



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Distribuciones Cuasiestacionarias y Procesos de Ramificación Multitipo

Analía Soledad Ferrari

Director: Pablo Groisman

25 de agosto de 2010

Agradecimientos

Gracias:

A Patu, por concederme el honor de ser mi director, por darme un gran empujón en estos últimos años, por conseguir que la matemática vuelva a ser divertida e interesante, por aguantarme por horas en su oficina y comprender cada una de mis caras, por responder cada uno de mis "no entiendo" con mucha paciencia y con una sonrisa.

A Marie, por aparecer en el momento justo, por enseñarme tanto sobre las probabilidades, por sus consejos, su amistad y por todo lo que aprendí gracias a ella.

A ambos, porque sin ustedes esta larga etapa de mi vida no hubiese llegado a su fin.

A los Julis, mis amigos, hermanos, y compañeros de tantas aventuras. Julieta por ponerle onda a las materias más espantosas, por contarme las cosas desde un lado divertido, por tranquilizarme en mis días de locura, por empujarme para dar la pila de finales. Julicho por ayudarme en todos estos años: estudiar con vos hizo que todo sea mas divertido, por tu paciencia para enseñarme como se comportan las cosas típicamente, por estar cada vez que te necesito, por tus consejos y tus idas y vueltas.

A mamá y papá, por su apoyo constante durante toda la carrera, por soportar mi locura y mi histeria antes de cada examen y por alegrarse con cada logro mío.

A Nuno y Nico, por su amistad, por empujarme para terminar y por explicarme y ser el help de esto llamado .tex.

A Maru, por sus charlas, consejos y horas de estudio.

A Constanza, por ser una gran docente, y porque varios años atrás, cuando todo parecía imposible, su paciencia fue fundamental.

A Fer, por compartirme su hogar cuando mi vida era una caos.

A Matías por su ayuda durante varios años.

A Sol y Dani, por compartir esos 2 años maravillosos, y por estar a mi lado siempre.

A las chicas de la facu: Pau, Lau, Georgi, Flor, Lu, Dany, Pablo, Marian, Iri, Carito, Ro, Mari... por compartir esta carrera, por las horas de estudio, por hacer que ir a facultad sea más divertido, por las charlas, y por todas las cosas que compartimos fuera y dentro de la facu.

A la gente que trabajó conmigo que hizo que el trabajo sea agradable y que soportó mis congresos, mis exámenes, y mis días de "tengo mucho que estudiar", en especial a Matías y a Lisi.

Y a muchos otros que no menciono pero forman parte de mi vida diaria.

Índice general

1. Procesos de Markov	11
1.1. Cadenas de Markov a tiempo discreto	11
1.2. Propiedad de Markov Fuerte	13
1.3. Construcción de una Cadena de Markov	14
1.4. Cadenas de Markov irreducibles y aperiódicas	15
1.5. Distribuciones estacionarias	17
1.6. Promedios empíricos: el Teorema Ergódico	24
2. Distribuciones Cuasiestacionarias	27
3. Procesos de Ramificación	33
3.1. Proceso de Galton-Watson	33
3.2. Procesos de ramificación a tiempo continuo	41
3.2.1. Construcción	41
3.2.2. Funciones generadoras	46
3.2.3. El proceso discreto inmerso	48
3.3. Tiempo continuo, multi-tipo	50
3.3.1. Construcción	50
3.3.2. Funciones generadoras	52
4. Aplicaciones	65
4.1. Proceso de urna	65
4.2. Convergencia empírica a la distribución cuasiestacionaria	67
5. Apéndice	75

Introducción

Uno de los temas centrales en el estudio de cadenas de Markov es el comportamiento de la cadena para tiempos largos. Bajo condiciones de irreducibilidad y aperiodicidad de estas cadenas se tiene existencia y unicidad del comportamiento asintótico, lo cual está dado por la medida invariante. Esta medida es la única con la siguiente propiedad: si el proceso comienza distribuido según ella, para todo los tiempos siguientes la medida del proceso es igual a la inicial. En términos de la matriz de transición, la medida invariante es caracterizada por el autovector a izquierda asociado al autovalor 1. Un ejemplo es el paseo al azar en los vértices de un cuadrado. Consideremos entonces los siguientes cuatro puntos de \mathbb{R}^2 : $(0,0), (1,0), (1,1), (0,1)$. Los enumeramos de 1 a 4 para simplificar la notación y construiremos un proceso de Markov en el espacio de estados $\mathcal{S} = \{1, 2, 3, 4\}$. Nos moveremos por estos cuatro puntos pasando de un vecino al otro. Una vez que estamos parados en un punto elegimos a uno de los dos vecinos con probabilidad $1/2$ para cada uno. Para cada n , X_n denota la esquina en la que estamos parados a tiempo n . O sea, $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov con probabilidad de transición

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

Por la simetría del modelo, la única medida invariante es la uniforme, que asigna $1/4$ de probabilidad a cada esquina. Es fácil ver que este es el autovector a izquierda asociado al autovalor 1.

No obstante, muchos procesos tienen un estado absorbente. Esto significa que tienen un estado del cual el proceso no sale, con lo cual si estudiamos el comportamiento de la cadena para tiempos grandes, la cadena estará en el estado absorbente. Por ejemplo, podemos considerar un proceso de nacimiento y muerte cuyo espacio de estados son los naturales. Con probabilidad de transición

$$p(n, n+1) = p, \quad p(n, n-1) = 1-p, \quad n \in \mathbb{N}$$

y $p(0, n) = 0, \forall n \in \mathbb{N}$ y $p(0, 0) = 1$. Si $p \geq \frac{1}{2}$ puede verse que, con probabilidad 1, la cadena será absorbida en el estado $n = 0$, y a partir de entonces se quedará constante igual a cero. Este estado corresponde a la situación donde murieron todos los individuos de la población.

Otro ejemplo es considerar la siguiente matriz de transición

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

Hay solamente dos estados $S = \{1, 2\}$, del primero se salta con probabilidad 1/2 para si mismo, y con probabilidad 1/2 se salta para el segundo. En el segundo se mantiene con probabilidad 1 en si mismo. Por lo tanto la única medida invariante es la que pone probabilidad total para el segundo estado.

Considerando un espacio de estados finito, bajo la hipótesis de irreducibilidad de la matriz sin el estado absorbente, este será alcanzado con probabilidad 1 y entonces la versión estacionaria del proceso carece de todo interés ya que se trata de un proceso constantemente igual al estado absorbente con probabilidad 1. Podemos entonces preguntarnos por la distribución del proceso a tiempos grandes condicionada a que no ha sido absorbido y analizar su comportamiento asintótico. A este límite lo llamamos *límite de Yaglom*. Definimos una distribución cuasiestacionaria como una medida de probabilidad que es invariante para la evolución condicionada a no ser absorbida. Esta distribución está dada por el autovector a izquierda asociado al autovalor de módulo máximo de la matriz de transición sin la fila y columna del estado absorbente. Además, probaremos que el límite de Yaglom es la medida cuasiestacionaria.

Tratamos este problema para procesos de Markov a tiempo discreto. En espacio de estados finito, la distribución cuasiestacionaria puede ser calculada usando herramientas del álgebra lineal, pero en el caso en que el espacio de estados sea muy grande, esto se torna impracticable. Aldous, Flannery y Palacios propusieron un método para obtener una muestra (aproximada) de la distribución cuasiestacionaria.

La idea central es, dada una cadena de Markov con un estado absorbente, construir otra cadena de Markov sin estados absorbentes, que nos ayude a aproximar la distribución cuasiestacionaria que buscamos. En palabras, esto se hace de la siguiente manera: dada la cadena de Markov con un estado absorbente, el nuevo proceso que creamos pasea por los estados hasta que llega al estado absorbente. En el siguiente paso se vuelve aleatoriamente a uno de los estados no absorbentes proporcionalmente a la cantidad de veces que ha pasado por cada estado. La cantidad normalizada de veces que ha pasado por cada estado se dice *medida empírica*. Considerando el vector aleatorio cuya primera entrada corresponde al estado que visita a tiempo n y la segunda, la medida empírica en ese tiempo, tendremos una cadena de Markov. Hecho esto, encontraremos una identificación entre la medida empírica en los tiempos de absorción y el proceso discreto inmerso en un proceso de ramificación multtipo a tiempo continuo en caso super crítico, del cual estudiamos la distribución asintótica de la población. Y con esto obtendremos la convergencia buscada.

La tesis está estructurada de la siguiente manera:

En el Capítulo 1, tratamos nociones básicas de cadenas de Markov a tiempo discreto en espacios de estados finito, hacemos una de las construcciones posibles, y vemos que bajo ciertas condiciones hay existencia y unicidad de la distribución estacionaria. Por último estudiamos la convergencia de los promedios empíricos.

En el Capítulo 2, damos una introducción al estudio de distribuciones cuasiestacionaria, analizamos existencia y unicidad.

En el Capítulo 3, tratamos procesos de ramificación. Damos la construcción del proceso, analizamos la extinción de la cadena y en el caso de que esto no ocurra obtenemos el comportamiento asintótico del proceso. Esto lo hacemos para distintos casos; comenzamos con tiempo discreto, luego tiempo continuo y al final analizamos el caso multidimensional a tiempo continuo.

En el Capítulo 4, estudiamos los procesos de urna y su equivalencia con los procesos de ramificación, ya que esto nos ayudará a estudiar la simulación propuesta por para la distribución cuasiestacionaria de cadenas de Markov con estados absorbentes.

Capítulo 1

Procesos de Markov

Vamos a comenzar introduciendo algunas nociones sobre procesos estocásticos. Un proceso estocástico es una familia de variables aleatorias $(X_t)_{t \in \mathbb{T}}$ definidas en un mismo espacio muestral (Ω, \mathcal{F}, P) donde

$$X_t : \Omega \rightarrow \mathcal{S}.$$

Usualmente \mathcal{S} se dice espacio de estados mientras que el índice t suele representar al tiempo. En algunos casos pensaremos en $\mathbb{T} = \mathbb{N}$ o $\mathbb{T} = \mathbb{N}_0$, dando lugar a un proceso estocástico a tiempo discreto, mientras que hablaremos de procesos a tiempo continuo cuando $\mathbb{T} = [0, \infty)$. Por ejemplo, una sucesión $(X_i)_{i \in \mathbb{N}}$ de variables aleatorias independientes idénticamente distribuidas (en adelante i.i.d.) es un proceso estocástico. En este capítulo, nos dedicaremos al estudio de procesos a tiempo discreto tomando valores en espacios de estados finitos, consideraremos $\mathcal{S} = \{s_1, \dots, s_k\}$. Está basado en las notas de Haggstrom [6] y en el libro de Brémaud [3].

1.1. Cadenas de Markov a tiempo discreto

Definición 1.1. Se dice que el proceso estocástico $(X_n)_{n \in \mathbb{N}}$ es una cadena de Markov si:

$$P(X_{n+1} = s_{n+1} | X_0 = s_0, \dots, X_n = s_n) = P(X_{n+1} = s_{n+1} | X_n = s_n) \quad \forall n \in \mathbb{N}$$

si $P(X_0 = s_0, \dots, X_n = s_n) \neq 0$.

Es decir que si conocemos el proceso hasta tiempo n , la posición del proceso a tiempo $n+1$ solo depende de la posición a tiempo n y no de todo el pasado.

Definición 1.2. El proceso $(X_n)_{n \in \mathbb{N}}$ se dice Markov homogéneo si además la probabilidad condicional no depende del tiempo, es decir, para todo $n \in \mathbb{N}$

$$P(X_{n+1} = s_j | X_0 = s_0, \dots, X_n = s_i) = P(X_{n+1} = s_j | X_n = s_i) = p(s_i, s_j) = P_{ij}.$$

En lo que sigue, nos restringiremos a considerar procesos homogéneos.

Definición 1.3. Sea P una matriz con entradas P_{ij} con $i, j \in \{1, \dots, k\}$. P se dice matriz de transición si verifica:

- $P_{ij} \geq 0 \quad \text{para todo } i, j \in \{1, \dots, k\}.$
- $\sum_{j=1}^k P_{ij} = 1 \quad \text{para todo } i \in \{1, \dots, k\}.$

Observemos que,

$$P(X_0 = s_0, \dots, X_n = s_n) = P(X_0 = s_0)P_{01}P_{12}\dots P_{n-1n}.$$

Definición 1.4. Llamamos μ o $\mu^{(0)}$ a la distribución inicial del proceso,

$$\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_k^{(0)}) = (P(X_0 = s_1), P(X_0 = s_2), \dots, P(X_0 = s_k)).$$

Conociendo la distribución inicial μ y la matriz de transición P queda caracterizado el proceso.

$$P(X_0 = i) = \mu_i,$$

$$P(X_{n+1} = s_j | X_n = s_i) = P_{ij}.$$

Los vectores, $\mu^{(1)}, \mu^{(2)}, \dots$ describen la distribución de la cadena de Markov a tiempo 1, 2, ..., donde

$$\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \dots, \mu_k^{(n)}) = (P(X_n = s_1), P(X_n = s_2), \dots, P(X_n = s_k))$$

para todo $n \in \mathbb{N}$.

Teorema 1.1. Sea $(X_n)_n \in \mathbb{N}$ una cadena de Markov definida en \mathcal{S} con distribución inicial $\mu^{(0)}$ y matriz de transición P . Para cada n , la distribución $\mu^{(n)}$ satisface:

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

Demostración. Veamos que la igualdad es cierta usando inducción.

Para $n = 1$, sea $s_j \in \mathcal{S}$, se tiene

$$\begin{aligned} \mu_j^{(1)} &= P(X_1 = s_j) \\ &= \sum_{i \in \mathcal{S}} P(X_1 = s_j | X_0 = s_i)P(X_0 = s_i) \\ &= \sum_{i \in \mathcal{S}} \mu_i^{(0)} P_{ij} \\ &= (\mu^{(0)}P)_j \end{aligned}$$

De donde,

$$\mu^{(1)} = \mu^{(0)}P$$

Supongamos que vale para $n = m$, veamos que es cierto para $n = m + 1$

$$\begin{aligned}
\mu_j^{(m+1)} &= P(X_{m+1} = s_j) \\
&= \sum_{i \in S} P(X_{m+1} = s_j | X_m = s_i) P(X_m = s_i) \\
&= \sum_{i \in S} \mu_i^{(m)} P_{ij} \\
&= (\mu^{(m)} P)_j
\end{aligned}$$

de donde, $\mu^{(m+1)} = \mu^{(m)} P$. Pero por la hipótesis inductiva, $\mu^{(m)} = \mu^{(0)} P^m$, entonces

$$\mu^{(m+1)} = \mu^{(0)} P^{m+1}.$$

□

1.2. Propiedad de Markov Fuerte

En la sección anterior vimos la propiedad de Markov, que nos dice que para predecir el futuro del proceso, conocer sólo el presente nos brinda tanta información como conocer toda la historia del proceso hasta el presente. Una formulación más general es la Propiedad de Markov Fuerte, aquí sólo la enunciaremos, pero para ello son necesarias unas definiciones previas.

Definición 1.5. *Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad. Sean $(\mathcal{F}_n)_{n \in \mathbb{N}}$, σ -álgebras. Se dice que $(\mathcal{F}_n)_{n \in \mathbb{N}}$ es un filtración, si*

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F} \quad \forall n \in \mathbb{N}.$$

La introducción de filtraciones es importante para el estudio de procesos estocásticos. Podemos pensar que \mathcal{F}_n representa toda la información disponible hasta tiempo n . Un proceso estocástico $X = (X_n)_{n \in \mathbb{N}}$ se dice que es adaptado a la filtración $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ si para cada $n \in \mathbb{N}$, X_n es \mathcal{F}_n -medible. (Notamos, $X_n \in \mathcal{F}_n$).

Definición 1.6. *Sea (Ω, \mathcal{F}) un espacio medible con una filtración $(\mathcal{F}_n)_{n \in \mathbb{N}}$.*

- *Un tiempo aleatorio T es una variable aleatoria \mathcal{F} -medible, con valores en $\mathbb{N} \cup \{+\infty\}$.*
- *Un tiempo aleatorio τ se dice tiempo de parada con respecto a la filtración si $\{\tau \leq n\} \in \mathcal{F}_n$, para todo $n \in \mathbb{N}$.*

Teorema 1.2 (Propiedad de Markov Fuerte). *Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov homogénea definida en un espacio de estados S y con matriz de transición P . Sea τ un tiempo de parada respecto a la filtración $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ para $n \in \mathbb{N}$. Dado cualquier estado $s_i \in S$, si $X_\tau = s_i$, entonces:*

- *El proceso después de τ y el proceso antes de τ son independientes.*

- *El proceso después de τ es una cadena de Markov homogénea con matriz de transición P .*

Demostración. Puede verse en Brémaud [3, Sección 7.1]. \square

La propiedad fuerte de Markov es, una extensión de la propiedad de Markov a tiempos de parada. Cabe aclarar que, en general, no vale el resultado para tiempos aleatorios cualesquiera. La estructura particular del tiempo de parada es la que permite al proceso que conocer la posición a tiempo τ , sea lo mismo que conocer toda la historia del proceso hasta τ .

1.3. Construcción de una Cadena de Markov

Vimos que la distribución inicial y la matriz de transición caracterizan una cadena de Markov. Surge entonces la siguiente pregunta; ¿Cómo simulamos una cadena de Markov $(X_n)_{n \in \mathbb{N}}$ en el espacio de estados \mathcal{S} , con distribución inicial μ y con matriz de transición P ?

Vamos a necesitar U_0, U_1, \dots variables aleatorias independientes con distribución uniforme en el intervalo $[0, 1]$ y dos funciones, la primera la llamamos función inicial; $\Psi : [0, 1] \rightarrow \mathcal{S}$, asumimos que cumple lo siguiente:

- particionamos el intervalo $[0, 1]$ en varios subintervalos, y Ψ es constante en cada uno.
- para cada $s \in \mathcal{S}$, el largo total de los intervalos con $\Psi = s$ es $\mu(s)$, es decir, $\int_0^1 1_{\{\Psi(x)=s\}} dx = \mu(s)$.

Con la función Ψ , podemos generar X_0 usando la primera de las variables uniformes. Sea $X_0 = \Psi(U_0)$, esto da una correcta distribución de X_0 pues, para cada $s \in \mathcal{S}$;

$$P(X_0 = s) = P(\Psi(U_0) = s) = \int_0^1 1_{\{\Psi(x)=s\}} dx = \mu(s).$$

Definamos la función inicial:

$$\Psi(x) = \begin{cases} s_1 & \text{si } x \in [0, \mu^{(0)}(s_1)) \\ s_2 & \text{si } x \in [\mu^{(0)}(s_1), \mu^{(0)}(s_1) + \mu^{(0)}(s_2)) \\ \vdots & \\ s_i & \text{si } x \in [\sum_{j=1}^{i-1} \mu^{(0)}(s_j), \sum_{j=1}^i \mu^{(0)}(s_j)) \\ \vdots & \\ s_k & \text{si } x \in [\sum_{j=1}^{k-1} \mu^{(0)}(s_j), 1]. \end{cases}$$

Claramente así definida verifica ambas condiciones, pues es constante en cada intervalo y además, $\int_0^1 1_{\{\Psi(x)=s_i\}} dx = \sum_{j=1}^i \mu(s_j) - \sum_{j=1}^{i-1} \mu(s_j) = \mu(s_i)$.

Ya sabemos como generar X_0 , nos falta generar X_{n+1} a partir de X_n para cualquier n . Para ello vamos a usar la v.a. U_{n+1} y una función $\Phi : \mathcal{S} \times [0, 1] \rightarrow \mathcal{S}$, es decir, la entrada de la función es un elemento de \mathcal{S} y un número entre 0 y 1 y la salida un elemento de \mathcal{S} . Necesitamos que Φ cumpla ciertas propiedades:

- para $s_i \in S$, la función $\Phi(s_i, x)$ es constante a trozos (mirándola solo como función de x).
- para $s_i, s_j \in S$ el largo total de los intervalos con $\Phi(s_i, x) = s_j$ es $p(s_i, s_j)$.

Al igual que la primer función que definimos, podemos reescribir la segunda condición:

$$\int_0^1 \mathbf{1}_{\{\Phi(s_i, x) = s_j\}} dx = P_{ij}.$$

Si Φ satisface esta condición, entonces $P(X_{n+1} = s_j | X_n = s_i) = P(\Phi(s_i, U_{n+1}) = s_j | X_n = s_i) = P(\Phi(s_i, U_{n+1}) = s_j) = P_{ij}$. Observemos que U_{n+1} es independiente de U_1, \dots, U_n y por lo tanto es independiente de X_n . Luego definimos la función actualizadora, para $s_i \in S$:

$$\Phi(s_i, x) = \begin{cases} s_1 & \text{si } x \in [0, P_{i1}) \\ s_2 & \text{si } x \in [P_{i1}, P_{i1} + P_{i2}) \\ \vdots & \\ s_i & \text{si } x \in [\sum_{j=1}^{i-1} P_{ij}, \sum_{j=1}^i P_{ij}) \\ \vdots & \\ s_k & \text{si } x \in [\sum_{j=1}^{k-1} P_{ij}, 1]. \end{cases}$$

Por como definimos Φ claramente cumple ambas condiciones, pues es constante a trozos y además, $\int_0^1 \mathbf{1}_{\{\Phi(s_i, x) = s_j\}} dx = \sum_{l=1}^j P_{il} - \sum_{l=1}^{j-1} P_{il} = P_{ij}$. De esta forma simulamos la cadena de Markov,

$$\begin{aligned} X_0 &= \Psi(U_0) \\ X_1 &= \Phi(X_0, U_1) \\ X_2 &= \Phi(X_1, U_2) \\ &\vdots \end{aligned}$$

y así sucesivamente.

Esto prueba la existencia de la cadena de Markov a partir de las variables aleatorias independientes con distribución uniforme definidas en un mismo espacio y de dos funciones, a las que llamamos función inicial (Ψ) y función actualizadora (Φ). Existen otras elecciones posibles para las funciones Φ , y Ψ que pueden tener distintas ventajas. Esta elección será muy útil en las siguientes secciones.

1.4. Cadenas de Markov irreducibles y aperiódicas

En la sección siguiente vamos a estudiar el comportamiento de las cadenas a largo plazo, para ello necesitamos hacer ciertas suposiciones sobre las cadenas de Markov. Vamos a considerar cadenas irreducibles y aperiódicas, y analizar sus propiedades. Comencemos con algunas definiciones.

Definición 1.7. Una cadena de Markov $(X_n)_{n \in \mathbb{N}}$ en \mathcal{S} con matriz de transición P , se dice irreducible si $\forall s_i, s_j \in \mathcal{S}$ existe $n \in \mathbb{N}$ tal que $(P^n)_{ij} > 0$.

El período de un estado, $s_i \in \mathcal{S}$ es $d(i) = \text{mcd}\{n \geq 1 : (P^n)_{ii} > 0\}$.

Un estado $s_i \in \mathcal{S}$ se dice aperiódico si $d(i)=1$.

Una cadena de Markov se dice aperiódica, si todos sus estados son aperiódicos.

Lema 1.1. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov aperiódica en $\mathcal{S} = \{s_1, \dots, s_k\}$ con matriz de transición P . Existe $N \in \mathbb{N}$ tal que $(P^n)_{ii} > 0$ para $i \in \{1, \dots, k\}, \forall n \geq N$.

Para demostrar este lema vamos a utilizar el siguiente resultado de teoría de números.

Lema 1.2. Sea $A = \{a_1, a_2, \dots\}$ un conjunto de números naturales, tal que

1. $\text{mcd} = \{a_1, a_2, \dots\} = 1$,
2. si $a \in A$ y $\tilde{a} \in A$ entonces $a + \tilde{a} \in A$ (cerrado bajo suma).

Entonces existe $N \in \mathbb{N}$ tal que $n \in A$ para todo $n \geq N$.

Demostración. Puede verse en el apéndice de Brémaud [3]. □

Demostración del Lema 1.1. Para cada $s_i \in \mathcal{S}$, sea $A_i = \{n \geq 1 : (P^n)_{ii} > 0\}$, es decir, A_i son los tiempos posibles de retorno del estado s_i . Como la cadena de Markov es aperiódica, el estado s_i es aperiódico: $\text{mcd}\{n \geq 1 : (P^n)_{ii} > 0\} = 1$. Veamos que es cerrado bajo suma, sean $a, \tilde{a} \in A$, por lo que $P(X_a = s_i | X_0 = s_i) > 0$ y $P(X_{\tilde{a}+a} = s_i | X_a = s_i) > 0$ luego,

$$\begin{aligned} P(X_{a+\tilde{a}} = s_i | X_0 = s_i) &\geq P(X_{a+\tilde{a}} = s_i, X_a = s_i | X_0 = s_i) \\ &= P(X_{a+\tilde{a}} = s_i | X_a = s_i, X_0 = s_i)P(X_a = s_i | X_0 = s_i) \\ &= P(X_{a+\tilde{a}} = s_i | X_a = s_i)P(X_a = s_i | X_0 = s_i) > 0 \end{aligned}$$

de donde, $a + \tilde{a} \in A$. Luego A satisface las hipótesis del lema previo, por lo tanto existe N_i tal que $(P^n)_{ii} > 0 \forall n \geq N_i$, tomando $N = \max\{N_1, N_2, \dots, N_k\}$ queda demostrado el teorema. □

Corolario 1.1. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica, en un espacio de estados $\mathcal{S} = \{s_1, \dots, s_k\}$ y P matriz de transición, entonces existe $M < \infty$ tal que $(P^n)_{ij} > 0 \forall i, j \in \{1, 2, \dots, k\}$ y $\forall n \geq M$.

Demostración. Por el teorema anterior, existe $N \in \mathbb{N}$ tal que $(P^n)_{ii} > 0 \forall n \geq N, \forall i \in \{1, 2, \dots, k\}$. Sean $s_i, s_j \in \mathcal{S}$. Como $(X_n)_{n \in \mathbb{N}}$ es irreducible, existe n_{ij} tal que $(P^{n_{ij}})_{ij} > 0$. Sea $M_{ij} = N + n_{ij}$, para m tal que $m - n_{ij} > N$

$$\begin{aligned} P(X_m = s_j | X_0 = s_i) &\geq P(X_{m-n_{ij}} = s_i, X_m = s_j | X_0 = s_i) \\ &\geq P(X_{m-n_{ij}} = s_i | X_0 = s_i)P(X_m = s_j | X_{m-n_{ij}} = s_i) \\ &> 0. \end{aligned}$$

Luego $(P^m)_{ij} > 0 \forall m \geq M_{ij}$, el corolario queda demostrado, tomando $M = \max\{M_{ij} : i, j \in \{1, 2, \dots, k\}\}$. □

1.5. Distribuciones estacionarias

Vamos a considerar una de las cuestiones principales en la teoría de Markov, la distribución de la cadena a largo plazo. ¿Qué podemos decir sobre la cadena de Markov, si la dejamos correr un largo tiempo? Para empezar veamos la definición de distribución estacionaria.

Definición 1.8. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov en $\mathcal{S} = \{s_1, \dots, s_k\}$ con matriz de transición P . El vector $\pi = (\pi_1, \dots, \pi_k)$ se dice que es una distribución estacionaria de la cadena si:

1. $\pi_i \geq 0 \quad \forall i = \{1, \dots, k\}$ y $\sum_{i=1}^k \pi_i = 1$.
2. $\pi P = \pi$, es decir, $\sum_{i=1}^k \pi_i P_{ij} = \pi_j \quad \forall j = \{1, \dots, k\}$.

La primera condición muestra que π es una probabilidad en \mathcal{S} y la segunda implica que si la distribución inicial $\mu^{(0)}$ es igual a π entonces la distribución de la cadena a tiempo 1 satisface: $\mu^{(1)} = \mu^{(0)}P = \pi P = \pi$, de donde $\mu^{(n)} = \pi$, para todo n .

Vamos a probar que si la cadena es irreducible y aperiódica existe una única distribución estacionaria, para ello debemos analizar ciertas propiedades de estas cadenas.

Definición 1.9. Sea T_{ij} la primera vez que la cadena visita el estado s_j , siendo que a tiempo 0 comenzó en s_i . Si $X_0 = s_i$, definimos:

$$T_{ij} = \min\{n \geq 1 / X_n = s_j\}$$

y $T_{ij} = \infty$ cuando la cadena nunca visita al estado s_j . También definimos el tiempo esperado hasta visitar el estado s_j , como $\tau_{ij} = E(T_{ij})$.

Lema 1.3. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica, definida en un espacio de estados $\mathcal{S} = \{s_1, \dots, s_k\}$, con matriz de transición P . Para cualquier estado $s_i, s_j \in \mathcal{S}$, si $X_0 = s_i$, entonces $P(T_{ij} < \infty) = 1$ y además, τ_{ij} es finito.

Demostración. Por el corolario anterior, existe $M < \infty$ tal que $(P^M)_{ij} > 0$ para todo $i, j \in \{1, \dots, k\}$. Sea $\alpha = \min\{(P^M)_{ij} / i, j \in \{1, \dots, k\}\}$. Notemos que α es positivo. Sean s_i, s_j dos estados de \mathcal{S} y supongamos que la cadena empieza en s_i . Entonces,

$$\begin{aligned} P(T_{ij} > M) &\leq P(X_0 = s_i, X_1 \neq s_j, \dots, X_M \neq s_j) \\ &\leq P(X_M \neq s_j) \\ &\leq 1 - \alpha. \end{aligned}$$

También, podemos acotar la probabilidad de no haber llegado a s_j a tiempo $2M$, condicionando sobre lo sucedido a tiempo M :

$$\begin{aligned} P(T_{ij} > 2M) &\leq P(T_{ij} > 2M | T_{ij} > M)P(T_{ij} > M) \\ &\leq P(X_{2M} \neq s_j | T_{ij} > M)P(T_{ij} > M) \\ &= P(X_{2M} \neq s_j | X_1 \neq s_j, \dots, X_M \neq s_j)P(T_{ij} > M) \\ &\leq P(X_{2M} \neq s_j | X_M \neq s_j)(1 - \alpha) \\ &\leq (1 - \alpha)^2. \end{aligned}$$

Iterando este argumento, para cualquier l tenemos,

$$\begin{aligned} P(T_{ij} > lM) &\leq P(T_{ij} > M)P(T_{ij} > 2M|T_{ij} > M) \cdots P(T_{ij} > lM|T_{ij} > (l-1)M) \\ &\leq (1-\alpha)^l. \end{aligned}$$

Tomando límite cuando $l \rightarrow \infty$, obtenemos que $P(T_{ij} = \infty) = 0$. Veamos que τ_{ij} es finito,

$$\begin{aligned} E(T_{ij}) &= \sum_{n=1}^{\infty} P(T_{ij} \geq n) \\ &= \sum_{l=0}^{\infty} \sum_{n=lM}^{(l+1)M-1} P(T_{ij} > n) \\ &\leq \sum_{l=0}^{\infty} \sum_{n=lM}^{(l+1)M-1} P(T_{ij} > lM) \\ &= M \sum_{l=0}^{\infty} P(T_{ij} > lM) \\ &\leq M \sum_{l=0}^{\infty} (1-\alpha)^l \\ &= M \frac{1}{1-(1-\alpha)} = \frac{M}{\alpha} < \infty, \end{aligned}$$

con lo cual queda demostrado el lema. \square

Definición 1.10. Si $X_0 = s_1$, definimos ρ_i , el número esperado de visitas al estado i antes de regresar al estado inicial, es decir, para $i = 1, \dots, k$,

$$\rho_i = E\left(\sum_{n=1}^{T_{11}} \mathbf{1}_{\{X_n=s_i\}}\right).$$

Observemos que $\rho_i = E\left(\sum_{n=1}^{T_{11}} \mathbf{1}_{\{X_n=s_i\}}\right) = E\left(\sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=s_i\}} \mathbf{1}_{\{T_{11}>n\}}\right)$ que podemos reescribirlo como $\sum_{n=1}^{\infty} P(X_n = s_i, T_{11} > n) = \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n)$.

Definición 1.11. Un estado $i \in S$ se dice recurrente si

$$P(T_{ii} < \infty) = 1.$$

En otro caso se dice transitorio. Un estado recurrente $i \in S$ se dice recurrente positivo si $E(T_{ii}) < \infty$. Una cadena de Markov se dice recurrente (positiva) si todos sus estados son recurrentes (positivos).

Observación 1.1. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica, definida en un espacio de estados finito. Entonces, $(X_n)_{n \in \mathbb{N}}$ es recurrente positiva.

Ahora sí estamos en condiciones de demostrar la existencia de una distribución estacionaria.

Teorema 1.3. *Para cualquier cadena de Markov irreducible y aperiódica, existe por lo menos una distribución estacionaria.*

Demostración. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica, en un espacio de estados $\mathcal{S} = \{s_1, \dots, s_k\}$ y matriz de transición P . Supongamos que la cadena empieza en el estado s_1 . El tiempo esperado de retorno τ_{11} es finito, y claramente $\rho_i < \tau_{11}$, por lo tanto ρ_i también es finito para todo $i \in \{1, \dots, k\}$. Proponemos,

$$\pi = (\pi_1, \dots, \pi_k) = \left(\frac{\rho_1}{\tau_{11}}, \frac{\rho_2}{\tau_{11}}, \dots, \frac{\rho_k}{\tau_{11}} \right).$$

Veamos que así definido, cumple las dos condiciones de la definición de distribución estacionaria.

- por como lo definimos, $\pi_i \geq 0$ para $i = 1, \dots, k$. Veamos que $\sum_{i=1}^k \pi_i = 1$;

$$\begin{aligned} \tau_{11} &= E(T_{11}) = \sum_{n=0}^{\infty} P(T_{11} > n) \\ &= \sum_{n=0}^{\infty} \sum_{i=1}^k P(X_n = s_i, T_{11} > n) \\ &= \sum_{i=1}^k \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n) \\ &= \sum_{i=1}^k \rho_i \end{aligned}$$

de modo que,

$$\sum_{i=1}^k \pi_i = \frac{1}{\tau_{11}} \sum_{i=1}^k \rho_i = 1.$$

- queremos ver que $\sum_{i=1}^k \pi_i P_{ij} = \pi_j$ para todo $j = 1, \dots, k$. Veamos primero para $j \neq 1$

$$\begin{aligned}
\pi_j &= \frac{\rho_j}{\tau_{11}} = \frac{1}{\tau_{11}} \sum_{n=0}^{\infty} P(X_n = s_j, T_{11} > n) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} P(X_n = s_j, T_{11} > n) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} P(X_n = s_j, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, X_n = s_j, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_j | X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_j | X_{n-1} = s_i)^1 \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=1}^k P_{ij} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{i=1}^k P_{ij} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{i=1}^k P_{ij} \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n) \\
&= \frac{1}{\tau_{11}} \sum_{i=1}^k P_{ij} \rho_i \\
&= \sum_{i=1}^k \pi_i P_{ij}.
\end{aligned}$$

Falta verificar esta condición para $j = 1$,

$$\rho_1 = \sum_{n=0}^{\infty} P(X_n = s_1, T_{11} > n) = P(T_{11} > 0) = 1.$$

¹ $\{T_{11} > n-1\} = \{X_1 \neq s_1, \dots, X_{n-1} \neq s_1\}$

Vimos que $P(T_{11} < \infty) = 1$, luego

$$\begin{aligned}
\rho_1 &= P(T_{11} < \infty) = \sum_{n=1}^{\infty} P(T_{11} = n) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, T_{11} = n) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, X_n = s_1, T_{11} > n-1) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_n = s_1 | X_{n-1} = s_i) P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^k P_{i1} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \sum_{i=1}^k P_{i1} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \sum_{i=1}^k P_{i1} \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n) \\
&= \sum_{i=1}^k P_{i1} \rho_i
\end{aligned}$$

por lo tanto,

$$\pi_1 = \frac{\rho_1}{\tau_{11}} = \sum_{i=1}^k \frac{\rho_i P_{i1}}{\tau_{11}} = \sum_{i=1}^k \pi_i P_{i1}.$$

Con lo cual, probamos que π es una distribución estacionaria. \square

Nos preguntamos si esta distribución estacionaria es única. Pero antes de analizar la unicidad de la distribución estacionaria, podemos responder a la pregunta inicial: ¿Qué le pasa a la cadena de Markov cuando la dejamos correr un tiempo suficientemente largo? Bajo ciertas hipótesis podemos ver que sin tener en cuenta cual sea la distribución inicial, la distribución a tiempos grandes, se parecerá a la distribución estacionaria. Para ver esto, debemos usar alguna métrica para calcular la distancia entre las distribuciones de probabilidad. Existen varias métricas, aquí usaremos una llamada distancia de variación total.

Definición 1.12. Sean los vectores $\mu^{(1)} = (\mu_1^{(1)}, \dots, \mu_k^{(1)})$ y $\mu^{(2)} = (\mu_1^{(2)}, \dots, \mu_k^{(2)})$ distribuciones de probabilidad en $S = \{s_1, \dots, s_k\}$. Definimos la distancia de variación total entre $\mu^{(1)}$ y $\mu^{(2)}$ como:

$$d_{VT}(\mu^{(1)}, \mu^{(2)}) = \frac{1}{2} \sum_{i=1}^k |\mu_i^{(1)} - \mu_i^{(2)}|.$$

Si $\mu^{(1)}, \mu^{(2)}, \dots$ y μ son distribuciones en \mathcal{S} , decimos que $\mu^{(n)}$ converge a μ en variación total cuando,

$$\lim_{n \rightarrow \infty} d_{VT}(\mu^{(n)}, \mu) = 0.$$

Notación: $\mu^{(n)} \xrightarrow[n \rightarrow \infty]{VT} \mu$.

Observemos que si $d_{VT}(\mu^{(1)}, \mu^{(2)}) = 0 \implies \mu^{(1)} = \mu^{(2)}$. Además, puede verse que

$$d_{VT}(\mu^{(1)}, \mu^{(2)}) = \sup_{A \in \mathcal{S}} |\mu^{(1)}(A) - \mu^{(2)}(A)|.$$

Teorema 1.4. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica en $\mathcal{S} = \{s_1, \dots, s_k\}$, con matriz de transición P y una distribución inicial μ^0 . Entonces para cualquier distribución estacionaria π para la matriz P ,

$$\mu^{(n)} \xrightarrow[n \rightarrow \infty]{VT} \pi.$$

Demostración. Vimos que la cadena de Markov (X_n) , podemos obtenerla a partir de variables aleatorias uniformes en $[0, 1]$ y dos funciones a las que llamamos, función de inicialización (Ψ_μ) y función de actualización (Φ). Vamos a usar el método de acoplamiento, es decir, construir varios procesos en el mismo espacio de probabilidad.

Sea $(X'_n)_{n \in \mathbb{N}}$ otra cadena de Markov, con ψ_π la función iniciadora para la distribución inicial π , sean (U'_1, U'_2, \dots) otra sucesión de variables aleatorias uniformes en $[0, 1]$ (independientes de las U_i). Simulamos X'_n :

$$\begin{aligned} X'_0 &= \Psi(U'_0) \\ X'_1 &= \Phi(X'_0, U'_1) \\ X'_2 &= \Phi(X'_1, U'_2) \\ &\vdots \end{aligned}$$

Como π es la distribución estacionaria, X'_n tiene distribución π para todo n . Además, las cadenas $(X_1, X_2, \dots), (X'_1, X'_2, \dots)$ son independientes. Queremos ver que con probabilidad 1 las cadenas se encuentran, es decir, que existe un n tal que $X_n = X'_n$. Para ello definamos el primer tiempo de encuentro:

$$T = \min\{n / X_n = X'_n\}$$

y notamos $T = \infty$ si las cadenas nunca se encuentran. Sabemos que existe M finito, tal que $(P^M)_{ij} > 0$ para todo $i, j = \{1, \dots, k\}$. Sea $\alpha = \min\{(P^M)_{ij} : j, i = \{1, \dots, k\}\}$, claramente $\alpha > 0$.

Acotemos la probabilidad de que las cadenas se encuentren antes del tiempo M ,

$$\begin{aligned}
 P(T \leq M) &\geq P(X_M = X'_M) \geq P(X_M = s_1, X'_M = s_1) \\
 &= P(X_M = s_1)P(X'_M = s_1) \\
 &= \sum_{i=1}^k P(X_0 = s_i, X_M = s_1) \sum_{i=1}^k P(X'_0 = s_i, X'_M = s_1) \\
 &= \sum_{i=1}^k P(X_M = s_1 | X_0 = s_i)P(X_0 = s_i) \sum_{i=1}^k P(X'_M = s_1 | X'_0 = s_i)P(X'_0 = s_i) \\
 &\geq \alpha \sum_{i=1}^k P(X_0 = s_i) \alpha \sum_{i=1}^k P(X'_0 = s_i) = \alpha^2.
 \end{aligned}$$

Luego, $P(T > M) \leq 1 - \alpha^2$. De la misma forma, condicionando a lo que pasó a tiempo M , podemos acotar $P(X_{2M} = X'_{2M})$, luego

$$P(X_{2M} \neq X'_{2M} | T > M) \leq 1 - \alpha^2.$$

Por lo tanto,

$$\begin{aligned}
 P(T > 2M) &= P(T > M)P(T > 2M | T > M) \\
 &\leq (1 - \alpha^2)P(X_{2M} \neq X'_{2M} | T > M) \\
 &\leq (1 - \alpha^2)^2.
 \end{aligned}$$

Iterando este argumento, para cualquier l ,

$$P(T > lM) \leq (1 - \alpha^2)^l.$$

Luego,

$$\lim_{n \rightarrow \infty} P(T > n) = 0.$$

Es decir, que las dos cadenas se encuentran con probabilidad 1. Vamos a construir una tercera cadena de Markov (X''_0, X''_1, \dots) . Sea $X''_0 = X_0$ y para cada n :

$$X''_{n+1} = \begin{cases} \Phi(X''_n, U_{n+1}) & \text{si } X''_n \neq X'_n \\ \Phi(X''_n, U'_{n+1}) & \text{si } X''_n = X'_n. \end{cases}$$

En otras palabras, esta nueva cadena es exactamente (X_0, X_1, \dots) hasta T , y luego (X'_0, X'_1, \dots) . Observemos que (X''_0, X''_1, \dots) es una cadena de Markov con matriz de transición P , pues las (U_n) son independientes de las $(U'_n)_{n \in \mathbb{N}}$.

La variable X''_0 tiene distribución inicial $\mu^{(0)}$. Luego, para todo n , X''_n tiene distribución $\mu^{(n)}$. Para cualquier $i \in \{1, \dots, k\}$

$$\begin{aligned}
 \mu^{(n)} - \pi_i &= P(X''_n = s_i) - P(X'_n = s_i) \leq P(X''_n = s_i, X'_n \neq s_i) \\
 &\leq P(X''_n \neq X'_n) = P(T > n).
 \end{aligned}$$

Cambiando los roles de X_n'' y X_n' , obtenemos

$$\pi_i - \mu_i^{(n)} \leq P(T > n).$$

Luego,

$$|\pi_i - \mu_i^{(n)}| \leq P(T > n) \xrightarrow{n \rightarrow \infty} 0.$$

Lo que implica,

$$\lim_{n \rightarrow \infty} d_{VT}(\mu^{(n)}, \pi) = \lim_{n \rightarrow \infty} \frac{1}{2} \sum_{i=1}^k |\pi_i - \mu_i^{(n)}| = 0.$$

□

Teorema 1.5 (Teorema de Unicidad). *Para cualquier cadena de Markov irreducible y aperiódica definida en un espacio de estados finito existe una única distribución estacionaria.*

Demostración. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica, con matriz de transición P . Sabemos que existe al menos una distribución estacionaria. Sean π y $\tilde{\pi}$ dos distribuciones estacionarias de P , veamos que $\pi = \tilde{\pi}$. Si la cadena empieza con distribución inicial $\mu^{(0)} = \tilde{\pi}$ entonces $\mu^{(n)} = \tilde{\pi}, \forall n \in \mathbb{N}$. Por otro lado, vimos que $\lim d_{VT}(\mu^{(n)}, \pi) = 0$. Como $\mu^{(0)} = \tilde{\pi}$ es lo mismo que $\lim d_{VT}(\tilde{\pi}, \pi) = 0$, como $d_{VT}(\tilde{\pi}, \pi)$ no depende de n ,

$$d_{VT}(\tilde{\pi}, \pi) = 0,$$

y por lo tanto,

$$\pi = \tilde{\pi}.$$

□

1.6. Promedios empíricos: el Teorema Ergódico

En esta sección trataremos el Teorema Ergódico para cadenas de Markov. Vamos a ver condiciones para garantizar la convergencia en probabilidad del promedio empírico.

Definición 1.13. Si $X_0 = s_1$, definimos el tiempo de parada T_{11}^k , como la k -ésima vez que la cadena vuelve al estado s_1 , es decir,

$$T_{11}^{k+1} = \min\{n > T_{11}^k / X_n = s_1\}.$$

Definición 1.14. Sea $C(n)$ la cantidad de veces que la cadena pasa por el estado s_1 antes del tiempo n .

$$C(n) = \sum_{i=1}^n \mathbf{1}_{\{X_i = s_1\}}.$$

Lema 1.4. Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y recurrente en $\mathcal{S} = \{s_1, \dots, s_k\}$ y sea $f : \mathcal{S} \rightarrow \mathbb{R}$. Para cualquier distribución inicial μ ;

$$\lim_{n \rightarrow \infty} \frac{1}{C(n)} \sum_{i=1}^n f(X_i) = \sum_{j=1}^k f(s_j) \rho_j \quad c.s.$$

Demostración. Sea $i \in \{1, \dots, k\}$, y supongamos $X_0 = s_1$. Consideremos las siguientes variables aleatorias,

$$U_P = \sum_{n=T_{11}^P+1}^{T_{11}^{P+1}} f(X_n),$$

U_P depende de $X_{T_{11}^P+1}, \dots, X_{T_{11}^{P+1}}$, por lo tanto las variables aleatorias $(U_P)_{P \geq 1}$ son i.i.d. para cualquier distribución inicial. Asumimos $f \geq 0$ y usando la Propiedad de Markov Fuerte 1.2,

$$\begin{aligned} E_\mu(U_P) &= E(U_1) = E\left(\sum_{n=1}^{T_{11}} f(X_n)\right) \\ &= E\left(\sum_{n=1}^{T_{11}} \sum_{j=1}^k f(s_j) \mathbf{1}_{\{X_n=j\}}\right) \\ &= \sum_{j=1}^k f(s_j) E\left(\sum_{n=1}^{T_{11}} \mathbf{1}_{\{X_n=j\}}\right). \end{aligned}$$

Luego,

$$E_\mu(U_P) = \sum_{j=1}^k f(s_j) \rho_j.$$

Como el espacio de estados es finito, y el tiempo esperado de retorno a cualquier estado también es finito, la esperanza de las U_P es finita. Entonces, estamos en condiciones de aplicar la Ley de los Grandes Números a las variables aleatorias $(U_P)_{P \in \mathbb{N}}$ y obtenemos

$$\frac{1}{n} \sum_{P=1}^n U_P \xrightarrow[n \rightarrow \infty]{c.s.} \sum_{j=1}^k f(s_j) \rho_j.$$

Además como,

$$\frac{1}{n} \sum_{P=1}^n U_P = \frac{1}{n} \sum_{P=1}^n \sum_{n=T_{11}^P+1}^{T_{11}^{P+1}} f(X_n) = \frac{1}{n} \sum_{n=T}^{T_{11}^{n+1}} f(X_n)$$

y como

$$T_{11}^{C(n)} \leq n \leq T_{11}^{C(n)+1},$$

resulta

$$\frac{\sum_{j=1}^{T_{11}^{C(n)}} f(X_j)}{C(n)} \leq \frac{\sum_{j=1}^n f(X_j)}{C(n)} \leq \frac{\sum_{j=1}^{T_{11}^{C(n)+1}} f(X_j)}{C(n)}.$$

Por otro lado, $(X_n)_{n \in \mathbb{N}}$ es recurrente por lo tanto $C(n) \xrightarrow{n \rightarrow \infty} \infty$. Luego,

$$\frac{\sum_{j=1}^n f(X_j)}{C(n)} \xrightarrow{n \rightarrow \infty} \sum_{j=1}^k f(s_j) \rho_j.$$

En el caso de una función arbitraria f , se puede considerar $f^+ = \max\{0, f\}$ y $f^- = \max\{0, -f\}$, tanto f^+ como $-f^-$ cumplen el lema, y de la diferencia obtenemos el resultado requerido. \square

Teorema 1.6 (Teorema Ergódico). *Sea $(X_n)_{n \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica, definida en un espacio de estados $S = \{s_1, \dots, s_k\}$ y con distribución estacionaria π y sea $f : S \rightarrow \mathbb{R}$. Luego, para cualquier distribución inicial μ :*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f(X_j) = \sum_{i=1}^k f(s_i) \pi_i = E_\pi(f(X_1)).$$

Demostración. Por el lema anterior,

$$\lim_{n \rightarrow \infty} \frac{1}{C(n)} \sum_{i=1}^n f(X_i) = \sum_{j=1}^k f(s_j) \rho_j.$$

Si $f \equiv 1$,

$$\frac{n}{C(n)} \xrightarrow{n \rightarrow \infty} \sum_{j=1}^k \rho_j = \tau_{11}.$$

Luego,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(X_i) = \lim_{n \rightarrow \infty} \frac{C(n)}{n} \frac{1}{C(n)} \sum_{i=1}^n f(X_i) = \frac{\sum_{j=1}^k f(s_j) \rho_j}{\tau_{11}} = \sum_{j=1}^k f(s_j) \pi_j.$$

\square

Capítulo 2

Distribuciones Cuasiestacionarias

Muchos procesos estocásticos tienen un estado absorbente, lo que significa que si el proceso llega al estado absorbente, se queda ahí para siempre. En un proceso de Markov con un estado absorbente, buscar la distribución estacionaria no tiene sentido, ya que si dejamos evolucionar el proceso un buen rato, el proceso estará en el estado absorbente. Veremos que bajo ciertas condiciones, si el proceso tiene un estado absorbente, con probabilidad 1 será absorbido. Por lo tanto, en estos casos nos interesa estudiar qué ocurre antes de ser absorbido. Si la cadena está definida en un espacio de estados \mathcal{S} , cualquier elemento o cualquier subconjunto de \mathcal{S} podrían ser estados absorbentes. En este trabajo el estado absorbente será el 0, sólo para simplificar la notación.

Definición 2.1. *Dada una cadena de Markov definida en un espacio de estados $\mathcal{S} \cup \{0\}$ con $\mathcal{S} = \{s_1, \dots, s_k\}$ y matriz de transición P , con entradas P_{ij} , $i, j \in \{0, 1, \dots, k\}$. Si existe $j \in \{1, \dots, k\}$ tal que $P_{j0} > 0$, decimos que 0 es un estado absorbente si $P_{0j} = 0$ para todo $j \in \{1, \dots, k\}$.*

Para analizar cuándo el proceso es absorbido, debemos definir el tiempo de absorción. Luego veremos que las hipótesis necesarias para asegurarnos que la cadena es absorbida con probabilidad 1, son irreducibilidad y aperiodicidad sobre \mathcal{S} .

Definición 2.2. *Definimos el tiempo de parada T_0 como el tiempo de absorción, es decir:*

$$T_0 = \min\{n \in \mathbb{N} / X_n = 0\}.$$

y notamos $T_0 = \infty$ si $X_n \neq 0$ para todo $n \in \mathbb{N}_0$.

Sea P_S la matriz P restringida a \mathcal{S} , es decir, la matriz P sin la fila y la columna correspondientes al 0.

Lema 2.1. *Sea una cadena de Markov, con un estado absorbente definida en el espacio de estados finito $\mathcal{S} \cup \{0\}$ con matriz de transición P y P_S irreducible y aperiódica. Dada μ una distribución inicial,*

$$P_\mu(T_0 = \infty) = 0.$$

Demostración. La matriz P_S es una matriz sub-estocástica, es decir, la suma de las coordenadas de cada fila es menor a 1, sin embargo como es irreducible y aperiódica podemos usar el Corolario 1.1 pues puede verse en su demostración que no es necesario que sea una matriz estocástica. Por lo tanto existe $N \in \mathbb{N}$ tal que $P_{ij}^n > 0, \forall n \geq N, \forall s_i, s_j \in S$. Por otro lado, la probabilidad de ir de algún estado de S al 0 es positiva, es decir, $P_{j0} > 0$ para algún $s_j \in S$. Luego, existe $M \in \mathbb{N}$ tal que $P_{i0}^M > 0, \forall s_i \in S$. Sean $\alpha = \min\{P_{i0}^M > 0 / j = \{1, \dots, k\}\}$ y μ la distribución inicial. Entonces para $n \in \mathbb{N}$, $P_\mu(T_0 > n) = \sum_{i=1}^k P_i(T_0 > n)\mu_i$. Acotemos $P_i(T_0 > M)$ para cada $i \in \{1, \dots, k\}$

$$\begin{aligned} P_i(T_0 > M) &\leq P_i(X_1 \neq 0, \dots, X_M \neq 0) \\ &\leq P_i(X_M \neq 0) \leq 1 - \alpha. \end{aligned}$$

Condicionando a lo sucedido a tiempo M obtenemos:

$$\begin{aligned} P_i(T_0 > 2M) &= P_i(T_0 > 2M | T_0 > M)P_i(T_0 > M) \\ &\leq P_i(X_{2M} \neq 0 | T_0 > M)P_i(T_0 > M) \\ &= P_i(X_{2M} \neq 0 | X_M \neq 0)P_i(T_0 > M) \\ &\leq (1 - \alpha)^2. \end{aligned}$$

Iterando,

$$P_i(T_0 > lM) \leq (1 - \alpha)^l \quad \forall i \in \{1, \dots, k\}.$$

Luego, $P_\mu(T_0 > lM) = \sum_{i=1}^k P_i(T_0 > n)\mu_i \leq (1 - \alpha)^l$ y por lo tanto,

$$P_\mu(T_0 = \infty) = 0.$$

□

Del lema se ve que la cadena es absorbida en un tiempo finito con probabilidad 1 y además que $P_\mu(T_0 > n)$ tiende a cero exponencialmente.

Corolario 2.1. *Todos los momentos de T_0 son finitos.*

Demostración. Sea μ la distribución inicial,

$$\begin{aligned} E_\mu(T_0^l) &= \sum_{m=1}^{\infty} m^l P_\mu(T_0 = m) \leq \sum_{m=1}^{\infty} m^l P_\mu(T_0 > m-1) \\ &= \sum_{n=0}^{\infty} (n+1)^l P_\mu(T_0 > n) \leq \sum_{n=0}^{\infty} (n+1)^l (1 - \alpha)^{[n/M+1]}. \end{aligned}$$

De donde, para todo $l \in \mathbb{N}$, $E_\mu(T_0^l)$ es finito. □

Para un proceso que empieza con distribución inicial μ , definimos la evolución del proceso condicionada a no ser absorbido como:

$$\varphi_j^\mu(n) = P_\mu(X_n = s_j | T_0 > n)$$

equivalentemente,

$$\begin{aligned} \varphi_j^\mu(n) &= \frac{P_\mu(X_n = s_j)}{P_\mu(T_0 > n)} = \frac{P_\mu(X_n = s_j)}{P_\mu(X_n \neq 0)} = \frac{P_\mu(X_n = s_j)}{\sum_{i=1}^k P_\mu(X_n = s_i)} \\ &= \frac{\sum_{z=1}^k P_\mu(X_n = s_i | X_0 = s_z) P_\mu(X_0 = s_z)}{\sum_{i=1}^k \sum_{z=1}^k P_\mu(X_n = s_i | X_0 = s_z) P_\mu(X_0 = s_z)} \\ &= \frac{\sum_{z=1}^k \mu_z(P^n)_{zj}}{\sum_{i=1}^k \sum_{z=1}^k \mu_z(P^n)_{zi}} \\ &= \frac{(\mu P^n)_j}{\sum_{i=1}^k (\mu P^n)_i}. \end{aligned}$$

De aquí se ve que $\sum_{j=1}^k \varphi_j^\mu(n) = 1$. Queremos analizar el comportamiento asintótico de φ_n^μ , es decir, $\lim_{n \rightarrow \infty} \varphi_n^\mu$.

Definición 2.3. *Cuando existe $\lim_{n \rightarrow \infty} \varphi_n^\mu$ y es una medida de probabilidad se lo denomina Límite de Yaglom para μ . Denotamos*

$$\mathbf{v}_j = \lim_{n \rightarrow \infty} \varphi_j^\mu(n).$$

Al igual que con las medidas invariantes, cuando este límite existe, es invariante para la evolución de la probabilidad condicionada. A estas medidas se las llama distribuciones cuasiestacionarias.

Definición 2.4. *El vector $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_k)$ se dice que es una distribución cuasiestacionaria si*

$$1. \quad \mathbf{v}_i \geq 0 \quad \forall i \in \{1, \dots, k\} \text{ y } \sum_{i=1}^k \mathbf{v}_i = 1,$$

$$2. \quad \varphi_j^\mathbf{v}(n) = \mathbf{v}_j \quad \forall n \in \mathbb{N}_0 \quad \forall j = \{1, \dots, k\}.$$

La primer condición nos dice que es una medida y la segunda que es invariante para el proceso

condicionado. Dada μ una distribución inicial;

$$\begin{aligned}
 \varphi_j^\mu(n+1) &= P_\mu(X_{n+1} = s_j | X_{n+1} \neq 0) = \frac{P_\mu(X_{n+1} = s_j)}{P_\mu(X_{n+1} \neq 0)} \\
 &= \frac{\sum_{i=1}^k P(X_{n+1} = s_j | X_n = s_i) P_\mu(X_n = s_i)}{\sum_{z=1}^k P_\mu(X_{n+1} = s_z)} \\
 &= \frac{\sum_{i=1}^k P(X_{n+1} = s_j | X_n = s_i) P_\mu(X_n = s_i)}{\sum_{z=1}^k \sum_{i=1}^k P(X_{n+1} = s_z | X_n = s_i) P_\mu(X_n = s_i)} \\
 &= \frac{\sum_{i=1}^k P_{ij} \varphi_i^\mu(n)}{\sum_{z=1}^k \sum_{i=1}^k P_{iz} \varphi_i^\mu(n)} = \frac{\sum_{i=1}^k P_{ij} \varphi_i^\mu(n)}{\sum_{i=1}^k (\varphi_i^\mu(n) \sum_{z=1}^k P_{iz})} \\
 &= \frac{\sum_{i=1}^k P_{ij} \varphi_i^\mu(n)}{\sum_{i=1}^k (\varphi_i^\mu(n)(1 - P_{i0}))} = \frac{\sum_{i=1}^k P_{ij} \varphi_i^\mu(n)}{1 - \sum_{i=1}^k \varphi_i^\mu(n) P_{i0}}.
 \end{aligned}$$

Luego se tiene

$$\varphi_j^\mu(n+1) = \sum_{i=1}^k P_{ij} \varphi_i^\mu(n) + \sum_{i=1}^k \varphi_i^\mu(n) P_{i0} \varphi_j^\mu(n+1).$$

Por lo tanto v es una distribución cuasiestacionaria sí y solo sí

$$v_j = \sum_{i=1}^k P_{ij} v_i + \sum_{i=1}^k P_{i0} v_i v_j.$$

Por otra parte si v es el límite de Yaglom, verifica:

$$v_j = \sum_{i=1}^k P_{ij} v_i + \sum_{i=1}^k P_{i0} v_i v_j.$$

y por lo tanto es una distribución cuasiestacionaria.

Queremos analizar existencia y unicidad de la distribución cuasiestacionaria, para ello vamos a utilizar herramientas de álgebra lineal, ya que el comportamiento de la cadena depende de la matriz de transición.

Teorema 2.1 (Perron-Frobenius). *Sea A una matriz no negativa e irreducible, de $r \times r$. Entonces, existe un autovalor λ_1 real, positivo, con multiplicidad 1 y tal que $\lambda_1 > |\lambda_j|$ para todo λ_j autovalor de A . Además sean u, v autovectores a izquierda y derecha respectivamente asociados a λ_1 , podemos elegirlos positivos y tal que $|v| = 1$ y $u^t v = 1$. Sean $\lambda_2, \dots, \lambda_r$ los demás autovalores de A , tales que $\lambda_1 > |\lambda_2| \geq \dots \geq |\lambda_r|$. Sea m_j la multiplicidad de λ_j . Luego,*

$$A^n = \lambda_1^n v u^t + O(n^{m_2-1} |\lambda_2|^n), \quad \forall n \in \mathbb{N}.$$

Demostración. Para la demostración ver Seneta [12]. □

Ahora sí estamos en condiciones de mostrar existencia de una distribución cuasiestacionaria.

Teorema 2.2. *Para una cadena de Markov, con un estado absorbente definida en un espacio de estados $S \cup \{0\}$, con $S = \{s_1, \dots, s_k\}$ y matriz de transición P , con P_S irreducible y aperiódica, existe una única distribución cuasiestacionaria, que es el límite de Yaglom.*

Demostración. Por el Teorema de Perron-Frobenius (2.1) aplicado a la matriz P_S , existe un autovalor λ_1 real, positivo, con multiplicidad 1 y tal que $\lambda_1 > |\lambda_j|$ para todo λ_j autovalor de P_S , y existen u y v autovectores a izquierda y derecha asociados a λ_1 tales que,

$$(P_S^n)_{ij} = \lambda_1^n v_i u_j + O(n^{m_2-1} |\lambda_2|^n) \quad \forall n \in \mathbb{N}. \quad (2.1)$$

Queremos ver que existe el límite de Yaglom, para ello busquemos una expresión para la probabilidad condicionada usando la igualdad de (2.1). Sea μ la distribución inicial,

$$\begin{aligned} P_\mu(X_n = s_j | T_0 > n) &= \frac{P_\mu(X_n = s_j)}{P_\mu(T_0 > n)} = \frac{\sum_{i=1}^k P_i(X_n = s_j) P(X_0 = s_i)}{\sum_{i=1}^k P_i(T_0 > n) P(X_0 = s_i)} \\ &= \frac{\sum_{i=1}^k (P_S^n)_{ij} \mu_i}{\sum_{i=1}^k \sum_{z=1}^k (P_S^n)_{iz} \mu_i} \\ &= \frac{\sum_{i=1}^k (\lambda_1^n v_i u_j + O(n^{m_2-1} |\lambda_2|^n)) \mu_i}{\sum_{i=1}^k \sum_{z=1}^k (\lambda_1^n v_i u_z + O(n^{m_2-1} |\lambda_2|^n)) \mu_i} \\ &= \frac{\sum_{i=1}^k \lambda_1^n v_i u_j \mu_i + O(n^{m_2-1} |\lambda_2|^n) \sum_{i=1}^k \mu_i}{\sum_{i=1}^k \sum_{z=1}^k \lambda_1^n v_i u_z \mu_i + O(n^{m_2-1} |\lambda_2|^n) \sum_{i=1}^k \sum_{z=1}^k \mu_i} \\ &= \frac{\sum_{i=1}^k \lambda_1^n v_i u_j \mu_i + O(n^{m_2-1} |\lambda_2|^n)}{\sum_{i=1}^k \sum_{z=1}^k \lambda_1^n v_i u_z \mu_i + O(n^{m_2-1} |\lambda_2|^n) k} \\ &= \frac{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i}{\lambda_1^n (\sum_{z=1}^k u_z) (\sum_{i=1}^k v_i \mu_i) + O(n^{m_2-1} |\lambda_2|^n) k} \\ &\quad + \frac{O(n^{m_2-1} |\lambda_2|^n)}{\sum_{i=1}^k \sum_{z=1}^k \lambda_1^n v_i u_z \mu_i + O(n^{m_2-1} |\lambda_2|^n) k} \\ &= \left[\frac{(\sum_{z=1}^k u_z) (\sum_{i=1}^k v_i \mu_i)}{u_j \sum_{i=1}^k v_i \mu_i} + \frac{O(n^{m_2-1} |\lambda_2|^n) k}{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i} \right]^{-1} + O(n^{m_2-1} \left| \frac{\lambda_2}{\lambda_1} \right|^n) \\ &= \left[\frac{\sum_{z=1}^k u_z}{u_j} + \frac{O(n^{m_2-1} |\lambda_2|^n) k}{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i} \right]^{-1} + O(n^{m_2-1} \left| \frac{\lambda_2}{\lambda_1} \right|^n). \end{aligned}$$

Expresemos de otra forma el primer término,

$$\left[\frac{\sum_{z=1}^k u_z}{u_j} + \frac{O(n^{m_2-1} |\lambda_2|^n) k}{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i} \right]^{-1} = \frac{u_j}{\sum_{z=1}^k u_z} - \frac{1}{\xi^2} \frac{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i}{O(n^{m_2-1} |\lambda_2|^n) k},$$

con $\xi \in \left(\frac{u_j}{\sum_{z=1}^k u_z}, \frac{u_j}{\sum_{z=1}^k u_z} + \frac{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i}{O(n^{m_2-1} |\lambda_2|^n) k} \right)$. Entonces se tiene

$$P_\mu(X_n = s_j | T_0 > n) = \frac{u_j}{\sum_{z=1}^k u_z} - \frac{1}{\xi^2} \frac{\lambda_1^n u_j \sum_{i=1}^k v_i \mu_i}{O(n^{m_2-1} |\lambda_2|^n) k} + O(n^{m_2-1} \left| \frac{\lambda_2}{\lambda_1} \right|^n).$$

Luego,

$$\lim_{n \rightarrow \infty} P_\mu(X_n = s_j | T_0 > n) = \frac{u_j}{\sum_{i=1}^k u_i} = v_j.$$

Por lo tanto, para cualquier distribución inicial, el Límite de Yaglom es el autovector a izquierda asociado a λ_1 de P_S :

$$v P_S = \lambda_1 v.$$

Veamos que v es una distribución cuasiestacionaria. Es una probabilidad ya que, $v_j > 0, \forall j = \{1, \dots, k\}$ y $\sum_{j=1}^k v_j = \frac{\sum_{j=1}^k u_j}{\sum_{i=1}^k u_i} = 1$, y es invariante para la distribución inicial,

$$\begin{aligned} \Phi_j^v(n) &= \frac{P_v(X_n = s_j)}{\sum_{z=1}^k P_v(X_n = s_z)} \\ &= \frac{\sum_{i=1}^k P_i(X_n = s_j) v_i}{\sum_{z=1}^k \sum_{i=1}^k P_i(X_n = s_z) v_i} \\ &= \frac{\sum_{i=1}^k (P_S^n)_{ij} \frac{u_i}{\sum_{z=1}^k u_z}}{\sum_{z=1}^k \sum_{i=1}^k \frac{u_i}{\sum_{z=1}^k u_z} (P_S^n)_{iz}} \\ &= \frac{\sum_{i=1}^k u_i (P_S^n)_{ij}}{\sum_{z=1}^k \sum_{i=1}^k u_i (P_S^n)_{iz}} \\ &= \frac{\lambda_1 u_j}{\sum_{z=1}^k \lambda_1 u_z} \\ &= \frac{u_j}{\sum_{z=1}^k u_z} \\ &= v_j. \end{aligned}$$

Con lo cual, probamos que existe una distribución cuasiestacionaria. Veamos que es única.

Sean v y \tilde{v} dos distribuciones cuasiestacionarias. Si la cadena empieza con distribución inicial $\mu = \tilde{v}$ entonces $\phi^\mu(n) = \tilde{v}, \forall n \in \mathbb{N}$.

Por otro lado vimos que para cualquier distribución inicial μ , como $\lim_{n \rightarrow \infty} \phi^\mu(n) = v$ resulta que $\lim_{n \rightarrow \infty} d_{TV}(\phi^\mu(n), v) = 0$. Tomando $\mu = \tilde{v}$ obtenemos $\lim d_{TV}(\tilde{v}, v) = 0$, como $d_{TV}(\tilde{v}, v)$ no depende de n ,

$$d_{TV}(\tilde{v}, v) = 0$$

y por lo tanto,

$$v = \tilde{v}.$$

□

Capítulo 3

Procesos de Ramificación

Este capítulo se basa en el estudio de procesos de ramificación, que nos serán de mucha utilidad para analizar el comportamiento de un método para simular las distribuciones cuasiestacionarias. Esta basado principalmente en los libros de Harris [7] y en Athreya-Ney [2].

El primer estudio de procesos de ramificación se dio porque existía la preocupación de que los apellidos aristocráticos se estaban extinguiendo. El origen de las investigaciones se atribuye a los estudios realizados, independientemente, por Bienaymé en Francia y por Galton y Watson en Inglaterra (1873).

Los procesos de ramificación han sido utilizados como modelos matemáticos para describir procesos empíricos, relacionados con la Biología, la Física nuclear, la Medicina, las ciencias de la Computación, la Demografía, etc.

La estructura de este capítulo es la siguiente, comenzamos con el proceso de ramificación llamado Galton-Watson, es un proceso a tiempo discreto que representa la cantidad de individuos de las generaciones sucesivas. Estudiamos varias propiedades, calculamos la probabilidad de extinción, y analizamos el comportamiento del proceso cuando no se extingue. Seguimos con una generalización, que es considerar el tiempo de vida continuo, analizamos propiedades y vemos la similitud con el caso discreto. Por último, consideramos distintos tipos de individuos, tratamos el tiempo continuo, y principalmente analizamos el comportamiento del proceso cuando no se extingue. Varios de los resultados de este último caso pueden encontrarse en Georgii-Baake [5].

3.1. Proceso de Galton-Watson

Suponemos que el proceso se inicia con un individuo, el cual constituye la generación cero, sus hijos forman parte de la primera generación, sus nietos la segunda, y así sucesivamente. De esta manera denotamos la variable aleatoria Z_n como el número de individuos de la n -ésima generación.

Sea p_i la probabilidad de que un individuo deje i descendientes, con $i \in \mathbb{N}_0$. Para cada $j \in \mathbb{N}$, $k \in \mathbb{N}_0$, ξ_j^k representa el número de hijos del j -ésimo miembro de la k -ésima generación.

Podemos expresar a Z_{n+1} a partir de $\{\xi_j^k, 0 \leq k \leq n, j \geq 1\}$,

$$Z_{n+1} = \xi_1^n + \dots + \xi_{Z_n}^n.$$

Asumimos que $\{\xi_j^k, k \geq 0, j \geq 1\}$ son variables aleatorias i.i.d., donde $P(\xi_1^1 = i) = p_i \quad \forall i \in \mathbb{N}_0$. Si comenzamos con i individuos, notamos $Z_n^{(i)}$, $n \in \mathbb{N}$. Este proceso es la suma de i procesos de ramificación independientes que se inician con un solo individuo. Para evitar situaciones triviales, vamos a suponer que $p_0 + p_1 < 1$. Es decir, la probabilidad de tener a lo sumo un individuo es menor que 1.

Estamos interesados en estudiar si la familia perdura en el tiempo, en este caso nos preguntamos si al pasar las generaciones el apellido seguirá existiendo. Para esto tendríamos que conocer la distribución de Z_n , para así poder calcular $P(Z_n = 0)$ para n suficientemente grande. Pero esto es complicado ya que la distribución de la cantidad de individuos en la generación n -ésima depende de las generaciones anteriores.

Una función que será muy útil en el estudio de procesos de ramificación es la función generadora de momentos de la variable Z_1 .

Sea $\Phi : [0, 1] \rightarrow \mathbb{R}_+$ definida por $\Phi(s) = E(s^{Z_1}) = E(s^{\xi_1^1})$. Consideraremos también sus composiciones,

$$\Phi_0(s) = s \quad \Phi_1(s) = \Phi(s), \tag{3.1}$$

$$\Phi_{n+1}(s) = \Phi(\Phi_n(s)) \quad n \in \mathbb{N}. \tag{3.2}$$

Proposición 3.1. *La función generadora de Z_n es $\Phi_n(s)$, es decir, componer n veces la función generadora de Z_1 .*

Demostración. Sea $\Phi_{(n)}$ la función generadora asociada a Z_n :

$$\Phi_{(n)}(s) = \sum_{k \in \mathbb{N}_0} P(Z_n = k) s^k.$$

Como $\Phi(s) = \Phi_{(1)}(s)$, basta probar que la sucesión de funciones $\Phi_{(n)}$ cumple una relación de recurrencia como (3.2). Por como la definimos,

$$\Phi_{(n+1)}(s) = E(s^{Z_{n+1}}) = E(s^{\xi_1^n + \dots + \xi_{Z_n}^n}) = \sum_{k \in \mathbb{N}_0} E(\mathbf{1}_{\{Z_n=k\}} s^{\xi_1^n + \dots + \xi_k^n}).$$

La variable Z_n es independiente de las variables $\xi_j^n, j \geq 1$, por lo tanto,

$$E(\mathbf{1}_{\{Z_n=k\}} s^{\xi_1^n + \dots + \xi_k^n}) = E(\mathbf{1}_{\{Z_n=k\}}) E(s^{\xi_1^n + \dots + \xi_k^n}) = P(Z_n = k) E(s^{\xi_1^n + \dots + \xi_k^n}).$$

Como las variables, ξ_1^n, \dots, ξ_k^n son independientes e idénticamente distribuidas

$$E(s^{\xi_1^n + \dots + \xi_k^n}) = \prod_{j=1}^k E(s^{\xi_j^n}) = (E(s^{\xi_1^1}))^k = (\Phi(s))^k.$$

Luego,

$$\Phi_{(n+1)}(s) = \sum_{k \in \mathbb{N}_0} P(Z_n = k)(\Phi(s))^k = \Phi_{(n)}(\Phi(s)).$$

□

Estas funciones nos van a ayudar para calcular los momentos del proceso. Cuando estos existan, los podemos expresar en términos de las derivadas de $\Phi_n(s)$ para $s = 1$.

Teorema 3.1. *Para el proceso de ramificación ya definido, si $E(Z_1) = m$, y $V(Z_1) = \sigma^2$ entonces*

$$E(Z_n) = m^n,$$

$$V(Z_n) = \begin{cases} \frac{\sigma^2 m^n (m^n - 1)}{m^2 - m} & \text{si } m \neq 1 \\ n\sigma^2 & \text{si } m = 1. \end{cases}$$

Demostración. Para la media de la primera generación tenemos,

$$E(Z_1) = \sum_{j \in \mathbb{N}_0} p_j j = \Phi'(1) = m.$$

Para la generación n -ésima,

$$\begin{aligned} E(Z_n) &= \sum_{j \in \mathbb{N}_0} P(Z_n = j) j = \Phi'_n(1) = (\Phi_{n-1}(\Phi(1)))' = \Phi'_{n-1}(\Phi(1))\Phi'(1) \\ &= \Phi'_{n-1}(1)\Phi'(1) = \dots = (\Phi'(1))^n = m^n. \end{aligned}$$

Para calcular la varianza,

$$\Phi''(1) = \sum_{k \in \mathbb{N}_0} k(k-1)P(Z_1 = k) = E(Z_1^2) - E(Z_1) = V(Z_1) + m(m-1).$$

Si $\sigma^2 = V(Z_1)$. Luego, $\Phi''(1) = \sigma^2 + m(m-1)$

$$\begin{aligned}
\Phi''_{n+1}(1) &= \Phi''(1)(\Phi'_n(1))^2 + \Phi'(1)\Phi''_n(1) \\
&= \Phi''(1)m^{2n} + m\Phi''_n(1) \\
&= \Phi''(1)m^{2n} + m[\Phi''(1)m^{2n-2} + m\Phi''_{n-1}(1)] \\
&= \Phi''(1)m^{2n} + \Phi''(1)m^{2n-1} + m^2\Phi''_{n-1}(1) \\
&= \dots \\
&= \Phi''(1)[m^{2n} + m^{2n-1} + \dots + m^{n+1}] \\
&= [\sigma^2 + m(m-1)]m^{n+1} \sum_{j=0}^n m^k \\
&= \sigma^2 m^{n+1} \sum_{j=0}^n m^k + m(m-1)m^n \sum_{j=0}^n m^k \\
&= \sigma^2 m^{n+1} \sum_{j=0}^n m^k + m^n(m^{n+2} - m).
\end{aligned}$$

Si $m = 1$,

$$\Phi''_{n+1}(1) = \sigma^2 n.$$

Si $m \neq 1$,

$$\Phi''_{n+1}(1) = \sigma^2 m^{n+1} \frac{m^{n+1} - 1}{m - 1} + m^n(m^{n+2} - m).$$

Como

$$\Phi''_{n+1}(1) = V(Z_{n+1}) - E(Z_{n+1}) + E(Z_{n+1})^2.$$

Obtenemos que si $m = 1$, $V(Z_{n+1}) = \sigma^2 n$. Si $m \neq 1$,

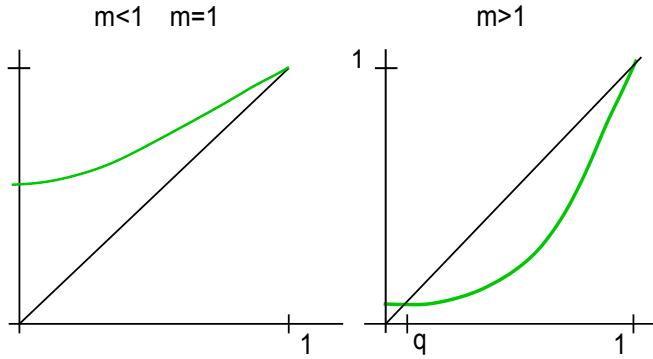
$$V(Z_{n+1}) = \sigma^2 m^{n+1} \frac{m^{n+1} - 1}{m - 1} + m^n(m^{n+2} - m) + m^{n+1} - [m^{n+1}]^2 = \sigma^2 m^{n+1} \frac{m^{n+1} - 1}{m - 1}.$$

□

Queremos analizar el comportamiento del proceso, para ello necesitamos algunas propiedades de la funciones generadoras. Luego, veremos que de lo único que depende que el proceso se extinga es de la media. Además cuando no se extingue, queremos analizar el comportamiento de la población.

Lema 3.1. *La función Φ cumple las siguientes propiedades:*

1. Φ es estrictamente creciente y convexa.
2. $\Phi(0) = p_0$, $\Phi(1) = 1$.
3. Si $m \leq 1$ entonces $\Phi(s) > s$, para todo $s \in [0, 1]$.



4. Si $m > 1$ entonces $\Phi(s) = s$ tiene una única solución en $[0, 1)$.

Demostración. 1. Veamos las derivadas de Φ , como todas las sumas son absolutamente convergentes para $|s| < 1$ podemos intercambiar los límites.

$$\Phi(s) = \sum_{k=0}^{\infty} p_k s^k,$$

$$\Phi'(s) = \sum_{k=1}^{\infty} p_k k s^{k-1} > 0 \quad \forall s \in (0, 1),$$

$$\Phi''(s) = \sum_{k=2}^{\infty} p_k k(k-1) s^{k-2} > 0 \quad \forall s \in (0, 1) \text{ (pues supusimos } p_0 + p_1 < 1\text{)}.$$

Por lo tanto Φ es estrictamente creciente y convexa

2. Es inmediato
 3. Consideremos la función

$$g : [0, 1] \rightarrow \mathbb{R} \quad g(s) = \Phi(s) - s.$$

Sus derivadas, $g'(s) = \Phi'(s) - 1$, $g''(s) = \Phi''(s)$. Como Φ es estrictamente convexa en $[0, 1]$ y $\Phi(1) = 1$, la función $g(s)$ también es estrictamente convexa y $g(1) = 0$. Luego la derivada $g'(s)$ es una función estrictamente creciente, y si suponemos $m \leq 1$, $g'(1) = \Phi'(1) - 1 = m - 1 \leq 0$, por lo tanto g' es estrictamente decreciente en $[0, 1]$. Como $g(1) = \Phi(1) - 1 = 0$, $g(s) > 0$ para $s \in [0, 1)$ y $\Phi(s) > s$ en $[0, 1)$.

4. Sea $m > 1$, y g la función definida anteriormente. Como $g''(s) > 0 \forall s \in (0, 1)$, g es convexa en $(0, 1)$ y g' es estrictamente creciente, por lo tanto g tiene a lo sumo un punto crítico en $(0, 1)$. Como $g'(1) = m - 1 > 0$ y $g(1) = \Phi(1) - 1 = 0$, existe $s_0 < 1$ con $g(s_0) < 0$. Por otro lado, $g(0) = p_0 > 0$. Luego existe $q \in (0, 1)$ tal que $g(q) = 0$, es decir, $\Phi(q) = q$.

□

Teorema 3.2. *Sea $(Z_n)_{n \geq 0}$ un proceso de ramificación. La probabilidad de que la familia se extinga es q , donde q representa la solución más chica de $\Phi(s) = s$.*

$$P\left(\bigcup_{n \in \mathbb{N}} \{Z_n = 0\}\right) = q.$$

Demostración. Primero veamos que, $\Phi_n(0) \nearrow q$, cuando $n \rightarrow \infty$. Observemos que $\Phi(0) > 0$, y Φ es creciente, por lo tanto, $0 < \Phi(0) < \Phi(q) = q$. Aplicando nuevamente Φ , $0 < \Phi(0) < \Phi_2(0) < q$. Repitiendo este procedimiento, obtenemos que $\Phi_n(0)$ es creciente y acotada. Sea $L = \lim_{n \rightarrow \infty} \Phi_n(0)$, $L \leq q$. Como Φ es continua,

$$\Phi_{n+1}(L) = \Phi\left(\lim_{n \rightarrow \infty} \Phi_n(0)\right) = \lim_{n \rightarrow \infty} \Phi_{n+1}(0) = L.$$

Luego, $L = q$. Ahora sí calculemos la probabilidad de extinción,

$$P(\exists n / Z_n = 0) = P\left(\bigcup_{n \in \mathbb{N}} \{Z_n = 0\}\right)$$

Como $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$ y $\Phi_n(0) = P(Z_n = 0)$ tenemos

$$P\left(\bigcup_{n \in \mathbb{N}} \{Z_n = 0\}\right) = \lim_{n \rightarrow \infty} P(Z_n = 0) = \lim_{n \rightarrow \infty} \Phi_n(0) = q.$$

□

Este teorema nos dice que si $m \leq 1$ el apellido de la familia se extingue en un tiempo finito con probabilidad 1. Cuando $m < 1$ se lo llama caso sub-crítico, y cuando $m = 1$ caso crítico. Por otro lado, si $m > 1$ (caso super-crítico) hay una probabilidad positiva de que los descendientes permanezcan en todas las generaciones futuras. Veremos que en este caso, no solo hay probabilidad positiva de no extinguirse, sino que además el tamaño de la población crece exponencialmente. Para ello debemos introducir el concepto de martingala.

Definición 3.1. *Sea $(X_n)_{n \in \mathbb{N}}$ un proceso estocástico. Se dice que X_n es una Martingala, respecto de la filtración $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si, para todo $n \geq 0$*

1. $E(|X_n|) < \infty$
2. $X_n \in \mathcal{F}_n$
3. $E(X_{n+1} | \mathcal{F}_n) = X_n$.

Para abreviar, decimos que X_n es \mathcal{F}_n -martingala.

Observación 3.1. *Si X_n es una martingala, $E(X_n) = E(X_0)$ para todo $n \in \mathbb{N}$.*

Teorema 3.3. *Sea $(X_n)_{n \in \mathbb{N}}$ una martingala no negativa, respecto a la filtración \mathcal{F}_n . Entonces, casi seguramente, $\lim_{n \rightarrow \infty} X_n$ existe y es finito.*

Demostración. La demostración la vamos a omitir. Puede verse en Durret [4, pág. 234-235]. \square

Lema 3.2. *Sea $\mathcal{F}_n = \sigma(\xi_i^k, i \geq 1, 1 \leq k \leq n)$, y $m = E(\xi_i^k)$, entonces $W_n = \frac{Z_n}{m^n}$ es una \mathcal{F}_n -martingala.*

Demostración. Claramente, $W_n \in \mathcal{F}_n$. Como m^n es una constante, $E(W_{n+1} | \mathcal{F}_n) = \frac{1}{m^n} E(Z_{n+1} | \mathcal{F}_n)$, y

$$E(Z_{n+1} | \mathcal{F}_n) = E\left(\sum_{i=1}^{Z_n} \xi_i^n | \mathcal{F}_n\right) = E\left(\sum_{i \in \mathbb{N}} \mathbf{1}_{\{i \leq Z_n\}} \xi_i^n | \mathcal{F}_n\right) = E\left(\lim_{N \rightarrow \infty} \sum_{i=1}^N \mathbf{1}_{\{i \leq Z_n\}} \xi_i^n | \mathcal{F}_n\right).$$

La suma es de variables positivas, por lo tanto

$$E(Z_{n+1} | \mathcal{F}_n) = \sum_{i \in \mathbb{N}} E(\mathbf{1}_{\{i \leq Z_n\}} \xi_i^n | \mathcal{F}_n).$$

Como $\mathbf{1}_{\{i \leq Z_n\}} \in \mathcal{F}_n$ y ξ_i^n es independiente de \mathcal{F}_n ,

$$E(Z_{n+1} | \mathcal{F}_n) = \sum_{i \in \mathbb{N}} \mathbf{1}_{\{i \leq Z_n\}} E(\xi_i^n) = m Z_n.$$

Luego,

$$E(W_{n+1} | \mathcal{F}_n) = \frac{1}{m^{n+1}} E(Z_{n+1} | \mathcal{F}_n) = \frac{1}{m^{n+1}} m Z_n = W_n.$$

Como $E(W_0) = 1$, $E(W_n) = 1$ para todo n . \square

Corolario 3.1. *Sea $W_n = \frac{Z_n}{m^n}$, existe W una variable aleatoria, tal que*

$$\lim_{n \rightarrow \infty} W_n = W \quad c.s.$$

De esta convergencia se ve que Z_n crece como $m^n W$, si $W \neq 0$. Si $P(W = 0) = 1$ nos dice simplemente que m^n crece mas rápido que Z_n . Pero en el caso en el cual estamos interesados, podemos ver que $P(W = 0) < 1$.

Teorema 3.4 (Convergencia en L^p). *Sea X_n una martingala con $\sup_{n \in \mathbb{N}} E(|X_n|^p) < \infty$ donde $p > 1$, entonces existe X tal que $X_n \xrightarrow[L^p]{n \rightarrow \infty} X$.*

Demostración. La demostración puede verse en Durret [4, pág. 252-253]. \square

Teorema 3.5. *Si $m > 1$, $V(Z_1) = \sigma^2 < \infty$ y $Z_0 = 1$. Entonces,*

1. $\lim_{n \rightarrow \infty} E(W_n - W)^2 = 0$.

2. $E(W) = 1$.

3. $P(W = 0) = P(Z_n = 0 \text{ para algún } n) = q$.

Demostración. Por el Teorema 3.1,

$$E(W_n^2) = \frac{E(Z_n^2)}{m^{2n}} = \frac{V(Z_n) + E(Z_n)^2}{m^{2n}} = \frac{\sigma^2 m^{n-1} (m^n - 1)}{(m-1)m^{2n}} + 1 = \frac{\sigma^2 (1 - m^{-n})}{(m-1)m} + 1.$$

Como $E(W_n^2)$ es una función creciente en n ,

$$\sup_{n \in \mathbb{N}} E(W_n^2) = \lim_{n \rightarrow \infty} E(W_n^2) = \frac{\sigma^2}{m^2 - m} + 1 < \infty.$$

Luego, por el Teorema 3.4 existe $W' \in L^2$ tal que W_n converge a W' en L^2 . Por lo tanto W_n converge casi seguramente a W' . Vimos que $W_n \xrightarrow[n \rightarrow \infty]{c.s.} W$, por unicidad del límite, $P(W = W') = 1$.

Como W_n converge a W en L^2 , $E(W_n) \rightarrow E(W)$ y $E(W_n) = 1$ para todo n así que $E(W) = 1$ y W no puede ser constantemente 0, es decir, $P(W = 0) < 1$. Sea $r = P(W = 0)$, tenemos

$$P(W = 0) = P\left(\lim_{n \rightarrow \infty} \frac{Z_n}{m^n} = 0\right) = \sum_{k=0}^{\infty} P\left(\lim_{n \rightarrow \infty} \frac{Z_n}{m^n} = 0 | Z_1 = k\right) P(Z_1 = k).$$

La distribución de la cantidad de hijos en la n ésima generación sabiendo que en la primer generación hubo k individuos, es igual a la distribución de la cantidad de hijos en la $n-1$ ésima generación de k familias independientes, luego

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} \frac{Z_n^{(k)}}{m^n} = 0\right) &= P\left(\lim_{n \rightarrow \infty} \frac{1}{m} \left[\frac{Z_{n-1}^1}{m^{n-1}} + \dots + \frac{Z_{n-1}^k}{m^{n-1}} \right] = 0\right) \\ &= P\left(\lim_{n \rightarrow \infty} W_{n-1}^1 + \dots + W_{n-1}^k = 0\right) \\ &= P\left(\bigcap_{i=1}^k \left\{ \lim_{n \rightarrow \infty} W_{n-1}^i = 0 \right\}\right) = \prod_{i=1}^k P\left(\lim_{n \rightarrow \infty} W_{n-1}^i = 0\right) \\ &= P(W = 0)^k = r^k. \end{aligned}$$

Por lo tanto,

$$r = P(W = 0) = \sum_{k=0}^{\infty} r^k P(Z_1 = k) = \Phi(r).$$

De donde, $P(W = 0) = q$, por el Teorema 3.2 $P(Z_n > 0 \quad \forall n \in \mathbb{N}) = 1 - q$. Por otro lado, $\{W > 0\} \subseteq \{Z_n > 0 \quad \forall n\}$. Por lo que,

$$P\left(\bigcap_{n \in \mathbb{N}} \{Z_n > 0\} \setminus \{W > 0\}\right) = P\left(\bigcap_{n \in \mathbb{N}} \{Z_n > 0\}\right) - P(W > 0) = 0.$$

□

Es decir, que para casi todo ω en el conjunto $\{Z_n > 0 \quad \forall n\}$ se tiene $W(\omega) > 0$ y por lo tanto $Z_n \sim m^n$.

3.2. Procesos de ramificación a tiempo continuo

Una generalización al proceso de Galton-Watson, es considerar el tiempo de vida de cada individuo como una variable aleatoria continua, al término de la cual se muere y se produce un número aleatorio de individuos descendientes. En lugar de considerar la cadena de Markov discreta, $\{Z_n, n \in \mathbb{N}\}$, tenemos que considerar el proceso, $\{Z(t), t \geq 0\}$, donde $Z(t)$ es el número de individuos a tiempo t . Además, suponemos que la distribución del tiempo de vida de cada individuo es la misma, exponencial de parámetro a y todos los individuos se reproducen según la distribución de la variable aleatoria N . Cada individuo es independiente de los otros individuos y de la historia del proceso.

Para empezar daremos la definición de proceso de Markov a tiempo continuo, y luego definiremos proceso de ramificación unidimensional a tiempo continuo.

Definición 3.2. *Daremos que el proceso estocástico $X = (X_t, t \geq 0)$ a valores en el espacio (finito o numerable) S es Markov si para todo $0 \leq s \leq t$ se tiene*

$$P(X_t = x | X_u, u \leq s) = P(X_t = x | X_s), \quad \text{para todo } x \in S.$$

Sea $P_{ij}(t)$ la probabilidad de tener j individuos a tiempo t , siendo que a tiempo 0 había i individuos.

Definición 3.3. *Un proceso estocástico $\{Z(t, \omega), t \geq 0\}$ en el espacio de probabilidad (Ω, \mathcal{F}, P) se llama proceso de ramificación markoviano unidimensional a tiempo continuo, si:*

1. *El espacio de estados es \mathbb{N}_0 .*
2. *Es una cadena de Markov homogénea con respecto a la σ -álgebra, $\mathcal{F}_t = \sigma\{Z(s, \omega); s \leq t\}$.*
3. *La probabilidad de transición $P_{ij}(t)$ satisface,*

$$\sum_{j=0}^{\infty} P_{ij}(t) s^j = \left[\sum_{j=0}^{\infty} P_{1j}(t) s^j \right]^i$$

para todo $i \geq 0$ y $|s| \leq 1$.

Esta última condición nos dice que considerar un proceso de ramificación con i individuos iniciales, es exactamente lo mismo que considerar la suma de i procesos de ramificación independientes con un individuo inicial.

3.2.1. Construcción

Vamos a construir el proceso de ramificación con un individuo inicial, por como definimos procesos de ramificación, si comenzamos con varios individuos, simplemente es agarrar varias copias de comenzar con uno solo. Cada individuo vive un tiempo exponencial de parámetro $a > 0$. Sea la variable aleatoria N con distribución p , notamos $p_i = P(N = i)$ con $i \in \mathbb{N}_0$. La cantidad

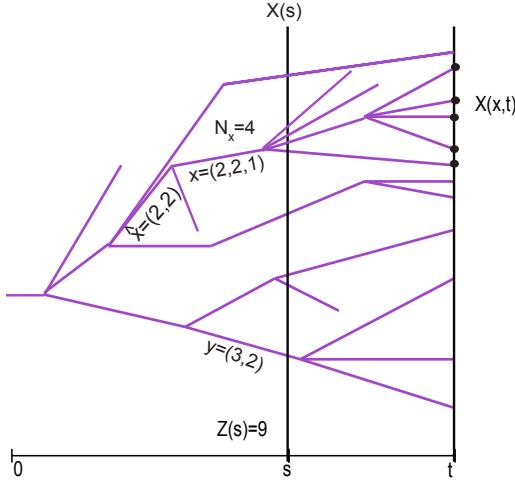


Figura 3.1: Una realización del proceso de ramificación. El conjunto $X(s)$ son las intersecciones de la linea vertical con las ramas a tiempo s . La cantidad de individuos vivos a ese tiempo es $Z(s)$. El conjunto $X(x,t)$ consiste en los descendientes de x que estan vivos a tiempo t (indicados con \circ). $X(x,t) = \{(3,1),(3,2),(3,3),(3,4),(4)\}$

de hijos que tiene cada individuo es una variable aleatoria con distribución p . Construimos el árbol genealógico de la misma forma que lo construye Georgii-Baake [5] para el caso multi-tipo, la construcción original puede verse en Harris ([7]). Para empezar consideramos el espacio de estados,

$$\mathbb{X} = \bigcup_{n \in \mathbb{N}_0} \mathbb{X}_n.$$

Donde \mathbb{X}_n describe la generación n -ésima. Esto es, $\mathbb{X}_0 = 1$, $\mathbb{X}_1 = \mathbb{N}_0$ un elemento $x \in \mathbb{X}_1$ tiene la forma $x = l_1$, donde l_1 indica que es el l_1 -ésimo descendiente de la raíz. Luego para $n \in \mathbb{N}$, definimos $\mathbb{X}_n = \mathbb{N}^n$ y $x = (l_1, \dots, l_n) \in \mathbb{X}_n$ es el l_n -ésimo hijo de $\hat{x} = (l_1, \dots, l_{n-1}) \in \mathbb{X}_{n-1}$. A cada $x \in \mathbb{X}$ le asociamos,

- un tiempo de vida aleatorio τ_x , con distribución exponencial de parámetro $a > 0$, y
- una variable aleatoria $N_x \in \mathbb{N}_0$ con distribución p , tal que la familia $\{\tau_x, N_x, x \in \mathbb{X}\}$ es independiente.

La variable N_x indican cuantos descendientes tiene $x \in \mathbb{X}$. Por lo tanto una realización del proceso es el conjunto $X = \bigcup_{n \in \mathbb{N}_0} X_n$, que queda definido recursivamente por,

$$X_0 = 1 \quad X_n = \{x = (\hat{x}, l_n) \in \mathbb{X}_n / \hat{x} \in \mathbb{X}_{n-1}, l_n \leq N_{\hat{x}}\}.$$

Para cada $x \in X$ sea T_x el momento en el que x muere y se reproduce, definido recursivamente por $T_x = T_{\hat{x}} + \tau_x$, con $T_{x_0} = \tau_1$ si $x_0 = 1 \in X_0$. Luego, el intervalo de vida del individuo x es

$[T_{\hat{x}}, T_x]$. Por lo tanto, $X(t) = \{x \in X / T_{\hat{x}} \leq t < T_x\}$ es la población a tiempo t . La familia completa queda determinada por el proceso $(X(t))_{t \geq 0}$. Cuando $0 < s < t$ y $x \in X(s)$, definimos $X(x, t)$ a los descendientes de x que están vivos a tiempo t . Consideramos $Z(t) = |X(t)|$, en palabras $Z(t)$ es la cantidad de individuos a tiempo t . Si no indicamos la cantidad de individuos iniciales, consideramos que se inicia con un individuo.

Al igual que en el caso discreto estamos interesados en averiguar si la familia se extingue cuando pasa el tiempo, es decir, analizar el comportamiento de $Z(t)$ para t suficientemente grande. Para ello debemos estudiar las probabilidades de transición; $P_{ij}(t)$ para $t > 0$. Pero tener varios individuos y que cada uno pueda tener hijos en cualquier momento, y estos volver a tener hijos, hace que calcular las probabilidades de transición a todo tiempo sea muy complicado. Es por ello que vamos a estudiar las probabilidades de transición en tiempos infinitesimales, estas son más sencillas de calcular ya que en un tiempo suficientemente pequeño solo un individuo podrá reproducirse. Luego, podremos ver que las probabilidades de transición para cualquier tiempo son solución de las ecuaciones de Kolmogorov.

Sea $P_{ij}(h, t+h)$ la probabilidad de transición $P(Z(t+h) = j | Z(h) = i)$. Asumimos que el proceso es homogéneo en el tiempo, es decir,

$$P_{ij}(h, t+h) = P_{ij}(t) \quad \forall t \geq 0.$$

Calculemos las probabilidades infinitesimales, $P_{ij}(h)$ para h suficientemente pequeño, $i, j \in \mathbb{N}$.

Si el proceso comienza con i individuos, el primer salto de la población ocurre cuando el primero de estos i individuos muere y se reproduce. Si llamamos Y_k a la duración aleatoria del tiempo de vida del k -ésimo individuo, sabemos que Y_k tiene distribución exponencial de parámetro a para todo k y son independientes, por lo tanto el primer salto de la población ocurre con la distribución de la variable aleatoria $\min\{Y_1, \dots, Y_i\}$, es decir con distribución exponencial de parámetro ia .

La probabilidad de que no muera ningún individuo en el intervalo de tiempo $(0, h)$, es e^{-aih} . Y de que muera uno sólo es $i(1 - e^{-ah})(e^{ah})^{i-1}$, ya que el tiempo de vida de cada individuo es independiente y por lo tanto consideramos que un individuo (entre los i posibles) muera en el intervalo $(0, h)$ y que los demás no mueran en ese intervalo. Esta probabilidad podemos expresarla como $iah + o(h)$.

Para h pequeño la probabilidad mueran 2 o más individuos en el intervalo $(0, h)$ es $1 - (e^{-aih}) - (i(1 - e^{-ah})(e^{ah})^{i-1})$ que es del orden de $o(h)$.

Por lo tanto,

$$P_{ij}(h) = o(h)$$

para $j < i - 1$.

La probabilidad de que exactamente uno muera antes de tiempo h es $iah + o(h)$ y luego el proceso salta al estado $j \geq i - 1$ con probabilidad p_{j-i+1} , luego

$$P_{ij}(h) = iahp_{j-i+1} + o(h)$$

si $j \geq i - 1, j \neq i$ y h pequeño.

Para calcular $P_{ii}(h)$ debemos considerar los siguientes eventos disjuntos, que no muera ningún individuo en $(0, h)$, que muera solo uno, y que mueran más de 2. Este último como dijimos antes es $o(h)$. La probabilidad de que ningún individuo muera es $e^{-aih} = 1 - iah + o(h)$. Cuando consideramos que muera un individuo, para seguir teniendo i individuos este debe tener un solo hijo. Por lo tanto,

$$P_{ii}(h) = 1 - iah + iahp_1 + o(h).$$

Teorema 3.6 (Ecuaciones de Kolmogorov). *Las probabilidades de transición $P_{ij}(t)$, con $t > 0$ son solución de las siguientes ecuaciones:*

$$(forward) \quad \frac{d}{dt}P_{ij}(t) = -jaP_{ij}(t) + a \sum_{k=1}^{j+1} kP_{ik}(t)p_{j-k+1}, \quad (3.3)$$

$$(backward) \quad \frac{d}{dt}P_{ij}(t) = -iaP_{ij}(t) + ia \sum_{k=i-1}^{\infty} P_{kj}(t)p_{k-i+1}, \quad (3.4)$$

con condiciones de contorno,

$$P_{ij}(0^+) = \delta_{ij}. \quad (3.5)$$

Demostración. Para demostrar ambas ecuaciones usaremos las probabilidades de transición infinitesimales ya calculadas.

$$\begin{aligned} P_{ij}(t+h) &= P_i(Z(t+h) = j) = \sum_{k=1}^{\infty} P(Z(t+h) = j | Z(t) = k) P_i(Z(t) = k) \\ &= \sum_{k=1}^{\infty} P_{kj}(h) P_{ik}(t) = \sum_{k=1}^{j+1} [kahp_{j-k+1} P_{ik}(t)] + (1 - jah) P_{ij}(t) + o(h). \end{aligned}$$

De donde,

$$\lim_{h \rightarrow 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = a \sum_{k=1}^{j+1} kp_{j-k+1} P_{ik}(t) - jaP_{ij}(t).$$

Por otra parte

$$\begin{aligned} P_{ij}(t+h) &= P_i(Z(t+h) = j) = \sum_{k=1}^{\infty} P(Z(t+h) = j | Z(h) = k) P_i(Z(h) = k) \\ &= \sum_{k=1}^{\infty} P_{kj}(t) P_{ik}(h) = \sum_{k=i-1}^{\infty} [iahp_{k-i+1} P_{kj}(t)] - iahP_{ij}(t) + o(h). \end{aligned}$$

Por lo tanto,

$$\lim_{h \rightarrow 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = ai \sum_{k=i-1}^{\infty} p_{k-i+1} P_{kj}(t) - iaP_{ij}(t).$$

□

Desafortunadamente, las ecuaciones de Kolmogorov siempre tienen solución si $\sum_j P_{ij}(t) \leq 1$, pero esta solución puede que no sea única. Este problema surge cuando es posible que el proceso tenga infinitos individuos en un tiempo finito, es decir, que explote. Esta situación existe cuando las ecuaciones (3.3), (3.4) tienen soluciones tal que $\sum_j P_{ij}(t) < 1$. Puede verse en Harris [7, cap. V, sec. 3], que si el proceso no explota, las ecuaciones de Kolmogorov tienen como única solución $P_{ij}(t)$ que satisfacen $\sum_j P_{ij}(t) = 1$.

Para ver que el proceso no explota calculemos $E(Z(t))$, para $t \in (0, \infty)$. Para ello consideremos la variable aleatoria Y con distribución exponencial de parámetro a . Supongamos que comenzamos el proceso con un solo individuo. La cantidad de individuos a tiempo t podemos considerarla de la siguiente forma; que el individuo inicial murió antes de tiempo t , es decir, que $Y < t$, y este dejó N hijos, o que el individuo permanece con vida a tiempo t , es decir, $Y > t$

$$Z(t) = \mathbf{1}_{\{Y < t\}} \sum_{i=0}^N Z^i(t - Y) + \mathbf{1}_{\{Y > t\}},$$

donde $Z^i(t - Y)$ denota un proceso de ramificación a tiempo continuo con 1 individuo inicial, observado a tiempo $t - Y$, la potencia i nota que es el proceso de ramificación del i -ésimo hijo. Como cada individuo tiene hijos de forma independiente, los procesos Z^i son independientes.

Luego la esperanza,

$$E(Z(t)) = E\left(\mathbf{1}_{\{Y < t\}} \sum_{i=0}^N Z^i(t - Y)\right) + E(\mathbf{1}_{\{Y > t\}})$$

de donde, condicionando a Y en el primer término, obtenemos;

$$\begin{aligned} E(Z(t)) &= \int_0^t E\left(ae^{-as} \sum_{i=0}^N Z^i(t-s)\right) ds + E(\mathbf{1}_{\{Y > t\}}) \\ &= \int_0^t ae^{-as} E\left(\sum_{i=0}^N Z^i(t-s)\right) ds + P(Y > t) \\ &= \int_0^t ae^{-as} E(N) E(Z(t-s)) ds + P(Y > t) \\ &= E(N)a \int_0^t e^{-as} E(Z(t-s)) ds + e^{-at}, \end{aligned}$$

la tercera igualdad es cierta pues la variable N es independiente de las variables $(Z^k)_{k=1..i}$ y las Z^k tienen todas igual distribución. Es decir que $E(Z(t))$ es solución de la ecuación diferencial $g(t) = E(N)a \int_0^t e^{-as} g(t-s) ds + e^{-at}$. Puede verse en Perko[11] que ésta ecuación tiene solución única. Veamos que la solución es $e^{\lambda t}$, con λ alguna constante adecuada.

$$e^{\lambda t} = E(N)a \int_0^t e^{-as} e^{\lambda(t-s)} ds + e^{-at}$$

$$\begin{aligned} e^{\lambda t} &= E(N)ae^{\lambda t} \int_0^t e^{-(a+\lambda)s} ds + e^{-at} \\ e^{\lambda t} &= E(N)ae^{\lambda t} \frac{1 - e^{-(a+\lambda)t}}{a+\lambda} + e^{-at} \\ 1 &= \frac{E(N)a}{a+\lambda} (1 - e^{(a+\lambda)t}) + e^{-(a+\lambda)t}. \end{aligned}$$

Por lo tanto, $e^{\lambda t}$ es solución de la ecuación sí y solo sí, $\lambda = a(E(N) - 1)$. Luego,

$$E(Z(t)) = e^{a(E(N)-1)t} \quad t \in (0, \infty). \quad (3.6)$$

Por lo tanto $Z(t)$ es finito con probabilidad 1 para todo $t \geq 0$.

De (3.6), podemos ver que:

1. Si $E(N) < 1$, $E(Z(t))$ decrece exponencialmente (caso sub-crítico).
2. Si $E(N) = 1$, $E(Z(t)) = 1$ (caso crítico).
3. Si $E(N) > 1$, $E(Z(t))$ crece exponencialmente (caso super-crítico).

3.2.2. Funciones generadoras

Definimos la función generadora de $Z(t)$,

$$F(s, t) = \sum_{k=0}^{\infty} P(Z(t) = k | Z(0) = 1) s^k = \sum_{k=0}^{\infty} P_{1k}(t) s^k = E_1(s^{Z(t)})$$

y las funciones $f(s) = \sum_{j=0}^{\infty} p_j s^j$, $|s| \leq 1$ y $u(s) = a(f(s) - s)$, donde a es la media del tiempo de vida de un individuo y p_k la probabilidad de que un individuo tenga k hijos.

Teorema 3.7 (Ecuaciones de Kolmogorov). *La función $F(s, t)$ definida anteriormente, cumple las siguientes ecuaciones:*

$$(backward) \quad \frac{\partial}{\partial t} F(s, t) = u(s) \frac{\partial}{\partial s} F(s, t), \quad (3.7)$$

$$(forward) \quad \frac{\partial}{\partial t} F(s, t) = u(F(s, t)), \quad (3.8)$$

con condición de contorno:

$$F(s, 0) = s. \quad (3.9)$$

Demostración. Primero demostremos (3.7), para ello usaremos (3.3). Para $i = 1$

$$\frac{\partial}{\partial t} P_{1k}(t) = -kaP_{1k}(t) + a \sum_{i=1}^{k+1} iP_{1i}(t)p_{k-i+1}.$$

Por lo tanto,

$$\begin{aligned}
\frac{\partial}{\partial t} F(s, t) &= \sum_{k=0}^{\infty} \frac{\partial}{\partial t} P_{1k}(t) s^k = \sum_{k=0}^{\infty} \left[-kaP_{1k}(t) + a \sum_{i=1}^{k+1} iP_{1i}(t)p_{k-i+1} \right] s^k \\
&= -a \sum_{k=0}^{\infty} kP_{1k}(t)s^k + a \sum_{k=0}^{\infty} \sum_{i=1}^{k+1} iP_{1i}(t)p_{k-i+1}s^k \\
&= -a \sum_{k=0}^{\infty} kP_{1k}(t)s^k + a \left(\sum_{i=0}^{\infty} p_i s^i \right) \left(\sum_{k=0}^{\infty} P_{1k}(t)ks^{k-1} \right) \\
&= a \sum_{k=0}^{\infty} P_{1k}(t)ks^{k-1} \left(\sum_{i=0}^{\infty} p_i s^i - s \right) \\
&= u(s) \frac{\partial}{\partial s} F(s, t).
\end{aligned}$$

Para probar (3.8), usaremos (3.4) para $i = 1$, y la condición (3) de la Definición 3.3.

$$\frac{d}{dt} P_{1k}(t) = -aP_{1k}(t) + a \sum_{l=0}^{\infty} P_{lk}(t)p_l.$$

Por lo tanto,

$$\begin{aligned}
\frac{\partial}{\partial t} F(s, t) &= \sum_{k=0}^{\infty} \left(-aP_{1k}(t) + a \sum_{l=0}^{\infty} P_{lk}(t)p_l \right) s^k \\
&= -a \sum_{k=0}^{\infty} P_{1k}(t)s^k + a \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} P_{lk}(t)p_ls^k \\
&= -a \sum_{k=0}^{\infty} P_{1k}(t)s^k + a \sum_{l=0}^{\infty} p_l \sum_{k=0}^{\infty} P_{lk}(t)s^k \\
&= -a \sum_{k=0}^{\infty} P_{1k}(t)s^k + a \sum_{l=0}^{\infty} p_l \left(\sum_{k=0}^{\infty} P_{1k}(t)s^k \right)^l \\
&= a(f(F(s, t)) - F(s, t)) \\
&= u(F(s, t)).
\end{aligned}$$

□

Observemos que la función $F(s, t)$ cumple una fórmula de iteración similar a $f_n(s)$ en el proceso de Galton-Watson.

$$F(s, t+u) = \sum_{k=0}^{\infty} P_{1k}(t+u)s^k,$$

condicionando a las posibles cantidades de individuos a tiempo u , y usando la homogeneidad de la cadena obtenemos,

$$F(s, t+u) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P_{1j}(u)P_{jk}(t)s^k,$$

intercambiando el orden de las sumas y usando la definición de proceso de ramificación a tiempo continuo,

$$F(s, t+u) = \sum_{j=0}^{\infty} P_{1j}(u) F^j(s, t).$$

Por lo tanto,

$$F(s, t+u) = F(F(s, t), u) \quad |s| \leq 1, \quad u \geq 0, \quad t \geq 0. \quad (3.10)$$

3.2.3. El proceso discreto inmerso

El proceso de ramificación a tiempo discreto nos va a ayudar a analizar el comportamiento del proceso de ramificación a tiempo continuo. Para ello veremos en el siguiente Teorema el proceso discreto inmerso en un proceso continuo.

Teorema 3.8. *Sea $\{Z(t, \omega), t \geq 0\}$ un proceso de ramificación markoviano unidimensional a tiempo continuo. Para todo $\delta > 0$ la sucesión $Z_n(\omega) = Z(n\delta, \omega)$ es un proceso de Galton-Watson con función generadora de momentos $\varphi(s) = F(s, \delta)$.*

Demostración. Dado $\delta > 0$, el proceso $\{Z_n\}_{n \in \mathbb{N}}$ es una cadena de Markov con probabilidad de transición $P(Z_{n+1} = i | Z_n = j) = P(Z(n\delta) = i | Z(n) = j) = P_{ji}(\delta)$. Para ver que es un proceso de Galton-Watson nos basta ver que la función generadora de Z_n es componer n veces la función generadora de Z_1 . La función generadora de momentos de Z_n es $\Phi_{(n)}(s) = E(s^{Z_n} | Z_0 = 1)$. Por (3.10)

$$\Phi_{(n)}(s) = \Phi(\Phi_{(n-1)}(s)),$$

por lo tanto $\Phi_{(n)}(s) = \Phi_n(s)$ que es la n -ésima composición de la función $\Phi(s)$. \square

Para decir cuando se extingue el proceso a tiempo continuo usaremos el proceso discreto inmerso con $\delta = 1$, ya que si existe t_0 tal que $Z(t_0) = 0$, para todo $t \geq t_0$ $Z(t) = 0$. Y vimos en la sección anterior que el proceso de Galton-Watson se extingue en un tiempo finito con probabilidad 1 en el caso sub-crítico ($m < 1$) y en el caso crítico ($m = 1$). Como $E(Z(t)) = e^{a(E(N)-1)}$, obtenemos que el proceso se extingue en un tiempo finito si $E(N) \leq 1$. En el caso super-crítico, cuando $E(N) > 1$, la expresión de la esperanza nos sugiere que la población crece exponencialmente, para demostrarlo necesitamos usar convergencia de martingalas. Para ello, debemos ver la versión a tiempo continuo de las definiciones de filtración y martingala.

Definición 3.4. 1. *Sea (Ω, \mathcal{F}) un espacio medible. Decimos que una familia $(\mathcal{F}_t)_{t \geq 0}$ de σ -álgebras contenidas en \mathcal{F} es una filtración si $\mathcal{F}_s \subseteq \mathcal{F}_t$ para todo $0 \leq s \leq t$.*

2. *Sea (Ω, \mathcal{F}) un espacio medible, equipado con una filtración $(\mathcal{F}_t)_{t \geq 0}$. Un proceso estocástico X definido en (Ω, \mathcal{F}) se dice adaptado a la filtración $(\mathcal{F}_t)_{t \geq 0}$ si para cada $t \geq 0$, X_t es \mathcal{F}_t -medible.*

3. *Sea $X = \{X_t : t \geq 0\}$ un proceso estocástico definido en un espacio de probabilidad (Ω, \mathcal{F}, P) y sea $(\mathcal{F}_t)_{t \geq 0}$ una filtración en dicho espacio. Decimos que X es una martingala a tiempo continuo con respecto a la filtración $(\mathcal{F}_t)_{t \geq 0}$ si satisface las siguientes condiciones:*

- X es adaptado a la filtración $(\mathcal{F}_t)_{t \geq 0}$;
- $E(|X_t|) < +\infty$ para todo $t \geq 0$;
- Para todo $0 \leq s \leq t$, se tiene $E(X_t | \mathcal{F}_s) = X_s$.

Al igual que en el caso discreto, para analizar el límite de la población necesitamos resultados de convergencia de martingalas.

Teorema 3.9 (Teorema de Convergencia de Martingalas). *Sea $\{X(t)\}_{t \geq 0}$ una \mathcal{F}_t -martingala continua a derecha. Si $\sup_{t \geq 0} \{E(X(t))\} < \infty$, entonces existe $X = \lim_{t \rightarrow \infty} X(t)$ c.s.*

Demostración. Puede verse en Kallenberg, [8, pag. 107]. □

Teorema 3.10. *La familia $\{Z(t)e^{-\lambda t}, \mathcal{F}_t = \sigma(Z(s), 0 \leq s \leq t), t \geq 0\}$ es una martingala.*

Demostración. Claramente $Z(t)e^{-\lambda t} \in \mathcal{F}_t$. Para calcular la esperanza de $Z(t)$, nos va a ser muy útil expresar la cantidad de individuos a tiempo t como la suma de $Z(s)$ procesos de ramificación independientes mirados a tiempo $t - s$. Es decir, dados s y t tales que $0 < s < t$,

$$Z(t) = \sum_{j=1}^{Z(s)} Z^j(t-s)$$

si $Z(s) > 0$, y $Z(t) = 0$ si $Z(s) = 0$. Por lo tanto,

$$E(Z(t) | \mathcal{F}_s) = E(Z(t) | Z_s) = E\left(\sum_{j=1}^{Z(s)} Z^j(t-s) | Z(s)\right) = Z(s)E(Z(t-s)) = Z(s)e^{(t-s)\lambda}.$$

Luego, $E(Z(t)e^{-\lambda t} | \mathcal{F}_t) = Z(s)e^{-\lambda s}$. Como $E(Z(0)) = 1$, $E(Z(t)e^{-\lambda t}) = 1$, y por lo tanto $Z(t)e^{-\lambda t}$ es una \mathcal{F}_t -martingala. □

Corolario 3.2. *Sea $X(t) = Z(t)e^{-\lambda t}$, existe X una variable aleatoria, tal que*

$$\lim_{t \rightarrow \infty} X(t) = X \quad \text{c.s.}$$

Otra vez la cuestión es ver si $X = 0$, para ello usaremos el caso discreto, ya que por el Corolario 3.1 para todo $\delta > 0$ tenemos,

$$W_\delta = \lim_{n \rightarrow \infty} Z(n\delta)e^{-n\lambda_\delta},$$

y para $\lambda_\delta > 0$ y $V(Z(1)) < \infty$ vimos que para casi todo ω en el evento $\{Z(n\delta, \omega) > 0, \forall n \in \mathbb{N}\}$ se tiene que $\{W_\delta(\omega) > 0\}$. Como $P(W_\delta = W) = 1$, para casi todo ω en $\{Z(t, \omega) > 0\}$ se tiene que $\{W(\omega) > 0\}$. Concluimos que $Z(t) \sim e^{t\lambda}$ con probabilidad 1.

3.3. Tiempo continuo, multi-tipo

Consideremos el proceso de ramificación a tiempo continuo, con distintos tipos de individuos (por ejemplo como muestra la figura, verde, rojo y azul). Al igual que en el caso anterior consideramos que el tiempo de vida de cada individuo tiene distribución exponencial, pero en este caso los distintos tipos de individuos podrían tener distintas medias. También vamos a considerar $Z(t)$ como la cantidad de individuos a tiempo t , pero en este caso si consideramos que hay k individuos distintos, $Z(t) = (Z_1(t), \dots, Z_k(t))$ donde $Z_i(t)$ es la cantidad de individuos de tipo i a tiempo t . Notamos $P^{\mathbf{i}}(\mathbf{j}, t)$ con $\mathbf{i} = (i_1, \dots, i_k), \mathbf{j} = (j_1, \dots, j_k) \in \mathbb{N}_0^k$, a la probabilidad de tener j_1 individuos de tipo 1, ..., j_k individuos de tipo k , a tiempo t , dado que a tiempo 0 el proceso tenía i_1 individuos de tipo 1, ..., i_k individuos de tipo k . Comencemos dando la definición de proceso de ramificación multtipo.

Definición 3.5. Un proceso estocástico $\{Z(t, \omega), t \geq 0\}$ en el espacio de probabilidad (Ω, \mathcal{F}, P) se llama proceso de ramificación markoviano k -dimensional a tiempo continuo, si:

1. El espacio de estados es \mathbb{N}_0^k .
2. Es una cadena de Markov homogénea con respecto a la σ -álgebra, $\mathcal{F}_t = \sigma\{Z(s, \omega); s \leq t\}$.
3. La probabilidad de transición $P^{\mathbf{i}}(\mathbf{j}, t)$ satisface,

$$\sum_{\mathbf{j} \in \mathbb{N}_0^k} P^{\mathbf{i}}(\mathbf{j}, t) \mathbf{s}^{\mathbf{j}} = \prod_{l=1}^k \left[\sum_{\mathbf{j} \in \mathbb{N}_0^k} P^{e_l}(\mathbf{j}, t) \mathbf{s}^{\mathbf{j}} \right]^{i_l}$$

para todo $\mathbf{i} \in \mathbb{N}_0^k$ y $|s_j| \leq 1$, $j \in \{1, \dots, k\}$, donde $\mathbf{s} = (s_1, \dots, s_k) \in \mathbb{R}^k$.

Esta última condición nos dice que considerar un proceso de ramificación con $\mathbf{i} = (i_1, \dots, i_k)$ individuos iniciales, es exactamente lo mismo que considerar la suma de i_j procesos de ramificación independientes con un individuo inicial de tipo j , y la suma sobre los distintos tipos de individuos.

3.3.1. Construcción

Sea $\mathcal{S} = \{1, \dots, k\}$ el espacio de los distintos tipos de individuos. Cada individuo $i \in \mathcal{S}$ vive un tiempo exponencial de parámetro $a_i > 0$. Sea N la matriz con entradas aleatorias N_{ij} con $i, j \in \{1, \dots, k\}$ donde N_{ij} es el número de hijos de tipo j que tiene el individuo de tipo i . $\mathbf{N}_i = (N_{ij})_{j \in \mathcal{S}}$ tiene distribución $\mathbf{p}_i = (p_{i1}, \dots, p_{ik})$, y media finita $m_{ij} = E(N_{ij})$, $\forall i, j \in \mathcal{S}$.

Asumimos que la matriz $M = (m_{ij})_{i, j \in \mathcal{S}}$ es irreducible. Construimos el árbol genealógico. Para empezar consideramos el espacio de estados

$$\mathbb{X} = \bigcup_{n \in \mathbb{N}} \mathbb{X}_n.$$

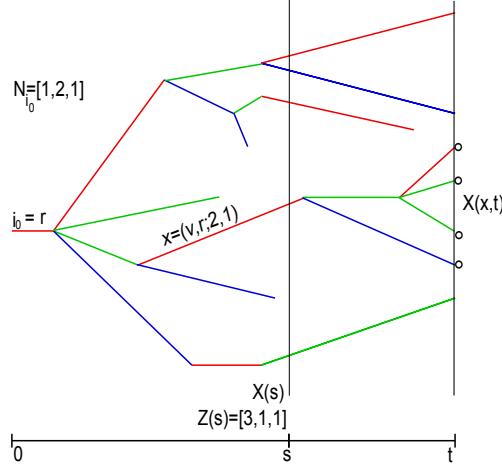


Figura 3.2: Una realización del proceso de ramificación. Donde $S = \{r, v, a\}$. El conjunto $X(s)$ son las intersecciones de la linea vertical a tiempo s . El tipo de individuos vivos a ese tiempo es $Z(t)$. El conjunto $X(x, t)$ consiste en los descendientes de x que estan vivos a tiempo t (indicados con \circ). $X(x, t) = \{(v, r; 1, 1), (v, v; 1, 1), (v, v; 1, 2), (a; 1)\}$

Donde \mathbb{X}_n es el espacio donde vive la generación n -ésima. Esto es $\mathbb{X}_0 = S$ y $i_0 \in \mathbb{X}$ es la raíz del árbol. El siguiente es $\mathbb{X}_1 = S \times \mathbb{N}$ y el elemento $x = (i_1, l_1) \in \mathbb{X}_1$ es el l_1 -ésimo hijo de tipo i_1 en la raíz. Finalmente, para $n > 1$, $\mathbb{X}_n = S^n \times \mathbb{N}^n$ y $x = (i_1, \dots, i_n, l_1, \dots, l_n) \in \mathbb{X}_n$ es l_n -ésimo hijo de tipo i_n del padre $\hat{x} = (i_1, \dots, i_{n-1}, l_1, \dots, l_{n-1})$. Escribimos $\sigma(x) = i_n$ para el tipo de $x \in \mathbb{X}_n$. Observemos que $\sigma(x) \in S$.

A cada $x \in \mathbb{X}$ le asociamos

- un tiempo de vida aleatorio τ_x , con distribución exponencial de parámetro $a_{\sigma(x)}$,
- descendencia aleatoria, $\mathbf{N}_x = (N_{xj})_{j \in S} \in \mathbb{N}_0^S$ con distribución $\mathbf{p}_{\sigma(x)}$ tal que las variables $\{\tau_x, N_x : x \in \mathbb{X}\}$ son independientes.

Las variables aleatorias N_x indican como son actualizados los individuos $x \in \mathbb{X}$. Por lo tanto el conjunto $X = \bigcup_{n \in \mathbb{N}} X_n$ esta definido recursivamente por

$$X_0 = \{i_0\} \quad X_n = \{x = (\hat{x}, i_n, l_n) \in \mathbb{X}_n : \hat{x} \in \mathbb{X}_{n-1}, l_n \leq N_{\hat{x}, i_n}\}.$$

La variable aleatoria τ_x es el tiempo de vida de $x \in X$. Sea T_x el instante en muere y se reproduce x . Lo definimos recursivamente por $T_x = T_{\hat{x}} + \tau_x$, con $T_{\hat{i}_0} = \tau_{\hat{i}_0}$ si $\hat{i}_0 \in \mathbb{X}_0$, es decir el tiempo en el que esta vivo $x \in \mathbb{X}$ es: $[T_{\hat{x}}, T_x]$. Por lo tanto, $X(t) = \{x \in X : T_x \leq t < T_x\}$ es la población a tiempo t .

El árbol completo esta determinado por el proceso $X(0, \infty) = (X(t))_{t \geq 0}$ definido en un espacio al que llamamos Ω . Cuando comenzamos con una distribución inicial v , la distribución de $X(t)$ en Ω es P^v y la esperanza es E^v , cuando $i_0 = i$ notamos, P^i y E^i .

Para cada $x \in X(s)$ consideramos, $X(x, t) = \{y \in X : xy \in X(t)\}$, en palabras el conjunto $X(x, t)$ son los descendientes de x que estan vivos a tiempo t .

Por la propiedad de falta de memoria de la distribución exponencial, los descendientes de $X(x, [s, \infty)) = (X(x, t))_{t \geq s}$ con $x \in X(s)$, son independientes de $X[0, s]$.

Consideremos las siguientes medidas,

$$Z(t) = \sum_{x \in X(t)} \delta_{\sigma(x)} \quad Z(x, t) = \sum_{y \in X(x, t)} \delta_{\sigma(y)}$$

Es decir, $Z(t) = (Z_1(t), \dots, Z_k(t))$ donde $Z_i(t)$ es la cantidad de individuos de tipo i que viven a tiempo t . En particular, $Z_i(t)$ es el cardinal de $X_i(t) = \{x \in X(t) : \sigma(x) = i\}$. Y la subpoblación descendiente de x ; $Z(x, t) = (Z_1(x, t), \dots, Z_k(x, t))$ con $Z_i(x, t)$ la cantidad de individuos de tipo i descendientes de x que estan vivos a tiempo t . El total de la población a tiempo t lo denotamos, $\|Z(t)\| = \sum_{j \in S} Z_j(t) = |X(t)|$.

3.3.2. Funciones generadoras

Consideremos las siguientes funciones,

$$f(\mathbf{s}) = (f^1(\mathbf{s}), \dots, f^k(\mathbf{s}))$$

donde $f^i(\mathbf{s}) = \sum_{\mathbf{j} \in \mathbb{N}_0^k} p^i(\mathbf{j}) \mathbf{s}^{\mathbf{j}}$ y $p^i(\mathbf{j})$ la probabilidad de que un individuo de tipo i tenga \mathbf{j} individuos, es decir, j_1 individuos de tipo 1, .., j_k individuos de tipo k .

$$u^i(\mathbf{s}) = a_i[f^i(\mathbf{s}) - s_i]$$

donde a_i es el tiempo medio de vida del individuo de tipo i . Por último, definimos la función generadora de $Z(t)$ cuando comenzamos con \mathbf{i} individuos, como

$$F(\mathbf{i}, \mathbf{s}, t) = \sum_{\mathbf{j} \in \mathbb{N}_0^k} P^{\mathbf{i}}(\mathbf{j}, t) \mathbf{s}^{\mathbf{j}}.$$

Denotamos

$$\begin{aligned} F(\mathbf{s}, t) &= (F(e_1, \mathbf{s}, t), \dots, F(e_k, \mathbf{s}, t)) \\ u(\mathbf{s}) &= (u^1(\mathbf{s}), \dots, u^k(\mathbf{s})) \end{aligned}$$

Teorema 3.11. *Las funciones $F(\mathbf{s}, t)$, y $u(\mathbf{s})$ definidas anteriormente cumplen las Ecuaciones de Kolmogorov:*

$$(forward) \quad \frac{\partial}{\partial t} F(e_i, \mathbf{s}, t) = \sum_{j=1}^k u^j(\mathbf{s}) \frac{\partial}{\partial s_j} F(e_i, \mathbf{s}, t), \quad (3.11)$$

$$(backward) \quad \frac{\partial}{\partial t} F(e_i, \mathbf{s}, t) = u^i(F(\mathbf{s}, t)), \quad (3.12)$$

con $i = 1, \dots, k$. Con condiciones de contorno,

$$P^i(j, 0^+) = \delta_{ij} \quad (3.13)$$

Demostración. Puede verse usando el Teorema 3.7 □

Al igual que en los otros procesos de ramificación queremos calcular $E(Z(t))$. Para ello consideramos las variables aleatorias Y_i con $i \in \{1, \dots, k\}$ independientes con distribución exponencial y media a_i . Cada Y_i es el tiempo de vida de un individuo de tipo i . Notamos $Z_j^i(t)$ a la cantidad de individuos de tipo j que hay a tiempo t considerando que el proceso comenzó con un solo individuo de tipo i .

Si comenzamos con un individuo de tipo i , la cantidad de individuos de tipo j a tiempo t podemos expresarla de la siguiente manera,

$$Z_j^i(t) = \mathbf{1}_{\{Y_i < t\}} \sum_{l=1}^k \sum_{n=1}^{N_{il}} Z_j^{l,n}(t - Y) + \mathbf{1}_{\{Y_i > t\}} \delta_{ij}$$

ya que podemos separar en 2 casos; que el individuo inicial sigue vivo a tiempo t , o que el individuo inicial muere antes de tiempo t , y este tiene N_i hijos, donde $N_i = (N_{i1}, \dots, N_{ik})$, por lo tanto en este caso la cantidad de hijos de tipo j a tiempo t lo podemos contar como la cantidad de hijos de tipo j a tiempo $t - Y$ de los procesos de ramificación $Z_j^{l,n}$, donde l es el tipo de individuo con el que comienza el proceso y n indica que es el n -ésimo hijo de tipo l del individuo inicial i . Tomando esperanza y condicionando a lo ocurrido con Y_i obtenemos,

$$E(Z_j^i(t)) = \int_0^t a_i e^{-a_i s} \sum_{l=1}^k E \left(\sum_{n=1}^{N_{il}} Z_j^{l,n}(t-s) \right) ds + P(Y_i > t) \delta_{ij}.$$

Como la cantidad de hijos del individuo inicial es independiente del comportamiento en el futuro de estos individuos,

$$E(Z_j^i(t)) = a_i \sum_{l=1}^k \int_0^t e^{-a_i s} E(N_{il}) E(Z_j^l(t-s)) ds + e^{-a_i t} \delta_{ij}.$$

Si llamamos $C_{ij}(t) = E(Z_j^i(t))$, la matriz $C(t)$ en lugar i, j cumple,

$$C_{ij}(t) = a_i \sum_{l=1}^k E(N_{il}) \int_0^t e^{-a_i s} C_{lj}(t-s) ds + e^{-a_i t} \delta_{ij} = a_i \sum_{l=1}^k E(N_{il}) e^{-a_i t} \int_0^t e^{a_i u} C_{lj}(u) du + e^{-a_i t} \delta_{ij}.$$

Puede verse en Perko [11] que existe una única solución de esta ecuación diferencial y es derivable. Por lo tanto,

$$\begin{aligned}
\frac{d}{dt}(C_{ij}(t)) &= a_i \sum_{l=1}^k E(N_{il}) e^{-a_it} \left[e^{a_it} C_{lj}(t) - a_i \int_0^t e^{a_i u} C_{ij}(u) du \right] - a_i e^{-a_it} \delta_{ij} \\
&= a_i \sum_{l=1}^k E(N_{il}) C_{lj}(t) - a_i a_i \sum_{l=1}^k E(N_{il}) \int_0^t e^{a_i u} C_{ij}(u) du - a_i e^{-a_it} \delta_{ij} \\
&= a_i \sum_{l=1}^k E(N_{il}) C_{lj}(t) - a_i C_{lj}(t) = a_i \sum_{l=1}^k E(N_{il}) C_{lj}(t) - a_i C_{lj}(t) \sum_{l=1}^k \delta_{il} \\
&= a_i \sum_{l=1}^k (E(N_{il}) - \delta_{il}) C_{lj}(t).
\end{aligned}$$

Si denotamos $A = a(E(N) - Id)$, la matriz $C_{ij}(t)$ cumple la siguiente ecuación diferencial;

$$\frac{d}{dt} C_{ij}(t) = \sum_{l=1}^k A_{il} C_{lj}(t).$$

Por lo tanto, para cada $t > 0$

$$\begin{aligned}
\frac{d}{dt} C(t) &= AC(t), \\
C(0) &= Id.
\end{aligned}$$

De la teoría de ecuaciones diferenciales se desprende que $C(t) = e^{tA}$. Para más detalle puede verse Norris [10, pág. 61,62]. De donde,

$$E(Z(t)) = e^{tA}.$$

La matriz $E(Z(t))$ tiene entradas positivas para todo $t > 0$, por la teoría de Perron-Frobenius, para cada t existe un autovalor $\rho(t)$ de $C(t)$ estrictamente positivo tal que cualquier otro autovalor $\hat{\rho}(t)$ de $C(t)$ satisface $|\hat{\rho}(t)| < \rho(t)$ y además $\rho(t)$ tiene multiplicidad 1. Como $C(t) = e^{tA}$ los autovalores de $C(t)$ están dados por $e^{\lambda_i t}$, $i = 1, \dots, k$ donde $\lambda_1, \lambda_2, \dots, \lambda_k$ son los autovalores de A . Los autovectores asociados de $C(t)$ coinciden con los de A . Podemos ordenar los autovalores de A de la siguiente manera

$$\lambda = \lambda_1 > Re(|\lambda_2|) \geq \dots \geq Re(|\lambda_k|).$$

Además λ tiene autovectores asociados a izquierda y derecha, π y h positivos, normalizados, tal que $|\pi| = 1$ y $\pi h^t = 1$. Por lo tanto $C(t)$ tiene a $e^{\lambda t}$ como autovalor de módulo máximo, de multiplicidad 1, con autovectores asociados π y h . Usando el autovector a izquierda podemos construirnos una martingala, y ésta al igual que en los casos anteriores nos ayudará a analizar el comportamiento de $Z(t)$.

Teorema 3.12. *La familia $\{\pi Z(t) e^{-\lambda t}, \mathcal{F}_t = \sigma(Z(s), 0 \leq s \leq t), t \geq 0\}$ es una martingala.*

Demostración. Claramente $\pi Z(t) e^{-\lambda t} \in \mathcal{F}_t$. Veamos que para todo $i \in \{1, \dots, k\}$, $\pi Z^i(t)$ es martingala. Para calcular la esperanza de $\pi Z^i(t)$, nos va a ser muy útil expresar la cantidad de individuos a tiempo t como la suma de $Z^l(s)$ procesos de ramificación independientes que comienzan con un individuo de tipo l mirados a tiempo $t-s$, y la suma sobre todos los tipos de individuos. Es decir, dados s y t tales que $0 < s < t$,

$$Z^i(t) = \sum_{l=1}^k \sum_{j=1}^{Z_l^i(s)} Z^{l,j}(t-s)$$

si existe n tal que $Z_n^i(s) > 0$, y $Z(t) = 0$ si $Z_n^i(s) = 0$ para todo $n \in \{1, \dots, k\}$. Por lo tanto,

$$\begin{aligned} E(\pi Z^i(t) | \mathcal{F}_s) &= E(\pi Z^i(t) | Z(s)) = \pi E \left(\sum_{l=1}^k \sum_{j=1}^{Z_l^i(s)} Z^{l,j}(t-s) \middle| Z(s) \right) \\ &= \pi \sum_{l=1}^k Z_l^i(s) E(Z^l(t-s)) = \sum_{l=1}^k Z_l^i(s) \pi_l e^{(t-s)\lambda} = \pi Z^i(s) e^{(t-s)\lambda}. \end{aligned}$$

Luego, $E(\pi Z^i(t) e^{-\lambda t} | \mathcal{F}_t) = \pi Z^i(s) e^{-\lambda s}$. Como $\pi E(Z^i(0)) = \pi_i$, $E(\pi Z^i(t) e^{-\lambda t}) = \pi_i$, y por lo tanto $\pi Z^i(t) e^{-\lambda t}$ es una \mathcal{F}_t -martingala para todo $i \in \{1, \dots, k\}$. \square

Usando que la matriz A puede descomponerse como $S + N$ donde S es una matriz diagonalizable y N una matriz nilpotente puede verse que $C(t) = e^{\lambda t} h \pi^t + O(t^r e^{|\lambda_2|t})$ para algún r positivo que no depende de t . Por lo tanto,

1. Caso sub-crítico.

Si $\lambda < 0$, $E(Z(t))$ decrece exponencialmente, (cada coordenada de la matriz tiende a 0 exponencialmente). Por lo tanto, para $\epsilon > 0$,

$$P(|Z(t)| > \epsilon) \leq \frac{E(|Z(t)|)}{\epsilon} \xrightarrow{t \rightarrow \infty} 0.$$

De donde $|Z(t)|$ converge en probabilidad a 0. Por lo tanto, existe una sub-sucesión de $|Z(t)|$ que converge a 0 casi seguramente. Como $|Z(t)| \in \mathbb{N}_0$ y si existe t_0 tal que $|Z(t_0)| = 0$ entonces $|Z(t)| = 0$ para todo $t \geq t_0$, tenemos que $\lim_{t \rightarrow \infty} |Z(t)| = 0$ casi seguramente.

Observemos que, $\lambda < 0$ sí y solo sí el autovalor de módulo máximo de la matriz $E(N)$ es menor que 1.

2. Caso crítico.

Si $\lambda = 0$, $E(Z(t)) = h \pi^t$ para todo $t \geq 0$. No nos da información suficiente para analizar el comportamiento de la población para tiempos suficientemente grandes. Sin embargo, por el Teorema 3.12 y usando el Teorema 3.9 aplicado a cada coordenada, tenemos que existe una variable aleatoria Z_∞ tal que $\lim_{t \rightarrow \infty} \pi Z(t) = Z_\infty$ c.s. Este límite no podría ser distinto de 0, pues la probabilidad de que todos los individuos tengan un único descendiente de su tipo es distinta de 1 y por lo tanto el proceso cambiaría de estado rápidamente. Para analizar este caso con más detalle puede verse Athreya [2, cap.V].

3. Caso super-crítico.

Si $\lambda > 0$, $E(Z(t))$ crece exponencialmente. Veamos que la cantidad de individuos de cada tipo crece exponencialmente. Si el proceso comienza con un individuo de tipo i , el proceso $Z_i^i(t)$ domina el proceso de ramificación discreto de un solo tipo de individuo con media mayor a 1, pues $Z_i^i(t)$ considera los individuos de tipo i , hijos de cualquier individuo $j \in \mathcal{S}$ y cuando decimos el proceso de ramificación de un solo individuo consideramos los individuos de tipo i descendientes de individuos de tipo i . Como la cantidad de individuos en este último proceso crece exponencialmente, también lo hace $Z_i^i(t)$ para todo $i \in \mathcal{S}$.

En este último caso el tamaño de la población tiende a infinito, pero queremos analizar cómo es la distribución de la población, si dividimos por la cantidad de individuos a tiempo t , es decir, estudiar el comportamiento de $\frac{Z(t)}{|Z(t)|}$ para t grande. En este caso,

$$E(Z(t)e^{-\lambda t}) \xrightarrow{t \rightarrow \infty} h\pi^t. \quad (3.14)$$

Es un hecho notable que el comportamiento del árbol este determinado por π, h, λ . Denotamos:

$$\Omega_s = \{X(t) \neq \emptyset \quad \forall t > 0\},$$

al evento en que la población sobrevive a todo tiempo. Para ver el comportamiento de la población, lo primero que queremos estudiar es una especie de ley de los grandes números para la distribución de la población, considerando el tiempo discreto, pero donde las unidades de tiempo son: $n\delta$, para $\delta > 0$. Para ello necesitamos el siguiente lema.

Lema 3.3. Sean $\{\xi_i^{(n)} : n, i \in \mathbb{N}\}$ variables aleatorias i.i.d. con valores en \mathbb{N} , con media $m > 1$ y segundo momento finito. Si $(V_n)_{n \in \mathbb{N}}$ son variables aleatorias con valores en \mathbb{N} tal que $V_{n+1} \geq \sum_{i=1}^{V_n} \xi_i^{(n)}$ para todo $n \in \mathbb{N}$. Entonces

$$\liminf_{n \rightarrow \infty} \frac{V_{n+1}}{V_n} \geq m \quad c.s.$$

en el evento $E = \{\lim_{n \rightarrow \infty} V_n \neq 0\}$.

Demostración. El proceso de ramificación simple definido por $Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)}$, cumple que Z_n crece exponencialmente si $\lim_{n \rightarrow \infty} Z_n \neq 0$. Por lo tanto por comparación, V_n crece exponencialmente en E .

Sea $m' < m$. Queremos acotar $P(V_{n+1} < m'V_n | V_n)$, para ello observemos que $\{V_{n+1} < m'V_n\} = \{V_{n+1} - mV_n < (m' - m)V_n\} \subseteq \{\sum_{i=1}^{V_n} \xi_i^{(n)} - mV_n < (m' - m)V_n\} \subseteq \{|\sum_{i=1}^{V_n} \xi_i^{(n)} - mV_n| > (m -$

$m')V_n\}$. Por lo tanto,

$$\begin{aligned} P(V_{n+1} < m'V_n | V_n) &\leq P\left(\left|\sum_{i=1}^{V_n} \xi_i^{(n)} - mV_n\right| > (m-m')V_n \mid V_n\right) \\ &\leq \frac{E(|\sum_{i=1}^{V_n} \xi_i^{(n)}|^2 | V_n)}{(m-m')^2 V_n^2} \\ &= \frac{V_n \text{Var}(\xi_i^{(n)})}{(m-m')^2 V_n^2} \leq \frac{C}{V_n}. \end{aligned}$$

Como suponemos que ocurre E , V_n crece exponencialmente, por lo tanto $P(\sum_{n=1}^{\infty} \frac{c}{V_n} < \infty | E) = 1$ y $\sum_{n=1}^{\infty} P(V_{n+1} < m'V_n | V_n) < \infty$ tiene probabilidad total en E .

Por el Lema de Borel Cantelli condicionado (ver Apéndice 5.1),

$$\left\{ \sum_{k_1}^{\infty} P(V_{k+1} < m'V_k | V_k) = \infty \right\} = \bigcap_{n_0} \bigcup_{k > n_0} \left\{ \frac{V_{k+1}}{V_k} < m' \right\}.$$

Por lo tanto,

$$P\left(\bigcap_{n_0} \bigcup_{k > n_0} \left\{ \frac{V_{k+1}}{V_k} < m' \right\} \mid E\right) = 0.$$

De donde,

$$1 = P\left(\bigcup_{n_0} \bigcap_{k > n_0} \left\{ \frac{V_{k+1}}{V_k} \geq m' \right\} \mid E\right) = P\left(\liminf_{k \rightarrow \infty} \frac{V_{k+1}}{V_k} \geq m' \mid E\right).$$

□

Ahora sí veamos como se comporta la distribución de la población, para ello vamos a seguir el procedimiento de Georgii-Baake[5]. El caso discreto fue hecho por Kurtz, Lyons, Pemantle y Peres ([9])

Proposición 3.2. *Sea $\delta, u > 0$, $i, j \in S$ y sea f una función medible tal que $c_j = E^j(f \circ X[0, u])$. Entonces,*

$$\lim_{n \rightarrow \infty} \frac{1}{Z_j(n\delta)} \sum_{x \in \mathbb{X}_j(n\delta)} f \circ X(x, [n\delta, n\delta + u]) = c_j \quad \text{c.s. en } \Omega_s$$

Demostración. Asumimos que δ es mayor que u y sea $\rho = E^j(Z_j(\delta)) > 1$. Tal δ existe porque $\lambda > 0$ y A es irreducible. Sea $\mathcal{F}_{n\delta}$ la σ -álgebra generada por $X[0, n\delta]$. Para cada $n \in \mathbb{N}$ consideremos las variables aleatorias

$$\varphi_{n,x} = f \circ X(x, [n\delta, n\delta + u])$$

con $x \in X_j(n\delta)$. $\varphi_{n,x}$ es $\mathcal{F}_{(n+1)\delta}$ -medibles, y condicional a $\mathcal{F}_{n\delta}$ son i.i.d. con media c_j . Esto implica que si construimos la sucesión $(\varphi_l)_{l \geq 1}$ en Ω_s enumerando primero $\{\varphi_{1,x} : x \in X_j(\delta)\}$ en algún orden, luego $\{\varphi_{2,x} : x \in X_j(2\delta)\}$, y así sucesivamente. Por la Ley de los Grandes Números,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{l=1}^k \varphi_l = c_j \quad \text{c.s. en } \Omega_s.$$

En particular, tomamos la subsucesión,

$$A_n = \frac{1}{\sum_{l=1}^n \psi_l} \sum_{l=1}^n \sum_{x \in X_j(l\delta)} \varphi_{l,x}.$$

Donde $\psi_l = Z_j(l\delta)$

La sucesión $(\psi_l)_{l \geq 1}$ domina el proceso de ramificación discreto de un solo tipo de individuo con media $p > 1$, pues ψ_l considera los individuos de tipo j , hijos de cualquier individuo $i \in S$ y cuando decimos el proceso de ramificación de un solo individuo consideramos los individuos de tipo j descendientes de individuos de tipo j . Como este último proceso sobrevive, también lo hace la sucesión $(\psi_l)_{l \geq 1}$.

Las variables ψ_l verifican las hipótesis del Lema 3.3 de donde,

$$\liminf_{l \rightarrow \infty} \frac{\psi_{l+1}}{\psi_l} \geq p \quad \text{c.s. en } \Omega_s.$$

Esto implica,

$$\limsup_{n \rightarrow \infty} \frac{\psi_l}{\psi_n} < \infty \quad \text{c.s. en } \Omega_s \quad l \in \{1, \dots, n-1\}.$$

Por lo tanto,

$$\limsup_{n \rightarrow \infty} \frac{\sum_{l=1}^{n-1} \psi_l}{\psi_n} < \infty \quad \text{c.s. en } \Omega_s.$$

Luego,

$$A_n + (A_n - A_{n-1}) \frac{\sum_{l=1}^{n-1} \psi_l}{\psi_n} \xrightarrow[n \rightarrow \infty]{\text{c.s. en } \Omega_s} c_j.$$

Veamos que esta es la convergencia que necesitamos.

$$\begin{aligned} (A_n - A_{n-1}) \frac{\sum_{l=1}^{n-1} \psi_l}{\psi_n} &= \left(\frac{1}{\sum_{l=1}^n \psi_l} \sum_{l=1}^n \sum_{x \in X_j(l\delta)} \varphi_{l,x} - \frac{1}{\sum_{l=1}^{n-1} \psi_l} \sum_{l=1}^{n-1} \sum_{x \in X_j(l\delta)} \varphi_{l,x} \right) \frac{\sum_{l=1}^{n-1} \psi_l}{\psi_n} \\ &= \frac{\sum_{l=1}^{n-1} \psi_l}{\psi_n \sum_{l=1}^n \psi_l} \sum_{l=1}^n \sum_{x \in X_j(l\delta)} \varphi_{l,x} - \frac{1}{\psi_n} \sum_{l=1}^{n-1} \sum_{x \in X_j(l\delta)} \varphi_{l,x}, \end{aligned}$$

de donde

$$\begin{aligned} A_n + (A_n - A_{n-1}) \frac{\sum_{l=1}^{n-1} \psi_l}{\psi_n} &= \frac{\psi_n + \sum_{l=1}^{n-1} \psi_l}{\psi_n \sum_{l=1}^n \psi_l} \sum_{l=1}^n \sum_{x \in X_j(l\delta)} \varphi_{l,x} - \frac{1}{\psi_n} \sum_{l=1}^{n-1} \sum_{x \in X_j(l\delta)} \varphi_{l,x} \\ &= \frac{1}{\psi_n} \sum_{l=1}^n \sum_{x \in X_j(l\delta)} \varphi_{l,x} - \frac{1}{\psi_n} \sum_{l=1}^{n-1} \sum_{x \in X_j(l\delta)} \varphi_{l,x} \\ &= \frac{1}{\psi_n} \sum_{x \in X_j(n\delta)} \varphi_{n,x} \end{aligned}$$

Concluimos que,

$$\frac{1}{\Psi_n} \sum_{x \in X_j(n\delta)} \varphi_{n,x} \xrightarrow[n \rightarrow \infty]{\text{c.s. en } \Omega_s} c_j.$$

Por lo tanto la proposición es cierta para δ suficientemente grande.

Si δ es arbitrario, podemos elegir $k \in \mathbb{N}$ tal que $\delta' = k\delta$ sea tan grande como necesitemos. Sea $0 \leq l < k$, aplicando el resultado al subárbol $X(x, [l\delta, \infty])$ con $x \in X(l\delta)$ y promediando, obtenemos:

$$\lim_{n \rightarrow \infty} \frac{1}{\Psi_{nk+l}} \sum_{x \in X_j((nk+l)\delta)} \varphi_{nk+l,x} = c_j \quad \text{c.s. en } \Omega_s.$$

Luego, queda demostrada la proposición. \square

Una aplicación de este resultado que usaremos más adelante, es el siguiente corolario.

Corolario 3.3. *Para cualquier $\delta, u > 0$ y para $i, j \in \mathcal{S}$*

$$C_{j,u}(n\delta) = \frac{1}{Z_j(n\delta)} \sum_{x \in X_j(n\delta)} Z(x, n\delta + u) \xrightarrow[n \rightarrow \infty]{\text{c.s. en } \Omega_s} E^j(Z(u)).$$

Demostración. Basta con tomar $f(x) = |x|$ y aplicar el Lema 3.2. \square

Este resultado nos dice como se comporta la cantidad de individuos cuando el tiempo tiende a infinito, pero estamos considerando el tiempo discreto, lo que veremos a continuación son algunas acotaciones para poder pasar a tiempo continuo. La idea de las acotaciones es ver que en un tiempo muy pequeño (δ) el proceso no puede variar mucho.

Lema 3.4. *Dado $\varepsilon > 0$ existe $\delta > 0$ tal que para todo $i, j \in \mathcal{S}$ y $k \in \mathbb{N}$, P^i -casi seguramente en Ω_s*

$$\limsup_{n \rightarrow \infty} \sup_{n\delta \leq s \leq (n+1)\delta} \frac{\|Z(s)\|}{\|Z(n\delta)\|} < 1 + \varepsilon, \quad (3.15)$$

$$\liminf_{n \rightarrow \infty} \inf_{n\delta \leq s \leq (n+1)\delta} \frac{Z_j(s)}{Z_j(n\delta)} > 1 - \varepsilon, \quad (3.16)$$

$$\liminf_{n \rightarrow \infty} \inf_{n\delta \leq s \leq (n+1)\delta} \inf_{k\delta \leq u \leq (k+1)\delta} \frac{\sum_{y \in X_j(s)} \|Z(y, u+s)\|}{\sum_{y \in X_j(n\delta)} \|Z(y, n\delta+k\delta)\|} > 1 - \varepsilon. \quad (3.17)$$

Demostración. Sea $n\delta \leq s \leq (n+1)\delta$. Podemos escribir,

$$\|Z(s)\| = \sum_{x \in X(n\delta)} |X(x, s)| \leq \sum_{x \in X(n\delta)} M(x, [n\delta, (n+1)\delta])$$

donde $M(x, [n\delta, (n+1)\delta]) = \max_{n\delta \leq s \leq (n+1)\delta} |X(x, s)|$. Claramente, $\|Z(n\delta)\| \geq \max_{j \in \mathcal{S}} Z_j(n\delta)$. Por lo tanto,

$$\sup_{n\delta \leq s \leq (n+1)\delta} \frac{\|Z(s)\|}{\|Z(n\delta)\|} \leq \max_{j \in \mathcal{S}} \frac{1}{Z_j(n\delta)} \sum_{x \in \mathbb{X}_j(n\delta)} M(x, [n\delta, (n+1)\delta]).$$

Por la Proposición 3.2, la última expresión converge a $m(\delta) = \max_{j \in S} E^j(M(0, [0, \delta]))$ c.s. en $\Omega_s(f = M)$.

El proceso $M(0, [0, \delta])$ está dominado por el tamaño total a tiempo δ del siguiente proceso de ramificación modificado: las variables aleatorias $N_{x, \sigma(x)}$ las reemplazamos por $\min\{N_{x, \sigma(x)}, 1\}$, para que cada individuo tenga al menos un hijo de su propio tipo. Este último proceso tiene una matriz generadora finita, que llamamos A^+ . Por lo tanto, $m(\delta) \leq \max_{j \in S} (e^{\delta A^+}) \xrightarrow{\delta \rightarrow 0} 1$. Luego, (3.15) queda demostrada.

Para demostrar 3.16, basta con tomar $u = k = 0$ en (3.17). Por lo tanto veamos que es cierta (3.17).

Sea $n\delta \leq s \leq (n+1)\delta$ y $k\delta \leq u \leq (k+1)\delta$. Consideremos sólo los individuos $y \in \mathbb{X}(s)$ que están vivos a tiempo $n\delta$ y siguen vivos a tiempo $(n+1)\delta$ y sólo sus descendientes que viven durante el período $[(n+k)\delta, (n+k+2)\delta]$, es decir, $z \in X(y, s+u)$. Podemos acotar de la siguiente forma;

$$\sum_{y \in X_j(s)} \|Z(y, s+u)\| \geq \sum_{x \in X_j(n\delta)} \mathbf{1}_{\{\tau_{x, n\delta} > \delta\}} \sum_{z \in X(x, (n+k)\delta)} \mathbf{1}_{\{\tau_{z, (n+k)\delta} > 2\delta\}}$$

donde $\tau_{x, t} = \inf\{u > 0 / x \notin X(t+u)\}$. Notamos τ_x a $\tau_{x, 0}$.

Luego,

$$\frac{\sum_{y \in X_j(s)} \|Z(y, u+s)\|}{\sum_{y \in X_j(n\delta)} \|Z(y, n\delta+k\delta)\|} \geq \frac{\sum_{x \in X_j(n\delta)} \mathbf{1}_{\{\tau_{x, n\delta} > \delta\}} \sum_{z \in X(x, (n+k)\delta)} \mathbf{1}_{\{\tau_{z, (n+k)\delta} > 2\delta\}}}{\sum_{y \in X_j(n\delta)} \|Z(y, n\delta+k\delta)\|}$$

Por la Proposición 3.2)

$$\frac{\sum_{x \in X_j(n\delta)} \mathbf{1}_{\{\tau_{x, n\delta} > \delta\}} \sum_{z \in X(x, (n+k)\delta)} \mathbf{1}_{\{\tau_{z, (n+k)\delta} > 2\delta\}}}{\sum_{y \in X_j(n\delta)} \|Z(y, n\delta+k\delta)\|} \xrightarrow[n \rightarrow \infty]{c.s. \text{ en } \Omega_s} \frac{E^j(\mathbf{1}_{\{\tau_x > \delta\}} \sum_{z \in X(x, k\delta)} \mathbf{1}_{\{\tau_{z, k\delta} > 2\delta\}})}{\sum_{y \in X_j(0)} \|Z(y, k\delta)\|}.$$

Donde τ_x es el tiempo de vida del individuo inicial, que es de tipo j . El último denominador es exactamente $E^j(|X(k\delta)|)$, y el numerador, es igual a

$$E^j \left(\mathbf{1}_{\{\tau_x > \delta\}} \sum_{z \in X(x, k\delta)} \exp[-2\delta a_{\sigma(z)}] \right).$$

Sea $a = \max_{i \in S} a_i$, luego esta expresión es mayor o igual que:

$$\exp[-2\delta a] E^j(\mathbf{1}_{\{\tau_x > \delta\}} |X(k\delta)|) = \exp[-2\delta a] (E^j(|X(k\delta)|) - E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)|)).$$

Por lo tanto,

$$\frac{E^j(\mathbf{1}_{\{\tau_x > \delta\}} \sum_{z \in X(x, k\delta)} \mathbf{1}_{\{\tau_{z, k\delta} > 2\delta\}})}{\sum_{y \in X_j(0)} \|Z(y, k\delta)\|} \geq e^{-2\delta a} \left(1 - \frac{E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)|)}{E^j(|X(k\delta)|)} \right). \quad (3.18)$$

Analicemos por separado el último cociente. En el numerador condicionamos a lo ocurrido con la variable τ_x

$$\begin{aligned} E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)|) &= \int_0^\infty E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)| \mid \tau_x = s) a_x e^{-a_x s} ds \\ &= \int_0^\delta E^j(|X(k\delta)| \mid \tau_x = s) a_x e^{-a_x s} ds. \end{aligned}$$

La variable τ_x es el instante en el que el individuo inicial se reproduce, éste tiene $N_x = (N_{x1}, \dots, N_{xk})$ descendientes. Notamos $X^{\mathbf{n}}(t)$ al proceso que comienza con $\mathbf{n} = (n_1, \dots, n_k)$ individuos iniciales,

$$\begin{aligned} E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)|) &= \int_0^\delta a_x e^{-a_x s} \sum_{\mathbf{n} \in \mathbb{N}_0^k} E(|X^{\mathbf{n}}(k\delta - s)|) P(N_x = \mathbf{n}) ds \\ &= \int_0^\delta a_x e^{-a_x s} \sum_{i=1}^k \sum_{n_i \in \mathbb{N}_0} n_i E^i(|X(k\delta - s)|) P(N_x = \mathbf{n}) ds \\ &= \int_0^\delta a_x e^{-a_x s} E\left(\sum_{i=1}^k N_{xi} E^i(|X(k\delta - s)|)\right) ds \\ &= \int_0^\delta a_x e^{-a_x s} \sum_{i=1}^k E(N_{xi}) E^i(|X(k\delta - s)|) ds. \end{aligned}$$

Además como $(E(N_{ij}))$ es finita para todo $i, j \in \{1, \dots, k\}$

$$E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)|) \leq \max_{i \in S} \{E(N_{xi})\} a_x \sum_{i=1}^k \int_0^\delta e^{-a_x s} E^i(|X(k\delta - s)|) ds.$$

Usando que h es autovector a derecha de la matriz $E(Z(t))$ obtenemos

$$\begin{aligned} E^i(|X(k\delta - s)|) &= E^i(||Z(k\delta - s)||) = \sum_{j=1}^k E^i(Z_j(k\delta - s)) = \sum_{j=1}^k \frac{E^i(Z_j(k\delta - s)) h_j}{h_j} \\ &\leq \frac{\sum_{j=1}^k E^i(Z_j(k\delta - s)) h_j}{\min_{j \in S} \{h_j\}} = \frac{E^i(Z(k\delta - s)h)}{\min_{j \in S} \{h_j\}} = \frac{h_i e^{\lambda(k\delta - s)}}{\min_{j \in S} \{h_j\}}. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} E^j(\mathbf{1}_{\{\tau_x \leq \delta\}} |X(k\delta)|) &\leq \max_{i \in S} \{E(N_{xi})\} a_x \sum_{i=1}^k \int_0^\delta e^{-a_x s} \frac{e^{\lambda(k\delta - s)} h_i}{\min_{j \in S} \{h_j\}} ds \\ &\leq \frac{\max_{i \in S} \{E(N_{xi})\} a_x \sum_{i=1}^k h_i}{\min_{j \in S} \{h_j\}} \int_0^\delta e^{-a_x s} e^{\lambda(k\delta - s)} ds \\ &= \frac{\max_{i \in S} \{E(N_{xi})\} a_x \sum_{i=1}^k h_i}{\min_{j \in S} \{h_j\}} e^{\lambda k \delta} \left(\frac{1}{\lambda + a_x} - \frac{e^{-(\lambda + a_x) \delta}}{\lambda + a_x} \right). \end{aligned}$$

Por otro lado el denominador

$$\begin{aligned} E^j(|X(k\delta)|) &= \sum_{i \in S} E^j(Z_i(k\delta)) = \sum_{i \in S} \frac{E^j(Z_i(k\delta)) h_i}{h_i} \\ &\geq \frac{E^j(Z(k\delta)h)}{\max_{i \in S} \{h_i\}} = \frac{e^{\lambda k \delta} h_j}{\max_{i \in S} \{h_i\}}. \end{aligned}$$

Luego para δ suficientemente pequeño, siguiendo con la desigualdad (3.18),

$$\begin{aligned} \frac{E^j(\mathbf{1}_{\{\tau_x > \delta\}} \sum_{z \in X(x, k\delta)} \mathbf{1}_{\{\tau_{z, k\delta} > 2\delta\}})}{\sum_{y \in X_j(0)} \|Z(y, k\delta)\|} &\geq e^{-2\delta a} \left(1 - \frac{\frac{\max_{i \in S} \{E(N_{xi})\} a_x \sum_{i=1}^k h_i}{\min_{j \in S} \{h_j\}} e^{\lambda k \delta} \left(\frac{1}{\lambda + a_x} - \frac{e^{-(\lambda + a_x)\delta}}{\lambda + a_x} \right)}{\frac{e^{\lambda k \delta} h_j}{\max_{i \in S} \{h_i\}}} \right) \\ &= e^{-2\delta a} \left(1 - \frac{\max_{i \in S} \{h_i\} \max_{i \in S} \{E(N_{xi})\} a_x \sum_{i=1}^k h_i \left(\frac{1}{\lambda + a_x} - \frac{e^{-(\lambda + a_x)\delta}}{\lambda + a_x} \right)}{h_j \min_{j \in S} \{h_j\}} \right) > 1 - \varepsilon. \end{aligned}$$

pues $E(N_x) < \infty$, y $h_i \in (0, 1)$.

□

Ahora sí, estamos en condiciones de analizar el comportamiento del proceso, cuando este no se extingue.

Teorema 3.13 (Kesten-Stigum). *Consideramos el caso super-crítico $\lambda > 0$. Para el proceso de ramificación ya definido $Z(t)$, con $Z(0) = e_j$ para $j \in S$,*

$$\frac{1}{|X(t)|} \sum_{x \in X(t)} \delta_{\sigma(x)} = \frac{Z(t)}{\|Z(t)\|} \xrightarrow{t \rightarrow \infty} \pi \quad \text{c.s. en } \Omega_S.$$

Demostración. Sea $\varepsilon > 0$, por (3.14) existe $\delta > 0$ tal que para $u \in \delta \mathbb{N}$ suficientemente grande,

$$\|E^j(Z(u)e^{-\lambda u}) - h_j \pi\| < \varepsilon$$

para todo $j \in S$. Por el Corolario 3.3, podemos elegir $s \in \delta \mathbb{N}$ grande, tal que

$$\|C_{j,u}(s)e^{-\lambda u} - h_j \pi\| < \varepsilon \quad \text{c.s. en } \Omega_S.$$

Donde $C_{j,u}(s) = \frac{1}{Z_j(s)} \sum_{x \in X_j(s)} Z(x, s+u)$. Si denotamos $\Pi(t) = \frac{Z(t)}{\|Z(t)\|}$ y $a(t) = \frac{\|Z(t)\| e^{-\lambda u}}{\langle Z(t-u, h) \rangle}$,

$$\begin{aligned} \|a(t)\Pi(t) - \pi\| &= \left\| \frac{e^{-\lambda u} Z(t)}{\langle Z(t-u, h) \rangle} - \pi \right\| \\ &= \frac{1}{\langle Z(t-u, h) \rangle} \left\| e^{-\lambda u} Z(t) - \langle Z(t-u, h) \rangle \pi \right\| \\ &= \frac{1}{\langle Z(t-u, h) \rangle} \left\| e^{-\lambda u} \sum_{x \in X(t-u)} Z(x, t) - \pi \sum_{i \in S} Z_i(t-u) h_i \right\| \\ &\leq \sum_{i \in S} Z_i(t-u) \frac{1}{\langle Z(t-u, h) \rangle} \left\| \frac{e^{-\lambda u} \sum_{x \in X(t-u)} Z(x, t)}{Z_i(t-u)} - \pi h_i \right\| \\ &= \sum_{i \in S} Z_i(t-u) \frac{1}{\langle Z(t-u, h) \rangle} \|e^{-\lambda u} C_{j,u}(t-u) - \pi h_i\| < \varepsilon. \end{aligned}$$

Concluimos que, $\|a(t)\Pi(t) - \pi\| < \varepsilon$. Como π y $\Pi(t)$, son vectores de norma 1 obtenemos,

$$\|\Pi(t) - \pi\| < \varepsilon$$

para todo $t \in \delta n$ suficientemente grande, casi seguramente en Ω_S . Falta ver que la convergencia es para $t > 0$, para ello usemos las desigualdades vistas en el Lema 3.4.

Por (3.16) $\frac{Z_j(t)}{Z_j(n\delta)} \geq 1 - \varepsilon$, por (3.15) $\frac{\|Z(n\delta)\|}{\|Z(t)\|} > \frac{1}{1+\varepsilon}$, luego

$$\frac{Z_j(t)}{\|Z(t)\|} = \frac{Z_j(n\delta)}{\|Z(n\delta)\|} \frac{\|Z(n\delta)\|}{\|Z(t)\|} \frac{Z_j(t)}{Z_j(n\delta)} > \frac{1-\varepsilon}{1+\varepsilon}(\pi - \varepsilon).$$

para t suficientemente grande. Es decir,

$$\liminf_{t \rightarrow \infty} \frac{Z_j(t)}{\|Z(t)\|} = \pi.$$

Tanto $\Pi(t)$, como π son probabilidades en un espacio de estados finito, por lo tanto el límite superior y el límite inferior deben coincidir, pues si alguna coordenada del límite superior es mayor que la del límite inferior, para que la suma siga siendo 1 tendría que haber alguna coordenada que sea menor, y esto no puede ocurrir.

Con lo cual probamos que,

$$\lim_{t \rightarrow \infty} \Pi(t) = \pi.$$

□

Capítulo 4

Aplicaciones

4.1. Proceso de urna

Consideremos el siguiente experimento, de una urna que contiene bolitas de colores, se extrae una bolita y se sustituye por un número aleatorio de bolitas de diferentes colores, la distribución de cantidad de bolitas que agrego sólo depende del color de la bolita extraída. Bajo ciertas condiciones, las proporciones de colores diferentes tendrán un límite. Este límite lo calcularemos viendo que este experimento podemos pensar como un proceso de ramificación. Describamos en detalle el experimento.

En una urna hay bolitas de k colores distintos. El vector aleatorio $X(n)$ es la composición de la urna después de n extracciones

$$X(n) = (X_1(n), \dots, X_k(n)),$$

donde X_i es el número de bolitas del color i . La distribución inicial $X(0)$ es arbitraria. En la $(n+1)$ -ésima extracción, la probabilidad de sacar una bolita de color i , es:

$$\frac{X_i(n)}{\sum_{i=1}^k X_i(n)}.$$

Una vez que saco la bolita i repongo con N_{ij} bolitas de color j , con $1 \leq j \leq k$. Las variables aleatorias $(N_{ij}^l, l \in \mathbb{N})$ son independientes e idénticamente distribuidas, l indica que es la l -ésima bolita que extraje de la urna. Además cumplen las siguientes hipótesis:

1.
 - $N_{ij} \in \mathbb{N}_0$ para todo $i, j \in \{1, \dots, k\}$.
 - $N_{ii} \geq 1 \quad \forall i \in \{1, \dots, k\}$, (siempre se repone una bolita del color que se extrajo).
 - $P(\sum_{j=1}^k N_{ij} > 1) > 0$ para algún i (excluimos el caso degenerado que es reponer siempre una bolita, ya que en ese caso se estaría manteniendo la cantidad de bolitas iniciales).
2. $E(N_{ij}^2) < \infty$, para todo $i, j \in \{1, \dots, k\}$.

3. Sea M la matriz con entradas $M_{ij} = E(N_{ij})$. Suponemos que M es irreducible.

Teorema 4.1. *Sea $N = (N_{ij})_{i,j \in S}$ una matriz aleatoria que satisface:*

1. $N_{ij} \in \mathbb{N}_0$ para todo $i, j \in \{1, \dots, k\}$, $N_{ii} \geq 1 \quad \forall i \in \{1, \dots, k\}$.
2. $P(\sum_{j=1}^k N_{ij} > 1) > 0$ para algún i .
3. $E(N)$ es irreducible.

La ecuación

$$\mu(E(N) - Id) = \alpha\mu$$

tiene una única solución con $\alpha > 0$ y μ una medida.

Demostración. Puede verse en Seneta [12]. □

Por como definimos el proceso el generador es:

$$P(X(n+1) = \mathbf{z} + \mathbf{y} - e_i | X(n) = \mathbf{z}) = P(\mathbf{N}_i = \mathbf{y}) \frac{X_i(n)}{\sum_{i=1}^k X_i(n)}.$$

Esto describe el proceso de una urna parametrizada con k vectores aleatorios $\mathbf{N}_i = (N_{ij})_{1 \leq j \leq k}$, $1 \leq i \leq k$. Sea $\hat{X}(n)$ el vector de proporciones después de n extracciones, es decir para $i = \{1, \dots, k\}$

$$\hat{X}_i(n) = \frac{X_i(n)}{\sum_{j=1}^k X_j(n)}.$$

Para analizar la convergencia de $\hat{X}_i(n)$, veamos que se comporta igual que un proceso discreto inmerso en un proceso de ramificación continuo multitypo. Sea $Z(t)$ un proceso de ramificación multitypo a tiempo continuo, tal que cada individuo tiene siempre un descendiente de su tipo y además el tiempo de vida de cada individuo tiene distribución exponencial de parámetro 1 para todo tipo de individuo. Este proceso está dominado por debajo por un proceso de ramificación continuo de un solo tipo de individuo con media mayor a 1, pues si el proceso comienza con un individuo de tipo i , $Z_i(t)$ considera los individuos de tipo i descendientes de cualquier individuo y cuando decimos el proceso de ramificación de un solo tipo consideramos los individuos de tipo i descendientes de individuos de tipo i . Como supusimos que cada individuo siempre deja un descendiente de su tipo, tenemos que la media de la cantidad de individuos en el proceso simple es mayor a 1, por lo tanto este proceso no se extingue y tampoco lo hace $Z(t)$. Consideraremos las variables $(\xi_n)_{n \in \mathbb{N}}$ donde ξ_i es el tiempo de la i -ésima ramificación. Como el proceso no se extingue, $\xi_n \xrightarrow{n \rightarrow \infty} \infty$. Observemos que $\{\xi_n\}_{n \in \mathbb{N}}$ son las discontinuidades de $Z(t)$ y además que $\lim_{n \rightarrow \infty} \frac{Z(\xi_n)}{|Z(\xi_n)|} = \lim_{t \rightarrow \infty} \frac{Z(t)}{|Z(t)|}$ pues este último límite existe c.s..

Teorema 4.2. *Sean $\{Y_n, n \in \mathbb{N}\}$ el proceso de urna y $\{Z(t), t \geq 0\}$ el proceso de ramificación, definidos anteriormente. Entonces los procesos $\{Y_n, n \in \mathbb{N}\}$ y $\{Z(\xi_n), n \in \mathbb{N}\}$ son equivalentes.*

Demostración. Lo primero que notamos es que $\{Y_n\}_{n \in \mathbb{N}}$ es un proceso a tiempo discreto y es una cadena de Markov con probabilidades de transición homogéneas. Por lo tanto, si ambos procesos tienen el mismo estado inicial, nos basta con mostrar que ambos tienen igual probabilidades de transición. Supongamos que ambos comienzan iguales: $Z(0) = Y(0)$ y miremos el primer paso. El proceso de ramificación comienza con $Z_i(0)$ individuos de tipo i , con $i \in \{1, \dots, k\}$. El tiempo de vida de cada individuo tiene distribución exponencial de media 1. Así que el proceso espera un tiempo exponencial de media $\sum_{i=1}^k Z_i(0)$ y luego se ramifica, la probabilidad de que se ramifique un individuo de tipo i es $\frac{Z_i(0)}{\sum_{j=1}^k Z_j(0)}$. En el proceso de urna la probabilidad de elegir una bolita de color i es $\frac{Y_i(0)}{\sum_{j=1}^k Y_j(0)}$. Como ambos procesos se inicializaron iguales,

$$\frac{Z_i(0)}{\sum_{j=1}^k Z_j(0)} = \frac{Y_i(0)}{\sum_{j=1}^k Y_j(0)}.$$

Después de esa ramificación se incrementa un número aleatorio de individuos o bolitas, N_{ij} y el proceso se inicializa de vuelta, por la falta de memoria del tiempo de vida de los individuos. Esto muestra que si $Y(0) = Z(0)$ entonces Y_1 tiene la misma distribución que Z_1 . Este mecanismo de transición es idéntico para todo el proceso. \square

Como vimos que ambos procesos son iguales, por el Teorema 3.13

$$\lim_{n \rightarrow \infty} \frac{X_i(n)}{\sum_{j=1}^k X_j(n)} = \pi. \quad c.s.$$

4.2. Convergencia empírica a la distribución cuasiestacionaria

Dado $(X_n)_{n \in \mathbb{N}}$ un proceso de Markov en un espacio de estados finito nos interesa estudiar como evoluciona el proceso. Es decir queremos saber si las variables X_n tienen algún límite. Vimos que si el espacio de estados \mathcal{S} es finito, y la matriz de transición P es irreducible y aperiódica, existe una única distribución estacionaria π que cumple:

- $\sum_{i \in \mathcal{S}} \pi_i p_{ij} = \pi_j \quad \forall j \in \mathcal{S}, \pi_j > 0$ y $\sum_{j \in \mathcal{S}} \pi_j = 1$.

De donde, si comenzamos con la medida invariante y dejamos evolucionar el proceso, la distribución de X_n no cambia, es decir,

$$P_\pi(X_n = s_j) = \pi_j \quad \forall n \in \mathbb{N}.$$

Luego, π es una medida invariante sí y solo sí $\pi P = \pi$, es decir, π es autovector a izquierda asociado al autovalor 1.

- Dada μ una distribución inicial,

$$\lim_{n \rightarrow \infty} P_\mu(X_n = s_j) = \pi_j.$$

Es decir, que no importa con qué distribución comencemos, si dejamos evolucionar el proceso, este tendrá aproximadamente la distribución de la medida estacionaria.

- Convergencia de la distribución empírica

$$N^{-1} \sum_{n=1}^N \mathbf{1}_{\{X_n=s_j\}} \xrightarrow[c.s.]{} \pi_j.$$

Esta convergencia se deduce del Teorema Ergódico 1.6 tomando $f(X_n) = \mathbf{1}_{\{X_n=j\}}$.

Si queremos calcular la distribución estacionaria, tenemos varias opciones

- buscar el autovector asociado al autovalor 1 de la matriz P ,
- simular la cadena en N pasos, y usar la convergencia de la distribución empírica para estimar la distribución estacionaria,
- usar simulación perfecta, no vamos a entrar en detalle sobre este método pero puede verse en: <http://dimacs.rutgers.edu/dbwilson/exact.html>/

En la primera opción calculamos exacta cual es la distribución, el problema es que si el espacio de estados es muy grande, calcular los autovectores es muy engorroso. La segunda opción es una buena forma de aproximar la distribución estacionaria.

Por otro lado, consideremos el proceso de Markov $(X_n)_{n \in \mathbb{N}}$ definido en el espacio de estados $\mathcal{S} \cup \{0\}$ donde 0 es un estado absorbente, y P es la matriz de transición con P_S es irreducible y aperiódica. Llamamos T al primer momento en que la cadena es absorbida. Por el Teorema 2.2, existe una única distribución cuasiestacionaria \mathbf{v} que satisface:

- $\mathbf{v}_i \geq 0 \quad \forall i \in \{1, \dots, k\}, \quad \sum_{i=1}^k \mathbf{v}_i = 1,$
- $P_{\mathbf{v}}(X_n = s_j | T > n) = \mathbf{v}_j \quad \forall n \in \mathbb{N}_0 \quad \forall j = \{1, \dots, k\}.$

Además vimos que cumple:

- $\mathbf{v}P_S = \lambda\mathbf{v}$, donde λ es el autovalor de módulo máximo de P_S , λ es real y $0 < \lambda < 1$. (Puede verse en Seneta [12])
- $\lim_{n \rightarrow \infty} P_{\mu}(X_n = s_j | T > n) = \mathbf{v}_j$, donde μ es la distribución inicial.

Nos gustaría calcular \mathbf{v} , para eso podemos hallar los autovalores y autovectores de P_S , pero si el espacio de estados es muy grande, no es práctico. En este caso, usar el método de rechazo no es bueno, es decir, simular para estimar \mathbf{v} corriendo la cadena y descontando las veces que toca al estado absorbente, es malo, pues $P(T > n)$ tiende a cero exponencialmente. Por lo tanto, queremos un método para simular \mathbf{v} . Vamos a dar el método propuesto por Aldous, Flannery y Palacios [1]. La idea es definir un proceso donde la distribución empírica converja a \mathbf{v} .

Primero suponemos que conocemos a \mathbf{v} y consideramos el proceso $V = (V_n)_{n \in \mathbb{N}}$ definido en \mathcal{S} de la siguiente manera, V_n es X_n mientras que $T > n$, cuando $T = N$ no contamos la transición y la cadena X vuelve a \mathcal{S} en el paso $N + 1$ con la distribución de \mathbf{v} .

Teorema 4.3. Consideremos los procesos $(X_n)_{n \in \mathbb{N}}$ y $(V_n)_{n \in \mathbb{N}}$ construidos anteriormente. Si ν es la distribución cuasiestacionaria de $(X_n)_{n \in \mathbb{N}}$ entonces ν es la distribución estacionaria de $(V_n)_{n \in \mathbb{N}}$.

Demostración. Para comenzar verifiquemos que $(V_n)_{n \in \mathbb{N}}$ es una cadena de Markov:

Claramente,

$$P(V_{n+1} = s_j | V_n = s_i, \dots, V_0 = s_0) = P(V_{n+1} = s_j | V_n = s_i).$$

Sea Q la matriz de transición de V , donde las entradas de Q son:

$$Q_{ij} = P_{ij} + P_{i0}\nu_j \quad \forall i, j \in \{1, \dots, k\}$$

pues la forma de que la cadena V salte de s_i a s_j es que X pase de s_i a s_j , o que pase de s_i al 0, y luego al ser redistribuida con ν , esta vaya a s_j .

Luego,

$$P(V_{n+1} = s_j | V_n = s_i) = Q_{ij}$$

Como ν es la distribución cuasiestacionaria de X_n cumple lo siguiente:

1. $\nu_i \geq 0 \quad \forall i \in \{1, \dots, k\}$ y $\sum_{i=1}^k \nu_i = 1$,
2. $\nu P_S = \lambda_1 \nu$.

Luego, para ver que ν es la distribución estacionaria, solo falta ver que $\nu Q = \nu$. Sea $j \in \{1, \dots, k\}$,

$$\begin{aligned} \sum_{i=1}^k \nu_i Q_{ij} &= \sum_{i=1}^k \nu_i (P_{ij} + P_{i0}\nu_j) \\ &= \sum_{i=1}^k \nu_i P_{ij} + \sum_{i=1}^k \nu_i P_{i0}\nu_j \\ &= \lambda_1 \nu_j + \nu_j \sum_{i=1}^k \nu_i \left(1 - \sum_{z=1}^k P_{iz}\right) \\ &= \lambda_1 \nu_j + \nu_j \left(\sum_{i=1}^k \nu_i - \sum_{z=1}^k \sum_{i=1}^k \nu_i P_{iz}\right) \\ &= \lambda_1 \nu_j + \nu_j \left(1 - \sum_{z=1}^k \lambda_1 \nu_z\right) \\ &= \lambda_1 \nu_j + \nu_j (1 - \lambda_1) \\ &= \nu_j. \end{aligned}$$

Con lo cual probamos que ν es la distribución estacionaria de V_n . \square

Sabemos que la distribución empírica θ_N converge a la distribución estacionaria v , donde

$$\theta_i(N) = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{V_n=s_i\}}.$$

Por supuesto que como no conocemos v no podemos simular este proceso. Sin embargo podemos adaptarlo; si X va al estado absorbente en el paso N , la cadena X en el paso $N+1$ vuelve a \mathcal{S} usando la distribución empírica θ_N . Demos la definición formal del proceso de Markov $(V_n, \xi(n))_{n \in \mathbb{N}}$, si $\xi(n)$ es la distribución empírica de V a tiempo n .

Definición 4.1. Sea \mathcal{M} un conjunto de medidas en \mathcal{S} y $(V_n, \xi(n))$ en $\mathcal{S} \times \mathcal{M}$ una cadena de Markov, con $V_1 = i_1$ y $\xi_1 = e_{i_1}$ con transiciones:

$$P(V_{n+1} = s_j, \xi(n+1) = \xi + e_j | V_n = s_i, \xi(n) = \xi) = P_{ij} + P_{i0} \frac{\xi_j}{|\xi|}.$$

donde $\xi = \sum_{j \in \mathcal{S}} \xi_j$.

El siguiente teorema muestra que para el proceso adaptado, la distribución empírica converge a v .

Teorema 4.4. Para el proceso definido anteriormente,

$$\lim_{n \rightarrow \infty} n^{-1} \xi_n = v$$

donde v es la distribución cuasiestacionaria de X_n .

Demostración. A la definición anterior le agregamos un contador. Sea $C(m)$, el número de veces que la cadena X fue absorbida por el 0 hasta el paso m ,

$$C(m) = \sum_{i=1}^m \mathbf{1}_{\{X_i=0\}}.$$

Lo inicializamos con $C(1) = 0$. Veamos como afecta este contador en el generador. Suponiendo que $V_n = s_i$, y que X fue C veces absorbida, hay dos formas distintas de que $V_{n+1} = s_j$,

- la primera es que en ese paso X no fue absorbida, y pasó de s_i a s_j ;

$$P(V_{n+1} = s_j, \xi(n+1) = \xi + e_j, C(n+1) = C | V_n = s_i, \xi(n) = \xi, C(n) = C) = P_{ij},$$

- y la segunda que X fue absorbida, y al volver a \mathcal{S} con ξ fue a s_j ;

$$P(V_{n+1} = s_j, \xi(n+1) = \xi + e_j, C(n+1) = C + 1 | V_n = s_i, \xi(n) = \xi, C(n) = C) = P_{i0} \frac{\xi_j}{|\xi|}.$$

Sea $S_n = \min\{m/C(m) = n\}$ es decir, S_n es el tiempo en que la cadena X fue absorbida por n -ésima vez. Sea $\beta(n) = \xi(S_n)$ la medida de conteo empírica a ese tiempo. Vimos que el proceso de urna es equivalente a el proceso discreto inmerso en un proceso de ramificación multitypo a tiempo continuo. Veamos ahora que $\beta(n)$ es un proceso de urna y por lo tanto podremos usar los resultados de procesos de ramificación. La cadena comienza en el estado i , esto es en el proceso de urna, que la urna se inicializa con una bolita de color i . La cadena pasea por los estados según la matriz de transición P hasta que llega al estado absorbente. Esto es equivalente a reponer bolitas en la urna, se reponen N_{ij} bolitas, donde N_{ij} tiene la distribución de $\sum_{j=1}^{T-1} \mathbf{1}_{\{V_j(n)\}}$. Notemos que la distribución de reponer bolitas en la urna solo depende de la bolita extraída. La cadena vuelve a un estado $s_j \in \mathcal{S}$, usando $\beta(1)$, en la urna esto es elegir una bolita entre las bolitas que estan en la urna que son $\beta(1)$. Ambos procesos continúan con este mecanismo.

Queremos ver que las variables N_{ij} cumplen las hipótesis del Teorema 4.1 para que el autovalor de módulo máximo de $E(N)$ sea mayor a 1, y así estar en las condiciones del caso super-crítico del proceso de ramificación.

1. ■ $\sum_{n=0}^{T-1} \mathbf{1}_{\{V_n=s_j\}} \in \mathbb{N}_0 \quad \forall j \in \{1, \dots, k\},$
■ $\sum_{j=1}^k \sum_{n=0}^{T-1} \mathbf{1}_{\{V_n=s_j\}} = \sum_{n=0}^{T-1} \sum_{j=1}^k \mathbf{1}_{\{V_n=s_j\}} = \sum_{n=0}^{T-1} 1 = T \geq 1$ pues $V_0 = s_i$,
■ $P_i(\sum_{j=1}^k \sum_{n=0}^{T-1} \mathbf{1}_{\{V_n=j\}} > 1) = P_i(T > 1) > 0$ pues $P_{i0} \neq 1$.
2. $E_i[(\sum_{n=0}^{T-1} \mathbf{1}_{\{V_n=s_j\}})^2] \leq E_i(T^2) < c.$
- 3.

$$M_{ij} = E(N_{ij}) = E\left(\sum_{n=0}^{T-1} \mathbf{1}_{\{V_n=s_j\}} | V_0 = s_i\right) = E\left(\sum_{n=0}^{\infty} \mathbf{1}_{\{V_n=s_j, T>n\}} | V_0 = s_i\right) = \sum_{n=0}^{\infty} (P_{\mathcal{S}}^n)_{ij}$$

es decir,

$$M = \sum_{n=0}^{\infty} P_{\mathcal{S}}^n.$$

La hipótesis de irreducibilidad de $P_{\mathcal{S}}$ implica irreducibilidad de M .

Por lo tanto $(\beta(n))_{n \in \mathbb{N}}$ esta bajo las hipótesis del Teorema 3.13, sólo nos falta verificar que v es la solución del Teorema 4.1, es decir la única solución de $\mu R = \alpha \mu$, con α positivo y μ una medida.

$$\begin{aligned}
\sum_{i \in S} v_i r_{ij} &= \sum_{i \in S} v_i (M_{ij} - \mathbf{1}_{\{ij\}}) \\
&= \sum_{i \in S} v_i M_{ij} - \sum_{i \in S} v_i \mathbf{1}_{\{ij\}} \\
&= \sum_{i \in S} v_i \sum_{n=0}^{\infty} (P^n)_{ij} - v_j \\
&= \sum_{n=0}^{\infty} \sum_{i \in S} v_i (P^n)_{ij} - v_j \\
&= \sum_{n=0}^{\infty} \lambda^n v_j - v_j \\
&= v_j \frac{1}{1-\lambda} - v_j \\
&= v_j \frac{\lambda}{1-\lambda}
\end{aligned}$$

Como $\alpha = \frac{\lambda}{1-\lambda} > 0$, v es la única solución de $vR = \alpha v$. Luego por el teorema (3.13),

$$\frac{\beta(n)}{\sum_{j \in S} \beta_j(n)} \xrightarrow[n \rightarrow \infty]{c.s.} v.$$

Como $\beta_j(n) = \xi_j(S_n)$, $\beta_j(n)$ es cuantas veces la cadena X pasó por el estado s_j antes de la n -ésima absorción. Si sumamos sobre todos los estados de S , es la cantidad de estados por los que pasó la cadena, es decir, S_n .

$$\sum_{j \in S} \xi_j(S_n) = S_n$$

Luego,

$$\frac{\xi(S_n)}{S_n} \xrightarrow[n \rightarrow \infty]{c.s.} v$$

Tenemos la convergencia que buscamos pero en lugar de para todo n sólo en los momentos en los que la cadena fue absorbida. Para ver que la convergencia es cierta para $n \in \mathbb{N}$ primero veamos que $\frac{S_{n+1}}{S_n} \rightarrow 1$,

$$P\left(\left|\frac{S_{n+1}}{S_n} - 1\right| > \epsilon\right) \leq \frac{E\left[\left(\frac{S_{n+1}-S_n}{S_n}\right)^2\right]}{\epsilon^2} \leq \frac{E\left[(S_{n+1}-S_n)^2\right]}{n^2 \epsilon^2}.$$

Además, $E\left[(S_{n+1}-S_n)^2\right] = E(E\left[(S_{n+1}-S_n)^2\right]|S_n) = E(E(T_0^2)) = E(T_0^2)$ que es finito. Sea $c > 0$ tal que $E(T_0^2) < c$. Luego,

$$P\left(\left|\frac{S_{n+1}}{S_n} - 1\right| > \epsilon\right) \leq \frac{c}{n^2 \epsilon^2}.$$

Como $\sum_{n=1}^{\infty} \frac{c}{n^2 \varepsilon^2}$ es finito

$$\frac{S_{n+1}}{S_n} \xrightarrow[n \rightarrow \infty]{c.s.} 1.$$

Sea m tal que, $S_n \leq m \leq S_{n+1}$. Con lo cual,

$$\xi(S_n) \leq \xi(m) \leq \xi(S_{n+1}).$$

Por un lado

$$\frac{\xi(m)}{S_n} \leq \frac{\xi(S_{n+1})}{S_n} = \frac{\xi(S_{n+1})}{S_{n+1}} \frac{S_{n+1}}{S_n} \rightarrow v,$$

por otro lado

$$\frac{\xi(m)}{S_{n+1}} \geq \frac{\xi(S_n)}{S_{n+1}} = \frac{\xi(S_n)}{S_n} \frac{S_n}{S_{n+1}} \rightarrow v.$$

Finalmente vimos que,

$$\lim_{n \rightarrow \infty} n^{-1} \xi(n) = v.$$

Por lo tanto demostramos que la distribución empírica de la cadena V converge a la distribución cuasiestacionaria de X . \square

Capítulo 5

Apéndice

Teorema 5.1. Sean X_1, X_2, \dots martingalas con $|X_{n+1} - X_n| < M$. Sean,

$$C = \{\lim_{n \rightarrow \infty} X_n \text{ existe y es finito}\},$$

$$D = \{\limsup_{n \rightarrow \infty} X_n = +\infty, \quad \liminf_{n \rightarrow \infty} X_n = -\infty\},$$

$$\text{Entonces, } P(C \cup D) = 1.$$

Demostración. Puede verse en Durret [4](pág. 239). □

Lema 5.1 (Borel Cantelli condicionado). Sea $(\mathcal{F}_n, n \in \mathbb{N})$ una filtración con $\mathcal{F}_0 = \{\emptyset, \Omega\}$ y $(A_n)_{n \in \mathbb{N}}$ una sucesión de eventos, tal que $A_n \in \mathcal{F}_n$. Entonces,

$$\bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} \{A_n\} = \left\{ \sum_{n \in \mathbb{N}} P(A_n | \mathcal{F}_{n-1}) = \infty \right\}.$$

Demostración. Consideremos las siguientes variables aleatorias, $X_0 = 0$ y

$$X_n = \sum_{j=1}^n \mathbf{1}_{\{A_j\}} - P(A_j | \mathcal{F}_{j-1}) \quad \forall n \in \mathbb{N}.$$

La variable X_n así definida es una martingala:

- $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, luego $A_j \in \mathcal{F}_n, \quad \forall j \leq n$ y por lo tanto $X_n \in \mathcal{F}_n$.

■

$$\begin{aligned} E(X_{n+1} | \mathcal{F}_n) &= E \left(\sum_{j=1}^{n+1} \mathbf{1}_{\{A_j\}} - P(A_j | \mathcal{F}_{j-1}) | \mathcal{F}_n \right) \\ &= E \left(\sum_{j=1}^{n+1} \mathbf{1}_{\{A_j\}} | \mathcal{F}_n \right) - \sum_{j=1}^{n+1} P(A_j | \mathcal{F}_{j-1}) \\ &= E \left(\sum_{j=1}^n \mathbf{1}_{\{A_j\}} | \mathcal{F}_n \right) + E(\mathbf{1}_{\{A_{n+1}\}} | \mathcal{F}_n) - \sum_{j=1}^n P(A_j | \mathcal{F}_{j-1}) - P(A_{n+1} | \mathcal{F}_n) \\ &= X_n. \end{aligned}$$

- $E(X_{n+1}) = E(X_n)$, y $E(X_{n+1}) = E(X_n) = E(X_0) = 0$.

Por otro lado, $|X_{n+1} - X_n| = |\mathbf{1}_{\{A_{n+1}\}} - P(A_{n+1}|\mathcal{F}_n)| \leq 1$. Por lo tanto estamos en las condiciones del teorema 5.1, lo que implica que $P(C \cup D) = 1$. Si ocurre $D = \{\limsup_{n \rightarrow \infty} X_n = +\infty, \liminf_{n \rightarrow \infty} X_n = -\infty\}$, como $X_n = \sum_{j=1}^n \mathbf{1}_{\{A_j\}} - P(A_j|\mathcal{F}_{j-1})$,

$$\sum_{j=1}^n \mathbf{1}_{\{A_j\}} \xrightarrow{n \rightarrow \infty} +\infty \quad \sum_{j=1}^n P(A_j|\mathcal{F}_{j-1}) \xrightarrow{n \rightarrow \infty} +\infty.$$

Si ocurre $C = \{\lim_{n \rightarrow \infty} X_n \text{ existe y es finito}\}$

$$\sum_{j=1}^n \mathbf{1}_{\{A_j\}} \xrightarrow{n \rightarrow \infty} +\infty \iff \sum_{j=1}^n P(A_j|\mathcal{F}_{j-1}) \xrightarrow{n \rightarrow \infty} +\infty.$$

Luego, $\{A_n \text{ ocurre infinitas veces}\} = \{\sum_{j=1}^{\infty} P(A_j|\mathcal{F}_{j-1}) = +\infty\}$. □

Bibliografía

- [1] D. Aldous, B. Flannery, and J.L. Palacios, *Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of markov chains.*, Probability in the Engineering and Informational Sciences. (1988), Vol. 2, no. 3 pp. 293–307. 68
- [2] K. B. Athreya and P. E. Ney, *Branching processes*, Dover Publications Inc., Mineola, NY, 2004, Reprint of the 1972 original [Springer, New York; MR0373040]. MR MR2047480 33, 55
- [3] Pierre Brémaud, *Markov chains*, Texts in Applied Mathematics, vol. 31, Springer-Verlag, New York, 1999, Gibbs fields, Monte Carlo simulation, and queues. MR MR1689633 (2000k:60137) 11, 14, 16
- [4] Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001) 39, 75
- [5] Hans-Otto Georgii and Ellen Baake, *Supercritical multitype branching processes: the ancestral types of typical individuals*, Adv. in Appl. Probab. **35** (2003), no. 4, 1090–1110. MR MR2014271 (2005a:60132) 33, 42, 57
- [6] Olle Häggström, *Finite Markov chains and algorithmic applications*, London Mathematical Society Student Texts, vol. 52, Cambridge University Press, Cambridge, 2002. MR MR1908939 (2003f:60002) 11
- [7] Theodore E. Harris, *The theory of branching processes*, Dover Phoenix Editions, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1963 original [Springer, Berlin; MR0163361 (29 #664)]. MR MR1991122 33, 42, 45
- [8] Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002) 49
- [9] Thomas Kurtz, Russell Lyons, Robin Pemantle, and Yuval Peres, *A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes*, Classical and modern branching processes (Minneapolis, MN, 1994), IMA Vol. Math. Appl., vol. 84, Springer, New York, 1997, pp. 181–185. MR MR1601737 (98j:60122) 57

- [10] J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144) 54
- [11] Lawrence Perko, *Differential equations and dynamical systems*, third ed., Texts in Applied Mathematics, vol. 7, Springer-Verlag, New York, 2001. MR MR1801796 (2001k:34001) 45, 53
- [12] E. Seneta, *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York, 2006, Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544]. MR MR2209438 30, 66, 68