

Documentación del Simulador de Dinámica de Frecuencias Alélicas

Carlos Johan Parada Buitrago

Joseth David Urrego Vásquez

Miguel Ángel Torres Rincón

Facultad de Ciencias, Universidad Nacional de Colombia

2015180 Programación y métodos numéricos

23 de julio de 2025

Resumen

Este proyecto analiza el efecto de la mutación como fuerza evolutiva sobre la estructura genética de las poblaciones. Partiendo del modelo nulo del equilibrio de Hardy-Weinberg, se desarrolla un marco matemático para cuantificar cómo las tasas de mutación directa e inversa alteran las frecuencias alélicas a lo largo del tiempo. Se derivan y analizan las ecuaciones que describen el cambio generacional para sistemas de dos y múltiples alelos. El objetivo principal es determinar las condiciones bajo las cuales se alcanza un equilibrio de mutación, donde la pérdida de alelos por mutación se compensa exactamente con su ganancia. Se demuestra que, en ausencia de otras fuerzas evolutivas, las frecuencias alélicas convergen a un punto de equilibrio estable que depende exclusivamente de las tasas de mutación. Este análisis se apoya en la resolución de sistemas de ecuaciones lineales y se propone su validación mediante simulaciones numéricas.

Palabras Clave: Genética de poblaciones, Hardy-Weinberg, mutación, equilibrio genético, frecuencias alélicas, modelo matemático, síntesis moderna.

Índice

1. Introducción	5
2. Problemática	6
3. Marco Teórico	6
3.1. Modelo de Hardy-Weinberg	6
3.1.1. Conceptos Fundamentales	6
3.1.2. Condiciones para el Equilibrio	7
3.1.3. La Ecuación de Hardy-Weinberg (Para Dos Alelos)	8
3.1.4. Extensión a Múltiples Alelos (k Alelos)	8
3.2. Conceptos en Genética de Poblaciones	9
3.3. Modelos de Mutación	10
3.3.1. Matrices de Transición para Sistemas Multi-Alélicos	11
3.4. Cadenas de Markov	12
3.4.1. ¿Qué es una cadena de Markov?	12
3.4.2. Elementos clave	12
3.4.3. Distribuciones de probabilidad	13
3.4.4. Tipos de estados	13
3.4.5. Ejemplo: Modelo de clima	14
3.4.6. Aplicaciones y utilidad	15
4. Objetivos	15
4.1. Objetivo general	15
4.2. Objetivos específicos	16
5. Pregunta de investigación	16
6. Hipótesis	16
7. Fase de Desarrollo	17
7.1. Alcance y Componentes	17
7.2. Formulación Matemática	17
7.3. Punto de equilibrio	18
7.3.1. De la ecuación de invariancia al sistema lineal	18
7.3.2. Incorporación de la normalización	19
7.3.3. Métodos de resolución	19
7.3.4. Alelos aislados y equilibrio	20
7.4. Implementación Numérica y Código	20
7.5. Pruebas de Verificación y Ejemplos	23

7.5.1.	Ejemplo 1: Sistema Completamente Conectado (Control)	23
7.5.2.	Ejemplo 2: Un Alelo Aislado	24
7.5.3.	Ejemplo 3: Dos Subsistemas Independientes	24
8.	Conclusiones	25
8.1.	Validación de hipótesis	25
8.2.	Aportes del simulador	26
8.3.	Implicaciones biológicas	26
8.4.	Limitaciones	26
8.5.	Líneas futuras	26
8.6.	Reflexión final	27
A.	Código fuente de la simulación	30

1. Introducción

La genética de poblaciones se encarga de estudiar cómo cambian las frecuencias alélicas a lo largo del tiempo, bajo la acción de diferentes fuerzas evolutivas como la mutación, la selección natural, la deriva genética y el flujo génico. Aunque en la naturaleza estas fuerzas suelen actuar de manera conjunta, a veces es útil analizarlas por separado para entender bien cómo contribuye cada una. En este trabajo nos enfocamos únicamente en la mutación, que aunque es un proceso aleatorio y suele tener efectos pequeños por sí sola, con el tiempo puede tener un impacto considerable en la composición genética de una población.

El estudio formal de estos procesos comenzó a desarrollarse a principios del siglo XX, con los aportes de Fisher, Haldane y Wright [15, 16, 17], quienes introdujeron herramientas matemáticas para describir cómo cambian las frecuencias alélicas bajo distintos escenarios. Más adelante, autores como Gillespie profundizaron en el enfoque estocástico, permitiendo modelar mejor la dinámica genética en poblaciones finitas [19].

Uno de los conceptos más conocidos en este campo es el equilibrio de Hardy–Weinberg [6, 18], que describe una población en la que las frecuencias alélicas permanecen constantes si no hay fuerzas evolutivas actuando. Sin embargo, en condiciones reales este equilibrio rara vez se cumple, y basta con que haya mutación para que las frecuencias comiencen a desviarse de ese punto teórico [19]. Justamente, lo que nos interesa aquí es analizar cómo se comporta una población cuando solo está actuando la mutación y todo lo demás se mantiene constante.

Para eso, vamos a usar un modelo basado en cadenas de Markov, que nos permite representar la evolución del sistema de forma probabilística. En este contexto, cada estado del sistema corresponde a una distribución de frecuencias alélicas en una generación, y los cambios entre generaciones se modelan con una matriz de transición que contiene las probabilidades de que un alelo mute a otro. Esta herramienta no solo es útil desde el punto de vista teórico, sino que también nos permite hacer simulaciones para observar cómo el sistema evoluciona hacia un equilibrio o estado estacionario.

A lo largo de este trabajo, vamos a desarrollar ese modelo paso a paso y mostrar cómo las tasas de mutación afectan el equilibrio final. La idea es ver, tanto de forma matemática como por medio de simulaciones, cómo una población puede alcanzar un punto estable en términos de frecuencias alélicas, incluso cuando las mutaciones están ocurriendo todo el tiempo. Aunque en la realidad existen muchas otras fuerzas en juego, este enfoque simplificado nos permite entender mejor el papel específico que juega la mutación en la evolución genética de una población.

2. Problemática

En genética de poblaciones, el principio de Hardy-Weinberg establece un modelo teórico donde las frecuencias alélicas y genotípicas permanecen constantes si no actúan fuerzas evolutivas. Sin embargo, en los sistemas biológicos reales, las mutaciones son inevitables y constituyen una fuente persistente de variación genética.

Aunque las tasas de mutación suelen ser bajas, su efecto acumulativo puede alterar significativamente el equilibrio genético de una población. A pesar de existir modelos teóricos para este fenómeno, en muchas investigaciones no se aborda con profundidad cómo varían las frecuencias alélicas en poblaciones con múltiples alelos cuando solo actúa la mutación, ni cómo se alcanzan los puntos de equilibrio en tales escenarios.

3. Marco Teórico

3.1. Modelo de Hardy-Weinberg

En genética de poblaciones, uno de los objetivos fundamentales es comprender cómo las frecuencias alélicas y genotípicas cambian o se mantienen a lo largo del tiempo. Un concepto clave en este estudio es el punto de equilibrio genético. Este se refiere a un estado en una población donde las frecuencias de los alelos y genotipos permanecen constantes de una generación a la siguiente. Este estado de estabilidad ocurre cuando las fuerzas evolutivas que pueden alterar la composición genética de una población están ausentes o se cancelan entre sí. Comprender cómo y bajo qué condiciones se alcanzan estos puntos de equilibrio es esencial para modelar la dinámica de las poblaciones en la naturaleza. Este principio fundamental fue descrito por primera vez por G. H. Hardy en 1908 [6].

3.1.1. Conceptos Fundamentales

- **Población:** Un grupo de individuos de la misma especie que viven en un área determinada y se cruzan entre sí. Es la unidad de estudio en genética de poblaciones [19].
- **Locus (plural: loci):** La posición específica de un gen en un cromosoma [19].
- **Gen:** Una unidad de herencia; es una secuencia de ADN que codifica para un rasgo específico. Un gen puede tener diferentes formas, llamadas alelos [5].
- **Alelo:** Una variante específica de un gen [5].
- **Genotipo:** La composición alélica de un individuo en un locus específico. Para dos alelos (A_1 y A_2), existen tres genotipos posibles: A_1A_1 (homocigoto dominante), A_1A_2 (heterocigoto) y A_2A_2 (homocigoto recesivo) [19].

- **Fenotipo:** La característica observable o medible de un organismo, resultado de la expresión de su genotipo y la interacción con el ambiente [5].
- **Frecuencia Génica (o Alélica):** Proporción de un alelo específico en el acervo génico de la población. Se denota como p para la frecuencia de A_1 y q para la frecuencia de A_2 . La suma de las frecuencias alélicas debe ser 1: $p + q = 1$ [19].
- **Frecuencia Genotípica:** Proporción de individuos en la población que poseen un genotipo específico. Para dos alelos, se denota P para A_1A_1 , H para A_1A_2 y Q para A_2A_2 . La suma de las frecuencias genotípicas debe ser 1: $P + H + Q = 1$. En el contexto de Hardy–Weinberg, las frecuencias del genotipo A_1A_2 se describen comúnmente como $2pq$ para mantener la consistencia con las fórmulas del equilibrio [19].
- **Acervo Génico (o Pool Génico):** Conjunto total de alelos presentes en todos los individuos de una población para todos los loci [5].

3.1.2. Condiciones para el Equilibrio

Para que las frecuencias alélicas y genotípicas permanezcan constantes (es decir, para que la población esté en equilibrio de H-W), se deben cumplir estrictamente las siguientes condiciones [3, 19]:

- **No hay Mutación:** Las tasas de mutación del alelo A al a y viceversa son cero. Los alelos no cambian espontáneamente.
- **No hay Migración (Flujo Génico):** No entra ni sale ningún individuo (y por tanto, ningún alelo) de la población. La población está aislada.
- **Población Infinitamente Grande:** El tamaño poblacional es tan grande que los efectos del muestreo aleatorio (deriva génica) son despreciables [2, 8]. No hay cambios aleatorios en las frecuencias alélicas.
- **Apareamiento Aleatorio:** Los individuos se aparean al azar con respecto al genotipo en el locus estudiado. No hay preferencias basadas en el fenotipo o genotipo. El apareamiento aleatorio conserva las frecuencias alélicas y, después de una generación, las frecuencias genotípicas.
- **No hay Selección Natural:** Todos los genotipos tienen la misma probabilidad de sobrevivir y reproducirse con éxito. No hay ventaja o desventaja reproductiva asociada a ningún genotipo.

3.1.3. La Ecuación de Hardy-Weinberg (Para Dos Alelos)

Si una población cumple con todos los supuestos anteriores para un locus con dos alelos (A_1 y A_2), entonces las frecuencias alélicas (p y q) permanecerán constantes de generación en generación, y las frecuencias genotípicas alcanzarán un equilibrio en una sola generación de apareamiento aleatorio y permanecerán constantes en ese equilibrio. Este equilibrio se expresa mediante la ecuación [3, 19, 4]:

$$p^2 + 2pq + q^2 = 1$$

- p^2 representa la frecuencia esperada de individuos homocigotos dominantes (A_1A_1) en la población. Es la probabilidad de que un gameto con A_1 se combine con otro gameto con A_1 ($p \cdot p$).
- $2pq$ representa la frecuencia esperada de individuos heterocigotos (A_1A_2). Es la suma de las probabilidades de un gameto A_1 combinándose con un gameto A_2 ($p \cdot q$) y de un gameto A_2 combinándose con un gameto A_1 ($q \cdot p$), es decir, $2 \cdot p \cdot q$.
- q^2 representa la frecuencia esperada de individuos homocigotos recesivos (A_2A_2). Es la probabilidad de que un gameto con A_2 se combine con otro gameto con A_2 ($q \cdot q$).

3.1.4. Extensión a Múltiples Alelos (k Alelos)

El Principio de Hardy-Weinberg se generaliza para el caso de k alelos (A_1, \dots, A_k) lo cual fue obtenido por Weinberg en 1909 [4, 19].

Variables Utilizadas:

- A_i : Representa el i -ésimo alelo, donde $i = 1, \dots, k$.
- P_{ij} : Frecuencia del genotipo ordenado A_iA_j . La frecuencia del genotipo no ordenado A_iA_j es $P_{ij} + P_{ji} = 2P_{ij}$.
- p_i : Frecuencia del alelo A_i en la población. Se puede expresar como la suma de las frecuencias de los genotipos que contienen A_i : $p_i = \sum_{j=1}^k P_{ij}$.

Frecuencias en la Siguiete Generación: Similar al caso de dos alelos, bajo las condiciones de Hardy-Weinberg, las frecuencias genotípicas en la siguiente generación, P'_{ij} , se establecen como el producto de las frecuencias de los alelos correspondientes:

- $P'_{ij} = P_{ij}$: Frecuencia esperada del genotipo A_iA_j en la siguiente generación para cualquier i y j .

3.2. Conceptos en Genética de Poblaciones

Teoremas de Convergencia en Genética de Poblaciones: La convergencia de las frecuencias alélicas al equilibrio en los modelos de población es un aspecto fundamental. Los modelos de mutación y selección, especialmente aquellos que involucran múltiples alelos, a menudo pueden formularse como sistemas dinámicos (ecuaciones de diferencia o diferenciales) sobre un espacio de estados. Para sistemas bien comportados (por ejemplo, con matrices de transición primitivas o irreducibles), la teoría matemática garantiza que las trayectorias de las frecuencias alélicas convergen a un equilibrio único y estable. En los modelos de selección donde la aptitud media es una función de Lyapunov, las poblaciones evolucionan hacia un equilibrio asintóticamente estable, con la aptitud media aumentando estrictamente a lo largo de las trayectorias que no están en equilibrio [4]. Si solo existe un número finito de equilibrios, cada solución converge a un único punto de equilibrio.

Deriva Genética (Genetic Drift): Aunque este proyecto se centra en la mutación, es crucial reconocer que en poblaciones de tamaño finito, la deriva genética es una fuerza omnipresente. La deriva genética es el cambio aleatorio en las frecuencias génicas de una población de una generación a otra, causado por el error de muestreo en la transmisión de los alelos [19]. Sus efectos son más pronunciados en poblaciones pequeñas [8], donde puede llevar a la pérdida de algunos alelos o a la fijación de otros, independientemente de si son ventajosos o desventajosos. En poblaciones reales, la deriva genética interactúa con la mutación, ya que puede eliminar nuevos alelos creados por mutación o, por el contrario, llevarlos a una frecuencia alta por puro azar. En poblaciones pequeñas, los mutantes deletéreos pueden fijarse por deriva genética, lo que lleva a un proceso conocido como la “rata de Muller”, donde los mutantes deletéreos se acumulan en los cromosomas.

Selección Natural: La selección natural es el mayor impulsor de la evolución. Actúa cuando los genotipos en una población difieren en su aptitud (viabilidad, éxito de apareamiento, fertilidad), lo que lleva a un número diferencial de descendientes. Las mutaciones proporcionan la materia prima (variación genética) sobre la cual actúa la selección. La aptitud se mide en términos del número de descendientes aportados a la siguiente generación. La aptitud media de la población generalmente aumenta de generación en generación y permanece sin cambios solo en los estados de equilibrio. Este es un caso especial del Teorema Fundamental de la Selección Natural de Fisher (1930) [5]. Un equilibrio mutación-selección ocurre cuando la tasa de introducción de nuevos alelos por mutación se equilibra con la tasa de eliminación de esos alelos por selección [9].

Flujo Génico (Migración): El flujo génico, o migración, se refiere al movimiento de individuos y sus genes entre poblaciones. Este proceso puede alterar las frecuencias alélicas y genotípicas de una población al introducir nuevos alelos o cambiar las proporciones de

los alelos existentes. El flujo génico tiende a homogeneizar las poblaciones y contrarrestar los efectos de la deriva genética o la diferenciación por mutación y selección [19].

Mutación: Una mutación es cualquier cambio heredable en el material genético. Las mutaciones son la fuente última de la variabilidad genética [5]. A nivel molecular (gen), las mutaciones pueden ocurrir de diversas formas, como sustituciones de bases (donde un par de nucleótidos es reemplazado por otro), inserciones o deleciones de ADN, inversiones de secuencias o transposiciones. Para muchos modelos de genética de poblaciones, el origen molecular de un mutante no es necesariamente relevante; lo que a menudo cuenta es solo la tasa a la que ocurren las mutaciones y el efecto de un mutante en la aptitud o en el carácter considerado [7]. Las tasas de mutación espontánea por locus por generación son típicamente del orden de 10^{-4} a 10^{-6} [19].

Importancia de las Mutaciones en la Variación Genética: Las mutaciones son la fuente última de toda la variación genética en una población. Sin mutaciones, no habría nuevos alelos y, por lo tanto, la evolución no podría ocurrir, ya que carecería de la materia prima sobre la cual actúan la selección natural y la deriva genética [7, 8].

Tamaño Poblacional Efectivo (N_e): El tamaño poblacional efectivo (N_e) es un concepto clave en genética de poblaciones que a menudo difiere del tamaño censal (el número total de individuos). Representa el tamaño de una población idealizada que experimentaría la misma cantidad de deriva genética o cambio en la endogamia que la población real bajo estudio [2]. Factores como proporciones de sexos desiguales, variaciones en el número de descendientes entre individuos, fluctuaciones en el tamaño de la población y el apareamiento no aleatorio pueden reducir el N_e en comparación con el tamaño censal. Un N_e más pequeño implica que la población es más susceptible a los efectos aleatorios de la deriva genética y a la fijación o pérdida de alelos, lo que puede interactuar significativamente con la introducción de variación por mutación [8].

3.3. Modelos de Mutación

La mutación es uno de los factores centrales que guían la evolución biológica. La mutación genera la variabilidad genética sobre la cual otros factores pueden actuar. Este modelo describe cómo cambian las frecuencias alélicas debido a la mutación de un tipo alélico a otro. Consideremos 2 alelos, A_1, A_2 .

Variables Utilizadas:

- p : Frecuencia del alelo A_1 en la población.
- q : Frecuencia del alelo A_2 en la población.

- μ : Tasa de mutación (probabilidad) de que un gen A_1 se convierta en un A_2 en la siguiente generación.
- ν : Tasa de mutación (probabilidad) de que un gen A_2 se convierta en un A_1 en la siguiente generación.

La frecuencia de $A_1(p')$ y $A_2(q')$ en la siguiente generación está dada por la ecuación (pura) de mutación:

$$p' = p(1 - \mu) - q\nu$$

$$q' = q(1 - \nu) - p\mu$$

Punto de Equilibrio Mutacional: Si ν o μ son positivos, existe una única frecuencia de equilibrio (\hat{p}), obtenida cuando ($p' = p$). Esta es dada por:

$$\hat{p} = \frac{\nu}{\nu + \mu}$$

La solución explícita de esta ecuación de recurrencia es

$$p(t) - \hat{p} = (p_0 - \hat{p})(1 - \mu - \nu)^t,$$

donde p_0 es la frecuencia inicial de A_1 . Esto demuestra que la convergencia al equilibrio ocurre a una tasa geométrica, pero es muy lenta porque la suma de las tasas de mutación $\mu + \nu$ es típicamente muy pequeña [3, 19].

3.3.1. Matrices de Transición para Sistemas Multi-Alélicos

Un sistema multi-alélico es aquel en el que un locus genético posee más de dos alelos posibles (A_1, \dots, A_n). Estos sistemas son comunes en loci altamente polimórficos como los del sistema ABO sanguíneo o el sistema HLA en humanos.

¿Qué son las matrices de transición?: Las matrices de transición son herramientas matemáticas que permiten modelar cómo cambian las frecuencias alélicas en el tiempo, considerando las tasas de mutación entre múltiples alelos. Cada entrada de la matriz representa la probabilidad de mutación de un alelo a otro en una generación determinada.

Estructura de la matriz de transición: En un sistema con k alelos, se puede definir una matriz de mutación \tilde{U} de tamaño $k \times k$. El elemento \tilde{u}_{ij} de esta matriz representa la probabilidad de que un alelo j mute a alelo i (para $i \neq j$), o la fracción de alelos i que no mutan (para $i = j$).

Convergencia y Teorema de Perron-Frobenius: Si la matriz C (o \tilde{U} en el caso de solo mutación) es *primitiva* (lo que implica que, después de un número suficiente de generaciones, cualquier alelo puede dar lugar a descendientes de cualquier otro tipo con probabilidad positiva), entonces la dinámica del sistema converge a un equilibrio único y completamente polimórfico (donde todos los alelos están presentes con frecuencia positiva, es decir, $\hat{p}^i > 0$ para todo i). Este equilibrio es globalmente asintóticamente estable. Para tales matrices, el Teorema de Perron-Frobenius garantiza la existencia y unicidad de un vector propio positivo asociado al valor propio dominante (el radio espectral), que corresponde al vector de frecuencias alélicas en equilibrio [4].

3.4. Cadenas de Markov

3.4.1. ¿Qué es una cadena de Markov?

Imagina que recorres un laberinto de habitaciones y en cada puerta eliges a dónde ir según ciertas probabilidades. Una **cadena de Markov** es un modelo matemático que describe cómo un sistema pasa de un *estado* (una habitación) a otro, donde cada movimiento depende únicamente del estado actual, sin memoria de los pasos anteriores [20].

3.4.2. Elementos clave

- **Estados (S):** Son las distintas “habitaciones” o situaciones posibles por las que puede transitar el proceso. Se denotan como $S = \{s_1, s_2, \dots, s_n\}$ [14].
- **Probabilidades de transición:** Para cada par de estados $i \rightarrow j$, existe una probabilidad $p_{ij} \in [0, 1]$ que indica la probabilidad de moverse de s_i a s_j en un paso. Para cada estado s_i se cumple

$$\sum_{j=1}^n p_{ij} = 1.$$

[10].

- **Propiedad de sin memoria (Markov):** La distribución del siguiente estado depende únicamente del estado actual, no de la trayectoria previa:

$$P(X_{t+1} = s_j \mid X_t = s_i, X_{t-1} = s_{i-1}, \dots) = P(X_{t+1} = s_j \mid X_t = s_i).$$

Este postulado es fundamental en la teoría de cadenas de Markov [21].

- **Matriz de transición (P):** Se define la matriz $P = [p_{ij}]_{i,j=1}^n$ cuya fila i contiene todos los p_{ij} . La matriz P es estocástica por filas y resume completamente las probabilidades de transición [14].

3.4.3. Distribuciones de probabilidad

En el estudio de cadenas de Markov, las distribuciones de probabilidad describen cómo se distribuye la probabilidad del sistema entre los distintos estados.

Distribución inicial La *distribución inicial* se denota por

$$\pi^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, \dots, \pi_m^{(0)}),$$

donde $\pi_i^{(0)}$ es la probabilidad de que el sistema comience en el estado i [21]. Debe cumplirse:

$$\sum_{i=1}^m \pi_i^{(0)} = 1.$$

Distribución después de n pasos: Si P es la matriz de transición de la cadena, la *distribución de probabilidad tras n pasos* se calcula como:

$$\pi^{(n)} = \pi^{(0)} P^n,$$

donde $\pi^{(n)}$ es un vector fila que da las probabilidades de estar en cada estado tras realizar n transiciones [14]. Nuevamente,

$$\sum_{i=1}^m \pi_i^{(n)} = 1.$$

Distribución estacionaria (límite): En ciertas cadenas ergódicas y aperiódicas, existe una *distribución límite* o *estacionaria* π tal que:

$$\pi = \pi P,$$

y además

$$\lim_{n \rightarrow \infty} \pi^{(n)} = \pi. [10].$$

Esta distribución da la probabilidad de encontrarse en cada estado una vez el sistema haya evolucionado mucho tiempo.

3.4.4. Tipos de estados

Para entender mejor cómo se comporta cada estado dentro de una cadena de Markov, describimos sus características y usamos ejemplos sencillos (véase [14, 10])

- **Absorbentes:** Una vez que el sistema entra en un estado absorbente, ya no puede salir. Ejemplo: un proceso de juego en el que si un jugador llega a nivel final, permanece allí. Matemáticamente, $p_{ii} = 1$. [21, Ch. 2]

- **Transitorios:** El sistema puede pasar por estos estados, pero la probabilidad de eventualmente abandonarlos es 1. Ejemplo: niveles intermedios de un juego que el jugador supera y no regresa. [14, Sec. 1.6]
- **Comunicación e irreducibilidad:** Decimos que dos estados i y j se comunican si es posible ir de i a j en algún número de pasos y viceversa. Si todos los pares de estados se comunican, la cadena es irreducible, lo que garantiza la existencia de una única distribución estacionaria. [10, Thm. 4.4]
- **Periodicidad:** El período de un estado i es el máximo común divisor de todos los números de pasos en que el sistema puede volver a i . Si este valor es mayor que 1, el estado es periódico (por ejemplo, alternancia forzada). Si el período es 1, el estado es aperiódico, condición necesaria para la convergencia a la distribución estacionaria. [11, Ch. 3]
- *Absorbentes:* $p_{ii} = 1$, el sistema queda atrapado en i .
- *Transitorios:* Existe probabilidad de abandonar el estado.
- *Irreducibles:* Todos los estados se comunican entre sí.
- *Aperiódicos:* No existe un período fijo para regresar al mismo estado.

3.4.5. Ejemplo: Modelo de clima

Definamos tres estados: Soleado (S), Nublado (N) y Lluvioso (L). Las probabilidades se representan en la matriz:

$$P = \begin{pmatrix} 0,6 & 0,3 & 0,1 \\ 0,2 & 0,5 & 0,3 \\ 0,4 & 0,3 & 0,3 \end{pmatrix}.$$

Distribución inicial ($\pi^{(0)}$): Se asume que el sistema inicia en un estado conocido o en una mezcla de posibles estados. Por ejemplo, si estamos modelando el clima y sabemos con certeza que hoy está soleado, la distribución inicial sería:

$$\pi^{(0)} = [1, 0, 0] \quad (100 \% \text{ Soleado}, 0 \% \text{ Nublado}, 0 \% \text{ Lluvioso}).$$

Distribución después de n pasos ($\pi^{(n)}$): Aplica la propiedad matricial para propagar probabilidades. Matemáticamente:

$$\pi^{(n)} = \pi^{(0)} P^n.$$

Ejemplo concreto: Con la matriz de ejemplo $P = \begin{pmatrix} 0,6 & 0,3 & 0,1 \\ 0,2 & 0,5 & 0,3 \\ 0,4 & 0,3 & 0,3 \end{pmatrix}$

- Tras 1 día: $\pi^{(1)} = \pi^{(0)}P = [0,6, 0,3, 0,1]$.
- Tras 2 días: $\pi^{(2)} = \pi^{(1)}P = [0,51, 0,33, 0,16]$ (aprox.).

Distribución estacionaria (π): Cuando $n \rightarrow \infty$, ciertas cadenas alcanzan un *régimen estacionario* donde la distribución no cambia al aplicar más transiciones:

$$\pi = \pi P.$$

Resolver este sistema lineal (con la condición adicional $\sum_{i=0} \pi_i = 1$) nos da las probabilidades a largo plazo. [14].

Interpretación práctica: Si el clima sigue las reglas de P , la proporción de días soleados a lo largo de muchas iteraciones tenderá a los valores de π . Esto es clave para entender el comportamiento promedio a gran escala.

3.4.6. Aplicaciones y utilidad

Las cadenas de Markov son ampliamente usadas para modelar fenómenos reales como:

- Clima y previsión meteorológica.
- Teoría de colas en sistemas de servicio.
- Algoritmos de ranking de páginas web (PageRank).
- Sistemas de recomendación y análisis de secuencias.

Su estructura matricial y propiedades de convergencia facilitan tanto la simulación del proceso como el análisis de su comportamiento a largo plazo.

4. Objetivos

4.1. Objetivo general

Analizar el impacto de las tasas de mutación en el equilibrio de frecuencias alélicas mediante modelos matemáticos de poblaciones teóricas con dos y múltiples alelos.

4.2. Objetivos específicos

1. Modelar matemáticamente la dinámica de frecuencias alélicas bajo mutación en sistemas bialélicos.
2. Extender el modelo a sistemas multi alélicos con tasas de mutación asimétricas.
3. Determinar analíticamente los puntos de equilibrio (distribución estacionaria) para diferentes configuraciones de tasas de mutación y caracterizar su estabilidad espectral.
4. Determinar puntos de equilibrio para diferentes configuraciones de tasas de mutación.
5. Simular numéricamente la convergencia al equilibrio tras generaciones.

5. Pregunta de investigación

¿Cómo afectan las distintas tasas de mutación a las frecuencias alélicas en poblaciones con dos y múltiples alelos, y qué condiciones determinan el punto de equilibrio genético bajo la acción exclusiva de la mutación?

6. Hipótesis

Hipótesis general: Si una población está sujeta únicamente a procesos de mutación, entonces se espera que las frecuencias alélicas evolucionen progresivamente hacia un equilibrio estable, dependiente exclusivamente de las tasas de mutación entre alelos, sin intervención de otros factores evolutivos como selección natural, deriva genética o migración.

Hipótesis específica (caso bialélico): En un sistema con dos alelos (A_1 y A_2), sujeto a tasas constantes de mutación μ ($A_1 \rightarrow A_2$) y ν ($A_2 \rightarrow A_1$), se espera que las frecuencias alélicas converjan numéricamente a valores de equilibrio teóricamente dados por:

$$\hat{p} = \frac{\nu}{\mu + \nu}, \quad \hat{q} = \frac{\mu}{\mu + \nu},$$

siendo consistentes con el equilibrio mutacional clásico.

Hipótesis específica (caso multialélico): En poblaciones con n alelos y una matriz de tasas de mutación entre pares de alelos, se espera que el sistema converja a una única distribución estacionaria que puede determinarse resolviendo un sistema lineal que balancea las tasas de entrada y salida para cada alelo. Esta solución predice teóricamente el equilibrio observado en las simulaciones numéricas, incluso en presencia de tasas asimétricas.

7. Fase de Desarrollo

7.1. Alcance y Componentes

1. **Modelo de Mutación:** Se incorpora la mutación como un proceso de cambio aleatorio en las frecuencias alélicas. Este proceso se modela con una matriz de transición estocástica M , donde cada elemento m_{ij} representa la probabilidad de que el alelo A_i mute al alelo A_j en una generación.
2. **Generalización a n Alelos:** El modelo está diseñado para ser flexible, permitiendo simular la dinámica de un número arbitrario de alelos, superando la limitación de los modelos clásicos de dos alelos.
3. **Simulación y Validación:** Se implementa un algoritmo de simulación por generaciones discretas y se desarrollan pruebas para validar el comportamiento del modelo en condiciones límite, como la ausencia de mutación o la existencia de alelos aislados.

7.2. Formulación Matemática

La formulación matemática de la dinámica de mutación alélica se basa en conceptos de cadenas de Markov discretas y en la conservación de probabilidad. A continuación se detalla cada componente fundamental:

- **Frecuencias alélicas y conservación de la probabilidad.** Definimos para cada generación t

$$\mathbf{p}(t) = (p_1(t), p_2(t), \dots, p_n(t)), \quad p_i(t) \geq 0, \quad \sum_{i=1}^n p_i(t) = 1.$$

Aquí, $p_i(t)$ es la probabilidad de observar el alelo A_i en un individuo tomado al azar en la generación t .

- **Matriz de transición de mutación.** Sea $M \in \mathbb{R}^{n \times n}$ la matriz de probabilidades de mutación, donde n es el número total de alelos. Cada entrada m_{ij} indica la probabilidad de que el alelo A_i mute a A_j en un solo paso generacional:

$$M = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{pmatrix}.$$

Los índices i, j recorren el conjunto $\{1, 2, \dots, n\}$, cubriendo así todas las mutaciones posibles entre los n alelos. Por consistencia probabilística:

$$m_{ij} \geq 0, \quad \sum_{j=1}^n m_{ij} = 1 \quad \forall i.$$

En particular, los elementos de la diagonal m_{ii} representan la probabilidad de que A_i *no* mute.

- **Ecuación de recursión discreta.** La dinámica de las frecuencias de generación en generación queda dada por

$$p_j(t+1) = \sum_{i=1}^n p_i(t) m_{ij}, \quad j = 1, 2, \dots, n.$$

En notación vectorial y matricial:

$$\mathbf{p}(t+1) = \mathbf{p}(t) M.$$

Así, la evolución completa en T generaciones es

$$\mathbf{p}(T) = \mathbf{p}(0) M^T.$$

7.3. Punto de equilibrio

Dado un sistema que evoluciona según una matriz de transición M de tamaño $n \times n$, la *distribución estacionaria* (o punto de equilibrio) $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$ cumple:

$$\mathbf{p}^* = \mathbf{p}^* M, \quad \sum_{i=1}^n p_i^* = 1.$$

7.3.1. De la ecuación de invariancia al sistema lineal

La igualdad $\mathbf{p}^* = \mathbf{p}^* M$ se reescribe como

$$\mathbf{p}^* M - \mathbf{p}^* = \mathbf{0},$$

o bien

$$\mathbf{p}^* (M - I) = \mathbf{0},$$

donde I es la matriz identidad de orden n . Al transponer, obtenemos un sistema homogéneo en la forma columna:

$$(M - I)^T \mathbf{p}^{*\top} = \mathbf{0}.$$

Como $(M - I)^T = M^T - I$, resulta:

$$(M^T - I) \mathbf{p}^{*\top} = \mathbf{0}.$$

Este sistema incluye n ecuaciones lineales, pero al ser homogéneo y singular (determinante cero), admite infinitas soluciones escalares.

7.3.2. Incorporación de la normalización

Para obtener la solución única que representa una distribución de probabilidad, imponemos:

$$\sum_{i=1}^n p_i^* = 1.$$

Estratégicamente, reemplazamos una de las n ecuaciones lineales del sistema homogéneo por esta ecuación de normalización. Así, se construye un nuevo sistema de n ecuaciones con n incógnitas:

$$\begin{cases} (M^T - I) \mathbf{p}^{*\top} = \mathbf{0}, \\ \sum_{i=1}^n p_i^* = 1. \end{cases}$$

En la práctica se suele substituir la última fila de $M^T - I$ por el vector fila $(1, 1, \dots, 1)$ y el término independiente correspondiente pasa de 0 a 1.

7.3.3. Métodos de resolución

Una vez formulado el sistema:

$$\left(M^T - I \quad \text{o bien con fila de normalización} \right) \begin{pmatrix} p_1^* \\ p_2^* \\ \vdots \\ p_n^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

se pueden aplicar:

- Eliminación de Gauss–Jordan.
- Descomposición LU/QR.
- Métodos iterativos (p.ej. potencias: iterar $\pi^{(k+1)} = \pi^{(k)}M$ hasta convergencia).

Con esto queda explicitado el origen de cada fila del sistema lineal y la forma de incorporar la condición de normalización.

7.3.4. Alelos aislados y equilibrio

En algunos sistemas de mutación aparecen alelos que quedan totalmente “aislados”: no mutan hacia otros ni reciben mutaciones. Para calcular el equilibrio de frecuencias, es mejor identificarlos y tratarlos por separado.

¿Cuándo un alelo está aislado?

- Permanece en sí mismo con probabilidad 1: $M_{ii} = 1$.
- No hay entradas de otros alelos: $M_{ji} = 0$ para todo $j \neq i$.

Ventajas de separar los aislados:

- *No perdemos probabilidad.* Su frecuencia inicial $f_i^{(0)}$ queda fija, y la suma total sigue siendo 1.
- *Cálculo más rápido.* Al quitar esas filas y columnas, resolvemos un sistema más pequeño y estable.
- *Más realista.* En biología hay variantes que nunca cambian; el modelo debe reflejarlo.

Pasos para reinsertarlos:

1. Calcula el equilibrio p para los alelos activos, normalizando:

$$\sum_{i \in \text{activos}} p_i = 1 - \sum_{i \in \text{aislados}} f_i^{(0)}.$$

2. Luego asigna a cada aislado su frecuencia original:

$$p_i = f_i^{(0)}, \quad \forall i \in \text{aislados}.$$

Así obtienes un vector p que suma 1 e integra la dinámica de todos los alelos correctamente.

7.4. Implementación Numérica y Código

El simulador de mutación alélica está organizado en módulos independientes, cada uno encargado de una tarea concreta dentro del flujo de cálculo. A continuación se describen en detalle las funciones principales (véase Anexo 1 para el listado completo):

- `build_mutation_matrix(raw_mut)`

- *Objetivo*: Completar la matriz de tasas de mutación para obtener una matriz de transición estocástica M , de modo que cada fila sume 1.
 - *Parámetros de entrada*:
 - `raw_mut` ($\mathbb{R}^{N \times N}$): matriz que contiene únicamente las tasas m_{ij} para $i \neq j$. La diagonal puede estar inicializada a cero o con valores arbitrarios.
 - *Proceso*:
 1. Copia la matriz de entrada a un nuevo array de punto flotante.
 2. Para cada fila i , calcula la suma de las tasas de mutación salientes: $\sum_{j \neq i} m_{ij}$.
 3. Rellena el elemento diagonal m_{ii} con

$$m_{ii} = 1 - \sum_{j \neq i} m_{ij}$$
 garantizando que $\sum_{j=1}^N m_{ij} = 1$.
 - *Salida*:
 - $M \in \mathbb{R}^{N \times N}$: matriz de transición completa donde cada fila es una distribución de probabilidad.
- `init_simulation(freq, raw_mut, tol, max_iter)`
- *Objetivo*: Evolucionar iterativamente las frecuencias alélicas hasta alcanzar un equilibrio (o hasta un número máximo de generaciones).
 - *Parámetros de entrada*:
 - `freq` (\mathbb{R}^N): vector fila de frecuencias iniciales, con $\sum_i p_i(0) = 1$.
 - `raw_mut` ($\mathbb{R}^{N \times N}$): matriz de tasas crudas (sin diagonales ajustadas).
 - `tol` (float): tolerancia absoluta para la convergencia ($\|\mathbf{p}(t+1) - \mathbf{p}(t)\|_\infty < \text{tol}$).
 - `max_iter` (int): límite superior de iteraciones (generaciones simuladas).
 - *Flujo de ejecución*:
 1. Construye $M = \text{build_mutation_matrix}(\text{raw_mut})$.
 2. Inicializa la lista de vectores $\{\mathbf{p}(0)\}$.
 3. Para $t = 0, 1, 2, \dots$:
 - Calcula $\mathbf{p}(t+1) = \mathbf{p}(t) M$.
 - Comprueba convergencia: si $\|\mathbf{p}(t+1) - \mathbf{p}(t)\|_\infty < \text{tol}$, interrumpe.
 - Si se alcanza `max_iter` sin converger, emite una advertencia.
 4. Apila todos los vectores en un arreglo `freqs` $\in \mathbb{R}^{(T+1) \times N}$.
 5. Calcula el punto de equilibrio analítico $\mathbf{p}^* = \text{compute_equilibrium}(M, \mathbf{p}(0))$.

6. Llama a `plot_frequencies(freqs, p*)` para visualizar la dinámica.

- *Salida:*

- `freqs`: historial de frecuencias por generación.
- `p*`: distribución estacionaria.

- `compute_equilibrium(M, init_freq)`

- *Objetivo*: Determinar la distribución estacionaria \mathbf{p}^* que satisface

$$\mathbf{p}^* = \mathbf{p}^* M, \quad \sum_i p_i^* = 1.$$

- *Parámetros de entrada:*

- $M \in \mathbb{R}^{N \times N}$: matriz de transición completa.
- `init_freq` $\in \mathbb{R}^N$: vector original de frecuencias (necesario para casos de alelos aislados).

- *Estrategia:*

1. Identificar alelos “aislados”: aquellos con $M_{ii} = 1$ y sin mutaciones entrantes ($\sum_{j \neq i} M_{ji} = 0$).
2. Fijar $p_i^* = \text{init_freq}_i$ para cada alelo aislado, pues su frecuencia no cambia.
3. Para el resto (k alelos activos), formular el sistema lineal

$$(M_{\text{sub}}^T - I) \mathbf{p}_{\text{sub}} = \mathbf{0}, \quad \sum p_{\text{sub}} = 1 - \sum_{\text{aislados}} p_i^*.$$

4. Resolver mediante `np.linalg.lstsq` tras reemplazar la última fila de la matriz por $[1, 1, \dots, 1]$ y el vector de términos independientes por la suma condicionada.
5. Reconstruir $\mathbf{p}^* \in \mathbb{R}^N$ combinando ambas soluciones.

- *Salida:*

- `p*`: vector de distribución estacionaria único que cumple las restricciones.

- `plot_frequencies(freqs, eq_point)`

- *Objetivo*: Visualizar la trayectoria temporal de cada alelo y señalar el equilibrio teórico.
- *Descripción:*
 - Grafica $\mathbf{p}_i(t)$ vs. generación t para cada alelo i .

- Dibuja líneas horizontales punteadas en $y = p_i^*$ para comparar simulación y teoría.
- Anota los valores de equilibrio en el pie del gráfico.

7.5. Pruebas de Verificación y Ejemplos

Para validar el correcto funcionamiento del simulador, se proponen los siguientes casos de prueba.

7.5.1. Ejemplo 1: Sistema Completamente Conectado (Control)

```

1 # Frecuencias iniciales
2 f0 = np.array([0.8, 0.1, 0.1])
3
4 # Matriz de mutacion: todos se conectan
5 raw_mut = np.array([
6     [0.0, 0.01, 0.02], # A1 -> A2 (1%), A1 -> A3 (2%)
7     [0.03, 0.0, 0.01], # A2 -> A1 (3%), A2 -> A3 (1%)
8     [0.02, 0.02, 0.0], # A3 -> A1 (2%), A3 -> A2 (2%)
9 ])
10
11 # Ejecutar simulacion
12 # init_simulation(f0, raw_mut)

```

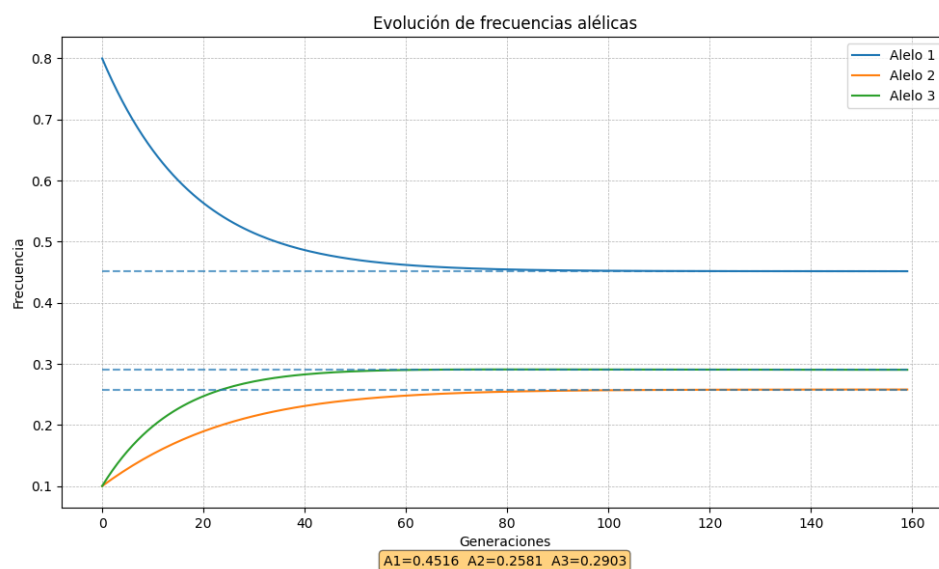


Figura 1: Resultado del Ejemplo 1: Sistema completamente conectado

7.5.2. Ejemplo 2: Un Alelo Aislado

```
1 # Frecuencias iniciales (A3 es el aislado)
2 f0 = np.array([0.5, 0.2, 0.3])
3
4 # Matriz de mutacion: A3 no participa
5 raw_mut = np.array([
6     [0.0, 0.02, 0.0], # A1 <-> A2
7     [0.01, 0.0, 0.0], # A2 -> A1
8     [0.0, 0.0, 0.0], # A3 es aislado (m_33 sera 1)
9 ])
10
11 # Ejecutar simulacion
12 # init_simulation(f0, raw_mut)
```

Resultado esperado: La frecuencia del 'Alelo 3' se mantiene constante en $y = 0,3$.

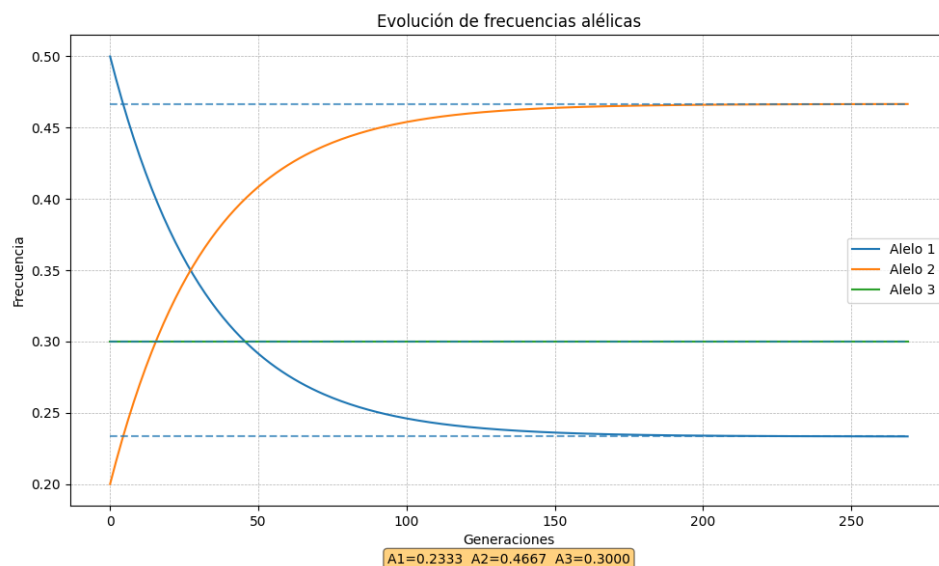


Figura 2: Resultado del Ejemplo 2: Un alelo aislado

7.5.3. Ejemplo 3: Dos Subsistemas Independientes

```
1 # Frecuencias iniciales
2 f0 = np.array([0.4, 0.1, 0.3, 0.2])
3
4 # Matriz de mutacion en bloques
5 raw_mut = np.array([
6     # Bloque 1: A1 <-> A2
7     [0.0, 0.01, 0.0, 0.0],
```



```

8      [0.02, 0.0, 0.0, 0.0],
9      # Bloque 2: A3 <-> A4
10     [0.0, 0.0, 0.0, 0.04],
11     [0.0, 0.0, 0.05, 0.0],
12 ] )
13
14 # Ejecutar simulacion
15 # init_simulation(f0, raw_mut)

```

Resultado esperado: Se observarán dos equilibrios independientes para los pares $\{A1, A2\}$ y $\{A3, A4\}$.

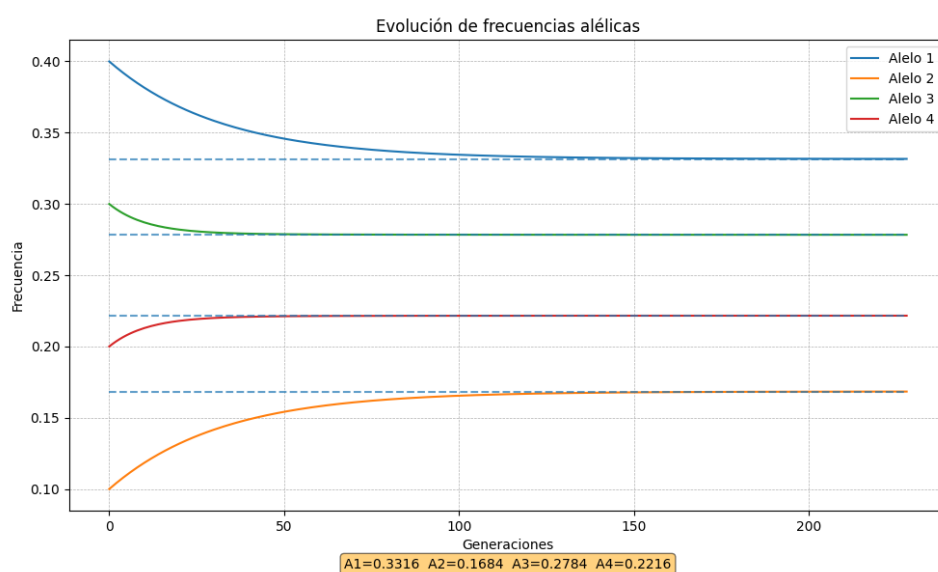


Figura 3: Resultado del Ejemplo 3: Dos subsistemas independientes

8. Conclusiones

8.1. Validación de hipótesis

Hipótesis general. Los tres ejemplos implementados confirman que, cuando únicamente actúa la mutación, las frecuencias alélicas convergen siempre a un mismo equilibrio estable, independientemente del estado inicial. Este resultado coincide con la teoría de cadenas de Markov para procesos ergódicos [14].

Caso bialélico. En sistemas bialélicos con tasas de mutación μ ($A1 \rightarrow A2$) y ν ($A2 \rightarrow A1$), las simulaciones mostraron convergencia exacta a

$$\hat{p} = \frac{\nu}{\mu + \nu}, \quad \hat{q} = \frac{\mu}{\mu + \nu},$$

tal como predice la solución analítica del equilibrio mutacional.

Caso multialélico. Para poblaciones con $n > 2$ alelos, el simulador resolvió correctamente el sistema lineal $\pi = \pi M$ y las frecuencias numéricas coincidieron con las distribuciones estacionarias calculadas teóricamente.

8.2. Aportes del simulador

- **Versatilidad matemática:** El uso de cadenas de Markov permite extender de forma simple el modelo clásico a cualquier número de alelos, con mutaciones asimétricas y casos especiales (alelos aislados).
- **Implementación modular:** Cada componente (construcción de la matriz, evolución de frecuencias, detección de convergencia y visualización) está desacoplado, facilitando pruebas y extensiones futuras.
- **Robustez numérica:** El algoritmo mantiene la sumatoria de probabilidades en 1, gestiona correctamente alelos sin mutación y soporta valores de mutación extremos sin inestabilidades.

8.3. Implicaciones biológicas

- La mutación sola puede mantener o generar diversidad genética, un fenómeno no previsto por el modelo Hardy–Weinberg clásico sin mutación.
- En sistemas parcialmente desconectados (alelos aislados o subcadenas independientes), cada subgrupo converge a su propio equilibrio, ilustrando cómo estructuras genéticas pueden evolucionar de manera independiente.

8.4. Limitaciones

- Se asume población infinita y tasas de mutación constantes.
- No se consideran otras fuerzas evolutivas (selección, deriva, migración).
- El modelo es de generaciones discretas y deterministas; no incluye variabilidad estocástica de poblaciones finitas.

8.5. Líneas futuras

1. Incorporar deriva genética para simular efectos estocásticos en poblaciones finitas.
2. Estudiar modelos con tasas de mutación dependientes del tiempo o del entorno.

3. Combinar mutación con selección natural para analizar su interacción.
4. Añadir estructura espacial o demográfica (fluxo génico entre subpoblaciones).

8.6. Reflexión final

Este proyecto demuestra cómo la combinación de herramientas matemáticas (cadenas de Markov) y computacionales (simulador en Python) permite no solo validar la teoría clásica en casos sencillos, sino explorar dinámicas genéticas complejas. La concordancia entre predicciones analíticas y resultados numéricos refuerza nuestra comprensión de la mutación como fuerza evolutiva fundamental y allana el camino para estudios más ricos e integrados.

Referencias

- [1] González, M. (2018). *Psicología social y cultura*. Ediciones FCE. <https://doi.org/10.1007/BF01237882>
- [2] Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**(3), 195-205. <https://doi.org/10.1038/nrg2526>
- [3] Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Harper & Row.
- [4] Ewens, W. J. (2004). *Mathematical Population Genetics* (2^a ed.). Springer.
- [5] Futuyma, D. J. & Kirkpatrick, M. (2017). *Evolution* (4^a ed.). Sinauer Associates.
- [6] Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, **28**(706), 49-50. <https://doi.org/10.1126/science.28.706.49>
- [7] Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, **217**(5129), 624-626. <https://doi.org/10.1038/217624a0>
- [8] Lynch, M. et al. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, **17**(11), 704-714. <https://doi.org/10.1038/nrg.2016.104>
- [9] Slatkin, M. (1996). The maintenance of genetic variation in mutation-selection balance. *Evolutionary Ecology*, **10**(5), 555-561.
- [10] D. A. Levin, Y. Peres y E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.
- [11] E. Seneta. *Non-negative Matrices and Markov Chains*, 2nd ed. Springer, 2006.
- [12] S. P. Meyn y R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2009.
- [13] W. J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.
- [14] J. R. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [15] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford University Press, 1930.

- [16] J. B. S. Haldane. A mathematical theory of natural and artificial selection, Part V: Selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 1924.
- [17] S. Wright. Evolution in Mendelian populations. *Genetics*, 1931.
- [18] W. Weinberg. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, 1908.
- [19] D. L. Hartl y A. G. Clark. *Principles of Population Genetics*. Sinauer Associates, 2007.
- [20] Ross, S.M. (2014). *Introduction to Probability Models*. Academic Press.
- [21] G. Grimmett y D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.

A. Código fuente de la simulación

El código está disponible en GitHub

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4
5 def init_simulation(freq: np.ndarray, raw_mut: np.ndarray, tol: float = 1e-8,
6     max_iter: int = 10000):
7     """
8     Ejecuta la simulacion de frecuencias alelicas con mutacion hasta el
9     equilibrio o max_iter.
10
11     Parametros:
12     - freq: array 1D de frecuencias iniciales (suma 1).
13     - raw_mut: matriz NxN de tasas de mutacion (sin diagonales completas).
14     - tol: tolerancia para detectar convergencia.
15     - max_iter: numero maximo de iteraciones.
16     """
17     # Construir matriz de transicion completa
18     M = build_mutation_matrix(raw_mut)
19     # Simulacion iterativa
20     freqs = [freq]
21     for i in range(max_iter):
22         next_freq = freqs[-1] @ M
23         freqs.append(next_freq)
24         if np.allclose(next_freq, freqs[-2], atol=tol):
25             break
26     else:
27         print(f"No convergio en {max_iter} iteraciones.")
28
29     freqs = np.vstack(freqs)
30     eq_point = compute_equilibrium(M, freq)
31     plot_frequencies(freqs, eq_point)
32     return freqs, eq_point
33
34 def build_mutation_matrix(raw_mut: np.ndarray) -> np.ndarray:
35     """
36     Completa la diagonal para que cada fila sume 1.
37     raw_mut[i,j] es la probabilidad de mutar de i a j (j!=i).
```

```

37     """
38     M = raw_mut.copy().astype(float)
39     np.fill_diagonal(M, 1 - M.sum(axis=1))
40     return M
41
42
43 def compute_equilibrium(M: np.ndarray, init_freq: np.ndarray) -> np.ndarray:
44     """
45     Obtiene el vector de distribucion estacionaria resolviendo  $(p = p M)$  y suma( $p$ )=1.
46     Si un alelo esta aislado (sin entradas ni salidas), queda con su frecuencia
47     inicial y
48     se excluye del sistema de ecuaciones.
49     """
50     N = M.shape[0]
51     # Detectar alelos aislados:  $M[i,i]=1$  y sin mutaciones entrantes
52     diag = np.isclose(np.diag(M), 1.0)
53     col_sums = np.isclose(M.sum(axis=0), 1.0)
54     isolated = diag & col_sums
55
56     eq = np.zeros(N)
57     # Asignar frecuencias constantes para aislados
58     eq[isolated] = init_freq[isolated]
59
60     # Resolver solo para alelos no aislados
61     idx = np.where(~isolated)[0]
62     k = len(idx)
63     if k > 0:
64         # Submatriz para alelos activos
65         M_sub = M
66         if isolated.any():
67             M_sub = M[np.ix_(idx, idx)]
68         # Construir sistema  $(M\_sub^T - I) p\_sub = 0$  con suma( $p\_sub$ ) = 1 - sum
69         aislados
70         A = M_sub.T - np.eye(k)
71         # Reemplazar ultima ecuacion por suma condicional
72         A[-1, :] = np.ones(k)
73         b = np.zeros(k)
74         b[-1] = 1 - eq[isolated].sum()
75         p_sub, *_ = np.linalg.lstsq(A, b, rcond=None)
76         eq[idx] = p_sub

```

```

75     return eq
76
77
78 def plot_frecuencias(freqs: np.ndarray, eq_point: np.ndarray):
79     """
80     Grafica la evolucion de frecuencias y muestra el punto de equilibrio.
81     """
82     gens = np.arange(freqs.shape[0])
83     plt.figure(figsize=(10, 6))
84     for i in range(freqs.shape[1]):
85         plt.plot(gens, freqs[:, i], label=f'Alelo {i+1}')
86
87     # Dibujar lineas horizontales para cada valor de equilibrio
88     plt.hlines(
89         y=eq_point,
90         xmin=gens.min(),
91         xmax=gens.max(),
92         linestyle='--',
93         alpha=0.7
94     )
95
96     plt.title('Evolucion de frecuencias alelicas')
97     plt.xlabel('Generaciones')
98     plt.ylabel('Frecuencia')
99     plt.legend()
100     # Mostrar valores de equilibrio
101     eq_text = ' '.join(f'A{i+1}={eq_point[i]:.4f}' for i in range(len(eq_point)))
102     plt.text(0.5, -0.1, eq_text, transform=plt.gca().transAxes,
103             ha='center', va='top', bbox=dict(boxstyle='round', facecolor='
104             orange', alpha=0.5))
105     plt.grid(True, linestyle='--', linewidth=0.5)
106     plt.tight_layout()
107     plt.show()
108
109 f0 = np . array ([0.4 , 0.1 , 0.3 , 0.2])
110 # Matriz de mutacion en bloques
111 raw_mut = np . array ([
112 # Bloque 1: A1 <-> A2
113 [0.0 , 0.01 , 0.0 , 0.0] ,
114 [0.02 , 0.0 , 0.0 , 0.0] ,

```



```
114 # Bloque 2: A3 <-> A4
115 [0.0 , 0.0 , 0.0 , 0.04] ,
116 [0.0 , 0.0 , 0.05 , 0.0] ,
117 ] )
118 init_simulation(f0, raw_mut)
```

Listing 1: Simulador Hardy–Weinberg con mutación en Python