Gamma(N+1)}{\Gamma(k+1) \Gamma(N-k+1)} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \\
 &= P(0) \frac{\Gamma(N+1)}{\Gamma(k+1) \Gamma(N-k+1)} \frac{\Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2} + k\right) / \Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right) / \Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right)} \\
 &= \frac{\Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N}{1-\alpha_1-\alpha_2}\right)} \frac{\Gamma(N+1)}{\Gamma(k+1) \Gamma(N-k+1)} \\
 &\quad \frac{\Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2} + k\right) / \Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right) / \Gamma\left(\frac{N(1-\alpha_2)}{1-\alpha_1-\alpha_2} - k\right)} \\
 &= \frac{\Gamma(N+1) \Gamma\left(\frac{N(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2} + k\right) \Gamma\left(\frac{(1-\alpha_1)}{1-\alpha_1-\alpha_2} - k\right)}{\Gamma(k+1) \Gamma(N-k+1) \Gamma\left(\frac{N\alpha_2}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{N\alpha_1}{1-\alpha_1-\alpha_2}\right)}.
 \end{aligned}$$

2.3 APROXIMACIÓN POR DIFUSIÓN

En esta sección se realiza una aproximación para el proceso que describe una población que evoluciona de acuerdo a los modelos de Wright-Fisher y Moran. También se presentará la distribución estacionaria correspondiente a cada modelo. El proceso de difusión que se obtiene para el modelo de Wright-Fisher resulta ser el mismo que se obtiene para el modelo de Moran.

2.3.1 APROXIMACIÓN POR DIFUSIÓN PARA EL MODELO DE WRIGHT-FISHER

Considérese un población de tamaño N con las hipótesis mencionadas para el modelo de Wright-Fisher y sea X_n el número de individuos de \mathbf{a} en la n -ésima generación con $n = 0, 1, 2, \dots$. Sean α y β fijas tales que

$$\alpha = \frac{\alpha_1}{N}, \quad \beta = \frac{\alpha_2}{N}$$

con $\alpha_1 > 0$ y $\alpha_2 > 0$.

Por lo tanto se tiene que la intensidad de mutación (tasa de mutación por individuo) del tipo **a** al tipo **A** es $N\alpha = \alpha_1$, y es positiva y finita, y similarmente del tipo **A** al tipo **a** es $N\beta = \alpha_2$. Considérese el proceso

$$Y_t^N = \frac{X_{[Nt]}}{N}, t \geq 0, \quad (2.26)$$

asociado al proceso $X = \{X_n : n \geq 0\}$, donde $[x]$ representa el mayor entero menor que x .

Para la unidad $t = 1$ se tiene que el proceso $Y^N = \{Y_t^N : t \geq 0\}$ corresponde a un lapso de N generaciones del proceso X . El proceso $Y^N = \{Y_t^N : t \geq 0\}$ representa la proporción de individuos del tipo **a** en la población al tiempo $[Nt]$ del proceso X . Mientras que α y β son las tasas de mutación correspondientes a dicho proceso.

Tenemos que la difusión $Y = \{Y_t : t \geq 0\}$ es resultado de tomar el límite de Y^N cuando $N \rightarrow \infty$.

Teorema 2.3.1 *Sea $Y = \{Y_t : t \geq 0\}$ el proceso de difusión asociado al proceso $\{X_n : n \geq 0\}$, entonces los parámetros infinitesimales del proceso de difusión están dados por:*

$$\mu(x) = -x\alpha + (1-x)\beta$$

$$\sigma^2(x) = x(1-x),$$

donde $x \in [0, 1]$.

Demostración. Sea $h = 1/N$, entonces el incremento del proceso en h es:

$$\Delta Y_{t,h}^N = Y_{t+h}^N - Y_t^N = \frac{X_{[Nt]+1} - X_{[Nt]}}{N}.$$

Tome $Y^N = \frac{i}{N}$, $i \in S$, entonces por definición de Y^N se tiene que el proceso $NY_{t+\frac{1}{N}}^N$ tiene distribución Binomial con parámetros N y p_i . Por lo

tanto calculando la esperanza y tomando en cuenta (2.4) se tiene:

$$\begin{aligned}
 E \left[\Delta Y_{t,h}^N | Y_t^N = \frac{i}{N} \right] &= E \left[(Y_{t+h}^N - Y_t^N) | Y_t^N(t) = \frac{i}{N} \right] \\
 &= E \left[Y_{t+h}^N | Y_t^N(t) = \frac{i}{N} \right] - E \left[Y_t^N | Y_t^N(t) = \frac{i}{N} \right] \\
 &= \frac{1}{N} E \left[N Y_{t+h}^N | Y_t^N(t) = \frac{i}{N} \right] - \frac{i}{N} \\
 &= \frac{1}{N} N p_i - \frac{i}{N} = \left[\frac{i}{N} (1 - \alpha_1) + \frac{(N-i)}{N} \alpha_2 \right] - \frac{i}{N} \\
 &= -\alpha_1 \frac{i}{N} + \left(1 - \frac{i}{N} \right) \alpha_2,
 \end{aligned}$$

dado que $\alpha = \frac{\alpha_1}{N}$ y $\beta = \frac{\alpha_2}{N}$ se tiene que

$$E \left[\Delta Y_{t,h}^N | Y_t^N = \frac{i}{N} \right] = \frac{1}{N} \left[-\alpha \frac{i}{N} + \left(1 - \frac{i}{N} \right) \beta \right]. \quad (2.27)$$

Tomando el límite en (2.27) y suponiendo que $i/N \rightarrow x$ como $N \rightarrow \infty$, se obtiene

$$\lim_{N \rightarrow \infty} N E \left[\Delta Y_{t,h}^N | Y_t^N = x \right] = -\alpha x + (1-x) \beta,$$

y la convergencia es uniforme para $x \in [0, 1]$.

Realizando un análisis similar para la varianza infinitesimal $\sigma^2(x)$ y tomando en cuenta los siguientes resultados:

1. Sea X y Z variables aleatorias y a real, entonces se tiene que $E[aX + Z] = aE[X] + E[Z]$.
2. Sea X variable aleatoria entonces $V[X] = E[X^2] - E^2[X]$, por lo tanto $E[X^2] = V[X] + E^2[X]$. Notése que $N Y_{t+\frac{1}{N}}^N$ tiene distribución Binomial con parámetros N y p_i , entonces $E \left[(N Y_{t+h}^N)^2 \right] = N p_i (1 - p_i) + (N p_i)^2$.
3. Obsérvese que p_i y q_i están dadas por (2.4) y (2.5) correspondientemente.
4. Notése que $\alpha = \frac{\alpha_1}{N}$ y $\beta = \frac{\alpha_2}{N}$.

Se tiene que:

$$\begin{aligned}
 \sigma^2(x) &= \lim_{h \rightarrow 0} \frac{1}{h} E \left[\{\Delta Y_{t,h}^N\}^2 | Y_t^N = x \right] = \lim_{h \rightarrow 0} \frac{1}{h} E \left[(Y_{t+h}^N - Y_t^N)^2 | Y_t^N = \frac{i}{N} \right] \\
 &\stackrel{1}{=} \lim_{h \rightarrow 0} \frac{1}{h} E \left[(Y_{t+h}^N)^2 | Y_t^N = \frac{i}{N} \right] - 2E \left[(Y_{t+h}^N) (Y_t^N) | Y_t^N = \frac{i}{N} \right] + \\
 &\quad + E \left[(Y_t^N)^2 | Y_t^N = \frac{i}{N} \right] \\
 &\stackrel{1}{=} \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N^2} E \left[(N Y_{t+h}^N)^2 | Y_t^N = \frac{i}{N} \right] + \\
 &\quad - \frac{2}{N} \frac{i}{N} E \left[(N Y_{t+h}^N) | Y_t^N = \frac{i}{N} \right] + \left(\frac{i}{N} \right)^2 \\
 &\stackrel{2}{=} \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N^2} (N p_i (1 - p_i) + N^2 p_i^2) - \frac{2i}{N^2} (N p_i) + \frac{i^2}{N^2} \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} p_i (1 - p_i) + p_i^2 - 2 \frac{i}{N} p_i + \left(\frac{i}{N} \right)^2 \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} p_i (1 - p_i) + \left(p_i - \frac{i}{N} \right)^2 \\
 &\stackrel{3}{=} \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} \left(\frac{i}{N} (1 - \alpha_1) + \left(\frac{N-i}{N} \right) \alpha_2 \right) \left(\frac{i}{N} \alpha_1 + \left(\frac{N-i}{N} (1 - \alpha_2) \right) \right) \\
 &\quad + \left(\left(\frac{N-i}{N} \right) \alpha_2 - \frac{i}{N} \alpha_1 \right)^2 \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} \left\{ \frac{i}{N} \left(\frac{N-i}{N} \right) [(1 - \alpha_1) (1 - \alpha_2) + \alpha_1 \alpha_2] + \left(\frac{i}{N} \right)^2 \alpha_1 (1 - \alpha_1) \right. \\
 &\quad \left. + \left(\frac{N-i}{N} \right)^2 (\alpha_2 (1 - \alpha_2) + N \alpha_2^2) \right\} \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} \left\{ \frac{i}{N} \left(\frac{N-i}{N} \right) (1 - \alpha_1 - \alpha_2 + 2\alpha_1 \alpha_2 - 2N \alpha_1 \alpha_2) + \right. \\
 &\quad \left. \left(\frac{i}{N} \right)^2 (\alpha_1 (1 - \alpha_1) + N \alpha_1^2) + \left(\frac{N-i}{N} \right) (\alpha_2 (1 - \alpha_2) + N \alpha_2^2) \right\} \\
 &\stackrel{4}{=} \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} \left\{ \frac{i}{N} \left(\frac{N-i}{N} \right) + \frac{i}{N} \left(\frac{N-i}{N} \right) (-\alpha_1 - \alpha_2 + 2\alpha_1 \alpha_2 + 2\alpha_2) \right. \\
 &\quad \left. + \left(\frac{i}{N} \right)^2 (\alpha_1 (1 - \alpha_1) + \alpha_1) + \left(\frac{N-i}{N} \right)^2 (\alpha_2 (1 - \alpha_2) + \alpha_2) \right\} \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{1}{N} \left\{ \frac{i}{N} \left(\frac{N-i}{N} \right) + \mathcal{O} \left(\frac{1}{N} \right) \right\}.
 \end{aligned}$$

donde $\lim_{N \rightarrow \infty} \frac{o(\frac{1}{N})}{\frac{1}{N}} = 0$.

Entonces como $h = 1/N$ se tiene que

$$\lim_{N \rightarrow \infty} NE \left[\{\Delta Y_{t,h}^N\}^2 \mid Y_t^N = x \right] = x(1-x)$$

esto pues $i/N \rightarrow x$ como $N \rightarrow \infty$, y la convergencia es uniforme para $x \in [0, 1]$.

Realizando un análisis similar se obtiene:

$$\lim_{N \rightarrow \infty} NE \left[\{\Delta Y_{t,h}^N\}^4 \mid Y_t^N = x \right] = 0,$$

y la convergencia es uniforme para $x \in [0, 1]$.

El resultado anterior sugiere que el proceso Y_t^N converge cuando $N \rightarrow \infty$ a un proceso de difusión Y , cuyos parámetros infinitesimales son de la forma:

$$\mu(x) = -x\alpha + (1-x)\beta$$

$$\sigma^2(x) = x(1-x),$$

donde $x \in [0, 1]$.

Distribución estacionaria

En esta sección se obtendrá la distribución estacionaria correspondiente al proceso de difusión $Y(t)$, de acuerdo al comportamiento de los puntos frontera del espacio de estados de $Y(t)$.

Puesto que $Y(t) = \{Y_t : t \geq 0\}$ tiene como parámetros infinitesimales a $\mu(x) = -x\alpha + (1-x)\beta$ y $\sigma^2(x) = x(1-x)$, $x \in [0, 1]$, y por la **Definición 1.12** se tiene que

$$\begin{aligned} s(x) &= \exp \left(- \int^x \frac{2\mu(\xi)}{\sigma^2(\xi)} d\xi \right) \\ &= x^{-2\beta} (1-x)^{-2\alpha} \end{aligned}$$

y por lo tanto

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

entonces la función escalar y la medida de velocidad están dadas por:

$$S[a, x] = \int_a^x s(\xi) d\xi = \int_a^x \frac{d\xi}{\xi^{2\beta} (1-\xi)^{2\alpha}} \quad \alpha \in [0, 1], a < x, \quad (2.28)$$

y

$$M[a, x] = \int_a^x m(u) du = \int_a^x u^{2\beta-1} (1-u)^{2\alpha-1} du. \quad (2.29)$$

Ahora se analiza el comportamiento para las fronteras 0 y 1, ambos análisis son similares por lo tanto sólo se desarrolla el caso en que la frontera es 0.

Sea $x \in (0, 1)$ y $\xi \in (0, x]$, entonces por (2.28) se tiene:

$$\begin{aligned} S(0, x] &= \lim_{a \rightarrow 0} S[a, x] \\ &= \lim_{a \rightarrow 0} \int_a^x \frac{d\xi}{\xi^{2\beta} (1-\xi)^{2\alpha}} \\ &= \int_0^x \frac{d\xi}{\xi^{2\beta} (1-\xi)^{2\alpha}}, \end{aligned} \quad (2.30)$$

cuando $2\alpha \geq 1$ y $2\beta \geq 1$ se tiene que $S(0, x] = \infty$.

De (2.29) se tiene que:

$$\begin{aligned} M(0, \xi] &= \lim_{a \rightarrow 0} M[a, \xi] \\ &= \lim_{a \rightarrow 0} \int_a^\xi u^{2\beta-1} (1-u)^{2\alpha-1} du \\ &= \int_0^\xi u^{2\beta-1} (1-u)^{2\alpha-1} du, \end{aligned} \quad (2.31)$$

cuando $2\alpha \geq 1$ y $2\beta \geq 1$ se tiene que $M(0, \xi] < \infty$.

Por (2.30) y (2.31) se obtiene $N(0)$, definida como:

$$\begin{aligned} N(0) &= \int_x^0 M(0, \xi] dS(\xi) \\ &= \int_x^0 \left[\left(\int_0^\xi u^{2\beta-1} (1-u)^{2\alpha-1} du \right) \frac{1}{\xi^{2\beta} (1-\xi)^{2\alpha}} \right] d\xi, \end{aligned} \quad (2.32)$$

cuando $\beta > 0$ se tiene que $N(0) < \infty$, y si $\beta = 0$ sucede que $N(0) = \infty$.

Observación: Note que si el $\lim_{b \rightarrow c} \int_b^d f(x) dx$ existe y es finito, entonces $\int_c^d f(x) dx = \lim_{b \rightarrow c} \int_b^d f(x) dx$ y se dice que la integral es convergente.

Por lo tanto se tiene que, si $2\alpha \geq 1$ y $2\beta \geq 1$ entonces $S(0, x] = \infty$ y $N(0) < \infty$, con $x \in (0, 1)$. De manera similar se obtiene que $S[x, 1) = \infty$ y $N(1) < \infty$, con $x \in (0, 1)$. Entonces si $2\alpha \geq 1$ y $2\beta \geq 1$ se tiene que 0 y 1 son fronteras de entrada del espacio de estados $Y(t)$, por lo tanto $Y(t)$ tiene una única distribución estacionaria.

Teorema 2.3.2 Sea $Y(t)$ un proceso de difusión regular homogéneo con parámetros infinitesimales $\mu(x) = -x\alpha + (1-x)\beta$ y $\sigma^2(x) = x(1-x)$, $x \in [0, 1]$. Entonces la distribución estacionaria $\varphi(y)$ para $2\alpha \geq 1$ y $2\beta \geq 1$ es única y está dada por:

$$\varphi(y) = \frac{\Gamma(2\alpha + 2\beta) (1-y)^{2\alpha-1} y^{2\beta-1}}{\Gamma(2\alpha) \Gamma(2\beta)}$$

Demostración. Como $2\alpha \geq 1$ y $2\beta \geq 1$ entonces se tiene que 0 y 1 son fronteras de entrada del espacio de estados $Y(t)$. Y por el **Teorema 1.4.1** se tiene que:

$$\begin{aligned} \varphi(y) &= \frac{m(y)}{\int_0^1 m(u) du} \\ &= \frac{(1-y)^{2\alpha-1} y^{2\beta-1}}{\int_0^1 u^{2\beta-1} (1-u)^{2\alpha-1} du} \\ &= \frac{(1-y)^{2\alpha-1} y^{2\beta-1}}{\int_0^1 u^{2\beta-1} (1-u)^{2\alpha-1} du} \\ &= \frac{\Gamma(2\alpha + 2\beta) (1-y)^{2\alpha-1} y^{2\beta-1}}{\Gamma(2\alpha) \Gamma(2\beta)} \end{aligned} \tag{2.33}$$

se observa que es la distribución Beta con parámetros 2β y 2α .

2.3.2 APROXIMACIÓN POR DIFUSIÓN PARA EL MODELO DE MORAN

Considérese un población de tamaño N con las hipótesis mencionadas para el modelo de Moran y sea X_t el número de individuos de \mathbf{a} al tiempo $t, t \geq 0$. Sean α y β fijas tales que

$$\alpha_1 = \alpha/N, \alpha_2 = \beta/N$$

con $\alpha_1 > 0$ y $\alpha_2 > 0$. Considérese el proceso

$$Z_t^N = \frac{X_t}{N}, \tag{2.34}$$

asociado al proceso de nacimiento y muerte $X = \{X_t : t \geq 0\}$.

Por lo tanto el proceso $Z^N = \{Z_t^N : t \geq 0\}$ representa la proporción de individuos del tipo \mathbf{a} en la población al tiempo $t, t \geq 0$ del proceso X .

Se tiene que la difusión $Z = \{Z_t : t \geq 0\}$ es resultado de tomar el límite de Z^N cuando $N \rightarrow \infty$.

Teorema 2.3.3 *Sea $Z = \{Z_t : t \geq 0\}$ el proceso de difusión asociado al proceso $\{X_t : t \geq 0\}$, entonces los parámetros infinitesimales del proceso de difusión están dados por:*

$$\mu(x) = -x\alpha + (1-x)\beta$$

$$\sigma^2(x) = x(1-x),$$

donde $x \in [0, 1]$.

Demostración. Sea $h = 1/N$ y $\lambda = N^2$, entonces el incremento del proceso en h es:

$$\Delta Z_{t,h}^N = Z_{t+h}^N - Z_t^N.$$

Por lo tanto calculando la esperanza se tiene:

$$\begin{aligned}
 E \left[\Delta Z_{t,h}^N | Z_t^N = \frac{i}{N} \right] &= E \left[(Z_{t+h}^N - Z_t^N) | Z_t^N = \frac{i}{N} \right] \\
 &= E \left[\frac{X_{t+h}}{N} - \frac{X_t}{N} | \frac{X_t}{N} = \frac{i}{N} \right] \\
 &= \frac{1}{N} E [X_{t+h} - X_t | X_t = i] \\
 &= \frac{1}{N} [P \{X_{t+h} - X_t = 1 | X_t = i\} - P \{X_{t+h} - X_t = -1 | X_t = i\}] \\
 &= \frac{1}{N} [P \{X_{t+h} = i + 1 | X_t = i\} - P \{X_{t+h} = i - 1 | X_t = i\}]
 \end{aligned}$$

y por (2.15), (2.16) y (2.17) se tiene que:

$$\begin{aligned}
 E \left[\Delta Z_{t,h}^N | Z_t^N = \frac{i}{N} \right] &= \frac{1}{N} \left\{ \lambda \left(\frac{N-i}{N} \right) p_i h + \mathcal{O}_1(h) - \lambda \left(\frac{i}{N} \right) q_i h + \mathcal{O}_2(h) \right\} \\
 &= \frac{1}{N} \left\{ \lambda h \left(\left(\frac{N-i}{N} \right) p_i - \frac{i}{N} q_i \right) + \mathcal{O}(h) \right\} \\
 &= \frac{1}{N} \left[\frac{N^2}{N} \left(\left(\frac{N-i}{N} \right) \left(\frac{i}{N} (1 - \alpha_1) + \frac{N-i}{N} \alpha_2 \right) \right) \right. \\
 &\quad \left. - \frac{i}{N} \left(\frac{i}{N} \alpha_1 + \frac{N-i}{N} (1 - \alpha_2) + \mathcal{O}(h) \right) \right] \\
 &= \left(\frac{N-i}{N} \right) \left(\frac{i}{N} \right) (1 - \alpha_1 - 1 - \alpha_2) + \left(\frac{N-i}{N} \right)^2 \alpha_2 + \\
 &\quad - \left(\frac{i}{N} \right)^2 \alpha_1 + \frac{\mathcal{O}(1/N)}{N} \\
 &= \frac{1}{N^2} \left[(N-i)(\alpha_2 - \alpha_1)i + (N-i)^2 \alpha_2 - i^2 \alpha_1 \right] + \frac{\mathcal{O}(1/N)}{N} \\
 &= \frac{1}{N^2} [Ni\alpha_2 - Ni^2\alpha_1 - i^2\alpha_2 + i^2\alpha_2 + N^2\alpha_2 \\
 &\quad - 2Ni\alpha_2 + i^2\alpha_2 - i^2\alpha_1] + \frac{\mathcal{O}(1/N)}{N} \\
 &= \frac{1}{N^2} [-Ni\alpha_1 - Ni\alpha_2 + N^2\alpha_2] + \frac{\mathcal{O}(1/N)}{N}
 \end{aligned}$$

y como $\alpha_1 = \alpha/N$ y $\alpha_2 = \beta/N$, se tiene

$$\begin{aligned}
 E \left[\Delta Z_{t,h}^N | Z_t^N = \frac{i}{N} \right] &= \frac{1}{N^2} \left[-Ni \left(\frac{\alpha}{N} \right) - Ni \left(\frac{\beta}{N} \right) + N^2 \left(\frac{\beta}{N} \right) \right] + \frac{\mathcal{O}(1/N)}{N} \\
 &= \frac{1}{N^2} [-\alpha i - \beta i + N\beta] + \frac{\mathcal{O}(1/N)}{N} \\
 &= \frac{1}{N} \left\{ -\alpha \frac{i}{N} - \beta \frac{i}{N} + \beta + \mathcal{O}(1/N) \right\} \\
 &= \frac{1}{N} \left[-\alpha \frac{i}{N} + \left(1 - \frac{i}{N} \right) \beta + \mathcal{O}(1/N) \right]
 \end{aligned}$$

nótese que $\lim_{N \rightarrow \infty} \frac{\mathcal{O}(\frac{1}{N})}{\frac{1}{N}} = 0$.

Tomando el límite en la ecuación anterior y suponiendo que $i/N \rightarrow x$ como $N \rightarrow \infty$, se obtiene

$$\lim_{h \rightarrow 0^+} NE \left[\Delta Z_{t,h}^N | Z_t^N = x \right] = -\alpha x + (1-x)\beta,$$

y la convergencia es uniforme para $x \in [0, 1]$.

Y por lo tanto realizando un análisis similar se obtiene que:

$$\begin{aligned}
 E \left[\{ \Delta Z_{t,h}^N \}^2 | Y_t^N = x \right] &= E \left[(Z_{t+h}^N - Z_t^N)^2 | Z_t^N = \frac{i}{N} \right] \\
 &= E \left[\frac{1}{N^2} (X_{t+h} - X_t)^2 | X_t = i \right] \\
 &= \frac{1}{N^2} E \left[(X_{t+h} - X_t)^2 | X_t = i \right] \\
 &= \frac{1}{N^2} [P \{ X_{t+h} - X_t = 1 | X_t = i \} + P \{ X_{t+h} - X_t = -1 | X_t = i \}] \\
 &= \frac{1}{N^2} [(1)^2 P \{ X_{t+h} = i + 1 | X_t = i \} + (-1)^2 P \{ X_{t+h} = i - 1 | X_t = i \}] \\
 &= \frac{1}{N^2} [P \{ X_{t+h} = i + 1 | X_t = i \} + P \{ X_{t+h} = i - 1 | X_t = i \}],
 \end{aligned}$$

y por (2.15), (2.16) y (2.17) se tiene que:

$$\begin{aligned}
 E \left[\{\Delta Z_{t,h}^N\}^2 | Y_t^N = x \right] &= \frac{1}{N^2} \left\{ \lambda h \left(\left(\frac{N-i}{N} \right) p_i + \frac{i}{N} q_i \right) + \mathcal{O}(h) \right\} \\
 &= \frac{1}{N^2} \left\{ \lambda h \left(\frac{N-i}{N} \right) \left(\frac{i}{N} (1-\alpha_1) + \left(\frac{N-i}{N} \right) \alpha_2 \right) + \right. \\
 &\quad \left. + \frac{i}{N} \left(\frac{i}{N} \alpha_1 + \left(\frac{N-i}{N} \right) (1-\alpha_2) \right) + \mathcal{O}(h) \right\} \\
 &= \frac{1}{N^2} \left\{ \frac{\lambda h}{N^2} \left(i(N-i)(1-\alpha_1) + (N-i)^2 \alpha_2 + i^2 \alpha_1 + \right. \right. \\
 &\quad \left. \left. + i(N-i)(1-\alpha_2) + \mathcal{O}(h) \right) \right\} \\
 &= \frac{1}{N^2} \left\{ \lambda h 2 \frac{i}{N} \left(\frac{N-i}{N} \right) - \frac{i}{N} \left(\frac{N-i}{N} \right) (\alpha_1 + \alpha_2) + \alpha_2 + \right. \\
 &\quad \left. - 2 \frac{i}{N} \alpha_2 + \left(\frac{i}{N} \right)^2 (\alpha_1 + \alpha_2) + \mathcal{O}(h) \right\}
 \end{aligned}$$

y como $\alpha_1 = \alpha/N, \alpha_2 = \beta/N, \lambda = N^2/2$ y $h = 1/N$ se tiene que

$$\begin{aligned}
 E \left[\{\Delta Z_{t,h}^N\}^2 | Y_t^N = x \right] &= \frac{N^2}{N^2(2N)} \left\{ 2 \frac{i}{N} \left(\frac{N-i}{N} \right) - \frac{i}{N} \left(\frac{N-i}{N} \right) \left(\frac{\alpha+\beta}{N} \right) + \right. \\
 &\quad \left. + \frac{\beta}{N} - 2 \frac{i}{N} \frac{\beta}{N} + \left(\frac{i}{N} \right)^2 \left(\frac{\alpha+\beta}{N} \right) \right\} + \frac{1}{N^2} \mathcal{O}(h) \\
 &= \frac{1}{N} \left(\frac{i}{N} \left(1 - \frac{i}{N} \right) \right) + \\
 &\quad + \frac{1}{2N} \left(- \frac{i}{N} \left(\frac{N-i}{N} \right) \left(\frac{\alpha+\beta}{N} \right) + \frac{\beta}{N} - 2 \frac{i\beta}{N^2} + \left(\frac{i}{N} \right)^2 \left(\frac{\alpha+\beta}{N} \right) \right) + \\
 &\quad + \frac{1}{N^2} \mathcal{O}(1/N).
 \end{aligned}$$

Entonces,

$$\lim_{N \rightarrow \infty} N E \left[\{\Delta Z_{t,h}^N\}^2 | Z_t^N(t) = x \right] = x(1-x)$$

esto pues $i/N \rightarrow x$ como $N \rightarrow \infty$, y la convergencia es uniforme para $x \in [0, 1]$.

Realizando un análisis similar se obtiene:

$$\lim_{N \rightarrow \infty} \frac{1}{h} E \left[\{\Delta Z_{t,h}^N\}^4 | Z_t^N = x \right] = 0.$$

El resultado anterior sugiere que el proceso Z_t^N converge cuando $N \rightarrow \infty$ a un proceso de difusión Z , cuyos parámetros infinitesimales son de la forma:

$$\mu(x) = -x\alpha + (1-x)\beta$$

$$\sigma^2(x) = x(1-x),$$

donde $x \in [0, 1]$.

En conclusión se tiene que los parámetros infinitesimales del proceso de difusión tanto para el modelo de Wright-Fisher como para el modelo de Moran son los mismos, en consecuencia la distribución estacionaria para la versión difusión del modelo de Moran es la misma que para la versión difusión del modelo de Wright-Fisher.

CAPÍTULO 3

MODELO GENÉTICO CON RECOMBINACIÓN

En este capítulo se analiza el modelo genético de Moran con recombinación en los casos univariado y multivariado. La recombinación genética es un proceso que lleva a la obtención de un nuevo genotipo¹ a través del intercambio de material genético entre secuencias de DNA en cromosomas homólogos de dos orígenes diferentes.

3.1 CASO UNIVARIADO

Considérese una población diploide con N individuos y cada individuo tiene un número par de cromosomas homólogos. Cuando dos loci² no están vinculados se tiene que los dos loci se segregan independientemente. Supóngase que el *locus 1* tiene dos posibles alelos denotados por **A** y **a**, y el *locus 2* tiene dos posibles alelos denotados por **B** y **b**. Un individuo puede expresar su genotipo de la siguiente manera:

$$\begin{array}{cc} l_1 & l_2 \\ \frac{A}{a} & \frac{B}{B} \end{array} \begin{array}{l} \longleftarrow \\ \longleftarrow \end{array} \begin{array}{l} \text{Cromosoma} \\ \text{Cromosoma} \end{array}$$

Donde l_1 y l_2 representan el *locus 1* y *locus 2* respectivamente.

Las alternativas de los pares de cromosomas son:

$$\frac{AB}{AB}, \frac{AB}{Ab}, \frac{Ab}{Ab}, \frac{AB}{aB}, \frac{AB}{ab}, \frac{Ab}{aB}, \frac{Ab}{ab}, \frac{aB}{aB}, \frac{aB}{ab}, \frac{ab}{ab}$$

¹Contenido genético de un individuo, en forma de ADN.

²Plural de locus.

En la formación de la siguiente generación, el individuo produce (segrega) gametos³. En general se tiene que dos gametos en la población van a unirse para formar un individuo de la siguiente generación. En la segregación de los gametos es donde puede ocurrir la recombinación. Así un $\frac{Ab}{aB}$ individuo puede producir un Ab -gameto o un aB -gameto, pero por recombinación se puede producir un AB - y ab -gametos. Cuando dos loci están ligados se tiene que la recombinación de gametos es menos probable que cuando no están ligados. Sea r el parámetro que cuantifica la vinculación, es decir, r es la probabilidad de una recombinación.

Supóngase que la población inicialmente tiene solamente haplotipos Ab y aB , donde todos los individuos son de los genotipos $\frac{Ab}{Ab}, \frac{Ab}{aB}$ o $\frac{aB}{aB}$.

Si r la probabilidad de recombinación es pequeña, entonces es posible que muchas generaciones pasen sin que aparezca un individuo con un cromosoma recombinado producido por un genotipo $\frac{Ab}{aB}$. Durante este tiempo, podría ser que la población alcanzara un estado en el cual cada individuo sea del tipo $\frac{Ab}{Ab}$ si la población no puede salir de ese estado, es decir, sólo si Ab -gametos pueden ser producidos, la población se fijó sobre Ab . Similarmente se puede fijar sobre aB .

Supóngase que la recombinación todavía no ha ocurrido y que existen $2N$ gametos de N individuos en la población. Además que existen i del tipo Ab , y por lo tanto $2N - i$ del tipo aB . Para describir a la población no sólo se necesita la frecuencia de los gametos si no también la frecuencia de los genotipos, es decir, como los gametos se aparean en los individuos. El apareamiento uniforme de los individuos en la generación previa, la cual es equivalente a la unión uniforme de sus gametos, implica la frecuencia del genotipo. Entonces, las probabilidades de producir los individuos de tipo $\frac{Ab}{Ab}, \frac{Ab}{aB}$ y $\frac{aB}{aB}$ son $\left(\frac{i}{2N}\right)^2, 2\left(\frac{i}{2N}\right)\left(1 - \frac{i}{2N}\right)$ y $\left(1 - \frac{i}{2N}\right)^2$ respectivamente.

De esta población, se tiene que un Ab -gameto puede ser producido por un $\frac{Ab}{Ab}$ con probabilidad 1 o por un $\frac{Ab}{aB}$ con probabilidad $\frac{1}{2}(1 - r)$. Esto se debe a que un $\frac{Ab}{Ab}$ individuo sólo producir un Ab -gameto, mientras que un $\frac{Ab}{aB}$ individuo puede produce cuatro tipos de cromosomas Ab, aB, AB y ab con probabilidades $\frac{1}{2}(1 - r), \frac{1}{2}(1 - r), \frac{1}{2}r$ y $\frac{1}{2}r$ respectivamente.

Proposición 3.1.1 *Sea p_i la probabilidad de que un gameto producido sea Ab y q_i la probabilidad de que un gameto producido sea aB y por lo tanto*

³Células sexuales.

la probabilidad de que un gameto producido sea de un tipo recombinante es $1 - (p_i + q_i)$, entonces

$$p_i = \left(\frac{i}{2N}\right) - r \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right) \quad (3.1)$$

y

$$q_i = 1 - \frac{i}{2N} - r \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right) \quad (3.2)$$

Demostración. Supóngase que el número de Ab tipos es i , entonces

$$\begin{aligned} p_i &= P[\text{un gameto producido es } Ab] \\ &= \left(\frac{i}{2N}\right)^2 + 2 \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right) \frac{1}{2} (1 - r) \\ &= \left(\frac{i}{2N}\right) - r \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right). \end{aligned}$$

De igual manera se tiene que

$$\begin{aligned} q_i &= P[\text{un gameto producido es } aB] \\ &= \left(1 - \frac{i}{2N}\right)^2 + 2 \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right) \frac{1}{2} (1 - r) \\ &= 1 - \frac{i}{2N} - r \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right). \end{aligned}$$

Finalmente la probabilidad de que un gameto producido sea de un tipo recombinante es

$$1 - (p_i + q_i) = 2 \left(\frac{i}{2N}\right) \left(1 - \frac{i}{2N}\right) r = \frac{i}{N} \left(1 - \frac{i}{2N}\right) r.$$

Teorema 3.1.1 Sea p_i la probabilidad de que un gameto producido sea Ab y q_i la probabilidad de que un gameto producido sea aB , y sean $2N$ los gametos que componen a los individuos de la siguiente generación, entonces la probabilidad de transición P_{ij} con j gametos del tipo Ab y $2N - j$ del tipo aB es

$$P_{ij} = \binom{2N}{j} p_i^j q_i^{2N-j}, \quad j = 0, 1, 2, \dots, 2N.$$

Demostración. Ver Karlin y Taylor (1981).

Observación: La probabilidad de que un gameto no recombinado aparezca en la siguiente generación es $(p_i + q_i)^{2N}$, por lo tanto la probabilidad de que uno o más gametos recombinados aparezcan es $1 - (p_i + q_i)^{2N}$.

Nótese que se asume que la población es de tamaño N para cada generación y cada individuo sobrevive en la edad de reproducción y puede contribuir a la siguiente generación.

Aproximación a un proceso de difusión con recombinación

Supóngase ahora que se tiene una cadena de Markov con $2N + 2$ estados, de los cuales $2N + 1$ representan los estados donde el gameto no recombinado tiene aun que aparecer, y una población que contiene j Ab -gametos y $(2N - j)$ aB -gametos con $j = 0, 1, 2, \dots, 2N$. El otro estado es donde el gameto recombinado apareció en un individuo. Se admite el caso en el cual un AB -o ab -gameto pudo aparecer y por lo tanto no entrar a la siguiente generación y nunca aparecer otra vez.

Se denota a R_i como la probabilidad de que se alcance el estado recombinante R , dado que la población inicia en el estado i . Por lo tanto sea $u_i = 1 - R_i$ la probabilidad que la población eventualmente se fije, entonces se tiene que $(u_0, u_1, \dots, u_{2N})$ satisface las siguientes ecuaciones:

$$u_i = \sum_{j=0}^{2N} P_{ij} u_j \quad i = 0, 1, 2, \dots, 2N$$

con condiciones

$$u_0 = u_{2N} = 1.$$

Para N suficientemente grande la solución a la ecuación es prácticamente imposible ya sea analítica o numéricamente. Por lo tanto se realiza una aproximación del **proceso de difusión con muerte**: siguiendo el proceso de la frecuencia del Ab gameto hasta que el proceso sea "terminado" por la aparición de un gameto recombinante o porque ocurre fijación en un estado.

Considérese Y_t^N el proceso en el cual una unidad de tiempo corresponde al paso de N generaciones de la población, es decir,

$$Y_t^N = X_{[Nt], t \geq 0},$$

donde $[x]$ representa el mayor entero menor que x .

Por lo tanto tomando el límite cuando N tiende a infinito ($N \rightarrow \infty$), se tiene que el proceso Y_t^N va a tender en el límite a un proceso de difusión Y_t con muerte.

Teorema 3.1.2 *Sea Y_t un proceso de difusión con muerte asociado al proceso $\{X_n : n \geq 0\}$, entonces los parámetros infinitesimales del proceso están dados por:*

$$\begin{aligned}\mu(x) &= 0 \\ \sigma^2(x) &= \frac{x(1-x)}{2} \\ k(x) &= 4\rho x(1-x).\end{aligned}$$

Demostración. Sea X_n la variable aleatoria que indica el número de individuos de tipo Ab , entonces en la n -ésima generación, la variable X_n puede tomar valores $X_n = j/2N, j = 0, 1, 2, \dots, 2N$. Nótese que X_{n+1} es el estado donde no aparece recombinación, entonces se distribuye como $\frac{1}{2N}Z$, donde Z tiene una distribución binomial con parámetros $(2N, p_j / (p_j + q_j))$. Como la probabilidad de que no aparezca recombinación es $(p_j + q_j)^{2N}$, entonces

$$\begin{aligned}E\left[(\Delta X_n) | X_n = \frac{j}{2N} = x\right] &\stackrel{1}{=} \{E[X_{n+1} - X_n | X_n = \frac{j}{2N} = x]\} (p_j + q_j)^{2N} \\ &\stackrel{2}{=} \{E[X_{n+1} | X_n = x] - E[X_n | X_n = x]\} (p_j + q_j)^{2N} \\ &= \{E[X_{n+1} | X_n = x] - x\} (p_j + q_j)^{2N} \\ &\stackrel{3}{=} \left(\frac{2N}{2N} \frac{p_j}{p_j + q_j} - x\right) (p_j + q_j)^{2N} \\ &= \left(\frac{p_j}{p_j + q_j} - x\right) (p_j + q_j)^{2N} \\ &= [p_j - x(p_j + q_j)] (p_j + q_j)^{2N-1}.\end{aligned}$$

Nótese que:

1. Es por el hecho de que condicionamos con el evento de que no aparezca recombinación.
2. Sea X y Z variables aleatorias y a real, entonces se tiene que $E[aX + Z] = aE[X] + E[Z]$.

3. Se sabe que X_{n+1} se distribuye como $\frac{1}{2N}Z$, donde Z tiene una distribución binomial con parámetros $(2N, p_j/(p_j + q_j))$ y por lo tanto se tiene que:

$$\begin{aligned} E[X_{n+1}] &= \frac{1}{2N}[2N(p_j/(p_j + q_j))] \\ V[X_{n+1}] &= \frac{1}{4N^2}[2N(p_j/(p_j + q_j))(q_j/(p_j + q_j))]. \end{aligned}$$

Realizando un análisis similar se tiene que

$$\begin{aligned} E\left[(\Delta X^2) | X_n = \frac{j}{2N} = x\right] &\stackrel{1}{=} \{E[(X_{n+1} - X_n)^2 | X_n = \frac{j}{2N} = x]\} (p_j + q_j)^{2N} \\ &\stackrel{2}{=} \{E[(X_{n+1})^2 | X_n = x] - 2E[(X_{n+1})(X_n) | X_n = x] \\ &\quad + E[(X_n)^2 | X_n = x]\} (p_j + q_j)^{2N} \\ &\stackrel{3}{=} \left\{ \frac{1}{4N} \left[2N \left(\frac{p_j}{p_j + q_j} \right) \left(\frac{q_j}{p_j + q_j} \right) \right] + \left[\frac{2N}{2N} \left(\frac{p_j}{p_j + q_j} \right) \right]^2 \right. \\ &\quad \left. - 2x \left(\frac{p_j}{p_j + q_j} \right) + x^2 \right\} (p_j + q_j)^{2N} \\ &= \left\{ \frac{1}{2N} \left(\frac{p_j}{p_j + q_j} \right) \left(\frac{q_j}{p_j + q_j} \right) + \left(\frac{p_j}{p_j + q_j} \right)^2 \right. \\ &\quad \left. - 2x \left(\frac{p_j}{p_j + q_j} \right) + x^2 \right\} (p_j + q_j)^{2N} \\ &= \left\{ \frac{1}{2N} \left(\frac{p_j}{p_j + q_j} \right) \left(\frac{q_j}{p_j + q_j} \right) + \left(\left(\frac{p_j}{p_j + q_j} \right) - x \right)^2 \right\} (p_j + q_j)^{2N} \\ &= \left(\frac{1}{2N} \left(\frac{p_j}{p_j + q_j} \right) \left(\frac{q_j}{p_j + q_j} \right) \right. \\ &\quad \left. + \left(\frac{p_j}{p_j + q_j} - x \right)^2 \right) (p_j + q_j)^{2N} \\ &= \left[\frac{p_j q_j}{2N} + (p_j - x(p_j + q_j))^2 \right] (p_j + q_j)^{2N-2} \end{aligned}$$

y la P [”Muerto” | $X_n = j/2N = x$] = $1 - (p_j + q_j)^{2N}$, donde ”Muerto” significa que el proceso termina.

Tomando el cuenta un nuevo parámetro de recombinación r definido como

$$r = \rho/N^2 \quad \rho \text{ constante,}$$

se tiene que:

$$\begin{aligned}
 E \left[\Delta X | X_n = \frac{i}{2N} = x \right] &= [p_j - x(p_j + q_j)](p_j + q_j)^{2N-1} \\
 &\stackrel{1}{=} \left[\left(\frac{i}{2N} \right) - \frac{\rho}{N^2} \left(\frac{i}{2N} \right) \left(1 - \frac{i}{2N} \right) \right. \\
 &\quad \left. - \frac{i}{2N} \left(1 - 2 \frac{\rho}{N^2} \left(\frac{i}{2N} \right) \left(1 - \frac{i}{2N} \right) \right) \right] * \\
 &\quad \left(1 - 2 \frac{\rho}{N^2} \left(\frac{i}{2N} \right) \left(1 - \frac{i}{2N} \right) \right)^{2N-1} \\
 &\stackrel{2}{=} \left[x - \frac{\rho}{N^2} x(1-x) - x + 2 \frac{\rho}{N^2} x^2(1-x) \right] * \\
 &\quad \left(1 - 2 \frac{\rho}{N^2} x(1-x) \right)^{2N-1} \\
 &= \left[-\frac{\rho}{N^2} x(1-x) + 2 \frac{\rho}{N^2} x^2(1-x) \right] * \\
 &\quad \left(1 - 2 \frac{\rho}{N^2} x(1-x) \right)^{2N-1} \\
 &= \left[-\frac{\rho}{N^2} \frac{i}{2N} \left(1 - \frac{i}{2N} \right) + 2 \frac{\rho}{N^2} \left(\frac{i}{2N} \right)^2 \left(1 - \frac{i}{2N} \right) \right] * \\
 &\quad \left(1 - 2 \frac{\rho}{N^2} \frac{i}{2N} \left(1 - \frac{i}{2N} \right) \right)^{2N-1} \\
 &\stackrel{3}{=} \mathcal{O}(1/N^2). \tag{3.3}
 \end{aligned}$$

1. La igualdad (1) es por la definición de p_j y q_j .
2. Se sabe que $x = \frac{i}{2N}$.
3. Cada término de la ecuación es un término de orden superior $\mathcal{O}(1/N^2)$.

Por otro lado se tiene que

$$\begin{aligned}
 E \left[(\Delta X^2) \mid X_n = \frac{j}{2N} = x \right] &= \left[\frac{p_j q_j}{2N} + (p_j - x(p_j + q_j))^2 \right] (p_j + q_j)^{2N-2} \\
 &= \left\{ \frac{1}{2N} (x - r(x1 - x))(1 - x - rx(1 - x)) \right. \\
 &\quad \left. + [x - rx(1 - x) - x(1 - 2rx(1 - x))]^2 \right\} * \\
 &\quad [1 - 2rx(1 - x)]^{2N-2} \\
 &= \left\{ \frac{1}{2N} [x(1 - x) - rx^2(1 - x) - rx(1 - x)^2 \right. \\
 &\quad \left. + r^2 x^2(1 - x)^2] + [x - rx + rx^2 - x + 2rx^2(1 - x)]^2 \right\} * \\
 &\quad [1 - 2rx(1 - x)]^{2N-2} \\
 &= \left\{ \frac{1}{2N} [x(1 - x) - rx^2(1 - x) - rx(1 - x)^2 \right. \\
 &\quad \left. + r^2 x^2(1 - x)^2] + [-rx + rx^2 + 2rx^2(1 - x)]^2 \right\} * \\
 &\quad [1 - 2rx(1 - x)]^{2N-2} \\
 &= x(1 - x)/2N + \mathcal{O}(1/N^2), \tag{3.4}
 \end{aligned}$$

Además que se tiene

$$\begin{aligned}
 P[\text{"Muerto"} \mid X_n = x] &= 1 - (p_j + q_j)^{2N} \\
 &\stackrel{1}{=} 1 - \left(\sum_{k=0}^{2N} \binom{2N}{k} (1)^{2N-k} (-2rx(1-x))^k \right) \\
 &= 1 - \left(1 - \binom{2N}{1} (-2rx(1-x)) + \dots + (-2rx(1-x))^{2N} \right) \\
 &= 2N(2rx(1-x)) - \left[\binom{2N}{2} (-2rx(1-x))^2 + \dots \right. \\
 &\quad \left. \dots + (-2rx(1-x))^{2N} \right] \\
 &\stackrel{2}{=} 2N \left(2 \left(\frac{\rho}{N^2} \right) x(1-x) \right) - \left[\binom{2N}{2} (-2rx(1-x))^2 + \dots \right. \\
 &\quad \left. \dots + (-2rx(1-x))^{2N} \right] \\
 &\stackrel{3}{=} 2N \left(2 \left(\frac{\rho}{N^2} \right) x(1-x) \right) + \mathcal{O}(1/N^2) \tag{3.5}
 \end{aligned}$$

1. Por el teorema del Binomio se tiene que $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$.

2. Por definición de un nuevo parámetro de recombinación $r = \rho/N^2$.
3. Por ser términos de orden superior $\mathcal{O}(1/N^2)$.

Tomando el límite en (3.3),(3.4) y (3.5) y observando que la escala del periodo tiempo es $1/N$ unidades y además que $\lim_{N \rightarrow \infty} \frac{N}{\mathcal{O}(1/N^2)} = 0$, se tiene

$$\begin{aligned}
 \mu(x) &= \lim_{N \rightarrow \infty} \left(\frac{1}{1/N} \right) E[\Delta X | X_n = x] = 0 \\
 &= \lim_{N \rightarrow \infty} \left(\frac{1}{1/N} \right) \mathcal{O}(1/N^2) \\
 &= \lim_{N \rightarrow \infty} \frac{N}{\mathcal{O}(1/N^2)} = 0, \\
 \sigma^2(x) &= \lim_{N \rightarrow \infty} \left(\frac{1}{1/N} \right) E[(\Delta X)^2 | X_n = x] \\
 &= \lim_{N \rightarrow \infty} \left(\left(\frac{1}{1/N} \right) x(1-x)/2N + \mathcal{O}(1/N^2) \right) \\
 &= x(1-x)/2 + \lim_{N \rightarrow \infty} (\mathcal{O}(1/N^2)) \\
 &= x(1-x)/2
 \end{aligned}$$

y

$$\begin{aligned}
 k(x) &= \lim_{N \rightarrow \infty} \left(\frac{1}{1/N} \right) P[\text{“Muerto”} | X_n = x] \\
 &= \lim_{N \rightarrow \infty} \left(\frac{1}{1/N} \right) (4\rho x(1-x)/N + \mathcal{O}(1/N^2)) \\
 &= 4\rho x(1-x) + \lim_{N \rightarrow \infty} \left(\frac{1}{1/N} \right) (\mathcal{O}(1/N^2)) \\
 &= 4\rho x(1-x).
 \end{aligned}$$

Por lo tanto los parámetros infinitesimales del proceso son:

$$\begin{aligned}
 \mu(x) &= 0 \\
 \sigma^2(x) &= \frac{x(1-x)}{2} \\
 k(x) &= 4\rho x(1-x),
 \end{aligned}$$

y entonces es un proceso de difusión con muerte en el intervalo $[0, 1]$.

3.2 CASO MULTIVARIADO

La siguiente sección se basa en el artículo de *Griffiths, R.C y Rodríguez, E. R.* (2008) llamado *The age of an allele in a population with recombination.*

Considérese una población fija de tamaño N con dos loci por individuo y cada locus con dos alelos. Supóngase que el *locus 1* tiene dos posibles alelos denotados por \mathbf{A} y \mathbf{a} , y el *locus 2* tiene dos posibles alelos denotados por \mathbf{B} y \mathbf{b} . Si el alelo presentó mutación se denotará por 1, y si no tuvo mutación por 0. En general un individuo de la población se denota por un par ordenado (i, j) con $i, j = 0, 1$, donde la primera coordenada denota el estado del alelo en el *locus 1* y la segunda coordenada denota el estado en el *locus 2*. Además se asume que al tiempo cero se tienen k individuos con haplotipo $(0, 1)$, $(N - k - 1)$ con haplotipo $(0, 0)$, uno con haplotipo $(1, 0)$ y ninguno con haplotipo $(1, 1)$.

Sea $X_t^{i,j}$ con $t \geq 0$, $i, j = 0, 1$ el número de individuos del tipo (i, j) en la población al tiempo t . Por lo tanto se tiene que la cadena $\mathbf{X} = \{\mathbb{X}_t, t \geq 0\}$, donde $\mathbb{X}_t = (X_t^{00}, X_t^{10}, X_t^{01}, X_t^{11})$ con $t \geq 0$ presenta los registros de la evolución de la población.

Sea n^{ij} con $i, j = 0, 1$, el indicador del número de individuos con haplotipo (i, j) , entonces el espacio de estados de \mathbf{X} es el conjunto:

$$\Delta_4^N = \left\{ \bar{n} = (n^{00}, n^{10}, n^{01}, n^{11}) : n^{ij} \geq 0, i, j = 0, 1 \quad \sum_{(i,j) \in \{0,1\} \times \{0,1\}} n^{ij} = N \right\}.$$

Sea e^{ij} con $i, j = 0, 1$ un vector con todas las coordenadas cero excepto en la entrada (i, j) con valor uno. También se define $n^{i\bullet} = \sum_{j=0}^1 n^{ij}$, $n^{\bullet j} = \sum_{i=0}^1 n^{ij}$ y $n^{\bullet\bullet} = \sum_{i=0}^1 \sum_{j=0}^1 n^{ij}$.

Supóngase que los eventos de muerte y nacimiento ocurren después de cierto tiempo que tiene una distribución exponencial con parámetro λ , y se describen de la siguiente forma: con probabilidad $r \in [0, 1]$ dos individuos de la población se recombinan y producen un descendiente que sustituye un individuo que es elegido para morir, con probabilidad $1 - r$ un individuo produce una copia de sí mismo y su descendiente sustituye a un individuo elegido para morir.

Teorema 3.2.1 Sea $\mathbf{X} = \{\mathbb{X}_t, t \geq 0\}$ una cadena con estado $\bar{n} \in \Delta_4^N$, entonces la tasa a la que un evento nacimiento-muerte ocurre es:

$$\nu_{\bar{n}} = \sum_{i,j=0}^1 \sum_{k,l=0}^1 \lambda_{\bar{n}\bar{n}+e^{ij}-e^{kl}} = \lambda,$$

donde la tasa de transición de \bar{n} a $\bar{n} + e^{ij} - e^{kl}$ es:

$$\lambda_{\bar{n}\bar{n}+e^{ij}-e^{kl}} = \lambda \frac{n^{kl}}{N} \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right].$$

Demostración. Debido a que el proceso \mathbf{X} se distribuye exponencial con parámetro λ , entonces tiene la siguiente probabilidad de transición infesimal:

$$P[X_{t+h} = \bar{n} + e^{ij} - e^{kl} | X_t = \bar{n}] = \lambda \frac{n^{kl}}{N} \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right] h + \mathcal{O}(h),$$

para $(i, j), (k, l) \in \{0, 1\} \times \{0, 1\}$, y es cero para otro caso.

Por lo tanto se tiene que la tasa de transición de \bar{n} a $\bar{n} + e^{ij} - e^{kl}$ es

$$\lambda_{\bar{n}\bar{n}+e^{ij}-e^{kl}} = \lambda \frac{n^{kl}}{N} \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right],$$

entonces la tasa a la que un evento nacimiento-muerte ocurre es

$$\begin{aligned} \nu_{\bar{n}} &= \sum_{i,j=0}^1 \sum_{k,l=0}^1 \lambda_{\bar{n}\bar{n}+e^{ij}-e^{kl}} \\ &= \sum_{i,j=0}^1 \sum_{k,l=0}^1 \lambda \frac{n^{kl}}{N} \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right] \\ &= \lambda \left(\sum_{k,l=0}^1 \frac{n^{kl}}{N} \right) \left(\sum_{i,j=0}^1 \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right] \right) \\ &= \lambda \frac{1}{N} \left(\sum_{k,l=0}^1 n^{kl} \right) \left(\frac{1}{N} \sum_{i,j=0}^1 n^{ij} - \frac{r}{N} \sum_{i,j=0}^1 n^{ij} + \frac{r}{N^2} \left(\sum_{i,j=0}^1 n^{i\bullet} n^{\bullet j} \right) \right) \\ &= \lambda \frac{1}{N} \left(\sum_{k,l=0}^1 n^{kl} \right) \left(\frac{1}{N} \sum_{i,j=0}^1 n^{ij} - \frac{r}{N} \sum_{i,j=0}^1 n^{ij} + \frac{r}{N^2} \left(\sum_{i=0}^1 n^{i\bullet} \right) \left(\sum_{j=0}^1 n^{\bullet j} \right) \right) \end{aligned}$$

y como $\bar{n} \in \Delta_4^N$ entonces

$$\nu_{\bar{n}} = \lambda \left(\frac{1}{N} N \right) \left(\frac{1}{N} N - \frac{r}{N} N + \frac{r}{N^2} N^2 \right) = \lambda.$$

Aproximación por difusión

Considérese el proceso estocástico

$$Y^N = Y_t^N = \frac{X_{[t/h^N]}}{N} : t \geq 0,$$

donde $[x]$ es el entero más pequeño que x , h^N es la longitud del subintervalo que se tiene que el intervalo $[0, t]$ y además se tiene que $X_{[t/h^N]} = (X_{[t/h^N]}^{00}, X_{[t/h^N]}^{10}, X_{[t/h^N]}^{01}, X_{[t/h^N]}^{11})$ para $t \geq 0$. Defínase Δ_4 como:

$$\Delta_4 = \left\{ \bar{y} = (y^1, y^2, y^3, y^4) : y^i \geq 0, \quad i = 1, 2, 3, 4 \quad \sum_1^4 y^i = 1 \right\}.$$

Teorema 3.2.2 *Sea $Y = \{Y_t, t \geq 0\}$ el proceso de difusión asociado al proceso $X = \{X_t, t \geq 0\}$, entonces los parámetros infinitesimales del proceso de difusión están dados por:*

$$\mu^{(i,j)}(\bar{x}) = -\rho(x^{ij} - x^{i\bullet} x^{\bullet j})$$

$$\sigma^{2(i,j)(k,l)}(\bar{x}) = 2x^{kl} \left(\delta^{(i,j)(k,l)} - x^{ij} \right),$$

donde

$$\delta^{(i,j)(k,l)} \begin{cases} 1 & \text{si } (i,j) = (k,l) \\ 0 & \text{en otro caso} \end{cases}$$

Demostración. Sea $\lambda = N$ y tomando $h^N = 1/N$ y sea $\bar{x} \in \Delta_4$, entonces los parámetros infinitesimales están dados por:

$$\mu^{(i,j)}(\bar{x}) = \lim_{h^N \rightarrow 0} \frac{1}{h^N} E \left[Y_{t+h^N}^{ij} - Y_t^{ij} \mid Y_t = \bar{x} \right] \quad (3.6)$$

$$\sigma^{2(i,j)(k,l)}(\bar{x}) = \lim_{h^N \rightarrow 0} \frac{1}{h^N} E \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right) \left(Y_{t+h^N}^{kl} - Y_t^{kl} \right) \mid Y_t = \bar{x} \right] \quad (3.7)$$

para $i, j, k, l \in \{0, 1\}$.

Entonces se tiene que:

$$\begin{aligned}
 \mu^{(i,j)}(\bar{x}) &= \lim_{h^N \rightarrow 0} \frac{1}{h^N} E \left[Y_{t+h^N}^{ij} - Y_t^{ij} \mid Y_t = \bar{x} \right] \\
 &= \lim_{h^N \rightarrow 0} \frac{1}{h^N} E \left[Y_{t+h^N}^{ij} - Y_t^{ij} \mid Y_t = \frac{1}{N} (n^{00}, n^{10}, n^{01}, n^{11}) \right] \\
 &\stackrel{1}{=} \lim_{h^N \rightarrow 0} \frac{1}{h^N} \frac{1}{N} E [X_{t/h^N+1}^{ij} - X_{t/h^N}^{ij} \mid NY_t = (n^{00}, n^{10}, n^{01}, n^{11})] \\
 &= \lim_{h^N \rightarrow 0} \frac{1}{h^N} \frac{1}{N} E [X_{t/h^N+1}^{ij} - X_{t/h^N}^{ij} \mid X_{t/h^N} = (n^{00}, n^{10}, n^{01}, n^{11})] \\
 &\stackrel{2}{=} \lim_{h^N \rightarrow 0} \frac{1}{h^N} \frac{1}{N} \left\{ \lambda \left(1 - \frac{n^{ij}}{N} \right) \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right] \right. \\
 &\quad \left. - \lambda \frac{n^{ij}}{N} \left[(1-r) \left(1 - \frac{n^{ij}}{N} \right) + r \left(1 - \frac{n^{i\bullet} n^{\bullet j}}{N} \right) \right] \right\} \\
 &= \lim_{h^N \rightarrow 0} \frac{1}{h^N} \frac{1}{N} \left\{ \lambda r \left(1 - \frac{n^{ij}}{N} \right) \frac{n^{i\bullet} n^{\bullet j}}{N} - \lambda r \frac{n^{ij}}{N} \left(1 - \frac{n^{i\bullet} n^{\bullet j}}{N} \right) \right\} \\
 &\stackrel{3}{=} \rho (1 - x^{ij}) x^{i\bullet} x^{\bullet j} - \rho x^{ij} (1 - x^{i\bullet} x^{\bullet j}) \\
 &= \rho x^{i\bullet} x^{\bullet j} - \rho x^{ij} x^{i\bullet} x^{\bullet j} - \rho x^{ij} + \rho x^{ij} x^{i\bullet} x^{\bullet j} \\
 &= -\rho (x^{ij} - x^{i\bullet} x^{\bullet j}).
 \end{aligned}$$

1. Por definición de proceso de difusión $Y^N = Y_t^N = \frac{X_{[t/h^N]}}{N}$.
2. Por definición de tasa de transición.
3. Se asume que $n^{ij}/N \rightarrow x^{ij}$, $n^{i\bullet}/N \rightarrow x^{i\bullet}$, $n^{\bullet j}/N \rightarrow x^{\bullet j}$, $(Nr) \rightarrow \rho$ como $N \rightarrow \infty$, $\lambda = N$ y tomando $h^N = 1/N$.

Ahora para encontrar $\sigma^{2(i,j)(k,l)}(\bar{x})$ se observan dos casos:

- i) Sea $(i,j), (k,l) \in \{0,1\}^2$ donde $(i,j) \neq (k,l)$. Y defínase Z^{ij} como:

$$Z^{ij} = X_{t/h^N+1}^{ij} - X_{t/h^N}^{ij} \quad i,j=0,1, \text{ con } t \geq 0,$$

entonces,

$$\begin{aligned}
 E \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right) \left(Y_{t+h^N}^{kl} - Y_t^{kl} \right) \mid Y_t = \frac{1}{N} (n^{00}, n^{10}, n^{01}, n^{11}) \right] \\
 = E \left[\frac{Z^{ij}}{N} \frac{Z^{kl}}{N} \mid X_{t/h^N} = (n^{00}, n^{10}, n^{01}, n^{11}) \right]
 \end{aligned}$$

Nótese que existen tres posibles valores para $Z^{ij}Z^{kl}$ que son: -1 si sucede que $Z^{ij} = 1$ y $Z^{kl} = -1$ o si $Z^{ij} = -1$ y $Z^{kl} = 1$; 0 si alguno de los dos es cero y 1 si sucede que $Z^{ij} = Z^{kl} = -1$ o si $Z^{ij} = Z^{kl} = 1$. La probabilidad del último caso es cero porque si se produce un aumento en una de las coordenadas de $(n^{00}, n^{10}, n^{01}, n^{11})$, entonces se tiene una disminución en el otro.

Por lo tanto se tiene que:

$$\begin{aligned} & E \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right) \left(Y_{t+h^N}^{kl} - Y_t^{kl} \right) \mid Y_t = \frac{1}{N} (n^{00}, n^{10}, n^{01}, n^{11}) \right] = \\ & = -\frac{1}{N^2} \left\{ \lambda \frac{n^{kl}}{N} \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right] + \lambda \frac{n^{ij}}{N} \left[(1-r) \frac{n^{kl}}{N} + r \frac{n^{k\bullet} n^{\bullet l}}{N} \right] \right\}. \end{aligned}$$

Sea $\lambda = N$, $h^N = 1/N$ y observando que $n^{ij} \rightarrow x^{ij}$, $n^{i\bullet}/N \rightarrow x^{i\bullet}$, $n^{\bullet j}/N \rightarrow x^{\bullet j}$, $(Nr) \rightarrow \rho$ así como $N \rightarrow \infty$ y denotando $\bar{x} = (x^{00}, x^{01}, x^{10}, x^{11})$, entonces se tiene que

$$\begin{aligned} \sigma^{2(i,j)(k,l)}(\bar{x}) &= \lim_{h^N \rightarrow 0} \frac{1}{h^N} E \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right) \left(Y_{t+h^N}^{kl} - Y_t^{kl} \right) \mid Y_t = \bar{x} \right] \\ &= \lim_{N \rightarrow \infty} NE \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right) \left(Y_{t+h^N}^{kl} - Y_t^{kl} \right) \mid Y_t = \bar{x} \right] \\ &= \lim_{N \rightarrow \infty} N \left(-\frac{1}{N^2} \right) \left\{ \lambda \frac{n^{kl}}{N} \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet} n^{\bullet j}}{N} \right] + \right. \\ &\quad \left. \lambda \frac{n^{ij}}{N} \left[(1-r) \frac{n^{kl}}{N} + r \frac{n^{k\bullet} n^{\bullet l}}{N} \right] \right\} \\ &= \lim_{N \rightarrow \infty} \left(-\frac{1}{N} \right) \left\{ Nx^{kl} [(1-r)x^{ij} + rx^{i\bullet}x^{\bullet j}] \right. \\ &\quad \left. + Nx^{ij} [(1-r)x^{kl} + rx^{k\bullet}x^{\bullet l}] \right\} \\ &= -\lim_{N \rightarrow \infty} \frac{1}{N} \{ 2Nx^{ij}x^{kl} - 2rNx^{ij}x^{kl} + rN(x^{ij}x^{k\bullet}x^{\bullet l} + x^{kl}x^{i\bullet}x^{\bullet j}) \} \\ &= -2x^{ij}x^{kl} + \lim_{N \rightarrow \infty} \frac{1}{N} (2\rho x^{ij}x^{kl} - \rho(x^{ij}x^{k\bullet}x^{\bullet l} + x^{kl}x^{i\bullet}x^{\bullet j})) \\ &= -2x^{ij}x^{kl}. \end{aligned}$$

ii) Sea $(i, j) = (k, l)$, por lo tanto se tiene:

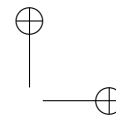
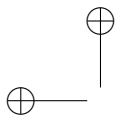
$$\begin{aligned} E \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right)^2 \mid NY_t = (n^{00}, n^{10}, n^{01}, n^{11}) \right] &= \\ &= E \left[\left(\frac{Z^{ij}}{N} \right)^2 \mid X_{t/h^N} = (n^{00}, n^{10}, n^{01}, n^{11}) \right] \end{aligned}$$

Sea $\lambda = N$, $h^N = 1/N$ y observando que $n^{ij} \rightarrow x^{ij}$, $n^{i\bullet}/N \rightarrow x^{i\bullet}$, $n^{\bullet j}/N \rightarrow x^{\bullet j}$, $(Nr) \rightarrow \rho$ así como $N \rightarrow \infty$ y denotando $\bar{x} = (x^{00}, x^{01}, x^{10}, x^{11})$, entonces se tiene que

$$\begin{aligned} \sigma^{2(i,j)(k,l)}(\bar{x}) &= \lim_{h^N \rightarrow 0} \frac{1}{h^N} E \left[\left(Y_{t+h^N}^{ij} - Y_t^{ij} \right)^2 \mid Y_t = \bar{x} \right] \\ &= \lim_{N \rightarrow \infty} N \frac{1}{N^2} \left\{ \lambda \left(1 - \frac{n^{ij}}{N} \right) \left[(1-r) \frac{n^{ij}}{N} + r \frac{n^{i\bullet}}{N} \frac{n^{\bullet j}}{N} \right] \right. \\ &\quad \left. + \lambda \frac{n^{ij}}{N} \left[(1-r) \left(1 - \frac{n^{ij}}{N} \right) + r \left(1 - \frac{n^{i\bullet}}{N} \frac{n^{\bullet j}}{N} \right) \right] \right\} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ 2\lambda \frac{n^{ij}}{N} \left(1 - \frac{n^{ij}}{N} \right) + r\lambda \left(1 - \frac{n^{ij}}{N} \right) \frac{n^{i\bullet}}{N} \frac{n^{\bullet j}}{N} \right. \\ &\quad \left. + r\lambda \left(1 - \frac{n^{i\bullet}}{N} \frac{n^{\bullet j}}{N} \right) \right\} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ 2\lambda x^{ij} (1 - x^{ij}) + r\lambda (1 - x^{ij}) x^{i\bullet} x^{\bullet j} \right. \\ &\quad \left. + r\lambda (1 - x^{i\bullet} x^{\bullet j}) \right\} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ 2N x^{ij} (1 - x^{ij}) + \rho (1 - x^{ij}) x^{i\bullet} x^{\bullet j} \right. \\ &\quad \left. + \rho (1 - x^{i\bullet} x^{\bullet j}) \right\} \\ &= 2x^{ij} (1 - x^{ij}) + \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ \rho (1 - x^{ij}) x^{i\bullet} x^{\bullet j} \right. \\ &\quad \left. + \rho (1 - x^{i\bullet} x^{\bullet j}) \right\} \\ &= 2x^{ij} (1 - x^{ij}) \end{aligned}$$

Entonces,

$$\begin{aligned} \mu^{(i,j)}(\bar{x}) &= -\rho (x^{ij} - x^{i\bullet} x^{\bullet j}) \\ \sigma^{2(i,j)(k,l)}(\bar{x}) &= 2x^{kl} \left(\delta^{(i,j)(k,l)} - x^{ij} \right), \end{aligned}$$



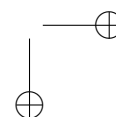
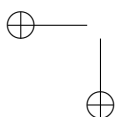
3. MODELO GENÉTICO CON RECOMBINACIÓN

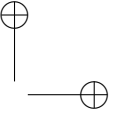
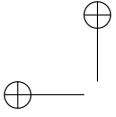
54

con

$$\delta^{(i,j)(k,l)} \begin{cases} 1 & \text{si } (i,j) = (k,l) \\ 0 & \text{en otro caso} \end{cases}$$

Por lo cual se puede concluir que Y^N converge en distribución a Y , donde Y es un proceso de difusión con coeficientes $\mu^{(i,j)}(\bar{x})$ y $\sigma^{2(i,j)(k,l)}(\bar{x})$.





CAPÍTULO 4

DOS-LOCUS ANCESTRALES

En un modelo de poblaciones de dos locus la recombinación puede presentarse de dos formas: una recombinación que proviene de un sólo ancestro o una recombinación que proviene de dos ancestros. En esta sección se toma el caso en que una recombinación que proviene de un sólo ancestro. El número de ancestros puede ser creciente o decreciente, el cual resulta ser un proceso de nacimiento o muerte.

4.1 EL MODELO DOS-LOCUS

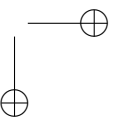
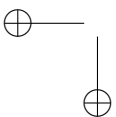
Un gen en un modelo de dos-locus se describe por un par ordenado $(x, y) \in [0, 1] \times [0, 1]$, cuya evolución estocástica sigue el modelo de Moran con recombinación, entonces

- i) Con probabilidad $1 - r$ un individuo elige un cromosoma de la generación previa y hereda los genes en el *locus 1* y en el *locus 2*
- ii) Con probabilidad r un individuo elige dos cromosomas de la generación previa y hereda un gen a al *locus 1* y el otro al *locus 2*.

Considérese el caso en que N es grande y r es de orden N^{-1} , para equilibrar los efectos de la deriva¹ y la recombinación. Defínase la tasa de recombinación como

$$\rho = \lim_{N \rightarrow \infty} 2Nr. \tag{4.1}$$

¹Cambio aleatorio en la frecuencia de alelos de una generación a otra.



En el modelo de un único locus defínase T_n como el tiempo más reciente en el cual apareció un ancestro en común de una muestra de n genes, entonces

$$T_n = M_n + M_{n-1} + \dots + M_2,$$

donde M_k son variables aleatorias exponenciales con parámetro $\binom{k}{2}$. Entonces la densidad T_n es conocida (ver Tavaré S.; Ofer Zeitouni (2001)) y está dada por:

$$f_{T_n}(t) = \sum_{k=2}^n (-1)^k \exp\left(-k(k-1)\frac{t}{2}\right) \frac{(2k-1)k(k-1)n_{[k]}}{2n_{(k)}},$$

donde $n_{[k]} = (n-1) \dots (n-k+1)$ y $n_{(k)} = n(n+1) \dots (n+k-1)$. Y por lo tanto se tiene que

$$E[T_n] = \sum_{r=2}^n E[M_r] = \sum_{r=2}^n \frac{2}{r(r-1)} = 2(1 - n^{-1}).$$

Entonces se tiene

$$1 = E[T_2] \leq E[T_n] \leq E[T_N] < 2,$$

donde T_N es el tiempo al cual la población total tiene sólo un ancestro en común. Notése que $E[T_n]$ es muy cercano a 2 para una cierta n . Además se tiene que

$$E[T_N - T_n] = 2\left(\frac{1}{n} - \frac{1}{N}\right) < \frac{2}{n},$$

es decir, la diferencia de medias es pequeña para una cierta n .

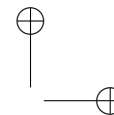
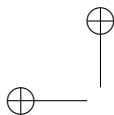
Teorema 4.1.1 *Sea W_n el tiempo de paro hasta que haya un ancestro en común de una muestra de n genes en un modelo dos-locus, entonces*

$$E[W_n] = 2\rho^{-1} \int_0^1 \frac{1 - v^{n-1}}{1 - v} \left(e^{\rho(1-v)} - 1\right) dv. \quad (4.2)$$

Demostración. Ver Griffiths, R.C. (1991).

Teorema 4.1.2 *Sea $W_{n,j}$ el tiempo en que el proceso ancestral de una muestra de tamaño n permanece en el estado j , entonces $W_n = \sum_{j=2}^{\infty} W_{n,j}$ y*

$$E[W_{n,j}] = 2 \sum_{k=2}^{\min(j,n)} (k-2)! \rho^{j-k} / j! \quad j=2,3,\dots \quad (4.3)$$



Demostración. Ver Griffiths, R.C. (1991).

Observación:

- i) Si $\rho = 0$, entonces $E[W_{n,j}] = 2/(j(j-1))$ y por lo tanto se tiene que $W_{n,j}$ tiene una distribución exponencial con parámetro $\binom{j}{2}$.
- ii) Si $j = 2$ se tiene que $E[W_{n,2}] = 1$, es decir, no hay dependencia de ρ .
- iii) Cuando $\rho \rightarrow \infty$ se tiene que $E[W_{n,j}] \approx 2\rho^{j-2}/(j(j-1))$.

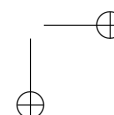
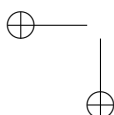
En general se puede obtener el número de ancestros de una muestra que pueden existir en un cierto tiempo antes de un ancestro en común en el modelo dos-locus ancestral.

4.2 ESTRUCTURA DEL GRÁFICO DE LA RECOMBINACIÓN ANCESTRAL

Sea \mathfrak{G} el gráfico de un modelo dos-locus, donde \mathfrak{G} está compuesto de dos árboles coalecentes marginales \mathcal{T}_1 y \mathcal{T}_2 que corresponden a *locus 1* y *locus 2* respectivamente. Se considera que los árboles marginales tiene un ancestro en común en la cima de \mathfrak{G} , aunque dicho ancestro puede ocurrir en un tiempo corto. Denótese $\mathcal{E}(\cdot)$ como el conjunto de borde de la gráfica, entonces se hace una partición de los bordes de \mathfrak{G} en cuatro conjuntos disjuntos, se obtienen

$$\begin{aligned} \mathcal{A} &= \mathcal{E}(\mathfrak{G}) \cap \mathcal{E}(\mathcal{T}_1) \cap \mathcal{E}(\mathcal{T}_2)^c & \mathcal{B} &= \mathcal{E}(\mathfrak{G}) \cap \mathcal{E}(\mathcal{T}_1)^c \cap \mathcal{E}(\mathcal{T}_2) \\ \mathcal{C} &= \mathcal{E}(\mathfrak{G}) \cap \mathcal{E}(\mathcal{T}_1) \cap \mathcal{E}(\mathcal{T}_2) & \mathcal{D} &= \mathcal{E}(\mathfrak{G}) \cap \mathcal{E}(\mathcal{T}_1)^c \cap \mathcal{E}(\mathcal{T}_2)^c. \end{aligned}$$

Observación: Los ancestros representados por los bordes en \mathcal{A} no tienen material genético que tenga influencia sobre el *locus 2*, de igual manera sucede de \mathcal{B} sobre el *locus 1*. Los bordes en \mathcal{C} corresponden a los ancestros que contribuyen al material genético de ambos loci, y entonces \mathcal{D} no va a tener material genético que tenga influencia sobre el *locus 1* o el *locus 2*.



Definiendo a $\mathcal{E}(\mathfrak{G}_t)$ los bordes de una sección representativa de \mathfrak{G} al tiempo t que puede ser dividida en cuatro tipos. Denotando

$$\begin{aligned} n_{\mathcal{A}}(t) &= |\mathcal{E}(\mathfrak{G}_t) \cap \mathcal{A}| \\ n_{\mathcal{B}}(t) &= |\mathcal{E}(\mathfrak{G}_t) \cap \mathcal{B}| \\ n_{\mathcal{C}}(t) &= |\mathcal{E}(\mathfrak{G}_t) \cap \mathcal{C}| \\ n_{\mathcal{D}}(t) &= |\mathcal{E}(\mathfrak{G}_t) \cap \mathcal{D}| \\ n(t) &= |E_t| \end{aligned}$$

donde $|\cdot|$ denota el número de elementos del conjunto. Se puede observar que

$$n_{\mathcal{A}}(t) + n_{\mathcal{B}}(t) + n_{\mathcal{C}}(t) + n_{\mathcal{D}}(t) = n(t),$$

además se tiene que

$$\begin{aligned} n_{\mathcal{A}}(t) + n_{\mathcal{C}}(t) &= |\mathcal{E}(\mathcal{T}_1(t))| \equiv A_n(t) \\ n_{\mathcal{B}}(t) + n_{\mathcal{C}}(t) &= |\mathcal{E}(\mathcal{T}_2(t))| \equiv B_n(t) \end{aligned}$$

donde $A_n(\cdot)$ y $B_n(\cdot)$ son los procesos ancestrales marginales para el loci *locus 1* y *locus 2* respectivamente.

Definiendo el proceso de Markov

$$m(t) = (n_{\mathcal{A}}(t), n_{\mathcal{B}}(t), n_{\mathcal{C}}(t), n_{\mathcal{D}}(t)) \text{ con } t \geq 0$$

donde los bordes de \mathfrak{G} pueden ser de la forma $(1,0),(0,1),(1,1)$ o $(0,0)$ dependiendo si pertenecen a $\mathcal{A}, \mathcal{B}, \mathcal{C}$ o \mathcal{D} respectivamente. Cuando la coalescencia ocurre para dos bordes del tipo (α, β) y (γ, δ) el ancestro que resulta es del tipo $(\max(\alpha, \beta), \min(\gamma, \delta))$, y si la recombinación ocurre en un borde del tipo (α, β) entonces los dos nuevos bordes son tipos $(\alpha, 0)$ y $(0, \beta)$.

En Ethier y Griffiths (1990) se demuestra que el proceso $m(t)$ es Markoviano, pues la evolución es consistente con el tipo de definiciones en la partición de \mathfrak{G} . Además se tiene que las tasa de transición del proceso son:

$$(a, b, c, d) \longrightarrow \begin{cases} (a+1, b+1, c-1, d) & c\rho/2 \\ (a-1, b-1, c+1, d) & ab \\ (a-1, b, c, d) & ac + a(a-1)/2 \\ (a, b-1, c, d) & bc + b(b-1)/2 \\ (a, b, c-1, d) & c(c-1)/2 \\ (a, b, c, d+1) & (a+b+d)\rho/2 \\ (a, b, c, d-1) & d(a+b+c) + d(d-1)/2 \end{cases}$$

con $a + b + c + d > 1$.

Estas tasa de transición se deben a que: por ejemplo para la primer transición $(a, b, c, d) \rightarrow (a + 1, b + 1, c - 1, d)$ sucede que si un evento recombinante ocurre en un borde del tipo $(1, 1)$; éste es resultado de la pérdida de un borde de $(1, 1)$, y de la adición de un borde de $(1, 0)$ y un borde de $(0, 1)$. La tasa de cambio es $c\rho/2$. Ahora se ve que si se tiene la transición $(a, b, c, d) \rightarrow (a - 1, b, c, d)$, se tiene que es resultado de la coalescencia de un borde del tipo $(0, 1)$ y de un borde del tipo $(1, 1)$, o de la coalescencia de dos bordes de tipo $(1, 0)$. La tasa de cambio con la que ocurre el primer evento es ac y la tasa de cambio con la que ocurre el segundo evento es $a(a - 1)/2$. De igual manera se pueden verificar las otras transiciones. La tasa de transición total es la suma de todas las tasas, si $a + b + c + d = n$ entonces está tasa es igual a d_n dada por $d_n = c\rho/2 + n(n - 1)/2$.

El proceso marginal $(n_A(t), n_B(t), n_C(t))$ con $t \geq 0$, también es Markoviano, pues no depende de $n_D(t)$. Por tanto las tasas de transición del proceso son:

$$(a, b, c) \rightarrow \begin{cases} (a + 1, b + 1, c - 1) & c\rho/2 \\ (a - 1, b - 1, c + 1) & ab \\ (a - 1, b, c) & ac + a(a - 1)/2 \\ (a, b - 1, c) & bc + b(b - 1)/2 \\ (a, b, c - 1) & c(c - 1)/2 \end{cases} \quad (4.4)$$

con $a + b + c > 1$.

Nótese que la recombinación ocurre sólo en los bordes de C . Aquí la tasa de transición total de un estado (a, b, c) con $n = a + b + c$ es igual que la del caso anterior, es decir, $d_n = c\rho/2 + n(n - 1)/2$.

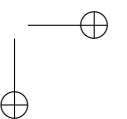
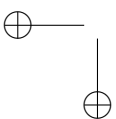
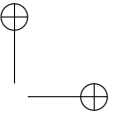
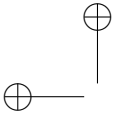
Sea $W_{A,n}$ y $W_{B,n}$ los tiempos de paro de una muestra de n genes que tiene ancestros en común en el locus A y B respectivamente.

Teorema 4.2.1 *Sea $\zeta(a, b, c; \rho)$ el tiempo de paro esperado hasta que $n_A(t) + n_C(t) = 1$ y $n_B(t) + n_C(t) = 1$ empezando con $n_A(0) = a, n_B(0) = b$ y $n_C(0) = c$, entonces*

$$E[\max(W_{A,n}, W_{B,n})] = \zeta(0, 0, n; \rho)$$

es el tiempo de paro esperado hasta que una muestra de n genes tiene ancestros en común en el locus A y B .

Demostración. Ver Griffiths, R.C. (1991).



CONCLUSIÓN

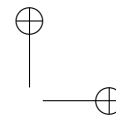
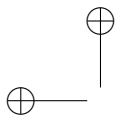
La motivación de la presente tesis se basó en la importancia de la modelación estocástica en genética de poblaciones, que es uno de los temas de mayor interés en el estudio de la evolución de poblaciones. Durante este proceso se pueden presentar eventos de recombinación, que es un factor importante en la evolución de algunos organismos, incluidos los humanos.

Mediante los análisis y desarrollos realizados en la presente tesis, fue posible observar a través de la demografía con la cual evolucionan el modelo de Wright-Fisher y el modelo Moran que el modelo de Moran es un modelo de nacimiento-muerte mientras que el modelo de Wright-Fisher es sólo de nacimiento. En el caso en que la evolución de la población sigue el modelo de Moran con mutación, resultó que la distribución estacionaria es única y es igual a la su distribución límite.

Por otra parte, el proceso de difusión que se obtuvo para el modelo de Wright-Fisher es el mismo que se obtuvo para el modelo de Moran. Esto se debió a que los parámetros infinitesimales del proceso de difusión para ambos modelos son los mismos, por lo tanto se tiene que la distribución estacionaria para la versión difusión del modelo de Moran es la misma que para la versión difusión del modelo de Wright-Fisher.

En cuanto a los modelos genéticos univariado y multivariado con recombinación, éstos se consideraron siguiendo una evolución de acuerdo al modelo de Moran, y además se realizó una aproximación del proceso de difusión con muerte.

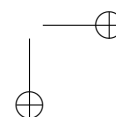
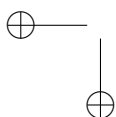
Respecto al modelo dos locus, se consideró que dicho modelo sigue una evolución estocástica de acuerdo al modelo de Moran con recombinación y que en este modelo, se puede obtener a su vez, el número de ancestros de una muestra en un cierto tiempo antes de un ancestro en común.



4. CONCLUSIÓN

62

Por último se pudo observar que debido a que la partición de la estructura de los bordes del gráfico puede realizarse en conjuntos disjuntos, el proceso que contiene a las particiones de la estructura tiene la propiedad Markoviana, y por lo tanto se puede utilizar las propiedades de este proceso para estudiar el comportamiento de los árboles genealógicos.



APÉNDICE

Algunos resultados auxiliares

Los siguientes resultados fueron tomados de Norris, J.R. (1997).

Lema .1 Sean $x, y \in \mathfrak{R}$, $M \in \mathbb{Z}^+$ y denotemos a $S_M(x, y)$ como la siguiente suma:

$$S_N(x, y) = \binom{N}{0} v_1 v_2 \cdots v_N + \binom{N}{1} u_0 v_2 \cdots v_N + \cdots + \binom{N}{N} u_0 u_1 \cdots u_{N-1}$$

donde

$$u_j = x + yj \quad y \quad v_j = 1 - x - yj,$$

entonces

$$S_N(x, y) = (1 - y)(1 - 2y) \cdots (1 - Ny) \tag{5}$$

Demostración. La demostración será por inducción sobre la N .

i) Probemos para $N = 2$, entonces por hipótesis se tiene que

$$\begin{aligned} S_2 &= \binom{2}{0} v_1 v_2 + \binom{2}{1} u_0 v_2 + \binom{2}{2} u_0 u_1 \\ &= v_1 v_2 + 2u_0 v_2 + u_0 u_1 \\ &= (1 - x - y)(1 - x - 2y) + 2x(1 - x - 2y) + x(x + y) \\ &= 1 - x - y - x - x^2 + xy - 2y + 2xy + 2y^2 + 2x - x^2 - 3xy \\ &= 2y^2 - 3y + 1 = (1 - y)(1 - 2y). \end{aligned}$$

ii) Supóngase que se cumple para $n = N$, por demostrar que se cumple para $n = N + 1$,

$$\begin{aligned}
 S_{N+1}(x, y) &= v_1 v_2 \cdots v_{N+1} + \binom{N+1}{1} u_0 v_2 \cdots v_{N+1} + \cdots + \\
 &\quad + \binom{N+1}{k-1} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \cdots + u_0 u_1 \cdots u_N \\
 &= v_1 v_2 \cdots v_{N+1} + \left[\binom{N}{1} + \binom{N}{0} \right] u_0 v_2 \cdots v_{N+1} + \cdots + \\
 &\quad + \left[\binom{N}{k} + \binom{N}{k-1} \right] u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \cdots + \\
 &\quad + \left[\binom{N}{N} + \binom{N}{N-1} \right] u_0 u_1 \cdots u_{N-1} v_N + u_0 u_1 \cdots u_N \\
 &= v_1 v_2 \cdots v_{N+1} + \binom{N}{1} u_0 v_2 \cdots v_{N+1} + \cdots + \\
 &\quad + \binom{N}{k} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \cdots + \binom{N}{N} u_0 u_1 \cdots u_{N-1} v_N + \\
 &\quad + \binom{N}{0} u_0 v_2 \cdots v_N + \cdots + \binom{N}{k-1} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \\
 &\quad + \binom{N}{N-1} u_0 u_1 \cdots u_{N-1} v_N + u_0 u_1 \cdots u_N.
 \end{aligned}$$

Ahora partiendo la ecuación anterior en dos sumas $S_{N+1}^1(x, y)$ y $S_{N+1}^2(x, y)$ donde

$$\begin{aligned}
 S_{N+1}^1(x, y) &= v_1 v_2 \cdots v_{N+1} + \binom{N}{1} u_0 v_2 \cdots v_{N+1} + \cdots + \\
 &\quad + \binom{N}{k} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \cdots + \binom{N}{N} u_0 u_1 \cdots u_{N-1} v_N. \\
 S_{N+1}^2(x, y) &= \binom{N}{0} u_0 v_2 \cdots v_{N+1} + \cdots + \binom{N}{k-1} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \\
 &\quad + \binom{N}{N-1} u_0 u_1 \cdots u_{N-1} v_N + u_0 u_1 \cdots u_N.
 \end{aligned}$$

y analizando por separado ambas sumas se tiene:

$$\begin{aligned}
 S_{N+1}^1(x, y) &= v_1 v_2 \cdots v_{N+1} + \binom{N}{1} u_0 v_2 \cdots v_{N+1} + \cdots + \\
 &\quad + \binom{N}{k} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \cdots + \binom{N}{N} u_0 u_1 \cdots u_{N-1} v_{N+1} \\
 &= v_{N+1} [v_1 \cdots v_N + \binom{N}{1} u_0 v_2 \cdots v_N + \cdots + \binom{N}{k} u_0 u_1 \cdots u_{k-1} v_k \cdots v_N + \\
 &\quad + \cdots + \binom{N}{N} u_0 u_1 \cdots u_{N-1}] \\
 &= v_{N+1} S_N(x, y). \tag{6}
 \end{aligned}$$

Nótese que si $z = x + y$, entonces

$$S_N(z, y) = w_1 w_2 \cdots w_N + \binom{N}{1} \nu_0 w_2 \cdots w_N + \cdots + \nu_0 \nu_1 \cdots \nu_{N-1},$$

donde

$$\begin{aligned}
 w_j &= 1 - z - jy = 1 - (x + y) - jy \\
 &= 1 - x - (j + 1)y \\
 &= v_{j+1}. \\
 \nu_j &= z + jy = x + y + jy \\
 &= x + (j + 1)y \\
 &= u_{j+1},
 \end{aligned}$$

entonces

$$\begin{aligned}
 S_N(z, y) &= \binom{N}{0} v_2 \cdots v_{N+1} + \cdots + \binom{N}{k-1} u_1 \cdots u_{k-1} v_k \cdots v_{N+1} \\
 &\quad + \binom{N}{N-1} u_0 u_1 \cdots u_{N-1} v_N + u_1 \cdots u_N
 \end{aligned}$$

y por lo tanto

$$\begin{aligned}
 S_{N+1}^2(x, y) &= \binom{N}{0} u_0 v_2 \cdots v_{N+1} + \cdots + \binom{N}{k-1} u_0 u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \\
 &\quad + \binom{N}{N-1} u_0 u_1 \cdots u_{N-1} v_N + u_0 u_1 \cdots u_N \\
 &= u_0 \left[\binom{N}{0} v_2 \cdots v_{N+1} + \cdots + \binom{N}{k-1} u_1 \cdots u_{k-1} v_k \cdots v_{N+1} + \right. \\
 &\quad \left. + \binom{N}{N-1} u_1 \cdots u_{N-1} v_N + u_1 \cdots u_N \right] \\
 &= u_0 S_N(x + y, y). \tag{7}
 \end{aligned}$$

Entonces de (A.2) y (A.3) se tiene

$$\begin{aligned}
 S_{N+1}(x, y) &= S_{N+1}^1(x, y) + S_{N+1}^2(x, y) \\
 &= v_{N+1} S_N(x, y) + u_0 S_N(x + y, y) \\
 &= v_{N+1} [(1-y)(1-2y) \cdots (1-Ny)] + u_0 [(1-y)(1-2y) \cdots (1-Ny)] \\
 &= [(1-y)(1-2y) \cdots (1-Ny)] [v_{N+1} + u_0] \\
 &= [(1-y)(1-2y) \cdots (1-Ny)] [(1-x - (N+1)y) + x] \\
 &= [(1-y)(1-2y) \cdots (1-Ny)] [(1 - (N+1)y)],
 \end{aligned}$$

es decir, se cumple para $n = N + 1$.

Lema .2 Sea $x > 0, r \in \mathbb{N}$ tal que $1 \leq r \leq x$, entonces

$$\Gamma(x) = \left[\prod_{k=1}^r (x - k) \right] \Gamma(x - r). \tag{8}$$

Demostración. La demostración se realiza por inducción sobre r , entonces

- i) Para $r = 1$ se cumple, pues para $x > 0$ se tiene que $\Gamma(x) = x\Gamma(x - 1)$.
- ii) Supóngase válida para $r < x$ por demostrar que se cumple para $r + 1 < x$

x ,

$$\begin{aligned} \Gamma(x) &= \left[\prod_{k=1}^r (x-k) \right] \Gamma(x-r) \\ &= \left[\prod_{k=1}^r (x-k)(x-(r+1)) \right] \Gamma(x-(r+1)) \\ &= \left[\prod_{k=1}^{r+1} (x-k) \right] \Gamma(x-(r+1)). \end{aligned}$$

Lema .3 Sea $N \geq 1$ y $\gamma_1, \gamma_2 \in \mathfrak{R}$ tal que $0 < \gamma_1, \gamma_2 \leq 1$, entonces

1. Para $j = 1, 2, \dots, N$ se tiene

$$\Gamma\left(\frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2}\right) = \left[\prod_{k=1}^j \left(\frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2} - k\right) \right] \Gamma\left(\frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2} - j\right),$$

2. También se cumple

$$\Gamma\left(\frac{N}{1-\gamma_1-\gamma_2}\right) = \left[\prod_{k=1}^N \left(\frac{N}{1-\gamma_1-\gamma_2} - k\right) \right] \Gamma\left(\frac{N(\gamma_1+\gamma_2)}{1-\gamma_1-\gamma_2}\right),$$

3. Para $j = 1, 2, \dots$ se cumple

$$\Gamma\left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2} + j\right) = \left[\prod_{k=1}^j \left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2} + j - k\right) \right] \Gamma\left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2}\right).$$

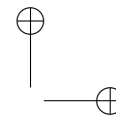
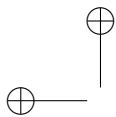
Demostración. Las demostraciones de 1, 2 y 3 son inmediatas del **Lema A.2**, ya que

1. Para $j < \frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2}$, con $j = 1, \dots, N$ se tiene

$$\Gamma\left(\frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2}\right) = \left[\prod_{k=1}^j \left(\frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2} - k\right) \right] \Gamma\left(\frac{N(1-\gamma_2)}{1-\gamma_1-\gamma_2} - j\right)$$

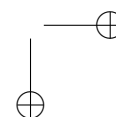
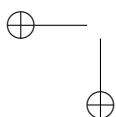
2. Ahora para $\frac{N}{1-\gamma_1-\gamma_2}$ tenemos

$$\begin{aligned} \Gamma\left(\frac{N}{1-\gamma_1-\gamma_2}\right) &= \left[\prod_{k=1}^N \left(\frac{N}{1-\gamma_1-\gamma_2} - k\right) \right] \Gamma\left(\frac{N}{1-\gamma_1-\gamma_2} - N\right) \\ &= \left[\prod_{k=1}^N \left(\frac{N}{1-\gamma_1-\gamma_2} - k\right) \right] \Gamma\left(\frac{N - N(1-\gamma_1-\gamma_2)}{1-\gamma_1-\gamma_2} - N\right) \\ &= \left[\prod_{k=1}^N \left(\frac{N}{1-\gamma_1-\gamma_2} - k\right) \right] \Gamma\left(\frac{N(\gamma_1+\gamma_2)}{1-\gamma_1-\gamma_2}\right). \end{aligned}$$



3. Y para $\left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2} + j\right)$ se tiene que

$$\Gamma\left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2} + j\right) = \left[\prod_{k=1}^j \left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2} + j - k\right) \right] \Gamma\left(\frac{N\gamma_2}{1-\gamma_1-\gamma_2}\right).$$



BIBLIOGRAFÍA

- [1] Alfaro, R.Y. (2003), *Estimación de líneas ancestrales en genética utilizando el coalecente de Kingman y métodos montecarlo*, Tesis de Licenciatura, Facultad de Ciencias-UNAM.
- [2] Donnelly, P.; Rodrigues, E. R. (2000), *Convergence to Stationarity in the Moran model*, **37**, pp. 705-717.
- [3] Ethier, S.N.; Griffiths, R.C. (1990), "On the two-locus sampling distribution" en *Journal Mathematical Biology*, **29**, pp. 131-159.
- [4] Griffiths, R.C. (1991), The two-locus ancestral graph, *Selected Proceedings of the Sheffield Symposium on Applied Probability*, IMS Lecture Notes - Monograph Series, **18**, Institute of Mathematical Statistics, Hayward, California, pp. 100-107.
- [5] Griffiths, R.C; Rodrigues, E. R. (2008), *The age of an allele in a population with recombination.*, Versión preliminar.
- [6] Griffiths, R.C; Tavaré, S. (1994a), *Simulating probability distributions in the coalescent*, *Theoretical Population Biology*, **46**, pp. 131-159.
- [7] Griffiths, R.C; Tavaré, S. (1994b), *Ancestral inference in population genetics*, *Statistical Science*, **9**, pp. 307-319.
- [8] Karlin, S.; McGregor, J. (1964), On some stochastic models in genetics, *Stochastic models in Medicine and Biology*, The University of Wisconsin Press, E.U.A.
- [9] Karlin, S.; Taylor, H. (1975), *A first course in stochastic processes*, Academic Press, Inc., E.U.A.

- [10] Karlin, S.; Taylor, H. (1981), *A second course in stochastic processes*, Academic Press, Inc., E.U.A.
- [11] Moran, P.A.P. (1958), *Random processes in genetics*, Proc. Cambridge Phil. Society, **54**, pp. 60-72.
- [12] Norris, J.R. (1997), *Markov chains*, Cambridge University Press, Reino Unido.
- [13] Ross, S. (1996), *Stochastic processes*, 2a. ed., John Wiley Sons, USA.
- [14] Ross, S. (1998), *A first course in probability*, 5a. ed., Prentice-Hall, Inc., USA.
- [15] Sánchez, R.P. (2002), *El proceso de líneas de descendencia en el estudio genético de poblaciones*, Tesis de Licenciatura, Facultad de Ciencias-UNAM.
- [16] Tavaré S.; Ofer Zeitouni. (2001), *Lectures on Probability Theory and Statistics*. Ecole d'Etés de Probabilité de Saint-Flour XXXI – 2001, ed. Picard J., Lecture Notes in Mathematics, Springer Verlag, New York.
- [17] Tudor, C. (2002), *Procesos estocásticos*, Aportaciones Matemáticas, Series Textos 2, Soc. Mat. Mex.