



Universidad Nacional Autónoma de México

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

Proceso de frecuencia simple asimétrico

T E S I S

que para optar por el grado de

Maestra en Ciencias

PRESENTA:

Fernanda López Eslava

Director:

Dr. Adrián González Casanova Sobreón,
Instituto de Matemáticas, UNAM.

Ciudad de México, Octubre, 2022



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.

Índice general

Agradecimientos	2
Introducción	3
1. Genealogía de poblaciones neutras	5
1.1. Modelo de Wright-Fisher	6
1.2. Modelo de Moran	9
1.3. Modelo Look-Down	10
1.4. Modelo de Cannings	11
1.5. Cadenas de Markov a tiempo continuo	11
1.6. Coalescente de Kingman	13
1.7. Modelo de Wright-Fisher y modelo de Moran con selección	16
1.8. Modelo de Wright-Fisher y modelo de Moran con mutación	18
1.9. Λ -coalescente	21
1.10. Construcción de Poisson	22
1.11. Construcción paintbox	23
1.12. Dualidad	24
1.12.1. El modelo de Eldon-Wakeley	26
1.13. Beta coalescentes.	27
2. El proceso de frecuencia simple asimétrico	29
2.1. Construcción	29
2.2. Definición	30
2.3. El proceso de ancestría del modelo simple asimétrico	32
2.4. Dualidad de los procesos asimétricos	35
3. El modelo simple asimétrico con mutaciones	36
3.1. Construcción	36
3.2. Definición	36
3.3. El proceso de ancestría asimétrico con mutaciones	38
4. Recursiones	41
4.1. Recursión para el modelo asimétrico con mutaciones	41
4.2. Probabilidad de fijación del modelo simple asimétrico	45
4.2.1. Algoritmo del modelo simple asimétrico	47
Conclusión	49
Referencias	50

Agradecimientos

Introducción

Los modelos matemáticos han sido importantes en la tarea de explicar la variabilidad genética (diferencias en las secuencias de genes entre organismos individuales de una especie) de seres vivos en distintas áreas de la ciencia. En el capítulo 1 se presentan los modelos fundamentales en genética de poblaciones y herramientas como la dualidad y el estudio de la genealogía de una muestra en una población para comprender mejor estos modelos. En este capítulo se busca dar un panorama general de los modelos fundamentales en genética de poblaciones, el modelo de Wright-Fisher neutro y el modelo de Moran neutro (nos referimos a un modelo como neutro cuando en el modelo no hay selección), se describe su genealogía, definimos el proceso de frecuencia asociado a estos modelos neutros y algunos resultados de cadenas de Markov para estos modelos. Posteriormente agregamos los supuestos de selección y mutación al modelo de Wright-Fisher y al modelo de Moran, esto servirá para comprender la construcción de los modelos de los siguientes capítulos. En el capítulo 1 también se presenta el coalescente de Kingman y el Λ -coalescente.

La selección en modelos genéticos resulta importante ya que es un supuesto que se acerca a la realidad de muchas poblaciones y también resulta interesante como este supuesto cambia las propiedades matemáticas de los procesos estocásticos a los cuales se les agrega esta fuerza evolutiva. En el capítulo 2 presentamos un modelo que llamamos *modelo simple asimétrico*, se construye el modelo en el mismo espíritu que el modelo de Moran, descrito en el capítulo anterior. El *modelo simple asimétrico* describe la evolución de una población en donde los individuos pueden tener más de dos descendientes, los individuos se reproducen uno a la vez, es decir no se permiten 2 o más reproducciones al mismo tiempo y es asimétrico en el sentido siguiente: la población se divide en 2 tipos, los individuos con selección tendrán un número aleatorio de hijos mayor al de los demás individuos de la población, es decir, la distribución para la cantidad de hijos es distinta para cada tipo.

Los modelos en donde los individuos tienen un gran número de descendientes han sido estudiados, por ejemplo, en [12] y [4]. Por otro lado, en [9] y en [5] se consideran poblaciones con selección, en estos trabajos también se tiene la construcción de un proceso que describe la genealogía de dichas poblaciones. Nuestro trabajo se inspira en los dos trabajos anteriores, la construcción del proceso genealógico de nuestro modelo de frecuencia simple asimétrico sigue la idea de [5], sin embargo, la selección de nuestro modelo no radica en las distintas tasas de reproducción para individuos con selección y sin ella, sino en la distinta distribución del número de hijos que cada uno de los dos tipos tiene. En [5] se considera un modelo multitypo, en el cual los individuos tienen un número aleatorio de posibles hijos con la misma distribución (r_a la probabilidad de tener a posibles hijos) y dada esta probabilidad, cada individuo tiene una número b de hijos viables (o que llegan a la madurez) con probabilidad denotada por v_{iab} . Además se describe la genealogía de dicha población, al suponer que el proceso que cuenta la cantidad de hijos de cada individuo tiene distribución

estacionaria φ (para lo cual es necesario agregar mutación al modelo) y al aplicar el generador de este proceso a funciones especiales se llega a un proceso de ancestría parecido al nuestro, cuyas tasas de transición se expresan en términos de la distribución estacionaria φ . En nuestro modelo no es necesario suponer una distribución estacionaria (y por lo tanto no es necesario agregar mutación en el modelo) para obtener un proceso genealógico dual al proceso de frecuencia del modelo simple asimétrico. Además, nuestro modelo se puede generalizar, en el sentido que, el número de hijos que tiene cada tipo de individuo en la población puede ser una distribución general para cada tipo (con ciertas consideraciones que no mencionaremos en este trabajo), es por ello que nuestro modelo resulta novedoso.

En el capítulo 3 agregamos mutación al *modelo simple asimétrico*. En ambos capítulos se describe la genealogía de los modelos. En el capítulo 4 se presenta una recursión para la probabilidad de tomar una muestra de una población que evoluciona según el *modelo simple asimétrico con mutación* del capítulo 3 y finalmente presentamos gráficas para la probabilidad de fijación de individuos con selección, esto con la construcción del modelo simple asimétrico del capítulo 2 programado en el lenguaje R.

Capítulo 1

Genealogía de poblaciones neutras

Modelos fundamentales en genética de poblaciones son el modelo de Wright-Fisher y el modelo de Moran que estudian la evolución de una población haploide, es decir, cada individuo tiene exactamente un padre, la población tiene tamaño fijo y todos los individuos tienen la misma posibilidad de reproducirse (población neutra). La diferencia entre los dos modelos es el tiempo, el cual es discreto para el primero y es continuo para el segundo. Cuando el tamaño de la población y el tiempo (reescalado i.e medido en unidades del tamaño de la población del modelo) tienden a infinito, obtenemos procesos de Markov que son soluciones de ecuaciones diferenciales estocásticas, Esto nos permite usar la teoría de difusiones para estudiar características interesantes de la población. La genealogía de la difusión de Wright-Fisher es descrita por el coalescente de Kingman [8], una cadena de Markov en las particiones de \mathbb{N} , cuyas transiciones, llamadas coalescencias, suceden según un proceso Poisson de parámetro 1. Supuestos como selección y mutación pueden ser agregados a los modelos neutros, dichas fuerzas evolutivas cambian la dinámica de la población, lo cual se refleja en la (EDE) ecuación diferencial estocástica correspondiente al tomar el límite y al mismo tiempo reescalar el tiempo de dichos modelos.

En el modelo de Wright-Fisher de tamaño $N \in \mathbb{N}$, en cada generación $g_i \in \mathbb{N}$ la población se renueva por completo, de la siguiente forma: cada uno de los N individuos de la nueva generación $l + 1$ elige de manera independiente un padre de la generación anterior l con la misma probabilidad. La probabilidad de que dos individuos elijan exactamente el mismo individuo de la generación anterior como padre en una población de tamaño N es $\frac{1}{N}$, y que tres individuos elijan exactamente el mismo individuo como padre el mismo padre es $\frac{1}{N^2}$, etcétera. Por lo que la probabilidad de que $j \geq 3$ individuos elijan exactamente el mismo individuo de la generación anterior como padre es $\frac{1}{N^j}$, y cuando el tamaño de la población es muy grande ($N \rightarrow \infty$), la probabilidad de que $j \geq 3$ individuos tengan el mismo padre en la generación anterior es muy pequeña, con lo cual estos modelos no permiten que un individuo en un evento de reproducción produzca más de dos hijos, esto será importante ya que en nuestro modelo más importante, el *modelo simple asimétrico* del capítulo 2, nos interesa modelar una población que tiene un gran número de descendientes, para así tener una población cuya genealogía es descrita por procesos con múltiples coalescencias. Suponer que en algunos eventos de reproducción hay individuos que pueden tener una reproducción de más de dos hijos, es una generalización a modelos como el de Wright-Fisher y el modelo de Moran. El Λ -coalescente es un proceso que describe la genealogía de poblaciones que tienen la característica de tener múltiples hijos en una reproducción, por lo que es una generalización al coalescente de Kingman. En el proceso genealógico de una población, consideramos una muestra de tamaño n de nuestra población de un tamaño

total de N individuos (o genes) y viajamos hacia atrás en el tiempo, buscando los ancestros en común de dicha muestra. Por ejemplo, en tiempo discreto, si en nuestra muestra de tamaño n tenemos dos individuos que son primos, entonces sus líneas ancestrales coalescen dos generaciones atrás, ya que no tienen los mismos padres pero si los mismos abuelos (por parte de la madre, por ejemplo). Para poder reconstruir la genealogía de una población, es necesario saber las reglas con las que la población se renueva, es decir, la forma en que la descendencia toma lugar en la población. Los modelos más simples, son el modelo de Wright-Fisher y el modelo de Moran, los cuales describen la evolución de una población que cumple con los siguientes supuestos:

- 1) el tamaño de la población permanece constante en el tiempo,
- 2) todos los individuos tienen la misma probabilidad de ser elegidos como padre.

Además en el modelo de Moran el tiempo es continuo y las generaciones se pueden superponer en el tiempo, mientras que en el modelo de Wright-Fisher el tiempo es discreto y en cada generación la población se renueva por completo. Ahora explicamos e ilustramos estos modelos de manera más detallada.

1.1. Modelo de Wright-Fisher

En una población de tamaño fijo N y con individuos (o genes) x_1, \dots, x_N en la generación l . Se construye la generación $l + 1$ de la siguiente forma: cada individuo de la generación $l + 1$ elige como padre a un individuo x_i de la generación l de manera independiente, uniformemente al azar y con reemplazo.

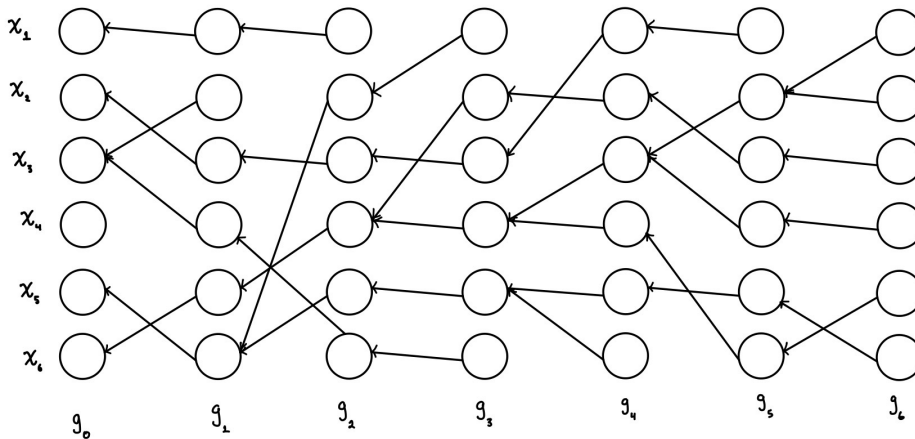


Figura 1.1: Representación gráfica del modelo de Wright-Fisher con tamaño de población 6, los círculos en cada generación representan los individuos, padre e hijo se conectan con líneas que unen a un individuo con otro de la generación anterior.

Definimos ahora la gráfica de Wright-Fisher] para definir sobre ella el proceso de frecuencia y el proceso de ancestría del modelo de Wright-Fisher neutro.

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

A los subconjuntos de V que coinciden en la primera coordenada les llamamos generaciones y los denotamos por $Gen(g)$, es decir, para cada $g \in \mathbb{Z}$, $Gen(g) = \{(g, i) \in V | i \in [N]\} \subset V$.

Definición 2. Sea $g_0 \in \mathbb{Z}$, $N \in \mathbb{N}$ y (V, E) la gráfica de Wright-Fisher con N individuos por generación. Consideramos una muestra de $n \in [N]$ individuos en la generación g_0 a la cual denotamos por S . Definimos el proceso de ancestría de $S \subset Gen(g_0) \subset V$ para cada $g \in \mathbb{Z}^+$ como

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

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

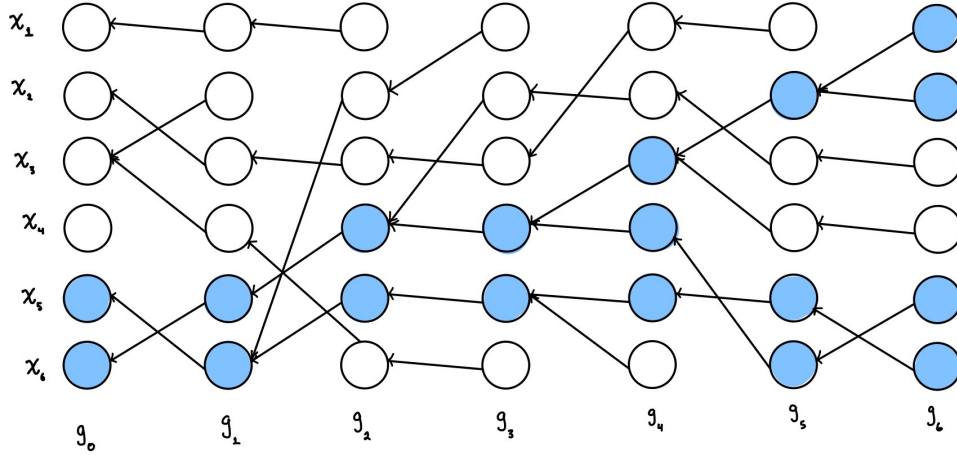


Figura 1.2: Representación gráfica del proceso de ancestría de Wright-Fisher.

Observaciones:

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

Supongamos que tenemos una población de tamaño $N \in \mathbb{N}$ en la generación g_0 y una muestra S de tamaño $n \in [N]$ y quisiéramos calcular cuál es la probabilidad de que los n

individuos no tengan el mismo padre, es decir, que cada uno tenga un ancestro diferente en la generación $g_0 - 1$,

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

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

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

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

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

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

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

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

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

Definición 3 (Proceso de frecuencia de Wright-Fisher neutro). *Sea $N \in \mathbb{N}$ fijo, (V, E) la gráfica de Wright-Fisher, $x \in [0, 1]$ y $\zeta \subseteq \text{Gen}(g_0) \subseteq V$ tal que $|\zeta| = x \cdot N$. Definimos el*

proceso de frecuencia de Wright-Fisher neutro para cada $g \in \mathbb{Z}^+$ como

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

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

Observaciones:

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

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

1.2. Modelo de Moran

En una población de tamaño fijo N con individuos $x_1(t), \dots, x_N(t)$ al tiempo $t \in \mathbb{R}$. Las reproducciones ocurren según un proceso Poisson de parámetro $\lambda = 1$. En un evento de reproducción se eligen uniformemente al azar dos individuos, de los cuales uno muere y es reemplazado por una copia del otro individuo.

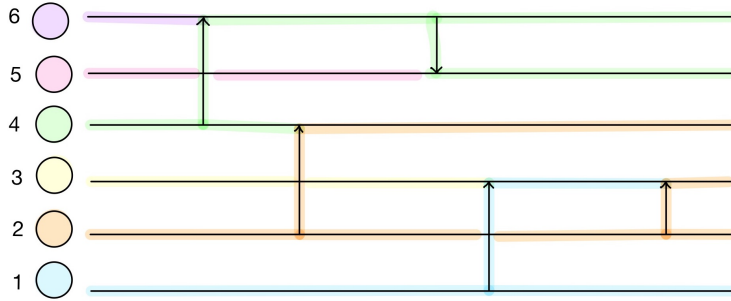


Figura 1.3: Representación gráfica del modelo de Moran, en donde el tipo de cada individuo se representa con un color. En cada evento de reproducción representado por flechas, el padre hereda su color al individuo al final de la flecha.

Definición 4 (Proceso de frecuencia del modelo de Moran). *En el modelo de Moran con tamaño de población N y formada por dos tipos de individuos (tipo 1 y tipo 0), definimos el proceso de frecuencia de individuos de tipo 0 del modelo de Moran neutro como*

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

y lo denotamos como $(X_t, t \geq 0)$

En cada evento de reproducción participan dos individuos, uno vive y se reproduce y el otro muere. Si la población esta formada por dos tipos de individuos k individuos de tipo 0 y $(N - k)$ individuos de tipo 1, se elige a un individuo de tipo 1 con probabilidad $\frac{N-k}{N}$ y se elige un individuo de tipo 0 con probabilidad $\frac{k}{N}$. Los individuos en la población tienen reproducciones, que suceden a tasa 1, es decir, el tiempo entre una reproducción y la siguiente (de un individuo) tiene distribución exponencial de parámetro 1. Podemos pensar que cada individuo de la población tiene un reloj exponencial de parámetro uno y el que suena primero es el que se reproduce, es por esto que la tasa total de reproducción en la población es N ya que $\min\{\exp_1(1), \dots, \exp_N(1)\} \sim \exp(N)$.

Para que el número de individuos de tipo 0 aumente y pase de k a $k + 1$, debe morir un individuo de tipo 1 (se elige un individuo de entre los $(N - k)$ que hay con probabilidad $\frac{N-k}{N}$) y se debe reproducir un individuo de tipo 0 (hay k individuos de tipo 0), por lo tanto, la probabilidad de transición es $N \frac{k}{N} \frac{(N-k)}{N}$. De manera similar el número de individuos de tipo 1 aumenta en uno a tasa $\frac{(N-k)k}{N}$. El número de individuos de tipo 0, equivalentemente de tipo 1, en la población cambia si los individuos elegidos en cada reproducción son de distinto tipo y si se eligen dos individuos del mismo tipo no hay ningún cambio.

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

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

y las tasas de transición del proceso de frecuencia de individuos de tipo 0 del modelo Moran neutro, que llamamos *proceso de frecuencia del modelo Moran neutro*, cuando $X_0 = \frac{i}{N}$ son

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

1.3. Modelo Look-Down

En una población de tamaño N con individuos (o genes) denotados por el vector $(x_1(t), \dots, x_N(t))$ indicando que cada individuo ocupa un nivel (de 1 hasta N) en la población. Los eventos de reproducción ocurren según un proceso Poisson de parámetro $\lambda = 1$, en los cuales se elige uniformemente al azar un nivel (equivalentemente un individuo en ese nivel), el j -ésimo por ejemplo, el cual copia el tipo del individuo elegido uniformemente al azar de entre los individuos menores que él, $\{1, 2, 3, \dots, j-1\}$, i.e el padre siempre busca en los niveles de abajo para elegir a su hijo.

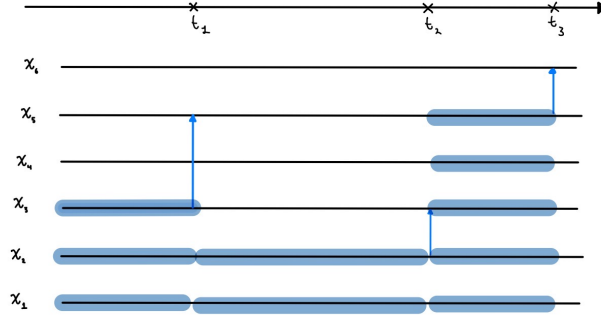


Figura 1.4: Representación gráfica del modelo Look-Down en donde el individuo $j = 5$ elige entre los individuos $\{1, 2, 3, 4\}$ al 3 como padre.

1.4. Modelo de Cannings

En una población de individuos haploides de tamaño N_j en la generación j . Denotamos v_i^j el número de hijos del individuo i en la generación j , en el modelo de Cannings la sucesión $(\mathbf{v}(j), j \in \mathbb{Z})$ es una sucesión de variables aleatorias i.i.d con $\mathbf{v}(j) = (v_1^j, \dots, v_{N_j}^j)$ y $\sum_{i=1}^{N_j} v_i^j = N_{j+1}$. Cada $\mathbf{v}(j)$ es intercambiable, es decir, para cualquier permutación σ en las permutaciones de N_j

$$(v_1^j, \dots, v_{N_j}^j) \stackrel{d}{=} (v_{\sigma(1)}^j, \dots, v_{\sigma(N_j)}^j).$$

El modelo de Wright-Fisher y el modelo de Moran son casos particulares de los modelos de Cannings cuando $N_j = N$ para toda generación j . Los individuos $1, 2, \dots, N$ en la generación j se reproducen para formar la siguiente generación $j + 1$ y tienen una cantidad $v_1^{j+1}, v_2^{j+1}, \dots, v_N^{j+1}$ de hijos respectivamente y también se pide que la sucesión $v_1^j, v_2^j, \dots, v_N^j$ tenga la misma distribución para toda generación j (intercambiable) pues todos los individuos tienen la misma probabilidad de ser elegidos como padres, indicando que no hay selección.

Definición 5. Decimos que un proceso evolutivo es neutro si el proceso de frecuencia correspondiente $(X_n, n \geq 0)$ satisface que $\mathbb{E}[X_{k+1} | X_k = i] = i \quad \forall i \in [0, 1]$.

1.5. Cadenas de Markov a tiempo continuo

Si consideramos los procesos de frecuencia, que indican la frecuencia en la población total de una muestra tomada al tiempo inicial del modelo de Wright-Fisher y del modelo de Moran, ambos son cadenas de Markov, a tiempo discreto y a tiempo continuo respectivamente. Además, en el límite, el proceso que describe su genealogía es el mismo. Recordaremos la definición de una cadena de Markov a tiempo continuo y de su generador, ya que teniendo este último podemos conocer cómo evoluciona un proceso de Markov en el tiempo.

Definición 6 (Cadena de Markov a tiempo continuo). Sea $(X_t, t \geq 0)$ un proceso estocástico tal que para cada $t \in \mathbb{R}^+$ la variable aleatoria X_t toma valores como sigue,

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

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

Proposición 7. Sea $(X_t, t \geq 0)$ una cadena de Markov a tiempo continuo. Entonces, para todo $t_0 \leq t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n \in \mathbb{R}^+$ y estados $i, j, k, \dots, l \in E$ tales que $\mathbb{P}[X_{t_{n-1}} = i, X_{t_{n-2}} = k, \dots, X_{t_0} = l] > 0$, se cumple que

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

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

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

Denotamos por $\mathbb{E}_y[X]$ a la esperanza condicional de X dado $X_0 = y$.

Definición 8 (Generador de un proceso de Markov). [6] Sea $(X_t, t \geq 0)$ un proceso de Markov a tiempo continuo, homogéneo y real valuado. Para una función $f : \mathbb{R} \rightarrow \mathbb{R}$ se define

$$Af(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}_x[f(X_t) - f(x)]}{t}, \quad (1.1)$$

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

La siguiente definición servirá para entender la relación de dualidad que existe entre el coalescente de Kingman y el proceso límite del proceso de frecuencia del modelo de Wright-Fisher, este proceso límite se conoce como la difusión de Wright-Fisher, en este trabajo no se pretende abordar el tema de difusiones de manera detallada.

Definición 9 (Difusión unidimensional). Una difusión unidimensional $(X_t, t \geq 0)$ es un proceso de Markov con generador infinitesimal

$$Af(x) = \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} + \mu(x)\frac{df}{dx}. \quad (1.2)$$

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

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

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

y la variancia del proceso de frecuencia es

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

Por lo tanto

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

y

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

Entonces su generador infinitesimal reescalando el tiempo por N es

$$A^{(N)} f(x) = \frac{1}{2} [x(1-x) + o(1)] \frac{d^2 f}{dx^2}, \quad (1.3)$$

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

$$\lim_{N \rightarrow \infty} A^{(N)} f(x) = \frac{1}{2} [x(1-x)] \frac{d^2 f}{dx^2} = A f(x), \quad (1.4)$$

Al proceso $(X_t, t > 0)$ cuyo generador es (1.4) se le conoce como difusión de Wright-Fisher $(X_t, t > 0)$

1.6. Coalescente de Kingman

La genealogía de una muestra de tamaño n en una población puede ser descrita por procesos en $\mathcal{P}^{(n)}$, las particiones de $[n] = \{1, 2, 3, \dots, n\}$. Los números que se encuentran en la misma clase corresponden a los individuos tienen un ancestro en común. Esto se ilustra en la Figura 1.5.

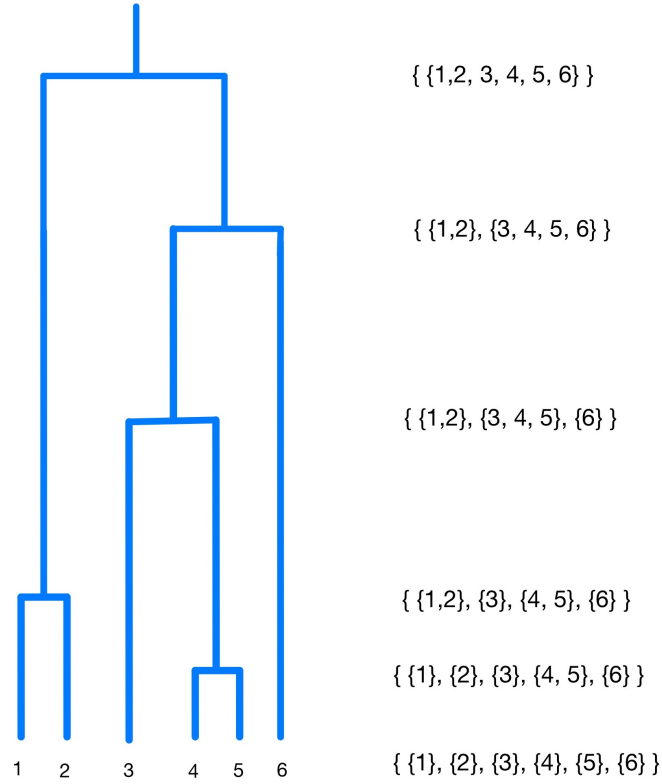


Figura 1.5: Representación gráfica de la genealogía de una muestra de 6 individuos.

Los n -coalescentes son procesos de Markov, $(\Pi_t^{[n]}, t > 0)$ con valores en $\mathcal{P}^{(n)}$, con tasas de transición, $q_{\xi, \eta}$ dadas por

$$q_{\xi, \eta} = \begin{cases} 1 & \text{si } \eta \text{ se obtiene al coalescer dos clases de equivalencia de } \xi, \\ 0 & \text{en otro caso,} \end{cases} \quad (1.5)$$

para toda ξ y $\eta \in \mathcal{P}^{(n)}$.

El *coalescente de Kingman* es un proceso de Markov en \mathcal{P} , las particiones de \mathbb{N} , tal que restringido a $\mathcal{P}^{(n)}$ resulta ser un n -coalescente. El coalescente de Kingman describe la genealogía de los procesos límite de los procesos de frecuencia del modelo de Wright-Fisher y del modelo de Moran. El proceso $|\Pi_t^{[n]}|$, llamado el *proceso de conteo de bloques*, es el proceso que indica el número de ancestros al tiempo t del n -coalescente de Kingman. El proceso $(|\Pi_t^{[n]}|, t \geq 0)$ es una cadena de Markov con valores en \mathbb{N} y tasas de transición de n a m dadas por

$$Q_{n, m} = \begin{cases} \binom{n}{2} & \text{si } m = n - 1 \\ -\binom{n}{2} & \text{si } m = n \\ 0 & \text{si en otro caso} \end{cases} \quad (1.6)$$

El proceso de conteo de bloques del coalescente de Kingman y la difusión de Wright-Fisher (el proceso límite del modelo de Wright-Fisher) tienen generadores

$$A^K f(x) = \binom{n}{2} [f(n-1) - f(n)], \quad n \in \mathbb{N}$$

y

$$A^{W-F} f(x) = \frac{1}{2} x(1-x) f''(x), \quad x \in [0, 1]$$

respectivamente. El n -coalescente de Kingman comienza con el estado $\{\{1\}, \{2\}, \{3\}, \dots, \{n\}\}$ y evoluciona con coalescencias binarias que suceden a tasa 1. El proceso $(|\Pi_t^{[n]}|, t > 0)$ es decreciente, comienza en el estado $N|\Pi_0^{[n]}| = n$ y pasa al estado $n-1$ después de un tiempo exponencial de parámetro $\binom{n}{2}$, cuando son elegidos dos bloques del proceso $(\Pi_t^{[n]}, t > 0)$ para coalescer en uno solo. Cuando $|\Pi_t^{[n]}| = 1$ significa que hemos llegado al ancestro común de la muestra de n individuos. Los tiempos de salto del proceso $(\Pi_t^{[n]}, t > 0)$ cuando $\tau_0^{[n]} = 0$ se definen como $\tau_r^{[n]} := \inf\{t > \tau_{r-1}^{[n]} \mid \Pi_{t-} \neq \Pi_t\}$ y el tiempo que transcurre hasta encontrar el ancestro común de una muestra de tamaño n denotado por $TMRCA(n)$ se define como

$$TMRCA(n) := \inf\{t > 0 \mid |\Pi_t^{[n]}| = 1\}$$

y para el n -coalescente de Kingman es

$$\begin{aligned} \mathbb{E}[TMRCA(n)] &\stackrel{d}{=} \sum_{k=1}^n \mathbb{E} \left[\exp \left(\binom{k}{2} \right) \right] \\ &= \sum_{k=2}^n \frac{1}{\binom{k}{2}} \\ &= \sum_{k=2}^n \frac{2}{k(k-1)} \\ &= 2 \sum_{k=2}^n \left(\frac{1}{1-k} - \frac{1}{k} \right) \\ &= 2 \left(1 - \frac{1}{n} \right). \end{aligned}$$

Por lo anterior, cuando $n \rightarrow \infty$ esperamos aproximadamente 2 unidades de tiempo para encontrar el ancestro común de la muestra. Se utiliza el término *comes down from infinity* para describir la característica muy particular de los coalescentes definidos en las particiones de \mathbb{N} . Se dice que un proceso de coalescencia baja de infinito cuando después de cualquier cantidad de tiempo $t > 0$ el proceso $(\Pi_t^{[n]}, t > 0)$ tiene una cantidad finita de bloques, es decir, $|\Pi_t^{[n]}| < \infty$. El n -coalescente de Kingman tiene esta propiedad. La prueba del siguiente teorema puede ser consultado en la en la sección 2.1.2 de [1].

Teorema 10. *Sea E el evento que para todo $t > 0$, $|\Pi_t^{[n]}| < \infty$. Entonces $\mathbb{P}(E) = 1$.*

Tomando una muestra de tamaño n en una población que evoluciona según el modelo de Cannings con tamaño constante N , definimos c_N como la probabilidad de que dos individuos elegidos aleatoriamente y sin reemplazo tengan el mismo padre, es decir, $c_N = \sum_{i=1}^N \frac{v_i(v_i-1)}{N(N-1)} = \frac{v_1(v_1-1)}{N-1}$ y el proceso de coalescencia que describe la genealogía de esta muestra es el n -coalescente de Kingman. La demostración del siguiente lema puede ser consultada en [11].

Lema 11 (Lema de Möhle). *Consideramos un modelo de Cannings definido por sucesiones i.i.d. $\{(v_i(t))_{i=1}^N, t \in \mathbb{Z}\}$. Si*

$$\frac{\mathbb{E}[v_1(v_1 - 1)(v_1 - 2)]}{N^2 c_N} \rightarrow 0 \text{ cuando } N \rightarrow \infty,$$

entonces $c_N \rightarrow 0$ y la genealogía converge al coalescente de Kingman.

1.7. Modelo de Wright-Fisher y modelo de Moran con selección

Hasta ahora hemos explicado modelos en donde los individuos son elegidos para participar en los eventos de reproducción de manera uniforme, es decir, sin selección. A continuación presentamos el modelo de Wright-Fisher y el modelo de Moran con selección, y con mutación 1.8. En modelos con selección, cobra sentido dividir a la población en dos tipos, pues ahora la diferencia no solo está en la etiqueta de los individuos (tipo 0 y tipo 1) sino en su forma de evolucionar en el tiempo.

Definición 12 (Modelo Wright-Fisher con selección). *Sean $s \in (0, 1)$, $g \in \mathbb{Z}^+$, $N \in \mathbb{N}$, N fijo y $k \in \{1, 2, \dots, N - 1\}$. En una población de tamaño N formada por k individuos de tipo 0 y $N - k$ de tipo 1, los individuos de la generación $g + 1$ eligen padre de manera independiente a los demás y con reemplazo, como sigue, con probabilidad*

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

elige un padre de tipo 0 y con probabilidad

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

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

Ahora definimos el proceso de frecuencia asociado al modelo del Wright-Fisher con selección.

Definición 13 (Proceso de frecuencia de Wright-Fisher con selección). *Sea $V = \mathbb{Z} \times [N]$, dado que conocemos los tipos de los individuos $(g, i) \in V \forall i \in [N]$ definimos*

$$X_g^{(N)} = \frac{1}{N} \sum_{i \in [N]} 1_{\{\text{individuo } i \text{ es tipo 0 en la generación } g\}}$$

la frecuencia de individuos de tipo 0 en la generación g en un modelo de Wright-Fisher con selección. Llamamos al proceso $(X_g^{(N)}, g \in \mathbb{Z}^+)$ como el proceso de frecuencia de Wright-Fisher con selección.

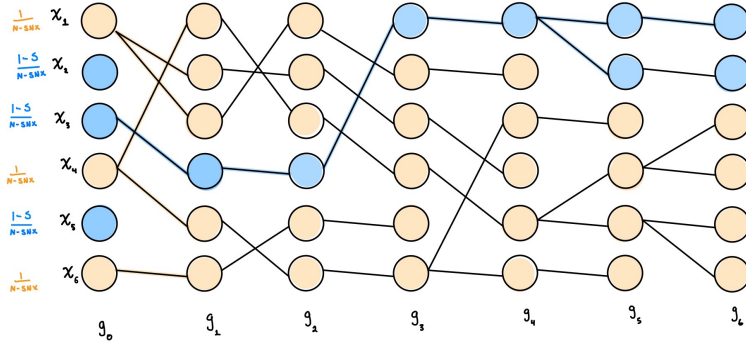


Figura 1.6: Representación gráfica del modelo de Wright-Fisher con selección. Un individuo con ventaja selectiva (naranja) es elegido con probabilidad $\frac{1}{N-Nsx}$ mientras que un individuo que no la tiene (azul) es elegido con probabilidad $\frac{1-s}{N-Nsx}$.

Observaciones:

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

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

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

Definición 14 (Modelo de Moran con selección). *Sea $N \in \mathbb{N}$ fijo y $s \in [0, 1]$. En el modelo de Moran con selección y tamaño de población constante N , consideramos dos tipos de individuos, tipo 0 y tipo 1. Cada individuo, ya sea de tipo 1 o tipo 0 se reproduce en tiempos aleatorios, los eventos de reproducción suceden según un proceso Poisson de parámetro 1, escogiendo al azar dos individuos de la población de manera uniforme, un individuo muere y el otro vive y se reproduce (teniendo un hijo como descendencia), ambos individuos tienen la misma probabilidad de vivir y reproducirse o morir. Además cada individuo, ya sea de tipo 1 o tipo 0 se reproduce en tiempos aleatorios, estos eventos de reproducción ocurren según un proceso Poisson de parámetro s , en estas reproducciones se escoge a un individuo de la población uniformemente al azar, si los individuos son del mismo tipo no sucede nada, de lo contrario el individuo de tipo 1 es el que siempre se reproduce y su hijo reemplaza al individuo de tipo 0 que muere.*

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

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

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

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

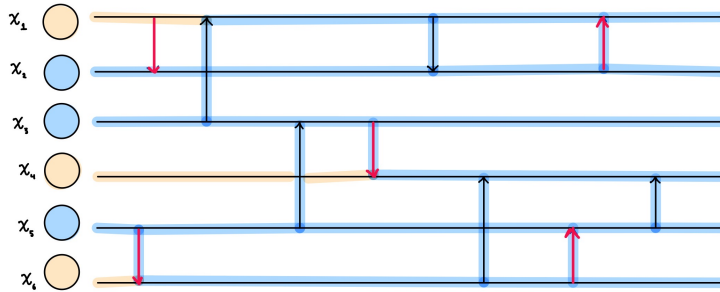


Figura 1.7: Representación gráfica del modelo de Moran con selección. Las flechas rojas representan los eventos de reproducción “extra” que tienen los individuos con ventaja selectiva (individuos azules).

Definición 15 (Proceso de frecuencia del modelo de Moran con selección). *En el modelo de Moran con selección y tamaño de población N formada por dos tipos de individuos (tipo 1 y tipo 0), definimos el proceso de frecuencia de individuos de tipo 0 del modelo de Moran con selección como*

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

Observaciones:

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

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

1.8. Modelo de Wright-Fisher y modelo de Moran con mutación

Una mutación se define como “un cambio al azar en en la secuencia del ADN”, en el modelo de Wright-Fisher con mutaciones, consideraremos que ocurre una mutación cuando

un individuo cambia de tipo. Los individuos de tipo 0 cambian a tipo 1 con probabilidad $\mu_{0,1}$ y los individuos de tipo 1 cambian a tipo 0 con probabilidad $\mu_{1,0}$.

Definición 16 (El modelo de Wright-Fisher con mutaciones). Sean $\mu_{0,1}, \mu_{1,0} \in (0,1)$, $g \in \mathbb{Z}^+$, $N \in \mathbb{N}$, N fijo y $k \in \{1, 2, \dots, N-1\}$. En una población de tamaño N formada por k individuos de tipo 0 y $N-k$ de tipo 1, los individuos de la generación g se reproducen al ser elegidos de manera independiente, uniformemente al azar y con reemplazo. Los individuos de la generación $g+1$ copian el tipo de su padre. Además con probabilidad $\mu_{(0,1)}$ los individuos de tipo 0 sufren una mutación y cambia a tipo 1 y con probabilidad $\mu_{(1,0)}$ los individuos de tipo 1 cambian de tipo 1 a tipo 0.

Definición 17 (Proceso de frecuencia de Wright-Fisher con mutaciones). Sea $V = \mathbb{Z} \times [N]$, dado que conocemos los tipos de los individuos $(g, i) \in V \forall i \in [N]$ definimos

$$X_g^{(N)} = \frac{1}{N} \sum_{i \in [N]} 1_{\{\text{individuo } i \text{ es tipo 0 en la generación } g\}}$$

la frecuencia de individuos de tipo 0 en la generación g en un modelo de Wright-Fisher con mutación. Llamamos al proceso $(X_g^{(N)}, g \in \mathbb{Z}^+)$ como el proceso de frecuencia de Wright-Fisher con mutaciones.

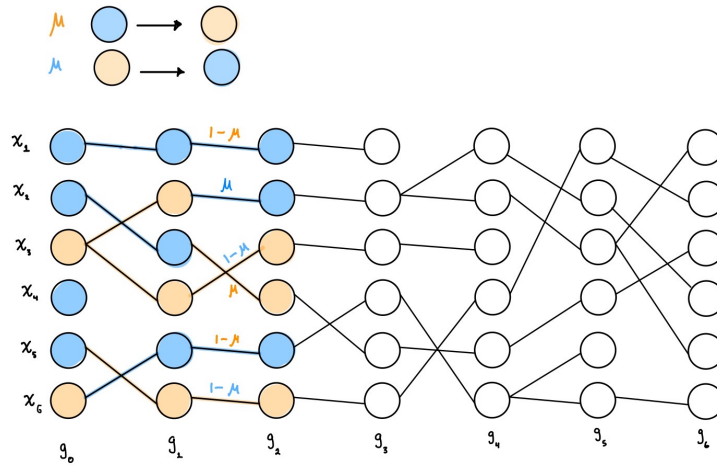


Figura 1.8: Representación gráfica del modelo de Wright-Fisher con mutaciones en donde al construir una nueva generación, los individuos pueden sufrir una mutación, cambiándolos de tipo (color).

Observaciones:

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

- La probabilidad de transición es

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

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

Definición 18 (Modelo de Moran con mutaciones). *Sea $N \in \mathbb{N}$ fijo y $\mu_{0,1}, \mu_{1,0} \in [0, 1]$. En el modelo de Moran con mutaciones y tamaño de población N , consideramos dos tipos de individuos, tipo 0 y tipo 1. Cada individuo, sin importar su tipo, se reproduce en tiempos aleatorios, dichos eventos de reproducción ocurren según un proceso Poisson de parámetro $\lambda = 1$, escogiendo a un individuo de la población uniformemente al azar, un individuo se reproduce y el otro muere. Además cada individuo tiene una mutación y cambia de tipo, las mutaciones que tienen los individuos de tipo 0 suceden según un proceso Poisson de parámetro $\mu_{0,1}$ y las mutaciones que tienen los individuos de tipo 1 suceden según un proceso Poisson de parámetro $\mu_{1,0}$.*

Definición 19 (Proceso de frecuencia del modelo de Moran con mutaciones). *En el modelo de Moran con mutaciones y tamaño de población N formada por dos tipos de individuos (tipo 1 y tipo 0), definimos el proceso de frecuencia de individuos de tipo 0 del modelo de Moran con mutaciones como:*

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

Observaciones:

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

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

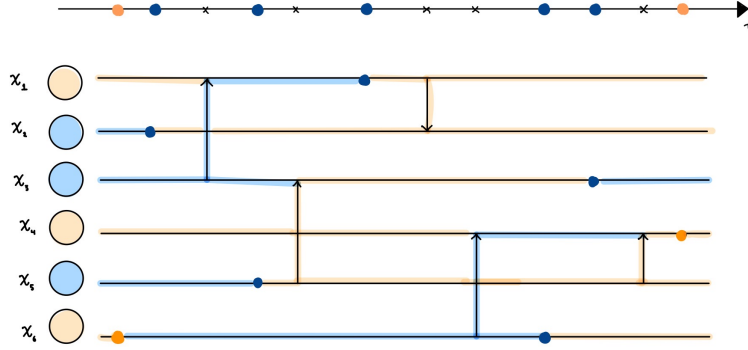


Figura 1.9: Representación gráfica del modelo de Moran con mutaciones. Los eventos de mutación y reproducción se representan en la recta real, la línea de la parte superior que representa el tiempo.

1.9. Λ -coalescente

Los procesos llamados Λ -coalescentes fueron introducidos en 1999 por Pitman [12] y Sagitov[13] de forma independiente. Un Λ -coalescente es un proceso de Markov que toma valores en las particiones de \mathbb{N} , intercambiable con eventos de coalescencias simples, es decir, no se permiten coalescencias simultáneas. A diferencia del coalescente de Kingman las coalescencias pueden ser de más de dos bloques o líneas ancestrales.

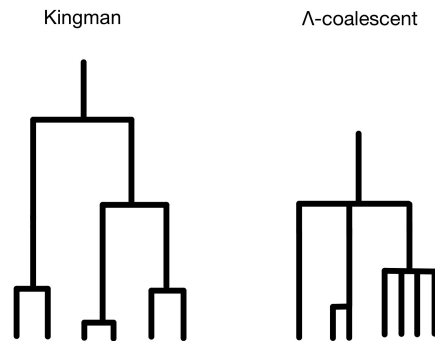


Figura 1.10: Representación gráfica del coalescente de Kingman y del Λ -coalescente.

Definición 20 (Λ -coalescente). Sea Λ una medida no negativa en $[0, 1]$. Un Λ -coalescente es un proceso de Markov $(\Pi^\infty(t), t \geq 0)$ con espacio de estados \mathcal{P}_∞ , el conjunto de las particiones de \mathbb{N} , tal que para cada n la restricción $\Pi^n := (\Pi^n(t), t \geq 0)$ de Π_∞ a $[n]$ es una cadena de Markov tal que k bloques, de un total de $b \in \mathbb{N}$ bloques, coalescen a un bloque a tasa

$$\lambda_{b,k} = \int_0^1 x^k (1-x)^{b-k} \frac{\Lambda(dx)}{x^2}, \text{ con } 2 \leq k \leq b. \quad (1.7)$$

La tasa total a la que k líneas coalescen dado que hay b es

$$\lambda_b = \sum_{k=2}^b \binom{b}{k} \lambda_{b,k}, \quad (1.8)$$

Así, las tasas de transición de Π^n están dadas por

$$q_{\xi,\eta} = \begin{cases} \int_0^1 x^k (1-x)^{b-k} \frac{\Lambda(dx)}{x^2} & \text{si } \eta \text{ se obtiene al coalescer } k \text{ clases de equivalencia de } \xi, \\ 0 & \text{en otro caso,} \end{cases} \quad (1.9)$$

con $\xi, \eta \in \mathcal{P}^{(n)}$.

Ejemplo 21. Sea $\Lambda = \delta_0(dx)$, en este caso la ecuación 1.7 es $\lambda_{b,k} = 0$ para toda $k \neq 2$ y $\lambda_{b,2} = 1$. Entonces el Λ coalescente correspondiente a esta medida Λ es el coalescente de Kingman, donde cada par de bloques coalesce a tasa 1.

Ejemplo 22. Sea $\Lambda = U(dx)$, la medida uniforme en $(0, 1)$. En este caso el Λ coalescente correspondiente es conocido como el coalescente de Bolthausen-Sznitman con

$$\lambda_{b,k} = \frac{(k-2)!(b-k)!}{(b-1)!}, \quad 2 \leq k \leq b.$$

Ejemplo 23. Sea $0 < \alpha < 2$. Si tenemos que $\Lambda(dx) = \text{Beta}(2-\alpha, \alpha)$

$$\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha} (1-x)^{\alpha-1} dx.$$

El coalescente es conocido como coalescente-Beta de parámetro α .

Ejemplo 24. Sea $\Lambda = \delta_1(dx)$, en este caso nada pasa durante un tiempo exponencial de media 1, hasta que todos los bloques coalescen en 1. El árbol de coalescencia es con forma de estrella.

Ejemplo 25. Sea $y \in [0, 1]$ fija y $\Lambda = y^2 \delta_y(dx)$ en este caso

$$\lambda_{b,k} = \int_0^1 x^{k-2} (1-x)^{b-k} y^2 \delta_y(dx) = \binom{b}{k} y^{k-2} (1-y)^{b-k} y^2 = \binom{b}{k} y^k (1-y)^{b-k},$$

conocido como coalescente de Eldon-Wakely.

1.10. Construcción de Poisson

Presentamos ahora la construcción dada por J. Pitman en *Coalescents with multiple collisions* (1999) [12].

Sea Λ una medida en $(0, 1]$. Consideramos una colección de puntos de un proceso de Poisson puntual N en $(0, \infty) \times \{0, 1\}^{\mathbb{N}}$ con intensidad $dt \times L(d\xi)$, con

$$L := \int_{(0,1]} x^{-2} \Lambda(dx) P_x,$$

$\xi = \{\xi_i\}_{i \geq 1}$ una sucesión de variables aleatorias Bernoulli de parámetro $P_x(\xi_i = 1) = x$ para toda $i \geq 1$. Si π es una partición de \mathbb{N} en conjuntos unitarios, definimos Π_0^n la restricción de

π a $[n]$. Sea (t, ξ_1, ξ_2, \dots) un punto del proceso N tal que $\sum_{i=1}^n \xi_i \geq 2$, entonces el proceso $\{\Pi_t^n\}_{t \geq 0}$ tiene un cambio de estado al tiempo t . Si $\Pi_{t-}^n = \{B_1, B_2, \dots, B_b\}$, entonces Π_t^n se obtiene de Π_{t-}^n al unir aquellos B_i (bloques) tales que $\xi_i = 1$. La transición sucede si y solo si $\sum_{i=1}^n \xi_i \geq 2$, es decir, si hay 2 o más bloques que participan ($\xi_i = 1$), ya que la coalescencia de un bloque consigo mismo no produce ningún cambio.

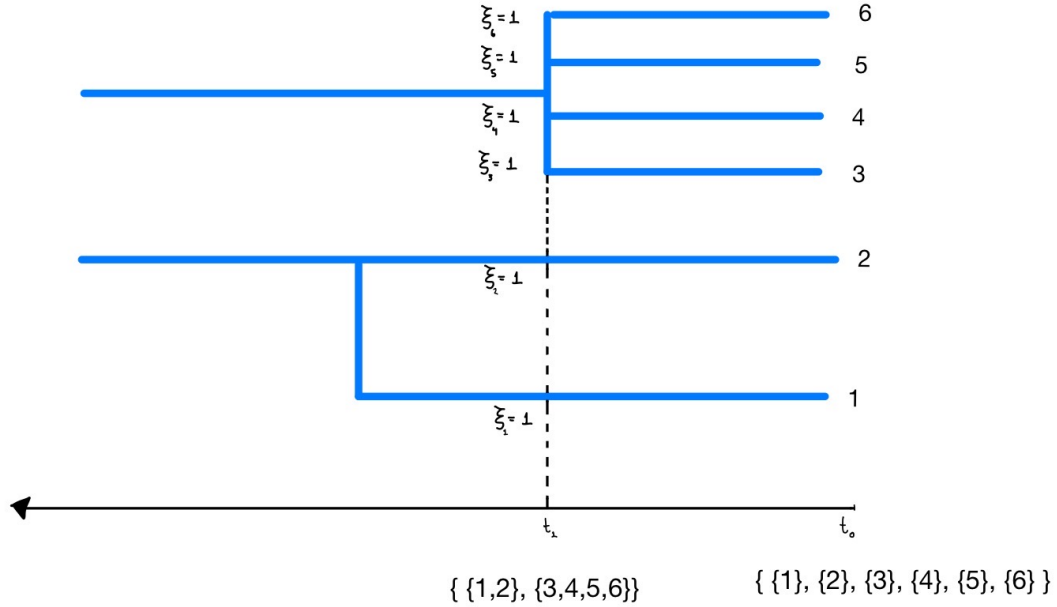


Figura 1.11: Evento de coalescencia.

1.11. Construcción paintbox

Sea \mathcal{S}_0 el espacio de las particiones del intervalo $(0, 1)$, si $s \in \mathcal{S}_0$, s es de la forma $s = (s_0, s_1, \dots)$ con $s_1 \geq s_2 \geq \dots \geq 0$ y $\sum_{i=0}^{\infty} s_i = 1$, la sucesión (s_0, s_1, \dots) , nos permite dividir al intervalo $(0, 1)$ en subintervalos de tamaño s_i , el orden en que acomodamos dichos subintervalos no es importante para esta construcción, ya que tomaremos variables aleatorias uniformemente distribuidas en $(0, 1)$ independientes entre ellas e independientes de la sucesión s , pero podemos pensar que los acomodamos de la siguiente forma $(0, s_0)$, $(s_0, s_0 + s_1)$, $(s_0 + s_1, s_0 + s_1 + s_2)$, etcétera. Podemos obtener una partición intercambiable en este espacio de la siguiente manera. Sea $s \in \mathcal{S}_0$ y U_1, U_2, \dots v.a independientes con distribución uniforme en $(0, 1)$ e independientes de s , para $0 < u < 1$ definimos $I(u)$ como el índice i del subintervalo $(s_0 + s_1 + \dots + s_{i-1}, s_0 + s_1 + \dots + s_i)$ al que pertenece u ,

$$I(u) = \inf \left\{ n \mid \sum_{i=0}^n s_i > u \right\}.$$

Consideramos la partición aleatoria Π con la relación $i \sim j$ si y solo si $I(U_i) = I(U_j) > 0$ o si $i = j$, es decir, dos índices distintos están relacionados si sus variables aleatorias uniformes correspondientes pertenecen al mismo subintervalo formado por la sucesión s . La parte $(0, s_0)$ corresponde a los índices que están relacionados con si mismos, es decir, son aquellos bloques que solo tienen un elemento (conjuntos unitarios de la partición). La partición Π así construida es una partición intercambiable ya que $(U_1, \dots, U_n) \stackrel{d}{=} (U_{\sigma(1)}, \dots, U_{\sigma(n)})$, para σ una permutación de $[n]$.

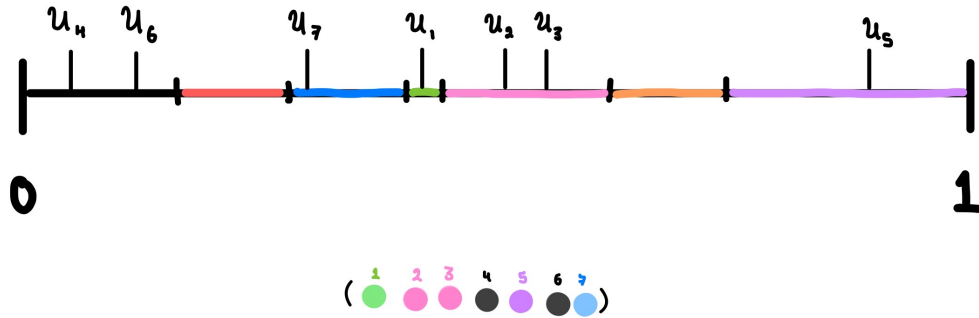


Figura 1.12: La partición asociada a la división del $(0, 1)$ en subintervalos que se muestra en la figura es $\Pi_{[7]} = (\{1\}, \{2, 3\}, \{4\}, \{5\}, \{6\}, \{7\})$, los índices 4 y 6 son conjuntos unitarios ya que las variables aleatorias u_4 y u_6 pertenecen al subintervalo $(0, s_0)$.

Usando esta construcción de particiones intercambiables, construimos el Λ -coalescente como sigue. Consideramos una colección contable de puntos de un proceso de Poisson puntual N en $(0, \infty) \times (0, 1)^{\mathbb{N}}$ con intensidad $dt \times x^{-2} \Lambda(dx) \times \prod_{i=1}^{\infty} U_i$, con U_i variables aleatorias uniformes en $(0, 1)$ independientes e idénticamente distribuidas y Λ una medida finita en $[0, 1]$ con $\Lambda(\{0\}) = 0$. Si π es una partición de \mathbb{N} , definimos Π_0^n la restricción de π a $[n]$. El proceso $\{\Pi_t^n\}_{t>0}$ tiene saltos en los tiempos $t > 0$ tales que en (t, x, U_1, U_2, \dots) punto del proceso N se toma $s_x \in \mathcal{S}_0$ como $s_x = (1-x, x)$ y por lo tanto la división del intervalo $(0, 1)$ es $(0, 1-x) \cup (1-x, 1)$. Para alguno de estos tiempos t si $\Pi_{t-}^n = \{B_1, B_2, \dots, B_b\}$, entonces Π_t^n se obtiene de Π_{t-}^n al unir aquellos B_i (bloques) tales que U_i pertenece al intervalo $(1-x, 1)$. Los bloques cuyas variables aleatorias U_i pertenecen al intervalo $(0, 1-x)$ no tienen ningún cambio. La transición sucede si y solo si hay más de 2 variables aleatorias que pertenecen al intervalo $(1-x, 1)$. Después de cada evento de coalescencia, el índice de los bloques corresponde al índice más pequeño que pertenece a dicho bloque después de la coalescencia, es decir, si $B_1 = \{1\}$ y $B_9 = \{9\}$ coalescen al tiempo t , entonces al tiempo t^+ B_1 es el bloque $B_1 = \{1, 10\}$.

1.12. Dualidad

En esta sección presentamos la definición de dualidad entre dos procesos de Markov respecto a una función de dualidad y presentamos un ejemplo de dualidad de momentos, cuando la función de dualidad es $H(x, n) = x^n$, esta dualidad es clásica en el campo de

genética de poblaciones. Esta técnica de dualidad nos permite relacionar dos procesos estocásticos, uno que describe la evolución de cierta población y otro que representa la genealogía de una muestra tomada de dicha población. La existencia del proceso dual para un proceso de Markov dado y un estudio detallado de la teoría de dualidad puede ser encontrado en [14]. La siguiente definición se encuentra en [14] página 61.

Definición 26 (Procesos H-duales). Sean $X = (X_t, t \geq 0)$ y $Z = (Z_t, t \geq 0)$ dos procesos de Markov con espacio de estados S_1 y S_2 respectivamente y $H : E \times F \rightarrow \mathbb{R}$ una función acotada y medible. Decimos que X y Z son duales con respecto a la función H o H-duales si para cada $x \in S_1$, $z \in S_2$ y $t > 0$, se cumple que

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

La siguiente proposición resulta útil cuando se quiere demostrar que dos procesos son duales respecto a una función H utilizando los generadores de dichos procesos.

Proposición 27. Sean $(X_t, t \geq 0)$ y $(Z_t, t \geq 0)$ dos procesos de Markov con generadores \mathcal{A}^X y \mathcal{A}^Z , sea $H : E \times F \rightarrow \mathbb{R}$ una función continua y acotada. Si $H(x, \cdot), P_t H(x, \cdot) \in D(\mathcal{A}^Z)$ $\forall x \in E, t \geq 0$ y $H(\cdot, z), Q_t H(\cdot, z) \in D(\mathcal{A}^X)$ $\forall z \in F, t \geq 0$, y si

$$\mathcal{A}^X H(\cdot, z)(x) = \mathcal{A}^Z H(x, \cdot)(z) \quad \forall x \in E, \forall z \in F,$$

entonces $(X_t, t \geq 0)$ y $(Z_t, t \geq 0)$ son duales con respecto a H .

Ejemplo 28. Sean $(X_t, t \geq 0)$ la difusión de Wright–Fisher y $(Z_t, t > 0)$ el proceso de conteo de bloques del coalescente de Kingman, estos procesos son momento duales.

Demostración. Sea A_1 el generador infinitesimal de $(X_t, t \geq 0)$ y $H : [0, 1] \times \mathbb{N} \rightarrow [0, 1]$, definida como $H(x, n) = x^n$. Como $H(x, n)$ y $\mathbb{E}[x^{Z_t}]$ son polinomios en x y $H(X, n)$ y $\mathbb{E}[Z_t^x]$ son acotadas en n , por la Proposición 27, basta probar que para toda $x \in [0, 1]$ y para toda $n \in \mathbb{N}$, se cumple que

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

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

□

1.12.1. El modelo de Eldon-Wakeley

El modelo, presentado por Bjarki Eldon y John Wakeley en 2006 [4], en donde un individuo elegido uniformemente al azar puede tener múltiples descendientes en un evento de reproducción con cierta probabilidad (ρ) y solo un descendiente con otra cierta probabilidad ($1 - \rho$), en el modelo las generaciones se superponen, por lo que puede considerarse como una generalización del modelo de Moran. La genealogía de este modelo es descrita por un proceso de coaliciones múltiples no simultáneas, es decir, un caso particular del proceso presentado por Pitman(1999)[12] y Sagitov)1999)[13].

Definición 29 (El modelo de Eldon-Wakeley). *Sea $N \in \mathbb{N}$ y sea U una variable aleatoria con distribución P_U en $\{0, 1, \dots, N\}$. En una población de tamaño constante N , ocurren reproducciones según un proceso Poisson de parámetro 1, en las cuales se elige un individuo uniformemente al azar para reproducirse. El individuo elegido tiene $U - 1$ número de descendientes y él vive, mientras que $N - (U - 1)$ individuos elegidos al azar mueren para dar lugar a los nuevos $U - 1$ individuos y así mantener una población constante.*

Un ejemplo es cuando $P_U(u) = 1 - \rho$ si $u = 2$ y $P_U(u) = \rho$ si $u = \lfloor N\nu \rfloor$, para $\nu \in [0, 1]$ fija.

Definición 30 (El proceso de frecuencia de Eldon-Wakeley). *Sea $N \in \mathbb{N}$ y sean x y $z \in [0, 1]$. Consideramos una población que evoluciona según el modelo de Eldon-Wakeley con $U = z$. La población tiene dos tipos de individuos, al tiempo inicial hay Nx individuos de tipo 0 y $N(1 - x)$ individuos de tipo 1. Definimos el proceso de frecuencia de Eldon-Wakeley de individuos de tipo 0 como*

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

denotado como $(X_t, t \geq 0)$.

Las tasas de transición del proceso de frecuencia de Eldon-Wakeley son

$$x \text{ a } \begin{cases} x + y(1 - x) & \text{a tasa } x, \\ x - yx & \text{a tasa } (1 - x). \end{cases}$$

por lo tanto su generador es

$$\mathcal{A}f(x) = x[f(x(1 - y) + y) - f(x)] + (1 - x)[f(x(1 - y)) - f(x)],$$

y si consideramos la función $f_n(x) = x^n$, se tiene que el generador para $f_n(x)$ es

$$\begin{aligned}
\mathcal{A} f_n(x) &= x[f_n(x(1-y) + y) - f_n(x)] + (1-x)[f_n(x(1-y)) - f_n(x)] \\
&= x[(x(1-y) + y)^n - x^n] + (1-x)[(x(1-y))^n - x^n] \\
&= x \left[\sum_{b=0}^n \binom{n}{b} y^b x^{n-b} (1-y)^{n-b} - x^n \right] + (1-x)[(x(1-y))^n - x^n] \\
&= x \left[\sum_{b=0}^n \binom{n}{b} y^b x^{n-b} (1-y)^{n-b} \right] - x^{n+1} \left[\sum_{b=0}^n \binom{n}{b} y^b (1-y)^{n-b} \right] + (1-x)[(x(1-y))^n - x^n] \\
&= \sum_{b=0}^n \binom{n}{b} y^b (1-y)^{n-b} [x^{n-b+1} - x^{n+1}] + (1-x)[(x(1-y))^n - x^n] \\
&= \sum_{b=0}^n \binom{n}{b} y^b (1-y)^{n-b} [x^{n-b+1} - x^{n+1}] + (x^{n+1} - x^n)[1 - (1-y)^n] \\
&= \sum_{b=2}^n \binom{n}{b} y^b (1-y)^{n-b} [x^{n-b+1} - x^{n+1}] + ny(1-y)^{n-1}[x^n - x^{n+1}] + (x^{n+1} - x^n)[1 - (1-y)^n] \\
&= \sum_{b=2}^n \binom{n}{b} y^b (1-y)^{n-b} [x^{n-b+1} - x^n] \\
&= \sum_{b=2}^n \binom{n}{b} y^b (1-y)^{n-b} [f_{n-b+1}(x) - f_n(x)] \\
&= \mathcal{A}^{\Lambda_{\delta_y}} f_n(x)
\end{aligned}$$

con $\mathcal{A}^{\Lambda_{\delta_y}}$ el generador del coalescente de Eldon-Wakely del ejemplo 25.

1.13. Beta coalescentes.

En esta sección consideramos la familia de coalescentes de un parámetro, llamada Beta-coalescentes. Sea $0 < \alpha < 2$. Si consideramos que $\Lambda(dx) = \text{Beta}(2 - \alpha, \alpha)$, entonces

$$\Lambda(dx) = \frac{1}{\Gamma(2 - \alpha)\Gamma(\alpha)} x^{1-\alpha} (1-x)^{\alpha-1} dx,$$

el coalescente asociado es conocido como Beta-coalescente de parámetro α . El proceso de conteo de bloques $(N_t, t \leq 0)$ del Beta $(2 - \alpha, \alpha)$ -coalescente tiene tasas de transición del estado $n \in \{2, 3, \dots\}$ al estado $k \in \{1, \dots, n-1\}$, dadas por

$$\begin{aligned}
g_{n,k} &= \binom{n}{k-1} \int_0^1 x^{n-k-1} (1-x)^{k-1} \Lambda(dx) \\
&= \binom{n}{k-1} \frac{1}{\Gamma(2 - \alpha)\Gamma(\alpha)} \int_0^1 x^{n-k-\alpha} (1-x)^{k+\alpha-2} dx \\
&= \binom{n}{k-1} \frac{\Gamma(n-k-\alpha+1)\Gamma(k+\alpha-1)}{\Gamma(2 - \alpha)\Gamma(\alpha)\Gamma(n)} \\
&= n \frac{\Gamma(n-k-\alpha+1)\Gamma(k+\alpha-1)}{\Gamma(2 - \alpha)\Gamma(\alpha)\Gamma(k)\Gamma(n-k+2)}
\end{aligned}$$

y tasa total $g_n = \sum_{k=1}^{n-1} g_{n,k}$ y la cadena de saltos del proceso $(N_t, t \leq 0)$ es $p_{n,k} = \frac{g_{n,k}}{g_n}$. Un proceso de Galton-Watson es una cadena de Markov que indica el número total de

individuos que hay en una población en la n -ésima generación cuando los individuos en la generación anterior tienen una cantidad aleatoria de hijos, todos con la misma distribución y de manera independiente. Sea $F=(p_k, k \in \mathbb{Z}_+)$ la distribución de progenie. Denotamos por Z_n el tamaño de la población al tiempo $n \geq 0$. El proceso de Galton-Watson tiene probabilidades de transición

$$\mathbb{P}[Z_{n+1} = j | Z_n = m] = (F^{*m})(j),$$

es decir, la convolución de m v.a iid con distribución F evaluada en j . La genealogía del proceso análogo con espacio de estados continuo y tiempo continuo (CSBPs) puede ser descrita por la familia de Beta-coalescentes, Bertoin y Le Gall usaron CSBPs para estudiar las propiedades a pequeña escala de dichos coalescentes en [2]. Los CSBPs $(\hat{Z}_t, t > 0)$ están caracterizados por la propiedad de ramificación siguiente:

si $p_t(x, \cdot)$ es la probabilidad de transición de Z cuando $Z_0 = x$, entonces $\forall x, y \in \mathbb{R}^+$

$$p_t(x + y, \cdot) = p_t(x, \cdot) * p_t(y, \cdot),$$

es decir, la distribución del proceso comenzando en $x+y$ es la misma que la suma (independiente) del proceso comenzando en x y del proceso comenzando en y . En [10] se muestra que un CSBP está caracterizado por una función $\psi : [0, \infty) \rightarrow \mathbb{R}$ llamada mecanismo de ramificación. La función ψ es el exponente de Laplace de un proceso de Lévy con saltos no negativos, es decir, existe una medida ν en $(0, \infty)$ tal que $\int_0^\infty (1 \wedge x^2) \nu(dx) < \infty$, $a \in \mathbb{R}$, $b \geq 0$ y $q \geq 0$ tales que

$$\psi(q) = aq + bq^2 + \int_0^\infty (e^{-qx} - 1 + qx1_{x < 1}) \nu(dx).$$

Consideramos Λ una medida finita en $[0, 1)$ y $(\hat{Z}_t, t > 0)$ un ψ -CSBP con mecanismo de ramificación

$$\psi(q) = \int_0^\infty (e^{-qx} - 1 + qx) x^{-2} \Lambda(dx),$$

El siguiente resultado puede ser encontrado en [1] en el capítulo 4 sección 4.3.1.

Teorema 31. *El Λ -coalescente baja de infinito (comes down from infinity) si y sólo si Z se extingue en un tiempo finito, es decir,*

$$\int_1^\infty \frac{dq}{\psi(q)} < \infty.$$

Si es así, se define $v(t)$ tal que $\int_{v(t)}^\infty \frac{dq}{\psi(q)} = t$. Entonces cuando $t \rightarrow 0$

$$\frac{N_t}{v(t)} \rightarrow 1$$

casi seguramente en L^p para cada $p \geq 1$.

Capítulo 2

El proceso de frecuencia simple asimétrico

En una población en la cual ocurren reproducciones no simultáneas, en donde el individuo que se reproduce produce múltiples sucesores, consideramos dos tipos de individuo cuyo número de hijos depende del tipo. Los individuos se etiquetan como tipo (-) y tipo (+), el número de hijos que tienen el individuo que se reproduce es decidido por una sucesión de $N - 1$ variables aleatoria uniformes en $[0, 1]$ asignadas a cada uno de los individuos de la población, exceptuando al individuo que se reproduce. Los individuos de tipo (+) se reproducen y sus hijos reemplazan a aquellos individuos de la población cuya variable aleatoria uniforme correspondiente resulta estar en el intervalo $[0, y + z]$, mientras que los individuos de tipo (-) se reproducen y sus hijos reemplazan a aquellos individuos de la población cuya variable aleatoria uniforme correspondiente pertenece al intervalo $[0, y]$. Las reproducciones de los individuos suceden según un proceso Poisson de parámetro 1. Para la construcción del proceso consideramos un proceso Poisson puntual. En cada evento de reproducción se tiene una muestra de $N - 1$ variables aleatorias uniformes en $[0, 1]$, i.e. u_1, u_2, \dots, u_{N-1} , quienes determinan la participación de los individuos en la reproducción. Si el individuo que se reproduce es tipo (-) la población aumenta una cantidad $\sum_{i=1}^{N-1} \mathbb{1}_{\{0 < u_i < y\} \cap \{\text{el individuo } i \text{ es tipo (+)}\}}$ de individuos de tipo (-) y si el individuo que se reproduce es de tipo (+) la población aumenta una cantidad $\sum_{i=1}^{N-1} \mathbb{1}_{\{0 < u_i < y+z\} \cap \{\text{el individuo } i \text{ es tipo (-)}\}}$ de individuos de tipo (+).

2.1. Construcción

Fijamos $N \in \mathbb{N}$ y $x, y, z \in [0, 1]$ con $y + z \leq 1$. Consideramos una población con $N \in \mathbb{N}$ individuos, que al tiempo inicial t_0 está formada por una fracción de x individuos de tipo (-) y $(1 - x)$ individuos de tipo (+), un individuo elegido uniformemente al azar se reproduce y produce

- 1) $\sum_{i=1}^{N-1} \mathbb{1}_{\{0 < u_i \leq y\}}$ número de hijos, si el individuo que se reproduce es tipo (-) y
- 2) $\sum_{i=1}^{N-1} \mathbb{1}_{\{0 < u_i \leq y+z\}}$ número de hijos, si el individuo que se reproduce es tipo (+)

2.2. Definición

Definición 32 (El modelo simple asimétrico). Sean $N \in \mathbb{N}$ y $y, z \in [0, 1]$ fijos con $y + z \leq 1$. Sea $\{(U_i^j)_{j \in [N-1]}\}_{i \in \mathbb{N}}$ una sucesión de variables aleatorias iid con distribución uniforme en $[0, 1]$. Sea $(W_t)_{t \leq 0}$ un proceso Poisson puntual en $\mathbb{R}^+ \times [0, 1]^{N-1}$ de intensidad $dt \times \prod_{j=1}^{N-1} dU_i^j$. Consideramos una población con $N \in \mathbb{N}$ individuos, para $W_{t_i}^{(N)} = (t_i, u_i^1, \dots, u_i^{N-1})$ y con una población conformada por Nx individuos de tipo (-) y $N(1-x)$ individuos de tipo (+) al tiempo t_i^- . Al tiempo t_i ocurre la i -ésima reproducción, en la cual un individuo elegido uniformemente al azar, denotado por J_i se reproduce. Si J_i es tipo (-) produce $\sum_{k=1}^{N-1} \mathbb{1}_{\{0 < u_i^k < y\}}$ número de hijos del mismo tipo los cuales reemplazan a individuos de la población. Si J_i es tipo (+) produce $\sum_{k=1}^{N-1} \mathbb{1}_{\{0 < u_i^k < y+z\}}$ número de hijos del mismo tipo los cuales reemplazan a individuos de la población.

El modelo de frecuencia simple asimétrico

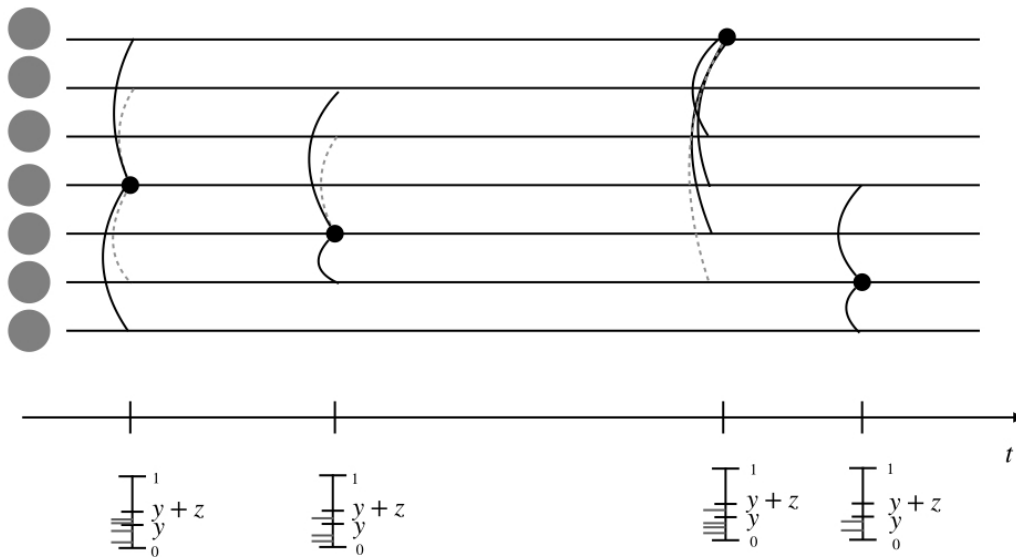


Figura 2.1: Representación gráfica del modelo asimétrico simple. En el primer evento en donde se toma una muestra de 6 variables aleatorias uniformes en $[0, 1]$, los individuos cuya v.a correspondiente pertenece al intervalo $[0, y]$ se unen al individuo que se reproduce con una línea negra mientras que los individuos cuya v.a correspondiente pertenece al intervalo $(y, y+z]$ se unen con líneas grises punteadas, estas líneas solo serán utilizadas si el individuo que se reproduce es tipo (+). Los individuos cuya v.a uniforme correspondiente pertenece al intervalo $(y+z, 1]$ no tienen ningún cambio y no se representan en la figura.

La selección de los individuos de tipo (+) les permite tener mayor probabilidad de tener hijos que los individuos de tipo (-).

Definición 33 (El proceso de frecuencia asimétrico). Sean $N \in \mathbb{N}$, $y, z \in [0, 1]$ fijos con $y + z \leq 1$. Definimos el proceso de frecuencia del modelo simple asimétrico como

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

denotado como $(X_t^{(N)}, t \geq 0)$.

Las transiciones del proceso de frecuencia simple asimétrico denotado por $(X_t^{(N)}, t \geq 0)$ cuando $X_{t^-}^{(N)} = i/N$ son

$$\frac{i}{N} \text{ a } \begin{cases} \frac{i}{N} + \frac{N-i}{N} \mathbb{1}_{U_i < y} & \text{a tasa } \frac{i}{N}, \\ \frac{i}{N} - \frac{1}{N} \sum_{i=1}^{N-i} \mathbb{1}_{U_j < y+z} & \text{a tasa } \frac{(N-i)}{N}, \end{cases}$$

Las transiciones del proceso límite del proceso de frecuencia simple asimétrico denotado por $(X_t, t \geq 0)$ cuando $X_{t^-} = x$ son

$$x \text{ a } \begin{cases} x + (1-x)y & \text{a tasa } x, \\ x(1-y-z) & \text{a tasa } (1-x). \end{cases} \quad (2.1)$$

Teorema 34. El proceso de frecuencia simple asimétrico $(X_t^{(N)}, t \geq 0)$ converge en el sentido de los generadores al proceso $(X_t, t \geq 0)$ con tasas de transición dadas por 2.1.

Demostración. Sea $f \in C^2([0, 1])$ y $x \in [0, 1]$, entonces el generador del proceso de fre-

cuencia simple asimétrico aplicado a f cumple que

$$\begin{aligned}
\mathcal{A}^{(N)} f(x) &= \mathbb{E}[f(X_t^{(N)}) - f(x) | X_{t^-}^{(N)} = x] \\
&= x \mathbb{E} \left[f\left(x + \frac{1}{N} \sum_{i=1}^{N(1-x)} \mathbb{1}_{U_i < y}\right) - f(x) \right] + (1-x) \mathbb{E} \left[f\left(x - \frac{1}{N} \sum_{i=1}^{Nx} \mathbb{1}_{U_i < y+z}\right) - f(x) \right] \\
&= x \mathbb{E} \left[f(x) + f'(x) \frac{1}{N} \sum_{i=1}^{N(1-x)} \mathbb{1}_{U_i < y} + f''(x) \frac{1}{2N^2} \left(\sum_{i=1}^{N(1-x)} \mathbb{1}_{U_i < y} \right)^2 + o\left(\frac{1}{N^2}\right) - f(x) \right] \\
&\quad + (1-x) \mathbb{E} \left[f(x) - f'(x) \frac{1}{N} \sum_{i=1}^{Nx} \mathbb{1}_{U_i < y+z} + f''(x) \frac{1}{2N^2} \left(\sum_{i=1}^{Nx} \mathbb{1}_{U_i < y+z} \right)^2 + o\left(\frac{1}{N^2}\right) - f(x) \right] \\
&= x \left[f(x) + f'(x) \frac{1}{N} (N(1-x)) \mathbb{P}[U_1 < y] + f''(x) \frac{1}{2N^2} \mathbb{E} \left[\left(\sum_{i=1}^{N(1-x)} \mathbb{1}_{U_i < y} \right)^2 \right] - f(x) \right] \\
&\quad + (1-x) \left[f(x) - f'(x) \frac{1}{N} (Nx) \mathbb{P}[U_1 < y+z] + f''(x) \frac{1}{2N^2} \mathbb{E} \left[\left(\sum_{i=1}^{Nx} \mathbb{1}_{U_i < y+z} \right)^2 \right] - f(x) \right] \\
&\quad + 2o\left(\frac{1}{N^2}\right) \\
&= x \left[f(x) + f'(x)(1-x)y + f''(x) \frac{1}{2N^2} \mathbb{E} \left[\left(\sum_{i=1}^{N(1-x)} \mathbb{1}_{U_i < y} \right)^2 \right] + o\left(\frac{1}{N^2}\right) - f(x) \right] \\
&\quad + (1-x) \left[f(x) - f'(x)x(y+z) + f''(x) \frac{1}{2N^2} \mathbb{E} \left[\left(\sum_{i=1}^{Nx} \mathbb{1}_{U_i < y+z} \right)^2 \right] + o\left(\frac{1}{N^2}\right) - f(x) \right] \\
&= x \left[f(x) + f'(x)(1-x)y + o\left(\frac{1}{N}\right) - f(x) \right] + (1-x) \left[f(x) - f'(x)x(y+z) + o\left(\frac{1}{N}\right) - f(x) \right] \\
&\xrightarrow{N \rightarrow \infty} x[f(x + (1-x)y) - f(x)] + (1-x)[f(x - x(y+z)) - f(x)].
\end{aligned}$$

en donde los términos $o\left(\frac{1}{N}\right)$ dependen de x , pero convergen uniformemente a 0 en x . \square

2.3. El proceso de ancestría del modelo simple asimétrico

La construcción de la genealogía de una muestra tomada de una población en donde se considera selección fue presentado por Neuhauser y Krone (1995) en [9], conocida como la Gráfica de Ancestría con Selección (The Ancestral Selection Graph ASG), en la cual se sigue la ancestría de una muestra, tomada de una población con dos tipos de individuos, A_1 y A_2 , la ventaja del individuo tipo 2 se debe a la tasa de reproducción, la cual es mayor que la tasa de reproducción de individuos de tipo 1 ($\lambda_2 > \lambda_1$). El proceso genealógico de dicha población resulta ser un proceso de coalescencia y ramificación. La estrategia que utilizamos para seguir las líneas ancestrales de una muestra en una población descrita por el *modelo simple asimétrico* sigue la misma idea que en [9].

El procedimiento que se utiliza es: al tiempo t_0 (en el presente) se toma una muestra de tamaño n , a tasa 1 sucede un evento de coalescencia o un evento de ramificación. En cada evento tendremos una sucesión de tamaño N de variables aleatorias uniformes en $[0, 1]$ independientes, cada una de estas variables correspondientes a cada individuo de la población, incluyendo a los individuos de la muestra de tamaño $n < M$, y se elige a un individuo al azar, el cual correspondería al individuo que se reprodujo en dicho evento. Seguiremos el linaje de los individuos en la muestra, la línea ancestral de los individuos

cuya variable aleatoria U_i se encuentra en el intervalo $(y, y + z]$ siempre se seguirá hacia atrás en el tiempo, se une con la línea del individuo que se eligió para reproducirse y también se sigue la línea de ancestría de este individuo elegido para reproducirse. La línea ancestral de los individuos en la muestra cuya variable aleatoria uniforme pertenece al intervalo $[0, y]$ coalesce (y se deja de seguir) con la línea ancestral del individuo que se eligió para reproducirse y se sigue la línea de ancestría de este último individuo. Las líneas ancestrales de los individuos cuya variable aleatoria pertenece al intervalo $(y + z, 1]$ permanecen sin ningún cambio.

Un evento que llamamos *coalescencia* sucede cuando: las líneas ancestrales de los individuos en la muestra se unen a la línea ancestral de su padre, esto es cuando hay uno o más individuos con variable aleatoria uniforme en el intervalo $[0, y]$ en la muestra. Ocurre un evento de *ramificación* cuando: las variables aleatorias correspondientes a los individuos de la muestra no pertenecen al intervalo $[0, y]$, hay al menos una variable aleatoria uniforme que pertenece al intervalo $(y, y + z]$ y el individuo que se eligió para reproducirse no pertenece a la muestra. Si conocemos la configuración de la población, es decir, el número de individuos de tipo $(-)$ y el número de individuos de tipo $(+)$ (o equivalentemente la frecuencia de individuos de tipo $(+)$ en la población), para algún tiempo t en el pasado podemos conocer la configuración de la muestra en el presente. Después de el primer evento, ya sea de coalescencia o de ramificación, nos referiremos a las líneas ancestrales de la muestra como ancestros potenciales, es decir, las líneas amarillas en la figura 2.2.

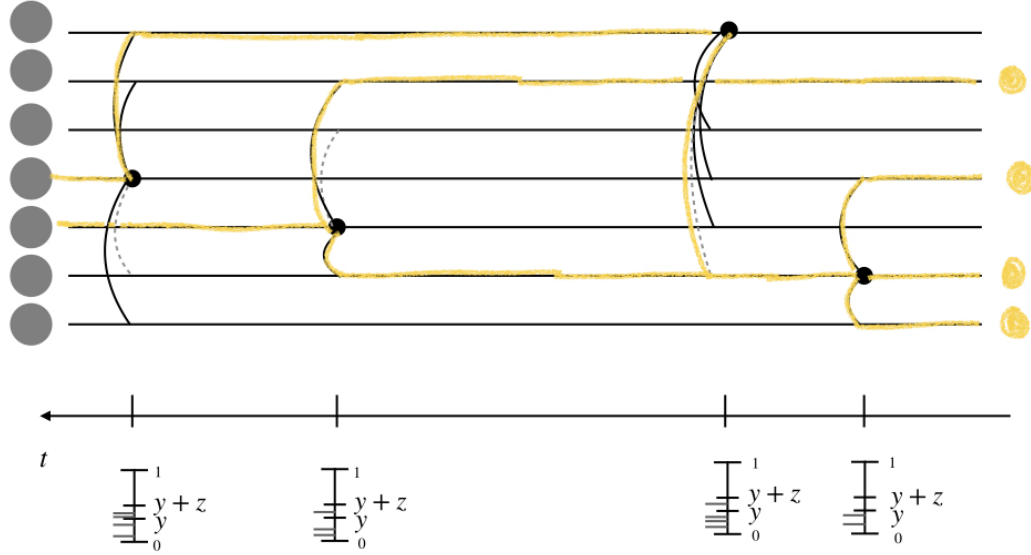


Figura 2.2: Representación gráfica de la genealogía de una muestra de 4 individuos (círculos amarillos) para una población que evoluciona según el modelo asimétrico simple. Las líneas amarillas representan los ancestros potenciales.

Definición 35 (El proceso de ancestría simple asimétrico). Sean $N, n \in \mathbb{N}$ y $y, z \in [0, 1]$ fijos con $y + z \leq 1$. Sean $\{(U_i^j)_{j \in [N]}\}_{i \in \mathbb{N}}$ una sucesión de variables aleatorias iid con distribución uniforme en $[0, 1]$ y $(B_i)_{i \in [n]}$ una sucesión de variables aleatorias con $B_i = \sum_{j=1}^n \mathbb{1}_{\{U_i^j < y\}}$ (distribución binomial $(n, \mathbb{P}[U_i^j < y])$). Sea $(Z_t^{(N)})_{t \leq 0}$ el proceso de ancestría (conteo de bloques o líneas) y supongamos que $Z_0^{(N)} = n$. Si $Z_t^{(N)} = j$, entonces a tasa 1 se elige un individuo uniformemente al azar, denotado por J_i , para ser el ancestro que se reproduce. En el i -ésimo evento que ocurre al tiempo t_i se tiene que:

- 1) (evento de coalescencia) el i -ésimo evento es de coalescencia si $B_i > 0$, entonces las líneas ancestrales de los individuos de la muestra tales que $U_i^j < y$ coalescen con la línea ancestral de J_i ,
- 2) (evento de ramificación) el i -ésimo evento es de ramificación si $B_i = 0$, el individuo J_i no pertenece a los ancestros potenciales y $U_i^k < y + z$ para al menos un individuo k en los ancestros potenciales, entonces el número de líneas ancestrales aumenta en 1, siguiendo también la línea del individuo J_i -ésimo.

Las transiciones del proceso de ancestría simple asimétrico cuando cuando $Z_{t^-}^{(N)} = n$

son

$$n \text{ a } \begin{cases} n-i+1 & \text{a tasa} & (N-n)\binom{n}{i}y^i(1-y)^{n-i}, \\ n-i & \text{a tasa} & n\binom{n}{i}y^i(1-y)^{n-i-1}, \\ n+1 & \text{a tasa} & (N-n)[(1-y)^n - (1-y-z)^n] \end{cases}$$

Las transiciones del proceso límite del proceso de ancestría simple asimétrico $(Z_t, t \geq 0)$ cuando $Z_{t-} = n$ son

$$n \text{ a } \begin{cases} n-i+1 & \text{a tasa} & [\binom{n}{i}y^i(1-y)^{n-i}], \\ n+1 & \text{a tasa} & [(1-y)^n - (1-y-z)^n] \end{cases}$$

2.4. Dualidad de los procesos asimétricos

Proposición 36. *El proceso límite del proceso de frecuencia simple asimétrico $(X_t, t \geq 0)$ y el proceso límite del proceso de ancestría simple asimétrico $(Z_t, t \geq 0)$ son momento duales.*

Demostración. Sea $f_n : [0, 1] \times \mathbb{N} \rightarrow [0, 1]$ definida como $f_n(x, n) = x^n$. Aplicando el generador del proceso $(X_t, t \geq 0)$ a la función f_n se tiene que

$$\begin{aligned} \mathcal{G}^X f_n(x) &= x[f_n(x(1-y) + y) - f_n(x)] + (1-x)[f_n(x(1-y-z)) - f_n(x)] \\ &= x[(x(1-y) + y)^n - x^n] + (1-x)[(x(1-y-z))^n - x^n] \\ &= x[(x(1-y) + y)^n - x^n] + (1-x)[(x(1-y-z))^n - x^n] \\ &\quad + (1-x)[(x(1-y))^n - x^n] \\ &\quad - (1-x)[(x(1-y))^n - x^n] \\ &= \sum_{i=2}^n \binom{n}{i} y^i (1-y)^{n-i} [x^{n-i+1} - x^n] + (1-x)[(x(1-y-z))^n - x^n] - (1-x)[(x(1-y))^n - x^n] \\ &= \sum_{i=2}^n \binom{n}{i} y^i (1-y)^{n-i} [x^{n-i+1} - x^n] + [(1-y)^n - (1-y-z)^n][x^{n+1} - x^n] \\ &= \sum_{i=2}^n \binom{n}{i} y^i (1-y)^{n-i} [f_{n-i+1}(x) - f_n(x)] + [(1-y)^n - (1-y-z)^n][f_{n+1}(x) - f_n(x)] \\ &= \mathcal{A}^{\Delta_y} f_n(x) + [(1-y)^n - (1-y-z)^n][f_{n+1}(x) - f_n(x)] \\ &= \mathcal{G}^Z f_n(x), \end{aligned}$$

con \mathcal{A}^{Δ_y} el generador del proceso de frecuencia de Eldon-Wakeley 29, \mathcal{G}^X el generador del proceso límite del proceso de frecuencia simple asimétrico $(X_t, t \geq 0)$ y \mathcal{G}^Z el generador del proceso límite del proceso de ancestría simple asimétrico $(Z_t, t \geq 0)$. \square

Capítulo 3

El modelo simple asimétrico con mutaciones

Agregamos mutaciones al *proceso simple asimétrico* de la definición 32. Los individuos de la población tienen eventos de mutación que ocurren según un proceso Poisson de parámetro $\lambda = 1$, cambiando su tipo. Con probabilidad $u_{(+)}$ un individuo de tipo $(-)$ cambian a tipo $(+)$ y con probabilidad $u_{(-)}$ un individuo de tipo $(+)$ cambia a tipo $(-)$.

3.1. Construcción

Fijamos $N \in \mathbb{N}$ y $u_{(+)}, u_{(-)}, y, z \in [0, 1]$ con $y + z \leq 1$. Consideramos una población con $N \in \mathbb{N}$ individuos, formada por dos tipos de individuos con distintas posibilidades de participar en cada evento de reproducción, un individuo elegido uniformemente al azar decide reproducirse y produce

- 1) $\sum_{i=1}^{N-1} \mathbb{1}_{\{0 < u_i \leq y\}}$ número de hijos, si el individuo que se reproduce es tipo $(-)$ y
- 2) $\sum_{i=1}^{N-1} \mathbb{1}_{\{0 < u_i \leq y+z\}}$ número de hijos, si el individuo que se reproduce es tipo $(+)$
- 3) Además los individuos de la población tienen eventos de mutación que ocurren según un proceso Poisson de parámetro $\lambda = 1$. Un individuo que tiene una mutación y es de tipo $(+)$ cambia a tipo $(-)$ con probabilidad $u_{(-)}$ y si el individuo que tiene una mutación es de tipo $(-)$, con probabilidad $u_{(+)}$ cambia a tipo $(+)$.

3.2. Definición

Definición 37 (El modelo simple asimétrico con mutaciones). Sea $N \in \mathbb{N}$, y sean $u_{(+)}, u_{(-)}, y, z \in [0, 1]$ fijos, con $y + z \leq 1$. Sean $\{(U_i^j)_{j \in [N-1]}\}_{i \in \mathbb{N}}$ y $\{V_i\}_{i \in \mathbb{N}}$ dos sucesiones de variables aleatorias iid con distribución uniforme en $[0, 1]$. Sea $(W_t)_{t \leq 0}$ un proceso Poisson puntual en $\mathbb{R}^+ \times [0, 1]^{N-1}$ de intensidad $dt \times \prod_{j=1}^{N-1} dU_i^j$ un procesos Poisson puntual en $\mathbb{R}^+ \times [0, 1]$ de intensidad $dt \times V_i$. Consideramos una población con $N \in \mathbb{N}$ individuos, a tasa dt ocurre un evento de reproducción. Para un punto $w_{t_i}^{(N)} = (t_i, u_i^1, \dots, u_i^{N-1})$ de $(W_t)_{t \leq 0}$, al tiempo t_i ocurre la reproducción i -ésima en la cual un individuo elegido uniformemente al azar y denotado por J_i se reproduce. Si J_i es tipo $(-)$ produce $\sum_{k=1}^{N-1} \mathbb{1}_{\{0 < u_i^k < y\}}$ número de hijos del mismo tipo los cuales reemplazan a individuos de la población. Si J_i es tipo

(+) produce $\sum_{k=1}^{N-1} \mathbb{1}_{\{0 < u_{\frac{k}{N}} < y+z\}}$ número de hijos del mismo tipo los cuales reemplazan a individuos de la población.

Además, los individuos tienen mutaciones que los cambian de tipo. Para un punto $\tilde{w}_{s_j} = (s_j, v_j)$ de $(\tilde{W}_t)_{t \leq 0}$ al tiempo s_j ocurre la j -ésima mutación y un individuo elegido uniformemente al azar tiene una mutación y cambia de tipo según la siguiente regla. Si el individuo es tipo (+) la mutación ocurre si $v_{s_j} < u_{(-)}$. Si el individuo es tipo (-) la mutación ocurre si $v_{s_j} < u_{(+)}$.

En la definición anterior, las mutaciones que ocurren a tasa $u_{(-)}$ le ocurren a individuos de tipo (+), y las mutaciones que ocurren a tasa $u_{(+)}$ son mutaciones que le ocurren a individuos de tipo (-), y los cambia a tipo (+).

Decimos que las mutaciones ocurren “a tasa” $u_{(+)}$ (equivalentemente “a tasa” $u_{(-)}$) pensando en el siguiente hecho. Consideramos un proceso Poisson de parámetro $\lambda = 1$ y una variable aleatoria U uniforme en $[0, 1]$, las llegadas del proceso Poisson de parámetro $\lambda = 1$ serán ignoradas si $U \geq u_{(+)}$, por lo que las llegadas ocurrirán cuando $U < u_{(+)}$ de esta manera obtenemos un proceso Poisson de parámetro $\lambda \cdot u_{(+)} = 1 \cdot u_{(+)} = u_{(+)}$.

Definición 38 (El proceso de frecuencia simple asimétrico con mutaciones). Sean $N \in \mathbb{N}$, y $y, z \in [0, 1]$ fijos con $y + z \leq 1$. Definimos el proceso de frecuencia del modelo simple asimétrico con mutaciones como el proceso que sigue la frecuencia de individuos de tipo (-) en una población que se reproduce según la definición 37, es decir,

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

denotado como $(X_t^{(N)}, t \geq 0)$

Las tasas de transición del proceso de frecuencia simple asimétrico con mutaciones $(X_t^{(N)}, t \geq 0)$ son

$$\frac{i}{N} \text{ a } \begin{cases} \frac{i}{N} + \frac{N-i}{N} \mathbb{1}_{U_i < y} & \text{a tasa } \frac{i}{N}, \\ \frac{i}{N} - \frac{1}{N} \sum_{i=1}^{Ni} \mathbb{1}_{U_j < y+z} & \text{a tasa } \frac{N-i}{N}, \\ \frac{i}{N} + \frac{1}{N} & \text{a tasa } N \frac{(N-i)}{N} u_{(-)}, \\ \frac{i}{N} - \frac{1}{N} & \text{a tasa } N \frac{i}{N} u_{(+)}, \end{cases}$$

Las transiciones del proceso de límite del proceso de frecuencia simple asimétrico con mutaciones $(X_t, t \geq 0)$ son

$$x \text{ a } \begin{cases} x + y(1-x) & \text{a tasa } x, \\ x - (y+z)x & \text{a tasa } (1-x), \\ x + \frac{1}{N} & \text{a tasa } (1-x)u_{(-)}, \\ x - \frac{1}{N} & \text{a tasa } xu_{(+)}, \end{cases}$$

por lo que su generador infinitesimal es

$$\begin{aligned} \mathcal{Q}^{(n)} f(x) = & x[f(x(1-y) + y) - f(x)] + (1-x)[f(x(1-y-z)) - f(x)] \\ & + (1-x)u_{(-)} \left[f\left(x + \frac{1}{N}\right) - f(x) \right] + xu_{(+)} \left[f\left(x - \frac{1}{N}\right) - f(x) \right]. \end{aligned}$$

3.3. El proceso de ancestría asimétrico con mutaciones

El proceso que sigue la genealogía de una muestra de una población que se reproduce según la definición 37, al igual que el proceso genealógico de la sección 2.3, es un proceso de ramificación y coalescencia, en donde, ocurren eventos de mutación que se distribuyen según un proceso Poisson de parámetro 1, un individuo elegido uniformemente al azar entre la población cambian de tipo, lo cual se notará al decidir los tipos de los individuos para algún tiempo $t > 0$, pero no afecta la estructura de coalescencia y ramificación del proceso.

Definición 39 (El proceso de ancestría simple asimétrico con mutaciones). Sean $N, n \in \mathbb{N}$, $y, z, u_{(+)} y u_{(-)} \in [0, 1]$ fijos con $y + z \leq 1$. Sean $\{(U_i^j)_{j \in [N]}\}_{i \in \mathbb{N}}$ una sucesión de variables aleatorias iid con distribución uniforme en $[0, 1]$ y $(B_i)_{i \in [n]}$ una sucesión de variables aleatorias con $B_i = \sum_{j=1}^n \mathbb{1}_{\{U_i^j < y\}}$ (distribución binomial $(n, \mathbb{P}[U_i^j < y])$). Sea $(Z_t^{(N)})_{t \leq 0}$ el proceso de ancestría (conteo de bloques o líneas) y supongamos que $Z_0^{(N)} = n$, entonces a tasa 1 se elige un individuo uniformemente al azar entre los N individuos de la población, denotado por J_i , para ser el ancestro que se reproduce. En el i -ésimo evento que ocurre al tiempo t_i se tiene que:

- 1) (evento de coalescencia) el i -ésimo evento es de coalescencia si $B_i > 0$, entonces las líneas ancestrales de los individuos de la muestra tales que $U_i^j < y$ coalescen con la línea ancestral de J_i ,
- 2) (evento de ramificación) el i -ésimo evento es de ramificación si $B_i = 0$, el individuo J_i no pertenece a los ancestros potenciales y $U_i^k < y + z$ para al menos un individuo k en los ancestros potenciales, entonces el número de líneas ancestrales aumenta en 1, siguiendo también la línea del individuo J_i -ésimo.

Además para $t > 0$ fijo, seguimos las reglas 1) y 2) del proceso obteniendo una estructura de ancestría. En cada línea de ancestría de la estructura obtenida desde el tiempo 0 hasta el tiempo t , colocamos puntos según un proceso Poisson de parámetro $\lambda = 1$ (representados gráficamente por triángulos negros), los cuales indicarán las mutaciones.

En la definición anterior, las mutaciones suceden sobre las líneas ancestrales y solo al saber los tipos de los individuos podemos determinar si la mutación ocurre o no. Si una línea ancestral tiene un triángulo negro, entonces podemos determinar si dicha mutación ocurre o no cuando sabemos el tipo del individuo. Si el individuo que encuentra una marca es tipo $(-)$, entonces cambia a tipo $(+)$ con probabilidad $u_{(+)}$, si el individuo que encuentra una marca es tipo $(+)$, entonces cambia a tipo $(-)$ con probabilidad $u_{(-)}$. Es por eso que la dinámica del proceso de ancestría simple asimétrico con mutaciones y del proceso de ancestría asimétrico es esencialmente igual cuando no conocemos los tipos.

El modelo de frecuencia simple asimétrico con mutación

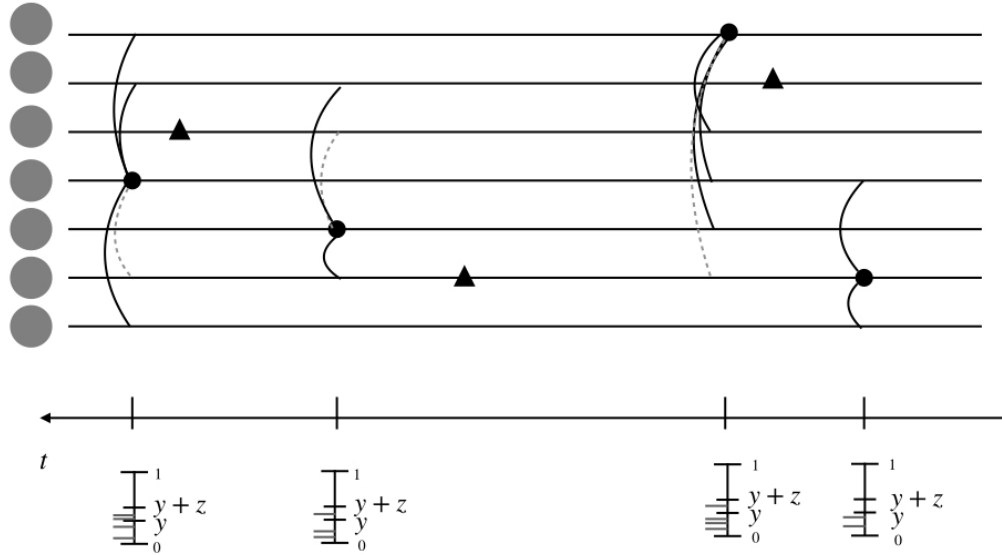


Figura 3.1: Representación gráfica del proceso de frecuencia simple asimétrico con mutaciones.

Las transiciones del proceso de ancestría asimétrico con mutaciones cuando $Z_0 = n$ son

$$n \text{ a } \begin{cases} n - i + 1 & \text{a tasa} & (N - n) \left[\binom{n}{i} y^i (1 - y)^{n-i} \right], \\ n - i & \text{a tasa} & n \left[\binom{n}{i} y^i (1 - y)^{n-i-1} \right], \\ n + 1 & \text{a tasa} & (N - n) \left[(1 - y)^n - (1 - y - z)^n \right] \end{cases}$$

Las transiciones del proceso de ancestría asimétrico con mutaciones de una muestra de n individuos de tipo (-) son:

$$n \text{ a } \begin{cases} n - i + 1 & \text{a tasa} & (N - n) \left[\binom{n}{i} y^i (1 - y)^{n-i} \right], \\ n - i & \text{a tasa} & n \left[\binom{n}{i} y^i (1 - y)^{n-i-1} \right], \\ n - 1 & \text{a tasa} & nu_{(-)}, \end{cases}$$

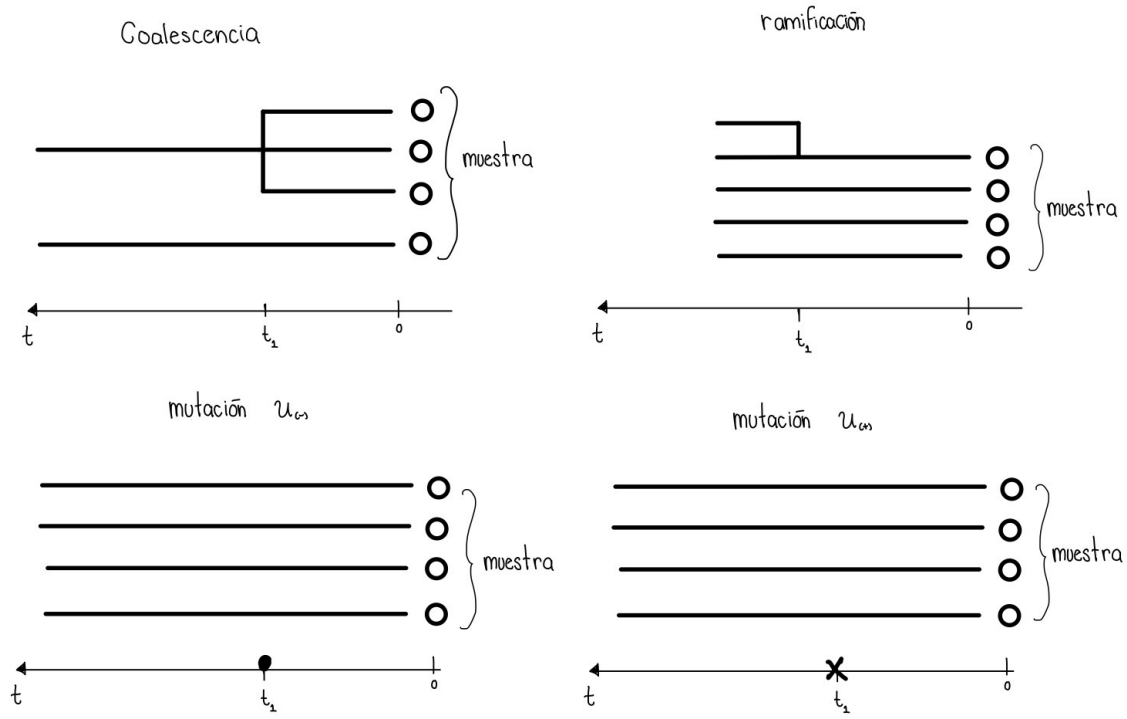


Figura 3.2: Representación gráfica de los eventos de coalescencia, ramificación y mutaciones.

Capítulo 4

Recursiones

En este capítulo presentamos una recursión para el modelo asimétrico con mutaciones. La recursión es calculada de manera similar que en [7] y en [9], en donde al tiempo inicial t_0 consideramos una muestra de una población que se reproduce según el modelo 32 de tamaño $n = n_1 + n_2$ con n_1 individuos de tipo (+) y n_2 individuos de tipo (-), recorriendo el linaje de dicha muestra, avanzamos hacia atrás en el tiempo hasta el primer evento, ya sea de ramificación, coalescencia o mutación, que ocurre al tiempo t_1 , al tiempo t_1^- tendremos una configuración de la población distinta a la que tenemos en presente, ya que en el tiempo t_1 ocurre un evento aleatorio (de ramificación, coalescencia o mutación). El objetivo de calcular estas recursiones es tener la probabilidad de tener una muestra en la que todos los individuos sean de tipo (-), por lo que calcularemos

$$p(0, n) := \mathbb{P}[\{\text{todos los individuos de una muestra con reemplazo de tamaño } n \text{ son de tipo } (-)\}].$$

4.1. Recursión para el modelo asimétrico con mutaciones

La tasa de transición a la cual la muestra de tamaño n tiene una mutación, coalescencia de tamaño k o ramificación son $nu_{(+)} + nu_{(-)}$, $\sum_{k=2}^n \binom{n}{k} y^k (1-y)^{n-k}$ y $(1-y)^n - (1-y-z)^n$ respectivamente. Por lo que la tasa total es $\mu_n = nu_{(+)} + nu_{(-)} + (1-y)^n - (1-y-z)^n + \sum_{k=2}^n \binom{n}{k} y^k (1-y)^{n-k}$, y así la probabilidad de que el primer evento que ocurre antes de tomar la muestra (n_1, n_2) al tiempo t_1 sea una coalescencia es $\frac{\sum_{k=2}^n \binom{n}{k} y^k (1-y)^{n-k}}{\mu_n}$, una ramificación $\frac{(1-y)^n - (1-y-z)^n}{\mu_n}$ o finalmente es una mutación de tipo (+) a tipo (-) con probabilidad $\frac{nu_{(-)}}{\mu_n}$ y es una mutación de tipo (-) a tipo (+) con probabilidad $\frac{nu_{(+)}}{\mu_n}$. La recursión que obtenemos para una muestra de tamaño n con una configuración $(0, n)$ es

$$p(0, n) = \frac{1}{\mu_n} \sum_{k=2}^n \binom{n}{k} y^k (1-y)^{n-k} p(0, n-k+1) + \frac{1}{\mu_n} [(1-y)^n - (1-y-z)^n] p(0, n+1) + \frac{1}{\mu_n} nu_{(-)} p(0, n-1) \tag{4.3}$$

como se vió en el capítulo anterior, son las transiciones del proceso de ancestría asimétrico con mutaciones de una muestra de n individuos,

$$n \text{ a } \begin{cases} n - k + 1 & \text{a tasa} & \binom{n}{k} y^k (1 - y)^{n-k}, \\ n + 1 & \text{a tasa} & (1 - y)^n - (1 - y - z)^n \\ n - 1 & \text{a tasa} & nu_{(-)} \\ \Delta & \text{a tasa} & nu_{(-)}. \end{cases}$$

el estado Δ es un estado absorbente, en el proceso entra y no vuelve a salir, esto sucede cuando a tasa $u_{(-)}$ todos los ancestros potenciales de la muestra (a lo más n ancestros potenciales) sufren una mutación.

Fijamos y, z, u y v los parámetros del modelo asimétrico con mutaciones 37 (en este caso escribimos $u := u_{(+)}$ y $v := u_{(-)}$ para escribirlos en R) y fijando el valor de $p(0, n) \in [0, 1]$, para alguna n , obtenemos un sistema de ecuaciones de $n-1$ incógnitas con $n-1$ ecuaciones que podemos resolver para obtener $p(0, 1), p(0, 2), \dots, p(0, n-1)$.

Ejemplo 40. *Fijamos $p(0, 5) = 1$ la probabilidad de obtener una muestra de tamaño 5 conformada únicamente por individuos de tipo $(-)$ en una población que se reproduce según el modelo asimétrico con mutaciones, utilizamos la recursión 4.3 y escribimos p_n en lugar de $p(0, n)$ para obtener el siguiente sistema de ecuaciones*

$$\begin{aligned} \mu_1 \cdot p_1 &= p_2[(1 - y) - (1 - y - z)] + p_0 \cdot v \\ \mu_2 \cdot p_2 &= p_3[(1 - y)^2 - (1 - y - z)^2] + p_1 \cdot 2v + \lambda_{(2,2)} p_1 \\ \mu_3 \cdot p_3 &= p_4[(1 - y)^3 - (1 - y - z)^3] + p_2 \cdot 3v + \lambda_{(3,2)} p_2 + \lambda_{(3,3)} p_1 \\ \mu_4 \cdot p_4 &= p_5[(1 - y)^4 - (1 - y - z)^4] + p_3 \cdot 4v + \lambda_{(4,2)} p_3 + \lambda_{(4,3)} p_2 + \lambda_{(4,4)} p_1 \\ p_5 &= 1 \end{aligned}$$

con $\lambda_{(n,k)} = \binom{n}{k} y^k (1 - y)^{n-k}$ y la siguiente representación matricial

$$\begin{pmatrix} -\mu_1 & (1 - y) - (1 - y - z) & 0 & 0 \\ \lambda_{(2,2)} + 2v & -\mu_2 & (1 - y)^2 - (1 - y - z)^2 & 0 \\ \lambda_{(3,3)} & \lambda_{(3,2)} + 3v & -\mu_3 & (1 - y)^3 - (1 - y - z)^3 \\ \lambda_{(4,4)} & \lambda_{(4,3)} & \lambda_{(4,3)} + 4v & -\mu_4 \end{pmatrix} \begin{pmatrix} -p_0 \cdot v \\ 0 \\ 0 \\ -p_5[(1 - y)^4 - (1 - y - z)^4] \end{pmatrix}$$

que resolvemos para este ejemplo con R y el comando `solve()` para valores $y = 0,5$, $z = 0,2$, $v = 0,1$ y $u = 0,1$, obtenemos

```
> E <- cbind(c(-mu_n(1), lambda_nk(2,2)+2*v, lambda_nk(3,3), lambda_nk(4,4)),
+           c((1-y)-(1-y-z), -mu_n(2), lambda_nk(3,2)+3*v, lambda_nk(4,3)),
+           c(0, (1-y)^2-(1-y-z)^2, -mu_n(3), lambda_nk(4,2)+4*v),
+           c(0, 0, (1-y)^3-(1-y-z)^3, -mu_n(4)))
> E
      [,1] [,2] [,3] [,4]
[1,] -0.4000 0.200 0.000 0.0000
```

```

[2,] 0.4500 -0.810 0.160 0.0000
[3,] 0.1250 0.675 -1.198 0.0980
[4,] 0.0625 0.250 0.775 -1.5419
> b <- c(-p_0*v, 0, 0, -p_5*((1-y)^4-(1-y-z)^4))
> b
[1] -0.1000 0.0000 0.0000 -0.0544
> solve(E, b)
[1] 0.3724209 0.2448419 0.1920780 0.1866186

```

Ahora mostramos algunos resultados de la recursión con distintos valores de los parámetros y , z , u y v .

```

> y=0.6
> z=0.1
> v=0.1
> u=0.1
> solve(E, b)
[1] 0.4395580 0.3186739 0.2620983 0.2319605

```

```

> y=0.1
> z=0.6
> v=0.1
> u=0.1
> solve(E, b)
[1] 0.3431170 0.2908227 0.3563543 0.5391867

```

```

> y=0.9
> z=0.1
> v=0.1
> u=0.1
> solve(E, b)
[1] 0.4616871 0.3850614 0.3470923 0.3178593

```

```

> y=0.9
> z=0.1
> v=0.1
> u=0.3
> solve(E, b)
[1] 0.22864957 0.14324786 0.11254601 0.09335087

```

```

> y=0.1
> z=0.9
> v=0.3
> u=0.1
> solve(E, b)
[1] 0.4707663 0.3466624 0.3387971 0.4680372

```

```

> y=0.1
> z=0.9
> v=0.4
> u=0.1
> solve(E, b)
[1] 0.5516542 0.4136844 0.3778588 0.4728188

```

Observación 1. Notamos que cuando incrementa el parámetro z que distingue a individuos de tipo (+), es menos probable obtener individuos tipo (-) en la muestra pues hay más cantidad de individuos de tipo (+) en la población. En el último ejemplo aun cuando $z = 0,9$ con el parámetro de mutación a tipo (-) de $u_{(-)} = 0,4$ la probabilidad de que la muestra esté formada por individuos tipo (-).

Ejemplo 41. Fijamos $p(0, 11) = 1$ la probabilidad de obtener una muestra de tamaño 11 conformada únicamente por individuos de tipo (-) utilizamos la recursión 4.3 para obtener el siguiente sistema de ecuaciones

$$\begin{aligned}
\mu_1 \cdot p_1 &= p_2[(1 - y) - (1 - y - z)] + p_0 \cdot v \\
\mu_2 \cdot p_2 &= p_3[(1 - y)^2 - (1 - y - z)^2] + p_1 \cdot 2v + \lambda_{(2,2)}p_1 \\
\mu_3 \cdot p_3 &= p_4[(1 - y)^3 - (1 - y - z)^3] + p_2 \cdot 3v + \lambda_{(3,2)}p_2 + \lambda_{(3,3)}p_1 \\
\mu_4 \cdot p_4 &= p_5[(1 - y)^4 - (1 - y - z)^4] + p_3 \cdot 4v + \lambda_{(4,2)}p_3 + \lambda_{(4,3)}p_2 + \lambda_{(4,4)}p_1 \\
\mu_5 \cdot p_5 &= p_6[(1 - y)^5 - (1 - y - z)^5] + p_4 \cdot 5v + \lambda_{(5,2)}p_4 + \lambda_{(5,3)}p_3 + \lambda_{(5,4)}p_2 + \lambda_{(5,5)}p_1 \\
\mu_6 \cdot p_6 &= p_7[(1 - y)^6 - (1 - y - z)^6] + p_5 \cdot 6v + \lambda_{(6,2)}p_5 + \lambda_{(6,3)}p_4 + \lambda_{(6,4)}p_3 + \lambda_{(6,5)}p_2 \\
&\quad + \lambda_{(6,6)}p_1 \\
\mu_7 \cdot p_7 &= p_8[(1 - y)^7 - (1 - y - z)^7] + p_6 \cdot 7v + \lambda_{(7,2)}p_6 + \lambda_{(7,3)}p_5 + \lambda_{(7,4)}p_4 + \lambda_{(7,5)}p_3 \\
&\quad + \lambda_{(7,6)}p_2 + \lambda_{(7,7)}p_1 \\
\mu_8 \cdot p_8 &= p_9[(1 - y)^8 - (1 - y - z)^8] + p_7 \cdot 8v + \lambda_{(8,2)}p_7 + \lambda_{(8,3)}p_6 + \lambda_{(8,4)}p_5 + \lambda_{(8,5)}p_4 \\
&\quad + \lambda_{(8,6)}p_3 + \lambda_{(8,7)}p_2 + \lambda_{(8,8)}p_1 \\
\mu_9 \cdot p_9 &= p_{10}[(1 - y)^9 - (1 - y - z)^9] + p_8 \cdot 9v + \lambda_{(9,2)}p_8 + \lambda_{(9,3)}p_7 + \lambda_{(9,4)}p_6 + \lambda_{(9,5)}p_5 \\
&\quad + \lambda_{(9,6)}p_4 + \lambda_{(9,7)}p_3 + \lambda_{(9,8)}p_2 + \lambda_{(9,9)}p_1 \\
\mu_{10} \cdot p_{10} &= p_{11}[(1 - y)^{10} - (1 - y - z)^{10}] + p_9 \cdot 10v + \lambda_{(10,2)}p_9 + \lambda_{(10,3)}p_8 + \lambda_{(10,4)}p_7 + \lambda_{(10,5)}p_6 \\
&\quad + \lambda_{(10,6)}p_5 + \lambda_{(10,7)}p_4 + \lambda_{(10,8)}p_3 + \lambda_{(10,9)}p_2 + \lambda_{(10,10)}p_1 \\
p_{11} &= 1
\end{aligned}$$

que resolvemos con el comando `solve()` en R para distintos valores de los parámetros y, z, u y v ,

```

> y=0.6
> z=0.1
> v=0.1
> u=0.1
> solve(E, b)
[1] 0.4395482 0.3186445 0.2618281 0.2232154 0.1934744 0.1694696 0.1496369
[8] 0.1329995 0.1188790 0.1068088

> y=0.1
> z=0.6
> v=0.1
> u=0.1
> solve(E, b)
[1] 0.151348222 0.035130962 0.010992862 0.004246914 0.002004075 0.001371326
[7] 0.001984249 0.005929293 0.026040283 0.145497109

```

```

> y=0.9
> z=0.1
> v=0.1
> u=0.1
> solve(E, b)
[1] 0.4616871 0.3850614 0.3470922 0.3178199 0.2929314 0.2712305 0.2521104
[8] 0.2351494 0.2200195 0.2064557

> y=0.9
> z=0.1
> v=0.1
> u=0.3
> solve(E, b)
[1] 0.7273230 0.6366152 0.5791088 0.5338000 0.4952623 0.4614568 0.4313621
[8] 0.4043449 0.3799537 0.3578390

> y=0.1
> z=0.9
> v=0.1
> u=0.3
> solve(E, b)
[1] 0.37349274 0.20615618 0.13104004 0.08911536 0.06297484 0.04561996
[7] 0.03372454 0.02603273 0.02690532 0.09464426

```

4.2. Probabilidad de fijación del modelo simple asimétrico

Para el modelo de simple asimétrico de la definición 32 consideramos el algoritmo descrito en la construcción 3,1. Con una población de $n = 10$ individuos y con una frecuencia inicial de individuos de tipo (+) igual a $x_0 = \frac{1}{2}$, los individuos son representados por ceros y unos para individuos tipo (-) e individuos tipo (+) respectivamente. Tomaremos una muestra de N variables aleatorias uniformes, una para cada individuo, se elige el individuo que se reproduce, es decir el que heredará su tipo. Aquellos individuos tipo (-) (o cero) cuyas variables aleatorias sean menor que el valor fijo y participan, al igual que aquellos individuos tipo (+) cuyas variables aleatorias uniformes correspondientes en la muestra sean menores que $y + z$, todos los individuos que participan copian el tipo del individuo elegido como padre, ya sea tipo (+) o tipo (-). Esto se repite hasta que la población este conformada por individuos de un solo tipo, es decir cuando uno de los dos tipos se fija. Tomamos una población de $N = 10$ individuos con frecuencia inicial $x_0 = \frac{1}{2}$ realizamos el algoritmo 1000 veces hasta que se fija un tipo, en el algoritmo tipo (+) es tipo 1 y tipo (-) es tipo 0, el promedio de veces que se fija el tipo 1 en 1000 realizaciones del algoritmo se muestra en las siguientes gráficas con distintos valores de y y z .

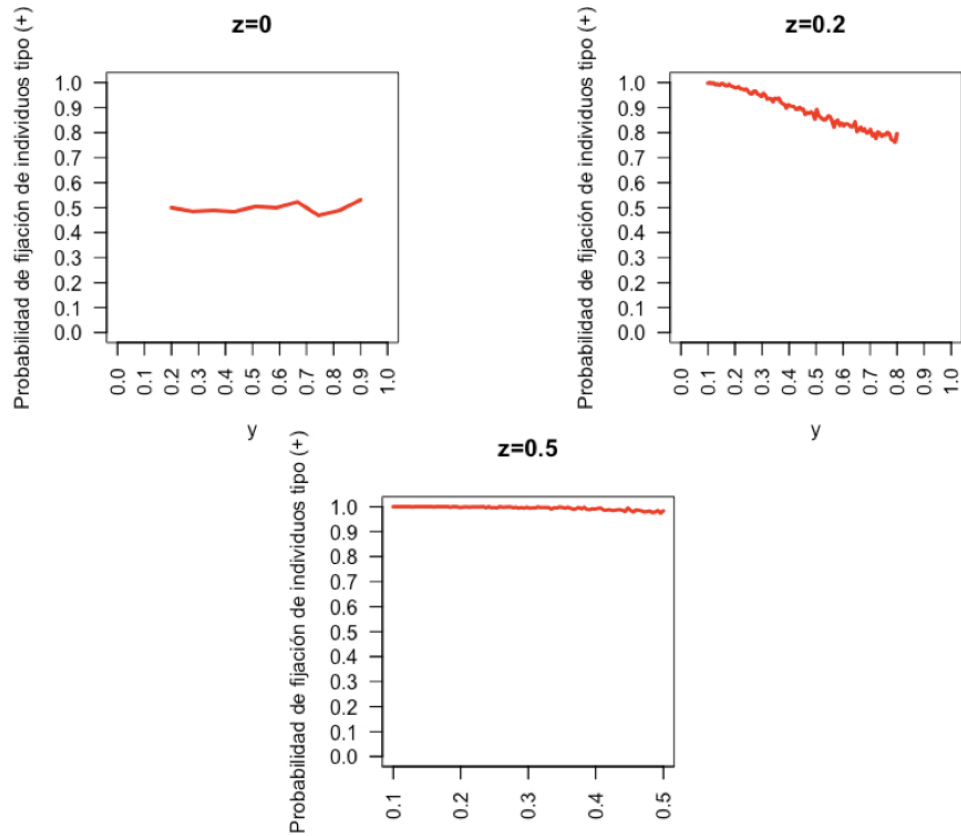


Figura 4.1: Probabilidad de fijación de individuos tipo (+) fijando z , cuando $z = 0$ no hay ventaja en la participación de individuos tipo (+), ambos tipos tienen la misma probabilidad de fijación. Cuando $z > 0$ la probabilidad de fijarse en la población es mayor.

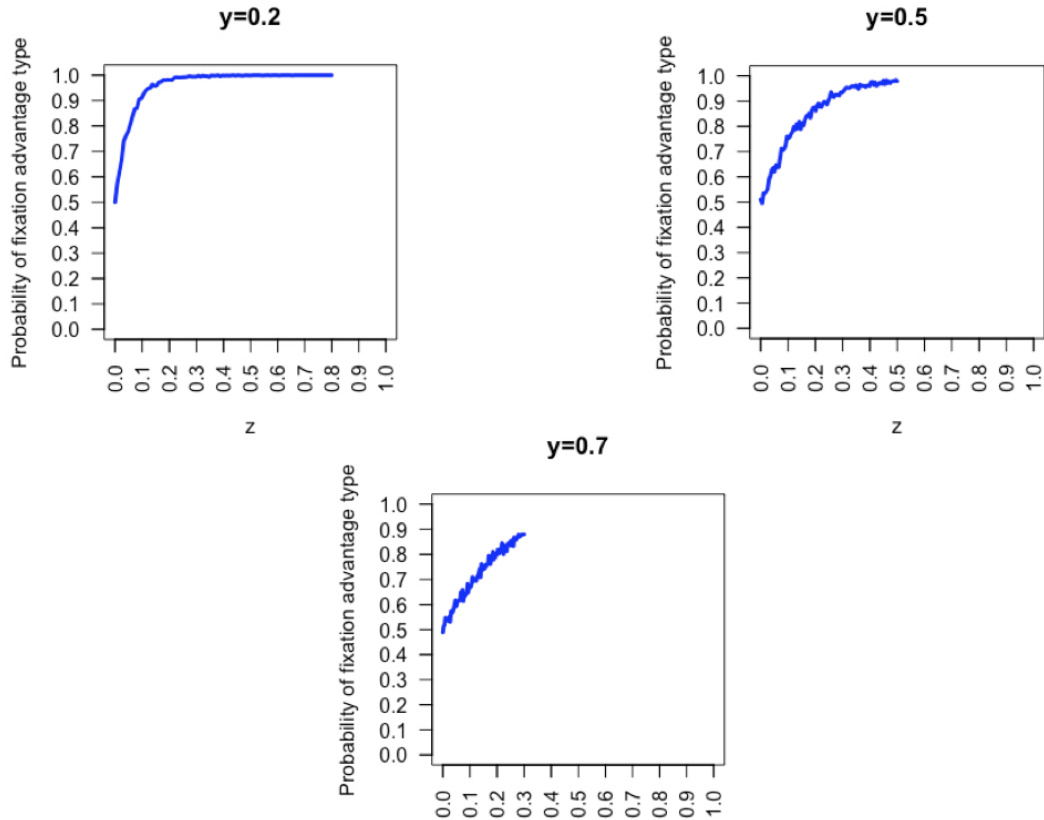


Figura 4.2: Probabilidad de fijación de individuos tipo (+) fijando y , cuando $y = 0,2$ la probabilidad de fijación es mayor en comparación con $y = 0,5$ y $y = 0,7$. Fijando y , la probabilidad de fijación de individuos tipo (+) aumenta cuando el parámetro z aumenta.

4.2.1. Algoritmo del modelo simple asimétrico

El algoritmo con el que se obtienen las gráficas de la sección anterior sigue tal cual la construcción de la sección 2,1. En una población de tamaño $M = 10$ y con frecuencia inicial $x = \frac{1}{2}$ (i.e configuración inicial $n_1 = 5$ y $n_2 = 5$) mientras la población no se fije, es decir mientras la población no sean todos unos o todos ceros, tomaremos una muestra de M variables aleatorias uniformes en $(0, 1)$, $(U_1, U_2, \dots, U_{10})$, el individuo en la posición J , con J el índice tal que $U_J = \min\{U_1, U_2, \dots, U_{10}\}$, hereda su tipo a los individuos que cuyas variables aleatorias uniformes correspondientes pertenecen al intervalo $[0, y]$ si el individuo en la posición J es 0 o hereda su tipo a los individuos que cuyas variables aleatorias uniformes correspondientes pertenecen al intervalo $[0, y+z]$ si el individuo en la posición J es 1. Este proceso se detiene hasta que el vector población (pob en el algoritmo) es una sucesión de 10 ceros o una sucesión de 10 unos, en este caso el algoritmo imprime un uno cuando se fija el 1 y un 0 cuando se fija el tipo 0. Realizamos este procedimiento 1000 veces y tomamos $\frac{\text{número de veces que gana el 1}}{1000}$ para las gráficas de la sección anterior.

```

FIX1_10=function(y,z){
  N=10
  x=1/2
  pob=rep(1,N*x)
  pob=c(pob,rep(0,N*(1-x)))

```

```

TYPE=sample(pob,N,replace = F)
##print(TYPE)
tim=c()
TYPE_new=c()
#####
while (length(unique(TYPE_new))!=1 ) {#while each individual
#has different type
  #e_clock=rexp(1,1) #TIME OF REPRODUCTION
  UI=runif(N,0,1) #DECISION VARIABLES
  J=which.min(UI) #individual who reproduces
  UI_Y=which((UI<y)|(UI==y)) #the individuals who do #
  participate
  UI_Z=which((y<UI)&((UI<y+z)|(UI==y+z))) #the individuals who
  #do participate if their father is a type +
  UI_NO=which(UI>y+z) #individuals who do not participate
  UI_YZ=which((UI<y+z)|(UI==y+z)) #all the individuals
  #who want to participate
  if(length(UI_YZ)!=0){#if there are individuals
  #who want participate
    if(TYPE[J]==1){#if the individual is of type 1
      #tim=c(e_clock,tim)
      TYPE_new=TYPE
      TYPE_new[UI_Y[which(TYPE[UI_Y]==0)]] =1
    }else if((TYPE[J]==0)&(length(UI_Y)!=0)){#if the individual
    #is of type 0
      #tim=c(e_clock,tim)
      TYPE_new=TYPE
      TYPE_new[UI_Y[which(TYPE[UI_Y]==1)]] =0
      TYPE_new[UI_Z[which(TYPE[UI_Z]==1)]] =0
    }else{TYPE_new=TYPE}
  }else{TYPE_new=TYPE}
  #print(TYPE_new)
}
return(TYPE_new[1])
}

```

Conclusión

El modelo de frecuencia simple asimétrico presentado en este trabajo resulta importante ya que se puede generalizar. Aquí consideramos la distribución del número de hijos como una distribución $Bin(N - 1, y)$ y $Bin(N - 1, y + z)$ con y y z fijos en $(0, 1)$, para individuos tipo $(-)$ y tipo $(+)$ respectivamente. Esta idea se generaliza al considerar que, en cada punto de proceso un Poisson que indica los tiempos de reproducción, los parámetros (Y_1 para el número de hijos de individuos de tipo $(-)$ y Y_2 para el número de hijos de individuos de tipo $(+)$) tienen una distribución dada por una medida $\Lambda_1(dy_1)$ y $\Lambda_2(dy_2)$, y así tener un proceso estocástico cuyo generador sea de la forma

$$Q(x) = x \int_0^1 [f(x + (1 - x)y) - f(x)] d\Lambda_1(y) + (1 - x) \int_0^1 [f(x(1 - y)) - f(x)] d\Lambda_2(y).$$

Y pensar que, así como el modelo de Eldon-Wakeley es un caso particular de los procesos de coalescencia múltiple (Λ -coalescentes) cuando $\Lambda(dx) = z^2 \delta_z(dx)$, el modelo de frecuencia simple asimétrico aquí presentado sea un caso particular de procesos genealógicos más generales.

Referencias

- [1] Berestycki N. (2009). Recent progress in coalescent theory. *Ensaio Matemáticos*. 35, 1–193.
- [2] Bertoin J., Le Gall J.-F. (2006). Stochastic flows associated to coalescent processes III: limit theorems. *Illinois J. Math.* 147–181.
- [3] Donnelly P., Kurtz T.G. (1999). Particle representations for measure-valued population models. *Ann. Probab.* 27, 66–205.
- [4] Eldon, B., J. Wakeley. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* 172: 2621–2633.
- [5] Etheridge A. M., Griffiths R. C., Taylor J. E. (2010). A coalescent dual process in a Moran model with genetic selection, and the lambda coalescent limit. *Theoretical Population Biology* 78, 77–92.
- [6] Etheridge A. (2011). Some mathematical models from population genetics. *École d’Été de Probabilités de Saint-Flour XXXIX-2009, Lecture Notes in Mathematics*, Springer.
- [7] Griffiths R.C., Tavaré S. (1994). Simulating Probability Distributions in the Coalescent, *Theoretical Population Biology*, 131-159.
- [8] Kingman J. F C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13(3): 235-248.
- [9] Krone S. M., Neuhauser C. (1997). Ancestral processes with selection. *Theoret. Population Biol.* 51, 210–237.
- [10] Lamperti J. (1967). The limit of a sequence of branching processes. *Z. Wahrsch. Verw. Gebiete* 7, 271–288.
- [11] Möhle M. (2000). Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.* 32, 983-993.
- [12] Pitman J. (1999). Coalescents with multiple collisions. *Ann. Probab.* 27 1870–1902.
- [13] Sagitov S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* 36, 1116–1125.
- [14] S. Jansen, N. Kurt. (2014). On the notion(s) of duality for Markov processes. *Probab. Surv.* 3-6.