



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍA
CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS

ESTIMACIÓN DE LA TASA DE MUTACIÓN HACIA FENOTIPOS MUTADORES

T E S I S

QUE PARA OBTENER EL TÍTULO DE

MAESTRO EN MATEMÁTICAS

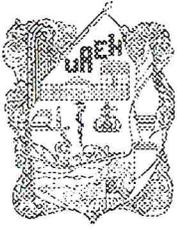
P R E S E N T A

ISAAC VÁZQUEZ MENDOZA

BAJO LA DIRECCIÓN DE

DRA. ERIKA ELIZABETH RODRÍGUEZ TORRES
DR. PHILIP JOHN GERRISH

MINERAL DE LA REFORMA, HIDALGO
ABRIL 2024



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

Instituto de Ciencias Básicas e Ingeniería

School of Engineering and Basic Sciences

Área Académica de Matemáticas y Física

Department of Physics and Mathematics

Mineral de la Reforma, Hgo., a 29 de enero de 2024

Número de control: ICBI-AAMyF/125/2024

Asunto: Autorización de impresión de tesis.

MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO
DIRECTOR DE ADMINISTRACIÓN ESCOLAR DE LA UAEH

El Comité Tutorial de la tesis titulada **“Estimación de la tasa de mutación hacia fenotipos mutadores”**, realizada por el sustentante **Isaac Vázquez Mendoza**, con número de cuenta **296000**, perteneciente a la **Maestría en Matemáticas**, una vez que ha revisado, analizado y evaluado el documento recepcional de acuerdo a lo estipulado en el Artículo 110 del Reglamento de Estudios de Posgrado, tiene a bien extender la presente:

AUTORIZACIÓN DE IMPRESIÓN

Por lo que el sustentante deberá cumplir los requisitos del Reglamento de Estudios de Posgrado y con lo establecido en el proceso de grado vigente.

Atentamente
“Amor, Orden y Progreso”

El Comité Tutorial

Dra. Erika Elizabeth Rodríguez Torres
Directora

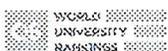
Dr. Phillip Gerrish
Codirector

Dr. Benjamín Alfonso Itzá Ortiz
Miembro del comité

Dr. Roberto Ávila Pozos
Miembro del comité



Ciudad del Conocimiento
Carretera Panuapa-Tulancingo km 4.5 Colonia
Carboneras, Mineral de la Reforma, Hidalgo,
México. C.P. 42184
Teléfono: +52 (771) 71720 05 ext. 2531
amayt_icbi@uaeh.edu.mx



Resumen

En este trabajo presentamos un par de modelos nuevos para simular la dinámica fenotípica en células bajo la hipótesis del fenotipo mutador. Ambos modelos dependen fuertemente del estudio del proceso estocástico de Yule-Furry y de la ley de los pequeños números. Aunado a lo anterior, el denominado modelo gráfico, recurre al concepto de distribución impropia para su consolidación; mientras que, el segundo modelo recurre al método de inversión y por tanto le denominamos modelo basado en el método de inversión.

Asimismo, presentamos un par de metodologías para estimar la tasa de mutación hacia fenotipos mutadores. Por una parte, la primera se basa en el método de mínimos cuadrados y, por otra parte, la segunda combina la transformada rápida de Fourier con los métodos de máxima verosimilitud.

A lo largo de este trabajo nos dedicamos a discutir la justificación, confiabilidad y alcances de los modelos y métodos de estimación presentados. En específico, se verifica que el modelo gráfico y el modelo basado en el método de inversión son ambos comparables con el modelo clásico y más ampliamente utilizado, el modelo Moran. Además, empleamos los métodos de estimación y al conjuntarlos con técnicas de bootstrap, jackknife y verosimilitud construimos intervalos de confianza para las estimaciones de la tasa de mutación hacia fenotipos mutadores.

Abstract

In this work, we present a pair of new models for simulating phenotypic dynamics in cells under the hypothesis of the mutator phenotype. Both models rely on the study of the Yule-Furry stochastic process and the law of small numbers. In addition to this, the graphical model resorts to the concept of improper distribution for its consolidation, while the second model relies on the inversion method and is therefore referred to as the model based on the inversion method.

Furthermore, we introduce a pair of methodologies for estimating the mutation rate to mutator phenotypes. On one hand, the first methodology is based on the least squares method, and, on the other hand, the second methodology combines the fast Fourier transform with maximum likelihood methods.

Along this work, we discuss the justification, reliability, and scope of the presented models and estimation methods. Specifically, it is verified that both the graphical model and the model based on the inversion method are comparable to the classic and widely used Moran model. Furthermore, we employ the estimation methods in conjunction with techniques such as bootstrap, jackknife, and likelihood to construct confidence intervals for the estimations of the mutation rate to mutator phenotypes.

Agradecimientos

Expreso mi completa gratitud al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) por haberme brindado la Beca Nacional para Estudios de Posgrado (No. CVU 1177174) a lo largo de mis estudios de maestría. El apoyo del CONAHCyT fue crucial para el desarrollo y culminación de este trabajo, mismo que me permitió crecer profesional, intelectual, personal y académicamente.

Asimismo, reconozco el papel tan fundamental que ha tenido la Universidad Autónoma del Estado de Hidalgo (UAEH) en mi vida desde la licenciatura y ahora en la maestría, al proporcionar un ambiente donde el respeto, la diversidad y el diálogo promueven el desarrollo humano y la colaboración científica. La Maestría en Matemáticas es un ejemplo tangible de lo antes mencionado y, por tanto, externo mi gratitud a las profesoras y profesores del cuerpo académico por compartir su experiencia en cada clase, seminario y conversación cotidiana.

Estoy agradecido con la Dra. Erika Elizabeth Rodríguez Torres por reconocer mi trabajo aquel mes de septiembre de 2021 e invitarme a formar parte de la maestría bajo su tutoría. Su apoyo y visión de la matemáticas me motivaron a explorar distintos paradigmas, los cuales me han aportado una mayor madurez matemática, la cual me ha mostrado la belleza, importancia y alcances de la interdisciplinariedad.

Quiero agradecer al Dr. Philip John Gerrish por haber compartido su tiempo, experiencia y apoyo conmigo. Al mostrarme el lado aplicado de las matemáticas *risueñas*, confirmé mi predilección por ellas. Además, le agradezco la confianza brindada, la cual me motivó a incursionar en nuevos horizontes personales, académicos y geográficos. Me siento completamente afortunado por haberle conocido; me impulsó a llegar cada vez más lejos, pese a las adversidades. Admiro su sencillez, carisma, paciencia, tenacidad y capacidad intelectual, y anhelo poder llegar tan lejos como él.

De igual manera, agradezco al resto del comité sinodal, el Dr. Benjamín Alfonso Itzá

Ortiz y el Dr. Roberto Ávila Pozos, sus notas, comentarios y observaciones fueron grandes aportaciones, las cuales pulieron notablemente este trabajo para cumplir con los estándares de calidad que distinguen a los egresados de la de la LIMA y de la Maestría en Matemáticas.

Aprovecho para agradecer al Dr. Rafael Villarroel Flores y al Dr. Jorge Viveros Rogel por su labor docente, misma que me ayudó a reflexionar sobre ideas matemáticas abstractas desde una perspectiva más completa, permitiéndome desarrollar un genuino interés en explorar y compartir las ideas discutidas en sus clases.

Me siento profundamente afortunado por haber conocido a tantas personas que me ayudaron a lo largo de estos años. Destaco el apoyo brindado por la Dra. Lindi Wahl y la M. en C. Mojgan Ezadian, su colaboración fue crucial para el desarrollo y la publicación de este trabajo. Agradezco al Dr. Francisco Javier Martínez Farías y al Dr. Pedro Amado Miranda Romagnoli por prestarme sus servidores para poder correr las simulaciones.

Asimismo, agradezco a la Society for the Study of the Evolution (SSE) por brindarme por su apoyo financiero a través de la Dra. Hanna Kokko y la Dra. Courtney Murren, para asistir y presentar parte este trabajo en la Evolution Meeting 2023. Mi participación en este congreso me motivó a continuar con mi educación y se lo debo a las personas que conocí ahí como el Dr. Stephen Proulx, Yadira García, Saúl Domínguez, Daniela Mardones y Xuening Zhang.

Reitero mi gratitud a quienes directa o indirectamente me ayudaron a lo largo de estos años para mejorar mi perfil profesional, mantener mi cabeza en orden, escucharme en los malos días y disfrutar de los días buenos como la Dra. Tracy Koslowski, el Dr. Rubén Alejandro Martínez Avendaño, la Dra. Liliana Peralta Hernández, Mauricio Islas Gómez, Marlene Godínez Galicia, Sonia Arely Sánchez González y Mónica Villarroel Ramírez.

Por último, le agradezco a mi familia y amigos, su apoyo me impulsó día con día para concluir satisfactoriamente esta etapa. En particular, quiero destacar el apoyo incondicional de mi mamá, Margarita, quien siempre ha creído en mí. Agradezco también a mi hermano Enrique por su apoyo profesional constante, a mi cuñada Carol por recordarme la importancia de la investigación en el cáncer, y a mi amiga Ciomara, cuya determinación y perseverancia en la consecución de sus metas son una constante inspiración para mí.

[...] the butterfly came and landed
right on the tip of his nose.

Miley Cyrus

Índice general

Resumen	II
Abstract	III
Introducción	1
1. Planteamiento del problema	4
1.1. Justificación	4
1.2. Pregunta de investigación	6
1.3. Objetivo general	6
1.4. Objetivos particulares	6
2. Preliminares	8
2.1. Conceptos generales de probabilidad	8
2.2. Conceptos generales de procesos estocásticos	17
3. Carcinogénesis	21
3.1. Mecanismos internos contra el cáncer	25

<i>ÍNDICE GENERAL</i>	VIII
4. Modelos para la dinámica fenotípica	34
4.1. Modelo gráfico	36
4.1.1. Proceso de Yule-Furry	36
4.1.2. Crecimiento poblacional	54
4.1.3. Algoritmo	57
4.2. Modelo basado en el método de inversión	71
4.2.1. Método de inversión	71
4.2.2. Algoritmo	75
5. Métodos de estimación	86
5.1. Distribución Luria-Delbrück	86
5.2. Ajuste directo	92
5.3. Ajuste por métodos de máxima verosimilitud	94
6. Resultados	96
6.1. Confiabilidad de los modelos	96
6.2. Confiabilidad de las estimaciones	100
7. Conclusiones	105
Referencias	107
A. Códigos en R	119
A.1. Modelo Moran	119

ÍNDICE GENERAL

A.2. Modelo gráfico	121
A.3. Modelo basado en el método de inversión	125
A.4. Estimación por ajuste directo	128
A.5. Estimación por métodos de máxima verosimilitud en tres dimensiones . . .	133
A.6. Estimación por métodos de máxima verosimilitud en dos dimensiones . . .	137

Introducción

La evolución es un proceso continuo y perpetuo el cual ha permitido, partiendo de organismos unicelulares, la aparición de relaciones colaborativas en organismos unicelulares lo que, a su vez, conllevó a la formación de formas de vida con mayor complejidad, como los organismos pluricelulares.

Ahora bien, como mencionamos, la aparición de los organismos pluricelulares sucede como consecuencia de la evolución y una componente crucial para que la evolución tenga lugar son las mutaciones, donde todo cambio en el material genético es considerado como una mutación [114]. Dichos cambios ocurren aleatoriamente y manera cotidiana [125].

Cada célula puede presentar mutaciones a lo largo de su vida y esto es esperado, pero la tasa con la que estas mutaciones aparecen, conocida con tasa de mutación, es sumamente pequeña. Luego, los organismos poseen diversos mecanismos para lidiar con las mutaciones por lo que, en general, la mayoría de las mutaciones no representan problemas para el organismo [45].

Con lo anterior se plantea que la dinámica fenotípica en una población celular se reduce a considerar la subpoblación de células silvestres, la subpoblación de células mutantes y las relaciones entre ellas; por ejemplo, las células silvestres dan origen tanto a células mutantes como a otras células silvestres; mientras que, las células mutantes dan origen exclusivamente a más células mutantes.

El paradigma que discutimos previamente era ampliamente aceptado hasta la publicación de Loeb *et al.* (1974) [87] en donde proponen la hipótesis del fenotipo mutador. Esta hipótesis hace referencia a la existencia de un tercer tipo de célula: las células mutadoras, las cuales se caracterizan por tener una tasa de mutación más alta en comparación con el resto de las células de la población.

En este trabajo nos proponemos estimar la tasa de mutación hacia fenotipos mutadores

ya que la presencia de las células mutadoras tiene como implicación que la aparición de mutaciones se favorece ocasionalmente lo cual resulta muy relevante ya que al haber más mutaciones se incrementa la probabilidad de que las mutaciones producidas representen algún problema en el organismo [51]; por ejemplo, bacterias resistentes a medicamentos, células con inestabilidad genética o mayor riesgo de desarrollar cáncer.

No obstante, la identificación de células mutadoras es, en general, un problema actualmente dado que no es posible distinguir a estas células de las células silvestres empleando las técnicas de laboratorio estándar desarrolladas hasta ahora [61]. Por otro lado, las células con alguna mutación sí pueden observarse con biomarcadores, por ejemplo. Es decir, la identificación se lleva a cabo a través del uso de algunas sustancias químicas, las cuales se infiltran al interior de las células y generan una reacción, como una tinción o luminiscencia, solo cuando las células poseen la característica buscada, en este caso, mutaciones.

Tomando ventaja de lo anterior desarrollamos modelos computacionales que emulan la reproducción celular e incorporar la aparición de mutaciones y, de esta manera, empleamos la función generadora de probabilidad empírica [97]

$$\mathfrak{F}_n(t) = \frac{1}{n} \sum_{i=1}^n t^{X_i},$$

con $t \in [0, 1]$ y $\{X_i\}_{i=1}^n$ una muestra aleatoria de tamaño $n \in \mathbb{N}$ proveniente de una variable aleatoria discreta X , para hacer inferencia sobre la presencia de células mutadoras a través del número de mutaciones que se obtuvieron en la simulaciones.

Ahora, a nivel celular podemos hallar grandes diferencias respecto a los organismos unicelulares y los organismos pluricelulares. En el caso de los organismos unicelulares están compuestos por una única célula que es capaz de realizar, de manera autónoma, diversas tareas; por ejemplo, procesar su propio alimento, responder a estímulos, reproducirse y, en específico, reproducirse tanto como el ambiente se lo permita.

Por otra parte, los organismos pluricelulares se componen por más de una célula y las células que componen a los organismos pluricelulares, en contraste con los organismos unicelulares, se organizan de forma colaborativa por lo que se han especializado a fin de preservar con vida al organismo a quien componen. En particular, lo anterior implica que las células de un organismo pluricelular están sujetas a un control reproductivo más estricto en comparación con los organismos unicelulares con el objetivo de mantener la fisionomía, salud y, en general, la homeostasis del organismo.

Cuando este control reproductivo de las células de un organismo pluricelular falla,

se está ante una amenaza de desarrollar cáncer. Así, finalizamos extrapolando nuestros resultados al contexto del cáncer, una condición que aparece a la par de los organismos pluricelulares, es exclusiva de ellos y resulta más agresiva con la aparición de mutadores [98]. Con esto, potencialmente, se lograría optimizar los tratamientos contra el cáncer puesto que, al contar con predicciones sobre si los mutadores están por aparecer o si ya han aparecido, podrían proponerse algunas modificaciones a lo largo de un tratamiento a fin de garantizar y mantener su utilidad para cada paciente.

Capítulo 1

Planteamiento del problema

1.1. Justificación

La existencia de células mutadoras en el organismo representa un tema de gran relevancia y plantea implicaciones clínicas potencialmente graves. Estas células, poseen una capacidad intrínseca que supone un mayor riesgo de encontrar mutaciones en una población celular, planteando serias preocupaciones en el ámbito de la salud. Sus efectos pueden abordarse desde diferentes perspectivas, pero una de las más notables es su papel en el surgimiento de bacterias resistentes a los medicamentos [36], lo que a su vez conlleva un riesgo latente en caso de infecciones bacterianas.

La resistencia a los antibióticos, un fenómeno de creciente preocupación actual [3], es un ejemplo destacado de cómo las células mutadoras pueden desencadenar problemas significativos. Cuando un organismo se enfrenta a una infección bacteriana, la elección de un tratamiento eficaz es esencial. Sin embargo, la aparición de linajes de bacterias resistentes a los medicamentos pueden obstaculizar, e incluso impedir, la respuesta de los tratamientos prescritos. Esto puede conducir a infecciones más persistentes y graves. Así, la presencia de células mutadoras, que favorecen la aparición de tales mutaciones, se convierte en un factor crítico que afecta directamente la capacidad de la medicina para combatir las infecciones.

Además de las implicaciones en términos de resistencia bacteriana, las mutaciones genéticas inducidas por las células mutadoras tienen consecuencias que van más allá. Pueden generar problemas en el organismo, como el cáncer o la inestabilidad genética, la cual afecta la aptitud de las células [98]. El cáncer, está caracterizado por la proliferación

descontrolada de células anómalas, a menudo se origina en mutaciones genéticas. Por lo tanto, es fundamental modelar y estudiar las tasas de mutación hacia fenotipos mutadores en el organismo, ya que esto puede ayudar a prevenir y tratar de manera más eficaz esta enfermedad.

Dentro de este contexto, la estimación de la tasa de mutación hacia fenotipos mutadores emerge como una tarea de vital importancia. A pesar de que existen investigaciones que se centran en fenotipos mutadores específicos, el desarrollo de métodos que puedan aplicarse de manera universal a todo tipo de fenotipos mutadores, representa una herramienta matemática de gran alcance. Estos métodos tienen la capacidad de monitorear y predecir la aparición de mutadores en el organismo, lo que podría expandir la forma en que se abordan las enfermedades relacionadas con mutaciones genéticas.

Los resultados de este tipo de investigación no solo tienen un impacto significativo en el contexto de la resistencia a los medicamentos y el cáncer, sino que también pueden abrir nuevas puertas en la prevención y el tratamiento de otras condiciones médicas relacionadas con mutaciones genéticas. Si es posible prever la aparición de mutadores en un paciente, podríamos tomar medidas preventivas para mantenerle estable al reducir drásticamente el riesgo de desarrollar enfermedades graves. Esta capacidad de predicción podría ser un avance crucial en la medicina, permitiendo una atención médica más personalizada.

La identificación temprana de la presencia de células mutadoras nuevas es esencial para el diagnóstico y el tratamiento efectivo del cáncer. Dado que esta enfermedad suele desarrollarse a lo largo del tiempo, el diagnóstico precoz puede marcar una diferencia significativa en las tasas de supervivencia y la calidad de vida de los pacientes. La investigación destinada a estimar la tasa de mutación hacia fenotipos mutadores puede proporcionar a los médicos y a los pacientes las herramientas necesarias para detectar y abordar el cáncer en sus etapas más tempranas.

En última instancia, la presencia de células mutadoras en el organismo humano plantea cuestiones complejas y fundamentales relacionadas con la salud y el bienestar de las personas. La resistencia a los medicamentos y el riesgo de cáncer son solo dos ejemplos de las implicaciones que estas células pueden tener en nuestra vida cotidiana. La investigación en torno a la estimación de la tasa de mutación hacia fenotipos mutadores representa un paso crucial en la comprensión y el control de este fenómeno.

1.2. Pregunta de investigación

En consecuencia de lo planteado anteriormente, a lo largo de este trabajo se pretende contestar a la siguiente pregunta. Partiendo de las metodologías para estimar la tasa de mutación de células silvestres a células mutante, ¿podemos generalizar dichas técnicas y obtener estimaciones confiables de la tasa de mutación de células silvestres a células mutadoras?

Además, se verificará el poder estadístico al construir intervalos de confianza para contrastar las hipótesis estadísticas siguientes.

H_0 : El valor real de la tasa de mutación de células silvestres a células mutadoras está contenido en el intervalo de confianza, o

H_1 : El valor real de la tasa de mutación de células silvestres a células mutadoras no es un elemento del intervalo de confianza.

1.3. Objetivo general

Formular e implementar una metodología estadística apegada a un protocolo experimental viable para estimar la tasa de mutación hacia fenotipos mutadores, probar el poder estadístico de los métodos desarrollados al generar datos *in silico* generados a través de modelos matemáticos fiables; de igual forma, explorar si existen diferencias significativas entre la distribución de probabilidad del número de células mutantes respecto a la distribución de probabilidad del número de células mutantes incluyendo al número de células mutantes con fenotipo mutador.

1.4. Objetivos particulares

- Construir dos modelos que sean capaces de emular la reproducción celular y consideren la aparición de distintos fenotipos.
- Verificar que no hay diferencias significativas de los resultados obtenidos con los modelos planteados y los resultados que pueden obtenerse con modelos clásicos y ampliamente utilizados.

- Desarrollar una metodología basada en la función generadora de probabilidad empírica que permita obtener estimaciones confiables para la tasa de mutación hacia fenotipos mutadores, en general.
- Validar los métodos de estimación propuestos utilizando datos *in silico* y generar regiones de confianza para las estimaciones.

Capítulo 2

Preliminares

En este capítulo presentamos las herramientas y resultados básicos necesarios para el desarrollo de este trabajo. La lectora o el lector familiarizado con los conceptos básicos de la teoría de probabilidad y procesos estocásticos puede comenzar la lectura en el capítulo siguiente. Además, para no desviarnos del objetivo de este trabajo hemos omitido algunos conceptos básicos relativos a teoría de la medida; sin embargo, recomendamos consultar el libro de Kolmogorov [74, Cap. 1-2] para todas aquellas personas que deseen profundizar en esta área en el contexto de probabilidad. Los resultados de probabilidad fueron tomados de libro de Resnick [115], mientras que, los resultados referentes a procesos estocásticos han sido tomados de los libros de Chiang, Ross y Taylor, [35], [118] y [137], respectivamente.

2.1. Conceptos generales de probabilidad

Para poder definir apropiadamente un espacio de probabilidad es conveniente introducir algunos conceptos de teoría de la medida.

Definición 1 (σ -álgebra) Sea $\Omega \neq \emptyset$ un conjunto. Una σ -álgebra \mathcal{F} es una colección de subconjuntos de Ω que satisface las siguientes propiedades:

- (a) $\Omega \in \mathcal{F}$.
- (b) Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.
- (c) Si $\{A_n\}_{n \in \mathbb{N}}$ es una colección numerable de conjuntos en \mathcal{F} , entonces $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

La pareja (Ω, \mathcal{F}) recibe el nombre de **espacio medible**.

Ahora, veamos algunos ejemplos de σ -álgebra.

Ejemplo 1 (σ -álgebra generada) Sea \mathcal{A} una colección de subconjuntos de Ω , entonces

$$\sigma(\mathcal{A}) = \bigcap \{ \mathcal{F} \mid \mathcal{F} \text{ es una } \sigma\text{-álgebra en } \Omega \text{ y } \mathcal{A} \subseteq \mathcal{F} \}$$

es la σ -álgebra generada por \mathcal{A} .

Un caso particular, y muy importante, de una σ -álgebra generada es la σ -álgebra de Borel. Si tomamos

$$\mathcal{A} = \{(a, b) \subseteq \mathbb{R} \mid a, b \in \mathbb{R}, a < b\}.$$

Entonces, la σ -álgebra generada por \mathcal{A} se conoce como σ -álgebra de Borel y se denota por $\mathcal{B}(\mathbb{R})$.

En todo espacio medible, la σ -álgebra nos dice qué conjuntos pueden ser medidos; sin embargo, no se ha fijado una forma de medir, lo que nos lleva a introducir el concepto de medida.

Definición 2 (Medida) Sea (Ω, \mathcal{F}) un espacio medible. Una medida en (Ω, \mathcal{F}) es una función $\mu : \mathcal{F} \rightarrow [0, \infty]$ que satisface:

- (a) $\mu(\emptyset) = 0$.
- (b) $\mu(A) \geq 0$, para todo $A \in \mathcal{F}$.
- (c) Si $\{A_n\}_{n \in \mathbb{N}}$ es una colección numerable de conjuntos disjuntos de \mathcal{F} , entonces

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Es importante recordar que las propiedades que satisfacen los conceptos previos serán omitidas; sin embargo, recomendamos a las personas interesadas consultar [115] y [69].

Obsevación 1 A la terna $(\Omega, \mathcal{F}, \mu)$ se le conoce como **espacio de medida** y $\mu(B)$, con $B \in \mathcal{F}$, como la medida del conjunto B . Luego, si $\mu(\Omega) < \infty$ se dice que μ es una medida finita y, en particular, si $\mu(\Omega) = 1$ se dice que μ es una medida de probabilidad y se denotará como P ; por lo que, en este caso, a $P(B)$ se le denomina como la probabilidad

del evento $B \in \mathcal{F}$. Además, como en este trabajo nos restringimos a emplear medidas de probabilidad, no es necesario hablar de la incorporación del infinito en la recta real.

Finalmente, al espacio de medida (Ω, \mathcal{F}, P) se le conoce como **espacio de probabilidad** y, en este contexto, a Ω se le conoce como **espacio muestral** y denota al conjunto de resultados de un experimento aleatorio.

Los espacios de probabilidad son más interesantes cuando definimos variables aleatorias en ellos, entonces procedamos con su definición.

Definición 3 (Variable aleatoria) Se dice que una función $X : \Omega \rightarrow \mathbb{R}$ es una función medible si $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$, para toda $x \in \mathbb{R}$. Las funciones medibles en un espacio de probabilidad se conocen como variables aleatorias.

Consideremos los espacios medibles (Ω, \mathcal{F}) y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, con $\mathcal{B}(\mathbb{R})$ la σ -álgebra de Borel, y sea X una variable aleatoria. Si $P : \mathcal{F} \rightarrow [0, 1]$ es una medida de probabilidad, diremos que $P_X(B) := P(\{\omega \in \Omega \mid X(\omega) \in B\})$, para todo $B \in \mathcal{B}(\mathbb{R})$, es la medida de probabilidad inducida por X (ver [43]).

En particular, la probabilidad del evento $X^{-1}((-\infty, x])$, para todo $x \in \mathbb{R}$, se llama función de distribución de la variable aleatoria X y la definición formal es la siguiente.

Definición 4 (Función de distribución) Sea X una variable aleatoria. La función de distribución, de X , es una función, $F_X : \mathbb{R} \rightarrow [0, 1]$, que está definida por

$$F_X(x) = P_X(X \leq x), \text{ para toda } x \in \mathbb{R}.$$

Con lo anterior, tenemos que si P_X denota la medida de probabilidad inducida por la variable aleatoria X , entonces hemos construido el espacio de medida $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ y, más fuerte aún, hemos construido un espacio de probabilidad, pues notemos que la función de distribución de X y P_X están relacionadas de la manera siguiente:

$$F_X(x) = P_X((-\infty, x]), \text{ para toda } x \in \mathbb{R}.$$

Ejemplo 2 (Variable aleatoria uniforme) Sea $(a, b) \subset \mathbb{R}$ un intervalo. Consideremos a la función de distribución $F : \mathbb{R} \rightarrow [0, 1]$, definida por

$$F(x) = \begin{cases} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } x \in (a, b) \\ 1, & \text{si } x \geq b. \end{cases} \quad (2.1)$$

La variable aleatoria X , inducida por F , es conocida como variable uniforme continua con parámetros a y b .

Diremos que una variable aleatoria X tiene distribución F , y en símbolos lo denotamos como $X \sim F$, si X es inducida por F ; así, si F es como en la ecuación (2.1) es frecuente llamarla distribución uniforme y para expresar que una variable aleatoria tiene dicha distribución se escribe $X \sim \text{Uniforme}(a, b)$.

Una vez aclarada la importancia de la función de distribución, estudiaremos algunas de sus propiedades, las cuales son fundamentales para el estudio de la probabilidad.

Teorema 1 Sea $F(x) = P(X \leq x)$ una función de distribución. Entonces se satisfacen las siguientes propiedades:

- (a) $\lim_{x \rightarrow \infty} F(x) = 1$.
- (b) $\lim_{x \rightarrow -\infty} F(x) = 0$.
- (c) Si $x \leq y$, entonces $F(x) \leq F(y)$.
- (d) F es continua por la derecha, es decir, si $h > 0$ y $F(x^+) := \lim_{h \rightarrow 0} F(x + h)$, entonces $F(x^+) = F(x)$.
- (e) Si $h > 0$ y $F(x^-) := \lim_{h \rightarrow 0} F(x - h)$, entonces $F(x^-) = P(X < x)$.

Demostración.

- (a) Sea $a = \{a_n\}_{n \in \mathbb{N}}$ una sucesión creciente de números reales tal que $\lim_{n \rightarrow \infty} a_n = \infty$. Sea $A = \{A_n\}_{n \in \mathbb{N}}$ una sucesión de eventos con $A_n = (X \leq a_n)$, para toda $n \in \mathbb{N}$. Como la sucesión a es creciente, entonces la sucesión A lo es; además, como en el límite de la sucesión a no converge, entonces

$$\lim_{n \rightarrow \infty} A_n = \Omega.$$

Así, por la continuidad de la medida de probabilidad (ver [43], [119]), se tiene que

$$\lim_{n \rightarrow \infty} F(a_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\Omega) = 1.$$

Y dado que de cualquier sucesión $\{x_n\}_{n \in \mathbb{N}}$ tal que $\lim_{n \rightarrow \infty} x_n = \infty$, podemos extraer una subsucesión creciente $\{x_{n_k}\}_{k \in \mathbb{N}}$ tal que $\lim_{k \rightarrow \infty} x_{n_k} = \infty$, pues \mathbb{R} con la métrica del valor absoluto es un espacio métrico, entonces se sigue el resultado.

- (b) Sea $a = \{a_n\}_{n \in \mathbb{N}}$ una sucesión decreciente de números reales tal que $\lim_{n \rightarrow \infty} a_n = -\infty$. Sea $A = \{A_n\}_{n \in \mathbb{N}}$ una sucesión de eventos con $A_n = (X \leq a_n)$, para toda $n \in \mathbb{N}$. Como la sucesión a es decreciente, entonces la sucesión A también lo es y esto implica que

$$\lim_{n \rightarrow \infty} A_n = \emptyset.$$

Entonces, nuevamente por la continuidad de la medida de probabilidad, se tiene que

$$\lim_{n \rightarrow \infty} F(a_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\emptyset) = 0.$$

Así, sin importar la forma en cómo los números tiendan a $-\infty$ se cumple que $\lim_{n \rightarrow \infty} F(x_n) = 0$.

- (c) Como $x \leq y$, entonces

$$\begin{aligned} F(x) &\leq F(x) + P(x < X \leq y) \\ &= P([X \leq x] \cup [x < X \leq y]) \\ &= P(X \leq y) \\ &= F(y). \end{aligned}$$

- (d) Sea $\{a_n\}_{n \in \mathbb{N}}$ una sucesión decreciente de números reales positivos tal que $\lim_{n \rightarrow \infty} a_n = 0$. Notemos que

$$F(x + a_n) = F(x) + P(x < X \leq x + a_n).$$

Entonces, tomando la sucesión de eventos $\{A_n\}$ donde $A_n = (x < X \leq x + a_n)$, para toda $n \in \mathbb{N}$, se tiene que la sucesión converge al conjunto vacío, por lo que usando la continuidad de la probabilidad obtenemos

$$\lim_{n \rightarrow \infty} F(x + a_n) = F(x) + \lim_{n \rightarrow \infty} P(A_n) = F(x),$$

de donde se sigue lo deseado.

- (e) Sea $\{a_n\}_{n \in \mathbb{N}}$ una sucesión decreciente de números reales positivos tal que $\lim_{n \rightarrow \infty} a_n = 0$. Tomando la sucesión de eventos $\{A_n\}_{n \in \mathbb{N}}$ con $A_n = (X < x - a_n)$, para toda $n \in \mathbb{N}$, se tiene que la sucesión converge al evento $(X < x)$, por lo que la continuidad de la probabilidad implica

$$P(X < a) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} F(x - a_n) = F(a^-).$$

■

Otra propiedad de la función de distribución se encuentra descrita en la siguiente definición.

Definición 5 (Variable aleatoria absolutamente continua) Decimos que una variable aleatoria X , con función de distribución F , es absolutamente continua si existe una función no negativa e integrable en el sentido de Riemman, $f : \mathbb{R} \rightarrow \mathbb{R}$, que denominaremos **función de densidad**, tal que

$$F(x) = \int_{-\infty}^x f(u) du.$$

En el caso que X sea una variable aleatoria discreta, es decir, una variable aleatoria cuyo rango es a lo sumo numerable el símbolo de integral se reemplaza por una suma sobre el conjunto de puntos en donde no es cero.

Cabe aclarar que solo en algunos casos, como en el Ejemplo 2, la función de densidad existe, y tanto esta como su función de distribución son conocidas bajo algún nombre específico, como distribución geométrica, binomial, normal, Poisson, etc., por tal razón, generalmente se trabaja con la función de distribución

Ahora presentamos algunos ejemplos que serán relevantes a lo largo de este trabajo.

Ejemplo 3 (Distribución Bernoulli) Sea $p \in (0, 1)$. Consideremos a la función de distribución $F : \mathbb{R} \rightarrow [0, 1]$ definida por

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - p, & \text{si } 0 \leq x < 1, \\ 1, & \text{en otro caso.} \end{cases}$$

Verificaremos que $f : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$f(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{si } x = 0, 1, \\ 0, & \text{en otro caso,} \end{cases}$$

es una función tal que

$$F(x) = \sum_{u \leq x} f(u).$$

Por la definición de f , para $x < 0$, se cumple que $f(x) = 0$, por lo que $F(x) = 0$. Luego, para $x = 0$ se tiene que $f(x) = 1 - p$; mientras que, para $0 < x < 1$ se satisface que $f(x) = 0$, en consecuencia, para $0 \leq x < 1$ se cumple que $F(x) = 1 - p$. Finalmente, $f(1) = p$ y si $x > 1$, entonces $f(x) = 0$; por lo tanto $F(x) = 1$ para $x \geq 1$, con lo que se verifica la afirmación.

Ahora, si una variable aleatoria X tiene distribución F , diremos que X tiene distribución Bernoulli con parámetro p y escribiremos $X \sim \text{Bernoulli}(p)$.

Ejemplo 4 (Distribución binomial) Sean $n \in \mathbb{N}$ y $p \in (0, 1)$. Consideremos la función de densidad $f: \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{si } x = 1, \dots, n, \\ 0, & \text{otro caso.} \end{cases}$$

Si una v.a, X , tiene función de densidad f , diremos que X tiene distribución Binomial con parámetros n y p , y escribiremos $X \sim \text{Binomial}(n, p)$.

Ejemplo 5 (Distribución binomial negativa) Sean $r \in \mathbb{N}$ y $p \in (0, 1)$. Consideremos la función de densidad $f: \{r, r+1, \dots\} \rightarrow \mathbb{R}$ definida por

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r}, & \text{si } x = r, r+1, \dots, \\ 0, & \text{en otro caso.} \end{cases}$$

Si una variable aleatoria X tiene función de densidad f , diremos que X tiene distribución binomial negativa con parámetros r y p , y escribiremos $X \sim \text{BN}(r, p)$.

Las variables aleatorias. poseen características numéricas entre las que destacan el valor esperado, la varianza y la covarianza. A continuación presentamos su definición.

Definición 6 Sea X una variable aleatoria absolutamente continua con función de densidad f .

- (a) El valor esperado, o esperanza, de X se define como la siguiente integral en el sentido de Riemann

$$E(X) = \int_{\mathbb{R}} x f(x) dx.$$

En el caso de que X sea una variable aleatoria discreta y absolutamente continua, sustituiremos el símbolo de integral por una suma sobre el conjunto de puntos en donde es distinta de cero. Además, cuando $E(X) = \pm\infty$ diremos que X no tiene esperanza finita.

- (b) La varianza de X se denota y define como

$$\text{Var}(X) = E([X - E(X)]^2).$$

(c) Si Y es una variable aleatoria, entonces la covarianza de X y Y se define como

$$\text{Cov}(X, Y) = E([X - E(X)][Y - E(Y)]).$$

La razón por la cual los tres conceptos previamente definidos destacan, es la interpretación que podemos asociarles. La esperanza es el número alrededor del cual se concentran los eventos con mayor probabilidad de ocurrir; por otro lado, la varianza se refiere a cuán dispersos se encuentran los eventos y, finalmente, la covarianza indica cómo es la relación entre dos variables aleatorias dadas. Para no desviarnos del objetivo de este trabajo sugerimos consultar [141], [115] y [69], para profundizar en las diversas propiedades que satisfacen los conceptos antes mencionados.

Ahora, introduciremos los vectores aleatorios, pues estos relacionan dos o más variables aleatorias. a fin de modelar experimentos aleatorios más sofisticados y la covarianza toma un papel fundamental en el estudio de ellos.

Definición 7 (Vector aleatorio) Sea $X : \Omega \rightarrow \mathbb{R}^n$ una función. Diremos que X es un vector aleatorio si

$$X^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) \in \mathcal{F}, \forall x_1, \dots, x_n \in \mathbb{R}.$$

De la definición anterior no es del todo claro que estemos hablando de un vector aleatorio; así, presentamos la siguiente proposición.

Proposición 1 Una función $(X_1, \dots, X_n)^T : \Omega \rightarrow \mathbb{R}^n$ es un vector aleatorio si, y sólo si, cada coordenada es una variable aleatoria.

Introduciremos ahora a las funciones de distribución conjuntas, las cuales son análogas a las funciones de distribución definidas previamente.

Definición 8 (Función de distribución conjunta) Sean X_1, \dots, X_n variables aleatorias. Diremos que la función de distribución conjunta del vector $X = (X_1, \dots, X_n)^T$, $F_X : \mathbb{R}^n \rightarrow [0, 1]$, se define por

$$F_X(x_1, \dots, x_n) = P_X(X_1 \leq x_1, \dots, X_n \leq x_n),$$

para cualquier $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Para vectores aleatorios existen definiciones y resultados análogos a los presentados para variables aleatorias. Recomendamos consultar [138], [115] y [69].

Ahora, extenderemos a vectores aleatorios el concepto de variables aleatorias absolutamente continuas y el concepto de función de densidad.

Definición 9 (Función de densidad conjunta) Sea $X = (X_1, \dots, X_n)^T$ un vector aleatorio con función de distribución F_X . Se dice que X es absolutamente continuo si existe una función no negativa e integrable, $f_X : \mathbb{R}^n \rightarrow \mathbb{R}$, tal que para todo $(x_1, \dots, x_n) \in \mathbb{R}^n$, se cumple que

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(u_1, \dots, u_n) du_n \dots du_1,$$

empleando la integral en el sentido de Riemann. Notemos que, similar a variables aleatorias, en el caso de que el vector aleatorio X esté formado de variables aleatorias discretas, sustituiremos los símbolos de integral por sumas sobre los correspondientes soportes. Además, a la función f_X la denominaremos función de densidad conjunta.

Dado un vector aleatorio, en ocasiones, es útil conocer el comportamiento estocástico de alguna colección de las variables aleatorias que lo componen, de esto surge el concepto que definimos a continuación.

Definición 10 (Función de densidad marginal) Sea $X = (X_1, \dots, X_n)^T$ un vector aleatorio con función de densidad f_X . Definimos a la función de densidad marginal $f_{X_i} : \mathbb{R} \rightarrow \mathbb{R}$, para cada $i = 1, \dots, n$, por

$$f_{X_i}(x_i) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{i-1}} \int_{-\infty}^{x_{i+1}} \cdots \int_{-\infty}^{x_n} f_X(u_1, \dots, u_n) du_n \dots du_{i+1} du_{i-1} \dots du_1,$$

empleando la integral en el sentido de Riemann.

Habiendo definido las funciones de densidad marginales podemos introducir el concepto de independencia en variables aleatorias.

Definición 11 (Independencia de variables aleatorias) Sean X y Y variables aleatorias con funciones de densidad f_X y f_Y , respectivamente. Diremos que X y Y son independientes si

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y), \quad (2.2)$$

para todos $x, y \in \mathbb{R}$. Donde $f_{(X,Y)}$ denota la función de densidad conjunta del vector aleatorio (X, Y) .

El concepto de independencia nos permite enunciar la definición de muestra aleatoria que presentamos a continuación.

Definición 12 (Muestra aleatoria) Sea X una variable aleatoria con función de distribución F_X . Sea $\{X_1, \dots, X_n\}$ una colección de n variables aleatorias. Si dichas variables son independientes e idénticamente distribuidas, es decir,

$$X_i \sim F_X,$$

para toda $i = 1, \dots, n$, diremos que $\{X_1, \dots, X_n\}$ es una muestra aleatoria de X de tamaño n . Además, diremos que x_1, \dots, x_n es una realización, u observación, de la muestra si

$$x_i = X_i(\omega),$$

para algún $\omega \in \Omega$ y para toda $i = 1, \dots, n$.

2.2. Conceptos generales de procesos estocásticos

Las variables y vectores aleatorios resultan de gran utilidad para modelar sucesos que ocurren una única ocasión. Luego, se recurre a las muestras aleatorias cuando repetimos dichos sucesos un número finito de veces.

Sin embargo, al reconocer que cada suceso puede estudiarse no solo considerando su resultado final sino también observando su evolución en el tiempo, surgen de esta manera los procesos estocásticos. A continuación, presentamos su definición.

Definición 13 (Proceso estocástico) Sea $X = \{X(t) \mid t \in T\}$ una colección de variables aleatorias definidas en un mismo espacio de probabilidad (Ω, \mathcal{F}, P) , indexadas por un conjunto $T \subseteq \mathbb{R}$, denominado *espacio parametral*, y que toman valores en un conjunto $S \subseteq \mathbb{R}$, denominado *espacio de estados*. Diremos entonces que X es un proceso estocástico.

Los procesos estocásticos pueden clasificarse por distintas propiedades. Las clasificaciones más elementales recaen en la forma del espacio parametral y del espacio de estados, y las presentamos a continuación.

Definición 14 (Clasificación basada en el espacio parametral) Sea $X = \{X(t) \mid t \in T\}$ un proceso estocástico.

- Si $|T| \leq |\mathbb{N}|$, se dice que X es un proceso a tiempo discreto.
- Si $|T| = |\mathbb{R}|$, se dice que X es un proceso a tiempo continuo.

Donde $|\cdot|$ denota la cardinalidad.

Definición 15 (Clasificación basada en el espacio de estados) Sea $X = \{X(t) \mid t \in T\}$ un proceso estocástico.

- Si $S \subseteq \mathbb{N}$, se dice que X es un proceso discreto.
- Si $S = \mathbb{R}$, se dice que X es un proceso continuo.

Para este trabajo nos interesan procesos estocásticos discretos y a tiempo continuo. Más aún, estudiaremos procesos que satisfagan la propiedad de Markov, la cual se enuncia a continuación.

Definición 16 (Propiedad de Markov) Sea $X = \{X(t) : t \geq 0\}$ un proceso estocástico a tiempo continuo. Diremos que X tiene la propiedad de Markov si

$$\begin{aligned} P(X(t+s) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s) \\ = P(X(t+s) = j \mid X(s) = i) \end{aligned}, \quad (2.3)$$

para todos los tiempos $s, t \geq 0$ y para todos los estados $i, j, x(u) \in S$, con $u \in [0, s)$. Si X tiene la propiedad de Markov es usual decir que X es un proceso Markoviano.

Luego, aquellos procesos que satisfacen las propiedades que nos interesan son llamados *cadena de Markov discretas a tiempo continuo*. Y, formalmente, su definición es la siguiente.

Definición 17 (Cadena de Markov discreta a tiempo continuo) Sea X un proceso estocástico discreto a tiempo continuo. Decimos que X es una cadena de Markov discreta a tiempo continuo si X tiene la propiedad de Markov.

Si X es un proceso Markoviano, la propiedad de Markov puede interpretarse como que la probabilidad de que ocurra un evento en X depende del evento anterior inmediato, únicamente, y no de todos los eventos previos.

De la propiedad de Markov, se desprende un resultado de suma importancia. Para presentarlo, es necesario discutir el siguiente lema.

Lema 2 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Dados los eventos $E_1, E_2, E_3 \in \mathcal{F}$, se cumple que

$$P(E_1, E_2, E_3) = P(E_3 \mid E_1, E_2)P(E_2 \mid E_1)P(E_1). \quad (2.4)$$

Demostración. Por definición de la probabilidad condicional se tiene que

$$P(E_3 | E_1, E_2) = P(E_3, E_2, E_1)[P(E_2, E_1)]^{-1} \text{ y}$$

$$P(E_2 | E_1) = P(E_2, E_1)[P(E_1)]^{-1}.$$

Así, al realizar el producto $P(E_3|E_1, E_2)P(E_2|E_1)P(E_1)$ el resultado es $P(E_3, E_2, E_1)$, lo deseado. ■

El lema previo será de utilidad para enunciar el teorema siguiente. El cual nos provee de herramientas que serán empleadas en la sección 4.1.1.

Teorema 3 (Ecuaciones de Chapman-Kolmogorov) Sea X una cadena de Markov discreta a tiempo continuo. Para todos los tiempos $h, t \geq 0$ y estados $j, k_0 \in \mathbb{N}$ se cumple que

$$\begin{aligned} P(X(t+h) = j | X(0) = k_0) \\ = \sum_{k=0}^{\infty} P(X(t+h) = j | X(t) = k)P(X(t) = k | X(0) = k_0). \end{aligned} \quad (2.5)$$

Demostración. Usando el Lema 2 y por considerar como $E_1 = \{X(0) = k_0\}$, $E_2 = \{X(t) = k\}$ y $E_3 = \{X(t+s) = j\}$ se tiene que

$$\begin{aligned} P(X(t+s) = j | X(0) = k_0) \\ &= P(X(t+s) = j, \Omega | X(0) = k_0) \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j, X(t) = k | X(0) = k_0) \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j, X(t) = k, X(0) = k_0)[P(X(0) = k_0)]^{-1} \\ &= \sum_{k=0}^{\infty} \left[P(X(t+s) = j | X(t) = k, X(0) = k_0)P(X(t) = k | X(0) = k_0) \right. \\ &\quad \left. P(X(0) = k_0)[P(X(0) = k_0)]^{-1} \right] \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j | X(t) = k, X(0) = k_0)P(X(t) = k | X(0) = k_0) \end{aligned} \quad (2.6)$$

Notemos que

$$P(X(t+s) = j | X(t) = k, X(0) = k_0) = P(X(t+s) = j | X(t) = k), \quad (2.7)$$

por la propiedad de Markov.

Ahora, al sustituir la Ecuación (2.7) en la Ecuación (2.6) obtenemos que

$$\begin{aligned} P(X(t+s) = j \mid X(0) = k_0) \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j \mid X(t) = k, X(0) = k_0) P(X(t) = k \mid X(0) = k_0) \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j \mid X(t) = k) P(X(t) = k \mid X(0) = k_0), \end{aligned}$$

lo deseado. ■

Capítulo 3

Carcinogénesis

De manera usual, se define a la célula como la unidad básica de la vida [54] y es que una célula es la forma más simple de vida, cada célula es capaz de procesar su alimento, responder a estímulos, moverse, crecer y reproducirse, además de que todo organismo vivo está compuesto de células.

Podemos estudiar a las células desde distintas perspectivas. Sin embargo, para este trabajo, nos enfocaremos en discutir la diversidad fenotípica. Posteriormente, contextualizaremos la relevancia que tienen los distintos fenotipos al presentar el desarrollo de cáncer.

Un fenotipo es una característica *perceptible* de un organismo [150]; por ejemplo, el tamaño del diámetro de una célula. En muchos caso, la detección de los distintos fenotipos puede realizarse experimentalmente a través de tinciones, como puede apreciarse en [131].

La diversidad fenotípica se refiere a la variabilidad en los fenotipos de los individuos dentro de una población o especie, donde tal variación puede ocurrir tanto por factores estocásticos, como mutaciones, y/o factores adaptativos, como respuestas al ambiente [80], [103].

Por un lado, los factores estocásticos representan una fuente constante de cambios en los fenotipos, los cuales carecen de intención [39]: Por otra parte, los factores adaptativos se desencadenan por la detección de cambios en el ambiente [55].

Tras el trabajo de Loeb *et al.* (1974) en [87], el paradigma en los fenotipos respectivo a mutaciones cambió, al proponer la *hipótesis del fenotipo mutador*. Esta hipótesis plantea la

existencia no solo de un fenotipo silvestre, coloquialmente llamado normal, y un fenotipo mutante, sino también la existencia de un fenotipo tal que las células con dicho fenotipo se caracterizan por poseer una tasa de mutación más alta que el resto de células, implicando que la mutagénesis puede ser favorecida.

En la Figura 3.1 podemos apreciar la dinámica fenotípica bajo la hipótesis del fenotipo mutador. Las células silvestres pueden dar origen a células mutantes, mutadoras y silvestres. Luego, las células mutadoras puede dar a origen a células mutantes y mutadoras. Por último, las células mutantes solo dan origen a más células mutantes.

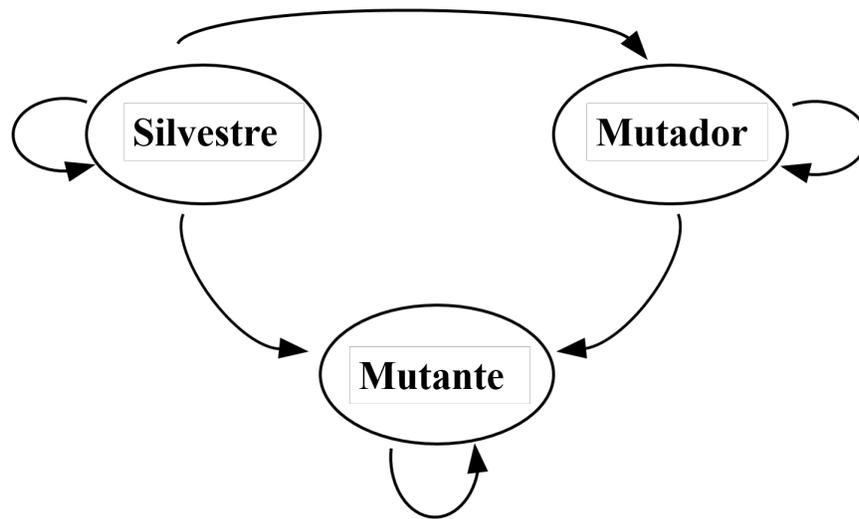


Figura 3.1: Paradigma de la dinámica fenotípica.

La presencia de células mutadoras juega un rol crucial. Está altamente relacionada con, por ejemplo, la inestabilidad genética que afecta la aptitud de las células, la resistencia a medicamentos y el desarrollo de cáncer. Más aún, la hipótesis del fenotipo mutador tiene sus orígenes en el estudio de cáncer [88], lo cual motiva la siguiente discusión sobre el desarrollo de cáncer.

Recordemos que los organismos compuestos por una única célula son llamados *unicelulares* y en otro caso, son llamados *pluricelulares*. Los animales, plantas y algunos hongos son ejemplos de organismos pluricelulares y, como se menciona en Aktipis (2020) [4], el origen de los organismos pluricelulares fue también origen del cáncer. A lo largo de este capítulo profundizaremos en el estudio del porqué el cáncer es tan antiguo como los organismos pluricelulares, ilustrando uno de los efectos mutacionales de mayor interés científico y en el cual nuestro trabajo tiene repercusiones.

Desde el punto de vista evolutivo, todo organismo tiene como objetivo la reproducción

exitosa ya que esto aumenta su aptitud. Esto es clave para el proceso de selección natural pues, de acuerdo con Pianka (2008) [110], esta opera considerando el éxito en la reproducción diferencial de los individuos, es decir, el número de descendientes que un organismo deja a lo largo de su vida [26]

Ahora, consideremos a una célula, de un organismo pluricelular, y considerémosla como un organismo unicelular; por ejemplo, una célula del estómago de algún humano entonces, por lo anterior, sería natural que esta célula buscara reproducirse tanto como el ambiente le permita; sin embargo, las células que conforman a un organismo pluricelular están siempre bajo un control de crecimiento poblacional muy estricto [143]. De esta manera, en el primer momento en que las células se organizan, por razones de cooperación [4], para formar a los organismos pluricelulares pierden esta la posibilidad de proliferación que naturalmente tenían, lo cual resulta contraintuitivo ya que esto disminuye su nivel de aptitud en términos de evolución y selección natural. Así, el cáncer aparece para que las células en los organismos pluricelulares sobrevivan y proliferen rápidamente, dejando más descendencia [4].

Se sigue de lo anterior que la evolución, de manera general, favorece a las células que proliferan rápidamente; lo cual es una propiedad que caracteriza a las células cancerígenas, también conocidas como células neoplásicas. Por otra parte, la evolución también favorece a los organismos con mejores mecanismos para evitar y/o suprimir las células cancerígenas [24]. De esta manera, ambos factores coexisten simultáneamente y los detallamos a continuación.

El cáncer es un grupo de enfermedades que surge cuando algunas células de un organismo pluricelular pierden el comportamiento cooperativo, proliferando y diseminándose sin restricciones [24], lo que potencialmente provoca comorbilidades e inclusive la muerte del organismo y para saber cómo se origina es necesario hablar sobre la *carcinogénesis*.

La carcinogénesis es el proceso por el cual las células sanas se transforman en células cancerígenas y se pueden reconocer tres etapas principales: iniciación, promoción y progresión [15].

En la etapa de iniciación ocurren las alteraciones, cambios o mutaciones en los genes que ocurren espontáneamente o de manera inducida que afectan la regulación de rutas bioquímicas asociadas con la proliferación celular, supervivencia y diferenciación; de esta manera, aparecen las células preneoplásicas, es decir, aquellas que tienen la capacidad de formar tumores [21], [128].

Luego, la etapa de promoción suele ser una etapa que se prolonga por varios años y donde las células preneoplásicas se acumulan y comienzan a desarrollar características

como la resistencia a la apoptosis, la muerte celular programada; la independencia de las señales de crecimiento, es decir, las células ya no dependen de estímulos externos para la reproducción [32]; y la falta de inhibición por contacto [21], esto es, usualmente, una célula deja de reproducirse cuando entra en contacto con otra célula del mismo tipo [108], lo cual mejora la habilidad de proliferación de las células, pues escapan de los mecanismos de control del organismo.

Finalmente, la progresión es la fase entre una lesión premaligna, donde la definición de lesión premaligna está basada usualmente en criterios epidemiológicos, clínicos y morfológicos; así, incluye a los tumores benignos y lesiones específicas como nódulos displásicos en el hígado y la displasia de Barret en el esófago [49], y el desarrollo del cáncer invasivo. La progresión es la etapa final de la transformación neoplásica, ocurre la proliferación celular; además, ocurren cambios genéticos y fenotípicos. Así, rápidamente el tumor aumenta de tamaño y las células pueden sufrir mutaciones, volviéndolas potencialmente invasivas y metastásicas [47], [128].

La carcinogénesis es un proceso multifactorial pues puede ocurrir por alteraciones genéticas, como anomalías en la replicación celular o en la reparación de genes y condiciones hereditarias [30]. Asimismo, también puede ocurrir por causas epigenéticas, como exposición a algunos agentes físicos, químicos o virus [76]. En ambos escenarios se da origen a mutaciones, pues el cáncer es resultado de ellas [9].

Recordemos que las mutaciones son alteraciones en la composición genética de un organismo que *pueden* resultar en un cambio o fenotipo heredable e identificable [71]. Las mutaciones tienen un rol crucial en el entendimiento del cáncer, son tan importantes que se utilizan para caracterizar a los diversos tipos de cáncer existentes a través de sus *firmas mutacionales*, los cuales son patrones en los procesos de mutación [101].

Ahora bien, no toda mutación origina cáncer [2] y para hablar de ello es necesario introducir a las células germinales o gametos, que se refiere a una subpoblación de células sexuales [18]; por ejemplo, los óvulos y los espermatozoides en los animales. Con lo anterior, podemos proceder a definir las células somáticas como el conjunto de todas las células que componen a un organismo pluricelular, exceptuando a las células germinales.

La razón para definir a las células somáticas es para hablar sobre las mutaciones que en ellas aparecen [148], es decir, las mutaciones somáticas dado que estas son las mutaciones capaces de dar origen al cáncer. Por otra parte, cuando ocurren las mutaciones en los gametos estas no ocasionan cáncer, aunque son capaces de heredar un mayor riesgo de desarrollar cáncer [91].

Las mutaciones en las células ocurren de manera espontánea y, en promedio, ocurre al menos una mutación que puede llegar a ser potencialmente cancerígena [125, Cap. 11]. Esto es, el riesgo de desarrollar cáncer siempre está presente; sin embargo, no es común desarrollarlo. Estimaciones mostraron que el riesgo de desarrollar cáncer antes de los 75 años es del 20%; mientras que, el riesgo de morir de cáncer es del 10% [50]. Asimismo, se ha estimado que aproximadamente el 20% de la población mundial será diagnosticada con algún tipo de cáncer a lo largo de su vida [106].

De esta manera, diariamente es posible que comience algún tipo de cáncer y, a la vez, no es tan frecuente llegar a desarrollar cáncer. Esto se debe a diversos factores, uno de ellos es que no todas las mutaciones somáticas convierten a un gen en un oncogén, es decir, un gen que fue alterado en el proceso de mutación de una célula somática no necesariamente resulta en un gen capaz de iniciar o favorecer la carcinogénesis [114].

3.1. Mecanismos internos contra el cáncer

Los organismos pluricelulares hemos desarrollado mecanismos intrínsecos que tienen, entre sus objetivos, la supresión de células cancerígenas [31]. El primero de ellos es la inhibición por contacto [1] la cual se refiere al proceso por el cual las células que pertenecen a un mismo tejido se autorregulan considerando dos principios fundamentales: la reproducción celular se detiene si la pared celular de cada célula del tejido está en contacto con la pared celular de las células del mismo tejido que están a su alrededor y las células del tejido que se reproduzcan fuera de la estructura de dicho tejido son sometidas a la apoptosis [31].

En la Figura 3.2 se ilustra el funcionamiento de los principios fundamentales que constituyen la inhibición por contacto. Si consideramos que las células pertenecen al mismo tejido y la línea punteada de color rojo como parte de la estructura que debe poseer dicho tejido entonces en la Figura 3.2a se muestran células que no están haciendo contacto con todas las células a su alrededor, por lo que la reproducción es necesaria a fin de que el tejido a formar esté completo. Luego, en la Figura 3.2b se muestra que las células han formado en su totalidad al tejido donde dado que la estructura del tejido debe mantenerse y que cada célula está en contacto con el resto de las células que le rodean, la reproducción está inhibida, salvo en caso de que ocurra la muerte de alguna de las células. Finalmente, en la Figura 3.2c se muestra que las células están fuera de la estructura que debe poseer el tejido y por tanto algunas de las células deben llevar a cabo la apoptosis.

Tras observar la Figura 3.2c podemos comprender mejor cuán relevante resulta la

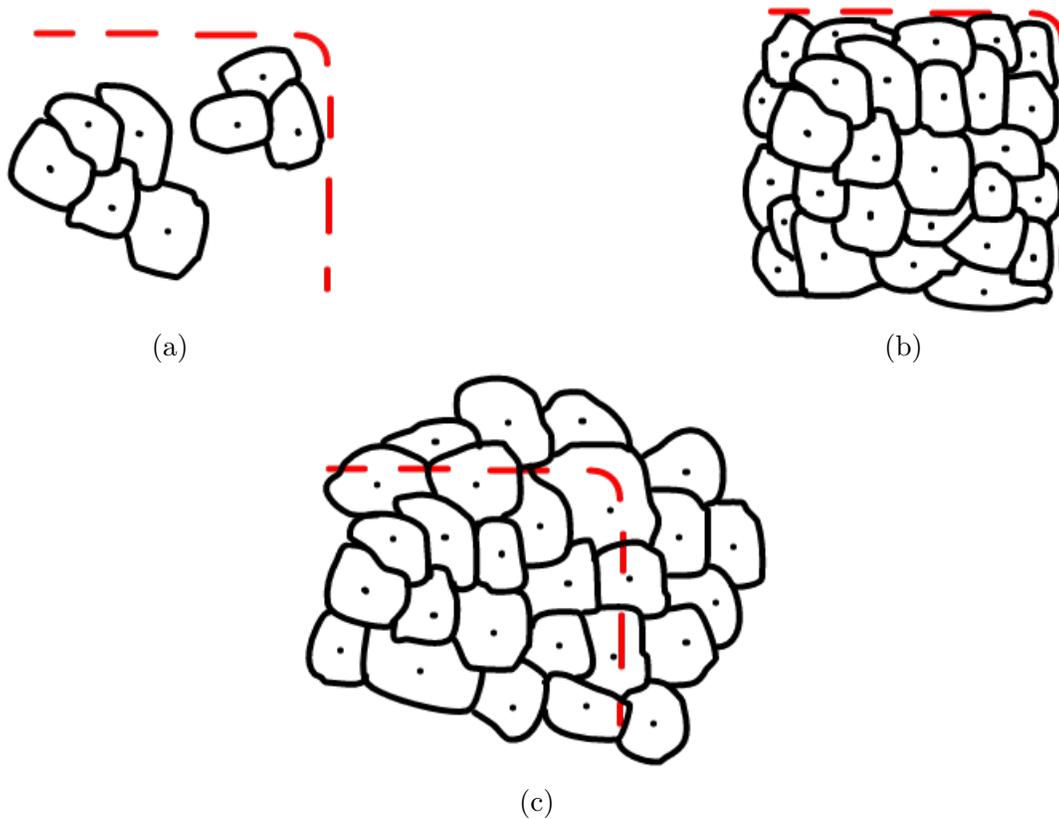


Figura 3.2: Mecanismo de inhibición por contacto.

inhibición por contacto en la carcinogénesis puesto que cuando una mutación en la célula interfiere con este mecanismo las restricciones que mantienen la estructura adecuada de los tejidos se ignoran y, si aunado a lo anterior, las células adquieren resistencia a la apoptosis entonces la proliferación de estas representa un riesgo potencial para la vida del individuo.

El segundo mecanismo del que hablaremos es conocido como senescencia celular y es un proceso de detención estable del ciclo celular que se inicia por señales o por diversos tipos de estrés; por ejemplo, señales de inflamación y/o daño de tejido, privación de nutrientes y desorganización del material del que están compuestos los cromosomas, es decir, desorganización de la cromatina [79]. Las células que son sometidas a la senescencia celular, que son llamadas células senescentes, sufren cambios en su expresión génica, desarrollan el fenotipo secretor complejo asociado a la senescencia (SASP, por sus siglas en inglés), el cual se refiere a que las células se mantienen activas metabólicamente, aunque con alteraciones, y secretan factores proinflamatorios en el cuerpo como citoquinas y proteasas [6], [34]. Además, la proliferación de las células senescentes se detiene, pero se mantienen vivas, por lo que es una alternativa a la apoptosis; sin embargo, aun cuando este mecanismo

está presente en diversos organismos, como en los humanos, existen algunos organismos específicos; por ejemplo, algunos árboles perennes, moluscos bivalvos o esturiones, cuyas células no poseen este mecanismo [22], [29].

Ahora, en el contexto del cáncer, recordemos que una de las características de las células neoplásicas es que son resistentes a la apoptosis por lo que la senescencia celular resulta crucial como mecanismo de prevención y defensa ante el cáncer en cualquiera de sus etapas de desarrollo.

Para introducir al tercer mecanismo que presentaremos es necesario definir a los telómeros. Los telómeros son estructuras que protegen los extremos de los cromosomas [146] y resultan cruciales dado que estos se rompan o se dañen. Sin embargo, tras cada vez que las células se reproducen ocurre una pérdida gradual de los telómeros que constituyen cada uno de los cromosomas de la célula, un proceso conocido como erosión o reducción de los telómeros [14], el cual puede desencadenar daños similares a rupturas de la doble hélice del ADN [65], consideradas como el daño más letal que pueden sufrir las células eucariotas [113].

Dadas las posibles implicaciones, una vez que los telómeros se han reducido hasta un determinado punto, la célula debe proceder a la apoptosis o a la senescencia; así es que después de que cada célula se haya reproducido algún determinado número de veces, el organismo se puede asegurar que es inofensiva. Por otra parte, se ha encontrado que existe una tercer posibilidad, las células somáticas pueden comenzar una transformación oncogénica [124]. En este caso, usualmente, la elongación de los telómeros se ve favorecida, por activación de la telomerasa [62]; por ejemplo, en humanos se estima que, aproximadamente, del 80 al 90 % de las células neoplásicas son capaces de elongar sus telómeros [66]. Así es que con la elongación de telómeros el organismo no recibe señales para neutralizar a la célula pese a que haya sobrepasado sus límites de reproducción, una de las razones por las cuales se considera al cáncer con la propiedad de inmortalidad [78].

En la Figura 3.3a, tomada de [99], podemos observar un cromosoma y en color naranja se aprecian los telómeros. Luego, en la Figura 3.3b, modificada de [149], se muestra a detalle la erosión de un telómero en función de la reproducción; además, con esta figura podemos entender también a la elongación de telómeros, ya que es el proceso inverso.

Ahora, presentamos el mecanismo de corrección de errores, también conocido como *proofreading*. Dentro de las células existen diversas enzimas las cuales son capaces de reconocer y reparar errores que se originan a lo largo del proceso de reproducción celular [114]; por ejemplo, la polimerasa I (Pol I), presente en las células procariotas, y su análoga en células eucariotas, la polimerasa β [109].

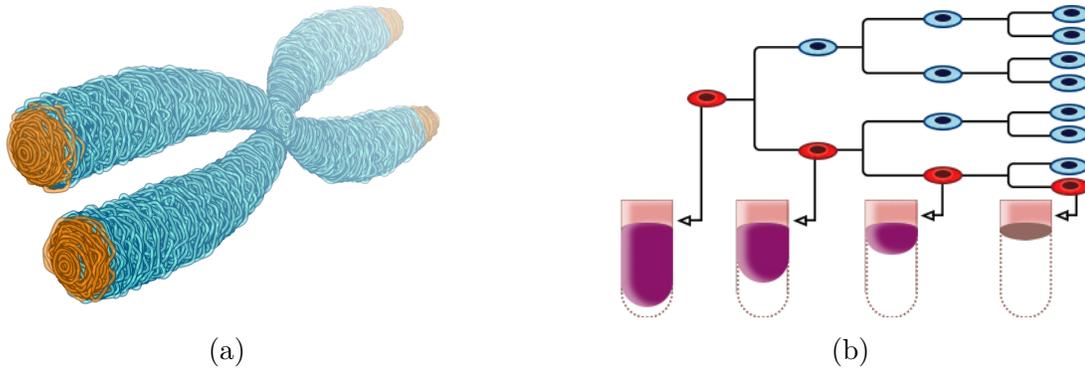


Figura 3.3: Telómeros como mecanismo de prevención del cáncer. 3.3a Estructura de un cromosoma mostrando en color naranja los telómeros, [99]. 3.3b Procesos de erosión de telómeros como efecto de la reproducción, modificada de [149].

Una vez identificado algún error, el éxito de la corrección de errores recae en la habilidad de elegir correctamente los nucleótidos (adenina, citocina, guanina, timina y uracilo) y extirpar los nucleótidos que se habían incorporado erróneamente en el genoma empleando exonucleasa [16], una enzima capaz de romper los enlaces que une los carbonos 3' - 5' de la desoxirribosa como se aprecia en la Figura 3.4, modificada del repositorio LabelMe [120], donde estos enlaces son los responsables de la unión entre un nucleótido y otro [94].

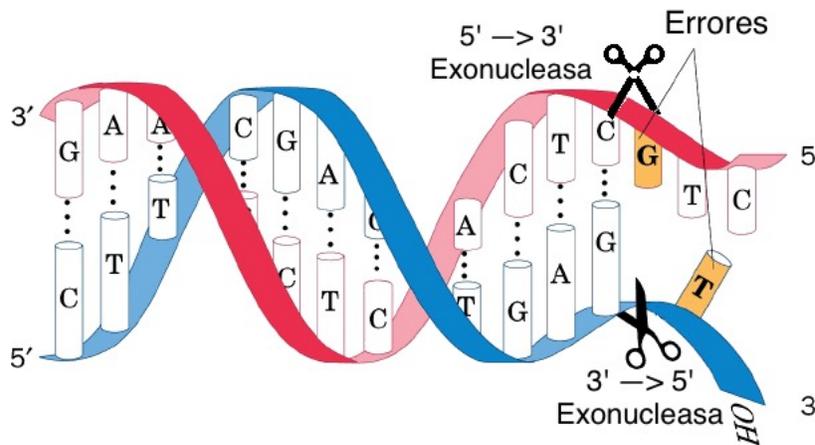


Figura 3.4: Función de la exonucleasa [120].

Los errores en la replications pueden causar diversas enfermedades; por ejemplo, deficiencia inmunodeficiencia, neurodegeneración, envejecimiento prematuro y cáncer [8]. Así, intuitivamente, se tiene que este es un mecanismo útil y eficiente en varios casos,

pero tiene sus limitaciones ya que, por otra parte, las células neoplásicas se caracterizan por su inestabilidad genética; es decir, una alta tendencia de alteraciones en el genoma durante la reproducción celular, por lo que se encuentran varias fallas a lo largo de un mismo genoma [151]. Ahora, consideremos lo siguiente, de acuerdo con Fan *et al.* (2007) [46] y Anandakrishnan *et al.* (2019) [9], existen modelos actuales [10], [12], [86], [89], [102], [105], [139], [153] en donde se asume que las mutaciones en todo punto del genoma son equiprobables; así que procediendo de esta manera, la probabilidad de que el mecanismo de corrección de errores detecte las n mutaciones, para luego corregirlas, es

$$\underbrace{\frac{1}{\ell_G} \frac{1}{\ell_G} \cdots \frac{1}{\ell_G}}_{n \text{ veces}} = \frac{1}{\ell_G^n}, \quad (3.1)$$

donde ℓ_G denota la longitud del genoma en nucleótidos.

Luego, si consideramos la ecuación anterior para el caso de los humanos, la Ecuación (3.1) es equivalente a

$$\frac{1}{(3200000000)^n},$$

pues la longitud del genoma humano, en nucleótidos, es de 3200000000 [27]. Por tanto, es prácticamente nula la probabilidad de que los mecanismos de corrección de errores sean efectivos contra las células neoplásicas.

Procedamos ahora a introducir el último de los mecanismos para la supresión de células cancerígenas que discutiremos, pero también el más conocido, el sistema inmune.

De manera general, el sistema inmune es el responsable de mantener saludable el organismo, por lo que se encarga no solo de prevenir la aparición del cáncer sino de brindar protección ante todo agente patógeno e infeccioso que pueda perjudicar y/o comprometer el bienestar de cada individuo [59]. Ahora bien, el sistema inmune está compuesto por diversos tipos de células, las cuales describimos a continuación.

Los macrófagos, un tipo específico de los glóbulos blancos, están equipados para detectar y responder a infecciones y lesiones de los tejidos además de que juegan un rol crucial en la reparación de tejidos y la homeostasis de ellos [83], es decir, la regulación dinámica de, por ejemplo, concentraciones de ácido, azúcar, energía, oxígeno, y temperatura [114]. Durante la carcinogénesis, actúan los macrófagos antitumorales que promueven la inflamación y participan en la eliminación de más células cancerígenas inmunogénicas [60], es decir, células cancerígenas que tienen la habilidad de inducir respuestas inmunes adaptativas similares a condiciones autoinmunes [20].

La inflamación crónica es un factor distintivo del cáncer [40] y los neutrófilos, glóbulos

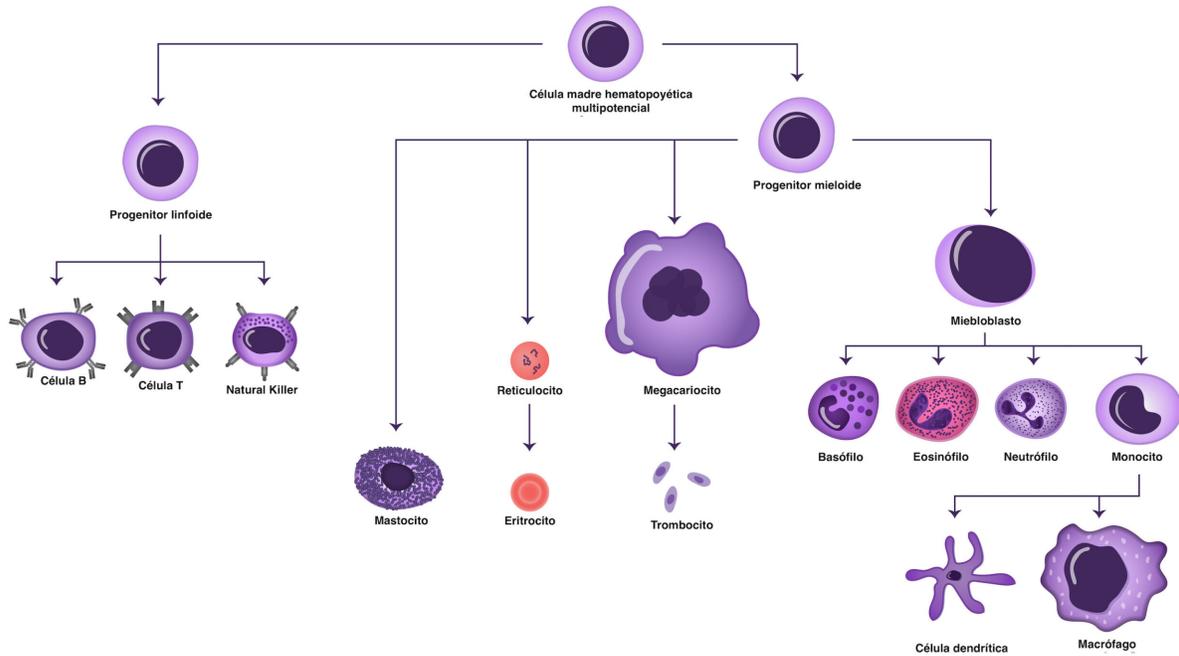


Figura 3.5: Células del sistema inmune [19].

blancos especializados, son de las primeras células inmunitarias en acudir cuando hay tejido dañado, por lo que representan una parte importante durante la inflamación dado que, además de eliminar patógenos, modulan la inflamación a través de diversos mecanismos tales como: secreción de proteínas antibacterianas o fagocitosis [73], es decir, la eliminación de microorganismos y sustancias de tamaño mayor a $0.5\mu m$ de diámetro [144].

Es importante mencionar que en pacientes con cáncer se han encontrado altos niveles de neutrófilos lo cual se ha asociado a pronósticos médicos adversos [42]. Además, se ha encontrado una variante de neutrófilos asociados a tumores, de la cual se han caracterizados dos subpoblaciones, una protumoral y, la otra, antitumoral [53]; sin embargo, se conoce muy poco acerca de la forma en la que los neutrófilos se vuelven protumorales y si es que es posible revertir dicho proceso [60].

Hablemos ahora sobre las células NK, conocidas también como *natural killer*. Estas células inmunológicas son de rápida respuesta contra células transformadas; por ejemplo, células que han sido infectadas o, bien, que hayan adquirido cambios malignos, o potencialmente malignos, como el caso de células neoplásicas [33]. Las células NK tienen una amplia variedad de receptores inhibidores y estimulantes a lo largo de su superficie, la membrana plasmática, los cuales son empleados para vigilancia inmunológica [60] pues, a partir de ellos, reconocen expresiones alteradas de proteínas y controlan la función citolítica

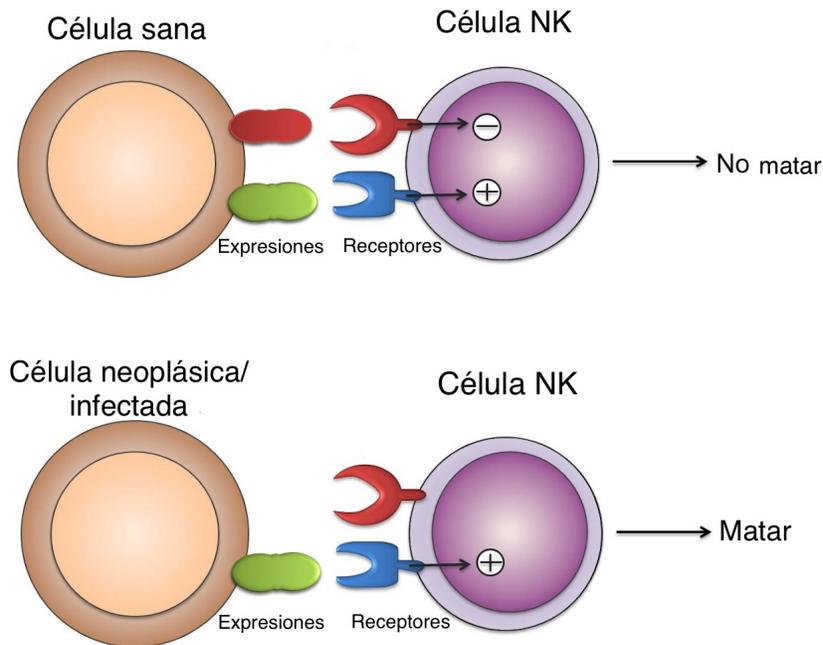


Figura 3.6: Mecanismo de reconocimiento de las células NK [129].

[107], esto es, la disolución de las células en sus componentes químicos [114].

En la Figura 3.6, modificada de [129], se ilustra el mecanismo a través del cual las células NK llevan a cabo la identificación de células y en caso de detectar alguna célula transformada proceder a su eliminación. Las células NK utilizan sus receptores para reconocer las expresiones que son comunicadas por las células. A manera de analogía, las expresiones de la célula funcionan como llaves; mientras que, los receptores toman el papel de cerraduras; así, dada una cerradura, se tienen los siguientes casos:

1. Se tiene una llave que entra en la cerradura y da acceso;
2. Se tiene una llave que entra en la cerradura, pero no da acceso; y
3. No se tiene llave que entre en la cerradura.

Y en cada caso, se tiene información. Por ejemplo, en el primer caso, una de las expresiones de la célula encaja con el determinado receptor y encajan de la manera adecuada por lo que dependiendo de qué receptor fue activado, la célula NK puede proceder a eliminar a la otra célula o no. Luego, en el segundo caso, se tiene que alguna de las expresiones encaja con el receptor, solo que no encajaron de la manera correcta lo que, en general, representa una potencial alerta. Por último, en el tercer caso, la célula NK no es

capaz de reconocer la expresión, pero a fin de prevenir un daño al organismo, la célula NK elimina a la célula e integra a sus memorias de respuesta este nuevo tipo de células dado que, en condiciones estándares, todas las células que resultan inocuas para el individuo son identificables por las células NK.

Como mencionamos, las células NK detectan, a través de sus receptores, a las células que potencialmente pueden dañar al organismo, pero solo a partir que estas expresan algún cambio, es decir, lo “comunican”. Las células neoplásicas han tomado ventaja de esto pues son capaces de incrementar la expresión de una molécula específica, el complejo mayor de histocompatibilidad de clase I (MHC-I), para inhibir las funciones citotóxicas, como la función citolítica [134]. Además, disminuyen la expresión de una proteína, ligandos de NKG2D, para tratar de evitar ser reconocidas por las células NK [58].

Ahora, introducimos las células dendríticas, mejor conocidas por DC. Estas son un tipo especializado de las células presentadoras de antígeno, estas son capaces de procesar un antígeno y exponerlos ante otras células del sistema inmune [60], es decir, desencadenan una respuesta inmune al generar una sustancia que puede ser reconocida por más células del sistema inmune [100].

En los humanos, las células dendríticas están presentes en todo el cuerpo, salvo en la parénquima cerebral [95]; sin embargo, se dividen en diversos subtipos, los cuales poseen distintas funciones inmunológicas detalladas, rutas migratorias y dependencia respecto a infecciones o estímulos inflamatorios [25].

De acuerdo con Böttcher *et al.* (2018) [23], se ha mostrado que, en ratones, las células neoplásicas de algunos tipos de cáncer afectan el reclutamiento de células dendríticas en el micro-ecosistema que rodea al tumor dentro del organismo secretando prostaglandina E2, mermando algunas funciones de las células NK contra el cáncer.

Hablaremos ahora sobre las células T, los linfocitos tímicos; estas pueden adquirir fenotipos funcionales cuyo efecto actúa directamente en condiciones inflamatorias [132]; además; representan el segundo tipo de células más frecuentes en el cuerpo, respondiendo a alérgenos, patógenos y tumores, al igual que mantienen la memoria inmunológica [77].

Las células T suelen trabajar de forma colaborativa con las células NK para degradar antígenos extraños empleando un complejo proteico conocido como inmunoproteasoma [114] y dado que muchas células neoplásicas producen antígenos [136], se ha observado que las células T son altamente efectivas contra el cáncer. Por otra parte, en algunos tipos de cáncer, la inmunoproteasoma es capaz de inhibir o promover la tumorigénesis [140]; asimismo, a través de estudios clínicos se ha sugerido que las células neoplásicas

toman ventaja de algunas propiedades inmunosupresoras de las células T, lo que afecta la capacidad de supervivencia e infiltración en los tumores que poseen las células T [60].

Para finalizar, hablaremos de los linfocitos B, también conocidos como células B. Estas se originan en la médula ósea, el tejido al interior de los huesos, y modulan la respuesta de inmunidad humoral, es decir, la producción de anticuerpos contra patógenos [60], [130].

Hasta el momento, la evidencia indica que las células B promueven la carcinogénesis cuando se tienen condiciones de inflamación crónica [145], así que la ausencia de células B maduras reducen la progresión tumoral [60]. Más aún, una característica importante del cáncer es que induce la generación de una citoquina, la linfotoxina, la cual promueve el crecimiento y metástasis de las células neoplásicas [126] y, por otra parte, las células B pueden generar linfotoxina y cuando se infiltran en los tumores, en efecto, la producen [152]. Por lo que la presencia de células B es de ayuda contra el cáncer, al igual que el resto de las células del sistema inmune y, a la par, promueve el desarrollo del cáncer.

Capítulo 4

Modelos para la dinámica fenotípica

Las células de los organismos pluricelulares pueden reproducirse de manera sexual y asexual las cuales son conocidas como meiosis y mitosis, respectivamente. La meiosis es el mecanismo de reproducción presente exclusivamente en los gametos [104], es decir, las células de reproducción que corresponden a los organismos; por ejemplo, óvulos y espermatozoides. Por otra parte, el resto de las células llevan a cabo la mitosis [121].

La meiosis es un proceso que puede prolongarse a lo largo de días [17] debido a que este proceso debe ser supervisado con sumo detalle a fin de prevenir errores posteriores [92] los cuales, potencialmente, resultarían dañinos para la descendencia del organismo.

La meiosis se divide en dos grandes etapas la meiosis I y la meiosis II. La meiosis I toma al rededor del 90% del tiempo total de duración de este proceso de reproducción [5]; mientras que, la meiosis II es un proceso similar a la mitosis [104]. De concluirse exitosamente, en la meiosis, se obtiene un total de cuatro células *únicas* dando lugar, cada una de ellas, a un gameto [123].

En contraste, para el resto de las células, la mitosis es un proceso más rápido prolongándose hasta los 180 minutos [85]. Además, en este caso, la célula progenitora también conocida como célula madre, se divide en dos nuevas células hijas genéticamente *idénticas* [122].

En lo que respecta a la modelación matemática de la reproducción celular, considerando la hipótesis del fenotipo mutador, existen ya algunos modelos ampliamente aceptados, siendo el modelo Moran el que se consolidó como el modelo más popular debido a su precisión biológica.

Omitiremos la discusión del modelo Moran pues está fuera del objetivo de este trabajo; sin embargo, invitamos a consultar [28, Fig. 5B] en donde realizan un breve resumen. No obstante, se anexa un código en R para generar simulaciones con dicho método en el Apéndice A.1.

En este capítulo, presentaremos brevemente a los modelos previamente mencionados para enfocarnos en los modelos desarrollados como parte de este trabajo.

Antes de comenzar será necesario introducir el concepto de tasa de mutación, el cuál aparecerá en repetidas ocasiones y en distintos contextos.

Definición 18 (Tasa de mutación) Dada una célula, diremos que su tasa de mutación es la probabilidad de que ocurran mutaciones en la célula por unidad de tiempo.

4.1. Modelo gráfico

En principio, las simulaciones que se generan usando el modelo Moran poseen una alta fidelidad al proceso biológico; sin embargo, se sabe que al generar simulaciones con dicho algoritmo es computacionalmente costoso, especialmente para simular tamaños finales de población “grande” [158]. Es decir, simular la dinámica fenotípica de una población que crece hasta, digamos, 10^9 individuos a partir de un único individuo inicial, requiere de al menos 10^9 decisiones, lo cual se traduce en un tiempo de espera prologando. Específicamente, obtener una simulación empleando el modelo Moran y con las condiciones antes descritas requiere de más de un mes ininterrumpido de tiempo computacional; mientras que, en un laboratorio con condiciones ambientales óptimas, una población celular de *E. coli* logra una población de tal magnitud en cuestión de, a lo más, un día [142].

Ahora, con el objetivo de optimizar el tiempo desarrollamos un modelo que hace uso de más herramientas deterministas y probabilísticas de manera tal que los resultados se obtuvieran con un menor tiempo de cómputo.

La siguiente sección está dedicada a las herramientas probabilistas a las que se recurre para el desarrollo del modelo gráfico.

4.1.1. Proceso de Yule-Furry

La siguiente discusión tiene como objetivo caracterizar el comportamiento estocástico del tamaño del linaje que un conjunto de células, de cardinalidad $k_0 \in \mathbb{N}$, es capaz de producir en el intervalo de tiempo $[t_0, t_f]$ dado que crecen a una tasa λ . Esto es, describiremos la variable aleatoria que cuenta el número total de células existentes al tiempo t_f en un cultivo celular que fue fundado, al tiempo t_0 , por k_0 células las cuales crecen a una tasa λ .

Para comenzar, introducimos siguiente concepto, el cual nos permitirá trabajar con el crecimiento celular de una manera abstracta que se desarrolla aleatoriamente en el tiempo.

Definición 19 (Proceso de ramificación) Sea $X = \{X(t) : t \in [0, \infty)\}$ una cadena de Markov a tiempo continuo y con espacio de estados $S = \mathbb{N}$. Diremos que X es un proceso de ramificación si X tiene probabilidades de transición estacionarias. Es decir, satisface el valor de

$$P(X(t+s) = j \mid X(s) = i) \quad (4.1)$$

es independiente de $s \geq 0$, para todo tiempo $t \geq 0$ y todos los estados $i, j \in \mathbb{N}$.

Una propiedad de gran relevancia que poseen los procesos de ramificación es la llamada propiedad de diferencia. Esta propiedad puede interpretarse como una característica de auto-similitud pues nos dice que dado un proceso de ramificación y considerar dicho proceso de ramificación a partir de un tiempo $s > 0$ el proceso estocástico resultante vuelve a ser un proceso de ramificación.

Teorema 4 (Propiedad de diferencia) Sea $X = \{X(t) \mid t \geq 0\}$ un proceso de ramificación. Definamos $Y = \{Y(t) = X(t+s) \mid t \geq 0\}$, para $s \geq 0$ dada, entonces Y es un proceso de ramificación.

Demostración. Sean $t_1, t_2 \geq 0$ tiempos y sean $i, j \in \mathbb{N}$ estados arbitrarios. Notemos que por definición se tiene que

$$\begin{aligned} P(Y(t_1 + t_2) = j \mid Y(t_1) = i, Y(u) \in \mathbb{N}, 0 \leq u < t_1) \\ &= P(X(t_1 + t_2 + s) = j \mid X(t_1 + s) = i, X(u + s) \in \mathbb{N}, 0 \leq u < t_1) \\ &= P(X(t_1 + t_2 + s) = j \mid X(t_1 + s) = i, X(v) \in \mathbb{N}, 0 \leq v < t_1 + s). \end{aligned}$$

Luego, por la propiedad de Markov de X obtenemos que

$$\begin{aligned} P(X(t_1 + t_2 + s) = j \mid X(t_1 + s) = i, X(v) \in \mathbb{N}, 0 \leq v < t_1 + s) \\ &= P(X(t_1 + t_2 + s) = j \mid X(t_1 + s) = i) \\ &= P(Y(t_1 + t_2) = j \mid Y(t_1) = i). \end{aligned}$$

Verificando que

$$\begin{aligned} P(Y(t_1 + t_2) = j \mid Y(t_1) = i, Y(u) = y(u), 0 \leq u < t_1) \\ &= P(Y(t_1 + t_2) = j \mid Y(t_1) = i), \end{aligned}$$

la propiedad de Markov de Y .

Sean $s_1, s_2 \geq 0$ tiempos arbitrarios y fijos. Dada $t \geq 0$ se cumple que

$$P(X(t + (s_1 + s)) = j \mid X(s_1 + s) = i) = P(X(t + (s_2 + s)) = j \mid X(s_2 + s) = i),$$

pues X tiene probabilidades de transición estacionarias.

Por tanto

$$P(Y(t + s_1) = j \mid Y(s_1) = i) = P(Y(t + s_2) = j \mid Y(s_2) = i).$$

Y ya que $s_1, s_2 \geq 0$ fueron tiempos arbitrarios se tiene que el valor de

$$P(Y(t + s^*) = j \mid Y(s^*) = i),$$

es independiente de $s^* \geq 0$. Verificando que Y tiene probabilidades de transición estacionarias. Más aún, en conjunción, lo anterior muestra que Y es un proceso de ramificación. ■

A manera de analogía, los procesos de ramificación admiten una representación en un diagrama similar al de un árbol genealógico. Así, el teorema anterior puede interpretarse de la siguiente manera: dado un árbol genealógico de un organismo y un ancestro en él, el subconjunto del árbol genealógico que consta de exclusivamente de los descendientes de dicho ancestro, ignorando los ancestros previos, es también un árbol genealógico, independiente del ancestro elegido. Esta propiedad tendrá un rol crucial en el planteamiento del modelo gráfico.

Para continuar, discutiremos el proceso de Yule-Furry. Existen diversas formas de presentar dicho proceso; por ejemplo, a través de una construcción axiomática [137, Cap. 6], eigenteoría [35, Cap. 3, 6] o de manera heurística [118, Cap. 6]. Para este trabajo nos hemos basado fuertemente en la introducción del proceso de Yule-Furry desde la perspectiva heurística y complementamos la argumentación de algunos resultados con lo presentado en la construcción axiomática.

Definición 20 (Proceso de Yule-Furry) Sea $X = \{X(t) \mid t \geq 0\}$ un proceso de ramificación. Si $X(t)$ denota el tamaño de una población al tiempo t y dicha población

- Comienza con $X(0) = k_0$ miembros,
- Sus miembros se reproducen de manera independiente, pero no mueren, y
- Crece exponencialmente a una tasa λ .

Entonces diremos que X es un proceso de Yule-Furry con parámetros λ y k_0 .

Cuando los parámetros λ y k_0 se sobreentienden, es usual omitirlos del nombre. Así es que, dado que a lo largo de esta sección dichos parámetros no toman valores, los omitiremos y nos referiremos al proceso de Yule-Furry con parámetros λ y k_0 como proceso de Yule-Furry, únicamente. Ahora, para familiarizarnos con este concepto, vemos el siguiente resultado.

Lema 5 Sea X un proceso de Yule-Furry. Si $p = P\left(\bigcup_{t \geq 0} \{X(t) \geq j\}\right)$ denota la probabilidad de que la población alcance un tamaño mínimo de j células, entonces

$$\lim_{h \rightarrow 0} \frac{1 - P(X(t+h) = i \mid X(t) = i)}{h} = i\lambda \quad (4.2)$$

y

$$\lim_{h \rightarrow 0} \frac{P(X(t+h) = j \mid X(t) = i)}{h} = i\lambda p, \quad (4.3)$$

para toda $t \geq 0$ y todas $i, j \in \{k_0, k_0 + 1, \dots\}$.

Demostración. Dado que $P(X(t+h) = i \mid X(t) = i)$ es la probabilidad de que en el intervalo de tiempo $(t, t+h)$ la población crezca de i a i individuos, entonces $1 - P(X(t+h) = i \mid X(t) = i)$ es la probabilidad de que al tiempo t , la población cuente con i individuos y luego de h unidades de tiempo crezca.

Sabemos que cada célula crece con una tasa λ por unidad de tiempo, por tanto λ es la tasa a la que ocurre una transición de estado. De esta manera, podemos interpretar a λ como la probabilidad de que una célula se reproduzca. Así, $1 - P(X(t+h) = i \mid X(t) = i)$ es igual al producto de la probabilidad de que las i células se reproduzcan, λi , y el tiempo durante el cual pueden reproducirse, h . Además, para incluir eventos posibles que contribuyan a la reproducción, pero que se consideran ínfimamente probables, se añade un sumando tal que es función del tiempo y que puede despreciarse, es decir, una $\mathcal{O}_1(h)$ la cual satisface que

$$\lim_{h \rightarrow 0} \frac{\mathcal{O}_1(h)}{h} = 0.$$

Lo anterior implica que

$$1 - P(X(t+h) = i \mid X(t) = i) = \lambda i h + \mathcal{O}_1(h) \quad (4.4)$$

y

$$\lim_{h \rightarrow 0} \frac{1 - P(X(t+h) = i \mid X(t) = i)}{h} = \lim_{h \rightarrow 0} \frac{\lambda i h + \mathcal{O}_1(h)}{h} = \lambda i, \quad (4.5)$$

lo cual prueba la primera parte.

Para probar la siguiente parte, el argumento es similar. La probabilidad de que en el intervalo de tiempo $(t, t+h)$ la población crezca de i a j individuos, $P(X(t+h) = j \mid$

$X(t) = i$), es igual al producto de la probabilidad de que las i células se reproduzca, λi , el tiempo que transcurre, h , y la probabilidad de que la población alcance un tamaño de j células, p , en h unidades de tiempo. Nuevamente, para incluir eventos posibles que contribuyan a la reproducción, pero que se consideran ínfimamente probables, es decir, una $\mathcal{O}_2(h)$ la cual satisface que

$$\lim_{h \rightarrow 0} \frac{\mathcal{O}_2(h)}{h} = 0.$$

Implicando que

$$P(X(t+h) = j \mid X(t) = i) = \lambda i h p + \mathcal{O}_2(h) \quad (4.6)$$

y

$$\lim_{h \rightarrow 0} \frac{P(X(t+h) = j)}{h} = \lim_{h \rightarrow 0} \frac{\lambda i h p + \mathcal{O}_2(h)}{h} = \lambda i p, \quad (4.7)$$

lo deseado. ■

El resultado anterior se presenta de una manera completamente heurística, en gran parte, para mantener una lectura fluida y los propósitos del trabajo. Sin embargo, la discusión puede formalizarse para dar pie a condiciones de regularidad que conllevan a la derivación de las ecuaciones diferenciales de Kolmogorov [35, p. 116]. Alternativamente, al añadir condiciones adicionales, el resultado anterior puede derivarse de proponer una construcción axiomática de un proceso estocástico [137, p. 334], lo que lleva a discutir la existencia de tal proceso.

Para proseguir, recordemos que nuestro objetivo es obtener la distribución de la variable aleatoria que cuenta el número de células, al tiempo t_f , que provienen de k_0 células iniciales que crecen a una tasa λ y cuyo cultivo celular ocurrió al tiempo t_0 .

Hallaremos dicha distribución tras resolver las ecuaciones retrospectivas de Kolmogorov para un proceso de Yule-Furry. Por tanto, exploremos el siguiente resultado.

Teorema 6 (Ecuación retrospectiva de Kolmogorov) Sea $X = \{X(t) : t \geq 0\}$ un proceso de ramificación tal que $X(0) = k_0$,

$$\lim_{h \rightarrow 0} \frac{1 - P(X(t+h) = j \mid X(t) = j)}{h} = u_j \quad (4.8)$$

y

$$\lim_{h \rightarrow 0} \frac{P(X(t+h) = k \mid X(t) = j)}{h} = v_{j,k}, \quad (4.9)$$

para todos los estados i, j .

Si $P_j(t) := P(X(t) = j \mid X(0) = k_0)$ entonces se cumple que

$$P'_j(t) = \sum_{k \neq j} v_{j,k} P_k(t) - u_j P_j(t), \quad (4.10)$$

para toda $t \geq 0$.

Demostración.

Por el Teorema 3, las ecuaciones de Chapman-Kolmogorov, sabemos que

$$\begin{aligned} P_j(t+h) &= P(X(t+h) = j \mid X(0) = k_0) \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j \mid X(t) = k) P(X(t) = k \mid X(0) = k_0) \\ &= \sum_{k=0}^{\infty} P(X(t+s) = j \mid X(t) = k) P_k(t). \end{aligned}$$

En consecuencia, se tiene lo siguiente

$$\begin{aligned} P_j(t+h) - P_j(t) &= \sum_{k=0}^{\infty} P(X(t+h) = j \mid X(t) = k) P_k(t) - P_j(t) \\ &= \sum_{k \neq j} P(X(t+h) = j \mid X(t) = k) P_k(t) \\ &\quad + P(X(t+h) = j \mid X(t) = j) P_j(t) - P_j(t) \\ &= \sum_{k \neq j} P(X(t+h) = j \mid X(t) = k) P_k(t) \\ &\quad - (1 - P(X(t+h) = j \mid X(t) = j)) P_j(t). \end{aligned} \quad (4.11)$$

Entonces, al tomar el límite cuando h tiende a 0 se obtiene que

$$\begin{aligned} &\lim_{h \rightarrow 0} \frac{1}{h} \sum_{k \neq j} P(X(t+h) = j \mid X(t) = k) P_k(t) \\ &\quad - \lim_{h \rightarrow 0} \frac{1}{h} (1 - P(X(t+h) = j \mid X(t) = j)) P_j(t) \\ &= \sum_{k \neq j} \lim_{h \rightarrow 0} \frac{1}{h} P(X(t+h) = j \mid X(t) = k) P_k(t) \\ &\quad - \lim_{h \rightarrow 0} \frac{1}{h} (1 - P(X(t+h) = j \mid X(t) = j)) P_j(t), \end{aligned} \quad (4.12)$$

donde el intercambio del límite y la serie se garantiza por el teorema de convergencia dominada de Lebesgue.

Luego, por las ecuaciones (4.9) y (4.8) se obtiene que

$$\lim_{h \rightarrow 0} \frac{1}{h} P(X(t+h) = j \mid X(t) = k) P_k(t) = v_{j,k} P_k(t) \quad (4.13)$$

y

$$\lim_{h \rightarrow 0} \frac{1}{h} (1 - P(X(t+h) = j \mid X(t) = j)) P_j(t) = u_j P_j(t), \quad (4.14)$$

respectivamente.

Al sustituir las ecuaciones anterior en la Ecuación (4.12) se obtiene que

$$\begin{aligned} P'_j(t) &= \lim_{h \rightarrow 0} \frac{P_j(t+h) - P_j(t)}{h} \\ &= \sum_{k \neq j} v_{j,k} P_k(t) - u_j P_j(t), \end{aligned} \quad (4.15)$$

lo deseado. ■

Del resultado anterior, tenemos como corolario las ecuaciones retrospectivas de Kolmogorov para el proceso de Yule-Furry.

Corolario 1 Sea X un proceso de Yule-Furry. Si

$$P_j(t) = P(X(t) = j \mid X(0) = k_0),$$

$j = k_0, k_0 + 1, \dots$ entonces

$$P'_{k_0}(t) = -\lambda k_0 P_{k_0}(t) \quad (4.16)$$

y

$$P'_j(t) = -\lambda j P_j(t) + \lambda(j-1) P_{j-1}(t), j > k_0, \quad (4.17)$$

denotan las ecuaciones retrospectivas de Kolmogorov para X .

Demostración. Por el Teorema 3, las ecuaciones de Chapman-Kolmogorov, se tiene lo

siguiente.

$$\begin{aligned}
P_j(t+s) &= P(X(t+s) = j \mid X(0) = k_0) \\
&= \sum_{k=0}^{\infty} P(X(t+s) = j \mid X(t) = k)P(X(t) = k \mid X(0) = k_0) \\
&= \sum_{k=0}^{\infty} P(X(t+s) = j \mid X(t) = k)P_k(t) \\
&= \sum_{k=0}^{\infty} P(X(t+s) - X(t) = j - k \mid X(t) = k)P_k(t) \\
&= \sum_{k=0}^j P(X(t+s) - X(t) = j - k \mid X(t) = k)P_k(t),
\end{aligned}$$

pues se cumple que $j - k \geq 0$, e implica que $k \leq j$.

Luego, para cada $k \in \{0, 1, \dots, j-2\}$ se cumple que

$$\begin{aligned}
P(X(t+s) - X(t) = j - k \mid X(t) = k) &\leq P(X(t+s) - X(t) \geq j - (j-2) \mid X(t) = k) \\
&= P(X(t+s) - X(t) \geq 2 \mid X(t) = k) \\
&= 1 - P(X(t+s) - X(t) < 2 \mid X(t) = k) \\
&= 1 - P(X(t+s) - X(t) = 0 \mid X(t) = k) \\
&\quad - P(X(t+s) - X(t) = 1 \mid X(t) = k).
\end{aligned}$$

Observemos que

$$P(X(t+s) = j \mid X(t) = k) = P(X(t+s) - X(t) = j - k \mid X(t) = k)$$

es la probabilidad de que la población crezca en $j - k$ células en s unidades de tiempo. Más aún, empleando las ecuaciones (4.4) y (4.6) se obtiene que

$$\begin{aligned}
P(X(t+s) - X(t) = 0 \mid X(t) = k) &= P(X(t+s) = k \mid X(t) = k) \\
&= 1 - \lambda is + \mathcal{O}_1(s)
\end{aligned}$$

y

$$\begin{aligned}
P(X(t+s) - X(t) = 1 \mid X(t) = k) &= P(X(t+s) = k + 1 \mid X(t) = k) \\
&= \lambda is + \mathcal{O}_2(s),
\end{aligned}$$

respectivamente.

Por tanto,

$$\begin{aligned} P(X(t+s) - X(t) < 2 \mid X(t) = k) &= P(X(t+s) - X(t) = 0 \mid X(t) = k) \\ &\quad + P(X(t+s) - X(t) = 1 \mid X(t) = k) \\ &= 1 + \mathcal{O}_1(s) + \mathcal{O}_2(s). \end{aligned}$$

Lo que implica que

$$P(X(t+s) - X(t) = j - k \mid X(t) = k) \leq -\mathcal{O}_1(s) - \mathcal{O}_2(s).$$

De esta manera, se obtiene que para $k = 0, 1, \dots, j-2$

$$P(X(t+s) - X(t) = j - k \mid X(t) = k) = \mathcal{O}_{3,k}(s), \quad (4.18)$$

donde $\lim_{s \rightarrow 0} \frac{\mathcal{O}_{3,k}(s)}{s} = 0$.

Luego,

$$\begin{aligned} P_j(t+s) &= \sum_{k=0}^j P(X(t+s) - X(t) = j - k \mid X(t) = k) P_k(t) \\ &= P(X(t+s) - X(t) = 0 \mid X(t) = j) P_j(t) \\ &\quad + P(X(t+s) - X(t) = 1 \mid X(t) = j-1) P_{j-1}(t) \\ &\quad + \sum_{k=1}^{j-2} \mathcal{O}_{3,k}(s) P_k(t) \\ &= (1 - \lambda j s + \mathcal{O}_1(s)) P_j(t) + (\lambda(j-1)s + \mathcal{O}_2(s)) P_{j-1}(t) \\ &\quad + \sum_{k=1}^{j-2} \mathcal{O}_{3,k}(s) P_k(t). \end{aligned}$$

Por tanto, al despejar a $P_j(t)$ obtenemos que

$$\begin{aligned} P_j(t+s) - P_j(t) &= (-\lambda j s + \mathcal{O}_1(s)) P_j(t) \\ &\quad + (\lambda(j-1)s + \mathcal{O}_2(s)) P_{j-1}(t) \\ &\quad + \sum_{k=1}^{j-2} \mathcal{O}_{3,k}(s) P_k(t). \end{aligned} \quad (4.19)$$

Implicando que

$$P'_j(t) = \lim_{s \rightarrow 0} \frac{P_j(t+s) - P_j(t)}{s} = -\lambda j P_j(t) + \lambda(j-1) P_{j-1}(t). \quad (4.20)$$

Para finalizar, notemos que si $j = k_0$ se tiene que $P_{k_0-1}(t) = P(X(t) = k_0 - 1 \mid X(0) = k_0) = 0$ pues es un proceso de nacimiento puro. Así que

$$P'_{k_0}(t) = -\lambda k_0 P_{k_0}(t), \quad (4.21)$$

obteniendo ambas ecuaciones deseadas. ■

Antes de proceder a hallar la solución de las ecuaciones retrospectivas de Kolmogorov para el proceso de Yule-Furry observemos que podemos dotar a dichas ecuaciones diferenciales de condiciones iniciales por realizar el siguiente análisis.

Lema 7 Sea $X = \{X(t) \mid t \geq 0\}$ un proceso de Yule-Furry y sea $j \in \mathbb{N}$. Entonces

$$P_j(0) = P(X(0) = j \mid X(0) = k_0) = \begin{cases} 1, & \text{si } j = k_0. \\ 0, & \text{en otro caso.} \end{cases} \quad (4.22)$$

Demostración. Sea $Y := X(0)$ y consideremos los eventos

$$E_i = \{\omega \in \Omega \mid Y(\omega) = i\},$$

con $i \in \{k_0, j\}$.

Notemos que E_{k_0} es no vacío, pues por hipótesis hemos supuesto que la población ha comenzado con k_0 células. Luego, si $j = k_0$ entonces $E_{k_0} \cap E_j \neq \emptyset$, dado que $E_{k_0} = E_j$ y E_{k_0} es no vacío.

Por otra parte, para $j \in \mathbb{N} \setminus \{k_0\}$, si existiera algún $\omega_0 \in E_{k_0}$ tal que $\omega_0 \in E_j$ implicaría que $j = Y(\omega_0) = k_0$, una contradicción puesto que $j \neq k_0$. Así, se tiene que

$$E_{k_0} \cap E_j = \begin{cases} E_{k_0}, & \text{si } j = k_0, \\ \emptyset, & \text{en otro caso.} \end{cases}$$

Lo anterior implica que

$$P(X(0) = k_0, X(0) = j) = P(E_{k_0} \cap E_j) = \begin{cases} P(X(0) = k_0), & \text{si } j = k_0, \\ 0, & \text{en otro caso.} \end{cases}$$

Luego, de la definición de probabilidad condicional obtenemos que

$$P_j(0) = P(X(0) = j \mid X(0) = k_0) = \begin{cases} 1, & \text{si } j = k_0, \\ 0, & \text{en otro caso.} \end{cases}$$

Mostrando lo afirmado. ■

Considerando la ecuaciones retrospectivas del Corolario 1 y el análisis previo podemos plantear los siguientes problemas de valores iniciales:

$$\begin{cases} P'_{k_0}(t) = -\lambda k_0 P_{k_0}(t), \\ P_{k_0}(0) = 1, \end{cases} \quad (4.23)$$

y

$$\begin{cases} P'_j(t) = -\lambda j P_j(t) + \lambda(j-1)P_{j-1}(t), \\ P_j(0) = 0, \end{cases} \quad (4.24)$$

para $j > k_0$, cuyas soluciones presentaremos en el siguiente teorema.

Teorema 8 Dadas $k_0 \in \mathbb{N}$ y $\lambda \in \mathbb{R}$, sea $f: \{k_0, k_0 + 1, \dots\} \times [0, \infty) \rightarrow \mathbb{R}$ una función definida por

$$f(j, t) = \binom{j-1}{k_0-1} e^{-k_0 \lambda t} (1 - e^{-\lambda t})^{j-k_0}. \quad (4.25)$$

La solución de los problemas de valores iniciales

$$\begin{cases} P'_{k_0}(t) = -\lambda k_0 P_{k_0}(t), \\ P_{k_0}(0) = 1, \end{cases} \quad (4.26)$$

y

$$\begin{cases} P'_j(t) = -\lambda j P_j(t) + \lambda(j-1)P_{j-1}(t), \\ P_j(0) = 0, \end{cases} \quad (4.27)$$

está dada por $f(k_0, t)$ y $f(j, t)$, respectivamente.

Demostración. Consideremos el problema de valores iniciales

$$\begin{cases} P'_{k_0}(t) = -\lambda k_0 P_{k_0}(t), \\ P_{k_0}(0) = 1. \end{cases}$$

Resolviendo la ecuación diferencial se tiene que

$$\log(P_{k_0}(t)) = -\lambda k_0 t + c,$$

con $c \in \mathbb{R}$ una constante. Lo anterior implica que

$$P_{k_0}(t) = c_1 e^{-\lambda k_0 t}, \quad (4.28)$$

donde $c_1 \in \mathbb{R}$ es una constante, la cual, al considerar la condición inicial se tiene que $c_1 = 1$. Así es que

$$\begin{aligned} P_{k_0}(t) &= e^{-\lambda k_0 t} \\ &= \binom{k_0 - 1}{k_0 - 1} e^{-\lambda k_0 t} (1 - e^{-\lambda t})^{k_0 - k_0} \end{aligned} \quad (4.29)$$

Para $j > k_0$, definamos $Q_j(t) := e^{j\lambda t} P_j(t)$, entonces

$$\begin{aligned} Q'_j(t) &= j\lambda e^{j\lambda t} P_j(t) + e^{j\lambda t} P'_j(t) \\ &= e^{j\lambda t} (j\lambda P_j(t) + P'_j(t)). \end{aligned}$$

Luego, de la Ecuación 4.17 en el Corolario 1 sabemos que

$$P'_j(t) = -\lambda j P_j(t) + \lambda(j-1)P_{j-1}(t).$$

Por tanto

$$Q'_j(t) = e^{j\lambda t} \lambda(j-1)P_{j-1}(t),$$

donde su solución está dada por

$$Q_j(t) = \lambda(j-1) \int_0^t e^{j\lambda x} P_{j-1}(x) dx \quad (4.30)$$

o equivalentemente,

$$P_j(t) = e^{-j\lambda t} \lambda(j-1) \int_0^t e^{j\lambda x} P_{j-1}(x) dx.$$

Como $j, k \in \mathbb{N}$ y $j > k$ entonces existe una única $l \in \mathbb{N}$ tal que $j = k_0 + l$. De esta manera, mostrar que

$$P_j(t) = \binom{j-1}{k_0-1} e^{-k_0 \lambda t} (1 - e^{-\lambda t})^{j-k_0},$$

para toda $j > k_0$, es equivalente a mostrar que

$$P_{k_0+l}(t) = \binom{(k_0+l)-1}{k_0-1} e^{-k_0\lambda t} (1 - e^{-\lambda t})^{(k_0+l)-k_0}, \quad (4.31)$$

para toda $l \in \mathbb{N}$.

Caso $l = 1$. La Ecuación 4.30 implica que para $j = k_0 + 1$ se tiene que

$$\begin{aligned} P_{k_0+1}(t) &= e^{-(k_0+1)\lambda t} \lambda k_0 \int_0^t e^{(k_0+1)\lambda x} P_{k_0}(x) dx \\ &= e^{-(k_0+1)\lambda t} \lambda k_0 \int_0^t e^{(k_0+1)\lambda x} e^{-\lambda k_0 x} dx \\ &= e^{-(k_0+1)\lambda t} k_0 (e^{\lambda t} - 1). \end{aligned}$$

Luego, como $k_0 = \binom{(k_0+1)-1}{k_0-1}$ y

$$\begin{aligned} e^{-(k_0+1)\lambda t} (e^{\lambda t} - 1) &= e^{-k_0\lambda t} - e^{-(k_0+1)\lambda t} \\ &= e^{-k_0\lambda t} (1 - e^{-\lambda t})^{(k_0+1)-(k_0)} \end{aligned}$$

se sigue que

$$P_{k_0+1}(t) = \binom{(k_0+1)-1}{k_0-1} e^{-k_0\lambda t} (1 - e^{-\lambda t})^{(k_0+1)-k_0}, \quad (4.32)$$

lo que concluye este caso.

Supongamos ahora que la Ecuación (4.31) es cierta para $l = l_0$ y procedamos a verificar el caso $l = l_0 + 1$.

Caso $l = l_0 + 1$. Aplicando la Ecuación 4.30, para $j = k_0 + l_0 + 1$ tenemos que

$$\begin{aligned} P_{k_0+l_0+1}(t) &= e^{-(k_0+l_0+1)\lambda t} \lambda (k_0 + l_0) \int_0^t e^{(k_0+l_0+1)\lambda x} P_{k_0+l_0}(x) dx \\ &= \binom{(k_0+l_0)-1}{k_0-1} e^{-(k_0+l_0+1)\lambda t} \lambda (k_0 + l_0) \int_0^t e^{(k_0+l_0+1)\lambda x} e^{-k_0\lambda x} (1 - e^{-\lambda x})^{l_0} dx \\ &= \binom{(k_0+l_0)-1}{k_0-1} e^{-(k_0+l_0+1)\lambda t} \lambda (k_0 + l_0) \int_0^t e^{(l_0+1)\lambda x} (1 - e^{-\lambda x})^{l_0} dx, \end{aligned}$$

por la hipótesis de inducción.

Luego, por el teorema del binomio sabemos que

$$(1 - e^{-\lambda x})^{l_0} = \sum_{m=0}^{l_0} \binom{l_0}{m} (-e^{-\lambda x})^m.$$

Más aún, aplicando la linealidad de la integral se obtiene que

$$\begin{aligned} \int_0^t e^{(l_0+1)\lambda x} (1 - e^{-\lambda x})^{l_0} dx &= \sum_{m=0}^{l_0} \binom{l_0}{m} \int_0^t e^{(l_0+1)\lambda x} (-e^{-\lambda x})^m dx \\ &= \sum_{m=0}^{l_0} \binom{l_0}{m} (-1)^m \int_0^t e^{(l_0+1)\lambda x} e^{-\lambda m x} dx \\ &= \sum_{m=0}^{l_0} \binom{l_0}{m} (-1)^m \int_0^t e^{(l_0+1-m)\lambda x} dx \end{aligned}$$

Donde

$$\int_0^t e^{(l_0+1-m)\lambda x} dx = \frac{1}{(l_0+1-m)\lambda} (e^{(l_0+1-m)\lambda t} - 1)$$

. Lo cual implica que

$$\begin{aligned} \int_0^t e^{(l_0+1)\lambda x} (1 - e^{-\lambda x})^{l_0} dx &= \sum_{m=0}^{l_0} \binom{l_0}{m} (-1)^m \frac{1}{(l_0+1-m)\lambda} (e^{(l_0+1-m)\lambda t} - 1) \\ &= \frac{1}{(l_0+1)\lambda} \left(\sum_{m=0}^{l_0} \binom{l_0+1}{m} (-1)^m (e^{(l_0+1-m)\lambda t} - 1) \right. \\ &\quad \left. + (-1)^{l_0+1} (1-1) - (-1)^{l_0+1} (1-1) \right) \\ &= \frac{1}{(l_0+1)\lambda} \sum_{m=0}^{l_0+1} \binom{l_0+1}{m} (-1)^m (e^{(l_0+1-m)\lambda t} - 1). \end{aligned}$$

Nuevamente, por aplicar el teorema del binomio se sigue que

$$\sum_{m=0}^{l_0+1} \binom{l_0+1}{m} (-1)^m e^{(l_0+1-m)\lambda t} = (-1 + e^{\lambda t})^{l_0+1}$$

y

$$\begin{aligned} \sum_{m=0}^{l_0+1} \binom{l_0+1}{m} (-1)^m &= \sum_{m=0}^{l_0+1} \binom{l_0+1}{m} (-1)^m 1^{l_0+1-m} \\ &= (-1 + 1)^{l_0+1} \\ &= 0. \end{aligned}$$

Por tanto

$$\int_0^t e^{(l_0+1)\lambda x} (1 - e^{-\lambda x})^{l_0} dx = \frac{1}{(l_0 + 1)\lambda} (-1 + e^{\lambda t})^{l_0+1}.$$

Implicando que

$$\begin{aligned} P_{k_0+l_0+1}(t) &= \binom{(k_0 + l_0) - 1}{k_0 - 1} e^{-(k_0+l_0+1)\lambda t} \lambda (k_0 + l_0) \int_0^t e^{(l_0+1)\lambda x} (1 - e^{-\lambda x})^{l_0} dx \\ &= \frac{(k_0 + l_0)!}{(k_0 - 1)!(l_0 + 1)!} e^{-k_0\lambda t} (e^{-\lambda t})^{l_0+1} (-1 + e^{\lambda t})^{l_0+1} \\ &= \binom{(k_0 + l_0 + 1) - 1}{k_0 - 1} e^{-k_0\lambda t} (1 - e^{-\lambda t})^{l_0+1}, \end{aligned}$$

lo enunciado.

Por último, verificamos que $P_j(0) = 0$, para toda $j > k_0$. Esto se obtiene directamente dado que el factor $(1 - e^{-\lambda t})^j$ es igual a cero al tiempo $t = 0$. Lo que concluye la demostración. ■

Del teorema previo se desprenden los siguientes corolarios.

Corolario 2 Sea $X = \{X(t) \mid t \in [0, \infty)\}$ un proceso de Yule-Furry. Entonces

$$(X(t) \mid X(0) = k_0) \sim \text{BN}(k_0, e^{-\lambda t}), \quad (4.33)$$

para toda $t \geq 0$.

Demostración. Sea $p: [0, \infty) \rightarrow \mathbb{R}$ una función definida por

$$p(t) = e^{-\lambda t}, \quad (4.34)$$

con $\lambda > 0$. Notemos que $0 \leq p(t) \leq 1$ para toda $t \in [0, \infty)$ pues p es una función positiva, monótona decreciente y al tiempo $t = 0$ se tiene que $p(0) = 1$.

Luego, recordemos que si una variable aleatoria X posee la siguiente función de masa de probabilidad

$$f_X(x) = \binom{x-1}{n-1} p^n (1-p)^{x-n}, \quad (4.35)$$

para $x = n, n+1, \dots$ entonces decimos que X tiene una distribución binomial negativa con parámetros $n \in \mathbb{N}$ y $0 \leq p \leq 1$, o bien, en símbolos escribimos $X \sim \text{BN}(n, p)$.

Por el Teorema 8, sabemos que dada $j \in \{k_0, k_0 + 1, \dots\}$

$$P(X(t) = j \mid X(0) = k_0) = \binom{j-1}{k_0-1} e^{-k_0\lambda t} (1 - e^{-\lambda t})^{j-k_0}$$

denota la probabilidad de que la población fundada a partir de k_0 células alcance un tamaño de j individuos al tiempo t .

De esta manera, variable aleatoria $X(t) \mid X(0) = k_0$ que cuenta tamaño al tiempo t de una población celular que se fundó inicialmente con k_0 individuos satisface que

$$(X(t) \mid X(0) = k_0) \sim \text{BN}(k_0, e^{-\lambda t}). \quad (4.36)$$

■

Y finalmente, enunciamos el resultado deseado que caracteriza la distribución de la variable aleatoria que cuenta el número total de células existentes al tiempo t_f en un cultivo celular el cual que fue fundado, al tiempo t_0 , con k_0 células iniciales que crecen a una tasa λ .

Corolario 3 Sea $X = \{X(t) \mid t \in [0, \infty)\}$ un proceso de Yule-Furry. Dado $t_0 \geq 0$ se cumple que

$$(X(t) \mid X(t_0) = k_0) \sim \text{BN}(k_0, e^{-\lambda(t-t_0)}), \quad (4.37)$$

para toda $t \geq t_0$.

Demostración. Sea $j \in \{k_0, k_0 + 1, \dots\}$ arbitraria y fija. Notemos que dad $t \geq t_0$

$$P(X(t) = j \mid X(t_0) = k_0) = P(X([t - t_0] + t_0) = j \mid X(t_0) = k_0).$$

Luego, ya que X tiene probabilidades de transición estacionarias, se sigue que

$$P(X([t - t_0] + t_0) = j \mid X(t_0) = k_0) = P(X(t - t_0) = j \mid X(0) = k_0),$$

donde

$$P(X(t - t_0) = j \mid X(0) = k_0) = \binom{j-1}{k_0-1} e^{-k_0\lambda(t-t_0)} (1 - e^{-\lambda(t-t_0)})^{j-k_0},$$

por el Corolario 2.

Finalmente, como $j \in \{k_0, k_0 + 1, \dots\}$ fue arbitraria se tiene que

$$P(X(t) = j \mid X(t_0) = k_0) = \binom{j-1}{k_0-1} e^{-k_0\lambda(t-t_0)} (1 - e^{-\lambda(t-t_0)})^{j-k_0}, \quad (4.38)$$

para toda $j \in \{k_0, k_0 + 1, \dots\}$. Concluyendo la prueba. ■

Para concluir esta sección, discutiremos algunos resultados sencillos, pero cruciales para el desarrollo del modelo gráfico.

Obsevación 2 Sea X la variable aleatoria que cuenta el número de células mutantes dentro de una población de N_f células. Si las células mutan de manera independiente y $0 \leq \mu \leq 1$ denota la probabilidad de que una célula mute, entonces modelar la probabilidad de que la célula c_i mute, $i = 1, \dots, N_f$, a través de una variable aleatoria X_i tal que

$$P(X_i = x) = \begin{cases} \mu, & \text{si } x = 1, \\ 1 - \mu, & \text{si } x = 0, \end{cases}$$

es decir, $X_i \sim \text{Bernoulli}(\mu)$, para cada célula c_i , donde el evento $\{X_i = 1\}$ significa que una mutación ocurre; mientras que, el evento $\{X_i = 0\}$ significa que una mutación no ocurre. De esta manera, se tiene que $X \sim \text{Binomial}(N_f, \mu)$.

Recurrir a la distribución binomial cuando el número de ensayos Bernoulli considerados es *suficientemente grande* suele representar un problema computacional, pues debe calcularse el factorial de números de la misma magnitud. Por ello, presentaremos el teorema de la ley de los pequeños números, el cual hace uso del siguiente resultado clásico.

Lema 9 Sea $x \in \mathbb{R}$. Entonces

$$\lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = e^x. \quad (4.39)$$

Demostración. Sabemos que $a^x = e^{x \log(a)}$, para toda $x \in \mathbb{R}$ y para toda $a \geq 0$. Así que $\left(1 - \frac{x}{n}\right)^n = \exp\left(n \log\left(1 - \frac{x}{n}\right)\right)$, lo cual implica que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = \lim_{n \rightarrow \infty} \exp\left(n \log\left(1 - \frac{x}{n}\right)\right) \quad (4.40)$$

$$= \exp\left(\lim_{n \rightarrow \infty} n \log\left(1 - \frac{x}{n}\right)\right) \quad (4.41)$$

$$= \exp\left(\lim_{n \rightarrow \infty} \frac{\log\left(1 - \frac{x}{n}\right)}{\frac{1}{n}}\right). \quad (4.42)$$

Aplicando la regla de L'Hopital obtenemos que

$$\lim_{n \rightarrow \infty} \frac{\log\left(1 - \frac{x}{n}\right)}{\frac{1}{n}} = \lim_{n \rightarrow \infty} \frac{\frac{-\frac{x}{n^2}}{1 - \frac{x}{n}}}{-\frac{1}{n^2}} \quad (4.43)$$

$$= \lim_{n \rightarrow \infty} \frac{x}{1 - \frac{x}{n}} \quad (4.44)$$

$$= x. \quad (4.45)$$

Tras sustituir la Ecuación (4.43) en la Ecuación (4.42) se tiene que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = e^x,$$

lo deseado. ■

Procedemos ahora a presentar el teorema de la ley de los pequeños números, que una aproximación a la distribución binomial de parámetros n y p , cuando n es *suficientemente grande*.

Teorema 10 (Ley de los pequeños números) Sean $n \in \mathbb{N}$ y $p \in [0, 1]$. Si $X \sim \text{Binomial}(n, p)$ y $\lambda = np$, entonces

$$\lim_{n \rightarrow \infty} P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda},$$

para cada $x \in \{0, 1, \dots\}$.

Demostración. Sea $x \in \{0, 1, \dots, n\}$ un número arbitrario y fijo. Sabemos que

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad (4.46)$$

dado que $X \sim \text{Binomial}(n, p)$.

Como $p = \frac{\lambda}{n}$, al tomar el límite cuando $n \rightarrow \infty$ en la Ecuación (4.46) se tiene que

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(X = x) &= \lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \lim_{n \rightarrow \infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
 &= \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
 &= \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \left(1 - \frac{0}{n}\right) \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{-x+1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
 &= \frac{\lambda^x}{x!} e^{-\lambda},
 \end{aligned}$$

al aplicar el Lema 9. ■

Obsevación 3 En contextos experimentales se tiene que N_f y μ , los parámetros discutidos en la Observación 2, satisfacen las hipótesis del Teorema 10 por lo que, sin pérdida de generalidad, podemos suponer que

$$X \sim \text{Poisson}(N_f \mu). \quad (4.47)$$

Más aún, esto permite considerar valores no enteros de N_f , es decir, extender a $N_f \in \mathbb{R}^+$, dado que la distribución Poisson, con parámetro λ , está definida para todo $\lambda > 0$.

4.1.2. Crecimiento poblacional

Biológicamente, en poblaciones celulares podemos distinguir cuatro fases:

1. Fase de retraso,
2. Fase de crecimiento exponencial,
3. Fase estacionaria, y
4. Fase de muerte.

La fase de retraso se caracteriza por la adaptación de las células al ambiente [117]. La duración de esta fase depende del número inicial de células en el ambiente, además de factores físicos, químicos y fisiológicos [135]; por ejemplo, la temperatura del ambiente,

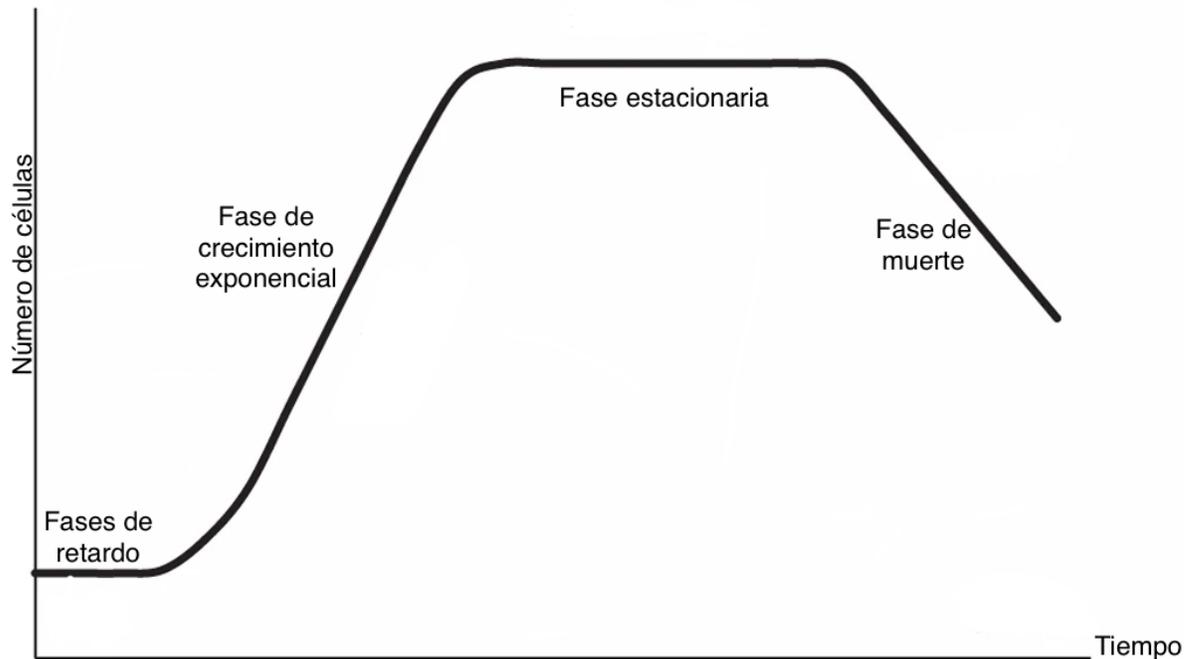


Figura 4.1: Curva de crecimiento celular [147].

el pH del ambiente y la actividad metabólica de las células, respectivamente. La fase de crecimiento exponencial comienza cuando las células comienzan el proceso de reproducción por fisión binaria y el número de células crece acorde a una función exponencial [93]. Posteriormente, en la fase estacionaria la reproducción cesa, pero las células permanecen metabólicamente activas [70]. Finalmente, en la fase de muerte se observa una pérdida neta en la cantidad de células en la población y, generalmente, esto también ocurre exponencialmente, pero más lento en comparación con la fase de crecimiento exponencial [93].

En la Figura 4.1, modificada de [147], podemos apreciar de manera gráfica el comportamiento de la población en cada fase y la transición entre fases a lo largo del tiempo. Notemos que en la fase de retardo el número de células se mantiene constante hasta antes de comenzar la fase de crecimiento exponencial basta considerar el comportamiento en dicha fase. Así que mantendremos este mismo esquema para obtener un modelo tal que las simulaciones construidas a partir de este sean *similares* a las que puede obtenerse con el clásico modelo Moran.

Intuitivamente, para modelar la fase de crecimiento exponencial en un cultivo celular se puede considerar el siguiente planteamiento [90].

Sea $N: [0, \infty) \rightarrow [N_0, \infty)$ una función tal que $N(t)$ indica el número de bacterias que crecen en un cultivo al tiempo t y $N_0 \in \mathbb{N}$ denota la cantidad inicial de células. Para emular el comportamiento que se observa en la fase de crecimiento exponencial, supongamos que la población celular crece a una tasa directamente proporcional a $N(t)$. Esto permite plantear el siguiente problema de valores iniciales

$$\begin{cases} \frac{dN(t)}{dt} = N(t), \\ N(0) = N_0. \end{cases} \quad (4.48)$$

Dado que la ecuación diferencial asociada al problema anterior es una ecuación de variables separables la solución a la ecuación diferencial es $N(t) = k_1 e^t$, con $k_1 \in \mathbb{R}$ una constante. Más aún, al considerar la condición inicial, se tiene que

$$N(t) = N_0 e^t \quad (4.49)$$

es la solución al problema de condiciones iniciales.

Supongamos que las células en la población poseen la misma tasa de mutación μ y que mutan de manera independiente; además, supongamos que inicialmente hay $M_0 \in \mathbb{N}$ células mutantes.

Sea $M: [0, \infty) \rightarrow \mathbb{R}$ una función tal que $M(t)$ denota el número de mutaciones al tiempo t . Notemos que el número de mutaciones incrementa por dos factores: las mutaciones que se han originado de las M_0 células mutantes iniciales y por el número esperado de células mutantes que aparecen aleatoriamente al tiempo t , el cual, por la Observación 3, se sabe que es

$$\mu N(t). \quad (4.50)$$

En principio, para incorporar la hipótesis de crecimiento exponencial podríamos suponer que M crece a una tasa $M(t)$. Sin embargo, para incluir el número de mutaciones que se originan por el azar se supondrá que M crece a una tasa $M(t) + \mu N(t)$. Lo cual, nos permite escribir el siguiente problema de valores iniciales

$$\begin{cases} \frac{dM(t)}{dt} = M(t) + \mu N(t), \\ M(0) = M_0. \end{cases} \quad (4.51)$$

Cuya solución es

$$M(t) = \mu t N(t) + M_0 e^t, \quad (4.52)$$

dado que al multiplicar por e^{-t} a la ecuación diferencial

$$\frac{dM(t)}{dt} = M(t) + \mu N(t)$$

se obtiene que

$$\frac{d}{dt} (e^{-t} M(t)) = e^{-t} \mu N(t). \quad (4.53)$$

Implicando que

$$\begin{aligned} e^{-t} M(t) &= \int e^{-t} \mu N(t) dt \\ &\stackrel{\text{Ec. (4.49)}}{=} \int \mu N_0 dt \\ &= \mu N_0 t + k_3, \end{aligned}$$

con $k_3 \in \mathbb{R}$, una constante. Luego, al despejar, se obtiene que

$$M(t) = \mu N_0 t e^t + k_3 e^t. \quad (4.54)$$

Finalmente, de acuerdo a la Ecuación (4.51), se satisface que $M(0) = M_0$, lo cual implica que $k_3 = M_0$ y por tanto

$$M(t) = \mu t N(t) + M_0 e^t, \quad (4.55)$$

lo deseado.

4.1.3. Algoritmo

Al analizar el planteamiento discutido en la sección previa, podemos notar que las siguientes suposiciones fueron realizadas:

1. La población crece en un espacio ilimitado,
2. Los recursos son inagotables, y

3. El experimento se observa de manera perpetua.

Dichas suposiciones resultan útiles, pero poco apegadas al contexto biológico. Es así que, a continuación, modificamos el planteamiento previo para darle un enfoque más realista.

Consideremos un cultivo celular que crece a una tasa constante g e inicia con una población de tamaño N_0 . Supongamos que las condiciones de espacio y nutrientes permiten la coexistencia de, a lo más, N_f células. Sea $t_f \in \mathbb{R}^+$ el tiempo requerido para que la población alcance el tamaño máximo; por convención, es usual medir el tiempo en generaciones y ya que las células se reproducen por fisión binaria se tiene que t_f es tal que $2^{t_f} = N_f$, o equivalentemente, $\log_2(N_f) = t_f$.

Partiendo de lo anterior, si $N: [0, t_f] \rightarrow [N_0, N_f]$ es una función tal que $N(t)$ denota el tamaño de la población al tiempo t , formulamos el siguiente problema de valores iniciales

$$\begin{cases} \frac{dN(t)}{dt} = gN(t), \\ N(0) = N_0. \end{cases} \quad (4.56)$$

Cuya solución es

$$N_0 e^{gt}. \quad (4.57)$$

Dado que, al integrar, se obtiene

$$\int \frac{dN(t)}{N(t)} dt = \log(N(t)) = \int g dt = gt + k_4,$$

con $k_4 \in \mathbb{R}$, una constante. Luego, al despejar a $N(t)$, se tiene que

$$N(t) = k_5 e^{gt}, \quad (4.58)$$

con $k_5 \in \mathbb{R}$, una constante. Por último, se requiere de que $N(0) = N_0$, lo cual se satisface al fijar $k_5 = N_0$ y determina la solución del problema de valores iniciales (4.56).

Notemos que al evaluar la función N al tiempo $t = t_f$ se cumple que

$$N(t_f) = N_f.$$

De donde se obtiene la relación

$$g = \frac{1}{t_f} \log \left(\frac{N_f}{N_0} \right), \quad (4.59)$$

la cual determina una expresión para la tasa de crecimiento.

Ahora, sea $M: [0, t_f] \rightarrow \mathbb{R}$ una función tal que $M(t)$ denota el número de mutaciones al tiempo t y sea M_0 la cantidad inicial de células mutantes. Procediendo como en el planteamiento para la Ecuación (4.51), el número de células mutantes crece a una tasa $\mu N(t) + gM(t)$, dado que el número esperado de células mutantes nuevas que aparecen aleatoriamente es $\mu N(t)$ y las mutaciones que se originan a partir de la descendencia de las M_0 células iniciales, las cuales crecen a una tasa g .

Con el planteamiento anterior podemos formular el siguiente problema de valores iniciales

$$\begin{cases} \frac{dM(t)}{dt} = \mu N(t) + gM(t), \\ M(0) = M_0. \end{cases} \quad (4.60)$$

Cuya solución es

$$M(t) = \mu t N(t) + M_0 e^{gt}, \quad (4.61)$$

puesto que al multiplicar por e^{-gt} a la ecuación diferencial

$$\frac{dM(t)}{dt} = \mu N(t) + gM(t)$$

obtenemos

$$\frac{d}{dt} (e^{-gt} M(t)) = \mu N(t). \quad (4.62)$$

Lo cual, al integrar, implica que

$$\begin{aligned} e^{-gt} M(t) &= \int e^{-gt} \mu N(t) dt \\ &\stackrel{\text{Ec. (4.49)}}{=} \int \mu N_0 dt \\ &= \mu N_0 t + k_6, \end{aligned}$$

con $k_6 \in \mathbb{R}$ constante.

Luego, al despejar a $M(t)$ se obtiene

$$M(t) = \mu t N(t) + k_6 e^{gt}$$

y al considerar la condición inicial se concluye que

$$M(t) = \mu t N(t) + M_0 e^{gt} \quad (4.63)$$

es la solución al problema de valores iniciales.

Observemos que el crecimiento de la población celular, descrito a través de la Ecuación (4.57), posee un comportamiento estrictamente determinista. Mientras que, el número de mutaciones, descrito a través de la Ecuación (4.61) se compone por un factor determinista y un factor aleatorio.

Por una parte, el factor determinista resulta análogo al comportamiento del crecimiento de la población y, por otra parte, el factor aleatorio se analizará con ayuda de las herramientas de probabilidad previamente discutidas.

Para proseguir, sean $\mu_{W \rightarrow R}$, $\mu_{W \rightarrow M}$ y $\mu_{M \rightarrow R}$ las tasas de mutación de células silvestres hacia células con fenotipo mutante, de células silvestres hacia células con fenotipo mutador y de células mutadoras hacia células mutantes con fenotipo mutador, respectivamente.

Supongamos que se inicia un cultivo celular con N_0 células silvestres y $M_0 = 0$ células con fenotipo mutante. Por tanto, la Ecuación (4.61) implica que el número esperado de mutaciones al tiempo $t = t_f$ es $t_f N_f \mu_{W \rightarrow R}$.

Dado que presentar una mutación, probabilísticamente, es un ensayo Bernoulli y el producto $t_f N_f$, en contextos biológico, es *suficientemente grande*, entonces el Teorema 10, la ley de los pequeños números, implica que si $M_{W \rightarrow R}$ es la variable aleatoria que cuenta el número de tipos de mutaciones en el cultivo al tiempo t_f , se cumple que

$$M_{W \rightarrow R} \sim \text{Poisson}(t_f N_f \mu_{W \rightarrow R}). \quad (4.64)$$

Biológicamente, no todas las mutaciones que ocurren en un organismo tienen efecto sobre dicho individuo. Esto, dado a que pueden ser neutrales, implicando que no hay una real diferencia con las células silvestres, o pueden ser corregidas de alguna manera.

Por tanto, se requiere de un método de decisión para determinar cuántos y cuáles tipos de mutaciones tienen presencia en el cultivo. Más aún, conocer el momento en que dichas mutaciones aparecen en la población. Para ello consideremos $f_{W \rightarrow R} : [0, t_f] \rightarrow [0, 1]$ una función definida por

$$f_{W \rightarrow R}(t) = \frac{1}{N_f} N_0 e^{gt}, \quad (4.65)$$

esto es, $f_{W \rightarrow R}(t)$ denota la proporción del tamaño poblacional al tiempo t respecto al tamaño máximo de la misma.

La función $f_{W \rightarrow R}$, vista desde la perspectiva probabilística, es una distribución impropia. Esto es, existe una constante $c \in \mathbb{R}$ tal que

$$\int_0^{t_f} f_{W \rightarrow R}(t) dt = c,$$

con $c \neq 1$. Por tanto, si se construye un histograma con una muestra aleatoria cuya distribución común sea $f_{W \rightarrow R}$, se cumple que dicho histograma queda bajo la gráfica de $f_{W \rightarrow R}$.

Lo anterior nos dota de un criterio de aceptación/rechazo en el que el propio comportamiento del crecimiento poblacional es el factor para realizar la decisión. De esta manera dispersamos $m_{W \rightarrow R}$ puntos, con $m_{W \rightarrow R}$ una realización de $M_{W \rightarrow R}$, en el conjunto $[0, t_f] \times [0, 1]$.

Para que la dispersión de los puntos sea insesgada procedamos de la manera siguiente. Sea $C = (T, P)$ un vector aleatorio tal que

$$T \sim \text{Uniforme}(0, t_f) \quad \text{y} \quad P \sim \text{Uniforme}(0, 1). \quad (4.66)$$

La primera entrada de C nos da información sobre el momento en que los tipos de mutaciones aparecen; mientras que, la segunda entrada de C modula la posición vertical de los puntos, indicando la proporción de la población a la cual dicha célula pertenece.

Luego, cada punto p_i , $i = 1, 2, \dots, m_{W \rightarrow R}$, se determina como una realización de C , es decir, $\{p_1 = (T_1, P_1), p_2 = (T_2, P_2), \dots, p_{m_{W \rightarrow R}} = (T_{m_{W \rightarrow R}}, P_{m_{W \rightarrow R}})\}$ es una muestra aleatoria de tamaño $m_{W \rightarrow R}$ del vector aleatorio C .

Lo anterior puede visualizarse en la Figura 4.2. La gráfica de la función $f_{W \rightarrow R}$ se aprecia en color negro, los puntos en color azul representan los tipos de mutaciones que son descartados y los puntos en color rojo representan los tipos de mutaciones presentes en el cultivo celular.

Tras el proceso de decisión, sea $m_{W \rightarrow R}^*$ el número de tipos de mutaciones presentes en la población celular, es decir, $m_{W \rightarrow R}^* \leq m_{W \rightarrow R}$ es el número de puntos bajo la gráfica de $f_{W \rightarrow R}$.

Reindexando, supongamos sin pérdida de generalidad que los puntos $p_1, p_2, \dots, p_{m_{W \rightarrow R}^*}$ denotan los tipos de mutaciones presentes en el cultivo celular. Notemos que el i -ésimo tipo

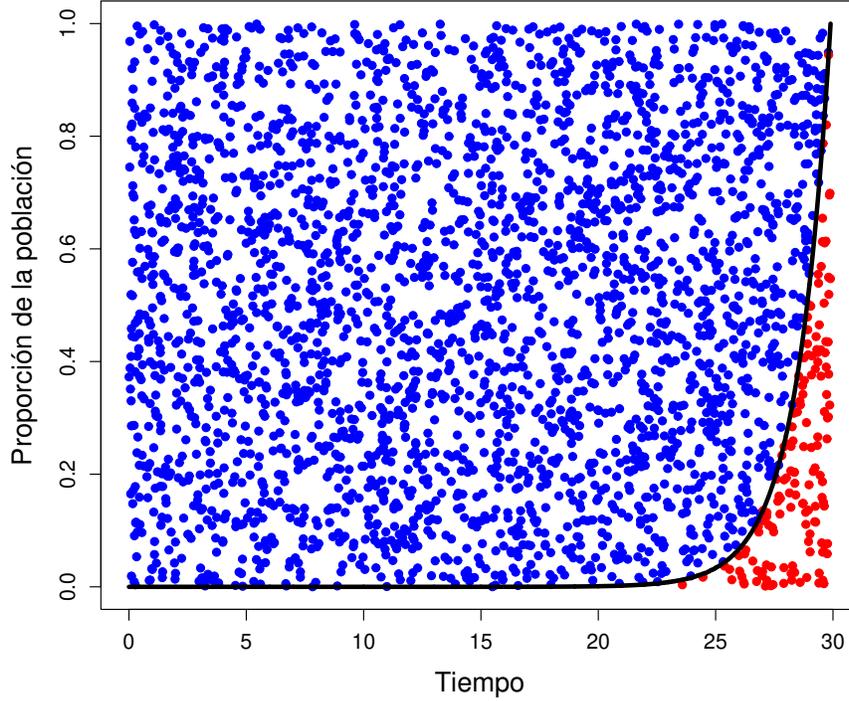


Figura 4.2: Método de decisión para determinar las mutaciones que se presentan en un cultivo celular. Para determinar las características del cultivo, los siguientes valores fueron considerados: $N_0 = 1$, $N_f = 10^9$ y $\mu_{W \rightarrow T} = 10^{-7}$.

de mutación, $i = 1, 2, \dots, m_{W \rightarrow R}^*$, está presente en una célula, digamos c_i , la cual cuenta desde el instante T_i al tiempo final t_f para dar origen a su descendencia clonal.

Notemos que al contar con una cantidad aleatoria de tiempo, el tamaño del linaje de la célula c_i es una variable aleatoria. Luego, ya que c_i mantiene la tasa de crecimiento g , se reproduce de manera independiente y hemos supuesto que las células no mueren, el Corolario 3 implica que la variable aleatoria, L_i , que cuenta el tamaño del linaje de la célula c_i es tal que

$$L_i \sim \text{BN} \left(1, e^{-g(t_f - T_i)} \right), \quad (4.67)$$

para $i = 1, 2, \dots, m_{W \rightarrow R}^*$.

La distribución anterior resulta de gran utilidad ya que, biológicamente, las células mutantes que aparecen en etapas tempranas tienen una mayor oportunidad de generar un linaje clonal robusto, en comparación con aquellas células mutantes que aparecen cerca del tiempo final. Más aún, con lo anterior se obtiene que el número total de células mutantes

en la población es

$$R = \sum_{i=1}^{m_{W \rightarrow R}^*} L_i. \quad (4.68)$$

Luego, supongamos que las células mutadoras provienen exclusivamente por mutaciones en las células silvestres. Así es que modelar la aparición de células mutadoras es análogo a lo presentado previamente. El cultivo se inicia con N_0 células silvestres y $M_0 = 0$ células con fenotipo mutador, que crecen exponencialmente a una tasa g .

La variable aleatoria $M_{W \rightarrow M}$ que cuenta los tipos de mutaciones que dan origen al fenotipo mutador en la población al tiempo t_f es tal que

$$M_{W \rightarrow M} \sim \text{Poisson}(t_f N_f \mu_{W \rightarrow M}), \quad (4.69)$$

por un argumento análogo al discutido para la Ecuación (4.64).

Notemos que el método de decisión que discutimos previamente no depende de la tasa de mutación $\mu_{W \rightarrow M}$. Por tanto, podemos aplicarlo directamente considerando una realización de $M_{W \rightarrow M}$ en lugar de una realización de $M_{W \rightarrow R}$.

Tras aplicar el método de decisión, se obtiene $m_{W \rightarrow M}^*$ el número de tipos de mutaciones presentes en la población, las cuales generan al fenotipo mutador, donde para cada tipo de mutación $p_i = (T_i, P_i)$, $i = 1, 2, \dots, m_{W \rightarrow M}^*$, la célula con dicha mutación, c_i , cuenta con $t_f - T_i$ unidades de tiempo para generar su linaje clonal de tamaño L_i , con L_i una variable aleatoria tal que

$$L_i \sim \text{BN}\left(1, e^{-g(t_f - T_i)}\right), \quad (4.70)$$

por un argumento análogo al usado en la Ecuación (4.67).

Esto implica que el número total de células mutadoras en el cultivo es

$$M = \sum_{i=1}^{m_{W \rightarrow M}^*} L_i, \quad (4.71)$$

una expresión similar la Ecuación (4.68).

Finalmente, para incorporar al modelo la aparición de células mutantes con fenotipo mutador, supongamos que dichas células aparecen como resultado de mutaciones dentro de células mutadoras, únicamente. Es decir, la dinámica anterior es un subproceso que ocurre

dentro del crecimiento de células mutadoras y por tanto, el Teorema 4 tiene un rol crucial en lo subsecuente.

Dado que las células mutantes con fenotipo mutador pueden aparecer exclusivamente dentro del linaje de células mutadoras, realizaremos un análisis individual a cada mutador.

Sea c_i , $i = 1, 2, \dots, m_{W \rightarrow M}$, la célula mutadora que posee el i -ésimo tipo de mutación que genera al fenotipo mutador y apareció al instante T_i . Considerando las ideas que motivan el Teorema 4, las células mutantes con un fenotipo mutador, cuyo ancestro es el mutador c_i , se producen en un cultivo celular que se inicia con $N_0 = 1$ células mutadoras y $M_0 = 0$ células mutantes con fenotipo mutador.

Bajo el planteamiento anterior, el problema de valores iniciales de la Ecuación (4.56) y su solución, se mantienen válidos, salvo que el rango y dominio de la función $N: [0, t_f] \rightarrow [N_0, N_f]$ se reescriben como $N: [0, t_f - T_i] \rightarrow [N_0, L_i]$.

Esto, debido a que L_i , la variable aleatoria que cuenta el tamaño de linaje de c_i , denota el tamaño máximo de células que se permiten en dicho cultivo y la población alcanza este tamaño en $t_f - T_i$ unidades de tiempo.

Luego, dado que $N(t_f - T_i) = L_i$ se obtiene una expresión para la tasa de crecimiento, g_i , que aplica dentro del ciclo de crecimiento del mutador c_i

$$g_i = \frac{1}{t_f - T_i} \log \left(\frac{L_i}{N_0} \right). \quad (4.72)$$

Más aún, el problema de valores iniciales en la Ecuación (4.60) y su solución también se mantienen válidas, salvo que, el rango y dominio de la función $M: [0, t_f] \rightarrow [M_0, \infty)$ se reescriben como $M: [0, t_f - T_i] \rightarrow [M_0, L_i]$.

Al evaluar la solución del problema de valores iniciales en la Ecuación (4.60) y argumentar como en la Ecuación (4.64), obtenemos que la variable aleatoria $M_{M_i \rightarrow R}$ que cuenta el número de tipos de mutaciones, al tiempo $t_f - T_i$, en el linaje de la célula que posee el i -ésimo tipo de mutación que genera al fenotipo mutador es tal que

$$M_{M_i \rightarrow R} \sim \text{Poisson} \left((t_f - T_i) L_i \mu_{M \rightarrow R} \right). \quad (4.73)$$

Notemos que la función en la Ecuación (4.65), que funge el papel discriminante en el proceso de decisión, depende del tamaño final de la población y la tasa de crecimiento. Y

dado que, en este caso, la población crece exponencialmente a una tasa g_i y el tamaño final de la población es L_i , consideremos $f_{M_i \rightarrow R} : [0, t_f - T_i] \rightarrow [0, 1]$ una función definida por

$$f_{M_i \rightarrow R}(t) = \frac{1}{L_i} e^{g_i t}. \quad (4.74)$$

Como previamente, la función $f_{M_i \rightarrow R}$ es una distribución impropia. Esto nos permite preservar la idea subyacente en el método de decisión. Dispersar $m_{M_i \rightarrow R}$ puntos, con $m_{M_i \rightarrow R}$ una realización de $M_{M_i \rightarrow R}$, en el conjunto $[0, t_f - T_i] \times [0, 1]$ y considerar a aquellos puntos bajo la gráfica de $f_{M_i \rightarrow R}$ como los tipos de mutaciones presentes en el ciclo de crecimiento de la célula mutador de tipo i .

Cada punto $p_{i,j}$, $j = 1, 2, \dots, m_{M_i \rightarrow R}$, queda determinado por una realización del vector aleatorio $D = (S, Q)$ tal que

$$S \sim \text{Uniforme}(0, t_f - T_i) \quad \text{y} \quad Q \sim \text{Uniforme}(0, 1). \quad (4.75)$$

Es decir, $\{p_{i,1} = (S_1, Q_1), p_{i,2} = (S_2, Q_2), \dots, p_{i,m_{M_i \rightarrow R}} = (S_{m_{M_i \rightarrow R}}, Q_{m_{M_i \rightarrow R}})\}$ es una muestra aleatoria de tamaño $m_{M_i \rightarrow R}$ del vector aleatorio D .

El procedimiento puede apreciarse en la Figura 4.3, que resulta similar a lo observado en la Figura 4.2. La gráfica de la función $f_{M_i \rightarrow R}$ se muestra en color negro, los puntos en color azul representan los tipos de mutaciones descartados y los puntos en color rojo representan los tipos de mutaciones que, en efecto, están presentes en la población clonal de la célula mutadora que posee el i -ésimo tipo de mutación que origina al fenotipo mutador.

Posterior al proceso de decisión, sea $m_{M_i \rightarrow R}^*$ el número de mutaciones presentes en la población. Esto es, $m_{M_i \rightarrow R}^* \leq m_{M_i \rightarrow R}$ denota el número de puntos bajo la gráfica de $f_{M_i \rightarrow R}$.

Sin pérdida de generalidad, supongamos que los puntos $p_{i,1}, p_{i,2}, \dots, p_{i,m_{M_i \rightarrow R}^*}$ representan los tipos de mutaciones presentes en el linaje clonal de la i -ésima célula mutada. Observemos que el j -ésimo tipo mutación que da origen a células mutantes con fenotipo mutador, $j = 1, 2, \dots, m_{M_i \rightarrow R}^*$, está presente en una célula mutadora, digamos $c_{i,j}$, la cual cuenta desde el instante $T_i + S_j$ al tiempo t_f para generar su linaje clonal.

Por el Teorema 4, sabemos que considerar al proceso de ramificación que se observa a partir del instante $T_i + S_j$, donde la célula mutadora $c_{i,j}$ se considera como la única célula inicial, es un proceso de Yule-Furry, nuevamente. Por lo que el Corolario 3 implica que la

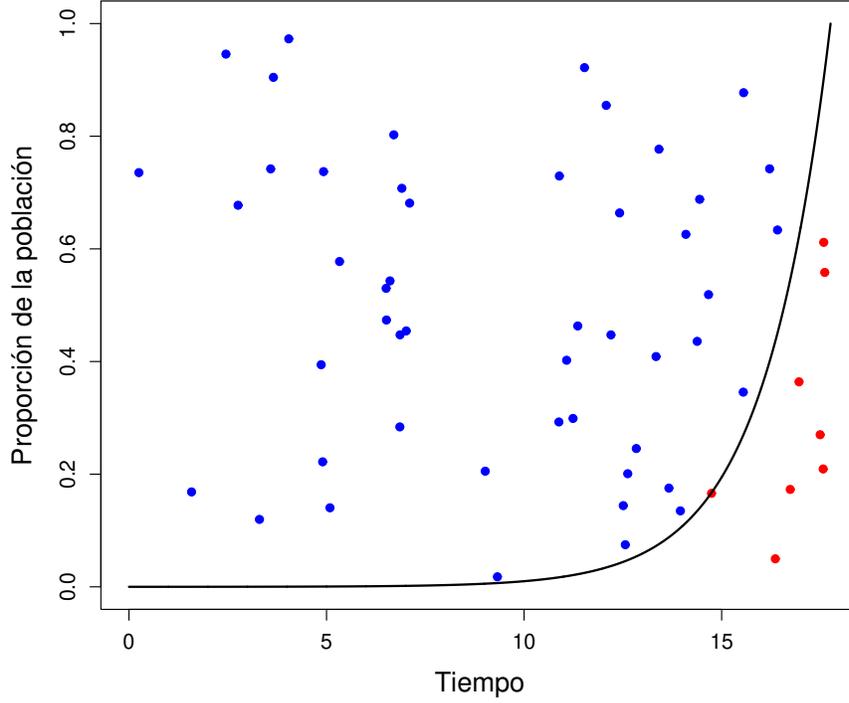


Figura 4.3: Método de decisión para determinar las mutaciones que se presentan en un cultivo celular iniciado con una célula mutadora. Se consideró un cultivo en donde $N_0 = 1$, $N_f = 10^9$, $\mu_{W \rightarrow R} = 0$, $\mu_{W \rightarrow M} = 10^{-7}$, $\mu_{M \rightarrow R} = 10^{-4}$ y con ninguna célula mutante, mutadora ni mutante con fenotipo mutador. El cultivo ocurrió en $t_f = \log_2(10^9) \approx 29.89735$ unidades de tiempo; mientras que la célula mutadora de la cual se observa el proceso de decisión para determinar mutaciones en su linaje apareció en el instante $t_i \approx 12$, con t_i la realización de T_i . Implicando que las mutaciones en su linaje cuentan con $t_f - t_i \approx 18$ unidades de tiempo para suscitarse.

variable aleatoria $L_{i,j}$ que cuenta las células en el linaje de la célula mutadora $c_{i,j}$ es tal que

$$L_{i,j} \sim \text{BN} \left(1, e^{-g_i(t_f - T_i - S_j)} \right). \quad (4.76)$$

Implicando que el número total de células mutantes con fenotipo mutante que aparecieron dentro del linaje clonal de la célula mutadora c_i es

$$\sum_{j=1}^{m_{M_i \rightarrow R}^*} L_{i,j}. \quad (4.77)$$

Más aún, por proceder como antes para cada una de las $m_{W \rightarrow M}^*$ células mutadoras

obtenemos que

$$M_R = \sum_{i=1}^{m_{W \rightarrow M}^*} \sum_{j=1}^{m_{M_i \rightarrow R}^*} L_{i,j}, \quad (4.78)$$

es el número total de células mutantes con fenotipo mutador, al finalizar la observación de la población celular. Lo que culmina el planteamiento del modelo gráfico.

A manera de resumen, presentamos el pseudocódigo que se desprende del modelo gráfico. La forma en la que hemos planteado el modelo gráfico permite su implementación computacional a través de una estructuración modular. Esto se debe a que los planteamientos propuestos para las células mutantes y las células mutadoras son independientes, pero el planteamiento para las células mutantes con fenotipo mutador depende del planteamiento para las células mutadoras.

De esta manera, el modelo gráfico se escribe en un par de módulos. En el Algoritmo 1, presentamos la sección del modelo gráfico que corresponde a la aparición de células mutantes y sus linajes clonales. Luego, en el Algoritmo 2, presentamos lo correspondiente a la aparición de células mutadoras, células mutantes con fenotipo mutador y sus respectivos linajes clonales.

Entradas:

- N_0 : Número de células silvestres con que se inicia la población.
- N_f : Tamaño final de la población.
- $\mu_{W \rightarrow R}$: Tasa de mutación de células silvestres hacia células mutantes.

$t_f = \log_2(N_f)$. Tiempo requerido para alcanzar el tamaño N_f .

$g = \log(N_f/N_0)/t_f$. Tasa de crecimiento.

m. Una realización de $M_{W \rightarrow R} \sim \text{Poisson}(t_f N_f \mu_{W \rightarrow R})$.

C. Una matriz de tamaño $m \times 2$

para $i=1$ **hasta** m **con incrementos de 1 hacer**

C[i,1]=t_i. Una realización de $T \sim \text{Uniforme}(0, t_f)$
 C[i,2]=p_i. Una realización de $P \sim \text{Uniforme}(0, 1)$

fin

D. Conjunto de filas i de C tal que $C[i,2] \leq (N_0 e^{gC[i,1]})/N_f$.

m*. Número de filas de D.

Mutantes=0. Inicializar variable contadora.

para $i=1$ **hasta** m* **con incrementos de 1 hacer**

d_t_i=t_f-C[1,i]. Tiempo disponible para generar el linaje clonal.
 l_i. Una realización de $L_i \sim \text{BN}(1, e^{-g d_t_i})$.
 Mutantes=Mutantes+l_i. Actualizar variable contadora¹.

fin

Salida: Mutantes.

Algoritmo 1: Módulo del modelo gráfico respectivo a las células mutantes.

¹Sustituya por Mutantes=Mutantes + l_i + 1, en caso que el lenguaje de programación a utilizar considere al conjunto $\{0, 1, 2, \dots\}$ como el soporte de una distribución binomial negativa, en vez del conjunto $\{1, 2, \dots\}$. Por ejemplo, al implementar este pseudocódigo en R [111, función `rnbinom`].

Entradas:

- N_0 : Número de células silvestres con que se inicia la población.
- N_f : Tamaño final de la población.
- $\mu_{W \rightarrow M}$: Tasa de mutación de células silvestres hacia células mutadoras.
- $\mu_{M \rightarrow R}$: Tasa de mutación de células mutadoras hacia células mutantes.

$t_f = \log_2(N_f)$. Tiempo requerido para alcanzar el tamaño N_f .

$g = \log(N_f/N_0)/t_f$. Tasa de crecimiento.

m. Una realización de $M_{W \rightarrow M} \sim \text{Poisson}(t_f N_f \mu_{W \rightarrow M})$.

C. Una matriz de tamaño $m \times 2$

para $i=1$ hasta m con incrementos de 1 hacer

- | C[i,1]=t_i. Una realización de $S \sim \text{Uniforme}(0, t_f)$
- | C[i,2]=p_i. Una realización de $Q \sim \text{Uniforme}(0, 1)$

fin

D. Conjunto de filas i de C tal que $C[i,2] \leq (N_0 e^{gC[i,1]})/N_f$.

m^* . Número de filas de D.

Mutadores=0. Inicializar variable contadora.

Mutantes_mutadores=0. Inicializar variable contadora.

para $i=1$ hasta m^* con incrementos de 1 hacer

| d_t_i=t_f-C[1,i]. Tiempo disponible para generar el linaje clonal.

| l_i. Una realización de $L_i \sim \text{BN}(1, e^{-g d_t_i})$.

Mutadores=Mutadores+l_i. Actualizar variable contadora. Nota al pie 1.

g_i=log(l_i)/d_t_i. Tasa de crecimiento del i-ésimo mutador.

m_i Una realización de $M_{M_i \rightarrow R} \sim \text{Poisson}(d_t_i \mu_{M \rightarrow R} l_i)$.

E. Una matriz de tamaño $m_i \times 2$

para $i=1$ hasta m_i con incrementos de 1 hacer

- | E[i,1]=t_i. Una realización de $T \sim \text{Uniforme}(0, d_t_i)$
- | E[i,2]=p_i. Una realización de $P \sim \text{Uniforme}(0, 1)$

fin

F. Conjunto de filas i de E tal que $E[i,2] \leq (e^{g_i E[i,1]})/l_i$

m^*_i . Número de filas de F.

Mutantes_Mutador_i=0. Inicializar variable contadora.

para $j=1$ hasta m^*_i con incrementos de 1 hacer

| d_t_j_i= d_t_i - E[1,j]. Tiempo para generar el linaje.

| l_i_j. Una realización de $L_{i,j} \sim \text{BN}(1, e^{-g_i d_t_j_i})$

| Mutantes_Mutador_i= Mutantes_Mutador_i + l_i_j. Nota al pie 1.

fin

Mutantes_mutadores=Mutantes_mutadores+Mutantes_Mutador_i.

fin

Salida: Mutadores; Mutantes_mutadores.

Algoritmo 2: Módulo del modelo gráfico respectivo a las células mutadoras y células mutantes con fenotipo mutador.

Ahora, en la Figura 4.4 se presentan los histogramas obtenidos a partir de considerar 150 simulaciones del modelo gráfico empleado los siguientes valores para los parámetros $N_f = 2^{33}$, $\mu_{W \rightarrow R} = 10^{-7}$, $\mu_{W \rightarrow M} = 10^{-6}$ y $\mu_{M \rightarrow M} = 10^{-5}$.

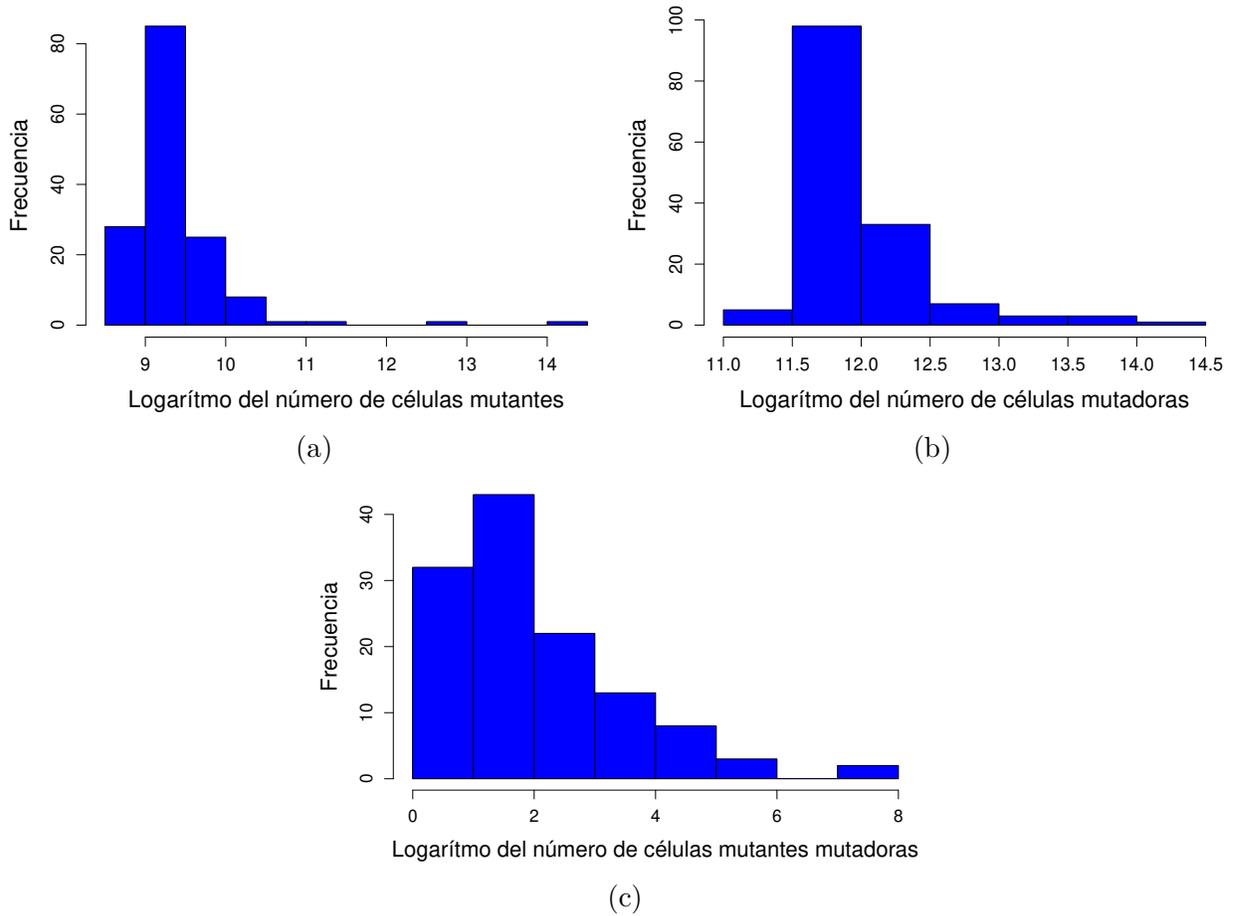


Figura 4.4: Histogramas resultantes de ejecutar 150 simulaciones.

Recordemos que estudiar un histograma da indicios de la distribución de probabilidad subyacente. En este caso, la característica más importante a considerar es que la Figura 4.4a se observa el siguiente comportamiento: el número de células mutantes producidas por células silvestres crece hasta alcanza un máximo y luego decae rápidamente.

Las distribuciones que presentan dicho comportamiento se conocen como *distribuciones de cola pesada* y la relevancia de esto recae en que la distribución Luria-Delbrück, que se presenta en el Capítulo 5, se caracteriza por ello.

4.2. Modelo basado en el método de inversión

El modelo gráfico es capaz de abstraer el comportamiento de las células a través de variables aleatoria. Esto permite obtener simulaciones en menor tiempo, pero la creación de los gráficos, por ejemplo Figura 4.2 y Figura 4.3, consumen tiempo de cómputo. Más aún, el proceso para determinar las mutaciones que se presentan en la población implica la producción de información redundante, dado que conocer las mutaciones que no se hacen presentes en el cultivo no resulta de interés, además que también implica un costo en tiempo computacional.

Para resolver lo anterior proponemos un par de soluciones. La primera es omitir la creación e impresión de tales gráficos, como se plantea en los algoritmos 1 y 2. Mientras que la segunda propuesta es enfocarnos en simular tiempos de aparición exclusivamente de células mutantes en la población directamente, en lugar de lidiar con tiempos de aparición de mutaciones que pueden o no dar paso a las células mutantes en la población.

4.2.1. Método de inversión

La idea de simular exclusivamente los tiempos de aparición de células mutantes motivó el desarrollo del modelo siguiente y para hablar de él será necesario introducir una definición que se origina de los trabajos de Francis Galton: la función cuantil [56], también conocida como la función cuantil de una función de distribución y cuya definición es la siguiente.

Definición 21 (Función cuantil) Sea F una función de distribución dada. La función cuantil de F es una función $F^- : [0, 1] \rightarrow [-\infty, \infty]$ dada por

$$F^-(x) := \inf\{y \in \mathbb{R} \mid F(y) \geq x\},$$

para $0 < x < 1$.

El concepto de función cuantil generaliza la idea de función inversa [44]; por tanto, escribimos F^- , y no F^{-1} , para hacer una distinción. Sin embargo, es importante mencionar que cuando se trabaja con una función de distribución continua y estrictamente creciente, se cumple que su función inversa existe y, mas aún, coincide con la función cuantil (ver [67, pág. 294])

Considerando la Definición 21 surge una pregunta natural sobre si la dicha función está bien definida para todo $x \in (0, 1)$. Lo cual motiva el siguiente teorema.

Teorema 11 Si F es una función de distribución dada, entonces su función cuantil, F^- , está bien definida.

Demostración. Sea $x \in (0, 1)$ un número arbitrario y fijo. Definamos $A_x = \{y \in \mathbb{R} \mid F(y) \geq x\}$ y $\varepsilon := 1 - x > 0$. Por el inciso (a) del Teorema 1 se tiene que

$$\lim_{y \rightarrow \infty} F(y) = 1.$$

Así, existe $N_0 \in \mathbb{N}$ tal que si $y > N_0$ entonces $|F(y) - 1| = 1 - F(y) < \varepsilon$, es decir, $x < F(y)$. Por tanto, como $N_0 + 1 > N_0$ se tiene que $F(N_0 + 1) > x$ y se sigue que $N_0 + 1 \in A_x$, por lo que $A_x \neq \emptyset$.

Luego, por el inciso (b) del Teorema 1 se tiene que

$$\lim_{y \rightarrow -\infty} F(y) = 0.$$

Así que para $\varepsilon_0 = x$ existe $M_0 \in \mathbb{N}$ tal que si $y < -M_0$ entonces $|F(y)| = F(y) < \varepsilon_0$. Por tanto, $-(M_0 + 1)$ es una cota inferior de A_x ya que de no ser así existiría $z \in A_x$ tal que $z < -(M_0 + 1)$. Sin embargo, como $z < -M_0$ entonces $F(z) < x = \varepsilon_0$ y dado que $z \in A_x$ también se cumple que $F(z) \geq x$, una contradicción.

De esta manera, como A_x es un conjunto no vacío y acotado inferiormente, $\inf A_x$ existe. Además, como x fue arbitraria, se tiene que $\inf A_x$ existe para cada $0 < x < 1$. ■

Para familiarizarnos con la función cuantil de una función de distribución discutiremos algunas propiedades que se utilizarán posteriormente. Sugerimos consultar el libro de Shorak y Wellmer [127] si se desea profundizar en este tema.

Teorema 12 Sea F una f.d. dada. Si F^- denota la función cuantil de F , entonces F^- satisface las siguientes propiedades.

- (a) $F(F^-(x)) \geq x$, para toda $x \in (0, 1)$.
- (b) $F(F^-(x)^-) \leq x$, para toda $x \in (0, 1)$. Donde $F(y^-)$, con $y = F^-(x)$, es el límite por la izquierda (vea inciso (e) del Teorema 1).

Demostración.

- (a) Sea $x \in (0, 1)$ arbitraria y fija. Como antes, definamos $A_x := \{y \in \mathbb{R} \mid F(y) \geq x\}$ y sea $a := F^-(x) = \inf A_x$, el cual existe por el Teorema 11. Dada $\varepsilon > 0$ se tiene que $a < a + \varepsilon$ y, más aún, $a + \varepsilon \in A_x$ pues:

si $a \in A_x$, necesariamente se cumple que $F(a) \geq x$ y dado que F es no decreciente se tiene que

$$F(a + \varepsilon) \geq F(a) \geq x.$$

Por otro lado, si $a \notin A_x$ y suponemos que $a + \varepsilon \notin A_x$, entonces $F(a) < x$ y $F(a + \varepsilon) < x$, por la definición de A_x ; no obstante, por la definición de ínfimo, se cumple que para la ε dada existe $\alpha \in A_x$ tal que $\alpha < a + \varepsilon$ lo que implica que $x \leq F(\alpha) \leq F(a + \varepsilon)$, una contradicción.

De esta manera, como para cada $\varepsilon > 0$ se cumple que $F(a + \varepsilon) \geq x$, entonces se tiene que

$$\lim_{\varepsilon \rightarrow 0} F(a + \varepsilon) = F(a^+) \geq x.$$

Y puesto que F es continua por la derecha se tiene que $F(a) = F(a^+)$, de donde se sigue que

$$F(a) = F(F^-(x)) \geq x.$$

Luego, como x fue arbitraria se muestra que el resultado es válido para cada $x \in (0, 1)$.

- (b) Sea $x \in (0, 1)$ arbitraria y fija. Definamos $A_x := \{y \in \mathbb{R} \mid F(y) \geq x\}$ y sea $a := \inf A_x = F^-(x)$, el cual existe por el Teorema 11. Dada $\varepsilon > 0$ se cumple que $a - \varepsilon < a$ lo que implica que $a - \varepsilon \notin A_x$, pues si $a - \varepsilon$ fuese un elemento de A_x necesariamente se cumpliría, por la definición de ínfimo, que $a \leq a - \varepsilon$, lo cual es una contradicción. Así, se sigue que $F(a - \varepsilon) < x$, para toda $\varepsilon > 0$, por lo que al tomar límite cuando $\varepsilon \rightarrow 0$ se tiene que

$$\lim_{\varepsilon \rightarrow 0} F(a - \varepsilon) = F(a^-) = F(F^-(x)^-) \leq x.$$

Notemos que como x fue arbitraria, se tiene que el resultado es válido para toda $x \in (0, 1)$, lo cual concluye la prueba. ■

Del teorema anterior se desprende el siguiente corolario. Este resultado nos permite reconocer explícitamente la forma que poseen los conjuntos $\{y \in \mathbb{R} \mid F(y) \geq x\}$, para toda $x \in (0, 1)$.

Corolario 4 Sea F es una función de distribución dada. Si $x \in (0, 1)$ y $A_x = \{y \in \mathbb{R} \mid F(y) \geq x\}$, entonces A_x es un intervalo de la forma (a, ∞) o $[a, \infty)$.

Demostración. Sea $x \in (0, 1)$ arbitraria y fija. Del Teorema 11 se tiene que A_x es acotado inferiormente; luego, usando un argumento similar a la demostración del inciso (a)

del Teorema 12 sabemos que para todo $\varepsilon > 0$ se tiene que $\inf A_x + \varepsilon \in A_x$ por lo que $(\inf A_x, \infty) \subseteq A_x$. Por otro lado, si $\inf A_x \notin A_x$ para $y \in A_x$ entonces $y > \inf A_x$ y se sigue que $y \in (\inf A_x, \infty)$, por lo que $A_x = (\inf A_x, \infty)$.

Note que si $\inf A_x \in A_x$ entonces $y \geq \inf A_x$, es decir, $y \in [\inf A_x, \infty)$. Además, si $y \in [\inf A_x, \infty)$ se cumple que $y = \inf A_x + (y - \inf A_x)$ de donde se sigue que $y \in A_x$, pues si $y > \inf A_x$ entonces tomando $\varepsilon_0 = y - \inf A_x > 0$ se tiene que $\inf A_x + \varepsilon_0 \in A_x$, mientras que si $y = \inf A_x$ entonces trivialmente $y \in A_x$. De esta manera se tiene que $A_x = [\inf A_x, \infty)$.

Con lo anterior se ha mostrado que A_x es intervalo no acotado superiormente y dado que x fue arbitraria el resultado es válido para toda $x \in (0, 1)$. ■

Obsevación 4 Desde que $x \in (0, 1)$, los incisos (a)-(c) del Teorema 1 implican que si $A_x = \{y \in \mathbb{R} \mid F(y) \geq x\}$, entonces $\inf A_x \in A_x$, por lo que

$$A_x = [\inf A_x, \infty).$$

O, equivalentemente, en términos de la función cuantil

$$A_x = [F^-(x), \infty).$$

La discusión anterior permite presentar el resultado en el que se basa el modelo que da nombre a esta sección.

Teorema 13 (Método de inversión) Sea F una f.d. dada y ξ una v.a. con distribución uniforme, es decir, $\xi \sim \text{Uniforme}(0, 1)$. Si $X = F^-(\xi)$, con F^- la función cuantil de la f.d. F , entonces $X \sim F$.

Demostración. Sea $x \in \mathbb{R}$ arbitraria y fija. Demostraremos a continuación que los eventos $E_1 := \{\omega \in \Omega \mid F(x) \geq \xi(\omega)\}$ y $E_2 := \{\omega \in \Omega \mid x \geq F^-(\xi(\omega))\}$ son equivalentes.

Sea $\omega^* \in E_1$. Como $\xi(\omega^*) \in (0, 1)$, pues $\xi \sim \text{Uniforme}(0, 1)$, entonces de la definición de la función cuantil se tiene que

$$F^-(\xi(\omega^*)) = \inf\{y \in \mathbb{R} \mid F(y) \geq \xi(\omega^*)\}.$$

Notemos que se cumple que $F(x) \geq \xi(\omega^*)$, puesto que $\omega^* \in E_1$, entonces $x \in \{y \in \mathbb{R} \mid F(y) \geq \xi(\omega^*)\}$; así, por la definición de ínfimo, se tiene que $F^-(\xi(\omega^*)) \leq x$, probando que $E_1 \subseteq E_2$.

Ahora, sea $\omega^* \in E_2$, entonces $x \geq F^-(\xi(\omega^*))$; de esta manera, aplicando el Corolario 4 y la Nota 4, se tiene que $x \in \{y \in \mathbb{R} \mid F(y) \geq \xi(\omega^*)\}$, por lo que $F(x) \geq \xi(\omega^*)$, mostrando que $E_2 \subseteq E_1$.

De lo anterior se tiene que $E_1 = E_2$ y ya que x fue arbitraria se tiene que el resultado es válido para toda $x \in \mathbb{R}$. De esta manera, se sigue que

$$P(X \leq x) = P(F^-(\xi) \leq x) = P(\xi \leq F(x)) = \int_0^{F(x)} \xi d\xi = F(x).$$

Es decir, $X \sim F$. ■

4.2.2. Algoritmo

Usando la discusión anterior y los resultados presentados en la Sección 4.1.2 como punto de partida, presentamos a continuación al modelo basado en el método de inversión.

Consideremos las hipótesis y notación del modelo gráfico que respectan a la cantidad inicial N_0 de células en la población, la cantidad final N_f de células en el cultivo, el crecimiento exponencial con tasa g y el intervalo de tiempo en el que se desarrolla el crecimiento poblacional $[0, t_f]$.

Como en el modelo gráfico, supongamos que las células crecen exponencialmente hasta el tiempo t_f , donde $t_f > 0$ denota el tiempo en el cual el espacio y los recursos están por agotarse. Esto es, para tiempos posteriores a t_f el crecimiento poblacional ha alcanzado la capacidad de carga asociada con las condiciones experimentales.

Además, nuevamente, sean $\mu_{W \rightarrow R}$, $\mu_{W \rightarrow M}$ y $\mu_{M \rightarrow R}$ las tasas de mutación de células silvestres hacia células con fenotipo mutante, de células silvestres hacia células con fenotipo mutador y de células mutadoras hacia células mutantes con fenotipo mutador, respectivamente.

Por la Observación 3, si $M_{W \rightarrow R}$ la variable aleatoria que cuenta el número de tipos de células mutantes entonces

$$M_{W \rightarrow R} \sim \text{Poisson}(N_f \mu_{W \rightarrow R}). \quad (4.79)$$

Diremos que en el cultivo existen $m_{W \rightarrow R}$ tipos de células mutantes, con $m_{W \rightarrow R}$ una realización de $M_{W \rightarrow R}$. Recordemos que nos interesa determinar el tiempo de aparición de

cada uno de las $m_{W \rightarrow R}$ tipos de células mutantes. Para ello, sabemos que

$$N_0 e^{gt},$$

denota la población al tiempo t , para toda $t \in [0, t_f]$. Por tanto, la Observación 3 implica que

$$\mu_{W \rightarrow R} N_0 e^{gt}$$

indica el número esperado de células mutantes al instante t .

Por un argumento similar que llevó al planteamiento de la Ecuación (4.65), la función anterior es una distribución impropia dado que su integral sobre el intervalo de tiempo $[0, t_f]$ es finita pero distinta de 1. No obstante, al considerar un factor de escalamiento apropiado, una constante $s \in \mathbb{R}$, se obtiene una distribución propia.

De esta manera, sea $s = \frac{g}{(N_f - N_0)\mu_{W \rightarrow R}}$ y $h_{W \rightarrow R} : [0, t_f] \rightarrow [s\mu_{W \rightarrow R} N_0, s\mu_{W \rightarrow R} N_f]$ dada por

$$h_{W \rightarrow R}(t) = s\mu_{W \rightarrow R} N_0 e^{gt}. \quad (4.80)$$

Entonces se cumple que

$$\begin{aligned} \int_0^{t_f} s\mu_{W \rightarrow R} N_0 e^{gt} dt &= \frac{N_0}{(N_f - N_0)} \int_0^{t_f} g e^{gt} dt \\ &= \frac{N_0}{(N_f - N_0)} (e^{gt_f} - 1) = 1 \end{aligned}$$

Más aún, como h es continua y no decreciente se tiene que h es una función de densidad cuya función de distribución es $H_{W \rightarrow R} : [0, t_f] \rightarrow [0, 1]$ dada por

$$H_{W \rightarrow R}(t) = \frac{N_0 s \mu_{W \rightarrow R}}{g} (e^{gt} - 1). \quad (4.81)$$

En casos como el anterior, en donde la distribución carece de un nombre común, se suele utilizar al nombre de la función de distribución, $H_{W \rightarrow R}$, para hacer referencia a tal distribución y escribir $X \sim H_{W \rightarrow R}$ para denotar que la variable aleatoria X tiene distribución $H_{W \rightarrow R}$.

Procedamos a hallar la función cuantil de $H_{W \rightarrow R}$ para simular los tiempos de aparición de las células mutantes a través del método de inversión. Como H es continua y monótona creciente, su inversa inversa existe y coincide con su función cuantil $H_{W \rightarrow R}^- : [0, 1] \rightarrow [0, t_f]$ dada por

$$H_{W \rightarrow R}^-(u) = \frac{1}{g} \log \left(1 + \frac{gu}{N_0 s \mu_{W \rightarrow R}} \right) \quad (4.82)$$

Sea $U_{W \rightarrow R} \sim \text{Uniforme}(0, 1)$. Entonces, el Teorema 13 implica que

$$\begin{aligned} V_{W \rightarrow R} &= \frac{1}{g} \log \left(1 + \frac{gU_{W \rightarrow R}}{N_0 s \mu_{W \rightarrow R}} \right) \\ &= \frac{1}{g} \log \left(1 + \frac{N_f - N_0}{N_0} U_{W \rightarrow R} \right) \end{aligned} \quad (4.83)$$

es una variable aleatoria tal que $V_{W \rightarrow R} \sim H_{W \rightarrow R}$.

Como se ha mencionado previamente, el método de inversión nos permite determinar el tiempo de aparición de cada célula mutante. Esto, a través de asociar una realización de la variable $V_{W \rightarrow R}$ a cada una de dichas células.

Ahora, presentamos en la Figura 4.5 una verificación visual de la precisión que nos proporciona el método de inversión para simular los tiempos de aparición. La gráfica de la función $h_{W \rightarrow R}$ se muestra en color negro, mientras que en color azul observamos un histograma construido a partir de 1000 realizaciones de la variable $V_{W \rightarrow R}$.

Más aún, al observar dicho histograma se aprecia que las mutaciones pueden ocurrir en todo momento; sin embargo, estas suelen aparecer con mayor frecuencia en tiempos grandes. Lo cual es congruente con el hecho que el envejecimiento y la aparición de mutaciones en las células están relacionados [37], [52], [96].

Luego, el i -ésimo tipo de célula mutante, $i = 1, 2, \dots, m_{W \rightarrow R}$ se originó al instante V_i , con V_i una muestra aleatoria de $V_{W \rightarrow R}$. Por tanto, dispone del instante V_i al final de la observación del experimento para generar su linaje clonal.

De esta manera, el tamaño del linaje de la célula mutante de tipo i es una variable aleatoria L_i tal que

$$L_i \sim \text{BN}(1, e^{-g(t_f - V_i)}), \quad (4.84)$$

por el Corolario 3.

Notemos que lo anterior es válido para toda $i = 1, 2, \dots, m_{W \rightarrow R}$. Por tanto, implica que el número total de células mutantes en el cultivo es

$$R = \sum_{i=1}^{m_{W \rightarrow R}} L_i. \quad (4.85)$$

Notemos que, al comparar la Ecuación (4.68) con la ecuación anterior ambas resultan ser similares. Sin embargo, una diferencia crucial se encuentra en el límite superior de la

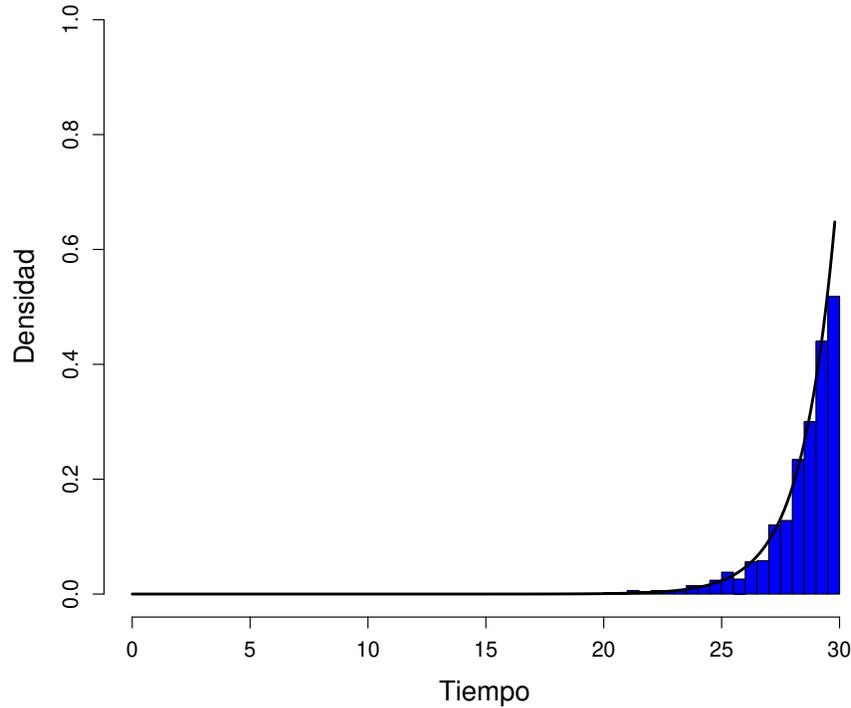


Figura 4.5: Método de inversión para determinar los tiempos de aparición de las células mutantes en un cultivo celular. Para determinar las características del cultivo, los siguientes valores fueron considerados: $N_0 = 1$, $N_f = 10^9$ y $\mu_{W \rightarrow T} = 10^{-7}$.

suma, en el caso de la Ecuación (4.85) aparece directamente el valor de la realización de una variable aleatoria que cuenta el número del tipo de células mutantes.

Mientras que en la Ecuación (4.68), el número que aparece en el límite superior es menor o igual, frecuentemente menor estricto, que la realización de una variable aleatoria que cuenta el número de tipos de mutaciones. Podemos interpretar lo anterior como que en el método gráfico se produce información redundante, lo cual en el método basado en el método de inversión se logra evadir.

Para proceder con la aparición de células mutadoras, procedamos de manera similar a la discutida en el método gráfico. Así, supongamos que las células mutadoras surgen a través de mutaciones en las células silvestres. Más aún, el cultivo se inicia con N_0 células silvestres y cero células con fenotipo mutador que crecen exponencialmente a una tasa g .

De esta manera, la variable aleatoria $M_{W \rightarrow M}$ que cuenta el número de tipos de

mutaciones que dan origen a células con fenotipo mutador es tal que

$$M_{W \rightarrow M} \sim \text{Poisson}(N_f \mu_{W \rightarrow M}), \quad (4.86)$$

por la Observación 3.

Sea $m_{W \rightarrow M}$ el número de tipos de mutaciones que originan células con fenotipo mutador, con $m_{W \rightarrow M}$ una realización de $M_{W \rightarrow M}$. Para determinar el tiempo de aparición de cada uno de los $m_{W \rightarrow M}$ tipos de mutación consideremos $s_1 = \frac{g}{(N_f - N_0) \mu_{W \rightarrow M}}$ y la función de densidad $h_{W \rightarrow M} : [0, t_f] \rightarrow [s_1 \mu_{W \rightarrow M} N_0, s_1 \mu_{W \rightarrow M} N_f]$ dada por

$$h_{W \rightarrow M}(t) = s_1 \mu_{W \rightarrow M} N_0 e^{gt}. \quad (4.87)$$

Más aún, lo anterior implica que la función $H : [0, t] \rightarrow [0, 1]$ definida por

$$H_{W \rightarrow M}(t) = \frac{N_0 s_1 \mu_{W \rightarrow M} (e^{gt} - 1)}{g} \quad (4.88)$$

es su función de distribución. Mientras que la función $H_{W \rightarrow M}^- : [0, 1] \rightarrow [0, t_f]$ dada por

$$H_{W \rightarrow M}^-(u) = \frac{1}{g} \log \left(1 + \frac{gu}{N_0 s_1 \mu_{W \rightarrow M}} \right) \quad (4.89)$$

es la función cuantil de H .

De esta manera, el método de inversión implica que si $U_{W \rightarrow M} \sim \text{Uniforme}(0, 1)$ entonces

$$\begin{aligned} V_{W \rightarrow M} &= \frac{1}{g} \log \left(1 + \frac{gU_{W \rightarrow M}}{N_0 s_1 \mu_{W \rightarrow M}} \right) \\ &= \frac{1}{g} \log \left(1 + \frac{N_f - N_0}{N_0} U_{W \rightarrow M} \right) \end{aligned} \quad (4.90)$$

es una variable aleatoria tal que $V_{W \rightarrow M} \sim H_{W \rightarrow M}$.

Se sigue entonces que el instante en el que surge el i -ésimo de tipo de mutación que origina al fenotipo mutador se determina con V_i , una realización de $V_{W \rightarrow M}$. Además, si c_i denota a la célula que posee el i -ésimo de tipo de mutación que origina al fenotipo mutador entonces c_i dispone desde el instante V_i al final del experimento t_f para generar a su linaje clonal.

Así que L_i la variable aleatoria que cuenta el número de células en el linaje de la célula mutadora c_i es tal que

$$L_i \sim \text{BN}(1, e^{-g(t_f - V_i)}), \quad (4.91)$$

por el Corolario 3. Implicando que el número total de células mutadoras es

$$M = \sum_{i=1}^{m_{W \rightarrow M}} L_i. \quad (4.92)$$

Ahora, procediendo de manera análoga al modelo gráfico, supondremos que las células mutantes con fenotipo mutador aparecen exclusivamente por consecuencia de mutaciones en las células mutadoras.

Por tanto, realicemos un análisis dentro del ciclo de crecimiento de la célula mutadora c_i , la cual posee el i -ésimo tipo de mutación que origina el fenotipo mutador. Considerando lo anterior, nos restringimos a un examinar un cultivo que inicia con $N_0 = 1$ célula mutadora, cero células mutantes mutadoras, admite un tamaño final de L_i células y se observará durante $t_f - V_i$ unidades de tiempo.

Sea $M_{M_i \rightarrow R}$ la variable aleatoria que cuenta el número de tipos de mutaciones que originan células mutantes y cuyo ancestro es la célula mutadora c_i . Entonces

$$M_{M_i \rightarrow R} \sim \text{Poisson}(L_i \mu_{M \rightarrow R}). \quad (4.93)$$

El número de tipos de mutaciones que aparecieron a lo largo del ciclo de crecimiento de la célula mutadora c_i se determina a través de $m_{M_i \rightarrow R}$, una realización de $M_{M_i \rightarrow R}$.

Luego, sea $s_i = \frac{g_i}{(L_i - 1) \mu_{M \rightarrow R}}$, con

$$g_i = \frac{1}{t_f - V_i} \log \left(\frac{L_i}{N_0} \right), \quad (4.94)$$

análogo a la Ecuación (4.72).

Dada la función de densidad $h_{M_i \rightarrow R} : [0, t_f - V_i] \rightarrow [s_i \mu_{M \rightarrow R}, s_i \mu_{M \rightarrow R} L_i]$, definida por

$$h_{M_i \rightarrow R}(t) = s_i \mu_{M \rightarrow R} e^{g_i t}, \quad (4.95)$$

y la función de distribución $H_{M_i \rightarrow R} : [0, t_f - V_i] \rightarrow [0, 1]$, tal que

$$H_{M_i \rightarrow R}(t) = \frac{s_i \mu_{M \rightarrow R}}{g_i} (e^{g_i t} - 1), \quad (4.96)$$

podemos determinar los tiempos de aparición de los $m_{M_i \rightarrow R}$ tipos de mutaciones que aparecieron a lo largo del ciclo de crecimiento de la célula mutadora c_i a través del Teorema 13 y la función cuantil de $H_{M_i \rightarrow R}$.

Sea $H_{M_i \rightarrow R}^- : [0, 1] \rightarrow [0, t_f - V_i]$ dada por

$$H_{M_i \rightarrow R}^-(u) = \frac{1}{g_i} \log \left(1 + \frac{g_i u}{s_i \mu_{M \rightarrow R}} \right), \quad (4.97)$$

la función cuantil de $H_{M_i \rightarrow R}$. Empleando el método de inversión tenemos que si $U_{M_i \rightarrow R} \sim \text{Uniforme}(0, 1)$ entonces

$$\begin{aligned} W_{M_i \rightarrow R} &= \frac{1}{g_i} \log \left(1 + \frac{g_i U}{s_i \mu_{M \rightarrow R}} \right) \\ &= \frac{1}{g_i} \log \left(1 + \frac{L_i - N_0}{N_0} U \right) \end{aligned} \quad (4.98)$$

es una variable aleatoria tal que $W_{M_i \rightarrow R} \sim H_{M_i \rightarrow R}$.

En la Figura 4.6 presentamos el método de inversión aplicado a la distribución $H_{M_i \rightarrow R}$, mostrando un resultado similar a lo que se observa en la Figura 4.5. La gráfica de $h_{M_i \rightarrow R}$ se aprecia en color negro, mientras que en color azul se observa un histograma construido a partir de 1000 realizaciones de la variable $W_{M_i \rightarrow R}$.

Lo anterior implica que el j -ésimo tipo de mutación que originan células mutantes con fenotipo mutador y cuyo ancestro es la célula mutadora c_i , $j = 1, 2, \dots, m_{M_i \rightarrow R}$, ha aparecido en el instante W_j , con W_j una muestra aleatoria de tamaño uno proveniente de $W_{M_i \rightarrow R}$.

Así es que la célula $c_{i,j}$ que posee el j -ésimo tipo de mutación que origina células mutantes con fenotipo mutador y que es descendiente de la célula mutadora c_i dispone del momento de su aparición $V_i + W_j$ a final del experimento t_f para generar su linaje clonal.

De esta manera, por aplicar el Corolario 3 sabemos que la variable aleatoria $L_{i,j}$ que cuenta el número de células en el linaje de la célula $c_{i,j}$ es tal que

$$L_{i,j} \sim \text{BN} \left(1, e^{-g_i(t_f - V_i - W_j)} \right). \quad (4.99)$$

Consecuentemente, el número de células mutantes con fenotipo mutador que se encuentran en el cultivo celular es

$$M_R = \sum_{i=1}^{m_{W \rightarrow M}} \sum_{j=1}^{m_{M_i \rightarrow R}} L_{i,j}. \quad (4.100)$$

Culminando el planteamiento del modelo basado en el método de inversión.

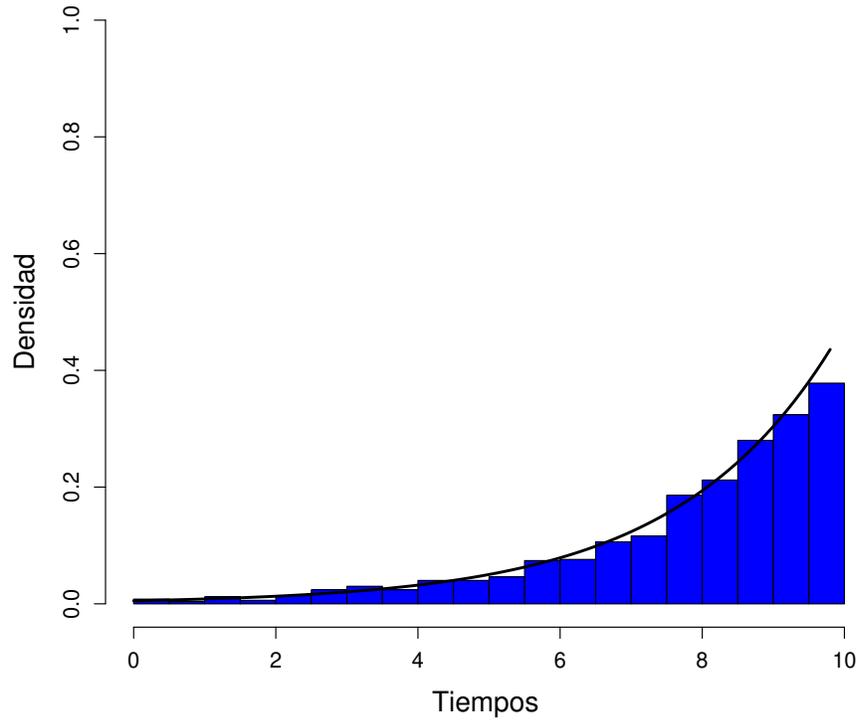


Figura 4.6: Método de inversión para determinar los tiempos de aparición de las células mutantes con fenotipo mutador en un cultivo celular. Para determinar las características del cultivo, los siguientes valores fueron considerados: $N_0 = 1$, $N_f = 10^9$ y $\mu_{W \rightarrow T} = 10^{-7}$. El cultivo ocurrió en $t_f = \log_2(10^9) \approx 29.89735$ unidades de tiempo; mientras que, la célula mutadora apareció en el instante $v_i \approx 20$, con v_i una realización de $V_{W \rightarrow M}$. Lo cual implica que las mutaciones en su linaje pueden aparecer en $t_f - v_i \approx 10$ unidades de tiempo.

A continuación, presentamos el pseudocódigo que se desprende del modelo anterior. Y, similar al modelo gráfico, podemos estructurar dicho pseudocódigo a través de dos módulos. Donde el primer módulo, Algoritmo 3, presenta la aparición y reproducción de células mutantes; mientras que, el segundo módulo, Algoritmo 4, expone la aparición de las células mutadoras, las células mutantes con fenotipo mutador y sus respectivas reproducciones.

Entradas:

- N_0 : Número de células silvestres con que se inicia la población
- N_f : Tamaño final de la población.
- $\mu_{W \rightarrow R}$: Tasa de mutación de células silvestres hacia células mutantes.

$t_f = \log_2(N_f)$. Tiempo requerido para alcanzar el tamaño N_f .

$g = \log(N_f/N_0)/t_f$. Tasa de crecimiento.

m. Una realización de $M_{W \rightarrow R} \sim \text{Poisson}(N_f \mu_{W \rightarrow R})$.

C. Una matriz de tamaño $m \times 1$.

Mutantes = 0. Inicializar variable contadora.

para $i=1$ **hasta** m **con incrementos de 1 hacer**

 u_i. Una realización de $U_{W \rightarrow R} \sim \text{Uniforme}(0, 1)$.

 C[i,1] = $\log(1 + ((N_f - N_0)u_i)/N_0)$. Tiempo de aparición.

 d_t_i = $t_f - C[1,i]$. Tiempo disponible para generar el linaje clonal.

 l_i. Una realización de $L_i \sim \text{BN}(1, e^{-gd_t_i})$.

 Mutantes = Mutantes + l_i. Actualizar variable contadora. Nota al pie 1.

fin

Salida: Mutantes.

Algoritmo 3: Módulo del modelo basado en el método de inversión respectivo a las células mutantes.

Entradas:

- N_0 : Número de células silvestres con que se inicia la población
- N_f : Tamaño final de la población.
- $\mu_{W \rightarrow M}$: Tasa de mutación de células silvestres hacia células mutadoras.
- $\mu_{M \rightarrow R}$: Tasa de mutación de células mutadoras hacia células mutantes.

$t_f = \log_2(N_f)$. Tiempo requerido para alcanzar el tamaño N_f .

$g = \log(N_f/N_0)/t_f$. Tasa de crecimiento.

m . Una realización de $M_{W \rightarrow R} \sim \text{Poisson}(N_f \mu_{W \rightarrow R})$.

D . Una matriz de tamaño $m \times 1$.

Mutadores = 0. Inicializar variable contadora.

Mutantes_mutadores = 0. Inicializar variable contadora.

para $i=1$ **hasta** m **con incrementos de 1** **hacer**

u_i . Una realización de $U_{M \rightarrow R} \sim \text{Uniforme}(0, 1)$.

$D[i,1] = \log(1 + ((N_f - N_0)u_i)/N_0)$. Tiempo de aparición.

$d_t_i = t_f - D[i,1]$. Tiempo disponible para generar el linaje clonal.

l_i . Una realización de $L_i \sim \text{BN}(1, e^{-gd_t_i})$.

 Mutadores = Mutadores + l_i . Actualizar variable contadora. Nota al pie 1.

$g_i = \log(l_i)/d_t_i$. Tasa de crecimiento del i -ésimo mutador.

m^* . Una realización de $M_{M_i \rightarrow R} \sim \text{Poisson}(l_i \mu_{W \rightarrow R})$.

E . Una matriz de tamaño $m^* \times 1$.

 Mutantes_Mutador_i = 0. Inicializar variable contadora.

para $j=1$ **hasta** m^* **con incrementos de 1** **hacer**

u_i_j . Una realización de $U_{M_i \rightarrow R}(0, 1)$.

$E[j,1] = \log(1 + (l_i - 1)u_i_j)$. Tiempo de aparición.

$d_t_j_i = d_t_i - E[j,1]$. Tiempo disponible para generar el linaje clonal.

 Mutantes_Mutador_i. Una realización de $L_{i,j} \sim \text{BN}(1, e^{-g_i d_t_j_i})$

 Mutantes_Mutador_i = Mutantes_Mutador_i + Mutantes_Mutador_i.

 Actualizar variable contadora. Nota al pie 1.

fin

 Mutantes_mutadores = Mutantes_mutadores + Mutantes_Mutador_i.

fin

Salida: Mutadores; Mutantes_mutadores.

Algoritmo 4: Módulo del modelo basado en el método de inversión respectivo a las células mutadoras y células mutantes con fenotipo mutador.

A continuación presentamos la Figura 4.7 con los histogramas obtenidos a partir de considerar 150 simulaciones del modelo gráfico empleado los siguientes valores para los parámetros $N_f = 2^{33}$, $\mu_{W \rightarrow R} = 10^{-7}$, $\mu_{W \rightarrow M} = 10^{-6}$ y $\mu_{W \rightarrow M} = 10^{-5}$.

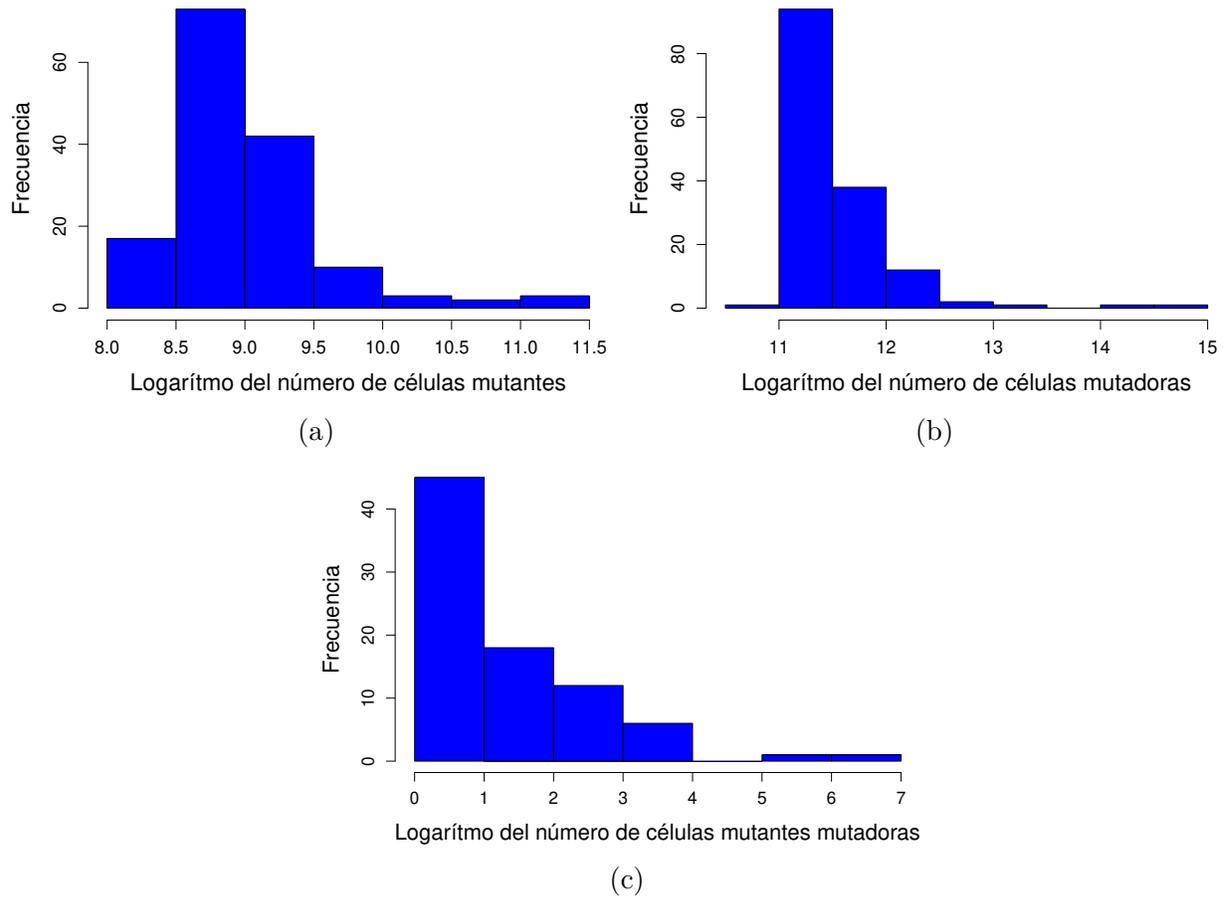


Figura 4.7: Histogramas resultantes de ejecutar 150 simulaciones.

Para referencia, se emplearon los mismos valores para generar las simulaciones que dan origen a los histogramas en la figura anterior y aquellas simulaciones de que dan origen a los histogramas en la Figura 4.4.

Aquí, nuevamente puede observarse el comportamiento de cola pesada en las distribuciones subyacentes en el número de células mutantes producidas por células silvestres, la Figura 4.7a y Figura 4.7b, respectivamente.

Capítulo 5

Métodos de estimación

En trabajos previos, dedicados a la estimación de la tasa de mutación de células silvestres a células mutantes [41], [63], [68], [72], [75], [81], [156], [157], han desarrollado metodologías con gran precisión.

En este trabajo, tomamos como referente a dicha bibliografía, la cual principalmente se fundamenta en la función generadora de probabilidad y la versión empírica de la misma [97]. Además, discutimos los métodos que desarrollamos para estimar la tasa de mutación hacia fenotipos mutadores.

El primero de ellos, implementa técnicas de ajuste directo a través del método de mínimos cuadrado, minimizando la *distancia* entre la función generadora de probabilidad y función generadora de probabilidad empírica, para producir los estimados deseados. Mientras que, el segundo método toma mayor ventaja del proceso de dilución, que se propone como parte del protocolo experimental, para emplear la transformada rápida de Fourier y los métodos de máxima verosimilitud para generar las estimaciones.

5.1. Distribución Luria-Delbrück

La distribución Luria-Delbrück, una distribución discreta, aparece a consecuencia del trabajo de Luria y Delbrück (1943) [90], en el que se dedican a estudiar cómo las células silvestre adquieren resistencia a medicamentos a través de mutaciones. No obstante, se determinó una expresión analítica para dicha distribución hasta el trabajo de Lea y Coulson

(1949) [84].

La expresión que se provee en [84], está dada en términos de una función generadora de probabilidad y no de una función de densidad, esto se debe a que una variable aleatoria con tal distribución no es absolutamente continua. Así es que la función de densidad de Luria-Delbrück no admite una expresión analítica [155], dado que no posee momentos finitos [154, p. 19]. Y es que, de conocer los momentos de la distribución, podemos recuperar la función generadora de momentos por el teorema de Taylor, ver [67, p. 74], y por tanto obtener la función de densidad.

Para proseguir, introduzcamos el concepto de función generadora de probabilidad.

Definición 22 (Función generadora de probabilidad) Sea X una variable aleatoria discreta y no negativa. Diremos que la función $G_X: \mathbf{C} \rightarrow \mathbf{C}$ definida por

$$G_X(\theta) = E(\theta^X) \quad (5.1)$$

es la función generadora de probabilidad de X .

Profundicemos en algunas propiedades de interés sobre la función generadora de probabilidad.

Lema 14 Sea X una variable aleatoria discreta y no negativa definida en el espacio de probabilidad (Ω, \mathcal{F}, P) y G_X su función generadora de probabilidad. Si R denota el rango de convergencia de G_X entonces se cumple que

$$\{\theta \in \mathbf{C}: |\theta| \leq 1\} \subseteq R. \quad (5.2)$$

Demostración. Sea $\theta \in \{\theta \in \mathbf{C}: |\theta| \leq 1\}$ un elemento arbitrario y fijo. Por el criterio de comparación sabemos que

$$\begin{aligned} \int_{\Omega} |\theta^X| dP &= \int_{\Omega} |\theta|^X dP \\ &\leq \int_{\Omega} dP \\ &= 1. \end{aligned}$$

Esto es, $E(\theta^X)$ es absolutamente convergente y, por tanto, el criterio de comparación implica su convergencia. De esta manera se tiene que $\theta \in R$; luego, como θ fue arbitrario se sigue lo deseado. ■

El teorema que presentamos a continuación será crucial en el desarrollo de la metodología de estimación.

Teorema 15 Sean X y Y variables aleatorias discretas con función de generadora de probabilidad G_X y G_Y , respectivamente. Entonces

$$G_{X+Y}(\theta) = G_X(\theta)G_Y(\theta). \quad (5.3)$$

Demostración. Dadas funciones u, v , se sabe que si X y Y son independientes y $E(u(X))$ y $E(v(Y))$ existen, entonces $E(u(X)v(Y)) = E(u(X))E(v(Y))$, ver [67, Teo. 2.4.4]. De esta manera, por considerar $u(x) = v(x) = \theta^x$ se tiene que

$$\begin{aligned} G_{X+Y}(\theta) &= E(\theta^{X+Y}) \\ &= E(\theta^X \theta^Y) \\ &= E(\theta^X)E(\theta^Y) \\ &= G_X(\theta)G_Y(\theta). \end{aligned} \quad (5.4)$$

■

Utilizando la discusión procedemos a introducir la distribución Luria-Delbrück a través de su función generadora de probabilidad.

Definición 23 (Distribución Luria-Delbrück) Sea X una variable aleatoria discreta. Diremos que X tiene distribución Luria-Delbrück con parámetro Λ si su función generadora de probabilidad es tal que

$$G_X(\theta) = e^{-\Lambda} \exp \left(\sum_{k=0}^{\infty} \frac{\Lambda \theta^k}{k(k+1)} \right), \quad (5.5)$$

para toda $\theta \in [0, 1]$.

Notemos que el rango de convergencia de la ecuación se restringe al intervalo $[0, 1]$, esto ocurre por la interpretación biológica que discutiremos más adelante. Luego, dado que la expresión en la Ecuación (5.5) resulta poco práctica, presentaremos una expresión equivalente. Para ello, enunciemos el siguiente lema.

Lema 16 Dado $\theta \in [0, 1]$ entonces

$$\sum_{j=1}^{\infty} \frac{\theta^j}{j(j+1)} = \begin{cases} 1 + \left(\frac{1-\theta}{\theta}\right) \log(1-\theta), & \text{si } \theta \in [0, 1). \\ 1, & \text{if } \theta = 1. \end{cases}$$

Demostración. Si $\theta = 1$ se tiene que

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{\theta^j}{j(j+1)} &= \lim_{k \rightarrow \infty} \sum_{j=1}^k \left(\frac{1}{j} - \frac{1}{j+1} \right) \\ &= \lim_{k \rightarrow \infty} \left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \cdots + \left(\frac{1}{k} - \frac{1}{k+1} \right) \\ &= \lim_{k \rightarrow \infty} 1 - \frac{1}{k+1} \\ &= 1. \end{aligned}$$

Para el caso $\theta \in [0, 1)$, recordemos que la serie de Taylor de $\log(1 + \zeta)$ [133, p. 586] está dada por

$$T(\zeta) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} \zeta^j, \quad \zeta \in (-1, 1).$$

Implicando que

$$\begin{aligned} \log(1 - \zeta) &= T(-\zeta) \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{2j+1}}{j} \zeta^j \\ &= - \sum_{j=1}^{\infty} \frac{\zeta^j}{j}, \end{aligned} \tag{5.6}$$

para cada $\zeta \in (-1, 1)$. Por tanto

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{\theta^j}{j+1} &= \frac{1}{\theta} \sum_{j=1}^{\infty} \frac{\theta^{j+1}}{j+1} \\ &= \frac{1}{\theta} \left(\theta - \theta + \sum_{j=2}^{\infty} \frac{\theta^j}{j} \right) \\ &= \frac{1}{\theta} \left(-\theta + \sum_{j=1}^{\infty} \frac{\theta^j}{j} \right) \\ &= -1 - \frac{1}{\theta} \log(1 - \theta). \end{aligned} \tag{5.7}$$

De la Ecuación (5.7) y la Ecuación (5.6) se sigue que

$$\begin{aligned}
 \sum_{j=1}^{\infty} \frac{\theta^j}{j(j+1)} &= \sum_{j=1}^{\infty} \left(\frac{1}{j} - \frac{1}{j+1} \right) \theta^j \\
 &= \sum_{j=1}^{\infty} \frac{\theta^j}{j} - \sum_{j=1}^{\infty} \frac{\theta^j}{j+1} \\
 &= -\log(1-\theta) + \frac{1}{\theta} \log(1-\theta) + 1 \\
 &= 1 + \left(\frac{1-\theta}{\theta} \right) \log(1-\theta),
 \end{aligned}$$

como se enunció. ■

Partiendo del lema anterior, se presenta una expresión alterna para la Ecuación 5.5, a través del siguiente teorema.

Teorema 17 Dada $\theta \in [0, 1]$ y bajo la convención $\lim_{(x,y) \rightarrow (0,0)} x^y = 1$ entonces

$$e^{-\Lambda} \exp \left(\sum_{j=1}^{\infty} \frac{\Lambda \theta^j}{j(j+1)} \right) = (1-\theta)^{\frac{\Lambda(1-\theta)}{\theta}}. \quad (5.8)$$

Demostración. Si $\theta = 1$, la Ecuación (5.8) se cumple por la convención.

Sea $\theta \in [0, 1)$ un número arbitrario y fijo. Además, sea $\psi = \sum_{j=1}^{\infty} \frac{\theta^j}{j(j+1)}$ entonces el Lema 16 implica que

$$\begin{aligned}
 e^{\Lambda\psi} &= \exp \left(\Lambda + \Lambda \left(\frac{1-\theta}{\theta} \right) \log(1-\theta) \right) \\
 &= e^{\Lambda} \exp \left(\Lambda \left(\frac{1-\theta}{\theta} \right) \log(1-\theta) \right) \\
 &= e^{\Lambda} \exp \left(\log \left((1-\theta)^{\Lambda \left(\frac{1-\theta}{\theta} \right)} \right) \right) \\
 &= e^{\Lambda} (1-\theta)^{\Lambda \left(\frac{1-\theta}{\theta} \right)}.
 \end{aligned}$$

Por tanto,

$$e^{-\Lambda} e^{\Lambda\psi} = (1-\theta)^{\Lambda \left(\frac{1-\theta}{\theta} \right)}.$$

■

Para proseguir, sea $h: [0, 1] \rightarrow \mathbb{R}$ dada por

$$h(\theta) = (1 - \theta)^{\left(\frac{1-\theta}{\theta}\right)}. \quad (5.9)$$

Si $\Lambda = N_f \mu_{W \rightarrow R}$, $(h(\theta))^\Lambda$ describe el número total de células mutantes producidas por células silvestre en una población. Luego, si $\Lambda = N_f \mu_{W \rightarrow M}$, $(h(\theta))^\Lambda$ denota el número total de células mutadoras producidas por células silvestre en la población. Más aún, si $\Lambda = \mu_{M \rightarrow R}$, $(h(\theta))^\Lambda$ describe el número total de células mutantes con fenotipo mutador producidas por una única células mutadora.

De esta manera, $(h(h(\theta))^{\mu_{M \rightarrow R}})^{N \mu_{W \rightarrow M}}$ es la función generadora de probabilidad que describe el número total de células mutantes con fenotipo mutador en una población. Entonces, por aplicar el Teorema 15 se tiene que

$$F(\theta) = (h(\theta))^{\Lambda_{W \rightarrow R}} (h(h(\theta))^{\mu_{M \rightarrow R}})^{\Lambda_{W \rightarrow M}}, \quad (5.10)$$

con $\Lambda_{W \rightarrow R} = N_f \mu_{W \rightarrow R}$ y $\Lambda_{W \rightarrow M} = N_f \mu_{W \rightarrow M}$, es la función generadora de probabilidad que indica el número total de células mutantes.

La ecuación anterior resulta crucial pues, en general, no existen técnicas de laboratorio para distinguir a las células mutadoras en un cultivo, lo que impide realizar una estimación directa de la tasa de mutación hacia fenotipos mutadores. Sin embargo, la Ecuación (5.10) permite realizar tal estimación a través del número de células mutantes y el número de células mutantes con el fenotipo mutador, para las cuales sí existen técnicas experimentales para su distinción dado que se detecta mutaciones y no al fenotipo mutador *per se*.

Finalmente, las funciones generadoras de probabilidad no dependen de los datos experimentales. Sin embargo, se debe *traducir* dichos datos para llevar a cabo las estimaciones; por tanto, presentamos a la función generadora de probabilidad empírica.

Definición 24 (Función generadora de probabilidad empírica) Sea $\{X_1, \dots, X_n\}$ una muestra aleatoria de tamaño n y proveniente de una variable aleatoria discreta. Si $\mathfrak{F}_n: [0, 1] \rightarrow \mathbb{R}$ es una función definida por

$$\mathfrak{F}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \theta^{X_i}, \quad (5.11)$$

diremos que \mathfrak{F}_n es la función generadora de probabilidad empírica asociada a la muestra $\{X_1, \dots, X_n\}$.

Antes de presentar los métodos de estimación, discutamos el protocolo experimental al que deben apegarse. Se inicia r^* cultivos con N_0 células silvestre y una vez que han alcanzado el tamaño máximo N_f , se toma una muestra de aleatorias de $b + \delta_1$ células en cada uno de los cultivos. Con las b células se inicia un siguiente ciclo de crecimiento; mientras que, las δ_1 células se colocan en una placa para contabilizar los mutantes presentes.

El proceso anterior se itera tantas veces como se desee. Dicho protocolo se ha propuesto con el objetivo de magnificar la presencia y efecto de las células mutadoras. Dado que en las b células que iniciarán el nuevo cultivo puede contener más que sólo células silvestre.

Ahora, para continuar presentamos los métodos de estimación que hemos propuesto.

5.2. Ajuste directo

Como se mencionó previamente, al trabajar con la función generadora de probabilidad de Luria-Delbrück se estila considerar el intervalo $[0, 1]$ exclusivamente. Esto se debe principalmente a cuestiones biológicas, al considerar valores cercanos a cero, por la derecha, reflejan el efecto de la observación de cantidades *pequeñas* de células mutantes; mientras que, valores cercanos a uno, por la izquierda, reflejan la influencia de cantidades *grandes* de células mutantes en los cultivos celulares [64].

En trabajos como [57] y [64] se limitan a estudiar la función generadora de probabilidad en los puntos $\theta \in \Theta$, con $\Theta = \{0.1, 0.8, 0.9\}$, para generar las estimaciones de la tasa de mutación hacia células mutantes. No obstante, para las estimar la tasa de mutación hacia fenotipos mutadores consideraremos los puntos $\theta \in \Theta^*$, con $\Theta^* = \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

Esto, debido a que emplearemos la función generadora de probabilidad de la suma de la variable aleatoria que cuenta el número de células mutantes producidas por células silvestres y aquella que cuenta el número de células mutantes producidas por células mutadoras. Dado que se espera observar cantidades *pequeñas* de células mutantes esporádicamente y al considerar el conjunto θ^* modulamos podemos reflejar dicho esquema.

Supongamos que se han considerado r^* réplicas del experimento, cada una con c^* ciclos de crecimiento. Sea $m_i \in \mathbb{R}^{r^*}$, $i = 1, \dots, c^*$, un vector cuya j -ésima entrada, $j = 1, \dots, r^*$, está formada por la suma del número células mutantes producidas por células silvestres y el número células mutantes producidas por células mutadoras.

Para acoplar la información al proceso de dilución del protocolo experimental, definamos

el vector $n_i \in \mathbb{R}^{r^*}$, $i = 1, \dots, c^*$, tal que su j -ésima entrada, $j = 1, \dots, r^*$, corresponde a una realización de la variable aleatoria S_j

$$S_j \sim \text{Binomial}(m_{i,j}, \delta_i),$$

es decir, un muestreo binomial. Donde $m_{i,j}$ y $\delta_i \in (0, 1)$ denotan la j -ésima entrada del vector m_i y la dilución considerada para i -ésimo ciclo de crecimiento, respectivamente.

Antes de introducir el método de estimación, proponemos intervalos $I_{W \rightarrow R} = [Nl_{W \rightarrow R}, Nu_{W \rightarrow R}]$, $I_{M \rightarrow R} = [l_{M \rightarrow R}, u_{M \rightarrow R}]$ e $I_{W \rightarrow M} = [l_{W \rightarrow M}, u_{W \rightarrow M}]$ tales que

$$\begin{aligned} \log_{10}(\mu_{W \rightarrow R}) &\in I_{W \rightarrow R}, \\ \log_{10}(\mu_{M \rightarrow R}) &\in I_{M \rightarrow R} \text{ y} \\ \log_{10}(\mu_{W \rightarrow M}) &\in I_{W \rightarrow M}. \end{aligned}$$

Dado que, para proceder computacionalmente, lo anterior nos ofrece mayor precisión en la búsqueda de los estimados.

Luego, para trabajar con un número finito de posibilidades, particionaremos los intervalos a fin de considerar solo $(p + 1)^3$ puntos dentro del conjunto $I := I_{W \rightarrow R} \times I_{M \rightarrow R} \times I_{W \rightarrow M}$. Así, en lugar de trabajar con el intervalo $[a, b]$, $(a, b) \in \{(Nl_{W \rightarrow R}, Nu_{W \rightarrow R}), (l_{M \rightarrow R}, u_{M \rightarrow R}), (l_{W \rightarrow M}, u_{W \rightarrow M})\}$, trabajaremos con el conjunto de puntos

$$\left\{ a \left(\frac{b}{a} \right)^{q/p} : q = 0, 1, \dots, p \right\} \subset [a, b]. \quad (5.12)$$

Sea $F_1: \Theta^* \times I \rightarrow \mathbb{R}$ dada por

$$F_1(\theta, \iota_1, \iota_2, \iota_3) = h(\theta)^{\iota_1} h(h(\theta)^{\iota_2})^{\iota_3} \quad (5.13)$$

y sea $F_i: F_{i-1}(\Theta^*) \times I \rightarrow \mathbb{R}$ dada por

$$F_i(\theta, \iota_1, \iota_2, \iota_3) = h(F_{i-1}(\theta))^{\iota_1} h(h(F_{i-1}(\theta))^{\iota_2})^{\iota_3}, \quad (5.14)$$

para $i = 2, \dots, c^*$.

De esta manera, los estimadores $\widehat{\Delta_{W \rightarrow R i}}$, $\widehat{\mu_{M \rightarrow R i}}$ y $\widehat{\Delta_{W \rightarrow M i}}$ de $\Delta_{W \rightarrow R}$, $\mu_{W \rightarrow R}$ y $\Delta_{W \rightarrow M}$ en el i -ésimo ciclo de crecimiento, respectivamente, son aquellos que satisfacen

$$\begin{aligned} &\sum_{\theta \in \Theta^*} \left(F_i(\theta, \widehat{\Delta_{W \rightarrow R i}}, \widehat{\mu_{M \rightarrow R i}}, \widehat{\Delta_{W \rightarrow M i}}) - \mathfrak{G}_i(\theta) \right)^2 \\ &= \min_{(\iota_1, \iota_2, \iota_3) \in I} \sum_{\theta \in \Theta^*} \left(F_i(\theta, N10^{\iota_1}, 10^{\iota_2}, N10^{\iota_3}) - \mathfrak{G}_i(\theta) \right)^2, \end{aligned}$$

con \mathfrak{G}_i la función generadora de probabilidad empírica construida con los r^* datos del i -ésimo ciclo de crecimiento, $i = 1, \dots, c^*$.

5.3. Ajuste por métodos de máxima verosimilitud

Recurrir a métodos de máxima verosimilitud requiere de trabajar directamente con funciones de masa de probabilidad. En este caso, implicaría trabajar con la expresión explícita de la distribución Luria-Delbrück, la cual no admite una expresión analítica [155].

Sin embargo, al considerar el planteamiento del método anterior hasta la Ecuación (5.14) se tiene la información necesaria para recuperar los estados de probabilidad de la distribución Luria-Delbrück suficientes para implementar los métodos de máxima verosimilitud. Lo anterior, a través de emplear la transformada rápida de Fourier como en [7] y [82, Ec. 5].

Sea $d_i = \max_{j \in \{1, \dots, r^*\}} n_{i,j}$, con $n_{i,j}$ la j -ésima entrada del vector n_i , para $i = 1, \dots, r^*$. Es decir, d_i es la máxima cantidad de células mutantes presentes en el i -ésimo ciclo de crecimiento tras la dilución.

De esta manera, dada $\iota = (\iota_1, \iota_2, \iota_3) \in I$ y definiendo $\mathcal{F}_i(\theta; \iota) := F_i(\theta, N10^{\iota_1}, 10^{\iota_2}, N10^{\iota_3})$, se tiene que la parte real de la transformada rápida de Fourier de \mathcal{F}_i ,

$$\check{\mathcal{F}}(k) = \frac{1}{d_i} \sum_{l=0}^{d_i-1} \mathcal{F}_i \left(\frac{2\pi i l}{d_i}; \iota \right) \exp \left(-\frac{2\pi i l k}{d_i} \right), \quad (5.15)$$

con $i^2 = -1$, aproxima al valor de

$$p_{X_i}(k; \iota) = P(X_i = k; \iota),$$

$k = 0, 1, \dots, d_i$. Donde X_i es la variable aleatoria cuya función generadora de probabilidad es \mathcal{F}_i .

Por tanto,

$$\mathcal{L}_i(\iota) = \prod_{j=1}^{r^*} p_{X_i}(n_{i,j}; \iota)$$

denota la función de verosimilitud asociada al i -ésimo ciclo de crecimiento.

Implicando que los estimadores de máxima verosimilitud $\widehat{\Delta}_{W \rightarrow R i}^*$, $\widehat{\mu}_{M \rightarrow R i}^*$ y $\widehat{\Delta}_{W \rightarrow M i}^*$ de $\Delta_{W \rightarrow R}$, $\mu_{M \rightarrow R}$ y $\Delta_{W \rightarrow M}$ en el i -ésimo ciclo de crecimiento, respectivamente, son aquellos que satisfacen

$$\mathcal{L}_i(\widehat{\Delta}_{W \rightarrow R i}^*, \widehat{\mu}_{M \rightarrow R i}^*, \widehat{\Delta}_{W \rightarrow M i}^*) = \max_{\iota \in I} \mathcal{L}_i(\iota). \quad (5.16)$$

Capítulo 6

Resultados

A lo largo de este trabajo desarrollamos un par de modelos para emular la dinámica fenotípica y un par de métodos para la estimación de la tasa de mutación hacia fenotipos mutadores. A continuación presentamos los resultados encontrados.

6.1. Confiabilidad de los modelos

Como se mencionó en el Capítulo 4, se observa que la distribución del número de células mutantes producidas por las células silvestre es una distribución de cola pesada [64]. Esto es esperado, dado que la distribución Luria-Delbrück debe ser la distribución que subyace en las figuras 4.4a y 4.7a.

Para verificar lo anterior, empleamos la herramienta web *bz-rates* [57] que recurre a la metodología de [64]. Empleando los modelo Moran (ver Apéndice A.1), gráfico y el basado en el método de inversión construimos un conjunto de 100 simulaciones por método.

Para cada una de las simulaciones, los valores de los parámetros se fijaron de la siguiente manera: $N_0 = 1$, $N_f = 10^6$, $\mu_{W \rightarrow R} = 10^{-7}$, $\mu_{W \rightarrow M} = 0$ y $\mu_{M \rightarrow R} = 0$. Esto, con la finalidad de estandarizar y equiparar las condiciones.

La herramienta *bz-rates* realiza una prueba χ^2 para bondad de ajuste [38] con el objetivo

de determinar cuál de las siguientes hipótesis estadísticas es cierta:

- H_0 : Los datos se ajustan a la distribución Luria-Delbrück, o
- H_1 : Los datos **no** se ajustan a la distribución Luria-Delbrück.

En nuestro caso, los p -valores que se obtuvieron fueron todos mayores que 0.05, ver Tabla 6.1. Por tanto, se acepta la hipótesis nula implicando que, en efecto, las variables aleatorias que aparecen en las ecuaciones (4.68) y (4.85) tienen una distribución Luria-Delbrück.

Lo anterior es consecuencia de que un p -valor es una medida de significancia estadística, la cual asigna valores pequeños, usualmente menores que 0.05, en caso de tener evidencia contra la hipótesis nula [116].

	Moran	Gráfico	Método de inversion
Estimación de $\mu_{W \rightarrow R}$	2.05063×10^{-7}	1.841444×10^{-7}	1.237306×10^{-7}
Error en la estimación	1.05063×10^{-7}	8.4144×10^{-8}	2.3730×10^{-8}
Cota inferior del intervalo del 95 % de confianza	1.262×10^{-7}	1.098×10^{-7}	6.346×10^{-8}
Cota superior del intervalo del 95 % de confianza	2.839×10^{-7}	2.585×10^{-7}	1.840×10^{-7}
χ^2 de Pearson	1.6333	4.1370	1.0446
p -valor	0.8027	0.3877	0.9029

Tabla 6.1: Resultados reportados por *bz-rates*.

Además, en la Figura 6.1 puede apreciarse una comparativa, generada por *bz-rates*, respecto al comportamiento de los datos y el comportamiento de la distribución Luria-Delbrück. El ajustes de los datos construidos por el modelo Moran, el modelo gráfico y el modelo basado en el método de inversión se muestra en las figuras 6.1a, 6.1b y 6.1c, respectivamente.

Como se mencionó previamente, el modelo Moran es ampliamente utilizado; sin embargo, se sabe que su ejecución computacional consume grandes cantidades de tiempo. Por otra parte, nuestros métodos representa una alternativa más rápida.

No obstante, es crucial verificar si existen diferencias significativas entre las distribuciones de los números de mutantes que se obtienen con el clásico modelo Moran y con nuestros modelos. Para esto, se fijaron las tasas de mutación $\mu_{W \rightarrow R} = 10^{-7}$, $\mu_{W \rightarrow M} = 0$ y $\mu_{M \rightarrow R} = 0$ y la cantidad inicial de células en el cultivo $N_0 = 1$.

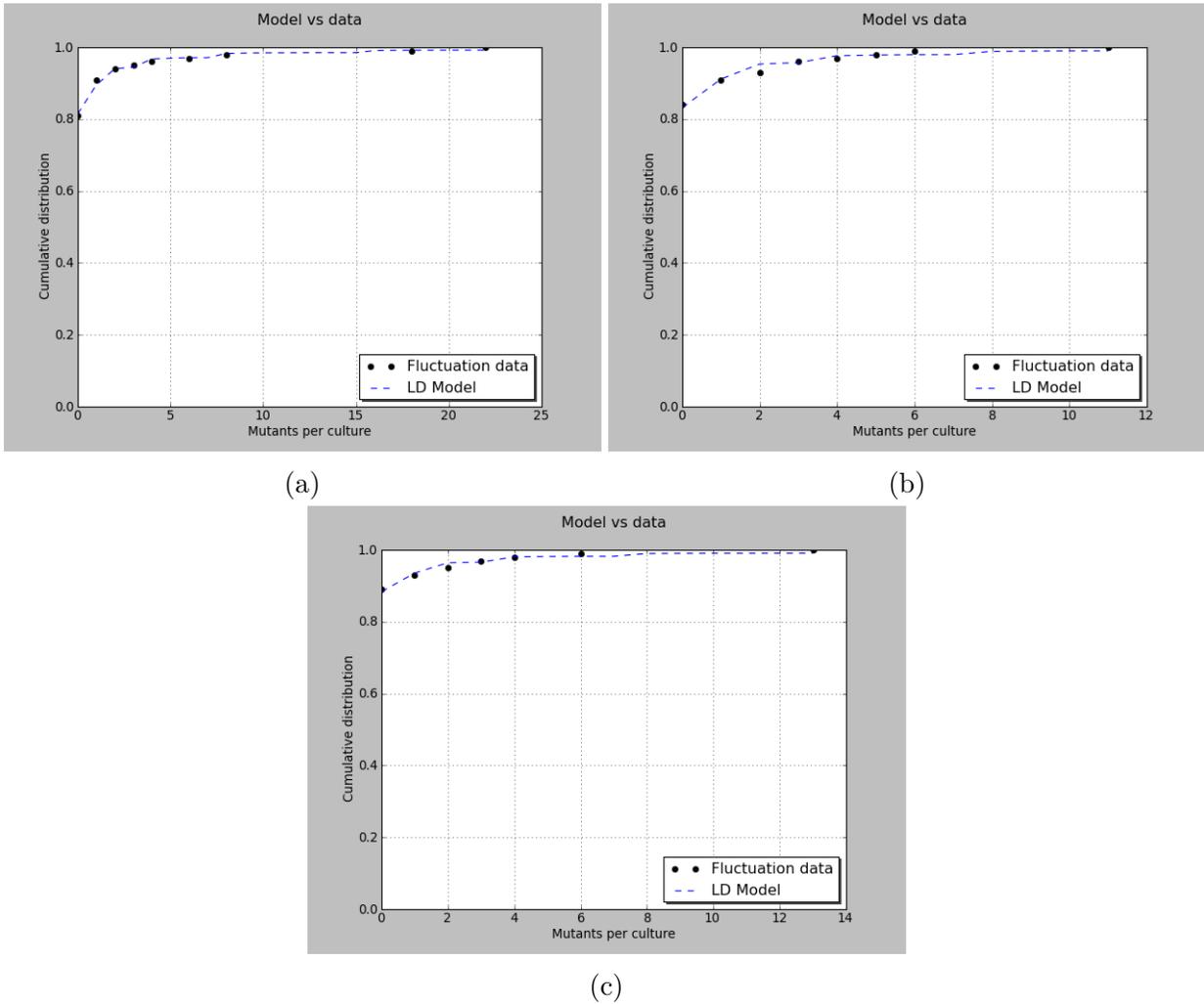


Figura 6.1: Ajuste de los datos a la distribución Luria-Delbrück.

En este caso, nuestro objetivo fue determinar si existen diferencias significativas al variar el tamaño final de la población. Por tanto, se consideró el siguiente conjunto de tamaños finales $A = \{20000, 40000, 60000, 80000, 100000, 200000, 300000, 400000, 500000, 600000, 700000, 800000, 900000, 1000000\}$.

De es manera, para cada $N_f \in A$ se generó un conjunto de 100 simulaciones por método. Luego, recurrimos a una modificación de la prueba Kolmogorov-Smirnov para dos muestras que provienen de variables aleatorias discretas [11]. Una prueba de bondad de ajuste la cual, dados dos conjuntos de datos D_1 y D_2 que provienen de algunas variables aleatorias discretas X y Y , respectivamente, determina cuál de las siguiente hipótesis estadísticas es

cierta:

H_0 : La distribución F_X y F_Y son iguales, o

H_1 : La distribución F_X y F_Y **no** son iguales

Donde F_X y F_Y denotan las funciones de distribución asociadas a las variables X y Y .

En este caso, los p -valores obtenidos fueron todos mayores que 0.05. Por tanto, se acepta la hipótesis nula e implicando que el modelo Moran, el modelo gráfico y el modelo basado en el método de inversión son equivalentes en cuanto la simulación del número de células mutantes producidas por células silvestre.

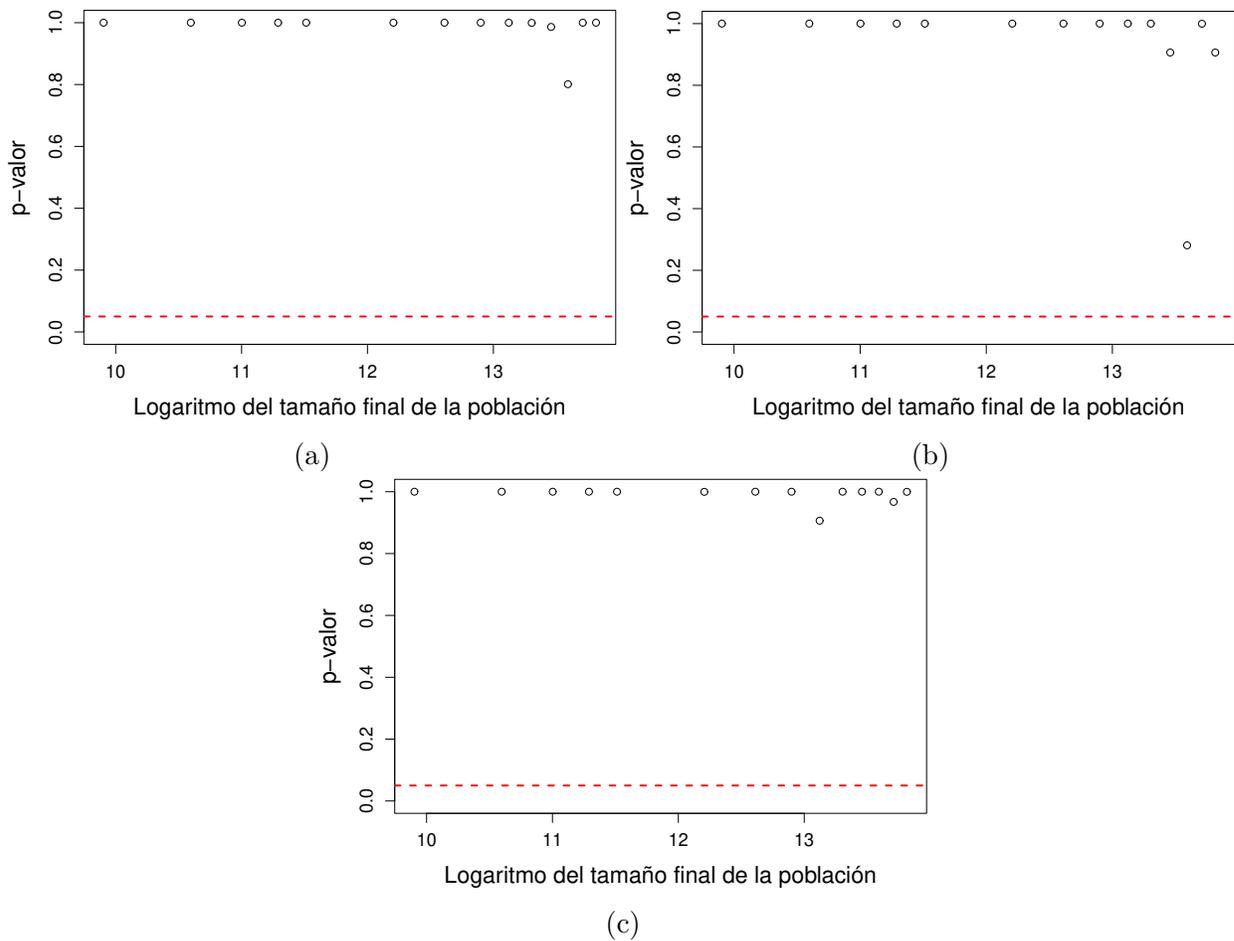


Figura 6.2: Comparaciones de las distribuciones de células mutantes obtenidas por los modelos Moran, gráfico y basado en el método de inversión.

Las comparaciones de los datos construidos con el modelo Moran respecto al modelo gráfico, el modelo Moran respecto al modelo basado en el método de inversión, y el modelo

gráfico respecto al modelo basado en el método de inversión se observan en las figuras 6.2a, 6.2b y 6.2c, respectivamente.

6.2. Confiabilidad de las estimaciones

Notamos, por inspección, que al considerar información de más de un ciclo de crecimiento las estimaciones no proveen mayor precisión. Por tanto, nos restringimos a discutir los resultados que se obtienen al utilizar un único ciclo de crecimiento.

Presentamos la estimación puntual y del intervalo del 95% de confianza para $\mu_{W \rightarrow M}$ a través de los siguientes métodos

1. Bootstrap del ajuste directo [112, Cap. 6].
2. Región de máxima verosimilitud ([48]; [155, Ec. 14]).
3. Jackknife del ajuste por métodos de máxima verosimilitud [112, Cap. 6].
4. Región de máxima verosimilitud suponiendo que

$$\mu_{M \rightarrow R} = k\mu_{W \rightarrow R},$$

para $k \in \mathbb{R}$ constante.

5. Jackknife del ajuste por métodos de máxima verosimilitud suponiendo que

$$\mu_{M \rightarrow R} = k\mu_{W \rightarrow R},$$

para $k \in \mathbb{R}$ constante.

Para estandarizar, dada una tasa de mutación η , $\eta \in \{\mu_{W \rightarrow R}, \mu_{W \rightarrow M}, \mu_{M \rightarrow R}\}$ denifimos al intervalo J , $J \in \{I_{W \rightarrow R}, I_{W \rightarrow M}, I_{M \rightarrow R}\}$, como

$$J = [\log_{10}(\eta) - 2, \log_{10}(\eta) + 2]. \quad (6.1)$$

Además, el parámetro para la discretización se fija como $p^* = 20$ y, como consideraremos un único ciclo de crecimiento, se define $c^* = 1$.

Los cinco métodos para la estimación puntual e intervalo de confianza se aplicarán a un par de conjuntos de datos, cada uno constituido por $r^* = 150$ simulaciones. El primer

conjunto de datos se construyó con el modelo gráfico empleando los siguientes parámetros: $N_0 = 1$, $N_f = 2^{33}$, $\mu_{W \rightarrow R} = 10^{-7}$, $\mu_{W \rightarrow M} = 10^{-6}$, $\mu_{M \rightarrow R} = 5 \times 10^{-5}$. Mientras que, el segundo conjunto de datos se construyó a través del modelo basado en el método de inversión considerando los siguientes parámetros: $N_0 = 1$, $N_f = 2^{33}$, $\mu_{W \rightarrow R} = 10^{-7}$, $\mu_{W \rightarrow M} = 10^{-5}$, $\mu_{M \rightarrow R} = 5 \times 10^{-4}$.

Para emplear el Método 1, se aplicó el método de ajuste directo 500 veces, considerando en cada una de ellas a $\delta_1 = 2^{20}/2^{33}$ como parámetro de dilución. Respecto al Método 2, el método de ajuste por métodos de máxima verosimilitud fue empleado una única ocasión y tomando $\delta_1 = 2^{20}/2^{33}$.

En tanto al Método 3, se aplicó el método de ajuste por métodos de máxima verosimilitud 500 veces, en cada ocasión se consideró un remuestreo con reemplazo de 50 cultivos y el parámetro de dilución se determinó a través de $\delta_1 = u$, con u una realización de una variable aleatoria $U \sim \text{Uniforme}(0.0000005, 2^{20}/2^{33})$.

Luego, para el Método 4, se empleó el método de ajuste por métodos de máxima verosimilitud fue empleado una única ocasión con $\delta_1 = 2^{20}/2^{33}$. Respecto a la constante k , es un parámetro denominado *fuerza del mutador* y se tomó $k = 500$ para el conjunto de datos construido con el método gráfico y $k = 5000$ para el conjunto de datos construido con el método basado en el método de inversión.

Por último, para el Método 5 se aplicó el método de ajuste por métodos de máxima verosimilitud 500 veces, en cada ocasión se consideró un remuestreo con reemplazo de 50 cultivos y el parámetro de dilución se determinó a través de $\delta_1 = u$, con u una realización de una variable aleatoria $U \sim \text{Uniforme}(0.0000005, 2^{20}/2^{33})$. Además, respecto a la fuerza del mutador se tomó $k = 500$ para el conjunto de datos construido con el método gráfico y $k = 5000$ para el conjunto de datos construido con el método basado en el método de inversión.

De esta manera, aplicamos los cinco métodos previamente discutidos a cada uno de los conjuntos de datos. En las tablas 6.2 y 6.3 se muestran los resultados obtenidos referentes a conjunto de datos construido con el modelo gráfico y el modelo basado en el método de inversión, respectivamente.

Método	Estimación de $\mu_{W \rightarrow M}$	Intervalo del 95 % de confianza
1	4.446193×10^{-6}	$(1 \times 10^{-8}, 1.165914 \times 10^{-5})$
2	8.652343×10^{-7}	$(1.874255 \times 10^{-7}, 1 \times 10^{-4})$
3	1.028965×10^{-6}	$(1 \times 10^{-8}, 3.434306 \times 10^{-6})$
4	1.656201×10^{-5}	$(2.168967 \times 10^{-6}, 9.211989 \times 10^{-5})$
5	1.542048×10^{-6}	$(1 \times 10^{-8}, 5.163048 \times 10^{-5})$

Tabla 6.2: Estimaciones e intervalos del 95 % de confianza para el conjunto de datos generado con el modelo gráfico, siendo $\mu_{W \rightarrow M} = 10^{-6}$ el valor real de la tasa de mutación.

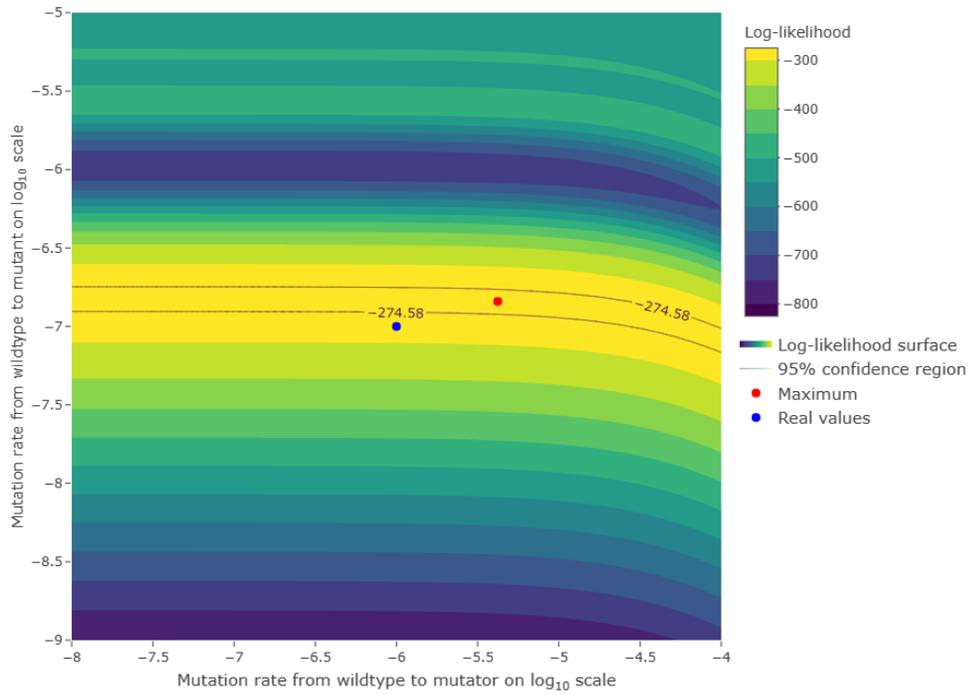
Método	Estimación de $\mu_{W \rightarrow M}$	Intervalo del 95 % de confianza
1	1.958892×10^{-4}	$(1 \times 10^{-7}, 1 \times 10^{-3})$
2	4.241429×10^{-7}	$(1 \times 10^{-8}, 1 \times 10^{-4})$
3	1.864772×10^{-6}	$(1 \times 10^{-7}, 7.416085 \times 10^{-4})$
4	2.604188×10^{-6}	$(1 \times 10^{-7}, 1.416185 \times 10^{-5})$
5	4.05051×10^{-5}	$(1 \times 10^{-7}, 1.572404 \times 10^{-5})$

Tabla 6.3: Estimaciones e intervalos del 95 % de confianza para el conjunto de datos generado con el modelo basado en el método de inversión, siendo $\mu_{W \rightarrow M} = 10^{-5}$ el valor real de la tasa de mutación.

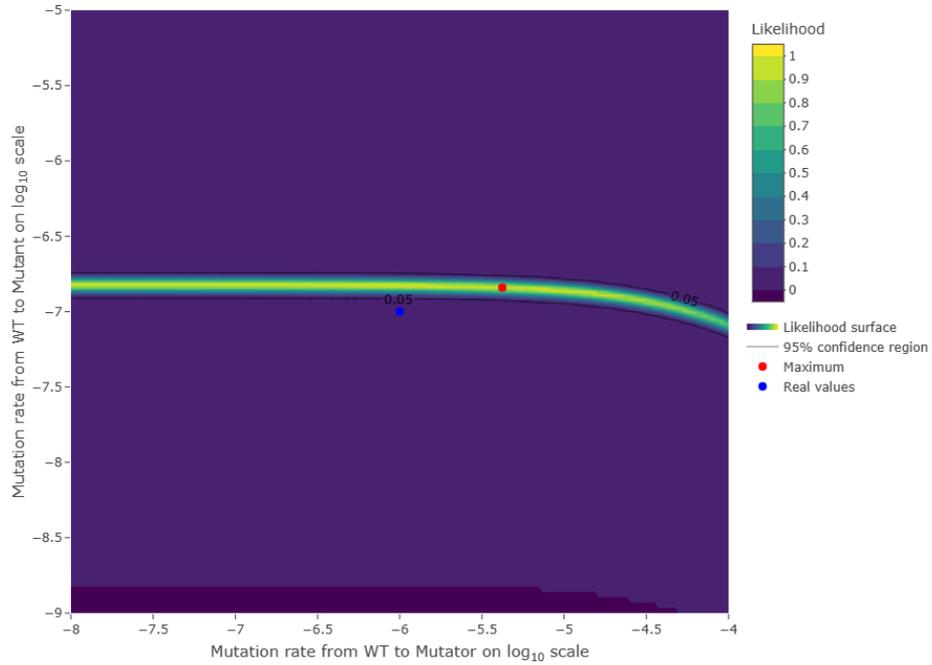
Más aún, ya que el Método 4 está dado en términos de dos variables, $\mu_{W \rightarrow R}$ y $\mu_{W \rightarrow M}$, y la función de verosimilitud entonces podemos observar su gráfica en \mathbb{R}^3 . Así, presentamos a continuación la gráfica de curvas de nivel de la superficie de verosimilitud, para las cuales consideramos $p^* = 150$ a fin de obtener una mejor resolución de imagen.

En las figuras 6.3 y 6.4, puede apreciarse las curvas de nivel del logaritmo de la función de verosimilitud, $\log(\mathcal{L}_1)$, y del re-escalamiento de la función de verosimilitud donde $(\max_{\iota \in I} \{\mathcal{L}_1(\iota)\})^{-1}$ se utilizo como factor de escalamiento de manera que $\mathcal{L}_1(I) = [0, 1]$.

Además, la región delimitada por las curvas en color negro denotan la región del 95 % de confianza, el punto en color rojo muestra los valores de las tasas de mutación, en logaritmo base diez, en las que se alcanzó el valor máximo de la función; mientras que, el punto en color azul representa los valores reales de las tasas de mutación, en logaritmo base diez.

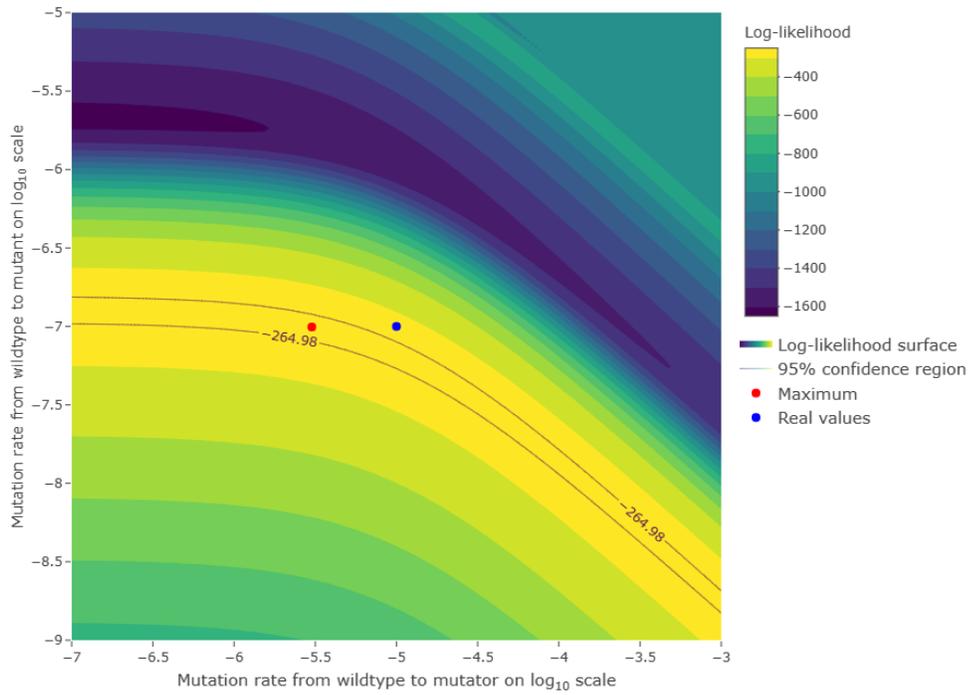


(a)

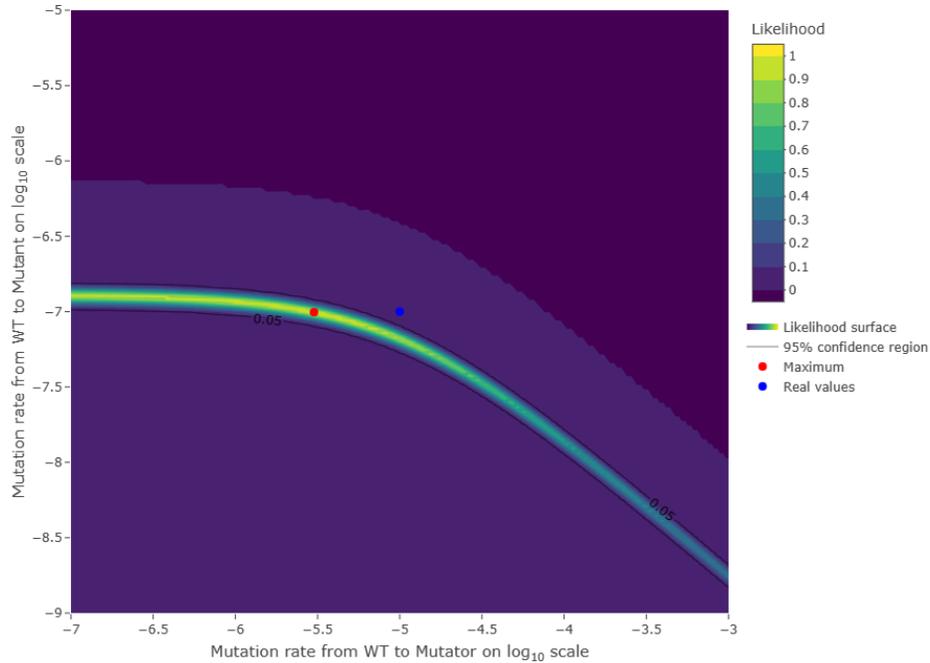


(b)

Figura 6.3: Curvas de nivel del logaritmo de la función de verosimilitud y del re-escalamiento al intervalo $[0, 1]$ de la función de verosimilitud asociadas al conjunto de datos construido con el modelo gráfico.



(a)



(b)

Figura 6.4: Curvas de nivel del logaritmo de la función de verosimilitud y del re-escalamiento al intervalo $[0, 1]$ de la función de verosimilitud asociadas al conjunto de datos construido con el modelo basado en el método de inversión.

Capítulo 7

Conclusiones

Respecto a los modelos propuestos en este trabajo, ofrecen una alternativa comparable al clásico modelo Moran y, además, su ejecución requiere de un menor tiempo computacional.

Asimismo, como se puede observar en la Tabla 6.1, las estimaciones de la tasa de mutación de células silvestre a células mutantes reportan un error más pequeño cuando se utilizan simulaciones construidas con el modelo gráfico y el modelo basado en el método de inversión en comparación con las simulaciones generadas mediante el modelo Moran.

Tras analizar la Figura 6.2, se concluye que el modelo gráfico y el modelo basado en el método de inversión son comparables al modelo Moran, incluso cuando se varían los tamaños de la población final. Además, como se observa en la Figura 6.2c, el modelo gráfico y el modelo basado en el método de inversión son comparables entre sí.

En cuanto a las estimaciones puntuales de la tasa de mutación hacia fenotipos mutadores y la obtención de los intervalos de confianza del 95 %, la metodología desarrollada ha demostrado ser efectiva.

Los métodos (1) de bootstrap del ajuste directo y (2) de la región de máxima verosimilitud contienen el valor real de $\mu_{W \rightarrow M}$. Lo anterior fue observado en las estimaciones que se obtienen en ambos casos, al considerar los datos generados mediante el modelo gráfico como aquellos datos generados mediante el modelo basado en el método de inversión.

No obstante, es importante destacar que los métodos bootstrap pueden fallar en caso de lidiar con distribuciones de cola pesada [13]. Por tanto, proponemos que el Método (2) es el

que ofrece el mejor desempeño, dado que la distribución Luria-Delbrück es una distribución con la característica antes mencionada.

Bibliografía

- [1] Abercrombie, M., “Contact inhibition in tissue culture,” *In Vitro*, vol. 6, n.º 2, págs. 128-142, sep. de 1970.
- [2] Adjiri, A., “DNA Mutations May Not Be the Cause of Cancer,” *Oncology and Therapy*, vol. 5, n.º 1, págs. 85-101, mayo de 2017.
- [3] Akram, F., Imtiaz, M. y Haq, I. ul, “Emergent crisis of antibiotic resistance: A silent pandemic threat to 21st century,” *Microbial Pathogenesis*, vol. 174, pág. 105923, 2023.
- [4] Aktipis, A., *The Cheating Cell: How Evolution Helps Us Understand and Treat Cancer*. Princeton University Press, 2020.
- [5] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. y Walter, P., *Molecular Biology of the Cell*. Garland Science, 2002, <https://www.ncbi.nlm.nih.gov/books/NBK26840/>.
- [6] Alessio, N., Aprile, D., Squillaro, T. et al., “The senescence-associated secretory phenotype (SASP) from mesenchymal stromal cells impairs growth of immortalized prostate cells but has no effect on metastatic prostatic cancer cells,” *Aging*, vol. 11, n.º 15, págs. 5817-5828, ago. de 2019.
- [7] Alexander, H. K., “Condition Distributions and Waiting Times in Multitype Branching Processes,” *Advances in Applied Probability*, vol. 45, n.º 3, págs. 692-718, 2013.
- [8] Alhmoud, J. F., Woolley, J. F., Moustafa, A.-E. A. y Malki, M. I., “DNA Damage/Repair Management in Cancers,” *Cancers*, vol. 12, n.º 4, pág. 1050, abr. de 2020.
- [9] Anandakrishnan, R., Varghese, R. T., Kinney, N. A. y Garner, H. R., “Estimating the number of genetic mutations (hits) required for carcinogenesis based on the distribution of somatic mutations,” *PLoS Computational Biology*, vol. 15, n.º 3, N. L. Komarova, ed., e1006881, mar. de 2019.

- [10] Armitage, P. y Doll, R., “The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis,” *British Journal of Cancer*, vol. 8, n.º 1, págs. 1-12, mar. de 1954.
- [11] Arnold, T. B. and Emerson, J. W., “Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions,” *R Journal*, vol. 3, n.º 2, 2011.
- [12] Ashley, D. J., “The two “hit” and multiple “hit” theories of carcinogenesis.,” *British Journal of Cancer*, vol. 23, n.º 2, págs. 313-328, jun. de 1969.
- [13] Athreya, K. B., “Bootstrap of the Mean in the Infinite Variance Case,” *The Annals of Statistics*, vol. 15, n.º 2, págs. 724-731, 1987.
- [14] Baird, D., “Telomeres,” en *Encyclopedia of Gerontology (Second Edition)*, J. E. Birren, ed., Second Edition, New York: Elsevier, 2007, págs. 593-601.
- [15] Basu, A., “DNA Damage, Mutagenesis and Cancer,” en, *International Journal of Molecular Sciences*, vol. 19, n.º 4, págs. 970, 23 de mar. de 2018.
- [16] Bebenek, A. y Ziuzia-Graczyk, I., “Fidelity of DNA replication—a matter of proofreading,” *Current Genetics*, vol. 64, n.º 5, págs. 985-996, mar. de 2018.
- [17] Bennett, M. D., “The time and duration of meiosis,” *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 277, n.º 955, págs. 201-226, mar. de 1977.
- [18] Berman, J. J., “Chapter 6 - Using time to solve cause and effect dilemmas,” en *Logic and Critical Thinking in the Biomedical Sciences*, J. J. Berman, ed., Academic Press, 2020, págs. 167-199.
- [19] Biotech Spain, *Visión general de las células del sistema inmunitario*, <https://biotech-spain.com/en/articles/visi-n-general-de-las-c-lulas-del-sistema-inmunitario/>, 2016.
- [20] Blankenstein, T., Coulie, P. G., Gilboa, E. y Jaffee, E. M., “The determinants of tumour immunogenicity,” *Nature Reviews Cancer*, vol. 12, n.º 4, págs. 307-313, mar. de 2012.
- [21] Bode, A. M. y Dong, Z., “Epigallocatechin,” en *Encyclopedia of Cancer*, M. Schwab, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, págs. 1012-1015.
- [22] Bodnar, A. G., “Cellular and molecular mechanisms of negligible senescence: insight from the sea urchin,” *Invertebrate reproduction & development*, vol. 59, n.º sup1, págs. 23-27, 2015.
- [23] Böttcher, J. P., Bonavita, E., Chakravarty, P. et al., “NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control,” *Cell*, vol. 172, n.º 5, 1022-1037.e14, feb. de 2018.

- [24] Boutry, J., Dujon, A. M., Gerard, A.-L. et al., “Ecological and Evolutionary Consequences of Anticancer Adaptations,” *iScience*, vol. 23, n.º 11, pág. 101716, 2020.
- [25] British Society for Immunology, *Dendritic cells. British Society for Immunology*, <https://www.immunology.org/public-information/bitesized-immunology/cells/dendritic-cells>, [Accesado en línea: 2022-10-15], 2022.
- [26] Brodie, E. D., Donoghue, M. J., Levin, S. A. et al., “Phenotypic Selection on Quantitative Traits,” en *The Princeton Guide to Evolution*. Princeton University Press, 2014, págs. 221-229.
- [27] Brown, T., *Genomes*, 2.^a ed. Oxford: Wiley-Liss, 2002.
- [28] Brunetti, M., Mackey, M. C. y Craig, M., “Understanding Normal and Pathological Hematopoietic Stem Cell Biology Using Mathematical Modelling,” *Current Stem Cell Reports*, vol. 7, n.º 3, págs. 109-120, jun. de 2021.
- [29] Calcinotto, A., Kohli, J., Zagato, E., Pellegrini, L., Demaria, M. y Alimonti, A., “Cellular Senescence: Aging, Cancer, and Injury,” *Physiological Reviews*, vol. 99, n.º 2, págs. 1047-1078, abr. de 2019.
- [30] Caldwell, J. C., “Chapter 13 - Alterations in cell proliferation, cell death, or nutrient supply,” en *Tumour Site Concordance and Mechanisms of Carcinogenesis: Volume 165*, ép. IARC scientific publications, R. Baan, B. Stewart y K. Straif, eds., World Health Organization, 2019.
- [31] Casás-Selves, M. y DeGregori, J., “How Cancer Shapes Evolution and How Evolution Shapes Cancer,” en, *Evolution: Education and Outreach*, vol. 4, n.º 4, págs. 624-634, dic. de 2011.
- [32] Catherine Sánchez, N., “Conociendo and comprendiendo la célula cancerosa: Fisiopatología del cáncer,” *Revista Médica Clínica Las Condes*, vol. 24, n.º 4, págs. 553-562, 2013, Tema central: Cáncer.
- [33] Cerwenka, A. y Lanier, L. L., “Natural killer cell memory in infection, inflammation and cancer,” *Nature Reviews Immunology*, vol. 16, n.º 2, págs. 112-123, ene. de 2016.
- [34] Chambers, C. R., Ritchie, S., Pereira, B. A. y Timpson, P., “Overcoming the senescence-associated secretory phenotype (SASP): a complex mechanism of resistance in the treatment of cancer,” *Molecular Oncology*, vol. 15, n.º 12, págs. 3242-3255, jul. de 2021.
- [35] Chiang, C., *Introduction to Stochastic Processes in Biostatistics* (Wiley Series in Probability and Statistics - Applied Probability and Statistics Section). Wiley, 1968.
- [36] Chopra, I., O’Neill, A. J. y Miller, K., “The role of mutators in the emergence of antibiotic-resistant bacteria,” *Drug Resistance Updates*, vol. 6, n.º 3, págs. 137-145, 2003.

- [37] Choudhury, S., Huang, A. Y., Kim, J. et al., “Somatic mutations in single human cardiomyocytes reveal age-associated DNA damage and widespread oxidative genotoxicity,” *Nature Aging*, vol. 2, n.º 8, págs. 714-725, ago. de 2022.
- [38] Cochran, W. G., “The χ^2 test of goodness of fit,” *The Annals of Mathematical Statistics*, págs. 315-345, 1952.
- [39] Collyer, M. L. y Adams, D. C., “Analysis of Two-State Multivariate Phenotypic Change in Ecological Studies,” *Ecology*, vol. 88, n.º 3, págs. 683-692, 2007.
- [40] Coussens, L. M. y Werb, Z., “Inflammation and cancer,” *Nature*, vol. 420, n.º 6917, págs. 860-867, dic. de 2002.
- [41] Crane, G., “A modified Luria-Delbrück fluctuation assay for estimating and comparing mutation rates,” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 354, n.º 2, págs. 171-182, jul. de 1996.
- [42] Donskov, F., “Immunomonitoring and prognostic relevance of neutrophils in clinical trials,” *Seminars in Cancer Biology*, vol. 23, n.º 3, págs. 200-207, jun. de 2013.
- [43] Durrett, R., *Probability: Theory and Examples*. Cambridge, Reino Unido: Cambridge University Press, 2019.
- [44] Embrechts, P. y Hofert, M., “A note on generalized inverses,” *Mathematical Methods of Operations Research*, vol. 77, n.º 3, págs. 423-432, abr. de 2013.
- [45] Eyre-Walker, A. y Keightley, P. D., “The distribution of fitness effects of new mutations,” *Nature Reviews Genetics*, vol. 8, n.º 8, págs. 610-618, ago. de 2007.
- [46] Fan, H. y Chu, J.-Y., “A Brief Review of Short Tandem Repeat Mutation,” *Genomics, Proteomics & Bioinformatics*, vol. 5, n.º 1, págs. 7-14, 2007.
- [47] Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A. y Fares, Y., “Molecular principles of metastasis: a hallmark of cancer revisited,” *Signal transduction and targeted therapy*, vol. 5, n.º 1, pág. 28, 2020.
- [48] Feng, Z. y McCulloch, C. E., “Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space,” *Statistics & Probability Letters*, vol. 13, n.º 4, págs. 325-332, 1992.
- [49] Feo, F., “Preneoplastic Lesions,” en *Encyclopedia of Cancer*, M. Schwab, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, págs. 2417-2423.
- [50] Ferlay, J., Colombet, M., Soerjomataram, I. et al., “Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods,” *International Journal of Cancer*, vol. 144, n.º 8, págs. 1941-1953, dic. de 2018.
- [51] Fox, E. J., Prindle, M. J. y Loeb, L. A., “Do mutator mutations fuel tumorigenesis?” *Cancer and Metastasis Reviews*, vol. 32, n.º 3-4, págs. 353-361, abr. de 2013.

- [52] Franco, I. y Eriksson, M., “Reverting to old theories of ageing with new evidence for the role of somatic mutations,” *Nature Reviews Genetics*, vol. 23, n.º 11, págs. 645-646, jun. de 2022.
- [53] Fridlender, Z. G., Sun, J., Kim, S. et al., “Polarization of Tumor-Associated Neutrophil Phenotype by TGF-beta: “N1” versus “N2” TAN,” *Cancer Cell*, vol. 16, n.º 3, págs. 183-194, sep. de 2009.
- [54] Garber, S., *Biology: A Self-Teaching Guide* (Self-teaching guide), 3rd. Jossey Bass, 2020.
- [55] Gienapp, P. y Merilä, J., “Evolutionary Responses to Climate Change,” en *Encyclopedia of the Anthropocene*, D. A. Dellasala y M. I. Goldstein, eds., Oxford: Elsevier, 2018, págs. 51-59.
- [56] Gilchrist, W., “Galton misrepresented,” *Significance*, vol. 2, n.º 3, págs. 136-137, 2005.
- [57] Gillet-Markowska, A., Louvel, G. y Fischer, G., “bz-rates: A Web Tool to Estimate Mutation Rates from Fluctuation Analysis,” *G3 Genes|Genomes|Genetics*, vol. 5, n.º 11, págs. 2323-2327, nov. de 2015.
- [58] Giuliani, M. y Poggi, A., “Checkpoint Inhibitors and Engineered Cells: New Weapons for Natural Killer Cell Arsenal Against Hematological Malignancies,” *Cells*, vol. 9, n.º 7, 2020.
- [59] Glencross, D. A., Ho, T.-R., Camiña, N., Hawrylowicz, C. M. y Pfeffer, P. E., “Air pollution and its effects on the immune system,” *Free Radical Biology and Medicine*, vol. 151, págs. 56-68, mayo de 2020.
- [60] Gonzalez, H., Hagerling, C. y Werb, Z., “Roles of the immune system in cancer: from tumor initiation to metastatic progression,” *Genes & Development*, vol. 32, n.º 19-20, págs. 1267-1284, oct. de 2018.
- [61] Gould, C. V., Sniegowski, P. D., Shchepetov, M., Metlay, J. P. y Weiser, J. N., “Identifying Mutator Phenotypes among Fluoroquinolone-Resistant Strains of *Streptococcus pneumoniae* Using Fluctuation Analysis,” *Antimicrobial Agents and Chemotherapy*, vol. 51, n.º 9, págs. 3225-3229, sep. de 2007.
- [62] Hahn, W. C. y Meyerson, M., “Telomerase activation, cellular immortalization and cancer,” *Annals of Medicine*, vol. 33, n.º 2, págs. 123-129, ene. de 2001.
- [63] Hall, B. M., Ma, C.-X., Liang, P. y Singh, K. K., “Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis,” *Bioinformatics*, vol. 25, n.º 12, págs. 1564-1565, abr. de 2009.
- [64] Hamon, A. e Ycart, B., “Statistics for the Luria-Delbrück distribution,” *Electronic Journal of Statistics*, vol. 6, n.º none, págs. 1251-1272, 2012.

- [65] Henriques, C. M. y Ferreira, M. G., “Consequences of telomere shortening during lifespan,” *Current Opinion in Cell Biology*, vol. 24, n.º 6, págs. 804-808, dic. de 2012.
- [66] Hirashima, K., Migita, T., Sato, S., Muramatsu, Y., Ishikawa, Y. y Seimiya, H., “Telomere Length Influences Cancer Cell Differentiation *In Vivo*,” *Molecular and Cellular Biology*, vol. 33, n.º 15, págs. 2988-2995, ago. de 2013.
- [67] Hogg, R. V., McKean, J. W. y Craig, A. T., *Introduction to Mathematical Statistics*. Boston, E.U.A.: Pearson Education, Inc., 2019.
- [68] Iglesia, F. de la, Martínez, F., Hillung, J. et al., “Luria-Delbrück Estimation of Turnip Mosaic Virus Mutation Rate *In Vivo*,” *Journal of Virology*, vol. 86, n.º 6, págs. 3386-3388, mar. de 2012.
- [69] Jacod, J. y Protter, P., *Probability Essentials* (Universitext), 1st. Springer Berlin Heidelberg, 2000.
- [70] Jaishankar, J. y Srivastava, P., “Molecular Basis of Stationary Phase Survival and Applications,” *Frontiers in Microbiology*, vol. 8, oct. de 2017.
- [71] Kirsch, D. R. y DiDomenico, B. J., “CHAPTER 6 - Mechanism-Based Screening for the Discovery of Novel Antifungals,” en *Discovery of Novel Natural Products with Therapeutic Potential*, V. P. Gullo, ed., Boston: Newnes, 1994, págs. 177-221.
- [72] Koch, A. L., “Mutation and growth rates from Luria-Delbrück fluctuation tests,” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 95, n.º 2-3, págs. 129-143, ago. de 1982.
- [73] Kolaczowska, E. y Kubes, P., “Neutrophil recruitment and function in health and inflammation,” *Nature Reviews Immunology*, vol. 13, n.º 3, págs. 159-175, feb. de 2013.
- [74] Kolmogorov, A. N., *Foundations of the Theory of Probability: Second English Edition*. Courier Dover Publications, 2018.
- [75] Kosterlitz, O., Tirado, A. M., Wate, C. et al., “Estimating the rate of plasmid transfer with an adapted Luria–Delbrück fluctuation analysis,” *bioRxiv*, págs. 2021-01, 2022.
- [76] Kulesz-Martin, M., Ouyang, X., Barling, A., Gallegos, J., Liu, Y. y Medler, T., “7.02 - Multistage Carcinogenesis: Cell and Animal Models,” en *Comprehensive Toxicology (Third Edition)*, C. A. McQueen, ed., Third Edition, Oxford: Elsevier, 2018, págs. 11-35.
- [77] Kumar, B. V., Connors, T. J. y Farber, D. L., “Human T Cell Development, Localization, and Function throughout Life,” *Immunity*, vol. 48, n.º 2, págs. 202-213, feb. de 2018.

- [78] kumar, N. y Sethi, G., "Telomerase and hallmarks of cancer: An intricate interplay governing cancer cell evolution," *Cancer Letters*, vol. 578, pág. 216-459, 2023.
- [79] Kumari, R. y Jat, P., "Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype," *Frontiers in Cell and Developmental Biology*, vol. 9, mar. de 2021.
- [80] Kussell, E. y Leibler, S., "Phenotypic diversity, population growth, and information in fluctuating environments," *Science*, vol. 309, n.º 5743, págs. 2075-2078, 2005.
- [81] Lang, G. I., "Measuring Mutation Rates Using the Luria-Delbrück Fluctuation Assay," en *Methods in Molecular Biology*, Springer New York, oct. de 2017, págs. 21-31.
- [82] Lange, K., "Calculation of the Equilibrium Distribution for a Deleterious Gene by the Finite Fourier Transform," *Biometrics*, vol. 38, n.º 1, págs. 79-86, 1982.
- [83] Lavin, Y., Mortha, A., Rahman, A. y Merad, M., "Regulation of macrophage development and function in peripheral tissues," *Nature Reviews Immunology*, vol. 15, n.º 12, págs. 731-744, nov. de 2015.
- [84] Lea, D. E. y Coulson, C. A., "The distribution of the numbers of mutants in bacterial populations," *Journal of Genetics*, vol. 49, n.º 3, págs. 264-285, dic. de 1949.
- [85] Lewis, W. H. y Lewis, M. R., "The duration of the various phases of mitosis in the mesenchyme cells of tissue cultures," *The Anatomical Record*, vol. 13, n.º 6, págs. 359-367, nov. de 1917.
- [86] Little, M. y Wright, E., "A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data," *Mathematical Biosciences*, vol. 183, n.º 2, págs. 111-134, 2003.
- [87] Loeb, L. A., Springgate, C. F. y Battula, N., "Errors in DNA replication as a basis of malignant changes," *Cancer research*, vol. 34, n.º 9, págs. 2311-2321, 1974.
- [88] Loeb, L. A., "Cancer Cells Exhibit a Mutator Phenotype," en *Advances in Cancer Research*, Elsevier, 1997, págs. 25-56.
- [89] Luebeck, E. G. y Moolgavkar, S. H., "Multistage carcinogenesis and the incidence of colorectal cancer," *Proceedings of the National Academy of Sciences*, vol. 99, n.º 23, págs. 15 095-15 100, nov. de 2002.
- [90] Luria, S. E. y Delbrück, M., "Mutations of Bacteria from Virus Sensitivity to Virus Resistance," *Genetics*, vol. 28, n.º 6, págs. 491-511, nov. de 1943.
- [91] Luzzatto, L., "Somatic mutations in cancer development," *Environmental Health*, vol. 10, n.º Suppl 1, 2011.

- [92] MacQueen, A. J. y Hochwagen, A., “Checkpoint mechanisms: the puppet masters of meiotic prophase,” *Trends in Cell Biology*, vol. 21, n.º 7, págs. 393-400, jul. de 2011.
- [93] Maier, R. M. y Pepper, I. L., “Chapter 3 - Bacterial Growth,” en *Environmental Microbiology (Third Edition)*, I. L. Pepper, C. P. Gerba y T. J. Gentry, eds., Third Edition, San Diego: Academic Press, 2015, págs. 37-56.
- [94] Mazouzi, A., Vigouroux, A., Aikeshv, B. et al., “Insight into mechanisms of 3'-5'; exonuclease activity and removal of bulky 8,5'-cyclopurine adducts by apurinic/aprimidinic endonucleases,” *Proceedings of the National Academy of Sciences*, vol. 110, n.º 33, E3071-E3080, 2013.
- [95] Mildner, A. y Jung, S., “Development and Function of Dendritic Cell Subsets,” *Immunity*, vol. 40, n.º 5, págs. 642-656, mayo de 2014.
- [96] Milholland, B., Auton, A., Suh, Y. y Vijg, J., “Age-related somatic mutations in the cancer genome,” *Oncotarget*, vol. 6, n.º 28, págs. 24 627-24 635, sep. de 2015.
- [97] Nakamura, M. y Pérez-Abreu, V., “Empirical probability generating function: An overview,” *Insurance: Mathematics and Economics*, vol. 12, n.º 3, págs. 287-295, 1993.
- [98] Natali, F. y Rancati, G., “The Mutator Phenotype: Adapting Microbial Evolution to Cancer Biology,” *Frontiers in Genetics*, vol. 10, ago. de 2019.
- [99] National Human Genome Research Institute, *Telomere. National Human Genome Research Institute*, <https://www.genome.gov/genetics-glossary/Telomere>, [Accesado en línea: 2022-10-07], oct. de 2022.
- [100] Nesterova, A. P., Klimov, E. A., Zharkova, M. et al., “Chapter 9 - Diseases of the respiratory system,” en *Disease Pathways*, A. P. Nesterova, E. A. Klimov, M. Zharkova et al., eds., Elsevier, 2020, págs. 391-442.
- [101] Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C. et al., “Mutational Processes Molding the Genomes of 21 Breast Cancers,” *Cell*, vol. 149, n.º 5, págs. 979-993, mayo de 2012.
- [102] Nordling, C. O., “A New Theory on the Cancer-inducing Mechanism,” *British Journal of Cancer*, vol. 7, n.º 1, págs. 68-72, mar. de 1953.
- [103] O’Connell, K. S., Hindley, G., Smeland, O. B. et al., “Chapter 16 - Shared heritability among psychiatric disorders and traits,” en *Psychiatric Genomics*, ép. Translational and Applied Genomics, E. E. Tsermpini, M. Alda y G. P. Patrinos, eds., Academic Press, 2022, págs. 341-360.
- [104] O’Connor, C., “Meiosis, genetic recombination, and sexual reproduction,” *Nature education*, vol. 1, n.º 1, pág. 174, 2008.

- [105] Ohta, T. y Kimura, M., “A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population,” *Genetical Research*, vol. 22, n.º 2, págs. 201-204, oct. de 1973.
- [106] ONU, *El cáncer mató a diez millones de personas en 2020, la mayoría en países de renta baja and media.* – ONU, <https://news.un.org/es/story/2021/02/1487492>, [Accesado en línea: 2022-06-07], 2021.
- [107] Paul, S. y Lal, G., “The Molecular Mechanism of Natural Killer Cells Function and Its Importance in Cancer Immunotherapy,” *Frontiers in Immunology*, vol. 8, sep. de 2017.
- [108] Pavel, M., Renna, M., Park, S. J. et al., “Contact inhibition controls cell survival and proliferation via YAP/TAZ-autophagy axis,” en *Nature Communications*, vol. 9, n.º 1, 27 de jul. de 2018.
- [109] Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. y Kraut, J., “Structures of Ternary Complexes of Rat DNA Polymerase beta;, a DNA Template-Primer, and ddCTP,” *Science*, vol. 264, n.º 5167, págs. 1891-1903, 1994.
- [110] Pianka, E., “Evolutionary Ecology: Overview,” en *Encyclopedia of Ecology*, S. E. Jørgensen y B. D. Fath, eds., Oxford: Academic Press, 2008, págs. 1481-1485.
- [111] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023.
- [112] Ramachandran, K. y Tsokos, C., *Mathematical Statistics with Applications*. Elsevier Science, 2009.
- [113] Rassool, F. V., “DNA double strand breaks (DSB) and non-homologous end joining (NHEJ) pathways in human leukemia,” *Cancer Letters*, vol. 193, n.º 1, págs. 1-9, abr. de 2003.
- [114] Rédei, G. P., *Encyclopedia of Genetics, Genomics, Proteomics, and Informatics* (Encyclopedia of Genetics, Genomics, Proteomics, and Informatics). Springer, 2008.
- [115] Resnick, S., *A Probability Path*. Nueva York, E.U.A.: Birkhäuser Boston, 1999.
- [116] Rohatgi, V.K. and Saleh, A.K.M.E., *An Introduction to Probability and Statistics* (Wiley Series in Probability and Statistics). John Wiley & Sons, Inc, 2015.
- [117] Rolfe, M. D., Rice, C. J., Lucchini, S. et al., “Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation,” *Journal of Bacteriology*, vol. 194, n.º 3, págs. 686-701, feb. de 2012.
- [118] Ross, S., *Introduction to Probability Models* (Introduction to Probability Models). Elsevier Science, 2007.
- [119] Royden, H. L., *Real analysis*, Third. Macmillan Publishing Company, New York, 1988.

- [120] Russell, B. C., Torralba, A., Murphy, K. P. y Freeman, W. T., “LabelMe: A Database and Web-Based Tool for Image Annotation,” *International Journal of Computer Vision*, vol. 77, n.º 1-3, págs. 157-173, oct. de 2007, http://labelme.csail.mit.edu/Release3.0/Images/static_web_tinyimagesdataset/e/exonuclease/; acceded 2022-10-14.
- [121] Sato, M., Kakui, Y. y Toya, M., “Tell the Difference Between Mitosis and Meiosis: Interplay Between Chromosomes, Cytoskeleton, and Cell Cycle Regulation,” *Frontiers in Cell and Developmental Biology*, vol. 9, abr. de 2021.
- [122] Scitable by Nature, *Mitosis. Learn Science at Scitable by Nature*, <https://www.nature.com/scitable/topicpage/mitosis-14046258/>, [Online, accessed 2022-10-26], 2014.
- [123] Scitable by Nature, *Replication and Distribution of DNA during Meiosis. Learn Science at Scitable by Nature*, <https://www.nature.com/scitable/topicpage/replication-and-distribution-of-dna-during-meiosis-6524853/>, [Online, accessed 2022-10-26], 2014.
- [124] Shammas, M. A., “Telomeres, lifestyle, cancer, and aging,” *Current Opinion in Clinical Nutrition and Metabolic Care*, vol. 14, n.º 1, págs. 28-34, ene. de 2011.
- [125] Sherwood, L. y Ward, C., *Human physiology: from cells to systems*. Cengage learning, 2018.
- [126] Sherwood, L. M., Parris, E. E. y Folkman, J., “Tumor Angiogenesis: Therapeutic Implications,” *New England Journal of Medicine*, vol. 285, n.º 21, págs. 1182-1186, nov. de 1971.
- [127] Shorak, G. R. y Wellner, J. A., *Empirical Processes with Application to Statistics*. Pensilvania, E.U.A.: Society for Industrial y Applied Mathematics, 2009.
- [128] Siddiqui, I. A., Sanna, V., Ahmad, N., Sechi, M. y Mukhtar, H., “Resveratrol nanoformulation for cancer prevention and therapy,” en, *Annals of the New York Academy of Sciences*, vol. 1348, n.º 1, págs. 20-31, 24 de jun. de 2015.
- [129] Siernicka, M., Firczuk, M. y Winiarska, M., *NK cells applications in immuno-oncology. Strategies towards Excellence in Immuno-Oncology Project*, <http://www.stream.wum.edu.pl/en/knowledge-base/96-nk-cells-applications-in-immuno-oncology>, [Accesado en línea: 2022-10-15], 2020.
- [130] Silva, N. S. D. y Klein, U., “Dynamics of B cells in germinal centres,” *Nature Reviews Immunology*, vol. 15, n.º 3, págs. 137-148, feb. de 2015.
- [131] Skala, S. L., Wang, X., Zhang, Y. et al., “Next-generation RNA Sequencing-based Biomarker Characterization of Chromophobe Renal Cell Carcinoma and Related Oncocytic Neoplasms,” *European Urology*, vol. 78, n.º 1, págs. 63-74, jul. de 2020.

- [132] Speiser, D. E., Ho, P. C. y Verdeil, G., “Regulatory circuits of T cell function in cancer,” *Nature Reviews Immunology*, vol. 16, n.º 10, págs. 599-611, ago. de 2016.
- [133] Spivak, M., *Calculus*, Second. Publish or Perish, 1996.
- [134] Sungur, C. M. y Murphy, W. J., “Positive and Negative Regulation by NK Cells in Cancer,” *Critical Reviews in Oncogenesis*, vol. 19, n.º 1-2, págs. 57-66, 2014.
- [135] Swinnen, I., “Predictive modelling of the microbial lag phase: a review,” *International Journal of Food Microbiology*, vol. 94, n.º 2, págs. 137-159, jul. de 2004.
- [136] Takahashi, Y., Demachi-Okamura, A., Oya, Y. et al., “Research advance in tumor specific antigens: a narrative review,” *AME Medical Journal*, vol. 6, n.º 0, 2020.
- [137] Taylor, H. y Karlin, S., *An Introduction to Stochastic Modeling*, 3rd. Elsevier Science, 1998.
- [138] Taylor, J. C., *An Introduction to Measure and Probability*. Springer-Verlag, New York, 1997.
- [139] Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G. y Vogelstein, B., “Only three driver gene mutations are required for the development of lung and colorectal cancers,” *Proceedings of the National Academy of Sciences*, vol. 112, n.º 1, págs. 118-123, 2015.
- [140] Tripathi, S. C., Vedpathak, D. y Ostrin, E. J., “The Functional and Mechanistic Roles of Immunoproteasome Subunits in Cancer,” *Cells*, vol. 10, n.º 12, pág. 3587, dic. de 2021.
- [141] Tucker, H. G., *A Graduate Course in Probability* (Probability and Mathematical Statistics, Vol. 2). Academic Press, Inc., New York-London, 1967.
- [142] Tuttle, A. R., Trahan, N. D. y Son, M. S., “Growth and Maintenance of *Escherichia coli* Laboratory Strains,” *Current Protocols*, vol. 1, n.º 1, ene. de 2021.
- [143] Tyson, J. J., “CHAPTER 3: Periodic Phenomena in Physarum,” en *Cell Biology of Physarum and Didymium*, H. E. Aldrich y J. W. Daniel, eds., Academic Press, 1982, págs. 61-110.
- [144] Uribe-Querol, E. y Rosales, C., “Phagocytosis: Our Current Understanding of a Universal Biological Process,” *Frontiers in Immunology*, vol. 11, jun. de 2020.
- [145] Visser, K. E. de, Korets, L. V. y Coussens, L. M., “De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent,” *Cancer Cell*, vol. 7, n.º 5, págs. 411-423, mayo de 2005.
- [146] Walsh, K. M. y R., W. M., “Epidemiology of Brain Tumors,” en *Youmans and Winn Neurological Surgery*, ép. Youmans Neurological Surgery, H. Winn, ed., Elsevier, 2017.

- [147] Wang, L., Fan, D., Chen, W. y Terentjev, E. M., “Bacterial growth, detachment and cell size control on polyethylene terephthalate surfaces,” *Scientific Reports*, vol. 5, n.º 1, oct. de 2015.
- [148] Winters-Miner, L. A., Bolding, P. S., Hilbe, J. M. et al., “Chapter 13 - Personalized Medicine,” en *Practical Predictive Analytics and Decisioning Systems for Medicine*, L. A. Winters-Miner, P. S. Bolding, J. M. Hilbe et al., eds., Academic Press, 2015, págs. 176-204.
- [149] Wojcicki, J., *Cord blood telomere length in Latino infants. Parents Guide Cord Blood Foundation*, <https://parentsguidecordblood.org/en/news/cord-blood-telomere-length-latino-infants>, [Accesado en línea: 2022-10-12], Parents Guide Cord Blood Foundation, ene. de 2016.
- [150] Wojczynski, M. K. y Tiwari, H. K., “Definition of Phenotype,” en *Genetic Dissection of Complex Traits*, ép. Advances in Genetics, vol. 60, Academic Press, 2008, págs. 75-105.
- [151] Yao, Y. y Dai, W., “Genomic Instability and Cancer,” *Journal of Carcinogenesis & Mutagenesis*, vol. 05, n.º 02, 2014.
- [152] Yuen, G. J., Demissie, E. y Pillai, S., “B Lymphocytes and Cancer: A Love–Hate Relationship,” *Trends in Cancer*, vol. 2, n.º 12, págs. 747-757, dic. de 2016.
- [153] Zhang, X. y Simon, R., “Estimating the number of rate limiting genomic changes for human breast cancer,” *Breast Cancer Research and Treatment*, vol. 91, n.º 2, págs. 121-124, mayo de 2005.
- [154] Zheng, Q., “Progress of a half century in the study of the Luria–Delbrück distribution,” *Mathematical Biosciences*, vol. 162, n.º 1-2, págs. 1-32, nov. de 1999.
- [155] Zheng, Q., “New algorithms for Luria-Delbrück fluctuation analysis,” *Mathematical Biosciences*, vol. 196, n.º 2, págs. 198-214, ago. de 2005.
- [156] Zheng, Q., “A new practical guide to the Luria–Delbrück protocol,” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 781, págs. 7-13, 2015.
- [157] Zheng, Q., “New approaches to mutation rate fold change in Luria–Delbrück fluctuation experiments,” *Mathematical Biosciences*, vol. 335, pág. 108 572, mayo de 2021.
- [158] Zhu, T., Hu, Y., Ma, Z.-M., Zhang, D.-X., Li, T. y Yang, Z., “Efficient simulation under a population genetics model of carcinogenesis,” *Bioinformatics*, vol. 27, n.º 6, págs. 837-843, ene. de 2011.

Apéndice A

Códigos en R

A.1. Modelo Moran

```
Moran <-function(population ,m0,m1,k){
  #Initial conditions; i.e. a sensitive wildtype (WT) cell with
  #mutation rate m0
  #First entry: 0 means sensitive to the antibiotic , 1 means
  #antibiotic-resistant cell produced by WT
  #Second entry: 0 means wildtype cell , 1 means mutator cell
  #Third entry: 0 means sensitive to the antibiotic , 1 means
  #resistant mutator
  m<-matrix(0,nrow=1,ncol=3)
  #Initial population: 1 cell
  p <-1
  while(p<population){
    #We randomly choose which cell is going to be a parent
    #(unbiased choice; i.e. every cell is equally likely to be
    #chosen)
    f <-sample(seq.int(1:p),1)
    #Parent has mutation rate m0 (as WT)
    if (m[f,2]==0){
      mu=m0
    }
    else{
      mu=m0*k
    }
  }
}
```

```

}
#Blank auxiliary matrix to save the daughter cells and
#their mutation rates
o ←matrix(0,2,3)
for(d in 1:2){
  #If the parent cell is sensitive to the antibiotic
  if (m[f,1]==0){
    r=runif(n=1, min=0, max=1)
    #Daughter of a WT becomes resistant
    if (r <= mu && mu==m0){
      o[d,1]=1
    }
    #The daughter of a mutator becomes resistant
    if (r <= mu && mu==k*m0){
      o[d,3]=1
    }
    #Otherwise, remains sensitive (both entries were zero
    #already)
  }
  #If the parent is resistant, its offspring it is as well
  if (m[f,1]==1){
    o[d,1]=1
  }
  #If the mutator parent is resistant, its offspring it is
  #as well
  if (m[f,3]==1){
    o[d,3]=1
  }
  #If the parent is not a mutator (which implies its
  #mutation rate is m0)
  if (m[f,2]==0){
    r=runif(n=1, min=0, max=1)
    #Cell becomes a mutator with probability m1, hence its
    #mutation rate increases from m0 to k*m0
    if (r<=m1){
      o[d,2]=1
    }
    #Otherwise, the mutation rate remains the same as the
    #parent
  }
}

```

```

    #If the parent has a mutation rate k*m0, its offspring
    #will have it as well
    if (m[f,2]==1){
      o[d,2]=1
    }
  }
  #Remove the parent
  m <-m[-f,]
  #We bind the original matrix and the matrix with the
  #information associated with the new two cells
  m <-rbind(m,o)
  #The population has increased by one cell
  p=p+1
}
colnames(m) <-c("Resistant", "Mutator", "Resistant_mutator")
return(m)
}

```

A.2. Modelo gráfico

Modulo I

```

graphicalmethod <-function(final_pop,t,mu){
  if(is.na(final_pop) || final_pop==0 || is.na(t) || t==0){
    result <-c(0,0)
  }
  else{
    #Sample of size one from a random variable with Poisson
    #distribution where lambda=final_pop*t*mu
    pop <-rpois(1,final_pop*t*mu)
    if(pop==0){
      result <-c(0,0)
    }
    else{
      #Sample of size pop from a uniform random variable on [0,1]
      y <-runif(pop)
      #Sample of size pop from a uniform random variable on [0,t]

```

```

x <-runif(pop,0,t)
#Bind the last two vectors
cells <-cbind(x,y)
#Create the graph of the function
#f(x)= (1/pop)exp(x*ln(pop)/t) on [0,t]
x_1 <-seq(0,t,0.001)
y_1 <-1/final_pop*exp(x_1*log(final_pop)/t)
#Determine the mutant cells
mutants <-subset(cells,
cells[,2] < 1/final_pop*exp(cells[,1]*log(final_pop)/t))
#if the mutants set is empty
if(length(mutants)==0){
  result <-c(0,0)
}
else{
  #Quantity of mutations
  mutations <-length(mutants[,1])
  #Display the mutant cells
  #mutants
  #Plot the cells, the graph of the discrimination
  #function, and the resulting mutants in red
  #Margins
  #par(mar = c(4.2, 4.2, 0.1, 0.1))

  #plot(cells, col = ifelse(cells[,2] <
  #1/final_pop*exp(cells[,1]*log(final_pop)/t), 'red', 'green'),
  #xlim=c(0,t), ylim=c(0,1), xlab='Time', ylab='Fraction of
  #final population', pch=19, lwd=0.75, cex.lab=1.5,
  #cex.axis=1.2, cex.main=0.1, cex.sub=0.1)
  #lines(x_1,y_1, col='black', lwd=2)
  #Time between the end of the experiment and the appearing
  #of the mutant cells
  time_dif <-t-mutants[,1]
  #Number of mutants raised by original mutant cells
  #Create an auxiliary vector with -1's
  mutants_offspring <-rep(-1,length(time_dif))
  #Number of mutants as a sample from a r.v. with negative
  #binomial distribution and parameters (1,
  #e^{-r(time_dif)}) plus one since the support of the
  #binomial distribution defined by R considers the zero,

```

```

    #but we are modeling a pure-birth process
    if(final_pop==1){
      result <-c(0,0)
    }
    else{
      for(i in 1:length(time_dif)){
        mutants_offspring[i] <-1+rnbinom(1,1,
          prob=exp(-log(final_pop)/t*time_dif[i]))
      }
      totalmutants <-sum(mutants_offspring)
      totalmutants
      #Original mutants and final raised mutants
      result <-c(mutations, totalmutants)
    }
  }
}
return(result)
}

```

Modulo II

```

graphmutators <-function(final_pop,t,mu1,mu2){
  #Sample of size one from a random variable with Poisson
  #distribution where lambda=final_pop*t*mu
  pop <-rpois(1,final_pop*t*mu1)
  if(pop==0){
    result <-c(0,0,0,0)
  }
  else{
    #Sample of size pop from a uniform random variable on [0,1]
    y <-runif(pop)
    #Sample of size pop from a uniform random variable on [0,t]
    x <-runif(pop,0,t)
    #Bind the last two vectors
    cells <-cbind(x,y)
    #Create the graph of the function f(x)=
    #(1/pop)exp(x*ln(pop)/t) on [0,t]
    x_1 <-seq(0,t,0.001)
  }
}

```

```

y_1 <-1/final_pop*exp(x_1*log(final_pop)/t)
#Determine the mutant cells
mutants <-subset(cells ,
cells[,2] < 1/final_pop*exp(cells[,1]*log(final_pop)/t))
#If the subset is empty
if(length(mutants)==0){
  result <-c(0,0,0,0)
}
else{
  #Quantity of mutants
  originalmutants <-length(mutants[,1])
  #Display the mutant cells
  #mutants
  #Margins
  #par(mar = c(4.2, 4.2, 0.1, 0.1))
  #Plot the cells , the graph of the discrimination function ,
  #and the resulting mutants in red
  #plot(cells , col = ifelse(cells[,2] <
  #1/final_pop*exp(cells[,1]*log(final_pop)/t), 'red', 'blue'),
  #xlim=c(0, t), ylim=c(0, 1), xlab='Time', ylab='Fraction of
  #final population', pch=19, lwd=0.75, cex.lab=1.5,
  #cex.axis=1.2, cex.main=0.1, cex.sub=0.1)
  #lines(x_1, y_1, col='black', lwd=3)
  #Time between the end of the experiment and the appearing
  #of the mutant cells
  time_dif <-t-mutants[,1]
  #Number of mutants raised by original mutant cells
  #Create an auxiliary vector with -1's
  mutants_offspring <-rep(-1,length(time_dif))
  #Number of mutants as a sample from a r.v. with negative
  #binomial distribution and parameters (1, e^{-r(time_dif)})
  #plus one since the support of the binomial distribution
  #defined by R considers the zero, but we are modelling a
  #pure-birth process
  #if final_pop==1 then log(final_pop)=0 and rnbino(1,1,
  #prob=exp(-log(final_pop)/t*time_dif[i]))=rnbino(1,1,0)=NA
  if(final_pop==1){
    result <-c(0,0)
  }
}

```

```

else{
  for(i in 1:length(time_dif)){
    mutants_offspring[i] <-1+rnbinom(1,1,
    prob=exp(-log(final_pop)/t*time_dif[i]))
  }
  totalmutants <-sum(mutants_offspring)
  totalmutants
  #Create an auxiliary vector with -1's for the mutator
  #mutants
  mutator_offspring <-rep(-1,length(time_dif))
  mutators <-rep(-1,length(time_dif))
  for(i in 1:length(time_dif)){
    a <-graphicalmethod(final_pop = mutants_offspring[i],
    t= time_dif[i], mu2)
    mutators[i] <-a[1]
    mutator_offspring[i] <-a[2]
  }
  total_mutator_mutants <-sum(mutator_offspring)
  total_mutators <-sum(mutators)
  result <-c(originalmutants, totalmutants, total_mutators,
  total_mutator_mutants)
}
}
}
return(result)
}

```

A.3. Modelo basado en el método de inversión

Modulo I

```

f1 <-function(Final_pop, final_time, mutation_rate){
if(is.na(Final_pop) || Final_pop==0 || is.na(final_time)
|| final_time==0){
  result <-c(0,0)
}
else{

```

```

#growth rate
#if Final_pop=1, then gr=0 and 1/gr is undefined
if(Final_pop==1){
  result <-c(0,0)
}
else{
  gr <- log(Final_pop)/final_time
  #Number of mutations
  mutations <-rpois(1,Final_pop*mutation_rate)
  if(mutations==0){
    result <-c(0,0)
  }
  else{
    #Sample of size 'mutations' from a uniform variable on
    #[0,1]

    uniform <-runif(mutations)
    #Time when mutants appeared
    t <-1/gr* log(uniform*(Final_pop-1)+1)
    #Time between the final time and the time when mutants
    #appeared
    time_diff <- final_time - t
    mutants_offspring <-rep(-1,mutations)
    for(i in 1:mutations){
      mutants_offspring[i] <-1+rnbinom(1,1,
        prob=exp(-gr*time_diff[i]))
      if(is.na(rnbinom(1,1,prob=exp(-gr*time_diff[i])))){
        print(time_diff[i])
      }
    }
  }
  totalmutants <-sum(mutants_offspring)
  totalmutants
  #Original mutants and final raised mutants
  result <-c(mutations, totalmutants)
}
}
return(result)
}

```

Modulo II

```

f2 <-function(Final_pop, final_time, mutation_rate,
mutation_rate_mutators){
  #r factor
  r <- log(Final_pop)/final_time
  #Number of mutations
  mutations <-rpois(1,Final_pop*mutation_rate)
  if(mutations==0){
    result <-c(0,0,0,0)
  }
  else{
    #Sample of size 'mutations' from a uniform variable on [0,1]
    uniform <-runif(mutations)
    #Time when mutants appeared
    t <-1/r* log(uniform*(Final_pop-1)+1)
    #Time between the final time and the time when mutants
    #appeared
    time_diff <- final_time - t
    #Number of mutants raised by each mutation
    #Create an auxiliary vector with -1's
    mutants <-rep(-1,length(time_diff))
    #Number of mutants as a sample from a r.v. with negative
    #binomial distribution and parameters (1,
    #exp(-r*time_diff[i]) )
    for(i in 1:mutations){
      #A pure-birth process
      mutants[i] <-1+rnbinom(1,1,prob=exp(-r*time_diff[i]))
    }
    totalmutants <-sum(mutants)
    totalmutants
    #Create an auxiliary vector with -1's for the mutator mutants

    mutator_offspring <-rep(-1,mutations)
    mutators <-rep(-1,mutations)
    for(i in 1:mutations){
      a <-f1(Final_pop = mutants[i], time_diff[i],
      mutation_rate_mutators)
      mutator_offspring[i] <-a[2]
    }
  }
}

```

```

    mutators[i] <-a[1]
  }
  total_mutator_mutants <-sum(mutator_offspring)
  total_mutators <-sum(mutators)
  #Original mutants, mutants at the end, mutators, mutator-
  #mutants
  result <-c(mutations, totalmutants, total_mutators,
    total_mutator_mutants)
}
return(result)
}

```

A.4. Estimación por ajuste directo

LSQ_Estimation=**function**(File, population_size, GC, dilution, discretization, WT2R_bounds, WT2M_bounds, M2R_bounds){

```

  #Reading the file
  td <- read.table(File, header = F, sep=" ")
  n_cycles <-GC
  nn1 <- population_size
  ni <- discretization
  lbm <- 10^WT2R_bounds[1]
  ubm <- 10^WT2R_bounds[2]
  lbm1 <- 10^WT2M_bounds[1]
  ubm1 <- 10^WT2M_bounds[2]
  lbmm <- 10^M2R_bounds[1]
  ubmm <- 10^M2R_bounds[2]
  ps1 <- dilution
  pb1 <- 2^(20)/(2^(33))

  h <- function(z) {
    (1 - z)^((1 - z) / z)
  }

  px <- function(z, Am, m1, Amm) {
    (h(z)^Am) * (h(h(z)^m1)^Amm)
  }

```

```

}

psb <- function(z, pb) {
  (1 - pb) + pb * z
}

ps <- function(z, p) {
  (1 - p) + p * z
}

pgb <- function(z, Am, m1, pb, Amm) {
  px(psb(z, pb), Am, m1, Amm)
}

pgs <- function(z, p, Am, m1, Amm) {
  px(ps(z, p), Am, m1, Amm)
}

#resample_size=50
#td=sample(td, resample_size, replace = T)
#head(td)

epgf_gc_j <-function(j, td, ps1){
  a=c()
  for (i in 1:nrow(td)){
    a=c(a, rbinom(n = 1, size = (td[, 2*j-1]+td[, 2*j])[i],
    prob = ps1))
    #a=sample(a, size=50, replace = T)
  }

  epgf <-function(x){
    return(sum(x^a)/nrow(td))
  }
  return(epgf)
}

zr <- seq(0.4, 0.9, by = 0.1)

```

```

mr <- (lbm * (ubm / lbm)^(0:ni / ni)) * nn1

m1r <- (lbm1 * (ubm1 / lbm1)^(0:ni / ni))

m1mr <- (lbmm * (ubmm / lbmm)^(0:ni / ni)) * nn1

for (i in 1:GC){
  aux=paste0('tls', i)
  assign(aux, c())
}
#tls1 <- c()
#tls2 <- c()
#tls3 <- c()

lls <- c()
tls <- c()

for (i in 1:GC){
  aux=paste0('epgf', i)
  assign(aux, epgf_gc_j(i, td, ps1))
}
#epgf1 <- epgf_gc_j(1)
#epgf2 <- epgf_gc_j(2)
#epgf3 <- epgf_gc_j(3)

for (i in 1:GC){
  aux=paste0('mne', i)
  assign(aux, 1000000)
}
#mne1 <- 1000000
#mne2 <- 1000000
#mne3 <- 1000000

for (mi in mr) {
  for (m1i in m1r) {

```

```

for (mmi in mmr) {
  #tp1=0
  #tp2=0
  #tp3=0
  for (i in 1:GC){
    aux=paste0('tp', i)
    assign(aux, 0)
  }
  for(z in zr){
    #pgs(z, p, Am, m1, Amm)           #Parameters to be used
    pg1 <- pgs(z, ps1, mi, mli, mmi)
    #pgb(z, Am, m1, pb, Amm)         #Parameters to be used
    #pg2 <- pgb(pg1, mi, m1i, pb1, mmi)
    #pg3 <- pgb(pg2, mi, m1i, pb1, mmi)
    if(GC>1){
      for (i in 2:GC){
        aux=paste0('pg', i)
        past=paste0('pg', i-1)
        past.var=get(past)
        assign(aux, pgb(past.var, mi, mli, pb1, mmi))
      }
    }
    #tp1=tp1+(pg1 - epgf1(z))^2
    #tp2=tp2+(pg2 - epgf2(z))^2
    #tp3=tp3+(pg3 - epgf3(z))^2
    for (i in 1:GC){
      theoretical=paste0('pg', i)
      theoretical.var=get(theoretical)
      empirical=paste0('epgf', i)
      empirical.var=get(empirical)
      aux=paste0('tp', i)
      assign(aux, get(aux) + (empirical.var(z) -
        theoretical.var)^2 )
    }
  }
  tpt1 <- tp1
  #tpt2 <- tp1 + tp2
  #tpt3 <- tp1 + tp2 + tp3
  if(GC>1){
    for (i in 2:GC){

```

```

    aux=paste0('tpt', i)
    tp_now=paste0('tp', i)
    tp_now.var=get(tp_now)
    past=paste0('tpt', i-1)
    past.var=get(past)
    assign(aux, past.var + tp_now.var)
  }
}
#if (tpt1 < mne1) {
# mne1 <- tpt1
# ls1 <- c(log(mi / nn1), log(m1i), log(mmi / nn1),
log(tpt1))
#}
#if (tpt2 < mne2) {
# mne2 <- tpt2
# ls2 <- c(log(mi / nn1), log(m1i), log(mmi / nn1),
#log(tpt2))
#}
#if (tpt3 < mne3) {
# mne3 <- tpt3
# ls3 <-c(log(mi / nn1), log(m1i), log(mmi / nn1),
#log(tpt3))
#}
for(i in 1:GC){
  tpt.name=paste0('tpt', i)
  tpt.var=get(tpt.name)
  bound.name=paste0('mne', i)
  bound.var=get(bound.name)
  if(tpt.var < bound.var){
    assign(bound.name, tpt.var)
    assign(paste0('ls', i), c(log(mi / nn1), log(m1i),
log(mmi / nn1), log(tpt.var)) )
  }
}

#tls <- cbind(tls, c(log(mi / nn1), log(m1i),
#log(mmi / nn1)),
#           c(log(tpt1), log(tpt2), log(tpt3))
#)
}

```

```

    }
  }

#tls1 <- cbind(tls1, ls1)
#tls2 <- cbind(tls2, ls2)
#tls3 <- cbind(tls3, ls3)
for(i in 1:GC){
  assign(paste0('tls',i), get(paste0('ls',i)))
}
#lls <- cbind(lls, tls)

aux=c(tls1[1:3])
if(GC>1){
  for(i in 2:GC){
    assign(paste0('aux'), cbind(aux, get(paste0('tls',i))[1:3]))
  }
}
#aux<-cbind(tls1[1:3,], tls2[1:3,], tls3[1:3,])
aux=t(exp(aux))
naming=c('Growth□cycle□1:')
if(GC>1){
  for(i in 2:GC){
    a=paste0('Growth□cycle□',i,':')
    naming=c(naming, a)
  }
}
row.names(aux)=naming
colnames(aux)=c('□WT□to□Mutant', '□WT□to□Mutator',
'□Mutator□to□Mutant')

return(aux)
}

```

A.5. Estimación por métodos de máxima verosimilitud en tres dimensiones

```

h <- function(z) {
  (1 - z)^((1 - z) / z)
}

px <- function(z, Am, m1, Amm) {
  (h(z)^{Am}) * (h(h(z)^m1)^{Amm})
}

psb <- function(z, pb) {
  (1 - pb) + pb * z
}

ps <- function(z, p) {
  (1 - p) + p * z
}

pgb <- function(z, Am, m1, pb, Amm) {
  px(psb(z, pb), Am, m1, Amm)
}

pgs <- function(z, p, Am, m1, Amm) {
  px(ps(z, p), Am, m1, Amm)
}

ML_Estimation=function(File, population_size, resampling,
  resample_size, GC, dilution, discretization,
  WT2R_bounds, WT2M_bounds, M2R_bounds, alpha){

  #Dilution process: Binomial sampling, with probability ps1, of
  #resistant cell in the j-th growth cycle
  binomial_sampling_GC_j <-function(td, j, ps1){
    a=c()
    for (i in 1:nrow(td)){
      a=c(a, rbinom(n = 1, size = (td[, 2*j-1]+td[, 2*j])[i],
        prob = ps1))
    }
    return(a)
  }

  #Recovery of the k-th probability state of the pmf

```

```

#WT2R Mutation rate from wildtype to resistant cells
#WT2M Mutation rate from wildtype to mutator cells
#pop Final population size
#N Total number of probability states
pmf_coeff <-function(k,WT2R,WT2M,M2R,pop,N,ps1){
  sum=0
  for(n in 0:(N-1)){
    #Auxiliary variable to transform the pgf into a #
    characteristic function
    aux_char=exp((2i*pi*n)/N)
    #Auxiliary variable to apply the FFT algorithm
    aux=(-1*2i*pi*n*k)/N
    #FFT
    sum=sum + pgs(aux_char,p=ps1,Am=10^{WT2R}*pop,m1=10^{WT2M},
    Amm=10^{M2R}*pop)*exp(aux)
  }
  #Real part
  return(Re(sum/N))
}

discretization_fun<-function(range, n){
  mu_min=range[1]
  mu_max=range[2]
  a=seq(0,n)
  result=mu_min*(mu_max/mu_min)^{a/n}
  return(result)
}

#log-likelihood function
ll_fun<-function(data,WT2R,WT2M,M2R,pop_size,N,ps1){
  sum(log(pmf_coeff(data+1,WT2R,WT2M,M2R,pop_size,N,ps1)))
}

#Reading the file
td <- read.table(File,header=T,sep=" ")
#Binomial sampling of resistant cells in the j-th growth cycle,
#j=1
data<-binomial_sampling_GC_j(td, j=GC,p=dilution)

if(resampling==T){

```

```

    data=sample(data, resample_size, replace = T)
  }

  #Maximum value in data
  N=max(data)
  N

  #Sample space discretization
  paramters_WT2R=WT2R_bounds
  WT2R=discretization_fun(paramters_WT2R, discretization)
  paramters_WT2M=WT2M_bounds
  WT2M=discretization_fun(paramters_WT2M, discretization)
  paramters_M2R=M2R_bounds
  M2R=discretization_fun(paramters_M2R, discretization)

  #Initial maximum
  max=ll_fun(data, WT2R = WT2R[1], WT2M = WT2M[1], M2R=M2R[1],
  population_size, N, dilution)

  powers=c()
  output=c()
  for(i in WT2R){
    for(j in WT2M){
      for(k in M2R){
        aux=ll_fun(data, WT2R = i, WT2M = j, M2R = k,
        population_size, N, dilution)
        output=rbind(output, c(10^{i}, 10^{j}, 10^{k}, aux))
        if(max<aux){
          max=aux
          powers=c(i, j, k)
        }
      }
    }
  }
}

#Maximum value
max
#Estimations on log_10 scale
names(powers)=c('Wildtype□to□Mutant', 'Wildtype□to□Mutator',

```

```

'Mutator to Mutant mutator')

df=as.data.frame(output)

df_confidence <-df[df[,4] >= max(qchisq(1-alpha,3)/2, ]

CI=matrix(0,3,2)
row.names(CI)=c('Wildtype to Mutant', 'Wildtype to Mutator',
'Mutator to Mutant mutator')
colnames(CI)=c('Lower bound', 'Upper bound')
CI[1,1]=min(df_confidence$V1)
CI[1,2]=max(df_confidence$V1)
CI[2,1]=min(df_confidence$V2)
CI[2,2]=max(df_confidence$V2)
CI[3,1]=min(df_confidence$V3)
CI[3,2]=max(df_confidence$V3)

out=list(10^{powers}, CI, output)
names(out)<-c('Estimations', 'ML_Confidence_Interval',
'Log_likelihood')
return(out)
}

```

A.6. Estimación por métodos de máxima verosimilitud en dos dimensiones

```

h <- function(z) {
  (1 - z)^((1 - z) / z)
}

px <- function(z,Am,m1,Amm) {
  (h(z)^{Am}) * (h(h(z)^m1)^{Amm})
}

psb <- function(z,pb) {

```

```

  (1 - pb) + pb * z
}

```

```

ps <- function(z,p) {
  (1 - p) + p * z
}

```

```

pgb <- function(z,Am,m1,pb,Amm) {
  px(psb(z,pb),Am,m1,Amm)
}

```

```

pgs <- function(z,p,Am,m1,Amm){
  px(ps(z,p),Am,m1,Amm)
}

```

```

ML_Bivariate=function(File, population_size, resampling,
                       resample_size, GC, dilution, discretization,
                       WT2R_bounds, WT2M_bounds, strength, alpha){
  #Dilution process: Binomial sampling, with probability ps1, of
  #resistant cell in the j-th growth cycle
  binomial_sampling_GC_j <-function(td,j,ps1){
    a=c()
    for (i in 1:nrow(td)){
      a=c(a,rbinom(n = 1, size = (td[, 2*j-1]+td[, 2*j])[i],
                  prob = ps1))
    }
    return(a)
  }
}

```

```

#Recovery of the k-th probability state of the pmf
#WT2R Mutation rate from wildtype to resistant cells
#WT2M Mutation rate from wildtype to mutator cells
#pop Final population size
#N Total number of probability states
pmf_coeff <-function(k,WT2R,WT2M,M2R,pop,N,ps1,strength){
  sum=0
  for(n in 0:(N-1)){
    #Auxiliary variable to transform the pgf into a
    #characteristic function

```

```

    aux_char=exp((2i*pi*n)/N)
    #Auxiliary variable to apply the FFT algorithm
    aux=(-1*2i*pi*n*k)/N
    #FFT
    sum=sum + pgs(aux_char ,p=ps1 ,Am=10^{WT2R}*pop ,m1=10^{WT2M} ,
    Amm=strength*10^{WT2R}*pop)*exp(aux)
  }
  #Real part
  return(Re(sum/N))
}

discretization_fun <-function(range, n){
  mu_min=range[1]
  mu_max=range[2]
  a=seq(0,n)
  result=mu_min*(mu_max/mu_min)^{a/n}
  return(result)
}

#log-likelihood function
ll_fun <-function(data, WT2R, WT2M, M2R, pop_size, N, ps1, strength){
  sum(log(pmf_coeff(data+1, WT2R, WT2M, M2R, pop_size, N, ps1,
  strength)))
}
#Reading the file
td <- read.table(File, header = F, sep=" ")
#Binomial sampling of resistant cells in the j-th growth cycle,
#j=1
data <-binomial_sampling_GC_j(td, j=GC, p=dilution)

if(resampling==T){
  data=sample(data, resample_size, replace = T)
}
#Maximum value in data
N=max(data)
N

#Sample space discretization
paramters_WT2R=WT2R_bounds

```

```

WT2R=discretization_fun(paramters_WT2R, discretization)
paramters_WT2M=WT2M_bounds
WT2M=discretization_fun(paramters_WT2M, discretization)

#Initial maximum
max=ll_fun(data, WT2R = WT2R[1], WT2M = WT2M[1], M2R=WT2R[1],
population_size, N, dilution, strength)

powers=c()
output=c()
for(i in WT2R){
  for(j in WT2M){
    aux=ll_fun(data, WT2R = i, WT2M = j, M2R = i,
population_size, N, dilution, strength)
    output=rbind(output, c(10^{i}, 10^{j}, strength*10^{i},
aux))
    if(max<aux){
      max=aux
      powers=c(10^{i}, 10^{j}, strength*10^{i})
    }
  }
}

#Maximum value
max

names(powers)=c('Wildtype□to□Mutant', 'Wildtype□to□Mutator',
'Mutator□to□Mutant□mutator')
df=as.data.frame(output)

df_confidence<-df[df[,4] >= max-qchisq(1-alpha,2)/2, ]

CI=matrix(0,3,2)
row.names(CI)=c('Wildtype□to□Mutant', 'Wildtype□to□Mutator',
'Mutator□to□Mutant□mutator')
colnames(CI)=c('Lower□bound', 'Upper□bound')
CI[1,1]=min(df_confidence$V1)
CI[1,2]=max(df_confidence$V1)
CI[2,1]=min(df_confidence$V2)

```