

Convergencia débil de procesos Galton–Watson

Carlos Uriel Herrera Espinanza Imanol Nuñez Morales

8 de junio de 2021

Resumen

En el presente trabajo se demuestra que al reescalar ciertas sucesiones de procesos de Galton–Watson, obtendremos que estos convergen débilmente a una difusión de Feller. Para lograr esto el trabajo se estructuró de la siguiente manera: primero se da una revisión sucinta de los procesos de Galton–Watson y se muestran simulaciones que dan pie a la convergencia en cuestión; luego se da una introducción al espacio D , de funciones càdlàg, el se dota de la topología de Skorokhod; posteriormente se da un primer acercamiento a los procesos de Markov para poder definir lo que es una difusión; finalmente se prueba el resultado principal, la convergencia débil y se muestran simulaciones que la ilustran.

1. El proceso de Galton–Watson

Dentro de los procesos de ramificación, el primero con el que solemos encontrarnos en cursos básicos de procesos estocásticos es aquel conocido como proceso de *Galton–Watson*. Este proceso fue nombrado en honor a Sir Francis Galton y Henry William Watson quienes fueron de los pioneros de la teoría de procesos de ramificación al estudiar la probabilidad de extinción de apellidos de alcurnia, tema que concernía a las clases altas de la Inglaterra Victoriana. El mismo tema fue estudiado años antes por Bienaymé, ver Heyde y Seneta (1977), por lo que tal proceso también es conocido como proceso de *Bienaymé–Galton–Watson*. A pesar de tener su origen en un problema tan particular, este proceso y sus subsecuentes generalizaciones nos permiten modelar una población de partículas que sigan ciertas leyes de reproducción. ¿A qué se refiere esto?

Para responder esta pregunta, consideremos que una partícula, al momento de morir, da lugar a una cantidad aleatoria de descendientes, posiblemente nula, los cuales tendrán el mismo comportamiento que la partícula original y serán independientes del resto de partículas en el sistema. Un ejemplo de este sistema se muestra en la figura 1, donde los tiempos de vida de todas las partículas son iguales y constantes. En este ejemplo se muestran las primeras tres generaciones de la población de partículas y se observa que una de las tres partículas de la segunda generación muere sin tener descendientes. Ciertos objetos de interés surgen naturalmente en este sistema, como la cantidad de partículas en un tiempo dado, o la probabilidad de que el sistema se quede sin partículas en algún momento –evento que se conoce como *extinción*–, entre otros.

Si suponemos que el tiempo de vida de cada una de las partículas es igual a una unidad de tiempo, el estudiar la cantidad de partículas en el sistema en los tiempos $n \in \mathbb{N}_0$ bastará para conocer la cantidad de partículas en el sistema en cualquier momento del tiempo. De esta manera es que surge el proceso de Galton–Watson, el cual es un proceso estocástico que modela el tamaño de una población de un sistema como el descrito anteriormente. Formalmente, si consideramos que $\Xi = \{\xi_{n,m} : n, m \in \mathbb{N}\}$ es un doble arreglo de variables aleatorias

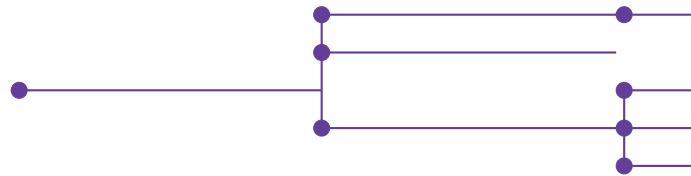


Figura 1: Ejemplo de un proceso de ramificación.

independientes e idénticamente distribuidas con valores en \mathbb{N}_0 , el proceso de Galton–Watson $\{Z_n : n \in \mathbb{N}_0\}$ se define mediante la relación de recurrencia

$$Z_{n+1} = \sum_{m=1}^{Z_n} \xi_{n+1,m}.$$

En este caso diremos que la progenie del proceso es Ξ . La relación anterior puede interpretarse como que la cantidad de individuos al tiempo $n + 1$ es la suma del número de descendientes que tuvo cada uno de los individuos que formaban el sistema al tiempo n . Visto de este modo, de la independencia del doble arreglo se obtiene inmediatamente

$$\begin{aligned} \mathbb{P}(Z_{n+1} = k_{n+1} \mid Z_n = k_n, \dots, Z_0 = k_0) &= \mathbb{P}\left(\sum_{m=1}^{k_n} \xi_{n+1,m} = k_{n+1} \mid Z_n = k_n, \dots, Z_0 = k_0\right) \\ &= \mathbb{P}\left(\sum_{m=1}^{k_n} \xi_{n+1,m} = k_{n+1}\right) = \mathbb{P}\left(\sum_{m=1}^{k_n} \xi_{n+1,m} = k_{n+1} \mid Z_n = k_n\right) \\ &= \mathbb{P}(Z_{n+1} = k_{n+1} \mid Z_n = k_n); \end{aligned}$$

es decir que el proceso de Galton–Watson es una cadena de Markov en tiempo discreto. Este mismo argumento revela que la matriz de probabilidades de transición $P = \{P_{i,j} : i, j \in \mathbb{N}_0\}$ tiene entradas $P_{i,j} = f^{*i}(j)$, donde f^{*i} es la i -ésima convolución de f consigo misma, siendo f la función de masa de probabilidad de $\xi_{1,1}$. Si definimos $\rho(s) = \mathbb{E}[s^{\xi_{1,1}}]$ y $\rho_n(s) = \mathbb{E}[s^{Z_n}]$ para $s \in [0, 1]$ y $n \in \mathbb{N}_0$, la definición recursiva nos permite deducir

$$\rho_{n+1}(s) = \mathbb{E}[\mathbb{E}[s^{Z_{n+1}} \mid Z_n]] = \mathbb{E}[\rho(s)^{Z_n}] = \rho_n(\rho(s)).$$

Por otra parte, una vez que conocemos el valor de Z_1 , podemos pensar en que cada una de las Z_1 partículas generan un nuevo proceso de Galton–Watson, los cuales son independientes, y entonces

$$\rho_{n+1}(s) = \mathbb{E}[\mathbb{E}[s^{Z_{n+1}} \mid Z_1]] = \mathbb{E}[\rho_n(s)^{Z_1}] = \rho_1(\rho_n(s)).$$

Cuando $Z_0 = 1$, las relaciones deducidas para las funciones generadoras de probabilidad nos permiten demostrar el siguiente resultado, cuya demostración puede consultarse en Resnick (1992, pág. 21).

Teorema 1.

Si $Z_0 = 1$, $\mathbb{E}[\xi_{1,1}] = \mu \leq 1$ y $\mathbb{P}(\xi_{1,1} = 1) < 1$, entonces el proceso se extingue con probabilidad uno. Si en cambio $\mu > 1$, la probabilidad de extinción es menor que uno y es la mínima solución no negativa a la ecuación $s = \rho(s)$.

La cantidad $\mathbb{E}[\xi_{1,1}] = \mu$, la cual llamaremos *media de reproducción*, es importante en el Teorema anterior. De manera intuitiva, que esta cantidad sea importante se debe a que controla la dinámica media de la población. En efecto, si $\mu < 1$, en promedio la población decrece; si $\mu = 1$

en promedio la población se mantiene constante; y si $\mu > 1$ la población crecerá en promedio. Esto expresado matemáticamente es

$$\mathbb{E}[Z_n] = \mu^n \mathbb{E}[Z_0],$$

expresión que puede derivarse mediante un argumento inductivo. Debido a esta dependencia de la media de reproducción, decimos que un proceso de Galton–Watson es *subcrítico* si $\mu < 1$, *crítico* si $\mu = 1$ y *supercrítico* si $\mu > 1$. Algunas propiedades asintóticas que dependen de este parámetro se mencionan en el Capítulo 3 de Kimmel y Axelrod (2015), aunque se omiten en este trabajo por no considerarse relevantes con respecto a lo que se busca desarrollar en el mismo.

Para motivar las siguientes secciones se presentan, en la figura 2, simulaciones de trayectorias del proceso Galton–Watson con $\xi_{1,1} \sim \text{Geom}(1/2)$ y variando la cantidad de partículas iniciales Z_0 . Puesto que $\mathbb{E}[\xi_{1,1}] = 1$, el Teorema 1 nos asegura que la probabilidad de extinción es uno, aunque el tiempo de extinción puede no ser observado en los tiempos considerados para las distintas simulaciones, como es el caso en las figura 2c y 2d. Hay que notar que en realidad, el proceso es en tiempo discreto, por lo que formalmente no existen las trayectorias escalonadas que se muestran. No obstante, si deseáramos conocer el tamaño de la población en un tiempo $t > 0$ que no sea entero, es razonable tomar $Z_{\lfloor t \rfloor}$, que es lo que representan las trayectorias. Observamos que conforme aumenta la cantidad de partículas iniciales en el sistema, el tiempo de extinción tiende a aumentar, lo que nos permite graficar una mayor cantidad de puntos. En este sentido, conforme más puntos agregamos, el proceso tiende a parecer una trayectoria continua errática. Este comportamiento nos lleva a intuir que, similar a como se aproxima el movimiento browniano mediante caminatas aleatorias reescaladas, si reescalamos convenientemente el proceso de Galton–Watson tanto en el tiempo como en el espacio, obtendremos un proceso estocástico en tiempo continuo cuyas trayectorias sean continuas.

Adelantándonos un poco, en la Sección 4 veremos que, en efecto, al reescalar de manera adecuada, tanto en el espacio como en el tiempo, los proceso $\{Y_t^n : t \geq 0\}$, definido mediante

$$Y_t^n = Z_{\lfloor t \rfloor}^n \quad \text{para } t \geq 0,$$

donde $\{Z_m^n : m \in \mathbb{N}_0\}$ son procesos de Galton–Watson críticos o supercríticos tales que Z_0^n sea asintóticamente equivalente con n , obtendremos una convergencia en distribución a un proceso estocástico con trayectorias continuas. Para poder hablar en forma de esta convergencia deberemos definir en qué sentido es que se da la misma, lo cual nos llevará a desarrollar un poco de la teoría del espacio D en la siguiente Sección. Asimismo deberemos de establecer ciertos resultados concernientes a los procesos de Markov conocidos como difusiones, pues la convergencia con la que trataremos será a este tipo de entes.

2. El espacio D

En la sección anterior se planteó una pregunta sobre la convergencia de ciertos procesos estocásticos. Para poder hablar de tal propiedad es necesario definir el espacio en el cual se da la convergencia y precisar qué tipo de convergencia es. Naturalmente, dado que estamos en un entorno aleatorio, la convergencia en ley resulta ser el candidato para el tipo de convergencia, dado que la convergencia en probabilidad es muy fuerte. Por otra parte, pensando de manera heurística, puesto que podemos representar gráficamente una cadena de Markov en tiempo discreto mediante trayectorias constantes por pedazos, no parece ser una idea descabellada el considerar que los procesos estocásticos que estudiemos tomen valores en el conjunto de

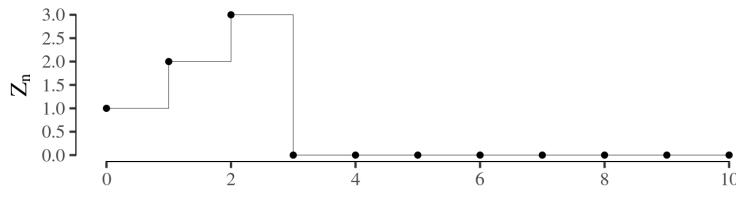
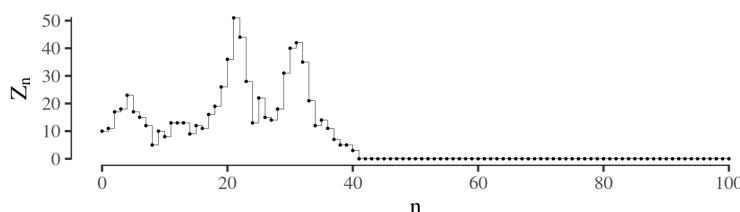
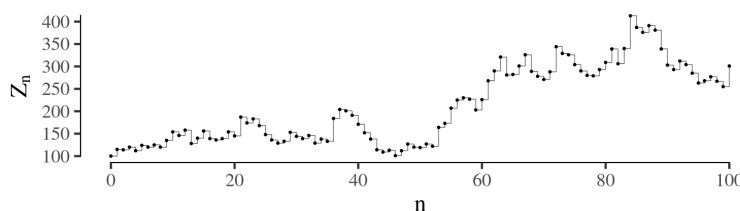
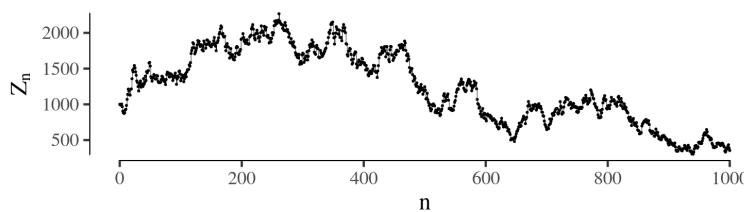
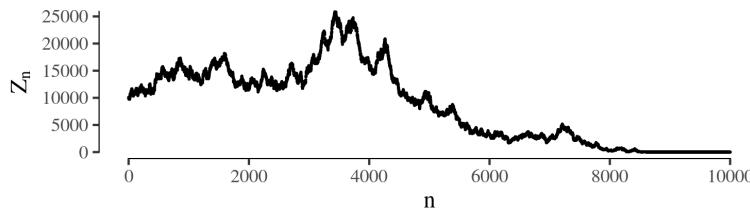
(a) $Z_0 = 1$ y $N = 10$ (b) $Z_0 = 10$ y $N = 100$ (c) $Z_0 = 100$ y $N = 100$ (d) $Z_0 = 1000$ y $N = 1000$ (e) $Z_0 = 10000$ y $N = 1000$

Figura 2: Simulaciones de procesos de Galton–Watson en los tiempos $n \in \{0, \dots, N\}$, tomando $\xi_{1,1} \sim \text{Geom}(1/2)$ y distintos valores iniciales Z_0 .

funciones $x : \mathbb{R}_+ \rightarrow \mathbb{R}$ continuas por la derecha y con límites por la derecha (*càdlàg*), el cual denotaremos por D .

Recordemos que se dice que una sucesión de elementos aleatorios $\{X_n : n \in \mathbb{N}\}$ con valores en un espacio métrico S converge en ley a X si y sólo si para toda función $f : S \rightarrow \mathbb{R}$ continua y acotada,

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

Esto nos deja en la necesidad de encontrar una manera de metrizar el conjunto D de manera que tengamos un espacio métrico completo y separable. ¿Por qué es que estas propiedades son deseables? La respuesta es dada por el Teorema de Prohorov, el cual enunciamos a continuación y cuya demostración puede consultarse en la Sección 5 de Billingsley (1999).

Teorema 2 (de Prohorov).

Sea (S, d) un espacio métrico y sea Π una colección de medidas de probabilidad sobre $\mathcal{B}(S)$. Si Π es tensa entonces es relativamente compacta. Adicionalmente, si (S, d) es completo y separable entonces el recíproco es cierto, es decir que Π sea relativamente compacta implica que es tensa.

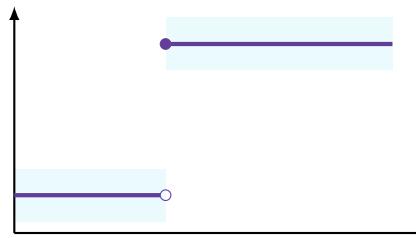


Figura 3: Vecindad de una función escalonada en $D[0, 1]$ bajo la métrica uniforme.

Del mismo teorema queda claro que deberemos de poder describir de alguna manera los conjuntos relativamente compactos en D con la métrica que definamos, por lo que de preferencia no deberá ser muy exótica. Para poder definir la métrica, primero nos restringiremos a encontrar una métrica con las propiedades deseadas sobre el conjunto de funciones $x : [0, 1] \rightarrow \mathbb{R}$ càdlàg, conjunto que de ahora en adelante se denotará por $D[0, 1]$. Una primera métrica en la que podemos pensar es aquella inducida por la norma uniforme, es decir

$$d(x, y) = \sup_{t \in [0,1]} |x(t) - y(t)| =: \|x - y\|_{u,1}, \quad x, y \in D[0, 1].$$

¿Qué problema presenta esta distancia? Notando que para cualesquiera $s, t \in [0, 1]$ distintos se cumple que $d(\mathbf{1}_{[s,1]}, \mathbf{1}_{[t,1]}) = 1$, es claro que el conjunto $D[0, 1]$ equipado con esta métrica no es separable. Más aún, este pequeño ejemplo revela que tendremos problemas con las discontinuidades de las funciones, pues con la métrica uniforme cualquier función que aproxime, por ejemplo, a $\mathbf{1}_{[1/2,1]}$ debe tener una discontinuidad más o menos del mismo tamaño en el mismo punto en el que $\mathbf{1}_{[1/2,1]}$ tiene una discontinuidad. Esta propiedad resulta ser muy restrictiva, por lo que habrá que proponer una métrica que nos permita relajar esta condición. ¿En qué sentido queremos la relajación? Lo que buscaremos es definir una métrica tal que force a que las funciones en una vecindad de otra tengan discontinuidades de magnitudes similares, pero que puedan estar ligeramente desfasadas en el tiempo.

Formalmente, sea Λ el conjunto de funciones $\lambda : [0, 1] \rightarrow [0, 1]$ estrictamente crecientes y continuas tales que $\lambda(0) = 0$ y $\lambda(1) = 1$. No es difícil demostrar que para $\lambda \in \Lambda$, λ^{-1} existe y también es un elemento de Λ . Para $x, y \in D[0, 1]$ definimos $d_\lambda(x, y)$ como el ínfimo de los valores de $\varepsilon > 0$ para los cuales existe $\lambda \in \Lambda$ tal que

$$(1) \quad \sup_{t \in [0,1]} |\lambda(t) - t| = \sup_{t \in [0,1]} |t - \lambda^{-1}(t)| < \varepsilon \quad \text{y} \quad \sup_{t \in [0,1]} |x(t) - y \circ \lambda(t)| = \sup_{t \in [0,1]} |x \circ \lambda^{-1}(t) - y(t)| < \varepsilon.$$

Intuitivamente, si λ es la deformación de tiempo considerada, entonces la primera cota nos dice que el desfase temporal permisible no puede ser grande, mientras que la segunda nos dice que bajo esta deformación en el tiempo, ambas funciones son similares. De manera más compacta definimos la *métrica* (J_1) de Skorokhod mediante

$$d_1(x, y) = \inf_{\lambda \in \Lambda} \{ \| \lambda - \text{id} \|_{u,1} \vee \| x - y \circ \lambda \|_{u,1} \} \quad x, y \in D[0, 1].$$

Para verificar que se trata de una métrica, notemos que (1) nos da directamente la simetría, mientras que es evidente que $x = y$ implica $d_1(x, y) = 0$. De manera inversa, si $d_1(x, y) = 0$ entonces existe una sucesión $\{\lambda_n\} \subset \Lambda$ tal que $\| \lambda_n - \text{id} \|_{u,1} \rightarrow 0$ y $\| x - y \circ \lambda_n \|_{u,1} \rightarrow 0$, lo cual combinado con

$$|x(t) - y(t)| \leq \|x - y \circ \lambda_n\|_{u,1} + |y \circ \lambda_n(t) - y(t)|,$$

implica que se da la igualdad $x(t) = y(t)$ en los puntos de continuidad de y . Puede probarse, ver el Lema 1 de la Sección 12 de Billingsley (1999), que los puntos de discontinuidad de $y \in D[0, 1]$ forman un conjunto a lo más numerable; de hecho para cada $\varepsilon > 0$ la cantidad de puntos $t \in [0, 1]$ tal que $|x(t) - x(t-)| \geq \varepsilon$ es finita. En consecuencia podemos deducir que x y y coinciden en un denso de $[0, 1]$ y por lo tanto son iguales. Esto es, $d_1(x, y) = 0$ implica $x = y$. Por último, notando que para cualesquiera $\lambda_1, \lambda_2 \in \Lambda$ y $x, y, z \in D[0, 1]$ se dan las relaciones

$$\begin{aligned} \| \lambda_1 \circ \lambda_2 - \text{id} \|_{u,1} &= \| \lambda_1 \circ \lambda_2 - \lambda_2^{-1} \circ \lambda_2 \|_{u,1} \leq \| \lambda_1 - \lambda_2^{-1} \|_{u,1} \\ &\leq \| \lambda_1 - \text{id} \|_{u,1} + \| \lambda_2^{-1} - \text{id} \|_{u,1} = \| \lambda_1 - \text{id} \|_{u,1} + \| \lambda_2 - \text{id} \|_{u,1} \end{aligned}$$

y

$$\|x - z \circ \lambda_1 \circ \lambda_2\|_{u,1} \leq \|x - y \circ \lambda_2\|_{u,1} + \|y \circ \lambda_2 - z \circ \lambda_1 \circ \lambda_2\|_{u,1} \leq \|x - y \circ \lambda_2\|_{u,1} + \|y - z \circ \lambda_1\|_{u,1},$$

es sencillo derivar la desigualdad del triángulo $d_1(x, z) \leq d_1(x, y) + d_1(y, z)$, con lo cual hemos asegurado que d_1 es una métrica. La topología inducida sobre $D[0, 1]$ por esta métrica es conocida como la *topología* (J_1) de Skorokhod y como se verá más adelante, el espacio resulta ser separable. En la figura 4 se muestra un ejemplo de una vecindad bajo esta métrica.

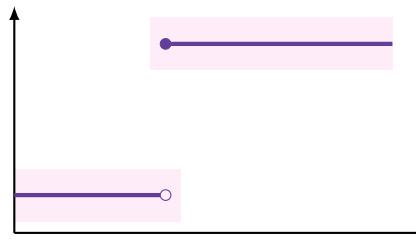


Figura 4: Vecindad de una función escalonada en $D[0, 1]$ bajo la métrica de Skorokhod.

Algunas de las propiedades de esta métrica se muestran a continuación. Supongamos que $\{x_n\} \subset D[0, 1]$ satisface $d_1(x_n, x) \rightarrow 0$, con $x \in D[0, 1]$. Es claro que entonces existe una sucesión $\{\lambda_n\} \subset \Lambda$ tal que $\| \lambda_n - \text{id} \|_{u,1} \rightarrow 0$ y $\| x \circ \lambda_n - x_n \| \rightarrow 0$. De este modo, puesto que para cualquier $t \in [0, 1]$ se satisface la desigualdad

$$|x(t) - x_n(t)| \leq \|x \circ \lambda_n - x_n\|_{u,1} + |x(t) - x \circ \lambda_n(t)|,$$

se sigue que $x_n(t) \rightarrow x(t)$ para cada punto de continuidad de x siempre que $d_1(x_m, x) \rightarrow 0$. Si adicionalmente x fuese continua, lo que implica que es uniformemente continua, de la

desigualdad

$$\|x - x_n\|_{u,1} \leq \|x_n - x_n \circ \lambda_n\|_{u,1} + \sup \left\{ \sup_{r,s \in [t,t+\|\lambda_n - \text{id}\|_{u,1}]} |x(s) - x(r)| : t \in [0,1 - \|\lambda_n - \text{id}\|_{u,1}] \right\}$$

obtenemos que $d_1(x_n, x) \rightarrow 0$ implica $d(x_n, x) \rightarrow 0$ para la métrica uniforme. De esta propiedad se desprende que la topología de Skorokhod relativa en $C[0,1]$ coincide con la topología de convergencia uniforme. Hasta el momento la métrica d_1 parece tener todas las propiedades que deseamos. No obstante, el siguiente ejemplo revela que el espacio $D[0,1]$ equipado la métrica d_1 no es completo. Para cada $n \in \mathbb{N}$ tomemos $x_n = \mathbf{1}_{[0,2^{-n}]}$ y

$$(2) \quad \lambda_n(t) = \frac{t}{2} \mathbf{1}_{[0,2^{-n}]}(t) + \left(\frac{1 - 2^{-n-1}}{1 - 2^{-n}} (t-1) + 1 \right) \mathbf{1}_{[2^{-n},1]}(t).$$

Entonces $\|x_{n+1} \circ \lambda_n - x_n\|_{u,1} = 0$ y $\|\lambda_n - \text{id}\|_{u,1} = 2^{-n-1}$ y en consecuencia $d_1(x_n, x_{n+1}) \leq 2^{-n-1}$. Si consideramos λ_n de tal forma que $\lambda_n(2^{-n-1}) \neq 2^{-n}$ entonces $\|x_{n+1} \circ \lambda_n - x_n\|_{u,1} = 1$. De ambos resultados es fácil ver que $d_1(x_n, x_{n+1}) = 2^{-n-1}$, de donde se sigue que $\{x_n\}$ es una sucesión de Cauchy con respecto a d_1 . Dado que $x_n(t) \rightarrow 0$ para cada $t > 0$, el límite debería ser $x \equiv 0$, mas $d_1(x_n, x) = 1$ para toda $n \in \mathbb{N}$, lo cual muestra que $\{x_n\}$ no es d_1 -convergente.

Dado que $D[0,1]$ con la métrica d_1 no es completo, deberemos seguir insistiendo en encontrar una métrica sobre este espacio de funciones que lo haga completo. Por otra parte, vimos que la topología generada por d_1 tiene buenas propiedades. La pregunta natural es la siguiente: ¿existe alguna métrica que genere la misma topología que d_1 y bajo la cual haya compleitez? La respuesta a esta pregunta es afirmativa. Para $\lambda : [0,1] \rightarrow [0,1]$ con $\lambda(0) = 0$ y $\lambda(1) = 1$ definamos

$$\gamma_1(\lambda) = \sup_{0 \leq s < t \leq 1} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|.$$

Observemos que si $\gamma_1(\lambda) < \infty$ entonces λ debe ser continua y estrictamente creciente, i.e. $\lambda \in \Lambda$, aunque no toda función de Λ satisface esta condición. Así, en la siguiente definición habrá que considerar solamente las funciones $\lambda \in \Lambda$ que satisfagan $\gamma_1(\lambda) < \infty$.

Definición 1.

Para $x, y \in D[0,1]$ sea $d'_1(x, y)$ el ínfimo de los valores de $\varepsilon > 0$ para los cuales existe $\lambda \in \Lambda$ tal que

$$\gamma_1(\lambda) = \gamma_1(\lambda^{-1}) < \varepsilon \quad y \quad \|x - y \circ \lambda\|_{u,1} = \|x \circ \lambda^{-1} - y\|_{u,1} < \varepsilon,$$

o de forma equivalente

$$d'_1(x, y) = \inf_{\lambda \in \Lambda} \{ \gamma_1(\lambda) \vee \|x - y \circ \lambda\|_{u,1} \}.$$

La prueba de que d'_1 es una métrica es esencialmente la misma que la realizada para d_1 por lo que la omitimos casi en su entereza. Lo único que probaremos es que para $\lambda, \mu \in \Lambda$, $\gamma_1(\lambda \circ \mu) \leq \gamma_1(\lambda) + \gamma_1(\mu)$. Efectivamente, tomando $s < t$ en $[0,1]$, tenemos que

$$\left| \log \frac{\lambda \circ \mu(t) - \lambda \circ \mu(s)}{t - s} \right| \leq \left| \log \frac{\lambda \circ \mu(t) - \lambda \circ \mu(s)}{\mu(t) - \mu(s)} \right| + \left| \log \frac{\mu(t) - \mu(s)}{t - s} \right| \leq \gamma_1(\lambda) + \gamma_1(\mu).$$

Al tomar el supremo sobre el lado izquierdo se obtiene la desigualdad deseada. Luego, sabiendo que d'_1 es una métrica, puede probarse que es equivalente a d_1 , en el sentido que $d_1(x_n, x) \rightarrow 0$ si y sólo si $d'_1(x_n, x) \rightarrow 0$. La demostración de este hecho la referimos al Teorema 12.1 de Billingsley (1999). Por lo tanto, d_1 y d'_1 generan la misma topología sobre $D[0,1]$ mas este espacio equipado con d'_1 sí sera completo.

Teorema 3.

El conjunto $D[0, 1]$ es separable bajo d_1 o d'_1 y completo bajo d'_1 .

Demostración. Para la separabilidad podemos considerar únicamente el caso de d_1 . Sea B_k el conjunto de funciones constantes por pedazos sobre los intervalos $((i-1)/k, i/k)$, con $i = 1, \dots, k$, cuyos valores son racionales y que además tomen valores racionales en $t = 1$. Es claro que $B = \bigcup_{k \geq 1} B_k$ es contable. Definamos las funciones

$$(3) \quad w_x(T) = \sup_{s, t \in T} |x(s) - x(t)|, \quad x \in D[0, 1], \quad T \subset [0, 1],$$

y

$$(4) \quad w'_x(\delta) = \inf_{\pi} \max_{1 \leq i \leq k} w_x([t_{i-1}, t_i]),$$

donde el ínfimo se extiende sobre todas las particiones $\pi = \{t_0 < t_1 < \dots < t_k\}$ de $[0, 1]$ tales que $\min\{t_i - t_{i-1}\} > \delta$. La ecuación (12.7) de Billingsley (1999) nos permite, dados $x \in D[0, 1]$ y $\varepsilon > 0$, encontrar $k \in \mathbb{N}$ tal que $1/k < \varepsilon/2$ y $w'_x(k) < \varepsilon/2$. Tomando $y \in D[0, 1]$ por medio de $y(t) = x((i-1)/k)$ para $t \in ((i-1)/k, i/k)$ y $y(1) = x(1)$ no es difícil ver que existe $z \in B$ tal que $d_1(y, z) < \varepsilon/2$. Por otra parte, por el Lema 3 de la Sección 12 de Billingsley (1999), $d_1(x, y) < \varepsilon/2$. Esto implica $d_1(x, z) < \varepsilon$ y en consecuencia, $D[0, 1]$ resulta separable.

Sigamos con la completitud del espacio bajo d'_1 . Tomemos una sucesión $\{x_n\} \subset D[0, 1]$ que sea de Cauchy con respecto a d'_1 . Es claro que entonces existe una subsucesión $\{y_n\} \subset \{x_n\}$ tal que $d'_1(y_n, y_{n+1}) < 2^{-n}$ para cada $n \in \mathbb{N}$. Para esta subsucesión, dada $n \in \mathbb{N}$, existe $\lambda_n \in \Lambda$ tal que $\gamma_1(\lambda_n) < 2^{-n}$ y $\|y_n - y_{n+1} \circ \lambda_n\|_{u,1} < 2^{-n}$. De las desigualdades $|u - 1| \leq \exp\{\log u\} - 1$ y $e^u - 1 \leq 2u$, válidas para $0 < u \leq 1/2$, obtenemos

$$\|\lambda_{n+m+1} \circ \lambda_{n+m} \circ \dots \circ \lambda_n - \lambda_{n+m} \circ \dots \circ \lambda_n\|_{u,1} = \|\lambda_{n+m+1} - \text{id}\|_{u,1} \leq e^{\gamma_1(\lambda_{n+m+1}) - 1} \leq 2\gamma_1(\lambda_{n+m+1}) < 2^{-n-m}.$$

Lo que hemos probado es que para $n \in \mathbb{N}$, la sucesión $\{\lambda_{n+m} \circ \dots \circ \lambda_n : m \in \mathbb{N}_0\}$ es uniformemente de Cauchy, por lo que existe un límite uniforme

$$\eta_n = \lim_{m \rightarrow \infty} \lambda_{n+m} \circ \dots \circ \lambda_n,$$

el cual es claramente creciente, continuo, y además satisface tanto $\eta_n(0) = 0$ como $\eta_n(1) = 1$. Para ver que $\eta_n \in \Lambda$, notemos que por definición de γ_1 y mediante un argumento inductivo tenemos las cotas, dados $0 \leq s < t \leq 1$,

$$\left| \log \frac{\lambda_{n+m} \circ \dots \circ \lambda_n(t) - \lambda_{n+m} \circ \dots \circ \lambda_n(s)}{t - s} \right| \leq \gamma_1(\lambda_{n+m} \circ \dots \circ \lambda_n) \leq \sum_{k=0}^m \gamma_1(\lambda_{n+k}) \leq 2^{-n+1}.$$

Al tomar el límite cuando $m \rightarrow \infty$ y posteriormente el supremo sobre $0 \leq s < t \leq 1$ podemos concluir $\gamma_1(\eta_n) < 2^{-n+1}$, con lo cual tenemos $\eta_n \in \Lambda$ para cada $n \in \mathbb{N}$. Finalmente, notemos que se da la relación $\eta_n = \eta_{n+1} \circ \lambda_n$ para cada $n \in \mathbb{N}$, o equivalentemente $\eta_{n+1}^{-1} = \lambda_n \circ \eta_n^{-1}$, de donde

$$\|y_n \circ \eta_n^{-1} - y_{n+1} \circ \eta_{n+1}^{-1}\|_{u,1} = \|y_n \circ \eta_n^{-1} - y_{n+1} \circ \lambda_n \circ \eta_n^{-1}\|_{u,1} = \|y_n - y_{n+1} \circ \lambda_n\|_{u,1} < 2^{-n}.$$

Por lo tanto, $\{y_n \circ \eta_n\}$ es uniformemente Cauchy y entonces converge de manera uniforme a un elemento $y \in D$. Puesto que $\gamma_1(\eta_n) \rightarrow 0$, se deduce entonces $d'_1(y_n, y) \rightarrow 0$ y en consecuencia $d'_1(x_n, y) \rightarrow 0$. \square

En vista de que la métrica d'_1 hace al espacio $D[0, 1]$ completo y separable, la idea será extenderla al espacio D , que recordemos es aquel de las funciones $x : \mathbb{R}_+ \rightarrow \mathbb{R}$ càdlàg. Para

la extensión deberemos definir algunos conceptos auxiliares. Tomando $t > 0$, consideremos $D_t \equiv D[0, t]$ como el espacio de funciones $x : [0, t] \rightarrow \mathbb{R}$ càdlàg y definamos

$$\begin{aligned}\|x\|_{u,t} &= \sup_{s \in [0,t]} |x(s)|, \\ \gamma_t(\lambda) &= \sup_{0 \leq s < r \leq t} \left| \log \frac{\lambda(r) - \lambda(s)}{r - s} \right|, \\ d'_t(x, y) &= \inf_{\lambda \in \Lambda_t} \{ \gamma_t(\lambda) \vee \|x - y \circ \gamma\|_{u,t} \}, \\ d_t(x, y) &= \inf_{\lambda \in \Lambda_t} \{ \|\lambda - \text{id}\|_{u,t} \vee \|x - y \circ \gamma\|_{u,t} \},\end{aligned}$$

donde únicamente hay que explicar que Λ_t es la colección de funciones $\lambda : [0, t] \rightarrow [0, t]$ continuas, estrictamente crecientes y tales que $\lambda(0) = 0$ y $\lambda(t) = t$. De manera similar se define Λ_∞ . Todo lo concerniente al caso $t = 1$ se puede extender al caso $t > 0$ general, modificando de manera adecuada los resultados.

Con las definiciones anteriores, no es difícil notar que la restricción de un elemento $x \in D$ a $[0, t]$ es un elemento de D_t . Por lo tanto, simulando la convergencia uniforme en compactos de C , pensariamos natural el definir la topología de Skorokhod de tal manera que $x_n \rightarrow x$ si y sólo si $d'_t(x_n, x) \rightarrow 0$ para cualquier valor de $t > 0$, donde se toman las restricciones de las funciones. No obstante, nuevamente las discontinuidades nos generan problemas pues $\mathbb{1}_{[1+1/n, \infty)}$ intuitivamente debería converger a $\mathbb{1}_{[1, \infty)}$, mas $d'_1(\mathbb{1}_{[1+1/n, \infty)}, \mathbb{1}_{[1, \infty)}) = 1$, por lo que este primer acercamiento no es el adecuado. Sin embargo, el siguiente lema nos permitirá deslindarnos parcialmente de estos problemas; para su demostración se puede consultar el Lema 1 de la Sección 16 de Billingsley (1999).

Lema 4.

Sean $\{x_n\} \subset D_t$ y $x \in D_t$ dados. Si $d'_t(x_n, x) \rightarrow 0$, $s < t$ y x es continua en s , entonces $d'_s(x_n, x) \rightarrow 0$.

Usando este lema es que extenderemos la métrica d' a D , considerando las métricas d'_m para valores enteros de m . Por el ejemplo visto, deberemos de tener cuidado en evitar discontinuidades en estos valores. Para esto definimos las funciones

$$g_m(t) = \mathbb{1}_{[0, m-1]}(t) + (m-t)\mathbb{1}_{(m-1, m]}(t), \quad m \in \mathbb{N},$$

y para $x \in D$ definimos $x^m = g_m x$. Es claro que $x^m \in D$ para cada m y además es una función continua en este tiempo. Con esta notación definimos, para $x, y \in D$,

$$d'(x, y) = \sum_{m=1}^{\infty} \frac{1}{2^m} \mathbb{1} \wedge d'_m(x^m, y^m).$$

Por lo realizado antes, vemos que si $d'(x, y) = 0$ entonces $x^m = y^m$ para cada $m \in \mathbb{N}$, lo que implica que para cada $m \in \mathbb{N}$ se satisface que $x(t) = y(t)$ siempre que $t \in [0, m-1]$. Se sigue pues que $x = y$. Las demás propiedades de métrica resultan claras para d' , por lo que tenemos que d' es una métrica sobre D . De manera análoga a D_1 , la topología inducida por esta métrica es conocida como topología de Skorokhod. Algunas métricas equivalentes se pueden consultar en Ethier y Kurtz (2005), Lindvall (1973) y Pollard (1984). Antes de continuar vale la pena mencionar que si sustituimos d'_m por d_m obtenemos otra métrica, equivalente a d' , bajo la cual D no será completo. Antes de probar que D con d' es completo y separable, en el siguiente Teorema se enuncian un par de condiciones equivalentes a convergencia con respecto a la métrica d' .

Teorema 5.

Sean $\{x_n\} \subset D$ y $x \in D$ arbitrarios. Los siguientes enunciados son equivalentes:

- (i) $d'(x_n, x) \rightarrow 0$;

(ii) existe $\{\lambda_n\} \subset \Lambda_\infty$ tal que

$$\sup_{t \in \mathbb{R}_+} |\lambda_n(t) - t| \rightarrow 0 \quad y \quad \|x_n \circ \lambda_n - x\|_{u,m} \rightarrow 0 \quad \text{para cada } m \in \mathbb{N};$$

(iii) $d'_t(x_n, x) \rightarrow 0$ para cada punto de continuidad $t > 0$ de x .

La prueba de este Teorema puede consultarse, de nueva cuenta, en Billingsley (1999). Notemos que el tercer numeral del mismo es similar a la convergencia uniforme en compactos de C mas omitiendo los puntos de discontinuidad de las funciones. En tanto, el segundo numeral nos dice que para la convergencia la distorsión que podemos tener el tiempo, debe ser uniformemente pequeña y deber ser tal que se dé convergencia uniforme sobre los conjuntos $[0, m]$, con $m \in \mathbb{N}$. De nueva cuenta, esto implicará que esta topología restringida a C coincide con aquella de convergencia en compactos. Dicho esto, veamos que D con la métrica d' es completo y separable.

Teorema 6.

D con la métrica d' es completo y separable.

Demostración. Para $x \in D$ y $m \in \mathbb{N}$ definamos $\psi_m(x)$ como la restricción de x^m a $[0, m]$. Dado que $d'_m(\psi_m(x), \psi_m(y)) = d'_m(x^m, y^m)$ es claro que $\psi_m : D \rightarrow D_m$ define un mapeo continuo. Consecuentemente el mapeo $\psi : D \rightarrow \prod_{m=1}^\infty D_m$ definido por $x \mapsto (\psi_1(x), \psi_2(x), \dots)$ es continuo cuando el espacio producto se equipa con la topología producto, la cual es inducida por la métrica

$$\rho(\alpha, \beta) = \sum_{m=1}^\infty \frac{1}{2^m} 1 \wedge d'_m(\alpha_m, \beta_m).$$

De hecho, es claro que $d'(x, y) = \rho(\psi(x), \psi(y))$, por lo que ψ es una isometría entre D y $\psi(D)$. Puesto que cada uno de los espacios D_m es separable y completo cuando se consideran con d'_m , el espacio $\prod_{m=1}^\infty D_m$ equipado con ρ es también separable y completo, por lo que si probamos que $\psi(D)$ es cerrado, entonces será completo y separable, lo que implicará el resultado deseado.

De esta manera, probaremos que si $\{x_n\} \subset D$ y $\alpha \in \prod_{m=1}^\infty D_m$ son tales que $\rho(\psi(x_n), \alpha) \rightarrow 0$ entonces existe $x \in D$ tal que $\alpha = \psi(x)$. Para cada m , el conjunto de puntos de discontinuidad de α_m es a lo más numerable, por lo que no es difícil comprobar la existencia de un conjunto $T \subset \mathbb{R}_+$ denso tal que α_m sea continuo en $T \cap [0, m]$. De la condición $\rho(\psi(x_n), \alpha) \rightarrow 0$, se sigue que $d_m(x_n^m, \alpha_m) \rightarrow 0$ para cada m . En virtud del Lema 4 esto último implica que para cada $m \in \mathbb{N}$,

$$\alpha_m(t) = \lim_{n \rightarrow \infty} x_n^m(t), \quad t \in T \cap [0, m],$$

y en consecuencia $\alpha_m(t) = \lim_{n \rightarrow \infty} x_n(t)$, $t \in T \cap [0, m-1]$. Por lo tanto, para cada $t \in T$ existe el límite $x(t) = \lim_{n \rightarrow \infty} x_n(t)$, el cual satisface la relación $x(t) = \alpha_m(t)$ para $t \in T \cap [0, m-1]$. Por ende, podemos extender x a una función càdlàg en cada intervalo $[0, m-1]$ y posteriormente en \mathbb{R}_+ . Finalmente, por la continuidad por la derecha de x , de α_m y de g_m , se sigue que $x^m = \alpha_m$ para cada m , o bien $\psi(x) = \alpha$. \square

En vista de los resultados probados, tenemos una métrica para el espacio D que lo hace completo y separable, condiciones que son importantes para el Teorema 2. Para poder hacer uso del mismo, deberemos de caracterizar los conjuntos relativamente compactos en D . Puesto que la métrica d' se construyó con base en la métrica d'_1 , es razonable caracterizar primero a los conjuntos relativamente compactos en D_1 para luego obtener la caracterización deseada en D . Para D_1 , el siguiente resultado tiene rasgos reminiscientes del Teorema de Arzelà-Ascoli, lo cual no debería de sorprendernos dado que la topología de Skorokhod es una extensión de la topología uniforme en $C[0, 1]$.

Teorema 7.

Un conjunto $A \subset D_1$ es relativamente compacto bajo la topología de Skorokhod si y sólo si

$$\sup_{x \in A} \|x\|_{u,1} < \infty \quad y \quad \lim_{\delta \rightarrow 0} \sup_{x \in A} w'_x(\delta) = 0,$$

donde $w'_x(\delta)$ se define como en (4).

Este resultado junto con la isometría entre D y un subconjunto cerrado de $\prod_{m=1}^{\infty} D_m$, probada la demostración del Teorema 6, nos permitirán dar una caracterización de los conjuntos relativamente compactos en D por medio de sistemas de funciones en D_m para distintos valores de m . Más aún, podremos dar una caracterización similar a la del Teorema anterior.

Teorema 8.

Para un subconjunto $A \subset D$ los siguientes enunciados son equivalentes:

- (i) A es relativamente compacto;
- (ii) para cada $m \in \mathbb{N}$, $\psi_m(A)$ es relativamente compacto en D_m ;
- (iii) para cada $m \in \mathbb{N}$,

$$\sup_{x \in A} \|x\|_{u,m} < \infty \quad y \quad \lim_{\delta \rightarrow \infty} \sup_{x \in A} w'_m(x, \delta) = 0,$$

donde

$$w'_m(x, \delta) = \inf_{\pi} \max_{1 \leq i \leq k} w_x([t_{i-1}, t_i]),$$

con el infimo siendo sobre todas las particiones $\pi = \{t_0 < t_1 < \dots < t_k\}$ de $[0, m]$ tales que $\min\{t_i - t_{i-1} : 1 \leq i < k\} > \delta$ y $w_x(T)$ se entiende como la generalización de (3)

Demostración. Supongamos primero que A es relativamente compacto. Entonces \bar{A} es compacto y en consecuencia $\psi_m(\bar{A})$ es compacto por ser ψ_m continua. Luego, $\psi_m(A) \subset \psi_m(\bar{A})$ es relativamente compacto. De manera inversa, si $\psi_m(\bar{A})$ es relativamente compacto para cada m , entonces $B = \prod_{m=1}^{\infty} \overline{\psi_m(A)}$ es compacto en $\prod_{m=1}^{\infty} D_m$ debido al Teorema de Tychonoff. Luego, por la demostración del Teorema 6, $B \cap \psi(D)$ resulta ser compacto. Ahora basta notar que $\psi(A) \subset B \cap \psi(D)$ es totalmente acotado, por lo que A también lo es, mostrando que es relativamente compacto. De esta manera tenemos la equivalencia entre (i) y (ii).

Resta probar que (iii) es equivalente con cualquiera de los otros enunciados. Notemos que por el Teorema 7, podemos reescribir el enunciado (ii) como sigue: para cada $m \in \mathbb{N}$,

$$\sup_{x \in A} \|\psi_m(x)\|_{u,m} < \infty \quad y \quad \lim_{\delta \rightarrow 0} \sup_{x \in A} \bar{w}_m(x, \delta) = 0,$$

donde $\bar{w}_x(x, \delta)$ es la extensión natural de (4) a D_m . Así pues, de las relaciones elementales $\|\psi_m(x)\|_{u,m} \leq \|x\|_{u,m} \leq \|\psi_{m+1}(x)\|_{u,m+1}$ y $w'_m(x, \delta) \leq \bar{w}_m(x, \delta)$ es evidente que (ii) implica (iii). Inversamente, bastará probar que si (iii) es cierta, entonces $\lim_{\delta \rightarrow 0} \sup_{x \in A} \bar{w}_m(x, \delta) = 0$. En este sentido suponiendo que se satisfacen las condiciones en (iii), sea $K_m = \sup_{x \in A} \|x\|_{u,m}$. Para $x \in A$ y $\delta < 1$, se tiene que $|\psi_m \circ x(t)| \leq \delta K_m$ para $t \in [m - \delta, m]$. Para $\varepsilon > 0$ dado, sea $\delta > 0$ tal que $\delta K_m < \varepsilon/4$ y

$$\sup_{x \in A} w'_m(x, \delta) < \frac{\varepsilon}{2}.$$

Dado $x \in A$, debe existir una partición π_x de $[0, m]$ que satisfaga las condiciones en (iii) y tal que $\max_{1 \leq i \leq k} w_x([t_{i-1}, t_i]) < \varepsilon/2$. Tomemos $j = \max\{i : t_{i-1} \leq m - \delta, t_{i-1} \in \pi_x\}$ y definamos la nueva partición de $[0, m]$, $\tilde{\pi} = \{t_0 < t_1 < \dots < t_{j-1} < m\}$. Por cómo tomamos δ y esta nueva partición, es fácil ver que $w_x([t_{j-1}, m]) < \varepsilon$ y en consecuencia $\bar{w}_m(x, \delta) < \varepsilon$. Puesto que $x \in A$ fue arbitrario, concluimos que

$$\lim_{\delta \rightarrow 0} \sup_{x \in A} \bar{w}_m(x, \delta) = 0.$$

□

Este último resultado, que caracteriza a los conjuntos relativamente compactos en D , en conjunto con el Teorema de Prohorov, nos da condiciones necesarias y suficientes para poder asegurar que una sucesión de medidas en de probabilidad sobre $\mathcal{B}(D)$ sea tensa. Omitimos la demostración por tratarse esencialmente de la misma que se hace en cuando el espacio es $C[0, 1]$ y consiste en traducir el inciso (iii) del Teorema 8 en términos estocásticos.

Teorema 9.

Una sucesión $\{\mathbb{P}_n\}$ de medidas de probabilidad sobre $\mathcal{B}(D)$ es tensa si, únicamente si,

- (i) para cada $m \in \mathbb{N}$,

$$\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}_n\{x \in D : \|x\|_{u,m} \geq a\} = 0, \quad y$$

- (ii) para cualesquiera $m \in \mathbb{N}$ y $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}_n\{x \in D : w'_m(x, \delta) \geq \varepsilon\} = 0.$$

Hasta este punto, hemos desarrollado suficientes herramientas para poder establecer la tensión de medidas de probabilidad sobre $\mathcal{B}(D)$ o, de manera equivalente, de elementos aleatorios con valores en D . No obstante, estos resultados únicamente nos permiten asegurar la tensión mas no nos dan un criterio de cuándo es que hay convergencia débil. Recordemos que si $\{X_n : n \in \mathbb{N}\}$ es una sucesión de elementos aleatorios en $C[0, 1]$ entonces para tener $X_n \Rightarrow X$, basta demostrar la tensión de $\{X_n : n \in \mathbb{N}\}$ y que las distribuciones finito-dimensionales de X_n converjan débilmente a aquellas de X . La parte final de esta sección se dedicará a dar un par de criterios de convergencia para elementos aleatorios en D , uno de los cuales es análogo al de $C[0, 1]$, así como un nuevo criterio de tensión para sucesiones de elementos aleatorios en D , conocido como el *criterio de Aldous*.

Para poder deducir condiciones bajo las cuales podemos asegurar que existe la convergencia débil $X_n \Rightarrow X$ con elementos aleatorios D -valuados, primero obtendremos una clase separante de $\mathcal{B}(D)$ que nos permitirá reducir la convergencia débil al sentido finito-dimensional en presencia de tensión. Para enunciar el siguiente resultado definimos las proyecciones naturales $\pi_{t_1, \dots, t_k} : D \rightarrow \mathbb{R}^k$ y $\pi_{t_1, \dots, t_k}^m : D_m \rightarrow \mathbb{R}^k$ por medio de

$$\pi_{t_1, \dots, t_k}(x) = (x(t_1), \dots, x(t_k)) \quad y \quad \pi_{t_1, \dots, t_k}^m(x) = (x(t_1), \dots, x(t_k)),$$

donde es claro que en el segundo caso $t_i \in [0, m]$ para cada i .

Teorema 10.

- (i) La proyección π_0 es continua y para $t > 0$, π_t es continua en $x \in D$ precisamente cuando x es continua en t .
- (ii) Para cada $t \in \mathbb{R}_+$, π_t es medible y en consecuencia π_{t_1, \dots, t_k} también lo es.
- (iii) Si $T \subset \mathbb{R}_+$ es denso entonces $\sigma(\pi_t : t \in T) = \mathcal{B}(D)$ y la clase

$$C = \left\{ \pi_{t_1, \dots, t_k}^{-1}(B) : k \geq 1, t_1, \dots, t_k \in T, B \in \mathbb{R}^k \right\}$$

es separante.

Demostración. Puesto que $\lambda(0) = 0$ para cada $\lambda \in \Lambda_\infty$, es claro que $d'(x_n, x) \rightarrow 0$ implica la convergencia $|\pi_0(x_n) - \pi_0(x)| \rightarrow 0$, dando la continuidad de π_0 . Suponiendo que x es continua en $t > 0$, un argumento similar al dado en (2) revela que $|\pi_t(x_n) - \pi_t(x)| \rightarrow 0$ siempre que $d'(x_n, x) \rightarrow 0$. Si en cambio x fuese discontinua en t , considerando

$$\lambda_n(s) = \frac{nt-1}{nt} s \mathbb{1}_{[0,t]}(s) + \left(s - \frac{1}{n} \right) \mathbb{1}_{(t-1/n, \infty)}(s)$$

y $x_n = x \circ \lambda_n$, es claro por el numeral (ii) del Teorema 5 que $d'(x_n, x) \rightarrow 0$ pero $\pi_t(x_n) \rightarrow \pi_t(x)$. Esto prueba el numeral (i).

Por lo ya demostrado, basta obtener la medibilidad de π_t cuando $t > 0$. Es claro que se da la relación $\pi_t = \pi_t^m \circ \psi_m$ cuando $t \leq m - 1$, lo cual aunado a la continuidad de ψ_m nos dará la medibilidad de π_t si probamos que π_t^m es medible. Para esto, definamos $h_\varepsilon^m(x) = \varepsilon^{-1} \int_t^{t+\varepsilon} x(s) ds$ y observemos que si $d'_m(x_k, x) \rightarrow 0$ entonces $h_\varepsilon^m(x_k) \rightarrow h_\varepsilon^m(x)$ en virtud del Teorema de Convergencia Dominada, lo que implica la continuidad de h_ε^m y de manera subsecuente su medibilidad. Resta notar que, debido a la continuidad por la derecha de las funciones en consideración, $h_\varepsilon^m \rightarrow \pi_t^m$ cuando $\varepsilon \rightarrow 0$. Así, π_t es medible por ser límite puntual de funciones medibles.

Para la tercera parte primeramente observamos que C es un π -sistema. Luego, de los numerales anteriores se sigue que únicamente deberemos probar $\mathcal{B}(D) \subset \sigma(\pi_t : t \in T)$ pues entonces tendremos la igualdad y en adición, se seguirá $\sigma(C) = \mathcal{B}(D)$ probando que C es una clase separante. Por el Teorema 12.5 de Billingsley (1999) sabemos que si $T_m = (T \cap [0, m]) \cup \{m\}$ entonces

$$(5) \quad \mathcal{B}(D_m) = \sigma(\pi_t^m : t \in T_m).$$

Observemos que para $t \leq m$ y $B \in \mathcal{B}(\mathbb{R})$,

$$\psi_m^{-1}((\pi_t^m)^{-1}(B)) = \{x \in D : \pi_t^m \circ \psi_m(x) \in B\} = \{x \in D : g_m(t)\pi_t(x) \in B\}.$$

Cuando $t = m$, $g_m(t) = 0$ por definición y en consecuencia

$$\psi_m^{-1}((\pi_t^m)^{-1}(B)) = \begin{cases} \emptyset & \text{si } 0 \notin B, \\ D & \text{si } 0 \in B, \end{cases}$$

por lo que en cualquier caso es un conjunto de $\sigma(\pi_t : t \in T)$. Si en cambio $t < m$, entonces tenemos la igualdad

$$\psi_m^{-1}((\pi_t^m)^{-1}(B)) = \{x \in D : \pi_t x \in g_m(t)^{-1}B\},$$

el cual es un elemento de $\sigma(\pi_t : t \in T)$ en vista del numeral (ii). Por ende, si $A = (\pi_t^m)^{-1}(B)$ con $t \in T_m$, entonces $\psi_m^{-1}(A) \in \sigma(\pi_t : t \in T)$. Por (5), esto último implica que ψ_m es medible con respecto a $\sigma(\pi_t : t \in T)$ y $\mathcal{B}(D_m)$. Por consiguiente, usando la continuidad de $\alpha_m \mapsto d'_m(\alpha_m, \beta_m)$ y la medibilidad antes probada, el mapeo $x \mapsto d'_m(\psi_m(x), \beta_m)$ es medible con respecto a las σ -álgebras $\sigma(\pi_t : t \in T)$ y $\mathcal{B}(\mathbb{R})$. Consecuentemente

$$x \mapsto \sum_{m=1}^{\infty} \frac{1}{2^m} 1 \wedge d'_m(\psi_m(x), \psi_m(y))$$

es medible con respecto a $\sigma(\pi_t : t \in T)$ y $\mathcal{B}(\mathbb{R})$ para $y \in D$ fija. Por lo tanto, $\sigma(\pi_t : t \in T)$ contiene a las bolas abiertas en D , de modo que, dado que es un espacio separable, contiene a $\mathcal{B}(D)$. \square

Hemos probado que cierta clase de cilindros forma una clase separante de $\mathcal{B}(D)$, ¿pero cómo es que usaremos esto para verificar la convergencia débil de una sucesión de elementos? Recordamos que se mencionó que para $x \in D_1$ y $\varepsilon > 0$, la cantidad de tiempos $t \in [0, 1]$ tales que $|x(t) - x(t-)| \geq \varepsilon$ es finita, situación que también ocurre con los elementos en D_m . Por ende, los elementos de $x \in D$ son continuos salvo en una colección contable de tiempos. Veremos que, en cierto sentido, esto también es cierto para las medidas de probabilidad sobre $\mathcal{B}(D)$. Efectivamente, dada \mathbb{P} , definamos $J_t(\varepsilon) = \{x \in D : |x(t) - x(t-)| \geq \varepsilon\}$. Dados $\varepsilon, \delta > 0$ y $m \in \mathbb{N}$, el conjunto de tiempos $t \in [0, m]$ tales que $\mathbb{P}(J_t(\varepsilon)) \geq \delta$ es a lo más finito, pues en caso contrario, existiría una sucesión $\{t_n\} \subset [0, m]$ tal que

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} J_{t_n}(\varepsilon)\right) \geq \delta,$$

implicando que el límite superior es no vacío, por lo que existe $x \in D_m$ tal que la colección de tiempos $|x(t) - x(t-)| \geq \varepsilon$ es infinita, lo cual sabemos no sucede. De esta manera es sencillo notar que el conjunto $\{t \in [0, m] : \mathbb{P}(J_t(\varepsilon)) > 0\}$ es contable, de modo que $\{t \in \mathbb{R}_+ : \mathbb{P}(J_t(\varepsilon)) > 0\}$ también lo es. Al considerar el límite cuando $\varepsilon \rightarrow 0$ obtenemos entonces que

$$\{t \in \mathbb{R}_+ : \mathbb{P}\{x \in D : |x(t) - x(t-)| > 0\} > 0\}$$

es un conjunto a lo más numerable. Así pues, se tiene que el conjunto de puntos $t \in \mathbb{R}_+$ para los cuales π_t es casi seguramente continua con respecto a \mathbb{P} , el cual denotaremos por $T_{\mathbb{P}}$, forma un conjunto denso en \mathbb{R}_+ . De esta manera, combinando este resultado y el Teorema 10, obtenemos el siguiente criterio de convergencia.

Teorema 11.

Sean $\{X_n\}$ y X elementos aleatorios D -valuados y sea \mathbb{P}_X la ley de X . Si $\{X_n\}$ es tensa y además

$$(X_n(t_1), \dots, X_n(t_k)) \Rightarrow (X(t_1), \dots, X(t_k))$$

para cualesquiera valores $t_1, \dots, t_k \in T_{\mathbb{P}_X}$, con $k \in \mathbb{N}$ arbitrario, entonces $X_n \Rightarrow X$.

Un segundo criterio de convergencia se enuncia a continuación, el cual corresponde al Teorema 16.7 de Billingsley (1999), aunque esta versión es para elementos aleatorios y no medidas.

Teorema 12.

Sean $\{X_n\}$ y X elementos aleatorios D -valuados y sea \mathbb{P}_X la ley de X . Una condición necesaria y suficiente para $X_n \Rightarrow X$ es que para cada punto $t \in T_{\mathbb{P}_X}$ se cumpla

$$\{X_n(s) : 0 \leq s \leq t\} \Rightarrow \{X(s) : 0 \leq s \leq t\} \quad \text{en } D_t.$$

Para cerrar la sección, presentamos el criterio de Aldous, que nos da condiciones probabilistas bajo las cuales podemos afirmar que una sucesión de elementos aleatorios con valores en D es tensa. Antes de dar el criterio, recordamos que un tiempo de paro con respecto a una filtración $\mathbb{F} = \{\mathcal{F}_t : t \in \mathbb{R}_+\}$ es una variable aleatoria τ con valores en \mathbb{R}_+ tal que $\{\tau \leq t\} \in \mathcal{F}_t$ para cada $t \in \mathbb{R}_+$. Además diremos que τ es un X_n -tiempo de paro si es un tiempo de paro con respecto a la filtración $\mathbb{F}_n = \{\mathcal{F}_t^n : t \in \mathbb{R}_+\}$, donde $\mathcal{F}_t^n = \sigma(X_n(s) : s \leq t)$. Dicho esto enunciamos el criterio de Aldous, cuya demostración puede consultarse en la Sección 16 de Billingsley (1999). Una versión de este resultado que involucra esperanzas se muestra en el Capítulo 2 de Kurtz (1981).

Teorema 13.

Sea $\{X_n\}$ una sucesión de elementos aleatorios D -valuados. Esta sucesión será tensa si se satisfacen las condiciones:

- (i) para cada $m \in \mathbb{N}$,

$$\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\|X_n\|_{u,m} \geq a) = 0,$$

- (ii) para cualesquiera valores $\varepsilon, \eta > 0$ y $m \in \mathbb{N}$ existen $\delta_0 > 0$ y $n_0 \in \mathbb{N}$, de tal manera que si $\delta \leq \delta_0$ y $n \geq n_0$, y además τ es un X_n -tiempo de paro discreto con $\tau \leq m$, entonces

$$\mathbb{P}(|X_n(\tau + \delta) - X_n(\tau)| \geq \varepsilon) \leq \eta.$$

En esta sección se desarrolló la teoría básica del espacio D , el cual resulta muy útil debido a que muchos procesos estocásticos que conocemos pueden verse como elementos aleatorios de este espacio. Asimismo se derivó un criterio de convergencia débil para elementos aleatorios con valores en D , el cual es similar al criterio que se tiene para aquellos elementos aleatorios con valores en el espacio de funciones continuas. Para ejemplos de la convergencia débil en

este espacio, uno puede consultar la Sección 14 de Billingsley (1999), o la Sección 4 de este mismo trabajo. Antes de poder dar este ejemplo de convergencia, que involucra al proceso de Galton–Watson, deberemos de dar algunas nociones asociadas a procesos de Markov en tiempo continuo, objetivo de la siguiente sección.

3. Procesos de Markov

En la primera sección se introdujeron los procesos de Galton–Watson de manera sucinta y entre sus propiedades se mencionó que forman parte de las cadenas de Markov. En la presente sección se introduce una generalización de este último concepto a procesos estocásticos que ya no necesariamente están indexados por \mathbb{N} o no toman valores en un espacio discreto. Intuitivamente, un proceso de Markov es un proceso estocástico con la propiedad de que el estado futuro del sistema sólo depende del estado actual y no de toda la historia del proceso. La complejidad de estudiar estos procesos radica en dos aspectos: el conjunto de índices y el espacio de estados. Usualmente, los procesos de Markov modelan la configuración de un sistema en un instante de tiempo determinado, de modo que los conjuntos de índices son entendidos como tiempo y los más comunes son \mathbb{N} o $T \subseteq \mathbb{R}_+$. Con esto en mente, tenemos las siguientes combinaciones posibles de tiempo-espacio:

- (i) Tiempo discreto - espacio discreto. En este contexto, los procesos de Markov son referidos como *Cadenas de Markov a tiempo discreto*, y son objetos que están determinados por una distribución inicial y una matriz de transición P .
- (ii) Tiempo continuo - espacio discreto. Estos modelos, al igual que los anteriores, son llamados cadenas de Markov, pero *a tiempo continuo*. Son procesos determinados por una distribución inicial y la Q-matriz. Dicha matriz tiene como entradas las tasas o intensidades con las que el proceso realiza un salto.
- (iii) Tiempo continuo - espacio abstracto. A diferencia de los dos procesos anteriores, estos objetos son más complicados de estudiar en el sentido de que las propiedades que podemos deducir están supeditadas a condiciones sobre el espacio de estados.

Como es usual, el punto de partida es un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. El conjunto de índices o del tiempo T es $[0, \infty)$. El conjunto de estados E está dotado de una σ -álgebra \mathcal{E} , de modo que la pareja (E, \mathcal{E}) es el espacio de estados. Generalmente, E posee una topología y \mathcal{E} es la σ -álgebra de Borel generada por los abiertos de dicha topología. Un supuesto sobre E es que su topología es Hausdorff, localmente compacta y tiene una base numerable. Estas condiciones son suficientemente generales para englobar una gran cantidad de procesos que son utilizados en las aplicaciones, y para obtener una teoría completa.

Varios espacios de funciones sobre E tendrán un papel fundamental en el desarrollo del trabajo. Denotamos por \mathcal{B} al conjunto de todas las funciones $f: E \rightarrow \mathbb{R}$ que son Borel medibles y acotadas. \mathcal{B} es un espacio vectorial y estará dotado de la norma del supremo.

Supongamos ahora que $X = \{X_t: t \in T\}$ es un proceso estocástico sobre $(\Omega, \mathcal{F}, \mathbb{P})$ con espacio de estados E . Entonces X_t es una variable aleatoria E -valuada, y como ya mencionamos arriba, X_t es el estado del sistema al tiempo t . Supondremos también que el espacio de probabilidad es filtrado; *i.e.*, posee una filtración $\mathbb{F} = \{\mathcal{F}_t: t \in T\}$ con la propiedad de que X es adaptado a \mathbb{F} . La mínima filtración respecto a la cual X es adaptado está dada por $\mathbb{F}^0 = \{\mathcal{F}_t^0: t \in T\}$ donde $\mathcal{F}_t^0 = \sigma(X_s: s \leq t)$. En muchas ocasiones, es necesaria una filtración mucho más grande que \mathbb{F}^0 , y usualmente asumimos que \mathbb{F} es *continua por la derecha*. Esto último significa que $\mathcal{F}_t = \mathcal{F}_{t+}$, donde $\mathcal{F}_{t+} = \bigcap \{\mathcal{F}_s: s > t\}$. Podemos construir una filtración continua por la derecha a partir de \mathbb{F}^0 tomando $\mathcal{F}_t = \bigcap \{\mathcal{F}_s^0: s > t\}$. Esta última σ -álgebra es llamada *refinamiento continuo por la derecha* de la filtración \mathbb{F}^0 . También es necesario asumir que la filtración \mathbb{F} es completa; en

otras palabras, si cumple que dado $A \in \mathcal{F}_t$ y $\mathbb{P}(A) = 0$, entonces $B \in \mathcal{F}_t$ para todo $B \subseteq A$. Dicha condición es equivalente a que \mathcal{F}_0 contiene a todos los subconjuntos de los conjuntos nulos de \mathbb{P} .

Con las condiciones discutidas arriba, estamos listos para la primera definición de esta sección.

Definición 2.

Sea $X = \{X_t : t \geq 0\}$ un proceso estocástico sobre $(\Omega, \mathcal{F}, \mathbb{P})$ con valores en E y adaptado a \mathcal{F} . X es un proceso de Markov relativo a \mathcal{F} si

$$(6) \quad \forall A \in \mathcal{E}: \mathbb{P}(X_{s+t} \in A \mid \mathcal{F}_s) = \mathbb{P}(X_{s+t} \in A \mid X_s)$$

La condición en (6) es llamada propiedad de Markov, y establece que la distribución condicional de X_{s+t} dado \mathcal{F}_s es la misma que la distribución condicional de X_{s+t} dado X_s . En otras palabras, conocer todo el pasado hasta el tiempo s es irrelevante en términos de predecir el estado futuro del sistema.

Técnicamente, las probabilidades condicionales están definidas por medio de la esperanza condicional, esto es,

$$\mathbb{P}(X_{s+t} \in A \mid \mathcal{F}_s) := \mathbb{E}[\mathbb{1}_A(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}[\mathbb{1}_A(X_{s+t}) \mid X_s] =: \mathbb{P}(X_{s+t} \in A \mid X_s),$$

las cuales son variables aleatorias, y por lo tanto, la igualdad en (6) debe entenderse en el sentido casi-seguro. Como es más común trabajar con esperanzas que con probabilidades condicionales, será muy útil tener una caracterización de la propiedad de Markov a través de la esperanza condicional, lo cual queda establecido en el siguiente teorema.

Teorema 14.

X es un proceso de Markov si y sólo si

$$(7) \quad \forall f \in \mathcal{B}: \mathbb{E}[f(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}[f(X_{s+t}) \mid X_s]$$

Demostración. Si f es una función indicadora, es sólo la definición. Por linealidad de la esperanza, la igualdad de cumple para funciones simples. Por convergencia monótona, se cumple para funciones no negativas. Luego, para f general se cumple al aplicar el caso de funciones no negativas a f^+ y f^- . \square

Definición 3.

Un proceso de Markov X es homogéneo si

$$(8) \quad \forall x \in E \text{ y } A \in \mathcal{E}: \mathbb{P}(X_{s+t} \in A \mid X_s = x) = \mathbb{P}(X_t \in A \mid X_0 = x)$$

Si X es homogéneo entonces el proceso $\hat{X} = \{X_{s+t} : t \in T\}$, dado que $X_s = x$, es el mismo (en distribución) que el proceso X . Por esta razón, si X cumple (8), la distribución inicial no suele especificarse; pues una vez que el proceso se encuentra en el estado x , éste se *reinicia*. Supondremos que X es homogéneo. La propiedad de ser homogéneo no es tan restrictiva como puede parecer. Un proceso no homogéneo puede convertirse en uno homogéneo si se “expande” el espacio de estados (Sharpe (1988)).

Si X es un proceso homogéneo, entonces se cumple

$$(9) \quad \forall f \in \mathcal{B}: \mathbb{E}[f(X_{s+t}) \mid X_s = x] = \mathbb{E}[f(X_t) \mid X_0 = x].$$

Cuando X modela la configuración de un sistema en un tiempo dado, es de interés observar el estado de dicho sistema en un tiempo aleatorio. En otras palabras, si τ denota el tiempo

aleatorio, queremos estudiar el estado X_τ del sistema en ese tiempo aleatorio. Sin embargo, hay un par de problemas con X_τ . El primero es que si el tiempo τ toma el valor ∞ , generalmente X_τ no está definida. Una solución en el contexto de procesos de Markov es añadir un punto p^* al espacio de estados de modo que $E_* = E \cup \{p^*\}$ está dotado con la topología definida mediante $\mathcal{T}_* = \mathcal{T} \cup \{A \cup \{p^*\} : A \in \mathcal{T}\}$, donde \mathcal{T} es la topología base de E , y $f \in \mathcal{B}$ se extiende a E_* definiendo $f(p^*) = 0$. El segundo problema es que X_τ no podría ser medible y necesitaremos herramientas técnicas que nos permitan considerar propiamente a X_τ como variable aleatoria.

Recordemos que τ es un tiempo de paro relativo a \mathbb{F} si para cada $t \in T$, el conjunto $\{\tau \leq t\} \in \mathcal{F}_t$, y que para cada tiempo de paro tenemos una σ -álgebra asociada \mathcal{F}_τ definida por

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ para todo } t \in T\}$$

que consiste de todos los eventos que podemos observar hasta el tiempo aleatorio τ . Como X es adaptado a \mathbb{F} es natural pensar que X_τ será medible respecto a \mathcal{F}_τ , pero esto no siempre es cierto a menos que supongamos que X es progresivamente medible respecto a \mathbb{F} , es decir, que el mapeo $X: \Omega \times [0, t] \rightarrow E$ es medible respecto a la σ -álgebra producto $\mathcal{F}_t \otimes \mathcal{B}([0, t])$, donde $\mathcal{B}([0, t])$ es la σ -álgebra de Borel sobre $[0, t]$. Un resultado estándar que no probaremos aquí nos dice que un proceso X es progresivamente medible si X tiene trayectorias càdlàg (ver Sharpe (1988)).

Con lo anterior en mente, y suponiendo que nuestro proceso X es suficientemente bondadoso definimos a los procesos fuertes de Markov.

Definición 4.

X es un proceso fuerte de Markov si para cualquier tiempo de paro τ se cumple que

$$\forall f \in \mathcal{B}: \mathbb{E}[f(X_{\tau+t}) \mid \mathcal{F}_\tau] = \mathbb{E}[f(X_{\tau+t}) \mid X_\tau]$$

No todo proceso de Markov es un proceso fuerte de Markov. Hay muchos contraejemplos disponibles y pueden consultarse en Fristedt y Gray (1997) y Sharpe (1988); pero la idea es sencilla: cuando paramos al proceso, y sólo observamos X_τ debe existir cierta “ambigüedad” en la evolución futura del proceso.

Nuestra siguiente discusión está relacionada con las probabilidades de transición del proceso de Markov X . Cuando trabajamos con cadenas de Markov a tiempo discreto, las probabilidades de transición están determinadas por la matriz P a través de sus potencias; y en el caso de cadenas de Markov a tiempo continuo, estas probabilidades satisfacen una ecuación diferencial en términos de su Q -matriz. Para nuestro caso general, buscamos caracterizar de la misma manera a nuestro proceso; lo que nos conduce a la siguiente definición.

Definición 5.

Dado $t \in T$, sea

$$P_t(x, A) := \mathbb{P}(X_t \in A \mid X_0 = x).$$

Entonces P_t es llamado kernel de transición del proceso X en el tiempo t .

Las primeras propiedades de $\{P_t : t \geq 0\}$ se establecen en la siguiente proposición.

Proposición 15.

Si $\{P_t : t \geq 0\}$ es el kernel de transición del proceso X entonces:

- (i) La función $x \mapsto \mathbb{P}(X_t \in A \mid X_0 = x)$ es medible,
- (ii) La función $A \mapsto \mathbb{P}(X_t \in A \mid X_0 = x)$ es una medida de probabilidad.

La función $\mathbb{P}_t(x, \cdot)$ es la distribución condicional de X_t dado $X_0 = x$. Por la propiedad de homogeneidad se cumple que $\mathbb{P}_t(x, \cdot)$ es también la distribución condicional de X_{s+t} dado $X_s = x$, es

dicir

$$\mathbb{P}_t(x, A) = \mathbb{P}(X_{s+t} \in A \mid X_s = x).$$

Notemos además que $P_0 = I$, donde I está definido como $I(x, A) = \mathbb{1}_A(x)$. Muchas veces, el espacio de estados posee una medida “natural” λ y en muchos casos, el kernel P_t tiene una densidad de transición con respecto de λ . Por ejemplo, si el espacio de estados coincide con $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ donde $\mathcal{B}(\mathbb{R})$ es la σ -álgebra de Borel y en este caso, la medida natural es la medida de Lebesgue. El siguiente teorema, de algún modo, justifica el nombre para la colección $\{P_t : t \geq 0\}$.

Teorema 16 (Chapman-Kolmogorov).

Si X es un proceso de Markov con kernel de transición $P = \{P_t : t \geq 0\}$, entonces

$$(10) \quad P_{s+t}(x, A) = \int_E P_s(x, dy) P_t(y, A)$$

Demostración. Condicionando sobre X_s tenemos que

$$\begin{aligned} P_{s+t}(x, A) &= \int_E \mathbb{P}(X_{s+t} \in A \mid X_s = y, X_0 = x) \mathbb{P}(X_s \in dy \mid X_0 = x) \\ &= \int_E \mathbb{P}(X_{s+t} \in A \mid X_s = y) P_s(x, dy) \quad (\text{prop. de Markov}) \\ &= \int_E \mathbb{P}(X_t \in A \mid X_s = y) P_s(x, dy) = \int_E P_s(x, dy) P_t(y, A) \end{aligned}$$

donde la penúltima igualdad se sigue de la homogeneidad del proceso X . \square

La igualdad (10) sólo nos dice que, partiendo de x , el proceso entra al conjunto A al tiempo $s + t$ si al tiempo s , pasó por el estado y ; siendo y cualquier elemento de E . También, (10) se conoce como la *propiedad de semigrupo* para el kernel P , esto es $P_{s+t} = P_s P_t$, donde el producto de kernels se define a través de

$$P_s P_t(x, A) := \int_E P_t(y, A) P_s(x, dy)$$

Observe que esta definición coincide con el producto matricial cuando consideramos a una cadena de Markov a tiempo discreto. Podemos considerar ahora cuando el semigrupo actúa sobre el espacio \mathcal{B} . Si $f \in \mathcal{B}$ entonces definimos

$$P_t f(x) := \int_E f(y) P_t(x, dy) = \mathbb{E}[f(X_t) \mid X_0 = x],$$

Notamos entonces que $P_t f$ es una función. Es decir, podemos pensar a P_t como un operador que mapea funciones en funciones. Más aún, no es difícil ver que para cada $t \in T$, P_t es un operador lineal y que es una contracción. Para esto, basta observar que si $f \in \mathcal{B}$, entonces

$$\begin{aligned} \|P_t f\| &= \sup \{|P_t f(x)| : x \in E\} \\ &= \sup \left\{ \left| \int_E f(y) P_t(x, dy) \right| : x \in E \right\} \\ &\leq \sup \left\{ \int_E |f(y)| P_t(x, dy) : x \in E \right\} \\ &\leq \sup \left\{ \int_E \sup_s |f(s)| P_t(x, dy) : x \in E \right\} \\ &= \sup_{s \in E} |f(s)| = \|f\| \end{aligned}$$

Esto implica que para cada $t \in T$, P_t mapea funciones acotadas en funciones acotadas y que es un operador continuo sobre \mathcal{B} .

Por otro lado, es posible calcular las distribuciones finito dimensionales del proceso X a través del kernel P , propiedad que enunciamos en la siguiente proposición

Proposición 17.

Si X es un proceso de Markov, $0 < t_1 < t_2 < \dots < t_n$ elementos de T y X_0 tiene distribución μ_0 , entonces

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) = \int_{A_0} \mu_0(dx_0) \int_{A_1} P_{t_1}(x_0, dx_1) \int_{A_2} P_{t_2-t_1}(x_1, dx_2) \cdots \int_{A_n} P_{t_n-t_{n-1}}(x_{n-1}, dx_n)$$

La Proposición 17 es muy importante para construir procesos de Markov. Si, en el contexto de nuestro problema, es sencillo definir el kernel P_t y si conocemos la distribución inicial μ_0 , entonces este resultado nos brinda una familia consistente de distribuciones y es posible aplicar el Teorema de Extensión de Kolmogorov. Muchas veces, esta construcción no es suficiente, ya que es usual requerir que X tenga trayectorias continuas, y dicha propiedad está fuera del alcance de las distribuciones finito-dimensionales. Este problema es sencillo de resolver, pero para tal propósito es necesario considerar un tipo especial de procesos de Markov y que resulta no muy restrictivo. Denotaremos por C al espacio de las funciones continuas y acotadas sobre E , y por C_0 al espacio

$$C_0 := \{f: E \rightarrow \mathbb{R} \text{ continua : } \forall \epsilon > 0 \exists K \text{ compacto: } \forall x \in K^c |f(x)| < \epsilon\},$$

por lo que las siguientes contenciones se cumplen $C_0 \subseteq C \subseteq \mathcal{B}$, y por lo tanto, C_0 y C son subespacios de \mathcal{B}

Definición 6 (Procesos de Feller).

Un proceso de Markov X es un proceso de Feller si las siguientes condiciones se cumplen:

- (i) Continuidad en el espacio: para $t \in T$ y $y \in E$, dado $X_0 = y$, la distribución condicional de X_t converge a la distribución condicional de X_t dado $X_0 = x$ cuando $x \rightarrow y$. Formalmente,

$$\forall f \in C: \mathbb{E}[f(X_t) \mid X_0 = x] \rightarrow \mathbb{E}[f(X_t) \mid X_0 = y] \text{ si } x \rightarrow y$$

- (ii) Continuidad en el tiempo: Dado $X_0 = x$ para $x \in E$, X_t converge en probabilidad a x cuando $t \downarrow 0$. Esto significa que,

$$\forall U \text{ vecindad de } x: \mathbb{P}(X_t \in U \mid X_0 = x) \rightarrow 1 \text{ si } t \downarrow 0.$$

Notemos que en el caso donde E es discreto, la condición (i) se satisface, mientras que si T es discreto (ii) se cumple. De este modo, toda cadena de Markov a tiempo discreto es un proceso de Feller. Ahora, damos una caracterización de los procesos de Feller en términos del semigrupo.

Teorema 18.

Si X es un proceso de Markov sobre E (que es localmente compacto, Hausdorff y segundo numerable) con kernels de transición $P = \{P_t: t \in T\}$, entonces X es un proceso de Feller si y sólo si:

- (i) Continuidad en el espacio: Si $f \in C_0$ y $t \in [0, \infty)$ entonces $P_t f \in C_0$
- (ii) Continuidad en el tiempo: Si $f \in C_0$ y $x \in E$ entonces $P_t f(x) \rightarrow f(x)$ cuando $t \downarrow 0$.

Un semigrupo $P = \{P_t: t \in T\}$ que cumple con las dos condiciones del Teorema 18 es llamado *semigrupo de Feller*. Así, un proceso de Markov X es un proceso de Feller si y sólo si su semigrupo es un semigrupo de Feller. La primera condición significa que C_0 es invariante bajo P_t , y la segunda condición implica una forma más fuerte de la continuidad en el tiempo.

Proposición 19.

Para $f \in C_0$, la función $t \mapsto P_t f$ es continua sobre T , esto es:

$$\|P_s f - P_t f\| = \sup\{|P_s f(x) - P_t f(x)| : x \in S\} \rightarrow 0 \text{ cuando } s \rightarrow t.$$

Como vemos, los procesos de Feller son bien portados en términos de continuidad. De hecho, son bastante bondadosos que podemos obtener información de sus trayectorias.

Teorema 20.

Sea X un proceso de Feller. Entonces

- (i) Existe una versión de X que tiene trayectorias càdlàg.
- (ii) Es un proceso fuerte de Markov respecto al refinamiento continuo por la derecha de la filtración natural \mathbb{F}^0 .

Un aspecto importante en la teoría de los procesos de Markov es el desarrollo de una metodología para entender el comportamiento global en términos de un operador que describe el comportamiento local del proceso. Esta característica es lo que permite considerarlos como la versión estocástica de sistemas descritos por ecuaciones diferenciales; lo que conduce a la siguiente definición.

Definición 7 (Generador Infinitesimal).

Si X es un proceso de Markov con kernel de transición P , el generador infinitesimal del proceso es un operador $G: \mathcal{D} \rightarrow C_0$ tal que

$$(11) \quad Gf := \lim_{t \downarrow 0} \frac{P_t f - f}{t}$$

y donde $\mathcal{D} \subseteq C_0$ es un conjunto donde el límite (11) existe.

El límite es calculado respecto a la norma del supremo en C_0 , esto es:

$$\left\| \frac{P_t f - f}{t} - g \right\| = \sup \left\{ \left| \frac{P_t f(x) - f(x)}{t} - g(x) \right| : x \in E \right\} \rightarrow 0 \text{ cuando } t \downarrow 0.$$

Dicho generador G es una generalización de la Q -matriz para cadenas de Markov a tiempo continuo, y así como la Q -matriz satisface las ecuaciones forward y backward de Kolmogorov, el generador G también.

Teorema 21.

Sea X es un proceso de Markov con semigrupo $P = \{P_t : t \in \mathbb{R}_+\}$ y generador G . Entonces, para toda $f \in \mathcal{D}$:

$$\forall x \in E: (P_t G)f(x) = (G P_t)f(x)$$

Bajo ciertas condiciones, se puede verificar que $P_t = e^{tG}$; que es la misma fórmula que se tiene para el caso de las cadenas de Markov a tiempo continuo. Más aún, si Q es la Q -matriz asociada a una cadena de Markov a tiempo continuo, se puede probar que

$$Gf(i) = \sum_j q_{ij}f(j) = (Qf)_i.$$

Terminaremos esta sección con un ejemplo que utilizaremos en el resto del trabajo. Un proceso de Markov con valores en $E = \mathbb{R}_+$ es llamado un *proceso de ramificación con espacio de estados continuo* si su semigrupo $P = \{P_t : t \in \mathbb{R}_+\}$ satisface la siguiente propiedad: para cualesquiera $x, y \in \mathbb{R}_+$ y $t \geq 0$,

$$P_t(x, \cdot) * P_t(y, \cdot) = P_t(x + y, \cdot),$$

donde $\nu * \mu$ es la convolución de las medidas ν y μ . La identidad anterior es conocida como la *propiedad de ramificación*; es decir, si X y Y son procesos de Markov independientes con el mismo semigrupo P , entonces $X + Y$ tiene el mismo semigrupo P . Supongamos que el semigrupo P satisface las siguientes dos propiedades:

- (i) $P_t(x, \{0\}) < 1$ para cada $x > 0$ y $t > 0$.
- (ii) $P_t(x, \cdot) \rightarrow \delta_x$ en el sentido débil.

Entonces se verifica que el semigrupo P es un semigrupo de Feller, y su transformada de Laplace satisface

$$\int_{\mathbb{R}_+} e^{-\lambda y} P_t(x, dy) = e^{-x\psi_t(\lambda)},$$

donde la familia de funciones $\{\psi_t : t \geq 0\}$ cumple que $\psi_t \circ \psi_s = \psi_{t+s}$. Si hacemos $P_t(x, dy)$ igual a la ley de $Y_1 + \dots + Y_N$, donde Y_i tiene una distribución exponencial de parámetro $1/t$, y N tiene ley Poisson de parámetro x/t e independiente de $\{Y_i\}$, entonces $P = \{P_t : t \geq 0\}$ es un semigrupo de Feller y además,

$$\int e^{-\lambda y} P_t(x, dy) = e^{-x\psi_t(\lambda)},$$

con $\psi_t(\lambda) = \lambda/(1+\lambda t)$. El proceso de Markov X asociado a P admite una versión con trayectorias continuas. A dicha versión de X se le conoce como *difusión de Feller* o *difusión de ramificación de Feller*. En la siguiente sección partimos de otra caracterización de esta misma difusión.

4. Aproximaciones a difusiones

En las secciones anteriores se vieron el espacio D , consistente de las funciones càdlàg, y algunas generalidades de los procesos de Markov. El objetivo de esta sección será el dar un ejemplo de la convergencia en el espacio D de ciertos procesos de Markov. Como se anticipó en la Sección 1, el ejemplo será similar al teorema de Donsker pero reescalando de manera adecuada procesos de Galton–Watson. Este límite fue originalmente conjeturado y estudiado por Feller (1951), por lo que se conoce como *difusión de Feller*. No obstante, pasaron años antes de que Jiřina (1969) y Lindvall (1972) demostraran formalmente, parece ser que de manera independiente, tal límite y en qué sentido se daba. La prueba que mostraremos usa la convergencia de semigrupos de procesos de Markov y se basa en resultados de Ethier y Kurtz (2005), Kurtz (1981) y en menor medida Kallenberg (2021).

La difusión de Feller se define como el proceso estocástico $\{X_t : t \in \mathbb{R}_+\}$ con trayectorias continuas que es solución la ecuación diferencial estocástica

$$(12) \quad dX_t = rX_t dt + \sqrt{\sigma^2 X_t} dB_t,$$

donde B es un movimiento browniano estándar. Aunque aquí le decimos difusión de Feller a (12), en la literatura es usual que tal nombre se restrinja al caso $r = 0$ y $\sigma^2 = 2$. Para obtener el generador infinitesimal de este proceso notemos que, por la Fórmula de Itô, ver el Teorema 7.20 de Çınlar (2011), para toda función $f \in C^2$ con soporte compacto,

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s \\ &= f(X_0) + \int_0^t \left(rX_s f'(X_s) + \frac{1}{2}\sigma^2 X_s f''(X_s) \right) ds + \int_0^t \sqrt{\sigma^2 X_s} f'(X_s) dB_s. \end{aligned}$$

Puede comprobarse que $\mathbb{E}[\int_0^t \sqrt{\sigma^2 X_s} f'(X_s) dB_s] = 0$, por lo que

$$\begin{aligned} P_t f(x) - f(x) &= \mathbb{E}_x \left[\int_0^t \left(r X_s f'(X_s) + \frac{1}{2} \sigma^2 X_s f''(X_s) \right) ds \right] \\ &= \int_0^t \mathbb{E}_x \left[r X_s f'(X_s) + \frac{1}{2} \sigma^2 X_s f''(X_s) \right] ds, \end{aligned}$$

y entonces, por el Teorema Fundamental del Cálculo y la Definición 7,

$$Gf(x) = \lim_{t \downarrow 0} \frac{P_t f(x) - f(x)}{t} = \mathbb{E}_x \left[r X_0 f'(X_0) + \frac{1}{2} \sigma^2 X_0 f''(X_0) \right] = r x f'(x) + \frac{1}{2} \sigma^2 x f''(x).$$

Para poder enunciar el resultado principal, deberemos hacer algunas observaciones acerca de los generadores infinitesimales. En general, se dice que un operador A con dominio $\mathcal{D} \subset E$, donde E es un espacio de Banach, es *cerrado* si la gráfica $\{(f, Af) : f \in \mathcal{D}\}$ es cerrada en E^2 . Cuando esta gráfica no es cerrada pero su cerradura corresponde a la gráfica de un operador \bar{A} entonces decimos que A es *cerrable*. En el caso en el cual A es cerrado, un *núcleo* para A es un subespacio lineal $D \subset \mathcal{D}$ tal que $A|_D$ tiene cerradura A , en cuyo caso este último operador queda determinado por su restricción a D . Por ejemplo, para los generadores dados anteriormente, el Teorema 2.1 del Capítulo 8 de Ethier y Kurtz (2005) asegura que C_c^∞ es un núcleo para G . Con estas observaciones, el siguiente resultado nos da un criterio de convergencia de cadenas de Markov a procesos estocásticos con índice de tiempo continuo. La prueba del mismo, en contextos más generales, puede consultarse en la Sección 6 del Capítulo 1 de Ethier y Kurtz (2005) o el Capítulo 17 de Kallenberg (2021).

Teorema 22.

Sean X^1, X^2, \dots cadenas de Markov, en tiempo discreto, con valores en \mathbb{R} y operadores de transición U_1, U_2, \dots , y sea X un proceso de Feller \mathbb{R} -valuado con semigrupo $\{P_t\}$ y generador G . Fijemos un núcleo D para G . Si definimos los operadores y procesos

$$G_n = n(U_n - \text{id}), \quad P_t^n = U_n^{\lfloor nt \rfloor}, \quad Y_t^n = X_{\lfloor nt \rfloor}^n,$$

entonces se da la equivalencia entre

- (i) para $f \in D$ existe $f_n \in \mathcal{D}_n$ tal que $f_n \rightarrow f$ y $G_n f_n \rightarrow Gf$;
- (ii) para cada $f \in C_0$, $P_t^n f \rightarrow P_t f$ de manera uniforme en $t > 0$;
- (iii) $Y_0^n \Rightarrow X_0$ en \mathbb{R} implica $Y_n \Rightarrow X$ en el espacio de Skorokhod.

Por el Teorema 17.2 de Kallenberg (2021) los operadores G_n que se definieron corresponden a las cadenas de Markov en tiempo continuo $\tilde{Y}_t^n = X^n \circ N_{nt}$, donde $\{N_t : t \geq 1\}$ es un proceso Poisson homogéneo de intensidad uno, pues notemos que para este proceso se tiene que

$$\begin{aligned} \mathbb{E}_x[f(\tilde{Y}_t^n)] &= \sum_{k=0}^{\infty} \mathbb{E}_x[f(X_k^n)] \mathbb{P}(N_{nt} = k) = \sum_{k=0}^{\infty} e^{-nt} \frac{(nt)^k}{k!} U_n^k f(x) \\ &= e^{nt(U_n - \text{id})} f(x) = e^{tG_n} f(x), \end{aligned}$$

mostrando que G_n es, efectivamente, el generador infinitesimal del proceso \tilde{Y}^n . Por otra parte, los operadores P_t^n esencialmente nos dicen que estamos acelerando el tiempo en el que se dan los saltos de las cadena de Markov X^n consideradas en el Teorema. Con esto finalmente estamos listos para enunciar el resultado principal de esta sección.

Resultado Principal.

Para cada $n \in \mathbb{N}$ sea $\Xi_n = \{\xi_{k,m}^n : k, m \in \mathbb{N}\}$ un arreglo doble de variables aleatorias independientes e idénticamente distribuidas, con valores en \mathbb{N}_0 , $\mathbb{E}[\xi_{1,1}^n] = 1 + r/n$, con $r \geq 0$, y $\text{Var}(\xi_{1,1}^n) = \sigma^2 \in (0, \infty)$.

Supongamos además que existe $\delta > 0$ tal que

$$(13) \quad \sup_{n \in \mathbb{N}} \mathbb{E}[(\xi_{1,1}^n)^{2+\delta}] < \infty.$$

Para $n \in \mathbb{N}$ sea X^n un proceso de Galton–Watson con progenie dada por Ξ_n . Si se tiene que $X_0^n = x_n$, con $\{x_n\}$ satisfaciendo $x_n/n \rightarrow x \in \mathbb{R}_+$, y definimos

$$Y_t^n = \frac{1}{n} X_{\lfloor nt \rfloor}^n, \quad t \geq 0,$$

entonces $Y^n \Rightarrow X$ en el espacio de Skorokhod, donde X es una difusión cuyo generador es la cerradura de

$$Gf(x) = rx f'(x) + \frac{1}{2} \sigma^2 x f''(x), \quad f \in C_c^\infty.$$

Como podrá sospecharse, la herramienta que usaremos para demostrar este resultado es el Teorema 22. Por las observaciones subsecuentes al mismo, en realidad lo que estamos viendo es que los procesos \tilde{Y}^n son los que convergen a la difusión dada, aunque el resultado se seguirá dado que conforme n tiende a infinito, los procesos Y^n y \tilde{Y}^n se parecen en un sentido que se aclara en el siguiente enunciado.

Lema 23.

Para cualesquiera $t > 0$ y $\epsilon > 0$, $\mathbb{P}(d'_t(Y^n, \tilde{Y}^n) > \epsilon) \rightarrow 0$.

Demostración. Para esta prueba usaremos la métrica d_t por ser más sencilla de manejar. Definamos $v_0^n = 0$ y de manera recursiva

$$v_k^n = \inf\{s > v_{k-1}^n : N_{nt} > N_{nv_{k-1}^n}\}, \quad k \in \mathbb{N}.$$

Sea además $m_n = \min\{k \in \mathbb{N}_0 : v_{k+1}^n > t\}$. Con estos tiempos definamos la función, aleatoria,

$$\lambda_n(s) = \sum_{k=0}^{m_n} (v_k^n + (ns - k)(v_{k+1}^n - v_k^n)) \mathbb{1}_{[k/d, (k+1)/d)}(s) + \left(v_{m_d}^n + \frac{ns - m_d}{nt - m_d} (t - v_{m_d}^n) \right) \mathbb{1}_{[m_n/d, t]}(s).$$

Esta función es, con probabilidad uno, continua, estrictamente creciente y tiene por puntos fijos a 0 y t , esto es $\lambda_n(0) = 0$ y $\lambda_n(t) = t$. Adicionalmente $\tilde{Y}_{\lambda_n(s)}^n = Y_s^n$ para $s \in [0, t]$ con probabilidad uno. Por consiguiente obtenemos las relaciones casi seguras

$$d_t(Y^n, \tilde{Y}^n) \leq \|\lambda_n - \text{id}\|_{u,t} + \|\tilde{Y}_{\lambda_n}^n - Y^n\|_{u,t} = \|\lambda_n - \text{id}\|_{u,t}.$$

Dado que λ_n es una función lineal por partes, se sigue que el supremo de su diferencia con la identidad debe suceder en alguno de los vértices, esto es

$$\|\lambda_n - \text{id}\|_{u,t} = \sup_{1 \leq k \leq N_{nt}} \left| \frac{k}{n} - v_k^n \right| = \frac{1}{n} \sup_{1 \leq k \leq N_{nt}} |k - nv_k^n| = \frac{1}{n} \sup_{1 \leq k \leq N_{nt}} |N_{nv_k^n} - nv_k^n| \leq \frac{1}{n} \sup_{0 \leq s \leq t} |N_{ns} - ns|.$$

Así pues, tenemos la desigualdad casi segura

$$d_t(Y^n, \tilde{Y}^n) \leq \frac{1}{n} \sup_{0 \leq s \leq t} |N_{ns} - ns|,$$

de donde se sigue el resultado en vista del Corolario 3.5 de Bhattacharya y Waymire (2016) pues, al ser $\{N_{ns} - ns : s \in \mathbb{R}_+\}$ una martingala continua por la derecha cuadrado-integrable, tal resultado nos dice que

$$\mathbb{P}\left(\sup_{0 \leq s \leq t} |N_{ns} - ns| > ne\right) \leq \frac{1}{n^2 e^2} \mathbb{E}[|N_{nt} - nt|^2] = \frac{t}{ne^2} \rightarrow 0. \quad \square$$

De esta manera, si logramos probar que $\tilde{Y}^n \Rightarrow X$, dado que X es un proceso con trayectorias continuas, por el Teorema 12 se sigue esto sucede si y sólo si para cada $t > 0$,

$$\{\tilde{Y}_s^n : 0 \leq s \leq t\} \Rightarrow \{X_s : 0 \leq s \leq t\} \quad \text{en } D_t.$$

En virtud del Lema anterior y el Teorema 3.1 de Billingsley (1999) esto implica que para cada $t > 0$,

$$\{Y_s^n : 0 \leq s \leq t\} \Rightarrow \{X_s : 0 \leq s \leq t\} \quad \text{en } D_t,$$

siendo esto equivalente a $Y^n \Rightarrow X$. Por lo tanto, establecer la convergencia de las cadenas de Markov en tiempo continuo nos dará el resultado deseado.

Para la prueba de la convergencia, primero notemos que Y^n , o bien \tilde{Y}^n , toma valores en el subconjunto de \mathbb{R} dado por $E_n = \{k/n : k \in \mathbb{N}_0\}$ por lo que hay que demostrar que para $f \in C_c^\infty$ dada,

$$\lim_{n \rightarrow \infty} \sup_{x \in E_n} \left| n(U_n f(x) - f(x)) - rx f'(x) - \frac{1}{2} \sigma^2 x f''(x) \right| = \lim_{n \rightarrow \infty} \sup_{x \in E_n} |G_n f(x) - G f(x)| = 0.$$

Fijando $n \in \mathbb{N}$ y $x \in E_n \setminus \{0\}$, notemos que U_n es el operador asociado a X^n/n , por lo que se tiene que

$$U_n f(x) = \mathbb{E} \left[f \left(\frac{1}{n} \sum_{k=1}^{nx} \xi_{1,k}^n \right) \right].$$

Por otra parte observemos que se dan las igualdades

$$rx = \sqrt{n} \mathbb{E} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right]$$

y

$$\begin{aligned} \sigma^2 x &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) - \frac{rx}{n} \right)^2 \right] \\ &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right)^2 \right] - 2 \frac{rx}{\sqrt{n}} \mathbb{E} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right] + \frac{r^2 x^2}{n} \\ &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right)^2 \right] - \frac{r^2 x^2}{n}. \end{aligned}$$

Por consiguiente deducimos que

$$\begin{aligned} G_n f(x) - G f(x) &= n \mathbb{E} \left[f \left(\frac{1}{n} \sum_{k=1}^{nx} \xi_{1,k}^n \right) - f(x) \right] - \sqrt{n} \mathbb{E} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right] f'(x) \\ &\quad - \frac{1}{2} \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right)^2 \right] f''(x) + \frac{r^2 x^2}{n} f''(x). \end{aligned}$$

Puesto que f tiene soporte compacto, es claro que existe $K > 0$ tal que $\sup_{x \in E_n} |r^2 x^2 f''(x)| \leq K$, por lo que únicamente deberemos establecer la convergencia $\lim_n \sup_x \epsilon_n(x) = 0$, donde se define $\epsilon_n(x)$ como

$$\left| n \mathbb{E} \left[f \left(\frac{1}{n} \sum_{k=1}^{nx} \xi_{1,k}^n \right) - f(x) \right] - \sqrt{n} \mathbb{E} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right] f'(x) - \frac{1}{2} \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1) \right)^2 \right] f''(x) \right|.$$

Una aplicación de integración por partes nos da la identidad,

$$\int_0^1 y^2 x(1-v) \left(f'' \left(x + v\sqrt{\frac{x}{n}}y \right) - f''(x) \right) dv = nf \left(x + \sqrt{\frac{x}{n}}y \right) - nf(x) - \sqrt{nx} y f'(x) - \frac{x}{2} y^2 f''(x).$$

Sustituyendo $S_{nx} = (nx)^{-1/2} \sum_{k=1}^{nx} (\xi_{1,k}^n - 1)$ en lugar de y en la expresión anterior nos da la igualdad

$$\epsilon_n(x) = \left| \mathbb{E} \left[\int_0^1 S_{nx}^2 x(1-v) \left(f'' \left(x + v\sqrt{\frac{x}{n}}S_{nx} \right) - f''(x) \right) dv \right] \right|.$$

Supongamos ahora que el soporte de f está contenido en $[0, M]$ y observemos que, dada la no negatividad de $\xi_{1,1}^n$,

$$x + v\sqrt{\frac{x}{n}}S_{nx} \geq x - v\frac{1}{n} \times nx = x(1-v).$$

La expresión $x(1-v)$ es mayor o igual que M si y sólo si $v \leq 1 - M/x$. En consecuencia, para los valores $0 \leq v \leq 0 \vee (1 - M/x)$ obtenemos

$$f'' \left(x + v\sqrt{\frac{x}{n}}S_{nx} \right) - f''(x) = 0,$$

de modo que

$$\begin{aligned} (14) \quad & \left| \int_0^1 S_{nx}^2 x(1-v) \left(f'' \left(x + v\sqrt{\frac{x}{n}}S_{nx} \right) - f''(x) \right) dv \right| \\ & \leq \int_{0 \vee (1-M/x)}^1 S_{nx}^2 x(1-v) \left| f'' \left(x + v\sqrt{\frac{x}{n}}S_{nx} \right) - f''(x) \right| dv \\ & \leq \int_{0 \vee (1-M/x)}^1 S_{nx}^2 x(1-v) 2 \|f''\| dv = S_{nx}^2 x \|f''\| \left(1 \wedge \frac{M}{x} \right)^2. \end{aligned}$$

Siguiendo las ideas de Ethier y Kurtz (2005), para demostrar $\lim_n \sup_x \epsilon_n(x) = 0$ demostraremos que para cualquier sucesión $\{x_n\}$ con $x_n \in E_n$ convergente, siendo ∞ un posible límite, $\lim_n \epsilon_n(x_n) = 0$. Puesto que

$$\epsilon_n(x) \leq x \|f''\| \left(1 \wedge \frac{M}{x} \right)^2 \mathbb{E}[S_{nx}^2] = x \|f''\| \left(1 \wedge \frac{M}{x} \right)^2 \left(\sigma^2 + \frac{r^2 x}{n} \right),$$

es claro que $\epsilon_n(x_n) \rightarrow 0$ cuando $x_n \rightarrow 0$ o cuando $x_n \rightarrow \infty$, por lo que podemos únicamente considerar el caso $x_n \rightarrow x$ con $x \in (0, \infty)$. Por la condición 13 y el Teorema 6.5.1 de Resnick (2014) se tiene que

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma^2} \int_{\{|\xi_{1,1}^n - 1 - r/n| \geq \varepsilon \sigma \sqrt{nx_n}\}} \left(\xi_{1,1}^n - 1 - \frac{r}{n} \right)^2 d\mathbb{P} = 0,$$

lo que a su vez implica, en vista del Teorema 27.2 de Billingsley (1995), la convergencia

$$\frac{1}{\sqrt{nx_n}} \sum_{k=1}^{nx_n} \left(\xi_{1,k}^n - 1 - \frac{r}{n} \right) \Rightarrow Z, \quad Z \sim N(0, \sigma^2).$$

Luego, del Teorema de Slutsky, $S_{nx_n} \Rightarrow Z$ y en consecuencia

$$V_n := \int_0^1 S_{nx_n}^2 x_n(1-v) \left(f'' \left(x_n + v\sqrt{\frac{x_n}{n}}S_{nx_n} \right) - f''(x_n) \right) dv \Rightarrow 0.$$

Finalmente, observando que

$$\begin{aligned}\mathbb{E} \left[x_n \|f''\| \left(1 \wedge \frac{M}{x_n} \right)^2 S_{nx_n}^2 \right] &= x_n \|f''\| \left(1 \wedge \frac{M}{x_n} \right)^2 \left(\sigma^2 + \frac{r^2 x_n}{n} \right) \\ &\rightarrow x \|f''\| \left(1 \wedge \frac{M}{x} \right)^2 \sigma^2 = \mathbb{E} \left[x \|f''\| \left(1 \wedge \frac{M}{x} \right)^2 Z^2 \right],\end{aligned}$$

de (14) y una generalización del Teorema de Convergencia Dominada para convergencia en probabilidad, se sigue que

$$\limsup_{n \rightarrow \infty} \epsilon_n(x_n) \leq \lim_{n \rightarrow \infty} \mathbb{E}[|V_n|] = 0,$$

lo cual concluye la prueba.

En la demostración que se acaba de realizar, el supuesto (13) se usó para poder justificar la convergencia débil de S_{nx} a una variable aleatoria normal. Si permitimos que $\Xi_n = \Xi_1$ para toda n , lo que en particular implica que la media de reproducción es 1, podemos dispensar de tal supuesto gracias al Teorema Central del Límite para variables aleatorias independientes e idénticamente distribuidas, que es lo que usan tanto Ethier y Kurtz (2005) y Kurtz (1981).

Finalmente, para tener una representación visual de la difusión y de qué tanto se parece el proceso de Galton–Watson reescalado a esta, se realizaron simulaciones con procesos de Galton–Watson tales que $\xi_{1,1}^n \sim \text{Geom}(1/2)$ para cada $n \in \mathbb{N}$. En este caso la difusión resultante resuelve la ecuación diferencial estocástica

$$dX_t = \sqrt{2X_t} dB_t,$$

misma ecuación que se usa para simular la difusión mediante el algoritmo de Euler–Maruyama. En la figura 5a se muestra una trayectoria del proceso Y_t , que es un reescalamiento de un proceso de Galton–Watson que comienza con 10000 individuos. En tanto, en la figura 5b se muestra una trayectoria del proceso X_t , correspondiente a la difusión. Mediante estas ilustraciones, podemos observar que ambas trayectorias tienen un comportamiento irregular parecido, por lo que es un primer indicio heurístico de que el resultado que se probó es cierto. Por otra parte, en la figura 6 se muestran 100 trayectorias como las que ya se describieron. Con estas gráficas se puede observar que tanto los procesos reescalados como las difusiones tienen una tendencia decreciente y que hay algunas trayectorias que se separan del resto, pero en ambos casos son los menos.

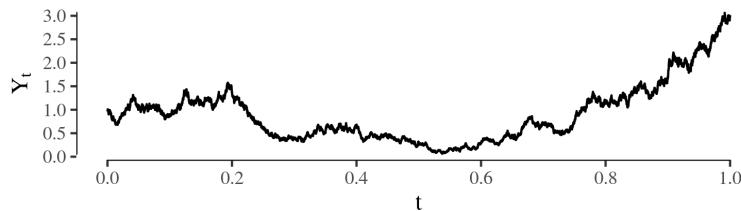
Para ilustrar el caso en el cual $r > 0$, donde las progenies ya no son constantes, se consideró que para cada $n \in \mathbb{N}$, $\xi_{1,1}^n$ tuviese distribución binomial negativa con parámetro de tamaño

$$k = \frac{(1 + r/n)^2}{\sigma^2 - (1 + r/n)}$$

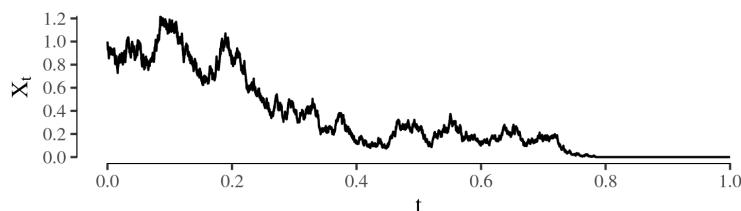
y parámetro de probabilidad

$$p = \frac{1 + r/n}{\sigma^2},$$

lo que nos asegura que se satisfacen las condiciones pedidas en el resultado principal, en particular $\mathbb{E}[\xi_{1,1}^n] = 1 + r/n$, $\text{Var}(\xi_{1,1}^n) = \sigma^2$ mientras que la condición (13) se cumple para $\delta = 1$, pues puede demostrarse que con los parámetros elegidos la sucesión $\{\mathbb{E}[(\xi_{1,1}^n)^3] : n \geq N_0\}$ es una sucesión convergente, donde se elige $N_0 \in \mathbb{N}$ de manera que $\sigma^2 - (1 + r/n) > 0$ se cumpla siempre que $n \geq N_0$. En la figura 7 se muestran 300 trayectorias simuladas tanto de los procesos reescalados de Galton–Watson como de las difusiones para $n = 10000$, $r = 1$ y $\sigma^2 = 2$. Mientras tanto, en la figura 8 se muestran 300 trayectorias de los mismos procesos que antes

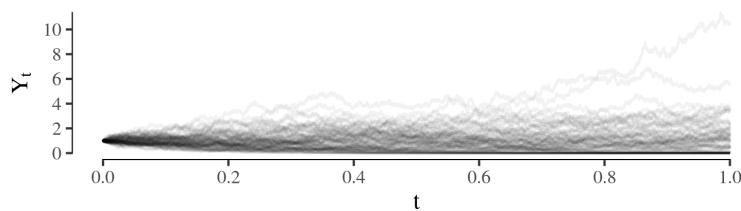


(a) Galton–Watson reescalado

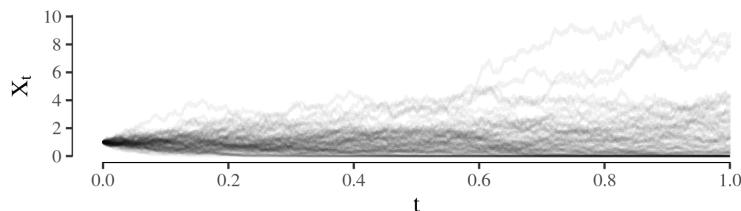


(b) Difusión de Feller

Figura 5: Simulación de una trayectoria de un proceso de Galton–Watson reescalado y una difusión de Feller.



(a) Galton–Watson reescalado



(b) Difusión de Feller

Figura 6: Simulación de 100 trayectorias de un proceso de Galton–Watson reescalado y una difusión de Feller.

considerando $r = 3$, dejando los demás parámetros constantes. La razón de simular más trayectorias que antes es, de manera intuitiva, que incrementar el parámetro r incrementa la varianza de los procesos de Galton–Watson, por lo que para ilustrar el comportamiento promedio se requiere de más simulaciones. Por esta misma razón, el incremento en las varianzas de los procesos, las trayectorias de los procesos reescalados presentan valores más altos que aquellos observados en la difusión. No obstante, puede observarse que los reescalamientos son similares a las difusiones por lo que esto ilustra la convergencia para el caso $r > 0$.

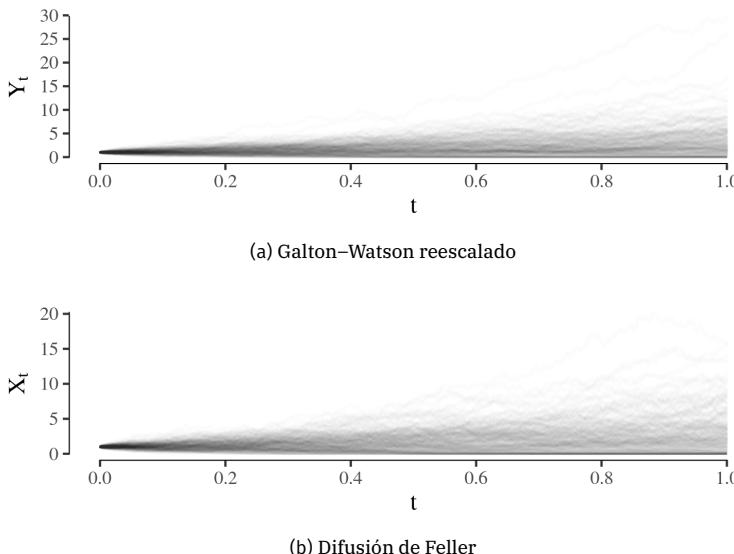


Figura 7: Simulación de 300 trayectorias de un proceso de Galton–Watson reescalado y una difusión de Feller con $r = 1$ y $\sigma^2 = 2$.

Para cerrar la sección, vale la pena mencionar que la convergencia del proceso de Galton–Watson a la difusión de Feller no es del todo fortuita, ya que esta difusión es uno de los procesos de ramificación en tiempo continuo y con espacio de estados continuo más sencillos que se conocen. Algunas de sus propiedades se discuten en Dawson (2017) y Lamperti (1967a). En Dawson (2017) también se encuentra una extensión de esta difusión al caso en el que hay inmigración, aunque no hay un resultado similar al de la convergencia. Finalmente, el mismo Feller (1951) conjecturó que el proceso de Wright–Fisher también puede reescalarse de manera que podamos aproximar otra difusión, aunque es de interés que la prueba de esta convergencia use herramientas distintas a las que se presentaron en este trabajo.

5. Conclusiones

Al principio de este trabajo se dio un breve resumen de los procesos de Galton–Watson, recordando que son cadenas de Markov en tiempo continuo que también son conocidos como procesos de ramificación, los cuales pueden verse como elementos aleatorios de D . Mediante simulaciones se motivó la idea de que conforme la población inicial aumenta, al reescalar de forma conveniente a los procesos, estos convergen a una difusión.

De este modo, se estudió el espacio D y se le dotó de la métrica de Skorokhod, la cual lo hace completo y separable. A lo largo del estudio de este espacio se remarcó que tanto el espacio como la topología que se construyeron son generalizaciones adecuadas del espacio de

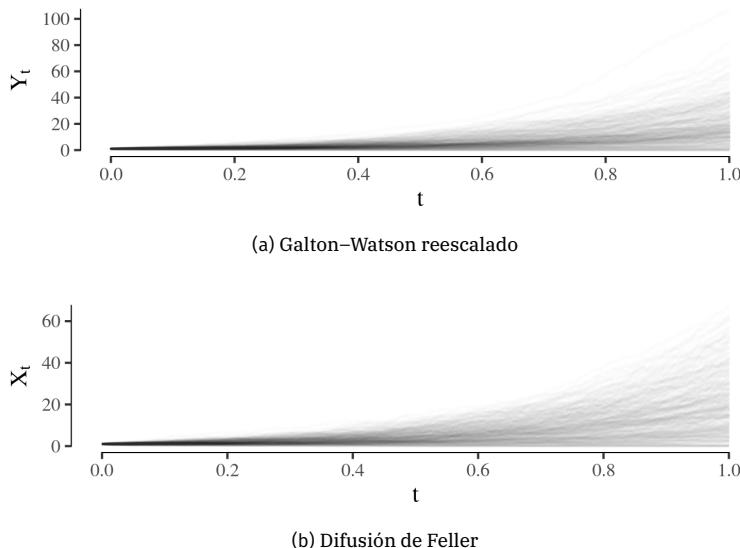


Figura 8: Simulación de 300 trayectorias de un proceso de Galton–Watson reescalado y una difusión de Feller con $r = 3$ y $\sigma^2 = 2$.

funciones continuas. En vista de que las cadenas de Markov pueden verse como elementos de D , también se dieron criterios para verificar la tensión de una sucesión de elementos aleatorios en este espacio. Asimismo se dieron un par de criterios, no exhaustivos, para poder asegurar la convergencia de elementos aleatorios en D .

Posteriormente se introdujeron algunas generalidades de los procesos de Markov. Esto se realizó con dos finalidades: el extender nociones que se tienen para cadenas de Markov tanto en tiempo discreto como en tiempo continuo. De particular interés para el trabajo son los conceptos de generador infinitesimal y de procesos de Feller, ya que la convergencia prevista desde la primera sección es a uno de estos procesos.

Finalmente se dio el resultado principal, es decir la convergencia de procesos de Galton–Watson reescalados al proceso conocido como difusión de Feller. La demostración que se muestra en este trabajo hace uso de la convergencia de ciertos generadores infinitesimales, extendiendo ligeramente el Teorema 1.3 del Capítulo 9 de Ethier y Kurtz (2005). En adición a la prueba formal, se mostraron algunas simulaciones realizadas para comprobar empíricamente que esta convergencia sí se da. Al final de la misma sección se mencionó que se conocen otros límites a difusiones, aunque estos rebasan el alcance del trabajo, misma razón por la cual ya no se incluyeron algunas de las propiedades de la difusión de Feller, como que su transformada de Laplace. No obstante, a lo largo del trabajo se mencionaron referencias que pueden consultarse si el lector tiene interés de profundizar lo visto en este escrito.

Referencias

- Bhattacharya, R. & Waymire, E. C. (2016). *A Basic Course in Probability Theory* (2.^a ed.). Springer International Publishing : Imprint: Springer. <https://doi.org/10.1007/978-3-319-47974-3>
- Billingsley, P. (1995). *Probability and Measure* (3.^a ed.). Wiley.
- Billingsley, P. (1999). *Convergence of Probability Measures* (2.^a ed.). Wiley.

- Çınlar, E. (2011). *Probability and Stochastics*. Springer.
- Dawson, D. A. (2017). Introductory Lectures on Stochastic Population Systems. *arXiv:1705.03781 [math]*. Consultado el 17 de mayo de 2021, desde <http://arxiv.org/abs/1705.03781>
- Ethier, S. N. & Kurtz, T. G. (2005). *Markov Processes: Characterization and Convergence*. Wiley Interscience.
- Feller, W. (1951). Diffusion processes in genetics. *Proc. Berkeley Sympos. Math. Statist. Prob.*, 227-246. University of California Press.
- Fristedt, B. & Gray, L. (1997). *A Modern Approach to Probability Theory*. Birkhäuser Boston, Inc., Boston, MA. <https://doi.org/10.1007/978-1-4899-2837-5>
- Heyde, C. C. & Seneta, E. (1977). *I.J. Bienaymé: Statistical Theory Anticipated*. Springer New York. Consultado el 18 de mayo de 2021, desde <https://doi.org/10.1007/978-1-4684-9469-3>
- Jiřina, M. (1969). On Feller's branching diffusion processes. *Časopis pro pěstování matematiky*, 094(1), 84-90. <https://doi.org/10.21136/CPM.1969.117652>
- Kallenberg, O. (2021). *Foundations of modern probability* (Vol. 99). Springer International Publishing. <https://doi.org/10.1007/978-3-030-61871-1>
- Kimmel, M. & Axelrod, D. E. (2015). *Branching Processes in Biology* (2.^a ed.). Springer.
- Kurtz, T. G. (1981). *Approximation of Population Processes*. Society for Industrial; Applied Mathematics.
- Lambert, A. (2008). Population Dynamics and Random Genealogies. *Stochastic Models*, 24, 45-163. <https://doi.org/10.1080/15326340802437728>
- Lamperti, J. (1967a). The limit of a sequence of branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 7(4), 271-288. <https://doi.org/10.1007/BF01844446>
- Lamperti, J. (1967b). Limiting distributions for branching processes. *Proc. Fifth Berkeley Sympos. math. Statist. Probability*, 225-241.
- Lindvall, T. (1972). Convergence of critical Galton–Watson branching processes. *Journal of Applied Probability*, 9(2), 445-450. <https://doi.org/10.2307/3212815>
- Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *Journal of Applied Probability*, 10(1), 109-121. <https://doi.org/10.2307/3212499>
- Munkres, J. R. (2014). *Topology* (2.^a ed.). Pearson.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag.
- Resnick, S. I. (1992). *Adventures in Stochastic Processes*. Birkhäuser.
- Resnick, S. I. (2014). *A Probability Path*. Birkhäuser/Springer.
- Sharpe, M. (1988). *General Theory of Markov Processes* (Vol. 133). Academic Press, Inc., Boston, MA.