



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

DISTRIBUCIONES TIPO FASE PARA UN MODELO
DE COALESCENCIA CON TRES ALELOS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN MATEMÁTICAS APLICADAS

P R E S E N T A :

ELÍAS GONZÁLEZ NIETO

TUTORA

DRA. MARÍA CLARA FITTIPALDI



CIUDAD UNIVERSITARIA, CDMX, 2025

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Persona

Agradecimientos

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Índice general

Agradecimientos	v
1. Motivación	1
2. Coalescencia	3
2.1. Cadenas de Markov a tiempo continuo (CMTC)	3
2.2. Modelo de Wright-Fisher	14
2.3. Modelo de Morán	16
2.4. Coalescente de Kingman	18
2.5. Λ -coalescente	20
3. Distribuciones Tipo Fase	23
3.1. Distribuciones Tipo Fase Discretas	23
3.2. Definición, Densidad y Distribución	31
3.3. Momentos	39
3.4. Convoluciones y Mezclas	49
3.5. Transformaciones Via Recompensas	52
3.6. Distribuciones Tipo Fase Multivariadas	57
3.6.1. Covarianzas entre las variables	58
4. Tiempos de Extinción en un Modelo de Tres Alelos	63
4.1. Modelo de Wright-Fisher con tres alelos	64
4.2. Análisis del tiempo de extinción por alelo	64
4.3. Coincidencias con el coalescente de Kingman	64
Apéndice A. Densidad de las Tipo Fase	65
Bibliografía	72

1 Motivación

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

2 Coalescencia

Existen modelos clásicos en teoría de genética de poblaciones, como aquellos de ramificación, en donde se estudia el desarrollo de las especies desde el pasado hacia el presente. Sin embargo, el modelo de coalescencia propone una manera diferente de estudiar este proceso: considera el presente y se mueve hacia atrás en el tiempo.

Al considerar los momentos en que dos genealogías tienen ancestros en común, los modelos de coalescencia no se interesan en describir a la descendencia de cierta generación, sino en encontrar al **ancestro común más reciente** (MRCA por sus siglas en inglés), al cual entenderemos como el último linaje después de que los demás hayan coalescido.

Según [15], *El patrón de polimorfismo, es decir, los estados alélicos de todas las copias de genes homólogos en una población, está determinado por la historia genealógica y mutacional de estas copias.* Además, este tipo de modelos supone que después de ver la historia de los genes, podemos imponer mutaciones neutrales y éstas serán buenas representaciones de la realidad. Estos dos son fuertes supuestos en los modelos de coalescencia.

En este capítulo se sientan las bases para comprender al coalescente de Kingman y los dos principales modelos que llevan a este: Wright-Fisher y Moran. A manera de introducción, se presentan resultados sobre teoría de cadenas de Markov a tiempo continuo, las cuales nos ayudarán para abstraer los modelos coalescentes. La primera sección está basada en [7] y [5], mientras que los modelos coalescentes se apoyan en [15] y [6].

2.1. Cadenas de Markov a tiempo continuo (CMTC)

En esta sección se estudia un tipo de proceso estocástico denominado Cadena de Markov a tiempo continuo. Al igual que su análogo en tiempo discreto, se caracteriza porque la evolución futura del proceso depende únicamente del estado presente y no de la trayectoria pasada. Este tipo de procesos tienen un papel fundamental en áreas como la teoría de colas, la biología matemática y la genética de poblaciones.

En lo siguiente nos referiremos a las cadenas de Markov a tiempo continuo como

cadenas de Markov o simplemente cadenas.

Definición 2.1. Decimos que un proceso estocástico $(X_t : t \geq 0)$ con espacio de estados a lo más numerable \mathcal{S} es una **cadena de Markov a tiempo continuo** si cumple que para cualesquiera $0 \leq s_0 < s_1 < \dots < s_n < s$, $t \geq 0$ y $j, i, i_0, i_1, \dots, i_n \in \mathcal{S}$, se tiene que

$$\mathbb{P}[X_{t+s} = j | X_s = i, X_{s_n} = i_n, X_{s_{n-1}} = i_{n-1}, \dots, X_{s_0} = i_0] = \mathbb{P}[X_{t+s} = j | X_s = i]$$

Siempre que estas probabilidades estén bien definidas. Si además se cumple que $\mathbb{P}[X_{t+s} = j | X_s = i] = \mathbb{P}[X_t = j | X_0 = i]$ se dice que la cadena es homogénea. En dicho caso, escribimos

$$p_t(i, j) := \mathbb{P}[X_{t+s} = j | X_s = i] = \mathbb{P}[X_t = j | X_0 = i]$$

para referirnos a la probabilidad de que el proceso tome el valor j , partiendo de i después de t unidades de tiempo.

Ejemplo 2.2. Consideremos un proceso Poisson $(N_t : t \geq 0)$ de tasa $\lambda > 0$. Veamos que constituye un ejemplo de una Cadena de Markov a tiempo continuo, pues dados $0 \leq t_0 < t_1 < \dots < t_n < t_{n+1}$ y $i_0, i_1, \dots, i_n, j, i \in \mathbb{N}$, al usar que los incrementos son estacionarios e independientes se sigue que:

$$\begin{aligned} \mathbb{P}[N_{t_{n+1}} = j | N_{t_n} = i, \dots, N_{t_0} = i_0] &= \frac{\mathbb{P}[N_{t_{n+1}} = j, N_{t_n} = i, \dots, N_{t_1} = i_1, N_{t_0} = i_0]}{\mathbb{P}[N_{t_n} = i, \dots, N_{t_1} = i_1, N_{t_0} = i_0]} \\ &= \frac{\mathbb{P}[N_{t_{n+1}} - N_{t_n} = j - i, N_{t_n} - N_{t_{n-1}} = i - i_{n-1}, \dots, N_{t_1} - N_{t_0} = i_1 - i_0, N_{t_0} = i_0]}{\mathbb{P}[N_{t_n} - N_{t_{n-1}} = i - i_{n-1}, \dots, N_{t_1} - N_{t_0} = i_1 - i_0, N_{t_0} = i_0]} \\ &= \mathbb{P}[N_{t_{n+1}} - N_{t_n} = j - i] \frac{\mathbb{P}[N_{t_n} - N_{t_{n-1}} = i - i_{n-1}, \dots, N_{t_1} - N_{t_0} = i_1 - i_0, N_{t_0} = i_0]}{\mathbb{P}[N_{t_n} - N_{t_{n-1}} = i - i_{n-1}, \dots, N_{t_1} - N_{t_0} = i_1 - i_0, N_{t_0} = i_0]} \\ &= \mathbb{P}[N_{t_{n+1}-t_n} = j - i] \end{aligned}$$

donde $N_{t_{n+1}-t_n} \sim Poi(\lambda(t_{n+1} - t_n))$, y por un procedimiento análogo, se obtiene que

$$\mathbb{P}[N_{t_{n+1}} = j | N_{t_n} = i] = \mathbb{P}[N_{t_{n+1}-t_n} = j - i]$$

Por tanto, el proceso Poisson es una Cadena de Markov a tiempo continuo.

Dado que en una cadena de Markov discreta y homogénea los pasos están bien definidos, se estudia la dinámica del proceso por medio de su matriz de transición P , la cual está definida como $P_{ij} = p(i, j)$. Con tiempo continuo también estudiamos esta probabilidad y tenemos resultados correspondientes, como las ecuaciones de Chapman-Kolmogorov:

Proposición 2.3. Para cualesquiera $i, j \in \mathcal{S}$ y $s, t \geq 0$, se tiene que

$$p_{s+t}(i, j) = \sum_{k \in \mathcal{S}} p_s(i, k) p_t(k, j)$$

La demostración de esta proposición se omite, ya que es análoga al caso discreto y se puede consultar en [5].

Sin embargo, para cadenas a tiempo continuo esta matriz no captura toda la información, ya que no tenemos definido un tamaño de paso fijo. Esto no lleva a tasas de salto, las cuales describen con qué rapidez cambia el proceso de un estado a otro. Para eso, introducimos la siguiente definición.

Definición 2.4. *Dados $i, j \in \mathcal{S}$ distintos, la tasa de salto de i a j está dada por*

$$q(i, j) := \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h}$$

Siempre que este límite exista.

Entendemos a las tasas de salto como la probabilidad infinitesimal de transición del estado i al j , pues como

$$p_0(i, j) = \delta_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

entonces si $i \neq j$ se tiene que

$$q(i, j) = \left. \frac{\partial p_t(i, j)}{\partial t} \right|_{t=0} = \lim_{h \rightarrow 0} \frac{p_h(i, j) - p_0(i, j)}{h} = \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h}$$

que coincide con la definición. Notemos que en particular si $i \in \mathcal{S}$, tenemos que:

$$q(i, i) = \lim_{h \rightarrow 0} \frac{p_h(i, i) - p_0(i, i)}{h} = \lim_{h \rightarrow 0} \frac{\left(1 - \sum_{j \neq i} p_h(i, j)\right) - 1}{h} = - \sum_{j \neq i} \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h}$$

Donde en la segunda igualdad usamos el hecho de que

$$p_h(i, i) = 1 - \mathbb{P}[\text{No ir de } i \text{ a } i \text{ en } h \text{ unidades de tiempo}] = 1 - \sum_{j \neq i} p_h(i, j)$$

Por lo tanto,

$$q(i, i) = - \sum_{j \neq i} q(i, j)$$

Llamaremos tasa de salida del estado i a $q_i = -q(i, i) = \sum_{j \neq i} q(i, j)$, que es la tasa con la cual es proceso abandona el estado i . Calculemos las tasas de salto del Ejemplo 2.2, primero notemos que dados $i \in \mathbb{N}$ y $h, t > 0$, tenemos que

$$p_h(i, i+1) = \mathbb{P}[N_{t+h} = i+1 | N_t = i] = \frac{\mathbb{P}[N_{t+h} - N_t = i+1 - i, N_t = i]}{\mathbb{P}[N_t = i]}$$

Y usando la independencia y estacionaridad de los incrementos, esto pasa a ser

$$\mathbb{P}[N_h = 1] \frac{\mathbb{P}[N_t = i]}{\mathbb{P}[N_t = i]} = \mathbb{P}[N_h = 1] = e^{-\lambda h} \lambda h$$

Ya que $N_h \sim Poi(\lambda h)$. De manera completamente análoga, podemos probar que si $k > 1$, entonces

$$p_h(i, i+k) = e^{-\lambda h} \frac{(\lambda h)^k}{k!}$$

Finalmente, dado que el proceso Poisson es no decreciente, tenemos que $p_h(i+k, i) = 0$, ya que no podemos regresar a un estado menor. Al calcular las tasas de salto, obtenemos que para $k > 1$:

$$\begin{aligned} q(i, i+1) &= \lim_{h \rightarrow 0} \frac{p_h(i, i+1)}{h} = \lim_{h \rightarrow 0} \frac{e^{-\lambda h} \lambda h}{h} = \lambda, \\ q(i, i+k) &= \lim_{h \rightarrow 0} \frac{p_h(i, i+k)}{h} = \lim_{h \rightarrow 0} \frac{e^{-\lambda h} (\lambda h)^k}{k! h} = 0 \\ q(i+k, i) &= \lim_{h \rightarrow 0} \frac{p_h(i+k, i)}{h} = \lim_{h \rightarrow 0} \frac{0}{h} = 0 \end{aligned}$$

Que coindice con lo que sabemos sobre el proceso Poisson, pues pasamos del estado i al $i+1$ con tasa λ , y a cualquier otro con tasa 0. Por otro lado, como $q(i, i) = -\sum_{j \neq i} q(i, j)$, entonces para este ejemplo $q_i = \lambda$.

En los modelos poblacionales, lo más común es describir las tasas de salto de la cadena y analizar su comportamiento, pues esto permite identificar con qué rapidez se modifican ciertas poblaciones o distribuciones de alelos. Notando que para cada par de estados $i \neq j$, $q(i, j) \geq 0$ y como $q(i, i) = -\sum_{j \neq i} q(i, j)$, entonces $\sum_{k \in \mathcal{S}} q(i, k) = 0$, podemos dar la siguiente definición, que será de mucha utilidad en el futuro.

Definición 2.5. Una matriz Q de tamaño $|\mathcal{S}| \times |\mathcal{S}|$ es la **matriz de tasas o generador infinitesimal** para la cadena si cumple que $Q_{ij} = q(i, j)$ para cada par de estados i, j .

Suponiendo que $|\mathcal{S}| = N$, la matriz se ve de esta forma

$$Q = \begin{pmatrix} -q_1 & q(1, 2) & \cdots & q(1, N) \\ q(2, 1) & -q_2 & \cdots & q(2, N) \\ \vdots & \vdots & \ddots & \vdots \\ q(N, 1) & q(N, 2) & \cdots & -q_N \end{pmatrix}$$

Este concepto puede extenderse a matrices infinitas (numerable) de manera natural, de la misma manera que hacíamos para las matrices de transición en la teoría de cadenas discretas. Siguiendo con el Ejemplo 2.2, la matriz de tasas del proceso Poisson es:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ 0 & -\lambda & \lambda & \dots \\ 0 & 0 & -\lambda & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Esta matriz cumple que sus entradas son no negativas y que sus filas suman 0. A la entrada Q_{ij} la entenderemos como la tasa de ir de i a j , y en su diagonal están las tasas de salida de cada estado.

Esqueleto Discreto

Habiendo estudiado las tasas de una cadena de Markov a tiempo continuo, es natural preguntarse cómo construir la cadena una vez que tenemos su matriz de tasas. En esta sección presentamos el esqueleto discreto de una cadena a tiempo continuo, la cual será un proceso de Markov a tiempo discreto donde cada paso será un cambio de estado en el proceso continuo.

Con este objetivo en mente, notamos que si un estado i cumple que $q_i = 0$, entonces

$$q(i, i) = \frac{\partial p_h(i, i)}{\partial h} = 0$$

Lo cual significa $p_h(i, i)$ es una función que no depende de h y como $p_0(i, i) = 1$, entonces una vez que la cadena llega a i , no deja este estado jamás¹. En términos de probabilidades, significa que dada $t \geq 0$,

$$\mathbb{P}[X_{t+s} = i | X_t = i] = 1 \quad \text{para todo } s \geq 0 \tag{2.1}$$

En particular, esto ocurre si la fila i de la matriz de tasa es solo de ceros, pues $q_i = 0$. Este concepto será fundamental para estudiar modelos generacionales que busquen el ancestro común más reciente.

Por otro lado, si $q_i = \infty$ en el sentido de límites, la cadena deja el estado i instantáneamente. En lo que sigue, supondremos que $0 < q_i < \infty$ y faremos la construcción del esqueleto discreto, definido como la cadena en tiempo discreto $(Y_n : n \in \mathbb{N})$ que registra los estados visitados por la cadena original, de acuerdo con el procedimiento presentado en [7].

Para empezar, dado que la distribución exponencial es la única que tiene la propiedad de pérdida de la memoria, y las cadenas de Markov cumplen que su futuro no depende de su pasado, sino solamente de su presente, esta es la única distribución que podemos usar para modelar los tiempos entre cada salto del proceso continuo.

¹ i es un estado absorbente, como en la teoría de cadenas discretas

Supongamos que en el tiempo t_0 , el proceso es tal que $X_{t_0} = i$. Luego, con una tasa de $q(i, k)$, el proceso cambia a otro estado k , es decir que lo hace en un tiempo $T \sim \exp(q(i, j))$. Sin embargo, dado que puede saltar a cualquier otro estado, consideraremos una colección de variables aleatorias $\tau_{ik} \sim \exp(q(i, k))$ independientes. Podemos pensar estas variables como relojes que son activados al mismo tiempo y que no sonarán sino hasta un tiempo aleatorio y que depende del estado. Así, del estado i se elegirá otro estado, el cual corresponderá al primer reloj en sonar. Esto es,

$$\tau_i := \min_{k \neq i} \tau_{i,k}$$

que corresponde al tiempo en que el proceso cambia del estado i . Sabemos que $\tau_i \sim \exp(q_i)$, pues $q_i = \sum_{i \neq j} q(i, j)$ y el mínimo de una colección independiente de variables aleatorias exponenciales es exponencial y su parámetro es la suma de los parámetros individuales.

Además, para conocer las probabilidades de transición del esqueleto discreto, notamos que si $j \in \mathcal{S}$ es un estado fijo, entonces la probabilidad de que la cadena tome ese valor es:

$$\begin{aligned} \mathbb{P}[\tau_i = \tau_{ij}] &= \mathbb{P}\left[\tau_{i,j} \leq \min_{k \neq i,j} \tau_{i,k}\right] = \int_0^\infty \mathbb{P}\left[z \leq \min_{k \neq i,j} \tau_{i,k}\right] f_{\tau_{i,j}}(z) dz \\ &= \int_0^\infty e^{-(q_i - q(i,j))z} q(i, j) e^{-q(i,j)z} dz = \frac{q(i, j)}{q_i} \int_0^\infty q_i e^{-q_i z} dz \\ &= \frac{q(i, j)}{q_i} \end{aligned}$$

Donde, además de usar las propiedades de las distribuciones condicionales y exponenciales, se usó que

$$\tau_i = \tau_{ij} \iff \tau_{ij} \leq \min_{k \neq i,j} \tau_{i,k}$$

Pues si $\min_{k \neq i} \tau_{i,k} = \tau_{ij}$, en particular τ_{ij} cumple con ser menor o igual que cualquier otro de los tiempos para acceder a otro estado. Así, usando que

$$\sum_{k \neq i,j} q(i, k) = \sum_{k \neq i} q(i, k) - q(i, j) = q_i - q(i, j)$$

se sigue que

$$\min_{k \neq i,j} \tau_{i,k} \sim \exp(q_i - q(i, j))$$

Esto nos indica que las probabilidades de transición del esqueleto discreto estarán dadas como sigue

$$\mathbb{P}[Y_1 = j | Y_0 = i] = \frac{q(i, j)}{q_i}$$

De este análisis obtuvimos un resultado que será de suma importancia en el desarrollo de este texto, y es que si la cadena se encuentra en el estado i , lo dejará en un

tiempo aleatorio $\tau_i \sim \exp(q_i)$. Esta es una de las características fundamentales de las cadenas de Markov a tiempo continuo y que las diferencia de su análogo discreto: el tamaño del paso es aleatorio. En particular son tiempos con distribución exponencial y cuyo parámetro es la tasa de salida.

Ecuaciones Forward y Backward de Kolmogorov

Aunque ya podemos construir una cadena discreta que resuma el comportamiento del proceso, aún no tenemos forma de encontrar las probabilidades de transición dada una matriz de tasas Q . A continuación usaremos fuertemente las ecuaciones de Chapman-Kolmogorov con ese fin. Notemos que dados $i, j \in \mathcal{S}$ y $t, h \geq 0$,

$$\begin{aligned} p_{t+h}(i, j) &= \sum_{k \neq i} p_h(i, k)p_t(k, j) + p_h(i, i)p_t(i, j) \\ &= \sum_{k \neq i} p_h(i, k)p_t(k, j) + (1 - \sum_{k \neq i} p_h(i, k))p_t(i, j) \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \frac{\partial}{\partial t} p_t(i, j) &= \lim_{h \rightarrow 0} \frac{p_{t+h}(i, j) - p_t(i, j)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\left[\sum_{k \neq i} p_h(i, k)p_t(k, j) + (1 - \sum_{k \neq i} p_h(i, k))p_t(i, j) \right] - p_t(i, j)}{h} \end{aligned}$$

Y notando que en el numerador se anulan algunos términos, esto pasa a ser

$$\begin{aligned} &= \lim_{h \rightarrow 0} \left[\sum_{k \neq i} \frac{p_h(i, k)}{h} p_t(k, j) - \sum_{k \neq i} \frac{p_h(i, k)}{h} p_t(i, j) \right] \\ &= \sum_{k \neq i} p_t(k, j) \lim_{h \rightarrow 0} \frac{p_h(i, k)}{h} - p_t(i, j) \lim_{h \rightarrow 0} \sum_{k \neq i} \frac{p_h(i, k)}{h} \\ &= \sum_{k \neq i} p_t(k, j) q(i, k) - p_t(i, j) \sum_{k \neq i} q(i, k) \\ &= \sum_{k \neq i} p_t(k, j) q(i, k) - q_i p_t(i, j) \end{aligned}$$

De hecho, hemos demostrado que

$$\frac{\partial}{\partial t} p_t(i, j) = \sum_{k \neq i} p_t(k, j) q(i, k) - q_i p_t(i, j) = \sum_{k \in \mathcal{S}} q(i, k) p_t(k, j)$$

Si denotamos por $P(t)$ a la matriz con probabilidades de transición en un tiempo t , esto quiere decir que:

$$P'(t) = QP(t), \quad (2.2)$$

la cual es conocida como la **Ecuación Backward de Kolmogorov**. Por un procedimiento completamente análogo obtenemos la **Ecuación Forward de Kolmogorov**,

$$P'(t) = P(t)Q \quad (2.3)$$

Como consecuencia tenemos que

$$QP(t) = P(t)Q \quad (2.4)$$

Para resolver las ecuaciones diferenciales 2.2 y 2.3, tenemos el siguiente resultado.

Teorema 2.6. *La matriz e^{tQ} definida como*

$$e^{tQ} := \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!} \quad \text{para cada } t \geq 0$$

es solución a las ecuaciones Forward y Backward de Kolmogorov.

Demostración. Sea $t \geq 0$. Veamos que e^{tQ} es solución de la ecuación 2.2

$$\begin{aligned} \frac{\partial}{\partial t} e^{tQ} &= \frac{\partial}{\partial t} \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n = \sum_{n=0}^{\infty} \frac{\partial}{\partial t} \frac{t^n}{n!} Q^n = \sum_{n=0}^{\infty} \frac{n t^{n-1}}{n!} Q^n \\ &= \sum_{n=0}^{\infty} \frac{t^{n-1}}{(n-1)!} Q^n = Q \sum_{n=0}^{\infty} \frac{t^{n-1}}{(n-1)!} Q^{n-1} = Q e^{tQ} \end{aligned}$$

Y dado que las matrices $P(t)$ y Q conmutan, también es solución de 2.3. \square

Por tanto, para calcular las probabilidades de transición de la cadena de Markov a tiempo continuo dada la matriz de tasas Q , basta con calcular $P(t) = e^{tQ}$.

Ejemplo 2.7. Hagamos un análisis de la cadena de Markov a tiempo continuo ($X_t : t \geq 0$) con espacio de estados $\{0, 1\}$ donde $X_0 = 0$ y permanece en él un tiempo $\exp(\lambda)$, mientras que permanece en el estado 1 un tiempo $\exp(\mu)$. Claramente, la matriz de tasas del proceso es

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

En la Figura 2.1 podemos ver una trayectoria con valores específicos para λ y μ . Como se puede ver, nos muestra que el proceso permanece más tiempo en el estado 1 debido a que su tasa de salida es menor que la de 0. Por otro lado, veamos cuáles son las probabilidades de transición usando el resultado obtenido en 2.6 y el desarrollo llevado a cabo en [19]:

$$P(t) = e^{tQ} = \begin{pmatrix} \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} & \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \\ \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} & \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} \end{pmatrix}.$$

Por tanto, tenemos que las probabilidades de transición de la cadena son:

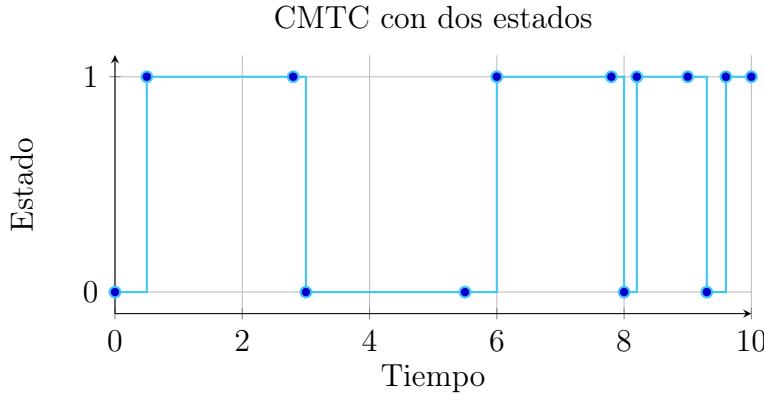


Figura 2.1: Trayectoria de una CMTC de dos estados con $\lambda = 2$ y $\mu = \frac{1}{2}$

$$\begin{aligned}
 p_t(0, 0) &= \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \\
 p_t(0, 1) &= \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \\
 p_t(1, 0) &= \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}, \\
 p_t(1, 1) &= \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}.
 \end{aligned}$$

Las cuales son sencillas de computar una vez que tenemos valores para λ y μ .

A continuación estudiaremos algunas de las propiedades de cadenas de Markov a tiempo continuo que son análogas a la teoría discreta. Esta sección se apoya en [5] y [14], que es donde se pueden consultar las pruebas de los resultados aquí expuestos.

Comunicación

Sea $(X_t : t \geq 0)$ una cadena de Markov a tiempo continuo con espacio de estados a lo más numerable \mathcal{S} . A continuación presentamos conceptos y resultados sobre la comunicación de estados, que son análogos a la teoría discreta y siguen siendo ciertos en tiempo continuo.

Definición 2.8. *Dados $i, j \in \mathcal{S}$, decimos que j es **accesible** con i si hay $t \geq 0$ tal que $p_t(i, j) > 0$, en este caso escribimos $j \rightarrow i$. Si ocurre que $j \rightarrow i$ y $i \rightarrow j$, entonces decimos que j e i se **comunican** y escribimos $j \leftrightarrow i$.*

Proposición 2.9. \leftrightarrow es una relación de equivalencia sobre \mathcal{S} .

A cada elemento de la partición inducida la llamaremos clase de comunicación, y si hay alguna propiedad que cumplan todos los miembros de una alguna clase, la llamaremos propiedad de clase.

Dado que en este trabajo estaremos interesados en estudiar cadenas a tiempo continuo con espacio de estados a lo más numberable, seguiremos diciendo que éstas son **irreducibles** cuando todos los estados se comuniquen entre ellos. Es decir que las cadenas irreducibles son aquellas que solo tienen una clase de comunicación.

Recurrencia y transitoriedad

Discutir la recurrencia y transitoriedad de los estados será fundamental para este trabajo, pues las cadenas que buscan el ancestro común más reciente serán absorbentes. En general, consideraremos cadenas de Markov a tiempo continuo con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1\}$ donde $1, 2, \dots, k$ son transientes y $k+1$ absorbente. Sin embargo, aunque intuitivamente el concepto es claro, en esta sección se introducen las definiciones y resultados, los cuales resultan ser análogos a la teoría discreta.

Sea $(X_t : t \geq 0)$ una cadena de Markov a tiempo continuo con espacio de estados a lo más numerable \mathcal{S} . Al considerar $i, j \in \mathcal{S}$, el tiempo que pasa desde que el proceso parte de i hasta llegar a j es aleatorio y depende de cada trayectoria. Por lo tanto, podemos definir a la variable aleatoria T_{ij} como

$$T_{ij} := \inf \{t \geq 0 : X_t = j \text{ con } X_0 = i\}.$$

donde convenimos que $\inf(\emptyset) = \infty$. En particular, se define a

$$T_j := \inf \{t > 0 : X_t = j \text{ y } X_s \neq j \text{ para todo } s < t, \text{ con } X_0 = j\}$$

Con estas variables podemos introducir las siguientes definiciones.

Definición 2.10. *Sea $j \in \mathcal{S}$. Decimos que:*

1. *j es transiente si $\mathbb{P}[T_j = \infty] > 0$.*
2. *j es recurrente si no es transiente, es decir si $\mathbb{P}[T_j = \infty] = 0$. En este caso:*
 - a) *j es recurrente nulo si $\mathbb{E}[T_j] = \infty$.*
 - b) *j es recurrente positivo si $\mathbb{E}[T_j] < \infty$.*

Tal como se anunciaba en 2.1, entenderemos a los estados absorbentes de la misma manera que en las cadenas discretas: un estado es recurrente si podemos regresar a él con probabilidad positiva, mientras que es transiente si pasa lo contrario.

Por su parte, el valor de $\mathbb{E}[T_j]$ es entendido como el **tiempo medio de retorno**, por esta razón, si este es infinito, en promedio regresamos a j en un tiempo muy grande (infinito), y por tanto el estado es recurrente nulo. De manera contraria, si este número se alcanza, entonces es recurrente positivo, pues, en media, regresamos en un tiempo finito. Tenemos resultados análogos a la teoría discreta.

Teorema 2.11. *La recurrencia nula y positiva son propiedades de clase*

Teorema 2.12. *Una cadena con espacio de estados finito no tiene estados recurrentes nulos.*

Este último teorema será especialmente útil cuando hablemos de modelos de coalescencia o cadenas que se absorban, pues en general siempre estarás un tiempo finito en cada estado que no sea absorbente.

Distribuciones Estacionaria

En el caso discreto, una distribución de probabilidad π se dice estacionaria para una cadena si cumple la ecuación de balance general $\pi P = \pi$; sin embargo, para el tiempo continuo tenemos un valor distinto de la matriz de transición para cada valor de $t \geq 0$. Por tanto, definir este concepto resulta más complicado. A continuación daremos las definiciones y resultados correspondientes al caso continuo.

Definición 2.13. *Una distribución de probabilidad π es una distribución estacionaria para la cadena de Markov $(X_t : t \geq 0)$ cuando*

$$\pi P(t) = \pi \text{ para toda } t \geq 0$$

El siguiente resultado nos da una caracterización de las distribuciones estacionarias.

Teorema 2.14. *Si $(X_t : t \geq 0)$ es una cadena de Markov a tiempo continuo con matriz de tasas Q , entonces una distribución de probabilidad π es una distribución estacionaria si y solo si $\pi Q = 0$*

Por lo que encontrar la distribución estacionaria se reduce a calcular el sistema de ecuaciones homogéneo inducido por $\pi Q = 0$.

Ejemplo 2.15. Si consideramos una cadena $(X_t : t \geq 0)$ con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1\}$ donde $1, 2, \dots, k$ son transientes y $k+1$ absorbente, entonces la proposición 1.28 de [14] nos permite afirmar que la única distribución estacionaria será δ_{k+1} tal que para cada $j \in \{1, 2, \dots, k, k+1\}$, se tiene que

$$\delta_{k+1}(j) = \begin{cases} 0 & \text{si } j \neq k+1 \\ 1 & \text{si } j = k+1 \end{cases}$$

Es decir que conforme pasa el tiempo, la distribución de las probabilidades de transición converge a δ_{k+1} . Finalmente, tenemos el siguiente resultado, que es análogo al de la teoría discreta y nos permitirá describir algunos modelos de genética de poblaciones.

Teorema 2.16. *Si $(X_t : t \geq 0)$ es una cadena de Markov a tiempo continuo, irreducible y con distribución estacionaria π , entonces para cualesquiera estados i, j ,*

$$\lim_{t \rightarrow \infty} p_t(i, j) = \pi(j)$$

En este caso, por la unicidad de límites, la distribución estacionaria es única.

2.2. Modelo de Wright-Fisher

Para este primer modelo tenemos los siguientes supuestos referentes a una población de tamaño N :

- Es panmítica (cada individuo tiene la misma probabilidad de reproducirse con cualquier otro).
- Es haploide (cada individuo carga solamente una copia de cada gen).
- Es neutral (cada individuo tiene la misma ventaja reproductiva sobre otro).
- El tiempo está discretizado y por tanto las generaciones no se traslanan.
- No hay generaciones con padres e hijos, pues una vez que un parente se reproduce, muere.
- Solo hay dos alelos: A y a .

La dinámica se describe a continuación: En cada paso del tiempo, los individuos de la generación n eligen a sus padres de manera aleatoria, uniforme e independiente de entre aquellos de la generación $n + 1$, heredando su alelo. Con esto en mente, se define al proceso estocástico $(X_n : n \in \mathbb{N})$ con espacio de estados $S = \{0, 1, \dots, N\}$ como aquel donde X_n es el número de individuos con alelo a en la generación n . Por lo tanto, las probabilidades de transición serán binomiales, es decir:

$$p_1(i, j) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

Es claro que este proceso es una cadena de Markov a tiempo discreto, pues el paso siguiente solo depende de su anterior, y se conoce como el **proceso de Wright-Fisher**. La interpretación en genética de poblaciones en esta cadena es que dado el presente, los individuos eligen a sus padres de manera uniforme e independiente y heredan el alelo que tengan. Un ejemplo de este proceso puede verse en la figura 2.2.

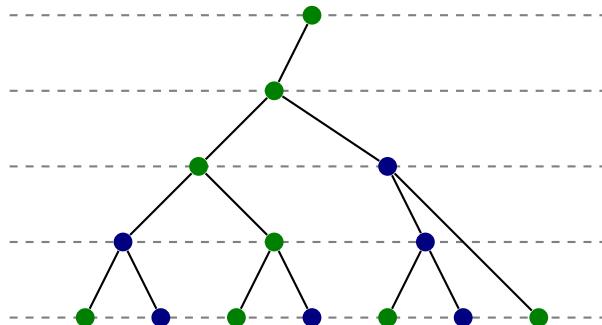


Figura 2.2: Ejemplo de un árbol coalescente usando el Modelo de Wright-Fisher

Notemos que este proceso es una martingala, pues al usar que $X_{n+1}|X_n \sim Bin(N, \frac{X_n}{N})$ y que es de Markov, se tiene que

$$\mathbb{E}[X_{n+1}|X_0, X_1, \dots, X_n] = \mathbb{E}[X_{n+1}|X_n] = N \frac{X_n}{N} = X_n$$

Esto quiere decir que, en media, el número de individuos con alelo a permanece igual a través del tiempo. Sin embargo, podemos notar que dado que la cadena es recurrente nula, pues 0 y N son absorbentes, entonces eventualmente alguno de los dos alelos se extinguirá y llegaremos al MCRA.

Ahora nos interesa presentar las herramientas para hablar del tiempo de coalescencia. Supongamos ahora que τ es la variable aleatoria que cuenta el tiempo hasta el último evento de coalescencia (o el tiempo que tardamos en llegar al MRCA) y que T_n es el tiempo de coalescencia dado que hay n linajes.

Analizaremos el caso cuando $n = 2$, pues eventualmente llegaremos a este, y después de un tiempo T_2 , llegaremos al MRCA. Notemos que para este caso,

$$\mathbb{P}[T_2 = 1] = \left(1 - \frac{1}{N}\right) = \frac{1}{N}$$

Pues el primer individuo elige un parente aleatoriamente (con probabilidad 1) y el segundo escoge a ese mismo con probabilidad $\frac{1}{N}$. Luego,

$$\mathbb{P}[T_2 = 2] = \left(1 - \frac{1}{N}\right) \frac{1}{N}$$

Pues quiere decir que en la generación anterior no lo encontraron, lo cual ocurre con probabilidad $1 - \frac{1}{N}$ y en la siguiente sí. En general, es fácil darse cuenta de que $T_2 \sim geo(\frac{1}{N})$, pues cuenta el número de generaciones que pasan hasta que coalescen, lo cual se puede ver como el número de fracasos antes del primer éxito. Así,

$$\mathbb{P}[T_2 = k] = \left(1 - \frac{1}{N}\right)^{k-1} \frac{1}{N} \quad \text{y} \quad \mathbb{E}[T_2] = N$$

Esto quiere decir que el tiempo de coalescencia es proporcional a la cantidad de individuos totales. Ahora analicemos el caso general, el cual es similar, pero como ahora no contamos con solo dos linajes, debemos considerar todas las formas en las que podemos elegir 2 elementos de nuestra muestra de n para llevar a cabo el evento de coalescencia. Sabemos que $\binom{n}{2}$ es el número de parejas que podemos formar de un total de n , y al elegirlas, la probabilidad de que coalezcan es de $\frac{1}{N}$, por lo tanto

$$\mathbb{P}[T_n = 1] = \left(1 - \frac{\binom{n}{2}}{N}\right) = \frac{\binom{n}{2}}{N}$$

Por un razonamiento análogo al caso cuando $n = 2$, se tiene que

$$\mathbb{P}[T_n = k] = \left(1 - \frac{\binom{n}{2}}{N}\right)^{k-1} \frac{\binom{n}{2}}{N} \tag{2.5}$$

Es decir que $T_n \sim geo\left(\frac{\binom{n}{2}}{N}\right)$ y por tanto

$$\mathbb{E}[T_n] = \frac{N}{\binom{n}{2}}$$

Esto quiere decir que mientras haya más linajes, será menor el tiempo de coalescencia. Esto era de esperarse, pues hay muchas más exponenciales y la primera en sonar lo hace en media el inverso de la suma de los parámetros individuales, lo cual es rápido. **ESTO ESTÁ ESCRITO HORRIBLE, HABRÁ QUE CORREGIRLO.**

2.3. Modelo de Morán

En esta sección introducimos el modelo de Morán, el cual es una cadena de Markov a tiempo continuo y nos será de mucha utilidad para construir el proceso coalescente. Dadas sus características, daremos una introducción a los procesos de nacimiento y muerte, pues conocerlos nos será de utilidad para comprender el modelo.

Procesos de Nacimiento y Muerte

Consideremos una cadena de Markov a tiempo continuo ($Z_t : t \geq 0$) donde el valor de las variable en cada momento del tiempo corresponde a la cantidad de individuos en una población en dicho tiempo. Por esta razón, tenemos que el espacio de estados del proceso es $\mathcal{S} = \mathbb{N}$. Consideraremos los siguientes supuestos:

- Las transiciones del proceso solo se dan un aumento de 1 (hay un nacimiento) o disminución de 1 (hay una muerte).
- Es neutral.
- Las generaciones pueden traslaparse.

La dinámica es sencilla: si suponemos que las variables aleatorias $T_1, T_2, T_3\dots$ representan los tiempos entre cada salto de la cadena, entonces solo puede pasar una de las siguientes para cada $n \in \mathbb{N}$:

1. $Z_{T_n} = Z_{T_{n-1}} + 1$ si hay un nacimiento.
2. $Z_{T_n} = Z_{T_{n-1}} - 1$ si hay una muerte.

Por lo tanto, la matriz de tasas del proceso será tridiagonal ², pues dado un estado j , solo tengo tasa positiva de ir a $j-1$, j o $j+1$. Para conocer estas tasas, supongamos que para cada $i \in \mathcal{S}$ tenemos una tasa de muerte μ_i y una tasa de nacimientos λ_i . Observemos que las tasas dependen del estado, pues mientras más individuos hay es

²Una matriz cuadrada cuyos elementos distintos de cero están en la diagonal y en las dos diagonales adyacentes

mucho más probable que haya reproducciones (y con ello, nacimientos) o que alguno muera. Así, la matriz de tasas será muy parecida a aquella del ejemplo 2.7

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2.6)$$

Por otro lado, al aplicar el resultado obtenido en 2.6 y siguiendo el resultado expuesto en el ejemplo 2.7 se tiene que las probabilidades de transición estarán dadas por

$$p(i, j) = \begin{cases} \frac{\lambda_i}{\lambda_i + \mu_i} & \text{si } j = i + 1 \\ \frac{\mu_i}{\lambda_i + \mu_i} & \text{si } j = i - 1 \\ 0 & \text{en otro caso} \end{cases} \quad (2.7)$$

Sabiendo cómo es la dinámica de los procesos de nacimiento y muerte, estudiemos cómo es la dinámica del modelo de Morán.

El modelo de Morán se basa en una población de tamaño N , panmítica, haploide y con dos alelos: A y a. Tiene dos diferencias principales frente al modelo de Wright-Fisher:

1. Las generaciones no son discretas, por tanto se podrán traslapar y el modelo será en tiempo continuo.
2. En este modelo solo se puede tener 0 o 2 descendientes en un tiempo exponencial con tasa 1.

La dinámica del modelo de Morán es la siguiente: en un tiempo exponencial de parámetro 1 se elige a dos individuos de la población, de donde uno muere (tiene 0 descendencia) y el otro se reproduce (y tiene 2 hijos). Esto se repite cambiando cada vez la selección de individuos y el tiempo que tardan en el proceso de nacimiento y muerte. Con esto en mente, pensemos al proceso estocástico ($Y_t : t \geq 0$) con espacio de estados $\mathcal{S} = \{0, 1, \dots, N\}$ como aquel donde Y_t es el número de individuos con alelo a al tiempo t .

Claramente esta cadena de Markov a tiempo continuo tiene como estados absorbentes al 0 y a N , pues una vez que todos se extinguen de algún alelo, entonces ya no puedes elegir de estos y el otro será el único en la población. Por lo tanto, dado que el espacio de estados es finito, sabemos que siempre se absorberá y llegaremos al MRCA. En la figura 2.3 puede verse un ejemplo del modelo de Morán.

En resumen, si suponemos que T_1, T_2, \dots, T_τ son los tiempos exponenciales de parámetro 1 donde se da el proceso de nacimiento y muerte (y donde τ es el número de veces que se da este procedimiento hasta llegar al MRCA), entonces solo puede pasar una de las siguientes para cada $n \in \mathbb{N}$:

1. $Y_{T_n} = Y_{T_{n-1}} + 1$ si el que se reproduce tenía alelo a y el que se muere A.
2. $Y_{T_n} = Y_{T_{n-1}} - 1$ si el que se reproduce tenía alelo A y el que se muere a.
3. $Y_{T_n} = Y_{T_{n-1}}$ si los dos que se eligieron tenían el mismo alelo.

Claramente se trata de una variación del proceso de nacimiento y muerte, por lo que la matriz de tasas es tridiagonal y tiene las siguientes tasas

$$q(i, j) = \begin{cases} \frac{i(N-i)}{N^2} & \text{si } j = i \pm 1 \\ -\frac{2i(N-i)}{N^2} & \text{si } j = i \\ 0 & \text{en otro caso} \end{cases} \quad (2.8)$$

Pues con probabilidad $\frac{i}{N}$ se elige un individuo con alelo a para reproducirse y con probabilidad $\frac{N-i}{N}$ se escoge uno con alelo A para morir. Esto es análogo para el caso cuando el que muere es a y se reproduce A. **AQUÍ PONER UNA FOTO**

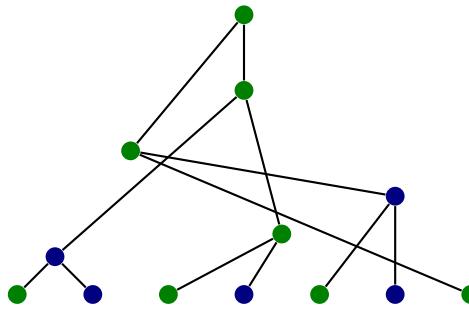


Figura 2.3: Ejemplo de un árbol coalescente bajo el Modelo de Morán.

2.4. Coalescente de Kingman

Finalmente, presentamos el modelo del coalescente de Kingman o n-coalescente, llamado así por John Kingman, quien desarrolló este modelo y sentó las bases de la teoría de coalescencia. El objetivo de este modelo matemático es encontrar el MRCA de una muestra dentro de una población, viendo cada evento de coalescencia (el momento donde los linajes se unen porque provienen del mismo alelo) como una transición en una cadena de Markov a tiempo continuo en el conjunto de las particiones del conjunto $\{0, 1, \dots, n\}$.

Consideremos una población de N individuos de donde tomamos una muestra de tamaño n y definamos una cadena de Markov a tiempo continuo $(X_t : t \geq 0)$ definida en \mathcal{E}_n , el cual es el conjunto de todas las particiones de $[n] := \{1, 2, \dots, n\}$ y con las siguientes tasas:

$$q(\mathcal{P}, \mathcal{Q}) = \begin{cases} 1 & \text{si } \mathcal{P} \prec \mathcal{Q} \\ -\lambda_{|\mathcal{P}|} & \text{si } \mathcal{P} = \mathcal{Q} \\ 0 & \text{en otro caso} \end{cases} \quad (2.9)$$

Donde $\mathcal{P} \prec \mathcal{Q}$ indica que podemos obtener \mathcal{Q} a través de la unión o coalescencia de dos elementos de \mathcal{P} y

$$\lambda_{|\mathcal{P}|} = |\{\mathcal{Q} \in \mathcal{E}_n : \mathcal{P} \prec \mathcal{Q}\}|$$

Dado que cada vez que hay una transición es porque se unen dos elementos de \mathcal{P} , entonces es sencillo ver que

$$\lambda_{|\mathcal{P}|} = \binom{|\mathcal{P}|}{2} = \frac{|\mathcal{P}|(|\mathcal{P}| - 1)}{2}$$

Para ejemplificar este proceso, se toma el supuesto de que siempre se empieza con los elementos de la población separados, es decir que $X_0 = \{\{1\}, \{2\}, \dots, \{n\}\}$. En la figura **AQUÍ PONER UNA FOTO** se puede ver un diagrama de cómo se desarrolla este proceso.

Con estas ideas en mente, podemos calcular el tiempo hasta el ancestro común más reciente de una muestra. Supongamos que T_k indica el tiempo en que dos linajes coalescen al tener k de ellos. Por tanto, sabemos que $T_k \sim \exp\left(\binom{k}{2}\right)$. Por lo tanto, el tiempo hasta el MRCA será

$$\tau = \sum_{k=2}^n T_k$$

la cual es la suma de variables aleatorias exponenciales de distinto parámetro. Finalmente, veremos la relación del coalescente de Kingman con el modelo de Wright-Fisher y del de Morán.

Relación entre Kingman y Wright-Fisher

En esta sección veremos la relación entre el modelo de Wright-Fisher y el coalescente de Kingman, y cómo el segundo es el límite del primero. Siguiendo un razonamiento parecido al hecho en 2.5, dados dos linajes, la probabilidad de que no coalezcan por t generaciones es $(1 - \frac{1}{N})^t$. Esto da lugar a un proceso a tiempo continuo, siempre que N sea suficientemente grande. Al reescalar el tiempo como si cada unidad fueran N generaciones obtenemos que la probabilidad de que no coalezcan después de t generaciones es

$$\left(1 - \frac{1}{N}\right)^{\lfloor Nt \rfloor} \rightarrow e^{-t} \text{ cuando } N \rightarrow \infty$$

Análogamente, si consideramos n linajes en la población, en total tenemos $\binom{n}{2}$ posibles eventos de coalescencia, por lo que la probabilidad de que no coalezcan después de t generaciones es de $\left(1 - \frac{\binom{n}{2}}{N}\right)^t$, y de manera análoga a 2.4, tenemos que

$$\left(1 - \frac{\binom{n}{2}}{N}\right)^{\lfloor Nt \rfloor} \rightarrow e^{-\binom{n}{2}t} \text{ cuando } N \rightarrow \infty$$

Por lo tanto, el tiempo de coalescencia dado que hay n linajes es una variable aleatoria exponencial de parámetro $\binom{n}{2}$, lo cual corresponde a las tasas de transición del coalescente de Kingman.

Por lo tanto, el tiempo esperado hasta el MRCA es:

$$\mathbb{E}[\tau] = \sum_{k=1}^n \mathbb{E}[T_k] = \sum_{i=1}^n \frac{1}{k(k-1)} = 2 \left(1 - \frac{1}{n}\right)$$

donde llegamos a la última igualdad por un desarrollo en fracciones parciales.

2.5. Λ -coalescente

El Λ -coalescente es un modelo basado en una población de tamaño N y de donde extraemos una muestra de n individuos. Consideremos una cadena de Markov a tiempo continuo $(X_t : t \geq 0)$ definida en el espacio \mathcal{E}_n , las particiones del conjunto $[n]$, y que modela la coalescencia para dicha población. Su principal diferencia con el modelo de Kingman es que se puede dar la coalescencia de más de 2 linajes. Supongamos que Λ es una medida finita en $[0, 1]$, entonces dado que hay b genealogías, la tasa para que k de ellas coalezcan está dada por

$$\lambda_{b,k} = \int_0^1 x^{k-2} (1-x)^{b-k} \Lambda(dx) \quad \text{para } k \in \{2, 3, \dots, b\} \quad (2.10)$$

En esta sección describiremos esta cadena y daremos expresiones para su matriz de tasas y sus características principales.

Para describir la matriz de tasas, notemos que si queremos pasar de tener i a j linajes, entonces es porque coalescieron $i-j+1$ de ellos, lo cual puede pasar de $\binom{i}{i-j+1}$ maneras distintas y ocurre con tasa $\lambda_{i,i-j+1}$. Por esta razón, tenemos que la matriz de tasas estará dada como sigue:

$$S_{ik} = \binom{i}{k} \lambda_{i,k} \quad \text{con } k = i - j + 1$$

Y

$$S_{ii} = - \sum_{k \neq i} S_{i,j}$$

Relación con el Coalescente de Kingman

Ahora veamos su relación con el coalescente de Kingman. Sabemos que la delta de Dirac δ_0 es una medida finita definida como

$$\delta_0(E) \begin{cases} 1 & \text{si } 0 \in E \\ 0 & \text{si } 0 \notin E \end{cases}$$

Si tomamos la restricción al $[0, 1]$, podemos suponer que $\Lambda = \delta_0$ y veamos cómo son las tasas en este caso:

- Si $k \neq 2$, entonces

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k} \delta_0(dx) = 0^{k-2}(1-0)^{b-k} = 0$$

- Si $k = 2$ y asumiendo que $0^0 = 1$, se tiene que

$$\lambda_{b,k} = \int_0^1 x^0(1-x)^{b-2} \delta_0(dx) = 0^0(1)^{b-2} = 1$$

En general $\lambda_{b,k} = \mathbf{1}_{\{2\}}(k)$ y las tasas fuera de la diagonal estarán dadas por

$$S_{ij} = \binom{i}{i-j+1} \mathbf{1}_{\{2\}}(i-j+1)$$

Sin embargo, notemos que $\mathbf{1}_{\{2\}}(i-j+1) \neq 0$ siempre que $j = i-1$, y en este caso

$$\binom{i}{i-j+1} = \binom{i}{i-(i-1)+1} = \binom{i}{i-i+1+1} = \binom{i}{2}$$

Por lo que al asignar esta medida hemos recuperado el coalescente de Kingman.

ψ -coalescente

Consideremos $\psi \in (0, 1)$ y $\Lambda = \delta_\psi$ en $[0, 1]$. En este caso, las tasas de coalescencia del modelo están dadas por

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k} \delta_\psi(dx) = \psi^{k-2}(1-\psi)^{b-k} \quad (2.11)$$

Por otro lado, la matriz de tasas estará dada por las siguientes expresiones

$$S_{ij} = \binom{i}{i-j+1} \psi^{i-j-1}(1-\psi)^{j-1} \quad \text{y} \quad S_{ii} = - \sum_{j \neq i} S_{ij} \quad (2.12)$$

En el ψ -coalescente, el parámetro ψ es la proporción de la población descendiente de un único ancestro en un evento dado, lo que determina cuántas líneas genealógicas se fusionan simultáneamente. Con esto en mente, si ψ es pequeña, entonces habrá muchos eventos pequeños de coalescencia, mientras que si ψ es grande, puede ocurrir que un individuo reemplace a un sector amplio de la población.

En las figuras AQUÍ PONER IMÁGENESSSS podemos ver algunos árboles coalescentes para valores distintos de ψ

β -coalescente

En este caso, supondremos que la medida Λ corresponde a la densidad $Beta(2 - \alpha, \alpha)$ para $\alpha \in [1, 2]$. Recordando que la función β está definida como

$$\beta(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx \quad \text{para } p, q > 0$$

Entonces podemos encontrar una expresión para las tasas de coalescencia en este modelo

$$\begin{aligned} \lambda_{b,k} &= \int_0^1 x^{k-2}(1-x)^{b-k}f_X(x)dx = \int_0^1 \frac{x^{k-2}(1-x)^{b-k}x^{2-\alpha-1}(1-x)^{\alpha-1}}{\beta(2-\alpha, \alpha)}dx \\ &= \frac{1}{\beta(2-\alpha, \alpha)} \int_0^1 x^{k-2+2-\alpha-1}(1-x)^{b-k+\alpha-1}dx \\ &= \frac{1}{\beta(\alpha, 2-\alpha)} \int_0^1 x^{k-\alpha-1}(1-x)^{b-k+\alpha-1}dx \\ &= \frac{\beta(k-\alpha, b-k+\alpha)}{\beta(\alpha, 2-\alpha)} \end{aligned}$$

Por lo tanto, en este caso

$$\lambda_{b,k} = \frac{\beta(k-\alpha, b-k+\alpha)}{\beta(\alpha, 2-\alpha)}$$

Este modelo sirve para describir árboles coalescentes con muestras de individuos muy grandes, como se discute en [10].

3 Distribuciones Tipo Fase

Como ya estudiamos, los modelos que buscan el ancestro común más reciente son cadenas de Markov a que eventualmente se absorben en un estado. Las distribuciones tipo fase serán esenciales para este tipo de comportamiento, pues son el tiempo de absorción de una cadena, a tiempo discreto o continuo. En particular, este tipo de distribuciones generalizan a las de Erlang¹, pues estas corresponden al tiempo de absorción en un Proceso Poisson que se extingue.

Las distribuciones tipo fase son de mucha utilidad en modelos de genética poblacional, pues, por ejemplo, las dimensiones del árbol coalescente tendrán esta distribución. Además, mediante operaciones de matrices podremos deducir sus propiedades básicas, como expresiones para su densidad, distribución o sus momentos y algunas otras como distribuciones conjuntas, covarianzas y coeficiente de correlación entre variables del proceso coalescente.

A su vez, estas distribuciones cumplen propiedades teóricas que las hacen destacar, ya que cualquier distribución no negativa se puede aproximar arbitrariamente cerca por una distribución de tipo fase. Para ver una discusión y la prueba de este hecho se puede consultar A.

En este capítulo se desarolla la teoría acerca de las distribuciones tipo fase para aplicarla en modelos generacionales, con el objetivo de dar un panorama general sobre estas distribuciones y ser capaces de aplicarlos. Los metodología aquí utilizada proviene, en su mayoría, de [2] y [22].

3.1. Distribuciones Tipo Fase Discretas

Sea $(X_n : n \in \mathbb{N})$ la cadena de Markov a tiempo discreto con espacio de estados $\mathcal{S} = \{1, 2, 3\}$, con distribución inicial $\boldsymbol{\pi} = (\pi_1, \pi_2, 0)$ y cuya matriz de transición es

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

¹Si $X \sim Erlang(n, \lambda)$, entonces su función de densidad es $f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \mathbf{1}_{\{x \geq 0\}}$, que coincide con la distribución $Gamma(n, \lambda)$.

Notamos que el estado 3 es absorbente y debido a la forma de nuestra distribución inicial, no podemos empezar en él. Nos interesa estudiar la variable aleatoria definida como

$$\tau := \inf\{n \in \mathbb{N} : X_n = 3\}$$

la cual entendemos como el tiempo de absorción de la cadena. En la figura 3.1 podemos ver un ejemplo de una trayectoria de esta cadena, y donde $\tau = 3$. A la distribución τ se le conoce como distribución tipo fase discreta, la cual tiene como parámetros a (π_1, π_2) y a \mathbf{T} , la cual es la submatriz estocástica (cuyas entradas son no negativas y suman a lo más 1) dada por

$$\mathbf{T} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

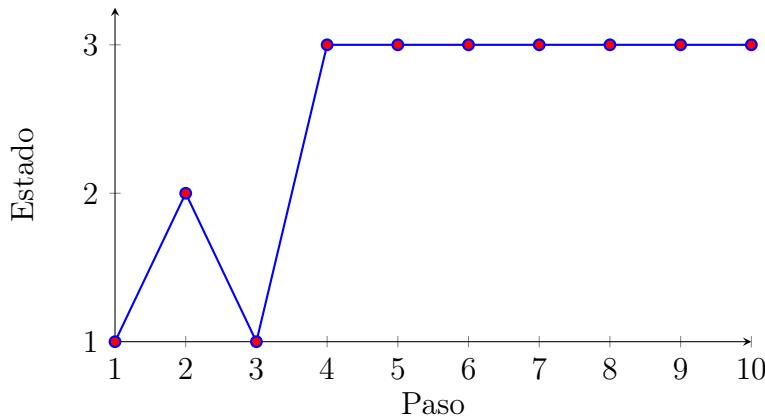


Figura 3.1: Trayectoria de la cadena donde $\tau = 3$

A continuación daremos la definición para cualquier cadena a tiempo discreto y el significado de los parámetros de la variable.

Definición y Ejemplos

Consideraremos una cadena de Markov a tiempo discreto $(X_n : n \in \mathbb{N})$, con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1\}$, donde los estados $1, 2, \dots, k$ son transientes y $k+1$ absorbente. Por lo tanto, la matriz de transición de la cadena es de la forma

$$P = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \quad (3.1)$$

donde \mathbf{T} es una submatriz estocástica, \mathbf{t} es un vector columna de tamaño k y $\mathbf{0}$ corresponde al vector de longitud k cuyas entradas son todas 0. Dado que P es una matriz de transición, sus columnas deben sumar 1. Esto es:

$$\sum_{j=1}^k T_{ij} + t_i = 1 \quad \text{para cada } i \in \{1, \dots, k\} \quad (3.2)$$

En términos matriciales, si \mathbf{e} es el vector columna de tamaño p con entradas todas 1 e \mathbf{I} la matriz identidad de $k \times k$, tenemos que:

$$\mathbf{t} + \mathbf{T}\mathbf{e} = \mathbf{e} \implies \mathbf{t} = (\mathbf{I} - \mathbf{T})\mathbf{e} \quad (3.3)$$

Ya que $\mathbf{T}\mathbf{e}$ corresponde al vector columna donde la entrada i es la suma de los elementos de la fila i de la matriz \mathbf{T} . A \mathbf{t} se le conoce como el vector de probabilidades de salida. Adicionalmente, consideremos a una distribución inicial $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$. Con estas herramientas listos para dar la siguiente definición.

Definición 3.1. Sea

$$\tau := \inf\{n \in \mathbb{N} : X_n = k + 1\}$$

el tiempo hasta absorción de la cadena. Decimos que τ tiene **distribución de tipo fase discreta** de parámetros $\boldsymbol{\pi}$ y \mathbf{T} . En dado caso, escribimos

$$\tau \sim DPH_k(\boldsymbol{\pi}, \mathbf{T}) \quad (3.4)$$

Al par $(\boldsymbol{\pi}, \mathbf{T})$ se le conoce como la **representación tipo fase** para la distribución.

En este trabajo, al vector de probabilidades de salida se le denotará como la letra en minúscula y en negritas correspondiente a aquella de la submatriz estocástica de la representación tipo fase. Usaremos a $(\boldsymbol{\pi}, 0)$ como distribución inicial de la cadena, es decir que $\tau \neq 0$, pues no podemos empezar desde el estado absorbente.

Ejemplo 3.2. Consideremos un experimento aleatorio donde podemos tener un éxito o un fracaso en cada paso del tiempo y con probabilidades p y $1 - p$, respectivamente. Al primer éxito, se termina el juego. Sea τ el número de pasos hasta el primer éxito, entonces τ tiene distribución tipo fase discreta, con $\boldsymbol{\pi} = (1)$ (quiere decir que siempre empezamos con un fracaso), $\mathbf{T} = (1 - p)$ y, por consecuencia, $\mathbf{t} = (p)$.

Nótese que $\tau \sim geo(p)$, lo cual tiene sentido, pues es la única distribución discreta con pérdida de la memoria, lo que hace que este proceso sea de Markov. En este caso particular, tenemos que para $x \in \mathbb{N}$,

$$f_\tau(x) = (1 - p)^{x-1} p \quad (3.5)$$

Propiedades

A continuación veremos algunos resultados relevantes respecto a las distribuciones tipo fase discretas.

Lema 3.3. Sea $n \in \mathbb{N}$, entonces

$$1. \sum_{k=0}^{n-1} \mathbf{T}^k \mathbf{t} = (\mathbf{I} - \mathbf{T}^n) \mathbf{e}$$

$$2. P^n = \begin{pmatrix} \mathbf{T}^n & (\mathbf{I} - \mathbf{T}^n) \mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix}$$

Demostración. Probemos 1. Usando que $\mathbf{t} = (I - \mathbf{T})\mathbf{e}$, tenemos:

$$\begin{aligned}\sum_{k=0}^{n-1} \mathbf{T}^k \mathbf{t} &= \sum_{k=0}^{n-1} \mathbf{T}^k (I - \mathbf{T})\mathbf{e} = \sum_{k=0}^{n-1} (\mathbf{T}^k - \mathbf{T}^{k+1})\mathbf{e} = \left(\sum_{k=0}^{n-1} \mathbf{T}^k - \sum_{k=0}^{n-1} \mathbf{T}^{k+1} \right) \mathbf{e} \\ &= \left(\sum_{k=0}^{n-1} \mathbf{T}^k - \sum_{k=1}^n \mathbf{T}^k \right) \mathbf{e} = (I - \mathbf{T}^n)\mathbf{e}.\end{aligned}$$

Para 2, procedemos por inducción sobre n . El caso base es claro. Supongamos que

$$P^n = \begin{pmatrix} \mathbf{T}^n & (I - \mathbf{T}^n)\mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{T}^n & \sum_{k=0}^{n-1} \mathbf{T}^k \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}$$

. Usando 1, tenemos que

$$\begin{aligned}P^{n+1} &= P^n P = \begin{pmatrix} \mathbf{T}^n & \sum_{k=0}^{n-1} \mathbf{T}^k \mathbf{T} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{T}^{n+1} + \mathbf{0} \sum_{k=0}^{n-1} \mathbf{T}^k \mathbf{T} & \mathbf{T}^n \mathbf{t} + \sum_{k=0}^{n-1} \mathbf{T}^k \mathbf{t} \\ \mathbf{0} \mathbf{T} + 1 \cdot \mathbf{0} & \mathbf{0} \mathbf{t} + 1 \cdot 1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{T}^{n+1} & \sum_{k=0}^n \mathbf{T}^k \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{T}^{n+1} & (I - \mathbf{T}^{n+1})\mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix}\end{aligned}$$

tal como queríamos. \square

Gracias a este lema, tenemos que si $i \in \{1, 2, \dots, k\}$, entonces $(P^n)_i = (\mathbf{T}^n)_i$. Es decir que siempre que estemos en los estados transientes, es lo mismo calcular P^n que \mathbf{T}^n . Gracias a esto podemos dar una expresión explícita de la densidad de la distribución

Teorema 3.4. Si $\tau \sim DPH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces $f_\tau(n) = \boldsymbol{\pi} \mathbf{T}^{n-1} \mathbf{t}$ para todo $n \in \mathbb{N}$.

Demostración. Usando probabilidad total dos veces, tenemos que

$$\begin{aligned}
f_\tau(n) &= \mathbb{P}[\tau = n] \\
&= \sum_{j=1}^k \mathbb{P}[\tau = n | X_{n-1} = j] \mathbb{P}[X_{n-1} = j] \\
&= \sum_{j=1}^k \mathbb{P}[\tau = n | X_{n-1} = j] \sum_{i=1}^k \mathbb{P}[X_{n-1} = j | X_0 = i] \mathbb{P}[X_0 = i] \\
&= \sum_{j=1}^k \sum_{i=1}^k \underbrace{\mathbb{P}[\tau = n | X_{n-1} = j]}_{t_j} \underbrace{\mathbb{P}[X_{n-1} = j | X_0 = i]}_{p_n(i,j)} \underbrace{\mathbb{P}[X_0 = i]}_{\pi_j} \\
&= \sum_{j=1}^k \sum_{i=1}^k t_j p_n(i,j) \pi_i \\
&= \sum_{j=1}^k \sum_{i=1}^k t_j (P^{n-1})_{ij} \pi_i \\
&= \sum_{j=1}^k \sum_{i=1}^k t_j (\mathbf{T}^{n-1})_{ij} \pi_i \\
&= \boldsymbol{\pi} \mathbf{T}^{n-1} \mathbf{t}
\end{aligned}$$

□

Ya tenemos una representación matricial de la densidad de la distribución. Veamos que se cumple siguiendo con el ejemplo 3.2, sabemos que en ese caso $\boldsymbol{\pi} = (1)$, $\mathbf{T} = (1 - p)$ y $\mathbf{t} = (p)$. Por tanto, dado $x \in \mathbb{N}$

$$f_\tau(x) = \boldsymbol{\pi} \mathbf{T}^{x-1} \mathbf{t} = 1(p)^{x-1}(1 - p) = (1 - p)^{x-1}p$$

que coincide con lo que habíamos obtenido en 3.5.

Teorema 3.5. Si $\tau \sim DPH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces $F_\tau(n) = 1 - \boldsymbol{\pi} \mathbf{T}^n \mathbf{e}$ para todo $n \in \mathbb{N}$.

*Demuestra*ción. La demostración es sencilla y basta con notar que $\tau > n$ es equivalente a que $X_n = j$ para algún $j \in \{1, 2, \dots, k\}$. Así, tenemos que:

$$\begin{aligned}
1 - F_\tau(n) &= \mathbb{P}[\tau > n] \\
&= \mathbb{P}[X_n \in \{1, 2, \dots, k\}] \\
&= \sum_{j=1}^k \mathbb{P}[X_n = j] \\
&= \sum_{j=1}^k \sum_{i=1}^k \mathbb{P}[X_n = j | X_0 = i] \mathbb{P}[X_0 = i] \\
&= \sum_{i=1}^k \sum_{j=1}^k \pi_i (\mathbf{T}^n)_{ij} \\
&= \boldsymbol{\pi} \mathbf{T}^n \mathbf{e}.
\end{aligned}$$

□

A los resultados 3.4 y 3.5 los podemos interpretar de la siguiente manera:

- En ambos casos, el término π en ambos resultados indica que iniciamos de manera aleatoria en algún estado, siguiendo la distribución inicial.
- Para el caso de la densidad, el término \mathbf{T}^{n-1} significa que el proceso llegó en el paso $n - 1$ a un estado transiente siguiendo la ley de probabilidad descrita por \mathbf{T} . Luego, el vector \mathbf{t} indica el momento en que el proceso salta al estado absorbente y termina el experimento. Esto es parecido a la función de masa de una variable aleatoria geométrica.
- Por otro lado, el término \mathbf{T}^n en la distribución, quiere decir que hasta el paso n , el proceso ha seguido en los estados transientes y no ha llegado al absorbente

Ahora nos interesa estudiar a una matriz que será de gran importancia en la teoría de distribuciones tipo fase discretas. Para ello, tenemos el siguiente teorema

Teorema 3.6. *Si $\tau \sim DPH_k(\pi, \mathbf{T})$, entonces la matriz $(\mathbf{I} - \mathbf{T})^{-1}$ existe.*

Demuestra. Como \mathbf{T} es una matriz subestocástica y todas sus entradas corresponden a estados transientes, se sigue que $\mathbf{T}^n \rightarrow \mathbf{0}$ cuando $n \rightarrow \infty$. Supongamos que x cumple que $x(\mathbf{I} - \mathbf{T}) = \mathbf{0}$ y entonces $x = x\mathbf{T}$. Por medio de un razonamiento inductivo, se puede probar que $x = x\mathbf{T}^n$ para todo $n \in \mathbb{N}$. Aplicando límite cuando $n \rightarrow \infty$, se tiene que $x = 0$. Hemos demostrado que la única solución a la ecuación $x(\mathbf{I} - \mathbf{T}) = \mathbf{0}$ fue la trivial, por lo tanto el núcleo de $(\mathbf{I} - \mathbf{T})$ es el conjunto $\{0\}$. Por el teorema de la dimensión, el rango de $\mathbf{I} - \mathbf{T}$ es k , por tanto la matriz es invertible. □

Este teorema nos permite tener la siguiente definición.

Definición 3.7. *Si $\tau \sim DPH_k(\pi, \mathbf{T})$, definimos su **matriz de Green asociada** como*

$$\mathbf{U} = (\mathbf{I} - \mathbf{T})^{-1} \tag{3.6}$$

Con el objetivo de dar una interpretación para la matriz de Green, en lo subsecuente denotaremos por $\mathbb{P}_i[\cdot] := \mathbb{P}[\cdot | X_0 = i]$ y $\mathbb{E}_i[\cdot] := \mathbb{E}[\cdot | X_0 = i]$. Usando que si $j \in \{1, 2, \dots, k\}$ y $X_n = j$, entonces $\tau > n$, se sigue que $[X_n = j] \subseteq [\tau > n]$ y por tanto

$$\mathbb{P}_i[X_n = j; \tau > n] = \mathbb{P}_i[X_n = j] = (P^n)_{ij}$$

Luego,

$$\mathbb{P}_i[X_n = j, \tau > n] = (\mathbf{T}^n)_{ij} \tag{3.7}$$

Ahora veamos cuál es la interpretación de la matriz de Green. Supongamos que $j \in \{1, 2, \dots, k\}$ y que Z_j es el tiempo que el proceso pasa en j antes de absorberse. Es decir que

$$Z_j = \sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n=j\}}$$

Por otro lado, afirmamos que:

$$\sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n=j\}} = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j; \tau>n\}} \quad (3.8)$$

Pues

$$\mathbf{1}_{\{X_n=j\}} = \mathbf{1}_{\{X_n=j; \tau>n\}} + \mathbf{1}_{\{X_n=j; \tau\leq n\}}$$

pero $\{X_n = j; \tau \leq n\} = Varnothing$ entonces $\mathbf{1}_{\{X_n=j\}} = \mathbf{1}_{\{X_n=j; \tau \geq n\}}$. Luego, tenemos que

$$\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j; \tau \geq n\}} = \sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n=j; \tau \geq n\}} + \sum_{n=\tau}^{\infty} \mathbf{1}_{\{X_n=j; \tau \geq n\}}$$

Pero esta segunda suma es 0, pues para cada $n \geq \tau$, se tiene que $X_n = k+1$ y $j \neq k+1$. Por lo tanto,

$$\sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n=j\}} = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j; \tau>n\}}$$

Notemos entonces que si usamos 3.7, 3.8 y las propiedades de esperanza, tenemos que

$$\mathbb{E}_i [Z_j] = \mathbb{E}_i \left[\sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n=j\}} \right] = \mathbb{E}_i \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j; \tau>n\}} \right] = \sum_{n=0}^{\infty} \mathbb{P}_i [X_n = j, \tau > n] = \sum_{n=0}^{\infty} (T^n)_{ij}.$$

Donde pudimos alternar la esperanza con la serie debido a que

$$\mathbb{E}_i [\mathbf{1}_{\{X_n=j, \tau>n\}}] = \mathbb{P}_i [X_n = j; \tau \geq n] = \mathbb{P}_i [X_n = j] = p_n(i, j) < \infty$$

Y,

$$\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j; \tau>n\}} = \sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n=j\}} \leq \sum_{n=0}^{\tau-1} 1 \leq \tau < \infty$$

Luego, como los estados son transientes, tenemos que $\mathbb{E}_i [Z_j] < \infty$ y por tanto la serie $\sum_{n=0}^{\infty} (T^n)_{ij} < \infty$. Ahora veamos que $\sum_{n=0}^{\infty} (T^n)_{ij} = (\mathbf{I} - \mathbf{T})^{-1}$ en términos de matrices. Sea $k \in \mathbb{N}$, entonces

$$\left(\sum_{n=0}^k T^n \right) (\mathbf{I} - T) = \sum_{n=0}^k T^n - \sum_{n=0}^k T^{n+1} = \mathbf{I} - T^{k+1} \quad (3.9)$$

Tomando límite cuando $k \rightarrow \infty$, se sigue el resultado, pues dado que la matriz \mathbf{T} es subestocástica, entonces $\mathbf{T}^k \rightarrow \mathbf{0}$. Como las matrices son cuadradas y $(\mathbf{I} - \mathbf{T})^{-1}$ es la inversa derecha de $\sum_{n=0}^{\infty} (T^n)_{ij}$, entonces es su inversa. Por tanto, tenemos que

$$\mathbb{E}_i [Z_j] = \sum_{n=0}^{\infty} (T^n)_{ij} = \mathbf{U}_{ij} \quad (3.10)$$

Con este procedimiento hemos demostrado el siguiente teorema.

Teorema 3.8. La entrada (i, j) de la matriz de Green asociada a una distribución de tipo fase discreta corresponde al valor esperado de tiempo que la cadena pasa en el estado j antes de absorberse y dado que empezó en i .

Veamos que la matriz de Green nos permite obtener una expresión para la esperanza de una distribución de tipo fase discreta.

Teorema 3.9. Sea $\tau \sim DPH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces

$$\mathbb{E}[\tau] = \boldsymbol{\pi} \mathbf{U} \mathbf{e} \quad (3.11)$$

Demostración. Observemos que si Z_j está definido como en el razonamiento anterior, entonces $\tau = \sum_{j=1}^k Z_j$, pues es la suma de los tiempos que pasamos en los estados transientes antes de ser absorbidos. Así, usando las propiedades básicas de esperanza,

$$\mathbb{E}[\tau] = \mathbb{E} \left[\sum_{j=1}^k Z_j \right] = \sum_{j=1}^k \mathbb{E}[Z_j] = \sum_{j=1}^k \sum_{i=1}^k \mathbb{E}_i[Z_j] \boldsymbol{\pi}_i = \sum_{j=1}^k \sum_{i=1}^k \mathbf{U}_{ij} \boldsymbol{\pi}_i = \boldsymbol{\pi} \mathbf{U} \mathbf{e}$$

□

Finalmente, veremos que la suma de dos de ellas vuelve a ser de tipo fase. La prueba de esto es muy común en este tipo de distribuciones, pues estudiaremos las cadenas de Markov subyacentes a los tiempos de absorción y construiremos una nueva. Este tipo de demostraciones serán muy usadas en este texto y se conocen como razonamientos probabilistas.

Teorema 3.10. Si $\tau_1 \sim DPH_k(\boldsymbol{\alpha}, \mathbf{S})$ y $\tau_2 \sim DPH_m(\boldsymbol{\beta}, \mathbf{T})$, entonces

$$\tau_1 + \tau_2 \sim DPH_{k+m} \left((\boldsymbol{\alpha}, \mathbf{0}), \begin{pmatrix} \mathbf{S} & s\boldsymbol{\beta} \\ \mathbf{0} & \mathbf{T} \end{pmatrix} \right)$$

Demostración. Sean $(X_n^1 : n \in \mathbb{N})$ y $(X_n^2 : n \in \mathbb{N})$ las cadenas de Markov asociadas a los valores de τ_1 y τ_2 , respectivamente. Construyamos una nueva cadena $(X_n : n \in \mathbb{N})$ con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1, \dots, k+m, k+m+1\}$ cuya distribución inicial sea $(\boldsymbol{\alpha}, \mathbf{0})$. Es decir que inicia en los primeros k estados de la cadena. Es fácil notar que el comportamiento de la cadena hasta τ_1 será igual que el de $(X_n^1 : n \in \mathbb{N})$ hasta τ_1 . Sabemos que $X_{\tau_1} = k+1$, entonces en este tiempo hacemos que la cadena siga, eligiendo el valor para X_{τ_1+1} de entre los estados $\{k+2, k+3, \dots, k+m\}$ y con la distribución $\boldsymbol{\beta}$. Así, la cadena seguirá hasta $X_{\tau_1+\tau_2}$ y los valores desde X_{τ_1+1} hasta el final seguirán la misma ley de probabilidad de $(X_n^2 : n \in \mathbb{N})$. Claramente, el tiempo de absorberse de la nueva cadena es $\tau_1 + \tau_2$ y por construcción tenemos que

$$P = \begin{pmatrix} \mathbf{S} & s\boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{t} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix}$$

Pues si $i, j \in \{1, 2, \dots, k\}$, la probabilidad de ir de i a j es \mathbf{S}_{ij} . Por otro lado, para ir de i a $j+m \in \{k+1, k+2, \dots, k+m\}$ es $s_i \boldsymbol{\beta}_j$, pues significa que estando en i ,

te absorbiste en la primera cadena y después elegir el estado j con probabilidad β_j . Finalmente, si $j + m \in \{k + 1, k + 2, \dots, k + m\}$ entonces llegamos a absorberlos con probabilidades del vector de salida \mathbf{t}_{j+m} . Finalmente, es claro que el estado $m + k + 1$ es absorbente. \square

Como pudimos ver, las distribuciones tipo fase discretas son una manera natural de hablar de tiempo de absorción de una cadena de Markov de esta naturaleza. Para el caso discreto, la distribución geométrica juega un papel crucial en su desarrollo, pues tiene la propiedad de pérdida de la memoria.

En las siguientes secciones se estudiarán las propiedades de las distribuciones tipo fase definidas sobre una cadena de Markov a tiempo continuo. Ahora trabajaremos con la matriz de tasas y tendremos resultados parecidos a los de la teoría discreta. Igualmente introduciremos conceptos nuevos, como transformaciones vía recompensas y covarianzas entre variables de tipo fase.

3.2. Definición, Densidad y Distribución

Consideremos una cadena de Markov a tiempo continuo $(X_t : t \geq 0)$ con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1\}$, donde $1, 2, \dots, k$ son transientes y $k+1$ es absorbente. Sabemos que la matriz de tasas del proceso tendrá la siguiente forma

$$Q = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix} \quad (3.12)$$

Dado que las filas de esta matriz deben sumar 0, entonces de manera análoga al caso discreto se tiene que

$$\sum_{j=1}^k T_{ij} + t = 0 \quad \text{para toda } i \in \{1, 2, \dots, k\}$$

Usando la notación introducida en la sección anterior, esto quiere decir que

$$\mathbf{t} + \mathbf{T}\mathbf{e} = 0 \implies \mathbf{t} = -\mathbf{T}\mathbf{e} \quad (3.13)$$

Como en el caso discreto, consideraremos una distribución de probabilidad $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ la cual servirá para asignar el estado inicial a la cadena de entre todos aquellos que son transientes. Con esto tenemos todo para dar la siguiente definición.

Definición 3.11. *Sea*

$$\tau := \inf\{t > 0 | X_t = p + 1\}$$

el tiempo de absorción de la cadena de Markov a tiempo continuo. Decimos que τ tiene distribución de tipo fase con parámetros la distribución inicial $\boldsymbol{\pi}$ y la submatriz de tasas \mathbf{T} . Lo denotamos por

$$\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$$

Y al par $(\boldsymbol{\pi}, \mathbf{T})$ lo llamamos una representación para la distribución tipo fase.

Ejemplo 3.12 (Distribución Exponencial). Notemos que si consideramos una máquina que deja de funcionar para siempre con una tasa λ y lo hace en un tiempo aleatorio T , entonces $T \sim \exp(\lambda)$ y además $T \sim PH_2((1), (-\lambda))$, pues el proceso se extingue (la cadena se absorbe), lo cual describe una distribución de tipo fase.

Ejemplo 3.13 (Distribución Erlang). Consideremos un proceso Poisson de tasa λ y supongamos que $k + 1 \in \mathbb{N}$ es un estado absorbente. Sabemos que existe una colección $\{T_i\}_{i \in \mathbb{N}}$ de variables aleatorias independientes e idénticamente distribuidas $\exp(\lambda)$, las cuales corresponden a los saltos del proceso. Si definimos a $S_n = \sum_{j=1}^n T_j$, entonces sabemos que en el tiempo S_{k+1} , el proceso se absorbe. Por las propiedades del proceso Poisson, tenemos que $S_{k+1} = \sum_{j=1}^{k+1} T_j$, que corresponde a una distribución de $Erlang(k + 1, \lambda)$. Entonces en este ejemplo, S_{k+1} tiene distribución de tipo fase con parámetros $\pi = (1, 0, 0, \dots, 0)$ y

$$\mathbf{T} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots & 0 \\ 0 & -\lambda & \lambda & 0 & \cdots & 0 \\ 0 & 0 & -\lambda & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\lambda \end{pmatrix}$$

Una trayectoria de este proceso se puede ver en la figura 3.2.

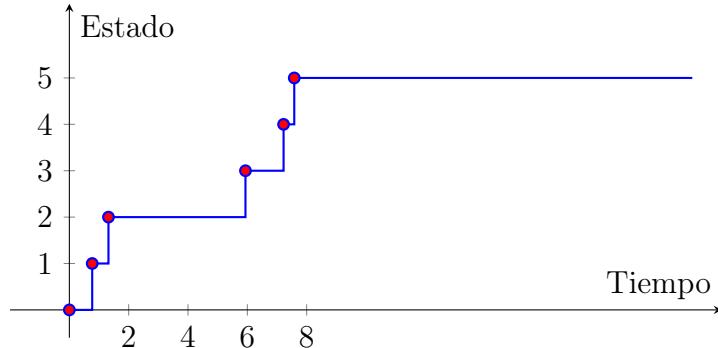


Figura 3.2: Trayectoria de un Proceso de Poisson con estado absorbente en $k + 1 = 5$, $\lambda = \frac{1}{2}$ y $S_5 = 7.5817$.

Ejemplo 3.14 (Suma de Exponentiales). Consideremos variables aleatorias independientes $X_i \sim \exp(\lambda_i)$ con $\lambda_i > 0$. Sea $Z = \sum_{j=1}^{n+1} X_j$ una caminata aleatoria. Podemos pensar a Z como el tiempo que tarda un cliente en pasar por n trámites, donde el tiempo que tarda en hacer el trámite i es una variable aleatoria exponencial de parámetro λ_i y una vez que llega al $n + 1$, termina y puede irse a casa. De esta forma, la

suma de variables exponenciales de distinto parámetro corresponde a una distribución de tipo fase con parámetros $\boldsymbol{\pi} = (1, 0, 0, \dots, 0)$ y

$$T = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_k \end{pmatrix}$$

Se puede ver la figura 3.3 para ver una representación gráfica de la dinámica de este proceso. Podemos notar que este proceso se desarrolla en serie, es decir que va siguiendo cada paso sucesivamente hasta llegar al final y absorberse.

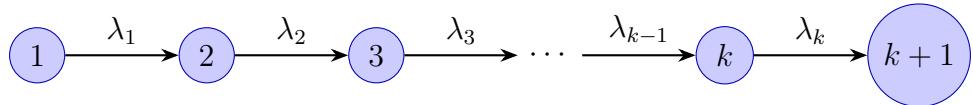


Figura 3.3: Dinámica en serie con estado absorbente en $k + 1$.

En general, sumar (o calcular la convolución) de variables aleatorias exponenciales de distinto parámetro no es tarea sencilla. Si ocurre que $X \sim \exp(\lambda)$ y $Y \sim \exp(\lambda)$, entonces sabemos que $X + Y \sim \text{Gamma}(2, \lambda)$, con función de densidad

$$f_{X+Y}(x) = \lambda^2 x e^{-\lambda x} \quad \text{para } s \geq 0$$

Por otro lado, si $X \sim \exp(\lambda_1)$ y $Y \sim \exp(\lambda_2)$, entonces según [9] ecuación 1,

$$f_{X+Y}(s) = \frac{\lambda_2}{(\lambda_2 - \lambda_1)} \lambda_1 e^{-\lambda_1 s} + \frac{\lambda_1}{(\lambda_1 - \lambda_2)} \lambda_2 e^{-\lambda_2 s} \quad \text{para } s \geq 0$$

Las distribuciones tipo fase son muy útiles para calcular tiempos de extinción de cadenas de Markov a tiempo continuo, donde frecuentemente lidiamos con sumas de este tipo.

Ejemplo 3.15 (Distribución Hiperexponencial). Como antes, consideremos $X_i \sim \exp(\lambda_i)$ con $\lambda_i > 0$ independientes. Supongamos que $\pi_1, \pi_2, \dots, \pi_k$ son probabilidades tales que $\sum_{j=1}^k \pi_j = 1$ y N una variable aleatoria discreta e independiente de lo demás con soporte $\{1, 2, \dots, k\}$ y distribución $(\pi_1, \pi_2, \dots, \pi_k)$. Consideremos a

$$X = \sum_{j=1}^k \mathbf{1}_{\{N=j\}} X_j$$

la cual es conocida como una **mezcla de variables aleatorias exponenciales** y sigue una distribución **hiperexponencial**. Calculemos la distribución y densidad usando la independencia de N y $\{X_i\}_{i \in \mathbb{N}}$,

$$F_X(x) = \mathbb{P} \left[\sum_{j=1}^k \mathbf{1}_{\{N=j\}} X_j \leq x \right] = \sum_{j=1}^k \mathbb{P}[N=j] \mathbb{P}[X_j \leq x] = \sum_{j=1}^k \pi_j F_{X_j}(x)$$

Derivando,

$$f_X(x) = \sum_{j=1}^k \pi_j f_{X_j}(x)$$

En este caso, X también corresponde a una distribución de tipo fase, ya que como cada salto es una variable aleatoria exponencial, el proceso puede verse como una cadena de Markov a tiempo continuo donde se elige en qué estado iniciar siguiendo la ley de probabilidad de $(\pi_1, \pi_2, \dots, \pi_k)$ y se absorbe de acuerdo a la matriz de tasas

$$Q = \begin{pmatrix} -\lambda_1 & 0 & 0 & 0 & \cdots & 0 & \lambda_1 \\ 0 & -\lambda_2 & 0 & 0 & \cdots & 0 & \lambda_2 \\ 0 & 0 & -\lambda_3 & 0 & \cdots & 0 & \lambda_3 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_k & \lambda_k \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

A este proceso también puede pensarse de una manera similar al ejemplo anterior. Si alguien hace un trámite que puede hacerse de diferentes formas, digamos $1, 2, \dots, k$, éste elegirá alguna manera de hacerlo siguiendo la ley de probabilidad de $(\pi_1, \pi_2, \dots, \pi_k)$. Así, habrá elegido cómo hacerlo y, una vez completado, finalizará todo el trámite. Una gráfica de la dinámica de este proceso está descrita en la figura 3.4, en donde podemos notar que ahora es un proceso en paralelo, pues puede elegir a cualquier estado irse y de ese absorberse.

Ejemplo 3.16 (Distribución Coxiana). La distribución Coxiana (o de Cox) es una generalización de los anteriores, pues podemos llegar al estado absorbente desde cualquier estado e igualmente yendo de manera sucesiva. Consideremos una cadena $(X_t : t \geq 0)$ con espacio de estados $\{1, 2, \dots, k, k+1\}$ donde $k+1$ es absorbente y con la siguiente submatriz de tasas:

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & p_1\lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & p_2\lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & -\lambda_{k-2} & p_{k-2}\lambda_{k-2} & 0 \\ 0 & 0 & \cdots & 0 & -\lambda_{k-1} & p_{k-1}\lambda_{k-1} \\ 0 & 0 & \cdots & 0 & 0 & -\lambda_k \end{pmatrix} \quad (3.14)$$

Donde $0 \leq p_i \leq 1$ para toda $i \in \{1, 2, \dots, k\}$. En el caso en que los coeficientes $p_i = 1$ siempre, se obtiene la distribución Erlang. Entonces si empezamos en 1, tenemos que $\boldsymbol{\pi} = (1, 0, 0, \dots, 0)$ y el tiempo de absorción τ sigue una distribución de tipo fase con

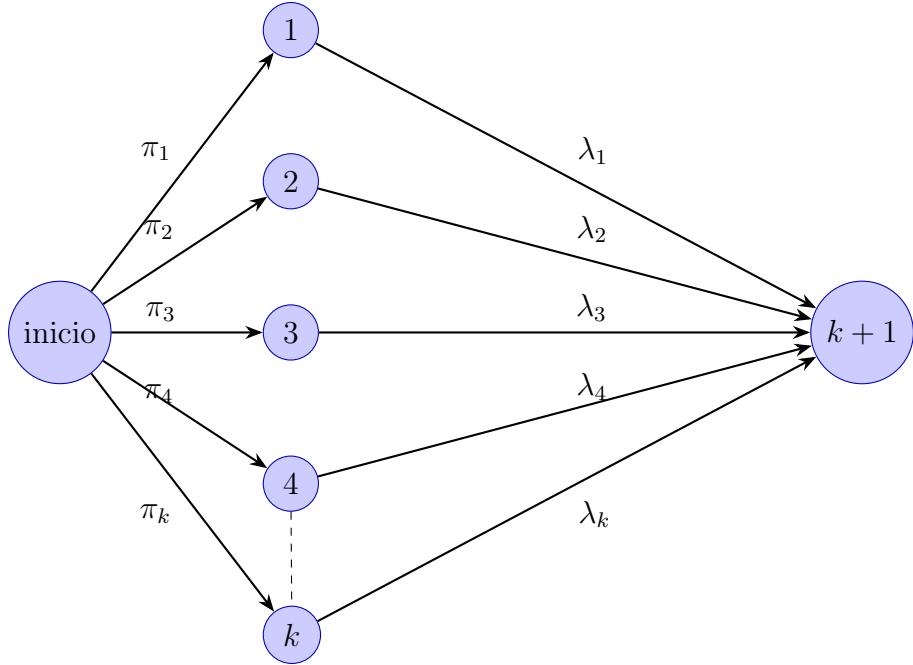


Figura 3.4: Dinámica en paralelo con estado absorbente en $k + 1$.

los parámetros antes mencionados. La dinámica de este proceso está descrita en la figura

En la siguiente sección calcularemos propiedades importantes acerca de las distribuciones tipo fase, como su densidad y distribución. Asimismo, veremos que las fórmulas usuales para la suma de variables aleatorias exponenciales son generalizadas por medio de estas operaciones matriciales, lo cual será muy útil en el futuro.

Densidad y Distribución

En la sección 2.1 estudiamos la matriz de transición $P(u)$ de una cadena de Markov a tiempo continuo. Gracias a 2.6, sabemos que si Q es la matriz de tasas, entonces

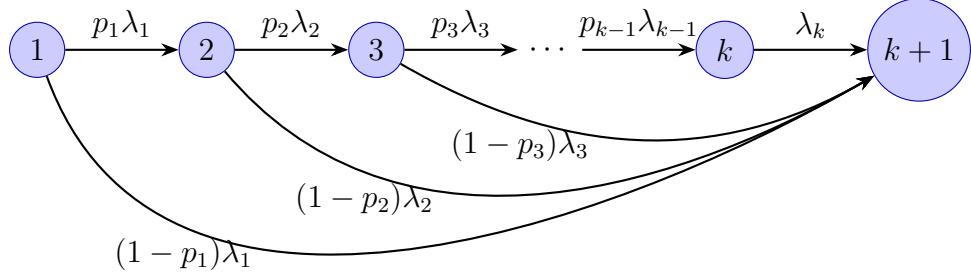
$$P(u) = e^{uQ} \quad \text{para todo } u \geq 0$$

donde

$$e^{uQ} := \sum_{k=0}^{\infty} \frac{(uQ)^k}{k!} \quad \text{para cada } u \geq 0$$

En el siguiente teorema vemos qué pasa si la matriz de tasas tiene la forma 3.12. En lo siguiente y para evitar confusiones, usaremos la notación \exp para referirnos a la función exponencial y la usaremos indistintamente junto con la notación habitual, recordando que e denota al vector columna de tamaño k cuyas entradas son todas 1.

Teorema 3.17. *Si Q corresponde a una matriz de tasas de la forma 3.12, entonces para todo $n \in \mathbb{N}$ y $u \geq 0$, entonces*

Figura 3.5: Dinámica de Cox con estado absorbente en $k + 1$.

$$1. Q^n = \begin{pmatrix} \mathbf{T}^n & -\mathbf{T}^n \mathbf{e} \\ \mathbf{0} & 0 \end{pmatrix}$$

$$2. \exp(uQ) = \begin{pmatrix} \exp(u\mathbf{T}) & \mathbf{e} - \exp(u\mathbf{T})\mathbf{e} \\ 0 & 1 \end{pmatrix}$$

Demuestra. Probaremos de manera inductiva el primer inciso. El caso base es claro. Supongamos que para cierto $n \in \mathbb{N}$ se tiene que

$$Q^n = \begin{pmatrix} \mathbf{T}^n & -\mathbf{T}^n \mathbf{e} \\ \mathbf{0} & 0 \end{pmatrix}$$

Veamos que

$$\begin{aligned} Q^{n+1} &= Q^n Q \\ &= \begin{pmatrix} \mathbf{T}^n & -\mathbf{T}^n \mathbf{e} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{T} & -\mathbf{T} \mathbf{e} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{T}^n \mathbf{T} + 0 & \mathbf{T}^n (-\mathbf{T} \mathbf{e}) + 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{T}^{n+1} & -\mathbf{T}^{n+1} \mathbf{e} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

que es lo que queríamos demostrar. Por otro lado, para $u \geq 0$ se cumple que

$$\begin{aligned} \exp(uQ) &= \sum_{n=0}^{\infty} \frac{u^n Q^n}{n!} \\ &= \mathbf{I}_{k+1} + \sum_{n=1}^{\infty} \frac{u^n Q^n}{n!} \\ &= \mathbf{I}_{k+1} + \left(\begin{array}{cc} \sum_{n=1}^{\infty} \frac{u^n \mathbf{T}^n}{n!} & -\sum_{n=1}^{\infty} \frac{u^n \mathbf{T}^n}{n!} \mathbf{e} \\ \mathbf{0} & 0 \end{array} \right) \\ &= \left(\begin{array}{cc} \mathbf{I}_k + \sum_{n=1}^{\infty} \frac{u^n \mathbf{T}^n}{n!} & -\sum_{n=1}^{\infty} \frac{u^n \mathbf{T}^n}{n!} \mathbf{e} \\ \mathbf{0} & 1 \end{array} \right) \\ &= \begin{pmatrix} \exp(u\mathbf{T}) & \mathbf{e} - \exp(u\mathbf{T})\mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix} \end{aligned}$$

□

Esto significa que siempre que $i, j \in \{1, 2, \dots, k\}$, y $u \geq 0$ entonces $(P(u))_{ij} = (e^{uT})_{ij}$. En particular, como ninguno de esos son el estado absorbente, entonces $[X_u = j] \subseteq [u \leq \tau]$. Así, tenemos que

$$\mathbb{P}_i[X_u = j, u \leq \tau] = (e^{uT})_{ij} \quad \text{para cada } i, j \in \{1, 2, \dots, k\} \quad (3.15)$$

Esto será de utilidad para demostrar el siguiente teorema

Teorema 3.18. *Si $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces para todo $u \geq 0$*

$$F_\tau(u) = 1 - \boldsymbol{\pi} e^{uT} \mathbf{e}.$$

Demuestra. Sean $(X_t : t \geq 0)$ la cadena de Markov asociada a la distribución de tipo fase τ y $u \geq 0$. Sabemos que $[\tau > u] = [X_u = j]$ para algún $j \in \{1, 2, \dots, k\}$. Así, tenemos que:

$$\begin{aligned} 1 - F_\tau(u) &= \mathbb{P}[\tau > u] = \mathbb{P}[X_u = j] = \sum_{i=1}^k \mathbb{P}[X_u = j | X_0 = i] \mathbb{P}[X_0 = i] \\ &= \sum_{i=1}^k (P(u))_{ij} \pi_i = \sum_{i=1}^k (e^{uT})_{ij} \pi_i = \boldsymbol{\pi} e^{uT} \mathbf{e} \end{aligned}$$

□

Por otro lado, para encontrar la densidad podemos derivar la distribución, usando la definición de la exponencial de una matriz y la linealidad de la derivada. Dada $u \geq 0$, tenemos que:

$$f_\tau(u) = \frac{dF_\tau(u)}{du} = \frac{d}{du} [1 - \boldsymbol{\pi} e^{uT} \mathbf{e}] = -\boldsymbol{\pi} \frac{de^{uT}}{du} \mathbf{e}$$

Veamos que:

$$\frac{de^{uT}}{du} = \frac{d}{du} \sum_{k=0}^{\infty} \frac{u^k \mathbf{T}^k}{k!} = \mathbf{T} \sum_{k=1}^{\infty} \frac{ku^{k-1} \mathbf{T}^{k-1}}{k!} = \mathbf{T} \sum_{k=0}^{\infty} \frac{u^k \mathbf{T}^k}{k!} = \mathbf{T} e^{uT}$$

Y usando que $\mathbf{t} = -\mathbf{T} \mathbf{e}$, hemos demostrado el siguiente teorema.

Teorema 3.19. *Si $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces para todo $u \geq 0$*

$$f_\tau(u) = \boldsymbol{\pi} e^{uT} \mathbf{t}$$

Veamos el caso del ejemplo 3.15, donde se definió a la variable aleatoria X y se dieron expresiones para su densidad y distribución. Notemos que en este caso, la submatriz de tasas es diagonal:

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_k \end{pmatrix}$$

Por lo tanto, dados $u \geq 0$ y $j, i \in \{1, 2, \dots, k\}$, al calcular $e^{u\mathbf{T}}$ obtenemos

$$(e^{u\mathbf{T}})_{ij} = \left(\sum_{n=0}^{\infty} \frac{u^n \mathbf{T}^n}{n!} \right)_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ \sum_{n=0}^{\infty} \frac{u^n (-\lambda_i)^n}{n!} & \text{si } i = j \end{cases} = \begin{cases} 0 & \text{si } i \neq j \\ e^{-u\lambda_i} & \text{si } i = j \end{cases}$$

Por lo que la matriz $e^{u\mathbf{T}}$ corresponde a la matriz diagonal donde

$$(e^{u\mathbf{T}})_{ii} = e^{-u\lambda_i}$$

Calculemos la distribución usando la expresión que obtuvimos:

$$\begin{aligned} F_X(u) &= 1 - \boldsymbol{\pi} e^{u\mathbf{T}} \mathbf{e} \\ &= 1 - (\pi_1, \pi_2, \dots, \pi_k) \begin{pmatrix} e^{-u\lambda_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & e^{-u\lambda_2} & 0 & 0 & \cdots & 0 \\ 0 & 0 & e^{-u\lambda_3} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & e^{-u\lambda_k} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \end{aligned}$$

Y es fácil ver que este último producto de matrices nos da como resultado

$$F_X(u) = 1 - \sum_{j=1}^k \pi_j e^{-u\lambda_i}$$

Y usando que como $X_i \sim \exp(\lambda_i)$, entonces $F_{X_i}(u) = e^{-u\lambda_i}$ y por tanto llegamos a la misma expresión que habíamos obtenido en 3.15. El caso de la densidad es parecido, pues:

$$f_X(u) = \boldsymbol{\pi} e^{u\mathbf{T}} \mathbf{t}$$

$$= (\pi_1, \pi_2, \dots, \pi_k) \begin{pmatrix} e^{-u\lambda_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & e^{-u\lambda_2} & 0 & 0 & \cdots & 0 \\ 0 & 0 & e^{-u\lambda_3} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & e^{-u\lambda_k} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix}$$

Nuevamente notamos que al hacer el producto matricial obtenemos que

$$f_X(u) = \sum_{j=1}^k \pi_j \lambda_j e^{-u\lambda_j}$$

Que coincide con lo que habíamos encontrado en 3.19, debido a que $f_{X_i}(u) = \lambda_i e^{-u\lambda_i}$.

A continuación presentamos métodos para calcular los momentos de las distribuciones tipo fase. Como en el caso discreto, tendremos una matriz de Green, donde cada entrada será el valor esperado de tiempo que pasamos en un estado dado que partimos de otro.

3.3. Momentos

En esta sección nos interesaremos por encontrar una expresión para los momentos de una distribución de tipo fase. Como en el caso discreto, habrá una matriz de Green, la cual tendrá propiedades análogas.

Matriz de Green

Para justificar la existencia de la matriz de Green, probaremos el siguiente resultado que es más general que 3.6, pues se basa en cualquier cadena de Markov a tiempo continuo.

Teorema 3.20. *Sea $(X_t : t \geq 0)$ una cadena de Markov a tiempo continuo con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1\}$ y cuya matriz de tasas tiene la forma*

$$Q = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}$$

Entonces son equivalentes:

1. $\{1, 2, \dots, k\}$ son transientes
2. \mathbf{T} es invertible

Observación 3.21. En la definición 3.11, supusimos que todos los estados son transientes; sin embargo, para este teorema no pedimos esto. Podríamos pensar que esta forma específica de la matriz de tasas obliga a que el estado $k + 1$ sea absorbente y lo demás transientes, pero este no es el caso. Un contrajejemplo sencillo es considerar una cadena con espacio de estados $\mathcal{S} = \{1, 2, 3\}$ y con la siguiente matriz de tasas:

$$Q = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \\ \lambda_2 & -\lambda_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Y con distribución inicial $(p, 1-p, 0)$ para alguna $p \in [0, 1]$. Claramente tiene la forma del teorema anterior, pero el estado 3, aunque es absorbente, no es accesible desde ninguno otro y desde la distribución inicial nunca podremos llegar a él.

Demostración. Demostraremos que 2 implica 1, y la prueba del recíproco se puede encontrar en [2], teorema 3.1.11, y la demostración es parecida a 3.1

Supongamos que \mathbf{T} es invertible y consideremos a a_i como la probabilidad de absorción en $k + 1$ empezando del estado i para cada $i \in \mathcal{S}$. Recordemos que, para el esqueleto discreto de la cadena a tiempo continuo, tenemos probabilidades de transición dadas por

$$p(i, j) = \frac{q(i, j)}{q_i}$$

Para cualquier par de estados i, j . Así, dado $i \in \{1, 2, \dots, k\}$ y condicionando al paso siguiente,

$$a_i = p(i, k + 1) + \sum_{\substack{j=1 \\ j \neq i}}^k p(i, j)a_j$$

Pero por definición de las probabilidades de transición, tenemos que

$$p(i, j) = \frac{q(i, j)}{q_i} = -\frac{q(i, j)}{q(i, i)}$$

Sustituyendo tenemos que,

$$a_i = -\frac{q(i, k + 1)}{q(i, i)} + \sum_{\substack{j=1 \\ j \neq i}}^k -\frac{q(i, j)}{q(i, i)}a_j$$

O bien,

$$-q(i, i)a_i = q(i, k + 1) + \sum_{\substack{j=1 \\ j \neq i}}^k q(i, j)a_j$$

Esto es,

$$0 = q(i, k+1) + \sum_{j=1}^k q(i, j)a_i$$

Como esto se cumple para toda $i \in \{1, 2, \dots, k\}$, se sigue que

$$\begin{pmatrix} q(1, k+1) \\ q(2, k+1) \\ \vdots \\ q(k, k+1) \end{pmatrix} + \begin{pmatrix} \sum_{j=1}^k q(1, j)a_j \\ \sum_{j=1}^k q(2, j)a_j \\ \vdots \\ \sum_{j=1}^k q(k, j)a_j \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Sabemos que $\mathbf{t}_i = q(i, k+1)$, pues es el vector de tasas de salida. Con esto, podemos escribir la expresión anterior en notación matricial, haciendo $\mathbf{a} = (a_1, a_2, \dots, a_k)$, se obtiene

$$\mathbf{t} + \mathbf{T}\mathbf{a} = 0$$

Usando que \mathbf{T} es invertible, y como $\mathbf{t} + \mathbf{T}\mathbf{a} = 0$, entonces $\mathbf{T}\mathbf{a} = -\mathbf{t}$ y por tanto

$$\mathbf{a} = \mathbf{T}^{-1}(-\mathbf{t}) = -\mathbf{T}^{-1}\mathbf{t} = -\mathbf{T}^{-1} - \mathbf{T}\mathbf{e} = \mathbf{e}$$

Por lo tanto, para cada $i \in \{1, 2, \dots, k\}$, se tiene que $a_i = 1$, entonces todos estos estados son transientes, pues con probabilidad 1 te absorbes en $k+1$ partiendo de ellos. \square

Corolario 3.22. *Si $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces \mathbf{T} y $-\mathbf{T}$ son invertibles.*

Por tanto, tenemos ahora todas las herramientas para definir la matriz de Green para el caso continuo.

Definición 3.23. *Sea $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, definimos su **matriz de Green asociada** como*

$$U := (-T)^{-1}$$

Como ya habíamos anunciado, esta matriz tiene propiedades análogas a las del caso discreto. Sea $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y supongamos que la variable Z_j denota el tiempo que la cadena subyacente $(X_t : t \geq 0)$ pasa en el estado j , entonces podemos escribirla como

$$Z_j = \int_0^\tau \mathbf{1}_{\{X_t=j\}} dt$$

Nos interesa calcular $\mathbb{E}_i[Z_j]$. Para ello, primero notemos que

$$\int_0^\tau \mathbf{1}_{\{X_t=j\}} dt = \int_0^\infty \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt \quad (3.16)$$

Ya que

$$\mathbf{1}_{\{X_t=j\}} = \mathbf{1}_{\{X_t=j; \tau > t\}} + \mathbf{1}_{\{X_t=j; \tau \leq t\}}$$

pero $\{X_t = j; \tau \leq t\} = Varnothing$ entonces $\mathbf{1}_{\{X_t=j\}} = \mathbf{1}_{\{X_t=j; \tau \geq t\}}$. Luego, tenemos que al integrar respecto de t se tiene que

$$\int_0^\tau \mathbf{1}_{\{X_t=j\}} dt = \int_0^\tau \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt$$

Y por otro lado, observemos que

$$\int_0^\infty \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt = \int_0^\tau \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt + \int_\tau^\infty \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt$$

Pero esta segunda integral es 0, pues para cada $t \geq \tau$, se tiene que $X_t = k+1$ y $j \neq k+1$. Por lo tanto,

$$\int_0^\infty \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt = \int_0^\tau \mathbf{1}_{\{X_t=j\}} dt$$

Usando esto calcularemos $\mathbb{E}_i[Z_j]$. Tenemos que:

$$\mathbb{E}_i[Z_j] = \mathbb{E}_i \left[\int_0^\tau \mathbf{1}_{\{X_t=j\}} dt \right] = \mathbb{E}_i \left[\int_0^\infty \mathbf{1}_{\{X_t=j; \tau \geq t\}} dt \right] \quad (3.17)$$

Como

$$\mathbb{E}_i \left[\mathbf{1}_{\{X_t=j; \tau \geq t\}} \right] = \mathbb{P}_i [X_t = j; \tau \geq t] = \mathbb{P}_i [X_t = j] = p_t(i, j) < \infty$$

y además,

$$\int_0^\infty \mathbf{1}_{\{X_t=j; \tau > t\}} dt = \int_0^\tau \mathbf{1}_{\{X_t=j\}} dt \leq \int_0^\tau dt = \tau < \infty$$

Entonces al aplicar el teorema de Fubini y usar 3.15, podemos escribir a 3.17 como

$$\int_0^\infty \mathbb{E}_i \left[\mathbf{1}_{\{X_t=j; \tau \geq t\}} \right] dt = \int_0^\infty \mathbb{P}_i [X_t = j; \tau \geq t] dt = \int_0^\infty (e^{t\mathbf{T}})_{ij} dt$$

Por otro lado, veamos que:

$$\int_0^\infty (e^{t\mathbf{T}})_{ij} dt = \int_0^\infty \sum_{n=0}^{\infty} \left(\frac{t^n \mathbf{T}^n}{n!} \right)_{ij} dt = \int_0^\infty \sum_{n=0}^{\infty} \frac{t^n}{n!} (\mathbf{T}^n)_{ij}$$

Y la primitiva de dicha función es

$$= \sum_{n=0}^{\infty} \frac{t^{n+1}}{(n+1)!} (\mathbf{T}^n)_{ij} \Big|_{t=0}^{t=\infty} = \sum_{n=1}^{\infty} \frac{t^n}{n!} (\mathbf{T}^{n-1})_{ij} \Big|_{t=0}^{t=\infty}$$

En términos matriciales, usando que como $\mathbf{T}^n = \mathbf{T}\mathbf{T}^{n-1}$ por definición de potencias y \mathbf{T} es invertible, entonces $\mathbf{T}^{-1}\mathbf{T}^n = \mathbf{T}^{n-1}$, se tiene que

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{T}^{n-1} \Big|_{t=0}^{t=\infty} &= \sum_{n=1}^{\infty} \frac{t^n}{n!} (\mathbf{T}^n \cdot \mathbf{T})^{-1} \Big|_{t=0}^{t=\infty} = \left(\sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{T}^n \right) \mathbf{T}^{-1} \Big|_{t=0}^{t=\infty} \\
&= \left(\sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{T}^n - \mathbf{I} \right) \mathbf{T}^{-1} \Big|_{t=0}^{t=\infty} = (e^{t\mathbf{T}} - \mathbf{I}) \mathbf{T}^{-1} \Big|_{t=0}^{t=\infty} = (\mathbf{T}^{-1} e^{t\mathbf{T}} - \mathbf{T}^{-1}) \Big|_{t=0}^{t=\infty}
\end{aligned}$$

Para calcular esto, notemos que si $i, j \in \{1, 2, \dots, k\}$, entonces

$$\mathbb{P}_i[X_t = j] \rightarrow 0 \text{ cuando } t \rightarrow \infty$$

Pues eventualmente la cadena será absorbida, ya que ambos estados son transientes. Esto quiere decir que

$$(e^{t\mathbf{T}})_{ij} \rightarrow 0 \text{ cuando } t \rightarrow \infty$$

O en términos matriciales,

$$e^{t\mathbf{T}} \rightarrow \mathbf{0} \text{ cuando } t \rightarrow \infty$$

Así, tenemos que

$$(\mathbf{T}^{-1} e^{t\mathbf{T}} - \mathbf{T}^{-1}) \Big|_{t=0}^{t=\infty} = (\mathbf{0} - \mathbf{T}^{-1}) - (\mathbf{T}^{-1} - \mathbf{T}^{-1}) = -\mathbf{T}^{-1} = \mathbf{U}$$

Gracias a esto, hemos demostrado el siguiente teorema.

Teorema 3.24. Si $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces U_{ij} representa el valor esperado de tiempo que pasa la cadena subyacente en j dado que partió de i y antes de absorberse; donde i, j son cualesquiera dos estados transientes.

Gracias a esto, tenemos el siguiente resultado.

Teorema 3.25. Si $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces $\mathbb{E}[\tau] = \boldsymbol{\pi} \mathbf{U} \mathbf{e}$

Demostración. La prueba es completamente igual a 3, por lo que se omite. □

Veamos el caso del ejemplo 3.16. **ACÁ ME FALTA UN EJEMPLO**

Vía Transformada de Laplace

A continuación presentaremos una forma explícita de calcular los momentos de una distribución de tipo fase por medio de su transformada de Laplace, la cual es parecido a la función generadora de momentos.

Definición 3.26. Si X es una variable aleatoria no negativa, su transformada de Laplace es una función L_X definida como

$$L_X(s) = \mathbb{E}[e^{-sX}]$$

Para todo $s \geq 0$ tal que dicha esperanza exista.

Si $n \in \mathbb{N}$, entonces al derivar esta función n veces obtenemos

$$L_X^{(n)}(s) = \mathbb{E}[-X^n e^{-sX}]$$

En donde la expresión de la izquierda denota la enésima derivada de L_X . Por tanto, al evaluar en 0 y multiplicar por un factor $(-1)^n$, obtenemos los momentos de X . Es decir, para todo $n \in \mathbb{N}$,

$$(-1)^n L_X^{(n)}(0) = \mathbb{E}[X^n] \quad (3.18)$$

Teorema 3.27. *Si $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, entonces para todo $s \geq 0$,*

$$L_\tau(s) = \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{t} = \boldsymbol{\pi}(s\mathbf{U} + \mathbf{I})^{-1}\mathbf{e}$$

*Demuestra*ción. Primero veamos que se cumple para $s = 0$. Por un lado, claramente $L_\tau(0) = 1$. Por otro lado, al usar la expresión de \mathbf{t} en términos de \mathbf{T} y que $\boldsymbol{\pi}$ es distribución de probabilidad, se tiene que

$$\boldsymbol{\pi}(0\mathbf{I} - \mathbf{T})^{-1}\mathbf{t} = \boldsymbol{\pi}(-\mathbf{T}^{-1})\mathbf{t} = \boldsymbol{\pi}(-\mathbf{T}^{-1})(-\mathbf{T}\mathbf{e}) = \boldsymbol{\pi}\mathbf{e} = \sum_{j=1}^k \boldsymbol{\pi}_j = 1$$

Por lo tanto, la identidad es cierta cuando $s = 0$. Supongamos que $s > 0$ y veamos que:

$$L_\tau(s) = \mathbb{E}[e^{s\tau}] = \int_0^\infty e^{-sx} f_\tau(x) dx = \int_0^\infty e^{-sx} \boldsymbol{\pi} e^{x\mathbf{T}} \mathbf{t} dx = \boldsymbol{\pi} \int_0^\infty e^{-sx} e^{x\mathbf{T}} d\mathbf{x} \quad (3.19)$$

Resolveremos la integral por partes, usando que $e^{x\mathbf{T}} \rightarrow \mathbf{0}$ cuando $x \rightarrow \infty$

$$\begin{aligned} \int_0^\infty e^{-sx} e^{x\mathbf{T}} dx &= -\frac{1}{s} e^{-sx} e^{x\mathbf{T}} \Big|_{x=0}^{x=\infty} - \int_0^\infty -\frac{1}{s} e^{-sx} \mathbf{T} e^{x\mathbf{T}} dx \\ &= \frac{1}{s} \mathbf{I} + \frac{1}{s} \mathbf{T} \int_0^\infty e^{-sx} e^{x\mathbf{T}} dx \end{aligned}$$

Por lo tanto,

$$(s\mathbf{I} - \mathbf{T}) \int_0^\infty e^{-sx} e^{x\mathbf{T}} dx = \mathbf{I}$$

Como ambas matrices son cuadradas, se tiene que

$$\int_0^\infty e^{-sx} e^{x\mathbf{T}} dx = (s\mathbf{I} - \mathbf{T})^{-1} \quad (3.20)$$

Al unir esto con 3.19, se tiene que

$$L_\tau(s) = \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{t}$$

Para verificar la segunda igualdad, notamos que $\mathbf{U}^{-1} = -\mathbf{T}$ y aplicamos el resultado sobre la inversa del producto de matrices,

$$\begin{aligned}
L_\tau(s) &= \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{t} = \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}(-\mathbf{T}\mathbf{e}) = \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T}^{-1})^{-1}(\mathbf{U}^{-1}\mathbf{e}) \\
&= \boldsymbol{\pi}((s\mathbf{I} - \mathbf{T})^{-1}\mathbf{U}^{-1})\mathbf{e} = \boldsymbol{\pi}([\mathbf{U}(s\mathbf{I} - \mathbf{T})]^{-1})\mathbf{e} = \boldsymbol{\pi}([s\mathbf{U} - \mathbf{T}\mathbf{U}]^{-1})\mathbf{e} \\
&= \boldsymbol{\pi}\left((s\mathbf{U} - \mathbf{T}(-\mathbf{T})^{-1})^{-1}\right)\mathbf{e} = \boldsymbol{\pi}(s\mathbf{U} + \mathbf{I})^{-1}\mathbf{e}
\end{aligned}$$

que es lo que deseábamos probar. \square

Con esto tenemos parte de lo necesario para dar una expresión explícita de los momentos de una distribución de tipo fase. Sin embargo, dado que derivaremos una expresión con matrices, nos auxiliaremos de los siguientes resultados, los cuales son presentados en ??.

Lema 3.28. Si $A(s)$ es una matriz cuadrada e invertible cuyas entradas son funciones que dependen de la variable s , entonces

$$\frac{d(A^{-1}(s))}{ds} = -A^{-1}(s) \frac{dA(s)}{ds} A^{-1}(s) \quad (3.21)$$

y

$$\frac{d(A^m(s))}{ds} = \sum_{j=0}^{m-1} A^j(s) \frac{dA(s)}{ds} A^{m-1-j}(s) \quad (3.22)$$

donde entendemos a $\frac{dA(s)}{ds}$ como la matriz tal que

$$\left(\frac{dA(s)}{ds} \right)_{ij} = \frac{dA(s)_{ij}}{ds}$$

Teorema 3.29. Sean $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y $n \in \mathbb{N}$. Entonces

$$\mathbb{E}[\tau^n] = n! \boldsymbol{\pi} \mathbf{U}^n \mathbf{e} \quad (3.23)$$

Demostración. Una vez que ya tenemos la transformada de Laplace, calcular los momentos se resume a aplicar las ideas expuestas en 3.18. Sabemos que:

$$\mathbb{E}[\tau^n] = (-1)^n L_\tau^{(n)}(s) \Big|_{s=0} = (-1)^n \boldsymbol{\pi} \left(\frac{d^n}{ds^n} (s\mathbf{I} - \mathbf{T})^{-1} \right) \Big|_{s=0} \mathbf{t} \quad (3.24)$$

Afirmamos que

$$\frac{d^k}{ds^k} (s\mathbf{I} - \mathbf{T})^{-1} = (-1)^k k! (s\mathbf{I} - \mathbf{T})^{-(k+1)} \text{ para cada } k \in \mathbb{N}$$

Procederemos por inducción sobre k . Veamos que se cumple para $k = 1$. Para empezar, es claro que $\frac{d(s\mathbf{I} - \mathbf{T})}{ds} = \mathbf{I}$, y además, usando 3.21 tenemos que

$$\begin{aligned}
\frac{d}{ds}(s\mathbf{I} - \mathbf{T})^{-1} &= -(s\mathbf{I} - \mathbf{T})^{-1} \frac{d(s\mathbf{I} - \mathbf{T})}{ds} (s\mathbf{I} - \mathbf{T})^{-1} \\
&= -(s\mathbf{I} - \mathbf{T})^{-1} (s\mathbf{I} - \mathbf{T})^{-1} \\
&= (-1)(s\mathbf{I} - \mathbf{T})^{-2}
\end{aligned}$$

Por lo tanto, se cumple si $k = 1$. Ahora supongamos que la afirmación es cierta hasta cierto k . Al aplicar nuevamente 3.21 y la hipótesis de inducción obtenemos que:

$$\begin{aligned}
\frac{d^{k+1}}{ds^{k+1}}(s\mathbf{I} - \mathbf{T})^{-1} &= \frac{d}{ds} \left(\frac{d^k}{ds^k}(s\mathbf{I} - \mathbf{T})^{-1} \right) \\
&= \frac{d}{ds}(-1)^k k! (s - \mathbf{T})^{-(k+1)} \\
&= (-1)^k k! \frac{d}{ds} [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} \\
&= (-1)^k k! - [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} \frac{d(s\mathbf{I} - \mathbf{T})^{k+1}}{ds} [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} \\
&= (-1)^{k+1} k! [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} \frac{d(s\mathbf{I} - \mathbf{T})^{k+1}}{ds} [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1}
\end{aligned}$$

Calculemos la derivada de en medio usando 3.22,

$$\begin{aligned}
\frac{d(s\mathbf{I} - \mathbf{T})^{k+1}}{ds} &= \sum_{j=0}^k (s\mathbf{I} - \mathbf{T})^j \frac{d(s\mathbf{I} - \mathbf{T})}{ds} (s\mathbf{I} - \mathbf{T})^{k-j} \\
&= \sum_{j=0}^k (s\mathbf{I} - \mathbf{T})^j (s\mathbf{I} - \mathbf{T})^{k-j} \\
&= \sum_{j=0}^k (s\mathbf{I} - \mathbf{T})^k \\
&= (k+1)(s\mathbf{I} - \mathbf{T})^k
\end{aligned}$$

Al unir este resultado con lo primero, obtenemos que

$$\begin{aligned}
\frac{d^{k+1}}{ds^{k+1}}(s\mathbf{I} - \mathbf{T})^{-1} &= (-1)^{k+1} k! [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} (k+1)(s\mathbf{I} - \mathbf{T})^k [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} \\
&= (-1)^{k+1} (k+1)! [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} (s\mathbf{I} - \mathbf{T})^k [(s\mathbf{I} - \mathbf{T})^{k+1}]^{-1} \\
&= (-1)^{k+1} (k+1)! (s\mathbf{I} - \mathbf{T})^{-(k+2)}
\end{aligned}$$

Ya que $-(k+1) + k - (k+1) = -(k+2)$. Con esto termina la inducción.

Sustituyendo lo obtenido en 3.24 se tiene que

$$\begin{aligned}
 \mathbb{E}[\tau^n] &= (-1)^n L_\tau^{(n)}(s) \Big|_{s=0} = (-1)^n \boldsymbol{\pi} \left(\frac{d^n}{ds^n} (s\mathbf{I} - \mathbf{T})^{-1} \right) \Big|_{s=0} \mathbf{t} \\
 &= (-1)^n \boldsymbol{\pi} (-1)^n n! (s\mathbf{I} - \mathbf{T})^{-(n+1)} \Big|_{s=0} \mathbf{t} \\
 &= (-1)^{2n} \boldsymbol{\pi} n! (-\mathbf{T})^{-(n+1)} (-\mathbf{T}\mathbf{e}) \\
 &= \boldsymbol{\pi} n! (-\mathbf{T})^{-n} \mathbf{e} \\
 &= \boldsymbol{\pi} n! \mathbf{U}^n \mathbf{e}
 \end{aligned}$$

Tal como buscábamos. \square

Esto nos permite obtener una forma explícita de la varianza de una distribución de tipo fase.

Corolario 3.30. *Sea $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$. Entonces*

$$Var(\tau) = 2\boldsymbol{\pi} \mathbf{U}^2 \mathbf{e} - (\boldsymbol{\pi} \mathbf{U} \mathbf{e})^2$$

Ejemplo 3.31.

ACÁ IGUAL ME GUSTARÍA PONER UN EJEMPLO DE LOS DEL PRINCIPIO, PERO LAS CUENTAS ESTÁN FEAS

Vía Primer Paso

En esta sección daremos una prueba alternativa para obtener una expresión de la esperanza y varianza de una distribución de tipo fase. Esta técnica es ampliamente usada en probabilidad y se conoce como el método de primer paso.

Consideremos $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ con cadena asociada $(X_t : t \geq 0)$. Sean A_i el tiempo hasta la absorción dado que empezamos en i y S_1 el tiempo del primer salto del proceso, entonces al condicionar al primer paso obtenemos la siguiente igualdad

$$\tau = S_1 + \sum_{j \neq i} \mathbf{1}_{\{X_{S_1}=j\}} A_j$$

Si llamamos $M_i = \mathbb{E}_i[\tau]$, entonces al aplicar esperanza dado que empezamos en i , obtenemos que

$$M_i = \frac{1}{-\mathbf{T}_{ii}} + \sum_{j \neq i} \frac{\mathbf{T}_{ij}}{-\mathbf{T}_{ii}} M_j \tag{3.25}$$

De donde

$$-\mathbf{T}_{ii} M_i = 1 + \sum_{j \neq i} \frac{\mathbf{T}_{ij}}{-\mathbf{T}_{ii}} M_j$$

O bien,

$$-\sum_{j \neq i} \mathbf{T}_{ij} M_j = 1$$

Si llamamos $M = (M_1, M_2, \dots, M_k)$, entonces en notación matricial esto es

$$-\mathbf{T}M = \mathbf{e}$$

Por lo tanto, como T es invertible, obtenemos que $M = (-\mathbf{T})^{-1}\mathbf{e} = \mathbf{U}\mathbf{e}$, donde \mathbf{U} es la matriz de Green. Finalmente, dado que la distribución inicial es $\boldsymbol{\pi}$, obtenemos que $\mathbb{E}[\tau] = \boldsymbol{\pi}\mathbf{U}\mathbf{e}$, tal como esperábamos.

Ahora veamos el caso de la varianza. Como ya obtuvimos la expresión de la esperanza, solamente calcularemos el segundo momento de la distribución usando el método del primer paso. Siguiendo con la notación usada antes, definamos $Q_i = \mathbb{E}_i[\tau^2]$, dado que se cumpla la igualdad

$$\tau = S_1 + \sum_{j \neq i} \mathbf{1}_{\{X_{S_1}=j\}} A_j$$

Tenemos que

$$\tau^2 = S_1^2 + 2S_1 \sum_{j \neq i} \mathbf{1}_{\{X_{S_1}=j\}} A_j + \left(\sum_{j \neq i} \mathbf{1}_{\{X_{S_1}=j\}} A_j \right)^2$$

Al aplicar esperanza dado que empezamos en i a esta igualdad, obtenemos que

$$Q_i = \mathbb{E}_i[S_1^2] + 2\mathbb{E}_i[S_1] \sum_{j \neq i} \frac{\mathbf{T}_{ij}}{-\mathbf{T}_{ii}} M_j + \sum_{j \neq i} \frac{\mathbf{T}_{ij}}{-\mathbf{T}_{ii}} Q_j$$

Al reordenar y usar que $\mathbb{E}[S_1] = \frac{1}{-\mathbf{T}_{ii}}$ y $\mathbb{E}[S_1^2] = \frac{2}{(-\mathbf{T}_{ii})^2}$ ya que $S_1 \sim \exp(\mathbf{T}_{ii})$ dado que empezamos en i , se tiene que

$$\mathbf{T}_{ii} Q_i + \sum_{j \neq i} \mathbf{T}_{ij} Q_j = \frac{2}{\mathbf{T}_{ii}} + 2 \sum_{j \neq i} \frac{\mathbf{T}_{ij}}{\mathbf{T}_{ii}} M_j = 2 \left(\frac{1}{\mathbf{T}_{ii}} + \sum_{j \neq i} \frac{\mathbf{T}_{ij}}{\mathbf{T}_{ii}} M_j \right) = -2M_i$$

Pues el paréntesis tiene la expresión de M_i que obtuvimos en 3.25. Por lo tanto, tenemos que si $Q = (Q_1, Q_2, \dots, Q_n)$, entonces

$$\mathbf{T}Q = -2M = -2\mathbf{U}\mathbf{e}$$

Y por tanto $Q = 2\mathbf{T}^{-1}\mathbf{U}\mathbf{e}$ y por definición de la matriz de Green, esto quiere decir que $Q = 2\mathbf{U}^2\mathbf{e}$. Por lo tanto, usando que la distribución inicial es $\boldsymbol{\pi}$, se obtiene la fórmula enunciada en 3.30,

$$\text{Var}(\tau) = 2\boldsymbol{\pi}\mathbf{U}^2\mathbf{e} - (\boldsymbol{\pi}\mathbf{U}\mathbf{e})^2$$

El método del primer paso es muy usado en probabilidad debido a su simplicidad y capacidad. Esta forma de calcular esperanzas será nuestra manera de calcular covarianzas al abordar el caso multivariado de las distribuciones tipo fase.

3.4. Convoluciones y Mezclas

En muchos contextos de genética de poblaciones resulta natural considerar no sólo una variable aleatoria con distribución de tipo fase, sino también sumas o combinaciones ellas. Por ello es importante comprender cómo se comportan las convoluciones y mezclas de distribuciones tipo fase, pues estas operaciones permiten hablar de nuevos modelos. Como veremos, las distribuciones de tipo fase son cerradas bajo estas transformaciones, por lo que la teoría acerca de éstas seguirá manteniéndose y será sencillo estudiarlas.

A continuación presentamos un teorema, cuya demostración será similar a aquella del teorema 3.10, por medio de una razonamiento probabilista.

Teorema 3.32. *Sean $\tau_1 \sim PH_k(\boldsymbol{\alpha}, \mathbf{S})$ y $\tau_2 \sim PH_m(\boldsymbol{\beta}, \mathbf{T})$ variables aleatorias independientes. Entonces*

$$\tau_1 + \tau_2 \sim PH_{k+m} \left((\boldsymbol{\alpha}, \mathbf{0}), \begin{pmatrix} \mathbf{S} & \mathbf{s}\boldsymbol{\beta} \\ \mathbf{0} & \mathbf{T} \end{pmatrix} \right)$$

Demostración. Sea $(Z_t : t \geq 0)$ la cadena de Markov a tiempo continuo con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1, \dots, k+m, k+m+1\}$ y con la siguiente matriz de tasas

$$Q = \begin{pmatrix} \mathbf{S} & \mathbf{s}\boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{t} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix} \quad (3.26)$$

y con distribución inicial $\boldsymbol{\pi} = (\boldsymbol{\alpha}, \mathbf{0})$. Es decir,

$$\boldsymbol{\pi}_i = \mathbb{P}[Z_0 = i] = \begin{cases} \boldsymbol{\alpha}_i & \text{si } i \in \{1, 2, \dots, k\} \\ 0 & \text{en otro caso} \end{cases}$$

Este proceso comienza su desarrollo en el conjunto $\{1, 2, \dots, k\}$, para después dar un salto hacia $\{k+1, k+2, \dots, k+m\}$ en un tiempo τ_1 . Notemos que si $i \in \{1, 2, \dots, k\}$ y $k+j \in \{k+1, k+2, \dots, k+m\}$, entonces por construcción de Q,

$$q(i, k+j) = Q_{ij} = \mathbf{s}_i \boldsymbol{\beta}_j$$

Esto tiene sentido, pues la tasa de salida del estado i es de \mathbf{s}_i , mientras que la probabilidad de que el salto sea a $k+j$ es $\boldsymbol{\beta}_j$. Después, el proceso sigue evolucionando de acuerdo a la matriz de tasas

$$\begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}$$

Hasta absorberse en el estado $k+m+1$ después de un tiempo τ_2 . Por tanto, el proceso $(Z_t : t \geq 0)$ se absorbe en un tiempo $\tau_1 + \tau_2$, que es lo que queríamos demostrar. \square

Algo natural es preguntarse si hay alguna manera en que la matriz de tasas de alguna transformación de τ_1 y τ_2 , definidos como en el teorema anterior, se exactamente la matriz

$$Q = \begin{pmatrix} \mathbf{S} & 0 \\ 0 & \mathbf{T} \end{pmatrix}$$

Para eso, introducimos el concepto de mezclas de variables aleatorias, el cual será de mucha utilidad de ahora en adelante.

Definición 3.33. Sean $r \in \mathbb{N}$ y $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_r)$ una distribución de probabilidad. Supongamos que para cada $i \in \{1, 2, \dots, r\}$ se tiene $\tau_i \sim PH_{k_i}(\boldsymbol{\pi}_i, \mathbf{T}_i)$. Una variable aleatoria Z tiene distribución **mezcla de tipo fase** si su función de densidad es:

$$f_Z(x) = \sum_{i=1}^r \pi_i f_{\tau_i}(x)$$

Notamos que este concepto generaliza el ejemplo 3.15. A continuación demostraremos que una mezcla de tipo fase sigue teniendo distribución de tipo fase. Empezaremos con el caso $r = 2$, y usaremos el siguiente lema para dar una interpretación de este las distribuciones mezcla de tipo fase en este caos.

Lema 3.34. *QUISIERA DEMOSTRAR QUE ESTO ES UN SI Y SOLO SI* Sean $\tau_1 \sim PH_k(\boldsymbol{\alpha}, \mathbf{S})$ y $\tau_2 \sim PH_m(\boldsymbol{\beta}, \mathbf{T})$ variables aleatorias independientes. Si U es tal que

$$U = \begin{cases} \tau_1 & \text{con probabilidad } p \\ \tau_2 & \text{con probabilidad } 1 - p \end{cases}$$

entonces U es una mezcla de τ_1 y τ_2 .

Demostración. Sean $a \leq b$. Sabemos que la función de densidad de U debe cumplir que,

$$\mathbb{P}[a \leq U \leq b] = \int_a^b f_U(x) dx$$

Probaremos que $\mathbb{P}[a \leq U \leq b] = \int_a^b p f_{\tau_1}(x) + (1 - p) f_{\tau_2}(x) dx$. Al usar probabilidad total:

$$\begin{aligned} \mathbb{P}[a \leq U \leq b] &= \mathbb{P}[a \leq U \leq b | U = \tau_1] \mathbb{P}[U = \tau_1] + \mathbb{P}[a \leq U \leq b | U = \tau_2] \mathbb{P}[U = \tau_2] \\ &= p \mathbb{P}[a \leq \tau_1 \leq b] + (1 - p) \mathbb{P}[a \leq \tau_2 \leq b] \\ &= p \int_a^b f_{\tau_1}(x) dx + (1 - p) \int_a^b f_{\tau_2}(x) dx \\ &= \int_a^b p f_{\tau_1}(x) + (1 - p) f_{\tau_2}(x) dx \end{aligned}$$

que es lo que queríamos demostrar. \square

Teorema 3.35. Sean $\tau_1 \sim PH_k(\boldsymbol{\alpha}, \mathbf{S})$ y $\tau_2 \sim PH_m(\boldsymbol{\beta}, \mathbf{T})$ variables aleatorias independientes y $p \in [0, 1]$. Entonces si la variable aleatoria U tiene como función de densidad

$$f_U(x) = pf_{\tau_1}(x) + (1-p)f_{\tau_2}(x)$$

entonces U tiene distribución de tipo fase con vector inicial $(p\boldsymbol{\alpha}, (1-p)\boldsymbol{\beta})$ y con submatriz de tasas

$$Q = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{pmatrix}$$

Demostración. Nuevamente haremos un razonamiento probabilista. Sea $(Z_t : t \geq 0)$ una cadena de Markov a tiempo continuo con espacio de estados $\mathcal{S} = \{1, 2, \dots, k, k+1, k+2, \dots, k+m, k+m+1, k+m+2\}$, con distribución inicial $\boldsymbol{\pi}$ definida como

$$\boldsymbol{\pi}_i = \begin{cases} p\boldsymbol{\alpha}_i & \text{si } i \in \{1, \dots, k\} \\ (1-p)\boldsymbol{\beta}_i & \text{si } i \in \{k+2, \dots, k+m+1\} \\ 0 & \text{si } i \in \{k+1, k+m+2\} \end{cases}$$

y matriz de tasas

$$\begin{pmatrix} \mathbf{S} & \mathbf{0} & \mathbf{s} \\ \mathbf{0} & \mathbf{T} & \mathbf{t} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Observemos que $\boldsymbol{\pi}$ es distribución de probabilidad, pues claramente sus entradas son no negativas y además

$$\sum_{i \in S} \boldsymbol{\pi}_i = \sum_{i \in S} p\boldsymbol{\alpha}_i + (1-p)\boldsymbol{\beta}_i = p \sum_{i \in S} \boldsymbol{\alpha}_i + (1-p) \sum_{i \in S} \boldsymbol{\beta}_i = p + 1 - p = 1$$

Para analizar la dinámica de este proceso tenemos los siguientes casos:

- Con probabilidad p , la cadena empieza en alguno de los estados $\{1, \dots, k\}$ y eventualmente se absorbe en $k+1$ en un tiempo τ_1
- Con probabilidad $1-p$, la cadena empieza en alguno de los estados $\{k+2, \dots, k+m+1\}$ y eventualmente se absorbe en $k+m+2$ en un tiempo τ_2

Por lo tanto, U es una mezcla de tipo fase, gracias a que

$$U = \begin{cases} \tau_1 & \text{con probabilidad } p \\ \tau_2 & \text{con probabilidad } 1-p \end{cases}$$

y al lema 3.34. Además, U corresponde al tiempo de absorción de la cadena $(Z_t : t \geq 0)$, que es lo que queríamos probar. **NO SÉ SI ESTA PRUEBA ESTÉ BIEN HECHA** :(□

Dado que esta variable vuelve a ser de tipo fase, entonces tenemos que las distribuciones de tipo fase son cerradas bajo convoluciones y mezclas de dos elementos. De manera inductiva, podemos demostrar el siguiente corolario.

Corolario 3.36. *Las distribuciones de tipo fase son cerradas bajo mezclas finitas*

Esto presenta una gran ventaja para las distribuciones, pues en general, la suma no es cerrada para la mayoría de distribuciones. Además, los modelos en genética de poblaciones que consideren sumas o mezclas de este tipo de variables, tal es el caso de las medidas de árboles coalescentes, seguirán siendo de tipo fase y podremos estudiar sus propiedades con la teoría aquí desarrollada.

3.5. Transformaciones Via Recompensas

Finalmente, abordaremos las transformaciones via recompensas, que son una de las variantes más importantes y útiles que tienen las distribuciones tipo fase para modelos en genética de poblaciones. En los árboles coalescentes, por ejemplo, la altura, el largo total de las ramas y otras medidas importantes se obtienen de esta forma de distribuciones tipo fase.

La idea es simple: asignar una recompensa no negativa a cada uno de los estados de la cadena de Markov asociada a una distribución de tipo fase. Con esto en mente, nos interesará conocer la recompensa total hasta la absorción, la cual también tiene distribución de tipo fase, y eso es lo que se presenta en esta sección.

Definición 3.37. *Sean $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y $(X_t : t \geq 0)$ la cadena de Markov asociada. Consideremos un vector $\mathbf{r} = (r(1), r(2), \dots, r(k))$ con entradas no negativas y al cual llamaremos vector de recompensas. Definimos a la variable aleatoria Y como sigue*

$$Y = \int_0^\tau r(X_t) dt$$

Entenderemos a Y como la recompensa total del proceso hasta el momento de absorción.

Observación 3.38. Observemos que si $\mathbf{r} = (1, 1, \dots, 1)$, entonces la variable Y corresponde al tiempo transcurrido hasta absorberse, es decir que $Y = \tau$. Por otro lado, si j es un estado cualquiera y las recompensas están dadas por $\mathbf{r}(i) = \mathbf{1}_{\{j\}}(i)$, entonces $Y = Z_j$, como se definió en 3.16.

El siguiente lema será utilizado en la sección siguiente.

Lema 3.39. *Si $X \sim \exp(\lambda)$ y $c > 0$, entonces $cX \sim \exp\left(\frac{\lambda}{c}\right)$*

En este mismo contexto, nos interesa estudiar la distribución de la recompensa total de un proceso hasta el momento de absorción como antes.

Recompensas Positivas

Sea $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y $(X_t : t \geq 0)$ la cadena de Markov a tiempo continuo asociada. Consideremos un vector de recompensas estrictamente positivas $\mathbf{r} = (r(1), r(2), \dots, r(k))$ como en la definición anterior y a la variable Y como la recompensa total del proceso. Por las propiedades de la cadena, sabemos que su espacio de estados es $\{1, 2, \dots, k-1, k\}$ y k es absorbente. Para cada $j \in \{1, 2, \dots, k-1\}$, definamos

$$E_j = \{X_t = j\} \subseteq (0, \infty)$$

Usando esta notación, si λ es la medida de Lebesgue, entonces para todo $j \in \{1, 2, \dots, k-1\}$, se tiene que $\lambda(E_j) = S_j$ es el tiempo de permanencia en el estado j , y sabemos que $S_j \sim \exp(-\mathbf{T}_{jj})$. Con esto en mente, tenemos que para cada $t \geq 0$, se cumple que

$$X_t = \sum_{j=1}^{k-1} j \mathbf{1}_{E_j}(t) + k \mathbf{1}_{(0, \infty) \setminus \bigcup_{j=1}^{k-1} E_j}(t)$$

Por lo tanto, para la recompensa total del proceso se tiene que

$$Y = \int_0^\tau r(X_t) dt = \int_0^\tau \sum_{j=1}^{k-1} r(j) \mathbf{1}_{E_j}(t) dt = \sum_{j=1}^{k-1} r(j) \int_0^\tau \mathbf{1}_{E_j}(t) dt = \sum_{j=1}^{k-1} r(j) \lambda(E_j) = \sum_{j=1}^{k-1} r(j) S_j$$

Usando el lema anterior, tenemos que para cada $j \in \{1, 2, \dots, k-1\}$,

$$r(j) S_j \sim \exp\left(-\frac{\mathbf{T}_{jj}}{r(j)}\right)$$

Lo cual es válido pues las recompensas no son cero. Por lo tanto, Y tiene distribución de tipo fase, ya que es suma de exponenciales de distintos parámetros. Además, notamos que la distribución inicial sigue siendo exactamente igual, ya que no interferimos en cómo inicia este experimento. Con este análisis hemos demostrado el siguiente teorema.

Teorema 3.40. *Sean $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$, $(X_t : t \geq 0)$ la cadena de Markov asociada, un vector de recompensa $\mathbf{r} = (r(1), r(2), \dots, r(k))$ y la recompensa total del proceso Y . Entonces $Y \sim PH_k(\boldsymbol{\pi}, \Lambda^{-1}(\mathbf{r})\mathbf{T})$ donde*

$$\Lambda_{ij}^{-1}(\mathbf{r}) = \begin{cases} \frac{1}{r(i)} & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$$

Este teorema es de suma utilidad, ya que algunas de las variables asociadas a procesos coalescentes corresponden a distribuciones tipo fase bajo transformaciones con recompensas, por lo que saber que estas tienen su misma distribución será de mucha utilidad. Aunque ya sabemos cuál es la esperanza de las distribuciones tipo fase en general, haremos un análisis con el método del primer paso para deducir la esperanza de la recompensa total del proceso, lo cual será útil en la siguiente sección.

Recompensas Positivas Vía Primer Paso

Sea $\tau \sim PH_k(\cdot, \mathbf{T})$ con cadena asociada $(X_t : t \geq 0)$, un vector de recompensas estrictamente positivas $\mathbf{r} = (r(1), r(2), \dots, r(k))$ y Y la recompensa total del proceso. Supongamos que el primer salto del proceso es en un tiempo aleatorio S_1 y que $A(i) = \mathbb{E}_i(Y)$. Condicionando al primer paso obtenemos la siguiente igualdad:

$$\begin{aligned} \int_0^\tau r(X_t) \mathbf{1}_{\{X_0=1\}} dt &= \int_0^{S_1} r(X_t) \mathbf{1}_{\{X_0=1\}} dt + \sum_{j \neq i} \int_{S_1}^\tau r(X_t) \mathbf{1}_{\{X_{S_1}=j\}} dt \\ &= r(i)S_1 + \sum_{j \neq i} \mathbf{1}_{\{X_{S_1}=j\}} \int_{S_1}^\tau r(X_t) dt \end{aligned}$$

Por lo tanto, al aplicar esperanza empezando desde i , se tiene que

$$\begin{aligned} A(i) &= r(i) \frac{1}{-T_{ii}} + \sum_{k \neq i} \mathbb{P}[X_{S_1} = k] A(k) \\ &= r(i) \frac{1}{-T_{ii}} + \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{\mathbf{T}_{ii}} A(k) \end{aligned}$$

Y al reordenar obtenemos que

$$-\mathbf{T}_{ii}A(i) = r(i) + \sum_{k \neq i} \mathbf{T}_{ik}A(k) \quad (3.27)$$

Por lo tanto, si llamamos $\mathbf{A} = (A(1), A(2), \dots, A(k))$, entonces en notación matricial tenemos que

$$(-\mathbf{T})\mathbf{A} = \mathbf{r} = \Lambda(\mathbf{r})\mathbf{e}$$

O bien,

$$\mathbf{A} = (-\mathbf{T})^{-1}\Lambda(\mathbf{r})\mathbf{e}$$

Donde

$$\Lambda_{ij}^{(\mathbf{r})} = \begin{cases} r(i) & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, usando la distribución inicial y llamando $\mathbf{U} = (-\mathbf{T})^{-1}$, obtenemos un resultado que dejamos plasmado en el siguiente corolario.

Corolario 3.41. *Sea $Y \sim PH_k(\boldsymbol{\pi}, \Lambda^{-1}(\mathbf{r})\mathbf{T})$ la recompensa total asociada a una transformación via recompensas y $\mathbf{U} = (-\mathbf{T})^{-1}$. Entonces*

- $\mathbb{E}[Y] = \boldsymbol{\pi}\mathbf{U}\Lambda(\mathbf{r})\mathbf{e}$
- $Var(Y) = 2\boldsymbol{\pi}(\mathbf{U}\Lambda(\mathbf{r}))^2\mathbf{e} - (\boldsymbol{\pi}\mathbf{U}\Lambda(\mathbf{r})\mathbf{e})^2$

Recompensas No Negativas

Como en el análisis anterior, consideremos $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y $(X_t : t \geq 0)$ la cadena de Markov a tiempo continuo asociada. Sea $\mathbf{r} = (r(1), r(2), \dots, r(k))$ un vector de recompensas no negativas y Y la recompensa total del proceso. Asimismo, el espacio de estados es $\mathcal{S} = \{1, 2, \dots, k-1, k\}$ y k es absorbente. Definamos a los conjuntos

$$S^+ = \{s \in \mathcal{S} : r(s) > 0\} \quad \text{y} \quad S^0 = \{s \in \mathcal{S} : r(s) = 0\}$$

que corresponden a los estados con recompensa positiva y aquellos con recompensa nula, respectivamente. A su vez, sea $(Y_n : n \in \mathbb{N})$ el esqueleto discreto asociado a la cadena. Definamos a la matriz \mathbf{P} como aquella tal que

$$\mathbf{P}_{ij} = -\frac{\mathbf{T}_{ij}}{\mathbf{T}_{ii}} \quad \text{para} \quad i \neq j$$

Supongamos que para cada $i \in S$,

$$p_i = 1 - \sum_{j=1}^k \mathbf{T}_{ik}$$

es la probabilidad de ir del estado i al estado absorbente y $\mathbf{p} = (p_1, p_2, \dots, p_{k-1})$. Así, tenemos que la matriz

$$\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{P} & \mathbf{p} \\ \mathbf{0} & 1 \end{pmatrix}$$

Es la matriz de transición del esqueleto discreto. Al hacer un reacomodo de índices, podemos pensar que la matriz $\hat{\mathbf{P}}$ puede expresarse como

$$\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{P}^{++} & \mathbf{P}^{+0} \\ \mathbf{P}^{0+} & \mathbf{P}^{00} \end{pmatrix}$$

Donde \mathbf{P}^{++} es una matriz cuyas filas como columnas pertenecen a S^+ , en \mathbf{P}^{+0} las filas son elementos de S^+ y las columnas de S^0 y es análogo para \mathbf{P}^{0+} y \mathbf{P}^{00} . Y a su vez, podemos definir

$$\hat{\boldsymbol{\pi}} = (\boldsymbol{\pi}^+, \boldsymbol{\pi}^0)$$

donde $\boldsymbol{\pi}^+$ y $\boldsymbol{\pi}^0$ son las distribuciones iniciales de estados en S^+ y en S^0 , respectivamente.

Con estas definiciones, consideremos una cadena de Markov a tiempo discreto $(Z_n : n \in \mathbb{N})$ definida en $S^* = S^+ \cup \{k+1\}$ y con distribución inicial $\boldsymbol{\pi}^+$ definida como $Z_n = Y_{M_n}$ donde M_n es la enésima visita de la cadena $(Y_n : n \in \mathbb{N})$ a S^+ . Es decir, la nueva cadena imita el comportamiento de la primera en S^+ . Así, tenemos que la matriz de transición de $(Z_n : n \in \mathbb{N})$ está dada por

$$\mathbf{P}^* = \begin{pmatrix} \mathbf{F} & \mathbf{f} \\ \mathbf{0} & 1 \end{pmatrix}$$

Donde \mathbf{F} es una submatriz de transición de tamaño $|S^+| \times |S^+|$ y el vector \mathbf{f} , de $|S^+| \times 1$, corresponde a las probabilidades de absorberse. Luego, para estados $i, j \in \mathcal{S}$ adoptaremos la siguiente notación

- p_{ij}^{++} será la probabilidad de transición de i a j cuando $i, j \in S^+$
- p_{ij}^{+0} será la probabilidad de transición de i a j cuando $i \in S^+$ y $j \in S^0$
- p_{ij}^{0+} será la probabilidad de transición de i a j cuando $i \in S^0$ y $j \in S^+$
- p_{ij}^{00} será la probabilidad de transición de i a j cuando $i, j \in S^0$
- En general, al utilizar el superíndice $(p^{00})_{k\ell}^{(n)}$, esto será la probabilidad de transición de i a j en n pasos.

Por lo tanto, es fácil ver que

$$f_{ij} = p_{ij}^{++} + \sum_{k \in S^0} p_{ik}^{+0} \sum_{n=0}^{\infty} \sum_{\ell \in S^0} (p^{00})_{k\ell}^{(n)} p_{\ell j}^{0+}$$

Ya que el primer sumando representa el caso donde vamos de i a j sin salir de S^+ , y el segundo representa los casos en los que llegas a estados de S^0 y estas en este conjunto hasta llegar a j . Es decir,

$$f_{ij} = p_{ij}^{++} + \mathbb{P} \left[\bigcup_{k \in S^0} \bigcup_{n=0}^{\infty} \bigcup_{l \in S^0} \{[\text{Ir de } i \text{ a } k] \cap [\text{Ir de } k \text{ a } l \text{ en } n \text{ pasos}] \cap [\text{Ir de } l \text{ a } j]\} \right]$$

En notación matricial, esto significa que

$$\mathbf{F} = \mathbf{P}^{++} + \mathbf{P}^{+0} \sum_{n=0}^{\infty} (\mathbf{P}^{00})^n \mathbf{P}^{0+} = \mathbf{P}^{++} + \mathbf{P}^{+0} (\mathbf{I} - \mathbf{P}^{00})^{-1} \mathbf{P}^{0+}$$

Pues por el análisis realizado en 3.9, sabemos que $\sum_{n=0}^{\infty} (\mathbf{P}^{00})^n = (\mathbf{I} - \mathbf{P}^{00})^{-1}$. Por otro lado, como \mathbf{f} corresponde al vector de salidas de una distribución de tipo fase discreta, entonces se tiene que $\mathbf{f} = \mathbf{e} - \mathbf{Fe}$.

De manera similar, hemos dividido la distribución inicial $\boldsymbol{\pi} = (\boldsymbol{\pi}^+, \boldsymbol{\pi}^0)$ y por un argumento análogo al anterior, tenemos que el proceso $(Z_n : n \in \mathbb{N})$ tiene como distribución inicial a $\boldsymbol{\alpha}$ definida como:

$$\boldsymbol{\alpha} = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0 (\mathbf{I} - \mathbf{P}^{00})^{-1} \mathbf{P}^{0+}$$

Sin embargo, notemos que el proceso original podría empezar en un estado de S^0 y absorberse sin haber pasado por alguno de S^+ , en este caso, la cadena $(Z_n : n \in \mathbb{N})$ tendrá probabilidad positiva de empezar en el estado $k+1$. En este caso, supondremos que $\alpha_{|S^+|+1} = 1 - \boldsymbol{\alpha}\mathbf{e}$ es dicha probabilidad. Con esto, hemos descrito completamente

a la cadena $(Z_n : n \in \mathbb{N})$.

Ahora describiremos la cadena de Markov a tiempo continuo que reflejará el comportamiento de las recompensas del proceso. Cuando $Z_n = i$, entonces el proceso $(X_t : t \geq 0)$ se queda en dicho estado un tiempo exponencial de parámetro $-\mathbf{T}_{ii}$; asimismo, adquiere una recompensa de tamaño $r(i)$, entonces por el mismo argumento que en el caso positivo, la recompensa total adquirida mientras $Z_n = i$ es exponencial de parámetro $-\frac{\mathbf{T}_{ii}}{r(i)}$. Si construimos una cadena de Markov a tiempo continuo con estas tasas de salto, obtenemos un proceso con matriz de tasas

$$\Lambda^* = \begin{pmatrix} \mathbf{T}^* & \mathbf{t}^* \\ \mathbf{0} & 0 \end{pmatrix}$$

donde

$$\mathbf{T}_{ij}^* = -\frac{\mathbf{T}_{ii}}{r(i)} f_{ij} \quad \text{y} \quad \mathbf{T}_i^* = -\frac{\mathbf{T}_{ii}}{r(i)} f_i \quad \text{para cada } i, j \in S^+.$$

Observemos que la construcción del proceso $(Z_n : n \in \mathbb{N})$ admite saltos de un estado a sí mismo (esto puede ocurrir para algún estado $i \in S^+$ si $X_n = i$ y después $X_{n+1}, X_{n+2}, \dots, X_{n+k} \in S^0$ para alguna $k \in \mathbb{N}$, y finalmente $X_{n+k+1} = i$), lo cual es un problema para el nuevo proceso, ya que es a tiempo continuo. Por esta razón, eliminaremos las transiciones de estados a sí mismos del proceso $(Z_n : n \in \mathbb{N})$, el cual es un procedimiento análogo al adelgazamiento de un Proceso Poisson.

Gracias a este análisis, hemos demostrado el siguiente teorema

Teorema 3.42. *Sean $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y $(X_t : t \geq 0)$ la cadena de Markov asociada y $\mathbf{r} = (r(1), r(2), \dots, r(k))$ un vector de recompensas no negativas. Si Y es la recompensa total del proceso, entonces corresponde a una mezcla con tamaño α_{d+1} en 0 y una distribución de tipo fase con parámetros $(\boldsymbol{\alpha}, \mathbf{T}^*)$ donde $d = |S^+|$ y*

$$\mathbf{T}_{ij}^* = -\frac{\mathbf{T}_{ii}}{r(i)} f_{ij} \quad \text{para } i \neq j \quad \text{y} \quad \mathbf{T}_i^* = -\frac{\mathbf{T}_{ii}}{r(i)} f_i$$

Este teorema será muy útil en el futuro, pues nos interesará hablar sobre transformaciones de recompensas que correspondan a indicadoras, y podremos restringirnos solamente a analizar la cadena de Markov en aquel estado donde esté concentrada la masa total de la recompensa.

3.6. Distribuciones Tipo Fase Multivariadas

En esta sección consideraremos la distribución conjunta de una colección finita de variables aleatorias de tipo fase, introduciremos la distribución de tipo fase multivariada y hablaremos acerca de sus propiedades, como la covarianza entre variables.

Definición 3.43. Sean $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ y $(X_t : t \geq 0)$ la cadena correspondiente. Sea $m \in \mathbb{N}$ y para cada $j \in \{1, 2, \dots, m\}$, un vector de recompensas de la forma $\mathbf{r}_j = (r_j(1), r_j(2), \dots, r_j(k))$. Denotaremos por \mathbf{R} a la matriz de $k \times m$ tal que

$$R_{ij} = r_j(i)$$

Definimos a la variable aleatoria Y_j como

$$Y_j = \int_0^\tau r_j(X_t) dt = \int_0^\tau \mathbf{R}_{X_t, j} dt$$

Decimos que el vector aleatorio $\mathbf{Y} = (Y_1, \dots, Y_m)$ tiene **distribución de tipo fase multivariada** de parámetros $\boldsymbol{\pi}$, \mathbf{T} y \mathbf{R} . Lo denotamos $\mathbf{Y} \sim MPH(\boldsymbol{\pi}, \mathbf{T}, \mathbf{R})$.

En el contexto de procesos de Markov, las distribuciones tipo fase multivariadas pueden interpretarse como el tiempo total que un proceso pasa en distintas “regiones” del espacio de estados antes de la absorción.

3.6.1. Covarianzas entre las variables

En esta sección obtendremos una expresión para la covarianza entre variables que correspondan a la recompensa total del proceso para la misma distribución de tipo fase τ y con dos vectores de recompensa $\mathbf{r}_1 = (r_1(1), \dots, r_1(k))$, $\mathbf{r}_2 = (r_2(1), \dots, r_2(k))$. Para esto, consideremos a Y_1 y Y_2 como estas variables y a $\mathbf{U} = (-\mathbf{T})^{-1}$ como la matriz de Green de τ . Sabemos que si $(X_t : t \geq 0)$ es la cadena asociada, entonces tenemos la siguiente expresión,

$$Y_1 = \int_0^\tau r_1(X_t) dt \quad \text{y} \quad Y_2 = \int_0^\tau r_2(X_t) dt$$

Dado que Y_1 y Y_2 tienen distribución de tipo fase, entonces tenemos expresiones para sus primeros y segundos momentos. Por otro lado, para calcular la covarianza obtendremos $\mathbb{E}[Y_1 Y_2]$ condicionando al primer paso con un proceso análogo al desarrollado en 3.3 y 3.5. Supongamos que el primer salto del proceso se da en tiempo S_1 , que i es un estado arbitrario no absorbente y notemos que

$$\begin{aligned} Y_1 Y_2 &= \int_0^\tau r_1(X_t) dt \int_0^\tau r_2(X_t) dt \\ &= r_1(i) r_2(i) S_1^2 + r_1(i) S_1 \left(\int_0^\tau r_2(X_t) dt - r_2(i) \right) + r_2(i) S_1 \left(\int_0^\tau r_1(X_t) dt - r_1(i) S_1 \right) \\ &\quad + \left(\int_0^\tau r_1(X_t) dt - r_1(i) S_1 \right) \left(\int_0^\tau r_2(X_t) dt - r_2(i) S_1 \right) \end{aligned}$$

La cual es una igualdad que se obtiene fácilmente al sumar convenientemente 0. Supongamos que $C_i = \mathbb{E}_i[Y_1 Y_2]$. Así, al aplicar esperanza y usar su linealidad, tenemos lo siguiente:

$$\begin{aligned}
\mathbb{E}[Y_1 Y_2] &= \mathbb{E}[r_1(i) r_2(i) S_1^2] + \mathbb{E} \left[r_1(i) S_1 \left(\int_0^\tau r_2(X_t) dt - r_2(i) \right) \right] \\
&\quad + \mathbb{E} \left[r_2(i) S_1 \left(\int_0^\tau r_1(X_t) dt - r_1(i) S_1 \right) \right] \\
&\quad + \mathbb{E} \left[\left(\int_0^\tau r_1(X_t) dt - r_1(i) S_1 \right) \left(\int_0^\tau r_2(X_t) dt - r_2(i) S_1 \right) \right] \\
&= r_1(i) r_2(i) \frac{2}{(-\mathbf{T}_{ii})^2} + r_1(i) \frac{1}{-\mathbf{T}_{ii}} \left(\int_0^\tau r_2(X_t) dt - r_2(i) \right) + r_2(i) S_1 \left(\int_0^\tau r_1(X_t) dt - r_1(i) S_1 \right) \\
&\quad + \left(\int_0^\tau r_1(X_t) dt - r_1(i) S_1 \right) \left(\int_0^\tau r_2(X_t) dt - r_2(i) S_1 \right)
\end{aligned}$$

No sé cómo se pasa de lo de arriba a lo de abajo, arriba me faltan poner algunas esperanzas. Está raro. Supongamos que $A_1(k)$ y $A_2(k)$ son la recompensa media acumulada dado que empezaste en k de \mathbf{r}_1 y \mathbf{r}_2 , respectivamente. Así, tenemos la siguiente igualdad

$$C_i = r_1(i) r_2(i) \frac{2}{(-\mathbf{T}_{ii})^2} + r_2(i) \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{-\mathbf{T}_{ii}} A_2(k) + r_2(i) \frac{1}{\mathbf{T}_{ii}} \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{-\mathbf{T}_{ii}} A_1(k) + \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{-\mathbf{T}_{ii}} C_k$$

O bien,

$$(-\mathbf{T}_{ii}) C_i = r_1(i) r_2(i) \frac{2}{(-\mathbf{T}_{ii})} + r_2(i) \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{-\mathbf{T}_{ii}} A_2(k) + r_2(i) \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{-\mathbf{T}_{ii}} A_1(k) + \sum_{k \neq i} \mathbf{T}_{ik} C_k$$

Notemos que al usar la igualdad 3.27, pues A_1 y A_2 son variables análogas a A , obtenemos que:

$$\begin{aligned}
\frac{r_1(i) r_2(i)}{(-\mathbf{T}_{ii})} + r_1(i) \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{(-\mathbf{T}_{ii})} A_2(k) &= \frac{r_1(i)}{(-\mathbf{T}_{ii})} \left[r_2(i) + \sum_{k \neq i} \mathbf{T}_{ik} A_2(k) \right] \\
&= \frac{r_1(i)}{(-\mathbf{T}_{ii})} [-\mathbf{T}_{ii} A_2(i)] = r_1(i) A_2(i)
\end{aligned}$$

De manera análoga,

$$\frac{r_1(i) r_2(i)}{(-\mathbf{T}_{ii})} + r_1(i) \sum_{k \neq i} \frac{\mathbf{T}_{ik}}{(-\mathbf{T}_{ii})} A_1(k) = r_2(i) A_1(i)$$

Por lo tanto, tenemos la expresión

$$(-\mathbf{T}_{ii}) C_i = r_1(i) A_2(i) + r_2(i) A_1(i) + \sum_{k \neq i} \mathbf{T}_{ik} C_k$$

De la cual se obtiene que

$$-\sum_{k=1}^k \mathbf{T}_{ik} C_k = r_1(i) A_2(i) + r_2(i) A_1(i)$$

Si llamamos $\mathbf{C} = (C_1, C_2, \dots, C_k)$, entonces en notación matricial y usando que $\mathbf{A}_1 = \mathbf{U}\Lambda(\mathbf{r}_1)\mathbf{e}$ y $\mathbf{A}_2 = \mathbf{U}\Lambda(\mathbf{r}_2)\mathbf{e}$ por el análisis hecho en 3.5, se sigue que

$$(-\mathbf{T})\mathbf{C} = \Lambda(\mathbf{r}_1)\mathbf{U}\Lambda(\mathbf{r}_2)\mathbf{e} + \Lambda(\mathbf{r}_2)\mathbf{U}\Lambda(\mathbf{r}_2)\mathbf{e}$$

Y por lo tanto

$$\mathbf{C} = \mathbf{U}\Lambda(\mathbf{r}_1)\mathbf{U}\Lambda(\mathbf{r}_2)\mathbf{e} + \mathbf{U}\Lambda(\mathbf{r}_2)\mathbf{U}\Lambda(\mathbf{r}_2)\mathbf{e}$$

Toda esta información la dejamos plasmada en el siguiente lema.

Lema 3.44. *Sea $\tau \sim PH_k(\boldsymbol{\pi}, \mathbf{T})$ con cadena asociada $(X_t : t \geq 0)$ y Y_1 y Y_2 las recompensas totales de esta cadena con vectores de recompensa \mathbf{r}_1 y \mathbf{r}_2 , respectivamente. Entonces*

$$\mathbb{E}[Y_1 Y_2] = \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_1) \mathbf{U} \Lambda(\mathbf{r}_2) \mathbf{e} + \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_2) \mathbf{U} \Lambda(\mathbf{r}_2) \mathbf{e}$$

Finalmente, ejemplificaremos estos resultados con un ejemplo.

Ejemplo 3.45. Sea $\tau \sim PH_4(\boldsymbol{\pi}, \mathbf{T})$ una variable aleatoria de tipo fase con submatriz de tasas y distribución inicial dadas por

$$\mathbf{T} = \begin{pmatrix} -3 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & 0 & -1 \end{pmatrix}, \quad \boldsymbol{\pi} = (1, 0, 0).$$

El estado 3 es absorbente, por lo que τ representa el tiempo hasta la absorción de una cadena de Markov de tres estados.

Definimos dos vectores de recompensa

$$\mathbf{r}_1 = (1, 0.5, 0), \quad \mathbf{r}_2 = (4, 1, 0),$$

y consideramos las variables aleatorias

$$Y_1 = \int_0^\tau r_1(X_t) dt, \quad Y_2 = \int_0^\tau r_2(X_t) dt,$$

que representan las recompensas acumuladas de la cadena en cada tipo de estado hasta la absorción. Por lo tanto, $\mathbf{Y} = (Y_1, Y_2)$ tiene una distribución de tipo fase multivariada, es decir,

$$\mathbf{Y} \sim MPH(\boldsymbol{\pi}, \mathbf{T}, \mathbf{R}), \quad \text{donde } \mathbf{R} = \begin{pmatrix} 1 & 4 \\ 0.5 & 1 \\ 0 & 0 \end{pmatrix}.$$

Usando los resultados obtenidos previamente, para $j = 1, 2$ se cumple que

$$\mathbb{E}[Y_j] = \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_j) \mathbf{e} \quad \text{con} \quad \mathbf{U} = (-\mathbf{T})^{-1}.$$

En este caso particular se tiene

$$\mathbf{U} = (-\mathbf{T})^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

por lo que

$$\mathbb{E}[Y_1] = \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_1) \mathbf{e} = (1, 0, 0) \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{e} = 0.5,$$

y análogamente

$$\mathbb{E}[Y_2] = \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_2) \mathbf{e} = (1, 0, 0) \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{e} = 1.666,$$

Por otra parte, el segundo momento cruzado se obtiene como

$$\mathbb{E}[Y_1 Y_2] = \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_1) \mathbf{U} \Lambda(\mathbf{r}_2) \mathbf{e} + \boldsymbol{\pi} \mathbf{U} \Lambda(\mathbf{r}_2) \mathbf{U} \Lambda(\mathbf{r}_1) \mathbf{e} = 1.3888,$$

por lo que la covarianza entre Y_1 y Y_2 resulta ser

$$Cov(Y_1, Y_2) = \mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_1] \mathbb{E}[Y_2] = 0.5555.$$

Este ejemplo muestra cómo es posible caracterizar la dependencia entre las componentes de \mathbf{Y} a partir de la estructura de la cadena subyacente y de los vectores de recompensa. En este caso, la covarianza positiva refleja que ambas recompensas tienden a acumularse simultáneamente en los mismos estados antes de la absorción.

4 Tiempos de Extinción en un Modelo de Tres Alelos

En los capítulos anteriores se establecieron las bases teóricas necesarias para comprender la estructura de las distribuciones tipo fase y su relación cadenas de Markov que se absorben. En particular, esto funciona en modelos clásicos de genética de poblaciones, tales como Wright–Fisher y Moran, pues permiten describir tiempos de extinción, fijación y coalescencia.

En este capítulo estudiaremos un caso particular de interés: la dinámica de un modelo neutro con tres alelos y los tiempos de extinción asociados a cada uno. Dados tres alelos presentes en una población finita, nos interesa analizar los tiempos τ_1 , τ_2 y τ_3 en los que cada uno de ellos desaparece de la población. Es interesante estudiar estos tiempos, pues aunque siempre habrá un alelo que se extinga primero, los dos tiempos restantes coincidirán una vez que el proceso alcance un estado con un solo tipo, lo cual impide que podamos estudiarlos vía estadísticos de orden.

Desde la perspectiva de distribuciones tipo fase, el tiempo hasta que cierto alelo se extingue puede verse como un tiempo de absorción para un subconjunto adecuado del espacio de estados del modelo. De este modo, cada variable τ_i puede interpretarse como una transformación vía recompensas del proceso subyacente, donde las recompensas asignadas dependen de la presencia o ausencia del tipo i . Sin embargo, al considerar la primera extinción entre los tres alelos, la estructura de tipo fase se pierde en general, pues el ordenamiento de los tiempos introduce dependencias que ya no pueden representarse mediante un solo generador de fase.

Un objetivo central de este capítulo es explorar si, una vez que uno de los tipos se extingue y el sistema queda reducido a dos alelos, la dinámica que queda se comporta como la de un modelo clásico de dos alelos, lo cual permitiría relacionar estos tiempos de extinción con las estructuras del coalescente de Kingman. Aunque se sabe que tal convergencia ocurre en el límite de poblaciones grandes y tiempos re-escalados, no es evidente que este comportamiento persista para tamaños finitos ni que las distribuciones resultantes conserven una representación tipo fase sencilla.

Para ello, comenzamos formalizando el modelo con tres alelos, describiendo el proceso de recuento como una cadena de Markov a tiempo continuo y caracterizando los

estados absorbentes asociados a la desaparición de cada tipo. Posteriormente, formularemos el modelo con la teoría de transformaciones vía recompensas para estudiar las variables τ_1, τ_2, τ_3 y analizaremos el comportamiento del proceso condicionado a la primera extinción. Finalmente, complementamos el desarrollo teórico con ejemplos numéricos y simulaciones que ilustran la estructura de estos tiempos y permiten comparar la dinámica observada con la predicha por los modelos de dos alelos y sus análogos coalescentes.

4.1. Modelo de Wright-Fisher con tres alelos

4.2. Análisis del tiempo de extinción por alelo

4.3. Coincidencias con el coalescente de Kingman

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

A Densidad de las Tipo Fase

Modelar el tiempo en el que ocurre cierto suceso es uno de los temas centrales al estudiar procesos aleatorios. En áreas de las matemáticas como el análisis de supervivencia, la teoría de colas o genética de poblaciones, es habitual trabajar con variables aleatorias que modelan tiempos de espera.

Este apéndice tiene como objetivo demostrar que cualquier variable aleatoria no negativa (como aquellas que representan tiempos) se puede aproximar en distribución por distribuciones de tipo fase. O, lo que es lo mismo, que las variables con distribución de tipo fase son densas en las no negativas. Este resultado es sumamente importante en la teoría y le da una ventaja importante a las variables de tipo fase, pues aproximan bien a cualquier distribución que modele tiempo. Este resultado está plasmado en el siguiente teorema

Teorema A.1. *Las variables aleatorias de tipo fase son densas en las distribuciones definidas en $[0, \infty)$*

Con el objetivo de llegar a la prueba de este resultado, recordemos la definición de convergencia en distribución para variables aleatorias.

Definición A.2. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias, donde cada X_n tiene función de distribución F_{X_n} , y sea X una variable aleatoria con función de distribución F . Decimos que $(X_n)_{n \in \mathbb{N}}$ converge en distribución a X si, para todo x que sea punto de continuidad de F , se cumple*

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F(x).$$

Esta convergencia se denota por

$$X_n \xrightarrow{\mathcal{D}} X.$$

A continuación, presentamos un resultado equivalente a la convergencia en distribución: el teorema de Portmentau. A este se refiere [4] como el teorema 11.1.1 y en este libro se puede consultar su prueba y discusión adicional.

Teorema A.3 (Teorema de Portmentau). *Para $(X_n)_{n \in \mathbb{N}}$ y X variables aleatorias, son equivalentes:*

1. $X_n \xrightarrow{\mathcal{D}} X$

2. Para toda f continua y acotada, $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$

Con miras de demostrar el teorema A.1, recordaremos algunos resultados clásicos de teoría de probabilidad y demostraremos una serie de lemas que nos serán útiles. Esta sección se apoya en lo expuesto en la sección 3.2.1 de [2] y en resultados de [20].

Lema A.4 (Desigualdad de Chebyshev). *Sea X una variable aleatoria con esperanza μ y varianza σ^2 , ambas finitas. Entonces para todo $\alpha > 0$,*

$$\mathbb{P}[|X - \mu| \geq \alpha] \leq \frac{\sigma^2}{\alpha^2}$$

Lema A.5 (Desigualdad de Jensen). *Dada φ una función convexa y X una variable aleatoria, se tiene que*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Lema A.6. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tal que para cada $n \in \mathbb{N}$, X_n tiene como función de distribución a F_{X_n} y $\mathbb{E}[X_n] = \theta$ para $\theta > 0$. Si*

$$\lim_{n \rightarrow \infty} \text{Var}(X_n) = 0,$$

entonces para toda función continua y acotada f , se satisface que

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = f(\theta)$$

Demostración. Sea $\varepsilon > 0$. Por un lado, gracias a la continuidad de f , existe $\delta > 0$ tal que

$$|f(x) - f(\theta)| < \frac{\varepsilon}{2} \text{ para toda } x \in B(\theta, \delta)$$

Donde $B(\theta, \delta)$ es la bola abierta con centro en θ y radio δ . Por otro lado, como f es acotada, existe $M > 0$ tal que para todo $x \in \mathbb{R}$,

$$|f(x)| \leq \frac{M}{2}$$

Esto implica que para cada $x, y \in \mathbb{R}$, $|f(x) - f(y)| \leq M$. Además, gracias al lema A.4, tenemos que para cada $n \in \mathbb{N}$,

$$\mathbb{P}[|X_n - \theta| > \delta] \leq \frac{\text{Var}(X_n)}{\delta^2}$$

Sea $n \in \mathbb{N}$. Notemos que al aplicar el lema A.5 y el hecho de que el valor absoluto es una función convexa obtenemos que

$$|\mathbb{E}[f(X_n)] - f(\theta)| = |\mathbb{E}[f(X_n) - f(\theta)]| \leq \mathbb{E}[|f(X_n) - f(\theta)|]$$

Por otro lado, al usar que f es continua y acotada, tenemos que

$$\begin{aligned} \mathbb{E}[|f(X_n) - f(\theta)|] &\leq \mathbb{E}[|f(X_n) - f(\theta)| \mathbf{1}_{\{|X_n - \theta| \leq \delta\}}] + \mathbb{E}[|f(X_n) - f(\theta)| \mathbf{1}_{\{|X_n - \theta| > \delta\}}] \\ &\leq \frac{\varepsilon}{2} + \mathbb{E}[|f(X_n) - f(\theta)| \mathbf{1}_{\{|X_n - \theta| > \delta\}}] \leq \frac{\varepsilon}{2} + M \mathbb{E}[\mathbf{1}_{\{|X_n - \theta| > \delta\}}] \\ &\leq \frac{\varepsilon}{2} + M \mathbb{P}[|X_n - \theta| \geq \delta] \leq \frac{\varepsilon}{2} + M \frac{\text{Var}(X_n)}{\delta^2}. \end{aligned}$$

Donde se usó A.4 en la última desigualdad. Luego, como

$$\lim_{n \rightarrow \infty} \text{Var}(X_n) = 0,$$

entonces existe $N \in \mathbb{N}$ tal que para toda $n \geq N$, $\text{Var}(X_n) < \frac{\delta^2}{2M}\varepsilon$. Así, suponiendo que $n \geq N$ y uniendo esto con lo anterior obtenemos que

$$|\mathbb{E}[f(X_n)] - f(\theta)| \leq \frac{\varepsilon}{2} + M \frac{\text{Var}(X_n)}{\delta^2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{para toda } n \geq N$$

lo cual prueba que

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = f(\theta)$$

□

Las hipótesis del teorema A.3 garantizan convergencia en distribución a la variable aleatoria constante θ para la sucesión descrita en el lema anterior. Esto tiene sentido, pues como la varianza tiende a cero conforme $n \rightarrow \infty$, entonces las variables cada vez se alejan menos de su media. A continuación presentamos otro resultado relevante.

Lema A.7. *Sean $\theta > 0$ y f continua y acotada. Entonces*

$$\lim_{n \rightarrow \infty} e^{-n\theta} \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) \frac{(n\theta)^k}{k!} = f(\theta)$$

*Demuestra*ción. La prueba es sencilla al usar el lema anterior. Para cada $n \in \mathbb{N}$, consideremos las variables aleatorias $Y_n \sim \text{Poi}(n\theta)$ y $X_n = \frac{Y_n}{n}$. Entonces por las propiedades de la distribución de Poisson se tiene que para cada $n \in \mathbb{N}$, $\mathbb{P}[Y_n = k] = e^{-n\theta} \frac{(n\theta)^k}{k!}$, $\mathbb{E}[Y_n] = n\theta$, $\text{Var}(Y_n) = n\theta$ y por tanto $\mathbb{P}[X_n = \frac{k}{n}] = e^{-n\theta} \frac{(n\theta)^k}{k!}$, $\mathbb{E}[X_n] = \theta$ y $\text{Var}(X_n) = \frac{\theta}{n}$. Al usar la propiedad del estadístico inconsciente para el valor esperado,

$$\mathbb{E}[f(X_n)] = \mathbb{E}\left[f\left(\frac{Y_n}{n}\right)\right] = \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) e^{-n\theta} \frac{(n\theta)^k}{k!} = e^{-n\theta} \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) \frac{(n\theta)^k}{k!}$$

Además, claramente la sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ cumple las hipótesis del lema anterior, ya que

$$\text{Var}(X_n) = \frac{\theta}{n} \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

Y al aplicar A.6,

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = f(\theta)$$

O bien,

$$\lim_{n \rightarrow \infty} e^{-n\theta} \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) \frac{(n\theta)^k}{k!} = f(\theta)$$

que es lo que queríamos demostrar. □

Observación A.8. Recordemos que si $(N_t, t \geq 0)$ es un proceso Poisson con tasa $\lambda > 0$, entonces podemos considerar al tiempo de la k -ésima llegada como una variable aleatoria S_k , la cual se distribuye $Erlang(k, \lambda)$ y cuya función de distribución evaluada en x estará denotada por $F_{S_k}(x; \lambda)$. Además, se da la siguiente igualdad de eventos

$$\{S_k > t\} = \{N_t < k\}$$

Pues si el tiempo de la k -ésima llegada es mayor que t , entonces en el tiempo t , no puede haber k o más llegadas; reciprocamente, si para el tiempo t hay menos de k llegadas, eso significa que el tiempo de la k -ésima tiene que ser mayor que t . Por lo tanto, tenemos que para todo $k \in \mathbb{N}$ y $t < 0$,

$$F_{S_k}(t; \lambda) = \mathbb{P}[S_k \leq t] = 1 - \mathbb{P}[S_k > t] = 1 - \mathbb{P}[N_t < k] = 1 - \sum_{i=0}^{k-1} \mathbb{P}[N_t = i]$$

Si denotamos por $p_t(i) = \mathbb{P}[N_t = i]$, la cual es conocida ya que sabemos la distribución de N_t , entonces

$$F_{S_k}(t; \lambda) = 1 - \sum_{i=0}^{k-1} p_t(i)$$

Lema A.9. Sean $\lambda > 0$ y X una variable aleatoria definida en $(0, \infty)$ y con distribución F . Supongamos que $(S_n)_{n \in \mathbb{N}}$ es una sucesión de variables aleatorias tales que para toda $k \in \mathbb{N}$, $S_k \sim Erlang(k, \lambda)$ y denotamos a sus distribuciones por $F_{S_k}(t; \lambda)$. Definimos

$$d_F(\lambda, k) = F\left(\frac{k}{\lambda}\right) - F\left(\frac{k-1}{\lambda}\right) \text{ para cada } k \in \mathbb{N}$$

Entonces para cada $t > 0$ se tiene que

$$\sum_{k=1}^{\infty} d_F(\lambda, k) F_{S_k}(t; \lambda) = e^{-\lambda t} \sum_{k=0}^{\infty} F\left(\frac{k}{\lambda}\right) \frac{(\lambda t)^k}{k!}.$$

Demostración. Para empezar, notemos que:

$$\begin{aligned} \sum_{k=1}^{\infty} d_F(\lambda, k) &= \sum_{k=1}^{\infty} F\left(\frac{k}{\lambda}\right) - F\left(\frac{k-1}{\lambda}\right) \\ &= \left(F\left(\frac{1}{\lambda}\right) - F\left(\frac{0}{\lambda}\right)\right) + \left(F\left(\frac{2}{\lambda}\right) - F\left(\frac{1}{\lambda}\right)\right) + \dots \\ &= \lim_{k \rightarrow \infty} F\left(\frac{k}{\lambda}\right) - F\left(\frac{0}{\lambda}\right) \\ &= 1 \end{aligned}$$

Pues es una suma telescópica, y como X está definida en $(0, \infty)$, entonces $F\left(\frac{0}{\lambda}\right) = 0$, y por ser función de distribución, $\lim_{k \rightarrow \infty} F\left(\frac{k}{\lambda}\right) = 1$. Usando esto y la observación anterior, tenemos que:

$$\begin{aligned}
\sum_{k=1}^{\infty} d_F(\lambda, k) F_{S_k}(t; \lambda) &= \sum_{k=1}^{\infty} d_F(\lambda, k) \left(1 - \sum_{i=0}^{k-1} p_t(i) \right) \\
&= \sum_{k=1}^{\infty} d_F(\lambda, k) - \sum_{k=1}^{\infty} d_F(\lambda, k) \sum_{i=0}^{k-1} p_t(i) \\
&= 1 - \sum_{k=1}^{\infty} d_F(\lambda, k) \sum_{i=0}^{k-1} p_t(i)
\end{aligned}$$

Ahora, usando que $\sum_{i=0}^{k-1} p_t(i) \leq 1$, se tiene que

$$0 \leq \sum_{k=1}^{\infty} d_F(\lambda, k) \sum_{i=0}^{k-1} p_t(i) \leq \sum_{k=1}^{\infty} d_F(\lambda, k) = 1$$

Entonces podemos aplicar el teorema de Fubini en el caso discreto y usar un argumento del tipo telescopico para obtener que

$$\begin{aligned}
\sum_{k=1}^{\infty} d_F(\lambda, k) \sum_{i=0}^{k-1} p_t(i) &= \sum_{i=0}^{\infty} p_t(i) \sum_{k=i+1}^{\infty} d_F(\lambda, k) = \sum_{i=0}^{\infty} p_t(i) \left(1 - F\left(\frac{i}{\lambda}\right) \right) \\
&= \sum_{i=0}^{\infty} p_t(i) - \sum_{i=1}^{\infty} p_t(i) F\left(\frac{i}{\lambda}\right) = 1 - \sum_{i=1}^{\infty} p_t(i) F\left(\frac{i}{\lambda}\right)
\end{aligned}$$

Al unir ambos resultados y usar que $p_t(i) = e^{-\lambda t} \frac{(\lambda t)^i}{i!}$, se sigue que

$$\sum_{k=1}^{\infty} d_F(\lambda, k) F_{S_k}(t; \lambda) = e^{-\lambda t} \sum_{k=0}^{\infty} F\left(\frac{k}{\lambda}\right) \frac{(\lambda t)^k}{k!}.$$

Tal como deseábamos. □

Con este último lema tenemos todo para demostrar el siguiente resultado

Teorema A.10. *Sea X una variable aleatoria definida en $(0, \infty)$ y con distribución F . Para cada $n \in \mathbb{N}$, sea X_n una variable aleatoria con distribución F_n dada por*

$$F_n(x) = \sum_{k=1}^{\infty} d_F(n, k) F_{S_k}(x; n) \text{ para cada } x \in \mathbb{R}$$

donde $d_F(n, k)$ está definida como en el lema A.9 y $F_{S_k}(x; n)$ corresponde a la distribución de una variable aleatoria Erlang(n, k) evaluada en x . Entonces

$$X_n \xrightarrow{\mathcal{D}} X.$$

Demostración. Empezamos probando que F_n funciona como una distribución distribución. Para comenzar, notemos que

$$\sum_{k=1}^{\infty} d_F(\lambda, k) = 1$$

Además, dado $k \in \mathbb{N}$, como $\frac{k-1}{\lambda} \leq \frac{k}{\lambda}$ y las funciones de distribución son no decrecientes, entonces $F(\frac{k-1}{\lambda}) \leq F(\frac{k}{\lambda})$, entonces efectivamente cada $d_F(n, k) \in [0, 1]$ y por tanto podemos usar a esta sucesión como distribución de probabilidad.

Finalmente, dado que las variables aleatorias de Erlang corresponden a distribuciones de tipo fase, por la teoría desarrollada en 3.4, tenemos que F_n es una mezcla de distribuciones de tipo fase y por tanto es una distribución. Por otro lado, consideremos $t \in \mathbb{R}$ tal que F es continua en t . Haciendo $\lambda = n$ en el lema A.9 y $f = F$, la cual es continua en t y acotada por ser de distribución, se obtiene que

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

Y por tanto

$$X_n \xrightarrow{\mathcal{D}} X$$

□

Pareciera que con este teorema hemos terminado la tarea de demostrar la densidad de las distribuciones tipo fase; sin embargo, el corolario 3.36 afirma que las distribuciones tipo fase son cerradas bajo mezclas finitas, y en el teorema no usamos una mezcla de este tipo, ya que está dada por una serie. Lo que sí conseguimos es el siguiente resultado

Corolario A.11. *Si X es una variable aleatoria en $(0, \infty)$, entonces existe una sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ que son una mezcla infinita de distribuciones de Erlang tal que*

$$X_n \xrightarrow{\mathcal{D}} X$$

A continuación demostraremos que podemos hacer un proceso análogo al anterior con mezclas finitas de variables aleatorias de tipo fase.

Observación A.12. Si X está definida en $[0, \infty)$ y es diferente a su restricción al $(0, \infty)$, entonces significa que la variable tiene probabilidad positiva de valer 0, supongamos que $\mathbb{P}[X = 0] = p > 0$. Con esta condición, podemos modificar el argumento anterior al considerar una mezcla con peso p en 0 y $1-p$ en una distribución que sea la mezcla numerable de aquellas de tipo Erlang. Por lo tanto, el corolario anterior es válido para variables aleatorias definidas en $[0, \infty)$.

Teorema A.13. *Sea X una variable aleatoria definida en $(0, \infty)$ con distribución F . Definimos*

$$F_{m,n}(t) = F\left(\frac{m}{n}\right)^{-1} \sum_{k=1}^m d_F(n, k) F_{S_k}(t; n)$$

Donde $d_F(n, k)$, $F_{S_k}(t; n)$ y F_n están definidos como antes. Entonces se tiene que

$$F_{m,n} \rightarrow F_n \text{ cuando } m \rightarrow \infty$$

uniformemente en $(0, \infty)$.

Demostración. Sea $t \in (0, \infty)$. Veamos que:

$$\begin{aligned} |F_n(t) - F_{m,n}(t)| &= \left| \sum_{k=1}^{\infty} d_F(n, k) F_{S_k}(t; n) - F\left(\frac{m}{n}\right)^{-1} \sum_{k=1}^m d_F(n, k) F_{S_k}(t; n) \right| \\ &= \left| \sum_{k=1}^m d_F(n, k) F_{S_k}(t; n) + \sum_{k=m+1}^{\infty} d_F(n, k) F_{S_k}(t; n) \right. \\ &\quad \left. - F\left(\frac{m}{n}\right)^{-1} \sum_{k=1}^m d_F(n, k) F_{S_k}(t; n) \right| \\ &= \left| \left(1 - F\left(\frac{m}{n}\right)^{-1}\right) \sum_{k=1}^m d_F(n, k) F_{S_k}(t; n) + \sum_{k=m+1}^{\infty} d_F(n, k) F_{S_k}(t; n) \right| \\ &\leq \left| \left(1 - F\left(\frac{m}{n}\right)^{-1}\right) \sum_{k=1}^m d_F(n, k) F_{S_k}(t; n) \right| + \left| \sum_{k=m+1}^{\infty} d_F(n, k) F_{S_k}(t; n) \right| \end{aligned}$$

Como $|F_{S_k}(t; n)| \leq 1$, entonces

$$\begin{aligned} |F_n(t) - F_{m,n}(t)| &\leq \left| \left(1 - F\left(\frac{m}{n}\right)^{-1}\right) \sum_{k=1}^m d_F(n, k) \right| + \left| \sum_{k=m+1}^{\infty} d_F(n, k) \right| \\ &\leq \left| \left(1 - F\left(\frac{m}{n}\right)^{-1}\right) F\left(\frac{m}{n}\right) \right| + \left| 1 - F\left(\frac{m}{n}\right) \right| \\ &= \left| F\left(\frac{m}{n}\right) - 1 \right| + \left| 1 - F\left(\frac{m}{n}\right) \right| \rightarrow |1 - 1| + |1 - 1| = 0 \end{aligned}$$

Donde en la última desigualdad usamos que

$$\sum_{k=1}^m d_F(n, k) = F\left(\frac{m}{n}\right)$$

Y como $\sum_{k=1}^{\infty} d_F(n, k) = 1$, entonces $\sum_{k=m+1}^{\infty} d_F(n, k) = 1 - F\left(\frac{m}{n}\right)$. Y finalmente

usamos que las funciones de distribución convergen a 1 cuando el argumento va a infinito. Como la última desigualdad no depende de t , entonces la convergencia es uniforme. \square

Gracias a este resultado tenemos el siguiente corolario.

Corolario A.14. *Sea X una variable aleatoria definida en $(0, \infty)$, con distribución F y t un punto de continuidad de esta función. Entonces*

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} F_{m,n}(t) = F(t)$$

donde $F_{m,n}$ está definida como en el teorema anterior.

Demostración. Dado $t \in (0, \infty)$ un punto de continuidad de F , el teorema anterior garantiza que

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} F_{m,n}(t) = \lim_{n \rightarrow \infty} F_n(t)$$

Finalmente, y gracias al teorema A.10, se sigue que

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

□

Gracias al corolario A.14 y al desarrollo realizado en este apéndice, hemos demostrado el teorema A.1, tal como queríamos.

Bibliografía

- [1] Odd O Aalen. Phase type distributions in survival analysis. *Scandinavian journal of statistics*, pages 447–463, 1995.
- [2] Mogens Bladt and Bo Friis Nielsen. *Matrix-exponential distributions in applied probability*, volume 81. Springer, 2017.
- [3] Michael Carter and Bruce Van Brunt. The lebesgue-stieltjes integral. In *The Lebesgue-Stieltjes Integral: A Practical Introduction*, pages 49–70. Springer, 2000.
- [4] Richard M Dudley. *Real analysis and probability*. Chapman and Hall/CRC, 2018.
- [5] Richard Durrett. *Essentials of stochastic processes*, volume 1. Springer, 2012.
- [6] Alison Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*. Springer, 2011.
- [7] María Clara Fittipaldi. Apunte procesos estocásticos 2, versión 2022-2. Material de curso, 2022. Apunte de la asignatura Procesos Estocásticos II.
- [8] Yun-Xin Fu. Exact coalescent for the wright–fisher model. *Theoretical population biology*, 69(4):385–394, 2006.
- [9] Asger Hobolth, Iker Rivas-González, Mogens Bladt, and Andreas Futschik. Phase-type distributions in mathematical population genetics: An emerging framework. *Theoretical Population Biology*, 157:14–32, 2024.
- [10] Götz Kersting, Jason Schweinsberg, and Anton Wakolbinger. The evolving beta coalescent. 2014.
- [11] Götz Kersting and Anton Wakolbinger. Probabilistic aspects of λ -coalescents in equilibrium and in evolution. *arXiv preprint arXiv:2002.05250*, 2020.
- [12] John Frank Charles Kingman. The coalescent. *Stochastic processes and their applications*, 13(3):235–248, 1982.
- [13] Fernando Lara and Fundación Universitaria Konrad Lorenz. Derivación matricial. *Bogotá: Fundación Universitaria KOnrad Lorenz*, 2007.
- [14] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

- [15] Magnus Nordborg. Coalescent theory. *Handbook of Statistical Genomics: Two Volume Set*, 2019.
- [16] James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.
- [17] Jim Pitman. Coalescents with multiple collisions. *Annals of Probability*, pages 1870–1902, 1999.
- [18] Luis Rincón. *Introducción a los procesos estocásticos*, volume 63. UNAM, Facultad de Ciencias, 2012.
- [19] Iván Irving Rosas Domínguez. Una aplicación de los procesos estocásticos a un problema de fisiología celular, 2023.
- [20] Sheldon M Ross. *A first course in probability*. Pearson Harlow, UK, 2020.
- [21] Ezequiel Severiano Guadalupe. Leyes de los grandes números en espacios de banach, 2018.
- [22] Roel Verbelen. Phase-type distributions & mixtures of erlangs. *A study of theoretical concepts, calibration techniques & actuarial applications university of leuven*, 2013.
- [23] John Wakeley. Coalescent models. In *Human population genomics: Introduction to essential concepts and applications*, pages 3–30. Springer, 2021.