



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

UN MODELO DE SELECCIÓN QUE INCLUYE COMPETENCIA
POR REDES ECOLÓGICAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

SOFÍA ROZANES VALENZUELA

TUTOR

DR. ADRIÁN GONZÁLEZ CASANOVA SOBERÓN



CIUDAD UNIVERSITARIA, Cd. Mx., MARZO DE 2021



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

*A mis tres mosqueteros:
Atto, Atreyu y Morisquito*

Agradecimientos

Gracias a mi tutor, amigo y Bruder, Adrián, por confiar en mí y ayudarme a confiar en mí misma; por ofrecerme trabajar en este proyecto que resultó no sólo una manera de, por fin, titularme, sino que además fue apasionante e interesante.

Gracias a mis sinodales de tesis —Clara, Verónica, Laura y Freddy— que aceptaron leer mi trabajo y que lo enriquecieron con sus valiosas observaciones y sugerencias.

Gracias a mis maestros de la carrera cuyas enseñanzas y entusiasmo dejaron huella en mí.

Gracias a los amigos queridos que conocí e hice en los corredores y salones de la Facultad de Ciencias y con los que he compartido no sólo clases, pero también viajes, anécdotas y muchas muchas discusiones.

Gracias a la Universidad Nacional Autónoma de México por haberme ofrecido a mí y a muchos mexicanos la oportunidad de estudiar una carrera profesional.

Gracias a mi mamá, a mi papá y a mi hermano por ser mi destino; el mejor que me hubiese podido tocar.

Y por último, gracias a mi esposo y a mi gata por ser mi familia y mi hogar.

Resumen

A través de un programa hecho en Python simulamos un modelo de Wright-Fisher con selección que incluye competencia por interacciones de redes ecológicas. Revisaremos varias de estas interacciones para ver cómo cada una de ellas afecta la evolución del perfil genético de una población; en particular, su estabilidad, es decir, la capacidad del sistema de regresar a su constitución original después de la presencia de una pequeña perturbación y de mantenerse en este estado por un largo periodo. La introducción de estas formas de competencia entre los alelos en el modelo de Wright-Fisher se trata de un tema bastante novedoso, [5].

En un proceso de Wright-Fisher neutral, en el que no existe ningún tipo de interacción, los individuos de la generación g eligen uniformemente a sus padres entre los individuos de la generación $g - 1$. Por tratarse de una población finita, sabemos que eventualmente alguno de los tipos alélicos se fijará, sin embargo, este proceso puede tomar numerosas generaciones.

Uno supone que cuando hay interacción, este proceso será mucho más rápido, porque puede uno pensar que siempre existe aquel alelo que sea más “fuerte” que los otros y que, por lo tanto, éste va a exterminar a los más débiles, como sucede en el caso de una interacción Transitiva. Sin embargo, como revisaremos más adelante, esto no siempre es así.

En una de las interacciones que probamos en el programa todos los alelos tienen un único contrinicante contra el cual siempre ganan; y, de igual modo, otro contra el cual siempre pierden. A esta interacción se le llama Órbita puesto que al hacer el diagrama que ilustra la red ecológica propia de la interacción, se forma un circuito de flechas; cada una de ellas va del alelo vencido al alelo vencedor. En el caso particular de tres alelos, la Órbita se comporta como el clásico juego de niños: ‘Piedra, Papel o Tijeras’.

Al simular esta interacción, se pudo observar en las gráficas un comportamiento muy bello en el que las frecuencias de los alelos aumentan y disminuyen y vuelven a aumentar en forma de ondas; mostrándonos así, que un decremento en la frecuencia de cierto alelo, no necesariamente se traduce en una desventaja, sino que, por el contrario, el alelo puede favorecerse por este declive momentáneo. Esta interacción retarda el tiempo de fijación en comparación con la interacción Transitiva. Además, podemos decir que, en esta interacción ningún tipo alélico tiene ventaja selectiva sobre los demás.

Abstract

By using a program built in Python, we have simulated a Wright-Fisher model with selection that includes competition by interactions of ecological networks. We will review some of these interactions in order to discover how, each one of these, affects the population's genetical profile evolution. Specifically, the stability of the population, i.e., the capacity of a system to go back to its original constitution after the presence of a small perturbation, and to remain in this state for a long period. The introduction of these kinds of competition between alleles to the Wright Fisher model is a fairly new topic [5].

In the neutral Wright-Fisher model, in which it doesn't exist any sort of interactions, the individuals choose their parents uniformly from the previous generation. Because we are dealing with a finite population, we know that eventually one of the alleles will fix; however, this process might take several generations. One might think, that when there is interaction, this process will be much faster, because one could suppose that there will always exist an allele which is "stronger" than the others, so this one will kill all the weaker alleles, as it happens with the Transitive interaction. Nonetheless, as we'll see later, this is not always the case.

In one of the interactions that we have tried, each allele wins against one of its opponents, and loses against another one, in a way that these relations are not repeated for any other allele. We call this interaction Orbit, because it is represented with an orbit made by arrows that go from the allele that loses to the one that wins. In the particular case of three alleles, the Orbit behaves like the kids' game "Rock-Paper-Scissors".

When we simulated this interaction in the program, we could observe in the graphics a beautiful behavior in the frequency processes of the alleles, which went up and down and up again like waves, proving that a decrement in an allele frequency doesn't necessarily mean this allele is in selective disadvantage, but in the contrary, this allele could be favored by this momentary frequency decline. As we'll see later on, this interaction causes a longer fixation time compared with the transitive interaction. Furthermore, we can say that in this interaction, none of the alleles have a selective advantage over the rest of the alleles.

Notación

Definamos la notación que vamos a utilizar de ahora en adelante:

\mathbb{N}^*	Números naturales: $\mathbb{N}^* := 1, 2, \dots, = \mathbb{Z}^+$
\mathbb{N}	Números naturales incluyendo al cero: $\mathbb{N} = \mathbb{N}^* \cup \{0\}$
$[n]$	sea $n \in \mathbb{N}^*$, entonces $[n] = \{1, 2, \dots, n\}$
$ D $	Denotará el cardinal de un conjunto discreto D
N	El tamaño de la población: $N \in \mathbb{N}^*$
k	Número de alelos: $k \in \mathbb{N}^*$, $k < \infty$
E	Espacio alélico: $E = \{1, 2, \dots, k\} = [k]$, $ E = k$
g	Generación: $g \in \mathbb{Z}$
V	Matriz en donde el vértice (g, i) corresponde al i -ésimo individuo de la generación g : $V = \mathbb{Z} \times [N]$
v	Individuo en la posición i y en la generación g , $v = (i, g)$
$\{X_g^a\}$	Proceso de frecuencia del alelo a . A veces también representado como: $\{X^a(g)\}$
X_0^a	También $X^a(0)$. Estado inicial del proceso de frecuencia del alelo a : $X_0^a = \frac{\sum_{i=1}^N \mathbb{1}_{\xi(i,0)=a}}{N}$
$\xi(i, g)$	El tipo alélico del individuo i en la generación g
$\{Y_g^a\}$	Proceso del número de individuos del tipo a en la generación g : $\{Y_g^a\} = \{X_g^a\}$
p_{rs}	Probabilidad de transición del proceso $\{Y_g^a\}$ del estado r al s : $p_{rs} = \mathbb{P}[Y_{g+1} = s Y_g = r]$
H^B	Tiempo de llegada hasta la condición B : $H^B = \inf\{n \geq 0 : X_n \in B\}$
t_i^B	Tiempo esperado hasta la condición B siendo que se comenzó en i al tiempo cero: $\mathbb{E}[H^B X_0 = i]$
τ_{fix}	Tiempo de fijación: $\tau_{fix} = \min\{g : X_g^i = 1 \text{ para alguna } i \in [k]\}$
$\{W_t\}_{t \geq 0}$	Movimiento Browniano estándar
ρ_n	El espacio de particiones en $[n]$
$A^N(v)$	Todos los ancestros de v : $A^N(v) := \{s \in V : s \text{ es un ancestro de } v\}$

$A^N(v_1, \dots, v_m)(r)$	Los ancestros de una muestra (v_1, \dots, v_m) que viven en la generación $g - r$
$\{D_r^N : r \in [N]\}$	El proceso ancestral de una muestra: $D_0^N = m$; $D_r^N(v) = A^N(v_1, \dots, v_m)(r) $
ω_m	Tiempo al ancestro común más reciente de una muestra de m individuos
$q_{\alpha, \beta}$	Tasas de transición del Coalescente de Kingman, donde $\alpha, \beta \in \rho_m$
P_v	Número de observaciones hechas por el individuo v
$\{R(v, p)\}$	Individuos observados en la generación anterior por v . Con reemplazo y distribución uniforme en $[N]$
H_v	Espacio de todos los posibles padres potenciales de v : $H = \{(z_1, z_2) : z_1 \in E; z_2 \in E \cup \{0\}; \text{si } z_2 \neq 0 \Rightarrow z_1 < z_2\}$
C_N	Regla de coloración o interacción selectiva
x	En lo relativo al programa y el código de éste, x será la probabilidad de hacer dos observaciones en lugar de sólo una
<i>seed</i>	Semilla aleatoria del código

Índice general

Agradecimientos	II
Resumen	III
Abstract	IV
Notación	v
1. Introducción	1
2. Marco teórico	5
2.1. Proceso de frecuencia de alelos	8
2.1.1. Probabilidad de Transición	10
2.1.2. Modelo de Moran	11
2.1.3. Fijación	13
2.1.4. Difusión de Wright-Fisher	16
2.2. Proceso ancestral	18
2.2.1. Coalescente de Kingman	22
2.3. Dualidad de los Momentos	23
2.4. Selección Natural	23
3. Construcción del modelo	26
3.1. Nuestro programa	32
3.1.1. Especificidades de nuestro programa	32
3.1.2. Reglas de coloración en nuestro programa	33
3.1.3. Funcionamiento del programa	36
3.1.4. Logística de operación del código del programa	38
4. Discusión	42
4.1. Analizando las gráficas	42

4.2. Estabilidad	51
4.2.1. Estabilidad con respecto al tiempo de fijación	52
4.2.2. Estabilidad con respecto a la varianza de las frecuencias alélicas	57
4.2.3. Varianza en los tiempos de extinción.	59
4.3. Otros asuntos	60
4.3.1. Más de dos observaciones	60
4.4. Conclusiones biológicas	61
A. Manual de usuario	63
A.0.1. User manual (In English)	64
B. Código del programa en Python	66

Índice de figuras

2.1. Ejemplo de un coalescente para una muestra de tamaño 5	19
3.1. Matriz V	27
3.2. Generación inicial	27
3.3. Coloración en modelo neutral	28
3.4. Observaciones	29
3.5. Naranja le gana siempre a azul	32
3.6. Diagrama de la regla transitiva	33
3.7. Diagrama de regla Gana el Otro	34
3.8. Diagrama de regla órbita con 3 alelos: Piedra, Papel o Tijera	34
3.9. Diagrama de la regla Órbita con 4 alelos. Las flechas van del tipo que tiene menos ventaja selectiva al que tiene más. Las aristas rojas representan competencias neutras en las que cualquiera de los dos tipos puede ganar con la misma probabilidad.	35
3.10. Diagrama de la regla Órbita Malita, $k=5$	36
3.11. Decisión de alelo ganador en competencia Personalizada. En rojo número real aleatorio en $[0,1]$	40
4.1. Generación inicial aleatoria y uniforme	43
4.2. $N= 500, k=4, x=0, seed=123$	43
4.3. $N= 500, k=4, x=1, seed=123$	43
4.4. Probabilidades pequeñas de observar dos padres	44
4.5. Transitivas con $k=3, x=.3, seed=315$	44
4.6. $N=500, k=5, x=0.2, seed=789$	45
4.7. Gana el otro	45
4.8. $N= 300, k=3, seed=891$	46
4.9. Promedios de número de generaciones hasta fijación variando x	47
4.10. Generación inicial aleatoria vs. manual	49
4.11. $N= 500, x=0.5, seed=789$	49

4.12. Desequilibrio de la Órbita de 4 alelos después de la primera extinción	50
4.13. $N= 300, k=3, seed=891$	50
4.14. $N=500, k=6, x=0.5, seed=789$	51
4.15. $N= 1200, k=3$	55
4.16. $N=1200, x=0.3$	56
4.17. $k=4, x=0.3$	56

Capítulo 1

Introducción

“It is not the strongest of the species that survives, nor the most intelligent, but rather the most responsive to change”

Charles Darwin

Queremos estudiar cómo evoluciona el perfil genético de una población con diferentes tipos alélicos que preserva su número de individuos a través del transcurso de generaciones y en la que los hijos heredan el tipo de su padre. Además, añadiremos que entre los distintos alelos existan varias reglas de selección que permitan la construcción de una red ecológica. Considerar este tipo de reglas de selección en un modelo Wright Fisher es algo sumamente novedoso, [5].

Para tales efectos, programamos en Python simulaciones de los procesos de frecuencia de alelos de un modelo Wright Fisher y, jugando con distintas variables —tamaño de la población, número de alelos, probabilidad de observar más de un padre potencial, reglas de coloración, etcétera— pusimos sobre la mesa una serie de discusiones sobre su comportamiento.

El programa se diseñó para que fuera accesible y sencillo de utilizar, pensando en que biólogos y otras personas interesadas en el campo de estudio de la genética de poblaciones, pudieran simplemente ingresar sus datos recolectados y obtener inmediatamente, sin requerir de fórmulas matemáticas complicadas, gráficas que describen detalladamente la evolución simulada de aquella muestra que registraron.

Sin ahondar mucho en los detalles por el momento, es importante aclarar algunas características del modelo con el que estamos trabajando:

- Cada alelo tiene su respectivo proceso de frecuencia. Proceso que cuenta con dos estados absorbentes. Uno que se alcanza cuando llega a cero y al que se le conoce como punto de extinción; y otro, cuando llega a uno, que se le llama punto de fijación.
- La probabilidad de alcanzar uno u otro estado de absorción depende de X_0 , el estado inicial de cada proceso.

- Los procesos estocásticos de frecuencia de alelos son cadenas de Márkov que dependen del tamaño de la población, N .
- El estado inicial del proceso de frecuencia del tipo alélico a , X_0^a es la frecuencia relativa del alelo a en la generación cero,

$$X_0^a = \frac{\sum_{i=1}^N \mathbb{1}_{\xi(v_i,0)=a}}{N}$$

donde la función $\xi(0, i)$ es el tipo del individuo i en la generación 0.

Por la manera en que está construido el modelo de Wright-Fisher neutral, es decir, el que no considera selección natural ni mutación, podemos asegurar que, si hay dos tipos alélicos, a y A , eventualmente todos los individuos serán de uno o del otro tipo; es decir, en alguna generación futura, el proceso de frecuencia de alguno de los alelos alcanzará el punto de fijación, lo que equivale a decir que el otro tipo se habrá extinguido. Al tiempo de llegada en el que esto sucede se le conoce como *tiempo de fijación*, τ_{fix} .

$$\tau_{fix} = \min \left\{ g \in \mathbb{N} : \{X_g^a = 1, X_g^A = 0\} \cup \{X_g^a = 0, X_g^A = 1\} \right\}$$

Si hay más de dos alelos –digamos que $E = \{1, 2, \dots, k\}$ es el espacio de tipos alélicos– entonces la abstracción anterior se puede generalizar iterativamente de modo que concluimos que en cualquier modelo Wright Fisher sin selección nuestra población está destinada a perder casi todos sus alelos, excepto aquel que se fija.

Como mencionamos anteriormente, en la presente tesis se utilizó un modelo Wright Fisher con selección que permite el estudio de redes ecológicas. Esto se traduce en nuestro modelo de la siguiente manera: al tiempo $g + 1$ cada individuo de la generación podrá observar de manera aleatoria y con repetición uno o dos padres potenciales de entre los N individuos en la generación anterior, g . El número de observaciones que hace cada individuo también es aleatorio. Una vez elegidos los padres potenciales de cada individuo, estos se enfrentarán en un duelo del que resultará ganador el alelo más apto según la regla selectiva que se determinó previamente. Únicamente el vencedor transmitirá al hijo su tipo alélico.

Al incluir el factor de selección natural las cosas cambiarán notoriamente en nuestros procesos de frecuencia. Por un lado, en el caso de las reglas transitivas, en las cuales es posible etiquetar a nuestros alelos de menor a mayor aptitud, $a_1 < a_2 < \dots < a_k$, el tiempo de fijación a favor del más apto, a_k , se acelerará considerablemente con respecto al que se tenía anteriormente, cuando la selección natural se mantenía fuera del juego. De igual forma, se puede esperar que con mucha probabilidad las extinciones de los demás tipos alélicos sucedan en el orden ascendente de sus subíndices. La selección transitiva representa un esquema clásico y tradicional de selección, pues anteriormente se suponía que siempre habría un alelo más fuerte que el otro, que a su vez sería más fuerte que otro,

etc.

En la actualidad, casos de competencia no transitiva han llamado la atención de investigadores, tanto matemáticos como biólogos y genetistas; puesto que se ha observado presentan procesos de frecuencia que progresan de formas muy interesantes y no tan sencillas de visualizar en primera instancia. En esta tesis, a parte de la interacción transitiva, simulamos tres reglas no transitivas. Una de ellas es la Órbita; a otra la llamamos Gana el Otro; y un error al programar la regla de la Órbita se tradujo en una nueva regla que conservamos porque resultó arrojar información que nos llamó la atención; a esta tercera, la llamamos Órbita Malita, únicamente haciendo referencia a su origen.

En la regla Órbita, si existen k alelos distintos, estos se pueden etiquetar de modo que, si $n = m + 1$, $m, n \in [k]$ o si $n = 1$ e $m = k$, entonces el alelo a_n le gana al alelo a_m ; en el resto de los casos, cuando las etiquetas no son subsecuentes, los dos alelos tendrán la misma probabilidad de resultar vencedores. En Órbita Malita, sean las etiquetas m y n tales que $n > m$, entonces si $m = 1$ y $n = k$ diremos que a_m le gana a a_n ; en otro caso, a_n le ganará a a_m . Finalmente, en la regla Gana el Otro, si hay tres alelos y compiten dos de ellos, ganará el tercero que no compite.

Además, el programa le da la opción al usuario de introducir su propia regla. En caso de elegir esta opción, el programa le solicitará al usuario que, para cada competencia posible entre dos alelos distintos y observados como padres potenciales, se introduzca cuál es la probabilidad de que, dada dicha competencia, se transmita cada uno de los alelos, incluyendo aquellos alelos que que no están compitiendo en esa particular contienda. Esta regla será llamada Personalizada.

Es importante mencionar que la interacción de Gana el Otro sólo marcha si hay tres y sólo tres tipos alélicos, no obstante, tanto la regla Transitiva como la Órbita y la Órbita Malita se pueden utilizar con cualquier cantidad de tipos alélicos. Al caso particular de la Órbita con tres alelos se le conoce como Piedra, Papel o Tijera porque funciona igual que el juego de niños con tal nombre. Cada uno de los tres alelos tiene un contrincante contra el cual gana y otro contra el cual pierde.

Un ejemplo existente en la naturaleza de la competencia Piedra, Papel o Tijera es la que ocurre entre los tres morfismos en machos de las lagartijas de costado manchado (*Uta stansburiana*) que habitan en los desiertos de California, EEUU, [9]. Los machos de garganta naranja que controlan extensos terrenos y que se aparean con varias hembras, se verán en una asombrosa desventaja selectiva ante los machos de garganta azul que tienen pocas hembras y territorios pequeños, cuando los machos del tercer morfismo, los nómadas de garganta amarilla, que son similares a las hembras, entren disimuladamente a su territorio. El fornido macho de garganta naranja, a diferencia del aparentemente más débil macho de garganta azul, no podrá prestar la suficiente vigilancia en sus amplios dominios cuando los machos de garganta amarilla lleguen haciéndose pasar por hembras y se apareen con su numeroso harem.

Este efecto cíclico de competencias retardará el tiempo de fijación de nuestros procesos estocás-

ticos de frecuencia comparado a cuando se utilizaba la regla Transitiva. Una población en la que ninguno de los fenotipos o diferentes especies que compiten entre sí tiene preferencia sobre los otros, es decir, en la que todos tienen la misma oportunidad de éxito reproductivo —como es el caso de las lagartijas de costado manchado—, se le llamará *población justa*.

En general, para cada regla o interacción selectiva nos interesa conocer cómo se difunden los tipos alélicos; con qué probabilidad se extinguen, con cuál se fijan, con qué velocidad pasan estos fenómenos, en qué condiciones resiste más un tipo, en qué situaciones se acelera la extinción, etc.

Finalmente, queremos proponer, para cada interacción selectiva aquí revisada, una noción de estabilidad. La estabilidad estará determinada por dos factores: uno de ellos, será el orden de su tiempo de fijación en relación al tamaño de la población. Entre más se tarde en cumplir la llegada a este punto absorbente, diremos que es una interacción selectiva más estable: pues seguiremos con nuestra red ecológica durante más tiempo. Entre más rápido se llegue a la fijación, diremos que es más inestable: pues significa que la mayor parte de nuestros tipos alélicos (excepto aquel que se fijó) no resistieron muchas generaciones a aquella injusta competencia.

Otro factor de estabilidad que revisaremos, será el de la varianza de los procesos de frecuencia de los diferentes alelos hasta el primer episodio de extinción. Diremos que una población que tiene muchos altibajos en sus procesos de frecuencia será menos estable que aquella cuyos procesos de frecuencia se mantienen imperturbables a través de las generaciones, es decir, que su varianza es baja.

Capítulo 2

Marco teórico

En este capítulo se bosquejarán el marco referencial y las consideraciones teóricas necesarias para la interpretación de resultados y discusiones derivados de nuestra simulación de perfiles genéticos. Para la elaboración de este marco teórico se consultaron los siguientes libros, artículos y páginas web: [4], [7], [2], [1], [3], [6], [5].

En lo subsecuente, hablaremos de un modelo matemático que represente la evolución genética de una población haploide con tamaño constante, $N \in \mathbb{N}^*$, de multitemo alélico y con generaciones discretas que no se traslapan entre sí. Este modelo, como elemento adicional, permitirá la selección con competencia por redes ecológicas.

En la evolución la aleatoriedad se encuentra de muchas maneras: en errores en la copia de material genético, en migraciones, cambios ambientales, etcétera. Además de estos elementos ya mencionados, existe una parte meramente fortuita del proceso que está inducida por el hecho de tratarse de la reproducción de poblaciones finitas. A este factor estocástico se le conoce como *deriva genética* y es el motivo por el cual, en ocasiones un alelo débil o no tan fuerte puede llegar a fijarse sobre otros tipos alélicos más fuertes, simplemente por una cuestión de azar.

Antes de continuar, para comprender de lo que estamos hablando, es necesario hacer una revisión por algunos términos de biología y ecología, así como algunas definiciones y teoremas matemáticos. Comencemos muy brevemente con la parte de biología y ecología:

- *Población* es un grupo de individuos de la misma especie que se pueden aparear entre sí y, por lo tanto, reproducir.
- *Locus* (loci en plural) es la ubicación específica en un cromosoma en la que se define tal o cual característica genética (marcador genético) de un individuo. Por ejemplo: color de ojos.
- *Alelos o tipos alélicos* son variaciones de un gen que producen diversificaciones en las características heredadas. Por ejemplo: ojos azules es un alelo, ojos verdes es otro, ojos amarillos otro

y ojos cafés otro.

- *Polimorfismo* es la presencia de dos o más alelos en un locus. El ejemplo del color de ojos en los humanos es un polimorfismo porque hay varios tipos.
- *Variación genética*: qué tanta diversidad muestran los individuos de la población en un gen en particular.
- *Haploides vs. diploides*: los primeros son organismos que tienen una sola copia de cada cromosoma, por ejemplo: las bacterias; los segundos son organismos, que como los humanos, tenemos dos copias, una de nuestros padres y otra de nuestras madres. En este caso, cada padre contribuye con una copia de su propio par. Existen, también, organismos con más de dos copias: cuatro (tetraploides), seis (hexaploides) o más (poliploides).
- *Mutación*: es la alteración en la información genética. Las mutaciones se pueden deber a un error en la copia del DNA durante la división celular o a la exposición a ciertas sustancias químicas o a la radiación. Una mutación, en ocasiones, puede ser beneficiosa para la adaptación del individuo, por lo que podría volverse más común con el paso del tiempo; aunque también pudiera ser dañina o neutra. La mayoría de las mutaciones son neutrales.
- *Aptitud*: Es el grado de ventaja o desventaja selectiva de un alelo. Matemáticamente está relacionada con la probabilidad de un tipo alélico para fijarse en el futuro en una población, condicionada a su frecuencia inicial.
- *Interacciones selectivas o reglas de coloración*: es cómo interactúan las medidas de aptitud de los distintos tipos alélicos. Quién le gana a quién y con qué probabilidad. Estas interacciones vinculan a los tipos alélicos de dos en dos. Al traducirlo a términos matemáticos, decimos que las interacciones selectivas serán una familia de funciones de probabilidad sobre el espacio de los tipos alélicos que determine cómo se relacionan los unos con los otros.
- *Red ecológica*: es la representación de un conjunto de interacciones biológicas de cualquier tipo en un ecosistema.
- *Estabilidad ecológica*: decimos que un ecosistema posee estabilidad ecológica si es capaz de regresar a un estado de equilibrio después de una perturbación o si a través del tiempo no experimenta grandes cambios en sus características.

Antes de proseguir, aclaremos dos detalles, uno referente al término *individuo* que estaremos usando continuamente en el resto de este texto y otro acerca de nuestra motivación para hacer uso de este modelo a la hora de describir el comportamiento de una población diploide, como puede ser la población humana.

Generalmente, cuando en un modelo de población hablamos de un individuo, no nos referimos al espécimen en sí, sino únicamente a una parte específica de su material genético. De ahora en adelante, utilizaremos el término individuo para referirnos a cierto locus de dicho individuo.

También, aunque en una población diploide los individuos tienen dos padres, cada gen puede ser rastreado a un único gen en la generación anterior que es su padre. Por lo anterior, se acostumbra en el modelo de Wright-Fisher tratar a las poblaciones diploides compuestas por N individuos, como poblaciones haploides de tamaño $2N$. Sin embargo, nosotros nos mantendremos con una población simplemente haploide de tamaño N . Es válido y tiene sentido hablar de poblaciones haploides en genética humana porque el humano siempre hereda el ADN mitocondrial de la madre el cromosoma Y y los hombres (sexo masculino) el cromosoma Y del padre (con excepción de una pequeña región en la punta del cromosoma que se recombina con el cromosoma X).

Las bases para el estudio de la genética evolutiva están fundamentadas en los trabajos precursores de Gregor Mendel y Charles Darwin. La teoría de Darwin de la evolución, descrita en su afamado y controversial libro *El Origen de las Especies* (1859), enunciaba que los rasgos hereditarios que propician la reproducción, se volverán más comunes en una población.

En 1866, Mendel —considerado actualmente como el padre de la genética moderna— anuncia, por su lado, la teoría sobre la heredabilidad, la cual obtiene muy poca recepción entre la comunidad científica de la época. Mendel planteaba que los rasgos están determinados por los genes; que cada gen puede ocurrir en una finita cantidad de tipos llamados alelos y que los padres le heredan a sus hijos su alelo.

Aunque actualmente ambas teorías —la de Mendel y la de Darwin— no sólo nos parecen compatibles, sino complementarias, en aquel entonces, los partidarios de una y de otra, no lograban llegar a un acuerdo. Fueron las matemáticas las que trajeron paz entre ambos bandos. Pero para esto tuvo que pasar un largo tiempo; fue hasta el año 1931, en que fue descrito, simultáneamente por el británico Ronald A. Fisher y por el estadounidense Sewell Wright el modelo de Wright-Fisher; un modelo matemático en el cual la teoría de Darwin y la de Mendel pudieron, al fin, convivir sin contradecirse.

Ahora bien, revisemos las matemáticas de este modelo clásico, así como las de nuevos descubrimientos y complementos derivados de éste mismo.

Definición 2.1. *(El modelo de Wright-Fisher neutral) El modelo de Wright-Fisher neutral es un modelo de genética de poblaciones de tiempo discreto en el cual las generaciones no se traslapan y en la que a cada generación se le identificará con un número entero según su sucesión cronológica, de manera que las generaciones corren del pasado hasta el futuro por todos los enteros, \mathbb{Z} , desde $-\infty$ hasta ∞ , pasando por el cero que representa, por convención, el presente o el momento de la evaluación.*

Sea $N \in \mathbb{N}^*$, un número finito, diremos que en cada generación habrá siempre N individuos y que

N será el tamaño de la población. Es decir, el tamaño de la población se mantendrá constante.

En el modelo de Wright-Fisher neutral la generación g se obtiene a partir de la generación $g - 1$, mediante la elección aleatoria y con reemplazo de N individuos de la generación $g - 1$. Es decir, el individuo (g, i) , con índice $i \in [N]$ y en la generación $g \in \mathbb{Z}$, elegirá de manera uniformemente aleatoria y con reemplazo a su padre entre los N individuos de la generación $g - 1$.

Digamos que una población *neutral* es aquella en la que todos los tipos tienen la misma oportunidad de éxito reproductivo. Decimos que es *panmíctica*, si todos los individuos se encuentran bajo las mismas condiciones y tienen la misma probabilidad de aparearse con cada uno de los demás integrantes de la población. En nuestro modelo, la población es tanto neutral como panmíctica.

Para esta tesis trabajaremos con un modelo multitipo, así que denotemos con $E = [k] = \{1, \dots, k\}$, $k < \infty$, a nuestro espacio de tipos alélicos; observemos que $|E| = k$. El índice de cada tipo alélico será, en principio, independiente de su aptitud y simplemente será una manera de etiquetarlo. $\xi(g, i) \in E$ será la función que devuelva el tipo alélico del individuo (g, i) .

Definición 2.2. (*Gráfica Aleatoria Wright-Fisher*) Sea $V = \mathbb{Z} \times [N]$ la matriz en donde el vértice (g, i) corresponde al i -ésimo individuo de la generación g . Sea A el conjunto de aristas que conectan a cada individuo con su padre, tal que:

$$A := \{(v, v') : v = (g, i) \in V, v' = (g - 1, U_v) \in V\}$$

donde $\{U_v\}_{v \in \mathbb{N}}$ es una familia de variables aleatorias independientes y uniformemente distribuidas en $[N]$.

Entonces la Gráfica aleatoria Wright-Fisher es $G := (V, A)$.

En una gráfica Wright-Fisher podemos estudiar a nuestra población tanto en dirección al futuro, analizando cómo cambia su perfil genético, como en dirección al pasado, estudiando las relaciones de ancestría. Esta dicotomía de la gráfica Wright-Fisher está ligada a una noción de dualidad como veremos posteriormente.

Aunque estaremos más concentrados en el modelo que va hacia adelante en el tiempo y que describe como evoluciona la población, es importante revisar su contraparte hacia el pasado porque, como veremos más adelante, varios resultados se derivan de esta dualidad.

Para revisar cómo evoluciona la población hacia el futuro nos enfocamos en los procesos de frecuencia de alelos.

2.1. Proceso de frecuencia de alelos

El desarrollo de los procesos de frecuencia alélica será nuestro mayor interés en el resto de esta tesis. Si tenemos una población de tamaño finito y constante, $N \in \mathbb{N}^*$, y k tipos alélicos, podemos

decir que para cada tipo alélico $j \in [k]$, existe un proceso estocástico $\{X^j(g) : g \geq 0\}$, con valor inicial $X^j(0) = x_j$ que describe el curso de frecuencias de dicho tipo a través de las generaciones. Estos procesos estocásticos son dependientes entre sí, pues en cada generación deben, entre todos, sumar uno, $\sum_{j=1}^k X^j(g) = 1$, para toda $g \in \mathbb{Z}$; y tienen la propiedad de Márkov con respecto al tiempo.

Desmenuemos lo enunciado arriba haciendo un breve repaso del curso de procesos estocásticos.

Definición 2.3. (*Proceso estocástico*) *Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ que toman valores en Ω , indexada a un conjunto T , que interpretaremos como el tiempo. Nosotros consideraremos que T será $\mathbb{N} = \{0, 1, 2, \dots\}$ o $[0, \infty)$, según si se trata de tiempo discreto o continuo.*

Sea F_t una sigma álgebra que tiene toda la información obtenida en el proceso hasta el tiempo t –por así decirlo, la historia del proceso hasta ese momento– entonces decimos que un proceso estocástico tiene la Propiedad de Márkov si su futuro condicionado a conocer todo F_t es el mismo que sólo condicionándolo a X_t , su estado al tiempo t .

Definición 2.4. (*Cadena de Márkov*) *Una cadena de Márkov es un proceso estocástico $\{X_n : n = 0, 1, \dots\}$, con espacio de estados discreto que satisface la propiedad de Márkov; esto es, que para cualquier entero $n \geq 0$, y para cualesquiera estados $x_0, \dots, x_{n+1} \in \Omega$, Ω discreto, se cumple que*

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n].$$

Ahora sí, podemos definir al proceso de frecuencia de la siguiente manera:

Definición 2.5. (*Proceso de frecuencia*) *Para cada tipo alélico $j \in [k]$, el proceso de frecuencia $\{X^j(g)\}_{g \geq 0}$ con valor inicial $X^j(0) = x_j$ se define como*

$$X^j(g) = \frac{\sum_{i=1}^N \mathbb{1}_{\{\xi(g,i)=j\}}}{N}$$

donde $\xi(g, i)$ es el tipo alélico del individuo con índice i de la generación g .

$X^j(g)$ nos dirá cuál es la proporción relativa del alelo j al tiempo o generación g , en particular, $X^j(g) \in \frac{[N]}{N}$ para toda g , lo que se traduce en que $X^j(g)$ tomará siempre valores en un conjunto finito. Las generaciones son números enteros no negativos.

Por la construcción del modelo, $\{X_{g \geq 0}^j : j \in [k]\}$ tiene la propiedad de Márkov con respecto a las generaciones, lo que quiere decir que la formación de una nueva generación dependerá únicamente de la frecuencia de alelos en el tiempo presente; el resto de la historia pasada será irrelevante para la predicción del futuro de la población.

Si $X^j(0) = x_j$, $j \in [k]$, entonces en la generación inicial hay $x_j N$ individuos de cada uno de los k

tipos alélicos, donde $x_j \in [0, 1]$, $\forall j \in [k]$, tal que

$$\sum_{j=1}^k x_j N = N.$$

Por la construcción del modelo, se puede decir lo mismo para las generaciones anteriores y subsecuentes a la inicial. De modo que,

$$\sum_{j=1}^k X_j(g)N = N, \quad \forall g \in \mathbb{Z}.$$

2.1.1. Probabilidad de Transición

Sea $\{X_g\}$ el proceso de frecuencia del alelo a , entonces $\{Y_g\} = N\{X_g\}$ será el proceso de número de individuos con el tipo a en cada generación g . Las *probabilidades de transición* de este proceso $\{Y_g\}$ están dadas por:

$$p_{r,s} = \mathbb{P}[X_{g+1} = s/N | X_g = r/N] = \mathbb{P}[Y_{g+1} = s | Y_g = r].$$

Proposición 2.1. *En nuestro modelo neutral Wright Fisher, para cualquier generación $g \in \mathbb{Z}$, las transiciones del proceso Y_g (como se definió anteriormente) a Y_{g+1} dado que $Y_g = r$, están caracterizadas por:*

$$Y_{g+1} | Y_g = r \stackrel{d}{=} \text{Bin}(N, r/N),$$

donde $\text{Bin}(N, r/N)$ es una variable aleatoria Binomial de parámetros N e r/N . Es decir,

$$p_{r,s} = \binom{N}{s} \left(\frac{r}{N}\right)^s \left(\frac{N-r}{N}\right)^{N-s}.$$

Demostración. Para ver porqué es esto, primero recordemos que un individuo en la generación $g+1$ tendrá el tipo a si y sólo si elige como padre a un individuo en la generación g del tipo a . La combinatoria representa cuáles serán los s individuos que en la generación $g+1$ elegirán al tipo a . Después tenemos que el siguiente factor es la probabilidad de que esos s elijan a padres con tipo a y que los $N-s$ restante elijan padres con diferente tipo a a . Nótese que estas probabilidades están dadas por $\frac{r}{N}$ y $\frac{N-r}{N}$ respectivamente, puesto que sabemos que en la generación g , de los N individuos, r son del tipo a y $N-r$ son de un tipo distinto; además, cada uno elige a su padre de manera aleatoria y con reemplazo entre los individuos de la generación anterior. \square

Como consecuencia de la proposición anterior tenemos que:

$$\mathbb{E}[Y_{g+1} | Y_g = r] = N * p_r = N * \frac{r}{N} = r \tag{2.1}$$

$$\text{Var}[Y_{g+1}|Y_g = r] = N * p_r * q_r = \frac{1}{N}r(N - r), \quad (2.2)$$

donde p_r es la probabilidad de que un individuo en la generación $g + 1$ tenga el alelo a dado que en la generación g había r individuos con el alelo a y q_r la probabilidad de que sea de otro alelo distinto a a .

Observemos que nuestro proceso $\{Y_g\}$ es una Martingala, i.e. para todo $m, g \in \mathbb{N}$

$$\mathbb{E}[Y_{g+m}|Y_g] \stackrel{\text{a.s.}}{=} Y_g.$$

Para entender esto veamos que un individuo en la generación $g + m$ tiene el tipo de su padre, que a su vez tiene el tipo de su padre, que a su vez... así hasta llegar a la generación g , de modo que este último es el antecesor en la generación g del individuo que estábamos viendo en primer lugar en la generación $g + m$. Como, por construcción, este antecesor se elige uniformemente al azar, la probabilidad de que este sea del tipo a es Y_g .

$$\mathbb{E}[Y_{g+m}|Y_g] = \frac{1}{N}\mathbb{E}[\sum_{i=1}^N \mathbb{1}_{\{\xi(g+m,i)=a\}}|Y_g] = \mathbb{E}[\mathbb{1}_{\{\xi(g+m,i)=a\}}|Y_g] = Y_g$$

porque $\mathbb{1}_{\{\xi(g+m,i)=a\}}|Y_g$ es una Bernoulli con parámetro Y_g .

El *Principio de Hardy-Weinberg*, descrito en 1908 por un matemático y un médico de apellidos Hardy y Weinberg, respectivamente, nos muestra las frecuencias de los alelos esperadas en una población hipotética que no evoluciona. El Equilibrio de Hardy Weinberg, un principio clásico en genética de poblaciones, dicta que la frecuencia de los alelos permanece constante de una generación a otra mientras no actúe la selección natural.

Proposición 2.2. (*Equilibrio de Hardy-Weinberg*) Para toda $g \in \mathbb{N}$ y $x \in [N]/N$ fija,

$$\lim_{N \rightarrow \infty} [X_{g+1}|X_g = x] \stackrel{\text{a.s.}}{=} x.$$

Demostración. Como $\forall g \in \mathbb{N}$ tenemos que $\mathbb{E}[|X_g|] < \infty$ y $\mathbb{E}[X_{g+1}|X_g = x] = x$, entonces la demostración es una aplicación de la Ley Fuerte de los Grandes Números. \square

El principio de Hardy-Weinberg muestra que si existen variaciones en la esperanza quiere decir que nuestro modelo no es neutral, ya sea porque alguno de los tipos alélicos tiene ventaja selectiva o está habiendo mutaciones.

2.1.2. Modelo de Moran

Existe, además del modelo Wright-Fisher, otro modelo matemático ampliamente utilizado para estudiar el proceso de evolución alélica; se llama el modelo de Moran. A pesar de que no lo abor-

daremos con profundidad en esta tesis, quisiéramos hacer mención de él porque está íntimamente ligado con el estudio del modelo Wright-Fisher.

Aunque es menos popular entre los biólogos, cuando hablamos de poblaciones lo suficientemente grandes, al revisar ambos modelos en dirección al pasado, estos se comportan de maneras muy similares.

Definición 2.6. *(Modelo de Moran)* Una población de N individuos evoluciona conforme al modelo de Moran si después de un tiempo distribuido exponencial con tasa 1, un par de individuos es muestreado y uno de ellos muere y el otro se duplica. Se puede observar que, a todo tiempo, se conserva el tamaño de la población. Después del evento de reproducción, el proceso vuelve a funcionar de igual forma que en el principio. Esto sucede de manera independiente a lo que pasó anteriormente, gracias a la propiedad de falta de memoria de la exponencial. Es decir, es una cadena de Márkov en tiempo continuo. El modelo está definido para toda $t \in \mathbb{R}^+$.

En el modelo de Moran, si se tienen r individuos del tipo a (a por decir cualquiera) en la generación g , la probabilidad de que en la siguiente generación, $g + 1$, haya $r + 1$ o la de que haya $r - 1$ individuos del tipo a , es la misma,

$$\mathbb{P}[Y^a(g + 1) = r + 1 | Y^a(g) = r] = \mathbb{P}[Y^a(g + 1) = r - 1 | Y^a(g) = r] = (N - r) \frac{r}{N} = r \frac{N - r}{N}.$$

Existen dos diferencias esenciales entre el modelo de Wright-Fisher y el modelo de Moran. La primera es que en el modelo de Wright-Fisher las generaciones no se traslapan —tal y como sucede con las plantas anuales o en algunas manadas de venados—; mientras que en el modelo de Moran, sí se traslapan y, por lo tanto, puede ser un modelo más preciso para representar poblaciones de especies como los humanos o la *Drosophila*, en las que las generaciones no están sincronizadas y en las que la población cambia por un individuo a la vez.

La segunda diferencia es que en el modelo de Wright-Fisher los individuos pueden llegar a tener hasta N hijos (el tamaño total de la población), cuando, en cambio, en el modelo de Moran máximo tienen un hijo a la vez. Además, aunque ambos son cadenas de Márkov, uno lo es en tiempo discreto y el otro en continuo, respectivamente.

El modelo de Moran se puede entender también como N procesos de Poisson homogéneos e independientes con tasa 1, $\{Z_t(i), t \in \mathbb{R}^+\}_{i=1}^N$. Cada que $Z_t(i)$ tenga un evento, diremos que el individuo i de la población murió y fue reemplazado por el descendiente de otro individuo de la misma población elegido de manera aleatoria.

Para poblaciones grandes, si medimos el tiempo en unidades de orden de N^2 —orden de la media en que tardan en reemplazarse todos los individuos de una generación, cada pareja elegida con probabilidad $\binom{N}{2}$, N reproducciones requeridas— la genealogía de una muestra tomada de un modelo de Moran converge al coalescente de Kingman, al igual que el modelo de Wright-Fisher,

únicamente que la coalescencia sucede más lento.

2.1.3. Fijación

Un estado absorbente es aquel en el que una vez que la cadena de Márkov entra en él, nunca más podrá dejarlo.

En el modelo de Wright-Fisher, 0 y 1 son estados absorbentes de los procesos de frecuencia de alelos. Por ejemplo, si el tipo alélico j se extingue en la generación m , i.e., si $X^j(m) = 0$; en las generaciones subsecuentes no volverá a aparecer el tipo alélico j , i.e., $X^j(n) = 0$ para toda $n \geq m$. Por otro lado, si el tipo alélico j se fija, digamos que en la generación μ , i.e., $X^j(\mu) = 1$; en las generaciones posteriores todos los individuos serán del tipo j , i.e. $X^j(\nu) = 1$ para toda $\nu \geq \mu$.

Si tenemos sólo dos alelos, a y A , y pensamos a a como una mutación del alelo A , una pregunta evidente que nos surge es: ¿cuál es la probabilidad de que esta mutación sobreviva? Es decir, ¿qué tan probable es que esta mutación invada a toda la población y los antiguos residentes, los A 's, desaparezcan? Y si llega a pasar, ¿cuánto tiempo tomará este proceso?

Como por el momento, estamos usando un modelo neutro, en el que todos los alelos (mutaciones y no mutaciones) tienen la misma probabilidad de sobrevivir, es decir en el que se eligen a los padres de manera uniformemente aleatoria y con reemplazo en la generación anterior, siempre es posible que en cierta generación extraigamos todas A 's o todas a 's. Por lo tanto, sabemos que eventualmente en nuestro proceso la proporción de A 's será 0, indicando la pérdida del alelo A , o 1, indicando la pérdida de a .

Definición 2.7. (*Fijación*) *Fijación es el evento en el cual el proceso estocástico de frecuencia de un tipo alélico alcanza el uno y así se queda por siempre.*

Intuitivamente un tiempo de llegada es el tiempo que tarda la trayectoria de un proceso hasta que se cumpla cierta condición. El instante en que se cumple dicha condición se registra el tiempo de llegada. Ahora demos la definición formal.

Definición 2.8. (*Tiempo de llegada*). *El tiempo de llegada, H^B , es la primera vez en que la cadena arriba a B , $B \in \Omega$, donde B es un conjunto medible que será la condición de llegada.*

$$H^B = \inf\{n \geq 0 : X_n \in B\}$$

($\inf\{\emptyset\} = \infty$).

Si decimos que $t_x^B = \mathbb{E}[H^B | X_0 = x]$ es el tiempo esperado hasta que se alcance la condición de llegada B siendo que el proceso comenzó en x al tiempo 0, tenemos que

Teorema 2.1. *El vector de tiempos de llegada esperados $t^B = \{t_x^B : x \in \Omega\}$ es la mínima solución no negativa al sistema de ecuaciones lineales:*

$$\begin{aligned} t_x^B &= 0, & \text{si } x \in B \\ t_x^B &= 1 + \sum_{\{y \in \Omega\}} p_{x,y} t_y^B, & \text{si } x \notin B. \end{aligned}$$

Lo anterior se puede entender como si ya llegó ahí se detiene, pero si no ha llegado a la condición de llegada, cuenta que dio un paso sin haber llegado y vuelve a revisar su situación desde donde ahora está parado.

Para el proceso de frecuencia de un determinado alelo, los estados absorbentes, 0 y 1, pueden ser vistos como condiciones de llegada.

Teorema 2.2. *Si en una población de tamaño N la frecuencia inicial del alelo a es $X(0) = x$, entonces el tiempo esperado hasta alcanzar un punto de absorción, $B = \{0, 1\}$ es $t_x^B + \mathcal{O}(\frac{1}{N})$ donde:*

$$t_x^B = t^B(x) = -2N(x \log x + (1-x) \log(1-x)).$$

Heurística. Utilizamos el argumento no demostrado de que t^B es continua y dos veces diferenciable. Decimos que si x es la proporción inicial del alelo a , entonces x se puede escribir, sin pérdida de generalidad, de la forma $x = r/N$, donde r es el número de individuos que tienen el alelo a en el tiempo cero. Sea $B = \{0, 1\}$ el conjunto de condiciones de llegada, entonces suponemos que la esperanza del tiempo hasta la absorción comenzando en x puede ser aproximada por una función dos veces diferenciable por x , que denominaremos $t^B(x)$.

Si en el tiempo cero tenemos r individuos del tipo a , por el Teorema 2.1 sabemos que el número de individuos del tipo a en la siguiente generación tiene una distribución $Bin(N, r/N)$, de modo que su esperanza es r y su varianza $r \frac{N-r}{N}$. Entonces el cambio en la proporción de alelos, dado que $x = \frac{r}{N}$, es $\delta(x) = [X_1 - X_0 | X_0 = r/N] = [\frac{Bin(N, r/N)}{N} - \frac{r}{N}]$ y tiene media 0 y varianza $\frac{1}{N^2} r(N-r) = \frac{1}{N}(x(1-x))$.

En el siguiente desarrollo utilizaremos indistintamente $t^B(x)$ y $t(x)$, esto para que la notación se vea más clara, pero recordemos que B es el conjunto de nuestras condiciones de llegada.

$$\begin{aligned} t^B(x) &= 1 + \sum_{\{r \in \Omega\}} p_{r,s} t_s^B \\ &= \sum_{\delta_x} \mathbb{P}[x \mapsto x + \delta_x] (t^B(x + \delta_x) + 1) && \text{por el Teorema 2.1} \\ &\approx \sum_{\delta_x} \mathbb{P}[x \mapsto x + \delta_x] (t(x) + \delta_x t'(x) + \frac{1}{2} \delta_x^2 t''(x) + 1) && \text{por el Teorema de Taylor} \\ &= t(x) + t'(x) \mathbb{E}[\delta_x] + \frac{1}{2} t''(x) \mathbb{E}[(\delta_x)^2] + 1 && \text{por definición de esperanza} \\ &= t(x) + \frac{1}{2N} (x(1-x)) t''(x) + 1 && \text{sustituyendo.} \end{aligned}$$

Despejando, lo anterior es igual a:

$$\begin{aligned}x(1-x)t''(x) &= -2N \\ t(0) = t(1) &= 0.\end{aligned}$$

Este sistema de ecuaciones diferenciales se puede resolver para que nos de lo que estamos buscando:

$$t^B(x) = -2N(x \log x + (1-x) \log(1-x)).$$

□

Definición 2.9. (*Tiempo de fijación*): El tiempo de fijación es el tiempo de llegada hasta que alguno de los alelos se fije, es decir, cuántas generaciones se tarda el proceso en que toda la población consista en un sólo tipo alélico.

$$\tau_{fix} = \min\{g : X^j(g) = 1 \text{ para alguna } j \in [k]\}.$$

El tiempo y la probabilidad de fijación o de extinción de un alelo, están determinados por la frecuencia inicial de dicho alelo. Esto se puede ver de manera intuitiva: si en una población sin selección natural iniciamos con una proporción pequeña de individuos del tipo a , lo más probable es que eventualmente el tipo a desaparezca; de igual modo si hay una gran proporción de individuos del tipo a en comparación a los demás tipos, uno podría esperar que el tipo a se fije en toda la población. Sin embargo, podría suceder que, por una cuestión azarosa, habiendo muy pocos a 's, estos se fijen, o que habiendo muchos, estos se extingan. A esta fuente de aleatoriedad, le llamamos, como ya habíamos mencionado, deriva genética.

Teorema 2.3. En un modelo de Wright-Fisher neutral con dos tipos alélicos, a y A , la probabilidad de que se fije el alelo a , con proceso de frecuencia absoluta $\{Y_g\}$ y que se extinga A , dado que la frecuencia inicial de a es $Y_0 = \lambda$, es

$$\mathbb{P}[Y_\tau = N | Y_0 = \lambda] = \frac{\lambda}{N}.$$

Demostración. Como el número de individuos es finito, siempre, hasta antes de llegar a los puntos absorbentes, es posible que un individuo tenga tanto un padre con alelo a como uno con alelo A ; por lo tanto, la fijación de alguno de ellos ocurrirá eventualmente.

Sea Y_g la frecuencia del tipo a en el tiempo g , $\mathbb{E}[Y_{g+1} | Y_g] = Y_g$ casi seguramente, ya que $\mathbb{E}[Y_{g+1} | Y_g]$ es una variable aleatoria que hereda su aleatoriedad de Y_g y por la ecuación 2.1 sabemos que se distribuyen igual. Sacando la esperanza en ambos lados, obtenemos que $\mathbb{E}[Y_{g+1}] = \mathbb{E}[Y_g]$, es decir, el valor esperado de Y_g se mantiene constante a través del tiempo.

Para probar qué pasa en el tiempo de llegada, sea $\tau = \min\{g : Y_g = 0 \text{ o } Y_g = N\}$ el tiempo de llegada, ya sea por extinción o fijación del alelo a , entonces, como sabemos que después de τ nuestro proceso se mantendrá igual en las generaciones subsecuentes, es decir, para toda $g > \tau$, $Y_g = Y_\tau$

$$\lambda = \mathbb{E}[Y_g | Y_0 = \lambda] = \mathbb{E}[Y_\tau | Y_0 = \lambda; g \geq \tau] \mathbb{P}[g \geq \tau] + \mathbb{E}[Y_g | Y_0 = \lambda; g < \tau] \mathbb{P}[g < \tau].$$

Si hacemos a $g \rightarrow \infty$, el primer término tiende a $\mathbb{E}[Y_\tau | Y_0 = \lambda]$ y el segundo a cero. Entonces,

$$\lambda = \mathbb{E}[Y_\tau | Y_0 = \lambda] = N \mathbb{P}[Y_\tau = N | Y_0 = \lambda] = N \mathbb{P}_\lambda(Y_\tau = N).$$

□

2.1.4. Difusión de Wright-Fisher

Aunque este tema va más allá del alcance de la presente tesis, le echamos un vistazo veloz y superficial en las siguientes referencias bibliográficas: [2] y [3].

Si dejamos a N , el tamaño de nuestra población, tender a infinito y, a la vez, aceleramos el tiempo a una tasa de orden N , entonces los procesos de frecuencia alélica convergen a un proceso de difusión.

Teorema 2.4 (Difusión de Wright-Fisher). *Si reescalamos apropiadamente el tiempo en unidades de tamaño N , de modo que el proceso evolucione en intervalos de tamaño $1/N$, en el límite cuando $N \rightarrow \infty$, el proceso de frecuencia de alelos converge a la difusión de Wright-Fisher, que es la solución de*

$$dX_t = \sqrt{X_t(1 - X_t)} dW_t,$$

donde $\{W_t\}_{t \geq 0}$ es el movimiento Browniano estándar.

Heurística. Supongamos que tenemos una población con dos tipos alélicos, a y A , y que $\{X_t\}$ es el proceso de frecuencia en concreto del alelo a y que al tiempo $t = 0$ tiene una proporción $X_0 = x$.

Para describir el comportamiento de $\{X_t\}$ checaremos cómo cambia, a través del tiempo, la esperanza de una función bien definida y evaluada en nuestro proceso, i.e., $\mathbb{E}[u(X_t)]$; $u : [0, 1] \rightarrow \mathbb{R}$.

Si en el tiempo presente, una proporción x de la población es del tipo a , entonces, como el proceso tiene transiciones binomiales, el número esperado de individuos del tipo a en la próxima generación es Nx y la varianza de ese número es $Nx(1 - x)$. Si damos un pasito de tamaño $1/N$, la esperanza de la frecuencia de alelos se mantiene igual a x y su varianza es $x(1 - x)/N$. Además $\mathbb{E}[(X_{1/N} - x)^k | X_0 = x] = \mathcal{O}(1/N^2)$ para toda $k \geq 3$, por lo que podemos decir que el proceso evoluciona de forma homogénea en el tiempo, así que podemos sólo fijarnos en qué ocurre cerca del

tiempo cero. Entonces,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[u(X_t)|X_0 = x]|_{t \geq 0} &\approx \frac{\{\mathbb{E}[u(X_{1/N}|X_0 = x) - u(x)]\}}{1/N} \\ &= N\{u'(x)\mathbb{E}[X_{1/N} - x|X_0 = x] + \frac{1}{2}u''(x)\mathbb{E}[(X_{1/N} - x)^2|X_0 = x] + \mathcal{O}(1/N^2)\} \\ &= \frac{1}{2}x(1-x)u''(x) + \mathcal{O}(1/N). \end{aligned}$$

El primer término del segundo renglón se anuló porque es igual a cero y por lo visto arriba $\mathbb{E}[(X_{1/N} - x)^2|X_0 = x] = x(1-x)/N$.

Entonces esperamos que la distribución de la frecuencia de alelos en el límite esté gobernada por la solución de la ecuación diferencial estocástica Wright-Fisher. \square

Un *proceso de difusión* de una dimensión, $\{X_t\}_{t \geq 0}$, como lo es nuestra difusión Wright-Fisher, es un proceso de Márkov que traza un camino continuo conforme el tiempo avanza y, más aún, cualquier función continua de $\{X_t\}_{t \geq 0}$ también. Su rango no necesariamente debe ser sobre todo \mathbb{R} , en realidad, en nuestros ejemplos de genética, estaremos interesados en los procesos de difusión cuyo rango está acotado en $[0, 1]$. Por el momento, generalicemos diciendo que nuestro proceso tiene un espacio de estados en (a, b) .

El *movimiento Browniano* es el más fundamental de los procesos de difusión. Para entender qué es el movimiento Browniano imaginemos variables aleatorias independientes idénticamente distribuidas $Z_{i \in \mathbb{N}}$, tal que $\mathbb{P}[Z_i = 1] = 1/2 = \mathbb{P}[Z_i = -1]$ sea $S_n = \sum_{i=1}^n Z_i$, si $S_0 = 0$, entonces $\{S_n\}_{n \geq 0}$ es el proceso conocido como la caminata aleatoria sobre Z .

Ahora si acertamos los pasos a pasitos de tamaño $\pm 1/\sqrt{n}$; y los damos más rapidito, en intervalos de $1/n$ y, bajo estos nuevos parámetros, reescalamos nuestra caminata aleatoria, tenemos un nuevo proceso, $B_t(n) = 1/\sqrt{n} S_{[nt]}$, donde $[nt]$ denota la parte entera de nt , que sigue siendo una caminata aleatoria.

Si hacemos a $n \rightarrow \infty$, por el Teorema del Límite Central

$$B_t(n) \rightarrow B_t \sim N(0, t)$$

Además como los pasos dados en intervalos de tiempo disjuntos son independientes, $B_t(n) - B_s(n)$ es independiente de $B_s(n) - B_0(n)$ para $0 < s < t$, propiedad que hereda el proceso límite. Y nuevamente por el Teorema del Límite Central, $B_t - B_s \sim N(0, t - s)$. Aunque las trayectorias del movimiento Browniano son continuas, no son diferenciables en ningún punto. A partir del Movimiento Browniano se forman las demás difusiones unidimensionales.

Las difusiones unidimensionales tienen la cualidad de que muchas propiedades del proceso pueden ser calculadas explícitamente. Esto es porque (salvo en a y en b , nuestros “puntos finales”) toda difusión unidimensional puede ser transformada a un Movimiento Browniano, a través, primero de

un cambio de la variable espacial (función escalar) y después con un cambio temporal (medida de la velocidad).

La ventaja de usar la difusión Wright Fisher es que se pueden calcular varias de sus propiedades de importancia práctica con fórmulas exactas, mientras que en el caso discreto, éstas son o difíciles o imposibles de calcular.

Para efectos de esta tesis dejaremos hasta aquí lo relativo a los procesos de difusión, pero invitamos al lector curioso e interesado en el tema a revisar la siguiente bibliografía en aras de encontrar más información al respecto: [2], [1], [3]. Nosotros, más adelante, buscaremos encontrar las propiedades de nuestro modelo con la simulación.

2.2. Proceso ancestral

Ahora queremos revisar el modelo en sentido opuesto, es decir, hacia el pasado; para esto será elemental determinar los árboles genealógicos que relacionan a los individuos de una población.

¿A qué nos referimos con árboles genealógicos? Intuitivamente, por lo visto en la naturaleza, uno esperaría ver pequeñas variaciones entre miembros de una misma familia, sobre todo si su ancestro común no se encuentra muy lejos en dirección al pasado. Puede ser que los genes del ancestro hayan evolucionado, quizá a causa de una mutación, pero si el ancestro es reciente, los cambios no deberían ser muchos, a diferencia de individuos que están relacionados con mucha distancia entre sí —por ejemplo, digamos que dos individuos comparten un abuelo de sus respectivos tatarabuelos, es decir, muchísimas generaciones atrás, ambos individuos tienen al mismo individuo como ancestro— donde uno esperaría mucha mayor variación entre uno y otro. Así pues, decimos que queremos saber, qué grado de cercanía genética hay entre nuestros individuos muestreados. *Genealogía* es la historia ancestral de un sujeto.

Para poder analizar el árbol genealógico, normalmente observamos la partición ancestral del proceso. Una partición π de \mathbb{N} es una relación de equivalencia en \mathbb{N} .

A cada generación g , el estado del coalescente puede ser representado como una partición. Sea para una generación g , la partición de $[N]$, $\pi_1 = A_1, \dots, A_m$, tal que la $\cup_{i=1}^m A_i = [N]$, donde si $i \neq j$ entonces A_i y A_j son conjuntos disjuntos, entonces los bloques $\{A_i\}_{i \in [m]}$; de la partición son las clases de equivalencia de esta relación.

Usaremos $a \sim b$ para denotar que a y b están en el mismo bloque de π . En palabras, cada A_i representará un subconjunto de linajes que ha coalescido. Para cualquier $n \in \mathbb{N}$, sea $[n] = \{1, \dots, n\}$, entonces ρ_n será el espacio de particiones en $[n]$.

Sea N , nuestro tamaño de población, nuestro proceso ancestral definido en un intervalo de tiempo $I = [-T, 0]$, donde T normalmente es ∞ , comenzará con una partición en que cada uno de los elementos esté separado de los demás, de ahí en adelante, cada coalescencia determinará una partición

π_t , perteneciente a el conjunto de particiones ρ_N , tal que $i \sim j$ si y sólo si v_i y v_j , $v_i, v_j \in \pi_0$, comparten un ancestro al tiempo $-t$. El proceso $\{\pi_t, 0 \leq t \leq T\}$ es un *proceso coalescente*.

El siguiente ejemplo es meramente ilustrativo.

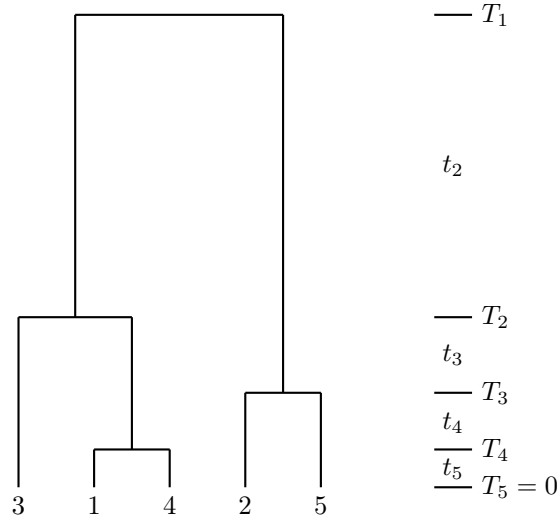


Figura 2.1: Ejemplo de un coalescente para una muestra de tamaño 5

T_1	$\{1, 2, 3, 4, 5\}$
T_2	$\{1, 3, 4\}\{2, 5\}$
T_3	$\{1, 4\}\{2, 5\}\{3\}$
T_4	$\{1, 4\}\{2\}\{3\}\{5\}$
$t = 0$	$\{1\}\{2\}\{3\}\{4\}\{5\}$

La partición al tiempo T_i tiene i número de bloques. En el tiempo $t = 0$ todos se muestran solos porque no han coalescido aún.

Definición 2.10. (*Ancestro*) Sea $V = \mathbb{Z} \times [N]$ la matriz en donde el vértice (g, i) corresponde al i -ésimo individuo de la generación g . Decimos que $(g-r, l) \in V$ es un ancestro de $(g, i) \in V$, $g \in \mathbb{Z}$; $r \in \mathbb{N}$; $l, i \in [N]$, si existe una trayectoria de r vértices conectados en la gráfica aleatoria Wright-Fisher, $G := (V, A)$ que comience en $(g-r, l)$ y termine en (g, i) .

Definición 2.11. (*Proceso ancestral o de conteo de bloques*) Sean:

$$A^N(v) := \{s \in V : s \text{ es un ancestro de } v\}$$

$$A^N(v_1, \dots, v_m)(r) = \{u \in \cup_{i=1}^m A^N(v_i) : g(u) = g-r\}.$$

Es decir, $A^N(v)$ son todos los ancestros de v y $A^N(v_1, \dots, v_m)(r)$ todos los ancestros de mi muestra (v_1, \dots, v_m) que viven en la generación $g - r$.

El proceso ancestral de una muestra v de m individuos en la generación g , $v = (v_1, \dots, v_m)$, es el proceso $\{D_r^N : r \in [N]\}$ que cuenta el número de ancestros en las generaciones previas, de los m individuos muestreados. Es decir, $D_0^N = m$ y

$$D_r^N(v) = |A^N(v_1, \dots, v_m)(r)|, \quad \text{donde } |A|, \text{ recordemos, denota la cardinalidad del conjunto } A$$

La probabilidad de que dos individuos muestreados en la misma generación compartan el mismo padre es de $1/N$. En el caso en que no compartan el mismo padre, la probabilidad de que compartan el mismo abuelo es de $1/N$ y así sucesivamente. Eventualmente, si viajamos lo suficiente hacia el pasado, los linajes de nuestros m individuos muestreados inicialmente, habrán coalescido en un solo ancestro, al que se le suele llamar, haciendo referencia al personaje bíblico, la *Eva ancestral* de la muestra.

Si etiquetamos a los m individuos muestreados, entonces podemos ver cómo nuestro proceso define un proceso continuo de Márkov $\{\Pi_t\}_{t \geq 0}$, en ρ_m , el espacio de particiones en $[m]$, que definirá las relaciones de equivalencia. Cada clase de equivalencia de Π_t corresponde a un ancestro vivo al tiempo t antes del presente, que consiste en todas las etiquetas de nuestros m individuos muestreados que desciende de dicho ancestro.

Definición 2.12 (Tiempo al Ancestro Común Más Reciente (TMRCA por sus siglas en inglés)). *En el caso de generaciones discretas que no se traslapan, el Tiempo para llegar al Ancestro Común Más Reciente de dos individuos muestreados en la misma generación tiene una distribución geométrica con probabilidad de éxito $1/N$. Es decir, la probabilidad que el Ancestro Común Más Reciente de estos dos individuos se encuentre T generaciones atrás es pq^{T-1} donde $p = 1/N$ y $q = 1 - p$. Nótese, que la esperanza del TMRCA es N .*

Como generalmente vamos a estar interesados en trabajar con poblaciones lo suficientemente grandes que tengan algo significativo que revelar, si el tamaño de la población lo hacemos tender a infinito, $N \rightarrow \infty$, y apresuramos el tiempo de forma que lo midamos en unidades de tamaño N —similar a como hicimos para conseguir la difusión de Wright-Fisher a partir de los procesos de frecuencia—, entonces el TMRCA de dos individuos muestreados se aproxima a una exponencial de parámetro 1.

Como la probabilidad de que 3 ó más individuos, así como la probabilidad de que dos parejas de individuos coalezcan al mismo tiempo, son ambas de orden $\mathcal{O}(1/N^2)$; entonces el tiempo que debemos de esperar para que ocurran cualquiera de estos dos eventos es de $\mathcal{O}(N^2)$; primero, con probabilidad tendiendo a 1, veremos que todos nuestros linajes coalesceran de dos en dos.

Por lo que el tiempo (en unidades de tamaño N) que deberemos recorrer hacia el pasado antes

de ver la primera coalescencia de linajes ancestrales de nuestra muestra de tamaño m , es aproximadamente una exponencial con tasa $\binom{m}{2}$, ya que es igualmente probable que la primera coalescencia sea entre cualquiera de las $\binom{m}{2}$ parejas posibles de nuestra muestra. Después el tiempo para la siguiente coalescencia será una exponencial con una tasa $\binom{m-1}{2}$, correspondiente a las $\binom{m-1}{2}$ parejas sobrantes. Y así consecutivamente. Vamos a decir que vagamente, este proceso de coalescencias de linajes es lo que llamamos el *Coalescente de Kingman*.

Podemos observar que, debido a que se trata de una población haploide, nuestro proceso recorrido hacia el pasado, dibuja árboles genealógicos que a cada coalescencia se van reduciendo más y más, a diferencia del concepto de árboles genealógicos o familiares (de poblaciones diploides) a los que estamos acostumbrados que se van bifurcando y ampliando.

Cuando medimos el tiempo en unidades de N generaciones, el tiempo que transcurre en el que hay exactamente k linajes, T_k , tiene aproximadamente una distribución exponencial con media $2/k(k-1)$. La probabilidad, cuando N es lo suficientemente grande de que durante las primeras n generaciones se mantengan los k distintos linajes iniciales es:

$$\approx \left(1 - \frac{k(k-1)}{2} \frac{1}{N}\right)^n \approx \exp\left(-\frac{k(k-1)}{2} \frac{n}{N}\right)$$

La primera aproximación es la probabilidad de no tomar dos individuos con el mismo padre y la segunda resulta de que cuando x es muy pequeña $(1-x) \approx e^{-x}$.

Si recordamos, la distribución exponencial con tasa λ está definida por $\mathbb{P}[T > t] = e^{-\lambda t}$ y tiene media $\frac{1}{\lambda}$. Si dejamos que el tamaño de la población tienda a infinito, $N \rightarrow \infty$, y expresamos el tiempo en términos de N generaciones, es decir, $t = n/N$, entonces el tiempo hasta la primera coalescencia converge a una distribución exponencial con media $2/k(k-1)$. Usando la terminología del proceso ancestral, k linajes coalescen a $k-1$ a una tasa $k(k-1)/2$. Como este razonamiento aplica a cualquier tiempo en que haya k linajes, obtenemos el resultado deseado.

Entonces podemos pensar en el siguiente lema.

Lema 2.1. *Si ω_m es el TMRCA de una muestra de m individuos cuya genealogía está determinada por un coalescente de Kingman, entonces,*

$$\mathbb{E}[\omega_m] = 2(1 - 1/m).$$

Demostración. Como $\omega_m = T_m + T_{m-1} + \dots + T_2$, donde T_i , $i \in \{2, 3, \dots, m\}$, como vimos anteriormente, son funciones exponenciales con tasa $\binom{i}{2}$, por lo que su esperanza es $\mathbb{E}[T_i] = 2/i(i-1)$, entonces

$$\begin{aligned} \mathbb{E}[\omega_m] &= \sum_{i=2}^m 2/i(i-1) && \text{La esperanza de una suma} \\ &= 2 \sum_{i=2}^m \left(\frac{1}{i-1} - \frac{1}{i} \right) && \text{porque } i-1-i=1 \\ &= 2(1-1/m) && \text{eliminando.} \end{aligned}$$

□

Lo que debe llamar nuestra atención sobre este resultado es que cuando $m \rightarrow \infty$, esta cantidad converge a 2. Podemos ver que la esperanza del TMRCA de una población con m infinito es tan solo el doble de tiempo a cuando sólo se muestrean 2 elementos y comenzamos con únicamente dos linajes distintos, es decir, cuando $m = 2$, en la que $\mathbb{E}[\omega_2] = 1$. Esto se refleja en nuestro diagrama en la Figura 2.1. Nótese que al principio en un periodo breve, las coalescencias suceden de manera efervescente, muchas parejas coalescen rápidamente, seguido por un periodo en el que los pocos ancestros sobrantes se toman su tiempo antes de coalescer. Por lo tanto, agregar muchísimos individuos a nuestro muestreo no significa agregar mucha información a nuestro proceso. Es más, cuando $m \rightarrow \infty$, el tiempo esperado para que ocurra la última coalescencia es siempre al menos la mitad del tiempo total esperado de la coalescencia total.

2.2.1. Coalescente de Kingman

Ahora sí, definamos formalmente el Coalescente de Kingman.

Definición 2.13 (Coalescente de Kingman). *Un coalescente de Kingman de una muestra de tamaño m es una cadena de Márkov en tiempo continuo en ρ_m , el espacio de particiones de $[m]$, con tasas de transición $q_{\alpha,\beta}$, ($\alpha, \beta \in \rho_m$) donde,*

$$q_{\alpha,\beta} = \begin{cases} 1, & \text{si } \beta \text{ se obtiene coalesciendo dos clases de equivalencia de } \alpha \\ 0, & \text{en otro caso.} \end{cases}$$

Por convención, la condición inicial es la partición en la que cada individuo está en un bloque distinto.

Teorema 2.5. *Sea una muestra de m individuos de la población, $m \geq 1$ fija, elegidos aleatoriamente al tiempo $t = 0$; si $\pi_t^{N,m}$ denota la partición ancestral al tiempo t de estos m individuos, tal que $i \sim j$ si y sólo si v_i y v_j comparten ancestro al tiempo $-t$; $v_i, v_j \in V_0$. Entonces conforme dejamos a $N \rightarrow \infty$, si mantenemos m fija y aceleramos el tiempo por un factor N ,*

$$(\pi_{\lfloor tN \rfloor}^{N,m}, t \geq 0) \rightarrow_d (\pi_t^m, t \geq 0),$$

donde $(\pi_t^m, t \geq 0)$ es el coalescente de Kingman. Es decir, el número de ancestros de una muestra converge en distribución al coalescente de Kingman.

Heurística. Si v_i y v_j , son dos individuos tomados aleatoriamente al tiempo $t = 0$, entonces el tiempo en que tardarán estos v_i y v_j en coalescer es una geométrica con probabilidad de éxito $p = 1/N$. Sea T_N una variable aleatoria geométrica con parámetro $1/N$. Como $\frac{1}{N}T_N \rightarrow_d \exp(1)$, una exponencial de parámetro 1, decimos que cuando aceleramos el tiempo conforme a N , (v_i, v_j) coalescen a una tasa aproximada de 1. Como esto es verdad para cualquier pareja de la población obtenemos el coalescente m de Kingman.

□

En varios modelos simples de población (de tamaño constante, neutros y panmícticos), conocidos como *modelos de Cannings*, la genealogía de una muestra de población converge al coalescente de Kingman. Siempre se llega a esta aproximación mediante el límite del proceso ancestral alcanzado cuando el tamaño de la población, N , tiende a infinito y la reparametrización del tiempo por un factor, convenientemente elegido, de N . Lo que nos revela que hasta cierto punto el coalescente de Kingman es universal, es decir, su manifestación es poco sensible a los detalles microscópicos detrás del modelo de probabilidad. Éste es un resultado gracias a Möhle, [1] y [3].

2.3. Dualidad de los Momentos

En la década de los ochenta, se demostró que la noción de dualidad entre los límites de los procesos, el de frecuencia de alelos y el de ancestría, es una gran herramienta para obtener nueva información. Combinar eficientemente las informaciones provenientes de la difusión de Wright Fisher y del coalescente de Kingman, es la clave para entender varios aspectos de los modelos biológicos.

Teorema 2.6 (Dualidad de Momentos). *La dualidad de Momentos con respecto a las leyes de la difusión Wright Fisher y las del coalescente Kingman, significa que para toda $x \in [0, 1]$ y $m \in N$*

$$\mathbb{E}[X_t^m | X_0 = x] = \mathbb{E}[x^{D_t} | D_0 = m] \quad (2.3)$$

La primera parte de la ecuación 2.3 revisa el proceso hacia el futuro y la segunda hacia el pasado, [1].

2.4. Selección Natural

Ésta es la parte esencial de la presente tesis. Como motivaciones están: el hecho de que investigadores han percibido que en ecosistemas oligotróficos (bajos en nutrientes) la competencia interespecífica —es decir, aquella que se da entre distintas especies que se disputan por los mismos

recursos— provoca mayor diversidad genética, esto debido a la selección natural [10]; y que, en ocasiones, el que un tipo alélico sea minoría puede verse como una ventaja selectiva en algunas redes ecológicas [5].

Una selección con *dependencia de frecuencia* es aquel proceso evolutivo de interacciones biológicas en el cual la aptitud de un alelo depende de la composición de una población, es decir, de la frecuencia de los otros tipos alélicos.

Diremos que una selección con *dependencia de frecuencia positiva* es aquella en la que la aptitud de un tipo alélico se incrementa cuando éste es más común en una población. Por, el contrario, una selección con *dependencia de frecuencia negativa* será aquella en la que la aptitud de un tipo alélico decrece si es más escaso en la población.

Las selecciones con dependencia de frecuencia, normalmente son el resultado de interacciones entre especies (depredación, parasitismo o competencia) o entre tipos alélicos (competitividad). Este tipo de selecciones permiten un equilibrio polimórfico, pues permiten la cooperación a través de la competencia.

Para ejemplificar esto con casos en la naturaleza tenemos, por un lado, los morfismos en machos de las lagartijas de costado manchado (*Uta stansburiana*), de los que hablamos en la introducción; y por el otro, está la investigación de la que se habla en el artículo [10], sobre la biodiversidad de Microfungi en la Cuenca de Cuatro Ciénegas (CCB), México. La CCB es un ecosistema ultra-oligotrófico (i.e., con bajos niveles de nutrientes) que alberga el más alto índice de endemismos (i.e., tener especies nativas que no se encuentran en ninguna otra parte del mundo) en Norte América y que incluye tres sistemas muy distintos entre sí y fuertemente divididos por fronteras de fósforo: Churince, Becerra y Pozas Rojas. Estos sistemas acuáticos están dominados por una alta diversidad procariótica (de organismos unicelulares como las bacterias). En el estudio que se detalla en [10] se registró poca diversidad microfúngica comparada con la de otros ecosistemas áridos, cuestión que los investigadores atribuyen, posiblemente, a las condiciones selectivas, tales como, la oligotrofia del ecosistema y las interacciones antagónicas con comunidades procarióticas con mayor competitividad.

En esta tesis trabajamos con un modelo de Wright-Fisher con selección que incluye competencia por redes ecológicas. Esto es la selección entendida como en el artículo [5], en donde un individuo tiene la posibilidad de primero hacer un número aleatorio de observaciones en la generación anterior y, posteriormente, de entre los distintos padres potenciales observados, elegir como padre real al más “apto” y heredar de éste su tipo alélico. Para esta tesis, en particular, se podrán hacer un máximo de dos observaciones y la función de probabilidad que determine el número de observaciones será una uniforme entre $\{1, 2\}$.

Vamos a considerar diversas formas de selección natural y después investigaremos qué efectos tiene cada una en la genealogía y en los patrones de variabilidad genética.

Para elegir al padre real, una vez que se conoce al conjunto de padres potenciales, introducimos

una *regla de coloración o interacción selectiva*, que puede depender de N y de las características del conjunto de padres potenciales (número y frecuencia de cada tipo) o que puede ser aleatoria.

Cuando decimos que un tipo alélico es más fuerte selectivamente que otro nos referimos a que su probabilidad de eventualmente fijarse es mayor.

Al permitir que en nuestro modelo los individuos puedan elegir con probabilidad positiva entre más de un padre potencial, hace que nuestro modelo ya no sea neutral, puesto que el padre real de cada individuo, ya no se va a elegir de manera uniforme entre los N individuos de la generación anterior, sino en base a las *interacciones selectivas*. Podemos decir que, la función de probabilidad que dicta el número de observaciones que hace cada individuo, parametriza por completo el mecanismo de selección.

Usualmente, por su sencillez, las interacciones selectivas que se utilizan en los modelos de genética de poblaciones son las del tipo *transitivo*. Por decir un ejemplo, sean a , b y c los tipos alélicos, entonces una interacción transitiva sería $a < b < c$; donde a es más débil que b y b a su vez más débil que c . De modo que si un individuo pudiera observar a un padre de cada tipo alélico: a , b y c ; siempre “ganaría” c . En este caso, es sencillo señalar que c es el más fuerte selectivamente de los tres tipos alélicos, seguido por b y que por último está a .

Pero nosotros, vamos a revisar también otras interacciones selectivas *no transitivas*; por ejemplo, la interacción llamada *Piedra, Papel o Tijera*. Donde, como en el juego de niños, “Piedra” pierde contra “Papel”, pero gana contra “Tijera” y, de este modo, cada uno de los alelos se encuentra en desventaja contra uno de sus opositores y en ventaja con respecto al otro. Aquí, se puede notar, que no hay ningún tipo que sea más fuerte que los otros, ni ninguno que sea más débil que los demás, todos tienen la misma probabilidad de extinguirse y la misma probabilidad de fijarse. En casos como éste diremos que se trata de una *regla justa*.

Capítulo 3

Construcción del modelo

Una vez que tenemos claro el sustento teórico (matemático y biológico), es importante desmenuzar el modelo específico que utilizamos para el programa de simulación de evolución genética —sus características, propiedades, comportamiento y, ¿por qué no?, sus limitaciones—; para así entender las gráficas y datos que dicho programa arrojó y poder abrir un espacio de discusión al respecto.

Con este propósito, basados en el artículo [6], trazaremos la gráfica aleatoria del modelo. Comenzamos mencionando que, como ya se podrán imaginar a estas alturas, se trata de un modelo de Wright-Fisher de tamaño constante, multitipo, sin mutaciones y con selección generalizada que permite la formación de redes ecológicas. En este modelo cada individuo podrá, de manera aleatoria, hacer 1 o 2 observaciones en la generación anterior antes de elegir a su padre real. Como se trata de una población haploide, cada individuo tendrá un único padre real.

Requerimos de un modelo multitipo, por lo tanto, tendremos $k \in \mathbb{N}^*$ tipos alélicos. Sea E el espacio de los tipos de alelos, entonces $|E| = k$. Cada individuo se “pintará” de un color o de otro según la carga genética heredada de sus ancestros, por lo que de ahora en adelante también nos referiremos a los tipos alélicos como colores.

Un individuo puede no tener descendencia o ser padre de uno o de más de un vástago; pero, eso sí, todo individuo debe tener uno, y sólo un, padre. Una vez que en alguna generación todos los individuos de cierto tipo alélico dejan de tener descendencia, es decir, una vez que el alelo se extingue, ya no podrá haber en la siguiente generación, ni en ninguna más adelante, descendencia con dicho color.

Cuando en una generación pasa que, por primera vez todos los individuos de la población tienen el mismo color, es decir, cuando se han extinguido los demás $k - 1$ colores, decimos que el tipo alélico “sobreviviente” se ha fijado. Una vez que esto ocurre, en las generaciones posteriores sólo habrá individuos de este color, por lo que la fijación de un alelo representará una condición de paro de nuestro programa de simulación. Como veremos claramente en nuestras simulaciones, la fijación

y la extinción de un alelo son procesos que dependen de N , el tamaño de la población.

La gráfica aleatoria del modelo consiste en un subconjunto de \mathbb{Z}^2 (generación x índice del individuo), con aristas también aleatorias, que denotarán la relación entre los individuos y sus padres potenciales. Esta gráfica ayudará a visualizar la dualidad entre el proceso de frecuencia y el proceso ancestral en el mismo espacio de probabilidad. A continuación enumeraremos los pasos para la construcción de la gráfica aleatoria relativa al modelo:

1. **Matriz:** habrá N individuos, un número natural constante, en cada generación, y las generaciones correrán por todos los enteros. Llamaremos a esta matriz V , $V := \mathbb{Z} \times [N]$.

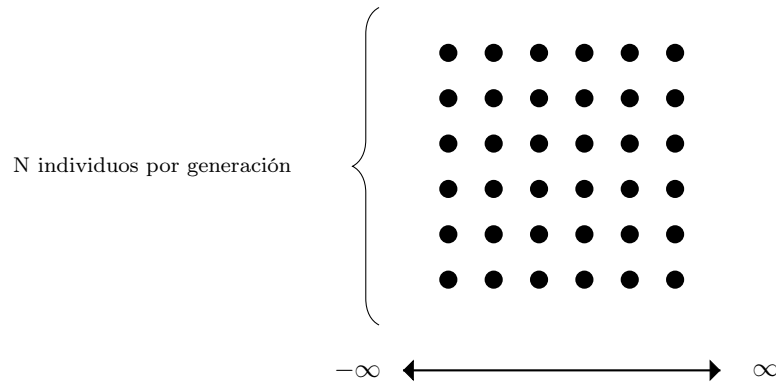


Figura 3.1: Matriz V

Cada uno de los vértices (puntitos) de nuestra gráfica representará a un individuo. Diremos que $v = (i, g) \in V$ será el i -ésimo individuo de la generación g .

2. **Generación inicial:** Fijamos una generación inicial $a \in \mathbb{Z}$; para fines prácticos, en este documento, utilizaremos la generación al tiempo cero, V_0 , como generación inicial. Arbitrariamente o en base a los resultados recolectados en una muestra, coloreamos a V_0 , es decir, le asignamos a cada uno de sus vértices, tipos alélicos. Por ejemplo, sea $k = 2$; $E = \{\text{naranja, azul}\}$:

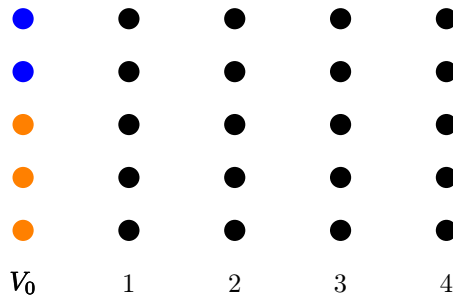


Figura 3.2: Generación inicial

Si se trata de un modelo neutral, una vez que se tienen los tipos alélicos de la primera generación, los padres de los individuos de la siguiente generación se eligen uniformemente al azar

entre los individuos de la generación cero; y así sucesivamente. Una vez que se conoce quién es padre de quién se puede colorear el resto de la matriz siguiendo la norma de que los padres pasan su color o tipo alélico a sus hijos. Como se ilustra en el siguiente diagrama:

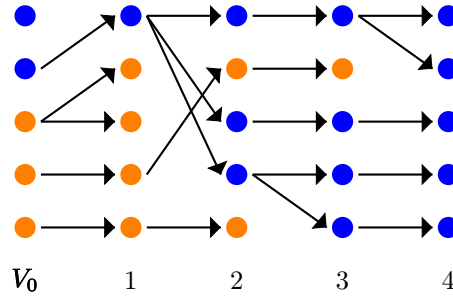


Figura 3.3: Coloración en modelo neutral

3. **Número de observaciones:** Para nuestro modelo que no es neutral, en cada generación $g \in \mathbb{Z}$ todo individuo i , $i \in [N]$, tal que $v = (i, g) \in V$, hará independientemente y con reemplazo $P(i, g)$ o P_v número de observaciones entre los N individuos de la generación $g-1$, donde $P(i, g)$ es una variable aleatoria que toma valores en $\{1, 2\}$, donde $\mathbb{P}[P(i, g) = 2] = x$ y $\mathbb{P}[P(i, g) = 1] = 1 - x$. Aunque en el artículo [6] se desarrolla el modelo para que cada individuo pueda hacer en la generación anterior cualquier número de observaciones con reemplazo, de los cuales obtendrá desde uno y hasta N (el total de la población) padres potenciales; además, de que el modelo del artículo [6] permite la mutación (es decir, la posibilidad de que ninguno de los padres potenciales observados sea el padre real); para efectos de esta tesis simplificaremos y diremos que un individuo puede hacer una o máximo 2 observaciones.

Diferenciaremos entre tres términos que se pueden confundir fácilmente:

- Una *observación* significa considerar a cierto individuo de la generación anterior como contendiente de la interacción selectiva, independientemente de su tipo alélico o del índice que tenga dentro de su propia generación. Se podrá observar más de una vez al mismo individuo de la generación anterior.
- Un *padre potencial* es cada uno de los diferentes tipos alélicos recolectados en las $P(i, g)$ observaciones hechas por el individuo (i, g) . Las observaciones dirigidas más de una vez hacia el mismo tipo equivalen a un solo padre potencial.
- Los *padres virtuales* son los distintos vértices en la generación $g-1$ a los que llegan las aristas de un individuo en la generación g . Puede haber más de uno del mismo tipo alélico. Llevar el registro de los padres virtuales no tiene ningún objetivo cuando se estudia el modelo Wright-Fisher de atrás para adelante, pero sí cuando se revisan los linajes en el proceso ancestral.

Lo anterior quiere decir que no necesariamente todas las observaciones que hace un individuo van dirigidas a diferentes individuos o tipos alélicos, estos se podrían repetir.

4. **Elección de padres potenciales:** Ya decidimos cuántas observaciones hará cada uno de los vértices en V , ahora queremos elegir hacia cuáles de los N individuos de la generación anterior se harán estas observaciones. Para tales instancias, llamaremos a otra familia de variables aleatorias independientes $\{R(v, p), v \in V, p \in \mathbb{N}^*, p \leq P_v\}$, con distribución uniforme en $[N]$, que serán las selecciones entre los N individuos de la generación anterior. Note que, esta elección es con reemplazo para satisfacer el hecho de que se puede observar más de una vez al mismo individuo en la generación anterior. De modo que, $v \in V_g$ hará observaciones de la siguiente forma: $\{(g - 1, R(v, p)); p \leq P_v\}$.

A continuación, dibujaremos una gráfica que trace aristas entre cada vértice y los individuos que observó. Vamos a ir recorriendo cada generación y checando qué observaciones hizo cada individuo en la generación anterior.

El conjunto de aristas dibujadas será el siguiente:

$$A = \{(v, (g - 1, R(v, p))), \text{ para toda } v = (g, l) = V_g \in V; \text{ tal que } 1 \leq l \leq N; 0 < p \leq P_v\}$$

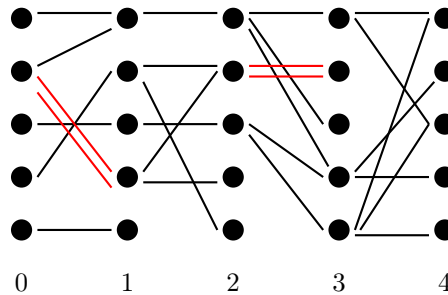


Figura 3.4: Observaciones

En el diagrama 3.4 se puede notar que hubo individuos que hicieron una observaciones, mientras que otros habrán hecho dos observaciones. También, resaltado con dos líneas paralelas en rojo, se denota cuando las dos observaciones que hizo un individuo se dirigieron hacia el mismo individuo en la generación anterior.

5. **Selección del padre real:** Una vez que un individuo realiza sus respectivas observaciones en la generación previa, deberá elegir a su padre real de entre sus padres potenciales. Será la regla de coloración, C_N , o interacción selectiva la cual determine esto. Cada vértice se coloreará dependiendo de qué tipos alélicos tengan los vértices de la generación anterior con los que se conecta mediante las observaciones. Esta regla de coloración puede ser aleatoria o puede depender de las características del conjunto de los padres potenciales observados y del tamaño de la población.

Tendremos un conjunto H que será el espacio de todo los posibles vectores en \mathbb{Z}^2 de padres potenciales que puede tener un individuo,

$$H = \{(z_1, z_2) : z_1 \in E \text{ y } z_2 \in E \cup \{0\}; \text{ si } z_2 \neq 0 \Rightarrow z_1 < z_2\}.$$

Diremos que $\{z_1; z_1 \in E\} = \{(z_1, 0), z_1 \in E\}$, será el caso cuando sólo haya un padre potencial, ya sea porque sólo se hizo una observación o porque se repitió el tipo alélico en las dos observaciones.

Una *regla de coloración* $C_N(N, E)$ es una familia de funciones de distribución sobre E tal que,

$$C_N = \{c_z(e) \mid e \in E\}_{z \in H}$$

donde,

$$c_z(e) = \mathbb{P}[\text{Tipo}(v) = e \mid v \text{ observó a } z \in H]$$

y sea $z \in E$, o de otra manera dicho, si z es del tipo $\{z_1, 0\} \in H$,

$$c_z(e) = \mathbb{P}[z = e],$$

con lo cual garantizamos que si sólo observa a un padre potencial éste será su padre real.

Volvamos a nuestro ejemplo sencillito de azules y naranjas. En este caso $H = \{(\text{naranja}, \text{azul}), (\text{azul}, \text{naranja}), (\text{naranja}, \text{naranja}), (\text{azul}, \text{azul}), (\text{azul}), (\text{naranja})\}$ Entonces,

$$\begin{aligned} C_N = \{ & \mathbb{P}[v \text{ sea naranja} \mid v \text{ observó } (\text{naranja}, \text{azul})], \\ & \mathbb{P}[v \text{ sea naranja} \mid v \text{ observó } (\text{azul}, \text{naranja})], \\ & \dots \\ & \mathbb{P}[v \text{ sea naranja} \mid v \text{ observó } (\text{naranja})], \\ & \mathbb{P}[v \text{ sea azul} \mid v \text{ observó } (\text{naranja}, \text{azul})], \\ & \mathbb{P}[v \text{ sea azul} \mid v \text{ observó } (\text{naranja}, \text{azul})] \} \end{aligned}$$

6. **Coloración:** Ahora vamos a ilustrar cómo es que se difunden los tipos alélicos tras las interacciones selectivas. Una gráfica de coloración (V, E) es una función f que irá pintando cada vértice con respecto al color que hereda de su padre real. Se puede entender a (V, E) como una familia de variables aleatorias que tienen a C_N como funciones de distribución.

Para cada $v \in V_g$, revisamos las aristas de sus observaciones hechas a la generación anterior para determinar sus padres potenciales. La función revisará qué padres potenciales tienen los

vértices y a partir de esto, usando la regla de coloración C_N , definirá qué tipo alélico adoptará cada vértice.

Sabemos cómo se ve V_0 porque hemos determinado desde el paso 2 sus tipos alélicos; ahora, definimos la función f para esta primera generación. Y después usando la función de coloración f de forma recursiva podremos ver cómo se propagan los tipos alélicos en las siguientes generaciones.

Sea $z_v \in H$ el conjunto de padres potenciales del vértice v , entonces, para cada tipo $i \in E$,

$$\mathbb{P}[f(v) = i] = c_{z_v}(i).$$

Es decir, la probabilidad de que el vértice v se coloree del tipo i es igual a la probabilidad de que gane el padre potencial tipo i dado que observó al subconjunto z_v de H .

$$z_v = (f(r_1), \dots, f(r_{P_v})).$$

Entonces, $\{r_i \mid i \in [P_v]\}$ son los índices de los individuos en la generación $V_{g-1} \in V$ que observó el vértice $v \in V_g \in V$, que llevan el color que les corresponde a partir de haberse sometido previamente a la misma función de coloración, f .

Por lo tanto, podemos decir que f , la función de coloración es una función recursiva, tal que, se puede definir como una familia de funciones $\{f_n\}$ donde

$$f_n : V_n \longrightarrow E$$

$$f_n(v) := \text{tipo de } v \text{ en } V_n.$$

Se puede observar que $\{f_n\}$ se trata de una cadena de Márkov en tiempo discreto, puesto que únicamente el estado de la generación inmediatamente anterior nos basta para conocer sus probabilidades en el futuro.

Para seguir ilustrando con nuestro ejemplo de naranjas y azul, digamos que cada vez que compiten naranja y azul, siempre gana naranja. Entonces así se propagarán los tipos alélicos.

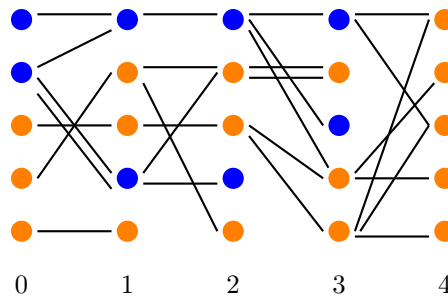


Figura 3.5: Naranja le gana siempre a azul

Así pues, esto que parece una pintura de Kandinsky, no es otra cosa que una gráfica de coloración (V, E) hasta la fijación del color naranja después de 4 generaciones. Entonces, la regla de coloración $C_N(N, E)$ va a señalar cuál es la probabilidad de que “gane” cada uno de los tipos alélicos en E dependiendo de contra quien compite. La regla de coloración que se propuso para este ejemplo evidentemente favorecía al tipo naranja, aunque el tipo azul resistió heroica y estoicamente su extinción hasta la generación 3.

Así es cómo coloreamos nuestra población de generación en generación. Para algunas reglas de coloración no será nada evidente identificar al tipo más fuerte selectivamente y muchas veces la fuerza o debilidad de un tipo radicará en el comportamiento de sus demás contrincantes.

3.1. Nuestro programa

Para escribir el código del programa que se presenta en esta tesis se consultaron los cursos de *Data Science* en Python de [8].

3.1.1. Especificidades de nuestro programa

Como hemos dicho hasta ahora, nuestro programa computacional simula la evolución del perfil genético de una población que mantiene constante su tamaño y que puede tener cualquier cantidad de tipos alélicos, aunque en la práctica se graficarán un máximo de diez diferentes tipos alélicos.

En el programa se grafican diagramas de barras apiladas donde cada color representa un tipo alélico diferente: sean k tipos alélicos, entonces la barra de hasta abajo es la asignada al tipo con la etiqueta "1", la que esté encima de ésta, es la del tipo "2" y así; hasta que la de más arriba es la barra del tipo " k ". Por el eje de las equis corren las generaciones de cero en adelante. La altura que alcanza la barra de algún color en específico en cierta generación indica la incidencia del tipo que representa en dicha generación. Siempre, por tratarse de un modelo Wright Fisher, la suma de las alturas de todas las barras de todos los colores en cada generación será N .

El programa le dará al usuario la oportunidad de observar las alteraciones en la composición genética de una población a partir de su estado inicial y a través del transcurso de las generaciones

hasta que se sucedan al menos una de las siguientes dos condiciones: que algún tipo alélico se fije (lo que es igual a que $k-1$ tipos se extingan) o que se llegue a la generación 2000 (esta segunda condición se agregó para ponerle fin a simulaciones que se seguían y seguían sin concluir después de demasiadas generaciones, sin embargo, se puede quitar borrando una línea de código). Lamentablemente, después de 150 generaciones la gráfica ya no se ve muy clara, por lo que, para efectos de las imágenes aquí presentadas, se hicieron pruebas con un límite superior de 150 generaciones.

3.1.2. Reglas de coloración en nuestro programa

La mayoría de las reglas de coloración que utilizaremos —a excepción de, como más adelante veremos, la de Órbita y la Personalizada— tendrán la característica de que, dada la competencia entre dos tipos alélicos diferentes, el ganador y el perdedor de dicha contienda estarán establecidos de antemano con probabilidad uno. Es decir, para cada par de tipos posible, la regla tiene estipulado cuál de los dos se transmitirá y cuál no. El vértice que los observa replicará el color del vencedor. Es decir, sean $i \neq j$ y $z_v = \{i, j\} \in H$ los padres potenciales del vértice v , entonces $c_{z_v}(i) = 1$ o $c_{z_v}(j) = 1$.

En todas nuestras reglas de coloración, no importará en qué orden se hicieron las observaciones. Sean $i \neq j$, si $z_{v_1} = \{i, j\}$ y $z_{v_2} = \{j, i\}$, $z_{v_1}, z_{v_2} \in H$, entonces $c_{z_{v_1}}(i) = c_{z_{v_2}}(i)$ y $c_{z_{v_1}}(j) = c_{z_{v_2}}(j)$.

Además, en el caso de haber un único tipo compitiendo por la heredabilidad, $z_v = \{z_1, 0\} \in H$, ya sea porque se observan en dos ocasiones al mismo tipo o porque ese vértice sólo hace una observación, resultará en la permanencia de dicho tipo. $c_{z_v}(z_1) = 1$.

Habiendo dicho lo anterior, expliquemos una por una cada regla presentada en este trabajo:

- **Transitiva:** Ya habíamos dicho que ésta es la regla de coloración que más comúnmente se utilizaba en la genética de poblaciones debido a su sencillez. Esta regla supone que existe una manera de ordenar los tipos alélicos de mayor a menor fuerza selectiva, tal que, si compete un tipo contra otro de menor denominación, gana el primero.

Esto significa la rápida fijación del más fuerte y la extinción del resto, como pudimos corroborar en nuestra experimentación.



Figura 3.6: Diagrama de la regla transitiva

En el diagrama 3.6 la raya de las flechas une por sus extremos los tipos que compiten y con la punta indica cuál es el ganador de dicha interacción, es decir, las flechas van del tipo que vence hacia aquel que es vencido. Esta convención la vamos a mantener a la hora de ilustrar con diagramas las demás reglas. En el diagrama 3.6 no se muestran explícitamente

las relaciones entre tipos que no son consecutivos, por ejemplo entre el tipo T_4 y el T_2 , esto es por hacer el diagrama más sencillo, pero esperamos se entienda que, como su nombre lo dice, las competencias son transitivas.

- Gana el otro: Esta regla sólo es posible cuando son tres y sólo tres tipos alélicos. En nuestro programa si se le pide simular esta regla con una cantidad de tipos distinta a tres, se señalará el error y se le pedirá al usuario que nuevamente confirme el número de tipos y la regla que se quieren modelar.

Esta regla consiste en que cuando compiten dos tipos alélicos distintos entre sí, digamos a y b , resulta ganador el tercer tipo, c , aquel que no compitió.

Según las simulaciones, en este caso, las frecuencias relativas de los tres tipos no varían mucho de un tercio y pareciera imposible que se extinguiera cualquiera, pues si llegase a haber un alza en algún tipo, esta misma alza será su condena porque estará más propenso a entrar en combate, lo cual favorecerá a cualquiera de los otros dos tipos y no al tipo en cuestión.

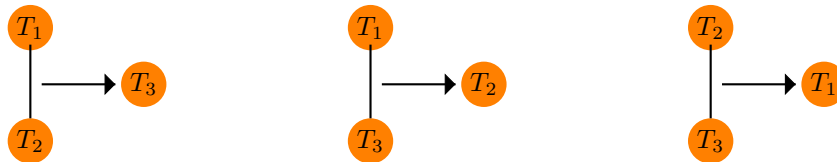


Figura 3.7: Diagrama de regla Gana el Otro

- Órbita: La regla de Piedra, Papel o Tijera, de la que habíamos hablado anteriormente, es el caso específico para tres tipos alélicos de una regla más general a la que llamaremos Órbita, en la que, hay una manera de enumerar a los tipos alélicos de forma que, si compiten dos tipos subsecuentes, gana el mayor de estos y si compiten el etiquetado con el número uno con el de la numeración mayor, k , gana el uno, lo que nos conduce a una dinámica cíclica.

Este tipo de regla de coloración sirven para describir una red alimenticia, en donde si j se come a i decimos que $j = i + 1$. En los diagramas de abajo podemos ver claramente porque lleva el nombre de Órbita.

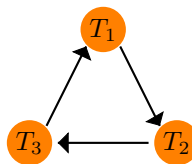


Figura 3.8: Diagrama de regla órbita con 3 alelos: Piedra, Papel o Tijera

En el caso de Piedra, Papel o Tijera el ciclo define todas las relaciones posibles entre los tipos, pero para $k > 3$ debemos ver cómo se definirán las “batallas selectivas” entre dos tipos

que no son subsecuentes. A estas interacciones las llamaremos *interacciones interorbitales*. En nuestro programa, definimos a estas interacciones interorbitales diciendo que con probabilidad de $1/2$ ganará uno y con la misma probabilidad ganará el otro, de modo que nuestra regla de coloración Órbita será siempre justa para todos los tipos.

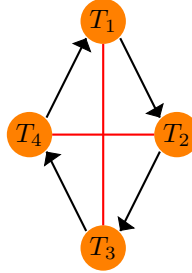


Figura 3.9: Diagrama de la regla Órbita con 4 alelos. Las flechas van del tipo que tiene menos ventaja selectiva al que tiene más. Las aristas rojas representan competencias neutras en las que cualquiera de los dos tipos puede ganar con la misma probabilidad.

Observemos en el diagrama 3.9 que, para cualquier cantidad de tipos alélicos, la órbita (las relaciones orbitales) define únicamente la relación de cada uno de los tipos con el siguiente y con el anterior, es decir con 2 de los otros tipos. Por lo que, sea $k \in \mathbb{N}^*$, $k > 3$, el número de alelos, entonces el número de interacciones “interorbitales” que cada tipo tendría con otros alelos sería $k - 3$. Pero como no importa el orden en que se observen los tipos, habrá un total de $k(k - 3)/2$ interacciones “interorbitales”.

Se puede notar que, si estas interacciones “interorbitales” estuvieran definidas de tal modo que uno de los alelos participantes resultara siempre ganador y el otro perdedor, como están definidas las interacciones orbitales; cuando k fuera impar, es decir, de la forma $k = 2n + 1$, $n \in \mathbb{N}^*$, existiría la manera de distribuir los resultados de las interacciones para que cada tipo gane contra el mismo número de tipos que contra los que pierde y de esta forma nuestra regla de coloración podría ser justa. Pues $k = 2n + 1 = (2(n - 1)) + 3$, entonces $k - 3 = 2(n - 1)$, el número de interacciones interorbitales que cada tipo tendría, se pueden repartir la mitad para que las gane y la mitad para que las pierda.

Pero al momento de que un alelo se extingue o en el caso cuando k es par desde un principio, es decir, cuando k es de la forma $k = 2n$, $n \in \mathbb{N}^*$, la cantidad de interacciones interorbitales resultará impar y, por lo tanto, la regla no puede, de ninguna manera, seguir siendo justa. Por lo que no existe Órbita justa con todas sus interacciones definidas que al extinguirse un tipo alélico pueda seguir siendo justa.

- Órbita malita: Esta regla es el producto de escribir erróneamente el código para la Órbita clásica, mas resultó tener un comportamiento y peculiaridades muy interesantes, por lo que se consideró que se quedara.

En esta regla, igual que en la Transitiva, existe una manera de enumerar a los tipos, de forma que al competir dos tipos alélicos diferentes siempre ganará el que tenga el rango mayor, a menos de que quienes compitan sean los tipos que tienen el número uno y el del rango mayor, k ; en ese caso, como en la Órbita tradicional, gana uno.

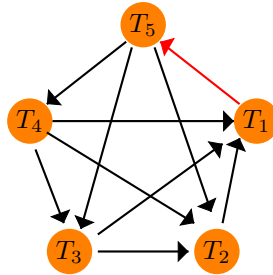


Figura 3.10: Diagrama de la regla Órbita Malita, $k=5$

Como se puede ver, el tipo del índice mayor, es decir, el tipo k le gana a todos excepto al más "débil", el que tiene etiqueta 1. ¿Quién ganará: el que está en desventaja ante un mayor número de alelos, pero es capaz de derrotar al más fuerte o el que es más fuerte ante casi todos los alelos y sólo pierde contra uno?

- Personalizada: En la regla personalizada, el usuario podrá determinar para cada combinación posible de parejas de padres potenciales a competir, cuáles son las probabilidades con que gana cada tipo alélico. Fíjese que, si compiten los alelos i y j , se le puede asignar al alelo k , $k \notin \{i, j\}$ una probabilidad positiva de ganar. Es decir, sea $z_v = \{i, j\}$, entonces $c_{z_v}(k) \geq 0$

3.1.3. Funcionamiento del programa

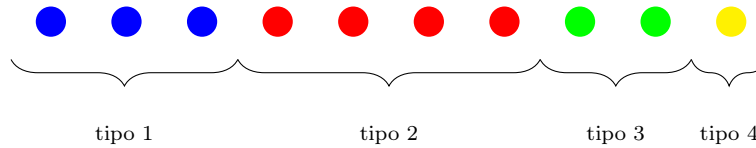
Para arrancar el programa requerimos determinar cómo estará conformada nuestra **generación inicial**, V_0 , (en el código nombrada `V_cero`). Se puede elegir que V_0 se produzca de manera aleatoria –en este caso, una vez que ya se conoce el tamaño de la población, N , y cuántos tipos alélicos, k , se mandan a arrojar N números enteros aleatorios con valores uniformemente distribuidos entre 1 y k –; o, la otra opción, es introducir manualmente las frecuencias absolutas de cada uno de los tipos, por ejemplo, en el caso de que alguien haya estado haciendo un registro de los tipos de cierta población y que ahora quiera simular su comportamiento genético en el futuro.

Por comodidad y debido a que la siguiente generación, V_1 , hará de manera aleatoria sus observaciones uniformemente en $[N]$, convenimos en que en el caso de introducir manualmente los tipos de la generación inicial, el tipo 1 ocupará las primeras casillas, después estará acomodado el tipo 2 y así hasta que las últimas casillas serán para el tipo k .

Así podría verse una generación inicial, V_0 , con un tamaño de población $N = 10$ y cuatro tipos alélicos, $k = 4$, cuando se produce aleatoriamente.



En este caso los tipos aparecen sin ningún orden en especial. Por otro lado, así se verá una generación V_0 en la que se introducen manualmente los datos, también con $N = 10$ y $k = 4$.



Cabe señalar que, en la práctica de la simulación, cuando se le pide a un ordenador que genere números aleatorios, en realidad, lo que éste produce es una sucesión de **números pseudoaleatorios**. Esto quiere decir que se genera una sucesión de números que, aunque son muy buena aproximación a un conjunto de números aleatorios, están contruidos a través de un algoritmo y determinados por un valor inicial. A este valor inicial se le conoce como **estado o semilla aleatoria**. Esto es muy apropiado para nuestra práctica, porque cada vez que se corre el programa utilizando la misma semilla, se producen la misma secuencia. De este modo podemos comparar resultados, dada la misma semilla, pero modificando otros parámetros.

En nuestro programa se le pide al usuario que ingrese los valores para las siguientes variables:

- k : = número de tipos alélicos
- N : = tamaño de la población (en caso de que V_{cero} se genere aleatoriamente, sino, en el caso de ingresar manualmente los datos, N se desprenderá de la suma de los individuos sobre todos los tipos).
- x : = probabilidad de que los individuos hagan dos observaciones en lugar de una, lo cual aumenta, a su vez, la probabilidad de que haya interacción selectiva. Debe estar entre el cero y el uno, en caso de que se pusiera un valor fuera de este rango se marcará un error y se volverá a solicitar x .
- *regla*: = El programa le da opción al usuario de seleccionar entre las cinco reglas de coloración integradas al programa:
 - Transitiva
 - Gana el Otro
 - Órbita
 - Órbita Malita
 - Personalizada.

Estas reglas se explicarán de manera detallada más adelante.

- *personalizado*:= En caso de elegir la regla *Personalizada* se le solicitará al usuario que rellene una matriz en la que por cada competencia posible entre dos de los alelos, tendrá que anotar cuáles son las probabilidades de que gane cada uno de los alelos existentes, incluso aquellos que no compiten. Estas probabilidades deben sumar uno para cada competencia.
- *seed*:= la semilla aleatoria.

Estuvimos jugando con nuestro programa para ver cómo en base a las manipulaciones de estas variables, las gráficas se modificaban o no.

3.1.4. Logística de operación del código del programa

Una vez que ya están determinadas las variables y de que se construyó de manera manual o aleatoria la generación inicial, V_0 , el programa hará el resto de las tareas automáticamente. Pero ¿cuáles son estas tareas?

1. Lista de V_{cero} : Todo comienza con la primera generación, V_{cero} , en forma de una *lista*. Esta *lista* es la serie de tipos alélicos que conforman V_{cero} ordenados según el índice de los individuos.
2. Cuentas: Ahora, V_{cero} se envía a una función que se llama “Cuentas” que analizará como está constituida. “Cuentas” registra en una nueva lista las frecuencias con que se observaron cada uno de los tipos. El primer valor de esta nueva lista será el correspondiente a la frecuencia del primer tipo, el segundo del segundo tipo y así consecutivamente hasta tener una lista con k entradas. “Cuentas” también revisa si algún tipo alélico se extinguió, es decir, si algún valor de la lista de frecuencias es un cero, y de ser así checa cuál.
3. Frecuencia de tipos: La lista de “Cuentas” se guarda a su vez en otra lista de listas, llamada “frecuencia_tipos”, en la que se irán archivando las listas de frecuencias de cada generación comenzando en V_0 .
4. Creación de nueva generación: Primero, se hace una nueva lista, V_{uno} , con N casillas, en las que adentro aún no sabemos qué tipo de individuos habitarán. Para saberlo tenemos que primero preguntarnos cuántas observaciones hará cada quien.
 - a) Cuantos: Como en nuestro programa la función de distribución del número de observaciones, es una función uniforme continua en la que, sea $x \in [0, 1] \in \mathbb{R}$ (la variable que introdujo el usuario):

$$\mathbb{P}[P_v = 2] = \mathbb{P}[U \leq x]$$

$$\mathbb{P}[P_v = 1] = \mathbb{P}[U \geq x].$$

Entonces solicitamos N números aleatorios entre cero y uno y si son menores a x decimos que ese individuo hará dos observaciones y de lo contrario únicamente hará una. De modo, que nos queda una lista de unos y dos, que indicarán cuántas observaciones harán los N individuos en V_uno .

- b) Padres: A continuación, se hará una lista de N vectores bidimensionales, denominada “Padres”, en las que si el i -ésimo individuo de V_uno hace una única observación, el primer término del i -ésimo vector será un número aleatorio en $[N]$ y el segundo será un cero; en caso contrario, si el individuo va a hacer dos observaciones, ambas coordenadas del vector serán números aleatorios en $[N]$. De esta forma conocemos los índices de los individuos en la generación anterior que observarán cada uno de los individuos en ésta.
- c) Regla de coloración: Finalmente, se manda la lista “Padres” a la función “regla de coloración” que funge como conmutador, para de ahí volver a enviar la lista “Padres” a la regla de coloración correspondiente a la que el usuario solicitó utilizar: transitiva, gana el otro, órbita, órbita malita o personalizada.

En cada una de estas reglas de coloración, se checará en la lista de “Padres” los tipos alélicos de los padres potenciales y se harán las respectivas batallas genéticas a fin de tener una lista con los N tipos vencedores. Esta lista constituirá nuestra nueva generación V_uno .

- d) Personalizada: Nos detendremos un momento para explicar cómo es que funciona esta regla de coloración en particular, puesto que puede parecer un poco confusa en un principio. Primero, lo que hace la regla cuando existen dos padres potenciales distintos, es ordenar a los dos contendientes, de forma que queda el de etiqueta menor primero, reetiquetado ahora como a y el de etiqueta mayor después, reetiquetado ahora como b .

Esto es porque, dado que el usuario determinó cierto número de alelos, k ; para enlistar todas las interacciones posibles, $\{(i, j)\}_{i < j, i, j \in E}$ y $E = [k]$, el programa primero fija a i como $i = 1$ y avanza j desde $i + 1$ hasta k , antes de cambiar la primera entrada, i , a 2, para ahora hacer avanzar a j de 3 hasta k y así hasta que la primera entrada llega a $k - 1$ y sólo puede competir contra k , de esta forma, se tienen todas las competencias posibles sin repetición entre los tipos, dado que el número de alelos es k .

En ese mismo orden se le solicita al usuario que rellene la matriz *Personalizado*, en la cual se introducen las probabilidades de que se convierta en padre real cada tipo alélico según se de el caso de cada una de las interacciones selectivas posibles. De forma que estas probabilidades se van ingresando en una matriz en cuyas filas corren todas las interacciones posibles, mientras que en las columnas están cada uno de los tipos alélicos. Entonces, cuando el programa entra a la regla de coloración *Personalizada* y tiene que saber quién resultará vencedor en un combate entre a y b , $a < b$, debe, antes que nada,

Probabilidad de que gane cada tipo					
Índice	Interacción	Tipo 1	Tipo 2	Tipo 3	Tipo 4
1	(1,2)	.2	.1	.2	.5
2	(1,3)	x	x	x	x
3	(1,4)	x	x	x	x
4	(2,3)	x	x	x	x
5	(2,4)	x	x	x	x
6	(3,4)	x	x	x	x

Tabla 3.1: Matriz Personalizado

buscar el índice de la fila en donde se encuentra, establecida por el usuario, la regla de probabilidades de éxito dada dicha competencia. Para esto debemos hacer la siguiente operación:

$$indice = ((a - 1) * k - suma) + b - a - 1,$$

donde $suma = \sum_{h=0}^{a-1} h$.

Una vez que sabemos con qué vector de probabilidad vamos a trabajar, se le pide al programa un número real aleatorio, uniformemente distribuido entre el cero y el uno. Según en qué fragmento de la línea entre cero y uno cae y comparado con mi vector de probabilidades, el individuo de la nueva generación adoptará el tipo del alelo que corresponda a este segmento.

Digamos que estamos con la interacción (1,2) —a la que, premeditadamente, le impusimos valores en la tabla 3.1— y que el número real aleatorio que arrojó el programa es .3567. Entonces:

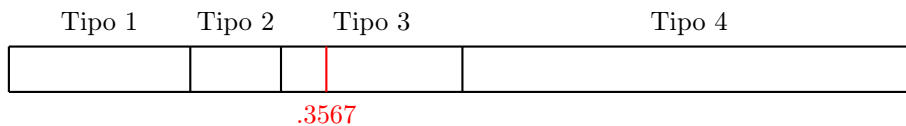


Figura 3.11: Decisión de alelo ganador en competencia Personalizada. En rojo número real aleatorio en $[0,1]$

Como se puede percibir en el diagrama 3.11, el número aleatorio arrojado por el programa cae en la sección del tipo 3 (nótese que los segmentos son de tamaño proporcional a las probabilidades agregadas en la matriz Personalizado), luego entonces resulta ganador en dicha competencia el tipo 3 y es este el tipo que adopta el individuo que hizo las observaciones (1,2). Este proceso lo hace N veces hasta obtener los tipos de los N individuos en V_uno .

- e) Revisión de V_uno : Una vez que conocemos a V_uno se debe analizar del mismo modo que se hizo con V_cero , por lo tanto, se manda V_uno a la función “Cuentas” para saber cuántos individuos hay de cada tipo, cuántos tipos se extinguieron y, si hubo extin-

ciones, cuáles fueron. Posteriormente se añade la lista de frecuencias a la lista de listas “Frecuencia_tipos”.

5. Loop: Antes de reiniciar con el loop, se dice que los valores de V_{uno} se vaciarán en V_{cero} , dejando a V_{uno} vacío provisionalmente, listo para en la siguiente ronda acoger a una nueva generación. En el antiguo V_{uno} , actual V_{cero} , estarán los nuevos padres potenciales.

De aquí el algoritmo entrará a un loop dentro del cual se mantendrá hasta que alguno de los tipos se fije, es decir, hasta que la función “Cuentas” registre $k-1$ ceros. En el loop se generarán cuantas generaciones sean necesarias hasta que la condición de paro se satisfaga.

6. Gráfica: En cuanto se cumple la condición de paro del loop, se arroja la gráfica con los datos de “frecuencia_datos”. Cada vector de esta lista de listas será una de las columnas/generaciones de la gráfica.

Capítulo 4

Discusión

Gracias al programa que construimos en Python pudimos obtener gráficas y tablas de datos que, a veces a simple vista y otras observando con mayor detenimiento, nos dieron la pauta para plantearnos una serie de discusiones que expondremos en el presente capítulo.

4.1. Analizando las gráficas

Para poder comparar los cambios en las gráficas arrojadas por el programa y poder asociar cierto comportamiento a las modificaciones de un parámetro en particular, adaptamos el código en Python, de forma que se le pidieran al usuario únicamente tres de los siguientes cuatro parámetros: tamaño de población, número de alelos, probabilidad de observar dos padres y semilla aleatoria. Entonces, ¿qué valor tomará el parámetro que no se le solicitó al usuarios? La respuesta es, que el programa, gracias a un *loop* que le añadimos, se repetirá varias veces, cada una cambiando el valor del parámetro que no solicitamos y éste irá tomando valores siempre dentro de su rango. Por ejemplo, N , el tamaño de población, puede que comience en 200 y vaya aumentando de 200 en 200 en cada iteración hasta llegar a 1000; o quizá x , la probabilidad de observar dos padres, podría variar de .2 en .2, desde 0, donde esta probabilidad es nula, y hasta alcanzar el 1. (Siempre se puede modificar el código para que cambien: con qué valor inicia y de qué tamaño son los “saltos” de estas variaciones). Nótese que tampoco se le pidió la regla de coloración, esto es porque el programa arrojará en cada iteración las cuatro gráficas correspondientes a las cuatro reglas preestablecidas (tres en caso de que el número de alelos sea distinto a tres): Transitiva, Gana el Otro, Órbita y Órbita Malita.

De este modo, se consiguieron gráficas cuyas diferencias sugieren las siguientes observaciones:

- Al crear la **generación inicial** de manera aleatoria y uniforme se puede percibir, tal y como se esperaba, que todos los alelos comienzan con una frecuencia inicial similar; cada una, aproximada a N/k . Como se puede ver en las siguientes imágenes, independientemente del número

de tipos y de la regla de coloración que utilice, al producir la generación inicial, cada alelo comienza aproximadamente con un número similar de individuos.

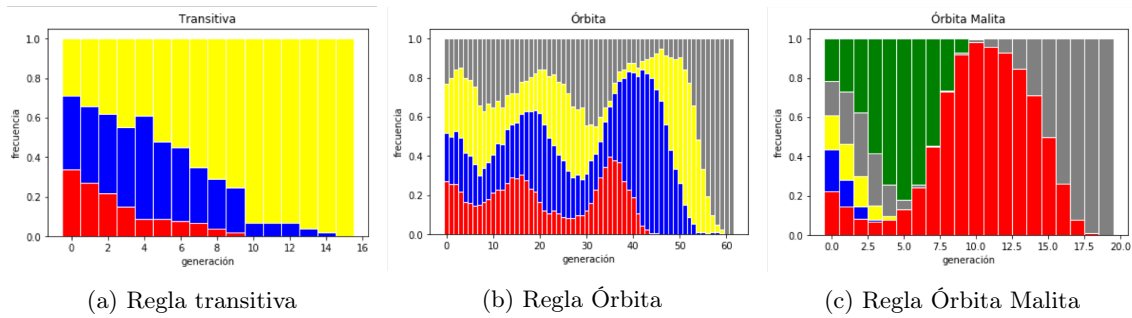


Figura 4.1: Generación inicial aleatoria y uniforme

- Cuando $x = 0$, es decir, cuando la probabilidad de observar más de un padre es igual a cero, todos los individuos en V observan un único padre potencial que será a la postre su padre real, esto quiere decir que no hay interacción selectiva; entonces las gráficas no siguen ninguna tendencia. Por eso, para una determinada semilla aleatoria, sea $x = 0$, el programa devolverá la misma gráfica para las cuatro reglas.

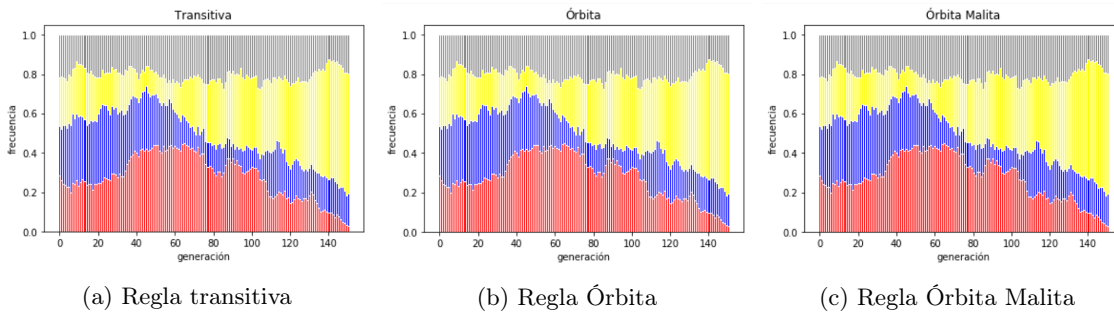


Figura 4.2: $N= 500, k=4, x=0, seed=123$

Por el contrario, cuando $x = 1$, todos los individuos en V harán dos observaciones. Aunque esto no necesariamente implique que siempre habrá competencia.

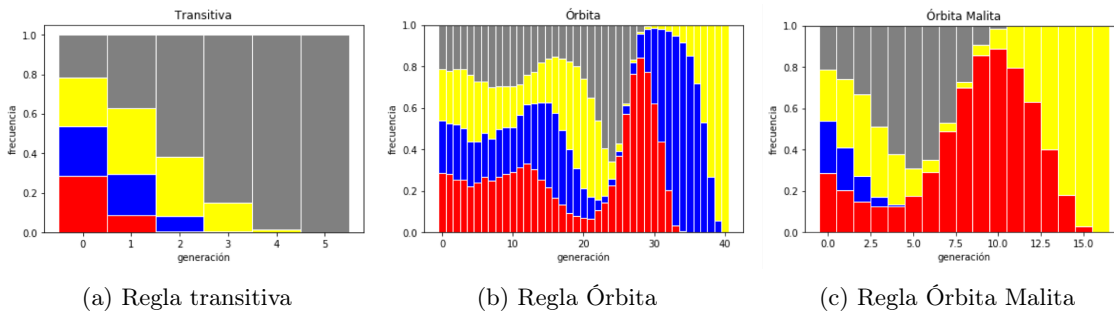


Figura 4.3: $N= 500, k=4, x=1, seed=123$

- En nuestra práctica, en la **regla Transitiva** siempre ganó el tipo “mayor” (el color de hasta arriba en las barras de las gráficas), que por construcción, sabíamos era el más fuerte selectivamente. Las únicas excepciones se deberán a que la probabilidad de interacción selectiva sea pequeña.

Por ejemplo, con $x = .01$ corrimos el programa con $N = 500$, $k = 8$ y ganó el tipo 7, en lugar del 8 en la generación 235. Cabe también mencionar que incluso desapareció primero el tipo 8 que el tipo 5. Pero era de esperarse puesto que la interacción selectiva era muy pequeña.

Aún así con $N = 500$, $k = 8$, $x = 0.05$ (x muy pequeña todavía) gana en la generación 148 el tipo 8, el más fuerte selectivamente.

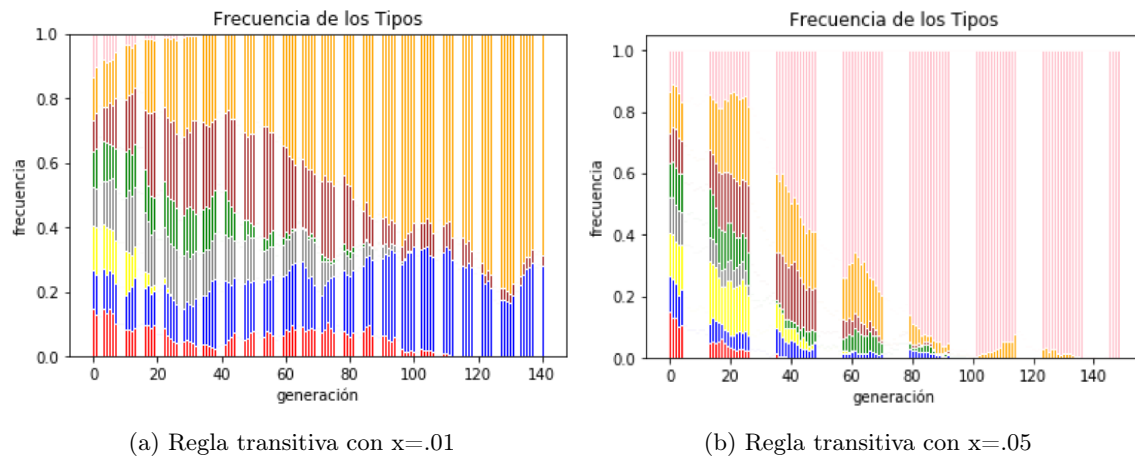


Figura 4.4: Probabilidades pequeñas de observar dos padres

- Me parece destacable el hecho de que en la **regla Transitiva** al aumentar el tamaño de la población, preservando los valores de las demás variables, casi no se altera el número de generaciones antes de alcanzar la fijación del tipo “mayor”. Aún cuando el tamaño de la población se aumente bastante, la fijación ocurre relativamente a las pocas generaciones de iniciado el programa.

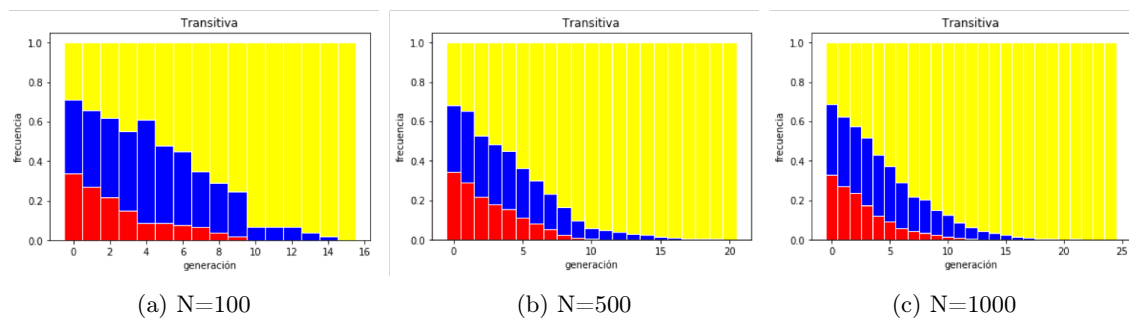


Figura 4.5: Transitivas con $k=3$, $x=.3$, $seed=315$

Esto me lo puedo explicar de la siguiente manera: independientemente del tamaño de la pobla-

ción, la proporción de encuentros a competir entre cada pareja posible de alelos distintos, será aproximadamente la misma siempre y en medida de esto, la evolución será aproximadamente a la misma velocidad. Si como en 4.5 el amarillo le gana siempre a azul y a rojo y, a su vez, el azul al rojo. Esperamos que un tercio de las batallas entre alelos sean de azul contra amarillo; otro tercio, de azul contra rojo y un último tercio, de amarillo contra rojo. Conforme haya más alelos, el alelo más “fuerte” va a ganar en más competencias posibles.

- En la **regla Transitiva**, generalmente, se extinguen en orden del más débil al más fuerte, pero hay veces en que, por cuestiones aleatorias, sobrevive hasta el final uno no tan fuerte. ¡Al final de la gráfica 4.6 se puede ver todavía rastros de nuestro tipo 1!

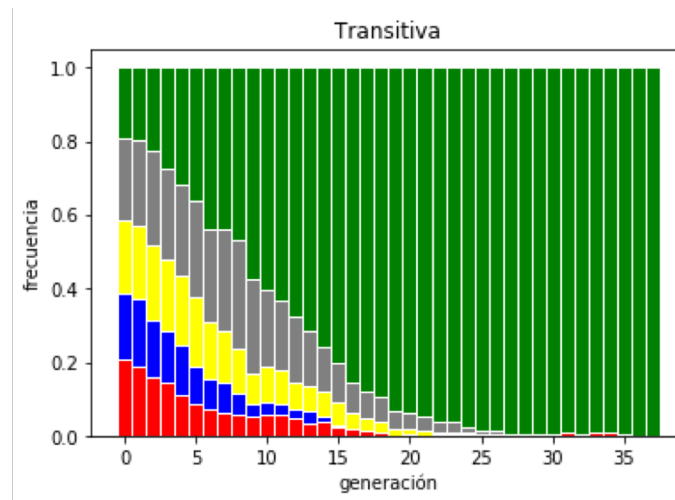


Figura 4.6: $N=500$, $k=5$, $x=0.2$, $seed=789$

- En la regla **Gana el Otro** observamos que si no le ponemos al programa, como condición de paro cierto número de generaciones, entonces el proceso se sigue y se sigue. En cierto intento, llegamos hasta la generación 10,000 antes de detener el programa manualmente, esto sin alcanzar la fijación.

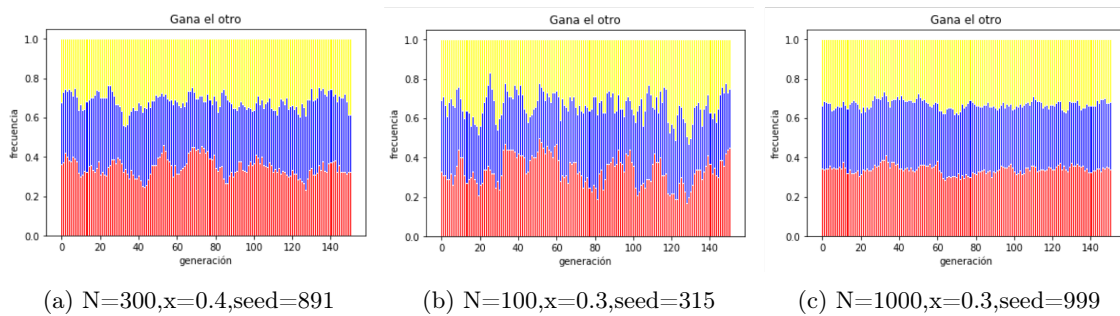


Figura 4.7: Gana el otro

- Entonces, podemos decir que en el caso de **Gana el Otro** el hecho de que x sea positiva significa un incremento en el tiempo hasta la fijación en comparación a cuando no hay *interacción selectiva*. En cambio, en la regla **Transitiva**, **Órbita** y **Órbita Malita**, pareciera que mientras mayor es x , el número de generaciones hasta la fijación decrece.

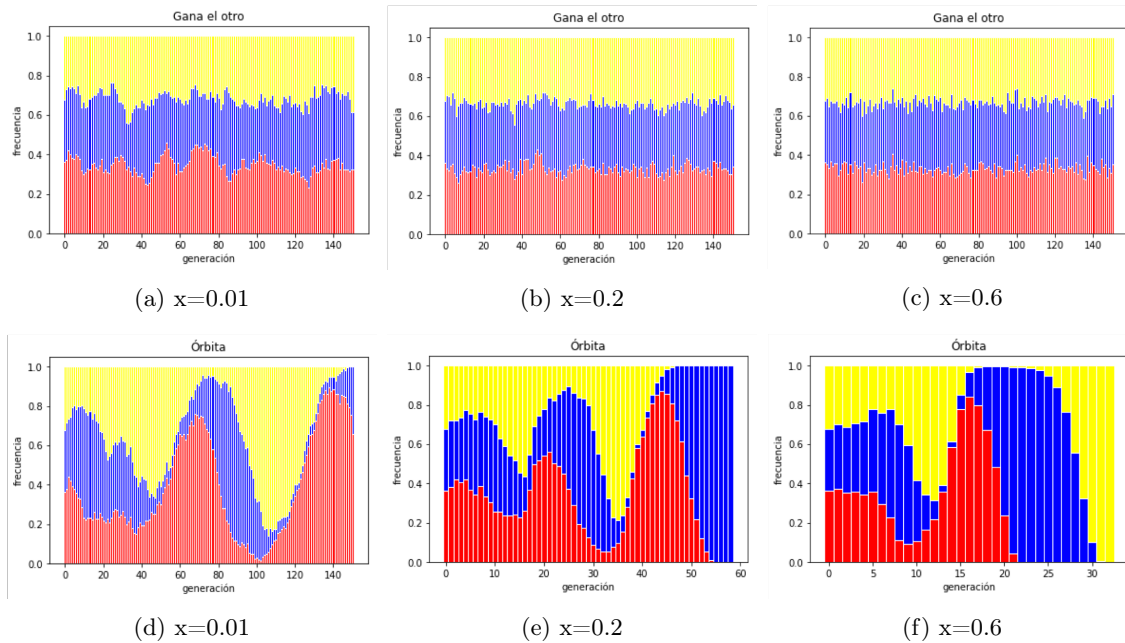


Figura 4.8: $N= 300$, $k=3$, $seed=891$

Aclarando, en la Figura 4.8, en (d), la Órbita cuando $x = 0.01$, no se alcanza la fijación dentro de las primeras 150 generaciones.

- Lo anterior es cierto en lo general, sin embargo, cuando x se encuentra muy cerca del cero y es, por ejemplo, de la forma $1/N$, se observa un comportamiento distinto. En la Tabla 4.1 revisamos el número de generaciones hasta la fijación con distintas semillas aleatorias, pero siempre con $N = 100$ y $x = \{0, 1/100\}$. Con azul se señala las ocasiones en que en la regla de Gana el Otro se registró un descenso en el número de generaciones hasta la fijación al pasar de $x = 0$ a $x = 0.01$. En rojo se señala las ocasiones en que en las demás reglas se registraron incrementos en el número de generaciones hasta la fijación cuando se pasa de $x = 0$ a $x = 0.01$. En otras palabras, estamos haciendo notar aquellos casos en los que no se cumple lo planteado en el punto anterior.

Podemos ver en la Tabla 4.1 que, con excepción de cuando la semilla aleatoria es 666, un muy pequeño cambio en x , se tradujo en una afectación importante en el número de generaciones hasta la fijación. También, llama la atención de que para algunas semillas (sobre todo en $seed = 666$ y $seed = 369$) cuando $x = 0.01$, los números de generaciones hasta la fijación en las cuatro reglas son muy cercanos los unos de los otros.

SEED=666				
Número de generaciones hasta la fijación				
REGLA	k=3		k=4	
	x=0	x=0.01	x=0	x=0.01
Transitiva	230	213	230	213
Gana el otro	230	221	-	-
Órbita	230	221	230	213
Órbita Malita	230	221	230	221

SEED=369				
Número de generaciones hasta la fijación				
REGLA	k=3		k=4	
	x=0	x=0.01	x=0	x=0.01
Transitiva	466	150	116	101
Gana el otro	466	155	-	-
Órbita	466	150	116	271
Órbita Malita	466	150	116	101

SEED=157				
Número de generaciones hasta la fijación				
REGLA	k=3		k=4	
	x=0	x=0.01	x=0	x=0.01
Transitiva	80	175	121	222
Gana el otro	80	233	-	-
Órbita	80	175	121	152
Órbita Malita	80	175	121	175

SEED=789				
Número de generaciones hasta la fijación				
REGLA	k=3		k=4	
	x=0	x=0.01	x=0	x=0.01
Transitiva	237	107	226	92
Gana el otro	237	95	-	-
Órbita	237	253	226	165
Órbita Malita	237	253	226	92

Tabla 4.1: Comparaciones en el tiempo de fijación entre $x = 0$ y $x = 1/N$

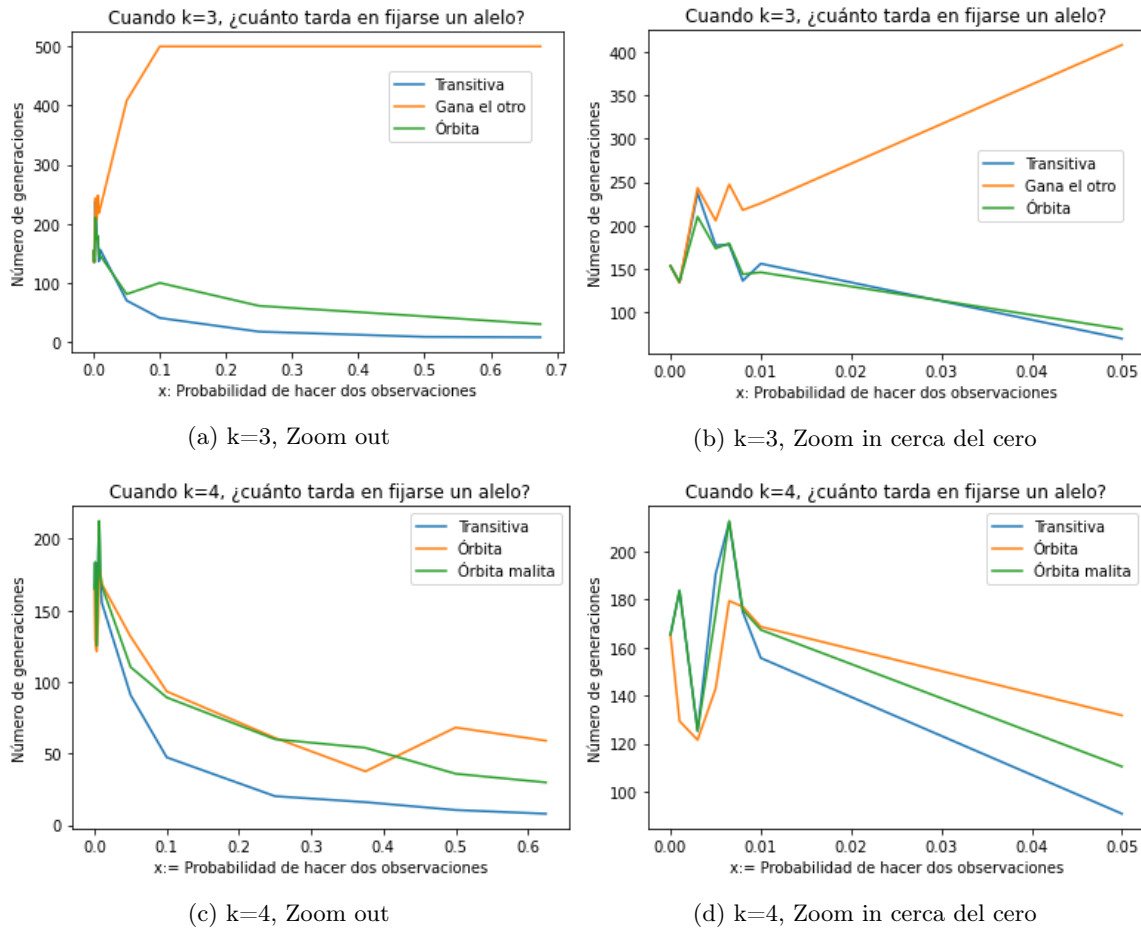


Figura 4.9: Promedios de número de generaciones hasta fijación variando x

Complementario, en la Figura 4.9 se observa la fluctuación de los promedios en el número de las generaciones hasta la fijación obtenidos de 100 procesos con diferentes semillas aleatorias, siempre con un tamaño de población $N = 100$ y para varias x 's, probabilidad de hacer dos observaciones. Esto para cada regla y los casos en que $k = 3$ y $k = 4$.

Se puede notar que cuando x se aproxima al cero, precisamente entre 0 y $\frac{1}{N} = \frac{1}{100} = 0.01$, el comportamiento es errático y distinto a las tendencias generales que posteriormente, para $x > 0.01$, siguen cada una de las reglas.

Todo lo anterior, lo podemos explicar esto diciendo que cuando $x = 0.01$, es muy baja la probabilidad de observar dos individuos, por lo tanto, es igual o aún más baja la probabilidad de que haya interacción selectiva (recordemos que cuando se hacen dos observaciones, éstas pueden ir dirigidas hacia el mismo individuo o ambos padres potenciales pueden ser del mismo tipo alélico, casos en los cuales no habría interacción selectiva).

- En la regla **Gana el Otro** se pueden “resucitar” tipos alélicos que ya estaban extintos. ¿Cómo es esto? Pues para explicarlo supongamos que a ya se extinguió y que siguen en el juego b y c . Entonces la próxima vez que exista una competencia selectiva, sí o sí, ésta debe ser entre b y c . Pero la regla dice, que en este caso, cuando compiten b contra c gana a , de modo que a vuelve a la vida. Así que, tal y como está el programa, la única forma en que se podría fijar uno de los tipos alélicos, sería que en la misma generación se extingan los otros dos alelos.
- Considerando el punto anterior, nos íbamos a dar a la tarea de “corregir” el programa, agregándole la condición de que, en el caso de la regla **Gana el Otro**, si un tipo alélico se extinguiera, éste ya no pudiera “resucitar” y, en caso de que quedaran sólo dos alelos y hubiera competencia, ambos rivales tuvieran $1/2$ de probabilidad de resultar ganadores; sin embargo, al revisar todas las gráficas correspondiente a la regla **Gana el Otro**, nos dimos cuenta de que, en realidad, ningún tipo alélico pareciera correr el riesgo de extinguirse, no, al menos, en las primeras 10,000 generaciones, por lo tanto, la probabilidad de resucitar también parece pequeña.
- Además, en la regla **Gana el Otro**, las frecuencias de los tres alelos se mantienen relativamente “equilibradas” de generación en generación. Esto intuitivamente lo podemos explicar diciendo que es porque la disminución de cierto tipo alélico, implica la alza en la frecuencia de los otros dos, lo cual los hace más propensos a competir y, por ende, al competir estos dos, termina ganando el que estaba a la baja de frecuencia. Cuando le pedimos al programa que la primera generación se cree de forma aleatoria, como mencionamos anteriormente, cada alelo comienza con una frecuencia similar a $1/3$ y esta constitución se conserva bastante semejante por generaciones. Divertido es cuando se le ingresa al programa la composición de la primera generación y deliberadamente se le da un amplísimo margen de ventaja a alguno de los alelos, con una frecuencia mucho mayor al de los demás; aún así prontamente, al cabo de unas cuantas

generaciones, esta disparidad se habrá corregido y tendrán cada uno de los tres alelos una frecuencia parecida a la de los demás.

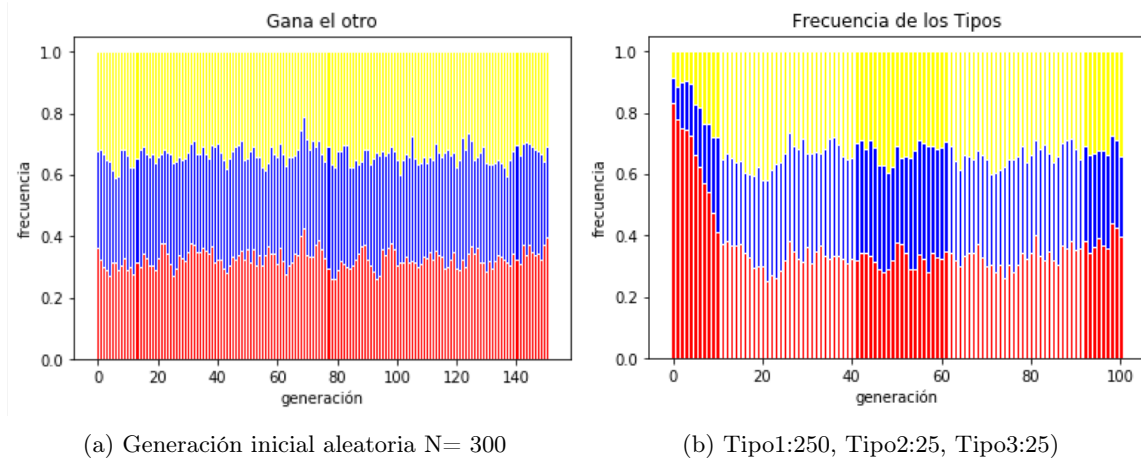


Figura 4.10: Generación inicial aleatoria vs. manual

- Cuando $k = 3$, la **Órbita tradicional** y la **Órbita Malita** son la misma regla. El tipo 1 le gana al 3, 3 al 2 y 2 al 1.
- En las **Órbitas** cuando un tipo se extingue, el tipo con la etiqueta inmediatamente inferior se ve fuertemente favorecido; mientras que el inmediatamente superior se ve en aprietos.

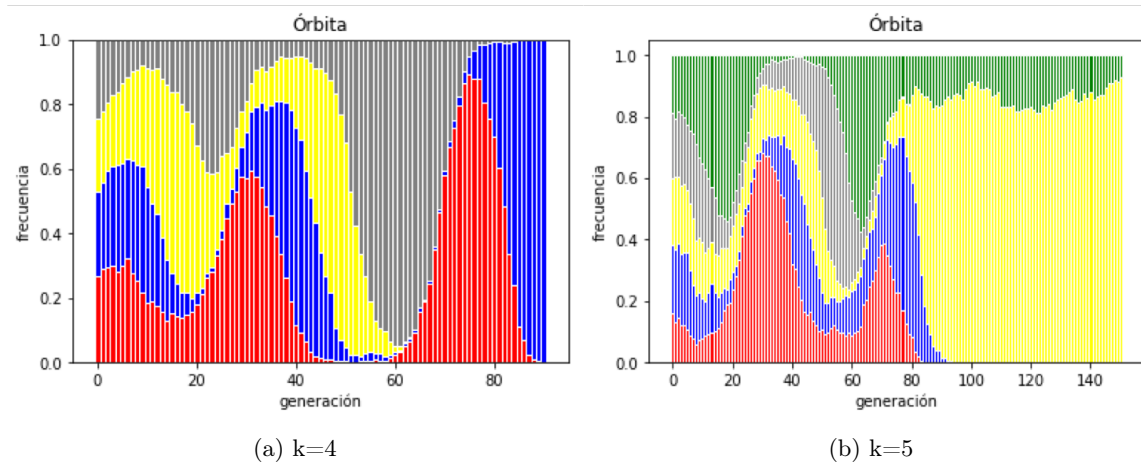


Figura 4.11: N= 500, x=0.5, seed=789

Esto se debe a que nuestra regla, que solía ser *justa*, es decir, ningún tipo alélico tenía ventaja sobre el resto, deja de serlo al momento en que sale del juego uno de los tipos. Entonces nuestra red ecológica se desequilibra y como se puede observar ilustrado en el diagrama 4.12 el alelo con la etiqueta posterior al que desapareció se ve favorecido por este desequilibrio y el anterior desfavorecido.



Figura 4.12: Desequilibrio de la Órbita de 4 alelos después de la primera extinción

- La **Órbita tradicional** y la **Órbita Malita** presentan un comportamiento cíclico muy hermoso, en el que se pueden observar que las gráficas tienen formas de ondas: elevaciones de frecuencias hasta cierto punto sin alcanzar la fijación, seguidas por descensos que no llegan a la extinción y nuevamente ascensos y así hasta que finalmente se alcanza alguna de las condiciones de paro. Pareciera que esta conducta cíclica se acentúa cuando x se acerca a .5 y es menos clara y evidente cuando x se aleja de .5 y está más cercana a 0 o a 1.

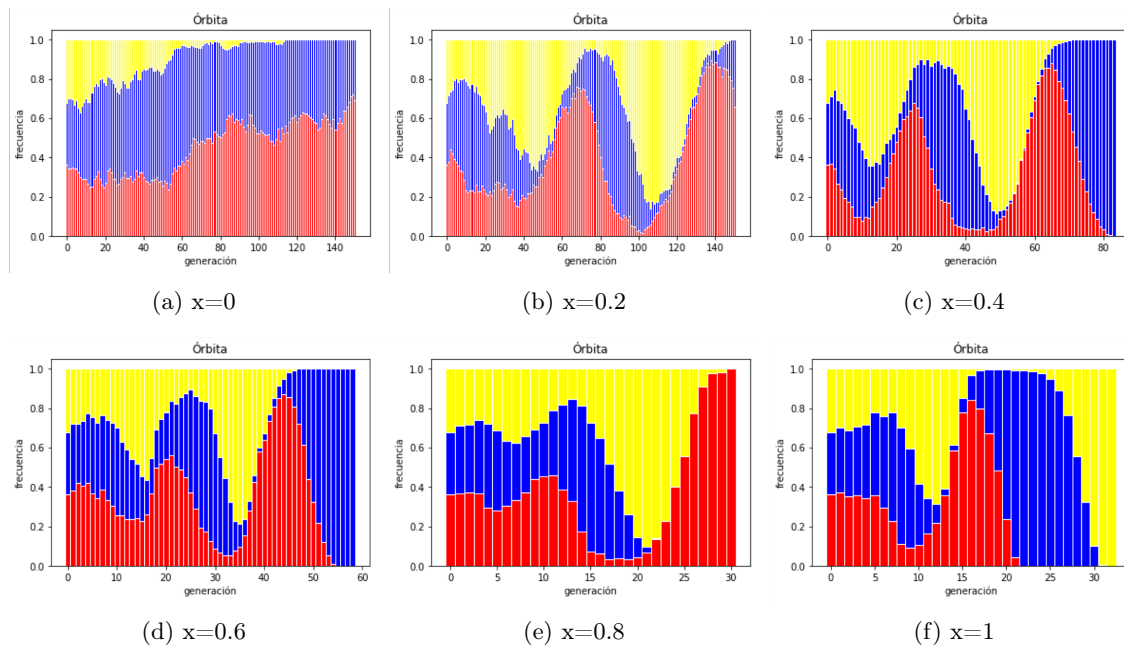


Figura 4.13: $N= 300, k=3, seed=891$

- Es muy interesante de ver el comportamiento que tiene la **Órbita Malita** en la que el tipo 1 al combatir contra el tipo k , gana; pero es derrotado por los otros $k - 2$ tipos alélicos. Si el tipo 1 al principio pelea muchas veces contra el tipo k y sale victorioso en cada combate —como dicta la regla— logrando que el tipo k se extinga; paradójicamente el tipo 1 estará firmando su sentencia de muerte, puesto que es el más débil de todos los demás tipos alélicos, entonces es muy probable (a menos que el azar lo socorra) de que perderá contra los demás alelos que hayan sobrevivido la extinción del tipo con etiqueta k . Curiosamente, para mejorar

su probabilidad de supervivencia, el tipo 1 requiere del tipo k , su única presa posible.

- En **Órbita malita**, cuando $k > 3$, los tipos en $\{2, \dots, k - 2\}$ se conducen como en la regla **Transitiva**, es decir, se extinguen después de muy pocas generaciones y en orden ascendente respecto a sus etiquetas: primero el tipo 2, luego el tipo 3 y así sucesivamente.

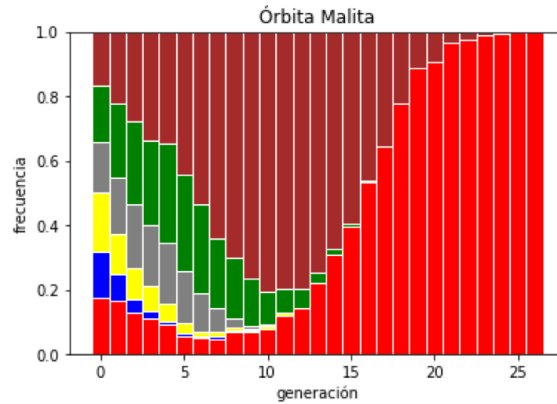


Figura 4.14: $N=500$, $k=6$, $x=0.5$, $\text{seed}=789$

4.2. Estabilidad

Ahora nos daremos a la tarea de revisar nociones de estabilidad para cada una de las reglas de coloración aquí vistas: Transitiva, Gana el otro, Órbita y Órbita malita.

Primero diremos que consideraremos la estabilidad en dos sentidos:

- Uno, revisará el número de generaciones que transcurren antes de que se fije alguno de los tipos alélico. Se compararán resultados con respecto al tamaño de la población, la probabilidad de observar dos padres y el número de alelos. Diremos que, en este sentido, aquella regla que tarde más generaciones en alcanzar la fijación, será la más estable.
- El otro, analizará las varianzas en los procesos de frecuencia alélicos. Es decir para cada alelo en cada interacción selectiva, se verá qué tanto varía su frecuencia desde la generación inicial y hasta el primer evento de extinción. Diremos que aquella regla cuyos alelos sufren más alteraciones (suben y bajan más veces) en su proceso de frecuencia, será menos estable que aquella en la que se mantienen las frecuencias alélicas menos cambiantes durante varias generaciones.

Finalmente, antes de terminar esta sección, inspeccionaremos sobre la varianza que hay entre los tiempos de extinción. Para entender porqué nos puede interesar esto, digamos que no es lo mismo una regla en la que después de extinto un tipo el resto de los alelos también se extinguen rápidamente;

a aquella regla en que las demás extinciones tardarán un tiempo similar o mayor al que tardó en sucederse la primera extinción.

Para conseguir una base de datos lo suficientemente interesante en la cual indagar los puntos de nuestro interés, hicimos que el programa repitiera cincuenta (50) veces el ciclo, cada vez cambiando la semilla aleatoria y para cada una de las reglas de coloración. Pusimos como condición de paro que se llegara a la generación dos mil (2000). Para cada vuelta de 50 ciclos, nosotros determinamos el tamaño de la población (N), la probabilidad de observar dos padres (x) y el número de tipos alélicos (k). Aunque cincuenta nos puede parecer un número no muy atractivo o grande, si consideramos que algunos procesos tardan más de mil generaciones antes de detenerse, algunos incluso llegan a los dos mil, no podíamos esperar que la computadora repitiera el ciclo en más ocasiones pues este proceso llegó a tomar hasta 20 minutos por vuelta. Sin embargo, consideramos que cincuenta es un número no despreciable y quedamos conformes con nuestros resultados.

4.2.1. Estabilidad con respecto al tiempo de fijación

Para revisar la estabilidad con respecto al tiempo de fijación sacamos el promedio y la varianza de los números de generaciones que se tardó la simulación en alcanzar la fijación en cada una de las 50 repeticiones. Esto independientemente del tipo alélico fijado.

Si hablamos de estabilidad con respecto al tiempo de fijación, en este sentido la **regla Transitiva**, a toda costa, es la regla menos estable. Como dijimos anteriormente, aún incrementando el tamaño de la población considerablemente, se alcanza la fijación a las pocas generaciones de comenzado el proceso.

También podemos observar en la tabla 4.2 que conforme aumenta la probabilidad de observar dos padres, x , el promedio de número de generaciones que necesitan transcurrir hasta la fijación, decrece, al igual que la varianza. Veamos que cuando: la población tiene un tamaño de 1200 individuos, hay 3 tipos alélicos y todos los individuos hacen dos observaciones (i.e., $x = 1$), con una varianza bajísima de 0.01 se alcanza la fijación después de un promedio de 4.8 generaciones. Nosotros hemos visto ya en las gráficas y lo podíamos suponer desde antes, que generalmente el tipo alélico que se fija es el tipo con el índice mayor, podemos así deducir que con probabilidad muy alta el tipo alélico que se fija después de menos de 5 generaciones es aquel con el índice mayor. No es cualquier cosa este resultado.

Asimismo, podemos notar que en la regla Transitiva el efecto que tiene el aumento de número de tipos alélicos sobre el promedio de generaciones hasta la fijación es bajísimo. Para $N = 1200$ y $x = 0.3$, podemos ver que cuando eran 3 tipos alélicos el proceso hasta que se fijara uno de ellos se tardó 23.5 generaciones en promedio, mientras que cuando fueron 6 tipos alélicos tomó 26.6 en promedio. Apenas 3 generaciones más correspondientes a los 3 tipos adicionales.

Tampoco, cuando hablamos de la regla Transitiva, el tamaño de la población parece ser un factor

REGLA TRANSITIVA				
Generaciones hasta la fijación				
N	x	k	Promedio	Varianza
200	0	3	361.4	22,928.52
		4	300.8	36,738.82
	0.3	3	17.3	4.81
		4	20.7	19.48
500	0	3	825.1	202,327.79
		4	851.1	260,387.71
	0.3	3	19.4	9.88
		4	21.7	4.21
800	0	3	1,384.8	378,279.31
		4	1,387.6	466,531.75
	0.3	3	22.2	18.58
		4	23.0	11.57
1200	0	3	1,544.9	213,445.39
		4	1,462.1	294,287.25
	0.1	3	70.3	122.83
		3	23.5	20.66
	0.3	4	25.3	31.94
		5	26.5	11.89
		6	26.6	16.25
1	3	4.8	0.01	

Tabla 4.2: Generaciones hasta la fijación. Transitiva.

que cambie el rumbo de la fijación significativamente. Sean $k = 4$ tipos alélicos y $x = 0.3$, entonces el promedio de generaciones que deben transcurrir para alcanzar la fijación en una población de 200 individuos es de 20.7, mientras que cuando se trata de una población de 1200 individuos (una población 6 veces más grande) es de 25.3.

Para la regla de **Gana el Otro** sería absurdo poner una tabla como lo hicimos para la Transitiva o haremos para **Órbita** y **Órbita Malita**, porque para casi todos los casos vistos (exceptuando aquellos de la Tabla 4.1 en que x era de la forma $1/N$) el programa nos arrojó que después de dos mil generaciones no se había alcanzado la fijación de ningún tipo. Así pues, podemos concluir que en este sentido ésta es la regla más estable.

Ahora revisemos la regla **Órbita**. Igual que en la Transitiva, a mayor probabilidad de hacer dos observaciones, el número de generaciones hasta la fijación disminuye muchísimo. Veamos que para $N = 1200$ y $k = 3$ pasa de un promedio de 537.1 generaciones hasta la fijación cuando la probabilidad de hacer dos observaciones es $x = 0.1$, a tan sólo 34.8 generaciones cuando $x = 1$. Además la varianza de este promedio decrece drásticamente mientras que x se eleva; como podemos observar en la tabla 4.3.

Por otro lado, a diferencia de la regla Transitiva, en la **Órbita**, a mayor número de tipos el número promedio de generaciones hasta la fijación sí aumenta bastante. Aunque la varianza de estos también se eleva muchísimo a cada alelo que se agrega.

REGLA ÓRBITA				
Generaciones hasta la fijación				
N	x	k	Promedio	Varianza
200	0	3	361.4	22,928.52
		4	300.8	36,738.82
	0.3	3	92.8	1,038.78
		4	119.9	11,847.78
500	0	3	825.1	202,327.79
		4	851.1	260,387.71
	0.3	3	135.4	2,084.32
		4	201.8	13,285.46
800	0	3	1,384.8	378,279.31
		4	1,387.6	466,531.75
	0.3	3	163.9	2,598.02
		4	325.9	55,587.26
1200	0	3	1,544.9	213,445.39
		4	1,462.1	294,287.25
	0.1	3	537.1	31,630.59
		3	197.1	2,011.93
	0.3	4	354.9	64,864.29
		5	679.0	102,992.37
		6	1098.7	346,450.64
	1	3	34.8	43.27

Tabla 4.3: Generaciones hasta la fijación. Órbita.

También, distinto a como se comporta la regla Transitiva, en la regla de Órbita el aumento en el tamaño de la población, N , se ve reflejado en un considerable incremento en el promedio de generaciones que se requieren para alcanzar la fijación; sin embargo, este incremento en el número de generaciones es mucho menor en relación a la proporción con la que creció la población.

Revisemos por último el caso de la regla **Órbita Malita**. Primero, como ya lo sabíamos, nótese que la regla Órbita y la Órbita Malita son exactamente iguales cuando $k = 3$. Después de hacer esta breve aclaración continuemos revisando características propias de la Órbita Malita. Por ejemplo, ésta regla, en modo totalmente opuesto a como funcionan las reglas Transitiva y Órbita —observemos en la tabla 4.4— a mayor número de tipos alélicos, menor es el promedio de las generaciones hasta la fijación. También el número de generaciones hasta la fijación presenta una varianza altísima cuando menos alelos hay y ésta se va reduciendo conforme aumentan los números de alelos. Esto es muy curioso y particular de esta interacción.

Para $k = 3$, idéntico que en la regla **Órbita**, a mayor N mayor es el número de generaciones promedio hasta la fijación. Pero, para $k = 4$ no se ve que el promedio de generaciones hasta la fijación tenga un patrón claro con respecto al tamaño de la población.

Observemos por último que en Órbita Malita al igual que en Órbita, cuando x es pequeña, la varianza en el número de generaciones hasta la fijación es enorme.

Para concluir con esta noción de estabilidad, mostramos en la Figura 4.15 gráficas que conjugan

REGLA ÓRBITA MALITA				
Generaciones hasta la fijación				
N	x	k	Promedio	Varianza
200	0	3	361.4	22,928.52
		4	300.8	36,738.82
	0.3	3	92.8	1,038.79
		4	75.6	757.39
500	0	3	825.1	202,327.79
		4	851.1	260,387.71
	0.3	3	135.4	2,084.32
		4	151.1	2,345.73
800	0	3	1,384.8	378,279.31
		4	1,387.6	466,531.75
	0.3	3	163.9	2,598.02
		4	135.1	3,338.34
1200	0	3	1,544.9	213,445.39
		4	1,462.1	294,287.25
	0.1	3	537.1	31,630.59
		3	197.1	2,011.93
	0.3	4	169.0	2,054.47
		5	122.1	1,330.26
		6	101.5	849.48
	1	3	34.8	43.27

Tabla 4.4: Generaciones hasta la fijación. Órbita Malita.

la información en las tablas 4.2, 4.3 y 4.4 de las tres reglas: Transitiva, Órbita y Órbita Malita para así poder compararlas.

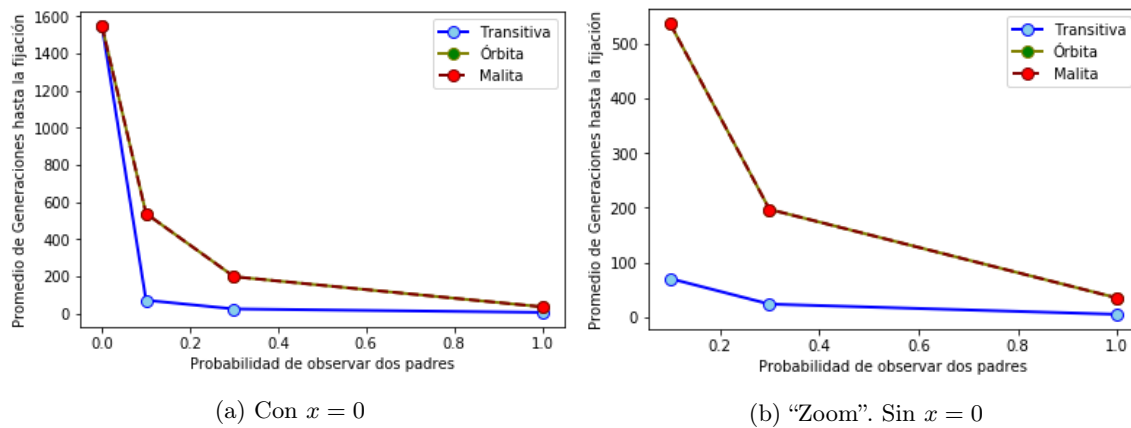


Figura 4.15: $N = 1200$, $k = 3$

La primera gráfica es el promedio de generaciones hasta la fijación de las cincuenta simulaciones con un tamaño de población de $N = 1200$ y tres tipos alélicos, $k = 3$, mientras que se varía la probabilidad de hacer dos observaciones, $x = \{0, 0.1, 0.3, 1\}$.

En las gráficas anteriores se traslapan la Órbita Malita y la Órbita por tratarse del caso en que $k = 3$. Podemos ver como la caída de generaciones hasta la fijación es mucho más abrupta en la

regla Transitiva que en las otras dos reglas. También podemos ver en las tablas 4.2, 4.3 y 4.4 que cuando $x = 0$, no sólo el promedio de generaciones hasta la fijación es mucho más alto, también la varianza es gigante.

Ahora, veamos una gráfica que nos muestre como para cada una de las reglas se ve afectado el promedio de generaciones hasta la fijación por el cambio en el número de tipos alélicos. Para esto fijamos a $N = 1200$ y $x = 0.3$.

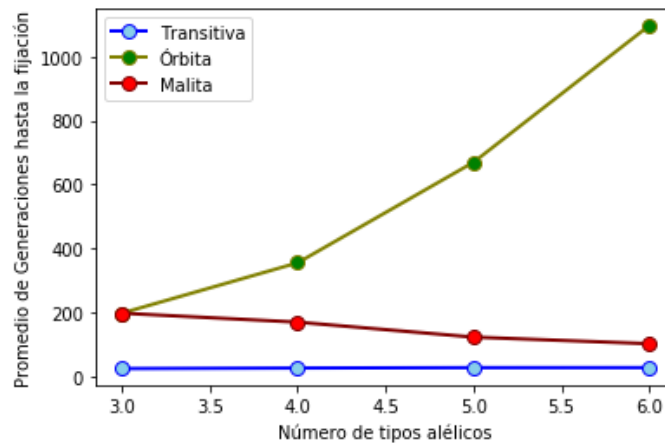


Figura 4.16: $N=1200$, $x=0.3$

¡Qué locura! Para la regla Transitiva, podemos ver como básicamente el número de tipos alélicos no afecta el promedio de generaciones hasta la fijación, mientras que para la Órbita, el aumento en el número de tipos alélicos representa también un aumento en este promedio y en la Órbita Malita a mayor tipos alélicos se observa un descenso en el promedio. Podemos también observar que la velocidad en que sube la línea de la Órbita es mucho mayor que con la que desciende la línea de Órbita Malita.

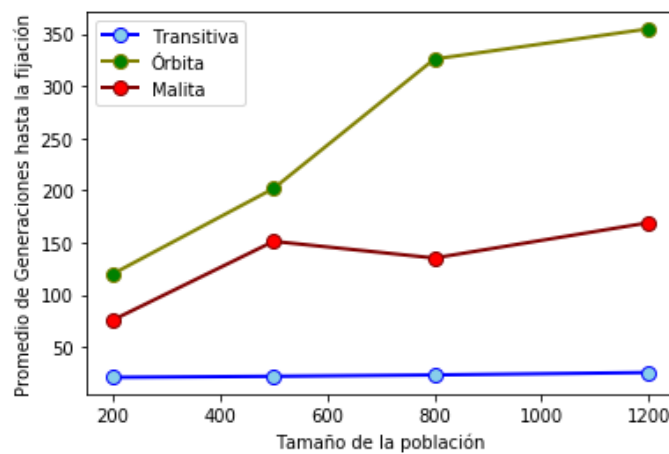


Figura 4.17: $k=4$, $x=0.3$

Finalmente, observemos la Figura 4.17 en la que se observa cómo influye el tamaño de la población en el número de generaciones que se requieren para alcanzar la fijación para cada una de las tres reglas. Ahora fijamos a $k = 4$ (para que se puede diferenciar la Órbita de la Órbita Malita) y a $x = 0.3$.

En la gráfica no se puede apreciar muy bien por la escala, pero la regla Transitiva sí presenta una elevación en el número de generaciones que se requieren para llegar a la fijación conforme va creciendo el tamaño de la población, sin embargo ésta elevación es mucho más agresiva en las otras dos reglas y por lo tanto más visible.

No obstante, no encuentro la razón por la que el promedio de generaciones hasta la fijación en la regla Malita decrece cuando $N = 800$, pero podría yo suponer que quizá si siguiéramos creciendo la población, el número de generación hasta la fijación se estancaría un poco; habría que ver.

4.2.2. Estabilidad con respecto a la varianza de las frecuencias alélicas

Esta medida revisa la varianza en las frecuencias que registró cada tipo alélico desde la generación inicial y hasta que ocurre la primera extinción. Todas las pruebas se hicieron con $N = 1200$, siendo esta una cifra tomada de manera totalmente arbitraria.

Además de ver la varianza, me interesó, pedirle al programa que me arrojara cuál era el promedio de estas frecuencias; así junto con la varianza, se puede ver la tendencia de cada tipo alélico según las condiciones que se preestablecen, es decir: la probabilidad de hacer dos observaciones, x , y el número de alelos, k . Igual que para el ejercicio anterior, para éste se hicieron 50 repeticiones, cada una cambiando la semilla aleatoria, y se utilizaron como condiciones de paro el primer evento de extinción o llegar a la generación dos mil (2,000).

Como las varianzas de las frecuencias son cuadrados de números menores a uno, resultan ser números muy pequeños. Por eso me permití multiplicarlos por 100, para que fueran más claros de entender.

En la tabla 4.5 no está la regla **Gana el otro**; esto se debe a que sólo tiene sentido hablar de esta regla cuando $k = 3$ y ya de por sí es una tabla muy grande. Pero lo que se observó en las primeras dos mil generaciones en la regla Gana el Otro, fue que para los tres tipos alélicos, independientemente de x , el promedio de las frecuencias fueron muy cercanas a un tercio y conforme x crecía, la varianza de estas frecuencias se desplomó. Cuando $x = 0.1$, esta varianza multiplicada por 100 (como se hizo con el resto de las varianzas), era 0.095; cuando $x = 0.3$ la varianza por 100 era 0.037 y cuando $x = 1$, la varianza multiplicada por 100 en la frecuencia de los tres alelos era 0.020.

Dejando la regla Gana el Otro a un lado, entremos en materia y observemos en la tabla 4.5 que para todos los casos de la regla **Transitiva** se ve que el tipo más pequeño (el que tiene el índice 1) es el de menor frecuencia y que el promedio de esta frecuencia va aumentando conforme el índice del tipo aumenta, siendo el tipo con el índice más alto el de mayor frecuencia, esto con un margen

importante; sin embargo, también es éste el alelo con mayor varianza.

En la regla **Órbita** los promedios de frecuencia para cada tipo se acercan mucho a $1/k$, independientemente de la probabilidad de hacer dos observaciones, x , y del número de alelos, k . Entre más alelos están compitiendo, es decir, cuando k es mayor, la varianza de todos los tipos es menor. De modo que, cuando $k = 3$ se presencia la mayor varianza en la frecuencia de los tipos alélicos.

Frecuencias alélicas de cada tipo								
			Regla Transitiva		Regla Órbita		Regla Malita	
k	x	Tipo	Promedio	Varianza*	Promedio	Varianza*	Promedio	Varianza*
3	0.1	1	0.086	0.89	0.340	4.67	0.340	4.67
		2	0.142	1.35	0.331	4.53	0.331	4.53
		3	0.772	4.25	0.329	4.43	0.329	4.43
	0.3	1	0.078	1.03	0.333	4.90	0.333	4.90
		2	0.139	1.52	0.338	5.12	0.338	5.12
		3	0.783	4.82	0.329	4.95	0.329	4.95
	1	1	0.110	2.33	0.319	4.05	0.319	4.05
		2	0.211	2.11	0.325	4.08	0.325	4.08
		3	0.679	7.74	0.356	4.93	0.356	4.93
4	0.1	1	0.069	0.50	0.237	2.03	0.312	2.42
		2	0.097	0.61	0.269	2.41	0.067	0.46
		3	0.162	0.77	0.232	1.92	0.276	2.10
		4	0.672	5.09	0.263	2.30	0.344	2.91
	0.3	1	0.058	0.58	0.247	2.64	0.310	3.03
		2	0.089	0.73	0.270	3.64	0.054	0.46
		3	0.153	0.99	0.233	2.67	0.295	2.91
		4	0.699	6.19	0.250	2.84	0.341	2.64
	1	1	0.080	1.40	0.248	3.10	0.333	5.57
		2	0.127	1.29	0.253	2.96	0.055	0.70
		3	0.228	0.83	0.251	3.14	0.290	5.82
		4	0.566	7.97	0.249	2.85	0.322	4.60
5	0.1	1	0.056	0.32	0.202	1.44	0.179	0.61
		2	0.069	0.39	0.209	1.53	0.061	0.35
		3	0.097	0.46	0.188	1.48	0.094	0.43
		4	0.172	0.53	0.205	1.45	0.218	0.54
		5	0.605	5.72	0.196	1.58	0.448	2.24
	0.3	1	0.049	0.38	0.200	2.17	0.219	2.36
		2	0.065	0.44	0.194	1.75	0.047	0.35
		3	0.099	0.56	0.198	1.93	0.081	0.49
		4	0.165	0.64	0.194	1.84	0.219	1.65
		5	0.623	6.80	0.214	2.42	0.434	3.64
	1	1	0.062	0.91	0.206	2.69	0.116	0.30
		2	0.086	0.81	0.198	2.11	0.065	0.66
		3	0.132	0.77	0.190	1.81	0.106	0.81
		4	0.236	0.46	0.208	2.50	0.229	0.96
		5	0.485	7.39	0.198	2.28	0.485	5.94

*Varianzas multiplicadas por 100

Tabla 4.5: Frecuencia de los tipos alélicos. Transitiva.

En la **Órbita**, una vez determinadas las condiciones al principio, es decir N , k y x , se observa que todos los alelos presentan una varianza similar entre ellos. Que es justo lo que uno puede apreciar

en las gráficas: todos los alelos en algún momento tienen una crecida y después una disminución, al igual que el resto de sus compañeros.

Por último, en la regla de **Órbita Malita**, con excepción del caso en que $k = 3$, en el que ya sabemos que se comporta igual que **Órbita**, el tipo 2 se encuentra en una franca desventaja, puesto que tiene un promedio de frecuencia bajísimo y una varianza también muy baja que le da muy pocas oportunidades de repuntar.

A diferencia de la regla **Órbita**, en la regla **Malita** los tipos no tienen una varianza similar entre sí. Por el contrario, el tipo con el índice más alto presenta una varianza muy alta, seguido por el tipo con el índice 1 o el del índice $k - 1$, pero el resto de los alelos presentan poca varianza. Similar a lo que sucede en el caso de la regla **Transitiva**.

Podemos decir, entonces, que nuevamente en este sentido de estabilidad, tal y como en el sentido anterior, la regla **Gana el Otro** es la más estable, puesto que sus tres alelos presentan la menor varianza en sus procesos de frecuencia. Sin embargo, decir en qué orden de estabilidad se encuentran el resto de las reglas con respecto a la varianza en sus procesos de frecuencia, me parece un poco complicado.

4.2.3. Varianza en los tiempos de extinción.

En esta parte revisaremos qué tan espaciadas o, por el contrario, consecutivas ocurren las extinciones de los tipos alélicos. Para esto también utilizamos poblaciones de 1,200 individuos, $N = 1200$, y nuevamente repetimos los procesos en ciclos de 50 simulaciones en las que en cada una va variando la semilla aleatoria. Manualmente cambiamos el tamaño del espacio alélico, k , y la probabilidad de hacer dos observaciones, x , para poder hacer comparaciones entre las cifras resultantes. Para este ejercicio la única condición de paro que utilizamos en el programa fue la de que se fijara un tipo alélico. No incluimos la regla **Gana el Otro** porque ya a estas alturas sabemos que no tiene mucho sentido hacerlo.

En la tabla 4.6 es muy sencillo de ver algo que ya esperábamos: en la regla **Transitiva** esta varianza es pequeña para cualquier número de alelos, sobre todo cuando $x = 1$. Esto quiere decir que una vez que se da la extinción de un alelo, las demás extinciones no se harán esperar.

Opuesto a lo anterior, la regla **Órbita** presenta la mayor varianza entre los tiempos de extinción, situación que se acentúa conforme se van agregando tipos alélicos a la regla.

La regla **Órbita Malita** sí presenta una varianza alta, pero no tanto como la de **Órbita**, esto se lo atribuimos a la semejanza que tiene el comportamiento de los alelos intermedios al de los alelos en una regla **Transitiva**.

En cualquier caso, para cualquier regla, el incremento en la probabilidad de hacer dos observaciones en lugar de sólo una, x , implica un descenso en la varianza entre los tiempos de extinción.

Varianza entre los tiempos de extinción			
k	Regla	x=0.3	x=1
3	Transitiva	12.10	1.94
	Órbita	438.98	24.73
	Malita	438.98	24.73
4	Transitiva	32.69	1.03
	Órbita	3,254.13	48.48
	Malita	3.095.29	39.64
5	Transitiva	28.62	2.51
	Órbita	35,123.81	17,123.83
	Malita	4,204.66.	51.70
6	Transitiva	39.47	2.41
	Órbita	111,500.15	137,271.69
	Malita	1,558.29.	57.75

Tabla 4.6: Varianza entre los tiempos de extinción.

4.3. Otros asuntos

Los aspectos a estudiar y las ramificaciones que puede tomar el estudio de esta materia son interminables. Podríamos complejizar las reglas de coloración y hacerlas más cercanas a lo que se observan en la naturaleza; también podríamos cambiar el tamaño de la población o hacer más número de repeticiones en cada vuelta para ampliar nuestra exactitud y nuestra base de datos; podríamos estudiar a fondo el fenómeno de porqué se ven más claramente las ondas en las gráficas de la regla Órbita cuando nos acercamos al 0.5; podríamos también incluir la opción de que mutaran, etc, etc. Como dije, las opciones son bastantes y variadas, pero la pretensión de esta tesis era obtener resultados interesantes arrancando de un modelo medianamente sencillo y simplificado.

Únicamente a modo de tentempié, en la siguiente sección planteo los problemas con los que se puede enfrentar aquel que intente generalizar el modelo para que los individuos puedan hacer cualquier número de observaciones antes de elegir a su padre real; también propongo algunas soluciones a estos conflictos.

4.3.1. Más de dos observaciones

Mientras que sigamos en el entendido de que en las interacciones selectivas no afectará la frecuencia con que se observa uno u otro padre potencial, sino sólo la "posición" que cada uno de estos tenga dentro de la red ecológica; para poder generalizar y darle al programa la oportunidad de hacer cualquier número de observaciones, nos tendremos que enfrentar, para ciertas reglas de coloración, a definir una serie de estatutos que el programa deberá de tomar de forma que pueda elegir al padre real de entre más de dos padres potenciales. Veamos qué sucede con cada una de las reglas de coloración al darle a cada individuo la oportunidad de hacer más de dos observaciones en la generación anterior y, por ende, la posibilidad de tener tres o más padres potenciales.

- En el caso Transitivo sería sencillo, puesto que los individuos en la generación g sólo tendrán que adoptar el tipo alélico del padre potencial con mayor denominación. Esto no implicaría mayor reto, ni confusión.
- En la regla de Gana el Otro, debido a que está limitado a un máximo de tres alelos, el hecho de que se pudieran hacer más de dos observaciones se reduciría a estas tres circunstancias:
 1. Que únicamente haya un padre potencial y por lo tanto éste será el padre real.
 2. Que haya dos padres potenciales y entonces gana el tercero que no está en discordia, tal y como dicta la regla original.
 3. Que los tres alelos sean padres potenciales, que se traduciría en anular la competencia y elegir al padre real de forma uniformemente aleatoria entre los tres.
- Es cuando entramos a las reglas de Órbita y Órbita Malita y que tenemos más de tres alelos —cuando se tienen tres alelos y tres padres potenciales, se anula la competitividad— que la cosa se puede complicar un poco más. En primer lugar porque habría que definir en qué orden se harán cada una de las competencias de dos en dos. Un orden distinto puede implicar un resultado distinto. Por ejemplo, si usamos la regla Órbita y tenemos cinco tipos alélico y resulta que cierto individuo v en la generación g ha observado a los siguientes padres potenciales en la generación $g - 1$: $z_v = \{1, 2, 5\} \in H$. Si compiten en este orden: $(5,1)$, $(1,2)$, $(2,5)$ entonces la final sería entre 2 y 5, cada uno con $1/2$ de probabilidad de resultar ganadores.

Mientras que si el orden fuera $(2,5)$, $(1,2)$, $(5,1)$ primero tendríamos que saber quién ganará entre 2 y 5 para saber quienes seguirán en la contienda y quienes no. Si entre 2 y 5 gana 2 pues tendremos que los padres potenciales se habrán reducido en la primera ronda a únicamente $\{1, 2\}$ y ya sabemos quién gana en este caso: $\{2\}$; pero si ganase el 5, entonces tendríamos cíclicamente a los mismos contendientes compitiendo y así hasta que suceda que gana 2.

Además es posible que haya más de dos competencias interorbitales, las cuales se deciden de manera aleatoria, cada uno de los competidores tiene 0.5 de probabilidad de resultar ganador. El orden en que se revisen estas competencias y los ganadores en dichas contiendas determinará el resultado final.

4.4. Conclusiones biológicas

Aunque no tengo formación académica en Biología, por ser ésta una tesis con un trasfondo de biología y genética, haré un esfuerzo por hacer algunas suposiciones sobre como los resultados aquí vistos se pueden traducir y aplicar para poblaciones de la naturaleza.

Por un lado, la experiencia nos dicta que las redes ecológicas sí existen —hay competencias no sólo entre morfismos alélicos, sino también entre distintas especies— pero aquí hemos visto que si la

probabilidad con que se hacen estas competencias es medianamente alta, implica una aceleración en la fijación de alguno de los tipos; sin embargo, las poblaciones en la naturaleza presentan diversificación y polimorfismos que perduran por varias generaciones. Por lo anterior, me inclino a pensar que la probabilidad de estas competencias es de orden $(1/N)$. También la biodiversidad se debe a constantes mutaciones genéticas a partir de las cuales surgen más alelos y con ellos, otras interacciones selectivas.

Asimismo, podría uno sospechar que una población obedece sistemas de redes ecológicas más complejos que los aquí simulados; es decir, con redes ecológicas en las que hay distintas reglas de competencia dentro de sí, que se entrelazan con otras redes y sistemas y que se transforman. Por ejemplo, algunos tipos alélicos se pueden relacionar entre sí con una Órbita y otros con una regla Transitiva, etc, etc. Las alternativas son muchas y variadas.

Por último quisiera agregar que aunque en general la gente tiende a pensar que la regla de Gana el Otro no es verosímil al hablar de ejemplos de la naturaleza, a mí me parece coherente considerar que si dos tipos están continuamente combatiendo entre sí, un tercer elemento se ve beneficiado por estos conflictos al encontrarse desvinculado de ellos.

Anexo A

Manual de usuario

Utilizar este programa es muy sencillo.

Requerimientos: Contar con una consola de IPython (yo utilicé Spyder Website para construir y correr el programa, la cual puedes bajar gratuitamente).

Pasos:

1. Descargar el archivo de esta tesis: `Codigo.py`
2. Abrir tu consola y correr el archivo `Codigo.py`
3. Elegir qué semilla aleatoria quieres utilizar. Puedes ingresar cualquier número natural, incluido el cero, que quieras. Presiona *Enter*.
4. Elige cuántos tipos alélicos tiene tu población. Puedes ingresar cualquier número natural mayor a uno. Sólo OJO: si pones más de diez, el programa no sabrá graficarlo. Presiona *Enter*.
5. Elige con qué probabilidad se harán dos observaciones. Recuerda que las probabilidades deben ser valores en $[0,1]$. Si te sales del rango te marcará error. Presiona *Enter*.
6. Ingresar el número que antecede en la siguiente lista a la regla que quieres utilizar:
 - a) Transitiva
 - b) Gana el Otro
 - c) Órbita
 - d) Órbita Malita
 - e) Personalizada Si pones un valor distinto a estos te marcará error. Si tu número de alelos es distinto a tres y eliges la regla Gana el Otro te marcará error.
Presiona *Enter*.

7. Si elegiste la regla Personalizada tendrás que proporcionarle al programa las probabilidades con que ganaría cada alelo en caso de cada una de las competencias posibles. Recuerda que las probabilidades deben ser valores en $[0,1]$. Si te sales del rango te marcará error. Recuerda que dada cada competencia posible, la suma de las probabilidades debe dar 1, si no marcará error. Presiona *Enter* después de cada vez que termines de digitar una probabilidad.
8. Si quieres que la primera generación se genere (valga la redundancia) de manera aleatoria, presiona el número 1 y después *Enter*. Si quieres ingresar manualmente cuántos individuos hay de cada tipo en la primera generación, presiona el número 2 y después *Enter*.
9. Si elegiste crear la primera generación de forma aleatoria, dile al programa de qué tamaño es tu población. Puedes ingresar cualquier número natural, excluido el cero. Después presiona la tecla *Enter*.
10. Si elegiste ingresar manualmente cuántos individuos hay de cada tipo en la primera generación, di cuántos individuos hay para cada tipo y después presiona *Enter*.
11. ¡Obtén tu gráfica! Recuerda que el tipo con la etiqueta más baja es la que se muestra más pegada al eje de las equis y la de la etiqueta más alta, se encuentra hasta arriba de los demás tipos.

A.0.1. User manual (In English)

It's very easy to use this program.

Requirements: A Python console (I used Spyder Website, both to build and to run the program. You can download it for free).

Steps to follow:

1. Download the folder in this thesis with the name `Codigo.py`
2. Open the Python console and run `Codigo.py`
3. Choose your random seed. You can use any natural number, including zero. Then press *Enter*.
4. Choose the number of allelic types for your population. It can be any natural number greater than one. Just BE CAREFUL: if you choose a number bigger than ten, the program will not know how to make the graphics. Press *Enter*.
5. Choose the probability with which an individual will make two observations instead of just one. Remember the probabilities must be in $[0,1]$. If you choose a number out of this range the program will mark an error. Press *Enter*.
6. Enter the number of the rule you would like to use:

- a) Transitive
- b) The Other One Wins
- c) Orbit
- d) Clumsy Orbit
- e) Personalized

If you enter a number different from those above, the program will mark an error. Also, if your number of alleles is not three and you choose The Other One Wins rule, the program will mark an error, you must have three alleles so The Other One Wins rule works. Press *Enter*.

7. If you choose the Personalized rule, you'll need to capture the probabilities with which each allele wins in each of the possible scenarios. Remember, because they are probabilities, these values must be in $[0,1]$, if you use a value out of this range, the program will mark an error. Remember that, in each possible competition, the sum of the probabilities must be one, if not it will mark an error. After entering each probability Press *Enter*.
8. If you want the first generation to be generated (forgive the redundancy) randomly, press number 1 and then *Enter*. Otherwise, if you want to enter manually the number of individuals that has each allele in the first generation, press number 2 and then *Enter*.
9. If you chose to randomly create the first generation, now you must tell the program what is the size of your population. You can enter any natural number, zero excluded. Afterwards, press *Enter*.
10. If you chose to enter manually the number of individuals of each allele in the the first generations, you must enter how many individuals there are of each allele, and then press *Enter*.
11. ¡Here is your graphic! Remember that the allele with the lowest label is the one presented closest to the X-axis, and you will find the one with the highest label above the other alleles, on the top.

Anexo B

Código del programa en Python

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Mon Mar  2 16:11:21 2020

@author: sofiarozanes
"""
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

def primera_aleatoria(M,h,list):
    print("\n.....")
    print("GENERACION 0")
    for n in range(M):
        n=np.random.randint(1,h+1)
        list.append(n)
    print("La generación cero es:")
    print(list)

def cuentas(list,h):
    cuenta=[]
    extintos=[]
    cuenta_ceros=0
    for a in range(1,h+1):
        cuenta_tipos=list.count(a)
        cuenta.append(cuenta_tipos)
        if cuenta_tipos==0:
            extintos.append(a)
            cuenta_ceros=cuenta_ceros+1
    return(cuenta,extintos,cuenta_ceros)

def tablita(list1,list2,h):
    diccionario={"TIPO":list(range(1,h+1)),"Cuantos":list1,"Frecuencia":list2}
    tablita=pd.DataFrame(diccionario)
    tablita.set_index('TIPO', inplace=True)
    print("\n")
    print(tablita)

def titulo(x):
    print("\n.....")
    print("GENERACION "+str(x))

def padres_a_observar(M,z):
    P_uno=[]
```

```

for n in range(M):
    y=np.random.rand()
    if y<z:
        p=2
    else:
        p=1
    P_uno.append(p)
return(P_uno)

def padres_potenciales(M,list):
    Padres=[]
    for n in range(M):
        papi=[]
        if list[n]==2:
            pa=np.random.randint(1,M+1)
            papi.append(pa)
            pi=np.random.randint(1,M+1)
            papi.append(pi)
        else:
            pa=np.random.randint(1,M+1)
            papi.append(pa)
            pi=0
            papi.append(pi)
        Padres.append(papi)
    return(Padres)

def regla_de_coloracion(r,list1,list2,M,h,list3):
    if r==1:
        V_uno=transitiva(list1,list2,M,h,list3)
    elif r==2:
        V_uno=gana_el_otro(list1,list2,M,h,list3)
    elif r==3:
        V_uno=orbita(list1,list2,M,h,list3)
    elif r==4:
        V_uno=orbita_malita(list1,list2,M,h,list3)
    elif r==5:
        V_uno=personalizada(list1,list2,M,h,list3)
    return V_uno

#gana siempre el "mayor"
def transitiva(list1,list2,M,h,list3):
    V_uno=[]
    raiz=0
    while raiz<M:
        cha=list1[list2[raiz][0]-1]
        che=list1[list2[raiz][1]-1]
        if (list2[raiz][1]==0):
            bebe=cha
        elif(cha<=che):
            bebe=che

```

```

        elif cha>che:
            bebe=cha
        V_uno.append(bebe)
        raiz=raiz+1
    return(V_uno)

#la competencia entre dos resulta en que gana el tercero
def gana_el_otro(list1,list2,M,h,list3):
    V_uno=[]
    raiz=0
    while raiz<M:
        cha=list1[list2[raiz][0]-1]
        che=list1[list2[raiz][1]-1]
        if (list2[raiz][1]==0):
            bebe=cha
        elif(cha==che):
            bebe=cha
        elif (cha==1 and che==2) or (cha==2 and che==1):
            bebe=3
        elif(cha==2 and che==3) or (cha==3 and che==2):
            bebe=1
        elif(cha==1 and che==3) or (cha==3 and che==1):
            bebe=2
        V_uno.append(bebe)
        raiz=raiz+1
    return(V_uno)

def orbita(list1,list2,M,h,list3):
    V_uno=[]
    raiz=0
    while raiz<M:
        cha=list1[list2[raiz][0]-1]
        che=list1[list2[raiz][1]-1]
        if (list2[raiz][1]==0):
            bebe=cha
        elif(cha==che):
            bebe=cha
        elif (cha==h and che==1) or (cha==1 and che==h):
            bebe=1
        elif(cha==che+1):
            bebe=cha
        elif(che==cha+1):
            bebe=che
        else:
            y=np.random.rand()
            if y<.5:
                bebe=cha
            else:
                bebe=che
        V_uno.append(bebe)

```



```

        raiz=raiz+1
    return(V_uno)

#gana siempre el mayor, a menos de que compitan el más chiquito con el
más grandote
def orbita_malita(list1,list2,M,h,list3):
    V_uno=[]
    raiz=0
    while raiz<M:
        cha=list1[list2[raiz][0]-1]
        che=list1[list2[raiz][1]-1]
        if (list2[raiz][1]==0):
            bebe=cha
        elif (cha==h and che==1) or (cha==1 and che==h):
            bebe=1
        elif(cha<=che):
            bebe=che
        elif cha>che:
            bebe=cha
        V_uno.append(bebe)
        raiz=raiz+1
    return(V_uno)

def personalizada(list1,list2,M,h,list3):
    V_uno=[]
    raiz=0
    while raiz<M:
        cha=list1[list2[raiz][0]-1]
        che=list1[list2[raiz][1]-1]
        if (list2[raiz][1]==0):
            bebe=cha
        elif cha==che:
            bebe=cha
        else:
            if cha<che:
                a=cha
                b=che
            else:
                a=che
                b=cha
            suma=0
            tope=0
            for tope in range(a):
                suma=suma+tope
            indice=((a-1)*h-suma)+b-a-1
            seccion=list3[indice]
            y=np.random.rand()
            index=0
            sumaprobas=0
            while y>sumaprobas:

```

```

        sumaprobas=sumaprobas+seccion[index]
        index=index+1
        bebe=index
        V_uno.append(bebe)
        raiz=raiz+1
    return(V_uno)

```

```

V_cero=[]
generacion=0
cuenta_ceros=0
cua_extintos=[]
frecuencias_tipos=[]
personalizado=[]

```

```

semilla=int(input('¿Qué semilla aleatoria quisieras utilizar?\n'))
np.random.seed(semilla)

```

```

k=int(input('¿Cuántos tipos alélicos hay?\n'))
for i in range(k):
    frecuencias_tipos.append([])

```

```

x=float(input('¿Qué probabilidad hay de que observe dos padres?\n'))
if x>1:
    while x>1:
        print('\n¡Este valor debe estar entre 0 y 1!')
        x=float(input('¿Qué probabilidad hay de que observe dos padre?
\n'))

```

```

regla=int(input('Elige la regla de coloración:\n1 Transitiva\n2 Gana
el otro\n3 Órbita\n4 Órbita Malita\n5 Personalizada\n'))

```

```

if regla==5:
    print('\nVas a determinar con qué probabilidad gana cada tipo si se
dan las siguientes competencias. \nRecuerda que éstas deben sumar 1.')
    s=1
    while s<k:
        for i in range(s+1,k+1):
            ganaproba=[]
            w=[s,i]
            print('\nSi compiten el tipo '+str(s)+' y el tipo '+str(i)
+':')
            sumas=0
            for a in range(1,k+1):
                ganap=float(input('¿Con qué probabilidad gana el tipo
'+str(a)+'?\n'))
                ganaproba.append(ganap)

```

```

        sumas=sumas+ganap
    if sumas!=1:
        while sumas!=1:
            ganaproba=[]
            print('\nError:iLa suma de las probabilidades debe
sumar 1!\nIntenta de nuevo')
            print('\nSi compiten el tipo '+str(s)+' y el tipo
'+str(i)+':')
            sumas=0
            for a in range(1,k+1):
                ganap=float(input('¿Con qué probabilidad gana
el tipo '+str(a)+'?\n'))
                ganaproba.append(ganap)
                sumas=sumas+ganap
            personalizado.append(ganaproba)
        s=s+1
    print(personalizado)

if regla==2 and k!=3:
    while regla==2 and k!=3:
        print("\niLa regla de Gana el Otro sólo es válida si hay 3
tipos alélicos!\nPrueba nuevamente")
        k=int(input('¿Cuántos tipos alélicos hay?\n'))
        regla=int(input('Elige la regla de coloración:\n1
Transitiva\n2 Gana el otro\n3 Órbita\n4 Órbita Malita\n5
Personalizada\n'))
        if regla>5 or regla<1:
            while regla>5 or regla<1:
                print("\niTenemos sólo reglas de coloración del uno al cinco!
\nPrueba nuevamente")
                regla=int(input('Elige la regla de coloración:\n1
Transitiva\n2 Gana el otro\n3 Órbita\n4 Órbita Malita5
Personalizada\n'))

rand=(int(input('1 Primera generación aleatoriamente\n2 Ingresar datos
para primera generación\n')))
if rand!=1 and rand!=2:
    while rand!=1 and rand!=2:
        print('\nniiii1 para aleatorio y 2 para manual. ¡No hay de
otra!')
        rand=(int(input('1 Primera generación aleatoriamente\n2
Ingresar datos para primera generación\n')))

#generar nuestra primera generación de manera aleatoria
if rand==1:
    N=int(input('¿De qué tamaño es tu población? \n'))
    primera_aleatoria(N,k,V_cero)

    cuentas(V_cero,k)

```

```

    cuantos=cuentas(V_cero,k)[0]

    cua_extintos=cuentas(V_cero,k)[1]

    cuenta_ceros=cuentas(V_cero,k)[2]

#generar nuestra primera generación manualmente
elif rand==2:
    R=[]
    N=0
    cuantos=[]
    for d in range(k):
        t=(int(input('¿Cuántos individuos en V_cero tienen el tipo
'+str(d+1)+'?\n')))
        y=[d+1]*t
        R.append(y)
        N=N+t
        cuantos.append(t)
        if t==0:
            cuenta_ceros=cuenta_ceros+1
            cua_extintos.append(d+1)

    for sublist in R:
        for item in sublist:
            V_cero.append(item)

#ya las dos versiones están parejas, ahora veamos cómo está
distribuida nuestra población
frecuencia=np.round(np.array(cuantos)/N,4)

for j in range(k):
    frecuencias_tipos[j].append(frecuencia[j])

tablita(cuantos,frecuencia,k)

print("\nHay "+str(cuenta_ceros)+" tipos alélicos desaparecidos hasta
la generacion "+str(generacion)+":")
print(cua_extintos)

#-----Aquí añadir loop
while (cuenta_ceros<k-1 and generacion<150):
    generacion=generacion+1
    titulo(generacion)

    cua=padres_a_observar(N,x)
    papas=padres_potenciales(N,cua)

    V_uno=regla_de_coloracion(regla,V_cero,papas,N,k,personalizado)

```

```

print("\nLa generación "+str(gereneracion)+" es:")
print(V_uno)

cuantos=cuentas(V_uno,k)[0]
frecuencia=np.round(np.array(cuantos)/N,4)

for j in range(k):
    frecuencias_tipos[j].append(frecuencia[j])

    cuenta_ceros=cuentas(V_uno,k)[2]
    print("\nHay "+str(cuenta_ceros)+" tipos alélicos desaparecidos
hasta la generacion "+ str(gereneracion)+":")
    print(cuentas(V_uno,k)[1])

    tablita(cuantos,frecuencia,k)
    V_cero=list(V_uno)

    print("\n")
    print(frecuencias_tipos)

if k<11:
    fig,ax=plt.subplots()
    barWidth=.8

colores=['red','blue','yellow','grey','green','brown','orange','pink',
'purple','salmon']
r=list(range(0,gereneracion+1))

ax.bar(r,frecuencias_tipos[0],color=colores[0],edgecolor='white',width
=barWidth,label='tipo uno')

ax.bar(r,frecuencias_tipos[1],bottom=frecuencias_tipos[0],color=colores
s[1],edgecolor='white',width=barWidth,label='tipo dos')
    if k>2:
        ax.bar(r,frecuencias_tipos[2],bottom=[i+j for i,j in
zip(frecuencias_tipos[0],frecuencias_tipos[1])],color=colores[2],edgec
olor='white',width=barWidth,label='tipo tres')
        if k>3:
            ax.bar(r,frecuencias_tipos[3],bottom=[i+j+k for i,j,k in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2])],c
olor=colores[3],edgecolor='white',width=barWidth,label='tipoc uatro')
            if k>4:
                ax.bar(r,frecuencias_tipos[4],bottom=[i+j+k+l for
i,j,k,l in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2],fre
cuencias_tipos[3])],color=colores[4],edgecolor='white',width=barWidth,
label='tipo cinco')
                if k>5:
                    ax.bar(r,frecuencias_tipos[5],bottom=[i+j+k+l+m

```

```

for i,j,k,l,m in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2],fre
cuencias_tipos[3],frecuencias_tipos[4]))],color=colores[5],edgecolor='w
hite',width=barWidth,label='tipo seis')
        if k>6:

ax.bar(r,frecuencias_tipos[6],bottom=[i+j+k+l+m+n for i,j,k,l,m,n in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2],fre
cuencias_tipos[3],frecuencias_tipos[4],frecuencias_tipos[5]))],color=co
lores[6],edgecolor='white',width=barWidth,label='tipo siete')
        if k>7:

ax.bar(r,frecuencias_tipos[7],bottom=[i+j+k+l+m+n+o for i,j,k,l,m,n,o
in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2],fre
cuencias_tipos[3],frecuencias_tipos[4],frecuencias_tipos[5],frecuencia
s_tipos[6]))],color=colores[7],edgecolor='white',width=barWidth,label='
tipo ocho')
                if k>8:

ax.bar(r,frecuencias_tipos[8],bottom=[i+j+k+l+m+n+o+p for
i,j,k,l,m,n,o,p in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2],fre
cuencias_tipos[3],frecuencias_tipos[4],frecuencias_tipos[5],frecuencia
s_tipos[6],frecuencias_tipos[7]))],color=colores[8],edgecolor='white',w
idth=barWidth,label='tipo nueve')
                        if k>9:

ax.bar(r,frecuencias_tipos[9],bottom=[i+j+k+l+m+n+o+p+q for
i,j,k,l,m,n,o,p,q in
zip(frecuencias_tipos[0],frecuencias_tipos[1],frecuencias_tipos[2],fre
cuencias_tipos[3],frecuencias_tipos[4],frecuencias_tipos[5],frecuencia
s_tipos[6],frecuencias_tipos[7],frecuencias_tipos[8]))],color=colores[9
],edgecolor='white',width=barWidth,label='tipo diez')
                plt.xlabel('generación')
                plt.ylabel('frecuencia')
                plt.title('Frecuencia de los Tipos')

                plt.show()

else:
    print('\n¡Eso no lo sabemos graficar! \n Intenta con menos de 11
tipos alélicos')

```

Bibliografía

- [1] Nathanaël Berestycki. *Recent progress in coalescent theory*. En: *Ensaïos Matematicos* 16.1 (2009), págs. 1-193.
- [2] Richard Durrett. *Probability models for DNA sequence evolution*. Springer Science & Business Media, 2008.
- [3] Alison Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*. Vol. 2012. Springer Science & Business Media, 2011.
- [4] Larry Gonick y Mark Wheelis. *Cartoon guide to genetics*. Harper Perennial, 1991.
- [5] Adrián González Casanova y Charline Smadi. *Multidimensional Wright-Fisher processes with general frequency-dependent selection*. <https://arxiv.org/abs/1903.06406>. 2019+.
- [6] Adrián González Casanova y Dario Spano. *Duality and fixation in Wright-Fisher processes with frequency-dependent selection*. En: *The Annals of Applied Probability* 28.1 (2018), págs. 250-284.
- [7] Sheldon M Ross, John J Kelly, Roger J Sullivan, William James Perry, Donald Mercer, Ruth M Davis, Thomas Dell Washburn, Earl V Sager, Joseph B Boyce y Vincent L Bristow. *Stochastic processes*. Vol. 2. Wiley New York, 1996.
- [8] Rick et al Scavett. *Data Camp*. URL: <https://www.datacamp.com>.
- [9] Barry Sinervo y Curt M Lively. *The rock-paper-scissors game and the evolution of alternative male strategies*. En: *Nature* 380.6571 (1996), pág. 240.
- [10] Patricia Velez, Jaime Gasca-Pineda, Edmundo Rosique-Gil, Luis E Eguiarte, Laura Espinosa-Asuar y Valeria Souza. *Microfungal oasis in an oligotrophic desert: diversity patterns and community structure in three freshwater systems of Cuatro Ciénegas, Mexico*. En: *PeerJ* 4 (2016), e2064.