

CIMAT

Centro de Investigación en Matemáticas, A.C.

Forma de la Distribución cuasiestacionaria
de una cadena de Markov bivariada:
Aplicación a un modelo genético

T E S I S

Que para obtener el grado de:
Maestro en Ciencias
con especialidad en

Probabilidad y Estadística

P r e s e n t a:

Valentina Castellanos Rodríguez

Guanajuato, Gto. México. Septiembre de 2006

Forma de la Distribución cuasiestacionaria de una
cadena de Markov bivariada: Aplicación a un modelo
genético

A mi adorada hija Váleriy Andrea

Agradecimientos

Agradezco primeramente a Dios por su infinita bondad, por haberme dado fuerzas en los momentos de debilidad y por todas las bendiciones que nos dio y nos sigue dando a mí y a mi hija.

A mi adorada hija Válerly por ser el motor que impulsa mi vida, por su apoyo y por su paciencia en mis horas de estudio y de desvelo, gracias a ella por haberme permitido robarle de su tiempo para dedicarlo al estudio.

Mi mayor agradecimiento es para el Dr. J. Armando Domínguez Molina por todo el apoyo que me brindó desde que estuve en el curso propedéutico hasta que terminé este proyecto. Mil gracias por su apoyo académico, tecnológico, económico, psicológico y moral, ya que sin su ayuda difícilmente hubiera terminado esta maestría.

A mis queridas compañeras Marina, Mónica y Claudia por apoyarme en todo.

Agradezco infinitamente a mi gran amiga Solsiré por acompañarme siempre con sus consejos y compañía durante el tiempo en que estaba haciendo la tesis, por soportar mis cambios bruscos de carácter y por ayudarme a cuidar a Válerly. Pero principalmente gracias por la persona tan grandiosa que es.

Un agradecimiento especial a la Dra. Graciela González Farías por haberme permitido entrar a este programa y por su gran calidad humana que, junto con el Dr. Rogelio Ramos Quiroga, hicieron de mi estancia aquí un lugar más placentero, por su trato tan amable y su apoyo incondicional.

Una mención muy especial para el Dr. Víctor Rivero, quien con sus comentarios y correcciones a la tesis, me ayudó mucho a sacar adelante el trabajo, gracias por haberme escuchado y apoyado.

Agradezco también a CONACyT por haberme apoyado económicamente con una beca de estudios de posgrado durante un tiempo, asimismo, agradezco en gran manera a CIMAT por haberme apoyado económicamente durante casi toda la maestría y en el tiempo en que hice esta tesis, mil gracias.

Gracias a Lolita y a Berenice por su trato tan amable y cariñoso hacia mi querida hija Válerly y por haberme apoyado cuando necesitaba documentos urgentes y por sus buenos deseos hacia nosotras.

A todos mis maestros por sus enseñanzas, que son invaluable y duraderas, y a todos aquellos que en algún momento me ayudaron moral o económicamente.

A todos Ustedes,

MIL GRACIAS

Contenido

Introducción	3
1 Preliminares	11
1.1 Nociones de Genética Poblacional	11
1.1.1 Conceptos básicos	12
1.1.2 Modelo de Wright-Fisher	17
1.2 Nociones del Coalescente	19
1.2.1 El coalescente para un tamaño de población constante	20
1.2.2 El coalescente en un ambiente de variación	23
2 Distribuciones Cuasiestacionarias, Semigrupos y Cadenas de Markov	27
2.1 Resultados sobre distribuciones cuasiestacionarias	28
2.2 Semigrupos de transición en ℓ^1	37
2.3 Producto tensorial y estabilidad asintótica	39
2.4 Absorción para una cadena de Markov simple	40
2.5 Dos tipos de absorción de cadenas de Markov	42
3 Comportamiento asintótico de $(X_1(t), X_2(t))$	48
3.1 Cadenas de Markov de mutaciones y deriva genética en poblaciones de tamaño variable	49
3.1.1 Modelo coalescente de Wright-Fisher-Moran	49
3.1.2 Simulación de la evolución de dos alelos bajo el modelo de Wright-Fisher-Moran	52
3.1.3 Distribución conjunta del proceso con operadores $R(t, s)$ en \mathcal{M}_s y semigrupos de transición	58
3.2 Comportamiento asintótico de las distribuciones de $(\tau_{\nu'}, \sigma_{\nu'})$	64
4 Mutación contra deriva	72
4.1 Simulación de la evolución de dos alelos bajo el modelo de Wright-Fisher-Moran en presencia de mutaciones	73
4.2 Forma de la DLC del modelo con deriva genética aleatoria	75
4.2.1 Forma de la DLC condicionada a no absorción en el origen	78

4.2.2	Forma de la DLC condicionada a no absorción en el margen	82
5	Conclusiones y discusión	93
	Bibliografía	100

Introducción

Considérese una población finita de genes que evoluciona por generaciones sin traslapes (i.e. individuos de generaciones distintas no pueden coexistir), el tamaño N de la población es el mismo en cada generación, y en la que cada individuo en la generación $k + 1$ elige aleatoriamente (con repetición) a un individuo en la generación k como su progenitor. Es bien conocido que tal mecanismo de reproducción induce una dinámica en las frecuencias genéticas de la población, conocida como deriva genética aleatoria, y que acarrea como consecuencia que después de un cierto número de generaciones la población esté constituida por descendientes de un único ancestro. Si la población estuviera constituida por alelos de un mismo gene, en ausencia de mutaciones la deriva genética conducirá la población a un estado en que todos los alelos (códigos alternativos de un gene) son del mismo tipo, y una vez que la población alcanza tal estado de fijación ésta permanece ahí. La rapidez con la que se fija o pierde un alelo depende del tamaño de la población. Así, la deriva genética aleatoria tiende a desaparecer la variación genética, mientras que las mutaciones restablecen la variación al introducir nuevos alelos en la población.

La deriva genética aleatoria se define como la selección aleatoria de gametos en el mecanismo de reproducción de una población finita. Vista hacia atrás en el tiempo,

la deriva genética es equivalente a la coalescencia de linajes y se puede modelar como un proceso markoviano de muerte puro (ver sección 1.2), el cual fue introducido por Kingman (1982) y ha sido investigado por Tajima (1983), Griffiths y Tavaré (1994), entre otros.

El coalescente de J.F. Kingman es un modelo sofisticado que ha sido de mucha utilidad para investigar la interacción de dos de las principales fuerzas genéticas que determinan la evolución de muchas poblaciones genéticas: la mutación y la deriva genética aleatoria. Dado que los tiempos de coalescencia son difíciles de obtener en la mayoría de los casos, se infieren usando la información de las mutaciones (cambios en la secuencia de ADN) a través de las generaciones. Una deficiencia de éste y de otros modelos es que no toma en cuenta las otras dos fuerzas genéticas importantes en poblaciones genéticas: recombinación y selección natural. En dichos modelos se considera la selección como neutral, es decir, se supone que todos los tipos de alelos impactan de igual forma la capacidad de reproducción o supervivencia. Similarmente se supone que las variaciones de frecuencias de alelos no son afectados por eventos de recombinación. Ejemplos clásicos de tales modelos son el modelo de Wright-Fisher y el de Wright-Fisher-Moran. Por otro lado, una justificación para ignorar las fuerzas de selección natural y de recombinación en muchos modelos genéticos es que se ve más claramente la interacción entre la deriva y las mutaciones.

El modelo de Wright-Fisher es una cadena de Markov con tiempo discreto y espacio de estados $\{0, 1, \dots, 2N\}$, el cual puede corresponder a una población diploide (cada individuo tiene una copia de su material genético) de tamaño N , o bien a una población haploide de tamaño $2N$, y donde un estado representa el número de alelos de un tipo específico. Por sencillez generalmente se supone hay sólo dos tipos de alelos en la población.

El modelo de Wright-Fisher-Moran es la versión con tiempo continuo de Wright-Fisher (ver secciones 1.1.2, 1.1.3 y 3.1.1).

Esta tesis trata de un modelo de Wright-Fisher-Moran donde el tamaño de la población al tiempo $t \geq 0$ está dado por una función $2N(t)$ que depende determinísticamente del tiempo, es decir, se supone que el tamaño de la población varía con el tiempo en forma no aleatoria. La población está constituida por alelos de un microsatélite, e interesa modelar el comportamiento de las longitudes del microsatélite. Los microsatélites son patrones de repetición de secuencia de nucleótidos cuya longitud (número de repeticiones) es inestable, i.e., puede variar de una generación a otra debido, en gran medida, a las mutaciones. La longitud de un microsatélite se refiere al número de repeticiones de la porción de código genético que se presentan con alto grado de variabilidad y ocurren con alta frecuencia en el ADN. Una de las razones por las que se elige el estudio del proceso de mutaciones en microsatélites es que precisamente en esas regiones de ADN no existe selección y la recombinación es negligible, lo cual permite emplear modelos menos complejos.

Durret et al. (1999) idearon uno de los primeros modelos para comprender el proceso de mutaciones en microsatélites. Dicho modelo fue concebido en un intento por explicar el hecho experimentalmente establecido de que la longitud X de un microsatélite generalmente no crece indeterminadamente, sino que permanece acotado. Un aspecto importante en dicho modelo es el mecanismo que gobierna el comportamiento del proceso de longitudes en el estado $X = 1$, donde se asume que después de alcanzar este estado el proceso permanece ahí por un tiempo distribuido exponencialmente y entonces salta a $X = 2$. Este mecanismo no es justificable porque se conoce muy poco acerca de cómo se crean

las secuencias repetidas de ADN. Bobrowski y Kimmel (1999) en un modelo modificado, proponen el estado $X = 1$ como un estado absorbente, lo cual implica que no existe una distribución estacionaria para el proceso por lo que es necesario condicionar el proceso a los estados donde no se da la absorción, y de esta manera conocer el estado del proceso hasta un tiempo t , antes de la absorción. Esta distribución condicionada en no absorción se conoce como distribución cuasiestacionaria (ver capítulo 2).

Los modelos de Durrett et al. (1999) y de Bobrowski y Kimmel (1999) no toman en cuenta la deriva genética. Más tarde, Bobrowski (2004) retoma el modelo de Bobrowski y Kimmel (1999) pero tomando en cuenta la deriva genética, e investiga la distribución cuasiestacionaria bajo diferentes comportamientos del tamaño de la población $2N(t)$. Uno de los pioneros en el estudio de modelos con tamaño de población variable es Beaumont (1999), quien procede a demostrar cómo diferencias en inferencias estadísticas dependen de cuál modelo demográfico es usado, en particular usa tamaño de población creciendo de manera exponencial, lineal y tamaño de población en declive.

El objetivo principal de este trabajo es hacer una descripción del modelo del proceso de mutaciones en microsatélites presentado por Bobrowski (2004) y del impacto del tamaño de población variable en la forma de la distribución cuasiestacionaria (ver sección 2.1) del proceso bivariado $(X_1(t), X_2(t))$ $t \geq 0$, donde $\{X_1(t), t \geq 0\}$ y $\{X_2(t), t \geq 0\}$ son cadenas de Markov que modelan el proceso de mutaciones en dos alelos seleccionados aleatoriamente de una población haploide de tamaño $2N(t)$, donde t es el tiempo presente y $X_i(t)$ toma valores en $S = \{0, 1, \dots\}$, $i = 1, 2$. Aquí los estados $x \in S$ representan longitud de microsatélites. También se identificará la función de velocidad de escape de masa al

margen de la distribución de $(X_1(t), X_2(t))$, condicionado a no absorción en $(0, 0)$ y en el margen $M = \{(i, j) \mid i = 0 \text{ o } j = 0\}$. En este modelo de mutaciones en microsatélites, estos dos tipos de absorción corresponden a la extinción de ambos microsatélites y de uno del par, respectivamente. La evolución de esos estados está modelada por un par de cadenas de Markov de tiempo continuo bajo el modelo de Wright-Fisher-Moran con mutaciones y deriva. Se considera un modelo donde las mutaciones siguen una cadena de Markov con estado absorbente en 0 (se escogió 0 en lugar de 1 porque facilita el empleo de funciones generadoras de probabilidad) y tal que al tiempo $t > 0$ la masa de probabilidad fuera de 0 es $O(e^{-\gamma t})$ para algún $\gamma > 0$.

Se hace una constante comparación de la forma de la distribución cuasiestacionaria del modelo sin deriva genética y del modelo con deriva genética. El modelo sin deriva genética se refiere al modelo donde las dos cadenas evolucionan de acuerdo a cierto semigrupo donde sólo interviene el mecanismo de mutación; por otro lado, el modelo con deriva genética se refiere al modelo donde se toma en cuenta tanto el mecanismo de mutación como la deriva genética a través del coalescente de Kingman (1982).

Este trabajo se basa, primordialmente, en los artículos de Bobrowski (2004) y de Bobrowski et al. (2001), quienes como se mencionó anteriormente, proponen como estado absorbente una longitud de microsatélite igual a 1. Bobrowski et al. (2001) presentan el comportamiento asintótico de ciertas estadísticas del modelo coalescente de Wright-Fisher-Moran tales como tamaño de población y mutación, entre otras, introduciendo una descripción matemática del modelo con teoría de semigrupos y tomando en cuenta la deriva genética aleatoria a través del coalescente de Kingman (1982), Bobrowski (2004) retoma lo

escrito por Bobrowski et al. (2001) y obtiene la forma de la distribución cuasiestacionaria del proceso bivariado correspondiente a la evolución de dos microsátélites.

La principal aportación de la autora de esta tesis es la simulación de las frecuencias de dos alelos bajo deriva genética aleatoria que pone de relieve el comportamiento asintótico de la frecuencia de microsátélites y que ayudan a comprender algunos de los resultados presentados por Bobrowski (2004) y Bobrowski et al. (2001). Se presentan simulaciones del modelo de Wright-Fisher-Moran, con diferente comportamiento del tamaño de la población: cuando $2N(t)$ tiende a cero, cuando $2N(t)$ crece rápidamente a infinito, cuando $2N(t)$ crece lentamente a infinito y cuando $2N(t)$ es asintóticamente constante (ver capítulo 3), que ayudan a comprender algunos de los resultados presentados en este trabajo, principalmente los relacionados con los efectos de la deriva genética en poblaciones que crecen a infinito cuando $t \rightarrow \infty$.

Es importante mencionar que aunque en este trabajo no se presentan resultados de la autora, el proceso de recolección, organización, análisis e interpretación de información sobre el tema en cuestión fue bastante arduo. Una de las principales complicaciones fue lo referente a la adquisición de material bibliográfico, sin embargo, la parte más complicada fue respecto a las interpretaciones genéticas a lo largo de todo el escrito, y el desarrollo matemático presentado en el capítulo 3. Las interpretaciones genéticas que se hacen son un tanto subjetivas pues se sabe poco acerca del origen y evolución de los microsátélites; la interpretación genética de las simulaciones y el uso de éstas para la interpretación de algunos resultados son, esencialmente, de la autora de este trabajo.

Un hecho notable es que se requirió de matemática avanzada como teoría de

semigrupos, producto tensorial y procesos estocásticos bivariados. Se requirió, además, de conocimientos básicos de programación en S-Plus para la programación y ejecución de las simulaciones, lo complicado fue integrar los conocimientos de genética y de probabilidad para la elaboración de las mismas.

Otra de las aportaciones en este trabajo fue el desarrollo explícito de demostraciones de teoremas, proposiciones y lemas presentados por algunos de los autores de los artículos mencionados en la bibliografía, principalmente por Bobrowski (2004) y Bobrowski et al. (2001), que también fue un proceso arduo y continuo.

En el Capítulo 1 se hace una breve descripción de conceptos básicos de genética de poblaciones y de los modelos de Wright-Fisher y de Moran. También se introducen nociones generales de la teoría de coalescencia.

En el Capítulo 2 se introducen los elementos que conforman un modelo del proceso de mutaciones en microsatélites, asimismo, se introducen las nociones de distribución cuasi-estacionaria y de distribución límite condicional (Pakes, 1995, Anderson, 1991). También se presentan algunos resultados acerca de la forma que toma la distribución cuasiestacionaria del proceso simple $\{X(t), t \geq 0\}$ y del proceso bivariado $(X_1(t), X_2(t))$ en el modelo sin deriva genética aleatoria.

En el Capítulo 3 se da una formulación del coalescente para el modelo de Wright-Fisher-Moran para la obtención del tiempo en que los dos microsatélites eran idénticos, i.e. estaban en el mismo estado, donde se toma en cuenta el proceso de mutaciones y la deriva genética aleatoria. Además se da la forma de la distribución conjunta del proceso $(X_1(t), X_2(t))$ bajo mutación y deriva genética en términos de semigrupos de transición.

También se desarrollan demostraciones sobre teoremas límite para la distribución conjunta de los tiempos τ , σ , donde σ es la cantidad de tiempo en que los dos microsátélites permanecen idénticos y τ es la cantidad de tiempo durante el cual los dos microsátélites evolucionan independientemente. Se presentan además, en la Sección 3.2, simulaciones de la evolución de dos alelos en una población con diferente comportamiento asintótico de $N(t)$.

En el Capítulo 4 se analiza la interacción de dos fuerzas genéticas: la mutación y la deriva genética aleatoria, por medio de la distribución cuasiestacionaria condicionada a no absorción y de la forma de la función de velocidad de escape, bajo diferentes escenarios de comportamiento asintótico del tamaño efectivo de la población $2N(\cdot)$.

Por último, en el Capítulo 5 se presentan algunas conclusiones relativas a distribuciones cuasiestacionarias del proceso bivariado según el comportamiento límite del tamaño de la población, así como algunas discusiones relativas al modelo.

Capítulo 1

Preliminares

1.1 Nociones de Genética Poblacional

La genética de poblaciones es una ciencia que ha proliferado a raíz de la teoría Darwiniana de evolución y del sistema hereditario Mendeliano. La idea que tenía Darwin del mecanismo hereditario era solamente la vaga idea de que “los hijos tienden a parecerse a sus padres ” y en ausencia de este conocimiento fue que desarrolló la teoría de evolución. Por otro lado, Mendel presenta una teoría de cómo se hereda el material genético de padres a hijos, describiéndolo como un proceso meramente aleatorio (Ewens, 2006).

El tema central en la teoría de genética de poblaciones es el estudio de los cambios en la genética de una población a través del tiempo como resultado de diversos factores, entre ellos la deriva genética aleatoria y las mutaciones. Sin embargo, la genética poblacional se enfoca en los cambios aleatorios en la genética de una población (Ewens, 2006, Hartl y Clark, 1997), debidos esencialmente a efectos de muestreo aleatorio en la transmisión de material genético de una generación a otra. El proceso aleatorio de transmitir uno de dos

genes de padres a hijos y los eventos aleatorios durante la vida, aseguran que dos individuos en condiciones iguales no necesariamente tienen la misma descendencia, por lo que el aspecto aleatorio o estocástico es componente importante en la teoría de genética poblacional, que es donde se ubica el material presentado en este trabajo.

Algunos términos genéticos estándares en la genética poblacional serán usados en esta tesis, asumiendo que el lector tiene nociones de algunos de los conceptos básicos tales como ADN y su composición, gene, teoría de Mendel y de Darwin. En la siguiente sección se presentan algunos conceptos de términos más frecuentes que incluye gene, alelo, frecuencia y probabilidad de alelos, locus, haploide, diploide, genotipo, genoma, microsatélite, recombinación, selección natural, entre otros, que aunque no profundizamos en el campo de la genética molecular o a nivel biológico, es importante tener una idea de lo que implican en la vida de un individuo.

1.1.1 Conceptos básicos

Antes de discutir la modelación de la evolución de secuencias de ADN, es necesario un conocimiento básico de lo que será modelado. Desde la perspectiva de la Genética poblacional (Hartl y Clark, 1997), un conjunto de individuos que comparten características similares, puede ser estudiado como una unidad. El punto de vista con que se abordan los conceptos de genética es en relación a la aleatoriedad del proceso de transmisión de genes de una generación a otra, es decir, no estamos interesados en profundizar sobre cuestiones de genética a nivel molecular, que aunque sería muy interesante y enriquecedor, no es el tema central del trabajo que se presenta.

El genoma de un organismo viviente (Brown, 2002) es la información biológica

necesaria para construir y mantener vivo a un espécimen de dicho organismo. Los genomas de la mayoría de los organismos vivientes son portados por moléculas de ácido desoxirribonucleico (ADN). El *ADN* consiste de dos cadenas complementarias que forman una doble hélice. Cada cadena es una secuencia lineal de cuatro nucleótidos llamados monómeros: (A) adenina, (G) guanina, (C) citosina y (T) timina.

El genoma humano se divide en genoma nuclear y mitocondrial (Brown, 2002) y es prototipo de los genomas de organismos superiores. En humanos, el genoma nuclear consta de aproximadamente 3200 millones de nucleótidos de ADN, el genoma mitocondrial está compuesto por moléculas circulares de ADN de 16 569 nucleótidos. Estas cifras nos dan una idea de la magnitud de la secuencia de nucleótidos que existen en el ADN y de la importancia que toma el proceso aleatorio en la transmisión de genes de una generación a otra.

Los términos cromosoma, genes y alelos (Durrett, 2002) tienden a confundirse e incluso se utilizan de manera indistinta por lo que es conveniente dar el concepto de cada uno de ellos y explicar en qué sentido se utilizarán en esta tesis. Los genes son regiones de código en las moléculas de ADN que en su mayoría codifican cadenas de aminoácidos que forman proteínas ubicados en un locus (sitio de la cadena de ADN); el *código* es la secuencia de nucleótidos en la región de ADN que ocupan los genes, por ejemplo, la secuencia *ATCGACGTACGAAT...* es parte de la secuencia de un gene que se diferencia de otro gene por la posición que ocupa en la cadena y los *alelos* son códigos alternativos que puede tener un gene o una colección de genes, por ejemplo, para un locus determinado se tiene el gene que determina el color de ojos de un individuo y todos los individuos de la misma

especie tienen ese gene, sin embargo existe variación respecto al color de ojos debido a que hay códigos alternativos en el locus, i.e. la secuencia de nucleótidos puede variar por diferentes causas. La palabra cromosoma viene del griego *chroma*, color y *soma*, cuerpo o elemento y es el material microscópico constituido del ADN y de proteínas especiales que se encuentran en el núcleo celular conocido (previo a la mitosis) como cromatina la cual se condensa durante el proceso de división y los cromosomas se hacen visibles como entidades independientes. Los cromosomas están compuestos por genes, en los seres humanos existen 23 pares de cromosomas, cada uno de ellos contiene miles de genes, en total 35 000 genes (consultar más detalles en <http://es.wikipedia.org>) a menudo se utiliza alelo o cromosoma indistintamente. (Para más referencias consulte Brown, 2002).

Una célula o un núcleo es *diploide* si contiene dos copias de cada cromosoma, y es *haploide* si contiene sólo una copia de cada cromosoma. Los cromosomas que determinan el sexo en diploides son una excepción de la regla de dos copias; en humanos, las mujeres tienen dos X cromosomas y los hombres uno X y uno Y.

Cuando los organismos haploides se reproducen, hay un padre que hereda copias de su material genético a sus descendientes. Cuando los diploides se reproducen, existen dos padres, cada uno de los cuales contribuye con uno de sus pares de cromosomas. De hecho, la contribución de uno de los padres puede ser una combinación de sus dos cromosomas: debido a que los pares homólogos sufren recombinación, un intercambio recíproco de material genético entre ellos puede ser diagramado como sigue (Durrett, 2002), donde las secuencias de la primera columna corresponden a los alelos del padre y de la madre respectivamente,

mientras que los alelos de la segunda columna corresponden a dos posibles alelos de un hijo

$$\begin{array}{cc}
 aaaaaaaaaaaaaaaaa & aaaaaavvvvvvvvv \\
 & \rightarrow \\
 vvvvvvvvvvvvvvv & vvvvvaaaaaaaaa.
 \end{array}$$

La recombinación es un fenómeno genético de gran importancia para la evolución porque acelera la tasa de formación de combinaciones de genes benéficos (para más referencias, ver Hartl y Clark, 1997). En el modelo que se analiza no se tomará en cuenta la recombinación pues el análisis matemático se complica. Sólo se tomarán en cuenta la deriva genética aleatoria y las mutaciones (se explican abajo).

Las diferencias entre genomas de individuos de una población de la misma especie se manifiestan, fundamentalmente, de dos maneras (Hartl y Clark, 1997, Brown, 2002, Gillespie, 1998):

- *Polimorfismos* de nucleótidos originados generalmente por mutaciones. Los polimorfismos son posiciones en el genoma donde algunos individuos pueden tener cierto nucleótido, por ejemplo: (A) adenina y otros tienen un nucleótido distinto, como (G). La mutación incluye un cambio en la secuencia de nucleótidos en un gene así como la formación de un rearrreglo de cromosomas, tal como una inversión o una translocación (sustitución de nucleótidos, eliminación o inserción). En el genoma humano se han identificado, aproximadamente, 1.4 millones de polimorfismos de nucleótidos.
- *Microsatélites* (Sibly, et al., 2001), que son series de repeticiones del código genético de longitud variable entre individuos; por ejemplo: ACGTAGTCTCTCTCAGTC, donde la longitud del microsatélite es cuatro, i.e. TC se repite cuatro veces.

La frecuencia de un tipo de alelo en una población dada es la proporción de alelos en dicha población que son del tipo especificado (Gillespie, 1998). Si en una población diploide de tamaño N hay dos alelos: A y a (respecto a un lugar en una porción de código genético) entonces hay 3 *genotipos* (composición del genoma de un individuo) posibles referente al sitio en cuestión: AA , Aa y aa . Sean x_{11} , x_{12} y x_{22} las frecuencias de los tres genotipos, respectivamente. Entonces $x_{11} + x_{12} + x_{22} = 1$, de modo que la frecuencia del alelo A en una población diploide de tamaño N es $p = x_{11} + \frac{1}{2}x_{12}$, mientras que $q = 1 - p$ es la frecuencia del alelo a . Se puede interpretar a p de dos formas:

- p es la proporción de A -alelos en la población
- p es la probabilidad de que un alelo elegido aleatoriamente de la población sea del tipo A .

Las principales fuerzas genéticas que influyen en la evolución de las especies son la mutaciones, la recombinación, la deriva genética aleatoria y la selección natural. Las tres primeras ya se introdujeron anteriormente. El concepto de *selección natural* tiene diferentes formulaciones:

- En todas las especies, se reproducen más descendientes que pueden posiblemente sobrevivir y reproducirse.
- Los organismos difieren en su habilidad para sobrevivir y reproducirse, en parte debido a las diferencias entre genotipos.
- En cada generación, los genotipos que promueven la sobrevivencia en el ambiente

actual están presentes en exceso en la edad reproductiva, por lo que se heredan desproporcionalmente a los descendientes en la siguiente generación.

1.1.2 Modelo de Wright-Fisher

En este modelo se considera un *locus* genético con dos alelos A y a que tienen la misma capacidad de reproducción (*fitness*) en una población diploide de tamaño constante N que evoluciona por generaciones que no se traslapan, bajo cruzamiento aleatorio. Para comprender mejor estos términos se da una explicación general (Durrett, 2002).

Locus genético es el lugar, en el genoma de un organismo, ocupado por una porción de código. Un ejemplo común es el lugar ocupado por una secuencia de nucleótidos que forman un gene. Los dos alelos A y a son sólo diferentes versiones de la información genética codificada en el *locus*.

Como ya se mencionó, el *fitness* de un individuo es una medida de la habilidad individual para sobrevivir y reproducirse. Aquí se considera el caso de *evolución neutral* en el cual las mutaciones cambian la secuencia del ADN pero no alteran el *fitness*.

Los individuos diploides poseen dos copias de su material genético en cada célula somática (de los 23 cromosomas en humanos, 1 es sexual y 22 somáticas). En general, se tratarán N individuos diploides como si fueran $2N$ copias del locus.

La condición de que las generaciones no se traslapen significa que un individuo en la generación n no estará presente en la siguiente generación, es decir, en la generación $n + 1$.

En el modelo de Wright-Fisher cada generación contiene un número fijo, N , de individuos diploides, aunque en la mayoría de las aplicaciones genéticas estos son los gametos

exitosos en el proceso de reproducción.

En la generación n , $2N$ gametos son muestreados de forma aleatoria de los padres, de los cuales i gametos portan el alelo A y $2N - i$ gametos portan el alelo a . Para construir la generación $n + 1$ se escoge aleatoriamente $2N$ veces con reemplazo. Este proceso aleatorio en la selección de gametos se conoce como deriva genética aleatoria y trae como consecuencias cambios en la frecuencia de alelos y disminución de la diversidad de alelos en la población. Si X_n denota el número de gametos del tipo A en la generación n , entonces, en ausencia de selección y mutación, el número de A 's al tiempo $n + 1$ es dado por la distribución binomial

$$P(X_{n+1} = j \mid X_n = i) = \binom{2N}{j} p_i^j (1 - p_i)^{2N-j}, \quad (1.1)$$

donde $p_i = \frac{i}{2N}$ es la probabilidad de seleccionar un alelo A .

Se obtiene así una cadena de Markov (X_n) con espacio de estados $\{0, 1, \dots, 2N\}$, que es conocida como el modelo de Wright-Fisher. En el límite cuando $n \rightarrow \infty$ eventualmente el número de A 's en la población será 0, indicando la pérdida del alelo A , o $2N$, indicando la pérdida de a . Una vez que un alelo se pierde de la población nunca regresa. Así, los estados 0 y $2N$ son estados absorbentes para la cadena de Markov (X_n) . Sea

$$\tau = \min \{n : X_n = 0 \text{ o } X_n = 2N\}$$

el tiempo de fijación, es decir, el tiempo en el cual aparece la primera generación en la que la población consiste sólo de un tipo de alelo: A o a . Como el número de individuos es finito, la fijación ocurrirá eventualmente. Debido a que la media de la ley binomial en (1.1) es $2Np_i$, se sigue que

$$E(X_{n+1} \mid X_n = i) = 2N \cdot \left(\frac{i}{2N} \right) = i = X_n, \quad (1.2)$$

o sea que (X_n) es una martingala. Esto tiene como consecuencia que:

$$P_i(X_\tau = 2N) = \frac{i}{2N}. \quad (1.3)$$

De hecho, si denotamos por P_i a la distribución de probabilidad del proceso X_n empezando de $X_0 = i$, y por E_i al valor esperado con respecto a P_i , entonces de la propiedad de martingala y del hecho que τ es tiempo de paro, resulta

$$E_i X_\tau = E_i X_0 = i. \quad (1.4)$$

Como X_τ es 0 o $2N$, tenemos $i = E_i X_\tau = 2N P_i(X_\tau = 2N)$, de donde se sigue (1.3).

El modelo de Wright-Fisher es un ejemplo de una cadena de Markov de tiempo discreto, en el cual es difícil calcular en forma cerrada cantidades de interés biológico tales como el tiempo esperado de fijación. Para poblaciones grandes es posible hacer aproximaciones por un proceso de difusión, de donde se demuestra que el tiempo esperado de fijación de un alelo en la población es $2N$ generaciones. (Para más detalles, ver Capítulo 7 de Hartl y Clark, 1997 o Karlin y Taylor, 1989).

El modelo de Wright-Fisher considera generaciones que no se traslapan, sin embargo, en muchas especies, e.g., humanos, las generaciones no están sincronizadas y es conveniente usar el modelo de Moran (para una introducción consultar Durrett, 2002). Una descripción completa del modelo de Wright-Fisher-Moran se presenta en el Capítulo 3.

1.2 Nociones del Coalescente

Una forma de modelar y analizar los efectos conjuntos de la deriva genética y de las mutaciones es a través del proceso coalescente (Norborg, 2003). Este proceso modela la

historia ancestral de una n -muestra de secuencias tomadas en el presente. Posibles eventos que ocurrieron en el pasado son coalescencias dejando ancestros comunes en el pasado, y mutaciones a lo largo de las ramas del árbol ancestral. El proceso coalescente es un proceso de Markov puro de muerte (Kingman, 1982a, 1982b, Griffiths y Tavaré, 1994a, 1994b y Bertoin, 2005).

1.2.1 El coalescente para un tamaño de población constante

Supondremos que la población evoluciona de acuerdo con el modelo de Wright-Fisher y que las mutaciones en el locus son selectivamente neutrales, i.e. las mutaciones no afectan la capacidad de reproducción del alelo (ver arriba, Sección. 1.1.2), que el tamaño de la población es constante en el tiempo y que no existe recombinación dentro del locus.

Para superimponer mutaciones, es necesario saber cuál linaje coalesce con cuál y cuándo. En otras palabras, se requiere conocer las longitudes de las ramas del árbol ancestral (los tiempos de coalescencia) y los tiempos entre eventos coalescentes. Esto es fácil de modelar debido a la neutralidad, i.e., las mutaciones cambian la secuencia del ADN pero no alteran el *fitness* (habilidad individual de reproducción), por tanto, todos los linajes tienen la misma probabilidad de coalescer. Es conveniente representar la topología del árbol ancestral como una secuencia de clases equivalentes de coalescencia: dos miembros de la muestra original son equivalentes en un cierto punto en el tiempo si tienen un ancestro común en ese tiempo (Balding, et al., 1997), como se muestra en la Figura 1.1.

Rastreando dos linajes en el pasado, hemos visto que bajo el modelo de Wright-Fisher los hijos elijen a sus padres aleatoriamente de la generación previa, de forma independiente. Entonces, la probabilidad de que dos linajes vengan del mismo padre y coalescan

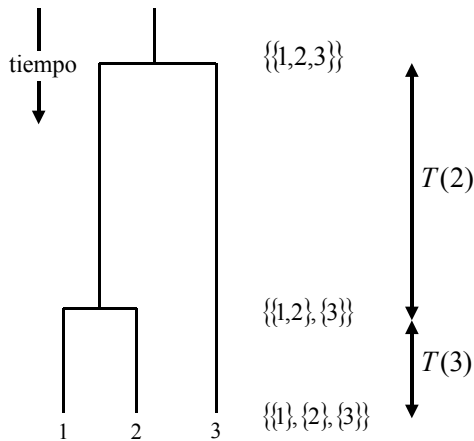


Figura 1.1:

en la generación anterior es $\frac{1}{N}$, y la probabilidad de que ellos vengan de distintos padres y permanezcan distintos es $(1 - \frac{1}{N})$. Como las generaciones son independientes, la probabilidad de que dichos linajes permanezcan distintos más de t generaciones en el pasado es $(1 - \frac{1}{N})^t$, así, el tiempo esperado de coalescencia es N generaciones (Bertoin, 2005). Reescalando el tiempo, de modo que en la nueva escala una unidad de tiempo corresponda a N generaciones, se ve que la probabilidad de que dos linajes permanezcan distintos por más de τ unidades de tiempo escalado es

$$\left(1 - \frac{1}{N}\right)^{[N\tau]} \rightarrow e^{-\tau},$$

cuando N tiende a ∞ (donde $[x]$ es el mayor entero menor o igual que x). Entonces, en el límite, el tiempo de coalescencia para un par de linajes es exponencialmente distribuido con media 1. La probabilidad de que un gameto se "extinga" de una generación a otra es

$$\left(1 - \frac{1}{N}\right)^{[N\tau]} \approx e^{-1} \approx 0.367879\dots$$

Considerando k linajes, la probabilidad de que ninguno de ellos coalesca en la

generación previa es

$$\prod_{i=0}^{k-1} \frac{N-i}{N} = \prod_{i=1}^{k-1} \left(1 - \frac{i}{N}\right) = 1 - \frac{\binom{k}{2}}{N} + O\left(\frac{1}{N^2}\right),$$

y la probabilidad de que más de dos coalescan es $O\left(\frac{1}{N^2}\right)$. Sea $T(k)$ el tiempo escalado en que se observa la primera coalescencia, dado que hay k linajes. Por el argumento de arriba, $T(k)$ es, en el límite, una variable aleatoria exponencialmente distribuida con media $\frac{2}{[k(k-1)]}$ (Bertoin, 2005).

Bajo la aproximación del coalescente, el número de linajes distintos en la ancestría de una muestra de tamaño finito n decrece en saltos de tamaño 1 al tomar el tiempo en reversa, y $T(k)$ es la cantidad de tiempo que al coalescente le toma para pasar de k a $k-1$ linajes (ver Figura 1.1).

En resumen, el coalescente modela la genealogía de una muestra de n individuos haploides como un árbol bifurcado aleatoriamente, donde los $n-1$ tiempos de coalescencia $T(n), T(n-1), \dots, T(2)$ son variables aleatorias mutuamente independientes y distribuidas exponencialmente. Cada par de linajes coalescen independientemente con tasa 1, así, la tasa total de coalescencia cuando existen k linajes es $\binom{k}{2}$. Una manera concisa de describir el coalescente es como un proceso de Markov de tiempo continuo con espacio de estados ε_n consistente de todas las relaciones de equivalencia en $\{1, \dots, n\}$, y de generador infinitesimal $Q = (q_{\xi\eta})_{\xi, \eta \in \varepsilon_n}$ definido por

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

donde $k := |\xi|$ es el número de clases de equivalencia en ξ , y $\xi \prec \eta$ si η se obtiene de ξ por

la coalescencia de dos clases equivalentes de ξ .

1.2.2 El coalescente en un ambiente de variación

Imagínese una población haploide evolucionando de acuerdo con el modelo de Wright- Fisher pero con tamaño de población variable y determinístico, i.e., a partir de un tiempo t_0 , tal que $t > t_0 > 0$, se conoce el tamaño de la población, el cual varía con el tiempo. Piénsese en el tiempo en dirección reversa a partir del instante actual, el cual se considerará como generación 0. Supóngase que existen $M_0 = N$ individuos ahora, M_1 en la generación 1, M_2 en la generación 2, y así sucesivamente. El mecanismo de reproducción tiene la siguiente dinámica: cada uno de los M_r individuos en la generación r escoge a sus padres uniformemente de manera aleatoria de los M_{r+1} individuos en la generación $r + 1$, independientemente de lo que se escogió en las generaciones $0, 1, \dots, r - 1$. Se define la función del tamaño relativo ν_N por

$$\nu_N(x) = \frac{M_t}{N}, \quad \frac{t}{N} \leq x \leq \frac{t+1}{N}, \quad x \geq 0, \text{ y } t = 0, 1, \dots$$

Obsérvese que $\nu_N(x)$ es el tamaño de la población hace x unidades de tiempo reescalado (i.e., donde una unidad de tiempo equivale a N generaciones) relativa al tamaño de ésta en la generación 0 (Griffiths y Tavaré, 1994a, 1994b). Para comprender el comportamiento del coalescente es de interés el caso en que cada generación es grande y se verifica que

$$\lim_{N \rightarrow \infty} \nu_N(x) = \nu(x)$$

existe y es estrictamente positivo para toda $x > 0$. A manera de ejemplo, si $M_r = (1-s)^r M_0$, donde $0 < s < 1$, se tiene un decrecimiento geométrico del tamaño de la población visto hacia atrás en el pasado. Si $s = \frac{\beta}{N}$ y el tiempo es medido en unidades de N generaciones,

entonces, la probabilidad de que dos linajes permanezcan distintos por más de x unidades de tiempo escalado está dado por

$$\lim_{N \rightarrow \infty} \nu_N(x) = \lim_{N \rightarrow \infty} \left(1 - \frac{\beta}{N}\right)^{[Nx]} = e^{-\beta x} \equiv \nu(x), \quad x > 0.$$

Cuando el tamaño de la población es aproximadamente constante, $\nu(x) \equiv 1$ para toda x .

Esto se sigue de que $\nu(x)$ es una función del tamaño relativo de la población, i.e.,

$$\nu(x) = \frac{M_x}{M_0},$$

donde M_x es el tamaño de la población en x unidades de tiempo y M_0 es el tamaño inicial de la población. En particular, si el tamaño de la población es constante, entonces $\nu(x) \equiv 1$.

Es conveniente definir la función de intensidad Λ del tamaño de la población (para así tomar en cuenta las fluctuaciones en el tamaño de la población) por

$$\Lambda(x) = \int_0^x \frac{dt}{\nu(t)}, \quad x > 0. \quad (1.5)$$

La densidad λ de Λ está dada por

$$\lambda(x) = \frac{1}{\nu(x)}, \quad x > 0, \quad (1.6)$$

que se puede pensar como el tamaño de la población hace Nx generaciones relativo al tamaño actual.

La estructura del coalescente con tamaño de población variable, dependiente del tiempo, puede ser descrito como sigue (Griffiths y Tavaré, 1994a). Considérese una muestra de n individuos tomados al tiempo presente $t = 0$, y sea $\{A_n(t), t > 0\}$ un proceso estocástico donde $A_n(t)$ es el número de distintos ancestros de la muestra al tiempo t . $A_n(\cdot)$ es un proceso de Markov de muerte no homogéneo que empieza en $A_n(0) = n$ y evoluciona en

saltos de tamaño -1 hasta alcanzar el valor 1 , tiempo en el cual la muestra tiene un ancestro común. Las probabilidades de transición del proceso $\{A_n(t), t > 0\}$ están determinadas por

$$P(A_n(t+h) = j \mid A_n(t) = i) = \begin{cases} \binom{i}{2} \lambda(t) h + o(h), & \text{si } j = i - 1 \\ 1 - \binom{i}{2} \lambda(t) h + o(h), & \text{si } j = i \\ 0, & \text{en otro caso,} \end{cases}$$

donde $\lambda(t)$ está dada por (1.6). Las distribuciones marginales de $\{A_n(t), t > 0\}$ pueden obtenerse mediante un cambio de tiempo determinístico del proceso $A_n(\cdot)$:

$$A_n(t) = \tilde{A}_n(\Lambda(t)), \quad t \geq 0. \quad (1.7)$$

$A_n(t)$ cuenta el número de distintos ancestros en la muestra al tiempo t y $\tilde{A}_n(\Lambda(t))$ es sólo un cambio de tiempo determinístico del modelo de tamaño de población constante, es decir, toma en cuenta el tiempo escalado y las fluctuaciones en el tamaño de la población (Griffiths y Tavaré, 1994b). Todas las propiedades del proceso de tamaño variable pueden ser calculadas a partir del modelo de población constante usando la representación (1.7).

Por ejemplo, esto muestra inmediatamente (ver Sección 3.1.3) que si $\Lambda(\infty) := \int_0^\infty \frac{dt}{\nu(t)} = \infty$, entonces cada par de individuos y la muestra tendrán un ancestro común con probabilidad 1.

También, de (1.7) es posible obtener la distribución conjunta de los tiempos en que el proceso ancestral permanece en cada estado (Griffiths y Tavaré, 1994a; i.e., sean T_n, T_{n-1}, \dots, T_2 las cantidades de tiempo durante las que el proceso ancestral permanece en los estados $n, n-1, \dots, 2$, respectivamente. La densidad conjunta de (T_n, \dots, T_2) es

$$g_n(t_n, \dots, t_2) = \prod_{j=2}^n \binom{j}{2} \lambda(s_j) \exp \left\{ - \binom{j}{2} (\Lambda(s_j) - \Lambda(s_{j+1})) \right\}, \quad (1.8)$$

para $0 \leq t_n, \dots, t_2 < \infty$, donde $s_{n+1} = 0$, $s_n = t_n$, $s_j = t_j + \dots + t_n$, $j = 2, \dots, n-1$ (ver Polanski y Kimmel, 2003). Al igual que en el caso de tamaño de población constante, el

coalescente con tamaño de población variable se puede pensar como un árbol cuyos nodos son los ancestros comunes de la muestra y con $n - 1$ ramas de longitud T_j , $j = 2, 3, \dots, n$.

En el Capítulo 3 se hará referencia a la función de densidad de probabilidad de coalescencia de una muestra de $n = 2$ individuos haploides, la cual se define a continuación.

De (1.8) se sigue que

$$g_2(t) = \lambda(t) \exp \{-\Lambda(t)\}.$$

Empleando (1.5) con $n = 2$ da

$$g_2(t) = \frac{1}{\nu(t)} \exp \left\{ - \int_0^t \frac{du}{\nu(u)} \right\},$$

donde $\nu(u) = 2N(u)$ (Griffiths y Tavaré, 1994a) por lo que

$$g_2(t) = \frac{1}{2N(t)} \exp \left\{ - \int_0^t \frac{du}{2N(u)} \right\} \quad (1.9)$$

y la probabilidad de no coalescencia para $n = 2$ en el intervalo $[t_0, t]$ es

$$p_2(t) = 1 - \int_{t_0}^t g_2(t, u) du = \exp \left\{ - \int_{t_0}^t \frac{1}{2N(u)} du \right\}. \quad (1.10)$$

(Ver Donnelly y Tavaré, 1994, Griffiths y Tavaré, 1994a).

Capítulo 2

Distribuciones Cuasiestacionarias, Semigrupos y Cadenas de Markov

La variación en la longitud de un microsatélite depende, esencialmente, del proceso de mutaciones. Una mutación que ocurre en un microsatélite origina un cambio en su secuencia de nucleótidos. Por ejemplo, sea `CGTCTCTCTCTCGAT` un microsatélite, donde el dinucleótido `TC` se repite 5 veces, entonces la longitud del microsatélite es $s = 5$. Si en la siguiente generación hay una mutación por inserción de algún nucleótido, entonces cambia la longitud del microsatélite; en el ejemplo anterior, si se inserta un nucleótido `G`, en el microsatélite: `CGTCTCTCGTCTCGAT`, entonces la longitud cambia de $s = 5$ a $s = 3$, tomando en cuenta las repeticiones del lado izquierdo de la inserción. En general, la longitud de un microsatélite puede cambiar cuando se cambia un nucleótido por otro o cuando uno de ellos se borra o cuando se inserta otro en la siguiente generación.

En este Capítulo se introducen los elementos que conforman un modelo del pro-

ceso de mutaciones en microsatélites, el cual se describe como una cadena de Markov de tiempo continuo $\{X(t), t \geq 0\}$ con espacio de estados $S = \{0, 1, \dots\}$ y estado absorbente 0, donde $X(t)$ representa la longitud del microsatélite en el tiempo $t \geq 0$. Asimismo, se introducen las nociones de distribución cuasiestacionaria y de distribución límite condicional (Pakes, 1995 y Anderson, 1991). También se analizará la forma que toma la distribución cuasiestacionaria del proceso $\{X(t), t \geq 0\}$ correspondiente a una cadena Markov simple y del proceso conjunto $\{(X_1(t), X_2(t)), t \geq 0\}$ correspondiente a la evolución conjunta de dos cadenas de Markov con el mismo espacio de estados S , cada una de las cuales modela la evolución de un microsatélite.

2.1 Resultados sobre distribuciones cuasiestacionarias

Supóngase que el comportamiento de un sistema está descrito por una cadena de Markov de tiempo continuo $\{X(t), t \geq 0\}$, con un espacio de estados de la forma $S \cup \{a\}$, donde a es un estado absorbente y S es una clase transitoria e irreducible. Supóngase que un análisis del proceso ha revelado la certeza de absorción eventual en a , pero que la evaluación explícita de la distribución del proceso $\forall t \geq 0$ es difícil de obtener o no es viable. Una manera de obtener información sobre el proceso es tratando de encontrar su distribución límite, condicionado a no absorción (Van Doorn, 1991). Esta distribución es llamada distribución cuasiestacionaria (DCE). En Anderson (1991), Secciones 3.2, 5.1 y 5.2 se da un ejemplo de un proceso de nacimiento y muerte con inmigración, que bajo ciertas condiciones no posee una distribución estacionaria pero sí tiene una distribución cuasiestacionaria. A continuación se presentan algunos resultados importantes, así como definiciones

de DCE y de distribución límite condicional (DLC). Un tratamiento más formal se da en Pakes (1995), en Ferrari et al. (1995) y en Anderson (1991, Secciones.5.1 y 5.2). Se omiten las demostraciones de los teoremas sobre distribuciones cuasiestacionarias presentados en esta sección debido a que se pueden consultar en Anderson (1991, Secciones.5.1 y 5.2).

Se considera un proceso de Markov de tiempo continuo $\{X(t), t \geq 0\}$ con espacio de estados $S = S_+ \cup \{0\}$, tal que $S_+ = \{1, 2, 3, \dots\}$ y 0 es estado absorbente. Sea $Q = [q_{ij}]$ el generador. Se supondrá que Q es estable, conservativa y regular, i.e.,

$$-q_{ii} = \sum_{i \neq j} q_{ij} < \infty,$$

y el proceso minimal $\{X(t)\}_{t \geq 0}$ correspondiente a Q es un proceso honesto, es decir, $P_\mu(X(t) \in S) = 1$ para toda $t \geq 0$ y para todas las distribuciones iniciales μ en S_+ . Se denota por $p_{ij}(t)$ a las probabilidades de transición del proceso con generador Q . Sea $P_i(\cdot) = P_i(\cdot | X(0) = i)$, $i \in S_+$, y, si ν es una medida finita en S_+ , sea $P_\nu = \sum \nu_i P_i$. Aquí y en lo sucesivo, cualquier suma será tomada sobre S_+ . Finalmente se asume que S_+ es irreducible y que 0 es accesible desde algún estado en S_+ .

Sea $T = \inf \{t \geq 0 : X(t) = 0\}$ el tiempo de absorción en 0. En lo sucesivo se supondrá siempre que $P_i(T < \infty) = 1$. Los resultados sobre distribuciones cuasiestacionarias están muy relacionados con el siguiente Teorema (Anderson, sec. 5.2) y la siguiente definición de medida invariante.

Definición 2.1 *Dada una función de transición $P_{ij}(t)$, un conjunto $u = \{u_i, i \in E\}$ de números no negativos tales que*

$$\sum_{i \in E} u_i P_{ij}(t) = u_j, \text{ para toda } j \in E, t \geq 0,$$

es llamada medida invariante para $P_{ij}(t)$. Si

$$\sum_{i \in E} u_i = 1$$

entonces u es llamada distribución estacionaria.

Cuando se toma como distribución inicial u , la probabilidad de que el proceso se encuentre en el estado j en el tiempo t es independiente del estado donde inicia el proceso. Otra interpretación que pudiera darse, está relacionada con el siguiente Teorema.

Teorema 2.2 *Suponga que $P_{ij}(t)$ es una función de transición irreducible*

(1) Entonces los límites $u_j = \lim_{t \rightarrow +\infty} P_{ij}(t)$ existen y son independientes de i para toda j en el espacio de estados E .

El vector $u = (u_i, i \in E)$ es una medida invariante y

(a) $u_j = 0$ para toda $j \in E$, o

(b) $u_j > 0$ para toda $j \in E$ y $\sum_{j \in E} u_j = 1$.

(2) Suponga que $w = (w_i, i \in E)$ es un vector de probabilidad tal que $wP(t) = w$ para alguna $t > 0$. Entonces $wP(t) = w$ para toda $t \geq 0$ (i.e., w es una distribución invariante) y $w = u$.

El inciso (a) del teorema anterior, indica que $P_{ij}(t)$ es una función de transición no honesta en el sentido de que $\sum_j P_{ij}(t) \leq 1$, y entonces todos los estados son transitorios y por tanto

$$\lim_{t \rightarrow \infty} P_{ij}(t) = 0, \text{ para toda } i, j \in E.$$

Para el caso del inciso (b) E es un espacio de estados irreducible y recurrente.

La existencia de un parámetro de decaimiento, que juega un papel importante en distribuciones cuasiestacionarias, se demuestra a partir de la siguiente proposición.

Proposición 2.3 (Anderson). T denota el conjunto $\{0, 1, 2, \dots\}$ de enteros no negativos, o el conjunto $[0, +\infty)$. Sea $\phi : T \rightarrow [0, +\infty]$ que satisface

- (i) Existe $a \geq 0$ tal que $\phi(t) < +\infty$ si $t \geq a$,
- (ii) Si $T = [0, +\infty)$, entonces ϕ es acotada en subintervalos acotados de $[a, +\infty)$,
- (iii) ϕ es subaditiva, i.e., $\phi(s + t) \leq \phi(s) + \phi(t)$ para $s, t \in T$.

Entonces $\lim_{t \rightarrow \infty} \frac{\phi(t)}{t}$ existe y

$$\lim_{t \rightarrow \infty} \frac{\phi(t)}{t} = \inf_{t > 0} \frac{\phi(t)}{t} < +\infty.$$

Teorema 2.4 (Kingman, 1963). Suponga que C es una clase irreducible para $P_{ij}(t)$.

(1) Existe un número $\lambda_c \geq 0$ llamado parámetro de decaimiento de C , tal que para cada par $i, j \in C$,

$$\frac{1}{t} \log P_{ij}(t) \rightarrow -\lambda_c, \text{ cuando } t \rightarrow \infty.$$

(2) Para cada $i \in C$ y $t > 0$,

$$P_{ii}(t) \leq e^{-\lambda_c t}.$$

(3) Para cada par $i, j \in C$, existe una constante $M_{ij} > 0$ tal que

$$P_{ij}(t) \leq M_{ij} e^{-\lambda_c t}.$$

(4) Si $\lambda_c > 0$, la clase C es transitoria.

La demostración se hace tomando la función $\phi(t) = -\log P_{ii}(t)$ de la proposición anterior anterior y con

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{\phi(t)}{t} = \inf_{t > 0} \frac{\phi(t)}{t} < +\infty.$$

Suponga que C es una clase transitoria para $P_{ij}(t)$. Si C fuera recurrente positiva, los límites ergódicos

$$u_j = \lim_{t \rightarrow \infty} P_{ij}(t), \quad i, j \in E,$$

formarían una distribución de probabilidad y darían información del comportamiento límite de la cadena en C . Como C es transitoria, de los Teoremas 2.2 y 2.4 se sigue que todas las $u_j = 0$.

Una estrategia que se sigue para conocer el estado del proceso hasta un tiempo $t > 0$, antes de la absorción es examinar el comportamiento límite de las probabilidades condicionadas a la no absorción, es decir, mientras el proceso está en los estados no absorbentes

$$\begin{aligned} \tilde{P}_{ij}(t) &:= P \{X(t) = j \mid X(0) = i, X(t) \in C\} \\ &= \frac{P_{ij}(t)}{\sum_{k \in C} P_{ik}(t)}, \quad i, j \in C. \end{aligned}$$

Cuando $t \rightarrow \infty$, en el término de la derecha el numerador tiende a 0. Pero por el Teorema 2.4 el parámetro de decaimiento λ_c es estrictamente positivo y tal que $P_{ij}(t)e^{\lambda_c t}$ tiende a un límite finito diferente de 0 cuando $t \rightarrow \infty$. Entonces

$$\tilde{P}_{ij}(t) = \frac{P_{ij}(t)e^{\lambda_c t}}{\sum_{k \in C} P_{ik}(t)e^{\lambda_c t}}$$

y $\tilde{P}_{ij}(\infty) = \lim_{t \rightarrow \infty} \tilde{P}_{ij}(t)$ existirán y serán estrictamente positivos. El conjunto de números

$$\tilde{P}_{ij}(\infty), \quad i, j \in C$$

es llamado Distribución cuasiestacionaria (DCE). Uno de los principales problemas es cómo determinar los límites

$$\lim_{t \rightarrow \infty} P_{ij}(t)e^{\lambda ct}$$

pero $P_{ij}(t)e^{\lambda ct}$ tiene las propiedades de semigrupo de $P_{ij}(t)$, y así, esperamos que los límites $\lim_{t \rightarrow \infty} P_{ij}(t)e^{\lambda ct}$ jueguen el rol de una medida invariante, como en el Teorema 2.2. Una definición más formal se da en Pakes (1995).

Definición 2.5 Una medida $M = \{m_i\}$ en S_+ es una DCE si $\sum m_i = 1$ y si para toda $j \in S_+$ y $t > 0$

$$P_M(X(t) = j) = m_j P_M(T > t). \quad (2.1)$$

Ferrari et al. (1994) mostraron que si tal M existe entonces

$$P_M(T > t) = e^{-\mu t} \quad \text{para algún } \mu \in (0, \infty),$$

y entonces la definición de distribución cuasiestacionaria coincide con la definición de medida μ -invariante:

Definición 2.6 Sea $\mu \geq 0$. Un conjunto $\{m_i, i \in C\}$ de números estrictamente positivos, tales que

$$\sum_{i \in C} m_i P_{ij}(t) \leq e^{-\mu t} m_j, \quad (2.2)$$

para toda $t \geq 0$ y toda $j \in C$, es llamado una medida μ -subinvariante para $P_{ij}(t)$ en C . Cuando la igualdad se cumple para toda $t \geq 0$ y toda $j \in C$, entonces $\{m_i, i \in C\}$ es llamada medida μ -invariante.

Nótese que no se pierde generalidad al suponer que $\tau_i \equiv P_i(T < \infty) = 1$, pues si $\tau_i < 1$, entonces se puede trabajar con el proceso restringido al evento de absorción; éste es el proceso cuyas probabilidades de transición son $\bar{p}_{ij}(t) = p_{ij}(t)\tau_j/\tau_i$ las cuales están bien definidas ya que por los supuestos anteriores es claro que $\tau_i > 0$ para toda i . Claramente,

$$P_i(X_{t_k} = j_k, k = 1, \dots, n \mid T < \infty) = \bar{p}_{i,j_1}(t_1) \prod_{k=1}^n \bar{p}_{j_{k-1},j_k}(t_k - t_{k-1}).$$

Uno de los principales cuestionamientos es ¿Qué garantiza la existencia de una distribución cuasiestacionaria? Haciendo referencia a dos Teoremas de Kingman (citados en Anderson, 1991, Sección 5.2) y algunas definiciones, se da respuesta a esta pregunta.

Definición 2.7 Sea $\mu \geq 0$. Un conjunto $\{x_i, i \in C\}$ de números estrictamente positivos, tales que

$$\sum_{j \in C} P_{ij}(t)x_j \leq e^{-\mu t}x_i \quad (2.3)$$

para toda $t \geq 0$ y toda $i \in C$, es llamado un vector μ -subinvariante para $P_{ij}(t)$ en C . Cuando la igualdad se cumple para toda $t \geq 0$ y toda $i \in C$, entonces $\{x_i, i \in C\}$ es llamado μ -invariante.

Teorema 2.8 (Kingman, 1963). Sea C una clase irreducible con parámetro de decaimiento $\lambda_c \geq 0$. Entonces existen medidas y vectores λ_c -subinvariantes para $P_{ij}(t)$ en C .

El estado $i \in C$ se dice que es λ_c -recurrente si

$$\int_0^{\infty} P_{ii}(t)e^{\lambda_c t} dt = +\infty,$$

y es λ_c -positivo si

$$\lim_{t \rightarrow \infty} e^{\lambda_c t} P_{ii}(t) > 0.$$

Uno de los principales Teoremas acerca de la existencia y unicidad de DCE's es el siguiente, donde se supone que C es una clase transitoria e irreducible.

Teorema 2.9 (Kingman, 1963). *Suponga que C tiene parámetro de decaimiento λ_c , y es λ_c -recurrente. Entonces la medida λ_c -subinvariante $\{m_i, i \in C\}$ y el vector λ_c -subinvariante $\{x_i, i \in C\}$ del teorema (2.8) son únicos y ambos son λ_c -invariantes. La clase C es λ_c -positiva si y sólo si*

$$\sum_{k \in C} m_k x_k < +\infty,$$

y entonces, cuando $t \rightarrow \infty$,

$$e^{\lambda_c t} P_{ij}(t) \rightarrow \frac{x_i m_j}{\sum_{k \in C} m_k x_k} \quad (2.4)$$

para toda $i, j \in C$.

La siguiente proposición puede ser usada para probar la existencia de ciertas DCE's, para lo cual definimos

$$\alpha_k = P \{X(t) \notin C \text{ para algún } t > 0 \mid X(0) = k\}, \quad k \in C.$$

Supondremos que $\alpha_k > 0$ para alguna $k \in C$.

Proposición 2.10 *Suponga que C es λ_c -positiva. Sea $\{m_k, k \in C\}$ y $\{x_k, k \in C\}$ la medida y el vector λ_c -invariantes, respectivamente. Suponemos que $\sum_{k \in C} m_k \alpha_k < +\infty$. Entonces*

$$\begin{aligned} \lim_{t \rightarrow \infty} P \{X(t) = j \mid X(0) = i, X(t) \in C, X(t+s) \notin C \text{ para alguna } s > 0\} \\ = \frac{m_j \alpha_j}{\sum_{k \in C} m_k \alpha_k}, \quad j \in C, \end{aligned}$$

y

$$\begin{aligned} \lim_{t \rightarrow \infty} \lim_{s \rightarrow \infty} P \{X(t) = j \mid X(0) = i, X(s+t) \in C, X(t+s+r) \notin C \text{ para algún } r > 0\} \\ = \frac{m_j x_j}{\sum_{k \in C} m_k x_k}, \quad j \in C. \end{aligned}$$

Es posible encontrar la distribución cuasiestacionaria cuando el proceso empieza de una distribución inicial arbitraria en lugar de un estado específico i . Por ejemplo, sea $\{\nu_i, i \in E\}$ una medida de probabilidad y definimos

$$P_{\nu j}(t) := \sum_{i \in E} \nu_i P_{ij}(t), \quad j \in E, \quad t \geq 0. \quad (2.5)$$

Sea C una clase transitoria para $P_{ij}(t)$, y definamos

$$\tilde{P}_{\nu j}(t) = \frac{P_{\nu j}(t)}{\sum_{k \in C} P_{\nu k}(t)} = P_{\nu} \{X(t) = j \mid X(t) \in C\}, \quad j \in C \quad (2.6)$$

donde P_{ν} denota la probabilidad dado que $X(0)$ tiene distribución ν .

Proposición 2.11 *Suponga que C es λ_c -positiva para $P_{ij}(t)$, y que $\{\nu_i, i \in E\}$ satisface*

$$\nu_i P_{ij}(t) = 0, \quad i \in E \setminus C, \quad j \in C.$$

Sean $\{m_k, k \in C\}$ y $\{x_k, k \in C\}$ la medida y vector λ_c -invariante, respectivamente, para $P_{ij}(t)$ en C , y supongamos que

- (i) $\sum_{k \in C} \nu_k x_k < +\infty$,
- (ii) $\text{Sup}_{k \in C} \nu_k / m_k < +\infty$.

Entonces

$$\lim_{t \rightarrow \infty} \tilde{P}_{\nu j}(t) = \begin{cases} \frac{m_j}{\sum_{k \in C} m_k} & \text{si } \sum_{k \in C} m_k < +\infty \\ 0 & \text{si } \sum_{k \in C} m_k = +\infty. \end{cases}$$

Uno de los resultados más usados en este estudio es el de Vere-Jones (citado en Pakes, 1995), quien demuestra que una DLC condicionada a la no absorción, es una DCE. En esta tesis, al hacer referencia a la DCE nos referiremos a la DLC y viceversa. En la terminología de Pakes (1995), la definición de DCE coincide con la de distribución límite condicional, que es también una DCE.

Definición 2.12 $M = \{M_j\}$ es una distribución límite condicional (DLC) de $\{X(t), t \geq 0\}$ si para cada $j \geq 1$,

$$M_j = \lim_{t \rightarrow \infty} P_i(X_t = j \mid T > t) \quad y \quad \sum M_j = 1, \quad (2.7)$$

siempre que el límite exista para alguna (y por tanto para toda) i .

2.2 Semigrupos de transición en ℓ^1

Sea $S = \{0, 1, \dots\}$ y sea ℓ^1 el espacio de sucesiones absolutamente sumables $x = (\xi_i)_{i \in S}$. Equipado con la norma $\|x\| = \sum_{i \in S} |\xi_i|$, ℓ^1 es un espacio de Banach. Las sucesiones de términos no negativos $x = (\xi_i)_{i \in S} \in \ell^1$, tales que $\|x\| = \|(\xi_i)_{i \in S}\| = \sum_{i \in S} \xi_i = 1$, serán llamadas distribuciones, y el conjunto de distribuciones será denotado por \mathcal{D}_1 .

Definición 2.13 Una familia $\{P(t), t \geq 0\}$ de operadores lineales acotados $P(t) : \ell^1 \rightarrow \ell^1$ es un semigrupo si y sólo si cumple

$$P(0) = I, \quad P(t+s) = P(t)P(s), \quad t, s \geq 0.$$

Definición 2.14 Un operador lineal acotado P en ℓ^1 es llamado positivo si mapea sucesiones no negativas en sucesiones no negativas. P es llamado operador de transición u operador de Markov si para toda sucesión no negativa $\xi \in \ell^1$, se cumple $\|P(\xi)\| = \|\xi\|$.

A cada cadena de Markov homogénea $\{X(t), t \geq 0\}$ en S le corresponde un semigrupo de operadores de transición $\left\{P(t) := [P_{ij}(t)]_{i,j \in S}, t \geq 0\right\}$, donde

$$P_{ij}(t) = P[X(t+s) = j \mid X(s) = i], \quad i, j \in S.$$

Debido a la propiedad de Markov, $P_{ij}(t)$ satisface

(a) $P_{ij}(t) \geq 0$,

(b) $\sum_{j \in S} P_{ij}(t) = 1, i, j \in S$

(c) $P_{ik}(t+s) = \sum_{j \in S} P_{ij}(s) P_{jk}(t)$ para $t, s \geq 0$ (ecuación de Chapman-Kolmogorov).

Se supondrá además que

(d) $\lim_{t \rightarrow 0^+} P_{ij}(t) = \delta_{i,j}$, donde $\delta_{i,j} = 1$ si $i = j$, y 0 en otro caso.

Debido a que $P(t) = [P_{ij}(t)]_{i,j \in S}$, la propiedad (c) puede escribirse compactamente en notación matricial como

$$P(t+s) = P(t)P(s), \quad t, s \geq 0. \tag{2.8}$$

La propiedad (d) establece que $P(t)$ es continua en $t = 0$, y de (2.8) se sigue que $P(0) = I$ (matriz identidad). También de (2.8) se deduce que $P(t)$ es continua para toda $t > 0$. De hecho poniendo $s = h > 0$ en (2.8) resulta, debido a (d),

$$\lim_{h \rightarrow 0^+} P(t+h) = P(t) \lim_{h \rightarrow 0^+} P(h) = P(t)I = P(t). \tag{2.9}$$

Por otro lado, para $t > 0$ y $0 < h < t$ se escribe (2.8) en la forma

$$P(t) = P(t-h)P(h).$$

Como $P(h)$ es cercana a la identidad I cuando h es suficientemente pequeña, e I es invertible, se sigue que $P(h)^{-1}$ existe y también se aproxima a la identidad I cuando

$h \downarrow 0$ Por tanto,

$$P(t) = P(t) \lim_{h \rightarrow 0^+} (P(h))^{-1} = \lim_{h \rightarrow 0^+} P(t+h) \quad (2.10)$$

Las igualdades (2.9) y (2.10) muestran que $P(t)$ es continua en t . Se sigue que la familia de operadores $\mathcal{P} = \{P(t), t \geq 0\}$ satisface la condición

$$\lim_{t \rightarrow 0^+} P(t)x = P(0)x = x, \quad x \in \ell^1,$$

donde el último límite es tomado en el sentido de la norma en ℓ^1 . Un resultado estándar de la teoría de semigrupos de operadores dice que esta condición implica que \mathcal{P} es una familia de operadores fuertemente continuos en ℓ^1 , i.e., que

$$\lim_{t \rightarrow t_0} P(t)x = P(t_0)x, \quad \text{para toda } x \in \ell^1 \text{ y } t_0 > 0.$$

Si x es la distribución inicial de una cadena de Markov con semigrupo de transición $\{P(t), t \geq 0\}$, entonces $x \cdot P(t)$ es la distribución de la cadena al tiempo t . Aquí, “ \cdot ” denota multiplicación matricial y se usará la notación $P(t)x := x \cdot P(t)$, donde $P(t)$ es vista como un operador en ℓ^1 . Por la definición 2.14 se tiene que $P(t)$ es una contracción para cada $t \geq 0$, i.e.,

$$\|P(t)\|_{\mathcal{L}(\ell^1)} = \sup_{\|x\|=1} \|P(t)x\| \leq 1$$

2.3 Producto tensorial y estabilidad asintótica

El objetivo de esta subsección es dar las herramientas matemáticas para describir las transiciones de una cadena de Markov bivariada.

Definición 2.15 Sea ℓ^1 un espacio de Banach como se definió anteriormente. Sean $x = (\xi_i)_{i \in S} \in \ell^1$ y $y = (\eta_j)_{j \in S} \in \ell^1$. El producto tensorial es denotado por

$$x \otimes y = (\xi_i \eta_j)_{i,j \in S}$$

$x \otimes y \in \ell^1 \otimes \ell^1$. Dados dos operadores lineales acotados A y B , actuando en el espacio ℓ^1 , se define su producto tensorial $A \otimes B$ como sigue: Para todo $x \otimes y \in \ell^1 \otimes \ell^1$,

$$(A \otimes B)(x \otimes y) = Ax \otimes By.$$

En particular, si $\{T(t), t \geq 0\}$ y $\{S(t), t \geq 0\}$ son semigrupos fuertemente continuos, de contracción, actuando en espacios de Banach ℓ^1 entonces $T(t) \otimes S(t)$ es un semigrupo fuertemente continuo, de contracción, actuando en $\ell^1 \otimes \ell^1$. Más aún, su generador E es igual a $A \otimes I + I \otimes B$, donde A, B son los generadores de $\{T(t), t \geq 0\}$ y $\{S(t), t \geq 0\}$ respectivamente. Para ser más específicos, si $x \in D(A)$ y $y \in D(B)$, entonces $(x \otimes y) \in D(E)$ y

$$E(x \otimes y) = (A \otimes I + I \otimes B)(x \otimes y) = Ax \otimes y + x \otimes By.$$

Definición 2.16 (Bobrowski, et al., 2001). Un semigrupo de operadores de transición $\{P(t), t \geq 0\}$, en ℓ^1 es llamado asintóticamente estable si para todas las distribuciones $(\pi_i)_{i \in S} \in \ell^1$, existe un límite $\lim_{t \rightarrow \infty} P(t)(\pi_i)_{i \in S}$. Si este límite no depende de la selección de $(\pi_i)_{i \in S}$, se dice que $\{P(t), t \geq 0\}$ es ergódico.

2.4 Absorción para una cadena de Markov simple

Recordemos que S es el conjunto $\{0, 1, \dots\}$ y $X = \{X(t), t \geq 0\}$ es una cadena de Markov de tiempo continuo con espacio de estados S que describe la longitud de un

microsatélite a través del proceso de mutaciones. $P_{ij}(t) = P[X(t) = j \mid X(0) = i]$ son las probabilidades de transición de la cadena X y la matriz de probabilidades de transición será denotada por $P(t)$ para $t \geq 0$. El generador infinitesimal de $\{P(t), t \geq 0\}$ será denotado por $Q = (q_{i,j})_{i,j \in S}$. Nótese que $\{P(t), t \geq 0\}$ son operadores de transición relativos al proceso X .

Se define $e_j = (\delta_{ij})_{i \in S} \in \ell^1$ para $j \in S$, donde δ_{ij} es el símbolo de Kronecker, así como el operador lineal $A \in \mathcal{L}(\ell^1)$, i.e., $A : \ell^1 \rightarrow \ell^1$, dado por $Ax = x - \xi_0 e_0$. En otras palabras, si $x = (\xi_0, \xi_1, \xi_2, \dots)$, entonces $Ax = (0, \xi_1, \xi_2, \dots)$.

Por la teoría de DCE's presentada en el Capítulo 2, podemos suponer, al igual que Bobrowski (2004), que existe un $x_0 \in \mathcal{D}_1$ y un número positivo γ tales que, para toda $x \in \mathcal{D}_1$, existe una constante no negativa $a(x)$ que cumple

$$\lim_{t \rightarrow \infty} e^{\gamma t} AP(t)x = a(x)x_0, \quad (2.11)$$

donde $a(x)$ es cero si y sólo si $x = e_0$. Esto implica que

$$\lim_{t \rightarrow \infty} P(t)x = e_0, \quad x \in \mathcal{D}_1, \quad (2.12)$$

y en particular que $P(t)e_0 = e_0$. Se interpreta (2.12) como que 0 es un estado absorbente para la cadena de Markov X , y (2.11) se interpreta como que x_0 es la DCE de la cadena condicionada a la no absorción. Estas últimas afirmaciones se demuestran directamente de la definición 2.5. Para ser más específicos, en la terminología estándar usada por Pakes (1995), x_0 es una distribución límite condicional. En dicha terminología se dice que x_0 es una DCE de $\{X(t), t \geq 0\}$ si $AP(t)x_0 = x_0 e^{-\gamma t}$ para algún $\gamma > 0$. Entonces

$$AP(t)AP(s) = AP(t+s), \quad t, s \geq 0. \quad (2.13)$$

Pues

$$\begin{aligned} [AP(t)AP(s)]x_0 &= x_0e^{-\gamma t}e^{-\gamma s} = x_0e^{-\gamma(t+s)}, \\ AP(t+s)x_0 &= x_0e^{-\gamma(t+s)}, \quad t, s \geq 0. \end{aligned}$$

Observación 2.17 *Note que γ es el parámetro de decaimiento del semigrupo de transición $\{P(t), t \geq 0\}$. La elección de la función exponencial en (2.11) es dictada por el siguiente resultado, el cual aparece en varias formas en la literatura sobre distribuciones cuasi-estacionarias (Ver Ferrari, et al., 1995 y Pakes, 1995). Sea $v(\cdot)$ una función positiva tal que $\lim_{t \rightarrow \infty} v(t) = \infty$ y supóngase que existe $x_0 \in \mathcal{D}_1$ tal que*

$$\lim_{t \rightarrow \infty} v(t)AP(t)x = a(x)x_0, \quad x \in \mathcal{D}_1,$$

donde $a(x)$ es cero si y sólo si $x = e_0$. Entonces (2.11) se cumple para algún $\gamma > 0$.

2.5 Dos tipos de absorción de cadenas de Markov

Sea \mathcal{M} el espacio de matrices $m = (\mu_{i,j})_{i,j \in S}$, tales que

$$\sum_{i,j \in S} |\mu_{i,j}| < \infty,$$

donde $m \in \ell^1 \otimes \ell^1$. El subconjunto de \mathcal{M} constituido por distribuciones, i.e., por matrices

$(\mu_{i,j})_{i,j \in S}$ que satisfacen

$$\mu_{i,j} \geq 0 \text{ y } \sum_{i,j \in S} \mu_{i,j} = 1,$$

es denotado por \mathcal{D}_2 . Las columnas $(\mu_{i,\cdot})_{i \in S}$ pertenecen a ℓ^1 . En esta notación, para $x =$

$(\xi_i)_{i \in S} \in \ell^1$ e $y = (\eta_i)_{i \in S} \in \ell^1$, se tiene $x \otimes y = (\xi_i \eta_j)_{i,j \in S} \in \mathcal{M}$.

Definición 2.18 Sea $\{P(t), t \geq 0\}$ semigrupo de transición con generador Q . Se define el producto tensor $\{P(t) \otimes P(t), t \geq 0\} = \{T(t), t \geq 0\} =: \mathcal{T}$, el cual es un semigrupo de operadores que actúa en $\ell^1 \otimes \ell^1$.

La acción de \mathcal{T} es descrita por

$$[P(t) \otimes P(t)]m = P^*(t)mP(t), \quad m \in \ell^1 \otimes \ell^1,$$

donde $*$ denota la transpuesta de una matriz. Si se representa una matriz m como el vector de sus columnas m_i , es decir, $m = (m_1, \dots, m_n, \dots)$, entonces

$$[P(t) \otimes P(t)]m = \sum_{i=1}^{\infty} P(t)m_i \otimes P(t)e_i; \quad (2.14)$$

además, cualquier $m \in \mathcal{M}$ puede ser representada como

$$m = \sum_{i,j \in S} \mu_{i,j} e_i \otimes e_j.$$

Entonces (2.14) es equivalente a

$$T(t)m = \sum_{i,j \in S} \mu_{i,j} P(t)e_i \otimes P(t)e_j.$$

El operador $T(t)$ mapea \mathcal{D}_2 en \mathcal{D}_2 . Mientras que $P(t)x$ describe la evolución de la distribución de una cadena de Markov con distribución inicial $x \in \mathcal{D}_1$, $T(t)m$ describe la evolución de dos cadenas de Markov independientes $(X_1(t), X_2(t))$, que al tiempo $t = 0$ fueron distribuidas de acuerdo a una matriz de probabilidad conjunta $m \in \mathcal{D}_2$. En el contexto de genética poblacional, para un par de microsatélites de una población de $2N(t)$ individuos haploides, los procesos $\{X_1(t), t \geq 0\}$ y $\{X_2(t), t \geq 0\}$ describen la longitud de cada microsatélite respectivamente que varía de acuerdo al proceso de mutaciones, los

cuales son descritos con cadenas de Markov de tiempo continuo con espacio de estados $S = \{0, 1, \dots\}$, donde cada $i, j \in S$ representa la longitud del microsátélite.

Sea \mathcal{M}_s el conjunto de matrices simétricas $m = (\mu_{i,j})$, i.e., tales que $\mu_{i,j} = \mu_{j,i}$. Nótese que \mathcal{M}_s es un subespacio de \mathcal{M} , y que los operadores $T(t)$ mapean \mathcal{M}_s en sí mismo porque, para $m \in \mathcal{M}_s$,

$$T(t)m = \sum_i \mu_{i,i} P(t)e_i \otimes P(t)e_i + \sum_{i < j} \mu_{i,j} [P(t)e_i \otimes P(t)e_j + P(t)e_j \otimes P(t)e_i] \quad (2.15)$$

y ambos sumandos $P(t)e_i \otimes P(t)e_i$ y $P(t)e_i \otimes P(t)e_j + P(t)e_j \otimes P(t)e_i$ pertenecen a \mathcal{M}_s .

La relación (2.12) implica que

$$\lim_{t \rightarrow \infty} T(t)m = e_0 \otimes e_0, \quad m \in \mathcal{D}_2,$$

i.e., el semigrupo $T(t)$ es asintóticamente estable y ergódico (ver definiciones 2.16 y 2.14), como se muestra en la siguiente proposición.

Proposición 2.19 *El semigrupo $\{T(t), t \geq 0\}$ es un semigrupo de Markov. Es asintóticamente estable o ergódico si $\{P(t), t \geq 0\}$ lo es.*

Demostración. Por la definición 2.18 $\{T(t), t \geq 0\}$ es no negativo, y por (2.14) para cualquier m no negativa se tiene que

$$\begin{aligned} \| T(t)m \| &= \left\| \sum_{i=1}^{\infty} P(t)m_i \otimes P(t)e_i \right\| = \sum_{i=1}^{\infty} \| P(t)m_i \otimes P(t)e_i \| \\ &= \sum_{i=1}^{\infty} \| P(t)m_i \| \| P(t)e_i \| = \sum_{i=1}^{\infty} \| P(t)m_i \| \\ &= \sum_{i=1}^{\infty} \| m_i \| = \| m \|, \end{aligned} \quad (2.16)$$

donde m_i es la i -ésima columna de m . Si $P(t)$ es asintóticamente estable, tenemos

$$\lim_{t \rightarrow \infty} T(t)m = \lim_{t \rightarrow \infty} \sum_{i=1}^{\infty} P(t)m_i \otimes P(t)e_i = \sum_{i=1}^{\infty} P m_i \otimes P e_i = (P \otimes P)m, \quad (2.17)$$

donde $P\pi = \lim_{t \rightarrow \infty} P(t)\pi$ debido al teorema de Scheffe (Billingsley, 1986). Si $\{P(t), t > 0\}$ es ergódico, entonces $P\pi = e_0$ no depende de la distribución π y

$$\lim_{t \rightarrow \infty} T(t)m = \sum_{i=1}^{\infty} \|m_i\| \pi_0 \otimes \pi_0,$$

como se deseaba. ■

Por tanto $\|T(t)\|_{\mathcal{L}(\mathcal{M})} \leq 1$.

Se introducen dos operadores en \mathcal{M}_s :

$$\begin{aligned} B &= (I - A) \otimes A + A \otimes I, \\ C &= A \otimes A. \end{aligned}$$

Para $m \in \mathcal{D}_2 \cap \mathcal{M}_s$, Bm describe la distribución conjunta inicial m del proceso $(X_1(t), X_2(t))$ fuera del origen $(0,0)$ y Cm describe la distribución de m fuera del margen $M = \{(i,j) \mid i = 0 \vee j = 0\}$ donde una de las dos cadenas de Markov es absorbida., es decir, aplicando el operador B a una matriz m se obtiene la matriz m pero con la primer entrada igual a cero, mientras que si le aplicamos el operador C se obtiene la matriz m con elementos en la primera columna y la primera fila igual a 0.

Lema 2.20 *El supuesto (2.11) implica que, para $m \in \mathcal{M}_s$, existen constantes no negativas $b(m)$ y $c(m)$ tales que*

$$\lim_{t \rightarrow \infty} e^{\gamma t} B T(t)m = b(m) [x_0 \otimes e_0 + e_0 \otimes x_0], \quad (2.18)$$

$$\lim_{t \rightarrow \infty} e^{2\gamma t} C T(t)m = c(m) x_0 \otimes x_0, \quad (2.19)$$

donde $b(m) = 0$ si y sólo si $Bm = 0$ y $c(m) = 0$ si y sólo si $Cm = 0$.

Demostración. Tomando $m = e_i \otimes e_j$, $T(t) = P(t) \otimes P(t)$ y $B = (I - A) \otimes A + A \otimes I$, tenemos que

$$e^{\gamma t} BT(t)m = (I - A)P(t)e_i \otimes e^{\gamma t} AP(t)e_j + e^{\gamma t} AP(t)e_i \otimes P(t)e_j,$$

donde, debido a (2.11), $e^{\gamma t} BT(t)m$ tiende a $a(e_j)e_0 \otimes x_0 + a(e_i)x_0 \otimes e_0$. Por tanto, (2.18) se sigue de (2.15) con

$$b(m) = \sum_{i \in S} \mu_{i,i} a(e_i) + \sum_{i < j, i \in S, j \in S} \mu_{i,j} [a(e_j) + a(e_i)].$$

La prueba de (2.19) es análoga. ■

En otras palabras, $b(m) [x_0 \otimes e_0 + e_0 \otimes x_0]$ es una distribución límite condicionada a la no absorción en el origen $(0, 0)$, y $x_0 \otimes x_0$ es una distribución límite condicionada a la no absorción en el margen M . Dado que $T(t)m$ es la distribución de las dos cadenas de Markov que al tiempo $t = 0$ fueron distribuidas de acuerdo con una matriz m de probabilidad conjunta, las distribuciones límite condicionales del Lema 2.20 pertenecen a un modelo sin deriva, i.e., el modelo sin deriva genética se refiere al modelo donde sólo consideramos la fuerza genética de la mutación en las probabilidades de transición en el cambio de longitud de cada microsatélite.

Finalizamos esta sección con el siguiente corolario el cual se usará en secciones posteriores.

Corolario 2.21 *Existe una constante $L > 0$ tal que*

$$\| e^{\gamma u} AP(u) \| \leq L, \quad \| e^{\gamma u} BT(u) \| \leq L, \quad \| e^{2\gamma u} CT(u) \| \leq L \quad (2.20)$$

para toda u no negativa.

Demostración . Las desigualdades de (2.20) se siguen directamente del principio de acotación uniforme. ■

Capítulo 3

Comportamiento asintótico de

$$(X_1(t), X_2(t))$$

En este Capítulo se da una formulación del coalescente correspondiente al modelo de Wright-Fisher-Moran, además se da la forma de la distribución conjunta del proceso $(X_1(t), X_2(t))$ bajo mutación y deriva genética en términos de los operadores de transición $P(t), T(t)$ y $R(t)$, donde $R(t) : \mathcal{M}_s \rightarrow \mathcal{M}_s$. También se demuestran Teoremas concernientes al comportamiento asintótico de la distribución conjunta de los tiempos σ y τ , donde σ es la cantidad de tiempo en que los dos microsátélites permanecen idénticos y τ es la cantidad de tiempo durante el cual los dos microsátélites evolucionan independientemente. Se presentan además, en la Sección 3.2, simulaciones de la evolución de dos alelos en una población con diferente comportamiento asintótico de $N(t)$.

3.1 Cadenas de Markov de mutaciones y deriva genética en poblaciones de tamaño variable

3.1.1 Modelo coalescente de Wright-Fisher-Moran

En el modelo de tiempo continuo de Moran se supone que la población está compuesta por un número constante de $2N$ individuos haploides. Cada individuo está sujeto a eventos de nacimiento y muerte de acuerdo a un proceso de Poisson con intensidad 1. En el momento en que ocurre un evento de nacimiento y muerte, un genotipo para el nuevo individuo es muestreado con remplazo de los $2N$ cromosomas presentes en ese momento.

En la formulación del modelo de Moran se piensa en términos de $2N$ individuos haploides, que es uno de los factores que diferencian al modelo de Moran del de Wright-Fisher. Más precisamente:

- Cada individuo es reemplazado con tasa 1. Es decir, el individuo x vive por una cantidad exponencialmente distribuida con media 1 y entonces es “reemplazado”.
- Para reemplazar al individuo x , se escoge aleatoriamente del conjunto de individuos (incluyendo a x).
- Un alelo a que es escogido muta a un alelo A con probabilidad u . Un alelo A que es escogido muta a un alelo a con probabilidad v .
- El nuevo alelo, posiblemente mutado, reemplaza al viejo.

A continuación se presenta una formulación del coalescente, correspondiente al modelo de Wright-Fisher-Moran para un par de individuos de una población de $2N$ individ-

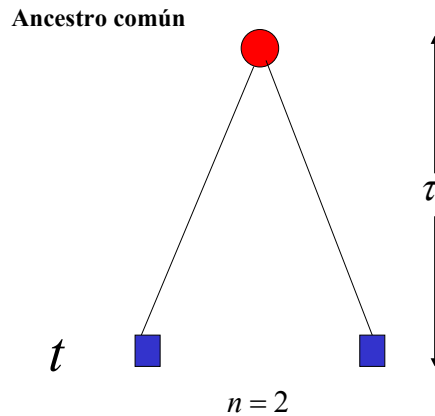


Figura 3.1:

uos haploides bajo deriva genética y mutaciones, las cuales siguen una cadena de Markov general de tiempo continuo (Bobrowski et al., 2001):

- *Coalescente con longitudes de rama independientes con distribución exponencial de parámetro $\frac{1}{2N}$.* El proceso coalescente puede verse como un árbol ancestral. La coalescencia de los dos individuos o cromosomas, es decir el ancestro común, es representado por el nodo del árbol; las ramas, en este caso dos, corresponden al tiempo que transcurre a partir del tiempo presente t hasta un tiempo τ en el pasado, en que ocurre el evento de coalescencia. Estos tiempos son independientes en cada cromosoma. La interpretación es que para cualesquiera dos individuos de la población, el tiempo de su ancestro común es una variable aleatoria τ con distribución exponencial de parámetro $\frac{1}{2N}$, como se muestra en la Figura 3.1. Esto se sigue de la Ecuación (1.9) con $2N(t)$

constante

$$\begin{aligned} g_2(t) &= \frac{1}{2N} \exp \left\{ - \int_0^t \frac{du}{2N} \right\} \\ &= \frac{1}{2N} \exp \left\{ - \frac{t}{2N} \right\}, \end{aligned}$$

donde $g_2(t)$ es la función de densidad del tiempo de coalescencia entre dos cromosomas.

- *El modelo de Markov de mutaciones con probabilidades de transición $P_{ij}(t)$ e intensidades Q_{ij} . Sean $\{X_1(t), t \geq 0\}$ y $\{X_2(t), t \geq 0\}$ dos procesos como se describen en la sección 2.5. Sean*

$$P_{ij}(t) = P[X_1(t) = j \mid X_1(0) = i]$$

las probabilidades de transición, es decir, si el alelo de un individuo es i al tiempo $t = 0$, entonces su alelo al tiempo t (o el alelo de su descendiente al tiempo t) es igual a j con probabilidad $P_{ij}(t)$. $Q := [Q_{ij}]$ es la matriz de intensidades de transición que cumple

$$Q_{ij} \geq 0, \quad i \neq j, \quad \sum_j Q_{ij} = 0, \quad \forall i.$$

En este escrito se usa el modelo de coalescente con deriva genética modificado con tamaño de población $N = N(t)$ dependiente del tiempo. A saber:

- El tiempo T_a (medido desde el tiempo presente t hacia atrás) al ancestro común de los dos individuos de la muestra tomada al tiempo t es una variable aleatoria tal que $\Pr[T_a > \varrho] = \exp \left\{ - \int_0^\varrho [2N(t-u)]^{-1} du \right\}$ (ver Sección 1.2).
- El modelo de mutación permanece sin cambio, i.e., tiene la forma de la cadena de Markov de tiempo continuo con probabilidades de transición $P_{ij}(t)$ e intensidades Q_{ij} .

Observación 3.1 *En el modelo de mutaciones Stepwise Mutation Model (SMM) con tamaño de población variable, se supone que un alelo en un locus considerado, está sujeto a mutaciones (que siguen un proceso de Poisson de intensidad ν que es directamente proporcional al tamaño de la población) las cuales reemplazan un alelo de tamaño X por un alelo de tamaño $X + U$, donde U es una variable aleatoria entero-valuada con distribución dada por $(\rho_n)_{n \in \mathbb{Z}}$, donde \mathbb{Z} es el conjunto de los enteros. La modelación de los procesos evolutivos de distribuciones de frecuencias alélicas, en su mayoría, están basados o son desarrollados del análisis de modelos de caminatas aleatorias introducidos por Moran (Sibly, et al., 2001) y son conocidos como modelos de mutaciones de salto (SMM) porque se supone que la longitud de un microsatélite sólo cambia en ± 1 en cada generación, debido a las mutaciones.*

3.1.2 Simulación de la evolución de dos alelos bajo el modelo de Wright-Fisher-Moran

El objetivo de esta Sección es mostrar gráficamente cuáles son los efectos de la deriva genética para la evolución de dos alelos, que evolucionan de manera independiente bajo el modelo de Wright-Fisher-Moran (ver Secciones 1.1.2, 1.1.3 y 3.1.1) con tamaño de población variable. En cada gráfica presentada a continuación, cada trayectoria representa la evolución de un tipo de alelo en una población bajo el efecto de la deriva genética aleatoria, donde si p es la probabilidad del alelo que aparece en las simulaciones, $1 - p$ es la probabilidad del otro alelo, inicialmente $p = 0.5$ y va cambiando debido a la deriva genética aleatoria y al mecanismo de reproducción. En cada gráfica se simulan 10 poblaciones con la finalidad de observar tendencias en la evolución del alelo. El tiempo está medido en N generaciones, y para fines ilustrativos, se tomarán en cuenta poblaciones con tamaño inicial entre 18 y

10000 y un máximo de 250 generaciones. La descripción gráfica de la evolución de estos dos cromosomas bajo diferente comportamiento límite de $N(t)$, se muestran en las siguientes Secciones.

Deriva genética con población decreciente

Cuando $\lim_{t \rightarrow \infty} 2N(t) = 0$, es decir, cuando la población decrece, o tiende a cero (en este caso de manera lineal), el efecto de la deriva genética es el mismo que en el caso de población constante: la fijación del alelo es más rápida en poblaciones pequeñas ($N(t)$ inicial igual a 18) con la diferencia que el tiempo esperado de fijación se reduce dependiendo del tamaño inicial de la población y del parámetro de decaimiento. En las gráficas de la Figura 3.2 se muestra, para las poblaciones iniciales de 100 con parámetros de decaimiento de 0.02 y 0.04, que los tiempos de fijación son aproximadamente $50N$ generaciones y $25N$ generaciones, respectivamente.

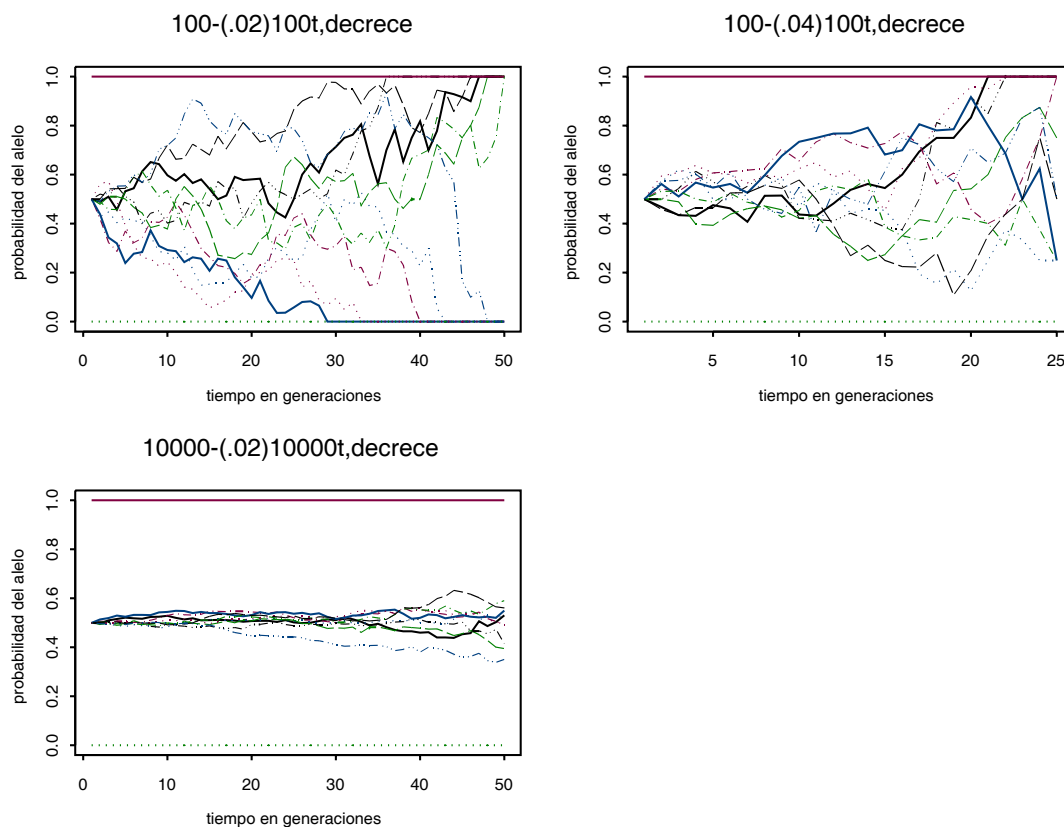


Figura 3.2.

Deriva genética con población constante

Considerando el caso cuando, en el límite $\lim_{t \rightarrow \infty} 2N(t) = N$, $0 \leq N < \infty$, i.e., la población tiende a ser constante en el límite, se observa en la Figura 3.3 que la deriva genética actúa más rápidamente en poblaciones pequeñas, es decir, un alelo tiende a fijarse o perderse en la población más rápido que en las poblaciones grandes; además, el tiempo esperado de fijación es de aproximadamente $2N$ generaciones (recordando que N es una población aloploide).

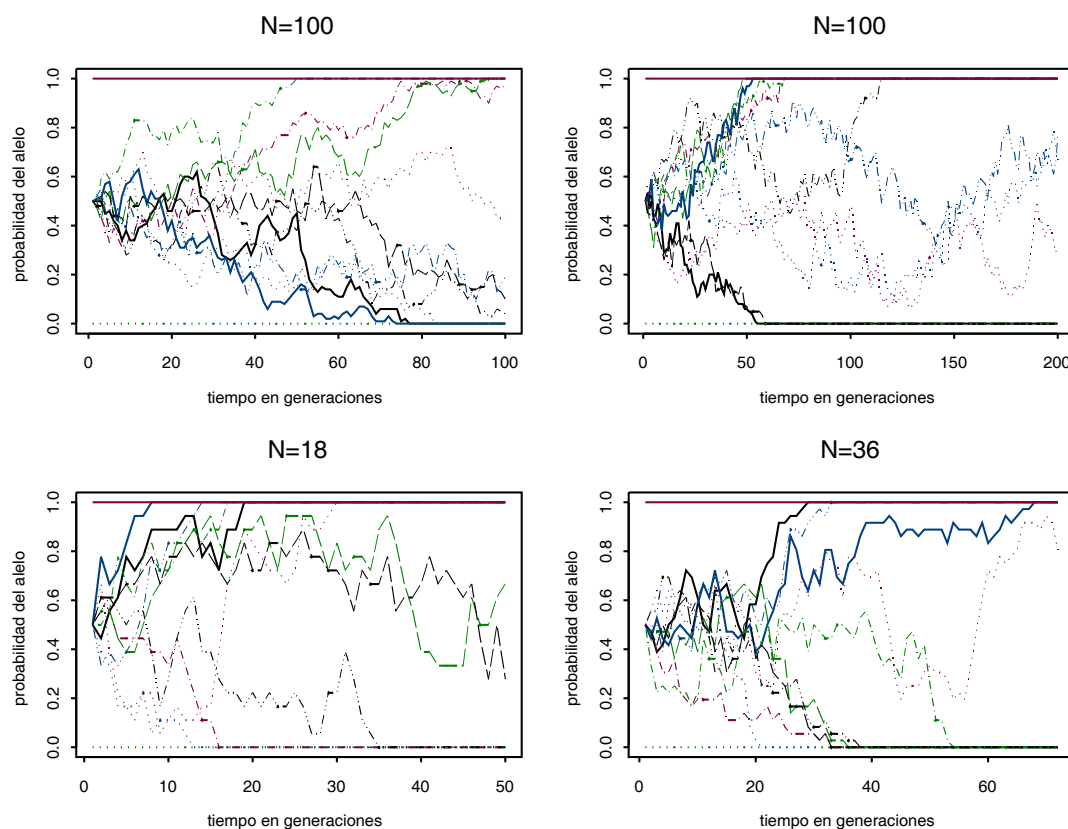


Figura 3.3

Deriva genética con población que crece linealmente

Si $\lim_{t \rightarrow \infty} 2N(t) = \infty$, y $\int_s^\infty \frac{dt}{2N(t)} = \infty$, es decir, cuando la población crece de manera lenta, en este caso linealmente, la deriva genética actúa de manera lenta en comparación con el caso de población constante y el caso cuando la población decrece. Contrario a lo que esperaríamos, aquí se ve claramente el efecto de la deriva genética. Suponemos que entre mayor sea el tamaño de la población los efectos de la deriva genética van disminuyendo y en este caso vemos que si la población crece a infinito pero de manera lenta (lineal) la deriva genética actúa de manera similar al caso de población constante sólo que no se observa

alguna tendencia para el tiempo esperado de fijación. Probablemente se requerirá de un número muy grande de generaciones para obtener la fijación del alelo (ver Figura 3.4).

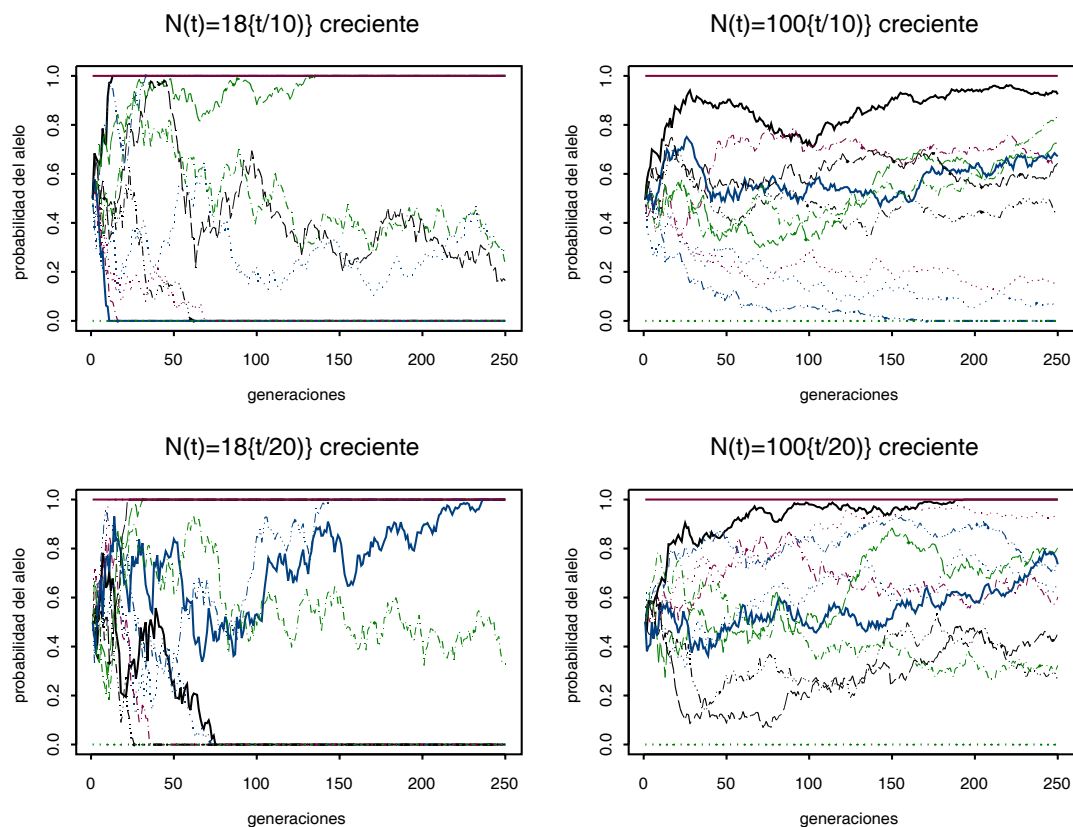


Figura 3.4

Deriva genética con población que crece exponencialmente

Cuando $\lim_{t \rightarrow \infty} 2N(t) = \infty$, y $\int_s^\infty \frac{dt}{2N(t)} < \infty$, i.e., la población crece a infinito de manera rápida (en este caso exponencialmente), a partir de una cierta generación tienden a estabilizarse las proporciones de los dos alelos en la población, conforme al principio de Hardy-Weinberg, i.e., la población considerada es en tamaño infinito por lo que los cambios aleatorios en frecuencias alélicas no son posibles y conforme a este principio la proporción de

cada uno de los genotipos de los individuos de una población tienden a permanecer estables (ver Hartl y Clark, 1997) y la deriva genética no tiene ningún efecto. En la Figura 3.5 se observa que la variación de la probabilidad del alelo es inversamente proporcional al tamaño inicial de la población (con igual parámetro de crecimiento), por otro lado, en poblaciones de igual tamaño y diferente parámetro de crecimiento se observa que a mayor parámetro de crecimiento, existe menor variabilidad en la probabilidad del alelo. Se observa, además, que el tiempo en que actúa la deriva genética es muy corto, mientras que el tiempo en que opera el principio de Hardy-Weinberg tiende a ser infinito.

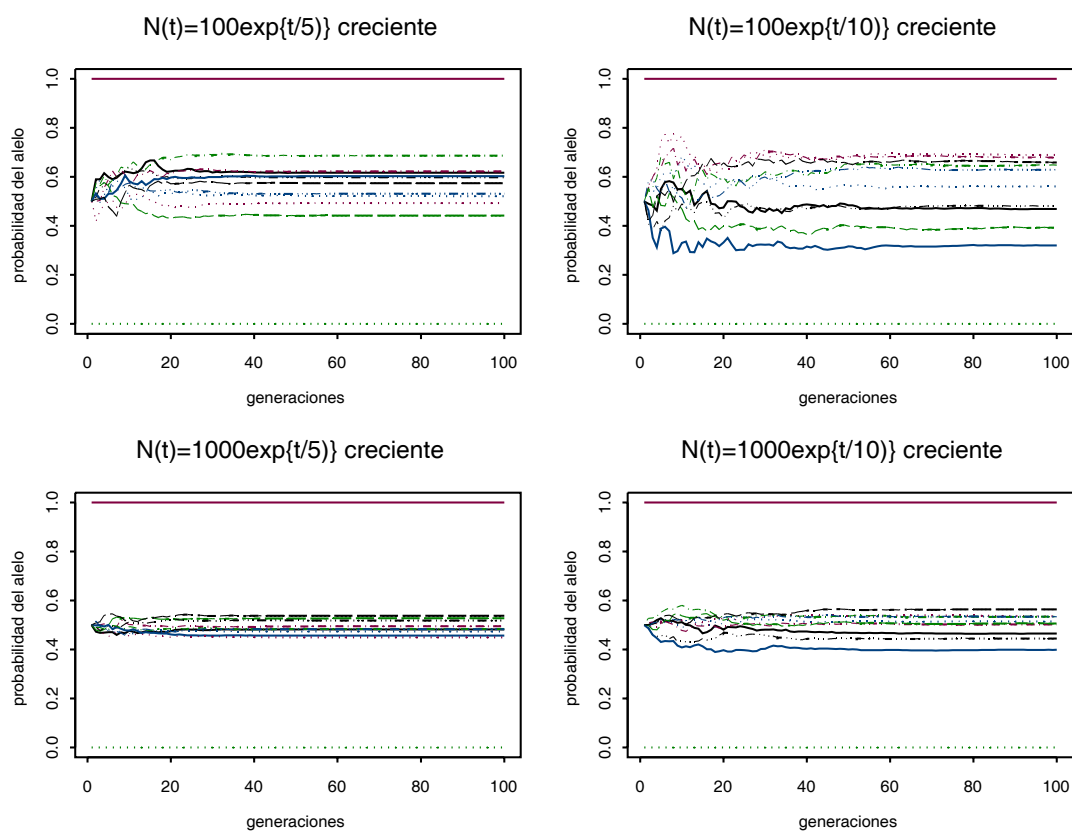


Figura 3.5

3.1.3 Distribución conjunta del proceso con operadores $R(t, s)$ en \mathcal{M}_s y semigrupos de transición

Aunque las mutaciones en dos cromosomas ocurren de forma independiente, los procesos acumulativos de mutaciones en esos cromosomas están relacionados debido al hecho de que los cromosomas pueden tener un ancestro común. Supóngase que a partir de un instante $t_0 \geq 0$, la evolución del tamaño de la población es conocida con las características que se dan en la siguiente hipótesis.

Hipótesis 3.2 *El tamaño $N(t)$ de la población al tiempo $t \geq t_0$, es es una función continua de t , la cual es estrictamente positiva en cada intervalo acotado, es decir,*

$$N(t) : [t_0, \infty) \rightarrow \mathbb{R}^+,$$

$$0 < \inf_{t_0 \leq t \leq t_0+h} N(t) \leq \sup_{t_0 \leq t \leq t_0+h} N(t) < \infty, \text{ para toda } h > 0.$$

Cuando el tamaño de la población $N(t)$ cumple la Hipótesis 3.2, el tiempo de coalescencia de un par de cromosomas (Sección 1.2) es una variable aleatoria impropia $\tau_2 \geq 0$ tal que

$$\Pr(\tau_2 \in B) = \int_B g_2(t, t-s) ds,$$

para todo subconjunto de Borel B de $[0, t-t_0)$. Por (1.9),

$$g_2(t, s) = \frac{1}{2N(s)} \exp \left\{ - \int_{t_0}^t \frac{1}{2N(u)} du \right\}. \quad (3.1)$$

En particular, la probabilidad de no coalescencia en el intervalo $[t_0, t)$ es

$$p_2(t) = 1 - \int_{t_0}^t g_2(t, u) du = \exp \left\{ - \int_{t_0}^t \frac{1}{2N(u)} du \right\}. \quad (3.2)$$

Sea $R_{jk}(t) = \Pr[X_1(t) = j, X_2(t) = k]$, la distribución conjunta de $(X_1(t), X_2(t))$ para $t \geq 0$ los cuales representan los procesos que determinan la longitud de los microsátélites seleccionados aleatoriamente al tiempo t . Si el ancestro común de X_1 y X_2 fue un alelo tipo i y existió hace τ unidades de tiempo, i.e.

$$X_1(t - \tau) = X_2(t - \tau) = i,$$

entonces

$$R_{jk}(t) = P_{ij}(\tau)P_{ik}(\tau), \quad t \geq \tau,$$

como se mostrará más abajo. El tipo de alelo del ancestro común en el tiempo $t - \tau$ es el estado de la cadena de Markov que modela el proceso de mutaciones, el cual es igual a i con probabilidad $\pi_i(t) = \Pr[X_1(t - \tau) = i] = \Pr[X_2(t - \tau) = i]$. Recordando que T_a es el tiempo de coalescencia de los dos cromosomas, medido del tiempo presente hacia atrás se sigue usando la propiedad de Markov fuerte que

$$\begin{aligned} R_{jk}(t) &= \int_0^\infty \sum_i \Pr[X_1(t) = j, X_2(t) = k \mid X_1(t - \tau) = i, T_a = \tau] \times \\ &\quad \Pr[X_1(t - \tau) = i] f_{T_a}(\tau) d\tau \\ &= \int_0^\infty \sum_i \pi_i(t - \tau) \Pr[X_1(t) = j \mid X_1(t - \tau) = i] \times \\ &\quad \Pr[X_2(t) = k \mid X_1(t - \tau) = i] f_{T_a}(\tau) d\tau. \end{aligned}$$

Aquí, $f_{T_a}(\tau)$ es la densidad del tiempo de coalescencia. Luego,

$$\begin{aligned} R_{jk}(t) &= \int_0^\infty \sum_i \pi_i(t - \tau) P_{ij}(\tau) P_{ik}(\tau) f_{T_a}(\tau) d\tau \\ &= \int_0^\infty \sum_i \pi_i(t - \tau) P_{ij}(\tau) P_{ik}(\tau) \left(\frac{1}{2N(t - \tau)} e^{-\int_0^\tau \frac{du}{2N(t-u)}} \right) d\tau. \end{aligned}$$

Denotando $R(t) = [R_{jk}(t)]$ y haciendo el cambio de variable $\sigma = t - \tau$, resulta

$$R(t) = \int_{-\infty}^t P^*(t - \sigma) \Pi(\sigma) P(t - \sigma) \left(\frac{1}{2N(\sigma)} e^{-\int_{\sigma}^t \frac{du}{2N(u)}} \right) d\sigma, \quad (3.3)$$

donde $\Pi(t) = \text{diag}[\pi_i(t)]$ y “*” denota transposición de matriz. Nótese que

$$\sum_{j,k} R_{jk}(t) = 1 - e^{-\int_0^{\infty} \frac{du}{2N(t-u)}},$$

por lo que la distribución $R(t)$ puede ser impropia si $\int_0^{\infty} \frac{du}{2N(t-u)} < \infty$. Esto significaría que X_1 y X_2 podrían no tener un ancestro común, siendo $p_2(t) = \exp\left\{-\int_{t_0}^t \frac{1}{2N(u)} du\right\}$ la probabilidad de no coalescencia en el intervalo $[t_0, t)$. Suponiendo que $t_0 = 0$, si $\int_0^{\infty} \frac{1}{2N(u)} du = \infty$, entonces $p_2(t) = \frac{1}{\exp\left\{\int_0^t \frac{1}{2N(u)} du\right\}} \rightarrow 0$ cuando $t \rightarrow \infty$, lo cual significaría que la probabilidad de coalescencia sería 1. Por otro lado, si $\int_0^{\infty} \frac{du}{2N(t-u)} < \infty$, entonces $0 < p_2(t) \leq 1$.

Nótese que la expresión (3.3) requiere que la cadena de Markov sea extendible indefinidamente en el pasado, i.e., que $\Pi(\sigma)$ exista para todo $\sigma \leq t$. Reescribiendo (3.3), dividiendo el intervalo de integración de (3.3) en dos partes: donde no hay coalescencia, es decir donde las dos cadenas de Markov ($X_1(t), X_2(t)$) fueron distribuidas de acuerdo a una matriz de probabilidad conjunta $m \in \mathcal{D}_2$, en $t = 0$, y la parte donde coalescen los dos procesos, a partir de que evolucionan como procesos independientes, resulta

$$R(t) = \left(\int_{-\infty}^0 + \int_0^t \right) P^*(t - \sigma) \Pi(\sigma) P(t - \sigma) \left(\frac{1}{2N(\sigma)} e^{-\int_{\sigma}^t \frac{du}{2N(u)}} \right) d\sigma \quad (3.4)$$

$$\begin{aligned} &= P^*(t) \left[\int_{-\infty}^0 P^*(-\sigma) \Pi(\sigma) P(-\sigma) \left(\frac{1}{2N(\sigma)} e^{-\int_{\sigma}^0 \frac{du}{2N(u)}} \right) d\sigma \right] \times \\ &P(t) e^{-\int_0^t \frac{du}{2N(u)}} + \int_0^t P^*(t - \sigma) \Pi(\sigma) P(t - \sigma) \left(\frac{1}{2N(\sigma)} e^{-\int_{\sigma}^t \frac{du}{2N(u)}} \right) d\sigma \\ &= P^*(t) R(0) P(t) e^{-\int_0^t \frac{du}{2N(u)}} \\ &+ \int_0^t P^*(t - \sigma) \Pi(\sigma) P(t - \sigma) \left(\frac{1}{2N(\sigma)} e^{-\int_{\sigma}^t \frac{du}{2N(u)}} \right) d\sigma. \end{aligned} \quad (3.5)$$

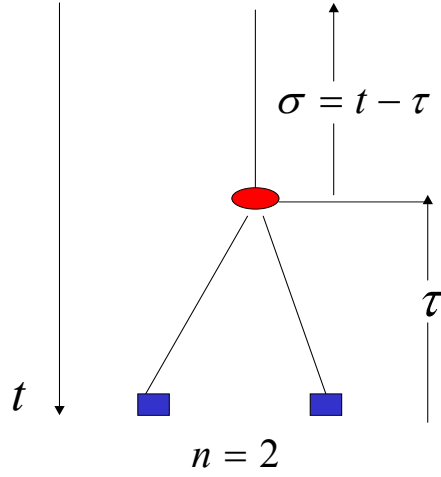


Figura 3.6:

Suponiendo que los dos individuos no coalescen entre $[0, t)$, si no entre $(-\infty, 0)$, los estados de sus alelos tienen distribución conjunta $R(0)$ dada por

$$R(0) = \int_{-\infty}^0 P^*(-\sigma) \Pi(\sigma) P(-\sigma) \left(\frac{1}{2N(\sigma)} e^{-\int_{\sigma}^0 \frac{du}{2N(u)}} \right) d\sigma \quad (3.6)$$

y distribuciones marginales $\pi(0)$ (ver Figura 3.6).

Definición 3.3 . $\Theta : \ell^1 \rightarrow \mathcal{M}_s$, operador positivo tal que $\|\Theta\| = 1$, está dado por

$$\Theta(\xi) = \sum_{i \in S} \xi_i e_i \otimes e_i = (\xi_i \delta_{i,j})_{i,j \in S}.$$

Este operador transforma un vector en una matriz diagonal con entradas diagonales idénticas a las del vector.

Definición 3.4 . $K : \mathcal{M}_s \rightarrow \ell^1$, operador positivo tal que $\|K\| = 1$. Para $\mu \in \mathcal{M}_s$ tal que

$\mu = (\mu_{i,j})_{i,j \in S}$, $K(\mu)$ está dado por

$$K(\mu) = \left(\sum_{j \geq 0} \mu_{i,j} \right)_{i \in S} = \left(\sum_{i \geq 0} \mu_{i,j} \right)_{j \in S}.$$

El operador K calcula la distribución marginal de la distribución conjunta.

Debido a que el tamaño de la población $N = N(t)$ depende de t , el proceso de Markov $(X_1(t), X_2(t))$, $t \geq 0$, en $S \times S$, no es homogéneo en el tiempo. Por tanto, para describir la historia del proceso $(X_1(t), X_2(t))$ se necesita la así llamada familia de evolución de operadores

$$R(t, s) : \mathcal{M}_s \rightarrow \mathcal{M}_t$$

dada por

$$\begin{aligned} R(t, s)m &= e^{-\int_s^t \frac{du}{2N(u)}} T(t-s)m \\ &+ \int_0^{t-s} \frac{1}{2N(t-u)} e^{-\int_{t-u}^t \frac{dv}{2N(v)}} T(u) \Theta P(t-s-u) K m du \end{aligned} \quad (3.7)$$

donde $R(t, s)$ es un operador tal que

$$R(t, s) = 1, \quad R(t, s)R(s, u) = R(t, u)$$

para $t \geq s \geq u \geq t_0$. $R(t, s)$ en \mathcal{M}_s . Si m es la distribución de $(X_1(t), X_2(t))$, $t \geq 0$ al tiempo s , entonces al tiempo presente $t > s$ su distribución está dada por $R(t, s)m$.

Para entender la Ecuación (3.3) observemos que existen dos posibilidades para la evolución del sistema en el intervalo de tiempo $[t_0, t)$. Si en este intervalo no hay coalescencia, lo cual ocurre con probabilidad $p_2(t)$, entonces el estado del proceso $(X_1(t), X_2(t))$ al tiempo t es el resultado de la acción de $T(t-t_0)$ en el proceso de mutaciones. Si la coalescencia ocurrió al tiempo $s \in [t_0, t)$, entonces el proceso $(X_1(t), X_2(t))$ es el resultado de la acción de Θ en el proceso de mutaciones seguido de la acción de $T(t-s)$.

Corolario 3.5 *Si $\{P(t), t \geq 0\}$ es una familia de operadores de Markov, es un operador que a $R(0) \mapsto R(t)$, es positivo si $R(0)$ lo es y $\|R(t)\|_{\mathcal{M}} = \|R(0)\|_{\mathcal{M}}$, para cualquier $R(0) \in$*

\mathcal{M} .

Demostración. Suponga que $R(0) \geq 0$ y $\|R(0)\|_{\mathcal{M}} = 1$. Ciertamente, por (3.5), $R(t) \geq 0$.

Por la Proposición 2.19,

$$\sum_{i,j} R_{i,j}(t) = e^{-\int_0^t \frac{du}{2N(u)}} \sum_{i,j} (R(0))_{i,j} + \int_0^t \frac{1}{2N(u)} e^{-\int_s^t \frac{du}{2N(u)}} \sum_{i,j} (\Theta\pi)_{i,j} ds.$$

Como $\sum_{i,j} (R(0))_{i,j} = 1$ y $\sum_{i,j} (\Theta\pi)_{i,j} = \sum_i \pi_i = 1$, resulta

$$\sum_{i,j} R_{i,j}(t) = e^{-\int_0^t \frac{du}{2N(u)}} + \int_0^t \frac{1}{2N(u)} e^{-\int_0^s \frac{du}{2N(u)}} ds = 1. \quad (3.8)$$

■

Introduciendo la fuerza de la deriva genética al modelo, no cambia la dinámica de las distribuciones marginales, solamente varía la dinámica de las distribuciones conjuntas, como se muestra en el siguiente resultado.

Lema 3.6 *Las distribuciones marginales de $R(t, s)m$ y $T(t, s)m$ son iguales a las distribuciones marginales de $P(t, s)Km$.*

Demostración. Poniendo $m = e_i \otimes e_j$ y $T(t) = P(t) \otimes P(t)$ se verifica que

$$KT(s)m = P(s)Km, \quad s \geq 0, \quad m \in \mathcal{M}_s, \quad (3.9)$$

y

$$K\Theta x = x, \quad x \in \ell^1, \quad K\Theta = I.$$

Así, de (3.9) y de las propiedades del semigrupo,

$$KT(t-u)\Theta P(u-s) = P(t-u)K\Theta P(u-s) = P(t-s).$$

Por lo tanto,

$$\begin{aligned}
KR(t, s)m &= e^{-\int_s^t \frac{du}{2N(u)}} KT(t-s)m \\
&+ \int_s^t \frac{1}{2N(u)} e^{-\int_u^t \frac{dv}{2N(v)}} KT(t-u)\Theta P(u-s)K m du \\
&= e^{-\int_s^t \frac{du}{2N(u)}} P(t-s)K m \\
&+ \int_s^t \frac{1}{2N(u)} e^{-\int_u^t \frac{dv}{2N(v)}} P(t-s)K m du \\
&= \left[e^{-\int_s^t \frac{du}{2N(u)}} + \int_s^t \frac{1}{2N(u)} e^{-\int_u^t \frac{dv}{2N(v)}} du \right] P(t-s)K m \\
&= P(t-s)K m,
\end{aligned}$$

pues, debido a (3.8), $e^{-\int_s^t \frac{du}{2N(u)}} + \int_s^t \frac{1}{2N(u)} e^{-\int_u^t \frac{dv}{2N(v)}} du = 1$. ■

3.2 Comportamiento asintótico de las distribuciones de $(\tau_{t'}, \sigma_{t'})$.

La manera en la cual la deriva genética influye en la distribución conjunta del proceso $(X_1(t), X_2(t))$, está muy relacionada con el comportamiento asintótico de la familia del par de variables $(\tau_{t'}, \sigma_{t'})$, donde t' es una representación abreviada de $t-s$ y la variable aleatoria $\tau_{t'}$ está definida en el intervalo $[0, t']$ de tal manera que

$$P[\tau_{t'} \geq u] = \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\},$$

(y en particular, $P[\tau_{t'} \geq t'] = \exp \left\{ - \int_s^t \frac{dv}{2N(v)} \right\}$). La variable aleatoria $\sigma_{t'}$ se define como $t' - \tau_{t'}$. Intuitivamente, $\sigma_{t'}$ es la cantidad de tiempo en el intervalo $[0, t']$ antes de que el proceso $(X_1(t), X_2(t))$ se dividiera en dos procesos independientes; mientras que $\tau_{t'}$ es la cantidad de tiempo durante el cual dos procesos, que inicialmente eran idénticos, i.e., al tiempo $t=0$, han evolucionado independientemente, como se muestra en la Figura 3.7. Es

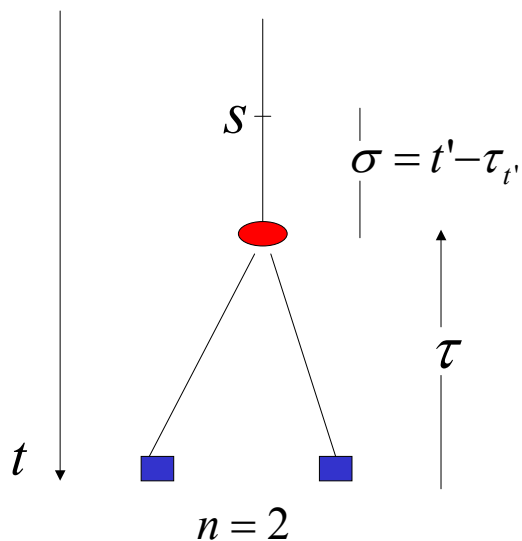


Figura 3.7:

importante mencionar que este proceso de ramificación se repite en futuras generaciones como se muestra en la Figura 3.8.

Para describir el comportamiento asintótico de las distribuciones del par de variables $(\tau_{t'}, \sigma_{t'})$ usaremos el teorema de representación de Riesz (ver Bartle, 1966, Capítulo 9), el cual establece una correspondencia biyectiva entre medidas regulares de Borel sobre un espacio compacto \mathbb{S} y funcionales positivos en el espacio $C(\mathbb{S})$ de funciones continuas. En la descripción del comportamiento asintótico de las distribuciones de $(\tau_{t'}, \sigma_{t'})$ las medidas se representan como funcionales.

Lema 3.7 Sea $\mathbb{S} = [0, \infty]^2$ y sea $p_{t'}$ la distribución conjunta del par de variables aleatorias $(\tau_{t'}, \sigma_{t'})$ en \mathbb{S} . Las medidas $\{p_{t'}\}$ convergen débilmente, cuando $t \rightarrow \infty$, a una medida p en \mathbb{S} que depende del comportamiento de la función $N(\cdot)$ de la siguiente manera:

- (a) Si $\lim_{t \rightarrow \infty} N(t) = 0$, entonces p es la medida puntual de Dirac $\delta_{\{0, \infty\}}$ con

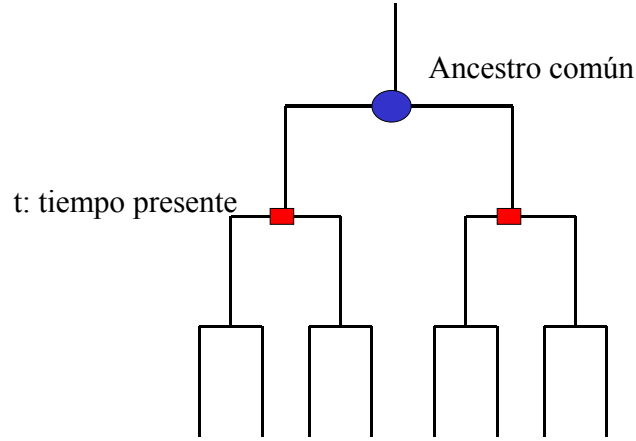


Figura 3.8:

masa en $\{0, \infty\} \in \mathbb{R}_+ \times \overline{\mathbb{R}}_+$.

(b) Si $\lim_{t \rightarrow \infty} N(t) = N$, donde $0 < N < \infty$, entonces p visto como un funcional en el espacio $C(\mathbb{S})$ de funciones continuas y acotadas en S , está dado por

$$f \mapsto \frac{1}{2N} \int_0^\infty e^{-\frac{t}{2N}} f(t, \infty) dt \quad \text{para } f \in C(\mathbb{S}).$$

(c) Si $\lim_{t \rightarrow \infty} N(t) = \infty$ y $\int_s^\infty \frac{dv}{2N(v)} = \infty$, entonces p es la medida puntual de Dirac $\delta_{\{\infty, \infty\}}$ con masa en $\{\infty, \infty\}$..

(d) Si $\lim_{t \rightarrow \infty} N(t) = \infty$ y $\int_s^\infty \frac{dv}{2N(v)} < \infty$, entonces p , visto como un funcional en el espacio $C(\mathbb{S})$, está dado por

$$f \mapsto \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} f(\infty, 0) + \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} f(\infty, u) du, \quad f \in C(\mathbb{S}).$$

Demostración. Para cualquier $v, w \geq 0$, y t suficientemente grande tal que $t - s - w \geq v$,

y por la Ecuación (3.2)

$$\begin{aligned}
p_{t'} [(v, \infty] \times (w, \infty)] &= P[v < \tau_{t'} < t - s - w] \\
&= P[v - t < \tau_{t'} - t < -(s + w)] \\
&= P[t - v > t - \tau_{t'} > s + w] \\
&= e^{-\int_{t-v}^t \frac{du}{2N(u)}} - e^{-\int_{s+w}^t \frac{du}{2N(u)}}, \tag{3.10}
\end{aligned}$$

donde $t - \tau_{t'}$ es el tiempo de coalescencia. Por tanto la expresión anterior se puede interpretar como la probabilidad de coalescencia en $[s + w, t - v]$. Entonces

$$\lim_{t \rightarrow \infty} p_{t'} [(v, \infty] \times (w, \infty)] = \lim_{t \rightarrow \infty} e^{-\int_{t-v}^t \frac{du}{2N(u)}} - \lim_{t \rightarrow \infty} e^{-\int_{s+w}^t \frac{du}{2N(u)}}. \tag{3.11}$$

En el caso (a), cuando el tamaño de la población decae hasta cero en el límite, para $v = 0$ el primer límite a la derecha de (3.11) tiende a uno pues la integral sobre un punto es cero. Para probar que el segundo límite en el lado izquierdo de (3.11) tiende a 0, es suficiente demostrar que $\int_{s+w}^t \frac{du}{2N(u)} = \infty$. Sea $t_n \uparrow \infty$ y sea $a_n = \int_{s+w}^{t_n} \frac{du}{2N(u)}$. Probaremos que para toda $M > 0 \exists a_n$ tal que $a_n > M$. Sea $u > s + w$ tal que $\frac{1}{2N(u)} > M$ para toda $v > u$, y sea n tal que $t_n > u + 1$. Entonces

$$\begin{aligned}
\int_{s+w}^{t_n} \frac{dv}{2N(v)} &= \int_{s+w}^u \frac{dv}{2N(v)} + \int_u^{t_n} \frac{dv}{2N(v)} \\
&\geq \int_{s+w}^u \frac{dv}{2N(v)} + \int_u^{t_n} M dv \\
&= \int_{s+w}^u \frac{dv}{2N(v)} + M(t_n - u) \geq M.
\end{aligned}$$

Haciendo $n \rightarrow \infty$, y tomando el límite cuando $t \rightarrow \infty$

$$\lim_{t \rightarrow \infty} \int_{s+w}^t \frac{du}{2N(u)} = \int_{s+w}^{\infty} \frac{du}{2N(u)} \geq \int_{s+w}^{t_n} \frac{du}{2N(u)} > M.$$

Haciendo $M \rightarrow \infty$ se obtiene lo deseado. En el caso cuando $v > 0$,

$$\begin{aligned} \lim_{t \rightarrow \infty} \exp \left\{ - \int_{t-v}^t \frac{du}{2N(u)} \right\} &= 0 \\ \lim_{t \rightarrow \infty} \exp \left\{ - \int_{s+w}^t \frac{du}{2N(u)} \right\} &= 0. \end{aligned}$$

Por un argumento similar al anterior, $\infty \leq \int_{t-v}^t \frac{du}{2N(u)} \leq \infty$. Entonces

$$\lim_{t \rightarrow \infty} p_{t'} [(v, \infty) \times (w, \infty)] = \begin{cases} 1, & v = 0 \\ 0 & v > 0, \end{cases}$$

o sea, $p_{t'} \rightarrow \delta_{(0, \infty)}(\cdot)$, i.e. $p_{t'}$ converge a la medida de Dirac en el punto $\{0, \infty\}$.

La interpretación en términos de los tiempos σ y τ , es que debido a que cuando $t \rightarrow \infty$, la población tiende a ser pequeña por lo que el tiempo de coalescencia de los dos alelos seleccionados es muy corto, que comparado con el tiempo transcurrido hasta t , es cero. A partir de ese momento los dos alelos son del mismo tipo durante un tiempo infinito. Este resultado indica que todos los individuos en la población tenderán a ser del mismo tipo de alelo, como se observa en la Figura 3.2 (Sección 3.1.2).

En el caso (b), con $2N(t)$ asintóticamente constante, (3.10) y (3.11) implican que

$$\begin{aligned} \lim_{t \rightarrow \infty} \exp \left\{ - \int_{t-v}^t \frac{du}{2N(u)} \right\} &= \exp \left\{ - \frac{v}{2N} \right\} \\ &= \int_v^\infty \frac{1}{2N} e^{-\frac{u}{2N}} du \times \delta_\infty(w, \infty], \end{aligned}$$

por tanto la medida $p_{t'}$ sobre \mathbb{S} tiene por límite débil a la medida

$$\frac{1}{2N} e^{-\frac{u}{2N}} du \times \delta_\infty(dv).$$

Lo anterior implica que para toda función continua y acotada $f : \mathbb{S} \rightarrow \mathbb{R}$ se tiene

$$\int \int p_{t'}(du, dv) f(u, v) \rightarrow \int \int f(u, v) \frac{1}{2N} e^{-\frac{u}{2N}} du \delta_\infty(dv).$$

El tamaño de la población tiende a ser constante y la variable aleatoria τ tiene función de distribución exponencial con parámetro $\frac{1}{2N}$. En la Figura 3.3 se muestran simulaciones del comportamiento de los dos alelos con $N(t)$ asintóticamente constante. El tiempo que tardarán en coalescer los dos alelos está distribuido exponencialmente y es finito.

En (c), cuando el tamaño de la población crece a infinito pero de manera lenta, i.e., $\int_s^\infty \frac{dv}{2N(v)} = \infty$, para toda v la Ecuación (3.11) se reduce a

$$\lim_{t \rightarrow \infty} \exp \left\{ - \int_{t-v}^t \frac{du}{2N(u)} \right\} = 1,$$

pués tenemos que

$$2N(u)_{u \in [t-v, t]} \geq 2 \min 2N(u).$$

De aquí se sigue que

$$\frac{1}{2N(u)} \leq \frac{1}{2 \min 2N(u)},$$

y aplicando el límite cuando $t \rightarrow \infty$ en ambos lados de la desigualdad y por la Hipótesis (3.2) se tiene

$$0 \leq \frac{1}{2N(u)} \leq \frac{1}{\infty} \geq 0,$$

por tanto, $\frac{1}{2N(u)} = 0$. De lo anterior se tiene que $p_{t'} \rightarrow \delta_{(\infty, \infty)}(\cdot)$, i.e. $p_{t'}$ converge a la medida de Dirac en el punto $\{\infty, \infty\}$. En este caso, aunque la población crece a infinito, de manera lenta (consideramos en las simulaciones un crecimiento lineal), la deriva genética está todavía presente. Debido a que la población tiende a infinito en un tiempo infinito, se requiere de $\tau = \infty$ para la coalescencia de los dos alelos en cuestión, sin embargo, el resultado sugiere que lo mismo pasará con el tiempo en que los dos alelos permanecen idénticos.

En (a), (b) y (c) se asume que el crecimiento poblacional es lento, i.e., $\int_s^\infty \frac{du}{2N(u)} = \infty$.

En (d), cuando la población crece a infinito pero de manera rápida (ver Figura 3.5), se tiene que $\lim_{t \rightarrow \infty} \exp \left\{ - \int_{t-v}^t \frac{du}{2N(u)} \right\} = 1$. Usando la hipótesis $\int_s^\infty \frac{du}{2N(u)} < \infty$, esto se demuestra como

$$\int_{t-v}^t \frac{du}{2N(u)} = \int_{t-v}^\infty \frac{du}{2N(u)} - \int_t^\infty \frac{du}{2N(u)},$$

y puesto que

$$\lim_{t \rightarrow \infty} \int_t^\infty \frac{du}{2N(u)} = 0,$$

se obtiene haciendo $t \rightarrow \infty$

$$\lim_{t \rightarrow \infty} \int_{t-v}^t \frac{du}{2N(u)} = \lim_{t \rightarrow \infty} \int_{t-v}^\infty \frac{du}{2N(u)} - \lim_{t \rightarrow \infty} \int_t^\infty \frac{du}{2N(u)} = 0.$$

Por tanto, tenemos

$$\lim_{t \rightarrow \infty} p_t' [(v, \infty] \times (w, \infty]] = 1 - \exp \left\{ \int_{s+w}^\infty \frac{du}{2N(u)} \right\}.$$

Este límite visto como un funcional sobre el espacio $C(\mathbb{S})$ de funciones continuas en \mathbb{S} , está determinado por la expresión siguiente (usando de nuevo el teorema de representación de Riesz):

$$f \mapsto \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} f(\infty, 0) + \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} f(\infty, u) du, \quad f \in C(\mathbb{S}).$$

■

Para este caso, resulta un tanto compleja la interpretación en términos de los tiempos τ y σ . Debido a que la población crece rápidamente a infinito (como se muestra

en la Figura 3.5, con población que crece exponencialmente), entre s e infinito puede no haber coalescencia o entre s e infinito puede haber coalescencia y los alelos permanecen en el mismo estado durante un tiempo finito. En términos de las probabilidades de los dos alelos, viendo los efectos de la deriva genética, según las simulaciones, observamos que la deriva genética actúa durante un tiempo corto y a partir de un determinado momento los efectos de la deriva genética son nulos, las frecuencias de los alelos se estabilizan conforme al principio de Hardy–Weinberg, permaneciendo así durante un tiempo infinito, por lo que es probable que los dos alelos no coalescan.

Capítulo 4

Mutación contra deriva

En esta Sección se estudia el papel que juegan dos fuerzas genéticas: la mutación y la deriva genética aleatoria. Esto se hace examinando tanto la distribución cuasiestacionaria condicionada a no absorción como la forma de la función de velocidad de escape, bajo diferentes escenarios de comportamiento asintótico del tamaño efectivo de la población $2N(\cdot)$.

Con este objetivo, en la Sección 4.2 se obtiene la función de velocidad de escape y la distribución cuasiestacionaria condicionada a no absorción en el estado $(0, 0)$ y en el margen. Además, se hace una comparación de éstas con las del modelo sin deriva, las cuales se presentaron en el Capítulo 2. Para comprender mejor cuáles son los efectos de la deriva genética en presencia de mutaciones, en la siguiente Sección se presentan simulaciones de la evolución de dos cromosomas, bajo el modelo de Wright-Fisher-Moran.

4.1 Simulación de la evolución de dos alelos bajo el modelo de Wright-Fisher-Moran en presencia de mutaciones

El objetivo de esta sección es ilustrar gráficamente cuáles son los efectos de la deriva genética en presencia de mutaciones para la evolución de dos cromosomas, bajo el modelo de Wright-Fisher-Moran (ver Secciones 1.1.2, 1.1.3 y 3.1.1). Aunque el proceso de mutaciones no corresponde al que comúnmente se usa en la modelación de microsátélites, en la Figura 4.1 se muestran los efectos de la deriva genética bajo mutación con la finalidad de describir la probabilidad del alelo. En cada gráfica presentada a continuación, cada trayectoria representa la evolución de un tipo de alelo en una población, en cada figura se simula la evolución del alelo en 10 poblaciones. El tiempo está medido en N generaciones.

En la Figura 4.1 se muestran los efectos de la deriva genética sin mutación y con mutaciones bajo el modelo de Wright-Fisher-Moran, tal que en la ecuación (1.1),

$$P(X_{n+1} = j \mid X_n = i) = \binom{2N}{j} p_i^j (1 - p_i)^{2N-j},$$

$$p_i = \frac{i}{2N} (1 - \nu) + \left(1 - \frac{i}{2N}\right) \mu$$

donde ν es la probabilidad de mutación del alelo tipo 1 al alelo tipo 2, y μ es la probabilidad de mutación del alelo tipo 2 al alelo tipo 1. En este ejemplo se tienen dos microsátélites de longitudes $s_0 = 2$ y $s_1 = 3$ respectivamente, cada uno con la misma proporción inicial en la población. En las gráficas de la segunda columna de la Figura 4.1 se presentan los efectos de la deriva genética aleatoria sin mutaciones para poblaciones de tamaño constante de 100 y 18, respectivamente. En la primera columna de la misma figura se muestran los efectos de las dos fuerzas genéticas, deriva y mutación. Cuando solamente actúa la fuerza de la

deriva genética en una población, un alelo se fijará o perderá más rápido en la población que en presencia de deriva genética y mutación, como se muestra en la Figura 4.1 (comparar gráficas de ambas columnas). La probabilidad de una mutación es del orden de 10^{-6} , aquí, para ilustrar lo anterior se toman las mutaciones con $\nu = 0.05$ y $\mu = 0.06$. Las mutaciones introducen diversidad, o variabilidad en las poblaciones.

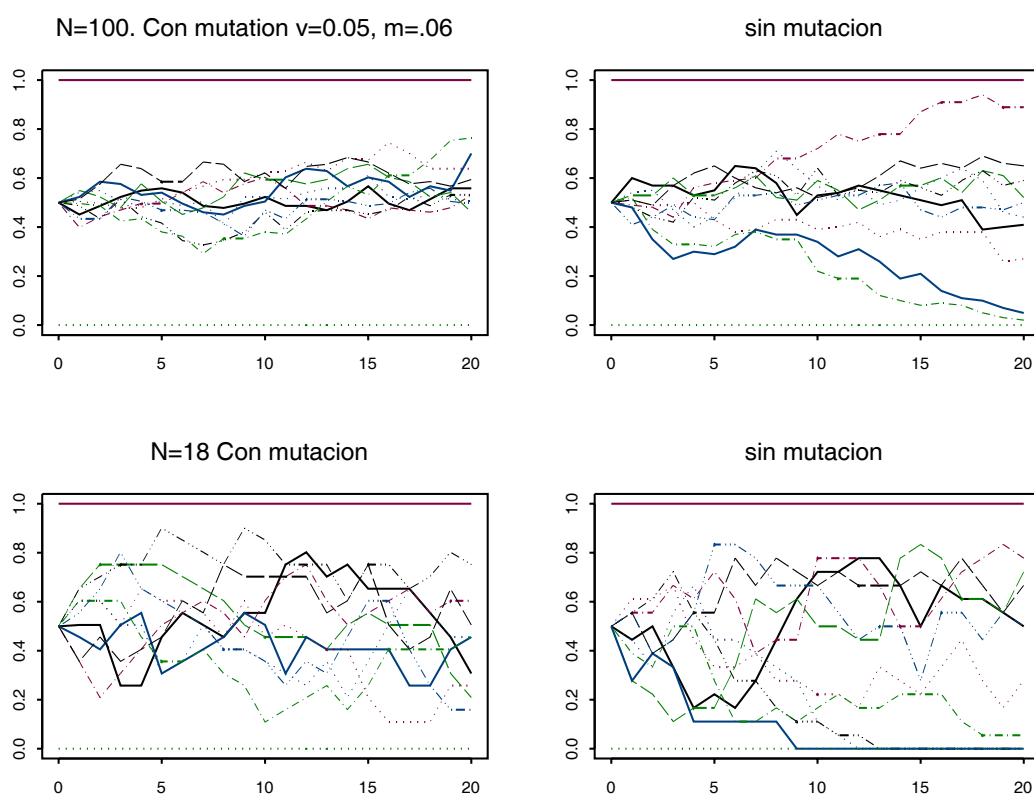


Figura 4.1

4.2 Forma de la DLC del modelo con deriva genética aleatoria

La demostración del siguiente teorema se basa en el Lema 3.7. En el Teorema 4.1 se presentan distribuciones límite en función de operadores lineales fuertemente continuos, bajo diferente comportamiento asintótico de $N(t)$. Es de gran utilidad para obtener la forma de la distribución límite condicionada a no absorción (tanto en el margen como en el origen) del proceso $(X_1(t), X_2(t))$ del modelo con deriva genética aleatoria.

Teorema 4.1 Sean \mathbf{X} y \mathbf{Y} dos espacios de Banach, y sean $V(t) \in \mathcal{L}(\mathbf{X})$, $W(t) \in \mathcal{L}(\mathbf{Y})$, $t \geq 0$, familias de operadores lineales fuertemente continuos en X e Y respectivamente tales que $V(0)x = x$, $W(0)y = y$, y los límites $\lim_{t \rightarrow \infty} V(t)x = V(\infty)x$ y $\lim_{t \rightarrow \infty} W(t)y = W(\infty)y$ existen para cada $x \in \mathbf{X}$ y cada $y \in \mathbf{Y}$. Más aún, sea $\Theta \in \mathcal{L}(\mathbf{X}, \mathbf{Y})$ y $K \in \mathcal{L}(\mathbf{Y}, \mathbf{X})$ operadores lineales acotados. Se define $U(t, s) \in \mathcal{L}(\mathbf{Y})$ como

$$U(t, s)y = \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} W(t-s)y + \int_0^{t-s} \frac{1}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} W(u)\Theta V(t-s-u)Kydu.$$

Entonces el límite de $U(t, s)y$ cuando $t \rightarrow \infty$ depende del comportamiento de $N(\cdot)$ de la siguiente manera:

- (a) Si $\lim_{t \rightarrow \infty} N(t) = 0$, entonces $\lim_{t \rightarrow \infty} U(t, s)y = \Theta V(\infty)Ky$.
- (b) Si $\lim_{t \rightarrow \infty} N(t) = N$, $0 < N < \infty$ entonces

$$\lim_{t \rightarrow \infty} U(t, s)y = \frac{1}{2N} \int_0^\infty e^{-\frac{t}{2N}} W(t)\Theta V(\infty)Kydt.$$

(c) Si $\lim_{t \rightarrow \infty} N(t) = \infty$ e $\int_s^\infty \frac{dt}{2N(t)} = \infty$, entonces

$$\lim_{t \rightarrow \infty} U(t, s)y = W(\infty)\Theta V(\infty)Ky.$$

(d) Si $\lim_{t \rightarrow \infty} N(t) = \infty$ e $\int_s^\infty \frac{dt}{2N(t)} < \infty$, entonces

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= W(\infty) \left[\exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} y \right. \\ &\quad \left. + \Theta \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} V(u)Ky du \right]. \end{aligned}$$

Demostración. Lo anterior es consecuencia directa del Lema 3.7 y de convergencia débil. De las hipótesis se tiene que $V(t)$ y $W(t)$, $t \geq 0$, son familia de operadores fuertemente continuos tales que $W(0) = I$, $V(0) = I$, y $\|W(t)\|_{\mathcal{L}(\mathbf{Y})} \leq 1$, $\|V(t)\|_{\mathcal{L}(\mathbf{X})} \leq 1$. Además, por (2.20), se tiene que $\sup_{t \geq 0} \max \{\|V(t)\|, \|W(t)\|\} < \infty$, por tanto $V(t)$ y $W(t)$ son equiacotados. Fijemos $y \in \mathbf{Y}$. La función $Z(v, w) = W(w)\Theta V(v)Ky$ es continua en $\mathbb{S} = [0, \infty] \times [0, \infty]$.

Más aún,

$$\begin{aligned} U(t, s)y &= \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} W(t-s)[y - \Theta Ky] + EZ(\tau_{tu}, \sigma_{tu}) \\ &= \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} W(t-s)[y - \Theta Ky] + \int_{\mathbb{S}} Z dp_{tu} \end{aligned}$$

tomando la derivada de la Ecuación (3.10)

$$dp_{tu} = \frac{1}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} + \frac{1}{2N(t-s-u)} \exp \left\{ - \int_{s+u}^t \frac{dv}{2N(v)} \right\}.$$

Sea

$$\begin{aligned} g_t(v) &= -\frac{d}{dv} \left(\exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} \right) \\ &= \frac{1}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\}, \quad 0 \leq u \leq t. \end{aligned}$$

Nótese que para todos los casos, el segundo término en $dp_{t'}$ tiende a 0 cuando $t \rightarrow \infty$, lo cual se demostró en el lema anterior, y el primer término de $U(t, s)y$ es cero para los casos (a), (b) y (c), pues $\lim_{t \rightarrow \infty} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} = 0$, por lo cual, para (a), (b) y (c),

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= \lim_{t \rightarrow \infty} \int_{\mathbb{S}} Z dp_{t'} = \lim_{t \rightarrow \infty} \int_s^t g_t(u)W(u)\Theta V(t-s-u)Ky du \\ &= \lim_{t \rightarrow \infty} \int_0^t g_t(u)W(u)\Theta V(t-s-u)Ky du. \end{aligned}$$

Para (a), usando los resultados del lema anterior, donde $\int_0^\infty g_t(u)du$ es igual a

$$\lim_{t \rightarrow \infty} \int_0^{t-s} -\frac{d}{du} \left(\exp \left\{ - \int_{t-v}^t \frac{du}{2N(u)} \right\} \right) du = 1,$$

y $W(0) = I$,

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= \lim_{t \rightarrow \infty} \int_0^{t-s} g_t(u)W(u)\Theta V(t-s-u)Ky du \\ &= \Theta V(\infty)Ky \int_0^\infty g(u)du \\ &= \Theta V(\infty)Ky. \end{aligned}$$

Para el caso (b), tomando el límite y sustituyendo $g_t(u) = \frac{1}{2N} e^{-\frac{u}{2N}}$,

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= \lim_{t \rightarrow \infty} \int_0^{t-s} g_t(u)W(u)\Theta V(t-s-u)Ky du \\ &= \frac{1}{2N} \int_0^\infty e^{-\frac{u}{2N}} W(u)\Theta V(\infty)Ky dt. \end{aligned}$$

En el caso (c),

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= \lim_{t \rightarrow \infty} \int_0^{t-s} g_t(u)W(u)\Theta V(t-s-u)Ky du \\ &= W(\infty)\Theta V(\infty)Ky \int_0^\infty g(u)du \\ &= W(\infty)\Theta V(\infty)Ky. \end{aligned}$$

Finalmente, para (d),

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= \lim_{t \rightarrow \infty} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} W(t-s)y \\ &\quad + \lim_{t \rightarrow \infty} \int_0^{t-s} g_t(u)W(u)\Theta V(t-s-u)Kydu. \end{aligned}$$

Haciendo el cambio de variable $u = t - s - u$ y aplicando el límite en ambos términos,

$$\begin{aligned} \lim_{t \rightarrow \infty} U(t, s)y &= \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} W(\infty)y \\ &\quad + \lim_{t \rightarrow \infty} \int_0^{t-s} g_t(s+u)W(t-s-u)\Theta V(u)Kydu \\ &= \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} W(\infty)y \\ &\quad + \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{s+u}^\infty \frac{dv}{2N(v)} \right\} W(\infty)\Theta V(u)Kydu \end{aligned}$$

de donde se sigue (d). ■

4.2.1 Forma de la DLC condicionada a no absorción en el origen

El análisis que se presenta en esta sección parte de la siguiente definición de distribución cuasiestacionaria dada por Bobrowski (2004).

Definición 4.2 *Una distribución $m_0 \in \mathcal{D}_2 \cap \mathcal{M}_s$ es una distribución cuasiestacionaria de dos cadenas de Markov $(X_1(t), X_2(t))$ relacionadas por la distribución de (τ_u, σ_u) , condicionada a la no absorción en $(0, 0)$ si*

$$\lim_{t \rightarrow \infty} v(t)BT(t)m = b(m)m_0$$

para alguna función v y $\forall m \in \mathcal{D}_2 \cap \mathcal{M}_s$, donde $b(m)$ es una constante la cual es diferente de cero a menos que $Bm = 0$.

Se dice que v es la velocidad de escape al origen $(0, 0)$. Por otro lado, m_0 es la distribución límite condicionada a no absorción al margen si se cumple

$$\lim_{t \rightarrow \infty} v(t)CT(t)m = c(m)m_0$$

para alguna función v y $\forall m \in \mathcal{D}_2 \cap \mathcal{M}_s$, donde $c(m) \geq 0$ que es sólo cuando $Cm = 0$ si y sólo si $C(m) = 0$. El Teorema 4.3 da la forma de la distribución límite condicionada a no absorción al origen $(0, 0)$, del proceso $(X_1(t), X_2(t))$, bajo diferente comportamiento asintótico de $N(t)$, en el modelo con deriva.

Teorema 4.3 *Para cualquier $m \in \mathcal{D}_2 \cap \mathcal{M}_s$, el límite $\lim_{t \rightarrow \infty} e^{\gamma(t-s)}BR(t, s)m$ existe y es igual a*

$$\begin{aligned} a(Km)\Theta x_0 & \quad \text{si } \lim_{t \rightarrow \infty} N(t) = 0, \\ \frac{a(Km)}{2N} \int_0^\infty e^{-\frac{t}{2N}} e^{\gamma t} BT(t)\Theta x_0 dt & \quad \text{si } \lim_{t \rightarrow \infty} N(t) = N, 0 < N < \infty, \\ r_1(m) [x_0 \otimes e_0 + e_0 \otimes x_0] & \quad \text{si } \lim_{t \rightarrow \infty} N(t) = \infty, \end{aligned}$$

donde $r_1(m)$ es una constante no negativa y $r_1(m) = 0$ si y sólo si $Bm = 0$.

Demostración. Se tiene que

$$\begin{aligned} T(t)(e_0 \otimes e_0) &= (P(t) \otimes P(t))(e_0 \otimes e_0) \\ &= P(t)e_0 \otimes P(t)e_0 \\ &= e_0 \otimes e_0. \end{aligned}$$

Además $B\Theta = \Theta A$, además $BT(t) = BT(t)B$, y consecuentemente

$$BT(t)\Theta P(t' - u) = BT(t)\Theta AP(t' - u).$$

Usando que B es acotado, se sigue de la Ecuación (3.7) que

$$\begin{aligned}
e^{\gamma(t-s)}BR(t,s)m &= \exp\left\{-\int_s^t \frac{du}{2N(u)}\right\} e^{\gamma t'}BT(t')m \\
&+ \int_0^{t'} \frac{1}{2N(t-u)} \exp\left\{-\int_{t-u}^t \frac{dv}{2N(v)}\right\} \\
&\times e^{\gamma u}BT(u)\Theta e^{\gamma(t'-u)}AP(t'-u)Kmd u. \tag{4.1}
\end{aligned}$$

Por tanto el operador $U(t,s) = e^{\gamma t'}BR(t,s)m$ tiene la forma descrita en el Teorema 4.1 con $V(t) = e^{\gamma t}AP(t)K$ y $W(t) = e^{\gamma t}BT(t)$. Debido a (2.11),

$$\begin{aligned}
\lim_{t \rightarrow \infty} e^{\gamma t}AP(t)Km &= \lim_{t \rightarrow \infty} V(t)m = V(\infty)m, \\
V(\infty)m &= a(Km)x_0, \tag{4.2}
\end{aligned}$$

donde $a(Km)$ es constante, y por el lema 2.20,

$$\begin{aligned}
\lim_{t \rightarrow \infty} e^{\gamma t}BT(t)m &= \lim_{t \rightarrow \infty} W(t)m = W(\infty)m \\
W(\infty)m &= b(m)[x_0 \otimes e_0 + e_0 \otimes x_0]. \tag{4.3}
\end{aligned}$$

Entonces por el Teorema 4.1(a)

$$\begin{aligned}
\lim_{t \rightarrow \infty} U(t,s)m &= \Theta V(\infty)m = \Theta a(Km)x_0 \\
&= a(Km)\Theta x_0,
\end{aligned}$$

y por la parte (b) del Teorema 4.1

$$\begin{aligned}
\lim_{t \rightarrow \infty} U(t,s)m &= \frac{1}{2N} \int_0^\infty e^{-\frac{u}{2N}} W(u)\Theta V(\infty)Kmd u \\
&= \frac{1}{2N} \int_0^\infty e^{-\frac{u}{2N}} e^{\gamma u}BT(u)\Theta a(Km)\Theta x_0 du \\
&= \frac{a(Km)}{2N} \int_0^\infty e^{-\frac{u}{2N}} e^{\gamma u}BT(u)\Theta x_0 du,
\end{aligned}$$

quedando demostrados los primeros dos casos del teorema. Por la parte (c) del Teorema

4.1

$$\begin{aligned}
\lim_{t \rightarrow \infty} U(t, s)m &= W(\infty)\Theta V(\infty)m \\
&= W(\infty)\Theta a(Km)x_0 \\
&= a(Km)W(\infty)\Theta x_0 \\
&= a(Km)W(\infty)m, \text{ donde } m = \Theta x_0 \\
&= a(Km)b(\Theta x_0)[x_0 \otimes e_0 + e_0 \otimes x_0] \\
&= r_1(m)[x_0 \otimes e_0 + e_0 \otimes x_0],
\end{aligned}$$

donde $r_1(m) = a(Km)b(\Theta x_0)$. Por (d) del Teorema 4.1,

$$\begin{aligned}
\lim_{t \rightarrow \infty} U(t, s)m &= W(\infty) \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} m \\
&\quad + W(\infty)\Theta \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} V(u)Kmdu \\
&= \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} W(\infty)m + \\
&\quad W(\infty) \left(\Theta \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} e^{\gamma u} AP(u)Kmdu \right) \\
&= \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} b(m) + \\
&\quad b \left(\Theta \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} e^{\gamma u} AP(u)Kmdu \right) \\
&= r_1(m)[x_0 \otimes e_0 + e_0 \otimes x_0],
\end{aligned}$$

donde

$$\begin{aligned}
r_1(m) &= \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} b(m) \\
&\quad + b \left(\Theta \int_0^\infty \frac{1}{2N(s+u)} \exp \left\{ - \int_{u+s}^\infty \frac{dv}{2N(v)} \right\} e^{\gamma u} AP(u)Kmdu \right),
\end{aligned}$$

con lo cual queda demostrado el último caso del Teorema. ■

La interpretación genética del Teorema 4.3 es la siguiente: la distribución límite condicional de dos cadenas de Markov condicionadas a no absorción en el estado $(0, 0)$, es igual a Θx_0 si $\lim_{t \rightarrow \infty} 2N(t) = 0$, es decir, la distribución se concentra en la diagonal, por lo que la deriva genética aleatoria alcanza su fin: fijación de alelos, es decir, uno de los alelos se perderá y el otro se fijará en la población. Si $\lim_{t \rightarrow \infty} 2N(t) = \infty$, entonces la distribución límite condicional es igual a la del modelo sin deriva, pues en el modelo sin deriva

$$\lim_{t \rightarrow \infty} e^{\gamma t} BT(t)m = b(m) [x_0 \otimes e_0 + e_0 \otimes x_0],$$

mientras que en el modelo con deriva

$$\lim_{t \rightarrow \infty} e^{\gamma t} BR(t, s)m = r_1(m) [x_0 \otimes e_0 + e_0 \otimes x_0].$$

Si $\lim_{t \rightarrow \infty} 2N(t) = 2N > 0$, entonces la distribución límite condicional es un múltiplo escalar de la transformada de Laplace de $t \mapsto e^{\gamma t} BT(t)\theta x_0$ en $\lambda = \frac{1}{2N}$.

4.2.2 Forma de la DLC condicionada a no absorción en el margen

Para encontrar límites relacionados con el operador C , el razonamiento y resultados son similares a los presentados en el Teorema 4.3. Por el Lema 2.20, las colas de las distribuciones de los tiempos de absorción en el origen y en el margen, es decir, de T_0 y T_M , respectivamente, son diferentes: las de T_M son mucho más ligeras. Por otro lado, la integral en el segundo sumando del lado derecho de la Ecuación (3.7) describe el hecho de que la evolución de distribuciones de dos procesos consisten de dos fases relacionadas con las variables aleatorias $\tau_{\iota}, \sigma_{\iota}$ consideradas en el Lema 3.7, es decir, existen dos posibilidades para la evolución del sistema en el intervalo de tiempo $[t_0, t)$. Si en este intervalo

no hay coalescencia, lo cual ocurre con probabilidad $p_2(t)$, entonces el estado del proceso $(X_1(t), X_2(t))$ al tiempo t es el resultado de la acción de $T(t - t_0)$ en el proceso de mutaciones. Si la coalescencia ocurrió al tiempo $s \in [t_0, t)$, entonces el proceso $(X_1(t), X_2(t))$ es el resultado de la acción de Θ en el proceso de mutaciones seguido de la acción de $T(t - s)$. Durante la primera fase, antes de la separación en el tiempo σ_{ν} , se podría pensar en dos procesos idénticos, su distribución está concentrada en la diagonal. El tiempo de absorción al margen para tal par "degenerado" es sólo el tiempo de absorción en el 0 para un proceso univariado, y las colas de la distribución de este tiempo de absorción son $O(e^{-\gamma t})$. Sin embargo, en la segunda fase, después de la separación, el tiempo de absorción al margen es mucho más corto: las colas de su distribución son $O(e^{-2\gamma t})$ cuando $t \rightarrow \infty$. Por tanto, la velocidad con la cual el proceso escapa al margen depende de la longitud de esas dos fases y, consecuentemente, de $2N(t)$. Este no fue el caso para el tiempo T_0 porque en ambas fases la velocidad de escape fue la misma. Para la absorción al margen, si τ_{∞} es finita, la velocidad de escape es, "la mayor parte del tiempo" $e^{\gamma t}$. Si τ_{∞} es infinita (ver Lema 3.7), σ_{ν} puede ser finita o infinita y las cosas se hacen un poco más complejas; pero al menos si $2N(t)$ crece suficientemente rápido, la velocidad de escape que domina es $e^{2\gamma t}$.

Teorema 4.4 (a) Para cualquier $m \in \mathcal{D}_2 \cap \mathcal{M}_s$, el límite $\lim_{t \rightarrow \infty} e^{\gamma(t-s)} CR(t, s)m$ existe y es igual a

$$a(Km) \quad \text{si } \lim_{t \rightarrow \infty} N(t) = 0$$

$$\frac{a(Km)}{2N} \int_0^{\infty} e^{-\frac{t}{2N}} e^{\gamma t} CT(t) \Theta x_0 dt \quad \text{si } \lim_{t \rightarrow \infty} N(t) = N, 0 < N < \infty.$$

(b) Si $\lim_{t \rightarrow \infty} N(t) = \infty$, $\int_s^{\infty} \frac{du}{2N(u)} < \infty$ y

$$\int_s^{\infty} \frac{1}{2N(u)} e^{\gamma u} \exp \left\{ - \int_u^{\infty} \frac{dv}{2N(v)} \right\} du < \infty,$$

entonces

$$\lim_{t \rightarrow \infty} e^{2\gamma(t-s)} CR(t, s)m = r_2(m)x_0 \otimes x_0, \quad (4.4)$$

donde $r_2(m)$ es una constante no negativa y $r_2(m) = 0$ si y sólo si $Cm = 0$.

Demostración. El operador $U(t, s) = e^{\gamma t'} CR(t, s)m$ es dado por (4.1), donde B es reemplazada por C . Como en la prueba del Teorema 4.3, por el Lema 2.20 se puede aplicar el Teorema 4.1 a $U(t, s)$ con $V(t) = e^{\gamma t} AP(t)K$ y $W(t) = e^{\gamma t} CT(t)$ para obtener (a).

Para demostrar (b), partimos de (3.7) poniendo $e^{\gamma t} V(t) = e^{2\gamma t} AP(t)K$ y $w = t - s - v$. Entonces

$$e^{2\gamma t'} CR(t, s)m = U_1(t, s)m + U_2(t, s)m$$

donde

$$U_1(t, s)m = \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} e^{2\gamma t'} CT(t')m$$

y

$$U_2(t, s)m = \int_0^{t'} \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^t \frac{du}{2N(u)} \right\} e^{2\gamma(t-s-v)} CT(t-s-v) \Theta e^{2\gamma v} AP(v)K m dv$$

tomando $w = t - s - v$

$$U_2(t, s)m = \int_0^{t'} \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^t \frac{du}{2N(u)} \right\} e^{2\gamma w} CT(w) \Theta e^{2\gamma v} AP(v)K m dv.$$

Aquí $e^{\gamma v} V(v) = e^{2\gamma v} AP(v)K$ y $W(w) = e^{2\gamma w} CT(w)$, por lo que

$$\begin{aligned} U_2(t, s)m &= \int_0^{t'} \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^t \frac{du}{2N(u)} \right\} e^{\gamma v} W(w) \Theta V(v) m dv \\ &= \int_S Z dq_t, \end{aligned}$$

donde $Z(v, w) = W(w)\Theta V(v)K$ y q_t es la medida en el espacio $S = [0, \infty]^2$ concentrada en la diagonal $t - s - v = w$, con marginal μ_t en el eje v dada por

$$\mu_t(dv) = \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^t \frac{du}{2N(u)} \right\} e^{\gamma v}.$$

Argumentando como en el Lema 3.7 vemos que, cuando $t \rightarrow \infty$, q_t converge débilmente a la medida q concentrada en $\mathbb{R} \times \{\infty\}$ dada por

$$\mu(dv) = \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^{\infty} \frac{du}{2N(u)} \right\} e^{\gamma v}.$$

Para toda $v_0, w_0 > 0$ y t suficientemente grande,

$$q_t((v_0, \infty], (w_0, \infty]) = \int_{v_0}^{t-s-w_0} \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^t \frac{du}{2N(u)} \right\} e^{\gamma v} dv,$$

y cuando $t \rightarrow \infty$, esto converge a

$$q((v_0, \infty], (w_0, \infty]) = \int_{v_0}^{\infty} \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^{\infty} \frac{du}{2N(u)} \right\} e^{\gamma v} dv.$$

De las expresiones anteriores se tiene

$$\lim_{t \rightarrow \infty} e^{2\gamma t'} CR(t, s)m = \lim_{t \rightarrow \infty} U_1(t, s)m + \lim_{t \rightarrow \infty} U_2(t, s)m,$$

entonces,

$$\begin{aligned} \lim_{t \rightarrow \infty} U_1(t, s)m &= \lim_{t \rightarrow \infty} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} e^{2\gamma t'} CT(t')m \\ &= \lim_{t \rightarrow \infty} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} W(t-s)m \\ &= \exp \left\{ - \int_s^{\infty} \frac{du}{2N(u)} \right\} W(\infty)m \\ &= c(m) \exp \left\{ - \int_s^{\infty} \frac{du}{2N(u)} \right\} x_0 \otimes x_0, \end{aligned}$$

ya que, por el Lema 2.20, $W(\infty) = c(m)x_0 \otimes x_0$,

$$\begin{aligned} \lim_{t \rightarrow \infty} U_2(t, s)m &= \lim_{t \rightarrow \infty} \int_0^{t'} \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^t \frac{du}{2N(u)} \right\} e^{\gamma v} W(w) \Theta V(v) m dv \\ &= W(\infty) \left[\int_0^\infty \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^\infty \frac{du}{2N(u)} \right\} e^{\gamma v} \Theta e^{\gamma v} AP(v) m dv \right] x_0 \otimes x_0, \end{aligned}$$

por lo que (4.4) se cumple con

$$\begin{aligned} r_2(m) &= c(m) \exp \left\{ - \int_s^\infty \frac{du}{2N(u)} \right\} \\ &\quad + c \left(\int_0^\infty \frac{1}{2N(s+v)} \exp \left\{ - \int_{s+v}^\infty \frac{du}{2N(u)} \right\} e^{\gamma v} \Theta e^{\gamma v} AP(v) m dv \right). \end{aligned}$$

■

Es importante resaltar el hecho de que las medidas q_t que aparecen en la demostración del Teorema 4.4 no son medidas de probabilidad. Por definición de convergencia débil dado un espacio medible $(S, B(S))$ donde $S = [0, \infty]^2$, $q_t : B(S) \rightarrow \mathbb{R}_+$, y $q_\infty : B(S) \rightarrow \mathbb{R}_+$ entonces se dice que $\forall A \in B$, tal que $q(\partial A) = 0$, $\partial A = \bar{A} \cap \bar{A}^c$, $q_t(A)$ converge débilmente a $q(A)$ si $\forall \varepsilon > 0 \exists N$ tal que si $t > N$ entonces $|q_t(A) - q(A)| < \varepsilon$.

Las hipótesis en (b) equivalen a $\int_s^\infty \frac{du}{2N_\gamma(u)} < \infty$, donde $2N_\gamma(u) = 2N(u)e^{-\gamma u}$. Entonces (b) puede establecerse como sigue: si el tamaño de la población modificada $2N_\gamma(\cdot)$ crece rápidamente a infinito, entonces la masa de probabilidad de la cadena de Markov escapa al margen con la velocidad $e^{2\gamma t}$ y la distribución cuasi-estacionaria es $x_0 \otimes x_0$. Debido a (2.19), tanto la velocidad de escape al margen como la distribución cuasiestacionaria del modelo con deriva genética (4.4) son las mismas que las del modelo sin deriva dada por (2.19), i.e.,

Forma de la DCE del Modelo sin deriva: $\lim_{t \rightarrow \infty} e^{2\gamma t} CT(t)m = c(m)x_0 \otimes x_0$
 Forma de la DCE del Modelo con deriva: $\lim_{t \rightarrow \infty} e^{2\gamma t} CR(t, s)m = r_2(m)x_0 \otimes x_0$.

Forma de la DLC condicionada a no absorción en el margen con crecimiento poblacional lento

Fuera de la absorción en el origen $(0, 0)$, los dos modelos son similares sin tomar en cuenta la manera en que la población crece a infinito, es decir, en este caso tanto el tamaño de la población modificada como el tamaño de la población crecen rápidamente a infinito, i.e., exponencialmente, sin considerar las posibles combinaciones, i.e., crecimiento poblacional rápido con crecimiento poblacional modificado lento por ejemplo., lineal. La Proposición 4.5 cubre un escenario de interés: cuando el crecimiento de la población es lento; la Proposición 4.6 cubre el escenario cuando el crecimiento de la población es rápido acompañado por crecimiento lento del tamaño de la población modificada ($2N_\gamma(u) = 2N(u)e^{-\gamma u}$). Es de interés ver cuál es la forma de la distribución cuasi-estacionaria en ambos casos y establecer una comparación entre los dos modelos: con deriva y sin deriva genética aleatoria.

Proposición 4.5 Si $2N(t) = \frac{t}{\alpha}$ para $t \geq s > 0$ y algún $\alpha > 0$, entonces

$$\lim_{t \rightarrow \infty} t e^{\gamma t'} CR(t, s)m = a(Km)\alpha \int_0^\infty e^{\gamma u} CT(u)\Theta x_0 du, \quad (4.5)$$

donde $t' := t - s$; y bajo el supuesto de que $m \in \mathcal{M}_s, Cm \neq 0$ y así, la distribución límite condicional es un múltiplo de la integral del lado derecho de (4.5), y la velocidad de escape es $t \cdot e^{\gamma t'}, \gamma > 0$.

Demostración. Usando la hipótesis $2N(t) = \frac{t}{\alpha}$ se tiene que

$$-\int_v^w \frac{du}{2N(u)} = -\int_v^w \frac{\alpha}{t} dt = \alpha \ln\left(\frac{v}{w}\right).$$

Entonces

$$\begin{aligned} t' e^{\gamma t'} CR(t, s)m &= t' e^{\gamma t'} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} CT(t')m + t' \int_0^{t'} \frac{1}{2N(t-u)} \times \\ &\quad \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} e^{\gamma t'} CT(u) \Theta P(t-s-u) K m du. \end{aligned} \quad (4.6)$$

Como $\exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} = \exp \left\{ \ln \left(\frac{s}{t} \right)^\alpha \right\} = \left(\frac{s}{t} \right)^\alpha$, entonces $\exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} = \exp \left\{ \ln \left(\frac{t-u}{t} \right)^\alpha \right\} = \left(\frac{t-u}{t} \right)^\alpha$, y $\frac{1}{2N(t-u)} = \frac{\alpha}{t-u}$ y $e^{\gamma t'} = e^{\gamma(t-s)} = e^{\gamma u} e^{\gamma(t-s-u)}$. Sustituyendo las expresiones anteriores en (4.6) obtenemos

$$\begin{aligned} t' e^{\gamma t'} CR(t, s)m &= t' e^{\gamma t'} \left(\frac{s}{t} \right)^\alpha CT(t-s)m \\ &\quad + \frac{t'}{t} \int_0^{t'} \frac{\alpha t}{t-u} \left(\frac{t-u}{t} \right)^\alpha e^{\gamma u} CT(u) \Theta e^{\gamma(t'-u)} P(t'-u) K m du. \end{aligned}$$

luego,

$$\begin{aligned} \lim_{t \rightarrow \infty} t' e^{\gamma t'} CR(t, s)m &= \lim_{t \rightarrow \infty} t' e^{\gamma t'} \left(\frac{s}{t} \right)^\alpha CT(t-s)m + \lim_{t \rightarrow \infty} \frac{t'}{t} \int_0^{t'} \frac{\alpha t}{t-u} \left(\frac{t-u}{t} \right)^\alpha \times \\ &\quad e^{\gamma u} CT(u) \Theta e^{\gamma(t'-u)} P(t'-u) K m du. \end{aligned} \quad (4.7)$$

Por el Lema 2.20 el primer término del lado derecho de la Ecuación (4.7) tiende a cero, por lo que $t' e^{\gamma t'} CR(t, s)m$ tiende al mismo límite que

$$\begin{aligned} &\frac{t'}{t} \int_0^{t'} \frac{\alpha t}{t-u} \left(\frac{t-u}{t} \right)^\alpha e^{\gamma u} CT(u) \Theta e^{\gamma(t'-u)} P(t'-u) K m du \\ &= \alpha \left(1 - \frac{s}{t} \right) \int_0^{t-s} \left(1 - \frac{u}{t} \right)^{\alpha-1} e^{\gamma u} CT(u) \Theta e^{\gamma(t'-u)} P(t'-u) K m du. \end{aligned}$$

Debido a las desigualdades en (2.20), el integrando de arriba está dominado por

$$\alpha L^2 \mathbf{1}_{[0, t']} \left(1 + \frac{u}{s} \right) e^{-\gamma u} \leq \alpha L^2 \left(1 + \frac{u}{s} \right) e^{-\gamma u}$$

entonces por el teorema de convergencia dominada de Lebesgue

$$\begin{aligned}
& \lim_{t \rightarrow \infty} \alpha \left(1 - \frac{s}{t}\right) \int_0^{t-s} \left(1 - \frac{u}{t}\right)^{\alpha-1} e^{\gamma u} CT(u) \Theta e^{\gamma(t'-u)} P(t'-u) K m du \quad (4.8) \\
&= \lim_{t \rightarrow \infty} \alpha \left(1 - \frac{s}{t}\right) \int_0^{t-s} \left(1 - \frac{u}{t}\right)^{\alpha-1} e^{\gamma u} CT(u) \Theta V(t'-u) m du \\
&= \alpha (1) \int_0^\infty (1)^{\alpha-1} e^{\gamma u} CT(u) \Theta V(\infty) m du \\
&= \alpha \int_0^\infty e^{\gamma u} CT(u) \Theta V(\infty) m du,
\end{aligned}$$

debido a que $V(\infty) = a(Km)x_0$; (4.8) es igual a

$$\alpha \int_0^\infty e^{\gamma u} CT(u) \Theta a(Km)x_0 du = a(Km)\alpha \int_0^\infty e^{\gamma u} CT(u) \Theta x_0 du, \quad (4.9)$$

como es deseado. ■

La hipótesis de la Proposición 4.5 indica que el crecimiento de la población no es tan rápido y la forma de la distribución cuasiestacionaria (4.9) es diferente a la del modelo sin deriva (2.19), contrario al resultado del Teorema 4.4 (b), es decir, que si la población crece rápidamente, i.e., exponencialmente con parámetro al menos γ , la distribución cuasiestacionaria y la velocidad de escape de la masa al margen son la misma como en el modelo sin deriva.

Proposición 4.6 *Supóngase que la función $2N(\cdot)$ es continua para argumentos suficientemente grandes y que existe una κ con $0 \leq \kappa < \gamma$ y tal que $\lim_{t \rightarrow \infty} \frac{2N(t)}{e^{\kappa t} \phi(t)}$ existe y es una constante distinta de cero, donde $\phi(\cdot)$ es diferenciable y positiva para argumentos grandes y $\lim_{t \rightarrow \infty} \frac{\phi'(t)}{\phi(t)} = 0$. Si $\int_s^\infty \frac{du}{e^{\kappa u} \phi(u)} < \infty$, entonces*

$$\lim_{t \rightarrow \infty} 2N(t) e^{\gamma t} CR(t, s) m = a(Km) \int_0^\infty e^{\kappa u} e^{\gamma u} CT(u) \Theta x_0 du, \quad m \in \mathcal{M}_s, Cm \neq 0.$$

Demostración. Para cualquier $\varepsilon > 0$ y t suficientemente grande

$$\frac{d}{dt}\phi(t)e^{-\varepsilon t} = \phi(t) \left[\frac{\phi'(t)}{\phi(t)} - \varepsilon \right] e^{-\varepsilon t} < 0,$$

lo cual implica que $\phi(t)e^{-\varepsilon t}$ es una función decreciente para t suficientemente grande. Por lo tanto el límite $\lim_{t \rightarrow \infty} \phi(t)e^{-\varepsilon t} = \lim_{t \rightarrow \infty} \frac{\phi(t)}{e^{\varepsilon t}}$ existe.

Aplicando la regla de L'Hopital y multiplicando por $\frac{\phi(t)}{\phi(t)}$, y por el supuesto de que $\lim_{t \rightarrow \infty} \frac{\phi'(t)}{\phi(t)} = 0$, obtenemos

$$\lim_{t \rightarrow \infty} \phi(t)e^{-\varepsilon t} = \lim_{t \rightarrow \infty} \frac{\phi'(t)}{\varepsilon e^{\varepsilon t}} \frac{\phi(t)}{\phi(t)} = \lim_{t \rightarrow \infty} \frac{\phi'(t)}{\phi(t)} \frac{\phi(t)}{\varepsilon e^{\varepsilon t}} = 0. \quad (4.10)$$

Tenemos que

$$\begin{aligned} \lim_{t \rightarrow \infty} 2N(t)e^{\gamma t} CR(t, s)m &= \lim_{t \rightarrow \infty} 2N(t)e^{\gamma t} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} CT(t')m \\ &\quad + 2N(t) \int_0^{t'} \frac{1}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} e^{\gamma t} \\ &\quad \times CT(u)\Theta P(t-s-u)K m du, \end{aligned} \quad (4.11)$$

por la Ecuación (4.10) y las desigualdades de (2.20) se tiene que

$$2N(t)e^{\gamma t} \exp \left\{ - \int_s^t \frac{du}{2N(u)} \right\} CT(t')m \rightarrow 0.$$

Entonces

$$\begin{aligned} &\lim_{t \rightarrow \infty} 2N(t)e^{\gamma t} CR(t, s)m = \\ &\int_0^{t'} \frac{2N(t)}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} e^{\gamma u} CT(u)\Theta e^{-\gamma u} e^{\gamma t'} AP(t'-u)K m du \\ &= \int_0^{t'} \frac{2N(t)}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} e^{\gamma u} CT(u)\Theta e^{\gamma(t'-u)} AP(t'-u)K m du \\ &= \int_0^{t'} \frac{2N(t)}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} e^{-\gamma u} e^{2\gamma u} CT(u)\Theta e^{\gamma(t'-u)} AP(t'-u)K m du \\ &= \int_0^{t'} g_t(u) Z_t(u) m du, \end{aligned} \quad (4.12)$$

donde

$$\begin{aligned} g_t(u) &= \frac{2N(t)}{2N(t-u)} \exp \left\{ - \int_{t-u}^t \frac{dv}{2N(v)} \right\} e^{-\gamma u}, \text{ y} \\ Z_t(u)m &= e^{2\gamma u} CT(u) \Theta e^{\gamma(t'-u)} AP(t'-u) Km. \end{aligned} \quad (4.13)$$

Basta demostrar que

$$\lim_{t \rightarrow \infty} 2N(t) e^{\gamma t} CR(t, s)m = \lim_{t \rightarrow \infty} \int_0^{t'} g_t(u) Z_t(u) m du$$

o bien, que el lado derecho de la Ecuación (4.12) converge a

$$\int_0^\infty a(Km) e^{-(\gamma-\kappa)u} e^{2\gamma u} CT(u) \Theta x_0 du.$$

Por otro lado tenemos que, por el teorema de convergencia dominada de Lebesgue,

$$\lim_{t \rightarrow \infty} \int_0^\infty e^{-(\gamma-\kappa)u} Z_t(u) m du = \int_0^\infty a(Km) e^{-(\gamma-\kappa)u} e^{2\gamma u} CT(u) \Theta x_0 du,$$

en efecto,

$$\lim_{t \rightarrow \infty} Z_t(u)m = e^{2\gamma u} CT(u) \lim_{t \rightarrow \infty} e^{\gamma(t'-u)} AP(t'-u) Km = a(Km) e^{2\gamma u} CT(u) \Theta x_0$$

y por (2.20) $\| Z_t(u)m \|_{\mathcal{M}} \leq L^2$. Por último

$$\left\| \int_0^\infty e^{-(\gamma-\kappa)u} Z_t(u) m du - \int_0^{t'} g_t(u) Z_t(u) m du \right\|_{\mathcal{M}} \leq L^2 \left\| e^{-(\gamma-\kappa)\cdot} - g_t(\cdot) \right\|_{L^1(0, \infty)},$$

donde $\| \cdot \|_{L^1(0, \infty)}$ es la norma en el espacio $L^1(0, \infty)$ de funciones absolutamente integrables en $(0, \infty)$. Como $\lim_{t \rightarrow \infty} g_t(u) = e^{-(\gamma-\kappa)u}$ y $g_t(\cdot) \geq 0$, por el teorema de Scheffé (ver Billingsley, 1986, pág. 218), es suficiente probar que

$$\lim_{t \rightarrow \infty} \int_0^{t'} g_t(u) du = \frac{1}{\gamma - \kappa}.$$

Para esto último sustituyendo $u = t - s - u$ y $g_t(u)$ con el lado derecho de la Ecuación (4.13) se tiene que

$$\begin{aligned} \int_0^{t'} g_t(u) du &= \lim_{t \rightarrow \infty} 2N(t) \int_0^{t'} \frac{1}{2N(s+u)} \exp \left\{ - \int_{s+u}^t \frac{dv}{2N(v)} \right\} e^{\gamma(t'-u)} du \\ &= \lim_{t \rightarrow \infty} \frac{2N(t)}{e^{\kappa t} \phi(t)} \frac{\int_0^{t'} \frac{1}{2N(s+u)} \exp \left\{ - \int_{s+u}^{\infty} \frac{dv}{2N(v)} \right\} e^{\gamma u} du}{\exp \left\{ - \int_t^{\infty} \frac{dv}{2N(v)} \right\} e^{(\gamma-\kappa)t} e^{-\ln \phi(t)}}, \end{aligned}$$

aplicando la regla de L'Hopital a a la segunda fracción del lado derecho de la ecuación anterior

$$\begin{aligned} &\lim_{t \rightarrow \infty} \frac{\int_0^{t'} \frac{1}{2N(s+u)} \exp \left\{ - \int_{s+u}^{\infty} \frac{dv}{2N(v)} \right\} e^{\gamma u} du}{\exp \left\{ - \int_t^{\infty} \frac{dv}{2N(v)} \right\} e^{(\gamma-\kappa)t} e^{-\ln \phi(t)}} \\ &= \lim_{t \rightarrow \infty} \frac{\frac{1}{2N(t)} \exp \left\{ - \int_t^{\infty} \frac{dv}{2N(v)} \right\} e^{\gamma t}}{\left[\frac{1}{2N(t)} + (\gamma - \kappa) - \frac{\phi'(t)}{\phi(t)} \right] \exp \left\{ - \int_t^{\infty} \frac{dv}{2N(v)} \right\} e^{\gamma t} e^{-\kappa t} \frac{1}{\phi(t)}} \\ &= \lim_{t \rightarrow \infty} \frac{e^{\kappa t} \phi(t)}{2N(t)} \frac{1}{\left[\frac{1}{2N(t)} + (\gamma - \kappa) - \frac{\phi'(t)}{\phi(t)} \right]} \end{aligned}$$

se prueba que

$$\begin{aligned} &\lim_{t \rightarrow \infty} \frac{e^{\kappa t} \phi(t)}{2N(t)} \frac{1}{\left[\frac{1}{2N(t)} + (\gamma - \kappa) - \frac{\phi'(t)}{\phi(t)} \right]} \\ &= \frac{1}{const.} \frac{1}{(\gamma - \kappa)}. \end{aligned}$$

■

Situaciones típicas donde aplica esta Proposición son $\phi(u) = u^\alpha$ con $(\alpha - 1)^2 + \kappa > 0$, y $\phi(u) = \ln u$ con $\kappa > 0$.

Capítulo 5

Conclusiones y discusión

Uno de los artículos pioneros que motivaron la teoría de los capítulos anteriores es el de Durrett et al. (1999), donde se introduce un modelo de cadenas de Markov de tiempo continuo para la evolución de microsatélites donde se prueba la existencia de una única distribución estacionaria, suponiendo que no existe un estado absorbente en la longitud de los microsatélites.

Bajo el supuesto de que en la evolución de los microsatélites existe un estado absorbente, cuando la longitud de éstos es uno, no hay una distribución estacionaria, pero puede existir una distribución cuasiestacionaria. Bajo este supuesto Bobrowski (2004) obtuvo tanto la función de velocidad de escape al margen y al origen del proceso bivariado en la evolución de los dos microsatélites, como la forma de la distribución cuasiestacionaria del modelo con deriva genética (tomando en cuenta el coalescente de J.F.Kingman) y del modelo sin deriva genética (sólo se considera el mecanismo de mutaciones). Más aún, estudió el impacto que tiene el tamaño de población variable $2N(t)$ en la forma de la distribución

cuasiestacionaria y en la función de velocidad de escape.

La forma de la distribución cuasiestacionaria y la función de velocidad de escape al origen $(0, 0)$ y al margen M , Ecuaciones (2.18) y (2.19) respectivamente, presentadas en el Lema 2.20 para el **modelo sin deriva** genética son las siguientes:

No absorción	DCE	$\mathbf{v}(\mathbf{t})$
$(0, 0)$	$b(m) [x_0 \otimes e_0 + e_0 \otimes x_0]$	$e^{\gamma t}, \gamma > 0$
$M = \{(i, j) : i = 0 \vee j = 0\}$	$c(m) [x_0 \otimes x_0]$	$e^{2\gamma t}, \gamma > 0,$

mientras que la forma de la distribución cuasiestacionaria y la función de velocidad de escape tanto en el origen como en el margen, presentadas en los Teoremas 4.3 y 4.4, respectivamente, para el **modelo con deriva genética aleatoria** bajo diferente comportamiento asintótico de $N(t)$ son las siguientes:

$\mathbf{N}(\mathbf{t})$	DCE en $(0, 0)$, $\mathbf{v}(\mathbf{t}) = \mathbf{e}^{\gamma t}$	DCE en M, $\mathbf{v}(\mathbf{t}) = \mathbf{e}^{2\gamma t}$
$\lim_{t \rightarrow \infty} N(t) = 0,$	$a(Km)\Theta x_0$	$a(Km)$
$\lim_{t \rightarrow \infty} N(t) = N,$	$\frac{a(Km)}{2N} \int_0^\infty e^{-\frac{t}{2N}} e^{\gamma t} BT(t)\Theta x_0 dt$	$\frac{a(Km)}{2N} \int_0^\infty e^{-\frac{t}{2N}} e^{\gamma t} CT(t)\Theta x_0 dt$
$\lim_{t \rightarrow \infty} N(t) = \infty,$	$r_1(m) [x_0 \otimes e_0 + e_0 \otimes x_0]$	$r_2(m)x_0 \otimes x_0,$
		$\int_s^\infty \frac{du}{2N(u)} < \infty,$ y $\int_s^\infty \frac{du}{2N(u)} e^{\gamma u} < \infty.$

La distribución límite condicional de dos cadenas de Markov condicionadas a no absorción en el estado $(0, 0)$ (Teorema 4.3), es igual a Θx_0 si $\lim_{t \rightarrow \infty} 2N(t) = 0$, es decir, la distribución se concentra en la diagonal, por lo que la deriva genética aleatoria alcanza su fin: fijación de alelos. Si $\lim_{t \rightarrow \infty} 2N(t) = \infty$, entonces la distribución límite condicional es igual a la del modelo sin deriva genética. Si $\lim_{t \rightarrow \infty} 2N(t) = 2N > 0$, entonces la distribución límite condicional es un múltiplo escalar de la transformada de Laplace $\mathcal{L}(\lambda)$ de $t \mapsto e^{\gamma t} BT(t)\theta x_0$ en $\lambda = \frac{1}{2N}$.

Por otro lado, los resultados del Teorema 4.4 (b) establecen que si tanto la población como el tamaño de población modificado $2N_\gamma(u) = 2N(u)e^{-\gamma u}$ crecen rápidamente a infinito (exponencialmente), i.e. se cumplen $\int_s^\infty \frac{du}{2N(u)} < \infty$, y $\int_s^\infty \frac{du}{2N(u)} e^{\gamma u} < \infty$, entonces la masa de probabilidad de la cadena de Markov escapa al margen con la velocidad $e^{2\gamma t}$ y la distribución cuasiestacionaria es $x_0 \otimes x_0$. La velocidad de escape al margen como la distribución cuasiestacionaria son las mismas que las del modelo sin deriva.

La condición necesaria para obtener los resultados previos en esos dos casos es que $2N(t)$ crezca tan rápidamente a infinito que

$$\int_s^\infty \frac{du}{2N(u)} < \infty. \quad (5.1)$$

En las Proposiciones 4.5 y 4.6, se demostró que si el crecimiento de $N(t)$ no es tan rápido, la condición (5.1) falla y tanto la DCE como la función de velocidad de escape son diferentes de las que aparecen en (2.19), es decir, en el modelo sin deriva genética.

Por los resultados de la Proposición 4.6, la condición (5.1) no garantiza proximidad de los modelos con y sin deriva, respecto a la forma de la DCE y de la función de velocidad de escape. Para tal proximidad es necesario el crecimiento rápido del tamaño de la población modificado $2N_\gamma(t) = 2N(t)e^{-\gamma t}$.

Como se muestra en la Proposición 4.6, aún si el tamaño de la población crece rápidamente a infinito pero el tamaño de la población modificado no cumple con esta condición, la velocidad de escape al margen y la DCE pueden ser diferentes a las del modelo sin deriva para diferentes escenarios de crecimiento de $N(t)$. La Proposición 4.6 demuestra que el tamaño efectivo de la población $2N(t)$ controla la función de la velocidad de escape. Para ser más específicos, escogiendo $2N(t)$ que sea una función que crezca más rápido que lin-

ealmente pero más lento que $e^{\gamma t}$, obtenemos la función de velocidad de escape en la forma $v(t) = 2N(t)e^{\gamma t}$.

Es interesante notar que, para poblaciones que se expanden más lento que $e^{\gamma t}$, las distribuciones cuasiestacionarias tienen una forma muy similar: todas son múltiplos escalares de la transformada de Laplace

$$R_\lambda \Theta x_0 = \int_0^\infty e^{-\lambda u} e^{2\gamma u} CT(u) \Theta x_0 du. \quad (5.2)$$

Resumimos las predicciones del modelo como sigue.

- La fuerza de la deriva genética actúa aún en poblaciones muy grandes si el crecimiento de esas poblaciones no es suficientemente rápido, como se muestra en las simulaciones para el caso de crecimiento poblacional lineal, Figura 3.4
- Para distribuciones cuasiestacionarias, la condición para que éstas sean similares en los modelos con y sin deriva es que se requiere que el tamaño de la población modificada $2N_\gamma(\cdot) = 2N(\cdot)e^{-\gamma t}$ crezca tan rápido que

$$\int_s^\infty \frac{du}{2N_\gamma(u)} < \infty$$

(compare con (5.1)).

- El patrón de crecimiento de la población controla la función de la velocidad de escape; si $2N(t)$ es una función que crece a infinito al menos linealmente pero más lento que $e^{\gamma t}$, entonces $v(t) = 2N(t)e^{\gamma t}$.
- La distribución cuasiestacionaria depende de $2N(\cdot)$; para poblaciones asintóticamente estables y poblaciones que se expanden con crecimiento lento, la DCE es un múltiplo

escalar de la transformada de Laplace con λ en la transformada de Laplace dependiente del tipo de crecimiento de $2N(\cdot)$; si la población crece al menos exponencialmente con exponente γ , la DCE es $x_0 \otimes x_0$.

Uno de los puntos importantes de esta tesis que valdría la pena discutir es acerca de las simulaciones de la evolución de dos alelos bajo el modelo de Wright-Fisher-Moran con diferente comportamiento asintótico del tamaño de población variable $2N(t)$. Aunque corresponden a la evolución de dos microsatélites (dos cadenas de Markov independientes) sólo bajo la acción de la deriva genética aleatoria, ayudan en la interpretación genética de resultados de Beaumont (1999), Bobrowski, et al. (2001), Bobrowski (2001) y Bobrowski (2004) y dan una descripción de los teoremas demostrados, específicamente del Lema 3.7 y Teoremas presentados en esta tesis, desde la perspectiva de la autora de la misma. Estas simulaciones podrían servir como base para la simulación de la evolución de dos microsatélites, por medio de cadenas de Markov dependientes e introduciendo un modelo específico para el proceso de mutaciones. Aunque la autora considera que los resultados o interpretaciones serían similares, queda la pregunta de si realmente se podría hacer un paralelismo en las interpretaciones.

Las interpretaciones genéticas que se dan en este trabajo son hasta cierto punto subjetivas en el sentido de que se sabe poco acerca del origen de los microsatélites así como de su evolución, y las interpretaciones que se hacen de las simulaciones con respecto a la distribución conjunta de τ y σ , son esencialmente de la autora de este trabajo.

Es importante mencionar que aunque el trabajo de recolección y organización de información sobre el tema en cuestión fue bastante arduo, la parte más complicada fue

respecto a las interpretaciones genéticas. Una de los puntos a discutir en este aspecto es el relacionado con el tamaño de población modificado $2N_\gamma(\cdot) = 2N(\cdot)e^{-\gamma t}$ que aparece en el Teorema 4.4. Matemáticamente, es una condición necesaria para la establecer la similitud de las distribuciones cuasiestacionarias para los modelos con y sin deriva genética, el cual se presenta en el desarrollo del Teorema 4.4; sin embargo, una interpretación genética de la misma no es clara.

El modelo que es tratado en esta tesis, con tamaño de población variable y determinístico, es uno de los supuestos que podrían discutirse. En la realidad, el tamaño de las poblaciones puede considerarse estocástico y no determinístico, pues se podrían dar casos donde, debido a fenómenos naturales o por migración una población se extinga repentinamente, o que su crecimiento sea más lento que lo que se espera o más rápido, entre otras causas.

También otro de los puntos a discutir es el relacionado con las otras dos fuerzas genéticas: selección natural y recombinación. Bajo el modelo de Wright-Fisher-Moran, no se consideran estas dos fuerzas genéticas, por lo que queda abierta la posibilidad de estudiar las distribuciones cuasiestacionarias en la evolución de microsatélites en presencia de las cuatro fuerzas genéticas descritas en esta tesis: mutación, deriva genética aleatoria, selección natural y recombinación, conjuntamente con la teoría del coalescente así como el desarrollo de nuevas simulaciones.

Aunque pareciera que el modelo desarrollado por Bobrowski, et al. (2004), no es del todo real, este modelo sirve como base para desarrollar nuevos modelos, así como el modelo de Wright-Fisher ha servido para describir el comportamiento asintótico de distribuciones

con tamaño de población variable.

Bibliografía

ANDERSON, W.A.(1991). *Continuos-Time Markov Chains*. Springer-Verlag.

BALDING, D.J., DONELLY, P., GRIFFITHS, R.C. AND TAVARÉ, S. (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505-518.

BARTLE, R. G. (1966). *The elements of integration and Lebesgue Measure*. Jhon Wiley & sons.

BEAUMONT, M. (1999). Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013-2029.

BERTOIN, J. (2005). Procesos de coagulación. *Apuntes de IV escuela de Probabilidad*, Morelia, Mich., México D.F.

BILLINGSLEY, p. (1986). *Probability and Measures*. Jhon Wiley & sons.

BOBROWSKI, A (2004). Quasi-stationary distributions of a pair of Markov chains related to time evolution of DNA locus. *Ad. Appl. Prob.* **36**, 57-77.

BOBROWSKI, A., KIMMEL, M. CHAKRABORTY, R. AND ARINO, O. (2001). A semi-group representation and asymptotic behavior of certain statistics of the Fisher-Wright-Moran coalescent. In *Handbook of Statistics*, Vol. 19, *Stochastics Processes: Theory and Methods*, eds C. R. Rao and D. N. Shanbhag, Elsevier Science, Amsterdam.

- BOBROWSKI, A., WANG, N., KIMMEL, M., CHAKRABORTY, R. (2002). Nonhomogeneous infinite sites model under demographic change: mathematical description and asymptotic behavior of pairwise distributions. *Math. Biosci.* **175**, 83-115.
- BROWN, T.A. (2002). *Genomes*. Jhon Wiley & sons.
- DONELLY, P. AND TAVARÉ, S. (1995). Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* **29**, 401-421.
- DURRET, R. AND KRUGYAK, S. (1999). A new stochastic model of microsatellite evolution. *J. Appl. Prob.* **36**, 621-631.
- DURRET, R.(2002). *Probability Models for DNA Sequence Evolution*. Springer-Verlag.
- ENCICLOPEDIA ELECTRONICA. <http://es.wikipedia.org>
- EWENS, W.J.. (2006). *Mathematical Population Genetics: Lecture Notes. 2nd. summer probability School*. Cornell University, New York, USA.
- FERRARI, P. A., KESTEN, H., MARTINEZ, S. AND PICCO. (1995). Existence of quasi-stationary distribution. A renewal dynamical approach. *The Annals of Probability*, Vol. **23**, 501-521.
- GILLESPIE, J. (1998). *Population Genetics: A concise guide*. The Jhon Hopkins. University Press.
- GRIFFITHS, R. C. AND TAVARÉ, S. (1994a). Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. London B* **344**, 403-410.
- GRIFFITHS, R. C. AND TAVARÉ, S. (1994b). Ancestral inference in population genetics. *Stat. Sci.* **9**, 307-319.
- HARTL, D. L. AND CLARK, A. G. (1997). *Principles of Population Genetics*. Sinauer,

Sunderland, Massachusetts.

KARLIN, S., TAYLOR, H. (1998). An introduction to Stochastic Modelling, Third edition, Academic Press, USA.

KARLIN, S., TAYLOR, H. (1998). A Second course of stochastic processes. Academic Press, USA.

KINGMAN, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13**, 235-248.

KINGMAN, J. F. C. (1982). On the genealogy of large populations. *J. Appl. prob.* **19**, 27-43.

NORDBORG, M. (2003). Coalescent theory. *Handbook of Statistical Genetics* **2**, Wiley, London, UK

POLANSKI, A, KIMMEL, M. (2003). Population Genetics models for the statistics of DNA samples under diferent demographic scenarios-Maximum Likelihood versus aproxi-mate methods. *Int. J. Appl. Math. Comput. Sci.*, **13**, 347-355.

PAKES, A. G. (1995). Quasi-stationary laws for Markov processes; exmples of an always proximate absorbing state. *Adv. Appl. Prob.* **27**, 120-145.

RESNICK, S. (1992). Adventures in Stochastic Processes. Birkhauser.

SIBLY, R. M., WHITTAKER, J. C. AND TALBOT, M. (2001). A maximum likelihood approach to fitting equilibrium models of microsatellite evolution. *Mol. Biol. Evol.* **18**, 413-417.

TAJIMA, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.

VAN DOORN, ERIK A. (1991). Quasi-stationary distributions and convergence to quasi-

stationary of birth-death processes. *Adv. Appl. Prob.* **23**, 683-700.

XU, H. AND FU, Y. (2004). Estimating effective population size or mutation rate with microsatellites. *Genetics* **166**, 555-563.