Notas de procesos estocásticos

Ramsés H. Mena

IIMAS-UNAM

Versión 2025

TEMA 1

INTRODUCCIÓN Y CONCEPTOS BÁSICOS

El estudio de fenómenos aleatorios que evolucionan en el *tiempo* es uno de los temas de más importancia en las Teorías de Probabilidad y Estadística, llamando la atención de investigadores y científicos aplicados. La literatura en el tema es amplia con excelentes exposiciones como las monografías por Gikhman and Skorokhod (1969); Ethier and Kurtz (1986); Williams (1991); Norris (1998).

Estas notas son una interpretación concisa del autor de varias de estas fuentes bibliográficas que, durante la duración del curso de Procesos de Markov, deberán ser usadas meramente como una guía y en conjunto con las exposiciones dadas en clase y las bibliografía correspondiente a cada tema. Todos los comentarios, críticas, correcciónes y/o modificaciones serán consideradas y siempre bien recibidas.

... motivación en clase ...

1.1. Marco general y definiciones

Desde cierta perspectiva, el principal objetivo de Estadística y Probabilidad se centra en la construcción y estudio de modelos para sucesiones de variables aleatorias $X_1, X_2, \ldots$ que describen cierto fenómeno de interés. Un enfoque general se formaliza mediante el concepto de ***procesos estocásticos***[1]

Definición 1 (Clásica). Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, $\mathcal{T}$ un conjunto índice y $(\mathbb{X}, \mathcal{X})$ un espacio medible. Un ***stochastic process*** es una colección de variables aleatorias $(\mathbb{X}, \mathcal{X})$-valuadas e indexadas en $\mathcal{T}$. Lo denotaremos por

$$\mathbf{X} = \{X(t, \omega); t \in \mathcal{T}, \omega \in \Omega\}$$

El término de procesos estocásticos se usa comúnmente cuando $\mathcal{T} \subseteq \mathbb{R}$, e.g. $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{R}$ or $\mathbb{R}^+$. Este será el caso que más trataremos en estas notas. A $(\mathbb{X}, \mathcal{X})$ se le denomina el ***espacio de estados*** del proceso $\mathbf{X}$, el cual asumiremos que es un espacio Polaco.

Nótese que de esta definición, se tiene que para cada $t \in \mathcal{T}$,

$$X(t, \omega) : \Omega \to \mathbb{X} \quad \text{and} \quad X^{-1}(t, \omega) : \mathcal{X} \to \mathcal{F},$$

en palabras para t fijo, $X(t, \omega)$ es una variable aleatoria $\mathcal{X}$-medible. También usaremos las notaciones alternativas para $X(t, \omega)$, i.e. $X_t(\omega)$ ó X_t, cuando no haya ambigüedad sobre el espacio de probabilidad subyacente. Asimismo, $X_{\mathcal{S}}$, para $\mathcal{S} \in \mathcal{T}$, se referirá a la sub-colección de variables aleatorias correspondiente..

Dependiendo de la naturaleza de $\mathcal{T}$, el proceso estocástico $\mathbf{X}$ recibe diferentes nombres en la literatura:

Variable aleatoria.- Una única variable aleatoria se puede ver como un proceso estocástico trivial con, *e.g.* $\mathcal{T} = \{1\}$.

[1] Como veremos a lo largo del curso existen diferentes definiciones hasta cierto punto equivalentes.

Vector aleatorio.- Sea $\mathcal{T} = \{1, 2, \ldots, k\}$, entonces $\mathbf{X}$ es un vector aleatorio con valores en $\mathbb{X}^k$.

Proceso aleatorio a tiempo continuo.- Con $\mathcal{T} = \mathbb{R}$ ó $\mathcal{T} = \mathbb{R}^+$.

Campo aleatorio- Cuando $\mathcal{T}$ es un subconjunto de un espacio multidimensional, *e.g.* $\mathbb{R}^d$.

Existe una visión alternativa para el proceso estocástico $\mathbf{X}$, para la cual necesitamos considerar el siguiente concepto.

Definición 2 (**Trayectorias, Espacio trayectorial** $\mathbb{X}^{\mathcal{T}}$). Sea $\mathbf{X}$ como en la Definición 1, y asumamos la siguiente notación

$$\mathbb{X}^{\mathcal{T}} := \{f : \mathcal{T} \to \mathbb{X} \mid f \text{ is a function } \}$$

Para $\omega \in \Omega$, al mapeo (elemento de $\mathbb{X}^{\mathcal{T}}$)

$$t \to X(t, \omega) \quad \text{también denotado} \quad X(t, \omega) : \mathcal{T} \to \mathbb{X}$$

se le denomina ***trayectoria***, ***camino*** ó ***realización*** del proceso $\mathbf{X}$. Si $\mathcal{T}$ es un conjunto finito, digamos $\mathcal{T} = \{1, 2, \ldots, n\}$ entonces una función f se puede identificar con el vector $(x_1, \ldots, x_n)$, formalmente el n-th $(x_1, \ldots, x_n)$ vector, en un sentido de teoría de conjuntos, esta definido como la función $f : \{1, \ldots, n\} \to \mathbb{X}$ dada por $f(j) = x_j$

Esto provee de la perspectiva alternativa para $\mathbf{X}$ dada por

$$\mathbf{X} : \Omega \to \mathbb{X}^{\mathcal{T}}, \quad \mathbf{X}^{-1} : \mathcal{X}^{\mathcal{T}} \to \mathcal{F}, \quad \omega \mapsto \mathbf{X}(\omega)$$

es decir, como una variable aleatoria $(\mathbb{X}^{\mathcal{T}}, \mathcal{X}^{\mathcal{T}})$-valuada.

Esto claramente induce una relación entree un proceso estocástico con valores en un espacio Polaco, $\mathbb{X}$, y medidas de probabilidad sobre $(\mathbb{X}^{\mathcal{T}}, \mathcal{X}^{\mathcal{T}})$.

En clase haremos preciso el significado de la σ-álgebra $\mathcal{X}^{\mathcal{T}}$, y cómo los conjuntos cilindros forman un álgebra que lo genera…

Definición 3 (**Ley del proceso**). La ley del proceso estocástico en la Definición 1 es la medida de probabilidad

$$\mu := \mathsf{P} \circ \mathbf{X}^{-1} \quad \text{on} \quad (\mathbb{X}^{\mathcal{T}}, \mathcal{X}^{\mathcal{T}}) \tag{1.1}$$

En otras palabras es la ley de la variable aleatoria $\mathbf{X}$ $(\mathbb{X}^{\mathcal{T}}, \mathcal{X}^{\mathcal{T}})$-medible.

Para un conjunto no vacío $\mathcal{S} \subset \mathcal{T}$, definamos $\pi_{\mathcal{S}}\mathbf{X} : \Omega \to \mathbb{X}^{\mathcal{S}}$ vía

$$(\pi_{\mathcal{S}}\mathbf{X})(\omega) := \pi_{\mathcal{S}}(\mathbf{X}(\omega)) = \mathbf{X}|_{\mathcal{S}} = \{X_t : t \in \mathcal{S}\}$$

y

$$\mu_{\mathcal{S}} := \mathsf{P} \circ (\pi_{\mathcal{S}}\mathbf{X})^{-1}$$

En otros términos

$$\mu_{\mathcal{S}}(B) = \mathsf{P}_{\mathbf{X}_{\mathcal{S}}}(B) = \mathsf{P}(\mathbf{X}_{\mathcal{S}} \in B) = \mathsf{P}(\mathbf{X} \in \pi_{\mathcal{S}}^{-1}(B)), \quad \forall B \in \mathcal{X}^{\mathcal{S}}$$

Definición 4 (**Distribuciones finito dimensionales**). Denotemos por $\text{Fidis}(\mathcal{T})$, el conjunto no vacío de sub-conjuntos de $\mathcal{T}$. Las medidas de probabilidad

$$\{\mu_{\mathcal{S}} : \mathcal{S} \in \text{Fidis}(\mathcal{T})\}$$

se denominan las distribuciones finito dimensionales de $\mathbf{X}$.

Definición 5 (**Compatibilidad/Consistencia de distribuciones finito dimensionales**). Si $\mathcal{U}, \mathcal{V} \in \text{Fidis}(\mathcal{T})$ y $\mathcal{U} \subseteq \mathcal{V}$, y si además $\pi^{\mathcal{V}}_{\mathcal{U}}$ denota el mapeo que restringe $\mathbb{X}^{\mathcal{V}}$ a $\mathbb{X}^{\mathcal{U}}$, entonces tenemos la ***condición de compatibilidad*** o ***propiedad proyectiva***

$$\mu_{\mathcal{U}} = \mu_{\mathcal{V}} \circ (\pi^{\mathcal{V}}_{\mathcal{U}})^{-1} \tag{1.2}$$

Propiedad 1.2 es equivalente (tarea) a

$$\mu_{X_{t_1},\ldots,X_{t_n}}(A_1,\ldots,A_n) = \mu_{X_{\varrho_1},\ldots,X_{\varrho_n}}(A_{\varrho 1},\ldots,A_{\varrho n}) \tag{1.3}$$
$$\mu_{X_{t_1},\ldots,X_{t_{n-1}},X_{t_n}}(A_1,\ldots,A_{n-1},\mathbb{X}) = \mu_{X_{t_1},\ldots,X_{t_{n-1}}}(A_1,\ldots,A_{n-1}) \tag{1.4}$$

para toda $n \geq 1$, $t_i \in \mathcal{T}$, permutación ϱ de $\{t_i\}_{i=1}^n$, y $A_i \in \mathcal{X}$.

Teorema 1 (Daniell – Kolmogorov). *Sea $\mathbb{X}$ un espacio Polaco, dotado con su σ-álgebra de Borel $\mathcal{X} = \mathscr{B}(\mathbb{X})$, y $\mathcal{T}$ un conjunto índice. Supóngase que para cada $\mathcal{S} \in$ Fidis$(\mathcal{T})$, existe una medida de probabilidad compatible/consistente $\mu_{\mathcal{S}}$ sobre $(\mathbb{X}^{\mathcal{S}}, \mathcal{X}^{\mathcal{S}})$. Entonces existe una única medida μ sobre $(\mathbb{X}^{\mathcal{T}}, \mathcal{X}^{\mathcal{T}})$ tal que*

$$\mu_{\mathcal{S}} = \mu \circ \pi_{\mathcal{S}}^{-1} \quad on \quad (\mathbb{X}^{\mathcal{S}}, \mathcal{X}^{\mathcal{S}})$$

donde $\pi_{\mathcal{S}}$ es el mapeo que restrige $\mathbb{X}^{\mathcal{T}}$ a $\mathbb{X}^{\mathcal{S}}$.

Demostración: Ver la monografía por González-Barrios Murguía (2011).

El Teorema de consistencia de Daniell – Kolmogorov (DK) proporciona un proceso canónico que tiene todas las posibles funciones en $\mathbb{X}^{\mathcal{T}}$ como posibles. Sin embargo, existen complicaciones cuando uno intenta desmarañar propiedades de las trayectorias como continuidad.

Discusión en clase sobre el caso sobre la μ inducida por $\mu_{\mathcal{S}}$ gaussiana con $m(t) = 0$ y $\sigma(s,t) = \min(s,t)$, e.g. si $\mathscr{C} := \mathbb{X}^{\mathcal{T}}\big|_{func.cts.}$, entonces $\mathscr{C}$ tiene medida interior μ, 0

Existen, al menos, dos aparentemente diferentes enfoques para el estudio de procesos estocásticos: “el enfoque distribucional” y el “el enfoque trayectorial”. Mucho de la teoría moderna se enfoca en este último, lo que conlleva a definiciones “indirectas” de procesos estocásticos, e.g. el uso de ecuaciones diferenciales estocásticas (EDE) para definir procesos de difusión[2]. Sin embargo, cada uno de estos dos enfoques tienen sus ventajas y desventajas, por ejemplo el definir un proceso solamente mediante sus propiedades distribucionales (ignorando sus propiedades trayectoriales, e.g. continuidad) tiene algunas ventajas, *e.g* el conjunto $\mathcal{T}$ puede ser arbitrario, la medibilidad de sus trayectorias no son parte de la definición, siempre se cuenta con la “ley” del proceso, lo cual puede ser útil para estimación y simulación, etc. Por otro lado, propiedades como monotonicidad de las trayectorias o distribuciones de eventos como $\sup_t X_t$ no siempre se pueden estudiar directamente con un enfoque solamente distribucional. Ver Ejemplo 1 abajo.

Dicho esto, es importante enfatizar que una de las principales interrogantes de la teoría de procesos estocásticos es investigar como la medida μ, construida vía el teorema de DK y definida en el σ-álgebra generado por los conjuntos cilindro, afecta las propiedades trayectoriales.

Example 1. *Sea $\Omega = [0,1]$ dotado con su σ-álgebra de Borel y la ley Uniforme sobre $[0,1]$ como $\mathbb{P}$. Dado $\omega \in \Omega$, definamos los procesos estocásticos*

$$Y_t(\omega) = 0, \ \forall t, \omega \qquad X_t(\omega) = 1(t = \omega)$$

Sea $A_t = \{\omega : X_t \neq Y_t\} = t$. Como tenemos que $\mathsf{P}(A_t) = 0$, entonces $\mathsf{P}(X_t = Y_t) = 1$ para cada t. Además, sea $A^N = \cup_{i=1}^N A_{t_i}$, entonces para cualquier elección de N y $t_i, \ldots, t_N$ tenemos $\mathsf{P}(A^N) = 0$. Entonces las distribuciones finito-dimensionales de $\{X_t\}$ son las mismas para $\{Y_t\}$, más aun en $N = \infty$ vemos que $\mathsf{P}(A^\infty) = 0$, lo que es equivalente a decir que $X_t(\omega)$ y $Y_t(\omega)$ concuerdan, casi seguramente, para cualquier colección fija de índices numerable. Por otro lado, algunas propiedades trayectoriales no concuerdan, e.g.

$$\begin{aligned} \mathsf{P}(\{\omega : (\sup\{X_t(\omega) : 0 \leq t \leq 1\}) \neq 0\}) &= 1 \\ \mathsf{P}(\{\omega : (\sup\{Y_t(\omega) : 0 \leq t \leq 1\}) \neq 0\}) &= 0 \end{aligned}$$

[2]Más adelante en el curso quedará más clara dicha relación

o

$$\begin{aligned} \mathsf{P}(\{\omega : t \mapsto X_t(\omega) \textit{ is continuous }\}) &= 0 \\ \mathsf{P}(\{\omega : t \mapsto Y_t(\omega) \textit{ is continuous }\}) &= 1. \end{aligned}$$

Ambos procesos tienen diferentes propiedades trayectoriales, aunque bajo la construcción de DK, son los mismos. Este ejemplo motiva los siguientes conceptos.

Definición 6 (Versión). Dos procesos estocásticos, $\mathbf{X}$ y $\mathbf{Y}$, se dicen uno ***versión*** del otro si comparten las mismas distribuciones finito dimensionales.

Definición 7 (Modificación). Dos procesos estocásticos, $\mathbf{X} = \{X_t\}_{t\in\mathcal{T}}$ y $\mathbf{Y} = \{Y_t\}_{t\in\mathcal{T}}$, se dicen uno ***modificación*** del otro, si comparten el mismo espacio de estados, $(\mathbb{X}, \mathcal{X})$, están definidos en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y satisfacen

$$\mathsf{P}(X_t = Y_t) = 1 \quad \text{para cada} \quad t \in \mathcal{T}. \tag{1.5}$$

Definición 8 (Indistinguible). Sean $\mathbf{X}$ y $\mathbf{Y}$ dos procesos como en la Definición 7. Decimos que son ***indistinguibles*** si

$$\mathsf{P}(X_t = Y_t, \text{ para todo} t \in \mathcal{T}) = 1. \tag{1.6}$$

Ejercicio 1. *Demuestra que sí $\mathbf{Y}$ es modificación de $\mathbf{X}$, entonces $\mathbf{Y}$ es también versión de $\mathbf{X}$.*

Nótese que, a diferencia de versiones, modificaciones están definidas en el mismo espacio de probabilidad.

1.1.1. Criterios de continuidad

Análogamente a la teoría de funciones, la continuidad de una función aleatoria es una propiedad típicamente relacionada con la convergencia de sucesiones. Entonces, diferentes tipos de convergencia para variables aleatorias inducen diferentes tipos de continuidad.

Definición 9 (Trayectorias continuas). Un proceso estocástico $\mathbf{X} = \{X_t(\omega); t \in \mathcal{T}\}$ tiene ***trayectoria continuas con probabilidad uno*** si

$$\mathsf{P}(\omega : t \mapsto X_t(\omega) \text{ es contunua }) = 1.$$

Esta condición se puede re-frasear en términos de convergencia en probabilidad. Sea $\mathcal{S} \subset \mathcal{T}$, si para toda sucesión $\{t_n\}$, $||t_n - t|| \to 0$ cuando $n \to \infty$, entonces

$$\mathsf{P}(\omega : |X_{t_n}(\omega) - X_t(\omega)| \to 0 \text{ as } n \to \infty) = 1 \qquad \text{for all } t \in \mathcal{S}$$

Un refinamiento de esta noción de continuidad, que especifica la suavidad máxima posible de las trayectorias, es la siguiente:

Definición 10. Un proceso estocástico X es ***localmente Hölder continuo con exponente*** γ si para algún $c < \infty$ y variable aleatoria $h(\omega) > 0$,

$$\mathsf{P}(\{\omega : \sup_{0\le s,t,\le T, |t-s|\le h(\omega)} \frac{|X_t(\omega) - X_s(\omega)|}{|t-s|^\gamma} \le c\}) = 1$$

Cuando $\gamma = 1$ hablamos de funciones ***Lipschitz continuas***.

Teorema 2 (Criterio de continuidad de Kolmogorov's). *Dado un proceso estocástico $\mathbf{X} = \{X_t, t \in [0, T]\}$, supongamos que existe $\alpha, \beta, c, h > 0$ tal que*

$$\mathsf{E}(|X_{t+h} - X_t|^\alpha) \le ch^{1+\beta}, \quad \textit{para toda} \quad 0 \le t \le t+h \le T. \tag{1.7}$$

Entonces, existe una modificación continua $\mathbf{Y}$ de $\mathbf{X}$ tal que $\mathbf{Y}$ también es localmente Hölder continua con exponente γ para cualquier $0 < \gamma < \beta/\alpha$.

Definición 11. Decimos que un proceso estocástico $\mathsf{X} = \{X_t(\omega)\}_{t\geq 0}$ tiene trayectorias ***continuas por la derecha con límites por la izquierda*** (CDLI), si $\omega \in \Omega$ casi en todos lados $\omega \in \Omega$, la trayectoria $t \mapsto X_t(\omega)$ es continuas por la derecha con límites por la izquierda en cualquier $t \geq 0$. (Es decir, para $h \downarrow 0$ ambos $X_{t+h}(\omega) - X_t(\omega)$ y el límite $X_{t-h}(\omega)$ existen)

A una modificación con trayectorias CDLI con probabilidad uno se le denomina una ***modificación CDLI del proceso estocástico*** $\mathbf{X}$. Al término CDLI, se le identifica también por Càdlàg, RCLL, Regular, dependiendo del idioma y/ó la fuente.

Proposición 1. *Las siguientes afirmaciones se cumplen:*

i) *Sean $\mathbf{X} = \{X_n : n \in \mathbb{N}_0\}$ y $\mathbf{Y} = \{Y_n : n \in \mathbb{N}_0\}$ dos procesos estocásticos* **a tiempo discreto** *que son modificaciones uno del otro, entonces también son indistinguibles.*

ii) *Sean $\mathbf{X} = \{X_t : t \geq 0\}$ y $\mathbf{Y} = \{Y_t : t \geq 0\}$ dos procesos estocásticos* **a tiempo contínuo***, ambos con trayectorias continuas por la derecha y valores en un espacio Polaco $(\mathbb{X}, \mathcal{X})$. Son modificaciones uno del otro si y solo si son indistinguibles.*

Proposición 2. *Las siguiente relaciones se cumplen para las trayectorias de cualesquiera procesos estocásticos.*

Hölder continuous $\Rightarrow$ Continuidad c.p.1 $\Rightarrow$ CDLI.

Dada la complicación y generalidad en la que se puede entender un proceso estocástico, y las implicaciones descritas en las proposiciones 1 y 2, parte del estudio de estos objetos estocásticos trabaja con modificaciones.

Además de las clasificaciones en términos de su conjunto de índices $\mathcal{T}$ mencionada antes, los procesos estocásticos pueden ser clasificados en términos de sus supuestos de dependencia/independencia subyacentes a $\mathbf{X}$.

Desde una perspectiva general, el estudio y aplicación de procesos estocásticos requiere del análisis y comprensión de, al menos, los siguientes aspectos:

i) El estudio de la ley de probabilidades que los regula así como la construcción de la misma.

ii) Su construcción y caracterización canónica vía versiones regulares.

iii) El estudio de sus propiedades de estabilidad así como la caracterización y/o construcción de procesos con estas características.

i) Nos permite conocer a los procesos desde un punto de vista probabilístico y por lo tanto poder estudiar las consecuencias de asumir cierto proceso como modelo para un fenómeno dado. También permite proceder con la calibración y estimación del mismo. ii) Nos permite entender las propiedades "trayectoriales" del proceso, e.g. trayectorias continuas, CDLI, etc. Con esto se puede recurrir a construcciones alternativas, e.g. vía ecuaciones estocásticas, y por lo tanto se amplia su espectro de aplicaciones y potenciales métodos de estudio y aplicación. iii) La mayoría de sus características distribucionales, métodos de estimación, de predicción y estudio se facilitan en la presencia de ciertas propiedades de estabilidad del proceso. Así pues el estudio de estas propiedades así como las condiciones bajo las cuales un proceso de Markov dado, satisface cierta estabilidad son de interés.

En relación a i) sabemos que la ley de una colección de variables aleatorias se simplifica considerablemente bajo ciertos supuestos de simetría. Veamos las siguientes

1.1.2. Procesos independientes

Muy probablemente la más usada de las simetrias distribucionales es la resultante de considerar un ***proceso estocástico independiente***, i.e. el caracterizado for distribuciones finito dimensionales del tipo

$$\mu_{t_0,\ldots,t_n}(A_0, \ldots, A_n) = \prod_{i=0}^{n} \mu_{t_i}(A_i) \tag{1.8}$$

En este contexto se desarrolla ***la teoría clásica de estadística***, en particular cuando μ toma una forma específica caracterizada por un parámetro. $\theta \in \Theta$, e.g.

$$\mathscr{F}_\theta := \{\mu_\theta(A); \forall A \in \mathcal{X}, \theta \in \Theta\}$$

1.1.3. Intercambiabilidad

Definición 12. Una colección de variables aleatorias $\{X_i\}_{i=1}^n$ se dice que es finitamente intercambiable si $\{X_i\}_{i=1}^n \stackrel{\mathrm{d}}{=} \{X_{\tau(i)}\}_{i=1}^n$ para cualquier permutación τ de $\{1,\ldots,n\}$. Una colección infinita $\{X_i\}_{i=1}^\infty$ se dice que es intercambiable si toda sub-colección es intercambiable.

En otras palabras, la propiedad de intercambiabilidad equivale a invarianza bajo permutaciones de los indices. Este tipo de simetría juega un papel fundamental en estadística, particularmente en estadística bayesiana.

1.1.4. Procesos estrictamente estacionarios

Una simetría distribucional que asegura cierta estructura de estabilidad es la siguiente.

Definición 13. Un proceso estocástico X se dice que es ***estrictamente estacionario*** si para $h, n, t_1, \ldots, t_n$ con $t_i, t_i+h \in \mathsf{T}$ sus distribuciones finito dimensionales son invariantes bajo traslaciones del tiempo. Esto es

$$\{X_{t_1}, \ldots, X_{t_n}\} \stackrel{\mathrm{d}}{=} \{X_{t_1+h}, \ldots, X_{t_n+h}\} \tag{1.9}$$

1.1.5. Proceso reversible en el tiempo

Otra propiedad de mucha utilidad y frecuentemente encontrada en el estudio de procesos markovianos estacionarios se sigue de la siguiente definición.

Definición 14. Un proceso estocástico $\mathbf{X}$ se dice que es ***reversible en el tiempo*** si

$$\{X_{t_1}, X_{t_2}, \ldots X_{t_n}\} \stackrel{\mathrm{d}}{=} \{X_{t_n}, X_{t_{n-1}}, \ldots X_{t_1}\} \tag{1.10}$$

se satisface para cualesquier $t_1, \ldots, t_n \in \mathsf{T}$ y $n = 1, 2, \ldots$

1.1.6. Procesos gaussianos

Definition 1. Una variable aleatoria G con valores en $\mathbb{R}^d$ se dice que es una ***variable aleatoria gaussiana*** (ó ***vector aleatorio gaussiano***) si existe un vector $\mu \in \mathbb{R}^d$ y una matriz $\Sigma \in \mathbb{R}^d \times \mathbb{R}^d$ tal que para todo $s \in \mathbb{R}^d$, $G \cdot s = \sum_{j=1}^d s^{(j)} G^j$ es una variable aleatoria gaussiana con media $\mu \cdot s$ y varianza $s'\Sigma s$. De aqui en adelante s denota un vector columna y s' su vector transpuesto. Al vector μ se le conoce como el ***vector media***, y a Σ la ***matrix de covarianza*** de G.

Con la notación $G \sim \mathsf{N}_d(\mu, \Sigma)$ indicaremos que G es un vector aleatorio gaussiano con valores en $\mathbb{R}^d$, media μ y matrix de covarianza Σ. Si Σ es no-singular ($\det\Sigma > 0$) la distribución de G es absolutamente continua con respecto a la medida de Lebesgue y su densidad en $x \in \mathbb{R}^d$ esta dada por

$$\frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'\right\}. \tag{1.11}$$

Equivalentemente, $G \sim \mathsf{N}_d(\mu, \sigma)$ si para toda $t \in \mathbb{R}^d$ y $\xi \in \mathbb{R}$

$$\mathsf{E}\left[\exp(i\xi(G \cdot t))\right] = \exp\left(i\xi \sum_{j=1}^d t^{(j)}\mu^{(j)} - \frac{\xi^2}{2}\sum_{j=1}^d\sum_{k=1}^d t^{(j)}t^{(k)}\sigma(i,j)\right), \tag{1.12}$$

donde $\sigma(i,j)$ denota el ij-ésimo elemento de Σ.

Teorema 3. *Dado un vector $\mu \in \mathbb{R}^d$ y una matrix $\Sigma \in \mathbb{R}^d \times \mathbb{R}^d$ simétrica positiva definida, se puede construir una variable aleatoria con valores en $\mathbb{R}^d$ dada por $G \sim \mathsf{N}_d(\mu, \Sigma)$. En cambio, si $G \sim \mathsf{N}_d(\mu, \Sigma)$ entonces Σ se puede tomar simétrica y es siempre real y positiva definida.*

Demostración. Usando álgebra matricidal elemental, $\Sigma = P'\Delta P$, donde P es una matriz $(d \times d)$ real y ortogonal y Δ es una matriz diagonal $(d \times d)$ de valores característicos con $\Delta^{(i,j)} \geq 0$ para todo $1 \leq i \leq d$. Sea $A = \Delta^{\frac{1}{2}} P$, donde $\Delta^{\frac{1}{2}}$ denota la matriz diagonal $(d \times d)$ cuya coordenada (i,j) es $\sqrt{\Delta^{(i,i)}}$. Nótese que A es una matriz real que satisface $\Sigma = A'A$. Sea Z un vector en $\mathbb{R}^d$ de variables aleatorias gaussianas independientes y definamos $G = \mu + A'Z$. Calculando la función característica de $t \cdot G$ para $t \in \mathbb{R}^d$, podemos ver que $G \sim \mathsf{N}_d(\mu, \Sigma)$ como lo establece el teorema. □

Corolario 1. *Supóngase que* $G \sim \mathsf{N}_d(\mu, \Sigma)$, *donde* $\Sigma^{(i,j)} = 0$ *para* $i \neq j$. *Entonces* $G^{(1)}, \ldots, G^{(d)}$ *son variables aleatorias gaussianas independientes con valores en* $\mathbb{R}$.

Definición 15. Un proceso estocástico X, se dice que es un ***proceso gaussiano*** si para toda elección finita $t_1, \ldots, t_k \in \mathsf{T}$ las *funciones finito dimensionales* son $\mathsf{N}_k(\mu, \Sigma)$. Si T es un subconjunto medible de $\mathbb{R}^n$ entonces a X se le identifica como un ***campo aleatorio gaussiano***.

Decimos que una función $f : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ es ***simétrica*** si para toda $s, t \in \mathcal{T}$, $f(s,t) = f(t,s)$. Es ***positiva definida*** si para toda $s_1, \ldots, s_n \in \mathcal{T}$ y toda $\xi_1, \ldots, \xi_n \in \mathbb{R}$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i f(s_i, s_j) \xi_j \geq 0. \tag{1.13}$$

Cualquier función simétrica, positiva definida f es no-negativa sobre la diagonal, en el sentido de que $f(x,x) \geq 0$ para toda x.

Teorema 4. *Dado un conjunto* $\mathcal{T}$, *una función arbitraria* $\mu : \mathcal{T} \to \mathbb{R}$, *una función simétrica y positiva definida* $\Sigma : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$, *existe un proceso gaussiano* $\mathbf{X}$ *con funciones de media y covarianza* μ *y* Σ *respectivamente.*

Ejercicio 2. *Demuestra el Teorema 4. Tip. Usa el hecho de que la forma cuadrática* $\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'$ *y que* $|\Sigma|$ *son invariantes bajo permutaciones de* x *y* $|\Sigma|$ $t_1, \ldots, t_n$

1.1.7. Procesos con incrementos independientes

Los procesos con incrementos independientes, también conocidos como ***procesos aditivos***, juega un papel fundamental en al teoría moderna de procesos estocásticos y su aplicaciones. Como veremos más adelante muchos de los procesos que estudiaremos caen dentro de esta amplia clase.

Definición 16. Un process estocástico, $\mathsf{X} := \{X_t; t \geq 0\}$, con valores en los reales se conoce como un ***proceso con incrementos independientes*** (PII) si

(i) las trayectorias $t \mapsto X_t$ son RCLL c.s. Es decir $X_{t+} = X_t$ y X_{t-} existen.

(ii) Para toda $n \geq 1$ y $0 \leq t_0 < t_1 < \cdots < t_n$, las variables aleatorias X_{t_0}, $\{X_{t_{j+1}} - X_{t_j}\}_{j=1}^{n}$ son independientes. (***incrementos independientes***)

Si adicionamos

(iii)* $\{X_{t+s} - X_t\} \overset{\mathrm{d}}{=} X_s$ para toda $s, t \geq 0$ (***incrementos estacionarios***)

(iv)* $X_0 = 0$ c.s.

entonces a $\mathbf{X}$ se le llama ***proceso de Lévy***.

La condición (i), es también conocida como la ***condición de trayectorias cádlág*** y se require para que solamente discontinuidades vía saltos puedan ocurrir. Decimos que X tiene un ***salto fijo*** en algún tiempo $t > 0$ si $\mathsf{P}(X_t \neq X_{t-}) > 0$. En el caso de un proceso de Lévy la condición de incrementos estacionarios excluye la posibilidad de saltos fijos. Condición (iv)* es implicita en la condición (iii)* (escoge $s = 0$).

TEMA 2

ELEMENTOS DE PROCESOS DE MARKOV

Los procesos de Markov son indudablemente los modelos estocásticos más importantes y socorridos en el estudio de fenómenos aleatorios que evolucionan en el tiempo. Sus propiedades de dependencia y continuidad así como sus diversas definiciones y construcciones plantean un compromiso adecuado entre generalidad y aplicabilidad.

En palabras, un proceso estocástico tiene la ***propiedad de Markov*** si el futuro y el pasado son condicionalmente independientes dados el presente. En lenguaje matemático tenemos la siguiente definición

Definición 17. Sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad y $\mathcal{F} = \{\mathcal{F}_t : t \in \mathsf{T}\}$ una filtración sobre este. Decimos que un proceso estocástico $\mathsf{X} := \{X_t; t \in \mathsf{T}\}$ definido en este espacio, $\mathcal{F}$ -adaptado, y con valores en un espacio de Borel $(\mathbb{X}, \mathcal{X})$ es un ***proceso de Markov*** con respecto a la filtración $\mathcal{F}$ si el pasado, $\mathcal{F}_t$, y el futuro, $\mathcal{F}^t := \sigma(X_s : s \geq t, s \in \mathsf{T})$, son condicionalmente independientes dados el presente $\sigma(X_t)$. Equivalentemente,

$$\mathbb{P}(A \cap B \mid \sigma(X_t)) = \mathbb{P}(A \mid \sigma(X_t))\mathbb{P}(B \mid \sigma(X_t)) \tag{2.1}$$

para cualquier $A \in \mathcal{F}_t$ y $B \in \mathcal{F}^t$. Si no se hace mención de las σ-algebras entonces se asume que X es un proceso de Markov con respecto a su filtración canónica.

Notemos que hemos asumido que el conjunto de incides, T, es un conjunto completamente ordenado. En particular, aquí asumiremos que $\mathsf{T} \subseteq \bar{\mathbb{R}}$, aunque existen otros casos más generales, e.g. $\mathsf{T} = \mathbb{R}^d$ que se estudia en la teoría de campos aleatorios.

De la Definición 17 se sigue la forma más clásica de la propiedad de Markov que establece que, para toda $s, t \geq 0$ y $A \in \mathcal{X}$

$$\mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_t) = \mathbb{P}(X_{t+s} \in A \mid X_t) \qquad \mathbb{P}\text{-c.s.} \tag{2.2}$$

o de manera equivalente

$$\mathbb{E}(f(X_{t+s}) \mid \mathcal{F}_t) = \mathbb{E}(f(X_{t+s}) \mid X_t) \tag{2.3}$$

para f una función $\mathcal{X}$-medible y acotada.

Al igual que en la teoría general de procesos estocásticos, los procesos de Markov se clasifican y estudian dependiendo de la naturaleza de su espacio de estados $\mathbb{X}$, el cual puede ser finito, numerable o no-numerable; de la naturaleza de T el cual puede ser discreto o continuo; así como de la estructura de sus trayectorias, e.g. continuas, continuas por la derecha con límites por la izquierda. Aquí nos referiremos a una ***Cadena de Markov*** cuando $\mathbb{X}$ tiene un carácter numerable, de no ser el caso lo llamaremos ***proceso de Markov***. Asimismo, si $\mathsf{T} \subseteq \mathbb{Z}$ diremos que se trata de un proceso o cadena de Markov ***a tiempo discreto***, de otra forma, e.g. $\mathsf{T} = [0, \infty)$, diremos que se trata de un proceso ***a tiempo continuo***. Estas clasificaciones de procesos de Markov caracterizan diferentes enfoques y técnicas usadas para su estudio.

2.1. Función de transición y la ecuación de Chapman-Kolmogorov

La igualdad característica (2.2) nos indica que que la ley de un proceso markoviano queda caracterizada por las probabilidades $\mathbb{P}(X_{t+s} \in A \mid \sigma(X_t))$ para toda $A \in \mathcal{X}$ y $t, s \geq 0$. Así pues, es de utilidad estudiar esta clase de funciones, las cuales se formalizan en la siguiente definición.

Definición 18. Sea $(\mathbb{X}, \mathcal{X})$ un espacio de Borel. Una función $P_{t,s}(x, A)$ definida para $0 \leq t < s < \infty$, $x \in \mathbb{X}$, $A \in \mathcal{X}$ se dice que es una ***función de transición markoviana*** sobre $(\mathbb{X}, \mathcal{X})$ si

i) $A \mapsto P_{t,s}(x, A)$ es una *medida de probabilidad* sobre $\mathcal{X}$ para todo $x \in \mathbb{X}$

ii) $x \mapsto P_{t,s}(x, A)$ es una función $\mathcal{X}$-medible para todo $A \in \mathcal{X}$

iii) Para todo $x \in \mathbb{X}, A \in \mathcal{X}$ y $0 \leq t < s < u$

$$P_{t,u}(x, A) = \int_{\mathbb{X}} P_{t,s}(x, dy)\, P_{s,u}(y, A), \tag{2.4}$$

A la ecuación (2.4) se le denomina la ***ecuación de Chapman-Kolmogorov***.

Si existe una función $P_t(x, A)$ tal que $P_{s-t}(x, A) = P_{t,s}(x, A)$ para toda t, s, x y A entonces se dice que la función de transición es ***temporalmente homogénea*** y la ecuación de Chapman-Kolmogorov se simplifica mediante

$$P_{t+s}(x, A) = \int P_t(x, dy)\, P_s(y, A). \tag{2.5}$$

De forma análoga, se dice que una función de transición es ***espacialmente homogénea*** si $P_{t,s}(x, A) = P_{t,s}(0, A - x)$, donde $A - x = \{y - x : y \in A\}$.

En ocasiones, en la Definición 18 se incluye

$$\lim_{t \downarrow 0} P_t(x, \cdot) = \delta_x(\cdot) \tag{2.6}$$

que claramente permite obtener $P_0(\cdot, \cdot)$. A una función de transición con dicha característica se le denomina como ***estándar*** o ***normal***.

Ejemplo 1. *Sea* $\mathbb{X} = \{0\} \cup (-\infty, -1] \cup [1, \infty)$ *y* $P_t(\cdot, \cdot)$ *una función de transición temporalmente homogénea dada por*

$$\begin{aligned}
&P_0(0, \{1\}) = \frac{1}{2} && P_0(0, \{-1\}) = \frac{1}{2}; \\
&P_t(x, \cdot) = \delta_{x+t}(\cdot), && \textit{si} \quad x \geq 1; t \geq 0 \\
&P_t(x, \cdot) = \delta_{x-t}(\cdot), && \textit{si} \quad x \leq -1; t \geq 0 \\
&P_t(0, \cdot) = \frac{1}{2}\{\delta_{1+t}(\cdot) + \delta_{-1-t}(\cdot)\}
\end{aligned}$$

Nótese que no se trata de una función de transición normal pero sigue cumpliendo las propiedades de la Definición (18)*.*

Ejercicio 3. *Demostrar que la función de transición del Ejemplo 1 satisface las condiciones de la Definición 18 y no es una función de transición normal.*

La propiedad (2.6) no es necesaria en general, e.g. cuando x es un punto de ramificación, como en el Ejemplo 1. Habiendo dicho esto, por el momento trabajaremos con funciones de transición estándares.

Dado un proceso de Markov para la filtración $\mathcal{F}$, X, se dice que admite una ***probabilidad de transición*** si

$$\mathbb{P}(X_t \in A \mid \mathcal{F}_s)) = P_{s,t}(X_s, A) \qquad \mathbb{P}\text{-c.s. para} \qquad s < t \tag{2.7}$$

Por lo que, en el caso homogéneo en el tiempo, la ecuación (2.3) se puede escribir como

$$\mathbb{E}(f(X_{t+s}) \mid \mathcal{F}_t) = \int f(x) P_s(X_t, dx) \quad c.s.$$

para f una función medible no-negativa.

De hecho para $s, t \geq 0$ y todo $A \in \mathcal{X}$, podemos verificar que

$$\begin{aligned} P_{t+s}(X_0, A) &= \mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_0) \\ &= \mathbb{E}[\mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_s) \mid \mathcal{F}_0] \\ &= \mathbb{E}[P_t(X_s, A) \mid \mathcal{F}_0] \\ &= \int_{\mathbb{X}} P_t(y, A) P_s(X_0, dy). \end{aligned} \tag{2.8}$$

La teoría de procesos de Markov esta fuertemente sesgada hacia el caso homogéneo en el tiempo. El argumento que justifica esta simplificación es que un proceso de Markov no-homogéneo en el tiempo, digamos $\mathsf{X} = (X_t; t \geq 0)$, puede ser estudiado mediante $\hat{X}_t = (t, X_t)$, el cual resulta en un proceso homogéneo en el tiempo. Es decir haciendo que el tiempo forme parte del espacio de estados. En estas notas adoptaremos esta convección y trabajaremos en su mayor parte con procesos homogéneos en el tiempo.

Ejercicio 4. *Sea* $\hat{\mathsf{X}} = (\hat{X}_t : t \geq 0)$*, con* $\hat{X}_t := (t, X_t)$ *y* $\mathsf{X} = (X_t; t \geq 0)$ *es un proceso de Markov no-homogéneo en el tiempo con espacio de estados* $(\mathbb{X}, \mathcal{X})$ *y función de transición no-homogénea* (2.4)*. Demuestra que* $\hat{\mathsf{X}}$ *es un proceso de Markov homogéneo en el tiempo y con valores en* $(\mathbb{R}_+ \cup \mathbb{X}, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{X})$*.*

Con lo anterior se puede definir un proceso de Markov, e.g. que satisface (2.2), en términos de una función de transición que, en particular, satisface (2.4). En el caso de procesos a tiempo discreto, e.g. cuando $\mathsf{T} = \mathbb{Z}_+$, (2.4) no presenta una restricción ya que de cualquier versión de la función de transición a un paso, i.e. $P(x, B) := P_{n-1,n}(x, B)$, se puede obtener ***probabilidad de transición en*** n***-pasos*** via la ecuación de Chapman-Kolmogorov

$$P^n(x, B) = \int_{\mathbb{X}} P^m(x, dy) P^{n-m}(x, B), \quad 0 \leq m \leq n, B \in \mathcal{X}$$

donde hemos adoptado la notación $P^n(x, A) := \mathsf{P}(X_n \in A \mid X_0 = x)$, cuando T es discreto. Una deducción análoga se sigue para el caso no-homogéneo en el tiempo.

Ejemplo 2. *Consideremos la función de transición con valores en* $(\mathbb{R}, \mathcal{B})$ *dada por*

$$P(x, A) = \int_A \mathsf{N}\left(y \mid \frac{x}{s+1}, \frac{(2+s)s}{(s+1)^2}\right) dy, \quad A \in \mathcal{B}$$

con $s > 0$*. En adelante adoptaremos la convención de denotar por* $\mathsf{D}(\theta)$ *a la distribución parametrizada por* θ *y a* $\mathsf{D}(\cdot \mid \theta)$ *por su función de densidad. Entonces* $\mathsf{N}(x \mid \mu, \sigma^2) \propto \exp\{-(x-\mu)^2/(2\sigma^2)\}$*. Claramente esta transición satisface (i) y (ii) de la Definición 18.*

Con esta notación notemos que, en general, si tenemos $P(x, \cdot) = \mathsf{N}(\cdot \mid xa, b)$ *con* $a \in \mathbb{R}$ *y* $b > 0$*, entonces usando la ecuación de Chapman-Kolmogorov se obtiene*

$$P^2(x, \cdot) = \mathsf{N}(\cdot \mid xa^2, b(a^2 + 1))$$

y por inducción

$$P^n(x, \cdot) = \mathsf{N}\left(\cdot \mid xa^n, b\sum_{i=1}^{n} a^{2(i-1)}\right). \tag{2.9}$$

Entonces, si sustituimos

$$a = \frac{1}{s+1} \quad \textit{and} \quad b = \frac{(2+s)s}{(s+1)^2}$$

y notamos que para $s > 0$, $(s+1)^{-1} < 1$

$$\sum_{i=1}^{n} (s+1)^{-2(i-1)} = \frac{(s+1)^2(1-s^2+2s+1)^{-n}}{s(s+2)}$$

obtenemos

$$P^n(x,\cdot) = \mathsf{N}\left(\cdot \mid \frac{x}{(s+1)^n}, 1 - \frac{1}{(s+1)^{2n}}\right) \tag{2.10}$$

que por construcción satisface la ecuación de Chapman-Kolmogorov. En otras palabras, para el caso a tiempo discreto, dada una función de transición que satisface (i) y (ii) de la Definición 18 siempre se puede construir su versión markoviana que conlleva al cumplimiento de la ecuación de Chapman-Kolmogorov. De hecho bajo esta filosofía, la función de transición (2.9) también es markoviana.

La construcción no es tan directa para el caso de T no-numerable, i.e. el caso a tiempo continuo, como lo ilustra el siguiente ejemplo.

Ejemplo 3. *Consideremos la función de transición con valores en* $(\mathbb{R}, \mathcal{B})$ *dada por*

$$P_t(x, A) = \int_A \mathsf{N}\left(y \mid \frac{x + \phi_t \mu}{1+\phi_t}, \tau[1-(1+\phi_t)^{-2}]\right) dy \tag{2.11}$$

con $\mu \in \mathbb{R}$ *y* $t \mapsto \phi_t$ *una función positiva. Al igual que en el Ejemplo 2 esta transición también satisface (i) y (ii) de la Definición 18. Sin embargo, la ecuación de Chapman-Kolmogorov no se satisface para cualquier función* ϕ_t*. En este caso, resulta relativamente fácil resolver (2.5) para* ϕ_t*.*

Primero veamos que si denotamos por $\mathcal{L}_Z(\lambda) = \mathbb{E}[e^{\lambda Z}]$, $\lambda \in \mathbb{C}$, *la transformada de Laplace generalizada para la variable aleatoria* Z, *y* $\mathsf{X} = (X_t; t \geq 0)$ *un proceso de Markov con funciones de transición homogéneas en el tiempo, entonces la ecuación de Chapman-Kolmogorov (2.5), se puede escribir como*

$$\mathcal{L}_{X_{t+s}|X_0=x}(\lambda) = \mathbb{E}(\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0 = x). \tag{2.12}$$

donde, en este caso, $\mathcal{L}_{X_t|X_0=x}(\lambda)$ *es la transformada correspondiente a (2.11), la cual esta dada por*

$$\mathcal{L}_{X_t|X_0=x}(\lambda) \quad = \quad \exp\left\{\frac{\lambda x}{(1+\phi_t)} + \frac{\lambda\mu\phi_t}{(1+\phi_t)} - \frac{\lambda^2\tau[1-(1+\phi_t)^{-2}]}{2}\right\} \tag{2.13}$$

ya que si $Z \sim \mathsf{N}(\eta, \xi)$ *tenemos que* $\mathcal{L}_Z(\lambda) = \exp\{\lambda\eta - \lambda^2\xi/2\}$*. Con esto se tiene que*

$$\begin{aligned}
\mathbb{E}(\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0 = x) &= \exp\left\{\frac{\lambda\mu\phi_t}{(1+\phi_t)} - \frac{\lambda^2\tau[1-(1+\phi_t)^{-2}]}{2}\right\} \mathcal{L}_{X_s|X_0=x}\left(\frac{\lambda}{1+\phi_t}\right) \\
&= \exp\left\{\lambda\mu\left[1-(1+\phi_t)^{-1}(1+\phi_s)^{-1}\right] + \lambda x_0(1+\phi_t)^{-1}(1+\phi_s)^{-1}\right\} \\
&\times \exp\left\{-\frac{\lambda^2\tau}{2}\left[1-(1+\phi_t)^{-2}(1+\phi_s)^{-2}\right]\right\}
\end{aligned} \tag{2.14}$$

que es igual a $\mathcal{L}_{X_{t+s}|X_0=x}(\lambda)$ *si y solo si*

$$\phi_{t+s} = \phi_t\phi_s + \phi_t + \phi_s$$

o equivalentemente

$$\begin{aligned} \phi_{t+s} + 1 &= \phi_t \phi_s + \phi_t + \phi_s + 1 \\ &= (\phi_t + 1)(\phi_s + 1). \end{aligned} \tag{2.15}$$

Entonces, haciendo $\psi_t := \phi_t + 1$, la ecuación de arriba es equivalente a $\psi_{t+s} = \psi_t \psi_s$, i.e. con solución $\psi_t = e^{\alpha t}$, para $\alpha > 0$. Entonces

$$\phi_t = e^{\alpha t} - 1.$$

Así pues, la probabilidad de transición

$$P_t(x, A) = \int_A \mathsf{N}\left(y \mid xe^{-\alpha t} + \mu[1 - e^{-\alpha t}], \tau[1 - e^{-2\alpha t}]\right) dy \tag{2.16}$$

satisface (i), (ii) y (iii) de la Definición 18.

Nota 1. *En la Sección 2.2 entenderemos el origen de la parametrización usada en las funciones de transición propuestas en los dos ejemplos anteriores.*

El significado intuitivo de la función de transición, $P_{t,s}(x, A)$, es el de la probabilidad condicional que $X_s \in A$ dado que $X_t = x$, cuando $0 \leq t < s$.

Si denotamos por ν la distribución de X_0, es decir, la ***la distribución inicial del proceso de Markov* X**. Entonces, usando la propiedad de Markov de forma repetida, podemos obtener las funciones de distribución finito dimensionales mediante

$$\mathbb{E}[f(X_{t_0}, \ldots, X_{t_n})] = \int \nu(dx_0) \int P_{0,t_1}(x_0, dx_1) \int \cdots \int P_{t_{n-1},t_n}(x_{n-1}, dx_n) f(x_1, \ldots, x_n) \tag{2.17}$$

para todo $0 = t_0 < t_1 < t_2 \cdots < t_n \in \mathsf{T}$, $f : \mathbb{X}^{n+1} \mapsto \mathbb{R}$. En particular tenemos

$$\mathsf{P}(X_0 \in dx_0, X_{t_1} \in dx_1, \ldots, X_{t_n} \in dx_n) = \nu(dx_0) P_{0,t_1}(x_0, dx_1) \cdots P_{t_{n-1},t_n}(x_{n-1}, dx_n). \tag{2.18}$$

Con la obvia variación en el caso homogéneo en el tiempo.

Nota 2. *Denotaremos por $\mathbb{P}_\nu$ y $\mathbb{E}_\nu$ la probabilidad y la esperanza del proceso iniciado en ν. Análogamente $\mathbb{P}_x$ y $\mathbb{E}_x$ cuando $\nu = \delta_x$.*

Ejercicio 5. *Sea X un proceso de Markov homogéneo en el tiempo con distribución inicial $\nu(dx_0) = \mathsf{N}(x_0 \mid \mu, \tau) dx_0$ y probabilidad de transición (2.16). Demuestra que (2.18) se reduce a una distribución normal multivariada. Es decir con una densidad de Lebesgue de la forma*

$$\mathsf{N}_d(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\sqrt{det\Sigma}} \exp\left\{-\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)'\right\}, \quad x \in \mathbb{R}^d, d \in \mathbb{N}_+ \tag{2.19}$$

con $\mu \in \mathbb{R}^d$ y Σ una matrix no-singular, positiva definida. Encuentra la forma de esta última para este ejemplo.

En ocasiones es de interés relajar la condición (i) de la Definición 18 a medidas no-negativas que no necesariamente integran uno, en particular que satisfacen

$$P_t(x, \mathbb{X}) \leq 1, \quad \text{para toda } t \in \mathsf{T} \text{ y } x \in \mathbb{X}$$

En este caso se habla de una ***función de transición submarkoviana***, de darse la igualdad se dice que la función de transición es ***conservativa*** o simplemente markoviana. El mecanismo usual para convertir el caso general al caso conservativo es agregar un nuevo punto $\partial \notin \mathbb{X}$ al espació de estados $\mathbb{X}$, i.e. $\mathbb{X}_\partial = \mathbb{X} \cup \{\partial\}$ y entonces construir una función de transición en este espacio aumentado mediante

$$\begin{aligned} &P'_t(x, A) = P_t(x, A), \quad x \in \mathbb{X}, A \in \mathcal{X} \\ &P'_t(x, \partial) = 1 - P_t(x, \mathbb{X}), \quad x \neq \partial, x \in \mathbb{X} \\ &P'_t(\partial, \mathbb{X}) = 0, \quad P'_t(\partial, \{\partial\}) = 1 \end{aligned} \tag{2.20}$$

la cual es claramente una transición markoviana sobre $(\mathbb{X}_\partial, \mathcal{X}_\partial)$. Intuitivamente, esto significa que el proceso de Markov, modulado por la función de transición (2.20), se mueve libremente en $\mathbb{X}$ hasta que muere, lo cual ocurre cuando toca el punto ***cementerio*** $\{\partial\}$, en un tiempo

$$\zeta(\omega) = \inf\{t \in \mathsf{T}; X_t(\omega) = \partial\}$$

conocido como el ***tiempo de vida*** del proceso.

En relación a las propiedades de estabilidad antes mencionadas y todavía en el caso de procesos de Markov, una medida π que satisface

$$\pi(A) = \int_{\mathbb{X}} P_t(x, A)\pi(dx) \tag{2.21}$$

para cualquier t y $A \in \mathcal{X}$ se dice que es una ***medida invariante*** o una ***distribución estacionaria*** (cuando $\pi(\mathbb{X}) = 1$).

Ejercicio 6. *Demuestra que cualquier proceso de Markov* X *con distribución estacionaria* π*, y función de transición homogénea en el tiempo dada, conlleva a un proceso estrictamente estacionario, dado que la distribución inicial es* π*.*

Ejercicio 7. *Demuestra que un proceso de Markov* X *con funciones de transición homogéneas en el tiempo y distribución inicial* π *es un proceso estocástico estrictamente estacionario si y solo si* π *es una medida de probabilidad invariante.*

Un proceso de Markov X, homogéneo en el tiempo, es reversible con respecto a π si sus probabilidades de transición satisfacen la siguiente condición

$$\int_B P_t(x, B')\pi(dx) = \int_{B'} P_t(x, B)\pi(dx) \tag{2.22}$$

para todo $B, B' \in \mathcal{X}$. Si (2.22) se satisface para una medida finita, $\pi(\mathbb{X}) < \infty$, decimos que P_t es una ***función de transición*** π***-reversible***

Una media de probabilidad π se denomina la ***distribución ergódica*** o ***distribución límite*** de un proceso de Markov homogéneo en el tiempo si

$$P_t(x, \cdot) \to \pi(\cdot)$$

cuando $t \to \infty$.

Ejercicio 8. *Encuentra la distribución límite de* (2.16)*. Constituye esta una medida de probabilidad invariante para* (2.16)*?*

2.2. Construcción de transiciones markovianas vía intercambiabilidad

En referencia al aspecto (i), mencionado en la Sub-sección1.1.1 y relacionado con la construcción de la ley que regulará o modelará nuestro proceso markoviano, un primer paso es la construcción de funciones de transición markovianas que contengan ciertas caraterísticas deseables. Aquí veremos una metodología que nos permite construir (hasta ahorita de manera canónica) un proceso de Markov. homogéneo en el tiempo, estacionario fuertemente y para una elección de distribución invariante arbitraria pero dada.

En los ejemplos (2) y (3) la construcción de funciones de transición markoviana resultaron relativamente fáciles por las propiedades de cerradura de la distribución Gaussiana. En general, obtener expresiones analíticas para las transiciones a n-pasos (en el caso a tiempo discreto) o transiciones que satisfagan la ecuación de Chapman-Kolmogorov (en el caso a tiempo continuo), no es una tarea fácil. Más aun, este objetivo aparenta ser más complicado si le ponemos la restricción de que queremos que nuestra función markoviana deje invariante cierta distribución, π, cf. igualdad (2.21).

A continuación usaremos la simetría que caracteriza una sucesión de variables aleatorias intercambiables, $\{X_i\}_{i=1}^{\infty}$, como un mecanismo para construir funciones de transición markovianas con marginales dadas. Con este fin, es pertinente enunciar el celebrado Teorema de representación de de Finetti.

Teorema 5. *(Bruno de , Hewitt and Savage) Sea* $(\mathbb{X}, \mathcal{X})$ *un espacio completo y separable. Denotemos por* $\mathcal{P}_{\mathbb{X}}$ *el espacio de todas la medidas de probabilidad sobre* $(\mathbb{X}, \mathcal{X})$*. Una sucesión de variables aleatorias* $\mathsf{X}_e = \{X_i\}_{i=1}^{\infty}$*, con valores en* $\mathbb{X}$*, es intercambiable si y solo si existe* Q *sobre* $\mathcal{P}_{\mathbb{X}}$ *tal que*

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} P(A_i) \mathsf{Q}(dP), \quad \textit{para todo } A_i \in \mathcal{X}, \textit{ y toda } n \geq 1. \tag{2.23}$$

Del teorema anterior se sigue que la propiedad de intercambiabilidad equivale a independencia condicional, i.e.

$$X_i \mid P \overset{\text{iid}}{\sim} P, \quad \text{con} \quad P \sim \mathsf{Q}$$

Si denotamos por

$$P_n(A) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(A)$$

la ***distribución empírica***, entonces Q es la distribución de la ***medida de probabilidad aleatoria*** P, donde $\mathbb{P}[P_n \to P] = 1$, i.e. $P \sim \mathsf{Q}$. Tal vez una de las consecuencias más importantes del Teorema 5 es que dada una sucesión X_e existe una única Q y viceversa. A Q se le identifica como la medida de de Finetti inducida por X_e.

En general existen mecanismos para construir distribuciones, Q, sobre espacios funcionales $\mathcal{P}_{\mathbb{X}}$, y por lo tanto caracterizar las distribuciones finito dimensionales de una sucesión de variables aleatorias intercambiables. Por ejemplo, la distribución en la Proposición 1.19 en las notas del Dr. G. Uribe, donde se explora el caso, $\mathbb{X} = \{0, 1\}$. Claramente, el caso donde $\mathbb{X}$ se asume no numerable, e.g. $\mathbb{X} = \mathbb{R}$, resulta más engorroso. Habiendo dicho esto, una tarea aparentemente más sencilla resulta de restringirse al conjunto de variables aleatorias construidas mediante

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathscr{Y}} \prod_{i=1}^{n} \mu_y(A_i) \pi^{\mathsf{Y}}(dy), \quad \text{para todo } A_i \in \mathcal{X}, \text{ y toda } n \geq 1. \tag{2.24}$$

donde $\mu_y(\cdot)$ y $\pi^{\mathsf{Y}}(\cdot)$ son medidas de probabilidad sobre $(\mathbb{X}, \mathcal{X})$ y $(\mathscr{Y}, \sigma(\mathscr{Y}))$ respectivamente, para toda $y \in \mathscr{Y}$. En terminología de estadística bayesiana a π^{Y} se le conoce como la ***distribución inicial*** para lo desconocido, i.e. el parámetro $y \in \mathscr{Y}$, y a $\mu_y(\cdot)$ el modelo para nuestras observaciones, el cuál, condicionado a un valor del parámetro, provoca que estas sean iid.

Esta ***restricción paramétrica***, i.e. $\mathcal{P}_{\mathbb{X}}|_{\mathsf{F}_{\mathscr{Y}}}$ (donde $\mathsf{F}_{\mathscr{Y}}$ denota la familia parametrica caracterizada por $\mathscr{Y}$), reduce el posible espectro de sucesiones intercambiables a aquellas inducidas por (2.24), y por lo tanto no se sigue la unicidad de la medida de de Finetti Q, como sucedia en la generalidad comprendida por (2.23). La construcción depende de las formas paramétricas asumidas para las medidas de probabilidad $\mu_y(\cdot)$ y $\pi^{\mathsf{Y}}(\cdot)$, las cuales es suficiente que cumplan con que el soporte de π^{Y} coincida con $\{y \in \mathscr{Y}; \mu_y(\cdot) > 0\}$.

Utilizando lo anterior, si tomamos cualesquiera dos variables aleatorias de la sucesión intercambiable X_e, digamos X_1 y X_2, tenemos lo siguiente:

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \int_{\mathscr{Y}} \mu_y(A_1)\mu_y(A_2)\pi^{\mathsf{Y}}(dy), \quad \text{para todo } A_1, A_2 \in \mathcal{X}. \tag{2.25}$$

de donde se sigue que la ***distribución predictiva*** esta dada por

$$\mathbb{P}(X_2 \in B \mid X_1 = x) = \int_{\mathscr{Y}} \mu_y(B)\tilde{\mu}_x(\mathrm{d}y), \quad \text{para todo } B \in \mathcal{X}. \tag{2.26}$$

con

$$\tilde{\mu}_x(\mathrm{d}y) = \frac{\mu_y(x)\pi^{\mathsf{Y}}(dy)}{\int_{\mathscr{Y}} \mu_y(x)\pi^{\mathsf{Y}}(dy)}. \tag{2.27}$$

la ***distribución posterior*** " habiendo observado " x. Por claridad en la exposición, hemos abusado de la notación al asumir que $\mu_y(\cdot)$ tiene densidad $\mu_y(x)$, con respecto a alguna medida de referencia.

Al denominador en (2.27) también lo denotaremos como

$$\pi^{\mathrm{X}}(dx) = \int_{\mathscr{Y}} \mu_y(dx)\pi^{\mathrm{Y}}(dy), \tag{2.28}$$

que en la literatura bayesiana se conoce como la ***distribución predictiva inicial*** correspondiente a X_e. De hecho, notemos que esta última, corresponde a la distribución marginal para toda X_i de la sucesión intercambiable X_e, i.e. $X_i \sim \pi^{\mathrm{X}}$ para toda $i = 1, 2, \ldots$

Regresando a nuestro objetivo inicial, i.e. el de construir una función de transición markoviana que preserve una medida de probabilidad invariante (cf. 2.21), tenemos que esta se puede definir precisamente vía la distribución predictiva (2.26). Así pues, podemos construir una función de transición markoviana (a un paso), que mantenga π^{X} como una medida invariante vía

$$P(x, A) = \int \mu_y(A)\tilde{\mu}_x(\mathrm{d}y), \quad \text{para } A \in \mathcal{X}. \tag{2.29}$$

Ejercicio 9. *Demostrar que π^{X} constituye una medida de probabilidad invariante para la función de transición* (2.29).

Debido a la simetría inducida por (2.25), existe una medida bivariada η, t.q.

$$\eta(dx, dy) = \mu_y(dx)\pi^{\mathrm{Y}}(dy) = \tilde{\mu}_x(dy)\pi^{\mathrm{X}}(dx)$$

y por lo tanto para tener las dos medidas condicionales, μ_y y $\tilde{\mu}_x$ podemos:

1. Escoger la forma de π^{X}, la medida que fungirá como medida invariante de nuestro proceso markoviano. Suponer una forma parametrica para $\tilde{\mu}_x(\cdot)$, que cumpla con la condición de soporte arriba descrita. Aplicando el Teorema de Bayes obtenemos $\mu_y(\cdot)$ y construimos la función de transición markoviana vía (2.29).

2. Dada nuestra elección de medidas de probabildiad $\mu_y(\cdot)$ y π^{Y}, obtenemos $\tilde{\mu}_x(\cdot)$ vía Teorema de Bayes y construimos la función de transición markoviana vía (2.29). En este caso la forma de la distribución invariante π^{X} queda determinada por (2.28).

Ejercicio 10. *Demostrar que la función de transición* (2.29) *tiene la propiedad de ser π^{X}-reversible.*

Ejemplo 4. *Siguiendo el mecanismo* 1*, arriba descrito supongamos que nuestro objetivo es construir una función de transición que deje invariante la distribución gaussiana en $\mathbb{X} = \mathbb{R}$. Entonces escogemos $\pi^{\mathrm{X}}(dx) = \mathsf{N}(x; \mu, \tau)dx$ y $\tilde{\mu}_x(dy) = \mathsf{N}(y; x, \phi\tau)dy$. Aplicando Teorema de Bayes, obtenemos*

$$\mu_y(dx) = \mathsf{N}\left(x; \frac{y + \phi\mu}{1 + \phi}, \frac{\tau\phi}{1 + \phi}\right).$$

Entonces la función de transición esta dada por

$$P(x, A) = \int_A \mathsf{N}\left(y \mid \frac{x + \phi\mu}{1 + \phi}, \tau[1 - (1 + \phi)^{-2}]\right) dy \tag{2.30}$$

Ejercicio 11. *Verifica los cálculos de $\mu_y(\cdot)$ y $P(x, A)$ correspondientes al Ejemplo 4.*

2.2.1. Caso a tiempo continuo

Del preámbulo del Ejemplo 2, se deduce que la construcción anterior es suficiente para construir una función de transición markoviana a tiempo discreto. Sin embargo, si el objetivo es construir una transición markoviana a tiempo continuo, hay varios mecanismos. El más elemental, y que aquí usaremos, es el utilizado en el Ejemplo 3, es decir:

1. Permitir que nuestra elección de medidas condicionales $\tilde{\mu}_x$ ó μ_y dependan de algún parámetro no existente en π^{X}. Por ejemplo, $\tilde{\mu}_x^{\phi}$ ó μ_y^{ϕ}.

2. Inducir la dependencia en el tiempo (recordemos que estamos en el caso homogéneo en el tiempo) vía el parámetro ϕ. Es decir, considerar una función $t \mapsto \phi_t$ y posteriormente encontrar la forma funcional de esta tal que la ecuación de Chapman-Kolmogorov se satisface (cf. Ejemplo 3)

Utilizando lo anterior, en el contexto del Ejemplo 4 uno obtiene la función de transición markoviana dada por (2.16).

Ejercicio 12. *(Continuación Ejemplo 3.) Demuestra que $\pi^{\mathsf{X}}(dx) = \mathsf{N}(x;\mu,\tau)dx$ es una distribución límite para la función de transición markoviana (2.16). Concluye que un proceso de Markov $\mathsf{X} := (X_t; t \geq 0)$ con función de transición (2.16) y distribución inicial $\nu = \pi^{\mathsf{X}}$, es un proceso fuertemente estacionario y reversible en el tiempo.*

2.3. Perspectiva de operadores y generadores

Denotemos por $\mathcal{C}(\mathbb{X})$ el espacio de funciones $f : \mathbb{X} \to \mathbb{R}$ continuas y acotadas. $\mathcal{C}(\mathbb{X})$ es un espacio de Banach con la norma de supremo
$$||f||_\infty = \sup_{x \in \mathbb{X}} |f(x)|$$
Así pues, decimos que una función $f \in \mathcal{C}(\mathbb{X})$ converge a cero cuando $x \to \partial$, si para cada $\epsilon > 0$ existe un subconjunto compacto E de $\mathbb{X}$ tal que
$$|f(x)| < \epsilon, \quad x \in \mathbb{X} \backslash E$$
y entonces se puede escribir $\lim_{x \to \partial} f(x) = 0$. Denotamos entonces
$$\mathcal{C}_0(\mathbb{X}) = \left\{ f \in \mathcal{C}(\mathbb{X}) : \lim_{x \to \partial} f(x) = 0 \right\}.$$
que es un subespacio cerrado de $\mathcal{C}(\mathbb{X})$ y por lo tanto Banach. Si $\mathbb{X}$ es compacto no se hace diferencia en la notación. Así pues cualquier función real f sobre $\mathbb{X}$ se extiende al espació compacto $\mathbb{X}_\partial$ haciendo $f(\partial) = 0$. Desde esta perspectiva, el espacio $\mathcal{C}_0(\mathbb{X}) \subset \mathcal{C}(\mathbb{X}_\partial)$ y por lo tanto también podemos escribir
$$\mathcal{C}_0(\mathbb{X}) = \{ f \in \mathcal{C}(\mathbb{X}_\partial) : f(\partial) = 0 \}$$

Cuando se estudia las propiedades distribucionales de los procesos de Markov, resulta menos engorroso via un enfoque funcional, al menos en el caso general. Esto es, mediante el estudio del cambio que induce una transición markoviana para una función dada, como en (2.3). Así pues, se puede definir el ***operador de transición***
$$\mathsf{P}_t f(x) = \int_{\mathbb{X}} f(y) P_t(x, dy) \tag{2.31}$$
para toda $f \in \mathcal{C}_0(\mathbb{X})$. Notemos que para $A \in \mathcal{X}$, $\mathsf{P}_t \mathbb{I}_A(x) = P_t(x, A)$. También en este contexto la ecuación de Chapman-Kolmogorov se traduce a
$$\mathsf{P}_{t+s} f(x) = \mathsf{P}_t \left(\mathsf{P}_s f(x) \right) = \mathsf{P}_s \left(\mathsf{P}_t f(x) \right) \tag{2.32}$$
para cada $f \in \mathcal{C}_0(\mathbb{X})$ medible. En otras palabras, los operadores (2.31) forman un ***semigrupo*** de operadores en el espacio $\mathcal{C}_0(\mathbb{X})$.

Con este enfoque, se puede decir que X es un ***proceso de Markov con distribución inicial*** ν si

- $\mathbb{P}_\nu(X_0 \in A) = \nu(A)$, para todo $A \in \mathcal{X}$
- Existe una función de transición markoviana tal que para todo $s, t \geq 0$ y toda función f medible y acotada
$$\mathbb{E}_\nu(f(X_{t+s} \mid \mathcal{F}_s)) = \mathsf{P}_t f(X_s)$$

Usando el teorema de extensión de Kolmogorov, se puede construir la versión canonica de un proceso de Markov con funciones de transición dadas, $(P_t : t \geq 0)$, y distribución inicial ν.

Lema 1. Sea $(P_t : t \geq 0)$ un operador de transición con la propiedad de semigrupo, i.e. (2.31), y ν una medida de probabilidad sobre $(\mathbb{X}, \mathcal{X})$, que se asume localmente completo y separable. Entones las distribuciones finito dimensionales

$$\mu_{t_0,t_1,\dots,t_n}(A_0 \times A_1 \times \cdots \times A_n) = \nu \mathbb{I}_{A_0} \mathsf{P}_{t_1-t_0} \mathbb{I}_{A_1} \cdots \mathsf{P}_{t_n-t_{n-1}} \mathbb{I}_{A_n}, \quad A_0, A_1, \dots, A_n \in \mathcal{X}$$

son una familia distribuciones consistente para cualesquiera $0 = t_0 < t_1 < \cdots < t_n$.

Ejercicio 13. *Demostrar el Lemma 1 y usar el Teorema de extensión de Kolmogorov para construir el proceso de Markov canónico asociado a dicha familia de distribuciones.*

Para referencia futura, formalicemos el concepto de semigrupo mediante la siguiente definición.

Definición 19. Sea $\mathcal{S}$ un espacio de Banach de funciones reales y medibles sobre un espacio Polaco $(\mathbb{X}, \mathcal{X})$ y $||\cdot||$ la norma correspondiente. Una familia de operadores lineales, $\mathsf{P}_t : \mathcal{S} \mapsto \mathcal{S}$, para $t \geq 0$, se dice que es un ***semigrupo fuerte de operadores*** en $\mathcal{S}$, abreviado como (SFO($\mathcal{S}$)), si

i) P_t es acotado para cada $t \geq 0$. En otras palabras

$$||\mathsf{P}_t|| := \sup_{f \in \mathcal{S}} \frac{||\mathsf{P}_t f||}{||f||} < \infty, \quad \text{para toda} \quad t \geq 0$$

ii) $\mathsf{P}_{t+s} = \mathsf{P}_t \mathsf{P}_s = \mathsf{P}_s \mathsf{P}_t$

iii)
$$\lim_{t \downarrow 0} ||\mathsf{P}_t f - f|| = 0 \quad \text{para cada} \quad f \in \mathcal{S}.$$

Si además se tienen que

$$||\mathsf{P}_t f|| \leq ||f|| \quad (||\mathsf{P}_t|| \leq 1) \quad \text{para toda } t \geq 0$$

se le denomina un ***SFO***$(\mathcal{S})$ ***de contracción.***

Dado $\mathcal{D}$ un subespacio lineal de $\mathcal{S}$ ($\mathcal{D} \subset \mathcal{S}$) y $\mathsf{T} : \mathcal{D} \mapsto \mathcal{S}$ un operador lineal. A $\mathcal{D}$ se le denomina el ***dominio*** de T. El conjunto $\mathcal{G}(\mathsf{T}) := \{(f, \mathsf{T}f) | f \in \mathcal{D}\} \subset \mathcal{S} \times \mathcal{S}$ se conoce como la ***gráfica*** de T. Nótese que $\mathcal{S} \times \mathcal{S}$ es también un espacio de Banach con la norma $||(f, g)|| = ||f|| + ||g||$. Se dice que T es ***cerrado*** si y solo si $\mathcal{G}(\mathsf{T}) = \overline{\mathcal{G}(\mathsf{T})}$.

Lema 2. Sea T un SFO($\mathcal{S}$) entonces es continuo.

Demostración. Unicamente hay que demostrar que

$$\lim_{n \to \infty} f_n = f \quad \Rightarrow \quad \mathsf{T}f_n \to \mathsf{T}f \quad \text{fuertemente}$$

pero por propiedades de la norma se sigue que

$$||\mathsf{T}f_n - \mathsf{T}f|| \leq ||\mathsf{T}(f_n - f)|| \leq ||\mathsf{T}||\,||f_n - f||$$

de donde se sigue el resultado.

Así pues, si $(\mathsf{P}_t : t \geq 0)$ es un SFO($\mathcal{S}$), entonces es un operador lineal cerrado para cada $t \geq 0$.

Ejercicio 14. *Sea* $\mathsf{X} := (X_t : t \geq 0)$ *un proceso de Markov homogéneo en el tiempo con probabilidad de transición*

$$P_t(x, A) = \int_A \mathsf{N}(y \mid x, t) \mathrm{d}y \tag{2.33}$$

sobre $(\mathbb{R}, \mathcal{B})$. *Demuestra que*

$$\mathsf{P}_t f(x) = \int_{\mathbb{R}} f(y) \mathsf{N}(y \mid x, t) \mathrm{d}y = \int_{\mathbb{R}} f(x + \sqrt{t} y) \mathsf{N}(y \mid 0, 1) \mathrm{d}y.$$

forman un SFO$(\mathcal{C}_0(\mathbb{R}))$.

Lema 3. Sea $(\mathsf{P}_t : t \geq 0)$ un SFO$(\mathcal{S})$. Entonces, existen constantes $M \geq 1$ y $\alpha \geq 0$ tal que $||\mathsf{P}_t|| \leq Me^{\alpha t}$.

Demostración. Primero notemos que existen constantes $M \geq 1$ y $t \leq t_0$ tal que $||\mathsf{P}_t|| \leq M$, $t \leq t_0$, ya que de no ser así entonces existiría una sucesión $t_n \downarrow 0$, $n \to \infty$ tal que $||\mathsf{P}_{t_n}|| \to \infty$ y (por Banach-Steinhaus) $\sup_n ||\mathsf{P}_{t_n} f|| = \infty$ para alguna $f \in \mathcal{S}$. Lo que contradice la continuidad. Entonces, existe $M \geq 1$ tal que $||\mathsf{P}_t|| \leq M$ para $t \leq t_0$. Finalmente, fijemos $\alpha = (\log M)/t_0$. Sea $t \in [0, \infty)$. Entonces para $[t/t_0]$ tenemos

$$||\mathsf{P}_t|| = ||\mathsf{P}_{kt_0}\mathsf{P}_{t-kt_0}|| \leq ||\mathsf{P}_{t_0}||^k ||\mathsf{P}_{t-kt_0}|| \leq e^{\alpha t}M. \quad \square$$

Corolario 2. *$t \mapsto \mathsf{P}_t f$ es continuo, i.e.*

$$\lim_{s \to t} ||\mathsf{P}_t f - \mathsf{P}_s f|| = 0, \quad \forall f \in \mathcal{S} \text{ y } t \geq 0.$$

Demostración. Probemos sólo la continuidad por la derecha. Sea $h > 0$, entonces para $f \in \mathcal{S}$ se tiene

$$||\mathsf{P}_{t+h} f - \mathsf{P}_t f|| = ||\mathsf{P}_t \mathsf{P}_h f - \mathsf{P}_t f|| \leq ||\mathsf{P}_t|| \, ||\mathsf{P}_h f - f|| \leq Me^{\alpha t} ||\mathsf{P}_h f - f|| \to 0, \quad h \downarrow 0.$$

Para $t \geq h \geq 0$ tenemos

$$\begin{aligned} ||\mathsf{P}_{t-h} f - \mathsf{P}_t f|| &= ||\mathsf{P}_{t-h} f - \mathsf{P}_{t-h+h} f|| = ||\mathsf{P}_{t-h} f - \mathsf{P}_{t-h}\mathsf{P}_h f|| \leq ||\mathsf{P}_{t-h}|| \, ||f - \mathsf{P}_h f|| \\ &\leq Me^{\alpha(t-h)} ||f - \mathsf{P}_h f|| \to 0, \quad h \downarrow 0, \end{aligned}$$

donde las últimas desigualdades de las dos expresiones anteriores se siguen de las cotas exponenciales en de un SFO, i.e. Lema 3. $\square$

Si observamos la propiedad de adición en el " parámetro " tiempo que característica un semigrupo, ésta es similar a la función exponencial en los reales, en donde, para cada $p \in \mathbb{R}$, $p^{t+s} = p^t p^s$. Esto sugiere que para cualquier p existe un número real a tal que $p^t = e^{at}$, donde

$$e^{at} = \sum_{i=0}^{\infty} \frac{(ta)^i}{i!}.$$

Esto sugiere la existencia de un operador $\mathcal{A}$, denominado el ***generador infinitesimal*** de $(P_t; t \geq 0)$. Nótese que en el caso de la función exponencial, dicho valor de a, se puede determinar como $a = \frac{d}{dt} e^{at}|_{t=0}$ o mediante integración

$$\int_0^{\infty} e^{-\lambda t} p^t dt = \frac{1}{\lambda - a}, \quad \lambda > a.$$

con $p^t = e^{ta}$.

Entonces, por analogía a la propiedad (2.32), la primera de estas formas de determinar el valor de a,sugiere la existencia de un operador infinitesimal que module la tendencia del proceso, i.e.

$$\frac{d}{dt} P_t = \lim_{s \downarrow 0} \frac{1}{s} (\mathsf{P}_{t+s} - \mathsf{P}_t) = \mathsf{P}_t \mathcal{A} = \mathcal{A}\mathsf{P}_t$$

donde a

$$\mathcal{A} := \lim_{s \downarrow 0} \frac{1}{s} (\mathsf{P}_s - \mathsf{I})$$

se le denomina el ***generador infinitesimal*** de $(P_t; t \geq 0)$. En otros términos el generador $\mathcal{A}f(x)$ describe el cambio infinitesimal de $f(X_t)$ dado que $X_t = x$. Es decir

$$\lim_{\epsilon \downarrow 0} \mathbb{E} \left(\frac{f(X_{t+\epsilon}) - f(X_t)}{\epsilon} \mid X_t = x \right)$$

Evidentemente, la utilidad de dicho generador radica, inicialmente, en entender su soporte, es decir el conjunto de funciones donde esta bien definido.

Ejemplo 5 (Generador movimiento browniano)**.** *Sea* $\mathsf{X} := (X_t : t \geq 0)$ *un proceso de Markov homogéneo en el tiempo con probabilidad de transición*

$$P_t(x, A) = \int_A \mathsf{N}(y \mid x, t)\mathrm{d}y \tag{2.34}$$

sobre $(\mathbb{R}, \mathcal{B})$*. Entonces*

$$\mathsf{P}_t f(x) = \int_{\mathbb{R}} f(y)\mathsf{N}(y \mid x, t)\mathrm{d}y = \int_{\mathbb{R}} f(x + \sqrt{t}y)\mathsf{N}(y \mid 0, 1)\mathrm{d}y.$$

Por lo tanto el generador infinitesimal esta dado por

$$\begin{aligned} \mathcal{A}f(x) &= \lim_{t \downarrow 0} \frac{1}{t} \{\mathsf{P}_t f(x) - f(x)\} \\ &= \lim_{t \downarrow 0} \frac{1}{t} \int_{\mathbb{R}} \Big[f(x + \sqrt{t}y) - f(x) \Big] \mathsf{N}(y \mid 0, 1)\mathrm{d}y. \end{aligned} \tag{2.35}$$

Ahora si usamos la expansión de Taylor de $f(x + \sqrt{t}y)$ *alrededor de* $y = 0$ *obtenemos*

$$f(x + \sqrt{t}y) = f(x) + f'(x)\sqrt{t}y + f''(x)t\frac{y^2}{2} + O(y^3)$$

por lo que

$$\begin{aligned} \mathcal{A}f(x) &= \lim_{t \downarrow 0} \frac{1}{t} \int_{\mathbb{R}} \left[f'(x)\sqrt{t}y + f''(x)t\frac{y^2}{2} \right] \mathsf{N}(y \mid 0, 1)\mathrm{d}y \\ &= \lim_{t \downarrow 0} \left[f'(x)\frac{\sqrt{t}}{t}\,\mathbb{E}[Y] + \frac{f''(x)}{2}\mathbb{E}[Y^2] \right] \\ &= \frac{f''(x)}{2} \end{aligned} \tag{2.36}$$

con $Y \sim \mathsf{N}(0, 1)$ *y por lo tanto* $\mathbb{E}[Y] = 0$ *y* $\mathbb{E}[Y^2] = 1$*. Claramente el dominio de este operador requiere que* $f', f'' \in \mathcal{C}(\mathbb{R})$*, por lo que* $\mathcal{D}(\mathcal{A}) = \mathcal{C}^2(\mathbb{R})$*.*

Ejercicio 15. *Completar el Ejemplo 5, demostrando que el ultimo término de la expansión deTaylor se va a cero en el calculo del generador infinitesimal* $\mathcal{A}$*.*

Ejercicio 16. *Sea* X *el proceso de Markov con probabilidad de transición (2.16). Calcula su generador infinitesimal.*

Definición 20. Dado un SFO($\mathcal{S}$), $(\mathsf{P}_t : t \geq 0)$ y

$$\mathcal{D} = \left\{ f \in \mathcal{S} | \, \exists \, g \in \mathcal{S} \text{ tal que } \lim_{t \downarrow 0} \left|\left| \frac{\mathsf{P}_t f - f}{t} - g \right|\right| = 0 \right\} \tag{2.37}$$

el operador lineal $\mathcal{A} : \mathcal{D} \to \mathcal{S}$ dado por

$$\mathcal{A}f = g = \lim_{t \downarrow 0} \frac{\mathsf{P}_t f - f}{t}$$

se le conoce como el ***generador infinitesimal*** asociado a $(\mathsf{P}_t : t \geq 0)$. A (2.37) se le denomina el ***dominio*** de $\mathcal{A}$, el cual también se identificara por $\mathcal{D}(\mathcal{A})$.

Teorema 6. *Sea* $(\mathsf{P}_t : t \geq 0)$ *un SFO*($\mathcal{S}$)*.*

i) Sea $f \in \mathcal{S}$*,* $t \geq 0$*. Entonces* $\int_0^t \mathsf{P}_s f ds \in \mathcal{D}(\mathcal{A})$ *y*

$$\mathsf{P}_t f - f = \mathcal{A} \int_0^t \mathsf{P}_s f ds$$

ii) Sea $f \in \mathcal{D}(\mathcal{A})$ y $t \geq 0$. Entonces $\mathsf{P}_t f \in \mathcal{D}(\mathcal{A})$. La función $t \mapsto \mathsf{P}_t f$ es derivable en $\mathcal{S}$ y las ecuaciones hacia atras **backward** *y hacia delante* **forward** *de Kolmogorov se satisfacen:*

$$\frac{d}{dt}\mathsf{P}_t f = \mathcal{A}\mathsf{P}_t f = \mathsf{P}_t \mathcal{A} f,$$

es decir

$$\lim_{h\downarrow 0}\left|\left|\frac{\mathsf{P}_{t+h}f - \mathsf{P}_t f}{h} - \mathsf{P}_t\mathcal{A}f\right|\right| = \lim_{h\downarrow 0}\left|\left|\frac{\mathsf{P}_{t+h}f - \mathsf{P}_t f}{h} - \mathcal{A}\mathsf{P}_t f\right|\right| = 0.$$

iii) Sea $f \in \mathcal{D}(\mathcal{A})$, $t \geq 0$. Entonces,

$$\mathsf{P}_t f - f = \int_0^t \mathsf{P}_s \mathcal{A} f ds = \int_0^t \mathcal{A}\mathsf{P}_s f ds$$

Demostración. Para la parte (i) notemos que

$$\frac{1}{h}(\mathsf{P}_h - \mathsf{I})\int_0^t \mathsf{P}_s f ds = \frac{1}{h}\int_0^t (\mathsf{P}_{s+h}f - \mathsf{P}_s f)\, ds$$

Haciendo el cambio de variable $s + h = u$, podemos escribir la primera integral como $\int_0^t \mathsf{P}_{s+h} f ds = \int_h^{t+h} \mathsf{P}_u f du$

$$= \frac{1}{h}\int_t^{t+h}\mathsf{P}_s f ds - \frac{1}{h}\int_0^h \mathsf{P}_s f ds$$

Notemos que para el segundo sumando de esta última igualdad tenemos

$$\left|\left|\frac{1}{h}\int_0^t \mathsf{P}_s f ds - f\right|\right| \leq \frac{1}{h}\int_0^h ||\mathsf{P}_s f - f|| ds \to 0, \quad h \downarrow 0$$

de manera análoga

$$\left|\left|\frac{1}{h}\int_t^{t+h} \mathsf{P}_s f ds - \mathsf{P}_t f\right|\right| \leq \frac{||\mathsf{P}_t||}{h}\int_0^h ||\mathsf{P}_s f - f|| ds \to 0, \quad h \downarrow 0$$

y por lo tanto el resultado se sigue. Para (ii) nótese que

$$\left|\left|\frac{\mathsf{P}_{t+h}f - \mathsf{P}_t f}{h} - \mathsf{P}_t\mathcal{A}f\right|\right| = \left|\left|\mathsf{P}_t\left(\frac{\mathsf{P}_h f - f}{h} - \mathcal{A}f\right)\right|\right| \leq ||\mathsf{P}_t||\left|\left|\frac{\mathsf{P}_h f - f}{h} - \mathcal{A}f\right|\right| \to 0, \quad h \downarrow 0$$

entonces

$$\lim_{h\downarrow 0}\left|\left|\frac{\mathsf{P}_{t+h}f - \mathsf{P}_t f}{h} - \mathsf{P}_t\mathcal{A}f\right|\right| = 0$$

Ahora, dado que $\mathcal{A}f \in \mathcal{S}$, $g = \mathsf{P}_t\mathcal{A}f \in \mathcal{S}$. Entonces re-escribiendo la última ecuación

$$\lim_{h\downarrow 0}\left|\left|\frac{\mathsf{P}_h(\mathsf{P}_t f) - \mathsf{P}_t f}{h} - g\right|\right| = 0$$

y por lo tanto $g = \mathcal{A}\mathcal{P}_t f$. Consecuentemente

$$\mathcal{P}_t\mathcal{A}f = \mathcal{A}\mathcal{P}_t f = \frac{d^+}{dt}\mathsf{P}_t f$$

Para ver que la derivada por la izquierda existe y es igual a la derivada derecha, observemos que para $h > 0$

$$\begin{aligned}\left|\left|\frac{\mathsf{P}_t f - \mathsf{P}_{t-h}f}{h} - \mathsf{P}_t\mathcal{A}f\right|\right| &\leq \left|\left|\frac{\mathsf{P}_t f - \mathsf{P}_{t-h}f}{h} - \mathsf{P}_{t-h}\mathcal{A}f\right|\right| + ||\mathsf{P}_{t-h}\mathcal{A}f - \mathsf{P}_t\mathcal{A}f|| \\ &\leq ||\mathsf{P}_{t-h}||\left|\left|\frac{\mathsf{P}_h f - f}{h} - \mathcal{A}f\right|\right| + ||\mathsf{P}_{t-h}\mathcal{A}f - \mathsf{P}_t\mathcal{A}f|| \to 0, \quad h \downarrow 0\end{aligned}$$

que se sigue por la continuidad fuerte y el hecho de que $\mathcal{A}f \in \mathcal{S}$. Para (iii), primero notemos que

$$\frac{d}{dt}\mathsf{P}_t f = \mathsf{P}_t \mathcal{A} f$$

es una función continua en t por Corolario 2. Por lo tanto es integrable y entonces

$$\mathsf{P}_t f - f = \int_0^t \frac{d}{ds}\mathsf{P}_s f ds = \int_0^t \mathcal{A}\mathsf{P}_s f ds = \int_0^t \mathsf{P}_s \mathcal{A} f ds$$

□

Nota 3. *El motivo por la terminología " backward " [" forward "] se debe a que la ecuación se obtiene con la perturbación del punto inicial [final resp.], i.e.* $\mathcal{A}\mathsf{P}_t f = \lim_{\epsilon\to 0} \frac{1}{\epsilon}(\mathsf{P}_\epsilon - \mathbb{I})\mathsf{P}_t f$ *[*$\mathcal{A}\mathsf{P}_t f = \lim_{\epsilon\to 0} \mathsf{P}_t(\frac{1}{\epsilon}(\mathsf{P}_\epsilon f - f))$ *resp]. A la ecuación " forward " también se le conoce como la* **ecuación de Fokker-Planck**.

Nota 4. *Del Teorema 6 podemos verificar que* π *es la distribución estacionaria de un proceso de Markov ssi* $\mathbb{E}\mathcal{A}f(X) = 0$, *con* $X \sim \pi$, *y para toda* f *para la cual* $\mathcal{A}f$ *este bien definido.*

Ejemplo 6. *Más adelante veremos que los operadores* $\mathcal{A}f(x) = f^{''}(x) - xf^{'}(x)$ *y* $\mathcal{A}f(x) = \lambda f(x+1) - xf(x)$, *corresponden a procesos de Markov con distribuciones invariantes* $\mathsf{N}(0,1)$ *y* $\mathsf{Po}(\lambda)$, *respectivamente (Ornstein-Uhlenbeck y un proceso de nacimiento y muerte con inmigración)*

Nota 5. *La ventaja del enfoque con generadores es que generalizaciones a espacios multivariados ó a espacios más complejos, son relativamente simples.*

Corolario 3. *Sea* $(\mathsf{P}_t : t \geq 0)$ *un SFO(*$\mathcal{S}$*). Entonces* $\overline{\mathcal{D}(\mathcal{A})} = \mathcal{S}$ *(i.e.* $\mathcal{D}(\mathcal{A})$ *es denso en* $\mathcal{S}$*) y* $\mathcal{A}$ *es un operador cerrado.*

Demostración. Teorema 6 (i) y el hecho de que, para cada $f \in \mathcal{S}$, se tiene

$$\left|\left|\int_0^t \frac{\mathsf{P}_s f ds}{t} - f\right|\right| \to 0, \quad t \downarrow 0$$

implican que $\overline{\mathcal{D}(\mathcal{A})} = \mathcal{S}$, por Teorema 3 (i).

Ahora, sea $\{f_n\}_n \subset \mathcal{D}(\mathcal{A})$ cualquier sucesión de funciones con la propiedad de que existen $f, g \in \mathcal{S}$ tal que $f_n \to f$ y $\mathcal{A}f_n \to g$ cuando $n \to \infty$. Necesitamos demostrar que $g = \mathcal{A}f$. Con esta finalidad, nótese que, gracias al Teorema 6 (iii), tenemos

$$\mathsf{P}_t f_n - f_n = \int_0^t \mathsf{P}_s(\mathcal{A}f_n)ds, \quad \text{para toda } t > 0.$$

Como

$$||(\mathsf{P}_t f_n - f_n) - (\mathsf{P}_t f - f)|| \to 0 \quad \text{y} \quad \left|\left|\int_0^t \mathsf{P}_s \mathcal{A} f_n ds - \int_0^t \mathsf{P}_s g ds\right|\right| \to 0, \quad \text{cuando } n \to \infty$$

necesariamente

$$\mathsf{P}_t f - f = \int_0^t \mathsf{P}_s g ds$$

para toda $t > 0$. Por lo tanto,

$$\left|\left|\frac{\mathsf{P}_t f - f}{t} - \frac{\int_0^t \mathsf{P}_s g ds}{t}\right|\right| = 0, \forall t > 0.$$

Entonces

$$\lim_{t\downarrow 0}\left|\left|\frac{\mathsf{P}_t f - f}{t} - g\right|\right| \leq \lim_{t\downarrow 0}\left|\left|\frac{\mathsf{P}_t f - f}{t} - \frac{\int_0^t \mathsf{P}_s g ds}{t}\right|\right| + \lim_{t\downarrow 0}\left|\left|\frac{\int_0^t \mathsf{P}_s g ds}{t} - g\right|\right| = 0$$

de tal forma que $g = \mathcal{A}f$. □

Cuando se estudia un proceso de Markov vía su generador infinitesimal, la pregunta que surge de inmediato es si éste caracteriza un semigrupo de operadores de transición, i.e., en que sentido podemos recuperar P_t mediante exponenciación de $\mathcal{A}$. La respuesta a esta pregunta la encontramos mediante el celebrado teorema de Hille-Yosida, el cual provee las condiciones suficientes para que un operador sea el operador infinitesimal de un semigrupo. Para poder enunciar este teorema, primero introduciremos otro operador asociado a un semigrupo. Recordemos la segunda manera de obtener a de $p^t = e^{at}$, i.e. vía

$$r_\lambda := \int_0^\infty e^{-\lambda t} p^t dt = \frac{1}{\lambda - a}, \quad \lambda > a.$$

con $p^t = e^{ta}$. . Nótese que podemos recuperar a de r_λ, mediante a_λ,

$$a_\lambda := \lambda(\lambda r_\lambda - 1) = \frac{a}{1 - \frac{a}{\lambda}}$$

ya que $a_\lambda \to a$ cuando $\lambda \uparrow \infty$.

Por el Corolario 2, $t \mapsto \mathsf{P}_t$ es continuo para cada $f \in \mathcal{S}$. También sabemos que que $||\mathsf{P}_t|| \leq M e^{\alpha t}$ para constantes $M \geq 1$ y $\alpha \geq 0$. Entonces podemos definir

$$\mathsf{R}_\lambda f(x) = \int_0^\infty e^{-\lambda t} \mathsf{P}_t f(x) dt, \tag{2.38}$$

es decir la transformada de Laplace del semigrupo. A dicho mapeo se le conoce como el ***resolvente de orden*** λ.

Lema 4. Sea $(\mathsf{P}_t : t \geq 0)$ un SFO$(\mathcal{S})$.

i) $||R_\lambda|| \leq M/(\lambda - \alpha)$

ii) La ***ecuación resolvente***

$$R_\mu - R_\lambda + (\mu - \lambda) R_\mu R_\lambda = 0. \tag{2.39}$$

es valida para cualesquiera $\lambda, \mu > \alpha$.

Demostración. La primera parte se sigue de inmediato de la prop. Para demostrar (ii), notemos que

$$e^{-\mu t} - e^{-\lambda t} = (\lambda - \mu) e^{-\lambda t} \int_0^t e^{-(\lambda - \mu)s} ds$$

Entonces

$$\begin{aligned} R_\mu f(x) - R_\lambda f(x) &= \int_0^\infty (e^{-\mu t} - e^{-\lambda t}) \mathsf{P}_t f(x) dt \\ &= (\lambda - \mu) \int_0^\infty e^{-\lambda t} \left(\int_0^t e^{(\lambda - \mu)s} \mathsf{P}_t f(x) ds \right) dt \\ &= (\lambda - \mu) \int_0^\infty e^{-\mu s} \left(\int_s^\infty e^{-\lambda(t-s)} \mathsf{P}_t f(x) dt \right) ds \end{aligned} \tag{2.40}$$

Ahora notemos que

$$\begin{aligned} \int_s^\infty e^{-\lambda(t-s)} \mathsf{P}_t f(x) dt &= \int_0^\infty e^{-\lambda u} \mathsf{P}_{s+u} f(x) du = \int_0^\infty e^{-\lambda u} \mathsf{P}_s \mathsf{P}_u f(x) du \\ &= \int_0^\infty e^{-\lambda u} \left(\int_{\mathbb{X}} \mathsf{P}_u f(y) P_s(x, dy) \right) du \\ &= \int_{\mathbb{X}} \left(\int_0^\infty e^{-\lambda u} \mathsf{P}_u f(y) du \right) P_s(x, dy) = \mathsf{P}_s \mathsf{R}_\lambda f(x) \end{aligned}$$

lo que, sustituyendo en (2.40), se reduce a la ecuación resolvente. □

Nótese que si P_t es honesto, entonces $R_\lambda \mathbb{I} = \lambda^{-1}$, para toda $\lambda > 0$.

Ejercicio 17. *Sea* $(\mathsf{P}_t : t \geq 0)$ *el semigrupo inducido por la función de transición*

$$P_t(x, A) = (1 - e^{-\alpha t})Q(A) + e^{-\alpha t}\delta_x(A), \quad \alpha > 0, A \in \mathcal{X} \tag{2.41}$$

y Q *una medida de probabilidad sobre* $(\mathbb{X}, \mathcal{X})$.

I. Es $(\mathsf{P}_t : t \geq 0)$ *un SFO*$(\mathcal{S})$*?*

II. Si $\mathsf{X} := (X_t : r \geq 0)$ *es un proceso de Markov con ley inducida por* $X_0 \sim Q$ *y* $(\mathsf{P}_t : t \geq 0)$*. Se trata de un proceso reversible y estrictamente estacionario?*

III. Simplifica la forma de su operador resolvente asociado.

El siguiente teorema establece la relación entre el semigrupo y generador infinitesimal, se puede considerar como la primera parte del Teorema de Hille-Yosida.

Teorema 7. *Sea* $(\mathsf{P}_t : t \geq 0)$ *un SFO*$(\mathcal{S})$ *con* $||\mathsf{P}_t|| \leq Me^{\alpha t}$ *y* $\mathcal{A}$ *su generador infinitesimal. Para toda* $\lambda > \alpha$ *se satisface*

I. $\mathsf{R}_\lambda \mathcal{S} = \mathcal{D}(\mathcal{A})$

II. $\lambda\mathbb{I} - \mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{S}$ *define un operador biyectivo con* $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A}) = \mathcal{S}$

III. $(\lambda\mathbb{I} - \mathcal{A})^{-1} : \mathcal{S} \to \mathcal{D}(\mathcal{A})$ *existe como un operador lineal acotado. En particular,* $(\lambda\mathbb{I} - \mathcal{A})^{-1} = \mathsf{R}_\lambda$

IV. $\mathsf{R}_\lambda(\lambda\mathbb{I} - \mathcal{A})f = f$ *para toda* $f \in \mathcal{D}(\mathcal{A})$.

V. $(\lambda\mathbb{I} - \mathcal{A})\mathsf{R}_\lambda g = g$ *para toda* $g \in \mathcal{S}$.

Demostración.

Paso 1. Demostrar (V)

Paso 2. Demostrar (IV)

Paso 3. (IV) $\Rightarrow$ (I) y la primera parte de (II)

Paso 4. (V) $\Rightarrow$ la segunda parte de (II).

Paso 5. (III) se sigue de combinar (IV,I)

Paso 1. Demostración (V). Sea $g \in \mathcal{S}$, entonces podemos escribir

$$\mathsf{P}_h\mathsf{R}_\lambda g = \int_0^\infty e^{-\lambda t}\mathsf{P}_{t+h} g dt$$

Entonces

$$\begin{aligned}\frac{\mathsf{P}_h\mathsf{R}_\lambda g - \mathsf{R}_\lambda g}{h} &= \frac{1}{h}\left[\int_0^\infty e^{-\lambda t}\mathsf{P}_{t+h} g dt - \int_0^\infty e^{-\lambda t}\mathsf{P}_t g dt\right] \\ &= \frac{e^{\lambda h} - 1}{h}\mathsf{R}\lambda g - \frac{1}{h}e^{\lambda h}\int_0^h e^{-\lambda t}\mathsf{P}_t g dt\end{aligned}$$

lo que converge a $\lambda\mathsf{R}_\lambda g - g$. Entonces se sigue que

$$\left|\left|\frac{\mathsf{P}_h\mathsf{R}_\lambda g - \mathsf{R}_\lambda g}{h} - (\lambda\mathsf{R}_\lambda g - g)\right|\right| \to 0, \quad h \downarrow 0.$$

Por definición, $\mathcal{R}_\lambda g \in \mathcal{D}(\mathcal{A})$ y

$$\mathcal{A}\mathsf{R}\lambda g = \lambda\mathsf{R}_\lambda g - g$$

de donde se sigue el resultado.
Paso 2. Demostración (IV). Sea $f \in \mathcal{D}(\mathcal{A})$, entonces for definición $\mathcal{A}f \in \mathcal{S}$. Tenemos entonces

$$\mathsf{R}_\lambda[\mathcal{A}f] = \int_0^\infty e^{-\lambda t}\mathsf{P}_t[\mathcal{A}f]dt = \int_0^\infty e^{-\lambda t}\mathcal{A}[\mathsf{P}_t f]dt = \mathcal{A}\int_0^\infty e^{-\lambda t}\mathsf{P}_t f dt = \mathcal{A}\mathsf{R}_\lambda f.$$

Donde la última igualdad se sigue del hecho que $\mathcal{A}$ es cerrado. □

Ejemplo 7. *Sea $\mathbb{X}$ numerable con $\mathcal{X} = 2^{\mathbb{X}}$. Sea P una matriz estocástica en $\mathbb{X} \times \mathbb{X}$. Definimos la función de transición mediante*

$$P_t(x, y) = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathsf{P}^{(n)}(x, y), \quad x, y \in \mathbb{X}$$

donde $\mathsf{P}^{(n)} = (\mathsf{P})^n$ es la n-ésima potencia de P, y $\mathsf{P}^{(0)} = \mathbb{I}$ es la matriz identidad. Por el Teorema de existencia de Kolmogorov, el proceso canonico en $(\mathbb{X}^{\mathbb{R}_+}, \mathcal{X}^{\mathbb{R}_+})$ con distribución inicial ν, es un proceso de Markov con respecto a su filtración natural.

Esta construcción se sigue de la transformación $X_t = Y_{N_t}$, donde $\{Y_n : n = 0, 1, \ldots\}$ es una cadena de Markov en $(\mathbb{X}, \mathcal{X})$ con matriz de transición P y distribución inicial δ_x, y $(N_t : t \geq 0)$ un proceso Poisson simple. Más formalmente deberíamos escribir

$$X_t = \sum_{n=0}^{\infty} \mathbf{1}_{\{N_t = n\}} Y_n.$$

Por construcción $(X_t : t \geq 0)$ tiene trayectorias continuas por la derecha.

El resolvente en este caso esta dado por

$$\mathsf{R}_\mu f = \frac{1}{\lambda + \mu}\sum_{n\geq 0}\left(\frac{\lambda}{\lambda + \mu}\right)^n \mathbf{P}^n f = ((\lambda + \mu)\mathbb{I} - \lambda\mathbf{P})^{-1} f, \quad f \in \mathcal{S}$$

para $\mu > (c - 1)\lambda$. Nótese que

$$\int_0^\infty t^n e^{-\xi t} dt = \frac{n!}{\xi^{n+1}}$$

Antes de enunciar el resultado de Hille-Yosida, es conveniente enfocarnos en SFO($\mathcal{S}$) de contracción. Como vimos por ser fuerte $||\mathsf{P}_t|| \leq Me^{\alpha t}$. Entonces, $\mathsf{P}_t e^{-\alpha t}$ es un SFO($\mathcal{S}$) con $||\mathsf{P}_t e^{-\alpha t}|| \leq M$. Entonces si definimos una nueva norma $||\cdot||^*$ por

$$||f||^* = \sup_{t\geq 0} ||\mathsf{P}_t e^{-\alpha t} f||$$

entonces $||f|| \leq ||f||^* \leq M||f||$. Por lo tanto $||\cdot||$ y $||\cdot||^*$ son normas equivalentes y $\mathcal{S}$ es Banach con respecto a $||\cdot||^*$. Se sigue de inmediato que $(\mathsf{P}_t e^{-\alpha t}; t \geq 0)$ es un SFO($\mathcal{S}$) con respecto a la nueva norma.

También necesitaremos la noción de ***disipatividad***, El operador lineal $B : V \to \mathcal{S}$, V un subespecie lineal de $\mathcal{S}$, es (α, M)-disipativo si

$$||(\lambda\mathbb{I} - B)f|| \geq \frac{\lambda - \alpha}{M}||f||, \quad \forall f \in V, \lambda > 0.$$

Teorema 8 (Hille-Yosida)**.** *Supóngase que existe un operador lineal $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{S}$, donde $\mathcal{D}(\mathcal{A})$ es un subespacio lineal de $\mathcal{S}$. Entonces $\mathcal{A}$ es el generador de un SFO($\mathcal{S}$) sí y solo si las siguientes tres propiedades se satisfacen*

i) $\overline{\mathcal{D}(\mathcal{A})} = \mathcal{S}$, *es decir $\mathcal{D}(\mathcal{A})$ es denso en $\mathcal{S}$.*

ii) $\mathcal{A}$ *es $(0, 1)$-disipativo. En otras palabras $||\lambda f - \mathcal{A}f|| \geq \lambda||f||$, para toda $f \in \mathcal{D}(\mathcal{A})$ y $\lambda > 0$*

iii) $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A}) = \mathcal{S}$ *para algún $\lambda > 0$.*

Demostración.

Parte de necesidad ($\Rightarrow$): Por el Corolario 3 y el inciso (II) del Teorema 7 se siguen (i) y (iii), por lo que unicamente queda demostrar (ii). Por el Teorema 7 (I) existe $g \in \mathcal{S}$, tal que $f = \mathsf{R}_\lambda g$ y por el inciso (V) del mismo teorema, $(\lambda\mathbb{I} - \mathcal{A})f = g$. Entonces, se tiene

$$||f|| = ||\mathsf{R}_\lambda g|| \leq \frac{1}{\lambda}||g|| = \frac{1}{\lambda}||\lambda f - \mathcal{A}f||$$

de donde despejando λ se sigue (ii).

Parte de suficiencia ($\Leftarrow$): Para esta parte necesitamos algunos lemas, que formularemos en el caso de $\mathcal{A}$ un operador (α, M)-disipativo para después usar el caso particular $(0, 1)$-disipativo.

Lema 5. Sea $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{S}$ un operador lineal (α, M)-disipativo con $\mathcal{D}(A) \subset \mathcal{S}$ un subespacio lineal, y $M \geq 1$, $\alpha \geq 0$. Entonces $\mathcal{A}$ es un operador cerrado si y solo si $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})$ es un conjunto cerrado , para algún .$\lambda > \alpha$

Demostración.

Supóngase que $\mathcal{A}$ es cerrado. Sea $\lambda > \alpha$ y $\{f_n\}_n \subset \mathcal{D}(\mathcal{A})$ una sucesión tal que $(\lambda\mathbb{I} - \mathcal{A})f_n \to h$, $n \to \infty$, para algún $h \in \mathcal{S}$, i.e. $h \in (\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})$. Por la (α, M)-disipatividad de $\mathcal{A}$ tenemos

$$||(\lambda\mathbb{I} - \mathcal{A})(f_n - f_m)|| \geq \frac{\lambda - \alpha}{M}||f_n - f_m||$$

entonces $\{f_n\}_n$ es una sucesión Cauchy en $\mathcal{S}$, y por lo tanto tiene un límite $f \in \mathcal{S}$. Se satisface entonces

$$\mathcal{A}f_n = -(\lambda\mathbb{I} - \mathcal{A})f_n + \lambda f_n \to -h + \lambda f, \quad n \to \infty.$$

Dado que $\mathcal{A}$ es cerrado, $f \in \mathcal{D}(\mathcal{A})$ y $h = (\lambda\mathbb{I} - \mathcal{A})f$. Entonces $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})$ es cerrado.

Ahora supóngase que $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})$ es cerrado en $\mathcal{S}$ para algún $\lambda > \alpha$. Sea $\{f_n\}_n \subset \mathcal{A}$, con $f_n \to f$ y $\mathcal{A}f_n \to g$, entonces

$$(\lambda\mathbb{I} - \mathcal{A})f_n \to \lambda f - g, \quad n \to \infty.$$

por lo tanto $\lambda f - g \in (\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})$. Por lo tanto, existe $h \in \mathcal{D}(\mathcal{A})$, tal que $\lambda h - \mathcal{A}h = \lambda f - g$. Por la disipatividad de $\mathcal{A}$ se tienen que

$$||(\lambda\mathbb{I} - \mathcal{A})(f_n - h)|| = ||\lambda(f_n - h) - \mathcal{A}(f_n - h)|| \geq \frac{\lambda - \alpha}{M}||f_n - h||.$$

Donde la parte del lado izquierdo converge a 0, cuando $n \to \infty$. Por lo tanto, $f_n \to h$ y entonces $f = h \in \mathcal{D}(\mathcal{A})$ y $g = \mathcal{A}f$, lo que demuestra que $\mathcal{A}$ es cerrado. $\square$

Lema 6. Sea $\mathcal{A} : \mathcal{D}(A) \to \mathcal{S}$ un operador (α, M)-disipativo, lineal y cerrado con $\mathcal{D}(\mathcal{A}) \subset \mathcal{S}$ un subespacio lineal, y $M \geq 1, \alpha \geq 0$. Sea

$$\lambda(\mathcal{A}) = \left\{ \lambda > \alpha \;\middle|\; \begin{array}{l} (\lambda\mathbb{I} - \mathcal{A}) \text{ es uno a uno} \\ (\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A}) = \mathcal{S} \\ (\lambda\mathbb{I} - \mathcal{A})^{-1} \text{ es un operador lineal acotado en } \mathcal{S} \end{array} \right\} \tag{2.42}$$

el ***conjunto resolvente***, i.e. el conjunto de las λ's para las cuales $(\lambda\mathbb{I} - \mathcal{A})$ tiene inversa continua, acotada y densa. . Entonces, $\lambda(\mathcal{A}) \neq \emptyset \Rightarrow \lambda(\mathcal{A}) = (\alpha, \infty)$.

Demostración. Basta con demostrar que $\lambda(\mathcal{A})$ es ambos, abierto y cerrado, en (α, ∞). Para demostrar que es abierto, sea $\lambda \in \lambda(A)$ y definamos la bola abierta en (α, ∞) con radio $||(\lambda\mathbb{I} - \mathcal{A})^{-1}||^{-1}$

$$B = \{\mu \in (\alpha, \infty) \mid |\lambda - \mu| < ||(\lambda\mathbb{I} - \mathcal{A})^{-1}||^{-1}\}.$$

Demostraremos que $B \subset \lambda(A)$. Sea $\mu \in B$. Entonces si la expansión de Neumann

$$\hat{\mathsf{R}} = \sum_{n=0}^{\infty}(\lambda - \mu)^n \left((\lambda\mathbb{I} - \mathcal{A})^{-1}\right)^{n+1} : \mathcal{S} \to \mathcal{S},$$

que define un operador lineal y acotado, converge, el operador es invertible (i.e. como sucede para la serie geométrica en $\mathbb{R}$). Afirmamos que $\hat{\mathsf{R}} = (\mu\mathbb{I} - \mathcal{A})^{-1}$, es decir la serie converge. Usemos que

$$\mu\mathbb{I} - \mathcal{A} = \lambda\mathbb{I} - \mathcal{A} + (\mu - \lambda)\mathbb{I}$$

para demostrar que $\hat{\mathsf{R}}(\mu\mathbb{I} - \mathcal{A}) = (\mu\mathbb{I} - \mathcal{A})\hat{\mathsf{R}} = \mathbb{I}$ y por lo tanto $\mu \in \lambda(\mathcal{A})$.

Ahora demostremos que es cerrado en (α, ∞). Con esta finalidad, sea $\{\lambda_n\}_n \subset \lambda(\mathcal{A})$, con $\lambda_n \to \lambda$, $n \to \infty$, para algún $\lambda \in \mathbb{R}$, $\lambda > \alpha$. Tenemos que demostrar que $\lambda \in \lambda(\mathcal{A})$. La demostración consiste de dos pasos

- 1: $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A}) = \mathcal{S}$
- 2: $(\lambda\mathbb{I} - \mathcal{A})$ es uno a uno.

Estos pasos implican que para cada $g \in \mathcal{S}$ existe precisamente un elemento $f \in \mathcal{D}(\mathcal{A})$ con $g = (\lambda\mathbb{I} - \mathcal{A})f$. Ahora, por la (α, M)-disipatividad

$$||(\lambda\mathbb{I} - \mathcal{A})^{-1}g|| = ||f|| \leq \frac{M}{\lambda - \alpha}||\lambda f - \mathcal{A}f|| = \frac{M}{\lambda - \alpha}||g||.$$

Debido a que $g \in \mathcal{S}$ es arbitrario,

$$||(\lambda\mathbb{I} - \mathcal{A})^{-1}|| \leq \frac{M}{\lambda - \alpha} < \infty$$

lo que demuestra que $\lambda \in \lambda(\mathcal{A})$.

Para demostrar el paso 1) arriba. Para cada $g \in \mathcal{S}$ y n, podemos definir $g_n = (\lambda\mathbb{I} - \mathcal{A})(\lambda_n\mathbb{I} - \mathcal{A})^{-1}g$. Entonces[1]

$$\begin{aligned} ||g_n - g|| &= ||(\lambda - \mathcal{A})\mathsf{R}_{\lambda_n}g - (\lambda_n - \mathcal{A})\mathsf{R}_{\lambda_n}g|| = ||(\lambda - \lambda_n)\mathsf{R}_{\lambda_n}g|| \leq |\lambda_n - \lambda|||\mathsf{R}_{\lambda_n}g|| \\ &\leq |\lambda_n - \lambda|\frac{M}{\lambda - \alpha}||g|| \to 0, \quad n \to \infty \end{aligned}$$

Por lo que el rango de $(\lambda_n\mathbb{I} - \mathcal{A})$ es denso en $\mathcal{S}$, i.e.

$$\overline{(\lambda_n\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})} = \mathcal{S}.$$

Debido a que $(\lambda_n\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A})$ es cerrado por el Lema 5, tenemos que $(\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A}) = \mathcal{S}$. El punto 2) se sigue de inmediato por la (α, M)-disipatividad. □

Lema 7. Sea $\mathcal{A} : \mathcal{D}(A) \to \mathcal{S}$ un operador (α, M)-disipativo, lineal y cerrado con $\mathcal{D}(\mathcal{A}) \subset \mathcal{S}$ un subespacio lineal, y $M \geq 1, \alpha \geq 0$. Supóngase que $\overline{\mathcal{D}(\mathcal{A})} = \mathcal{S}$ y $\lambda(\mathcal{A}) = (\alpha, \infty)$. Definimos por

$$\mathcal{A}_\lambda = \lambda\mathcal{A}(\lambda\mathbb{I} - \mathcal{A})^{-1}, \quad \lambda > \alpha$$

la ***aproximación de Yosida***. $\mathcal{A}_\lambda$ tiene las siguientes propiedades

a) $\mathcal{A}_\lambda$ es un operador lineal acotado $\mathcal{S} \to \mathcal{S}$ y
$$e^{t\mathcal{A}_\lambda} = \sum_{n=0}^{\infty} \frac{t^n\mathcal{A}_\lambda^n}{n!}$$
es un SFO($\mathcal{S}$) con generador $\mathcal{A}_\lambda$.

b) $\mathcal{A}_\lambda\mathcal{A}_\mu = \mathcal{A}_\mu\mathcal{A}_\lambda$

c) $||\mathcal{A}f - \mathcal{A}_\lambda f|| \to 0$, $\lambda \to \infty$ para toda $f \in \mathcal{D}(\mathcal{A})$.

[1] Aqui hemos abusado un poco de la notación R_λ, al usarlo para el caso (α, M)-disipativo.

Demostración. Para todo $\lambda > \alpha$ y sea $\mathsf{R}_\lambda = (\lambda\mathbb{I} - \mathcal{A})^{-1}$. En la demostración del lema anterior vimos que $||\mathsf{R}_\lambda|| \leq M/(\lambda-\alpha)$, ademas se tiene que $\mathsf{R}_\lambda\mathsf{R}_\mu = \mathsf{R}_\mu\mathsf{R}_\lambda$. Entonces,

$$\begin{aligned} (\lambda\mathbb{I} - \mathcal{A})\mathsf{R}_\lambda &= \mathbb{I}, \quad \text{en } \mathcal{S} \\ \mathsf{R}_\lambda(\lambda\mathbb{I} - \mathcal{A}) &= \mathbb{I}, \quad \text{en } \mathcal{D}(\mathcal{A}). \end{aligned} \tag{2.43}$$

Demostraremos a), entonces podemos escribir $\mathcal{A}_\lambda = \lambda\mathcal{A}\mathsf{R}_\lambda$, de donde se sigue

$$\mathcal{A}_\lambda = \lambda^2\mathsf{R}_\lambda - \lambda\mathbb{I}, \quad \text{en } \mathcal{S} \tag{2.44}$$

$$= \lambda\mathsf{R}_\lambda\mathcal{A}, \quad \text{en } \mathcal{D}(\mathcal{A}), \tag{2.45}$$

que implica que $\mathcal{A}_\lambda$ es acotado con

$$||e^{t\mathcal{A}_\lambda}|| \leq e^{-t\lambda}e^{t\lambda^2||\mathsf{R}_\lambda||} \leq e^{t\lambda^2 M/(\lambda-\alpha)-t\lambda}. \tag{2.46}$$

Además, para toda $f \in \mathcal{S}$

$$||e^{t\mathcal{A}_\lambda}f - f|| \leq \sum_{n=1}^{\infty}\frac{t^n}{n!}||\mathcal{A}_\lambda||^n||f|| \to 0, \quad t \downarrow 0$$

entonces es un SFO($\mathcal{S}$). De manera análoga se verifica que $\mathcal{A}_\lambda$ es el generador infinitesimal correspondiente.

La demostración de b) se sigue de (2.44) y de $\mathsf{R}_\lambda\mathsf{R}_\mu = \mathsf{R}_\mu\mathsf{R}_\lambda$. Para demostrar c) primero observemos que para toda $f \in \mathcal{S}$, $||\lambda\mathsf{R}_\lambda f - f|| \to 0$, $\lambda \to \infty$ ya que, usando (2.43), tenemos

$$||\lambda\mathsf{R}_\lambda f - f|| = ||\mathsf{R}_\lambda\mathcal{A}f|| \leq \frac{M}{\lambda-\alpha}||\mathcal{A}f|| \to 0, \quad \lambda \to \infty \quad \text{para cada } f \in \mathcal{D}(\mathcal{A}).$$

Sea $f \in \mathcal{S}$ y $\{f_n\}_n \subset \mathcal{D}(\mathcal{A})$ una sucesión que converge a f. Entonces para toda n

$$\limsup_{\lambda\to\infty}||\lambda\mathsf{R}_\lambda f - f|| \leq \limsup_{\lambda\to\infty}\left[||\lambda\mathsf{R}_\lambda f_n - f_n|| + \frac{M\lambda}{\lambda-\alpha}||f_n - f|| + ||f_n - f||\right]$$

El primer término converge a 0, el segundo a $M||f_n - f||$ y el tercero a $||f_n - f||$. Dejando $n \to \infty$, obtenemos que tiende a 0. El inciso c) se sigue de combinar con (2.45). □

Lema 8. Supóngase que B, C son operadores lineales acotados en $\mathcal{S}$, conmutativos, i.e. $BC = CB$, y $||e^{tB}|| \leq 1$ y $||e^{tC}|| \leq 1$. Entonces

$$||e^{tB}f - e^{tC}f|| \leq t||Bf - Cf||, \quad \text{para cada} \quad f \in \mathcal{S} \quad \text{y} \quad t \geq 0$$

Demostración. El resultado se sigue de la identidad

$$e^{tB}f - e^{tC}f = \int_0^t \frac{d}{ds}[e^{sB}e^{(t-s)C}]f ds = \int_0^t e^{sB}(B-C)e^{(t-s)C}f ds = \int_0^\infty e^{sB}e^{(t-s)C}(B-C)f ds$$

donde en la última igualdad se uso la conmutatividad de B y C. □

Regresemos a la demostración de la parte de suficiencia ($\Leftarrow$) del Teorema 8 [Hille-Yosida]. La idea es definir un SFO($\mathcal{S}$), ($\mathsf{P}_t : t \geq 0$) y demostrar que tiene generador $\mathcal{A}$.

Las condiciones (ii), (iii) y Lema 5 implican que $\mathcal{A}$ es cerrado y por los argumentos dados en en Lema 6, $\lambda(\mathcal{A}) \neq \emptyset$ y $\lambda(\mathcal{A}) = (0,\infty)$. 1Con la notación del Lema 7, definamos para cada $\lambda > 0$ el SFO($\mathcal{S}$), $\{e^{t\mathcal{A}_\lambda}\}_t$. Por (2.46) (con $\alpha = 0, M = 1$), $||e^{t\mathcal{A}_\lambda}|| \leq 1$ y por los Lemas 7 (b) y 8

$$||e^{t\mathcal{A}_\lambda}f - e^{t\mathcal{A}_\mu}f|| \leq t||\mathcal{A}_\lambda f - \mathcal{A}_\mu f||, \quad \forall t \geq 0 \text{ y } f \in \mathcal{S}.$$

Por el Lema 7 (b), $\lim_{\lambda\to\infty} e^{t\mathcal{A}_\lambda}$ existe para toda $t \geq 0$ uniformemente en $t \in [0,T]$ para cada $T > 0$ y toda $f \in \mathcal{D}(\mathcal{A})$. Definamos $\mathsf{P}_t f = \lim_{\lambda\to\infty} e^{t\mathcal{A}_\lambda}f, \forall f \in \mathcal{D}(\mathcal{A})$. El hecho de que $\overline{\mathcal{D}(\mathcal{A})} = \mathcal{S}$ permite definir $\mathsf{P}_t f$ sobre todo el espacio $\mathcal{S}$ y toda $t \geq 0$. Usando la identidad

$$\mathsf{P}_{t+s}f - \mathsf{P}_t\mathsf{P}_s f = \mathsf{P}_{t+s}f - e^{(t+s)\mathcal{A}_\lambda}f + e^{s\mathcal{A}_\lambda}(e^{t\mathcal{A}_\lambda}f - \mathsf{P}_t f) + (e^{s\mathcal{A}} - \mathsf{P}_s)\mathsf{P}_t f$$

nos permite concluir que las ecuaciones de Chapman-Kolmogorov se satisfacen y verificar que $(\mathsf{P}_t : t \geq 0)$ es un SFO($\mathcal{S}$) de contracción. Falta demostrar que $\mathcal{A}$ es el generador de $(\mathsf{P}_t : t \geq 0)$, por el Teorema 6 (iii)

$$e^{t\mathcal{A}_\lambda} f - f = \int_0^t e^{s\mathcal{A}_\lambda} \mathcal{A}_\lambda f ds, \quad \forall f \in \mathcal{S}, t \geq 0, \lambda > 0 \tag{2.47}$$

Para cada $f \in \mathcal{D}(\mathcal{A})$ y $t \geq 0$

$$e^{s\mathcal{A}_\lambda} \mathcal{A}_\lambda f - \mathsf{P}_s \mathcal{A} f = e^{s\mathcal{A}_\lambda}(\mathcal{A}_\lambda f - \mathcal{A} f) + (e^{s\mathcal{A}_\lambda} - \mathsf{P}_s)\mathcal{A} f$$

Por Lema 7 c) implica que

$$||e^{s\mathcal{A}_\lambda} \mathcal{A}_\lambda f - \mathsf{P}_s \mathcal{A} f|| \to 0, \quad \lambda \to \infty$$

uniformemente en $s \in [0, t]$. Consequentemente, (2.47) conlleva a

$$\mathsf{P}_t f - f = \int_0^t \mathsf{P}_s \mathcal{A} f ds, \quad \forall f \in \mathcal{D}(\mathcal{A}) \text{ y } t \geq 0.$$

Supóngase que $(\mathsf{P}_t : t \geq 0)$ tiene generador $\mathcal{H}$, con dominio $\mathcal{D}(\mathcal{H})$. Entonces, $\mathcal{D}(\mathcal{H}) \supset \mathcal{D}(\mathcal{A})$ y $\mathcal{H} = \mathcal{A}$, es decir $\mathcal{H}$ es una ***extensión*** de $\mathcal{A}$. Por el Teorema 7, $\lambda\mathbb{I} - \mathcal{H}$ es uno a uno para $\lambda > 0$. Dado que $\mathcal{S} = (\lambda\mathbb{I} - \mathcal{A})\mathcal{D}(\mathcal{A}) = (\lambda\mathbb{I} - \mathcal{H})\mathcal{D}(\mathcal{A})$, $\mathcal{D}(\mathcal{H}) \backslash \mathcal{D}(\mathcal{A}) = \emptyset$ de otra forma contradecimos el hecho de que $\lambda\mathbb{I} - \mathcal{H}$ es uno a uno. □

2.3.1. Procesos de Feller

En general, la Definición 17 incluye una gran clase de procesos y estudiar su ley, e.g. mediante el enfoque funcional arriba mencionado, plantea retos importantes. En lo que sigue del curso, nos enfocaremos en una clase de procesos de Markov que incluye todos los ejemplos que hemos estudiado y que estudiaremos y que se enmarca en un caso donde el Teorema de Hille-Yosida funciona. En particular, nos concentraremos en el espacio de Banach $\mathcal{C}_0(\mathbb{X})$, es decir el espacio de funciones continuas que se desvanecen en el infinito (o en ∂). Para $\mathbb{X}$ asumiremos que es un sub-espacio abierto o cerrado de $\mathbb{R}^d$ o $\mathbb{Z}^d$.

Nota: En la Teoría clásica de procesos de Feller, $\mathcal{C}_0(\mathbb{X})$ es remplazado por $\mathcal{C}_b(\mathbb{X})$, es decir el espacio de funciones continuas y acotadas. Sin embargo, esta elección de espacio de funciones es más adecuada cuando el proceso no tiene explosiones o escapes al infinito y permanece en una región acotada. En contraste, $\mathcal{C}_0(\mathbb{X})$, es útil en problemas donde el proceso puede escapar al infinito o donde el infinito actúa como un estado absorbente. Con algunas condiciones de recurrencia, estos espacios funcionales pueden coincidir. Debido a esto, en el estudio de procesos de difusión, el espacio $\mathcal{C}_0(\mathbb{X})$ es preferido, ya que los procesos de difusión en espacios no acotados pueden escapar al infinito en tiempo finito o infinito.

Definición 21. Decimos que $(\mathsf{P}_t : t \geq 0)$ es un ***semigrupo de Feller*** si

i) [***Propiedad de Feller–continuidad en*** $x \in \mathbb{X}$] $f \in \mathcal{C}_0(\mathbb{X}) \Rightarrow \mathsf{P}_t f \in \mathcal{C}_0(\mathbb{X})$ para toda $t \geq 0$.

ii) [***Continuidad fuerte en*** t] $\mathsf{P}_t f(x) \to f(x), t \downarrow 0$ para cada $f \in \mathcal{C}_0(\mathbb{X})$ y $x \in \mathbb{X}$.

A un proceso de Markov con semigrupo de Feller se le conoce como un ***Proceso de Feller***. Se puede ver que un semigrupo de Feller es un SFO($\mathcal{C}_0(\mathbb{X})$) de contracción, ya que en $\mathcal{C}_0(\mathbb{X})$ la convergencia puntual implica continuidad fuerte.

Ejercicio 18. *Sea* $(\mathsf{P}_t : t \geq 0)$ *un semigrupo como en el Ejercicio 17 con* $\mathbb{X}$ *subespacio de* $\mathbb{R}^d$, $d \in \mathbb{N}_+$. *Es un semigrupo de Feller?*

Una propiedad deseable de los procesos de Markov, es que puedan partir el pasado y futuro en tiempos de paro.

Definición 22. Un proceso X se dice tiene la ***propiedad fuerte de Markov*** si para cada función medible y acotada f, cualquier tiempo de paro σ adaptado, cualquier distribución inicial ν y $t \geq 0$

$$\mathbb{I}_{\sigma<\infty}\mathbb{E}_{\nu}(f(X_{\sigma+t}) \mid \mathcal{F}_{\sigma}) = \mathbb{I}_{\sigma<\infty}\mathbb{E}_{X_{\sigma}}f(X_t), \quad \mathbb{P}_{\nu}c.s. \tag{2.48}$$

Nótese que hemos excluido el evento $\{\sigma = \infty\}$, ya que X_t puede no tener límite cuando $t \to \infty$.

Lema 9. Sea X un proceso de Feller $(\mathbb{X}, \mathcal{X})$-valuado. Entonces (2.48) se satisface para cualquier medida inicial ν y tiempo de paro σ, para el cual existe un subconjunto numerable $S \subset [0, \infty)$ tal que $\sigma \in S \cup \{\infty\}$.

Ejercicio 19. *Demuestra el Lema 9.*

Una aplicación del Teorema de Hille-Yosida establece la conexión entre el generador infinitesimal correspondiente y un semigrupo de Feller. El siguiente teorema establece la existencia de una modificación *càdlàg*

Teorema 9. *Sea $(\mathsf{P}_t : t \geq 0)$ un semigrupo de Feller entonces, para cada $\nu \in \mathcal{P}_{\mathbb{X}}$, existe un proceso (modificación en $(\Omega, \mathcal{F})$) de Markov fuerte con valores en $\mathbb{X}_{\partial}$ y distribución inicial ν con trayectorias càdlàg .*

La idea de la demostración es aplicar el teorema de regularización para submartingalas y después encontrar una clase grande de funciones continuas de procesos de Feller que resulten en submartingalas para finalmente aplicar el teorema de regularización y obtener el resultado. Ver Teorema 2.7 en el libro de Ethier and Kurtz (1985).

Proposición 3. *Sea $\mathcal{A}$ el generador infinitesimal de un proceso de Feller $\mathsf{X} = (X_t; t \geq 0)$. Si $f \in \mathcal{D}(\mathcal{A})$ y el proceso tiene medida inicial ν, entonces el proceso*

$$\mathsf{M}_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{A}f(X_s)ds$$

es una $(\mathcal{F}_t, \mathbb{P}_{\nu})$-martingala.

Ejercicio 20. *Demuestra la Proposición 3. Hint: usa el Teorema1 6.*

Teorema 10. *Sea $(\mathsf{P}_t : t \geq 0)$ un semigrupo de Feller entonces, con función de transición $P_t(x, A)$ que satisface*

$$\lim_{t\to 0} \frac{1}{t} P_t(x, B(x, \epsilon)^c) = 0, \qquad \textit{para cada } x \in \mathbb{X} \textit{ y } \epsilon > 0 \tag{2.49}$$

Entonces el proceso, X, dado por el Teorema 9, satisface $\mathbb{P}(\mathsf{X} \in \mathcal{C}_{\mathbb{X}}[0, \infty)) = 1$, donde $\mathcal{C}_{\mathbb{X}}[0, \infty)$ denota el espacio de funciones continuas $f : [0, \infty) \to \mathbb{X}$. A este proceso se le conoce como un **proceso de Feller continuo** *o* **proceso de difusión**.

Para la demostración ver Proposición 2.9 (p. 171) en el libro de Ethier and Kurtz (1985).

Si asumimos $\mathsf{X} = (X_t; t \geq 0)$ es el proceso resultante del Teorema 10 y, por simplicidad, $\mathbb{X} \subset \mathbb{R}$, la condición (2.49) equivale a

$$\lim_{h\to 0} \frac{1}{h} \mathbb{P}[|X_{t+h} - X_t| > \epsilon \mid X_t = x] = 0,$$

con $\epsilon > 0$. Por la desigualdad de Chebyshev tenemos que dicha condición se satisface si

$$\lim_{h\to 0} \frac{1}{h} \mathbb{E}[|X_{t+h} - X_t|^p \mid X_t = x] = 0 \tag{2.50}$$

para algún $p > 2$. Si esta condición se satisface, la media y varianza infinitesimales,

$$\mu(x) = \lim_{h\to 0} \frac{1}{h}\mathbb{E}[X_{t+h} - X_t \mid X_t = x]$$

y

$$\sigma^2(x) = \lim_{h\to 0} \frac{1}{h}\mathbb{E}[|X_{t+h} - X_t|^2 \mid X_t = x]$$

caracterizan el proceso, es decir su generador infinitesimal esta dado por

$$\mathcal{A}f(x) = \mu(x)f'(x) + \frac{\sigma^2(x)}{2}f''(x), \quad f \in \mathcal{C}^2(\mathbb{R})$$

A $\mu(x)$ se le conoce como el coeficiente de ***deriva*** y a $\sigma(x)$ como el coeficiente de ***difusión***. Se puede ver que este proceso coincide con la solución de una ecuación diferencial estocástica dada por

$$dX_t = \mu(X_t)dt + \sigma(X_t)d\mathsf{B}_t$$

donde $(\mathsf{B}_t; t \geq 0)$ denota un movimiento browniano estándar, i.e. un proceso Lévy con función de transición (2.34).

Ejercicio 21. *Sea X el proceso de Feller inducido por las funciones de transición* (2.16)*. Cumple con la condición* (2.50)*? Justifica tu respuesta.*

Ejercicio 22. *Sea X el proceso de Feller inducido por las funciones de transición* (2.41)*, con $Q = \mathsf{N}(\mu, \tau)$, una distribución gaussiana en $\mathbb{R}$ con $\mu \in \mathbb{R}$ y $\tau > 0$. Cumple con la condición* (2.50)*? Justifica tu respuesta.*

Ejercicio 23. *En el contexto de la construcción en la Sección 2.2. Sea $\pi^{\mathsf{X}}(dx) = \mathsf{Ga}(x; a, b)dx$, es decir una distribución Gamma con media a/b, $a, b > 0$, y $\tilde{\mu}_x(y) = \mathsf{Po}(y; \phi x)$, $\phi > 0$.*

a) Demuestra que $\mu_y(dx) = \mathsf{Ga}(x; y + a, b + \phi)dx$ y exhibe la forma de la función de transición a un paso (2.29)*, en términos de la función modificada de Bessel del primer tipo, e.g. $\mathbf{I}_a(x)$. Hint: utiliza la expresión en series de esta última.*

b) Usando la transformada de Laplace para la distribución $\mathsf{Ga}(a, b)$ y $\mathsf{Po}(\eta)$, encuentra la transformada de Laplace de la función de transición encontrada en el paso anterior.

c) Utilizando el inciso anterior y cambiando el parámetro ϕ por una función $t \mapsto \phi_t$, demuestra que la ecuación de Chapman-Kolmogorov se satisface si y solo si

$$\phi_t = \frac{b}{e^{ct} - 1}, \quad c > 0. \tag{2.51}$$

Hint: utiliza la forma de esta última en términos de su transformada de Laplace, i.e. como en (2.12)*.*

d) (*) Define la función de transición markoviana $P_t(x, A)$ sobre $(\mathbb{R}_+, \sigma(\mathbb{R}_+))$, como la encontrada en el inciso a) con ϕ remplazado por (2.51)*. Es su semigrupo asociado de Feller?*

e) (*) Encuentra el generador infinitesimal del semigrupo del inciso anterior.

f) Si denotamos por $\mathsf{X} := (X_t; t > 0)$ el proceso de Markov regular modulado por la función de transición del inciso anterior y con distribución inicial $\mathsf{Ga}(a, b)$. Se trata de un proceso de difusión? Encuentra sus coeficientes de deriva y difusión infinitesimal, i.e. $\mu(x)$ y $\sigma(x)$.

g) Se trata de un proceso estacionario y reversible? Cual es su distribución estacionaria?

h) Dada la esta construcción, cómo simularías una trayectoria de este proceso, en $[0, T]$?

TEMA 3

ESTABILIDAD DE PROCESOS DE MARKOV

Uno de los principales objetos de estudio dentro de la teoría de procesos de Markov, es el de las propiedades y/o condiciones bajo las cuales un proceso dado es "estable" en algún sentido. Como se mencionó anteriormente, la mayoría de los procesos usados en la practica, e.g. en teoría de series de tiempo, procesos de difusión unidimensionales, métodos numéricos en estadística, exigen o se favorecen de algún aspecto de estabilidad trayectorial.

En el contexto de un proceso de Markov, X, vimos que el término ***ergodicidad*** se refiere a algún tipo de convergencia en el tiempo a una distribución estacionaria, la cual resulta independiente del estado inicial del proceso. Así pues, el término ***estabilidad estocástica*** se puede entender como la sensibilidad de estas distribuciones ante variaciones de las características locales del proceso.

En la Sección 2.2 se presentó una metodología para construir procesos markovianos y reversibles con distribuciones estacionarias dadas, pero arbitrarias. Sin embargo, en el caso cuando uno comienza con un modelo markoviano sin saber si es estacionario y/o sin conocer su distribución estacionaria, surge el cuestionamiento acerca de sus propiedades de estabilidad.

Como antes, dada una función de transición, $P_t(x, A)$, denotamos por

$$\mathsf{P}_t f(x) = \int_{\mathbb{X}} f(y) P_t(x, dy), \quad \text{para toda } x \in \mathbb{X}$$

el operador de transición y también usaremos la notación

$$\mu \mathsf{P}_t(A) = \int_{\mathbb{X}} P_t(x, A)\mu(dx), \quad \text{para toda } A \in \mathcal{X}. \tag{3.1}$$

Con esta notación, decimos que una medida, π, es ***invariante*** si $\pi\mathsf{P}_t(A) = \pi(A)$ para toda $A \in \mathcal{X}$. Cuando $\mathcal{X}$ es numerable, la igualdad anterior establece un sistema de ecuaciones, conocido como las ***ecuaciones de equilibrio***.

Intuitivamente, una distribución estacionaria, π, para X describe el comportamiento de X_t, para t grande. En efecto, supóngase que el proceso X, que comienza en $X_0 \sim \nu$, converge en algún sentido, a π_ν

$$\lim_{t\to\infty} \mathbb{P}_\nu(X_t \in A) = \pi_\nu(A).$$

Nótese que, en principio, dicha distribución podría depender de la distribución inicial. Entonces,

$$\begin{aligned}
\pi_\nu(A) &= \lim_{t\to\infty} \int_{\mathbb{X}} \nu(dx) P_t(x, A) \\
&= \lim_{t\to\infty} \int_{\mathbb{X}} \nu(dx) \int_{\mathbb{X}} P_{t-s}(x, dy) P_s(y, A) \\
&= \int_{\mathbb{X}} \pi_\nu(dy) P_s(y, A) = \pi_\nu \mathsf{P}_t(A).
\end{aligned} \tag{3.2}$$

Es decir la distribución estacionaria constituye una medida de probabilidad invariante.

Un concepto de mucha ayuda cuando se estudia la estabilidad de un proceso de Markov es el de la ***estructura de comunicación*** inducida por P_t sobre el espacio de estados $(\mathbb{X}, \mathcal{X})$. En otras palabras, resulta de interés la estructura de relación $x \longrightarrow A$ sobre $\mathbb{X} \times \mathcal{X}$ dada por

$$x \longrightarrow A \quad \text{si y solo si} \quad P_t(x, A) > 0 \quad \text{para algún } t \geq 0, \tag{3.3}$$

cuando esto ocurre se dice que el conjunto $A \in \mathcal{X}$ es ***accesible*** desde $x \in \mathbb{X}$. La idea detrás de este concepto es entender la forma en que el conjunto A, es " visitado " por el proceso de Markov regulado por $P_t(x, \cdot)$, e.g. si será visitado o no, que tan frecuente, etc.

Definición 23. Dado un proceso de Markov $\mathsf{X} := (X_t; t \geq 0)$ con valores en un espacio Polaco $(\mathbb{X}, \mathcal{X})$, sean

$$\eta_A := \int_0^\infty \mathbf{1}(X_t \in A) dt, \quad \text{y} \quad \tau_A := \inf\{t \geq 0; X_t \in A\} \tag{3.4}$$

el ***tiempo de ocupación*** del proceso X en A y el tiempo a la ***primera visita*** de X a A. El proceso X se dice

i) ***ψ-irreducible*** si existe una medida σ-finita, ψ, sobre $(\mathbb{X}, \mathcal{X})$ tal que

$$\psi(A) > 0 \quad \Rightarrow \quad \mathbb{E}_x(\eta_A) > 0, \quad \text{para toda} \quad x \in \mathbb{X}, A \in \mathcal{X}. \quad . \tag{3.5}$$

A la medida ψ se le identifica como la ***medida de irreducibilidad*** del proceso. En particular, una medida ψ se dice de ***máxima irreducibilidad*** para X si para cualesquiera otra medida irreducible, ϕ, se tiene que $\psi(A) = 0 \Rightarrow \phi(A) = 0$, para toda $A \in \mathcal{X}$.

ii) ***Recurrente*** si es ψ-irreducible y si

$$\psi(A) > 0 \quad \Rightarrow \quad \mathbb{E}_x(\eta_A) = \infty, \quad \text{para toda} \quad x \in \mathbb{X}, A \in \mathcal{X}. \tag{3.6}$$

Si $\mathbb{E}_x(\eta_A) < \infty$ se dice que el proceso es ***transitorio***.

iii) ***Harris recurrente*** si es ψ-irreducible y si

$$\psi(A) > 0 \quad \Rightarrow \quad \mathbb{P}_x(\tau_A < \infty) = 1 \quad \text{para toda} \quad x \in \mathbb{X}, A \in \mathcal{X}. \tag{3.7}$$

Se puede ver que la condición (3.7) es equivalente a $\mathbb{P}_x(\eta_A = \infty) = 1$ y que a su vez implica que el proceso es recurrente, i.e. $\mathbb{E}_x(\eta_A) = \infty$.

Que un proceso sea transitorio, i.e. el tiempo total de ocupación en un conjunto $A \in \mathcal{X}$ sea finito, significa que existe un $t > 0$ para el cual el proceso deja A para nunca regresar. Este $t > 0$, claramente dependerá de la trayectoria del proceso, $\omega(t)$. De otra forma, que el proceso sea recurrente, se traduce a que el conjunto $A \in \mathcal{X}$ se visita una y otra vez. Claramente es de interés identificar los conjuntos para los cuales una trayectoria es transitoria o recurrente.

Se puede ver que si un proceso es Harris recurrente, entonces existe una única medida invariante π. Si dicha medida es finita, i.e. se puede convertir en una medida de probabilidad vía normalización, entonces se dice que el proceso es ***positivo recurrente***. En otro caso se dice que X es un proceso recurrente nulo o transitorio.

Un proceso ψ-irreducible, X, se dice que tiene periodo d si $\mathbb{X}$ se puede particionar en conjuntos disjuntos de ψ-medida cero, $D_1, \ldots, D_d \in \mathcal{X}$ tal que

$$P_t(x, D_{i+1}) = 1, \quad \forall x \in D_i, i = 0, \ldots, d-1 \quad \text{y} \quad P_t(x, D_1) = 1 \quad \text{para } x \in D_d.$$

En particular, X se dice ***periódico*** si $d \geq 2$ y ***aperiódico*** si $d = 1$.

TEMA 4

CADENAS DE MARKOV

Debido a su importancia histórica y su impacto en aplicaciones, es de particular interés enfocarse en procesos de Markov con espacio de estados $\mathbb{X}$ numerable o finito, es decir, el caso de ***cadenas de Markov***. Para enfatizar dicha naturaleza del espacio de estados utilizaremos la notación $\mathbb{X} = \mathbb{S} \subseteq \mathbb{N}_0$. Desde un punto de vista matemático la teoría de cadenas de Markov es ambas, simple y compleja, ya que por un lado permite captar los conceptos esenciales de procesos estocásticos mientras por otro su profundo entendimiento requiere de técnicas sofisticadas. Parte del material que aquí presentamos es tomado de Stroock (2005). Veamos primero el caso de una cadena de Markov a tiempo discreto, i.e. $\mathsf{T} \subseteq \mathbb{Z}$. Es decir nos concentraremos en el proceso estocástico $\mathsf{X} := \{X_n : n \in T\}$ con valores en $(\mathbb{S}, \mathcal{S})$. Así pues, usaremos la notación

$$\mathsf{p}_{ij}^{n,n+1} = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad i, j \in \mathbb{S}$$

para referirnos a la probabilidad de transición a un paso del estado i al estado j al tiempo n. En el caso homogéneo en el tiempo, i.e. independiente de la posición n, utilizaremos la notación simplificada

$$\mathsf{p}_{ij} = \mathsf{p}_{ij}^{0,1}$$

para referirnos a la ***probabilidad de transición a un paso***. Para cadenas de Markov resulta de ayuda expresar estas probabilidades en forma de matrices, i.e. $\mathsf{P} = \{p_{ij}\}_{i,j \in \mathbb{S}}$. En dicho contexto se habla de una ***matriz estocástica*** si

- $\mathsf{p}_{ij} \geq 0$, $i, j \in \mathbb{S}$
- $\sum_{j \in \mathbb{S}} \mathsf{p}_{ij} = 1$, para todo $i \in \mathbb{S}$

De manera análoga, en el contexto de espacio de estados discreto, podemos usar la notación $\boldsymbol{\nu} = \{\nu_i\}_{i \in \mathbb{S}}$ para referirnos al ***vector distribución inicial***. Así pues, las ***distribuciones finito dimensionales*** (en el caso homogéneo en el tiempo) se pueden escribir como

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu_{i_0} \mathsf{p}_{i_0,i_1} \mathsf{p}_{i_1,i_2} \cdots \mathsf{p}_{i_{n-1},i_n}.$$

Para simplificar su futura referencia, abreviaremos por $\mathsf{CM}_{\mathbb{S}}(\boldsymbol{\nu}, \mathsf{P})$ a la cadena de Markov homogénea en el tiempo (discreto), con distribución inicial $\boldsymbol{\nu}$, matriz de transición P y valores en $\mathbb{S}$.

Proposición 4. *Sea* X *una* $\mathsf{CM}_{\mathbb{S}}(\boldsymbol{\nu}, \mathsf{P})$. *Entonces para toda* $n, m \geq 0$

(i) $\mathbb{P}(X_n = j) = (\boldsymbol{\nu} \mathsf{P}^n)_j$

(ii) $\mathbb{P}(X_n = j \mid X_0 = i) = \mathbb{P}(X_{n+m} = j \mid X_m = i) = \mathsf{p}_{ij}^{(n)}$

donde

$$(\boldsymbol{\nu} \mathsf{P})_j = \sum_{i \in \mathbb{S}} \nu_i \mathsf{p}_{ij}, \quad (\mathsf{P}^2)_{ik} = \sum_{j \in \mathbb{S}} \mathsf{p}_{ij} \mathsf{p}_{jk}$$

Demostración: Para (i) basta notar que

$$\mathbb{P}(X_n = j) = \sum_{i_0\in\mathbb{S}} \cdots \sum_{i_{n-1}\in\mathbb{S}} \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = j) \tag{4.1}$$

$$= \sum_{i_0\in\mathbb{S}} \cdots \sum_{i_{n-1}\in\mathbb{S}} \nu_{i_0} \mathsf{p}_{i_0,i_1} \mathsf{p}_{i_1,i_2} \cdots \mathsf{p}_{i_{n-1},j} = (\boldsymbol{\nu}\mathsf{P}^n)_j \tag{4.2}$$

(ii) Por la propiedad de Markov, condicional al evento $\{X_m = i\}$, $\{X_{m+n}, n = 0, 1, \ldots\}$ es Markov modulado por $\mathsf{CM}_{\mathbb{S}}(\boldsymbol{\delta}_i, \mathsf{P})$, entonces basta tomar $\boldsymbol{\nu} = \boldsymbol{\delta}_i$ en (i), es decir el vector que asigna masa uno al estado $i \in \mathbb{S}$.

La propiedad (i) se puede interpretar como que la (pre) multiplicación de un vector renglón por P. Usaremos la notación $(\mathsf{P}^n)_{ij} := \mathsf{p}^n_{ij}$ y $(\boldsymbol{\nu})_i := \nu_i$ cuando se quiera hacer más énfasis desde un punto de vista matricial. En este sentido si f es una función no negativa y acotada, denotamos por f su aplicación vectorial, i.e. $(\mathsf{f})_i = f(\{i\})$. Entonces,

- $\mathbb{E}_\nu(f) = \boldsymbol{\nu}\mathsf{f} = \sum_{i\in\mathbb{S}} f(\{i\})\nu(\{i\})$
- $\mathbb{E}(f(X_n) \mid X_0 = i) = (\mathsf{P}^n\mathsf{f})_i$
- $\mathbb{E}(f(X_n)) = \boldsymbol{\nu}\mathsf{P}^n\mathsf{f}$,

Una forma adecuada de medir la longitud de vectores, e.g. $\boldsymbol{\rho}$, cuando estos representan medidas, es mediante

$$||\boldsymbol{\rho}||_{\mathsf{v}} := \sum_{i\in\mathbb{S}} |\rho_i|$$

lo que corresponde a la ***norma de variación total***. Nótese que bajo esta norma se tiene que $||\boldsymbol{\rho}\mathsf{P}||_{\mathsf{v}} \leq ||\boldsymbol{\rho}||_{\mathsf{v}}$ lo que es diferente que la norma euclideana, e.g. $||\boldsymbol{\rho}||_2 = \left(\sum_{i\in\mathbb{S}} \rho_i^2\right)^{1/2}$. Cuando $\#\mathbb{S} < \infty$, estas normas son comparables mediante $||\boldsymbol{\rho}||_2 \leq ||\boldsymbol{\rho}||_{\mathsf{v}} \leq \sqrt{\#\mathbb{S}}||\boldsymbol{\rho}||_2$. Asimismo, cuando un vector (columna) representa una función, la manera apropiada de medir su longitud es mediante su ***norma uniforme***

$$||\mathsf{f}||_{\mathsf{u}} = \sup_{j\in\mathbb{S}} |f(\{j\})|,$$

ya que $||\boldsymbol{\mu}\mathsf{f}||_{\mathsf{u}} \leq ||\boldsymbol{\mu}||_{\mathsf{v}}\, ||\mathsf{f}||_{\mathsf{u}}$, para una medida μ en $\mathbb{S}$, y por lo tanto $||\mathsf{P}\mathsf{f}||_{\mathsf{u}} \leq ||\mathsf{f}||_{\mathsf{u}}$.

Ejemplo 8. *Sea* $\mathsf{X} := \{X_n\}_{n=0,1,\ldots}$ *una* $\mathsf{CM}_{\{0,1\}}(\boldsymbol{\nu}, \mathsf{P})$. *El caso completamente parametrizado en este caso puede ser representado por*

$$\mathsf{P} = \begin{vmatrix} 1-p & p \\ q & 1-q \end{vmatrix}$$

Una de las primeras preguntas que surgen es ¿qué sucede después de n*-pasos? Por la homogeneidad en el tiempo* $\mathsf{p}^{(n)}_{ij} = \mathbb{P}(X_{n+m} = j \mid X_m = i)$*, entonces usando Chapman-Kolmogorov, i.e.* $\mathsf{P}^{n+1} = \mathsf{P}^n\mathsf{P}$*, obtenemos*

$$\begin{vmatrix} \mathsf{p}^{(n+1)}_{00} & \mathsf{p}^{(n+1)}_{01} \\ \mathsf{p}^{(n+1)}_{10} & \mathsf{p}^{(n+1)}_{11} \end{vmatrix} = \begin{vmatrix} \mathsf{p}^{(n)}_{00} & \mathsf{p}^{(n)}_{01} \\ \mathsf{p}^{(n)}_{10} & \mathsf{p}^{(n)}_{11} \end{vmatrix} \begin{vmatrix} 1-p & p \\ q & 1-q \end{vmatrix}$$

lo que implica

$$\mathsf{p}^{(n+1)}_{00} = (1-p)\, \mathsf{p}^{(n)}_{00} + q\, \mathsf{p}^{(n)}_{01}$$

Ahora, como se tiene que $\mathsf{p}^{(n)}_{00} + \mathsf{p}^{(n)}_{01} = 1$ *se obtiene la relación de recurrencia*

$$\mathsf{p}^{(n+1)}_{00} = (1-p-q)\mathsf{p}^{(n)}_{00} + q, \quad \mathsf{p}^{(0)}_{00} = 1$$

Nótese que, en general, para resolver la relación de recurrencia no-homogénea

$$x_{n+1} = ax_n + b$$

primero buscamos la solución constante, entonces tenemos $x = ax + b$, lo que para $a \neq 1$, conlleva a $x = b/(1-a)$. Ahora, si definimos la ecuación en diferencias homogénea $y_n = x_n - b/(1-a)$ que resuelve $y_{n+1} = ay_n$, entonces $y_n = a^n y_0$ y por lo tanto la solución general cuando $a \neq 1$ esta dada por

$$x_n = Aa^n + b/(1-a),$$

para A constante. Esto implica que

$$\mathsf{p}_{00}^{(n)} = \frac{q}{p+q} + A\,(1-q-p)^n, \quad \mathsf{p}_{00}^{(0)} = 1.$$

Ahora, como

$$\mathsf{p}_{00}^{(0)} = 1 = \frac{q}{p+q} + A,$$

tenemos que, $A = p/(q+p)$ y

$$\mathsf{p}_{00}^{(n)} = \begin{cases} \frac{q}{q+p} + \frac{p}{q+p}\,(1-p-q)^n & \textit{para } q+p > 0 \\ 1 & \textit{para } q+p = 0 \end{cases}$$

Ahora si $|1-p-q| < 1$ tal que $\lim_{n\to\infty}(1-p-q)^n \to 0$. Tenemos que, en general (excluyendo los casos cuando $p = q = 0$ y $p = q = 1$)

$$\lim_{n\to\infty} \mathsf{P}^n = \begin{vmatrix} \frac{q}{q+p} & \frac{p}{q+p} \\ \frac{q}{q+p} & \frac{p}{q+p} \end{vmatrix}$$

Entonces, la distribución límite para esta cadena esta dada por

$$\boldsymbol{\pi} = \left| \frac{q}{q+p}, \frac{p}{q+p} \right|$$

Se puede checar fácilmente que esta también constituye una distribución (vector) invariante, es decir satisface $\boldsymbol{\pi}\mathsf{P} = \boldsymbol{\pi}$.

Ejercicio 24. *Sea $\{Z_n\}_{n=1,\dots}$ una sucesión de variables aleatorias iid con valores en $\mathbb{Z}$, $\mathbb{S}$ un espacio numerable y finito y $f_n : \mathbb{S} \times \mathbb{Z} \to \mathbb{S}$, funciones medibles. Considérese la sucesión de variables aleatorias $\mathsf{X} = \{X_n\}_{n=0,1,\dots}$ definidas mediante la relación de recurrencia*

$$X_{n+1} = f_{n+1}(X_n, Z_{n+1}) \tag{4.3}$$

donde X_0 es una variable aleatoria $\mathbb{S}$-valuada independiente de $\{Z_n\}_{n=1,\dots}$. Demuestra que

i) *X define una cadena de Markov.*

ii) *Cualquier matriz de transición P sobre $\mathbb{S}$ tiene una representación como (4.3).*

Al igual que en el caso general de procesos de Markov, el estudio de los aspectos (i)-(iii) [cf. 7] ocupan una parte considerable en la literatura de cadenas de Markov. Con referencia a (ii) y (iii), dígase la construcción, caracterización y estudio de sus propiedades de estabilidad, la metodología de la Sección 2.2 funciona de igual manera para el caso de cadenas de Markov reversibles y estacionarias, basta con escoger la distribución estacionaria con soporte en $\mathbb{S}$.

Ejercicio 25. *Sea $Y_1, Y_3, Y_5, \dots$ una sucesión de variables aleatorias iid tal que*

$$\mathbb{P}(Y_{2k+1} = -1) = \mathbb{P}(Y_{2k+1} = 1) = \frac{1}{2}, \quad k = 0, 1, 2, \dots$$

y define $Y_{2k} = Y_{2k-1}Y_{2k+1}$ para $k = 1, 2, \dots$

a) *Demuestra que $Y_2, Y_4, \dots$ tiene la misma distribución que $Y_1, Y_3, \dots$*

b) *Demuestra que que $Y_1, Y_2, Y_3, \dots$ son independientes a pares.*

c) Usando el inciso de arriba deduce que

$$(\mathsf{P}^n)_{ij} := \mathbb{P}(Y_{m+n} = j \mid Y_m = i) = \frac{1}{2}, \quad \forall n \text{ y } i,j = \pm 1$$

d) ¿ Se satisfacen las ecuaciones de Chapman-Kolmogorov ?

e) ¿ Se satisface la propiedad de Markov ?

Ejemplo 9. *Supongamos que se quiere construir una cadena de Markov* $\mathsf{X} = \{X_n\}_{n=1}^{\infty}$ *con distribución estacionaria dada por* $\pi^{\mathsf{X}}(x) = \mathsf{Po}(x;\lambda)$*, es decir un proceso que toma valores en* $\mathbb{S} = \mathbb{N}_0$*. Para este fin, y siguiendo la construcción de la Sección 2.2 supongamos que*

$$\tilde{\mu}_x(y) = \mathsf{Bin}(y; x, \xi), \quad 0 < \xi < 1$$

Entonces una simple aplicación del Teorema de Bayes conlleva a

$$\mu_y(x) = \frac{[\lambda(1-\xi)]^{x-y}}{(x-y)!}\, e^{-\lambda(1-\xi)}\mathbb{I}_{[y,\infty)}(x),$$

es decir la distribución correspondiente a la variable aleatoria Poisson desplazada por y*, i.e.* $y + \mathsf{Po}(\lambda(1-\xi))$*. Siguiendo con la construcción del proceso, las probabilidades a un paso están definidas como*

$$\begin{aligned}
\mathsf{p}_{ij} &= \sum_{y=0}^{\infty} \mu_y(j)\tilde{\mu}_i(y) \\
&= \sum_{y=0}^{\infty} \frac{[\lambda(1-\xi)]^{j-y}}{(j-y)!}\, e^{-\lambda(1-\xi)}\mathbb{I}_{[y,\infty)}(j)\binom{i}{y}\xi^y(1-\xi)^{i-y}\,\mathbb{I}_{[0,i]}(y) \\
&= e^{-\lambda(1-\xi)}\,(1-\xi)^{i+j}\,\lambda^j\, i! \sum_{y=0}^{i\wedge j} \frac{\left[\frac{\xi}{(1-\xi)^2\lambda}\right]^y}{(i-y)!(j-y)!y!}, \quad i,j \in \mathbb{N}_0
\end{aligned} \tag{4.4}$$

Si utilizamos la definición de la función hipergeométrica generalizada dada por

$${}_q\mathsf{F}_p(\mathsf{n};\mathsf{d};z) := \sum_{k=0}^{\infty} \frac{z^k}{k!}\frac{\prod_{i=1}^{p}(\mathsf{n}_i)_k}{\prod_{j=1}^{q}(\mathsf{d}_j)_k}$$

donde $(z)_a = \Gamma(z+a)/\Gamma(a)$*, entonces las probabilidades de transición (4.4) se pueden re-expresar como*

$$\mathsf{p}_{ij} = \mathsf{Po}(j;\lambda(1-\xi))\,(1-\xi)^i\,{}_2\mathsf{F}_0\left(-i,-j;\frac{\xi}{(1-\xi)^2\lambda}\right) \tag{4.5}$$

donde, como antes, $\mathsf{Po}(x;\lambda)$ *denota la densidad de masa de una distribución Poisson. Entonces, con* $X_1 \sim \mathsf{Po}(\lambda)$*,* X *es una* $\mathsf{CM}_{\mathbb{N}_0}(\mathsf{Po}(\lambda),\mathsf{P})$*, con la la posición* $\{i,j\}$ *de* P *dada por (4.5).*

Ejercicio 26. *Demostrar que la* $\mathsf{CM}_{\mathbb{N}_0}(\mathsf{Po}(\lambda),\mathsf{P})$ *del Ejemplo 9 tiene la representación*

$$X_{n+1} = \xi \circ X_n + \varepsilon_{n+1}, \quad n = 1,2\ldots \tag{4.6}$$

con $\varepsilon_n \sim \mathsf{Po}(\lambda(1-\xi))$ *independientes de* $(X_n)_{n=1,2,\ldots}$ *y* $\xi \circ X$ *el* ***operador de adelgazamiento binomial****, i.e. dado* X*, se tiene que* $\xi \circ X \sim \mathsf{Bin}(X,\xi)$*.*

Claramente las probabilidades de transición dadas por (4.5) quedan todas caracterizadas por dos parámetros, $\lambda > 0$ y $0 < \xi < 1$. En general, la construcción de una $\mathsf{CM}_{\mathbb{S}}(\nu,\mathsf{P})$, completamente parametrizada requeriría un número grande (o infinito si $\#\mathbb{S} = \infty$) de parámetros en P, lo cual resulta poco manejable.

Por construcción (cf. Ejercicio 9) sabemos que $\mathsf{Po}(x;\lambda)$ es una medida invariante para X en el Ejemplo 9. Sin embargo, dicha observación no resulta evidente de las probabilidades de transición (4.5), e.g. ¿podrías obtener una expresión para la transición a n pasos como en el Ejercicio 8?, para entonces obtener la medida límite y verificar que sea la invariante.

En algunas ocasiones la representación canónica del Ejercicio 24 surge de manera todavía más natural que la del Ejercicio 26. Veamos el siguiente ejemplo.

Ejemplo 10 (***Caminata aleatoria discreta***)**.** *Sea μ una medida de probabilidad concentrada en un conjunto a lo más numerable de $\mathbb{S}$. Si $\mathsf{p}_{ij} = \mu(j-i)$, la $\mathsf{CM}_{\mathbb{S}}(\nu, \mathsf{P})$ resultante se conoce como* **caminata aleatoria** *con distribución de salto μ. Si μ se concentra en vecinos de cero dentro de la reticula $\mathbb{S}$, e.g. $\{-1,1\}$ cuando $\mathbb{S} = \mathbb{Z}$ o $\{(1,0),(0,1),(-1,0),(0,-1)\}$ cuando $\mathbb{S} = \mathbb{Z}^2$, hablamos de una* **caminata aleatoria simple**. *Ésta será* **simétrica** *si a cada vecino se le asigna la misma probabilidad. En particular, para el caso de una caminata aleatoria simple y simétrica en $\mathbb{Z}$, se tiene que $\mathsf{p}_{ij} = \frac{1}{2}\delta_{i-1}(j) + \frac{1}{2}\delta_{i+1}(j)$, $i, j \in \mathbb{Z}$. Entonces, una representación del tipo (4.3) esta dada por*

$$X_n = X_{n-1} + Z_n$$

con $Z_n \sim \mathsf{Ber}_{\{-1,1\}}(\frac{1}{2})$ y donde $\mathsf{Ber}_{\{-1,1\}}(p)$ denota una distribución Bernoulli soportada en $\{-1,1\}$ con probabilidad de éxito p. Claramente, en este caso se puede escribir

$$X_n = X_0 + S_n$$

donde $S_n = \sum_{i=1}^n Z_n$.

Ejercicio 27. *Exhibe la forma explicita de p_{ij} en el caso de una caminata aleatoria simple y simétrica en $\mathbb{Z}^d$ y obtén una representación del tipo (4.3). Encuentra las probabilidad de transición en n-pasos. Se trata de una Martingala?*

Ejemplo 11 (***Cadena de nacimiento y muerte a tiempo discreto***)**.** *Sea X una $\mathsf{CM}_{\mathbb{N}_0}(\nu, \mathsf{P})$, con*

$$\mathsf{p}_{ij} = p_i\delta_{i+1}(j) + q_i\delta_{i-1}(j) + (1 - p_i - q_i)\delta_i(j), \tag{4.7}$$

donde $p_i, q_i \geq 0$ y $p_i + q_i \leq 1$. A p_i se le identifica como la **tasa de nacimiento** *y a q_i como la* **tasa de muerte**. *Para asegurar a cero como condición de frontera, usualmente se requiere que $q_0 = 0$ y $p_0 > 0$. Nótese que en este último caso se tendría que $\mathsf{p}_{00} < 1$.*

Un caso particular es el caso simple de un **modelo de colas**, *es decir donde X_n representa el numero de personas en una fila esperando por algún servicio. Asumimos que las personas llegan (nacen!) con una tasa λ, i.e. $p_i = \lambda$, para toda $i \in \mathbb{N}_0$. Si hay un sólo servidor que atiende personas a tasa μ (mueren!) entonces tenemos $q_i = \mu$, para toda $i \in \mathbb{N}_0$ excepto en el caso $q_0 = 0$. Si hay $k > 1$ servidores y cada uno atiende con tasa μ entonces*

$$q_i = i\mu\,\delta_{[0,k]}(i) + k\mu\,\delta_{[k,\infty)}(i).$$

Claramente repetir el procedimiento del Ejemplo 8 en casos donde la carnalidad de $\mathbb{S}$ sea finita pero grande puede resultar engorroso. Afortunadamente, teoría elemental de matrices nos permite calcular potencias de P de una manera más elegante. Por ejemplo si tenemos una $\mathsf{CM}_{\{1,2,3\}}(\mu, \mathsf{P})$ con

$$\mathsf{P} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{vmatrix}$$

Para encontrar P^n, diagonalizamos P. Calculemos los valores propios de P escribiendo su polinomio característico $0 = \det(\lambda\mathbb{I} - \mathsf{P})$. Entonces

$$\lambda\mathbb{I} - \mathsf{P} = \begin{vmatrix} \lambda - \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \lambda & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \lambda - \frac{1}{2} \end{vmatrix}$$

Entonces

$$\begin{aligned}
\det(\lambda\mathbb{I} - \mathsf{P}) &= \left(\lambda - \frac{1}{2}\right)\det\begin{vmatrix} \lambda & -\frac{1}{2} \\ -\frac{1}{2} & \lambda - \frac{1}{2} \end{vmatrix} + \frac{1}{2}\det\begin{vmatrix} -\frac{1}{2} & 0 \\ -\frac{1}{2} & \lambda - \frac{1}{2} \end{vmatrix} \\
&= \left(\lambda - \frac{1}{2}\right)\left(\lambda - \frac{\lambda}{2} - \frac{1}{4}\right) - \frac{1}{4}\left(\lambda - \frac{1}{2}\right) \\
&= \left(\lambda - \frac{1}{2}\right)(\lambda - 1)\left(\lambda + \frac{1}{2}\right)
\end{aligned}$$

Entonces P tiene tres distintos valores propios

$$\lambda_1 = 1, \quad \lambda_2 = \frac{1}{2}, \quad \lambda_3 = -\frac{1}{2}$$

Entonces P es diagonalizable (lo cual en este caso era ya claro pues P es simétrica). Nótese que 1 es siempre un valor característico para P ya que, para $\eta = (1, \dots, 1)^{\mathsf{T}}$, se tiene $\mathsf{P}\eta = \eta$, pues P es estocástica. Ahora encontremos los vectores característicos correspondientes, ν_1, ν_2, ν_3 con $\mathsf{P}\nu_i = \lambda_i \nu_i$. Lo que se puede hacer resolviendo simultáneamente el sistema de ecuaciones $(\mathsf{P} - \lambda_i \mathbb{I})\nu_i = 0$. En ejemplo, la matriz U con los vectores característicos como columnas está dada por

$$\mathsf{U} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{vmatrix}$$

Entonces $\mathsf{U}^{-1}\mathsf{P}\mathsf{U} = \Lambda$

$$\Lambda := \begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{vmatrix}$$

y por lo tanto $\mathsf{P}^2 = \mathsf{U}\Lambda\mathsf{U}^{-1}\mathsf{U}\Lambda\mathsf{U}^{-1} = \mathsf{U}\Lambda^2\mathsf{U}^{-1}$ y

$$\mathsf{P}^n = \mathsf{U} \begin{vmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & \lambda_3^n \end{vmatrix} \mathsf{U}^{-1}.$$

Para el ejemplo se tiene

$$\mathsf{U}^{-1} = \frac{1}{6} \begin{vmatrix} 2 & 2 & 2 \\ 3 & 0 & -3 \\ 1 & -2 & 1 \end{vmatrix}.$$

Así pues si quisiéramos calcular $\mathbb{P}[X_3 = 2 \mid X_0 = 1]$, solo hay que identificar la entrada correspondiente en la matriz P^3, que en este caso resulta

$$\mathsf{P}^3 = \begin{vmatrix} \frac{3}{8} & \frac{3}{8} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \end{vmatrix}.$$

Si quisiéramos la probabilidad $\mathbb{P}[X_3 = 2]$, entonces necesitaríamos una distribución inicial, e.g. $\mu = (1, 0, 0)$ la cadena comienza en $X_0 = 1$, y usando (i) de la Proposición 4, $\mathbb{P}(X_3 = 2) = (\boldsymbol{\mu}\mathsf{P}^3)_2 = 3/8$.

Dado que, $\lambda_1 = 1$ siempre es un valor característico para P, $\mathsf{traza}(\mathsf{P}) = \sum_i \lambda_i$ y $\mathsf{det}(\mathsf{P}) = \prod_i \lambda_i$ también podríamos proceder observando que $1 + \lambda_2 + \lambda_3 = 1$ lo que implica que $\lambda_2 = -\lambda_3$, y por $\lambda_2\lambda_3 = -1/4$ y por lo tanto $\lambda_2 = 1/2$ y $\lambda_3 = -1/2$. Entonces la matriz es diagonalizable y por lo tanto existen matrices, $\{\mathsf{W_i}\}_{i=1}^N$, en el caso donde $\#\mathbb{S} = N$, con

$$\mathsf{P}^n = \sum_{i=1}^{N} \lambda_i^n \mathsf{W}_i.$$

Nótese, que esto se sabe sin la necesidad de diagonalizar. En nuestro ejemplo tenemos que

$$\mathsf{W}_1 + \mathsf{W}_2 + \mathsf{W}_3 = \mathsf{P}^0 = \mathbb{I}$$

$$\mathsf{W}_1 + \frac{1}{2}\mathsf{W}_2 - \frac{1}{2}\mathsf{W}_3 = \mathsf{P}$$

$$\mathsf{W}_1 + \frac{1}{4}\mathsf{W}_2 + \frac{1}{4}\mathsf{W}_3 = \mathsf{P}^2$$

lo que resolviendo las ecuaciones simultáneas matriciales se reduce a

$$\mathsf{P}^n = \begin{vmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{vmatrix} + \left(\frac{1}{2}\right)^n \begin{vmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{vmatrix} + \left(-\frac{1}{2}\right)^n \begin{vmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{vmatrix} \tag{4.8}$$

Entonces podemos obtener, por ejemplo

$$\mathbb{P}(X_n = 2 \mid X_0 = 0) = (\mathsf{P}^n)_{02} = \frac{1}{3} - \frac{1}{2}\left(\frac{1}{2}\right)^n + \frac{1}{6}\left(-\frac{1}{2}\right)^n$$

.

Ejercicio 28. *Encuentra los valores de* W_i*,* $i = 1, 2, 3$ *y deduce* (4.8).

Ejercicio 29. *Sea* $\mathsf{X} := \{X_n\}_{n=0,1,\ldots}$ *una* $\mathsf{CM}_{\{1,2,3\}}(\nu, \mathsf{P})$ *con*

$$\mathsf{P} = \begin{vmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \end{vmatrix}.$$

Deduce una formula matricial para P^n.

Ejercicio 30. *Usando el enfoque de diagonalización, ó el alternativo usando propiedades de que,* $\lambda_1 = 1$ *siempre es un valor característico para* P*,* $\mathsf{traza}(\mathsf{P}) = \sum_i \lambda_i$ *y* $\det(\mathsf{P}) = \prod_i \lambda_i$*, obtén una forma para* P^n *en el caso del Ejemplo 8. ¿Qué ocurre con la distribución límite cuando* $p = q = 1$*?*

4.1. Método vía funciones generadoras

Este método es de particular interés cuando se quiere encontrar las probabilidades de transición a n pasos para solamente algunos estados $i, j \in \mathbb{S}$, en vez de toda la matriz P^n. Nótese que análogo a la serie geométrica, i.e. $1 + z + z^2 + \cdots = (1-z)^{-1}$ para toda $|z| < 1$, se tiene que para una matriz cuadrada y *finita* A, $\mathbb{I}+\mathsf{A}+\mathsf{A}^2+\cdots = (\mathbb{I}-\mathsf{A})^{-1}$, sí $\lim_{n\to\infty} \mathsf{A}^n = 0$ y $(\mathbb{I} - \mathsf{A})$ es invertible. Este es el caso si todo los valores propios de A tienen modulo < 1. Una matriz estocástica no cumple con esto ya que siempre tenemos que un valor propio es igual a uno. Sin embargo, se tiene el siguiente resultado

Lema 10. Si P es la matriz de transición de una $\mathsf{CM}_{\mathbb{S}}(\nu, \mathsf{P})$, con $\mathbb{S}$ finito entonces todos sus valores propios tienen modulo menor o igual a uno.

Entonces podemos usar la expansión en series de $\theta\mathsf{P}$ para toda $|\theta| < 1$. La idea es usar la expansion de $(\mathbb{I} - \theta\mathsf{P})^{-1}$ como una serie de potencias en θ, alrededor de $\theta = 0$ y encontrar P^n comparando los coeficientes de θ^n.

Ejemplo 12. *Sea $\mathsf{X} := \{X_n\}_{n=0,1,\dots}$ una $\mathsf{CM}_{\{1,2,3\}}(\nu, \mathsf{P})$ con P como en el Ejercicio 29. Recordemos que la inversa de la matriz A esta dada por*

$$(\mathsf{A}^{-1})_{ij} = \frac{(-1)^{i+j}}{\det \mathsf{A}} \det \mathsf{M}_{\mathsf{ji}} \tag{4.9}$$

donde M_{ji} es la matriz A con la renglón j y la columna i removidas.

Entonces

$$(\mathbb{I} - \theta\mathsf{P})^{-1}_{01} = \frac{\theta}{(1-\theta)\left(1+\frac{1}{2}\theta\right)}$$

Para expander esto en series de potencias en θ podemos usar fracciones parciales

$$\frac{\theta}{(1-\theta)\left(1+\frac{1}{2}\theta\right)} = \frac{a}{1-\theta} + \frac{b}{\left(1+\frac{1}{2}\theta\right)}$$

entonces $\theta = a\big(1 + \frac{1}{2}\theta\big) + b(1-\theta)$. Usando para $\theta = 1$ resulta en $a = 2/3$ y para $\theta = 0$ se obtiene que $b = -2/3$. Entonces

$$(\mathbb{I} - \theta\mathsf{P})^{-1}_{01} = \frac{2}{3}(1-\theta)^{-1} - \frac{2}{3}(1-(-\theta/2))^{-1} = \frac{2}{3}\left(1+\theta+\theta^2+\dots\right) - \frac{2}{3}\left(1 - \frac{\theta}{2} + \frac{\theta^2}{4} - \frac{\theta^3}{8} + \right)$$

donde para la última expansión usamos la serie geométrica. Ahora se pueden comparar los coeficientes con los provenientes de la serie de potencias matricial,

$$(\mathbb{I} - \theta\mathsf{P})^{-1}_{01} = \sum_{n=0}^{\infty} \theta^n (\mathsf{P}^n)_{01} = \sum_{n=0}^{\infty} \theta^n \mathsf{p}^{(n)}_{01}$$

y obtener, para toda $n \in \mathbb{N}$,

$$\mathbb{P}[X_n = 1 \mid X_0 = 0] = \mathsf{p}^{(n)}_{01} = \frac{2}{3} - \frac{2}{3}\left(-\frac{1}{2}\right)^n$$

Ejercicio 31. *En el contexto del Ejemplo 12, y utilizando el método vía funciones generadoras, encuentra $\mathsf{p}^{(n)}_{00}$.*

4.2. Estabilidad de cadenas de Markov a tiempo discreto

Utilizando un notación análoga a la usada en la Definición 23, denotaremos por

$$\eta_i := \sum_{n=0}^{\infty} \mathbf{1}(X_n = i), \quad \text{y} \quad \tau_i := \inf\{n \geq 0; X_n = i\}. \tag{4.10}$$

El tiempo que el proceso gasta en el estado i y el tiempo de primer regreso a i, respectivamente.

Asimismo, en un sentido análogo a (3.3), decimos que los estados $i, j \in \mathbb{S}$ se comunican ($i \longleftrightarrow j$) si son accesibles entre ellos, i.e. $i \longrightarrow j$ y $j \longrightarrow i$. Entonces, $\longleftrightarrow$ satisface las condiciones de una relación de equivalencia en $\mathbb{S}$, particionando este último en ***clases de comunicación***. Se dice que una clase, C, es ***cerrada*** si $i \in C$, $i \longrightarrow j$ implica $j \in C$. Entonces, una clase cerrada es aquella de la que no se puede escapar. Un estado $i \in \mathbb{S}$ se dice ***absorbente*** si $\{i\}$ forma una clase cerrada. Si existen clases cerradas $\{C_k\}_{k\in\mathbb{J}}$, la restricción de P a cada C_k es markoviana, además si cada una tiene como medida invariante π_k, entonces cualquier combinación lineal $\sum_{k\in\mathbb{J}} c_k\pi_k$ es una medida invariante sobre $\mathbb{S}$. Claramente si cada $||\boldsymbol{\pi}_k||_{\mathsf{v}} = 1$, entonces se habla de distribuciones invariantes.

Por lo tanto, cualquier resultado de unicidad de medidas (o distribuciones) invariantes excluye este tipo de situación. Esto exige que $\mathbb{S}$ constituya una sola clase de comunicación, es decir que la matriz de transición P sea ***irreducible***. En el contexto de (3.5) se requiere que, para todo $i, j \in \mathbb{S}$,

$$\mathbb{E}_i[\eta_j] = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j) > 0$$

lo que se satisface si, para todo $i, j \in \mathbb{S}$, existe algún $n := n_{ij} \geq 1$, tal que $\mathsf{p}_{ij}^n > 0$.

Ejercicio 32. *Sea X una $\mathsf{CM}_{\mathbb{S}}(\nu, \mathsf{P})$ con P dada por (4.7), es decir una caminata aleatoria en $\mathbb{S}$. Demuestra que P, y por lo tanto X, es irreducible*

- *en $\mathbb{S} = \mathbb{N}_0$ si y solo si $p_i, q_{i+1} > 0$ para toda $i \geq 0$ y $q_0 = 0$.*
- *en $\mathbb{S} = \{0, 1, \ldots, T\}$, con $T \in \{1, 2, \ldots\}$ si y solo si $p_i > 0, i \in \{0, \ldots, T-1\}$, $p_T = q_0 = 0$ y $q_i > 0$ para $i \in \{1, \ldots, T\}$*
- *en $\mathbb{S} = \mathbb{Z}$ si y solo si $p_i, q_i > 0$, para todo $i \in \mathbb{Z}$.*

Al igual que en el caso de espacio de estados general y a tiempo continuo, usando la estructura de comunicación que induce P se pueden estudiar las propiedades de recurrencia y transitividad de una cadena de Markov. De manera natural cuando P es irreducible sobre un espacio finito $\mathbb{S}$, tiene sentido pensar que la distribución inicial tiende a olvidarse, i.e. $\boldsymbol{\nu}\mathsf{P}^n$ se vuelve independiente de $\boldsymbol{\nu}$ para n suficientemente grande. Esto significa que, $\boldsymbol{\nu}\mathsf{P}^n = (\boldsymbol{\nu}\mathsf{P}^{n-m})\mathsf{P}^m$ es casi $\boldsymbol{\nu}\mathsf{P}^m$, para m grande y por el criterio de convergencia de Cauchy, $\boldsymbol{\pi} = \lim_{n\to\infty} \boldsymbol{\nu}\mathsf{P}^n$ existe. Claramente, si este es el caso $\boldsymbol{\pi} = \lim_{n\to\infty} \boldsymbol{\nu}\mathsf{P}^{n+1} = \lim_{n\to\infty}(\boldsymbol{\nu}\mathsf{P}^n)\mathsf{P} = \boldsymbol{\pi}\mathsf{P}$, como en (3.2). Es decir, $\boldsymbol{\pi}$ es el vector característico con valor propio 1.

Aunque, la existencia de una distribución invariante toma una forma más intuitiva en el caso de $\mathbb{S}$ finito, hay situaciones en las que la existencia de una única distribución invariante se extiende al caso de $\mathbb{S}$ infinito. El siguiente teorema establece una condición bajo la cual dicha existencia de $\boldsymbol{\pi}$ se da.

Teorema 11 (Döblin 1)**.** *Sea P una matriz de transiciones sobre $\mathbb{S}$, con la propiedad de que, para algún estado $j_0 \in \mathbb{S}$ y $\epsilon > 0$, $p_{ij_0} \geq \epsilon$ para toda $i \in \mathbb{S}$. Entonces P tiene un vector estacionario único $\boldsymbol{\pi}$, $\boldsymbol{\pi}_{j_0} \geq \epsilon$, y, para toda distribución inicial $\boldsymbol{\nu}$,*

$$||\boldsymbol{\nu}\mathbf{P}^n - \pi||_v \leq 2(1-\epsilon)^n, \quad n \geq 0$$

Digamos, todos los valores característicos de P diferentes de 1 tienen valor absoluto dominado por $(1-\epsilon)$.

Condición 1 (***Condición de Döblin***)**.** Existe un estado que se puede accesar desde cualquier otro con una tasa uniformemente rápida.

Demostración. Si $\boldsymbol{\rho} \in \mathbb{R}^{\mathbb{S}}$ es un vector renglón t.q. $||\boldsymbol{\rho}||_{\mathsf{v}} < \infty$ entonces

$$\begin{aligned} \sum_{j\in\mathbb{S}} (\boldsymbol{\rho}\mathsf{P})_j &= \sum_{j\in\mathbb{S}} \left(\sum_{i\in\mathbb{S}} (\boldsymbol{\rho})_i (\mathsf{P})_{ij} \right) = \sum_{i\in\mathbb{S}} \left(\sum_{j\in\mathbb{S}} (\boldsymbol{\rho})_i (\mathsf{P})_{ij} \right) = \sum_{j\in\mathbb{S}} (\boldsymbol{\rho})_i \\ \sum_{i\in\mathbb{S}} (\boldsymbol{\rho})_i = 0 \quad &\Rightarrow \quad ||\boldsymbol{\rho}\mathsf{P}^n||_{\mathsf{v}} \leq (1-\epsilon)^n ||\boldsymbol{\rho}||_{\mathsf{v}}, \ \text{ para } n \geq 1. \end{aligned} \tag{4.11}$$

Donde, este último resultado basta checarlo para $n = 1$, y por un argumento de inducción, verificar que

$$\begin{aligned}
|(\boldsymbol{\rho}\mathsf{P})_j| &= \left|\sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i(\mathsf{P})_{ij}\right| \\
&= \left|\sum_{i\in\mathbb{S}}(\boldsymbol{\rho})_i((\mathsf{P})_{ij}-\epsilon\delta_{j,j_0})\right| \\
&\le \sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i|((\mathsf{P})_{ij}-\epsilon\delta_{j,j_0}),
\end{aligned}$$

y por lo tanto

$$\begin{aligned}
||\rho\mathsf{P}||_v &\le \sum_{j\in\mathbb{S}}\left(\sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i|((\mathsf{P})_{ij}-\epsilon\delta_{j,j_0})\right) \\
&= \sum_{i\in\mathbb{S}}|(\boldsymbol{\rho})_i|\left(\sum_{j\in\mathbb{S}}((\mathsf{P})_{ij}-\epsilon\delta_{j,j_0})\right) \\
&= (1-\epsilon)||\boldsymbol{\rho}||_v.
\end{aligned}$$

Ahora, si $\boldsymbol{\mu}$ es un vector de probabilidad y definimos $\boldsymbol{\mu}_n=\boldsymbol{\mu}\mathsf{P}^n$. Entonces $\boldsymbol{\mu}_n=\boldsymbol{\mu}_{n-m}\mathsf{P}^m$ and $\sum_{i\in\mathbb{S}}((\boldsymbol{\mu}_{n-m})_i-(\boldsymbol{\mu})_i)=0$, y por lo tanto

$$||\boldsymbol{\mu}_n-\boldsymbol{\mu}_m||_{\mathsf{v}}\le(1-\epsilon)^m||\boldsymbol{\mu}_{n-m}-\boldsymbol{\mu}||_{\mathsf{v}}\le 2(1-\epsilon)^m,\qquad 1\le m<n$$

Entonces, $\{\boldsymbol{\mu}_n\}_{n=1}^{\infty}$ es Cauchy convergente, y entonces existe $\boldsymbol{\pi}$ tal que $||\boldsymbol{\mu}_n-\boldsymbol{\pi}||_{\mathsf{v}}\to 0$.

Por lo tanto $\boldsymbol{\pi}=\lim_{n\to\infty}\boldsymbol{\mu}\mathsf{P}^{n+1}=\lim_{n\to\infty}(\boldsymbol{\mu}\mathsf{P}^n)\mathbf{P}=\boldsymbol{\pi}\mathsf{P}$ (que es de probabilidad pues $\boldsymbol{\mu}_n$ lo es) y entonces $\boldsymbol{\pi}$ es estacionario. En particular,

$$(\boldsymbol{\pi})_{j_0}=\sum_{i\in\mathbb{S}}(\boldsymbol{\pi})_i(\mathsf{P})_{ij_0}\ge\epsilon\sum_{i\in\mathbb{S}}(\boldsymbol{\pi})_i=\epsilon$$

Finalmente, si $\boldsymbol{\nu}$ es un vector de probabilidad, entonces

$$||\boldsymbol{\nu}\mathsf{P}^m-\boldsymbol{\pi}||_{\mathsf{v}}=||(\boldsymbol{\nu}-\boldsymbol{\pi})\mathbf{P}^m||_{\mathsf{v}}\le 2(1-\epsilon)^m,$$

lo que prueba la afirmación y la unicidad de $\boldsymbol{\pi}$ como el único vector estacionario para P. □

Si $\#\mathbb{S}<\infty$ entonces el espectro (el complemento del conjunto resolvente (2.42) en el contexto general de operadores visto en la Sección 2.3) de una matriz de transiciones irreducible, P, se constituye de los valores característicos y se puede aplicar el Teorema de Perron-Frobenius para asegurar la existencia de una única medida invariante, $\boldsymbol{\mu}$ positiva. Sin embargo, dicho enfoque no podría predecir que dicha distribución invariante tuviese entradas no-negativas como sucede con el Teorema de Döblin.

Como una extensión del Teorema 11 se tiene que para cualesquiera $M\ge 1$ y $\epsilon>0$

$$\sup_j\inf_i(\mathbf{P}^M)_{ij}\ge\epsilon\Rightarrow||\boldsymbol{\nu}\mathbf{P}^n-\boldsymbol{\pi}||_{\mathsf{v}}\le 2(1-\epsilon)^{[\frac{n}{M}]}\qquad\text{(Döblin 2)}\tag{4.12}$$

para todo vector de probabilidad $\boldsymbol{\nu}$ y vector estacionario único π. Esto, ya que si $\boldsymbol{\pi}$ denota la distribución estacionaria de P^M, que existe por el Teorema 11, entonces para cualquier vector de probabilidad $\boldsymbol{\mu}$, $m\in\mathbb{N}$ y $0\le r\le M$

$$\left|\left|\mu\mathsf{P}^{mM+r}-\boldsymbol{\pi}\right|\right|_{\mathsf{v}}=\left|\left|(\mu\mathsf{P}^r-\boldsymbol{\pi})\mathsf{P}^{mM}\right|\right|_{\mathsf{v}}\le 2(1-\epsilon)^m.$$

De donde se sigue el resultado, con la unicidad de $\boldsymbol{\pi}$ dada como en el Teorema 11. Aunque la condición de Döblin no exige que $\mathbb{S}$ sea finito, esta rara vez se cumple para $\mathbb{S}$ infinito. Inclusive, podría no satisfacerse para $\#\mathbb{S}<\infty$. Por ejemplo, si X es una $\mathsf{CM}_{\{1,2\}}(\nu,\mathsf{P})$ con

$$\mathsf{P}=\left|\begin{array}{cc}0&1\\1&0\end{array}\right|,$$

se tiene que la cadena va de un estado a otro en un paso, sin embargo $(\mathsf{P}^n)_{ij} = 0$ si $i = j$ y n es impar ó si $i \neq j$ y n es par y por lo tanto no se cumple la condición de Döblin (4.12) ni su conclusión. En efecto, se puede checar que aunque $(1/2, 1/2)$ es el único vector estacionario para P

$$\left|\left|(1,0)\mathsf{P}^n - \left(\frac{1}{2}, \frac{1}{2}\right)\right|\right|_{\mathsf{v}} = 1, \text{para toda } n \geq 0$$

Obviamente el problema aquí es que $\mathbb{S}$ está constituido por más de una clase de comunicación, $(\mathsf{P}^n)_{11} > 0$ sólo si n es par. A pesar de esto, es razonable pensar que esta cadena se estabiliza en algún sentido.

Definamos $\mathsf{A}_n := \frac{1}{n}\sum_{m=0}^{n-1} \mathsf{P}^m$ como la matriz de transición que mide el promedio de tiempo que la cadena gasta en ciertos estados. Es decir, por

$$(\mathsf{A}_n)_{ij} = \frac{1}{n}\sum_{m=0}^{n-1} \mathbb{P}(X_m = j \mid X_0 = i) = \mathbb{E}\left[\frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}(X_m = j) \mid X_0 = i\right],$$

se entiende como el valor esperado que la cadena gasta en j durante $[0, n-1]$, dado que fue accesado desde i. Usando esta matriz de promedios en vez de P en el ejemplo anterior, vemos que para cualquier vector de probabilidad $\boldsymbol{\mu}$

$$\left|\left|\boldsymbol{\mu}\mathsf{A}_n - \left(\frac{1}{2}, \frac{1}{2}\right)\right|\right|_{\mathsf{v}} \leq \frac{1}{n}, \quad \text{para} \quad n \geq 1$$

lo cual claramente se "estabiliza" en el sentido de la norma de variación total.

Teorema 12 (Döblin 3). *Supongamos que* P *es una matriz de transición sobre* $\mathbb{S}$. *Si para algún* $M \in \mathbb{N}$, $j_0 \in \mathbb{S}$ *y* $\epsilon > 0$, $(\mathsf{A}_M)_{ij_0} \geq \epsilon$, *para toda* $i \in \mathbb{S}$, *entonces existe un único vector estacionario* $\boldsymbol{\pi}$ *para* P, *con* $(\boldsymbol{\pi})_{j_0} \geq \epsilon$, *y*

$$||\boldsymbol{\mu}\mathsf{A}_n - \boldsymbol{\pi}||_{\mathsf{v}} \leq \frac{M-1}{n\epsilon},$$

para cualquier vector de probabilidad $\boldsymbol{\mu}$.

Demostración. Sea $\boldsymbol{\pi}$ el único vector estacionario, garantizado por (Döblin 2) para A_M. Entonces, debido a que cualquier vector $\boldsymbol{\mu}$ que es estacionario para P es estacionario para A_M, es claro que $\boldsymbol{\pi}$ es el único candidato para P-estacionariedad. Observemos que

$$(\boldsymbol{\pi}\mathsf{P})\mathsf{A}_M = (\boldsymbol{\pi}\mathsf{A}_M)\mathsf{P} = \boldsymbol{\pi}\mathsf{P}$$

Por lo tanto, $\boldsymbol{\pi}\mathsf{P}$ es estacionaria para A_M y por lo tanto, por unicidad, igual a $\boldsymbol{\pi}$. Es decir $\boldsymbol{\pi} = \boldsymbol{\pi}\mathsf{P}$. Ahora, observemos que para cualquier vector de probabilidad $\boldsymbol{\mu}$

$$\begin{aligned} ||\boldsymbol{\mu}\mathsf{A}_n\mathsf{A}_m - \boldsymbol{\mu}\mathsf{A}_n||_{\mathsf{v}} &= \frac{1}{m}\left|\left|\sum_{k=0}^{m-1}(\boldsymbol{\mu}\mathsf{A}_n\mathsf{P}^k - \boldsymbol{\mu}\mathsf{A}_n)\right|\right|_{\mathsf{v}} \\ &\leq \frac{1}{m}\sum_{k=0}^{m-1}\left|\left|\boldsymbol{\mu}\mathsf{A}_n\mathsf{P}^k - \boldsymbol{\mu}\mathsf{A}_n\right|\right|_{\mathsf{v}} \end{aligned}$$

y para cada $k \geq 0$

$$\boldsymbol{\mu}\mathsf{A}_n\mathsf{P}^k - \boldsymbol{\mu}\mathsf{A}_n = \frac{1}{n}\sum_{l=0}^{n-1}(\boldsymbol{\mu}\mathsf{P}^{l+k} - \boldsymbol{\mu}\mathsf{P}^l) = \frac{1}{n}\left(\sum_{l=k}^{n+k-1}\boldsymbol{\mu}\mathsf{P}^l - \sum_{l=0}^{n-1}\boldsymbol{\mu}\mathsf{P}^l\right)$$

y por lo tanto $\left|\left|\boldsymbol{\mu}\mathsf{A}_n\mathsf{P}^k - \boldsymbol{\mu}\mathsf{A}_n\right|\right|_{\mathsf{v}} \leq \frac{2k}{n}$.

Entonces

$$||\boldsymbol{\mu}\mathsf{A}_n\mathsf{A}_m - \boldsymbol{\mu}\mathsf{A}_n||_{\mathsf{v}} \leq \frac{2}{mn}\sum_{k=0}^{m-1} k = \frac{m-1}{n} \tag{4.13}$$

Ahora, supongamos que $(\mathsf{A}_M)_{ij_0} \geq \epsilon$ para toda i, y sea $\boldsymbol{\pi}$ el único vector estacionario para P. Entonces, $\boldsymbol{\pi}$ el único vector estacionario para A_M y por (4.11) aplicado a A_M

$$||\boldsymbol{\mu}\mathsf{A}_n\mathsf{A}_M - \boldsymbol{\pi}||_{\mathsf{v}} = ||(\boldsymbol{\mu}\mathsf{A}_n - \boldsymbol{\pi})\mathsf{A}_M||_{\mathsf{v}} \leq (1-\epsilon)\,||\boldsymbol{\mu}\mathsf{A}_n - \boldsymbol{\pi}||_{\mathsf{v}}$$

que en conjunción con (4.13) resulta en

$$\begin{aligned} ||\boldsymbol{\mu}\mathsf{A}_n - \boldsymbol{\pi}||_{\mathsf{v}} &\leq ||\boldsymbol{\mu}\mathsf{A}_n - \boldsymbol{\mu}\mathsf{A}_n\mathsf{A}_M||_{\mathsf{v}} + ||\boldsymbol{\mu}\mathsf{A}_n\mathsf{A}_M - \boldsymbol{\pi}||_{\mathsf{v}} \\ &\leq \frac{M-1}{n} + (1-\epsilon)\,||\boldsymbol{\mu}\mathsf{A}_n - \boldsymbol{\pi}||_{\mathsf{v}}, \end{aligned}$$

De donde se sigue el resultado. □

Entonces, bajo ciertas condiciones, $\boldsymbol{\mu}\mathsf{P}^n$ o $\boldsymbol{\mu}\mathsf{A}_n$, convergen y el límite, $\boldsymbol{\pi}$, satisface $\boldsymbol{\pi} = \boldsymbol{\pi}\mathsf{P}$ y es único. Ahora si recordamos que $||\boldsymbol{\mu}\mathsf{f}||_{\mathsf{u}} \leq ||\boldsymbol{\mu}||_{\mathsf{v}}\,||\mathsf{f}||_{\mathsf{u}}$, entonces

$$\sup_j \inf_i (\mathsf{P}^M)_{ij} \geq \epsilon \Rightarrow ||\mathsf{P}^n\mathbf{f} - \boldsymbol{\pi}\mathbf{f}||_{\mathsf{u}} \leq 2(1-\epsilon)^{[\frac{n}{M}]}||\mathbf{f}||_{\mathsf{u}}$$

y

$$\sup_j \inf_i (\mathsf{A}_M)_{ij} \geq \epsilon \Rightarrow ||\mathsf{A}_n\mathbf{f} - \boldsymbol{\pi}\mathbf{f}||_{\mathsf{u}} \leq \frac{M-1}{n\epsilon}||\mathbf{f}||_{\mathsf{u}} \tag{4.14}$$

cuando $\mathbf{f}$ es acotado. Entonces, si X denota la $\mathsf{CM}_{\mathbb{S}}(\boldsymbol{\nu},\mathsf{P})$ y denotamos por

$$\bar{\eta}_j^{(n)} := \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{1}(X_m = j) \tag{4.15}$$

el ***tiempo promedio de ocupación en j antes de n***, se tiene que $(\boldsymbol{\nu}\mathsf{A}_n)_j = \mathbb{E}[\bar{\eta}_j^{(n)}]$, con $\boldsymbol{\nu}$ la distribución inicial. Así pues, el Teorema 12, cuando válido, implica que

$$\lim_{n\to\infty} \mathbb{E}[\bar{\eta}_j^{(n)}] = (\boldsymbol{\pi})_j.$$

Una pregunta que surge de inmediato es si esta convergencia ocurre de una forma más fuerte, e.g. en un sentido casi seguro, es decir $\bar{\eta}_j^{(n)} \to (\boldsymbol{\pi})_j$. En este sentido se tiene el siguiente resultado.

Teorema 13 (Teorema ergódico)**.** *Bajo las mismas hipótesis del Teorema 12 (Döblin 3)*

$$\sup_{j\in\mathbb{S}} \mathsf{E}\left[(\bar{\eta}_j^{(n)} - (\boldsymbol{\pi})_j)^2\right] \leq \frac{2(M-1)}{n\epsilon} \quad \textit{para toda} \quad n \geq 1$$

Más generalmente, para cualquier función acotada f sobre $\mathbb{S}$ y $n \geq 1$

$$\mathsf{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\mathbf{f}\right)^2\right] \leq \frac{2(M-1)||\mathbf{f}||_{\mathsf{u}}^2}{n\epsilon}.$$

Demostración. Sea $\bar{\mathbf{f}}$ un vector columna determinado por la función $\bar{f} = f - \boldsymbol{\pi}\mathbf{f}$. Dado que

$$\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\mathbf{f} = \frac{1}{n}\sum_{m=0}^{n-1} \bar{f}(X_m)$$

y por lo tanto

$$\begin{aligned} \left(\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\mathbf{f}\right)^2 &= \frac{1}{n^2}\left(\sum_{m=0}^{n-1} \bar{f}(X_m)\right)^2 = \frac{1}{n^2}\sum_{k,l=0}^{n-1} \bar{f}(X_k)\bar{f}(X_l) \\ &= \frac{2}{n^2}\sum_{0\leq k\leq l<n} \bar{f}(X_k)\bar{f}(X_l) - \frac{1}{n^2}\sum_{k=0}^{n-1} \bar{f}(X_k) \\ &\leq \frac{2}{n^2}\sum_{0\leq k\leq l<n} \bar{f}(X_k)\bar{f}(X_l) \end{aligned} \tag{4.16}$$

Entonces,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) - \boldsymbol{\pi}\mathbf{f}\right)^2\right] &\leq \frac{2}{n^2}\sum_{k=0}^{n-1}\mathbb{E}\left[\bar{f}(X_k)\sum_{l=0}^{n-k-1}\bar{f}(X_{k+l})\right] \\ &= \frac{2}{n^2}\sum_{k=0}^{n-1}\mathbb{E}\left[\bar{f}(X_k)\sum_{l=0}^{n-k-1}(\mathsf{P}^l\bar{\mathbf{f}})_{X_k}\right] \\ &= \frac{2}{n^2}\sum_{k=0}^{n-1}(n-k)\mathsf{E}\left[\bar{f}(X_k)(\mathsf{A}_{n-k}\bar{\mathbf{f}})_{X_k}\right]\end{aligned}$$

Pero, por (4.14), $||\mathsf{A}_{n-k}\bar{\mathbf{f}}||_{\mathsf{u}} \leq \frac{M-1}{(n-k)\epsilon}||\bar{\mathbf{f}}||_{\mathsf{u}}$, y dado que $||\bar{\mathbf{f}}||_{\mathsf{u}} \leq ||\mathbf{f}||_{\mathsf{u}}$

$$(n-k)\mathsf{E}\left[\bar{f}(X_k)(\mathsf{A}_{n-k}\bar{\mathbf{f}})_{X_k}\right] \leq \frac{(M-1)||\mathbf{f}||_{\mathsf{u}}^2}{\epsilon}$$

Que sustituyendo arriba obtenemos el segundo enunciado del teorema, el primero se reduce a $f = \mathbf{1}_{\{j\}}$, en cuyo caso $||\mathbf{f}||_{\mathsf{u}} \leq 1$. □

Adicional a los tiempos de primer regreso a un estado dado (4.10), cuando se estudia las propiedades de recurrencia de una cadena, es de utilidad considerar el ***tiempo del*** m***-ésimo regreso a*** j dado por

$$\begin{aligned}\rho_j^{(m)} &= \infty \qquad \text{si} \qquad \rho_j^{(m-1)} := \infty \text{ or } X_n \neq j \text{ para } n > \rho_j^{(m-1)} \\ \rho_j^{(m)} &= \inf\{n > \rho_j^{(m-1)} : X_n = j\} \qquad\qquad \text{e.o.c.}\end{aligned}$$

con $\rho_j^{(0)} := 0$. En esta notación se dice que el estado j es recurrente si $\mathbb{P}_j(\rho_j^{(1)} < \infty) = 1$ ó transitorio si no. Como veremos cuando j es recurrente las trayectorias del proceso se forman por ***épocas*** cortadas por los regresos a j. Claramente, $\rho_j := \rho_j^{(1)}$, denota el ***tiempo de la primera llegada*** a j.

Nótese que $\rho_j \geq 1$ y para $n \geq 1$

$$\mathbf{1}_{(n,\infty]}(\rho_j) = F_{n,j}(X_0, \ldots, X_n)$$

donde

$$F_{n,j}(i_0, \ldots, i_n) := \begin{cases} 1 & \text{si } i_m \neq j \text{ for } 1 \leq m \leq n \\ 0 & \text{e.o.c} \end{cases}$$

lo que muestra que el evento $\{\rho_j > n\}$ es una función medible de $(X_0, \ldots, X_n)$ y por lo tanto un tiempo de paro. Igualmente, debido a que

$$\mathbf{1}_{(n,\infty]}(\rho_j^{(m+1)}) = \mathbf{1}_{[n,\infty]}(\rho_j^{(m)}) + \sum_{l=1}^{n-1}\mathbf{1}_{\{l\}}(\rho_l^{(m)})F_{n-l,j}(X_l, \ldots, X_n),$$

entonces para cada $m, n \in \mathbb{N}_0$, $\{\rho_j^{(m)} > n\}$ es una función medible de $(X_0, \ldots, X_n)$, y por lo tanto un también tiempos de paro.

Teorema 14 (Recurrencia 1)**.** *Para toda* $m \in \mathbb{N}_0$ *y* $i, j \in \mathbb{S}$

$$\mathbb{P}(\rho_j^{(m)} < \infty \mid X_0 = i) = \mathbb{P}(\rho_j < \infty \mid X_0 = i)\,\mathbb{P}(\rho_j < \infty \mid X_0 = j)^{m-1}$$

En particular, si j *es recurrente, entonces* $\mathbb{P}(\rho_j^{(m)} < \infty \mid X_0 = j) = 1$ *para toda* $m \in \mathbb{N}_0$*. De hecho, si* j *es recurrente, entonces, condicional a* $X_0 = j$*,* $\{\rho_j^{(m)} - \rho_j^{(m-1)} : m \geq 1\}$ *es una sucesión de v.a.'s iid con la misma distribución que* ρ_j*.*

Demostración.

$$
\begin{aligned}
&\mathbb{P}(\rho_j^{(m)} < \infty \mid X_0 = i) = \sum_{n=1}^{\infty} \mathbb{P}(\rho_j^{(m-1)} = n\,,\, \rho_j^{(m)} < \infty \mid X_0 = i) \\
&= \sum_{n=1}^{\infty} \lim_{N\to\infty} \mathsf{E}\left[1 - F_{N,j}(X_n, \dots, X_{n+N}), \rho_j^{(m-1)} = n \mid X_0 = i\right] \\
&= \sum_{n=1}^{\infty} \lim_{N\to\infty} \mathsf{E}\left[1 - F_{N,j}(X_0, \dots, X_N) \mid X_0 = j\right] \mathbb{P}(\rho_j^{(m-1)} = n \mid X_0 = i) \\
&= \sum_{n=1}^{\infty} \lim_{N\to\infty} \mathbb{P}(\rho_j \le N \mid X_0 = j)\mathbb{P}(\rho_j^{(m-1)} = n \mid X_0 = i) \\
&= \mathbb{P}(\rho_j < \infty \mid X_0 = j)\mathbb{P}(\rho_j^{(m-1)} < \infty \mid X_0 = i) = \dots = \\
&= \mathbb{P}(\rho_j < \infty \mid X_0 = i)\, \mathbb{P}(\rho_j < \infty \mid X_0 = j)^{m-1}.
\end{aligned}
$$

Ahora, observemos que

$$
\begin{aligned}
&\mathbb{P}(\rho_j^{(m+1)} > n + n_m \mid X_0 = j, \rho_j^{(1)} = n_1, \dots, \rho_j^{(m)} = n_m) \\
&= \mathbb{E}[F_{n,j}(X_{n_m}, \dots, X_{n_m+n}) \mid X_0 = j, \rho_j^{(1)} = n_1, \dots, \rho_j^{(m)} = n_m] \\
&= \mathbb{E}[F_{n,j}(X_0, \dots, X_n) \mid X_0 = j] \\
&= \mathbb{P}(\rho_j > n \mid X_0 = j)
\end{aligned} \tag{4.17}
$$

Por lo tanto

$$\mathbb{P}(\rho_j^{(m)} < \infty \mid X_0 = j) = 1. \quad \square$$

Entonces, debido a que

$$
\mathbb{P}(\eta_j > m \mid X_0 = i) = \begin{cases} \mathbb{P}(\rho_j^{(m)} < \infty \mid X_0 = j) & \text{if } i = j \\ \mathbb{P}(\rho_j^{(m+1)} < \infty \mid X_0 = i) & \text{if } i \neq j \end{cases}
$$

Con η_j el tiempo total de ocupación en j como definido en (4.10).

Por lo tanto del Teorema 14 se sigue que

$$
\mathbb{E}[\eta_j \mid X_0 = i] = \delta_{i,j} + \frac{\mathbb{P}(\rho_j < \infty \mid X_0 = i)}{\mathbb{P}(\rho_j < \infty \mid X_0 = j)} \tag{4.18}
$$

$$
\mathbb{E}[\eta_j \mid X_0 = j] = \infty \iff \mathbb{P}(\eta_j = \infty \mid X_0 = j) = 1 \quad (\, j \text{ recurrente}) \tag{4.19}
$$

$$
\mathbb{E}[\eta_j \mid X_0 = j] < \infty \iff \mathbb{P}(\eta_j < \infty \mid X_0 = j) = 1 \quad (\, j \text{ transitorio}) \tag{4.20}
$$

En particular, bajo las condiciones del Teorema 12 (Döblin 3), sabemos que

$$(\mathsf{A}_n)_{j_0 j_0} \to (\boldsymbol{\pi})_{j_0} > 0,$$

y por lo tanto

$$\mathbb{E}[\eta_{j_0} \mid X_0 = j_0] = \sum_{m=0}^{\infty} (\mathsf{P}^m)_{j_0 j_0} = \lim_{n\to\infty} n\,(\mathsf{A}_n)_{j_0 j_0} = \infty$$

Esto decir, las condiciones Teorema 12 (Döblin 3) implican que j_0 es recurrente. Recordando el concepto de accecibilidad (3.3), que en el contexto de cadenas se traduce a decir que $i \longrightarrow j$ si $(\mathsf{P}^n)_{ij} > 0$ para algún $n \geq 0$, se puede decir mucho más. Notemos primero que si $i \longrightarrow j$ entonces $i = j$ ó $\mathbb{P}_i(\rho_j < \infty) > 0$.

Teorema 15. *Supongamos que* $\inf_i(\mathsf{A}_M)_{ij_0} \geq \epsilon$ *para algún* $M \geq 1$, $j_0 \in \mathbb{S}$, *y* $\epsilon > 0$. *Entonces* j *es recurrente si y solo si* $j_0 \to j$. *Mas aún, si* $j_0 \to j$, *entonces* $\mathbb{E}[\rho_j^p \mid X_0 = j] < \infty$ *para toda* $p \in (0, \infty)$.

Con el resultado anterior una pregunta que surge es como identificar la distribución estacionaria.

Lema 11. Bajo las condiciones del Teorema 12 (Döblin 3) se tiene que

$$\sup_{M\geq 1}\sup_{j\in\mathbb{S}}\inf_{i\in\mathbb{S}}(\mathsf{A}_M)_{ij} > 0 \Rightarrow (\boldsymbol{\pi})_j = \frac{1}{\mathbb{E}[\rho_j \mid X_0 = j]}$$

donde

$$\frac{1}{\mathbb{E}[\rho_j \mid X_0 = j]} := 0 \text{ si } j \text{ es transitiorio}$$

Demostración. La idea de la demostración es que por un lado se tiene que

$$\mathbb{E}\left[\bar{\eta}_j^{(n)} \mid X_0 = j\right] = (\mathsf{A}_n)_{jj} \to (\boldsymbol{\pi})_j$$

y por el otro

$$X_0 = j \Rightarrow \bar{\eta}_j^{(\rho_j^{(m)})} = \frac{1}{\rho_j^{(m)}}\sum_{l=0}^{\rho_j^{(m)}-1}\mathbf{1}(X_l = j) = \frac{m}{\rho_j^{(m)}}$$

Entonces, como $\rho_j^{(m)}$ es la suma de m copias independientes de ρ_j, lo anterior combinado con el la ley débil de los grandes números conlleva a

$$(\boldsymbol{\pi})_j = \lim_{m\to\infty}\mathbb{E}\left[\bar{\eta}_j^{(\rho_j^{(m)})} \mid X_0 = j\right] = \frac{1}{\mathbb{E}[\rho_j \mid X_0 = j]}$$

Para llevar a cabo este procedimiento, demostraremos un resultado más fuerte, *e.g.*

$$\mathbb{P}\left(\lim_{n\to\infty}\bar{\eta}_j^{(n)} = \frac{1}{\mathbb{E}[\rho_j \mid X_0 = j]}\middle| X_0 = j\right) = 1 \tag{4.21}$$

y en particular, dado que $0 \leq \bar{\eta}^{(n)} \leq 1$, el teorema de convergencia dominada dice que

$$(\boldsymbol{\pi})_j = \lim_{n\to\infty}(\mathsf{A}_{jj}) = \lim_{n\to\infty}\mathbb{E}\left[\bar{\eta}_j^{(n)} \mid X_0 = j\right] = \frac{1}{\mathbb{E}[\rho_j \mid X_0 = j]}$$

se sigue de (4.21). Entonces, únicamente necesitamos probar (4.21). Sean $j_0, M, \epsilon > 0$, tal que $(\mathsf{A}_M)_{ij_0} \geq \epsilon$ para toda i.

Si $j_0 \nrightarrow j$, entonces, por el Teorema 15, j es transitorio y $\mathbb{P}(\eta_j < \infty \mid X_0 = j) = 1$. Por lo tanto, condicional a $X_0 = j$, $\bar{\eta}_j^{(n)} \leq \frac{1}{n}\eta_j \to 0$ con probabilidad 1. Similarmente, debido a que j es transitorio, $\mathbb{P}(\rho_j = \infty \mid X_0 = j) > 0$y $\mathbb{E}[\rho_j \mid X_0 = j] = \infty$.

Si $j_0 \to j$. Entonces, por Teorema 15, $\mathbb{E}[\rho_j^4 \mid X_0 = j] < \infty$ y, condicional a $X_0 = j$, $\{\rho_j^{(m)} - \rho_j^{(m-1)}\}$ son iid con distribución igual a la de ρ_j. En particular, por la Ley Fuerte de los Grandes Numeros(LFGN) se tiene

$$\mathbb{P}\left(\lim_{m\to\infty}\frac{\rho_j^{(m)}}{m} = r_j \mid X_0 = j\right) = 1$$

donde $r_j = \mathbb{E}[\rho_j \mid X_0 = j]$. Por otro lado, para cualquier $m \geq 1$

$$|\bar{\eta}_j^{(n)} - r_j^{-1}| \leq |\bar{\eta}_j^{(n)} - \bar{\eta}_j^{(\rho_j^{(m)})}| + |\bar{\eta}_j^{(\rho_j^{(m)})} - r_j^{-1}|$$

y

$$
\begin{aligned}
|\bar{\eta}_j^{(n)} - \bar{\eta}_j^{(\rho_j^{(m)})}| &\leq \frac{|\bar{\eta}_j^{(n)} - \bar{\eta}_j^{(\rho_j^{(m)})}|}{n} + \left|1 - \frac{\rho_j^{(m)}}{n}\right| \bar{\eta}_j^{(\rho_j^{(m)})} \\
&\leq 2\left|1 - \frac{\rho_j^{(m)}}{n}\right| \leq 2\left|1 - \frac{mr_j}{n}\right| + \frac{2m}{n}\left|\frac{\rho_j^{(m)}}{n} - r_j\right|
\end{aligned}
$$

Debido a que $\rho_j^{(m)} \geq m$,

$$
|\bar{\eta}_j^{(\rho_j^{(m)})} - r_j^{-1}| \leq \frac{1}{r_j}\left|\frac{\rho_j^{(m)}}{n} - r_j\right|
$$

Entonces,

$$
|\bar{\eta}_j^{(n)} - r_j^{-1}| \leq 2\left|1 - \frac{mr_j}{n}\right| + \left(\frac{2m}{n} + \frac{1}{r_j}\right)\left|\frac{\rho_j^{(m)}}{n} - r_j\right|
$$

Finalmente si hacemos $m_n = [n/r_j]$ obtenemos

$$
|\bar{\eta}_j^{(n)} - r_j^{-1}| \leq \frac{2}{n} + \frac{3}{r_j}\left|\frac{\rho_j^{(m_n)}}{mn} - r_j\right| \to 0 \quad \text{cuando } n \to \infty
$$

lo que implica que, con probabilidad 1, el tiempo promedio que una trayectoria de la cadena gasta en cada estado tiende a la probabilidad que la distribución estacionaria le asigna en ese estado. □

Ejercicio 33. *Supongamos que* P *es una matriz de transición sobre un conjunto finito* $\mathbb{S}$ *y que* P *es doblemente estocástica, i.e. sus columnas también suman* 1*. Bajo la condición de que todo estado es accesible desde cualquier otro, muestra que*

$$
\mathbb{E}[\rho_j \mid X_0 = j] = \#\mathbb{S}, \quad \textit{para cada } j \in \mathbb{S}
$$

donde $\rho_j := \inf\{n > 0 : X_n = j\}$.

Con esta notación, si $\mathbb{E}[\rho_i \mid X_0 = i] < \infty$ se dice que i es un estado ***recurrente positivo***. En este contexto, el ***periodo*** del estado i esta dado por $d(i) := \mathsf{gcd}\{n \geq 0 : (\mathsf{P}^n)_{ii} > 0\}$. Si $d(i) = 1$, entonces se dice que i es ***aperiódico***.

Ejercicio 34. *Supóngase que* j *es un estado recurrente positivo, y sea* $C := \{i : i \leftrightarrow j\}$*. Dado un vector de probabilidad* $\boldsymbol{\mu}$ *con la propiedad que* $\sum_{i \neq C}(\boldsymbol{\mu})_i = 0$*. Demuestra que en general* $(\boldsymbol{\mu}\mathsf{A}_n)_i \to (\boldsymbol{\pi})_i$ *y, cuando* j *es aperiódico,* $(\boldsymbol{\mu}\mathsf{P}^n)_i \to (\boldsymbol{\pi})_i$ *para cada* $i \in C$*. Recuerda que* $\mathsf{A}_n := \frac{1}{n}\sum_{n=0}^{m-1}\mathsf{P}^n$.

Ejercicio 35. *Sea* i *un estado recurrente y, para* $k \in \mathbb{S}$ *sea* μ_k *el número esperado de veces que la cadena visita el estado* k *antes de regresar al estado* i *dado que la cadena comenzó en el estado* i*, i.e.*

$$
\mu_k = \mathbb{E}_i\left[\sum_{m=0}^{\rho_i - 1}\mathbb{I}(X_m = k)\right] \in [0, \infty]
$$

Define el vector renglón $\boldsymbol{\mu}$ *por* $(\boldsymbol{\mu})_k = \mu_k$.

a) Demuestra que, para toda $j \in \mathbb{S}$,

$$
(\boldsymbol{\mu}\mathsf{P})_j = \mathbb{E}_i\left[\sum_{m=1}^{\rho_i}\mathbb{I}(X_m = j)\right] = \mu_j.
$$

Entonces, sin la necesidad de suposiciones adicionales acerca de i*,* $\boldsymbol{\mu}$ *es* P*-invariante, i.e.* $\boldsymbol{\mu} = \boldsymbol{\mu}\mathsf{P}$.

b) *Claramente se tiene que $\mu_i = 1$ y $\sum_j \mu_j = \infty$ a menos de que el estado i sea recurrente positivo. Sin embargo, demuestra que si $i \not\leftrightarrow j$ entonces $\mu_j = 0$ y que si $i \leftrightarrow j$, $\mu_j \in (0, \infty)$. Hint: Demuestra que $\mathbb{P}_i(\rho_j^{(m)} < \rho_i) = \mathbb{P}_j(\rho_j < \rho_i)^{m-1}\mathbb{P}_i(\rho_j < \rho_i)$.*

c) *Si el estado i es recurrente positivo, demuestra que*

$$\bar{\boldsymbol{\mu}} := \frac{\boldsymbol{\mu}}{\sum_k \mu_k} = \boldsymbol{\pi}^C$$

con C una clase de estados recurrentes positivos. Equivalentemente, cuando i es recurrente positivo,

$$(\boldsymbol{\pi}^C)_j = \frac{\mathbb{E}_i\left[\sum_{m=0}^{\rho_i - 1} \mathbb{I}(X_m = j)\right]}{\mathbb{E}_i[\rho_i]}$$

En palabras, $(\boldsymbol{\pi}^C)_j =$ es la cantidad esperada relativa de tiempo que la cadena gasta en el estado j antes de regresar a i.

Ejercicio 36. *Sea X una $\mathsf{CM}_{\{0,1,\ldots,n\}}(\nu, \mathsf{P})$ con las entradas de P dadas por*

$$\mathsf{p}_{ij} = p_i\delta_{i+1}(j) + q_i\delta_{i-1}(j) + r_i\delta_i(j), \tag{4.22}$$

con $p_i + q_i + r_i = 1$ y $q_0 = p_n = 0$.

a) *Si $p_0 = 0$, $p_k, q_k > 0$ con $p_k + q_k < 1$. ¿Cómo se clasifican los estados de este proceso? ¿Cuántas clases de comunicación hay? ¿Qué ocurre si el proceso llega al estado 0? ¿Se puede decir que la cadena es aperiódica?*

b) *Contesta las preguntas del inciso anterior cuando $r_k = 0$ para toda k.*

c) *Sea $\boldsymbol{\mu}$ una función sobre $\mathbb{S}$ y $r_k, p_k, q_k > 0$, tal que $(\boldsymbol{\mu})_0 = 1$ y*

$$(\boldsymbol{\mu})_k = \prod_{i=1}^{k} \frac{p_{i-1}}{q_i}, \quad 1 \le k \le n.$$

Si

$$\boldsymbol{\pi} = \frac{\boldsymbol{\mu}}{\sum_{k\in\mathbb{S}} \mu_k}$$

demuestra que la cadena es reversible y estacionaria con respecto a $\boldsymbol{\pi}$.

4.3. Cadenas de Markov a tiempo continuo

Ahora analicemos el caso de una Cadena de Markov a tiempo continuo con espacio de estados $\mathbb{S}$ numerable, parte del material aquí lo tomaremos de Norris (1998). Al igual que en el caso general y el caso de $\mathsf{CM}_{\mathbb{S}}(\boldsymbol{\nu}, \mathsf{P})$, aquí también podemos aplicar la construcción de la Sección 2.2

Ejemplo 13 (*Continuación Ejemplo 9*)**.** *En el Ejemplo 9 vimos como construir una* $\mathsf{CM}_{\mathbb{N}_0}(\pi^{\mathsf{X}}, \mathsf{P})$*,* $\mathsf{X} = \{X_n\}_{n=1}^{\infty}$*, con distribución invariante* $\pi^{\mathsf{X}}(x) = \mathsf{Po}(x; \lambda)$ *mediante la construcción de la Sección 2.2. Para esto suponíamos que la dependencia estaba inducía por*

$$\tilde{\mu}_x(y) = \mathsf{Bin}(y; x, \xi), \quad 0 < \xi < 1$$

lo cual, después de aplicar Bayes, nos daba

$$\mu_y(x) = \frac{[\lambda(1-\xi)]^{x-y}}{(x-y)!}\, e^{-\lambda(1-\xi)} \mathbb{I}_{[y,\infty)}(x),$$

es decir la distribución correspondiente a la variable aleatoria Poisson desplazada por y*, i.e.* $y + \mathsf{Po}(\lambda(1-\xi))$*. Con esto encontrábamos que las entradas de la matriz de transición* P *estaban dadas por* (4.5).

Al igual que el mecanismo utilizado para construir la funciones de transición a tiempo continuo, e.g. (2.16) *y* (2.41)*, en este caso también podemos substituir* ξ *por una función continua* $t \mapsto \xi_t$*, con* $0 < \xi_t < 1$ *para toda* $t \geq 0$*, y encontrar la forma de esta que satisfacen Chapman-Kolmogorov. Una vez más las probabilidades de transición son más manejable vía su transformada de Laplace.*

Recordemos que si $Z \sim \mathsf{Po}(\lambda)$ *entonces* $\mathcal{L}_Z(\phi) = e^{\lambda(e^{\phi}-1)}$ *y si* $Z \sim \mathsf{Bin}(N, p)$ *entonces* $\mathcal{L}_Z(\phi) = (1 - p + pe^{\phi})^N$*. Con esto se tiene inmediatamente que la marginal de* Y*, en la construcción del Ejemplo 9, es* $\mathsf{Po}(\lambda\xi)$*. Esto debido a que*

$$\mathcal{L}_Y(\phi) = \mathsf{E}\{\mathsf{E}[e^{\phi Y} \mid X]\} = \mathsf{E}[(1-\xi+\xi e^{\phi})^X] = \mathcal{L}_X(\log\psi) = e^{\lambda(\psi-1)} = e^{\lambda\xi(e^{\phi}-1)}$$

donde ψ *se usó temporalmente para expresar* $(1 - \xi + \xi e^{\phi})$*.*

De la misma manera se puede ver que

$$\mathcal{L}_{Y|X=x}(\phi) = (1-\xi+\xi e^{\phi})^x \quad \text{y} \quad \mathcal{L}_{X|Y=y}(\phi) = e^{y\phi} e^{\lambda(1-\xi)(e^{\phi}-1)}$$

y por lo tanto

$$\begin{aligned}
\mathcal{L}_{X_t|X_0=i}(\phi) &= \mathsf{E}\left[\mathcal{L}_{X_t|Y}(\phi) \mid X_0 = i\right] \\
&= e^{\lambda(1-\xi_t)(e^{\phi}-1)} \mathcal{L}_{Y|X_0=i}(\phi) \\
&= e^{\lambda(1-\xi_t)(e^{\phi}-1)}\,(1-\xi_t+\xi_t e^{\phi})^i, \quad \text{para toda} \quad i \in \mathbb{N}_0
\end{aligned} \tag{4.23}$$

Ahora en términos de las transformadas las ecuación de Chapman-Kolmogorov se satisfacen si y solo si

$$\mathsf{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\phi) \mid X_0 = i\right] = \mathcal{L}_{X_{t+s}|X_0=i}(\phi) \tag{4.24}$$

Ejercicio 37. *Verifica que* $\xi_t = e^{-\alpha t}$*, con* $\alpha > 0$ *es la única solución a* (4.24).

Así pues, la construcción de la Sección 2.2 conlleva a la cadena de Markov $(X_t)_{t\geq 0}$ *con probabilidades de transición dadas por (4.5) (con* ξ *sustituida por* $e^{-\alpha t}$ *y* $x_1 = j$*), es decir*

$$\mathsf{p}_{ij}(t) = \mathsf{Po}(j; \lambda(1-e^{-\alpha t}))\,(1-e^{-\alpha t})^i\, {}_2\mathsf{F}_0\left(-i, -j; \frac{e^{-\alpha t}}{(1-e^{-\alpha t})^2\lambda}\right) \tag{4.25}$$

donde $\mathsf{p}_{ij}(t) := \mathbb{P}(X_t = j \mid X_0 = i)$*. Por construcción tenemos una cadena de Markov con distribución estacionaria* $\mathsf{Po}(\lambda)$ *y* $\mathsf{Po}(\lambda)$*-reversible. Al igual que en el caso a tiempo discreto, se tiene la representación*

$$X_t = e^{-\alpha t} \circ X_0 + \varepsilon_t, \quad \text{con} \quad \varepsilon_t \sim \mathsf{Po}(\lambda(1-e^{-\alpha t})) \tag{4.26}$$

con ε_t *independiente de* $(X_t)_{t\geq 0}$ *y* $\xi \circ X$ *el operador de adelgazamiento binomial.*

Ejercicio 38. *Dado una elección de tiempos $t_1, \ldots, t_n$, has un programa (en el lenguaje de programación de tu preferencia), que simule la cadena de Markov a tiempo continuo dada por (4.26). Explora para varios valores de (α, λ) y gráfica tus resultados.*

Ejercicio 39. (Punto extra en la tarea) *Dada una de las realizaciones del Ejercicio 38, e.g. $\boldsymbol{x}^{(n)} := (x_{t_1}, \ldots, x_{t_n})$, realiza un programa que estime los parámetros (α, λ). Por ejemplo, puedes maximizar numéricamente la verosimilitud inducida por (4.25) y $X_{t_1} \sim \mathsf{Po}(\lambda)$.*

De manera análoga al caso de procesos de Markov con espacio de estados generales, en el caso a tiempo continuo hablar de la transición de un estado a otro " en t pasos " no tiene sentido, i.e. ¿Cuál es la unidad de tiempo?, Claramente se requiere de una característica infinitesimal que describa la " intensidad de transición ", es decir

$$\mathsf{q}_{ij} = \frac{\mathrm{d}}{\mathrm{d}t}\mathsf{p}_{ij}(t)\Big|_{t=0} \tag{4.27}$$

Para esto, será necesario introducir el concepto de las matrices Q.

Definición 24. Sea $\mathbb{S}$ un conjunto numerable. Una ***matriz*** Q sobre $\mathbb{S}$, es una matriz $\mathsf{Q} = \{q_{ij}\} := (q_{ij} : i, j \in \mathbb{S})$ tal que sus entradas satisfacen

(i) $0 \leq -\mathsf{q}_{ii} < \infty$ para toda i

(ii) $\mathsf{q}_{ij} \geq 0$ para toda $i \neq j$

(iii) $\sum_{j \in \mathbb{S}} \mathsf{q}_{ij} = 0$ para toda i

Entonces, en cada renglón se puede escoger los elementos fuera de la diagonal como entradas no-negativas sujetos a la restricción de que $\mathsf{q}_i = \sum_{j \neq i} \mathsf{q}_{ij} < \infty$, con $\mathsf{q}_{ii} := -\mathsf{q}_i$, cantidad que toma una carácter especial en el estudio de cadenas de Markov a tiempo continuo. Claramente los elementos de la diagonal de Q son no-positivos y la suma sus renglones son cero.

Bajo el supuesto de regularidad que impone un semigrupo normal (2.6), i.e.

$$\lim_{t \downarrow 0} \mathsf{p}_{ij}(t) = \delta_{ij}, \tag{4.28}$$

que en contexto matricial equivale a $\mathsf{P}(0) = \mathbb{I}$, se puede ver que la matriz de transición $\mathsf{P}(t) := \{\mathsf{p}_{ij}(t)\}_{i,j \in \mathbb{S}}$ es diferenciable, en particular se tienen sus derivadas por la derecha.

Teorema 16. *Sea $\{\mathsf{P}(t)\}_{t \geq 0}$ el semigrupo de transición sobre $\mathbb{S}$.*

a) Para cualquier estado i,

$$\mathsf{q}_i := \lim_{h \downarrow 0} \frac{1 - \mathsf{p}_{ii}(h)}{h} \in [0, \infty]$$

existe y satisface

$$\mathsf{p}_{ii}(t) \geq e^{-t\mathsf{q}_i}.$$

b) Si $\mathsf{q}_i < \infty$, entonces para cualquier $(i, j) \in \mathbb{S}^2$, $i \neq j$,

$$\mathsf{q}_{ij} := \lim_{h \downarrow 0} \frac{\mathsf{p}_{ij}(h)}{h} \in [0, \infty)$$

existe y $\sum_j \mathsf{q}_{ij} \leq 0$.

c) Si para algún $i \in \mathbb{S}$, $\mathsf{q}_{ii} \leq 0$ y $\sum_j \mathsf{q}_{ij} = 0$, entonces $\mathsf{p}_{ij}(t)$ es continua y diferenciable en t, para ese i y todo $j \in \mathbb{S}$, y satisface la ecuación backward de Kolmogorov.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{p}_{ij}(t) = \sum_k \mathsf{q}_{ik}\mathsf{p}_{kj}(t) \tag{4.29}$$

Demostración: Para toda $t \geq 0$ y toda $n \geq 1$, tenemos $\mathsf{P}(t) = [\mathsf{P}(\frac{t}{n})]^n$ y, dado que (4.28) se satisface, existe $\epsilon > 0$ t.q. $\mathsf{p}_{ii}(h) > 0$ para toda $h \in [0, \epsilon]$. Por lo tanto, para toda $i \in \mathbb{S}$

$$\mathsf{p}_{ii}(t) \geq \left[\mathsf{p}_{ii}\left(\frac{t}{n}\right)\right]^n.$$

Para n suficientemente grande, $\frac{t}{n} \in [0, \epsilon]$, por lo tanto $\mathsf{p}_{ii}\left(\frac{t}{n}\right) > 0$ y $\mathsf{p}_{ii}(t) > 0$, además podemos definir

$$f_i(t) := -\log \mathsf{p}_{ii}(t)$$

Esta es una función real, no-negativa, tal que $\lim_{h\downarrow 0} f_i(h) = 0$. De forma equivalente vemos que (usando la prop. de semigrupo) $\mathsf{p}_{ii}(t+s) \geq \mathsf{p}_{ii}(t)\mathsf{p}_{ii}(s)$ y por lo tanto f_i es subaditiva, i.e.

$$f_i(t+s) \leq f_i(t) + f_i(s) \quad \forall s, t \in \mathbb{R}_+$$

Entonces

$$\mathsf{q}_i = \lim_{h\downarrow 0} \frac{f(t)}{t} \in [0, \infty]$$

y por lo tanto

$$\lim_{h\downarrow 0} \frac{1 - \mathsf{p}_{ii}(h)}{h} = \lim_{h\downarrow 0} \frac{1 - e^{-f_i(h)}}{f_i(h)} \frac{f_i(h)}{h} = \mathsf{q}_i$$

debido a que

$$\frac{1 - \mathsf{p}_{ii}(h)}{h} = \frac{1 - e^{-f_i(h)}}{f_i(h)} \frac{f_i(h)}{h} \leq \frac{f_i(h)}{h}$$

y $\forall\, h > 0$

$$\frac{1 - \mathsf{p}_{ii}(h)}{h} \leq \mathsf{q}_i$$

Entonces $a)$ queda demostrado.

Para la segunda igualdad, sean $i \neq j$. Debido a que $\mathsf{p}_{ii}(t)$ y $\mathsf{p}_{jj}(t)$ tienden a 1 cuando $t \to 0$, para cualquier $c \in (\frac{1}{2}, 1)$ $\exists\, \delta > 0$ t.q. para $t \in [0, \delta]$, $\mathsf{p}_{ii}(t) > c$ y $\mathsf{p}_{jj}(t) > c$. Sea $n > 0$ un entero y $h > 0$ t.q. $0 \leq nh \leq \delta$. Denotemos por $\{X_n\}_{n\geq 0}$ la CM a tiempo discreto definida por $X_n = X(nh)$, con matriz de transición $\mathbf{P}(h)$. Una forma de ir de $X_0 = i$ a $X_n = j$ es ir de $X_0 = i$ a $X_r = i$, $r = 0, \ldots, n-1$, sin pasar por j y entonces ir de $X_r = i$ a $X_{r+1} = j$ y de $X_{r+1} = j$ a $X_n = j$.

Las trayectorias correspondientes a valores distintos de r son diferentes, pero no son exhaustivas dentro de todas las posibilidades que hay de ir de $X_0 = i$ a $X_n = j$, entonces

$$\mathsf{p}_{ij}(nh) \geq \sum_{r=0}^{n-1} \mathbb{P}_i\left(\cap_{l=1}^{r-1}\{X_l \neq j\}, X_r = i\right) \mathsf{p}_{ij}(h) \mathbb{P}(X_n = j \mid X_{r+1} = j).$$

Los parámetros δ, n y h son tales que $\mathbb{P}(X_n = j \mid X_{r+1} = j) \geq c$. También

$$\begin{aligned}
&\mathbb{P}_i\left(\cap_{l=1}^{r-1}\{X_l \neq j\}, X_r = i\right) \\
&= \mathbb{P}_i(X_r = i) - \sum_{k<r} \mathbb{P}_i\left(\cap_{l=1}^{k-1}\{X_l \neq j\}, X_k = j\right) \mathbb{P}(X_r = i \mid X_k = j) \\
&\geq c - (1-c) \sum_{k<r} \mathbb{P}_i\left(\cap_{l=1}^{k-1}\{X_l \neq j\}, X_k = j\right) \geq c - (1-c) = 2c - 1 \qquad (4.30)
\end{aligned}$$

donde observamos que para $i \neq j$,

$$\mathbb{P}(X_r = i \mid X_k = j) + \mathbb{P}(X_r = j \mid X_k = j) \leq 1$$

y por lo tanto

$$\mathbb{P}(X_r = i \mid X_k = j) \leq 1 - \mathbb{P}(X_r = j \mid X_k = j) \leq 1 - c$$

Entonces

$$\mathsf{p}_{ij}(nh) \geq c(2c-1)n\mathsf{p}_{ij}(h)$$

Ahora, sea $t < \delta$ y $h < \delta$, y tomemos n como la parte entera de t/h. De la desigualdad anterior obtenemos

$$\frac{\mathsf{p}_{ij}(h)}{h} \leq \frac{1}{c(2c-1)} \frac{\mathsf{p}_{ij}(nh)}{nh}$$

y debido a que $\lim_{h \downarrow 0} nh = t$, vemos que $\lim_{h \downarrow 0} \frac{\mathsf{p}_{ij}(nh)}{nh} = \frac{\mathsf{p}_{ij}(t)}{t}$ y entonces

$$\limsup_{h \downarrow 0} \frac{\mathsf{p}_{ij}(h)}{h} \leq \frac{1}{c(2c-1)} \frac{\mathsf{p}_{ij}(t)}{t} < \infty$$

y dado que c se puede escoger arbitrariamente cercano a 1, tenemos

$$\limsup_{h \downarrow 0} \frac{\mathsf{p}_{ij}(h)}{h} \leq \liminf_{t \downarrow 0} \frac{\mathsf{p}_{ij}(t)}{t} < \infty$$

lo que implica la existencia y finites de $\lim_{h \downarrow 0} \frac{\mathsf{p}_{ij}(h)}{h}$. Para (c), notemos que por CK

$$\mathsf{p}_{ij}(t+s) = \sum_k \mathsf{p}_{ik}(s)\mathsf{p}_{k,j}(t)$$

entonces

$$\frac{\mathsf{p}_{ij}(t+s) - \mathsf{p}_{ij}(t)}{s} - \sum_k \mathsf{q}_{ik}\mathsf{p}_{kj}(t) = \sum_k \left[\frac{\mathsf{p}_{ik}(s) - \mathsf{p}_{ik}(0)}{s} - \mathsf{q}_{ik}\right] \mathsf{p}_{kj}(t)$$

donde cada término en la suma tiende a 0 cuando $s \downarrow 0$ por las partes (a) y (b). Por lo tanto, necesitamos controlar las colas de la suma. Sea $\mathbb{T} \subset \mathbb{S}$ finito que contiene a i, y nótese que

$$\sum_{k \notin \mathbb{T}} \left|\frac{\mathsf{p}_{ik}(s)}{s} - \mathsf{q}_{ik}\right| \mathsf{p}_{kj}(t) \leq \sum_{k \notin \mathbb{T}} \frac{\mathsf{p}_{ik}(s)}{s} + \sum_{k \notin \mathbb{T}} \mathsf{q}_{ik} = s^{-1}\left[1 - \sum_{k \in \mathbb{T}} \mathsf{p}_{ik}(s)\right] - \sum_{k \in \mathbb{T}} \mathsf{q}_{ik} \to -2\sum_{k \in \mathbb{T}} \mathsf{q}_{ik}$$

cuando $s \downarrow 0$. El límite se puede hacer arbitrariamente pequeño, haciendo $\mathbb{T}$ grande, ya que $\sum_k \mathsf{q}_{ik} = 0$. Entonces

$$\sum_k \left[\frac{\mathsf{p}_{ik}(s) - \mathsf{p}_{ik}(0)}{s} - \mathsf{q}_{ik}\right] \mathsf{p}_{kj}(t) \to 0$$

cuando $s \downarrow 0$. Entonces la derivada por la derecha existe y coincide con (4.29). Para ver que la derivada por los dos lados existe es suficiente notar que el lado derecho de (4.29) es continuo en t y dado que para cualquier función de transición $|\mathsf{p}_{ij}(t) - \mathsf{p}_{ij}(s)| \leq 1 - \mathsf{p}_{ii}(|t-s|)$, y como una función continua con derivada continua es derivable se sigue el resultado. □

A la matrix Q definida con q_{ij} como en el teorema anterior se le conoce como el ***generador infinitesimal*** de la cadena de Markov a tiempo continuo. De forma más compacta, se puede escribir

$$\mathsf{Q} = \lim_{h \downarrow 0} \frac{\mathsf{P}(h) - \mathbf{P}(0)}{h}$$

es decir la derivada derecha en 0 de la función matricial $t \mapsto \mathbf{P}(t)$.

Ejercicio 40. (Punto extra en la tarea). *Demostrar que para el ejemplo con probabilidades de transición dadas por (4.25), el generador infinitesimal esta dado por*

$$\mathsf{q}_{ij} = \begin{cases} -\lim_{t \downarrow 0} \frac{1 - p_t(i,i)}{t}, & j = i \\ \lim_{t \downarrow 0} \frac{p_t(i,j)}{t}, & i \neq j \end{cases} = \begin{cases} -\alpha(i+\lambda), & j = i \\ \alpha\lambda, & j = i+1 \\ i\alpha, & j = i-1 \\ 0, & \textit{e.o.c}, \end{cases}$$

Teorema 17 (Caso $\#\mathbb{S} < \infty$). *Sea* Q *una matriz sobre un espacio finito* $\mathbb{S}$ *y definamos* $\mathsf{P}(t) = e^{t\mathsf{Q}}$. *Entonces* $\mathsf{P}(t)$ *satisface la siguientes propiedades:*

(i) Propiedad de semigrupo: $\mathsf{P}(t+s) = \mathsf{P}(t)\mathsf{P}(s)$ *para toda* t, s

(ii) $(\mathsf{P}(t) : t \geq 0)$ *es la única solución a la ecuación forward de Kolmogorov*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{P}(t) = \mathsf{P}(t)\mathsf{Q}, \qquad \mathsf{P}(0) = \mathbf{I}$$

(iii) $(\mathsf{P}(t) : t \geq 0)$ *es la única solución a la ecuación backward de Kolmogorov*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{P}(t) = \mathsf{Q}\mathsf{P}(t), \qquad \mathsf{P}(0) = \mathbf{I}$$

(iv) Para $k = 0, 1, 2, \ldots$

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k \Big| \mathsf{P}(t) = \mathsf{Q}^k.$$

Demostración: Para cada t, s, $t\mathsf{Q}$ y $s\mathsf{Q}$ conmutan entonces la propiedad (i) se sigue de inmediato. Ahora,

$$\mathsf{P}'(t) = \sum_{k=1}^{\infty} \frac{t^{k-1}\mathsf{Q}^k}{(k-1)!} = \mathsf{P}(t)\mathsf{Q} = \mathsf{Q}\mathsf{P}(t),$$

por lo tanto satisface las ecuaciones backward y forward. Más aun si diferenciamos término a término se obtiene (iv). Por otro lado si $\mathsf{M}(t)$ satisface la ec. *forward* entonces

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{M}(t)e^{-t\mathsf{Q}}) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{M}(t)\right)e^{-t\mathsf{Q}} + \mathsf{M}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}e^{-t\mathsf{Q}}\right) \\ &= \mathsf{M}(t)\mathsf{Q}e^{-t\mathsf{Q}} + \mathsf{M}(t)(-\mathsf{Q})e^{-t\mathsf{Q}} \end{aligned}$$

entonces $\mathsf{M}(t)e^{-t\mathsf{Q}}$ es constante (=$\mathbb{I}$) entonces $\mathsf{M}(t) = e^{-t\mathsf{Q}} = \mathsf{P}(t)$. Un argumento análogo se sigue para la ecuación *backward*.

Teorema 18. *Una matriz* Q *sobre un espacio finito* $\mathbb{S}$ *es una matriz* Q *sii* $\mathsf{P}(t) = e^{t\mathsf{Q}}$ *es una matriz estocástica para toda* $t \geq 0$.

Demostración: Cuando $t \downarrow 0$ tenemos

$$\mathsf{P}(t) = \mathbf{I} + t\mathsf{Q} + \mathcal{O}(t^2)$$

Entonces $\mathsf{q}_{ij} \geq 0$ para $i \neq j$ sii $\mathsf{p}_{ij}(t) \geq 0$ for all i, j y $t \geq 0$ suficientemente pequeño. Debido a que $\mathsf{P}(t) = \mathsf{P}(t/n)^n$ para toda n, se sigue que $\mathsf{q}_{ij} \geq 0$ para $i \neq j$ ssi $\mathsf{p}_{ij}(t) \geq 0$ para toda i, j y toda $t \geq 0$. Si los renglones de Q suman cero entonces los renglones de Q^n también

$$\sum_{k\in\mathbb{S}} \mathsf{q}_{ik}^{(n)} = \sum_{k\in\mathbb{S}} \sum_{j\in\mathbb{S}} \mathsf{q}_{ij}^{(n-1)} q_{jk} = \sum_{j\in\mathbb{S}} \mathsf{q}_{ij}^{(n-1)} \sum_{k\in\mathbb{S}} \mathsf{q}_{jk} = 0$$

Entonces

$$\sum_{i\in\mathbb{S}} \mathsf{p}_{ij}(t) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{j\in\mathbb{S}} \mathsf{q}_{ij}^{(n)} = 1$$

Por otro lado, si $\sum_{j\in\mathbb{S}} \mathsf{p}_{ij}(t) = 1$ para toda $t \geq 0$, entonces

$$\sum_{j\in\mathbb{S}} \mathsf{q}_{ij} = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \sum_{j\in\mathbb{S}} \mathsf{p}_{ij}(t) = 0.$$

4.3.1. Relación con la medida estacionaria

Otra propiedad importante que se deriva directamente del generador infinitesimal Q es su relación con la distribución estacionaria. Esto es si $\boldsymbol{\pi}$ es una distribución estacionaria para el semigrupo de transición $\mathbf{P}(t)$ entonces

$$\boldsymbol{\pi}\mathsf{Q} = \frac{\boldsymbol{\pi}\mathsf{P}(t) - \boldsymbol{\pi}}{t} = 0$$

De la misma forma si $\boldsymbol{\pi}\mathsf{Q} = 0$ para alguna distribución $\boldsymbol{\pi}$ entonces

$$\boldsymbol{\pi}\mathsf{P}(t) = \boldsymbol{\pi}e^{t\mathsf{Q}} = \boldsymbol{\pi}(\mathbf{I} + t\mathsf{Q} + \frac{t^2\mathsf{Q}^2}{2!} + \ldots) = \boldsymbol{\pi}$$

Entonces $\boldsymbol{\pi}$ es una distribución estacionaria para $\mathsf{P}(t)$.

4.3.2. Posibles comportamientos trayectoriales

Recordemos que el supuesto de continuidad por la derecha conlleva a que para toda $\omega \in \Omega$ y $t \geq 0$ existe $\epsilon > 0$ tal que

$$X_s(\omega) = X_t(\omega), \quad \text{para} \quad t \leq s \leq t + \epsilon$$

Es decir, cada trayectoria de un proceso continuo por la derecha permanece constante por un tiempo antes de hacer una transición a otro estado. Existen, al menos tres posibilidades para las trayectorias:

- Un número infinito de brincos, pero finito en $[0, t]$
- Un número finito de brincos, y entonces queda atrapado en algún estado para siempre
- Un número infinito de brincos en un intervalo finito. En este caso, después de un ***tiempo de explosión*** ζ, el proceso puede comenzar de nuevo y explotar y explotar (posiblemente un número infinito de veces) ó solo vuelve a explotar un número finito de veces (incluyendo 0 veces).

Ver pág. 68-70 de Norris (1998).

Tiempos de brincos

Denotaremos mediante $J_0, J_1, \ldots$ los ***tiempos de brincos*** de $(X_t)_{t\geq 0}$, con

$$J_0 = 0, \qquad J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\} \quad \text{para } n = 0, 1, \ldots$$

donde $\inf \emptyset = \infty$, y por

$$S_n = \begin{cases} J_n - J_{n-1} & \text{si } J_{n-1} < \infty \\ \infty & \text{e.o.c.} \end{cases}$$

los ***tiempos de espera*** o ***tiempos de ocupación***. La continuidad por la derecha implica que $S_n > 0$ para toda n. Si $J_{n+1} = \infty$ para algún n, $X_\infty := X_{J_n}$. El (primer) ***tiempo de explosión*** ζ esta dado por

$$\zeta := \sup_n J_n = \sum_{n=1}^{\infty} S_n$$

Diremos que la cadena no explota si el proceso es ***honesto***, es decir $\mathsf{P}(\zeta = \infty) = 1$, en este caso $\sum_k \mathsf{p}_{ik}(t) = 1$ para toda $t \geq 0$ (pensando en la cadena sobre $\mathbb{S}$). Se dice que la Q ***es explosiva*** si, para la cadena de Markov asociada

$$\mathbb{P}_i(\zeta < \infty) > 0 \quad \text{para algún} \quad i \in \mathbb{S}.$$

De otra forma Q es ***no-explosiva***.

Al proceso *a tiempo discreto* $(Y_n)_{n\geq 0}$ definido por $Y_n := X_{J_n}$ se le denomina ***el proceso de brincos*** asociado a $(X_t)_{t\geq 0}$. Este proceso consiste simplemente de los valores que toma $(X_t)_{t\geq 0}$ hasta el momento de explosión (***los estados visitados***).

Por el momento, no consideraremos que ocurre después de una explosión. Por lo tanto será conveniente considerar un espacio de estados aumentado $\mathbb{S}^m = \mathbb{S} \cup \{\infty\}$ y considerar el proceso

$$X_t^m = \begin{cases} X_t & \text{si } t < \zeta \\ \infty & \text{si } t \geq \zeta \end{cases}$$

Al proceso X_t^m se le conoce como ***proceso minimal***.

Lema 12. Sea $S_n \overset{\text{ind}}{\sim} \text{Exp}(\lambda_n)$ y $0 < \lambda_n < \infty$ para toda n.

$$\text{(i) Si } \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty, \text{ entonces } \mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$$
$$\text{(ii) Si } \sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty, \text{ entonces } \mathbb{P}\left(\sum_{n=1}^{\infty} S_n = \infty\right) = 1$$

Ejercicio 41. *Demostrar el Lema 12.*

Lema 13. Sea $\mathbb{S}$ un conjunto numerable y $\{T_k\}_{k\in\mathbb{S}}$ variables aleatorias iid con distribución $\text{Exp}(q_k)$ con $0 < q := \sum_{k\in\mathbb{S}} q_k < \infty$. Sea $T = \inf_k T_k$. Entonces este ínfimo se alcanza en un único valor aleatorio K con prob. 1. Más aun, T y K son independientes, con $T \sim \text{Exp}(q)$ y la distribución de K esta dada por

$$\mathbb{P}(K = k) = \frac{q_k}{q}.$$

Demostración: Fijemos $K = k$ si $T_k < T_j$ para toda $j \neq k$, de otra forma K se no es propia. Entonces

$$\begin{aligned} \mathbb{P}(K = k, T \geq t) &= \mathbb{P}(T_k \geq t \text{ y } T_j > T_k \, \forall j \neq k) \\ &= \int_t^{\infty} q_k e^{-q_k s} \mathbb{P}(T_j > s \, \forall j \neq k) ds \\ &= \int_t^{\infty} q_k e^{-q_k s} \prod_{j\neq k} e^{-q_j s} ds \\ &= \int_t^{\infty} q_k e^{-qs} ds = \frac{q_k}{q} e^{-qt} \end{aligned}$$

Entonces $\mathbb{P}(K = k$ para algún k$) = 1$ y T y K son independientes con la distribución establecida por el enunciado.

Lema 14. Si $S \sim \text{Exp}(\lambda)$ y $R \sim \text{Exp}(\mu)$ independientes. Entonces para $t \geq 0$, tenemos

$$\mu\mathbb{P}(S \leq t < S + R) = \lambda\mathbb{P}(R \leq t < R + S)$$

Demostración:

$$\mu\mathbb{P}(S \leq t < S + R) = \mu \int_0^t \int_{t-s}^{\infty} \lambda\mu e^{-\lambda s} e^{-\mu r} dr ds = \lambda\mu \int_0^t e^{-\lambda s} e^{-\mu(t-s)} ds$$

de donde la igualdad se cumple por simetría.

4.3.3. La construcción de cadenas de Markov vía la matriz de brincos y tiempos de espera

La matriz Q define una matriz estocástica conocida como la ***matriz de brincos*** $\mathbf{J} := (\mathsf{J}_{ij} : i, j \in \mathbb{S})$, definida por

$$\mathsf{J}_{ij} = \begin{cases} \mathsf{q}_{ij}/\mathsf{q}_i & \text{si } j \neq i \text{ y } \mathsf{q}_i \neq 0 \\ 0 & \text{si } j \neq i \text{ y } \mathsf{q}_i = 0 \end{cases}$$

$$\mathsf{J}_{ii} = \begin{cases} 0 & \text{si } \mathsf{q}_i \neq 0 \\ 1 & \text{si } \mathsf{q}_i = 0 \end{cases}$$

Ejercicio 42. *Demuestra que* $\mathbf{J}$ *es una matriz estocástica.*

Un proceso minimal continuo por la derecha $(X_t)_{t\geq 0}$ es una cadena de Markov con distribución inicial μ y matriz generadora Q si su cadena de brincos $(Y_n)_{n\geq 0}$ es una cadena de Markov a tiempo discreto modulada por $\mathsf{CM}_{\mathbb{S}}(\mu, \mathbf{J})$ ***y si*** para cada $n \geq 1$, condicional a $\mathcal{F}^Y_{n-1}$, sus tiempos de espera $S_1, \ldots, S_n$ son exponenciales independientes con parámetros $\mathsf{q}(Y_0), \ldots, \mathsf{q}(Y_{n-1})$ respectivamente [con $\mathsf{q}(x) = \mathsf{q}_x$]. Para simplificar, diremos que $(X_t)_{t\geq 0}$ es una $\mathsf{CM}_{\mathbb{S}}(\mu, \mathsf{Q})$, y que a su vez puede ser construida como sigue:

Sean $T_i \overset{\text{iid}}{\sim} \text{Exp}(1)$, independientes de $(Y_n)_{n\geq 0}$ y

$$A_n := \sum_{r=0}^{n} \mathsf{q}(Y_r)^{-1} T_i, \quad \tau_t := \inf\{n : A_n > t\}$$

Entonces $X_t := Y(\tau_t)$, es decir

$$X_t = \sum_{n=0}^{\infty} \mathbb{I}_{[A_n, A_{n+1})}(t)$$

que equivalentemente significa que $X_t = Y_n$ si $A_n \leq t < A_{n+1}$ para algún n y ∞ e.o.c.

Teorema 19. *Sea* $(X_t)_{t\geq 0}$ *con matriz generadora* Q *y distribución inicial* μ*. Entonces* $(X_t)_{t\geq 0}$ **no explota** *si cualquiera de las siguientes condiciones se satisface*

(i) $\mathbb{S}$ *es finito*

(ii) $\sup_{i\in\mathbb{S}} q_i < \infty$

(iii) $X_0 = i$*, y* i *es un estado recurrente para la cadena de brincos.*

Ejercicio 43. *Demostrar Teorema 19.*

Ejercicio 44. *Sea* $\{X_t : t \geq 0\}$ *una la cadena de Markov a tiempo continuo con distribución inicial* $\pi = \mathsf{Po}(\lambda)$ *y probabilidades de transición (4.25), o equivalentemente con generador infinitesimal dado por el Ejercicio 40. Describe la matriz de brincos asociada. Cuál es la distribución de los tiempos de brincos y los tiempos de espera?*

4.3.4. Otra construcción

Comenzamos en el estado inicial $X_0 = Y_0$ con distribución inicial μ y $\{T_n^j\}_{n\geq 1, j\in\mathbb{S}} \overset{\text{iid}}{\sim} \text{Exp}(1)$. Ahora de manera inductiva para $n = 0, 1, 2, \dots$ ***si*** $Y_n = i$ sea

$$S_{n+1}^j = T_{n+1}^j / q_{ij}, \text{ para } j \neq i$$

$$S_{n+1} = \inf_{j\neq i} S_{n+1}^j,$$

$$Y_{n+1} = \begin{cases} j & \text{si } S_{n+1}^j = S_{n+1} < \infty \\ i & \text{si } S_{n+1} = \infty \end{cases}$$

Entonces, condicional a $Y_n = i$, las variables aleatorias S_{n+1}^j son ind. $\text{Exp}(q_{ij})$, y por Lema 13, S_{n+1} son exponenciales con parámetro q_i, Y_{n+1} tiene distribución $\{J_{ij} : j \in \mathbb{S}\}$, y S_{n+1} y Y_{n+1} son independientes, e independientes también de $\{Y_i\}_{i=0}^n$ y $\{S_i\}_{i=0}^n$. Esta construcción explica el porque la matriz Q recibe la interpretación

- q_i la tasa con la que se deja el estado i
- q_{ij} la tasa con la que se va del estado i al estado j.

Teorema 20. *Sea $(X_t)_{t\geq 0}$ con matriz generadora Q y ζ, su tiempo de explosión. Para un estado inicial $i \in \mathbb{S}$, sea $\mathsf{z}_i = \mathbb{E}_i(e^{-\theta\zeta})$, $\theta > 0$. Entonces $\mathsf{z} := (\mathsf{z}_i : i \in \mathbb{S})$ satisface*

(i) $|\mathsf{z}_i| \leq 1$ *para toda* $i \in \mathbb{S}$

(ii) $\mathsf{Q}\mathsf{z} = \theta\mathsf{z}$.

Demostración: Ver Teorema 2.7.2. en Norris (1998)

Corolario 4. *Para $\theta > 0$ las siguientes afirmaciones son equivalentes:*

a) Q *no es explosiva*

b) $\mathsf{Q}\mathsf{z} = \theta\mathsf{z}$ *y* $|\mathsf{z}_i| \leq 1$, *para toda* i, $\Rightarrow \mathsf{z} = 0$.

Demostración: Ver Corolario 2.7.3. en Norris (1998)

Ejercicio 45 (Proceso Poisson)**.** *Supóngase que tenemos una cadena de Markov a tiempo continuo, $\mathsf{X} = (X_t)_{t\geq 0}$ con $X_0 = 0$ y*

$$\mathsf{Q} = \begin{vmatrix} -\lambda & \lambda & 0 & 0 & \cdots & \\ 0 & -\lambda & \lambda & 0 & \cdots & \\ 0 & 0 & -\lambda & \lambda & 0 & \\ 0 & \cdots & 0 & \ddots & \ddots & \ddots \end{vmatrix}$$

Demuestra que $\mathbb{P}[X_t = n] = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$ para $n = 0, 1, 2, \dots$

Ejercicio 46 (Modelo M/M/1)**.** *Supóngase que la cadena de Markov $\mathsf{X} = (X_t)_{t\geq 0}$ representa el número de personas haciendo fila al tiempo t para ser atendidas por un sólo servidor (incluyendo aquellas que están siendo servidas). Además, supóngase que individuos llegan a la fila de acuerdo a un proceso Poisson con tasa λ y que cada persona requiere un tiempo servicio independiente, exponencialmente distribuido con tasa $1/\mu$. Verifica que*

$$\mathsf{Q} = \begin{vmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ 0 & 0 & \ddots & \ddots & \ddots \end{vmatrix}.$$

Ejercicio 47. (Optativo). *Lectura de las secciones 3.2 a 3.8 del libro Norris (1998). Estas secciones, presentan algunos conceptos ya vistos de forma más general, sin embargo pueden ayudar a resolver los ejercicios optativos (52) abajo.*

4.3.5. Operaciones usando el resolvente

Si en vez de operar con $\mathsf{P}(t) = e^{\mathsf{Q}t}$ operamos con su resolvente

$$\mathsf{R}(\lambda) = (\lambda\mathbb{I} - \mathsf{Q})^{-1}, \quad \lambda > 0$$

donde claramente la inversa de $(\lambda\mathbb{I} - \mathsf{Q})$ se puede calcular para los valores de λ que no son valores propios de Q, i.e. la inversa no existe si y solo si $\det(\lambda\mathbb{I} - \mathsf{Q}) = 0$.

Ahora si $\mathsf{R}(\lambda) = (\lambda\mathbb{I} - \mathsf{Q})^{-1}$ existe, entonces podemos recuperar $\mathsf{P}(t)$ invirtiendo la transformada de Laplace. Nótese que el resolvente tiene la siguiente interpretación: Sea $A \sim \mathsf{Exp}(\lambda)$ una *alarma* independiente de la cadena $\{X_t\}_{t\geq 0}$. Si nos preguntamos ¿Cuál es el estado de la cadena cuando dicha alarma suena?, entonces tenemos

$$\mathbb{P}_i(X_A = j) = \int_0^\infty \mathbb{P}_i(X_t = j, A \in (t, t+dt)) = \int_0^\infty \mathsf{p}_{ij}\lambda e^{-\lambda t}dt = \lambda \mathsf{r}_{ij}(\lambda),$$

donde $\mathsf{r}_{ij}(\lambda)$ denota la ij-ésima entrada de $\mathsf{R}(\lambda)$. Con la finalidad de ilustrar el tipo de cálculos que podemos hacer con el resolvente consideremos el siguiente ejercicio.

Nota 6. *Para el resolvente usamos la notación usual de la transformada de Laplace, ver* (2.38).

Ejemplo 14. *Sea* X *una* $\mathsf{CM}_{A,B,C}(\mu, \mathsf{Q})$ *con*

$$\mathsf{Q} = \begin{vmatrix} -3 & 1 & 2 \\ 0 & -2 & 2 \\ 2 & 1 & -3 \end{vmatrix}.$$

i) Calcula $\mathbb{P}_A[X_t = A]$

ii) Calcula $\mathbb{P}_C[X_t = B]$

iii) Calcula $\mathbb{P}_C[X_T = B]$ *con* $T \sim \mathsf{Exp}(4)$.

Para resolver (i), *notemos que para* p_{AA} *corresponde un* $r_{AA}(\lambda)$ *como la entrada correspondiente de* $\mathsf{R}(\lambda) = (\lambda\mathbb{I} - \mathsf{Q})^{-1}$. *Entonces*

$$(\lambda\mathbb{I} - \mathsf{Q}) = \begin{vmatrix} \lambda+3 & -1 & -2 \\ 0 & \lambda+2 & -2 \\ -2 & -1 & \lambda+3 \end{vmatrix}.$$

con $\det(\lambda\mathbb{I} - \mathsf{Q}) = \lambda(\lambda+3)(\lambda+5)$. *Ahora de* (4.9) *tenemos que la inversa en la entrada "AA" esta dada por*

$$r_{AA}(\lambda) = (\lambda\mathbb{I} - \mathsf{Q})^{-1}_{AA} = \frac{\det\begin{vmatrix} \lambda+2 & -2 \\ -1 & \lambda+3 \end{vmatrix}}{\det(\lambda\mathbb{I} - \mathsf{Q})} = \frac{\lambda^2 + 5\lambda + 4}{\lambda(\lambda+3)(\lambda+5)}.$$

Para invertir la transformada de Laplace usamos fracciones parciales, i.e. resolver

$$\lambda^2 + 5\lambda + 4 = \alpha(\lambda+3)(\lambda+5) + \beta\lambda(\lambda+5) + \gamma\lambda(\lambda+3)$$

lo cual resulta en $\alpha = 4/15$ *(con* $\lambda = 0$*),* $\beta = 1/3$ *(con* $\lambda = -3$*) y* $\gamma = 2/5$ *(con* $\lambda = -5$*). Por lo tanto*

$$r_{AA}(\lambda) = \frac{4}{15\lambda} + \frac{1}{3(\lambda+3)} + \frac{2}{5(\lambda+5)}.$$

Ahora, recordemos que la inversa de la transformada de Laplace de $1/(\lambda+\beta)$ *es* $e^{-\beta t}$, *por lo que*

$$\mathsf{p}_{AA}(t) = \frac{4}{15} + \frac{1}{3}e^{-3t} + \frac{2}{15}e^{-rt}.$$

Nótese que podemos verificar que $\mathsf{p}_{AA}(0) = 1$.

De manera análoga para resolver (ii) *tenemos que*

$$r_{CB}(\lambda) = \frac{-\mathsf{det}\begin{vmatrix} \lambda+3 & -1 \\ -2 & -1 \end{vmatrix}}{\lambda(\lambda+3)(\lambda+5)} = \frac{1}{3}\left(\frac{1}{\lambda} - \frac{1}{\lambda+3}\right)$$

lo que invirtiendo la transformada conlleva a

$$\mathbb{P}_C[X_t = B] = \mathsf{p}_{CB}(t) = \frac{1}{3} - \frac{1}{3}e^{-3t}.$$

Ahora, si quisiéramos calcular la probabilidad de que la cadena se encuentre en un estado dado después de que un tiempo independiente $T \sim \mathsf{Exp}(4)$ *a transcurrido, basta con evaluar el resolvente, pues tenemos la interpretación arriba señalada. Por ejemplo, podemos verificar que* $\mathsf{p}_{CA}(T) = 4/21$, $\mathsf{p}_{CB}(T) = 1/7$ *y* $\mathsf{p}_{CC}(T) = 2/3$.

Otra de las utilidades que se le puede dar al operador resolvente es para la densidad de la variable aleatoria que representa el tiempo de primera llegada a al estado j. Específicamente, si como antes, denotamos por

$$\rho_j = \inf\{t > 0 : X_t = j\}$$

y definimos

$$\mathsf{F}_{ij}(t) = \mathbb{P}_i(\rho_j < t)$$

la probabilidad de que la cadena toque el estado j, por primera vez, antes de $t \geq 0$, entonces tenemos la función matricial

$$\mathsf{F}(t) = (\mathsf{F}_{ij}(t) : i, j \in \mathbb{S}, t \geq 0)$$

con $\mathsf{F}(0) = \mathbb{I}$ y

$$\mathsf{F}_{ij}(t) = \int_0^t \mathsf{f}_{ij}(s)ds = \int_0^t \mathbb{P}_i(\rho_j \in ds).$$

La matriz $\mathsf{f}(t) = (\mathsf{f}_{ij}(t); t \geq 0)$ y $\mathsf{P}(t)$ se relacionan mediante

$$\begin{aligned}
\mathsf{p_{ij}(t)} &= \mathbb{P}_i(X_t = j) = \mathbb{P}_i(X_t = j, \rho_j \leq t) = \int_0^t \mathbb{P}_i(X_t = j, \rho_j \in ds) \\
&= \int_0^t \mathbb{P}_i(\rho_j \in ds)\mathbb{P}_j(X_{t-s} = j) = \int_0^t \mathsf{f}_{ij}(s)\mathsf{p}_{jj}(t-s)ds \\
&= \mathsf{f}_{ij} * \mathsf{p}_{jj}(t)
\end{aligned}$$

donde $*$ denota el operador convolución entre dos funciones. Así pues, la transformada de Laplace para la densidad de $\mathsf{f}_{ij}(t)$ esta dada por

$$\hat{f}_{ij}(\lambda) = \frac{r_{ij}(\lambda)}{r_{jj}(\lambda)}, \quad \text{for} \quad i, j \in \mathbb{S} \quad \text{y} \quad \lambda > 0. \tag{4.31}$$

En el contexto del Ejemplo 14, podríamos contestar preguntas del tipo

$$\mathbb{P}_C(\rho_A \leq t) = \mathsf{F}_{CA}(t) = \int_0^t f_{CA}(s)ds$$

encontrando primero la transformada de Laplace $\hat{f}_{CA}$ de f_{CA}, usando r_{CA} y r_{AA},

$$\hat{f}_{CA}(\lambda) = \frac{2(\lambda+2)}{\lambda^2 + 5\lambda + 4} = \frac{2}{3(\lambda+1)} + \frac{4}{3(\lambda+4)}$$

donde la última igualdad se sigue por fracciones parciales. Invirtiendo la transformada de Laplace, obtenemos

$$f_{CA}(t) = \frac{2}{3}e^{-t} + \frac{4}{3}e^{-4t}, \quad t \geq 0$$

e integrando resulta en

$$\mathbb{P}_C(\rho_A \leq t) = 1 - \frac{2}{3}e^{-t} - \frac{1}{3}e^{-4t}.$$

Como en el casó del resolvente, la transformadad de Laplace (4.31) admite también una interpretación probabilística directa como

$$\hat{f}_{ij}(\lambda) = \mathbb{P}_i(\rho_j < A), \quad \text{para} \quad i,j \in \mathbb{S}, \lambda > 0 \quad \text{y} \quad A \sim \mathsf{Exp}(\lambda).$$

lo que se puede calcular sin necesidad de invertir la transformada de Laplace. Por ejemplo, si $U \sim \mathsf{Exp}(\mu = 2.5)$

$$\mathbb{P}_B(\rho_C < U) = \hat{f}_{BC}(\mu) = \frac{r_{BC}(\mu)}{r_{CC}(\mu)} = \frac{2}{\mu + 2}$$

que es igual a $4/9$ cuando $\mu = 2.5$.

De manera análoga al caso de cadenas de Markov a tiempo discreto, si suponemos que nuestra cadena a tiempo continuo $(X_t)_{t\geq 0}$ es no-explosiva, i.e. Q es no-explosiva y $\boldsymbol{\pi}$ resuelve

$$\sum_{i\in\mathbb{S}} (\boldsymbol{\pi})_i \mathsf{q}_{ij} = 0, \quad (\boldsymbol{\pi})_i > 0 \quad \text{para todo} \quad j \in \mathbb{S} \quad \text{y} \quad \sum_{i\in\mathbb{S}} (\boldsymbol{\pi})_i = 1$$

entonces

$$\lim_{t\to\infty} \mathsf{p}_{ij}(t) = (\boldsymbol{\pi})_j, \quad \text{para todo} \quad i \in \mathbb{S}$$

y si $\eta_j(t) = \int_0^t \mathbf{1}(X_s = j)ds$ denota el tiempo de ocupación en el estado j antes de t, tenemos

$$\lim_{t\to\infty} \frac{\eta_j(t)}{t} = (\boldsymbol{\pi})_j \quad \text{con probabilidad uno.}$$

De la misma manera si $\mathsf{m} = (\mathsf{m}_i : i \in \mathbb{S})$, con $\mathsf{m}_i > 0$ es simetrizable, i.e. satisface las ecuaciones de balance detallado

$$\mathsf{m}_i \mathsf{q}_{ij} = \mathsf{m}_j \mathsf{q}_{ji}, \quad \text{para toda} \quad i,j \in \mathbb{S}.$$

con $M := \sum_{i\in\mathbb{S}} \mathsf{m}_i < \infty$ entonces

$$(\boldsymbol{\pi})_i = \frac{\mathsf{m}_i}{M}$$

define una única distribución invariante.

Ejercicio 48. *Sea* X *una cadena de Markov a tiempo continuo con* $\mathbb{S} = \mathbb{N}_0$ *y matriz generadora* Q *como en el Ejercicio 46. Encuentra para que valores de* μ *y* λ *existe la distribución invariante* $\boldsymbol{\pi}$ *y exhibe su forma en dichos casos.*

Ejercicio 49 (Proceso de nacimiento puro)**.** *Supóngase que tenemos una cadena de Markov a tiempo continuo,* $\mathsf{X} = (X_t)_{t\geq 0}$ *con* $X_0 = 1$ *y*

$$\mathsf{Q} = \left| \begin{array}{cccccc} -\lambda & \lambda & 0 & 0 & \cdots & \\ 0 & -2\lambda & 2\lambda & 0 & \cdots & \\ 0 & 0 & -3\lambda & 3\lambda & 0 & \\ 0 & \cdots & 0 & \ddots & \ddots & \ddots \end{array} \right|$$

Demuestra que $\mathbb{P}_1[X_t = n] = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$ *usando el método vía el operador resolvente.*

Ejercicio 50. *Sea* $(X_t)_{t\geq 0}$ *una cadena de Markov con espacio de estados* $\mathbb{S} = \{1,2,3\}$ *y matriz infinitesimal*

$$\mathsf{Q} = \begin{pmatrix} -4 & 2 & 2 \\ 1 & -3 & 2 \\ 1 & 0 & -1 \end{pmatrix}$$

a) *Encuentra* $\mathsf{P}(X_t = 2 \mid X_0 = 1)$.

b) *Si* ρ_3 *es la primera vez que la cadena entra al estado* 3*, encuentra* $\mathsf{P}_2(\rho_3 \leq t)$. *Hint. Puedes usar la solución al problema anterior.*

c) *Sea* $T \sim$ *Exp*(λ) *independiente de la cadena de Markov. Dado que* $X_0 = 3$ *encuentra la distribución de la cadena al tiempo* T*, i.e.*

$$\mathbb{P}_3(X_T = i) \quad \textit{para} \quad i \in 1,2,3$$

d) *Calcula las probabilidades del inciso anterior cuando (i)* $\lambda = 2$*, y cuando (ii)* $\lambda \downarrow 0$. *¿Cómo se compara este último caso con la distribución límite?*

Ejercicio 51. *Una planta de producción obtiene su energía de dos generadores. El número de generadores activos al tiempo* t *esta representado por una cadena de Markov a tiempo continuo* $\{X_t : t \geq 0\}$ *con espacio de estados* $\mathbb{S} = \{0,1,2\}$ *y matriz* Q *dada por*

$$\mathsf{Q} = \begin{pmatrix} -6 & 6 & 0 \\ 1 & -7 & 6 \\ 0 & 2 & -2 \end{pmatrix}$$

a) *Encuentra la proporción de tiempo a largo plazo de que los dos generadores estén trabajando. Si cada generador produce 2.5 MW por unidad de tiempo, ¿Cuál es la cantidad promedio de energía producida a largo plazo por unidad de tiempo?*

b) *¿Cuál es la probabilidad de que ambos generadores estén trabajando al tiempo* t *si solamente uno trabajaba al tiempo 0?*

c) *Si* ρ_2 *denota la primera vez que ambos generadores estén trabajando al mismo tiempo, encuentra la distribución* $F_{02}(X) = \mathsf{P}_0(\rho_2 \leq x)$*, es decir la distribución de que ambos generadores trabajen por primera vez, dado que los dos estaban sin trabajar al tiempo* $t = 0$.

Ejercicio 52. *Resolver los ejercicios 2.4.1., 2.4.3., 2.4.5., 2.7.1., 3.3.1. y 3.6.2. del libro "Norris, J.R. (1997). Markov chains. Cambridge University Press."*

TEMA 5

UNA BREVE DISCUSIÓN DE PROCESO POISSON Y ALGUNAS GENERALIZACIONES

Hemos visto diferentes formas de definir un proceso Poisson, para distinguirlo, lo denotaremos mediante $\mathsf{N} := (N_t; t \geq 0)$, donde N representa un número de eventos. Veamos otra definición de este proceso, como caso particular de la Definición 16.

Definición 25. Un proceso de Lévy, $\mathsf{N} := (N_t; t \geq 0)$, con $\mathbb{P}[N_t = n] = \mathsf{Po}(n; \lambda t)$ se dice que es un proceso Poisson simple con intensidad λ, $\mathsf{PPS}(\lambda)$.

Ejercicio 53. *Demuestra las siguientes propiedades de la distribución Poisson:*

1. *Si $z \in \mathbb{C}$ o $x \in \mathbb{R}$ con $|z| \leq 1$, la variable aleatoria z^X, $X \sim \mathsf{Po}(\mu)$, es acotada y $\mathbb{E}\left[z^X\right] = e^{\mu(z-1)}$*
2. *Si $(x)_{n\downarrow} := x(x-1)\cdots(x-n+1)$ entonces $\mathbb{E}\left[(X)_{n\downarrow}\right] = \mu^n$*
3. $$\sum_{k=0}^{n} \mathsf{Po}(k, \mu) = 1 - \int_0^{\mu} \mathsf{Po}(n; \lambda) \mathrm{d}\lambda$$
4. *Si $X_i \overset{ind}{\sim} \mathsf{Po}(\mu_i)$, $i = 1, \ldots, n$ entonces*
$$S_n = \sum_{i=1}^{n} X_i \sim \mathsf{Po}(\mu),$$
donde $\mu := \sum_{i=1}^{n} \mu_i$ y
$$\mathbb{P}\{X_1 = r_1, X_2 = r_2, \ldots, X_n = r_n \mid S_n = s\} = \frac{s!}{r_1!\, r_2! \ldots r_n!} \left(\frac{\mu_1}{\mu}\right)^{r_1} \left(\frac{\mu_2}{\mu}\right)^{r_2} \cdots \left(\frac{\mu_n}{\mu}\right)^{r_n}, \tag{5.1}$$
siempre que $\sum_{i=1}^{n} r_i = s$. Un vector aleatorio, n-dimensional, con ley (5.1) se dice que sigue una distribución multinomial, $\mathsf{Mult}(s; p_1, \ldots, p_n)$, donde $p_i := \frac{\mu_i}{\mu}$.
5. *Si $X_i \overset{ind}{\sim} \mathsf{Po}(\mu_i)$, $i = 1, \ldots$ y $\sigma = \sum_{i=1}^{\infty} \mu_i$ converge, entonces $S := \sum_{i=1}^{\infty} X_i \overset{p1}{\to} \mathsf{Po}(\sigma)$. De lo contrario, si σ diverge, entonces S diverge con probabilidad 1.*

De la Definición 25, para cada trayectoria de un $\mathsf{PPS}(\lambda)$, se pueden definir los tiempos de brincos, $0 = \mathsf{T}_0 < \mathsf{T}_1 < \cdots < \mathsf{T}_n$, donde $\mathsf{T}_k := \inf\{t \geq 0 : \mathsf{N}_t \geq k\}$, o equivalentemente $N_t = \max\{k \geq 0; \mathsf{T}_k \leq t\}$. Usando la notación $\Delta X_t := X_{t+} - X_{t-}$, y de la misma Definición 25, se puede ver que $\Delta N_{T_n} = 1$ c.s., con $T_n \sim \mathsf{Ga}(n, \lambda)$. En efecto, de la Ejercicio 45 vemos que, denotando por S_n en n-ésimo tiempo interarribo, se satisface $S_n \overset{\text{iid}}{\sim} \mathsf{Exp}(\lambda)$, y por lo tanto $\mathsf{T}_n = \sum_{j=1}^{n} S_j$. De hecho, de la distribución conjunta de (T_n, S_{n+1}) se ve inmediatamente que $\mathbb{P}[N_t = n] = \mathbb{P}[\mathsf{T}_n \leq t < \mathsf{T}_n + S_{n+1}] = \mathsf{Po}(n; \lambda t)$.

Ejercicio 54. *Verificar que* $\mathbb{P}[N_t = n] = \mathbb{P}[\mathsf{T}_n \leq t < \mathsf{T}_n + S_{n+1}] = \mathsf{Po}(n; \lambda t)$.

Ejercicio 55. *Sea* $\tilde{N}_t := N_t - \lambda t$*, i.e. el* **proceso Poisson compensado***. Demostrar que* $\tilde{\mathsf{N}} = (\tilde{N}_t)_{t\geq 0}$ *es también un proceso con incrementos independientes, que tiene la propiedad de martingala, y que su versión re-escalada* $\tilde{N}_t/\sqrt{\lambda}$ *tiene media cero y varianza* t*.*

Es decir, el proceso Poisson simple tiene brincos de tamaño uno a tiempos exponenciales con intensidad λ. En este sentido el $\mathsf{PPS}(\lambda)$ se puede ver como un ***proceso de conteo*** de eventos que ocurren en un intervalo dado, i.e. N_t cuenta el número de saltos en $(0, t]$. Sus realizaciones son no-decrecientes, i.e. $N_t \leq N_{t+h}$ y, de la Definición 25, $N_t \stackrel{\mathrm{d}}{=} N_{t+s} - N_s$, es decir el número de eventos en el intervalo $(s, t+s]$ tiene la misma distribución que el número de eventos en el intervalo $(0, t]$. Evidentemente, $\mathbb{E}[N_t] = \lambda t$ y $\mathsf{Var}(N_t) = \lambda t$, lo que da la interpretación a λ como el promedio del número de eventos (brincos) por unidad de tiempo, e.g. $t = 1$.

Claramente, el proceso no tiene trayectorias continuas, sin embargo $\mathbb{P}(N_{t+h} - N_t \geq 1) = 1 - e^{-\lambda h} \to 0$, cuando $h \downarrow 0$. Esto implica que $\mathbb{P}(T_k = t) = 0$, para toda $t \geq 0$ y $k = 1, 2, \ldots$, y por lo tanto N no tiene puntos fijos de discontinuidad. Esto último caracteriza a los procesos Poisson dentro de la familia de procesos de saltos markovianos.

Hemos visto anteriormente el $\mathsf{PPS}(\lambda)$ sirve para definir otros procesos de interés. En la construcción (7), teníamos que dada una cadena de Markov $\{Z_n\}$ con espacio de estados $(\mathbb{X}, \mathcal{X})$ y un $\mathsf{PPS}(\lambda)$, podíamos construir un proceso regular a tiempo continuo definido mediante $X_t := Z_{N_t}$ y modulado por probabilidades de transición dadas por

$$P_t(x, A) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathsf{P}^{(n)}(x, A), \quad x \in \mathbb{X} \quad y \quad A \in \mathcal{X} \tag{5.2}$$

Un caso particular de esta construcción es el siguiente

Definición 26. Sea N un $\mathsf{PPS}(\lambda)$ y $\{Y_i\}_{i\geq 0}$ una sucesión de variables aleatorias iid con distribución F sobre $\mathbb{R}_+$, independiente de N. Al proceso definido por

$$X_t := \sum_{i=1}^{N_t} Y_i \tag{5.3}$$

se le denomina ***proceso Poisson compuesto*** con intensidad λ y distribución de brincos F, $\mathsf{PPC}(\lambda, F)$.

Ejercicio 56. *Sea la caminata aleatoria* $Z_n := Y_1 + \cdots + Y_n$*, con* $Z_0 := 0$*,* $Y_i \stackrel{iid}{\sim} F$ *y* F *soportada en* $\mathbb{R}_+$*. Demuestra que el proceso definido como* $X_t := Z_{N_t}$*, "la posición de la caminata después de un número aleatorio de pasos modulado por* N_t*", coincide con el PPC en la Definición 26 y su probabilidad de transición es un caso particular de (5.2). Exhibe la expresión para esta última.*

En vez de tener saltos de tamaño uno como en el caso del $\mathsf{PPS}(\lambda)$, el $\mathsf{PPC}(\lambda, F)$ tiene saltos de tamaño aleatorio, con distribiución F. De hecho, si $Y_i := 1$ para toda $i \geq 1$, se recupera el $\mathsf{PPS}(\lambda)$.

Otro ejemplo que surge del $\mathsf{PPS}(\lambda)$, es de su versión compensada y re-escalada (cf. Ejercicio 55), cuando se toma el límite $\lambda \to \infty$. En efecto, se puede checar que

$$\left(\frac{\tilde{N}_t}{\sqrt{\lambda}}\right)_{t\in[0,T]} \stackrel{\lambda \to \infty}{\Rightarrow} (W_t)_{t\in[0,T]} \tag{5.4}$$

donde $\mathsf{W} := (W_t)_{t\geq 0}$ es un proceso de difusión (cf. Teorema 10) con media y varianza infinitesimal 0 y 1, respectivamente, conocido como el movimiento browniano (ver también Ejemplo 5 para su generador). El resultado (5.4) se puede ver como un teorema central del límite en el espacio de funciones regulares en $[0, T]$.

Además de que estos tres procesos, Poisson simple, Poisson compuesto y movimiento browniano, son procesos con incrementos independientes, también se puede verificar que son procesos de Feller. Tomemos el caso de un PPC, del Ejercicio 56 se puede verificar que la caminata aleatoria Z_n tiene incrementos independientes y estacionarios, i.e. Y_i's por unidad de tiempo. El incremento en m unidades de tiempo esta dado por $Z_{n+m} - Z_n$, es decir, sumas de m variables aleatorias iid. Así pues, querer caracterizar la distribución de procesos de Lévy, equivale a caracterizar las variables aleatorias con la característica que se puedan descomponer en sumas de variables aleatorias iid. Nótese que no todas las distribuciones tienen esta característica.

Antes de continuar nuestra discusión sobre procesos Poisson, daremos algunos conceptos, que nos ayudarán a contextualizar su importancia.

Definición 27. Una variable aleatoria Y es ***infinitamente divisible*** (ID) si para cada $m \geq 1$, podemos escribir

$$Y \stackrel{\mathrm{d}}{=} Y_1^{(m)} + \cdots + Y_m^{(m)}, \tag{5.5}$$

para una sucesión de variables aleatorias iid $Y_1^{(m)}, \ldots, Y_m^{(m)}$.

Si la igualdad (5.5) es válida cuando las $Y_i^{(m)}$s son independientes pero no necesariamente idénticamente distribuidas, entonces Y se dice que es ***infinitamente descomponible***. Si la igualdad no se cumple para toda $m \geq 1$, sino solamente para un m fijo, entonces se tiene que Y es m***-divisible***, el mismo argumento aplica para m***-descomponible***.

Ejemplo 15. *Si $X \sim Bin(3,p)$, entonces X es 3-divisible pero no ID, en particular no es 2-divisible, ya que si lo fuese se valdría $X \stackrel{d}{=} Y_1 + Y_2$ con $Y_1 \stackrel{d}{=} Y_2 \stackrel{d}{=} Y$. Entonces, necesariamente $Y \in [0, \frac{3}{2}]$ c.s. con $P(Y = 0) > 0$ y $P(Y = \frac{3}{2}) > 0$, lo cual daría $\frac{3}{2}$ para X (que debería tener soporte en $0, 1, 2, 3$). Entonces la variable aleatoria Binomial no es ID.*

De este ejemplo es claro que $(n+1)-$divisibilidad no implica $(n)-$divisibilidad. Una distribución F is ID si y solo si para cada $n \in \mathbb{N}$ se puede ver como la n-ésima convolución de una distribución F_n consigo misma. Asimismo, la función característica $\mathcal{C}$ es ID si y solo si para cada $n \in \mathbb{N}$ se puede ver como la n-ésima potencia de una función característica ϕ_n.

$$F = F_n^{*n}, \quad \mathcal{C}(u) = \{\mathcal{C}_n(u)\}^n, \text{ for } n \in \mathbb{N} \tag{5.6}$$

Nótese que la propiedad de ID no es la misma a la de cerradura bajo convolución. En particular se puede demostrar que la distribución log-normal es ID y ésta distribución no es cerrada bajo sumas, es decir sumas de log-normales no son log-normal.

Ejercicio 57. *Demostrar que si X y Y son variables aleatorias independientes e ID entonces aX y $X + Y$ lo son, con $a \in \mathbb{R}$.*

La propiedad de distribuciones infinitamente divisible se está relacionada con los procesos con incrementos independientes (PII), por ejemplo si nos concentramos en las distribuciones finito dimensionales correspondientes a un PII $\{X_t; t \geq 0\}$, notamos que para un $t > 0$ fijo

$$X(t) = \sum_{j=1}^{n} \{X(jt/n) - X((j-1)t/n)\},$$

donde los sumandos son iid, y entonces, por definición ID. Denotemos por $\mathcal{C}_t$ la función característica de $X(t)$. Dado que se tiene $X_{s+t} = X_s + \{X_{s+t} - X_s\}$, entonces

$$\mathcal{C}_{s+t}(u) = \mathcal{C}_s(u)\mathcal{C}_t(u), \quad t, s > 0.$$

Ahora para cada $u \in \mathbb{R}$ la función $t \mapsto \mathcal{C}_t(u)$ es continua en $\mathbb{R}_+$, y por lo tanto se tiene que

$$\mathcal{C}_t(u) = \{\mathcal{C}_1(u)\}^t, \quad t \geq 0. \tag{5.7}$$

Finalmente, si $0 = t_0 < t_1 < \cdots < t_n$ con $n \geq 2$, entonces por la propiedad de incrementos independiente y estacionarios tenemos

$$\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} = \{Y_1, Y_1 + Y_2, \ldots, Y_1 + Y_2 + \cdots + Y_n\}, \tag{5.8}$$

donde los Y_is son iid con $Y_i \stackrel{\mathrm{d}}{=} X(t_i - t_{i-1})$ para $i = 1, \ldots, n$.

Con (5.7) y (5.8) se puede enunciar los siguientes resultados.

Proposición 5. *La distribución de un proceso de Lévy $(X_t)_{t\geq 1}$ esta completamente determinada por la distribución de $X(1)$, que es ID.*

Proposición 6. *Sea Y una variable aleatoria ID, entonces existe un proceso de Lévy X con $X(1) \stackrel{d}{=} Y$.*

Teorema 21 (Representación Lévy-Khintchine)**.** *La función característica de un proceso con incrementos independientes (PII) –ver Definición 16–, $\mathsf{X} := (X_t; t \geq 0)$ sobre $\mathbb{R}$ esta dada por*

$$\mathbb{E}\left[e^{i\eta X_t}\right] = \mathbb{E}\left[e^{i\eta X_0}\right] \exp\{-\psi_t(\eta)\} \tag{5.9}$$

donde ψ_t se denomina el **exponente característico** *y esta dado por*

$$\psi_t(\eta) = i\eta d_t + \frac{\eta}{2}G_t + \int_{\mathbb{R}} (e^{i\eta x} - 1 - h(x))\nu_t(dx) \tag{5.10}$$

Aquí: d_t es una función continua en $\mathbb{R}$ denominada el **término de deriva***; G_t es un operador lineal simétrico no-negativo, con $G_{t_2} - G_{t_1}$ no-negativo para $t_1 < t_2$, conocido como el* **término gaussiano***; ν_t es una medida positiva que satisface $\int_{\mathbb{R}}(1 \wedge x^2)\nu_t(dx) < \infty$ conocida como la* **medida de Lévy***; y h una función de truncamiento acotada, con soporte compacto y que se comporta como x cerca del origen. Esta última, típicamente se remplaza por $x\,\mathbb{I}(|x| \leq 1)$. A un PII con estas características se le refiere colectivamente mediante la terna (d_t, G_t, ν_t). Si se trata de un PII que comienza en $X_0 = 0$ c.s., se tiene $(d_t, G_t, \nu_t) = (td, tG, t\nu)$, y el exponente característico se simplifica como $\psi_t = t\psi$, donde ψ esta dado por (5.10) con t removida. En el caso de procesos de Lévy, a ψ se le conoce como el exponente característico.*

En este sentido, y en relación con la Proposición 6, la dinámica distribucional de los procesos de Lévy están caracterizada por distribuciones ID, cuya función característica esta dada por $e^{-\psi(\eta)}$.

Proposición 7 (Theorem 9.1 (Sato, 1999))**.** *Si $(X_t)_{t\geq 1}$ es un PII que comienza en 0, entonces para cada t, la distribución de X_t es infinitamente divisible.*

Este resultado establece la equivalencia distribucional de medidas ID con la de PII que comienzan en cero. Como hemos visto en el curso, la equivalencia en ley no necesariamente garantiza la correspondencia con un proceso con trayectorias regulares. Sin embargo, en el caso de un proceso como el aquí enunciado se puede ver existe una modificación regular.

Proposición 8 (Theorem 11.5 Sato (1999))**.** *Si $(X_t)_{t\geq 1}$ es un PII que comienza en 0 en ley, entonces tienen una modificación regular que corresponde a un PII con $X_0 = 0$.*

De hecho como mencionamos anteriormente, se puede demostrar que estas clases de procesos, son de Feller y por lo tanto satisfacen todas las propiedades de esta clase general.

La teoría de PII y distribuciones ID es basta, con excelentes referencias bibliográficas como los libros de Sato (1999) y Steutel and Van Harn (2004). Claramente, su estudio no es el objetivo central de este curso y únicamente discutimos algunos aspectos elementales. En este sentido únicamente consideraremos PII crecientes y definidos sobre $\mathbb{R}_+$. En este caso es menos engorroso trabajar con la transformada de Laplace. En efecto, la transformada de Laplace[1] en este caso esta dada por

$$\mathbb{E}\left[e^{-\eta X_t}\right] = \mathbb{E}\left[e^{-\eta X_0}\right] \exp\left\{-d_t\eta - \int_0^\infty (1 - e^{-\eta x})\nu_t(dx)\right\} \tag{5.11}$$

donde $\int_{\mathbb{R}_+}(1 \wedge x)\nu_t(dx) < \infty$. Para el caso de distribuciones ID no-negativas la representación de Lévy-Khintchine simplifica (5.11) con $X_0 = 0$. Al parámetro d, resultante de esta última simplificación, y en el contexto de variables aleatorias ID, se le conoce como la extremidad izquierda, dígase el soporte arranca desde ese valor.

[1] En partes previas del curso usamos el argumento de la transformada en los complejos, aquí lo usaremos de la forma “clásica”, i.e. $\mathbb{E}e^{-\lambda X}$, con $\lambda > 0$.

Si definimos $F(\mathrm{d}x) := \nu(\mathrm{d}x)/\lambda$ con $\lambda := \nu(\mathbb{R}_+) < \infty$, entonces se puede ver que un $\mathsf{PPC}(\lambda, F)$ es un caso particular de un proceso de Lévy con terna $(0, 0, \nu)$. En efecto, por la Proposición 6, basta con considerar la variable aleatoria Poisson compuesta

$$X = \sum_{i=1}^{N} Y_i,$$

donde $\{Y_i\} \overset{\text{iid}}{\sim} F$ y $N \sim \mathsf{Po}(\lambda)$.

$$\begin{aligned}
\mathcal{L}_X(\xi) &= \mathbb{E}\left\{[\mathcal{L}_Y(\xi)]^N\right\} = \exp\left\{\lambda\left[\mathcal{L}_y(\xi) - 1\right]\right\} \\
&= \exp\left\{\lambda\left[\int_{\mathbb{R}_+} e^{-\xi y}\, F(\mathrm{d}y) - 1\right]\right\} \\
&= \exp\left\{\lambda\left[\int_{\mathbb{R}_+} \lambda^{-1}\, e^{-\xi y}\, \nu(\mathrm{d}y) - \int_{\mathbb{R}_+} \lambda^{-1}\nu(\mathrm{d}y)\right]\right\} \\
&= \exp\left\{\int_{\mathbb{R}_+} (e^{-\xi y} - 1)\, \nu(\mathrm{d}y)\right\} \\
&= \exp\left\{-\psi(\xi)\right\} \qquad (5.12)
\end{aligned}$$

con

$$\psi(\xi) = \int_{\mathbb{R}_+} (1 - e^{-\xi y})\, \nu(\mathrm{d}y).$$

Es decir X es ID, y equivalentemente el $\mathsf{PPC}(\lambda, F)$ un proceso de Lévy caracterizado por la terna $(0, 0, \nu)$, o equivalentemente por la transformada de Laplace $e^{-t\psi(\xi)}$. Claramente en este caso la medida de Lévy, i.e. proporcional a F, controla los saltos del proceso y satisface $\nu(\mathbb{R}_+) < \infty$, y por lo tanto $\int_{\mathbb{R}_+}(1 \wedge x)\nu(dx) < \infty$. De hecho, debido a esto, podemos decir que el número de brincos que un PCC hace en un intervalo finito es finito y esta regulado por ν.

Sin embargo, unos de los casos particulares de procesos de Lévy, mas atractivos en aplicaciones reales, son cuando el número de brincos en un intervalo finito es infinito. Analicemos el siguiente ejemplo.

Ejemplo 16. *Sea $\{X_t; t \geq 0\}$ un* **proceso gamma** *con $X_0 := 0$, es decir un proceso de Lévy con densidad de transición $p_t(0, x) = \mathsf{Ga}(x; t, b)$, $b > 0$. Entonces*

$$\begin{aligned}
\mathcal{L}_{X_t}(\xi) &= \mathbb{E}[e^{-\xi X_t}] = \int_0^\infty e^{-\xi x}\mathsf{Ga}(x; t, b) = (1 + \xi/b)^{-a} \\
&= \exp\left\{-t\ln(1 + \xi/b)\right\} = \exp\left\{-t\int_0^\xi \frac{\mathrm{d}z}{b + z}\right\} \\
&= \exp\left\{-t\int_0^\xi \mathrm{d}z \int_0^\infty e^{-bx - zx}\mathrm{d}x\right\} \\
&= \exp\left\{-t\int_0^\infty (1 - e^{-\xi x})\frac{e^{-bx}}{x}\mathrm{d}x\right\}, \qquad \textit{ya que } \left(\int_0^\xi e^{-zx}\mathrm{d}z = \frac{(1 - e^{-\xi x})}{x}\right) \qquad (5.13)
\end{aligned}$$

Por lo tanto, la medida de Lévy correspondiente esta dada por $\nu(\mathrm{d}x) = e^{-bx}/x\,\mathrm{d}x$. Nótese que en este caso $\mathrm{d} = 0$, y por lo tanto se trata de un proceso de Lévy con terna $(0, 0, \nu)$.

Aunque se cumple la condición de que $\int_{\mathbb{R}_+}(1 \wedge x)\nu(\mathrm{d}x) < \infty$, $\nu(\mathbb{R}_+) = \infty$ pues hay demasiados brincos pequeñitos, i.e. $\nu([0, \epsilon]) = \int_0^\epsilon \nu(\mathrm{d}x) = \infty$, no importa que tan pequeño sea $\epsilon > 0$. Claramente este proceso no corresponde a un PPC, ya que tiene demasiados brincos para ser capturados por una medida finita.

Si $\lambda := \nu(\mathbb{R}_+) = \infty$, podemos definir $\nu_\varepsilon(B) := \nu((\varepsilon, \infty) \cap B)$ y equivalentemente $\lambda_\varepsilon := \nu_\varepsilon(\mathbb{R}_+)$, $F_\varepsilon(\mathrm{d}x) := \nu_\varepsilon(\mathrm{d}x)/\lambda_\varepsilon$ y la variable Poisson compuesta correspondiente X_ε. La elección de ε se podría hacer mediante una ponderación de los valores de λ_ε y la variable error ϵ_ε con transformada de Laplace dada por $\exp\{\int_{(0,\varepsilon]}(e^{-\xi y} - 1)\, \nu(\mathrm{d}y)\}$.

Es importante señalar que el proceso gamma del Ejemplo 16 es un proceso con tasa de saltos infinita, a veces referidos como procesos de Lévy con ***actividad infinita***, y es muy diferente a un PCC con distribución de brincos F gamma, ya que este último tiene ***actividad finita***.

Ejercicio 58. *Sea* $\{X_t; t \geq 0\}$ *un proceso de Lévy con valores en* $\mathbb{R}_+$ *no-decreciente y con terna* $(0, 0, \nu)$*, donde* $\nu(\mathrm{d}x) = x^{\alpha-1}e^{-\beta x}\mathrm{d}x$*, con* $-1 < \alpha < \infty$ *y* $\beta > 0$

1. *¿Para que valores de* α*,* $\{X_t; t \geq 0\}$ *se puede ver como un proceso Poisson compuesto? En caso afirmativo, ¿cuál es la distribución de los " brincos " correspondientes?.*

2. *Si* $\beta = 0$ *y* $\alpha \in (-1, 0)$*, ¿Cuál es el exponente característico,* ψ*, correspondiente?*

3. *¿Para cuales valores de* α*,* $\{X_t; t \geq 0\}$ *es un proceso de Lévy con " actividad infinita "?*

Ejercicio 59. *Sea* $\{X_t; t \geq 0\}$ *un proceso de Lévy con valores en* $\mathbb{R}_+$ *no-decreciente y con terna* $(0, 0, \nu)$*, donde*

$$\nu(\mathrm{d}x) = \frac{(1 - e^{-\gamma x})}{(1 - e^{-x})} \frac{e^{-x}}{x} \mathrm{d}x, \quad \gamma > 0.$$

1. *¿Es* $\nu(\mathbb{R}_+) < \infty$*?*

2. *¿Cuál es* $\mathbb{E}[X_1]$*, cuando* $\gamma = 1/2, 1, 2$*?*

3. *Si* $\gamma \in \mathbb{N}$*, ¿Puedes decir algo acerca de la distribución de* X_1*?*

Ejercicio 60 (Ejercicio optativo). *Escoge entre los ejemplos/ejercicios anteriores (o propón otro nuevo) un proceso de Lévy con terna característica* $(0, 0, \nu)$ *y con actividad infinita. Realiza un programa de simulación e ilustra diferentes escenarios de sus trayectorias.*

Bibliography

Ethier, S. N. and T. G. Kurtz (1986). *Markov processes – characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc.

Gikhman, I. I. and A. V. Skorokhod (1969). *Introduction to the Theory of Random Processes*. Dover.

González-Barrios Murguía, J. M. (2011). *Lecture notes on probability theory.* Monografías-IIMAS. Universidad Nacional Autónoma de México.

Norris, J. R. (1998). *Markov chains.* Cambridge series in statistical and probabilistic mathematics. Cambridge University Press.

Sato, K. I. (1999). *Lévy processes and infinitely divisible distributions.* Cambridge: Cambridge University Press.

Steutel, F. W. and K. Van Harn (2004). *Infinite divisibility of probability distributions on de real line*. New York: Marcel Dekker.

Stroock, D. W. (2005). *An Introduction to Markov Processes.* Springer.

Williams, D. (1991). *Probability with Martingales.* Cambridge mathematical textbooks. Cambridge University Press.