

# Distribuciones cuasi-estacionarias y procesos de Fleming-Viot en espacios finitos

Pablo Groisman  
Dpto. de Matemática, FCEN-UBA - CONICET

## Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Cadenas de Markov a tiempo continuo</b>	<b>3</b>
2.1. Definición y Construcción . . . . .	3
2.2. Generador . . . . .	6
2.3. Medidas invariantes - Ergodicidad . . . . .	8
<b>3. Cadenas de Markov con estados absorbentes</b>	<b>9</b>
3.1. Ejemplos. . . . .	10
3.1.1. Modelo SIR . . . . .	10
3.1.2. Modelo logístico estocástico . . . . .	12
3.1.3. Paseos al azar en grafos: Internet . . . . .	12
<b>4. Distribuciones cuasi-estacionarias</b>	<b>13</b>
4.1. Evolución condicionada . . . . .	13
4.2. Distribuciones cuasi-estacionarias . . . . .	14
4.3. El tiempo de absorción . . . . .	16
4.3.1. La condición de Ferrari-Kesten-Martínez-Picco . . . . .	17
<b>5. El Proceso de Fleming-Viot</b>	<b>17</b>
5.1. Introducción . . . . .	17
5.2. Correlaciones . . . . .	19
5.3. Convergencia . . . . .	19
<b>6. Espacios numerables</b>	<b>22</b>

1	INTRODUCCIÓN	2
6.1.	Procesos de Ramificación subcríticos y el límite de Yaglom . . . . .	22
6.2.	Procesos de Nacimiento y Muerte . . . . .	23
6.3.	La condición de Ferrari-Kesten-Martínez-Picco . . . . .	23
<b>A.</b>	<b>Exponenciales de Matrices</b>	<b>23</b>
<b>B.</b>	<b>Teorema de Perron-Frobenius</b>	<b>25</b>
<b>C.</b>	<b>Desigualdad de Gronwall</b>	<b>25</b>

## Peligro! Notas en construcción

Lo primero para decir de estas notas es que están en construcción. Fueron escritas contra reloj para el minicurso con el mismo nombre (o parecido) del X Congreso Dr. Antonio Monteiro y no son, en varios aspectos, satisfactorias, aunque posiblemente sean mejores que la ausencia de las mismas.

Las notas contienen el material que se dará en el curso, muchos errores de tipeo y tal vez alguno de los otros. Además son incompletas, sobre todo para la parte de espacio de estados numerables. Esto se debe simplemente a falta de tiempo.

Así mismo, estas notas apelan exhaustivamente al Teorema de Perron-Frobenius. Esto no es deseable como se menciona en las mismas. Para el caso de cadenas de Markov, no hay ninguna necesidad de recurrir a este teorema y pueden encontrarse demostraciones muy buenas de la ergodicidad y convergencia al equilibrio en los lugares citados en este texto.

Para el caso de cadenas con estados absorbentes la existencia de la distribución cuasi-estacionaria y del límite de Yaglom no es evidente que pueda ser probada con argumentos más constructivos.

## 1. Introducción

Muchos procesos de Markov tienen un estado absorbente. Esto significa que existe un estado del cual el proceso no sale. Bajo condiciones de irreducibilidad este estado absorbente es alcanzado con probabilidad uno y entonces la versión estacionaria del proceso carece de todo interés: se trata de un proceso constantemente igual al estado absorbente con probabilidad uno.

Una buena forma de entender el comportamiento asintótico de estos modelos es preguntarse por la evolución del proceso condicional a que no ha sido absorbido y analizar el comportamiento asintótico de la evolución condicionada así como también distribuciones invariantes para esta evolución.

En estas notas tratamos este problema para procesos de Markov a tiempo continuo en espacio de estados finito. La condición de espacio de estados finito simplifica

muchísimo la situación y puede darse una teoría bastante cerrada. Cuando el espacio de estados es numerable, o incluso no numerable, la situación se torna mucho mas compleja y no existe una teoría unificada. Al final de las notas damos algunos ejemplos y mostramos las grandes diferencias que tienen los modelos en espacios de estados infinitos respecto del caso finito.

Comenzamos estas notas construyendo cadenas de Markov a tiempo continuo en espacio de estados finito y probando sus principales propiedades. Esta exposición es extremadamente breve y esta acompañada de referencias donde se puede hallar una exposición completa sobre este tópico.

A continuación tratamos el caso de las cadenas de Markov con estados absorbentes. Damos ejemplos y probamos que efectivamente la cadena es absorbida con probabilidad uno.

Para estos procesos introducimos la evolución condicionada y vemos cómo calcular su comportamiento asintótico que se conoce como el límite de Yaglom. Así mismo damos la noción de distribución cuasi-estacionaria (invariante para la evolución condicionada) y probamos el Teorema de Darroch y Seneta: existencia, unicidad y convergencia a la medida cuasi-estacionaria.

Por último introducimos el proceso de Fleming-Viot asociado a un proceso de Markov con estado absorbente. Este proceso intenta reproducir la dinámica de la evolución condicionada. Consiste de  $N$  partículas que se mueven en el espacio de estados del proceso original de forma independiente y todas siguiendo la ley del proceso original. Cuando una partícula es absorbida, inmediatamente es reencarnada en una de las partículas que están vivas dentro de la caja (elegida aleatoriamente) y el proceso continúa.

En el espacio de estados del proceso original consideramos la medida (aleatoria) dada por la cantidad de partículas que hay en cada sitio (normalizada para ser una medida de probabilidad) y esperamos que esta medida aproxime a la evolución condicionada.

En la Sección 5 definimos este proceso y probamos sus propiedades: ergodicidad, estimación de correlaciones y convergencia a la distribución cuasi-estacionaria (bajo la medida invariante) o a la evolución condicionada.

## 2. Cadenas de Markov a tiempo continuo

### 2.1. Definición y Construcción

Un proceso estocástico es una familia de variables aleatorias  $X = (X_t, t \in I)$  definidas en un mismo espacio muestral  $(\Omega, \mathcal{F}, \mathbb{P})$ . En nuestro caso, como estamos interesados en procesos a tiempo continuo, tomaremos a lo largo de todas estas notas como conjunto de índices  $I = \mathbb{R}_{\geq 0}$ .

De la definición surge que un proceso estocástico es una función definida en  $\Omega \times \mathbb{R}_+$  con valores en cierto espacio de estados  $\Lambda$ , i.e.  $X: \Omega \times \mathbb{R}_+ \rightarrow \Lambda$ , lo cual da dos interpretaciones posibles de  $X$ .

- Para cada  $t \geq 0$  tenemos una variable aleatoria  $X_t: \Omega \rightarrow \Lambda$ .
- Para (casi) todo  $\omega \in \Omega$  tenemos una *trayectoria* del proceso  $X(\omega): \mathbb{R}_+ \rightarrow \Lambda$ .

Usaremos a menudo ambas interpretaciones. En la segunda, hablamos de las trayectorias del proceso y las pensamos como funciones definidas en  $\mathbb{R}_+$  o alternativamente en todo  $\mathbb{R}$ .

**Definición 2.1.** Diremos que el proceso estocástico  $X = (X_t, t \geq 0)$  a valores en el espacio (finito)  $\Lambda$  es Markov si para todo  $0 \leq s \leq t$  se tiene

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

Es decir que si conocemos la trayectoria del proceso hasta el tiempo  $s$ , la posición del proceso a tiempo  $t$  sólo depende de la posición del proceso a tiempo  $s$  y no de todo el pasado.

**Ejemplo 2.1.** Un ejemplo de un proceso de Markov es el Proceso de Poisson (aunque el espacio de estados no es finito, es  $\mathbb{N}$ ). Dada una sucesión de variables aleatorias exponenciales independientes de parámetro  $\lambda$ ,  $E_1, \dots, E_n, \dots$  construimos el siguiente proceso. Pensamos a estas variables, por ejemplo, como el tiempo que van llegando personas a una cola.  $E_1$  es el tiempo en que llega el primer individuo,  $E_2$  es el tiempo entre la llegada del primer individuo y la del segundo, etc. En este caso, el proceso de Poisson es el que cuenta la cantidad de llegadas que hubo hasta tiempo  $t$ .

$$X_t = \sum_{n \geq 1} \mathbf{1}_{[\sum_{i=1}^n E_i, +\infty)}(t).$$

La markovianidad de este proceso sigue de la propiedad de falta de memoria de las variables exponenciales.

**Ejemplo 2.2.** Un ejemplo de un proceso que NO es Markov es el siguiente: Dada una sucesión de variables aleatorias i.i.d con distribución uniforme en  $[0, 1]$ , construimos el proceso

$$X_t = \sum_{n \geq 1} \mathbf{1}_{[\sum_{i=1}^n U_i, +\infty)}(t).$$

Acá el espacio de estados tampoco es finito pero puede reconstruirse este ejemplo fácilmente para que viva en un espacio de estados finito.

Dada una cadena de Markov a tiempo continuo denotaremos con  $\mathbb{P}^x$  a la probabilidad condicional a que  $X_0 = x$ , es decir

$$\mathbb{P}^x(X_t = y) = \mathbb{P}(X_t = y | X_0 = x),$$

y definimos para cada tiempo  $t \geq 0$  la matriz de transición

$$p_t(x, y) = \mathbb{P}^x(X_t = y).$$

Alternativamente, y abusando levemente de la notación. Dada una medida de probabilidad  $\mu$  en  $\Lambda$ , usaremos  $\mathbb{P}^\mu$  para indicar que el proceso a tiempo  $t = 0$  está distribuido según  $\mu$ . O sea

$$\mathbb{P}^\mu(X_t = y) = \sum_{x \in \Lambda} \mathbb{P}^x(X_t = y)\mu(x) = \sum_{x \in \Lambda} p_t(x, y)\mu(x) = (\mu P_t)(x).$$

En la última expresión pensamos a  $\mu$  como vector fila y a  $P_t$  como matriz con entradas  $p_t(x, y)$ .

Para construir una cadena de Markov en tiempo continuo observemos que, como el espacio de estados  $\Lambda$  es finito, la cadena debe moverse a saltos. Entonces espera un tiempo aleatorio en el estado  $x$  y luego salta a un estado  $y$  elegido al azar, luego vuelve a esperar una cantidad de tiempo aleatoria y vuelve a saltar al estado  $z$  y así sucesivamente. Entonces, para describir la cadena debemos especificar (1) la distribución de los tiempos de espera y (2) el mecanismo mediante el cual saltamos de un estado a otro. La markovianidad implica que los tiempos de espera tienen que ser variables aleatorias exponenciales, ya que deben ser variables aleatorias continuas con la propiedad de falta de memoria (¿por qué?). El parámetro de estas variables no puede depender del tiempo transcurrido en el estado ni de otros estados distintos al que estamos parados, por lo tanto consideraremos que tenemos una función  $q: \Lambda \rightarrow \mathbb{R}_{>0}$  y si el proceso está en  $x$  esperaremos un tiempo exponencial de parámetro  $q(x)$  para luego saltar a un nuevo estado  $y$ . Llamaremos a  $q(x)$  tasa de salida de  $x$ . Una vez que hemos decidido saltar, debemos elegir uno de los estados  $y \in \Lambda$  mediante un procedimiento aleatorio que solo debe depender del punto. Entonces consideraremos una nueva función  $p: \Lambda \times \Lambda \rightarrow [0, 1]$  de forma tal que la probabilidad de que si estamos en el estado  $x$  y decidimos hacer un salto, saltamos al estado  $y$  con probabilidad  $p(x, y)$ . La matriz  $p = p(x, y)_{x, y \in S}$  es lo que se denomina una matriz estocástica, es decir

- $p(x, y) \geq 0$  para todo  $x, y \in S$ .
- $\sum_{y \in S} p(x, y) = 1$ .

En estas notas consideraremos procesos en donde tanto  $q$  como  $p(\cdot, \cdot)$  no dependen de  $t$ . A esta propiedad se la denomina homogeneidad temporal.

La markovianidad del proceso  $X$  puede ser expresada también mediante la siguiente igualdad

$$\mathbb{P}^x(X_{t+s} = y | X_u, u \leq t) = \mathbb{P}^{X_t}(X_s = y) \quad \text{c.s.} \quad (1)$$

Del lado izquierdo tenemos una probabilidad condicional que es una variable aleatoria  $\sigma(X_u, u \leq t)$ -medible. Del lado derecho también tenemos una variable aleatoria que es función de  $X_t$ .

La ecuación (1) da otra interpretación de la markovianidad: condicionar a todo el pasado y mirar hacia adelante desde el tiempo  $t$  es lo mismo que volver a empezar el proceso desde el tiempo  $t$  tomando como dato inicial  $X_t$ .

La fórmula (1) requiere una demostración que no daremos aquí.

Para mas detalles (y mas rigor!) sobre la construcción de estos procesos y sus propiedades (ergodicidad, etc. ) ver por ejemplo el libro de Ferrari y Galves [6].

## 2.2. Generador

Un objeto fundamental para el estudio de las cadenas de Markov a tiempo continuo es el denominado generador infinitesimal. Consideremos el operador lineal  $L$  definido sobre las funciones de  $\Lambda$  en  $\mathbb{R}$  (que identificamos con  $\mathbb{R}^{|\Lambda|}$ ) por

$$Lf(x) = q(x) \sum_{y \in \Lambda} p(x, y)[f(y) - f(x)].$$

Este operador lleva consigo toda la información necesaria sobre los saltos de la cadena  $(X_t, t \geq 0)$ . Una forma de ver esto es la siguiente: estando en el estado  $x$ , el siguiente salto llega a tasa  $q(x)$  y cuando el salto ocurre, el nuevo estado será  $y$  con probabilidad  $p(x, y)$ . Este salto hace que la función  $f$  cambie en  $f(y) - f(x)$ . Es decir que  $Lf(x)$  representa la variación infinitesimal esperada de  $f(X_t)$  cuando  $X_t$  está en  $x$ . Más rigurosamente

**Teorema 2.1.** *Para todo  $x \in \Lambda$*

1. *Continuidad.*

$$\lim_{t \rightarrow 0^+} |\mathbb{E}^x f(X_t) - f(x)| = 0.$$

2. *Diferenciabilidad.*

$$\lim_{t \rightarrow 0^+} \left| \frac{\mathbb{E}^x f(X_t) - f(x)}{t} - Lf(x) \right| = 0. \quad (2)$$

3. *Más aún*

$$\frac{d}{dt} \mathbb{E}^x f(X_t) = \mathbb{E}^x Lf(X_t) \quad (3)$$

Acá  $\mathbb{E}^x$  denota esperanza calculada con  $\mathbb{P}^x$ , es decir condicional a que el proceso empezó en  $x$ . La demostración de este teorema puede hallarse en [6].

Observemos que si pensamos a  $f$  como un vector columna, el generador infinitesimal puede ser escrito de la forma

$$Lf(x) = (Qf)(x),$$

donde  $Q = (q(x, y), x, y \in \Lambda)$  con  $q(x, y) = q(x)p(x, y)$  si  $x \neq y$  y  $q(x, x) = -q(x)$ . Es decir que  $Q$  es la matriz (en la base canónica) de la transformación lineal  $L$ . Otra forma de representar a  $L$  es

$$Lf(x) = \sum_{y \in \Lambda} q(x, y)[f(y) - f(x)].$$

Decimos entonces que  $Q$  es la matriz de tasas del proceso  $X$  ya que representa la tasa a la que el proceso pasa al estado  $y$  cuando está en el estado  $x$ . Observar que tomando  $f = \mathbf{1}_y$  en (2) obtenemos

$$\mathbb{P}(X_{t+h} = y | X_t = x) = p_h(x, y) = q(x, y)h + o(h),$$

que es una buena forma de entender el significado de la matriz de tasas  $Q$ : infinitesimalmente, la probabilidad de estar en  $y$  a tiempo  $t+h$  dado que estamos en  $x$  a tiempo  $t$  viene dada por  $hq(x, y)$  (más términos de orden superior). Así mismo

$$\frac{d}{dt}p_t(x, y)|_{t=0} = q(x, y),$$

y de forma similar, por la propiedad de Markov y la homogeneidad

$$\begin{aligned} p_{t+h}(x, y) - p_t(x, y) &= \sum_{z \in \Lambda} [\mathbb{P}^x(X_{t+h} = y | X_t = z) - \mathbb{P}^x(X_t = y | X_t = z)] \mathbb{P}^x(X_t = z) \\ &= \sum_{z \in \Lambda} [p_h(z, y) - \mathbf{1}_z(y)] \mathbb{P}^x(X_t = z) = \sum_{z \in \Lambda} hq(z, y)p_t(x, z) + o(h) \\ &= h(P_t Q)(x, y) + o(h). \end{aligned}$$

Es decir que la matriz  $P_t$  verifica la ecuación diferencial ordinaria

$$\begin{cases} \frac{d}{dt}P_t = P_t Q, \\ P_0 = I. \end{cases}$$

Observar que tanto  $P_t$  como  $Q$  con matrices. La derivada del lado izquierdo debe ser entendida coordenada a coordenada. Estas ecuaciones se denominan *Ecuaciones de Kolmogorov*.

De la teoría de ecuaciones diferenciales ordinarias se desprende que

$$P_t = e^{tQ}, \quad \text{para todo } t \geq 0,$$

y por lo tanto

$$\mathbb{P}^\mu(X_t = y) = \sum_{x \in \Lambda} \mu(x) P_t(x, y) = \mu e^{tQ}(y)$$

Alternativamente, si escribimos  $\mu_t^\mu(y) := \mathbb{P}^\mu(X_t = y)$  para la distribución del proceso a tiempo  $t$  con condición inicial  $X_0$  distribuida según  $\mu$ , podemos usar las ecuaciones de Kolmogorov para obtener que  $\mu_t^\mu(y) = \mu P_t$  verifica

$$\begin{cases} \frac{d}{dt}\mu_t^\mu = \mu_t^\mu Q, \\ \mu_0^\mu = \mu. \end{cases} \quad (4)$$

Estas ecuaciones nos dan mucha información sobre cómo evoluciona el proceso cuando empieza con una condición inicial elegida aleatoriamente según la distribución  $\mu$ . Las utilizaremos en la próxima sección.

### 2.3. Medidas invariantes - Ergodicidad

Dado un proceso de Markov, una de las principales preguntas que intentamos responder es la siguiente: Si dejamos evolucionar el proceso durante un buen rato y luego lo observamos, ¿qué vemos? Es decir que queremos saber si las variables  $(X_t, t \geq 0)$  tienen un límite (en algún sentido) cuando  $t \rightarrow \infty$ .

Al estudio de estos límites se ha dedicado buena parte del esfuerzo de los probabilistas desde la aparición de estos procesos. Una forma de entenderlos es mirar la evolución del vector (distribución, medida de probabilidad, etc.)  $\mu_t^\mu$ .

De las ecuaciones de Kolmogorov (4) y de la continuidad de las soluciones de estas respecto del dato inicial, se desprende que si  $\nu = \lim_{t \rightarrow \infty} \mu_t^\mu$  entonces  $\nu$  debe satisfacer

$$\nu Q = 0.$$

Es decir que  $\nu$  es un equilibrio de (4). A esas medidas de probabilidad se las denomina *medidas invariantes* ya que, como se desprende de las ecuaciones verifican

$$\mu_t^\nu = \nu, \quad \text{para todo } t \geq 0.$$

Si tomamos una medida invariante  $\mu$  y sorteamos el dato inicial según ella, dejamos evolucionar el proceso y nos preguntamos por la distribución de  $X_t$  obtenemos que no ha cambiado, i.e.

$$\mathbb{P}^\mu(X_t = x) = \mu(x), \quad \text{para todo } t \geq 0.$$

Por eso es muy importante estudiar las medidas invariantes, porque es lo que veremos luego de que el proceso evoluciona un tiempo y entra en equilibrio. Equivalentemente, si  $\mu$  es invariante para la evolución ( $\mu_t^\mu = \mu$ ) entonces  $\mu Q = 0$  (¿por qué?).

Observemos que  $\mu$  es invariante si y sólo si  $\mu P_t = \mu$  para todo  $t > 0$ . Es decir que para todo  $t$ ,  $\mu$  es un autovector a derecha de  $P_t$  con autovalor 1.

El Teorema de Perron-Frobenius (ver Apéndice) nos garantiza la existencia de un único vector en estas condiciones si  $P_t$  tiene todas sus entradas positivas (o sea, si hay probabilidad positiva de ir de cualquier estado a cualquier otro). Es bueno notar que esta propiedad de la matriz  $P_t$  puede chequearse mirando la matriz de tasas (¿cómo?).

**Teorema 2.2.** *Sea  $X = (X_t, t \geq 0)$  un proceso de Markov a valores en el espacio de estados finito  $\Lambda$  con matriz de tasas  $Q$ . Si la matriz  $P_t = e^{tQ}$  tiene todas las entradas positivas entonces  $X$  admite una única medida invariante  $\mu$  que verifica  $\mu Q = 0$ .*

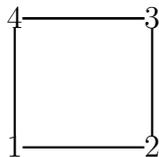
*Demostración.* Por el Teorema de Perron-Frobenius (ver apéndice) la matriz  $P_t$  tiene un autovalor de Perron-Frobenius  $r$ . Este autovalor no puede ser mayor que 1 ya que  $P_t$  manda medidas de probabilidad (vectores con entradas positivas que suman 1) en medidas de probabilidad. Por otra parte, como las filas de  $P_t$  suman 1, resulta que el vector  $w = (1, \dots, 1)$  es autovector a derecha de  $P_t$  con autovalor 1 para todo  $t \geq 0$ . Por lo tanto  $r = 1$  es el autovalor de Perron-Frobenius de  $P_t$  con autovector a izquierda  $v$  (normalizado para ser una probabilidad). En principio  $v$  podría depender de  $t$  pero veremos que no es el caso.

Como  $v$  verifica  $vP_t = ve^{tQ} = v$ , tenemos que  $v(tQ) = 0$  (¿por qué?). Viceversa, si tenemos un vector  $\mu$  que verifica  $\mu Q = 0$  entonces vale  $\mu P_t = \mu$ . Pero entonces para no contradecir al Teorema de Perron-Frobenius,  $v$  debe ser el único vector que verifica  $vQ = 0$  y por lo tanto esta es la única medida invariante del proceso  $X$ . ■

**Observación 2.1.** *Esta no es la mejor demostración que se puede dar del Teorema 2.2 por varias razones (que en realidad son la misma): en primer lugar, apela al Teorema de Perron-Frobenius que es puramente algebraico, construye al autovector de forma abstracta pero no a la medida de probabilidad. Esto hace que de la prueba no pueda extraerse ninguna información sobre la medida invariante.*

*Lo que se busca entonces es una demostración que construya la medida invariante usando información sobre la forma en que evoluciona el proceso  $X$  y que por ende va a tener que estar basada en argumentos probabilísticos. Demostraciones (distintas) que cumplen estos requisitos pueden verse en [6, 12]*

**Ejemplo 2.3** (Paseo al azar en un cuadrado). *Consideremos los cuatro vértices del cuadrado de unitario de  $\mathbb{R}^2$ :  $(0, 0)$ ,  $(1, 0)$ ,  $(1, 1)$ ,  $(0, 1)$ . Los numeramos de 1 a 4 para simplificar la notación y construiremos un proceso de Markov en el espacio de estados  $\Lambda = 1, 2, 3, 4$ . Nos moveremos por estos cuatro puntos pasando de un vecino al otro. Una vez que estamos parados en un punto esperamos un tiempo exponencial de parámetro 2 y elegimos a uno de los dos vecinos con probabilidad  $1/2$  para cada uno. Una vez elegido el lugar del salto, realizamos el salto y repetimos el procedimiento.*



*Este es un proceso de Markov con matriz de tasas dada por*

$$Q = \begin{pmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 1 & 0 & 1 & -2 \end{pmatrix}$$

Por la simetría del modelo, la única medida invariante es la uniforme en los 4 nodos y es fácil ver que este es el único autovector a derecha que tiene al 0 como autovalor.

### 3. Cadenas de Markov con estados absorbentes

Muchos procesos estocásticos tienen un estado absorbente, es decir que si la matriz de tasas viene dada por  $Q = q(x, y)_{x, y \in \Lambda \cup \{0\}}$ , entonces  $q(0, y) = 0$  para todo  $y \in \Lambda$  (si el proceso llega al estado 0, se queda ahí para siempre). En estos casos decimos que 0 es un estado absorbente. Si la matriz  $Q_\Lambda = Q$  restringida a  $\Lambda$  es irreducible y  $q(x, 0) > 0$  para algún  $x \in \Lambda$ , entonces la única medida invariante es  $\delta_0$ . Más aún existe un tiempo

de parada  $T^x$ , que denominamos tiempo de absorción de la cadena comenzada en  $x$  tal que  $X_t = 0$  para todo  $t \geq T^x$ .

De hecho, como veremos más adelante

$$\mathbb{P}^\mu(X_t > 0) = \sum_{y \in \Lambda} \sum_{x \in \Lambda} \mu(x) p_t(x, y) = (\mu P_t) \cdot (1, \dots, 1)^T \leq \sqrt{d} \|\mu e^{tQ}\| \leq C e^{\lambda_1 t}$$

Donde  $C$  es una constante,  $\|\cdot\|$  es la norma euclídea y  $\lambda_1$  es el mayor autovalor de  $Q$ , que es real y negativo. De este hecho se desprende que  $X_t \rightarrow 0$  cuando  $t \rightarrow \infty$  con probabilidad 1 (¿Por qué?), pero como  $X_t$  toma solo finitos valores entonces debe valer cero a partir de un momento. Es decir que existe un tiempo aleatorio  $T$ , que es finito con probabilidad 1, en que la cadena es absorbida.

**Ejercicio 3.1.** Probar el siguiente análogo del Teorema de Borel-Cantelli para tiempo continuo: sea  $X_t$  un proceso a puro salto, probar que si  $\int_0^\infty \mathbb{P}(X_t \neq 0) dt < \infty$  entonces

$$\lim_{t \rightarrow \infty} X_t = 0 \quad \text{c.s.}$$

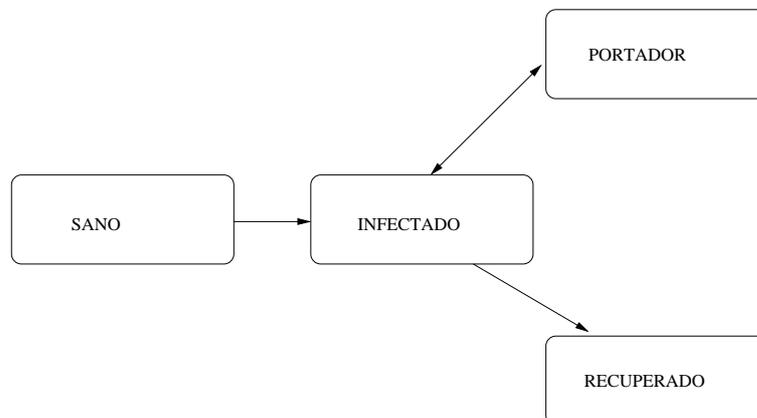
### 3.1. Ejemplos.

#### 3.1.1. Modelo SIR

Un modelo SIR es un modelo epidemiológico que calcula la evolución en el tiempo de la cantidad teórica de personas infectadas con una enfermedad contagiosa en una población cerrada. El nombre del modelo proviene del hecho de que involucra tres cantidades:  $S$  para el número de personas sanas,  $I$  para el número de infectados y  $R$  para el número de recuperados. Estos últimos son la clase de los inmunes. Incluye tanto a las que se han recuperado como a los que murieron.

Para algunas enfermedades como la tuberculosis, algunas personas infectadas nunca se recuperan completamente y llevan consigo la infección aunque no la sufren ellos mismos. Potencialmente podrían volver al estado de infectados y sufrir síntomas o podrían continuar en su estado de portadores e infectar a otros pero sin sufrir síntomas.

El diagrama que refleja esta situación es el siguiente



Una forma muy simple de este modelo (y poco realista) es pensar que cada uno de los integrantes de la población pasa de un estado al otro con ciertas tasas y que cada miembro de la población le ocurre esto de forma independiente.

Otro posible diagrama en el que las personas pueden recuperarse de la enfermedad pero esto no las hace inmunes es este.



En [2, 7] consideraron este ejemplo con tasas igual a 1 para pasar de un estado al otro (de los que están marcados). Si identificamos 2 = Sano, 1 = Infectado, 0 = Recuperado, obtenemos la siguiente matriz de tasas que tiene al 0 como estado absorbente.

$$Q = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

Observar que  $Q$  tiene autovalores  $\lambda_0 = 0$ ,  $\lambda_1 = -(3 - \sqrt{5})/2$ ,  $\lambda_2 = -(3 + \sqrt{5})/2$  cuyos autovectores (a izquierda) correspondientes son  $v_0 = (1, 0, 0)$ ,  $v_1 = (1, \frac{\sqrt{5}-3}{2}, \frac{-1+\sqrt{5}}{2})$  y  $v_2 = (1, -\frac{3+\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2})$ .

Entonces resulta que  $\exp(tQ) = C^{-1} \text{diag}[e^{\lambda_1 t}, e^{\lambda_2 t}, e^{\lambda_3 t}] C$ , donde  $C$  es la matriz de cambio de base que diagonaliza a  $Q$ . De aquí se desprenden también las probabilidades asintóticas. Como  $\lambda_0 = 0$  y los otros dos autovalores son negativos, la probabilidad de encontrar al proceso en 1 o en 2 tiende a cero exponencialmente mientras que la de encontrarlo en 0 tiende a 1 exponencialmente.

Si empezamos con una población muy grande, después de un tiempo largo la gran mayoría habrá sido absorbida. Cabe preguntarse igualmente como estarán distribuidos los que no han sido absorbidos.

Este problema podemos tratarlo desde un punto de vista analítico o probabilístico. Desde el punto de vista analítico, observemos que si miramos la matriz  $Q_\Lambda$  que es la restricción de  $Q$  a los estados transientes

$$Q_\Lambda = \begin{pmatrix} -2 & 1 \\ 1 & -1 \end{pmatrix},$$

esta matriz tiene como autovectores (a izquierda) a la restricción de  $v_2$  y  $v_3$  a estos mismos estados (con idénticos autovalores), ya que la primera fila es nula.

Entonces  $e^{tQ}$  tiende a cero a velocidad exponencial dada por  $e^{\lambda_1 t}$ . Podemos entonces normalizar dividiendo por esta cantidad y estudiar este límite. Si ese límite es no nulo, luego podemos normalizarlo para tener una medida de probabilidad en  $\{1, 2\}$  y esa sería la distribución buscada.

Observemos que  $e^{tQ_\Lambda} = \tilde{C} \text{diag}[e^{\lambda_1 t}, e^{\lambda_2 t}] \tilde{C}^{-1}$  y por lo tanto

$$e^{-\lambda_1 t} e^{tQ_\Lambda} = \tilde{C} \text{diag}[1, e^{(\lambda_2 - \lambda_1)t}] \tilde{C}^{-1}.$$

Como la matriz  $\tilde{C}$  tiene como columnas a  $v_1$  y  $v_2$  y  $\lambda_2 - \lambda_1 < 0$  tenemos

$$\lim_{t \rightarrow 0} e^{-\lambda_1 t} e^{tQ_\Lambda} = [\alpha \tilde{v}_1^T | \beta \tilde{v}_1^T]$$

para ciertas constante  $\alpha$  y  $\beta$  que dependen de  $\tilde{C}$  pero que no nos interesa calcular.

Es decir que la respuesta a la pregunta de cómo están distribuidos los que no han sido absorbidos después de un largo período de tiempo está dada por el vector  $\tilde{v}_1$ , normalizado para que sea una probabilidad en el  $\{1, 2\}$ . O sea,  $\nu(1) = (3 - \sqrt{5})/2$ ,  $\nu(2) = (\sqrt{5} - 1)/2$ . En la próxima sección veremos que podemos hacer esto para cualquier cadena con un estado absorbente.

### 3.1.2. Modelo logístico estocástico

Otro modelo epidemiológico es el modelo logístico estocástico. Consiste en un proceso de Markov que representa el número de individuos infectados  $I_t$  a tiempo  $t$  en una población constante formada por  $M$  individuos. El número de individuos susceptibles está representado por  $M - I_t$ .

Con la notación de la Sección 2.1 tenemos que  $(I_t, t \geq 0)$  es un Proceso de Markov con espacio de estados  $\Lambda = \{0, 1, 2, \dots, M\}$  y matriz de tasas

$$q(x, y) = \begin{cases} \lambda_x & y = x + 1, \\ \mu_x & y = x - 1, \\ -(\lambda_x + \mu_x) & y = x, \\ 0 & y \notin \{x - 1, x, x + 1\}, \end{cases}$$

para  $1 \leq x \leq M - 1$ . Finalmente  $q(0, y) = 0$  para todo  $y \in \Lambda$ ,  $q(M, M - 1) = \mu_M$  y  $q(M, y) = 0$  para todo  $y \neq M - 1$ . Las tasas de infección de los susceptibles y de recuperación de los infectados están dadas respectivamente por

$$\lambda_x = \lambda x(M - x), \quad \mu_x = \mu x,$$

donde  $\lambda, \mu > 0$  son los parámetros de infección y recuperación respectivamente.

Este proceso se mueve en el espacio de estados  $\Lambda$  creciendo en uno o decreciendo en uno. La tasa a la que lo hace depende el estado en el que esté. A medida que se acerca al total de la población, la tasa con la que crece se hace mas chica y la tasa con la que decrece se hace mas grande. Cuando la cantidad de infectados es  $M$  el proceso solo puede decrecer. Cuando la cantidad de infectados es 0, ahí se queda para siempre. El 0 es un estado absorbente.

### 3.1.3. Paseos al azar en grafos: Internet

Dado un conjunto (finito) de nodos  $V$  y un conjunto de aristas  $E \subset V \times V$ , obtenemos un grafo orientado  $G = (V, E)$ . Decimos que  $y \in V$  es vecino de  $x \in V$  si  $(x, y) \in E$ . Lo denotamos  $x \mapsto y$ .

Dado un grafo orientado, construimos un proceso de Markov  $X = (X_t, t \geq 0)$  denominado *Paseo al azar en  $G$* . Este proceso tiene espacio de estados  $\Lambda = V$  y matriz de tasas

$$q(x, y) = \frac{\mathbf{1}_{\{y \sim x\}}}{q(x)}, \quad q(x) = \sum_{y \in V} \mathbf{1}_{\{y \sim x\}}.$$

Si un nodo  $x$  no tiene ningún vecino entonces ese estado es absorbente y cuando el proceso llega ahí queda ahí para siempre.

Observar que los dos ejemplos anteriores caen en este contexto (¿cómo?).

Un ejemplo de esta situación se da cuando consideramos a Internet como un grafo. Decimos que la página web  $y$  es vecina de la página web  $x$  si en la página  $x$  hay un enlace a la página  $y$ .

El paseo al azar entonces modela a un navegante que en cada página que entra se pasa un tiempo exponencial, luego elige al azar entre todos los enlaces que encuentra en la página, hace click y pasa a la siguiente página.

Este es un proceso de Markov en un espacio de estados finito de tamaño estimado en aproximadamente 1.000.000.000.000 de sitios.

## 4. Distribuciones cuasi-estacionarias

### 4.1. Evolución condicionada

Cuando nos encontramos con un proceso de Markov con un estado absorbente la cuestión de las medidas invariantes se vuelve un poco sin sentido: con probabilidad 1, si esperamos un buen tiempo hallaremos al proceso en el estado 0. Sin embargo, en muchos casos antes de ocurrir esto el proceso alcanza un “metaequilibrio”. Estos estados pueden ser descriptos (al igual que con las medidas invariantes) como los posibles límites de la evolución condicionada. Mas precisamente, para un proceso que empieza a tiempo 0 con condición inicial distribuida según  $\mu$  definimos la evolución (del proceso) condicionada como

$$\varphi_t^\mu(x) = \mathbb{P}_\mu(X_t = x | X_t \neq 0) = \frac{\mathbb{P}_\mu(X_t = x)}{\mathbb{P}_\mu(X_t \neq 0)} = \frac{\mathbb{P}_\mu(X_t = x)}{\sum_{z \in \Lambda} \mathbb{P}_\mu(X_t = z)}, \quad x \in \Lambda.$$

Equivalentemente

$$\varphi_t^\mu(x) = \frac{\sum_{y \in \Lambda} \mu(y) p_t(y, x)}{\sum_{z \in \Lambda} \sum_{y \in \Lambda} \mu(y) p_t(y, z)} = \frac{(\mu P_t)(x)}{\sum_{z \in \Lambda} (\mu P_t)(z)}. \quad (5)$$

Como  $\mu P_t$  es derivable podemos diferenciar en (5) y usando (4) obtenemos

$$\frac{d}{dt} \varphi_t^\mu(x) = \sum_{y \in \Lambda} q(y, x) \varphi_t^\mu(y) + \sum_{y \in \Lambda} q(y, 0) \varphi_t^\mu(y) \varphi_t^\mu(x) \quad x \in \Lambda. \quad (6)$$

Con la condición inicial  $\varphi_0^\mu(x) = \mu(x)$ . Estas ecuaciones pueden ser interpretadas en forma similar a las ecuaciones de Kolmogorov: La variación de  $\varphi_t^\mu(x)$  esta dada por la matriz de tasas  $Q_\Lambda$ . Se agrega lo que está entrando a  $x$  que es  $\sum_{y \neq x} q(y, x)\varphi_t^\mu(x)$  y se resta lo que está saliendo de  $x$  que es  $-q(x, x)$ . Además aparece un término adicional que dice que también puede llegar masa a  $x$  proveniente de la masa que fue a parar a 0. Esta masa se reparte entre todos los estados  $x \in \Lambda$  según  $\varphi_t^\mu$  y surge de la normalización que hicimos al condicionar en (5).

Esta observación es muy importante para poder entender la evolución condicionada y es la que dará lugar a los procesos de Fleming-Viot que utilizaremos para aproximar la evolución condicionada y su límite.

## 4.2. Distribuciones cuasi-estacionarias

Al estudiar la evolución condicionada, nuevamente estamos interesados en el comportamiento asintótico  $\lim_{t \rightarrow \infty} \varphi_t^\mu(x)$  y para estudiar este límite debemos conocer el comportamiento en infinito de la matriz  $P_t^\Lambda = e^{tQ_\Lambda}$ . Cuando este límite existe y es una medida de probabilidad, se lo denomina *límite de Yaglom* para la medida  $\mu$ . Al igual que con las medidas invariantes, cuando este límite existe, es invariante para la evolución de la probabilidad condicionada  $\varphi_t^\mu$ . A estas medidas se las denomina *distribuciones cuasi-estacionarias*.

**Definición 4.1.** *Dada una medida de probabilidad  $\nu$  en  $\Lambda$ , decimos que es una distribución cuasi-estacionaria si  $\varphi_t^\nu = \nu$  para todo  $t \geq 0$ .*

De (6) se desprende que una medida de probabilidad  $\nu$  es una distribución cuasi-estacionaria si y sólo si

$$0 = \sum_{y \in \Lambda} q(y, x)\varphi_t^\mu(y) + \sum_{y \in \Lambda} q(y, 0)\varphi_t^\mu(y)\varphi_t^\mu(x) \quad x \in \Lambda. \quad (7)$$

Es decir que una distribución cuasi-estacionaria es un autovector a derecha de la matriz  $Q_\Lambda$  de autovalor  $-\sum_{y \in \Lambda} q(y, 0)\varphi_t^\mu(y)$ . Más adelante veremos que la distribución cuasi-estacionaria viene dada por el autovector a izquierda de Perron-Frobenius de  $e^{tQ_\Lambda}$ .

Observando (7) podemos ver que si tenemos una medida de probabilidad que es autovector a izquierda de  $Q_\Lambda$  entonces necesariamente el autovalor asociado es  $-\sum_{y \in \Lambda} q(y, 0)\varphi_t^\mu(y)$ .

En 1967, Darroch y Seneta ([4]) probaron que si la matriz  $Q_\Lambda$  es irreducible entonces existe una única distribución cuasi-estacionaria que coincide con el límite de Yaglom de cualquier distribución inicial.

**Teorema 4.1** (Darroch-Seneta). *Sea  $X = (X_t, t \geq 0)$  un proceso de Markov en espacio de estados  $\Lambda \cup \{0\}$  con matriz de tasas  $Q$ . Si la matriz  $Q_\Lambda$  es irreducible entonces  $X$  admite una única distribución cuasi-estacionaria  $\nu$ . Más aún, para cualquier distribución inicial  $\mu$  se tiene*

$$\|\varphi_t^\mu - \nu\| \leq e^{-\theta t}. \quad (8)$$

Donde  $\theta$  es cualquier número menor que el módulo de la diferencia entre el autovalor de Perron-Frobenius y el siguiente autovalor (el agujero espectral de  $Q_\Lambda$ ).

Por la teoría de Perron-Frobenius, como  $P_t^\Lambda$  es una matriz irreducible y positiva, existe un autovalor real y simple  $\alpha_t$  que acota superiormente al módulo de todos los autovalores de  $P_t^\Lambda$ . Se puede ver que  $\lambda_t$  es autovalor de  $P_t^\Lambda$  si y sólo si  $\lambda_t = e^{\rho t}$  con  $\rho$  autovalor de  $Q_\Lambda$ . Como  $P_t^\Lambda$  es estrictamente subestocástica, tenemos  $\lambda_t < 1$  y por lo tanto  $Q_\Lambda$  tiene un autovalor simple  $\rho < 0$  de forma tal que el autovalor de Perron-Frobenius de  $P_t^\Lambda$  viene dado por  $e^{\rho t}$ , es decir que todos los demás autovalores de  $Q$  tienen módulo menor que  $|\rho|$ .

Para calcular  $e^{tQ_\Lambda}$  utilizamos el siguiente Lema.

**Lema 4.1.** *Sea  $P_t^\Lambda = e^{tQ_\Lambda}$  una matriz subestocástica con entradas positivas y autovalores  $\rho_1, \dots, \rho_d$  ( $d = |\Lambda|$ ) ordenados de tal forma que  $\rho_1 > |\rho_2| \geq \dots > |\rho_d|$ . Sean  $v_1$  y  $w_1$  los autovectores de Perron-Frobenius a izquierda y derecha respectivamente normalizados de forma tal que  $\sum v_x = 1$ ,  $v_1^T w_1 = 1$ . Entonces existe  $k \in \mathbb{N}$  tal que*

$$e^{tQ} = e^{\rho t} w_1 v_1^T + O(t^k e^{t|\rho_2|}) \quad (9)$$

*Demostración.* Para calcular  $e^{tQ_\Lambda}$  procedemos como se indica en el Apéndice A:  $Q_\Lambda$  tiene una descomposición de la forma

$$Q_\Lambda = S + N,$$

donde  $S$  es diagonalizable en  $\mathbb{C}$  y  $N$  es nilpotente, además  $S$  y  $N$  conmutan y los autovalores de  $Q_\Lambda$  son los autovalores de  $S$  (con mismos autovectores, los restantes autovectores de  $S$  son autovectores generalizados de  $Q_\Lambda$ ). Entonces existe un polinomio  $p: \mathbb{R} \rightarrow \mathbb{R}$  tal que

$$e^{tQ_\Lambda} = e^{tS} p(tN).$$

Es de suponer entonces que el comportamiento en infinito de  $e^{tQ_\Lambda}$  está determinado por  $e^{tS}$ . Ahora bien, como  $S$  es diagonalizable, tenemos una matriz  $W$ , formada por filas  $w_1, \dots, w_d$  de autovectores a derecha de  $S$  y una matriz  $V$ , formada por columnas  $v_1, \dots, v_d$  de autovectores a izquierda de  $S$  tal que  $W = V^{-1}$  y

$$\text{diag}[\rho_1, \dots, \rho_d] = VSW,$$

es decir que

$$e^{tS} = W \text{diag}[e^{t\rho_1}, \dots, e^{t\rho_d}] V.$$

Entonces

$$\begin{aligned} e^{tQ_\Lambda} &= W \text{diag}[e^{t\rho_1}, 0, \dots, 0] V + W \text{diag}[0, e^{t\rho_2}, \dots, e^{t\rho_d}] V p(tN) \\ &\quad + W \text{diag}[e^{t\rho_1}, 0, \dots, 0] V (p(tN) - I). \end{aligned}$$

Donde  $I$  es la matriz identidad. Analicemos ahora término por término. El primer término del lado derecho es igual a  $e^{t\rho_1} w_1 v_1^T$  que es lo que queríamos obtener. El

segundo término puede ser acotado por ( $\|\cdot\|$  es la norma matricial asociada a la norma euclídea en  $\mathbb{R}^{|\Lambda|}$ ).

$$\|W \operatorname{diag}[0, e^{t\rho_2}, \dots, e^{t\rho_d}] V p(tN)\| \leq e^{t\rho_2} \|W\| \|V\| \|p(tN)\| \leq C t^k e^{t\rho_2},$$

para cierta constante  $C$  y  $k$  que no dependen de  $t$ . Por último el tercer término es exactamente cero ya que el rango de la matriz  $N$  está contenido en el núcleo de  $W \operatorname{diag}[e^{t\rho_1}, 0, \dots, 0] V$ . Por lo tanto

$$e^{tQ} = e^{t\rho_1} w_1 v_1^T + O(t^k e^{t\rho_2}),$$

que es lo que queríamos probar.

Estamos en condiciones de calcular el límite de Yaglom, tenemos

$$\begin{aligned} \varphi_t^\mu(x) &= \frac{\sum_{y \in \Lambda} \mu(y) p_t(y, x)}{\sum_{z \in \Lambda} \sum_{y \in \Lambda} \mu(y) p_t(y, z)} = \frac{(\mu P_t)(x)}{\sum_{z \in \Lambda} (\mu P_t)(z)} = \frac{\mu^T [e^{t\rho_1} w_1 v_1^T + O(t^k e^{t\rho_2})](x)}{\sum_{z \in \Lambda} \mu^T [(e^{t\rho_1} w_1 v_1^T + O(t^k e^{t\rho_2}))](z)} \\ &= \frac{(\mu^T w_1) v_1(x)}{\sum_{z \in \Lambda} ((\mu^T w_1) v_1)(z)} + O(t^k e^{t(\rho_2 - \rho_1)}) = v_1(x) + O(t^k e^{t(\rho_2 - \rho_1)}) \end{aligned}$$

Por lo tanto, poniendo  $\nu(x) = v_1(x)$  tenemos que para todo  $x \in \Lambda$

$$|\varphi_t^\mu(x) - \nu(x)| \leq O(t^k e^{t(\rho_2 - \rho_1)})$$

Esto prueba que el límite de Yaglom de cualquier distribución inicial  $\mu$  es el autovector de Perron-Frobenius de  $P_t$  y por lo tanto esa es la única medida quasi-estacionaria que denotaremos  $\nu$ . ■

### 4.3. El tiempo de absorción

Una característica muy importante de los procesos con estados absorbentes es el tiempo transcurrido hasta la absorción. Por eso definimos para cada estado  $x \in \Lambda$  el tiempo de parada

$$T^x := \inf\{t > 0, X_t = 0\}.$$

De la misma forma, para una medida inicial  $\mu$  definimos el tiempo de absorción de  $\mu$  como

$$T^\mu := \sum_{x \in \Lambda} T^x \mu(x).$$

El tiempo de absorción de una medida quasi-estacionaria tiene una distribución que es muy fácil de describir como muestra el siguiente lema.

**Lema 4.2.** *Sea  $X = (X_t, t \geq 0)$  un proceso de Markov a valores en  $\Lambda \cup \{0\}$  con el 0 como estado absorbente. Si  $\nu$  es una distribución quasi-estacionaria para  $X$  entonces  $T^\nu$  tiene distribución exponencial.*

*Demostración.* Probaremos que  $T^\nu$  tiene la propiedad de falta de memoria. De hecho

$$\begin{aligned} \mathbb{P}(T^\nu > t + s | T^\nu > t) &= \mathbb{P}^\nu(T > t + s | T > t) = \mathbb{P}^\nu(X_{t+s} \in \Lambda | X_t \neq 0) \\ &= \mathbb{P}^\nu(X_s \in \Lambda) = \mathbb{P}(T^\nu > s). \end{aligned}$$

■

En la sección anterior vimos que cuando  $\Lambda$  es finito (y  $Q$  irreducible) siempre existe una única distribución cuasi-estacionaria y el límite de Yaglom es esta distribución para toda distribución inicial  $\mu$ .

Cuando  $\Lambda$  es infinita (numerable) el panorama es completamente distinto: pueden existir infinitas distribuciones cuasi-estacionarias, una (como en el caso finito) o ninguna.

Este resultado, además de caracterizar la distribución del tiempo de absorción bajo la medida cuasi-estacionaria, es sumamente útil cuando se trata con espacios de estados infinitos porque dice que si hay una medida cuasi-estacionaria, el tiempo de absorción tiene que tener distribución exponencial (bajo la cuasi) pero como las medidas cuasi estacionarias son positivas en todos los puntos de  $\Lambda$  se tiene que  $\mathbb{E}(e^{\alpha \nu(x) T^x}) \leq \mathbb{E}(e^{\alpha \sum_x T^x \nu(x)}) < \infty$  si  $\alpha$  es pequeño. Como  $\nu(x) > 0$  tenemos que una condición necesaria para la existencia de una medida cuasi-estacionaria es que el tiempo de absorción de cada uno de los  $x \in \Lambda$  tenga un momento exponencial. Esto es equivalente a que ocurra para algún  $x$ .

### 4.3.1. La condición de Ferrari-Kesten-Martínez-Picco

En 1995, Ferrari, Kesten, Martínez y Picco ([5]) probaron que esta condición no solo es necesaria si no que además es (bajo ciertas condiciones) suficiente.

**Teorema 4.2.** *Sea  $\Lambda = \mathbb{N}$  y  $T$  el tiempo de absorción de la cadena  $X = (X_t, t \geq 0)$ . Asumamos  $\lim_{x \rightarrow \infty} \mathbb{P}^x(T > t) = 0$  y que para algún  $x \in \Lambda$ ,  $\mathbb{P}^x(T < \infty) = 1$ . Entonces existe una distribución cuasi-estacionaria si y sólo si existe  $\theta > 0$  tal que  $\mathbb{E}^x e^{\theta T} < \infty$*

## 5. El Proceso de Fleming-Viot

### 5.1. Introducción

Llamaremos Proceso de Fleming-Viot a un proceso que es ligeramente distinto al introducido por Fleming y Viot en [10], sin embargo tiene grandes similitudes y desde que ha sido introducido en [2] relacionado con las distribuciones cuasi-estacionarias ha quedado establecido llamar Proceso de Fleming-Viot (FV) al proceso que describiremos a continuación.

Dada una cadena de Markov  $X = (X_t, t \geq 0)$  a tiempo continuo con espacio de estados  $\Lambda \cup \{0\}$  finito, matriz de tazas  $Q$  y 0 como estado absorbente, le asociamos a  $X$  un proceso de FV que denotaremos  $\xi = (\xi_t, t \geq 0)$ . Fijada la cantidad de partículas

(individuos)  $N$ , el proceso FV asociado a  $X$  con  $N$  individuos es una cadena de Markov a tiempo continuo con espacio de estados  $\Lambda^N$ . El estado de la cadena  $\xi_t = (\xi_t^1, \dots, \xi_t^N)$  representa la posición (estado, tipo) de cada uno de los individuos. Cada uno de los individuos evoluciona en  $\Lambda$  de forma independiente y gobernado por la matriz de tasas  $Q$ . Cuando un individuo es absorbido (muere) instantáneamente elige a uno de los individuos “vivos” uniformemente y toma su posición (reencarna). Observar que como la cadena es a tiempo continuo, no hay muertes simultáneas (¿por qué?). Entre tiempos de absorción, cada una de las partículas evoluciona en  $\Lambda$  de forma independiente gobernada por  $Q$ .

Dada una configuración de partículas  $\xi = (\xi^1, \dots, \xi^N)$ , denotamos con  $\xi^{i,x}$  a la configuración que resulta de fijar a la partícula  $i$  en el estado  $x$  (y dejar al resto de las partículas en su lugar). Es decir  $\xi^{i,x}(i) = x$  y  $\xi^{i,x}(j) = \xi(j)$  para todo  $j \neq i$ . Con esta notación podemos describir al proceso de FV en términos de sus tasas. Este proceso pasa de la configuración  $\xi$  a la configuración  $\xi^{i,x}$  ( $x \neq \xi(i)$ ) a tasa

$$q(\xi(i), x) + q(\xi(i), 0) \frac{\sum_{j \neq i} \mathbf{1}_{\{\xi(j)=x\}}}{N-1}. \quad (10)$$

El primer término surge de que la partícula  $i$  evoluciona (mientras no es absorbida) como una cadena de Markov con matriz de tasas  $Q$ . El segundo término surge de que si la cadena es absorbida (esto ocurre a tasa  $q(\xi(i), 0)$ ) entonces elige a una de las partículas que está en  $\Lambda$ . Cada una de ellas con probabilidad  $1/(N-1)$ . Podemos escribir entonces el generador de este proceso como

$$L^N f(\xi) = \sum_{i=1}^N \sum_{x \in \Lambda \setminus \{\xi(i)\}} \left[ q(\xi(i), x) + q(\xi(i), 0) \frac{\sum_{j \neq i} \mathbf{1}_{\{\xi(j)=x\}}}{N-1} \right] (f(\xi^{i,x}) - f(\xi)). \quad (11)$$

Vamos a querer considerar también el proceso definido por la cantidad de partículas que hay en cada uno de los sitios  $x$ ,  $\eta(\xi, x)$ . También utilizaremos  $m(\xi)$  para la medida empírica inducida por la configuración  $\xi \in \cup_N \Lambda^N$

$$\eta(\xi, x) := \sum_{i=1}^N \mathbf{1}_{\{\xi(i) = x\}} \quad y \quad m(\xi) := \frac{\sum_{x \in \Lambda} \eta(\xi, x) \delta_x}{\sum_{x \in \Lambda} \eta(\xi, x)} \quad (12)$$

También usamos  $m_x(\xi)$  para denotar a  $m(\xi)(x)$ . Con esta notación podemos calcular, utilizando el generador (11) y la ecuación (2), la derivada temporal de  $\mathbb{E}_\xi^N[m_x(\xi_t)]$  (el super índice  $N$  indica la cantidad de partículas que tiene el proceso y el subíndice  $\xi$  indica que el proceso de FV comenzó con la configuración inicial  $\xi$ ). Obtenemos entonces

$$\frac{d\mathbb{E}_\xi^N[m_x(\xi_t)]}{dt} = \sum_{y \in \Lambda} q(y, x) \mathbb{E}_\xi^N[m_y(\xi_t)] + \frac{N}{N-1} \sum_{y \in \Lambda} q(y, 0) \mathbb{E}_\xi^N[m_y(\xi_t) m_x(\xi_t)]. \quad (13)$$

La idea detrás del proceso de Fleming-Viot es que cada una de las partículas está reproduciendo la evolución condicionada y por lo tanto si promediamos las posiciones de las partículas (medida empírica inducida por las partículas) obtendremos una aproximación de la evolución condicionada, y lo mismo para los límites temporales: si comenzamos el proceso de FV con una configuración  $\xi$  sorteada con la medida invariante de

FV entonces cuando  $N \rightarrow \infty$  si miramos la medida empírica inducida por las partículas deberíamos ver algo cercano a la distribución cuasi-estacionaria.

Observar que esto da una forma de simular a la distribución cuasi-estacionaria.

El objetivo de esta sección es probar estos resultados. No daremos todos los detalles, estos pueden ser encontrados en [1].

Los procesos de FV para aproximar distribuciones condicionadas y cuasi-estacionarias fueron introducidos por Burdzy, Hollyst y March en [2]. Ferrari y Maric ([7]) fueron los primeros en considerar el problema para espacio de estados numerables y en probar cotas para las correlaciones que, como veremos más adelante, son fundamentales para probar los resultados deseados.

## 5.2. Correlaciones

Suponiendo que para la configuración inicial  $\xi$  vale que  $\mu$  está cerca de  $m(\xi)$  y comparando la ecuación (13) con(6) puede esperarse que  $\mathbb{E}_\xi^N[m(\xi_t)]$  converja a  $\varphi_t^\mu$  cuando  $N \rightarrow \infty$ . También parece razonable que para probar esto último sea necesario probar que

$$\mathbb{E}_\xi^N[m_y(\xi_t)m_x(\xi_t)] - \mathbb{E}_\xi^N[m_y(\xi_t)]\mathbb{E}_\xi^N[m_x(\xi_t)] \rightarrow 0 \quad (N \rightarrow \infty).$$

Dicho en palabras, que las correlaciones entre las medidas de ocupación de dos sitios distintos tienden a cero cuando el número de partículas tiende a infinito. A esta propiedad se la denomina *Propagación del Caos*.

La prueba de esta propiedad es delicada. Involucra una construcción gráfica del proceso y debe ser probado el hecho de que si bien la posición de una partícula a tiempo  $t$  depende de la posición de las partículas sobre las que saltó alguna vez y las partículas sobre las que estas últimas saltaron, etc. el total de estas partículas es pequeño comparado con  $N$ . De hecho es de orden  $e^{kt}/N$  como muestra la siguiente proposición. La demostración puede hallarse en [1]

**Proposición 5.1.** *Para cada  $t > 0$ , y cualquier  $x, y \in \Lambda$*

$$\sup_{\xi \in \Lambda^N} |\mathbb{E}_\xi^N[m_x(\xi_t)m_y(\xi_t)] - \mathbb{E}_\xi^N[m_x(\xi_t)]\mathbb{E}_\xi^N[m_y(\xi_t)]| \leq \frac{\mathbf{1}_{\{x=y\}}}{N} + \frac{2}{N}(e^{2Ct} - 1). \quad (14)$$

Para cada  $N \geq 2$  el proceso de FV es irreducible (¿por qué?) y por lo tanto tiene una única medida invariante y comenzado con cualquier medida, el proceso converge a la medida invariante. Denotaremos a esta medida de probabilidad en  $\Lambda^N$  con  $\lambda^N$ .

## 5.3. Convergencia

La independencia asintótica (14) implica la convergencia de las medias empíricas en el proceso de FV a la evolución condicionada  $\varphi_t^{m(\xi)}$ , uniformemente en  $\xi$ . Más aún, el Teorema de Darroch y Seneta muestra que la evolución condicionada está cerca de la distribución condicionada uniformemente en  $\xi$ . Estos dos hechos dan lugar al siguiente resultado.

**Teorema 5.1.** *Para todo  $t > 0$*

$$\lim_{N \rightarrow \infty} \sup_{\xi \in \Lambda^N} \mathbb{E}_\xi^N [\|m(\xi_t) - \varphi_t^{m(\xi)}\|] = 0. \quad (15)$$

Más aún,

$$\lim_{N \rightarrow \infty} \int_{\Lambda^N} \|m(\xi) - \nu\| d\lambda^N(\xi) = 0. \quad (16)$$

Aquí  $\|\cdot\|$  denota la norma euclídea en  $\mathbb{R}^{|\Lambda|}$ .

**Observación 5.1.** *Notar que el límite (16) es consecuencia de (15) usando (8):*

$$\begin{aligned} \int_{\Lambda^N} \|m(\xi) - \nu\| d\lambda^N(\xi) &= \int_{\Lambda^N} \mathbb{E}_\xi^N [\|m(\xi_t) - \nu\|] d\lambda^N(\xi) \\ &\leq \int_{\Lambda^N} \mathbb{E}_\xi^N [\|m(\xi_t) - \varphi_t^{m(\xi)}\|] d\lambda^N(\xi) + \int_{\Lambda^N} \|\varphi_t^{m(\xi)} - \nu\| d\lambda^N(\xi) \\ &\leq \sup_{\xi \in \Lambda^N} \mathbb{E}_\xi^N [\|m(\xi_t) - \varphi_t^{m(\xi)}\|] + \sup_{\mu \in \mathcal{M}} \|\varphi_t^\mu - \nu\|. \end{aligned} \quad (17)$$

**Demostración del Teorema 5.1.** Teniendo en cuenta (17), debemos estimar la esperanza  $\mathbb{E}_\xi^N [\|m(\xi_t) - \varphi_t^{m(\xi)}\|]$ . Notar que

$$\mathbb{E}_\xi^N [\|m(\xi_t) - \varphi_t^{m(\xi)}\|] \leq \mathbb{E}_\xi^N [\|m(\xi_t) - \mathbb{E}_\xi^N [m(\xi_t)]\|] + \|\mathbb{E}_\xi^N [m(\xi_t)] - \varphi_t^{m(\xi)}\|. \quad (18)$$

Tomando  $y = x$  en (14) obtenemos

$$\mathbb{E} [(m_x(\xi_t) - \mathbb{E}[m_x(\xi_t)])^2] \leq \frac{2e^{2Ct}}{N}. \quad (19)$$

Utilizando la desigualdad de Jensen llegamos a

$$[\mathbb{E}_\xi^N \|m(\xi_t) - \mathbb{E}_\xi^N [m(\xi_t)]\|]^2 \leq \mathbb{E}_\xi^N [\|m(\xi_t) - \mathbb{E}_\xi^N [m(\xi_t)]\|^2] \leq \frac{2|\Lambda|e^{2Ct}}{N}. \quad (20)$$

El segundo término en (18) lo tratamos con el siguiente lema.

**Lema 5.1.** *Para cualquier  $T > 0$ ,*

$$\lim_{N \rightarrow \infty} \max_{0 \leq t \leq T} \max_{\xi \in \Lambda^N} \|\mathbb{E}_\xi^N [m(\xi_t)] - \varphi_t^{m(\xi)}\| = 0. \quad (21)$$

*Demostración.* Introducimos un poco de notación para simplificar.

$$u_x(t) = \mathbb{E}_\xi^N [m_x(\xi_t)], \quad y \quad v_x(t) = \varphi_t^{m(\xi)}(x). \quad (22)$$

Mostraremos que hay una constante  $B$  tal que para todo  $t > 0$

$$\frac{d}{dt} \|u(t) - v(t)\|^2 \leq B \|u(t) - v(t)\|^2 + \frac{4e^{2Ct} \sum_y q(y, 0)}{N}. \quad (23)$$

Como  $\|u(0) - v(0)\| = 0$ , el resultado se sigue de la Desigualdad de Gronwall que puede verse en el Apéndice C.

Fijamos  $t > 0$  y omitiremos escribir la dependencia temporal. De (13) obtenemos

$$\frac{du_x}{dt} = \sum_{y \in \Lambda} q(y, x)u_y + \sum_{y \in \Lambda} q(y, 0)u_xu_y + R_x(\xi, t) \quad (24)$$

$$\text{donde } R_x(\xi, t) = \sum_{y \in \Lambda} q(y, 0) \left[ \frac{N}{N-1} \mathbb{E}_\xi^N [m_y(\xi_t)m_x(\xi_t)] - \mathbb{E}_\xi^N [m_y(\xi_t)] \mathbb{E}_\xi^N [m_x(\xi_t)] \right].$$

La Proposición 5.1, implica que

$$\sup_{x \in \Lambda} \sup_{\xi} |R_x(\xi, t)| \leq \frac{2e^{2Ct} \sum_y q(y, 0)}{N}. \quad (25)$$

Por otra parte, de (6)

$$\frac{dv_x}{dt} = \sum_{y \in \Lambda} q(y, x)v_y + \sum_{y \in \Lambda} q(y, 0)v_xv_y. \quad (26)$$

Restando (26) de (24), obtenemos

$$\frac{d(u_x - v_x)}{dt} = \sum_{y \in \Lambda} q(y, x)(u_y - v_y) + \sum_{y \in \Lambda} q(y, 0)(u_xu_y - v_xv_y) + R_x(\xi, t).$$

Ahora,

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|u(t) - v(t)\|^2 &= \sum_x \frac{d(u_x - v_x)}{dt} (u_x - v_x) \\ &= \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(x, y)(u_y - v_y)(u_x - v_x) + \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(y, 0)(u_xu_y - v_xv_y)(u_x - v_x) \\ &\quad + \sum_x R_x(u_x - v_x). \end{aligned} \quad (27)$$

Tratamos cada término en el lado derecho de (27) por separado. Primero,

$$\left| \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(x, y)(u_y - v_y)(u_x - v_x) \right| \leq \left( \sum_{x, y \in \Lambda} q^2(x, y) \right)^{1/2} \|u - v\|^2. \quad (28)$$

Para tratar al segundo término observamos que  $u_xu_y - v_xv_y = v_x(u_y - v_y) + u_y(u_x - v_x)$  y por lo tanto el segundo término en (27) vale

$$\sum_{x \in \Lambda} \sum_{y \in \Lambda} q(y, 0)v_x(u_y - v_y)(u_x - v_x) + \sum_{x \in \Lambda} \sum_{y \in \Lambda} q(y, 0)u_y(u_x - v_x)(u_x - v_x). \quad (29)$$

Como  $v$  es acotada, el primer término en (29) puede ser tratado como (28) con  $q(x, y)$  reemplazada por  $\tilde{q}(x, y) := q(y, 0)v_x$ . El segundo término en (29) vale

$$\|u - v\|^2 \sum_y q(y, 0)u_y \leq \|u - v\|^2 \sup_y q(y, 0).$$

Finalmente, teniendo en cuenta que  $\sum_x v_x = \sum_x u_x = 1$ , el último término en (27) lo podemos acotar por

$$2 \sup_x |R_x| \leq \frac{4e^{2Ct} \sum_y q(y, 0)}{N}$$

Juntando todas estas cuentas obtenemos (23). ■

## 6. Espacios numerables

Si relajamos la condición de que  $\Lambda$  sea finito y permitimos que sea numerable todo un nuevo mundo se abre. Todo lo dicho en estas notas deja de valer en general. No es inminente ni la existencia ni la unicidad de la distribución quasi-estacionaria y tampoco el comportamiento de Fleming-Viot.

A continuación esbozamos solo dos problemas a modo de ejemplo para ver algunos de los problemas interesantes que aparecen en este contexto y algunos de los resultados conocidos con referencias a cada uno de ellos. Esta lista no es ni exhaustiva ni representativa.

### 6.1. Procesos de Ramificación subcríticos y el límite de Yaglom

Los procesos de ramificación, o procesos de Galton-Watson fueron introducidos por Francis Galton, quien investigaba sobre la posibilidad de extinción de los apellidos. En su modelo un hombre tenía cierta cantidad (aleatoria) de hijos varones y a su vez cada uno de ellos tenía una cantidad aleatoria de hijos varones. El proceso continúa mientras halla hijos varones (no se extinguen). El modelo supone que todos los hijos varones tienen hijos con la misma distribución y en forma independiente. Para una primera definición de este proceso conviene pensar al tiempo como discreto, representando a cada una de las generaciones. Puede darse una definición equivalente para tiempo continuo pero no lo haremos aquí. La definición del proceso es

$$X_0 = 1, \quad X_{n+1} = \sum_{i=1}^{X_n} \xi_i^{(n+1)},$$

donde  $(\xi_j^n, j \geq 1, n \geq 1)$  con variables aleatorias independientes e idénticamente distribuidas.

El comportamiento asintótico de este proceso depende de  $E(\xi_1^1)$ . Si  $E(\xi_1^1) \leq 1$  el proceso se extingue con probabilidad uno (salvo que  $\mathbb{P}(\xi_1^1 = 1) = 1$ ). En cambio, si  $E(\xi_1^1) > 1$ , con probabilidad positiva el proceso sobrevive para todo  $n \geq 1$ .

Para el caso  $\mathbb{E}(\xi_1^1)$ , en 1947 Yaglom ([19]) probó la existencia del límite de la evolución condicionada. A partir de ahí se le dio el nombre de *límite de Yaglom* a este tipo de límites.

## 6.2. Procesos de Nacimiento y Muerte

Los procesos de nacimiento y muerte son una familia bastante amplia de proceso con espacio de estados  $\mathbb{N}$ . Sirven para modelar diversos fenómenos comenzado por modelos de nacimiento y muerte como indica su nombre pero también para paseos al azar en  $\mathbb{N}$ , colas, modelos epidemiológicos, etc.

Las tasas de este proceso están dadas por

$$q(x, x+1) = p_x, \quad q(x, x-1) = q_x, \quad q(x, y) = 0 \text{ si } x \neq x+1, x-1,$$

para  $x \in \mathbb{N}$  y por ultimo  $q(0, y) = 0$  para todo  $y \in \mathbb{N}$ .

Bajo condiciones sobre  $p_x, q_x, x \in \mathbb{N}$  este proceso es absorbido con probabilidad uno. La mas simple es  $p_x = p, q_x = q$  para todo  $x \in \mathbb{N}$  con  $p < q$ .

En el caso en que el proceso es absorbido con probabilidad uno, varios autores estudiaron los límites de Yaglom. Entre ellos [3, 8, 9, 11, 14, 16, 17, 18]. En estos trabajos los autores muestran que, dependiendo de los parámetros puede haber una única distribución cuasi-estacionaria, infinitas (parametrizadas por un parámetro) o ninguna. Además encuentran el dominio de atracción de cada una de las distribuciones cuasi-estacionarias.

## 6.3. La condición de Ferrari-Kesten-Martínez-Picco

Como mencionamos anteriormente, la existencia de las distribuciones cuasi-estacionarias en espacios numerables no siempre es un hecho. Un trabajo fundamental en este sentido es el de Ferrari, Kesten, Martínez y Picco, [5] donde encuentran una condición que garantiza la existencia una distribución cuasi-estacionaria. Bajo ciertas condiciones esta condición es además necesaria, con lo que el resultado es casi óptimo. El teorema de Ferrari y compañía es el que mencionamos en la Sección 4.3.

**Agradecimientos.** A Pablo Ferrari por introducirme en el tema. Por su disposición y generosidad interminables. A Cristian Coletti por charlas productivas que ayudaron a escribir este apunte.

## A. Exponenciales de Matrices

En este apéndice haremos un rápido repaso por la teoría de exponenciación de matrices, que puede verse con detalle en [13].

Dada una matriz  $P$ , se define su exponencial como

$$e^P = \sum_{k=0}^{\infty} \frac{P^k}{k!}.$$

Sin mayor dificultad, se prueban las siguientes propiedades.

**Proposition A.1.** Sean  $P, S, T, N \in \mathbb{R}^{n \times n}$ , entonces

1. La serie que define  $e^P$  es absolutamente convergente.
2. Si  $Q = PTP^{-1}$  entonces  $e^P = Pe^T P^{-1}$ .
3. Si  $T = \text{diag}[\lambda_1, \dots, \lambda_n]$  entonces  $e^T = \text{diag}[e^{\lambda_1}, \dots, e^{\lambda_n}]$ .
4.  $\|e^P\| \leq e^{\|P\|}$ .
5. Si  $ST = TS$  entonces  $e^{S+T} = e^S e^T$ .
6.  $e^{-S} = (e^S)^{-1}$ .
7. Si existe  $k$  tal que  $N^k = 0$  (nilpotente) entonces  $e^N = I + N + \frac{1}{2}N^2 + \dots + \frac{1}{k!}N^k$ .

El siguiente teorema es una herramienta fundamental para poder calcular exponenciales de matrices.

**Teorema A.1.** Para toda matriz  $A \in \mathbb{R}^{n \times n}$  existen matrices únicas  $S$  y  $N$  tales que

1.  $A = S + N$ ,
2.  $SN = NS$ ,
3.  $S$  es semisimple (diagonalizable en  $\mathbb{C}$ ).
4.  $N$  es nilpotente (existe  $k \in \mathbb{N}$  tal que  $N^k = 0$ ).
5. Los autovectores de  $A$  pertenecen al núcleo de  $N$  (y por lo tanto son autovectores de  $S$  con mismo autovalor)

**Corolario A.1.** En las condiciones del teorema anterior, existe un polinomio  $p$  de grado menor o igual que  $n$  tal que

$$e^A = e^S p(N).$$

## B. Teorema de Perron-Frobenius

A lo largo de este apéndice  $P = (P_{ij}, 1 \leq i, j \leq n)$  es una matriz cuadrada con entradas reales no negativas. Es decir  $P_{ij} \geq 0$  para todo  $i, j$ . Escribimos  $P \geq 0$  y decimos que  $P$  es no-negativa. Si tenemos  $P_{ij} > 0$  para todo  $i, j$  escribimos  $P > 0$  decimos que  $P$  es una matriz positiva.

**Teorema B.1.** (Perron-Frobenius) *Sea  $P$  una matriz positiva, entonces existe un autovalor  $r$  de  $P$  tal que*

- (a)  $r$  es real y positivo.
- (b) Se le pueden asociar a  $r$  un autovector a derecha  $w$  y uno a izquierda  $v$  con entradas estrictamente positivas. Estos autovectores son únicos salvo constante multiplicativa.
- (c)  $r > |\lambda|$  para todo autovalor  $\lambda \neq r$ .
- (d) Si  $0 \leq B \leq P$  (componente a componente) y  $\beta$  es un autovalor de  $B$  entonces  $|\beta| \leq r$ . Más aún,  $|\beta| = r$  implica  $B = P$
- (e)  $r$  es una raíz simple del polinomio característico de  $P$ .

*Demostración.* Ver [15]. ■

## C. Desigualdad de Gronwall

El siguiente lema es una herramienta muy importante para acotar soluciones de ecuaciones diferenciales e integrales. Se usa entre otras cosas para probar la unicidad de las soluciones de ecuaciones diferenciales ordinarias. La idea fundamental detrás del lema es que soluciones de ecuaciones parecidas, son parecidas. Una prueba de este lema se puede hallar en cualquier libro de ecuaciones diferenciales ordinarias, por ejemplo [13].

**Lema C.1** (Gronwall). *Sea  $u: [0, T] \rightarrow \mathbb{R}$  continua y no negativa y supongamos que existen constantes  $C, K \geq 0$  tales que para todo  $0 \leq t \leq T$*

$$u(t) \leq C + \int_0^t K u(s) ds,$$

*entonces para todo  $0 \leq t \leq T$*

$$u(t) \leq Ce^{Kt}.$$

## Referencias

- [1] Amine Asselah, Pablo A. Ferrari, and Pablo Groisman. Quasi-stationary distributions and Fleming-Viot processes on finite spaces. *arXiv:0904.3039v2*.
- [2] Krzysztof Burdzy, Robert Holyst, and Peter March. A Fleming-Viot particle representation of the Dirichlet Laplacian. *Comm. Math. Phys.*, 214(3):679–703, 2000.
- [3] James A. Cavender. Quasi-stationary distributions of birth-and-death processes. *Adv. Appl. Probab.*, 10(3):570–586, 1978.
- [4] J.N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probability*, 4:192–196, 1967.
- [5] P. A. Ferrari, H. Kesten, S. Martinez, and P. Picco. Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Probab.*, 23(2):501–521, 1995.
- [6] Pablo A. Ferrari and Antonio Galves. *Construction of Stochastic Processes, Coupling and Regeneration*. <http://www.ime.usp.br/~pablo/book/abstract.html>.
- [7] Pablo A. Ferrari and Nevena Marić. Quasi stationary distributions and Fleming-Viot processes in countable spaces. *Electron. J. Probab.*, 12:no. 24, 684–702 (electronic), 2007.
- [8] Pablo A. Ferrari, Servet Martínez, and Pierre Picco. Existence of nontrivial quasi-stationary distributions in the birth-death chain. *Adv. in Appl. Probab.*, 24(4):795–813, 1992.
- [9] Pablo A. Ferrari, Servet Martínez, and Pierre Picco. Existence of nontrivial quasi-stationary distributions in the birth-death chain. *Adv. in Appl. Probab.*, 24(4):795–813, 1992.
- [10] Wendell H. Fleming and Michel Viot. Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.*, 28(5):817–843, 1979.
- [11] Phillip Good. The limiting behavior of transient birth and death processes conditioned on survival. *J. Austral. Math. Soc.*, 8:716–722, 1968.
- [12] Olle Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.
- [13] Morris W. Hirsch, Stephen Smale, and Robert L. Devaney. *Differential equations, dynamical systems, and an introduction to chaos*, volume 60 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2004.

- [14] Masaaki Kijima and E. Seneta. Some results for quasi-stationary distributions of birth-death processes. *J. Appl. Probab.*, 28(3):503–511, 1991.
- [15] E. Seneta. *Nonnegative matrices and Markov chains*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.
- [16] Erik A. van Doorn. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. in Appl. Probab.*, 23(4):683–700, 1991.
- [17] Erik A. van Doorn and Pauline Schrijner. Geometric ergodicity and quasi-stationarity in discrete-time birth-death processes. *J. Austral. Math. Soc. Ser. B*, 37(2):121–144, 1995.
- [18] Erik A. van Doorn and Pauline Schrijner. Ratio limits and limiting conditional distributions for discrete-time birth-death processes. *J. Math. Anal. Appl.*, 190(1):263–284, 1995.
- [19] A. M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)*, 56:795–798, 1947.