

Procesos Estocásticos I

Capacitación técnica especializada en el nuevo marco de Solvencia

Gerónimo Uribe Bravo

Instituto de Matemáticas

Universidad Nacional Autónoma de México

CAPÍTULO 1

Introducción

1. Definiciones y clasificación básica de procesos estocásticos

Un **proceso estocástico** es una colección de variables aleatorias $(X_t)_{t \in T}$ indexadas por un conjunto T y definidas en algún espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Interpretamos al conjunto de índices T como un parámetro temporal; para nosotros T será $\{0, \dots, n\}$, \mathbb{N} , algún intervalo $[0, t]$ ó $[0, \infty)$. Interpretamos a un proceso estocástico como la evolución en el tiempo de algún fenómeno cuya dinámica se rige por el azar. Un ejemplo sencillo de esto es la cantidad de soles que vamos acumulando al participar en un juego de volados. Otro ejemplo es la evolución en el tiempo de la reserva de una compañía de seguros. En el primer ejemplo, se puede indexar al proceso por algún intervalo de naturales, en cuyo caso hablaremos de un proceso estocástico a **tiempo discreto**. Además, dicho proceso toma valores en los naturales, por lo que también se trata de un proceso con **espacio de estados discreto**. En el segundo caso, se puede pensar en un modelo indexado por un subintervalo de $[0, \infty)$ y hablaremos de un proceso estocástico a **tiempo continuo**. Además, en principio el valor de la reserva podría ser cualquier real no-negativo y por lo tanto hablamos de un proceso con **espacio de estados continuo**.

Uno de los primeros resultados generales dentro de la teoría de los procesos estocásticos es el teorema de consistencia de Kolmogorov que nos permite construir procesos estocásticos a partir de colecciones vectores aleatorios (que satisfacen la condición técnica de ser consistentes). La prueba de este teorema se puede hacer basándose en la existencia de una sucesión de variables aleatorias uniformes. Antes de analizar por qué existe una sucesión de variables uniformes independientes, ejemplificaremos cómo se pueden construir algunos de los procesos estocásticos que analizaremos en este curso.

EJEMPLO 1.1 (Caminatas aleatorias simples y el problema de la ruina). Imaginemos la siguiente situación: tengo un capital de 20 pesos al tiempo cero y cada instante de tiempo apuesto un peso en un volado, ganando si cae águila. ¿cómo puedo estudiar matemáticamente a la evolución de mi capital en el tiempo? De particular interés es la variable aleatoria que nos indica el instante en que me arruino por primera vez, misma que a priori podría ser infinita si jamás me arruino.

El modelo matemático es el siguiente: consideremos variables aleatorias

$$U_1, U_2, \dots$$

uniformes en $(0, 1)$ e independientes. A la variable aleatoria $\mathbf{1}_{U_i \leq 1/2}$, que toma los valores cero y uno, la interpretaremos como indicándonos si el resultado del i -ésimo volado es águila (cuando toma el valor uno) y por lo tanto, la variable $2\mathbf{1}_{U_i \leq 1/2} - 1$ toma los valores 1 si cae águila y -1 si cae sol.

EJERCICIO 1.1. Con el objeto de verificar que comprendemos la noción de independencia, probar que las variables aleatorias $\mathbf{1}_{U_1 \leq 1/2}, \mathbf{1}_{U_2 \leq 1/2}, \dots$ son independientes y con distribución Bernoulli de parámetro $1/2$.

Finalmente, podemos definir

$$X_0 = 20 \quad \text{y} \quad X_{n+1} = X_n + 2\mathbf{1}_{U_{n+1} \leq 1/2} - 1.$$

El siguiente código en R simula la evolución de mi fortuna.

```
C<-20 #C es un vector cuya entrada i será mi capital al tiempo i
aux<-C #Esta variable me dice cuál es el último valor de mi capital
while (aux>0) { #Mientras no me haya arruinado
  aux<-aux+2*(runif(1)<1/2)-1 #actualizo mi capital al sumarle una variable
  #que toma valores -1 y 1 con probabilidad 1/2
  C<-c(C,aux) #Agrego el último valor de mi fortuna al vector C
}
plot(C)
```

LISTING 1.1. Ruina.R

En la Figura 1 podemos apreciar un ejemplo del resultado de correr el código anterior.

EJEMPLO 1.2 (Apostando con prisa). Modificaremos el ejemplo anterior como sigue: tengo un capital de 20 pesos al tiempo cero y cada instante de tiempo apuesto en un volado ya sea la mitad de mi fortuna si tengo más de 6 pesos o 2 pesos si mi fortuna es menor o igual a 6, ganando si cae águila.

Un modelo matemático es el siguiente: consideremos variables aleatorias U_1, U_2, \dots uniformes en $(0, 1)$ e independientes y definamos

$$X_0 = 20 \quad \text{y} \quad X_{n+1} = X_n + (2 * \mathbf{1}_{U_{n+1} \leq 1/2} - 1) \begin{cases} \lfloor X_n/2 \rfloor & X_n > 6 \\ 2 & X_n \leq 6 \end{cases}.$$

El modelo anterior se puede implementar fácilmente en R con el código siguiente.

```
C<-20 #C es un vector cuya entrada i será mi capital al tiempo i
fortuna<-C #Esta variable me dice cuál es el último valor de mi capital
while (fortuna>0){ #Mientras no me haya arruinado
  monto<-2*(fortuna<=6)+floor(fortuna/2)*(fortuna>6) #Calculo el monto que
  #apostaré, que es la mitad de mi fortuna cuando tengo más de 6 pesos y
  #si no, dos pesos.
  C<-c(C,monto)
}
```

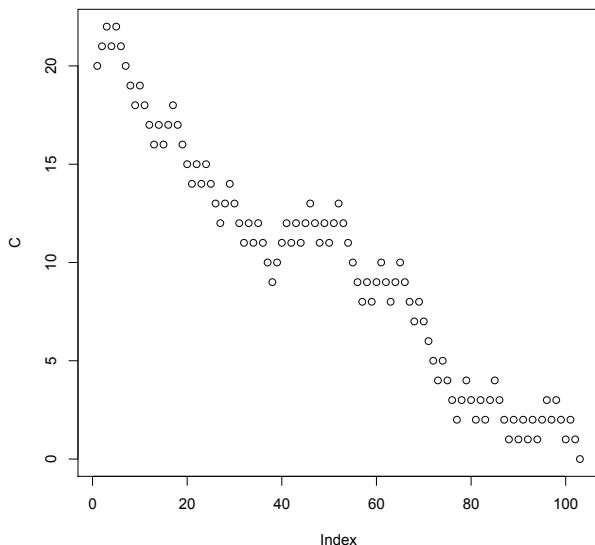


FIGURA 1. Trayectoria de una caminata aleatoria simple que comienza en 20 y es detenida al llegar a cero

```

fortuna<-fortuna+monto*(2*(runif(1)>1/2)-1) #actualizo mi capital al sumarle
una variable que toma valores -1 y 1 con probabilidad 1/2 multiplicada
por el monto de la apuesta
C<-c(C,fortuna)      #Agrego el último valor de mi fortuna al vector C
}
plot(C)

```

LISTING 1.2. Prisa.R

Por supuesto, esperamos que esta estrategia nos lleve más rápido a la ruina. En la Figura 2 podemos apreciar dos ejemplos de trayectorias simuladas de la evolución de la fortuna bajo este esquema de apuestas.

Los dos ejemplos anteriores corresponden a procesos con tiempo y espacio discreto. Ahora analizaremos un modelo a tiempo continuo y espacio discreto.

EJEMPLO 1.3 (Conteos aleatorios). Imaginemos que queremos modelar los tiempos sucesivos en que cambiamos un foco en nuestro lugar de trabajo. Supondremos que en cuanto se funde un foco lo cambiamos (instantáneamente) por uno nuevo.

Es natural asumir que podemos modelar los tiempos de vida de los sucesivos focos mediante una sucesión de variables aleatorias independientes. El supuesto

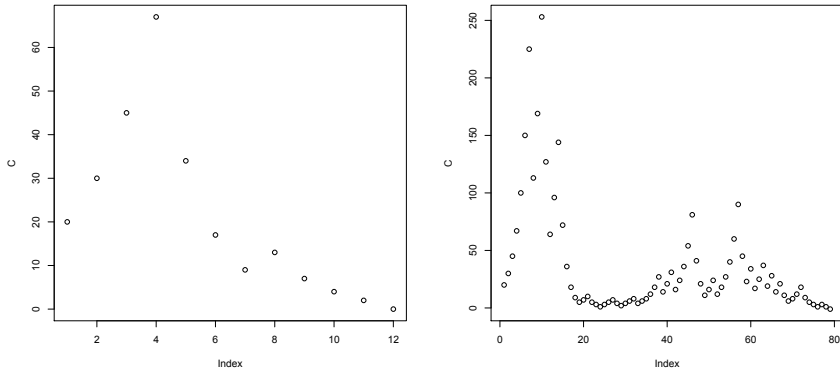


FIGURA 2. Dos trayectorias que muestran la evolución de un capital al someterlo a un esquema de apuestas arriesgadas

adicional que impondremos es que éstas tienen distribución exponencial de tasa $\lambda > 0$, donde la tasa es el recíproco de la media. Sean U_1, U_2, \dots variables independientes de distribución uniforme en $(0, 1)$ y definamos a $T_i = -\log(U_i) / \lambda$. Entonces T_1, T_2, \dots son variables aleatorias exponenciales independientes de tasa λ . La variable T_i la interpretamos como el tiempo de vida del i -ésimo foco.

Puesto que la distribución exponencial está caracterizada por la propiedad de pérdida de memoria

$$\mathbb{P}(S_i > t + s | S_i > t) = \mathbb{P}(S_i > s),$$

el suponer que el tiempo de vida de un foco tiene distribución exponencial puede ser cuestionable puesto que debe haber un efecto de desgaste en su tiempo de vida. Sin embargo, lo que se espera con el modelo es que capture la esencia del fenómeno que queremos modelar.

El proceso estocástico de interés es el que va midiendo la cantidad de focos que hemos cambiado en el intervalo de tiempo $[0, t]$ que mediremos en años. Sean

$$T_0 = 0, \quad T_{n+1} = T_n + S_{n+1} \quad \text{y} \quad N_t = \sum_{i=1}^{\infty} \mathbf{1}_{T_i \leq t}.$$

Entonces se interpreta a T_n como el instante de tiempo en el que cambiamos el n -ésimo foco y a N_t como la cantidad de focos que hemos cambiado en $[0, t]$.

Se puede simular a la función aleatoria $t \mapsto N_t$ en el intervalo $[0, 1]$ mediante el siguiente código.

```
lambda=24*360/1000 # Media, en años, del tiempo de vida de un foco
xi=rexp(1,lambda)   # xi representa el tiempo en el que se cambió el último
                    # foco
```

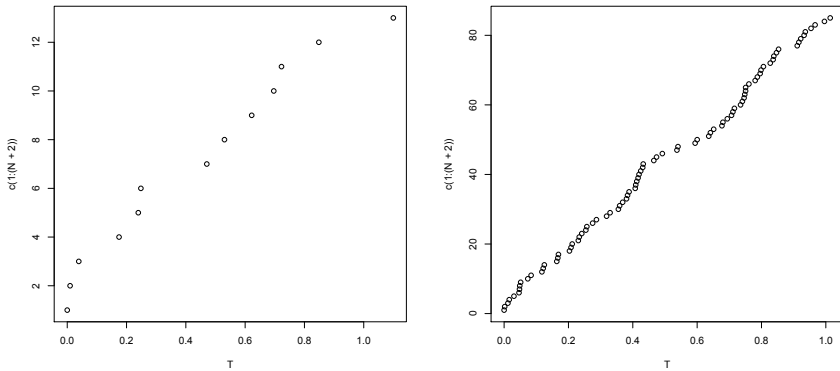


FIGURA 3. Los tiempos sucesivos en los que se cambia un foco a lo largo de 1 y 8 años

```

T=c(0,xi)                # El vector T irá acumulando los tiempos en que vamos
                           cambiando los focos
N=0                       # N nos dirá cuantos focos hemos cambiado al final de un año
while(xi<1){              # Mientras no haya pasado un año
  N<-N+1                  # Aumentamos el número de focos cambiados en uno
  xi<-xi+rexp(1,lambda)   # Vemos el tiempo en el que debemos cambiar el
                           siguiente foco
  T=c(T,xi)               # Aumentamos un evento temporal
}
plot(T,c(1:(N+2)))

```

LISTING 1.3. Poisson.R

En la Figura 3 podemos observar dos trayectorias con los tiempos en los que se cambian los focos, primero en un año y luego en ocho. Podemos observar una cierta regularidad, como si hubiera cierta tendencia determinista (una línea recta) y unas fluctuaciones aleatorias que capturan los sucesivos tiempos de cambio de foco.

EJEMPLO 1.4 (Tiempos de espera). Imaginemos la fila de un banco. Supongamos que los clientes van llegando a tiempos aleatorios y que cada uno requiere un servicio que es también una variable aleatoria. Lo que se quiere medir es: al correr el sistema, si un cliente llega al momento t , ‘¿Cuánto debe esperar para salir del banco?’

Un modelo posible se enfoca en los tiempos entre los arribos de los clientes y supone que éstos son variables aleatorias exponenciales de algún parámetro λ_i . Además, podemos suponer que los tiempos de servicio son variables aleatorias independientes con distribución común, que fijaremos como la exponencial de tasa

λ_s para fijar ideas. Se supone que todas las variables en cuestión son independientes. Este modelo, aunque sea criticable en su supuesto de pérdida de memoria heredado de las variables exponenciales, tiene la particularidad de que se pueden hacer cálculos explícitos que no son posibles en modelos más generales. Además, ejemplifica algunas características de los modelos más generales.

Sean S_1, S_2, \dots variables exponenciales independientes de parámetro λ_i y ξ_1, ξ_2, \dots variables exponenciales independientes (entre si y de las S_i) de parámetro λ_s . Si

$$T_0 = 0, \quad T_{n+1} = T_n + S_{n+1}, \quad N_t = \sum_{n=1}^{\infty} \mathbf{1}_{T_n \leq t}, \quad R_0 = 0 \quad \text{y} \quad R_{n+1} = R_n + \xi_{n+1},$$

definimos entonces los procesos

$$X_t = R_{N_t} - t \quad \text{y} \quad Q_t = X_t - \min_{s \leq t} X_s.$$

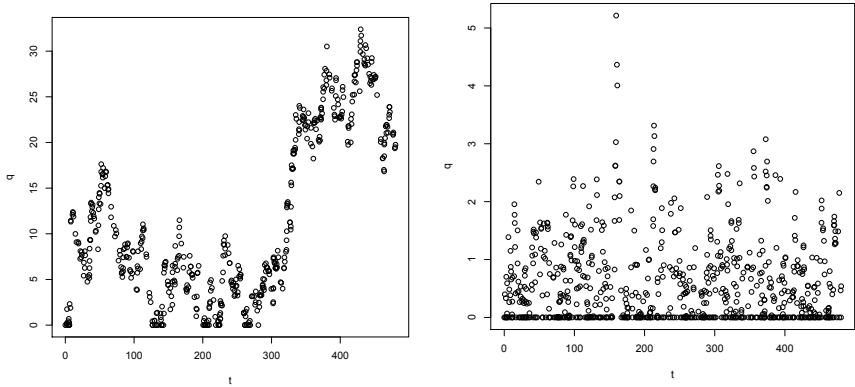
Entonces Q_t representa el tiempo de servicio necesario para atender a los clientes que se encuentran presentes en el banco al tiempo t .

Por otra parte, podemos simular al modelo matemático de la cola mediante el siguiente código en *R*.

```
li<-1           # Tasa interarribo
ls<-2           # Recíproco de la media de servicio
T=8*60         # Tiempo de la simulación
t<-c(0, rexp(1, li)) # Inicialización del vector de eventos temporales
q<-c(0, rexp(1, ls)) # Inicialización del vector de estado de la cola
while(tail(t,1)<T){ # Mientras no haya sobrepasado el umbral temporal
  taux=rexp(1, li) # Me fijo en cuánto falta para la llegada del próximo
                  cliente
  if(taux<tail(q,1)){ # En particular si el próximo cliente llega antes de que
                    la cola se vacíe
    t<-c(t, tail(t,1)+taux) #En cuyo caso agrego el evento de llegada
    q<-c(q, tail(q,1)-taux+rexp(1, ls)) # Junto con el tiempo de
    servicio que requiere menos el que ya he realizado
  }
  else{           # Si el próximo cliente llega después de que la cola se vacíe
    t<-c(t, tail(t,1)+tail(q,1), tail(t,1)+taux) #Agrego dos eventos
    temporales: cuando se vacía la cola y cuando llega el próximo
    cliente
    q<-c(q, 0, rexp(1, ls)) #Agrego además un estado de cola=0 más el
    servicio del próximo cliente que llega
  }
}
plot(t, q)
```

LISTING 1.4. Cola.R

Al ejecutar el código se obtienen gráficos como los de la Figura 4.

FIGURA 4. Estado de la cola cuando $\lambda_i = 1$ y $\lambda_s = 1, 2$

2. La construcción fundamental de una sucesión de variables aleatorias independientes

Como vimos en los ejemplos anteriores, y es cierto en gran generalidad, podemos construir procesos estocásticos muy generales a partir de sucesiones de variables aleatorias independientes. En cierto sentido, dichas sucesiones son los ejemplos más sencillos de procesos estocásticos, en los que no hay realmente una evolución. Al ser, sin embargo, los bloques fundamentales con los que se construyen todos los demás, nos detendremos en su construcción matemática.

2.1. El modelo matemático de una sucesión de volados. Primero se ejemplificará la construcción de una sucesión de variables aleatorias independientes a partir de una sola variable.

Consideremos al espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ en el que $\Omega = (0, 1]$, $\mathcal{F} = \mathcal{B}_{(0,1]}$ y \mathbb{P} es la medida de Lebesgue restringida a Ω . Definamos a las variables aleatorias $d_n : \Omega \rightarrow \mathbb{R}$ como sigue: a cada $\omega \in \Omega$, se le puede asignar su expansión diádica con colas infinitas de tal forma que

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n},$$

donde cada d_n es cero o uno. Aunque la expansión diádica todavía no esté bien definida, puesto que por ejemplo a $1/2$ se le podría asociar ya sea $(1, 0, 0, \dots)$ o $(0, 1, 1, \dots)$, la expansión diádica con colas infinitas, que es la segunda en nuestro ejemplo, sí lo está. Más formalmente, definamos

$$d_1(\omega) = \begin{cases} 0 & \text{si } \omega \in (0, 1/2] \\ 1 & \text{si } \omega \in (1/2, 1] \end{cases}.$$

Notemos que si $\omega_1 = 2\omega - d_1(\omega)$, entonces $\omega_1 \in (0, 1]$; recursivamente, definimos

$$d_{n+1}(\omega) = d_1(\omega_n) \quad \text{y} \quad \omega_{n+1} = 2\omega_n - d_1(\omega_n) \in (0, 1].$$

Es fácil ver que de hecho,

$$\begin{aligned} d_2(\omega) &= \mathbf{1}_{(1/4, 2/4]} + \mathbf{1}_{(3/4, 4/4]}, \\ d_3(\omega) &= \mathbf{1}_{(1/8, 2/8]} + \mathbf{1}_{(3/8, 4/8]} + \mathbf{1}_{(5/8, 6/8]} + \mathbf{1}_{(7/8, 8/8]} \end{aligned}$$

y en general

$$d_n(\omega) = \sum_{i=1}^{2^{n-1}} \mathbf{1}_{((2i-1)/2^n, 2i/2^n]}.$$

Esto implica inmediatamente que si $u_1, \dots, u_n \in \{0, 1\}$ entonces el conjunto

$$\{d_1 = u_1, \dots, d_n = u_n\}$$

es un intervalo de longitud $1/2^n$ y que por lo tanto d_1, \dots, d_n son variables aleatorias independientes de distribución Bernoulli de parámetro $1/2$.

2.2. Una sucesión de variables aleatorias uniformes independientes.

Ahora demostraremos que si $(X_i)_{i \geq 1}$ son variables aleatorias independientes con distribución Bernoulli de parámetro $1/2$ (definidas en otro espacio de probabilidad) entonces $U = \sum_{i \geq 1} X_i/2^i$ tiene distribución uniforme. En efecto, puesto que

$$\mathbb{P}(X_1 = u_1, \dots, X_n = u_n) = 1/2^n = \mathbb{P}(d_1 = u_1, \dots, d_n = u_n),$$

vemos que

$$\mathbb{P}\left(\sum_{i=1}^n X_i/2^i < x\right) = \mathbb{P}\left(\sum_{i=1}^n d_i/2^i < x\right).$$

Por otro lado,

$$U = \lim_n \sum_{i=1}^n X_i/2^i,$$

y de hecho la sucesión de variables aleatorias es creciente. Por lo tanto U es variable aleatoria y

$$\mathbb{P}(U < x) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i=1}^n X_i/2^i < x\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i=1}^n d_i/2^i < x\right) = \mathbb{P}((0, x]) = x.$$

Así, vemos que U es una variable uniforme.

Ahora utilizaremos lo anterior para mostrar que existe un espacio de probabilidad en el que están definidas una sucesión de variables aleatorias uniformes independientes. De hecho el espacio de probabilidad que consideraremos es el mismo $(\Omega, \mathcal{F}, \mathbb{P})$ que en la Subsección 2.1. Como \mathbb{Z}_+ y \mathbb{Z}_+^2 tienen la misma cardinalidad,

consideremos una biyección de $\phi : \mathbb{Z}_+^2 \rightarrow \mathbb{Z}_+$. Definamos $d_i^n = d_{\phi(n,i)}$ y para cada $n \in \mathbb{Z}_+$, sea

$$U_n = \sum_{i \geq 1} \frac{d_i^n}{2^i}.$$

Como $(d_i^n)_{i \geq 1}$ son variables aleatorias independientes de distribución Bernoulli de parámetro $1/2$, se sigue que U_n tiene distribución uniforme para cada $n \in \mathbb{N}$. Se afirma ahora que las variables $(U_n)_{n \geq 1}$ son independientes. En efecto, esto es consecuencia del siguiente lema, un tanto más general. Notemos que U_n es medible respecto de la σ -álgebra generada por $(d_i^n)_{i \geq 1}$, a la cual llamaremos \mathcal{F}_n .

LEMA 1. Sean $\mathcal{F}_{n,i}, i \geq 1, n \geq 1$ σ -álgebras independientes y definamos

$$\mathcal{F}_n = \sigma(\mathcal{F}_{n,1}, \mathcal{F}_{n,2}, \dots).$$

Entonces $\mathcal{F}_n, n \geq 1$ son σ -álgebras independientes.

DEMOSTRACIÓN. Debemos mostrar que para todo $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$, se tiene que

$$(1) \quad \mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).$$

Sea

$$\mathcal{C}_n = \{A_1 \cap \dots \cap A_m : m \geq 1 \text{ y } A_j \in \cup_{i \geq 1} \mathcal{F}_{n,i} \text{ para } j = 1, \dots, m\}.$$

Puesto que

$$\bigcup_{i \geq 1} \mathcal{F}_{n,i} \subset \mathcal{C}_n \subset \mathcal{F}_n,$$

vemos que

$$\sigma(\mathcal{C}_n) = \mathcal{F}_n.$$

Por otra parte, es fácil ver que \mathcal{C}_n es un π -sistema.

Consideremos ahora la clase

$$\mathcal{M}_1 = \{A \in \mathcal{F}_k : \mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) \text{ si } B = B_1 \cap \dots \cap B_n \text{ con } B_j \in \mathcal{C}_j\}.$$

Es fácil ver que \mathcal{M}_k es un λ -sistema que contiene, por hipótesis a \mathcal{C}_1 . Por lo tanto

$$\mathcal{M}_1 = \mathcal{F}_1.$$

Ahora consideramos a

$$\begin{aligned} \mathcal{M}_2 = \{A \in \mathcal{F}_2 : \mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) \\ \text{si } B = B_1 \cap B_3 \cap \dots \cap B_n \text{ con } B_1 \in \mathcal{F}_1 \text{ y } B_j \in \mathcal{C}_j \text{ para } j \geq 3\}. \end{aligned}$$

Se prueba entonces que \mathcal{M}_2 es un λ -sistema que por hipótesis y la igualdad $\mathcal{M}_1 = \mathcal{F}_1$ contiene a \mathcal{C}_2 . Al aplicar este razonamiento sucesivamente, obtenemos la igualdad (1). \square

2.3. Una sucesión de variables aleatorias independientes con distribuciones arbitrarias. Ahora utilizaremos la construcción de la sucesión de variables aleatorias uniformes independientes para demostrar el siguiente resultado:

TEOREMA 1.1. *Sean μ_n , $n \geq 1$ medidas de probabilidad en \mathbb{R} . Entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y una sucesión de variables aleatorias independientes $X_n : \Omega \rightarrow \mathbb{R}$, $n \geq 1$ tales que la distribución de X_n es μ_n .*

La herramienta principal de la construcción será el siguiente lema: (tomado de Billingsley p. 190). Recordemos que una función $F : \mathbb{R} \rightarrow [0, 1]$ es la función de distribución de una variable aleatoria real si y sólo si es no decreciente, continua por la derecha (y con límites por la izquierda) tal que $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.

DEFINICIÓN. La **función de cuantiles** de una función de distribución F es la función $\phi : (0, 1) \rightarrow \mathbb{R}$ dada por

$$\phi(u) = \inf \{x \in \mathbb{R} : u \leq F(x)\}.$$

La función de cuantiles satisface la igualdad

$$\phi(u) \leq x \Leftrightarrow u \leq F(x)$$

que se demostrará posteriormente. De esta igualdad se deduce la medibilidad de ϕ .

PRUEBA DEL TEOREMA 1.1. Sabemos que existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ en el que existe una sucesión de variables aleatorias independientes $(U_n)_{n \geq 1}$ uniformes en $(0, 1)$. Sea F_n la función de distribución asociada a la medida de probabilidad μ_n y ϕ_n la función de cuantiles de \mathcal{F}_n . Como ϕ_n es una función medible, $X_n = \phi_n(U_n)$ es una variable aleatoria. Además, como las variables $U_n, \neq 1$ son independientes, también lo son las variables $X_n, n \geq 1$:

$$\{X_1 \in A_1, \dots, X_n \in A_n\} = \{U_1 \in \phi_1^{-1}(A_1), \dots, U_n \in \phi_n^{-1}(A_n)\}.$$

Finalmente:

$$\mathbb{P}(X_i^{-1}((-\infty, x])) = \mathbb{P}(U_i^{-1}(\phi_i^{-1}((-\infty, x]))) = \mathbb{P}(U_i^{-1}((0, F_i(x)])) = F_i(x),$$

por lo que X_i tiene distribución μ_i . □

Ahora demostremos las propiedades de la función de cuantiles ϕ asociada a la función de distribución F . Sea $u \in (0, 1)$; entonces el conjunto $\{x \in \mathbb{R} : u \leq F(x)\}$ es no vacío y como F es no decreciente, es un intervalo ya sea de la forma $[\phi(u), \infty)$ o $(\phi(u), \infty)$, ya que $\phi(u)$ es el ínfimo del conjunto considerado. La segunda opción se descarta al notar que F es continua por la derecha. Por lo tanto, $u \leq F(x)$ si y sólo si $\phi(u) \leq x$.

CAPÍTULO 2

Cadenas de Markov a tiempo discreto

La Figura 1 representa un laberinto. Imaginemos que colocamos una rata en la esquina inferior izquierda y un plato de comida en la esquina superior derecha. Para modelar la trayectoria que sigue la rata hasta encontrar la comida, supongamos que cuando se encuentra en un cuarto del laberinto, la rata va a cualquier otro con la misma probabilidad. Un modelo matemático para esta situación es el siguiente. Enumeremos los cuartos del laberinto de izquierda a derecha, de abajo a arriba, por lo que la rata comienza en el cuarto 1 y encuentra la comida en el cuarto 9. Definamos $P_{i,j}$ como la probabilidad con que la rata pasa del cuarto i al cuarto j ; por ejemplo, vemos que

$$P_{5,j} = \begin{cases} 1/4 & j \in \{2, 4, 6, 8\} \\ 0 & j \notin \{2, 4, 6, 8\} \end{cases}.$$

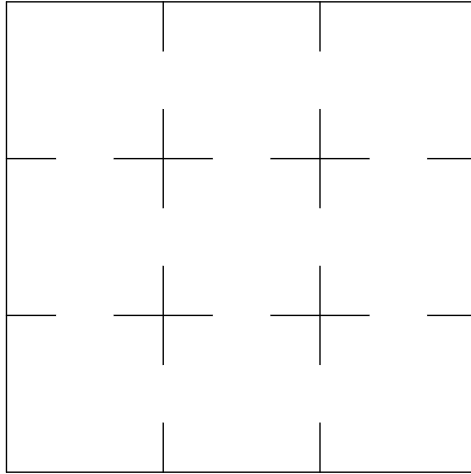


FIGURA 1. Laberinto para un experimento aleatorio

Esta información se puede organizar de forma matricial de la siguiente manera:

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

Entonces la probabilidad de que la rata siga la trayectoria 1, 4, 7, 8, 9 para encontrar la comida es $P_{1,4}P_{4,7}P_{7,8}P_{8,9}$. Notemos que

- (1) $P_{i,j} \geq 0$ para todo i y j y
- (2) $\sum_j P_{i,j} = 1$ para toda i .

A una matriz con estas dos características se le llama **matriz estocástica**. La segunda condición nos dice que $P_{i,1}, \dots, P_{i,9}$ es una distribución de probabilidad sobre el conjunto $\{1, \dots, 9\}$. Si definimos a ϕ_i como la función de cuantiles asociada, tendremos que $\phi_i(U_j)$ es una variable aleatoria con la misma distribución que el cuarto al que pasa la rata si está en el cuarto j . Es por esto que si definimos

$$X_0 = 1 \quad \text{y} \quad X_{n+1} = \phi_{X_n}(U_{n+1}),$$

las variables X_0, X_1, \dots nos modelan el movimiento de la rata por el laberinto. Para poder obtener la trayectoria de la rata detenida hasta que encuentre la comida, podríamos modificar la matriz P en

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

que difiere de P salvo en el último renglón, en el cual especificamos que una vez que la rata llegue al cuarto 9 se quede ahí. Si $\tilde{\phi}_i$ son las funciones de cuantiles asociadas a los renglones de \tilde{P} , podemos entonces modelar la trayectoria de la rata, detenida cuando alcanza la comida mediante la sucesión

$$\tilde{X}_0 = 1 \quad \text{y} \quad \tilde{X}_{n+1} = \tilde{\phi}_{\tilde{X}_n}(U_{n+1}).$$

Se presenta a continuación un código en R para simular la trayectoria de la rata.

```
P=matrix(c(0,1/2,0,1/2,0,0,0,0,0,1/3,0,1/3,0,1/3,0,0,0,0,1/2,0,0,0,1/2,0,0,0,1/3,0,0,0,1/3,0,1/3,0,0,0,1/4,0,1/4,0,1/4,0,1/4,0,0,0,1/3,0,1/3,0,0,0,1/3,0,0,0,1/2,0,0,0,1/2,0,0,0,1/2,0,0,0,0,0,1/3,0,1/3,0,1/3,0,0,0,0,1/2,0,1/2,0),9)
# Genera la matriz de transición para la rata en un laberinto

X<-1 # El vector X acumulará la trayectoria que sigue la rata; comienza
en el cuarto 1.
N<-0 # Paso en el que vamos
while(tail(X,1)!=9){ #Mientras la rata no encuentre la comida del cuarto 9
  X<-c(X,sample(c(1:9),1,prob=P[tail(X,1),])) # Escogemos un cuarto al azar a
  partir del que se encuentra
  N<-N+1 # Especificamos que se ha dado un paso más
}
```

LISTING 2.1. Rata.R

Como un ejemplo, se obtuvieron las siguientes dos trayectorias simuladas

- 1 2 3 2 1 2 1 2 1 4 1 4 1 4 1 2 3 6 5 6 5 2 3 2 3 6 3 2 3 6 3 2 1 2 1 2 3 6 9
- 1 4 5 6 3 6 9

A continuación presentamos una serie de preguntas para las cuales la teoría subsecuente encuentra una respuesta.

- ¿Cuánto tarda la rata en promedio en encontrar la comida si comienza en el cuarto i ?
- Si quitamos la comida y nada más seguimos la trayectoria de la rata, ¿Cuál es la probabilidad de que se encuentre en el cuarto j en el paso n si comienza en i ? Parte de la teoría que veremos nos dice que la probabilidad se estabiliza conforme $n \rightarrow \infty$.
- Si de nuevo seguimos solamente la trayectoria sin comida, ¿estamos seguros de regresar al punto inicial?
- Si agregamos la comida, ¿Cuántas veces regresará la rata al cuarto inicial antes de encontrar la comida?

A continuación daremos un marco teórico que permite generalizar al modelo anterior. Se trata de las cadenas de Markov cuyo estudio abarcará este capítulo.

Sea E un conjunto a lo más numerable al que llamaremos **espacio de estados**. Consideremos a una colección numérica $P = (P_{x,y})_{x,y \in E}$ a la que pensaremos como una matriz indexada por E . Supongamos que

- (1) $P_{x,y} \geq 0$ para todo x y y y
- (2) $\sum_y P_{x,y} = 1$ para toda x .

A P le llamamos matriz estocástica. Consideremos también una distribución de probabilidad π sobre E , que podemos pensar como un vector (digamos rengón) $\pi = (\pi_x)_{x \in E}$.

DEFINICIÓN. Una **cadena de Markov** con matriz de transición P y distribución inicial π es un proceso estocástico $(X_n)_{n \in \mathbb{N}}$ con valores en E tal que si

$x_0, \dots, x_1 \in E$ entonces

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \pi_{x_0} P_{x_0, x_1} \cdots P_{x_{n-1}, x_n}.$$

TEOREMA 2.1. *Dada una matriz de transición P y una distribución inicial π existe un espacio de probabilidad en el que están definidas una sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ definidas en él que conforman una cadena de Markov con matriz de transición P y distribución inicial π .*

La demostración del teorema es importante pues nos provee de un algoritmo de simulación para cadenas de Markov. Representa otra ilustración del hecho de que cualquier proceso estocástico se puede construir mediante variables uniformes independientes.

DEMOSTRACIÓN. Al enumerar a los elementos de E , podemos pensar que $E = \{0, \dots, n\}$ ó $E = \mathbb{N}$. Sea ϕ_i la función de cuantiles asociada al renglón i de P , ϕ la función de cuantiles de π y sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad en el que están definidas una sucesión $(U_i)_{i \in \mathbb{N}}$ de variables uniformes independientes. Definimos a

$$X_0 = \phi(U_0) \quad \text{y} \quad X_{n+1} = \phi_{X_n}(U_{n+1}).$$

Por definición de función de cuantiles:

$$\mathbb{P}(\phi_x(U_j) = y) = P_{x,y},$$

por lo que se sigue que si $x_0, \dots, x_n \in E$ entonces

$$\begin{aligned} \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}(\phi(U_0) = x_0, \phi_{x_0}(U_1) = x_1, \dots, \phi_{x_{n-1}}(U_n) = x_n) \\ &= \mathbb{P}(\phi(U_0) = x_0) \prod_{i=1}^n \mathbb{P}(\phi_{x_{i-1}}(U_i) = x_i) \\ &= \pi_{x_0} P_{x_0, x_1} \cdots P_{x_{n-1}, x_n}. \end{aligned}$$

□

Una de las características principales de las cadenas de Markov es la propiedad de Markov:

PROPOSICIÓN 2.1 (Propiedad de Markov). *Sea X una cadena de Markov de distribución inicial π y transición P . Si $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$ entonces*

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

La interpretación es que la evolución futura de la cadena sólo depende del pasado a través del presente.

DEMOSTRACIÓN. Calculemos el lado izquierdo:

$$\mathbb{P}(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) = \pi_{x_0} P_{x_0, x_1} \cdots P_{x_n, x_{n+1}}$$

por lo cual

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P_{x_n, x_{n+1}}.$$

Por otra parte, puesto que

$$\{X_n = x_n\} = \bigcup_{x_0, \dots, x_{n-1}} \{X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n\}$$

donde la unión es disjunta, se sigue que

$$\mathbb{P}(X_n = x_n) = \sum_{x_0, \dots, x_{n-1}} \pi_{x_0} P_{x_0, x_1} \cdots P_{x_{n-1}, x_n}$$

y que

$$\mathbb{P}(X_n = x_n, X_{n+1} = x_{n+1}) = \sum_{x_0, \dots, x_n} \pi_{x_0} P_{x_0, x_1} \cdots P_{x_{n-1}, x_n} P_{x_n, x_{n+1}}$$

por lo que

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) = P_{x_n, x_{n+1}}.$$

□

Esta misma técnica de descomposición del espacio de estados nos lleva a lo que se conoce como las ecuaciones de Chapman-Kolmogorov.

PROPOSICIÓN 2.2. *Sea X una cadena de Markov de distribución inicial π y transición P . Si $P_{x,y}^n = \mathbb{P}(X_n = y | X_0 = x)$ entonces*

$$P_{x,z}^{n+m} = \sum_{y \in E} P_{x,y}^m P_{y,z}^n.$$

La ecuación anterior recuerda mucho a la de multiplicación de matrices. En efecto, lo que nos dice es que $P_{x,y}^n$ es la n -ésima potencia de la matriz de transición P .

DEMOSTRACIÓN. Al generalizar la idea de la prueba anterior, vemos que si definimos $x_0 = x$ y $x_{n+m} = z$ entonces

$$\begin{aligned} & P_{x,z}^{n+m} \\ & \mathbb{P}(X_{n+m} = z | X_0 = x) \\ &= \sum_{x_1, \dots, x_{n+m-1} \in E} P_{x_0, x_1} \cdots P_{x_{n+m-1}, x_{n+m}} \\ &= \sum_{y \in E} \sum_{x_1, \dots, x_{m-1} \in E} P_{x_0, x_1} \cdots P_{x_{m-1}, y} \sum_{x_{m+1}, \dots, x_{n+m-1} \in E} P_{y, x_{m+1}} \cdots P_{x_{n+m-1}, z} \\ &= \sum_{y \in E} P_{x,y}^m P_{y,z}^n. \end{aligned}$$

□

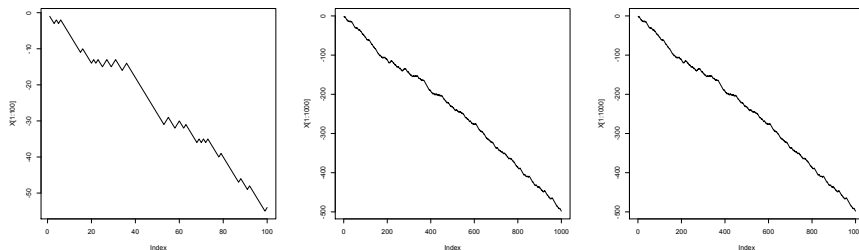


FIGURA 2. 100, 1000 y 10000 pasos de una caminata aleatoria simple con $p = 1/4$

Por supuesto, en general no es posible calcular explícitamente las potencias de la matriz de transición. Sin embargo, un paquete como R es capaz de realizar este producto de manera numérica y así poder resolver problemas de orden práctico que se modelen mediante cadenas de Markov. A continuación, se presentan algunos ejemplos de cadenas de Markov.

EJEMPLO 2.1. La caminata aleatoria simple es una cadena de Markov cuyo espacio de estados es \mathbb{Z} y es tal que $P_{i,i+1} = 1 - P_{i,i-1} = p$ para alguna $p \in (0, 1)$. Este es uno de los ejemplos introductorios. Basta entonces mencionar que se puede simular una trayectoria de longitud fija n (de hecho 2 trayectorias) mediante el siguiente código.

```
# Código para simular una trayectoria de n pasos de una caminata aleatoria
# simple de parámetro p
p<-1/2
n<-10000
U<-runif(n)
Y<-2*(U<p)-1
X<-cumsum(Y)
plot(X[1:100],type="l")
quartz() #usar x11() en UNIX y windows() en Windows, esto es para mac.
plot(X[1:1000],type="l")
quartz()
plot(X[1:10000],type="l")
```

LISTING 2.2. CAS1.R

Se pueden obtener entonces trayectorias como las de las Figuras 2 y 3 en las que se examinan trayectorias de 100, 1000 y 10000 pasos respectivamente para los parámetros $1/4$ y $1/2$. En la primera se aprecia la ley fuerte de los grandes números.

Por otra parte, se pueden calcular numéricamente las probabilidades de transición a n pasos mediante el código:

```
pa<-0.5 #Probabilidad de ir de i a i+1
n<-6 #Cantidad de pasos que daremos
```

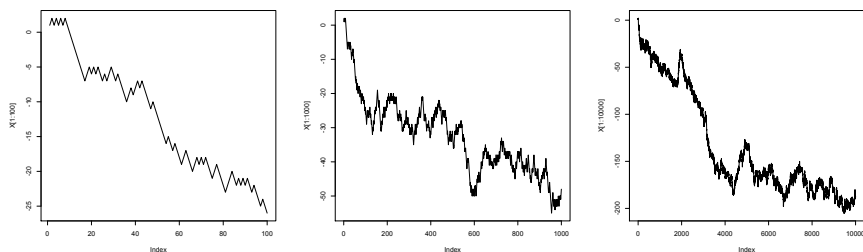


FIGURA 3. 100, 1000 y 10000 pasos de una caminata aleatoria simple con $p = 1/2$

```
p<-matrix(0,n+1,2*n+1) #La entrada P[i,j] nos da la probabilidad de
                          #encontrarnos en j-i al paso i-1
p[1,1]<-1 #Iniciación: comenzamos en 0 al paso 0 con probabilidad 1
for(i in 1:n){ #Con cada paso actualizamos nuestras probabilidades
  p[i+1,]=(1-pa)*p[i,]+pa*c(0,0,p[i,1:(2*n-1)])
}
```

LISTING 2.3. CASnPasos.R

Podemos entonces obtener la matriz P tal que $P_{i,j}$ nos da la probabilidad de que una caminata aleatoria simple esté en el estado $j-i$ al paso $i-1$. Para que cupiera en la página sólo se corrió con $n = 6$, pero computacionalmente $n = 1000$ no representa ningún problema. Una vez almacenados estos datos se pueden utilizar para obtener numericamente la media, varianza, o la esperanza de alguna otra función de la variable aleatoria que nos mide la posición después de n pasos.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.25	0.00	0.50	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.12	0.00	0.38	0.00	0.38	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00
5	0.06	0.00	0.25	0.00	0.38	0.00	0.25	0.00	0.06	0.00	0.00	0.00	0.00
6	0.03	0.00	0.16	0.00	0.31	0.00	0.31	0.00	0.16	0.00	0.03	0.00	0.00
7	0.02	0.00	0.09	0.00	0.23	0.00	0.31	0.00	0.23	0.00	0.09	0.00	0.02

EJEMPLO 2.2 (Cadena de nacimiento y muerte). Se trata de una cadena de Markov cuyo espacio de estados es $E = \{0, \dots, n\}$ ó $\mathbb{N} = \{0, 1, 2, \dots\}$ con probabilidades de transición son $P_{i,i+1} = p(i)$ y $P_{i,i-1} = q(i)$ donde $1 - q(i) = p(i) \in [0, 1]$. (Definimos $q(0) = 0$ y si $E = \{0, \dots, n\}$ entonces $p(n) = 0$.)

EJEMPLO 2.3 (Cadena de Ehrenfest). En este ejemplo hay dos urnas, con bolas numeradas del 1 al n repartidas entre ambas. A cada instante de tiempo se escoge

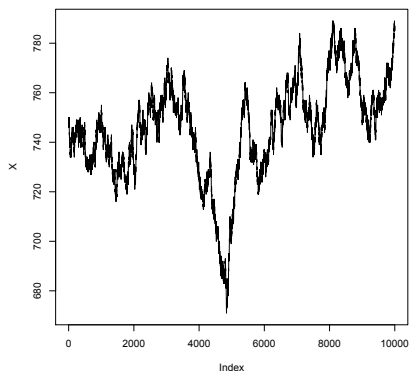


FIGURA 4. Trayectoria simulada de 1000 pasos de la cadena de Ehrenfest con $n = 1000$

un número al azar entre 1 y n y la bola con ese número se cambia de urna. Lo que se mide es la cantidad de bolas en la urna 1 (digamos). Esta será una cadena de Markov con espacio de estados $E = \{0, \dots, n\}$ y matriz de transición P dada por

$$P_{0,1} = 1 = P_{n,n-1}, \quad P_{i,i+1} = 1 - i/n \text{ si } i < n \quad \text{y} \quad P_{i,i-1} = i/n \text{ si } i > 0.$$

Este es un caso particular de la cadena de nacimiento y muerte con espacio de estados finito. Se puede simular la cadena mediante un código como el siguiente:

```
# Código para simular una trayectoria de m pasos de una cadena de Ehrenfest con
# espacio de estados {0,...,n}.
n<-1000
m<-1000
U<-runif(m)
X<-n/2
for(i in 2:m){
  aux<-tail(X,1)
  if(aux==0){X<-c(X,1)}
  else if (aux==n){X<-c(X,n-1)}
  else {X<-c(X,aux+1-2*(U[i]<aux/n))}
}
plot(X,type="l")
```

LISTING 2.4. Ehrenfest.R

Con él, se obtuvo la Figura 4.

Como ejemplo final, el lector puede verificar la liga <http://www.r-bloggers.com/basics-on-markov-chain-for-parents/> a un blog en el que se interpreta al juego de serpientes y escaleras en términos de cadenas de Markov con código en R para simular el desarrollo del juego.

1. Clases de comunicación

Sean P una matriz de transición sobre E y X una cadena de Markov con matriz de transición \mathbb{P} y distribución inicial ν tal que $\nu_x > 0$ para toda $x \in E$. Denotaremos por \mathbb{P}_x a \mathbb{P} condicionada por $X_0 = x$. Puesto que

$$\mathbb{P}_x(X_n = y) = \sum_{i_1, \dots, i_{n-1} \in E} \mathbb{P}_x(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = y),$$

vemos que

$$\mathbb{P}_x(X_n = y) = \sum_{i_1, \dots, i_n} P_{x, x_1} \cdots P_{x_{n-1}, y}.$$

Por lo tanto, si se introducen a las potencias de la matriz de transición $P^n, n \geq 1$ (y se define $P^0_{x,y} = \delta_{x,y}$) vemos que

$$\mathbb{P}_x(X_n = y) = P^n_{x,y}.$$

Sean x y y dos estados de E . Diremos que x conduce a y si existe $n \geq 0$ tal que $P^n_{x,y} > 0$. Claramente esto ocurre si y sólo si existen x_0, \dots, x_n con $x_0 = x$ y $x_n = y$ tales que $P_{x_{k-1}, x_k} > 0$. Cuando x conduce a y y y conduce a x , diremos que x y y se comunican y lo denotaremos mediante $x \sim y$.

PROPOSICIÓN 2.3. *La relación $x \sim y$ es una relación de equivalencia en E .*

A las clases de equivalencia inducidas por la relación \sim les llamaremos clases de comunicación.

DEMOSTRACIÓN.

Reflexividad: Puesto que $P^0_{x,x} = 1$, vemos que $x \sim x$.

Simetría: Por definición $x \sim y$ si y sólo si $x \sim y$.

Transitividad: Si $x \sim y$ y $y \sim z$, sean m y n en \mathbb{N} tales que $P^m_{x,y} > 0$ y $P^n_{y,z} > 0$. Puesto que

$$P^{n+m}_{x,z} \geq P^n_{x,y} P^m_{y,z} > 0,$$

vemos que x conduce a z y un argumento análogo muestra que entonces $x \sim z$.

□

Se dice que una cadena de Markov es **irreducible** si tiene una sola clase de comunicación. A la clase de comunicación a la que pertenece el estado $x \in E$ la denotamos por C_x ; explícitamente:

$$C_x = \{y \in E : x \sim y\}.$$

El concepto de clase de comunicación nos permite dar una primera descomposición del espacio de estados. Ésta se puede refinar al introducir el concepto

de clase de comunicación abierta y cerrada. Este concepto es útil pues se puede reducir el espacio de estados de una cadena de Markov a una clase de comunicación cerrada.

DEFINICIÓN. Sea C un subconjunto del espacio de estados E . Decimos que C es un **conjunto cerrado** si para toda $y \in E \setminus C$, x no conduce a y . Un **conjunto abierto** es aquel que no es cerrado.

2. La propiedad de Markov fuerte

La propiedad de Markov fuerte es una extensión de la propiedad de Markov a ciertos tiempos aleatorios. Es una herramienta de gran utilidad. En particular nos servirá para estudiar los conceptos de transitoriedad y recurrencia. Antes de pasar a la propiedad de Markov fuerte, veamos la siguiente extensión de la propiedad de Markov.

PROPOSICIÓN 2.4. *Sea A cualquier subconjunto de E^n tal que $\mathbb{P}(A \cap \{X_n = y\}) > 0$. Entonces, condicionalmente a $A \cap \{X_n = y\}$, el proceso $(X_{n+m}, m \geq 0)$ es una cadena de Markov que comienza en y y tiene matriz de transición P .*

DEMOSTRACIÓN. Al descomponer al conjunto A como unión de eventos elementales de la forma $\{(x_0, \dots, x_{n-1})\}$, vemos que

$$\begin{aligned} & \mathbb{P}(A, X_n = y, X_{n+1} = y_1, \dots, X_{n+m} = y_m) \\ &= \sum_{(x_0, \dots, x_{n-1}) \in A} \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = y, X_{n+1} = y_1, \dots, X_{n+m} = y_m) \\ &= \sum_{(x_0, \dots, x_{n-1}) \in A} P_{x_1, x_1} \cdots P_{x_{n-1}, y} P_{y, y_1} \cdots P_{y_{m-1}, y_m} \end{aligned}$$

Así, se obtiene

$$\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid A, X_n = y) = P_{x_{n-1}, y} P_{y, y_1} \cdots P_{y_{m-1}, y_m}.$$

□

Ahora verificaremos que la propiedad de Markov se extiende a ciertos tiempos aleatorios. Un **tiempo aleatorio** es una variable aleatoria $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$. Dicho tiempo aleatorio es finito si $T(\omega) \in \mathbb{N}$ para toda $\omega \in \Omega$ y es acotado si existe $K \in \mathbb{N}$ tal que $T(\omega) \leq K$ para todo $\omega \in \Omega$.

DEFINICIÓN. Un tiempo aleatorio T es un **tiempo de paro** si para toda $n \in \mathbb{N}$ existe $A_n \subset E^{n+1}$ tal que

$$\{T = n\} = \{(X_0, \dots, X_n) \in A_n\}.$$

Intuitivamente un tiempo de paro es un tiempo que obtenemos de observar la trayectoria hasta que se cumpla una condición. El instante en que se cumple es el tiempo de paro.

Nuestro primer ejemplo de un tiempo de paro es el tiempo T_1 en que una cadena X regresa a su estado inicial. En otras palabras:

$$T_1 = \begin{cases} \infty & X_n \neq X_0 \text{ para toda } n \\ \min \{n \geq 1 : X_n = X_0\} & \text{en otro caso} \end{cases}.$$

En efecto es un tiempo de paro puesto que

$$\{T_1 = 1\} = \{X_1 = X_0\}$$

y para $n \geq 2$

$$\{T_1 = n\} = \{X_1 \neq X_0, \dots, X_{n-1} \neq X_0, X_n = X_0\}.$$

De igual manera, el tiempo T_n en que ocurre la n -ésima visita al estado inicial es un tiempo de paro. Esto se prueba por inducción al notar que ya hemos verificado la base inductiva $n = 1$ y por otro lado

$$\{T_{n+1} = m\} = \bigcup_{l < m} \{T_n = l\} \cap \{X_{l+1} \neq X_0, \dots, X_{m-1} \neq X_0, X_m = X_0\}.$$

Otro ejemplo de un tiempo de paro es la primera vez H_A en que la cadena accede a un subconjunto A del espacio de estados. En otras palabras:

$$H_A = \begin{cases} \infty & X_n \in E \setminus A \text{ para toda } n \\ \min \{n \geq 0 : X_n \in A\} & \text{en otro caso} \end{cases}.$$

TEOREMA 2.2 (Propiedad de Markov fuerte). *Sea A cualquier subconjunto de E^{n+1} tal que $\mathbb{P}(A, X_n = y, T = n) > 0$. Entonces, condicionalmente a $A \cap \{T = n, X_n = y\}$, el proceso $(X_{n+m}, m \geq 0)$ es una cadena de Markov que comienza en y y tiene matriz de transición P .*

DEMOSTRACIÓN. Sea $A_n \subset E^{n+1}$ tal que

$$\{T = n\} = \{(X_0, \dots, X_n) \in A_n\}.$$

Al descomponer al conjunto A como unión de eventos elementales de la forma $\{(x_0, \dots, x_{n-1})\}$, vemos que

$$\begin{aligned} & \mathbb{P}(A, T = n, X_n = y, X_{n+1} = y_1, \dots, X_{n+m} = y_m) \\ &= \sum_{(x_0, \dots, x_n) \in A \cap A_n} \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = y, X_{n+1} = y_1, \dots, X_{n+m} = y_m) \\ &= \sum_{(x_0, \dots, x_n) \in A \cap A_n} P_{x_1, x_1} \cdots P_{x_{n-1}, y} P_{y, y_1} \cdots P_{y_{m-1}, y_m} \end{aligned}$$

Así, se obtiene

$$\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid A, T = n, X_n = y) = P_{x_{n-1}, y} P_{y, y_1} \cdots P_{y_{m-1}, y_m}.$$

□

3. Transitoriedad y recurrencia

Pasaremos ahora al análisis de dos conceptos que permiten hacer una distinción entre los estados de una cadena de Markov, de acuerdo a si siempre serán revisitados o no.

Sea $x \in E$. Definamos a la cantidad de visitas al estado x como la variable aleatoria

$$V_x = \sum_{n=0}^{\infty} \mathbf{1}_{X_n=x}.$$

Esta variable aleatoria podría tomar el valor infinito. Sin embargo, un resultado curioso es que si toma el valor infinito con probabilidad positiva, entonces toma el valor infinito con probabilidad 1. En caso de que V_x sea infinita con probabilidad 1 bajo \mathbb{P}_x hablamos de un **estado recurrente** y en caso contrario de un **estado transitorio**. Analicemos ahora por qué el conjunto $\{V_x = \infty\}$ tiene probabilidad cero o uno. Para esto, definamos a T_0, T_1, \dots como los instantes sucesivos que X visita al estado x . Bajo la medida \mathbb{P}_x ,

$$T_0 = 0, \quad T_1 = \min \{n > 0 : X_n = x\}$$

y T_{n+1} es la primera vez que la cadena de Markov $(X_{T_n+m}, m \geq 0)$ regresa a x . Hemos visto que cada T_n es un tiempo de paro.

Notemos que

$$V_x = \sum_{n=1}^{\infty} \mathbf{1}_{T_n < \infty}.$$

Se afirma ahora que bajo \mathbb{P}_x , V_x es una variable aleatoria geométrica de parámetro $\mathbb{P}_x(T_1 < \infty)$. En efecto, por una parte se tiene que

$$\{V_x \geq n\} = \{T_n < \infty\}$$

y por otra, la propiedad de Markov fuerte nos permite afirmar que para cada $n \geq 1$:

$$\mathbb{P}_x(T_{n+1} < \infty) = \mathbb{P}_x(T_{n+1} < \infty, T_n < \infty) = \mathbb{E}_x(\mathbf{1}_{T_n < \infty} \mathbb{P}_x(T_1 < \infty)),$$

por lo cual

$$\mathbb{P}_x(T_n < \infty) = \mathbb{P}_x(T_1 < \infty)^n.$$

El caso en que $\mathbb{P}_x(T_1 < \infty) = 1$ ocurre si y sólo si V_x es infinita \mathbb{P}_x casi seguramente. Si no, V_x es geométrica de parámetro $\mathbb{P}_x(T_1 < \infty)$ y por lo tanto su esperanza es finita. Esto nos proporciona una equivalencia, en términos de la matriz de transición, para que un estado sea recurrente.

PROPOSICIÓN 2.5. *El estado x es recurrente si y sólo si $\sum_x P_{x,x}^n = \infty$.*

DEMOSTRACIÓN. La afirmación se sigue de notar que

$$\mathbb{E}_x(V_x) = \sum_n \mathbb{E}_x(\mathbf{1}_{X_n=x}) = \sum_n P_{x,x}^n.$$

□

Ahora veremos que la transitoriedad o recurrencia es de hecho una propiedad de clase.

PROPOSICIÓN 2.6. *Si x y y se comunican entre sí e x es transitorio entonces y es transitorio.*

DEMOSTRACIÓN. Sean m y n tales que $P_{x,y}^m > 0$ y $P_{x,y}^n > 0$. Entonces

$$P_{x,x}^{m+l+n} \geq P_{x,y}^m P_{y,y}^l P_{y,x}^n.$$

Por lo tanto:

$$\text{si } \sum_n P_{x,x}^{m+l+n} < \infty \text{ entonces } \sum_n P_{y,y}^l < \infty.$$

□

La conclusión que obtenemos es que en una clase o todos los estados son recurrentes o todos son transitorios y que por lo tanto podemos hablar de **clases recurrentes** y de **clases transitorias**. Hay una forma fácil de saber si una clase es transitoria.

PROPOSICIÓN 2.7. *Sea $C \subset E$ una clase abierta. Entonces C es transitoria.*

DEMOSTRACIÓN. En efecto, puesto que C es una clase abierta, existe $x \in C$, $y \in E \setminus C$ y $m \geq 0$ tal que $P_{x,y}^m > 0$ mientras que $P_{y,x}^n = 0$ para toda $n \geq 0$. Por lo tanto

$$\mathbb{E}_y(V_x) = \sum_{n=0}^{\infty} \mathbb{E}_y(\mathbf{1}_{X_n=x}) = \sum_{n=0}^{\infty} P_{y,x}^n = 0$$

y puesto que V_x es una variable aleatoria no-negativa, entonces

$$\mathbb{P}_y(V_x = 0) = 1.$$

Así, vemos que

$$\mathbb{P}_x(V_x < \infty) \geq \mathbb{P}_x(V_x(X_m, X_{m+1}, \dots) = 0, X_m = y) = P_{x,y}^m > 0$$

por lo que x es transitorio. □

Veremos ahora que las conclusiones anteriores nos permiten clasificar a las clases de cadenas de Markov con espacio de estados finito. En efecto,

PROPOSICIÓN 2.8. *Si el espacio de estados es finito, una clase es recurrente si y sólo si es cerrada.*

DEMOSTRACIÓN. Sólo hace falta verificar que si C es cerrada entonces es recurrente. Puesto que C es cerrada, vemos que para cualquier $x \in C$,

$$1 = \mathbb{P}_x(X_n \in C \text{ para toda } n \geq 0).$$

Por otra parte, al ser E finito, lo anterior fuerza a que exista $y \in C$ que se visita infinitas veces bajo \mathbb{P}_x :

$$0 < \mathbb{P}_x(V_y = \infty).$$

Si T denota a la primera visita a y , vemos que

$$0 < \mathbb{P}_x(T_y < \infty) \mathbb{P}_y(V_y = \infty)$$

de acuerdo a la propiedad de Markov fuerte. Por lo tanto, vemos que y es recurrente y que así la clase C es recurrente. \square

La primera conclusión es que una cadena irreducible con espacio de estados finitos tiene a todos los estados recurrentes. Un ejemplo muy concreto sería el de la cadena de Ehrenfest. Aún más, en una cadena irreducible y recurrente, de cualquier estado se accede a cualquier otro.

PROPOSICIÓN 2.9. *Si la cadena es irreducible entonces $\mathbb{P}_x(V_y = \infty) = 1$ para toda $x, y \in E$*

En particular, si recordamos la cadena de Markov que modela el movimiento de una rata por el laberinto ilustrado en la Figura 1, si colocamos comida en alguna celda, entonces la rata la encontrará con probabilidad 1.

DEMOSTRACIÓN. Recordemos que $\mathbb{P}_y(V_y = \infty) = 1$ para toda $y \in E$. Por otra parte, al aplicar la propiedad de Markov al instante n , vemos que

$$1 = \mathbb{P}_y(V_y = \infty) = \sum_{x \in E} P_{y,x}^n \mathbb{P}_x(V_y = \infty).$$

Del lado derecho tenemos un promedio ponderado de las cantidades $\mathbb{P}_x(V_y = \infty) \leq 1$. El promedio es igual a 1 si y sólo si $\mathbb{P}_x(V_y = \infty) = 1$ para toda x tal que $P_{y,x}^n > 0$. Por irreducibilidad, para toda y existe n tal que $P_{y,x}^n > 0$ y por lo tanto $\mathbb{P}_x(V_y = \infty) = 1$ para toda $x, y \in E$. \square

Para la cadena de la ruina del jugador, donde el espacio de estados es $\{0, \dots, N\}$ y la matriz de transición es

$$P_{i,j} = \begin{cases} p & i < N, j = i + 1 \\ 1 - p & i > 0, j = i - 1 \\ 1 & i = 0, N \end{cases},$$

vemos que 0 y N son absorbentes y que de cualquier estado se accede a 0 y a N . Hay por lo tanto 3 clases de comunicación: $\{0\}$, $\{1, \dots, N - 1\}$, $\{N\}$. La primera y la última son cerradas y por lo tanto recurrentes mientras que la segunda es abierta y por lo tanto transitoria. Cabe la pregunta de si en esta cadena alcanzamos alguno de los estados 0 y N con probabilidad 1. La respuesta es por supuesto afirmativa y se puede generalizar como sigue:

PROPOSICIÓN 2.10. *Sean*

$$A = \{x \in E : C_x \text{ es abierta}\}, \quad C = \{x \in E : C_x \text{ es cerrada}\}$$

y

$$H_C = \begin{cases} \infty & \text{si } \{n \in \mathbb{N} : X_n \in C\} = \emptyset \\ \min \{n \geq 0 : X_n \in C\} & \text{en caso contrario} \end{cases}.$$

Si A es finita entonces para todo $x \in E$, $\mathbb{P}_x(H_C < \infty) = 1$.

DEMOSTRACIÓN. Si $x \in C$ entonces $1 = \mathbb{P}_x(H_C = 0) \leq \mathbb{P}_x(H_C < \infty)$.

Por la propiedad de Markov fuerte, si $x \in E$ y $y \in A$, se tiene que

$$\mathbb{P}_x(V_y = \infty) = \mathbb{P}_x(T_y < \infty) \mathbb{P}_y(V_y = \infty).$$

El segundo factor del lado derecho es cero puesto que y es transitorio. Puesto que E es finito, se concluye que $\mathbb{P}_x\left(\sum_{y \in A} V_y = \infty\right) = 0$ para todo $x, y \in A$.

Por otra parte, ya que $\sum_{y \in E} V_y = \infty$, se deduce que para $x \in A$:

$$\mathbb{P}_x\left(\sum_{y \in C} V_y = \infty\right) = 1.$$

Como

$$\mathbb{P}_x(H_C < \infty) \geq \mathbb{P}_x\left(\sum_{y \in C} V_y = \infty\right),$$

se deduce el resultado deseado. \square

El análisis de recurrencia y transitoriedad de cualquier cadena con espacio de estados finito es como el de los dos ejemplos anteriores. Sin embargo, cuando el espacio de estados es infinito, el análisis de recurrencia y transitoriedad es mucho más delicado. Se presenta un ejemplo célebre de este tipo de análisis. Para más ejemplos, necesitaremos profundizar en la relación entre la propiedad de Markov y las relaciones de recurrencia, en la siguiente sección.

EJEMPLO 2.4 (Recurrencia y transitoriedad de la caminata aleatoria simple y el teorema de Pólya). Sea P dada por $P_{i,i+1} = p = 1 - P_{i,i-1}$ para $i \in \mathbb{Z}$, donde $p \in (0, 1)$ y sea $S = (S_n)$ una cadena de Markov con esta matriz de transición.

Esta cadena es irreducible y por lo tanto, basta ver la recurrencia o transitoriedad del cero y los demás estados compartirán esta característica.

Si $p \neq 1/2$, la ley fuerte de los grandes números nos dice que $S_n/n \rightarrow 2p - 1$ conforme $n \rightarrow \infty$ casi seguramente. Por lo tanto, para $\varepsilon < p \wedge 1 - p$, existe N tal que para toda $n \geq N$, se tiene que $(2p - 1 - \varepsilon)n < S_n < (2p - 1 + \varepsilon)n$. Vemos que entonces el número de visitas al estado cero es finito casi seguramente y por lo tanto 0 es transitorio. Note que este argumento no es válido cuando $p = 1/2$.

Si $p = 1/2$, entonces se puede calcular explícitamente $P_{0,0}^{2n} = \binom{2n}{n} 2^{-2n}$. Ahora utilizamos la fórmula de Stirling, la cual afirma que

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1,$$

vemos que

$$\frac{P_{0,0}^n}{1/\sqrt{\pi n}}.$$

Por lo tanto

$$\sum_n P_{0,0}^n = \infty$$

y así vemos que S es una cadena recurrente.

Consideremos ahora una caminata aleatoria simple y simétrica en dimensión d , que tomaremos igual a 2 ó 3. Si $i, j \in \mathbb{Z}^d$, definiremos $P_{i,j} = 1/2d$ si $\sum_l |i_l - j_l| = 1$. Mediante un argumento combinatorio y la fórmula de Stirling, se puede probar que $P_{0,0}^n/n^{d/2} \rightarrow C_d$ donde $C_d \in (0, \infty)$. Entonces, vemos que si $d = 2$ la caminata aleatoria simple y simétrica es recurrente mientras que si $d = 3$, es transitoria. Éste último resultado se conoce como Teorema de Pólya para la caminata aleatoria simple.

4. Ejemplos de utilización de la propiedad de Markov

EJEMPLO 2.5 (El problema de la ruina). Consideremos la matriz de transición $P_{i,i+1} = 1 - P_{i,i-1} = p \in (0, 1)$ que da lugar a una caminata aleatoria simple en \mathbb{Z} . Sea X una cadena de Markov con transición P y tal que comienza en i bajo \mathbb{P}_i . Consideremos $N \geq 2$ y $m \in \{1, \dots, N-1\}$. Definamos al siguiente tiempo de paro:

$$T = \min \{n \geq 0 : X_n \in \{0, N\}\}.$$

El objetivo del ejemplo es utilizar la propiedad de Markov, mediante el llamado análisis del primer paso para calcular la probabilidad siguiente:

$$(2) \quad \mathbb{P}_m(X_T = N).$$

La interpretación es la siguiente: un jugador tiene un capital inicial de m pesos y se enrola en un juego de volados en el que gana un peso con probabilidad p (y los juegos son independientes). No deja de jugar hasta que pasan dos cosas: acumula un capital objetivo de N pesos o pierde toda su fortuna. Lo que se quiere determinar es la probabilidad de que termine de jugar con N pesos. Claramente, para este problema, es igual trabajar con la caminata aleatoria simple, y su espacio de estados infinito, que con la caminata aleatoria con absorción en 0 y en N que tiene espacio de estados $\{0, \dots, n\}$ y matriz de transición Q dada por

$$Q_{0,0} = Q_{N,N} = 1 \quad , \quad Q_{i,i+1} = p = 1 - Q_{i,i-1} \text{ si } 1 \leq i \leq N-1.$$

La cantidad $Q^l m, 0$ nos da la probabilidad haber llegado a cero antes del instante l y de visitar el estado N . El siguiente código R permite hacer un cálculo de esta cantidad cuando $N = 10$, $l = 26$, $m = 5$ y $p = 5/11$.

```
# Objetivo: encontrar, numéricamente, las probabilidades de transición y de
ruina para la cadena de la ruina del jugador
```

```

N<-10    #Capital objetivo
p<- 5/11    #Probabilidad de ir hacia arriba
m<-floor(N/2) #Capital inicial

P<-matrix(0,N+1,N+1)    #Matriz de transicion
P[1,1]<-1
P[N+1,N+1]<-1
for(i in 2:N) {P[i,(i-1):(i+1)]<-c(1-p,0,p)}

p<-matrix(0,1,N+1)    #Distribucion inicial
p[m+1]<-1
aux<-p
l<-26    # Cantidad de pasos que se haran
M<-matrix(0,1+1,N+1)
M[1,]<-aux

for(i in 1:l){
  aux<-aux%*%P
  M[i+1,]<-aux
}

library(xtable) #Cargar el paquete que genera la tabla
print(xtable(M),type="latex",file="Ruina2.tex") #Genera el TeX de la tabla

```

LISTING 2.5. Ruina2.R

Se escogió n pequeño para poder presentar los resultados en la Tabla 1. Sólo se imprimen dos decimales.

Antes que nada, probaremos que $T < \infty$ \mathbb{P}_m casi seguramente. Al notar que si hay N incrementos de X igual a 1 (lo cual sucede con probabilidad $p^N > 0$, entonces $T < \infty$). Sin embargo, como bajo \mathbb{P}_m los incrementos son independientes y toman el valor 1 con probabilidad p , esto sucedera casi seguramente (como se ve por ejemplo al utilizar Borel-Cantelli).

Escribamos

$$q_m = \mathbb{P}_m(X_T = N)$$

y determinemos ahora el valor de q_m . Hay dos casos sencillos:

$$q_0 = 0 \quad \text{y} \quad q_n = 1.$$

Por otro lado, observamos que si $m \leq n - 2$ entonces $1 < T$ por lo que

$$q_m = pq_{m+1} + (1-p)q_{m-1}.$$

Así, la probabilidad de interés queda determinada por una relación de recurrencia con valores de frontera.

Afortunadamente, la solución a la relación de recurrencia se conoce. En efecto, escribamos la relación de recurrencia en la forma

$$pq_m + qq_m = pq_{m+1} + qq_{m-1} \quad \text{ó} \quad pq_m + qq_m = pq_{m+1} + qq_{m-1}$$

	1	2	3	4	5	6	7	8	9	10	11
1	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.55	0.00	0.45	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.30	0.00	0.50	0.00	0.21	0.00	0.00	0.00
4	0.00	0.00	0.16	0.00	0.41	0.00	0.34	0.00	0.09	0.00	0.00
5	0.00	0.09	0.00	0.30	0.00	0.37	0.00	0.20	0.00	0.04	0.00
6	0.05	0.00	0.20	0.00	0.34	0.00	0.28	0.00	0.12	0.00	0.02
7	0.05	0.11	0.00	0.27	0.00	0.30	0.00	0.19	0.00	0.05	0.02
8	0.11	0.00	0.20	0.00	0.29	0.00	0.24	0.00	0.12	0.00	0.04
9	0.11	0.11	0.00	0.25	0.00	0.26	0.00	0.17	0.00	0.05	0.04
10	0.17	0.00	0.19	0.00	0.26	0.00	0.21	0.00	0.11	0.00	0.07
11	0.17	0.10	0.00	0.22	0.00	0.23	0.00	0.16	0.00	0.05	0.07
12	0.22	0.00	0.17	0.00	0.23	0.00	0.19	0.00	0.10	0.00	0.09
13	0.22	0.09	0.00	0.20	0.00	0.21	0.00	0.14	0.00	0.04	0.09
14	0.27	0.00	0.15	0.00	0.21	0.00	0.17	0.00	0.09	0.00	0.11
15	0.27	0.08	0.00	0.18	0.00	0.19	0.00	0.13	0.00	0.04	0.11
16	0.32	0.00	0.14	0.00	0.18	0.00	0.15	0.00	0.08	0.00	0.13
17	0.32	0.07	0.00	0.16	0.00	0.17	0.00	0.11	0.00	0.04	0.13
18	0.36	0.00	0.12	0.00	0.17	0.00	0.14	0.00	0.07	0.00	0.14
19	0.36	0.07	0.00	0.15	0.00	0.15	0.00	0.10	0.00	0.03	0.14
20	0.40	0.00	0.11	0.00	0.15	0.00	0.12	0.00	0.06	0.00	0.16
21	0.40	0.06	0.00	0.13	0.00	0.13	0.00	0.09	0.00	0.03	0.16
22	0.43	0.00	0.10	0.00	0.13	0.00	0.11	0.00	0.06	0.00	0.17
23	0.43	0.05	0.00	0.12	0.00	0.12	0.00	0.08	0.00	0.03	0.17
24	0.46	0.00	0.09	0.00	0.12	0.00	0.10	0.00	0.05	0.00	0.18
25	0.46	0.05	0.00	0.11	0.00	0.11	0.00	0.07	0.00	0.02	0.18
26	0.48	0.00	0.08	0.00	0.11	0.00	0.09	0.00	0.05	0.00	0.19
27	0.48	0.04	0.00	0.09	0.00	0.10	0.00	0.07	0.00	0.02	0.19

TABLA 1. El renglón i representa la distribución en el i -ésimo paso de la cadena de la ruina del jugador

ó inclusive

$$(q_m - q_{m+1}) = (q/p) (q_{m-1} - q_m) .$$

Se sigue entonces que

$$q_m - q_{m+1} = -q_1 (q/p)^m$$

y por lo tanto

$$q_m = \begin{cases} q_1 \frac{1-(q/p)^m}{1-(q/p)} & \text{si } q \neq p \\ q_1 m & \text{si } q = p = 1/2 \end{cases} .$$

Al utilizar la igualdad $q_n = 1$ obtenemos finalmente

$$q_m = \begin{cases} \frac{1-(q/p)^m}{1-(q/p)^N} & \text{si } q \neq p \\ m/N & \text{si } q = p \end{cases}.$$

Cuando $p = q = 1/2$, podemos adicionalmente calcular $v_m = \mathbb{E}_m(T)$. En efecto, por la propiedad de Markov y el hecho de que $T \circ \theta_1 = \mathbf{1}_{T \geq 1} + T$, vemos que

$$v_0 = v_N = 0 \quad \text{y} \quad 2v_m = 2 + v_{m+1} + v_{m-1}.$$

La última ecuación se puede escribir, definiendo $d_m = v_m - v_{m-1}$ como una igualdad matricial:

$$\begin{pmatrix} d_{m+1} \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_m \\ -2 \end{pmatrix},$$

por lo que

$$\begin{pmatrix} d_{m+1} \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^m \begin{pmatrix} d_1 \\ -2 \end{pmatrix}.$$

La potencia de la matriz resultante se puede calcular y nos devuelve

$$d_{m+1} = d_1 - 2m.$$

Puesto que $d_1 = v_1$ y $v_0 = 0$, vemos que

$$v_m = d_1 + \cdots + d_m = v_1 - v_1 - 2 - \cdots - v_1 - 2(m-1) = mv_1 - m(m-1).$$

Al utilizar $v_N = 0$, vemos que

$$v_1 = N - 1$$

y que por lo tanto

$$v_m = m(N - m).$$

EJEMPLO 2.6 (Recurrencia y transitoriedad de una caminata aleatoria simple con barrera absorbente). El ejemplo anterior nos permite establecer la recurrencia o transitoriedad de una cadena de Markov con espacio de estados infinito que está ligada a la caminata aleatoria simple. Sea $P_{i,i+1} = p$ y $P_{i,i-1} = q$ si $i \geq 1$ y $P_{0,1} = 1$. Sea X una cadena de Markov con transición P tal que comienza en i bajo \mathbb{P}_i . Notemos que P se comporta como una caminata aleatoria si se encuentra en $i \geq 1$, pero si llega a cero automáticamente pasa a 1. Puesto que esta cadena y la del ejemplo anterior tienen el mismo comportamiento hasta que llegan a cero o a n , vemos que si $p \neq q$:

$$\mathbb{P}_i(T_n < T_0) = \frac{1 - (q/p)^i}{1 - (q/p)^n}.$$

Al tomar el límite conforme $n \rightarrow \infty$, obtenemos

$$\mathbb{P}_i(T_0 = \infty) = \begin{cases} 1 - (q/p)^i & q < p \\ 0 & q \geq p \end{cases}.$$

Así, la caminata aleatoria simple con barrera reflejante en cero es transitoria si $p > 1/2$ y recurrente si $p \leq 1/2$.

EJEMPLO 2.7 (Cadenas de nacimiento y muerte). Para una sucesión de probabilidades de éxito $p_i \in (0, 1)$ (con $q_i = 1 - p_i$), sea $(\mathbb{P}_i, i \geq 0)$ la familia markoviana asociada a un proceso de nacimiento y muerte con matriz de transición determinada por $P_{i,i+1} = p_i$. Haremos un análisis de primer paso para calcular $h_i = \mathbb{P}_i(T_0 < \infty)$. Notemos que $h_0 = 1$.

De nuevo, la propiedad de Markov nos permite obtener la siguiente relación de recurrencia para $i \geq 1$:

$$h_i = p_i h_{i+1} + q_i h_{i-1}.$$

De nueva cuenta, si escribimos $d_i = h_i - h_{i+1}$, se tiene que

$$p_i d_i = q_i d_{i-1}$$

y sabemos que $d_0 = 1 - h_1$. Vemos que entonces

$$d_i = \gamma_i d_0 \quad \text{donde} \quad \gamma_i = \frac{p_i \cdots p_1}{q_i \cdots q_1}.$$

Así, podemos escribir

$$d_0 (1 + \gamma_1 + \cdots + \gamma_{i-1}) = d_0 + \cdots + d_{i-1} = 1 - h_i.$$

Ahora debemos hacer un análisis más preciso para determinar a la constante faltante d_0 .

Si $\sum_i \gamma_i = \infty$, puesto que $1 - h_i \in [0, 1]$, vemos que $d_0 = 1 - h_1 = 0$ y por lo tanto $h_i = 1$ para toda $i \geq 1$.

Si $\sum_i \gamma_i < \infty$, definamos

$$\tilde{h}_0 = 1 \quad \text{y} \quad \tilde{h}_i = 1 - a(1 + \gamma_0 + \cdots + \gamma_i),$$

por lo que $\tilde{h}_i \in [0, 1]$ y

$$\tilde{h}_i = p_i \tilde{h}_{i+1} + q_i \tilde{h}_{i-1} \quad \text{si } i \geq 1.$$

Vemos que entonces

$$\tilde{h}_i = \mathbb{E}_i(\tilde{h}_{X_n})$$

para toda $n \geq 1$. Sin embargo:

$$\mathbb{E}_i(\tilde{h}_{X_n}) = \mathbb{P}_i(T_0 \leq n) + \mathbb{E}_i(\tilde{h}_{X_n} \mathbf{1}_{n < T_0}) \geq \mathbb{P}_i(T_0 \leq n).$$

Como lo anterior es válido para toda n , vemos que

$$\tilde{h}_i \geq h_i.$$

Así, h_i es la mínima solución no-negativa a la relación de recurrencia y por lo tanto $d_0 = 1/(1 + \sum_i \gamma_i)$. Vemos que por lo tanto $h_i < 1$ para toda $i \geq 1$.

EJEMPLO 2.8 (Tiempos de arribo para la caminata aleatoria simple unidimensional). Sea $(\mathbb{P}_k, k \in \mathbb{Z})$ la familia Markoviana asociada a la caminata aleatoria simple con matriz de transición $P_{i,i+1} = 1 - P_{i,i-1} = p \in (0, 1)$. Sea T_0 el primer arribo a cero dado por

$$T_0 = \min \{n : X_n = 0\}.$$

Nuestro objetivo será determinar a

$$\phi(s) = \mathbb{E}_1(s^{T_0}).$$

Comenzando en 2, la caminata aleatoria simple debe pasar por uno para llegar a cero, y la trayectoria de 2 a 1 tiene la misma distribución que la de 1 a cero. Por lo tanto:

$$\mathbb{E}_2(s^{T_0}) = \phi(s)^2.$$

Por otra parte, la propiedad de Markov al instante 1 nos dice que

$$\phi(s) = \mathbb{E}_1(s^{T_0}) = (1-p)s + ps\mathbb{E}_2(s^{T_0}) = (1-p)s + ps\phi(s)^2.$$

Así, puesto que $\phi(s) \in (0, 1)$ para $s \in (0, 1)$, vemos que

$$\phi(s) = \frac{1 - \sqrt{1 - 4p(1-p)s^2}}{2ps}.$$

Esto tiene varias implicaciones. La primera es que

$$\mathbb{E}_1(T_0 < \infty) = \lim_{s \rightarrow 1} \mathbb{E}_1(s^{T_0}) = \frac{1 - |1 - 2p|}{2p} = \begin{cases} 1 & p < 1/2 \\ \frac{q}{p} & p \geq 1/2 \end{cases}.$$

La segunda es el cálculo, para $p \leq 1/2$ de la esperanza de T_0 : al derivar la ecuación que satisface ϕ vemos que

$$0 = 1 - (2p - 1)\phi'(1-)$$

EJEMPLO 2.9 (El mínimo de caminatas aleatorias skip-free). Sea $\mathbb{P}_k, k \in \mathbb{Z}$ la familia markoviana asociada a una caminata aleatoria cuyos saltos pertenecen a $\{-1, 0, 1, \dots\}$. Dichas caminatas se llaman sin saltos a la izquierda (skip-free to the left) aunque el punto es que el mínimo acumulativo tiene como imagen un intervalo de enteros. Sea

$$-I = \min_{n \geq 0} X_n.$$

Obviamente $I = 0$ si $\mathbb{P}_0(X_1 = -1) = 0$. Supongamos por lo tanto que este no es el caso. Veamos que entonces bajo \mathbb{P}_0 , I es una variable aleatoria geométrica (aunque posiblemente degenerada en infinito). Recordemos que las variables geométricas están caracterizadas por la propiedad de pérdida de memoria, por lo que es suficiente verificar que

$$\mathbb{P}_0(I \geq m + n) = \mathbb{P}_0(I \geq n) \mathbb{P}_0(I \geq m).$$

Sin embargo, notemos primero que si

$$T_{-m} = \min \{n \in \mathbb{N} : X_n = -m\},$$

entonces $\{I \geq m\} = \{T_{-m} < \infty\}$. Segundo, sobre el conjunto $\{T_{-m} < \infty\}$ se tiene que $I = I \circ \theta_{T_{-m}}$. La propiedad de Markov fuerte nos permite afirmar que

$$\mathbb{P}_0(I \geq m+n) = \mathbb{P}_0(I \geq m) \mathbb{P}_{-m}(I \geq n+m).$$

Finalmente, puesto que \mathbb{P}_0 es la distribución de $X+m$ bajo \mathbb{P}_{-m} , vemos que

$$\mathbb{P}_{-m}(I \geq n+m) = \mathbb{P}_0(I \geq n).$$

Ahora determinemos el parámetro de la distribución de I , al que denotaremos por p y escribimos $q = 1-p$. Al aplicar la propiedad de Markov al tiempo 1, vemos que

$$q = \mathbb{P}_0(I \geq 1) = \mathbb{P}_0(\mathbb{P}_{X_1}(I \geq 1)) = \mathbb{E}_0(q^{1+X_1}).$$

Si ϕ denota a la función generadora de la distribución de salto, vemos que

$$1 = \phi(q).$$

Por la desigualdad de Hölder, es fácil ver que la función $\phi(e^{-\lambda})$ es log-convexa. Por lo tanto la ecuación $\phi(s) = 1$ sólo se satisface cuando $s = 1$ si $\phi'(1-) = \mathbb{E}_0(X_1) \leq 0$ (lo cual dice que I es casi seguramente infinito) mientras que admite dos soluciones, digamos \tilde{q} y 1 si $\mathbb{E}_0(X_1) > 0$. Ahora veremos que $\tilde{q} = q$ en este último caso. Basta mostrar que $q \leq \tilde{q}$. Puesto que

$$\mathbb{E}_k(\tilde{q}^{X_1}) = \tilde{q}^k \mathbb{E}_0(\tilde{q}^{X_1}) = \tilde{q}^k \phi(\tilde{q}) = \tilde{q}^k,$$

la propiedad de Markov nos permite probar que

$$\mathbb{E}_k(\tilde{q}^{X_n}) = \tilde{q}^k.$$

Por lo tanto, al utilizar la propiedad de Markov

$$\begin{aligned} \tilde{q} = \mathbb{E}_0(\tilde{q}^{1+X_n}) &= \sum_{k=0}^n \mathbb{E}_0(\mathbf{1}_{T_{-1}=k} \mathbb{E}_{-1}(\tilde{q}^{1+X_{n-k}})) + \mathbb{E}_0(\mathbf{1}_{T_{-1}>n} \tilde{q}^{1+X_n}) \\ &= \mathbb{P}_0(T_1 \leq n) + \mathbb{E}_0(\mathbf{1}_{T_{-1}>n} \tilde{q}^{1+X_n}) \\ &\geq \mathbb{P}_0(T_1 \leq n). \end{aligned}$$

Al tomar el límite conforme $n \rightarrow \infty$, vemos que $\tilde{q} \geq q$ y por lo tanto, q es la solución mínima en $[0, 1]$ a la ecuación $\phi(s) = 1$.

EJEMPLO 2.10 (El principio de reflexión). Consideremos la familia markoviana $(\mathbb{P}_k)_{k \in \mathbb{Z}}$ asociada a la caminata aleatoria simple y sea

$$T_m = \min \{n \in \mathbb{N} : X_n = m\}.$$

Definamos

$$\tilde{X}_n = \begin{cases} X_n & \text{si } n \leq T_m \\ 2m - X_n & \text{si } n > T_m \end{cases}.$$

El principio de reflexión, que ahora demostraremos, afirma que la distribución de \tilde{X} bajo \mathbb{P}_0 es precisamente \mathbb{P}_0 . En efecto, el principio de reflexión es consecuencia de la propiedad de Markov fuerte y el hecho de que la distribución de $-X$ bajo \mathbb{P}_0 es también \mathbb{P}_0 .

Como una aplicación, notemos que si $k \geq 0$ entonces

$$\mathbb{P}_0(T_m \leq n) = \mathbb{P}_0(T_m \leq n, X_n \geq m) + \mathbb{P}_0(T_m \leq n, X_n < m).$$

Puesto que $\{T_m \leq n\} \subset \{X_n \geq m\}$ y

$$\{T_m \leq n, X_n \leq m\} = \{\tilde{X}_n \geq m\},$$

se deduce que

$$\mathbb{P}_0(T_m \leq n) = \mathbb{P}_0(X_n = m) + 2\mathbb{P}_0(X_n > m).$$

5. Medidas y distribuciones invariantes

Comenzaremos con un sistema lineal importante que se obtiene de una aplicación de la propiedad de Markov fuerte.

EJEMPLO 2.11 (Cantidad esperada de visitas antes de la recurrencia). Consideremos de nuevo a los tiempos de primera visita

$$T_y = \min \{n \in \mathbb{N} : X_n = y\}.$$

Fijemos a $x \in E$ y definamos

$$\nu_x^x = 1 \quad \text{y} \quad \nu_y^x = \mathbb{E}_x \left(\sum_{i=0}^{T_x} \mathbf{1}_{X_i=y} \right) \quad \text{para } y \neq x.$$

TEOREMA 2.3. *Para una cadena irreducible y recurrente, ν^x satisface $\nu_x^x = 1$, $0 < \nu_y^x < \infty$ y*

$$(3) \quad \sum_{z \in E} \nu_y^z P_{y,z} = \nu_y^x.$$

Una **medida invariante** es un vector renglón $\nu_y, y \in E$ con entradas no-negativas tal que $\nu P = \nu$ o más explícitamente,

$$\sum_{z \in E} \nu_y^z P_{y,z} = \nu_y.$$

Por lo tanto, el vector ν^x representa una construcción probabilística de una medida invariante.

DEMOSTRACIÓN. Claramente $\nu_x^x = 1$. Para ver que ν^x satisface (3), utilizamos la propiedad de Markov de manera similar al análisis del primer paso. Supongamos primero que $y \neq x$. Entonces, puesto que T_x es finito \mathbb{P}_x casi seguramente

$$\mathbb{E}_x \left(\sum_{n < T_x} \mathbf{1}_{X_n=y} \right) = \mathbb{E}_x \left(\sum_{n \leq T_x} \mathbf{1}_{X_n=y} \right) = \sum_{n=1}^{\infty} \mathbb{E}_x (\mathbf{1}_{X_n=y} \mathbf{1}_{T_x \leq n}).$$

El sumando $n = 1$ es fácil de calcular puesto que bajo \mathbb{P}_x , $T_x \geq 1$:

$$\mathbb{E}_x (\mathbf{1}_{X_1=y} \mathbf{1}_{T_x \geq 1}) = \mathbb{E}_x (\mathbf{1}_{X_1=y}) = P_{x,y}.$$

Para los sumandos con $n \geq 2$, aplicamos la propiedad de Markov al instante $n - 1$ (haciendo una especie de análisis de último paso), notando que $\{T \leq n\} \in \mathcal{F}_{n-1}$:

$$\begin{aligned} \mathbb{E}_x (\mathbf{1}_{X_n=y} \mathbf{1}_{T_x \leq n}) &= \sum_{z \neq x} \mathbb{E}_x (\mathbf{1}_{X_{n-1}=z} \mathbf{1}_{X_n=y} \mathbf{1}_{T_x \leq n}) \\ &= \sum_{z \neq x} \mathbb{E}_x (\mathbf{1}_{X_{n-1}=z} \mathbf{1}_{T_x \leq n}) P_{z,y}. \end{aligned}$$

Así, vemos que

$$\begin{aligned} \nu_y^x &= \mathbb{E}_x \left(\sum_{n < T_x} \mathbf{1}_{X_n=y} \right) = P_{x,y} + \sum_{z \neq x} \sum_{n=1}^{\infty} \mathbb{E}_x (\mathbf{1}_{X_n=y} \mathbf{1}_{T_x > n}) P_{z,y} \\ &= \nu_x^x P_{x,y} + \sum_{z \neq x} \mathbb{E}_x \left(\sum_{n=0}^{T_x-1} \mathbf{1}_{X_n=y} \right) P_{z,y} \\ &= \sum_{z \in E} \nu_z^x P_{z,y}. \end{aligned}$$

Ahora veremos que $1 = \nu_x^x = \sum_y \nu_y^x P_{y,x}$. En efecto, basta descomponer respecto al valor de T_x y de X_{T_x-1} , recordando que T_x es finito \mathbb{P}_x casi seguramente:

$$\begin{aligned} 1 &= \sum_{n=1}^{\infty} \mathbb{P}_x(T_x = n) \\ &= P_{x,x} + \sum_{n=2}^{\infty} \sum_{y \neq x} \mathbb{P}_x(T_x = n, X_{n-1} = y) \\ &= P_{x,x} + \sum_{n=2}^{\infty} \sum_{y \neq x} \mathbb{P}_x(T_x \geq n, X_{n-1} = y) P_{x,y} \\ &= P_{x,x} + \sum_{y \neq x} \nu_y^x P_{y,x}. \end{aligned}$$

Finalmente, si consideramos m tal que $P_{x,y}^m > 0$ y n tal que $P_{y,x}^n > 0$ entonces por una parte

$$\nu_y^x = \sum_z \nu_z^x P_{z,y}^m \geq P_{x,y}^m > 0$$

y por otra

$$1 = \nu_x^x = \sum_z \nu_z^x P_{x,z}^n \geq \nu_y^x P_{x,y}^n$$

implica que $\nu_y^x < \infty$. □

DEFINICIÓN. Un estado x de una cadena de Markov es **positivo recurrente** si $\mathbb{E}_x(T_x) < \infty$. Denotamos por m_x a dicha cantidad, a la que nos referiremos como tiempo medio de recurrencia de x .

Notemos que, de acuerdo a nuestra definición de ν^x , se tiene que

$$\sum_y \nu_y^x = \mathbb{E}_x \left(\sum_y \sum_{n < T_x} \mathbf{1}_{X_n = y} \right) = \mathbb{E}_x(T_x).$$

Si x es positivo recurrente, podemos entonces definir al vector $\pi^x = (\nu_y^x / \mathbb{E}_x(T_x))$ que satisface

$$\pi_y^x \geq 0, \quad \sum_y \pi_y^x = 1 \quad \text{y} \quad \sum_{y \in E} \pi_y^x P_{y,z} = \pi_z^x.$$

Una **distribución invariante** es un vector renglón que satisface las 3 condiciones anteriores. Una propiedad importante y útil de una distribución invariante es que si X_0 tiene distribución π , entonces X_n tendrá distribución π para toda n . Por otra parte, en espacio de estados finito es fácil ver que existen distribuciones invariantes.

COROLARIO 1. *En una cadena irreducible con espacio de estados finito, todos los estados son positivo recurrentes.*

DEMOSTRACIÓN. Sea x cualquier estado de la cadena. Por el teorema anterior $\nu_y^x < \infty$ para toda y y como E es finito, entonces

$$\mathbb{E}_x(T_x) = \sum_{y \in E} \nu_y^x < \infty. \quad \square$$

En caso de que exista un único vector de probabilidad invariante, se pueden calcular tiempos medios de recurrencia al resolver un sistema de ecuaciones. A continuación profundizaremos en esta idea.

TEOREMA 2.4. *Si ν es una medida invariante para una matriz de transición irreducible P y $\nu_x = 1$ entonces $\nu \geq \nu^x$. Si además P recurrente entonces $\nu = \nu^x$.*

DEMOSTRACIÓN. Al aplicar la invariancia de ν se obtiene

$$\nu_y = \sum_{y_1 \in E} \nu_{y_1} P_{y_1,y} = \nu_x P_{x,y} + \sum_{y_1 \neq x} \nu_{y_1} P_{y_1,x} = P_{x,y} + \sum_{y_1 \neq x} \nu_{y_1} P_{y_1,y}.$$

Al volverla a aplicar se obtiene

$$\nu_y = P_{x,y} + \sum_{y_1 \neq x} P_{x,y_1} P_{y_1,y} + \sum_{y_1, y_2 \neq x} \nu_{y_2} P_{y_2,y_1} P_{y_1,y}$$

y al continuar repetidamente, vemos que

$$\begin{aligned} \nu_y &= P_{x,y} + \sum_{y_1 \neq x} P_{x,y_1} P_{y_1,y} + \cdots + \sum_{y_1, \dots, y_n \neq x} P_{x,y_n} P_{y_n, y_{n-1}} \cdots P_{y_2, y_1} P_{y_1, y} \\ &+ \sum_{y_1, \dots, y_{n+1} \neq x} \nu_{y_{n+1}} P_{y_{n+1}, y_n} \cdots P_{y_2, y_1} P_{y_1, x}. \end{aligned}$$

Si $y \neq x$, encontramos la cota

$$\nu_y \geq \sum_{m=0}^n \mathbb{P}_x(X_m = y, T_x \geq m)$$

y el lado derecho converge conforme $n \rightarrow \infty$ a ν_y^x , por lo que

$$\nu_y \geq \nu_y^x.$$

Por otra parte, si la cadena es recurrente además de irreducible entonces ν^x es invariante, por lo cual $\mu = \nu - \nu^x$ es una medida invariante con $\mu_x = 0$. Por irreducibilidad, para toda $y \in E$ existe $n \geq 0$ tal que $P^n y, x > 0$ y entonces

$$0 = \mu_x = \sum_z \mu_z P_{z,x}^n \geq \mu_y P_{y,x}^n,$$

por lo que $\mu_y = 0$. □

TEOREMA 2.5. *Para una cadena irreducible las siguientes condiciones son equivalentes.*

- (1) *Todos los estados son positivo recurrentes*
- (2) *Algún estado es positivo recurrente*
- (3) *La cadena admite una distribución invariante.*

En este caso, la distribución invariante es única y asigna a x el recíproco de su tiempo medio de recurrencia.

DEMOSTRACIÓN. Primero probaremos que si la cadena admite una distribución invariante ν entonces todos los estados son positivo recurrentes. Para esto, notemos primero que $\nu_x > 0$ para toda $x \in E$. En efecto, lo debe ser para alguna x y al utilizar la irreducibilidad para encontrar n tal que $P_{y,x}^n > 0$, vemos que

$$\nu_y = \sum_{z \in E} P_{z,y}^n \nu_z \geq \nu_x P_{x,y}^n > 0.$$

Si $x \in E$, entonces ν/ν_x es una medida invariante que asigna 1 a x , por lo cual $\nu/\nu_x \geq \nu^x$. Puesto que ν es una distribución:

$$m_x = \mathbb{E}_x(T_x) = \sum_y \nu_y^x \leq \sum_y \frac{\nu_y}{\nu_x} = \frac{1}{\nu_x} < \infty.$$

Así, todos los estados son positivo recurrentes. Al ser en particular recurrentes, hemos visto que $\nu_x = \nu_x^x = m_x$ y como ν_x es una distribución entonces $m_x < \infty$. Esto termina la demostración de las implicaciones y además nos produce la fórmula requerida para la distribución invariante. \square

EJEMPLO 2.12 (La cadena de Ehrenfest). Sea $E = \{0, \dots, N\}$ y $P_{i,i+1} = 1 - i/N = 1 - P_{i,i-1}$. Esta cadena es irreducible y con espacio de estados finito, por lo que todos los estados son positivo recurrentes y por lo tanto existe un único vector de probabilidad invariante π . Además, $\pi_x = 1/\mathbb{E}_x(T_x)$. El siguiente código nos permite obtener numéricamente estas cantidades.

```
N<-10 #Cantidad de bolas para la cadena de Ehrenfest
P<-matrix(0,N+1,N+1)
P[1,2]<-1
P[N+1,N]<-1
for(i in 2:N){
  P[i,i+1]<-1-(i-1)/N
  P[i,i-1]<-(i-1)/N
}
I<-diag(1,N+1,N+1)
Null<-function(M) #Calcula una base del kernel de la matriz M mediante
  factorizacion QR
{
  tmp <- qr(M)
  set <- if(tmp$rank == 0) 1:ncol(M) else - (1:tmp$rank)
  qr.Q(tmp, complete = TRUE)[, set, drop = FALSE]
}
v<-t(Null(I-P)) # El kernel de I-P tiene dimension 1 por ser la cadena
  irreducible y finita
v<-v/sum(v) #Se obtiene el vector de probabilidad invariante
m<-1/v #Se obtienen los tiempos medios de recurrencia
library(xtable) #Cargar el paquete que genera tablas en TeX
print(xtable(v),type="latex",file="EhrenfestInvariantDist.tex") #Genera el TeX
  de la tabla
print(xtable(m),type="latex",file="EhrenfestMeanRecurrence.tex") #Genera el TeX
  de la tabla
```

LISTING 2.6. EhrenfestMatrix.R

El resultado se puede visualizar en las tablas siguientes

	1	2	3	4	5	6	7	8	9	10	11
1	0.00	0.01	0.04	0.12	0.21	0.25	0.21	0.12	0.04	0.01	0.00

	1	2	3	4	5	6	7	8	9	10	11
1	1024.00	102.40	22.76	8.53	4.88	4.06	4.88	8.53	22.76	102.40	1024.00

Por otra parte, para esta cadena particular, podemos calcular explícitamente el vector de probabilidad invariante y por lo tanto los tiempos medios de recurrencia. En efecto, sea $\pi_j = \binom{N}{j} 2^{-N}$ y notemos que si $j \in \{1, \dots, N-1\}$ entonces

$$\begin{aligned}
\sum_i \pi_i P_{i,j} &= \pi_{j-1} P_{j-1,j} + \pi_{j+1} P_{j+1,j} \\
&= \binom{N}{j-1} 2^{-N} \frac{N-j+1}{N} + \binom{N}{j+1} 2^{-N} \frac{j+1}{N} \\
&= 2^{-N} N! \left[\frac{1}{(j-1)!(N-j+1)!} \frac{N-j+1}{N} + \frac{1}{(j+1)!(N-j-1)!} \frac{j+1}{N} \right] \\
&= 2^{-N} \frac{N!}{(j-1)!(N-j-1)!} \left[\frac{1}{N(N-j)} + \frac{1}{jN} \right] \\
&= 2^{-N} \frac{N!}{(j-1)!(N-j-1)!} \left[\frac{1}{j(N-j)} \right] \\
&= 2^{-N} \binom{N}{j} \\
&= \pi_j.
\end{aligned}$$

Por lo tanto la distribución binomial de parámetros N y $1/2$ es invariante para la cadena de Ehrenfest con espacio de estados $(0, \dots, N)$. Como ésta es irreducible, los tiempos medios de recurrencia son

$$m_i = \frac{2^N}{\binom{N}{i}}.$$

Si ahora utilizamos la fórmula de Stirling, notamos que conforme $N \rightarrow \infty$

$$m_{N/2} \sim \frac{1}{\sqrt{\pi N}} \quad \text{pero} \quad m_0 = 2^N.$$

6. Comportamiento a tiempos grandes

El objetivo de esta sección es estudiar cómo se comporta una cadena de Markov recurrente cuando la dejamos correr un tiempo grande. Una aplicación de este tipo de ideas sirve para desplegar el resultado de una búsqueda en Google en orden descendiente de importancia de las páginas. Para describir cómo se define la importancia de una página, podemos pensar en la siguiente situación: se considera a la red como un conjunto de nodos (páginas web) y aristas (hipervínculos) que los conectan. Si N_v denota la cantidad de hipervínculos que hay en la página v , se

puede definir la importancia de una página v como la cantidad I_v que satisface el sistema de ecuaciones:

$$I_v = \sum_u \frac{I_u}{N_u}, \quad \sum_v I_v = 1.$$

La interpretación es que una página importante transmite su importancia a cada una de las ligas que está presente en la página y que la suma de las importancias es 1. Esta es la base del algoritmo PageRank de Google. Para interpretar lo anterior en términos de cadenas de Markov, consideremos la siguiente estrategia para navegar en esta red: al encontrarme en la página web, escojo un hipervínculo al azar y lo recorro. Entonces mi trayectoria por la red será una cadena de Markov con matriz de transición

$$P_{u,v} = \begin{cases} \frac{1}{N_u} & \text{si existe hipervínculo de } u \text{ a } v \\ \frac{1}{\text{Total de páginas}} & \text{no hay hipervínculos en } u \\ 0 & \text{otro caso} \end{cases}.$$

Notemos que la importancia de una página es simplemente una solución I a la ecuación $I = IP$. Por otra parte, mediante la estrategia de navegación postulada, podríamos pensar a la importancia de una página web como la fracción de tiempo asintótica que le dedico cuando recorro la web al azar. En esta sección veremos que de hecho, son la misma idea. En la página <http://www.nd.edu/~networks/resources.htm> el los autores Albert, Jeong y Barabási proporcionan sus datos sobre la estructura de la red que recopilaron para un estudio de 1999. Por supuesto ha crecido mucho desde entonces, pero es un buen ejemplo de datos que se han hecho de acceso público. Un pequeño detalle es que los fundadores de Google se dieron cuenta de cuestiones numéricas que aparecen al tratar de encontrar la importancia de una página, por lo que modificaron la definición de importancia a:

$$I_v = \alpha \sum_u \frac{I_u}{N_u} + \frac{1 - \alpha}{N},$$

donde $\alpha \in [0, 1]$ y N es la cantidad de páginas web que hay. La interpretación en términos de estrategia de navegación es que, con probabilidad α sigo la estrategia de navegación anterior y con probabilidad $1 - \alpha$ salto a una página al azar. Así, se estará considerando la matriz de transición

$$P^\alpha = \begin{cases} \frac{\alpha}{N_u} & \text{si existe hipervínculo de } u \text{ a } v \\ \frac{1}{\text{total de páginas}} & \text{no hay hipervínculos en } u \\ \frac{1 - \alpha}{N} & \text{otro caso} \end{cases}$$

El valor utilizado por Google es $\alpha = .85$. La diferencia entre P y P^α es que la segunda es irreducible y aperiódica mientras la primera no tiene por qué serlo.

Como ejemplo concreto utilicemos la gráfica que, el 21 de Mayo del 2012, se encuentra en la página de Wikipedia de [PageRank](#). Los vértices son serán etiquetados del 1 al 11 y la gráfica se presenta en la Figura 5. Con $\alpha = .85$, la

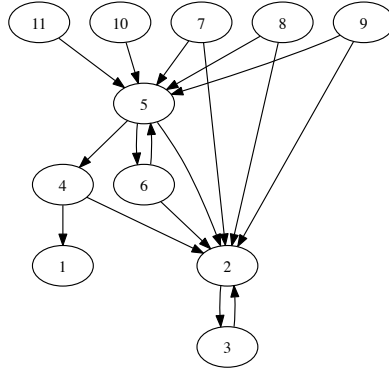


FIGURA 5. Ilustración sobre la definición de importancia de páginas web

matriz de transición es

$$100 * Q = \begin{pmatrix} 9.09 & 9.09 & 9.09 & 9.09 & 9.09 & 9.09 & 9.09 & 9.09 & 9.09 & 9.09 & 9.09 \\ 1.36 & 1.36 & 86.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 86.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 43.8 & 43.8 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 29.70 & 1.36 & 29.70 & 1.36 & 29.70 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 43.8 & 1.36 & 1.36 & 43.8 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 43.8 & 1.36 & 1.36 & 43.8 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 43.8 & 1.36 & 1.36 & 43.8 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 43.8 & 1.36 & 1.36 & 43.8 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 1.36 & 1.36 & 1.36 & 86.3 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \\ 1.36 & 1.36 & 1.36 & 1.36 & 86.3 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 & 1.36 \end{pmatrix}$$

El vector de probabilidad invariante, que nos da la importancia de cada página, es

$$100 * \pi = (3.28, 38.44, 34.29, 3.91, 8.09, 3.91, 1.62, 1.62, 1.62, 1.62, 1.62)$$

Como comentamos, otra manera de pensar a la importancia de un vértice es mediante la fracción de tiempo que paso en el. La fracción esperada de tiempo que paso en j comenzando en i después de n pasos es

$$\frac{1}{n} \sum_{l=0}^n Q_{i,j}^n.$$

Sin embargo, al calcular numéricamente la entrada 50 de la cadena, vemos que es igual a

$$Q^{50} = \begin{pmatrix} 3.28 & 38.44 & 34.29 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.43 & 34.30 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.45 & 34.28 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.45 & 34.28 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.44 & 34.29 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.45 & 34.28 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.45 & 34.28 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.45 & 34.28 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.45 & 34.28 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.44 & 34.29 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \\ 3.28 & 38.44 & 34.29 & 3.91 & 8.09 & 3.91 & 1.62 & 1.62 & 1.62 & 1.62 & 1.62 \end{pmatrix}$$

Vemos que desde el punto de vista numérico, Q^{50} es ya una matriz cuyos renglones son iguales al vector de probabilidad invariante. Esto no es ninguna coincidencia y uno de nuestros objetivos será explicarlo. Se utilizó el siguiente código para generar a la distribución invariante y a la potencia 50 de la matriz.

```
N<-11

A=matrix(c(
  1,1,1,1,1,1,1,1,1,1,1,
  0,0,1,0,0,0,0,0,0,0,0,
  0,1,0,0,0,0,0,0,0,0,0,
  1,1,0,0,0,0,0,0,0,0,0,
  0,1,0,1,0,1,0,0,0,0,0,
  0,1,0,0,1,0,0,0,0,0,0,
  0,1,0,0,1,0,0,0,0,0,0,
  0,1,0,0,1,0,0,0,0,0,0,
  0,1,0,0,1,0,0,0,0,0,0,
  0,1,0,0,1,0,0,0,0,0,0,
  0,0,0,0,1,0,0,0,0,0,0,
  0,0,0,0,1,0,0,0,0,0,0),N,N,byrow=TRUE) #Define la matriz de adyacencia de la
grafica

P=A/rowSums(A) #Normaliza los renglones para obtener una matriz estocastica
a=.85
Pa=a*P+(1-a)/N #Se produce la matriz irreducible

l<-50
aux<-Pa
for(i in 1:l){ #Se calcula la potencia 50 de la matriz de transicion
  aux<-aux%*%Pa
}

I<-diag(1,N,N)
tmp <- qr(I-Pa) # Se obtiene la factorizacion QR
set <- if(tmp$rank == 0) 1:ncol(M) else - (1:tmp$rank)
Null<-qr.Q(tmp, complete = TRUE)[, set, drop = FALSE]
v<-t(Null) # El kernel de I-P tiene dimension 1 por ser la cadena irreducible y
finita
```

```
v<-v/sum(v) #Se obtiene el vector de probabilidad invariante

library(xtable) #Cargar el paquete que genera tablas en TeX
print(xtable(Pa),type="latex",file="PageRankMatrix") #Genera el TeX de la tabla
print(xtable(v),type="latex",file="PageRankInvariantDistribution.tex")
print(xtable(aux),type="latex",file="PageRankMatrixPower.tex")
```

LISTING 2.7. PageRank.R

Vale la pena comparar lo anterior con la cadena de Ehrenfest: con $N = 10$, a continuación se presentan las potencias 50 y 51 de la matriz de transición:

$$1000 * P^{50} = \begin{pmatrix} 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \end{pmatrix}$$

$$1000 * P^{51} = \begin{pmatrix} 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \\ 2 & 0 & 88 & 0 & 410 & 0 & 410 & 0 & 88 & 0 & 2 \\ 0 & 20 & 0 & 234 & 0 & 492 & 0 & 234 & 0 & 20 & 0 \end{pmatrix}$$

Casi lo mismo se repite al calcular potencias superiores. En este caso, parecería haber convergencia de las potencias pares y de las potencias impares a un límite distinto. La diferencia entre la matriz de transición de PageRank y la de la cadena de Ehrenfest se encuentra en el siguiente concepto.

DEFINICIÓN. Sea P una matriz de transición. Definimos al **periodo** de un estado x como el máximo común divisor de $\{n \geq 0 : P_{x,x}^n > 0\}$. Decimos que la cadena es **aperiódica** si el periodo de todo estado es 1.

Si x tiene periodo d , entonces $P_{x,x}^{kd} > 0$ salvo para una cantidad finita de índices k mientras que $P_{x,x}^{kd+1} = \dots = P_{x,x}^{kd+(d-1)} = 0$.

Es fácil verificar que el periodo es el mismo para todos los elementos de una clase de comunicación. Entonces, para verificar que una cadena irreducible es aperiódica, basta verificar que un algún estado tiene periodo 1.

Cuando el espacio de estados es finito, una cadena es irreducible y aperiódica si y sólo si existe $N > 0$ tal que $P_{x,y}^N > 0$ para todo $x, y \in E$.

DEFINICIÓN. Decimos que una cadena finita es **regular** si existe una potencia de la matriz de transición con todas las entradas positivas.

Puesto que la matriz PageRank con $\alpha < 1$ tiene todas las entradas positivas, entonces es regular. Sin embargo, la cadena de Ehrenfest tiene periodo 2, por lo que no es regular. Esta es la gran diferencia entre ambos tipos de cadenas.

El resultado más importante que relaciona a las potencias de la matriz de transición con el vector de probabilidad invariante es el siguiente.

TEOREMA 2.6 (Teorema fundamental de convergencia). *Si P es la matriz de transición de una cadena de Markov irreducible, aperiódica y positivo recurrente y π es la única distribución invariante de P entonces $\sum_y |P_{x,y}^n - \pi_y| \rightarrow 0$ conforme $n \rightarrow \infty$.*

Utilizando el hecho de que si una sucesión converge también converge la sucesión de sus promedios, vemos

$$\lim_{n \rightarrow \infty} \frac{\mathbf{1}_{x=y} + P_{x,y} + \cdots + P_{x,y}^n}{n} = \pi_y.$$

Por lo tanto la entrada y del vector de probabilidad también se puede interpretar como la fracción esperada asintótica de veces en que la cadena se encuentra en el estado y , independientemente de dónde comenzó.

Hay dos tipos de prueba muy distintos para este resultado. El primero se enfoca en espacio de estado finito y utiliza el hecho de que la matriz de transición de la cadena será regular, así como un argumento de punto fijo. El segundo se basa en una idea probabilística conocida como acoplamiento, en el que se prueba que para dos cadenas de Markov independientes que satisfacen las hipótesis del teorema toman el mismo valor en un tiempo aleatorio y finito.

A continuación exploraremos la prueba del teorema fundamental de convergencia en el caso de espacio de estados finito. Comenzaremos con el caso en que no sólo P es regular sino que además P tiene todas las entradas positivas.

Definimos a

$$\rho = \min_{x,y \in E} P_{x,y}.$$

Puesto que el espacio de estados es finito, el mínimo anterior está definido y es positivo. Sea y un vector columna. Probaremos a continuación que las entradas de $P^n y$ se van acercando entre si conforme n crece. Para esto, sean m_n el mínimo de las entradas de $P^n y$ y M^n el máximo. Primero mostraremos que $m_0 \leq m_1 \leq M_1 \leq M_0$ y que $M_1 - m_1 \leq (1 - 2\rho)(M_0 - m_0)$. La idea, que se puede justificar con cálculo

diferencial, es que para maximizar el promedio $\sum_j P_{i,j} y_j$ sujeto a la restricción de que $\max_j y_j = M_0$ y $\min_j y_j = m_0$ formamos un vector con todas sus coordenadas iguales a M_0 salvo una, que será igual a m_0 y que colocamos con el menor peso posible, es decir, en la coordenada \underline{j} tal que $P_{i,\underline{j}} = \min_j P_{i,j} \geq \rho$. Así, obtendremos

$$\sum_j P_{i,j} y_j \leq M_0(1 - P_{i,\underline{j}}) + m_0 P_{i,\underline{j}} \leq M_0(1 - \rho) + m_0 \rho.$$

Con un argumento similar se obtiene

$$\sum_j P_{i,j} y_j \geq m_0(1 - \rho) + M_0 \rho$$

y por lo tanto

$$M_1 - m_1 \leq (1 - 2\rho)(M_0 - m_0)$$

como se anunció.

Al aplicar lo anterior con $P^{n-1}y$ en vez de y , vemos que $m_0 \leq m_1 \leq \dots \leq m_n \leq M_n \leq \dots \leq M_1 \leq M_0$ y que

$$(M_n - m_n) \leq (1 - 2\rho)^n (M_0 - m_0) \rightarrow 0.$$

Por lo tanto, $P^n y$ converge a un vector con todas sus entradas iguales.

Si aplicamos lo anterior al vector e_j que tiene 1 en la entrada j y cero en cualquier otra entradas, vemos que $(P^n e_j)_i$ (la i -ésima entrada de $P^n e_j$) es igual a $P_{i,j}^n$ que esta cantidad converge a u_j para cualquier i . Notemos además que $m_1 > 0$ y por lo tanto $u_j > 0$ para toda j . Así, la matriz P^n converge a una matriz cuyos renglones son todos iguales al vector cuya entrada j es u_j . Finalmente, notamos que u es una distribución invariante. Es una distribución pues, como el espacio de estados es finito, entonces

$$1 = \lim_{n \rightarrow \infty} \sum_j P_{i,j}^n = \sum_j \lim_{n \rightarrow \infty} P_{i,j}^n = \sum_j u_j.$$

Por otro lado, puesto que $P^{n+1} = P^n P$, vemos que

$$u_j = \lim_{n \rightarrow \infty} P_{i,j}^{n+1} = \lim_{n \rightarrow \infty} (P^n P)_{i,j} = \lim_{n \rightarrow \infty} \sum_k P_{i,k}^n P_{k,j} = \sum_k u_k P_{k,j}.$$

Finalmente, aprovechamos la ocasión para dar otra prueba de por qué en este caso, hay a lo más una distribución invariante. Si v es una distribución invariante entonces $vP^n = v$ y por otro lado

$$v_j = (vP^n)_j = \sum_i v_i P_{i,j}^n \rightarrow \sum_i v_i u_j = u_j.$$

Por lo tanto $u = v$.

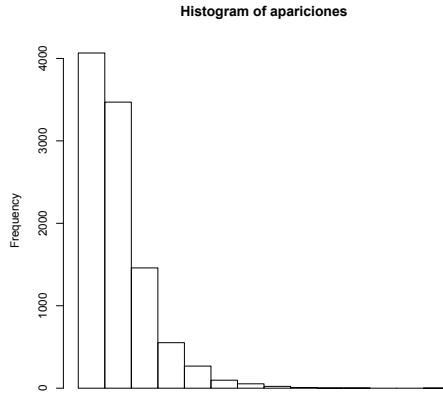


FIGURA 6. Histograma de una muestra del tiempo de aparición de la palabra 101

7. Cadenas absorbentes

Consideremos el siguiente problema: se tiene un abecedario fijo (digamos para fijar ideas igual a $\{0, 1\}$) con cuyos elementos conformamos palabras (como podría ser 1001). Si se extraen las letras del abecedario con igual probabilidad y con independencia entre las extracciones, ¿Cuánto se debe esperar para que aparezca la palabra escogida? Más generalmente: ¿con qué probabilidad aparece la palabra por primera vez al observar n letras? Un caso particular de este problema es cuando la palabra tiene sólo una letra (digamos p); el tiempo de aparición de la palabra es entonces geométrico de parámetro p .

El siguiente código toma una muestra de tamaño 10000 de la distribución del tiempo de aparición de la palabra 101 cuando las letras 1 y 0 aparecen con probabilidades $2/3$ y $1/3$. Se obtiene el histograma de la muestra (ver la Figura 6) así como la media empírica, que resultó ser de 8.26.

```
# Obtener una muestra de tamaño 10000 del tiempo de aparición de la palabra 101.
# 1 aparece con probabilidad 2/3
p<-2/3 #Probabilidad de que salga 1
N<-10000 #Tamaño de muestra
llevo<-0 #Cuántas simulaciones llevo
apariciones<-c() #En que tiempos aparecen las palabras
while(llevo<N){ #Mientras no lleve N simulaciones
  llevo<-llevo+1 #Realizo una mas
  n<-0 #Inicializando a $n$ en cero
  palabraAleatoria<-c() #Con un nuevo repositorio de letras
  aux<-0 #Y una variable que indica si ya he visto la palabra
  while(aux==0){ #Mientras no haya visto la palabra
    n<-n+1 #Espero una unidad de tiempo
    palabraAleatoria<-c(palabraAleatoria,rbinom(1,size=1,prob=2/3)) #Obtengo
    una letra nueva
```

```

if(n>2){if(palabraAleatoria[n-2]==1 & palabraAleatoria[n-1]==0 &
palabraAleatoria[n]==1){ aux<-1 }}
} #Si he visto la palabra, la variable aux nos sacar\'a del ciclo
apariciones<-c(apariciones,n) #Agrego el nuevo tiempo de aparici\'on
}
hist(apariciones)
mean(apariciones)

```

LISTING 2.8. Palabra.R

Consideremos otro problema similar: para el laberinto de la Figura 1, si la rata comienza su trayectoria en el cuarto i , ¿Cuál es la cantidad de tiempo esperada para que llegue a la celda 9? Estas preguntas pueden responderse con la teoría de cadenas de Markov absorbentes. Ya hemos discutido por qué el segundo problema está relacionado con cadenas de Markov, en particular con una cadena de Markov en la que el estado 9 es absorbente. El primero parecería no estarlo. Sin embargo, asociemos una cadena de Markov a nuestro ejemplo particular.

El espacio de estados es $\{0, 1, 2, 3, 4\}$ donde el estado i representa que han aparecido i letras correctas de la palabra 1001. El estado 4 será absorbente y nos señalará que ya ha aparecido por primera vez la palabra. Si p representa la probabilidad de que la n -ésima letra sea 1 y $q = 1 - p$, la matriz de transición de la cadena será

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

Si por ejemplo tuvieramos la palabra 11011, la matriz de transición sería

$$\begin{pmatrix} q & p & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 & 0 \\ 0 & 0 & p & q & 0 & 0 \\ q & 0 & 0 & 0 & p & 0 \\ q & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

puesto que si tengo 2 letras correctas es por que estoy viendo la palabra 11; si sale 1 seguiré observando 11 mientras que si sale 0, observaré 110.

Ahora interpretemos una entrada particular de la matriz de transición: $P_{0,4}^n$, para la palabra 1001. Es la probabilidad de que la cadena se encuentre en el estado al tiempo n , que equivale a que la cadena se haya absorbido en 4 antes del tiempo n , que significa a su vez que haya aparecido la palabra a lo más en la n -ésima extracción.

Pasemos a la abstracción común de ambos problemas.

DEFINICIÓN. Una **cadena absorbente** es una cadena absorbente con espacio de estados finito, que tiene al menos un estado absorbente y que desde cualquier estado se puede acceder a un estado absorbente.

Los estados de una cadena absorbente los podemos dividir en no-absorbentes (T , pues son transitorios) y absorbentes (A). Si los enumeramos del 1 al n y ponemos al final a los absorbentes, la matriz de transición tomará la forma

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix},$$

donde Q es una matriz de tamaño $m \times m$ ($m < n$) en donde se encuentran las probabilidades de transición entre estados no-absorbentes, R es una matriz de tamaño $m \times n$ correspondiente a las probabilidades de transición de estados no-absorbentes a absorbentes, mientras que I es la matriz identidad de tamaño $(n - m) \times (n - m)$. Si i es no-absorbente y j es absorbente, la entrada $P_{i,j}^n$ representa la probabilidad de que comenzando en i la cadena se haya absorbido por el estado j antes del instante n , mientras que $\sum_{j \leq m} Q_{i,j}^n$ representa la probabilidad de que comenzando en i , la cadena todavía no se haya absorbido en el instante n . Es usual escribir

$$\sum_{j \leq m} Q_{i,j}^n = Q^n \mathbf{1}$$

donde $\mathbf{1}$ es un vector columna de longitud m y entradas iguales a 1.

PROPOSICIÓN 2.11. *Sea*

$$T = \begin{cases} \infty & \{n \geq 0 : X_n \in A\} = \emptyset \\ \min \{n \geq 0 : X_n \in A\} & \text{en otro caso} \end{cases}$$

el tiempo de absorción de una cadena absorbente. Entonces $\mathbb{P}_i(T < \infty) = 1$ para toda $i \in E$ y $\lim_{n \rightarrow \infty} Q^n = \mathbf{0}$.

En el enunciado anterior, $\mathbf{0}$ es una matriz de $m \times m$ cuyas entradas son iguales a cero.

DEMOSTRACIÓN. En la Proposición 2.10 hemos probado que $\mathbb{P}_i(T < \infty) = 1$ puesto que T es una clase abierta y finita. Por otra parte, si $i, j \leq m$:

$$0 = \mathbb{P}_i(T = \infty) = \lim_{n \rightarrow \infty} \mathbb{P}_i(T \geq n) \geq \lim_{n \rightarrow \infty} \mathbb{P}_i(X_n = j, T \geq n) = \lim_{n \rightarrow \infty} Q_{i,j}^n.$$

□

Ahora mostraremos un resultado que nos permite encontrar la cantidad esperada de visitas a los diversos estados absorbentes.

PROPOSICIÓN 2.12. *La matriz $I - Q$ es invertible. Si $M = (I - Q)^{-1}$ entonces*

$$M_{i,j} = I + Q + Q^2 + Q^3 + \cdots = \mathbb{E}_i \left(\sum_{n=0}^{\infty} \mathbf{1}_{X_n=j} \right).$$

Si $t_i = \mathbb{E}_i(\sum_{n=0}^{\infty} \mathbf{1}_{X_n \in T})$ entonces

$$t = M\mathbf{1}.$$

Si $B_{i,j} = \mathbb{P}_i(X_T = j)$ entonces

$$B = MR.$$

DEMOSTRACIÓN. Debemos mostrar que si $(I - Q)x = 0$ entonces $x = 0$. Sin embargo, si $x = Qx$ entonces $x = Q^n x \rightarrow 0$, por lo que $x = 0$. Por otra parte, puesto que

$$(I - Q)(1 + Q + Q^2 + \cdots + Q^n) = I - Q^{n+1},$$

al multiplicar por $M = (I - Q)^{-1}$ vemos que

$$1 + Q + Q^2 + \cdots + Q^n = M(I - Q^{n+1})$$

y al tomar el límite conforme $n \rightarrow \infty$, vemos que

$$M_{i,j} = I + Q + Q^2 + Q^3 + \cdots.$$

Al utilizar el teorema de convergencia dominada, vemos que

$$\sum_n Q_{i,j}^n = \sum_n \mathbb{E}_x(\mathbf{1}_{X_n=j}) = \mathbb{E}_x\left(\sum_n \mathbf{1}_{X_n=j}\right).$$

Notemos que

$$\sum_{j \in T} \sum_n \mathbf{1}_{X_n=j} = \sum_n \mathbf{1}_{X_n \in T}$$

es la cantidad de veces que X está en el conjunto T , que es igual al tiempo de absorción. Al sacar esperanza, vemos que

$$t_i = \mathbb{E}_i(T) = \sum_j M_{i,j} = M\mathbf{1}.$$

Finalmente, notemos que

$$\begin{aligned} \mathbb{P}_i(X_T = j) &= \sum_n \sum_{j \in T} \mathbb{P}_i(X_T = j, T = n, X_{n-1} = z) \\ &= \sum_n \sum_{k \in T} P_{i,k}^{n-1} P_{k,j} \\ &= \sum_{k \in T} M_{i,k} P_{k,j}. \end{aligned}$$

□

Apliquemos ahora la proposición anterior al tiempo de aparición de la palabra 101 cuando $p = 2/3$. La matriz de transición y la matriz Q asociadas son:

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 0 & 2/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad Q = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 2/3 & 1/3 \\ 1/3 & 0 & 0 \end{pmatrix}.$$

El siguiente código nos permite calcular numéricamente la matriz fundamental, la matriz B y el vector t .

```
P=matrix(0,4,4)
P[c(1,3,10)]<-1/3
P[c(5,6,15)]<-2/3
P[4,4]<-1
Q=P[1:3,1:3]
R=P[1:3,4]
I=diag(1,3,3)
M=solve(I-Q)
uno=matrix(1,3,1)
t=M%%uno
B=M%%R
```

LISTING 2.9. PalabraFundamental.R

Se obtiene

$$M = \begin{pmatrix} 2.25 & 4.50 & 1.50 \\ 0.75 & 4.50 & 1.50 \\ 0.75 & 1.50 & 1.50 \end{pmatrix} \quad \text{y} \quad t = \begin{pmatrix} 8.25 \\ 6.26 \\ 3.75 \end{pmatrix},$$

mientras que la matriz B tiene obviamente las entradas iguales a 1. Notemos que el tiempo esperado para la aparición de la palabra coincide (a nivel numérico) con el obtenido por simulación con el código Palabra.R.

En el caso de la rata y el laberinto de la Figura 1, se utilizó el código siguiente para determinar la matriz fundamental M y al vector t .

```
P=t(matrix(c(0,1/2,0,1/2,0,0,0,0,0,1/3,0,1/3,0,1/3,0,0,0,0,1/2,0,0,0,1/
  2,0,0,0,1/3,0,0,0,1/3,0,1/3,0,0,0,1/4,0,1/4,0,1/4,0,0,0,1/3,0,1/
  3,0,0,0,1/3,0,0,0,1/2,0,0,0,1/2,0,0,0,0,0,1/3,0,1/3,0,1/3,0,0,0,0,1/2,0,1
  /2,0),9) )
P[9,6]<-0
P[9,8]<-0
P[9,9]<-1 # Genera la matriz de transición (absorbente) para la rata en un
  laberinto

Q=P[1:8,1:8]
I=diag(1,8,8)
M=solve(I-Q)
uno=matrix(1,8,1)
t=M%%uno
```

LISTING 2.10. LaberintoFundamental.R

Se obtuvieron los siguientes resultados numéricos:

$$M = \begin{pmatrix} 3.00 & 3.00 & 1.50 & 3.00 & 3.00 & 1.50 & 1.50 & 1.50 \\ 2.00 & 3.62 & 1.75 & 2.37 & 3.00 & 1.62 & 1.25 & 1.37 \\ 1.50 & 2.62 & 2.50 & 1.87 & 2.50 & 1.87 & 1.00 & 1.12 \\ 2.00 & 2.37 & 1.25 & 3.62 & 3.00 & 1.37 & 1.75 & 1.62 \\ 1.50 & 2.25 & 1.25 & 2.25 & 3.50 & 1.50 & 1.25 & 1.50 \\ 1.00 & 1.62 & 1.25 & 1.37 & 2.00 & 2.12 & 0.75 & 0.87 \\ 1.50 & 1.87 & 1.00 & 2.62 & 2.50 & 1.12 & 2.50 & 1.87 \\ 1.00 & 1.37 & 0.75 & 1.62 & 2.00 & 0.87 & 1.25 & 2.12 \end{pmatrix} \quad \text{y} \quad t = \begin{pmatrix} 18.00 \\ 17.00 \\ 15.00 \\ 17.00 \\ 15.00 \\ 11.00 \\ 15.00 \\ 11.00 \end{pmatrix}$$

CAPÍTULO 3

Procesos de renovación

Retomaremos el ejemplo de contéos aleatorios del Capítulo 1 (Ejemplo 1.3).

Consideremos variables aleatorias independientes e idénticamente distribuidas S_1, S_2, \dots con valores estrictamente positivos. Podemos pensar en que S_i es el tiempo de vida de un componente crucial para el funcionamiento de cierto sistema y que al fallar se debe reemplazar. Los tiempo de reemplazo serán

$$T_0 = 0, \quad T_1 = S_1, \quad T_2 = S_1 + S_2, \dots$$

Por otra parte la cantidad de componentes que han fallado hasta el tiempo t sería

$$N_t = \min \{n : T_n > t\} = \sum_{n=0}^{\infty} n \mathbf{1}_{T_n \leq t < T_{n+1}}.$$

A este modelo general se le llama modelo de renovación. La sucesión S será la **sucesión de tiempos de vida**, la sucesión T la de **tiempos de renovación** y la sucesión N será el **proceso de contéo asociado**.

Hay un par de procesos adicionales que son útiles e interesantes: el proceso de **tiempo residual** (hasta el próximo reemplazo) es

$$R_t = T_{N_t+1} - t,$$

el **proceso de edad** es

$$A_t = t - T_{N_t}.$$

Su suma es el **proceso de tiempos totales**

$$L_t = S_{N_t+1}.$$

En la Figura 1 se ilustran las definiciones asociadas a los procesos de renovación.

Imaginemos ahora que en vez de cambiar al componente crucial en cuanto falla, se realiza una revisión diaria en la que se ve si se cambia o no. Entonces es más conveniente medir el tiempo en días en vez de continuamente. En términos del modelo, se puede imponer simplemente que el tiempo de vida S_i sea una variable aleatoria con valores en $\{1, 2, \dots\}$ y las definiciones tienen sentido como las hemos puesto, salvo que R_n, A_n y L_n , con $n \in \mathbb{N}$, toman valores en \mathbb{N} y determinan a los procesos R, A y L . En este caso hablamos de un proceso de renovación aritmético.

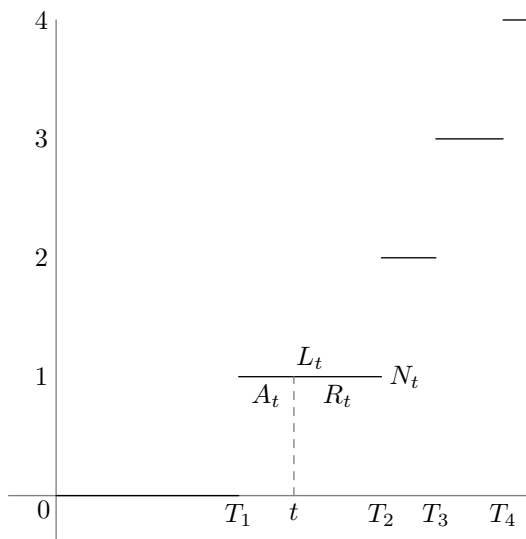


FIGURA 1. Ilustración de las definiciones de proceso de renovación

En este capítulo se hará una introducción a los procesos de renovación. Se enunciarán los resultados tanto en el caso general como en el aritmético, pero las justificaciones se harán en general en el caso aritmético.

EJEMPLO 3.1. Imaginemos la trayectoria de la rata en el laberinto de la Figura 1, que hemos modelado mediante una cadena de Markov. Supongamos que inicialmente le damos de comer a la rata en la casilla central y que, al terminar, la rata se va a dar un paseo aleatorio. Cuando regresa a la casilla central, encontrará la comida de nuevo dispuesta. En este caso, nuestros tiempos entre sucesos S_i serán la cantidad de pasos entre dos comidas de la rata y nuestros tiempos de reemplazo (o de renovación) serán los instantes de las visitas sucesivas de la rata a la casilla central. La variable R_n se puede interpretar como la cantidad de tiempo que le falta a la rata, después de n pasos, para volver a comer. La variable A_n es la cantidad de tiempo que lleva la rata sin comer al paso n mientras que L_n es la cantidad total de pasos que pasará la rata sin comer desde la última vez que comió anterior al paso n , hasta la siguiente vez que lo hará. No es completamente trivial verificar que la situación descrita corresponde a un fenómeno de renovación, pues no hemos discutido por qué los tiempos entre las visitas sucesivas de la rata a la casilla central conforman una sucesión de variables independientes e idénticamente distribuidas. Sin embargo, la propiedad de Markov fuerte nos permite ver que así es.

El tipo de preguntas a las que responde la teoría de renovación en este contexto son las siguientes: Al instante n : ¿cuántas veces ha comido la rata? ¿Qué pasa conforme $n \rightarrow \infty$? ¿Qué pasa en promedio? ¿Cuál es la distribución del tiempo que le falta para volver a comer (R_n)? ¿Qué le pasa a esta distribución conforme $n \rightarrow \infty$? ¿Cuál es la probabilidad de que al paso n la rata coma? ¿Qué pasa con dicha probabilidad conforme $n \rightarrow \infty$?

El ejemplo anterior es caso particular de uno mucho más general: si $0 = T_0 < T_1 < \dots$ son los instantes sucesivos en que una cadena de Markov recurrente visita a su estado inicial, que fijamos igual a x , entonces las variables $T_i - T_{i-1}$ son independientes e idénticamente distribuidas por lo que conforman un fenómeno de renovación. Si $X_0 = y \neq x$, entonces la distribución de S_1 es distinta a la de S_2, S_3, \dots , aunque aún así son iid. En este caso hablamos de un proceso de renovación demorado.

EJEMPLO 3.2 (El proceso Bernoulli). Se trata de un proceso de renovación aritmético en el cual los tiempos entre sucesos tienen distribución geométrica:

$$\mathbb{P}(S_i = k) = p(1-p)^{k-1}$$

para $k = 1, 2, \dots$. En este caso particular se pueden calcular las distribuciones de los procesos asociados al fenómeno de renovación. Además, admite una interpretación adicional: sean B_1, B_2, \dots variables aleatorias Bernoulli de parámetro p y sean

$$T_0 = 0 \quad \text{y} \quad T_{n+1} = \min \{i > T_n : B_i = 1\}.$$

Entonces T es un proceso Bernoulli en el sentido de que la sucesión $(T_{n+1} - T_n)$ es iid con distribución geométrica (concentrada en $\{1, 2, \dots\}$) de parámetro p .

Comencemos con la distribución de T_n : de acuerdo a nuestra interpretación, T_n es el tiempo en el que ocurre el n -ésimo éxito de la sucesión Bernoulli B , por lo que T_n es binomial negativa de parámetros n y p con valores en $\{n, n+1, \dots\}$, es decir:

$$\mathbb{P}(T_n = k) = \binom{k-1}{k-n} p^n (1-p)^{k-n}.$$

Al proceso de conteo asociado también se le puede calcular la distribución exacta: notemos que N_n es la cantidad de unos en la sucesión B_1, \dots, B_n , y como estas variables toman los valores cero y uno, pues $N_n = \sum_{i=1}^n B_i$. Así, N_n tiene distribución binomial de parámetros n y p . Esto además implica que $\mathbb{E}(N_n) = np$ y por lo tanto $\mathbb{E}(N_n/n) \rightarrow p$ conforme $n \rightarrow \infty$, que es un caso particular del teorema de renovación elemental que demostraremos más adelante.

Respecto al proceso de tiempos resiguales, notemos que

$$\{R_n = k\} = \{B_{n+1} = 0, \dots, B_{n+k-1} = 0, B_{n+k} = 1\}$$

de donde concluimos que

$$\mathbb{P}(R_n = k) = p(1-p)^{k-1} \quad k = 1, 2, \dots$$

En otras palabras, R_n es geométrica de parámetro p con valores en $\{1, 2, \dots\}$.

Un argumento similar funciona para el proceso de edad pues

$$\{A_n = k\} = \{B_n = 0, B_{n-1} = 0, \dots, B_{n-k+1} = 0, B_{n-k} = 1\}$$

por lo que

$$\mathbb{P}(A_n = k) = p(1-p)^k \quad k = 0, 1, \dots$$

En otras palabras, A_n es geométrica de parámetro p con valores $0, 1, 2, \dots$.

De hecho, las variables A_n y R_n son independientes pues al ser las B_i independientes vemos que

$$\begin{aligned} \mathbb{P}(A_n = j, R_n = k) &= \mathbb{P}(B_n = 0, \dots, B_{n-j-1} = 0, B_{n-j} = 1, B_{n+1} = 0, \dots, B_{n+k-1} = 0, B_{n+k} = 1) \\ &= \mathbb{P}(B_n = 0, \dots, B_{n-j-1} = 0, B_{n-j} = 1) \mathbb{P}(B_{n+1} = 0, \dots, B_{n+k-1} = 0, B_{n+k} = 1) \\ &= \mathbb{P}(A_n = j) \mathbb{P}(R_n = k) \end{aligned}$$

Finalmente, la distribución del proceso de tiempos totales se puede entonces calcular; recordemos que la suma de dos geométricas independientes es binomial negativa. Sólo hay que tener cuidado pues este cálculo asume que ambas variables geométricas toman valores en el mismo conjunto. El resultado es que

$$\mathbb{P}(L_n = m) = mp^2(1-p)^{m-1}. \quad m = 1, 2, \dots$$

A continuación exploraremos otra liga entre cadenas de Markov y procesos de renovación.

PROPOSICIÓN 3.1. *En procesos de renovación aritméticos, el proceso de tiempos residuales es una cadena de Markov. Si M es el supremo del soporte de la distribución de S_1 , la matriz de transición de R es irreducible en $\{i \in \mathbb{N} : i \leq M\}$. R es aperiódica si $\{n \geq 1 : \mathbb{P}(S_1 = n) > 0\} \not\subset h\mathbb{N}$ para toda $h \geq 2$.*

Un proceso de renovación aritmético aperiódico es aquel para el cual se satisface la anterior condición para que R sea aperiódica.

DEMOSTRACIÓN. Consideremos a la función

$$f(i, r) = \begin{cases} i-1 & i > 1 \\ r & i = 1 \end{cases}.$$

Al definir a la sucesión \tilde{R} mediante

$$\tilde{R}_0 = 1 \quad \text{y} \quad \tilde{R}_{n+1} = f(\tilde{R}_n, S_{n+1}),$$

vemos que \tilde{R} es una cadena de Markov con matriz de transición

$$P_{i,j} = \begin{cases} 1 & i > 1, j = i-1 \\ \mathbb{P}(S_1 = j) & i = 1 \end{cases}.$$

Sin embargo, puesto que la sucesión S es iid, R y \tilde{R} tienen la mismas distribuciones finito-dimensionales en el sentido siguiente:

$$\mathbb{P}(R_0 = i_0, \dots, R_n = i_n) = \mathbb{P}(\tilde{R}_0 = i_0, \dots, \tilde{R}_n = i_n).$$

En efecto, si $i_0, \dots, i_n \in \{1, 2, \dots\}$ e $i_l = 1$ si y sólo si $l \in I \subset \{0, \dots, n\}$ e $i_l = i_{l-1} - 1$ si $l \notin I$ y digamos que $I = \{l_1, \dots, l_m\}$ entonces

$$\begin{aligned} \mathbb{P}(\tilde{R}_0 = i_0, \dots, \tilde{R}_n = i_n) &= \mathbb{P}(S_{i_{l_k}} = i_{l_k}, k \leq m) \\ &= \mathbb{P}(S_k = i_{l_k}, k \leq m) \\ &= \mathbb{P}(R_0 = i_0, \dots, R_n = i_n). \end{aligned}$$

Esto prueba que R es una cadena de Markov con matriz de transición P .

Si M es el supremo del soporte de la distribución de S_1 y $M < \infty$, entonces $\mathbb{P}(S_1 = M) > 0$ y $\mathbb{P}(S_1 = n) = 0$ para toda $n \geq M$. Entonces de 1 se accede a M , a partir del cual se accede a $M - 1, \dots, 1$, por lo que $\{0, \dots, M\}$ es una clase de comunicación, que de hecho es cerrada pues $\mathbb{P}(S_1 = n) = 0$ para toda $n \geq M$. Esto hace que la cadena en $\{1, \dots, M\}$ sea irreducible. Si $M = \infty$, entonces para toda M existe $n \geq M$ tal que $\mathbb{P}(S_1 = n) > 0$, por lo que 0 se comunica con M vía n . Es decir, la cadena es irreducible en $\{0, 1, \dots\}$.

Si $\{n \geq 1 : \mathbb{P}(S_1 = n) > 0\} \not\subset h\mathbb{N}$ para $h \geq 2$ entonces existen n_1, n_2, k_1 y k_2 naturales tal que $n_1 k_1 = 1 + n_2 k_2$ y tales que $\mathbb{P}(S_1 = n_1), \mathbb{P}(S_1 = n_2) > 0$. Vemos entonces que es posible ir de cero a cero en $n_1 k_1 = n_2 k_2 + 1$ pasos y en $n_2 k_2$ pasos, por lo que el periodo de 1 es 1 y por lo tanto la cadena es aperiódica. \square

La respuesta a cuántas veces a comido la rata al paso n se puede interpretar en términos de N_n en el caso de procesos de renovación aritméticos o de N_t en el caso general. Los siguientes dos resultados nos permiten dar una posible respuesta: para tiempos grandes, la variable N_t se puede predecir determinísticamente.

PROPOSICIÓN 3.2 (Ley fuerte de los grandes números). *Si $\mu = \mathbb{E}(S_i) < \infty$ entonces $N_t/t \rightarrow 1/\mu$ casi seguramente*

DEMOSTRACIÓN. Por la ley fuerte de los grandes números, sabemos que $T_n/n \rightarrow \mu$ casi seguramente. Notemos que si $t \in [T_n, T_{n+1})$ entonces

$$\frac{T_n}{n} \frac{n}{n+1} \leq \frac{t}{N_t} \leq \frac{T_{n+1}}{n+1} \frac{n+1}{n}.$$

Los extremos de la desigualdad convergen al mismo límite, μ , conforme $n \rightarrow \infty$ de manera casi segura. Por lo tanto $N_t/t \rightarrow 1/\mu$. \square

Nuestro siguiente resultado involucra a la llamada función de renovación. Es la función $m : [0, \infty) \rightarrow [0, \infty)$ dada por $m(t) = \mathbb{E}(N_t)$.

PROPOSICIÓN 3.3 (Teorema de renovación elemental). *Si $\mu = \mathbb{E}(S_1) < \infty$ entonces $m(t)/t \rightarrow 1/\mu$.*

La demostración utilizará un resultado sobre camintas aleatorias conocido como la identidad de Wald y que se prueba como una aplicación de la teoría de martingalas. Se trata de una generalización de la fórmula $\mathbb{E}(T_n) = n\mathbb{E}(T_1)$ al caso en que n ya no es una cantidad fija sino aleatoria. En nuestro contexto, veremos que

$$(4) \quad \mathbb{E}(T_{N_t+1}) = \mathbb{E}(N_t + 1) \mathbb{E}(T_1) = \mathbb{E}(N_t + 1) \mu.$$

La demostración en este caso particular es la siguiente: notamos que

$$\{N_t + 1 = n\} = \{N_t = n - 1\} = \{T_{n-1} \leq t < T_n\}.$$

Se sigue que $\{N_t + 1 > n\}$ pertenece a la σ -álgebra generada por S_1, \dots, S_n y es por lo tanto independiente de S_{n+1} . Al aplicar el teorema de Tonelli (dos veces):

$$\begin{aligned} \mathbb{E}(T_{N_t+1}) &= \mathbb{E}\left(\sum_i \mathbf{1}_{i=N_t+1} T_i\right) \\ &= \mathbb{E}\left(\sum_i \sum_{j \leq i} \mathbf{1}_{i=N_t+1} S_j\right) \\ &= \mathbb{E}\left(\sum_j \sum_{i \geq j} \mathbf{1}_{i=N_t+1} S_j\right) \\ &= \mathbb{E}\left(\sum_j \mathbf{1}_{N_t+1 \geq j} S_j\right) \\ &= \sum_j \mathbb{E}(S_j) \mathbb{P}(N_t + 1 \geq j) \\ &= \mu \sum_j \mathbb{P}(N_t + 1 \geq j) \\ &= \mu \mathbb{E}(N_t + 1). \end{aligned}$$

PRUEBA DEL TEOREMA DE RENOVACIÓN ELEMENTAL. Ocupémonos primero de obtener una cota inferior. A partir de la identidad de Wald, vemos que

$$t \leq \mathbb{E}(T_{N_t+1}) \leq \mathbb{E}(N_t + 1) \mu.$$

Por lo tanto

$$\frac{1}{\mu} - \frac{1}{t} \leq \frac{\mathbb{E}(N_t)}{t}.$$

Notemos que

$$\lim_{t \rightarrow \infty} \frac{1}{\mu} - \frac{1}{t} = \frac{1}{\mu}.$$

Para obtener una cota superior, definimos a $\tilde{S}_i = S_i \wedge b$. Con la notación obvia, notemos que $\tilde{T}_i \leq T_i$ y que por lo tanto $N_t \leq \tilde{N}_t$. Además,

$$\mathbb{E}(\tilde{S}_1 \wedge b) \mathbb{E}(N_t + 1) = \mathbb{E}(\tilde{T}_{\tilde{N}_t+1}) \leq t + b,$$

por lo que

$$\frac{\mathbb{E}(N_t)}{t} \leq \frac{t + b}{t\mathbb{E}(S_1 \wedge b)}.$$

Al utilizar $b = \sqrt{t}$, el teorema de convergencia monótona nos dice que

$$\mathbb{E}(S_1 \wedge \sqrt{t}) \rightarrow \mu,$$

por lo que

$$\lim_{t \rightarrow \infty} \frac{t + \sqrt{t}}{t\mathbb{E}(S_1 \wedge \sqrt{t})} = \frac{1}{\mu}.$$

Podemos entonces deducir que

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}(N_t)}{t} = \frac{1}{\mu}.$$

□

1. La medida de renovación en el caso aritmético

Continuaremos el estudio de procesos de renovación al introducir a la medida de renovación. En realidad se introducirá a partir de la función u dada por

$$u_n = \mathbb{P}(\exists m, T_m = n) = \sum_m \mathbb{P}(T_m = n).$$

En términos de la rata en el laberinto, u_n es la probabilidad de que la rata coma al instante n . El comportamiento asintótico de u_n se conoce y es el objeto del teorema de renovación para procesos de renovación aritméticos de Erdős, Feller y Pollard.

TEOREMA 3.1. *Para un proceso de renovación aritmético, aperiódico y con media finita:*

$$\lim_{n \rightarrow \infty} u_n \rightarrow \frac{1}{\mu}.$$

DEMOSTRACIÓN. Primero probemos que el proceso de tiempos residuales R tiene una distribución invariante. En efecto, sean

$$\pi_i = \mathbb{P}(S_1 \geq i) / \mu$$

y P la matriz de transición de R . Entonces

$$\begin{aligned}
 (\pi P)_j &= \sum_i \pi_i P_{i,j} \\
 &= \pi_{j+1} P_{j+1,j} + \pi_1 P_{1,j} \\
 &= \frac{\mathbb{P}(S_1 \geq j+1)}{\mu} P_{j+1,j} + \frac{1}{\mu} \mathbb{P}(S_1 = j) \\
 &= \frac{\mathbb{P}(S_1 \geq j)}{\mu} = \pi_j.
 \end{aligned}$$

Al ser la cadena irreducible y aperiódica y con una distribución invariante, es positivo recurrente y se puede aplicar el teorema fundamental de convergencia para concluir que

$$\lim_{n \rightarrow \infty} P_{i,1}^n = \frac{1}{\mu}.$$

Por otra parte, podemos calcular explícitamente $P_{1,1}^n$ pues

$$P_{1,1}^n = \mathbb{P}(R_n = 1) = \mathbb{P}(n = T_m \text{ para alguna } m) = u_n.$$

Así, vemos que $u_n \rightarrow 1/\mu$ conforme $n \rightarrow \infty$. □

2. La medida de renovación en el caso continuo

Ahora presentaremos el resultado análogo al teorema de renovación aritméticos en el caso de procesos no aritméticos. No se abordará la prueba pues depende de resultados análogos al teorema fundamental de convergencia pero para cadenas de Markov a tiempo discreto y espacio continuo (en este caso, el espacio de estados del proceso de tiempos residuales es $(0, \infty)$).

Sea S un proceso de renovación no-aritmético, es decir una sucesión de variables aleatorias independientes e idénticamente distribuidas con valores en $(0, \infty)$ y tal que no toman valores en $\delta\mathbb{N}$. Se puede pensar que S_i es una variable aleatoria con densidad. Sea T la sucesión de tiempos entre sucesos y N el proceso de conteo asociado.

DEFINICIÓN. La **medida de renovación** es la medida U en $[0, \infty)$ tal que

$$U([0, t]) = \mathbb{E}(N_t) = \sum_{n=1}^{\infty} \mathbb{P}(T_n \leq t).$$

Se utilizará la notación U_t para $U([0, t])$. En general, para procesos de renovación no aritméticos, la cantidad $\mathbb{P}(\exists n, T_n = t)$ es cero para casi toda t . Es por eso que se utiliza la medida de renovación en vez de la función de renovación del caso aritmético. En efecto, si $[a, b]$ es un intervalo muy pequeño, puede que haya un punto T_n en dicho intervalo, pero sería raro que hubiera más de uno. Por lo tanto $U([a, b]) \approx \mathbb{P}(\exists n, T_n \in [a, b])$. El siguiente teorema puede entonces interpretarse como es una extensión del teorema de renovación para procesos aritméticos.

TEOREMA 3.2 (Teorema de renovación de Blackwell). *Para todo $h > 0$:*

$$\lim_{t \rightarrow \infty} U_{t+h} - U_t = \frac{h}{\mu}.$$

3. La ecuación de renovación

La ecuación de renovación aparece en el cálculo de algunas cantidades de la teoría de renovación. En general se trata de un análisis al primero instante de renovación al notar que el proceso T' dado por $0, T_2 - T_1, T_3 - T_1, \dots$ es un proceso de renovación idéntico en distribución a T_0, T_1, T_2, \dots e independiente de T_1 . Ejemplos de su utilidad es que nos permite estudiar el comportamiento asintótico de las distribuciones de los procesos de tiempo residual, de edad ó de tiempos totales. Sin embargo, también se puede entender como una relación de recurrencia para cantidades de interés en teoría de renovación. Nos concentraremos en el caso aritmético, al ser técnicamente menos complicado, al comenzar con algunos ejemplos. Escribiremos

$$p_k = p(k) = \mathbb{P}(S_1 = k).$$

EJEMPLO 3.3. Sea u_k la densidad de renovación para un proceso de renovación aritmético:

$$u_k = \mathbb{P}(\exists n \geq 0 \text{ tal que } T_n = k).$$

Escribiremos a dicha cantidad como $u(k)$ cuando la tipografía lo requiera.

Notemos que $u_0 = 1$, mientras que para $k \geq 1$, claramente $\mathbb{P}(T_0 = k) = 0$ y por lo tanto:

$$\begin{aligned} u_k &= \mathbb{P}(\exists n \geq 1 \text{ tal que } T_n = k) \\ &= \sum_{j=1}^k \mathbb{P}(S_1 = j, \exists n \geq 0 \text{ tal que } T_n - j = k - j) \\ &= \sum_{j=1}^k \mathbb{P}(S_1 = j, \exists n \geq 0 \text{ tal que } T'_n = k - j) \\ &= \sum_{j=1}^k p_j \mathbb{P}(\exists n \geq 0 \text{ tal que } T_n = k - j) \\ &= \sum_{j=1}^k p_j u_{k-j} \\ &= \mathbb{E}(u(k - S_1)), \end{aligned}$$

donde por supuesto se utiliza el hecho que por definición $u_j = 0$ si $j < 0$.

La ecuación

$$u_k = \delta_0(k) + \mathbb{E}(u(k - S_1))$$

que también se escribe

$$u_k = \delta_0(k) + \sum p_j u_{k-j}$$

es la llamada ecuación de renovación para la densidad de renovación.

EJEMPLO 3.4. Sea $L_n, n \geq 0$ el proceso de tiempos totales de un proceso de renovación aritmético. Para $r \geq 0$, sea $z(n) = \mathbb{P}(L_n = r)$ para $n \geq 0$ y $z(n) = 0$ si $n < 0$. Notemos que si $n \leq r$ entonces

$$z(n) = \mathbb{P}(S_1 = r) = p_r.$$

Por otra parte, si $n > r$, nos fijamos en la primera renovación $T_1 = S_1$, que si $L_n = r$ es necesariamente es más chica que n . Entonces se obtiene:

$$z(n) = \sum_{j \leq n} \mathbb{P}(S_1 = j, L_{n-j}(T') = r) = \sum_{j \leq n} p_j z(n-j).$$

La ecuación de renovación resultante es

$$z(n) = \mathbf{1}_{n \leq r} p_r + \sum_j p_j z(n-j).$$

EJEMPLO 3.5. Sea R_n el proceso de tiempos residuales de un proceso de renovación aritmético. Sea $z(n) = \mathbb{P}(R_n = r)$ si $n \geq 0$ y $z(n) = 0$ si $n < 0$. El cálculo de R_n se puede dividir en dos casos, dependiendo de si $S_1 \leq n$ o no:

$$z(n) = \mathbb{P}(R_n = r) = p_{n+r} + \sum_{j=1}^n \mathbb{P}(S_1 = r) z(n-j).$$

En general, la ecuación de renovación es la siguiente: dada una función b (tal que $b_n = 0$ si $n < 0$) se busca una función z tal que $z(n) = 0$ si $n < 0$ y

$$z(n) = b(n) + \sum_{j \leq n} p_k z(n-j).$$

La solución se puede encontrar al iterar la ecuación de renovación para escribir:

$$\begin{aligned} z(n) &= b(n) + \mathbb{E}(z(n - S_1)) \\ &= b(n) + \mathbb{E}(b(n - S_1)) + \mathbb{E}(b(n - S_1 - S_2)) \\ &= b(n) + \mathbb{E}(b(n - S_1)) + \mathbb{E}(b(n - S_1 - S_2)) + \mathbb{E}(b(n - S_1 - S_2 - S_3)) = \dots \end{aligned}$$

Puesto que $b(n - T_m) = 0$ si $m \geq n$, vemos que

$$z(n) = \sum_m \mathbb{E}(b(n - T_m))$$

y al sumar sobre los posibles valores de T_m , se obtiene

$$z(n) = \sum_m \sum_x b(n-x) \mathbb{P}(T_m = x) = \sum_x u_x b_{n-x}.$$

Una vez establecida la relación entre soluciones a la ecuación de renovación y la densidad de renovación, se tienen los elementos clave para probar el siguiente resultado.

TEOREMA 3.3 (Teorema clave de renovación en el caso discreto). *Si z es solución a la ecuación de renovación*

$$z(n) = b(n) + \sum_{j \leq n} p_j z(n-j).$$

y b es sumable entonces z está dada por $z(n) = \sum_x b(x) u(n-x)$ y $z(n) \rightarrow \sum_x b(x) / \mu$ conforme $n \rightarrow \infty$.

DEMOSTRACIÓN. Sólo hace falta verificar el comportamiento asintótico de z . Sin embargo, si b es sumable, puesto que u_n es una probabilidad, se tiene la cota

$$\sum_x b(x) u(n-x) \leq \sum_x b(x) < \infty$$

y por lo tanto el teorema de convergencia dominada y el teorema de renovación de Erdős-Feller-Pollard (EFP) nos dicen que

$$\lim_{n \rightarrow \infty} z(n) = \lim_{n \rightarrow \infty} \sum_x b(x) u(n-x) = \sum_x b(x) / \mu.$$

□

Veamos ahora algunas aplicaciones del teorema de renovación clave. En el caso del tiempo total asintótico, vemos que

$$\lim_{n \rightarrow \infty} \mathbb{P}(R_n = r) = \sum_x p_{r+x} / \mu = \mathbb{P}(S_1 \geq r) / \mu,$$

aunque en realidad esto ya lo sabíamos y lo utilizamos en la prueba del teorema de renovación de EFP.

Un ejemplo más interesante es el del tiempo total:

$$\lim_{n \rightarrow \infty} \mathbb{P}(L_n = r) = \sum_{x \leq r} p_x / \mu = r p_r / \mu.$$

CAPÍTULO 4

Procesos de Poisson

Imaginemos la siguiente situación: al tiempo cero (inicial) una serie de personas contratan un seguro por el que pagan una prima. La compañía de seguros tienen entonces la obligación de dar una compensación económica a los contratantes en caso de que suceda un cierto evento llamado siniestro. En muchos casos, el monto que debe pagar la compañía por el siniestro también es aleatorio.

El modelo matemático se realiza en dos etapas: primero nos enfocamos en los instantes en que suceden los siniestros y luego en sus montos. Si $0 = T_0 < T_1 < T_2 < \dots$ son los instantes en que acaecen los siniestros, se introduce al proceso de contéo asociado $N = (N_t, t \geq 0)$, donde N_t nos dice cuántos siniestros han sucedido al instante t , mismo que está caracterizado por satisfacer las identidades

$$N_t = \sum_n n \mathbf{1}_{T_n \leq t < T_{n+1}},$$
$$\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$$

y

$$\{N_t \geq n\} = \{T_n \leq t\}.$$

En general, un **proceso de contéo** es un proceso con trayectorias constantes por pedazos que toma valores enteros y va incrementando de uno en uno. A $S_i = T_i - T_{i-1}$ se le llama i -ésimo tiempo interarribo. Es más natural imponer supuestos sobre N que sobre los tiempos T_i . Los supuestos se pueden traducir como sigue:

Incrementos estacionarios: La distribución de $N_{t+s} - N_t$ sólo depende de s (en otras palabras, si los siniestros se presentan homogéneamente en el tiempo, se tiene la igualdad en distribución; en particular no hay fenómenos estacionales).

Incrementos independientes: Si $0 = t_0 \leq t_1 \leq \dots \leq t_n$, las variables $N_{t_1} - N_{t_0}, \dots, N_{t_n} - N_{t_{n-1}}$ son independientes. (Así, el que ocurran siniestros por la mañana no afecta lo que ocurre por la tarde.

DEFINICIÓN. Un **proceso de Poisson** es un proceso de contéo con incrementos independientes y estacionarios.

Esta definición captura la idea de que el riesgo de que ocurra un siniestro se distribuye homogéneamente tanto en el tiempo como en la población. El primer objetivo de este capítulo es mostrar como una tal definición tiene implicaciones prácticas y teóricas que hacen que podamos hacer cálculos con el proceso de Poisson que no es posible hacer explícitamente con otros modelos. En efecto, hay una definición alternativa de proceso Poisson que automáticamente nos deja hacer cálculos.

TEOREMA 4.1. *Un proceso de contéo N es de Poisson si y sólo si existe $\lambda > 0$ tal que los tiempos interarribo son variables exponenciales independientes de parámetro λ .*

Al parámetro λ se le conoce como intensidad del proceso de Poisson. Antes de probar el teorema, veámos algunas de sus consecuencias. La primera es que N_t es el proceso de contéo asociado a un proceso de renovación con tiempos interarribo exponenciales. Por lo tanto, se satisface la ley fuerte de los grandes números en particular. Además, se pueden calcular explícitamente las distribuciones al tiempo t del proceso de edad, de tiempos residuales y de tiempos totales, aún en el contexto de procesos a tiempo continuo.

Una segunda consecuencia es que se conoce la distribución exacta del proceso de contéo.

PROPOSICIÓN 4.1. *Sea N un proceso de Poisson de parámetro λ . Entonces N_t tiene distribución Poisson de parámetro λt .*

DEMOSTRACIÓN. Puesto que los tiempos interarribo son exponenciales independientes de parámetro λ , entonces T_n tiene distribución Γ de parámetros λ y n y es independiente de S_{n+1} . Así, vemos que la densidad conjunta de (T_n, S_{n+1}) está dada por

$$f_{T_n, S_{n+1}}(t, s) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{n-1!} \frac{e^{-\lambda s}}{\lambda}.$$

Por lo tanto:

$$\begin{aligned} \mathbb{P}(N_t = n) &= \mathbb{P}(T_n \leq t < T_n + S_{n+1}) \\ &= \int_0^t \int_{t-t_1}^{\infty} f_{T_n, S_{n+1}}(t_1, s) \, ds \, dt_1 \\ &= \int_0^t \int_{t-t_1}^{\infty} \frac{\lambda^n t_1^{n-1} e^{-\lambda t_1}}{n-1!} \lambda e^{-\lambda s} \, ds \, dt_1 \\ &= \int_0^t \frac{\lambda^n t_1^{n-1} e^{-\lambda t_1}}{n-1!} e^{-\lambda(t-t_1)} \, dt_1 \\ &= e^{-\lambda t} \frac{(\lambda t)^n}{n!} \end{aligned}$$

□

Como corolario, podemos calcular las distribuciones finito dimensionales de N : sean $0 = t_0 < t_1 < \dots < t_m$ y $n_0 = 0 \leq n_1 \leq \dots \leq n_m$ por lo que

$$\begin{aligned} \mathbb{P}(N_{t_1} = n_1, \dots, N_{t_m} = n_m) \\ = \mathbb{P}(N_{t_1} - N_{t_0} = n_1 - n_0, \dots, N_{t_m} - N_{t_{m-1}} = n_m - n_{m-1}) \end{aligned}$$

y al utilizar la independencia de incrementos

$$= \prod_{i=1}^m \mathbb{P}(N_{t_i} - N_{t_{i-1}} = n_i - n_{i-1})$$

así como su estacionariedad

$$\begin{aligned} &= \prod_{i=1}^m \mathbb{P}(N_{t_i - t_{i-1}} = n_i - n_{i-1}) \\ &= \prod_{i=1}^m \frac{e^{-\lambda(t_i - t_{i-1})} (\lambda(t_i - t_{i-1}))^{n_i - n_{i-1}}}{(n_i - n_{i-1})!}. \end{aligned}$$

Ya con esto, podemos hacer un primer código para simular al proceso de Poisson. De hecho se trata simplemente del código Poisson.R que ya se había introducido: se trata de generar variables exponenciales e ir las sumando hasta que se sobrepase el tiempo t :

```
lambda=10 # Intensidad
xi=rexp(1,lambda) # xi representa el tiempo del primer siniestro
T=c(0,xi) # El vector T irá acumulando los tiempos en que van
            ocurriendo siniestros
N=0 # N nos dirá cuantos siniestros ocurren hasta el tiempo 1
while(xi<1){ # Mientras no se haya sobrepasado el instante 1
  N<-N+1 # Aumentamos el número de siniestros
  xi<-xi+rexp(1,lambda) # Vemos el tiempo en el que ocurre el siguiente
                        siniestro
  T=c(T,xi) # Aumentamos un evento temporal
}
plot(T,c(1:(N+2)))
```

LISTING 4.1. Poisson2.R

Para ver otro esquema de simulación para el proceso de Poisson, en el que el ciclo while se substituye por un ciclo for, haremos algunos cálculos adicionales. La primera pregunta es: si sabemos que han ocurrido n siniestros en $[0, t]$, ¿en dónde han caído estos saltos? Para responder a dicha pregunta, calcularemos la densidad

conjunta de T_1, \dots, T_n dado que $N_t = n$. Puesto que

$$\begin{aligned} \mathbb{P}(T_1 \in dt_1, \dots, T_n \in dt_n, N_t = n) &= \mathbb{P}(T_1 \in dt_1, \dots, T_n \in dt_n, t_n \leq t \leq t_n + S_{n+1}) \\ &= \prod_{i=1}^n \lambda e^{-\lambda(t_i - t_{i-1})} e^{-\lambda(t - t_n)} \end{aligned}$$

vemos que la densidad condicional buscada es:

$$\begin{aligned} f_{T_1, \dots, T_n | N_t = n}(t_1, \dots, t_n) &= \prod_{i=1}^n \lambda e^{-\lambda(t_i - t_{i-1})} e^{-\lambda(t - t_n)} \frac{n!}{(\lambda t)^n e^{-\lambda t}} \\ &= \frac{n!}{t^n}. \end{aligned}$$

Se concluye que condicionalmente a $N_t = n$, las variables T_1, \dots, T_n los valores ordenados de n uniformes independientes en $[0, t]$. Por lo tanto, hay otra manera de simular la trayectoria de un proceso de Poisson en el intervalo de tiempo $[0, t]$:

```
l<-10          % Intensidad
t<-3           % Umbral de tiempo
N<-rpois(1,l*t) % Cantidad de siniestros
u<-runif(N)*t   % Tiempos en que ocurren (desordenados)
u<-sort(u)      % Tiempos de siniestros ordenados
plot(u)
```

LISTING 4.2. PoissonConUniformes.R

Antes de continuar con el modelo completo de reclamaciones en una compañía de seguros, veremos algunos procesos estocásticos importantes asociados al proceso de Poisson. Se definirá a

$$\mathcal{F}_t = \sigma(X_s : s \leq t).$$

EJEMPLO 4.1 (El proceso de Poisson compensado). El proceso de Poisson compensado es el proceso $M_t = N_t - \lambda t$. Este proceso satisface la siguiente propiedad: si $s \leq t$

$$\mathbb{E}(M_t \mid \mathcal{F}_s) = M_s.$$

En efecto, al notar que $M_t = M_t - M_s + M_s$ vemos que

$$\mathbb{E}(M_t \mid \mathcal{F}_s) = \mathbb{E}(M_t - M_s \mid \mathcal{F}_s) + \mathbb{E}(M_s \mid \mathcal{F}_s) = \mathbb{E}(M_t - M_s) + M_s = M_s$$

puesto que $M_t - M_s$ tiene media cero y es independiente de \mathcal{F}_s .

La propiedad anterior nos dice que M es una martingala. Las martingalas son fundamentales en el desarrollo de la probabilidad moderna.

EJEMPLO 4.2 (La martingala exponencial). Consideremos al proceso

$$E_t = e^{-qN_t + t\lambda(1 - e^{-q})}.$$

Este proceso también es una martingala. Para verificarlo, primero calculemos

$$\mathbb{E}(e^{-qN_t}) = \sum_{n=0}^{\infty} e^{-qn} \frac{e^{-\lambda t} (\lambda t)^n}{n!} = e^{-\lambda t} e^{\lambda t e^{-q}} = e^{-t\lambda(1-e^{-q})}.$$

Luego utilizamos la independencia entre $N_t - N_s$ y \mathcal{F}_s para obtener:

$$\begin{aligned} \mathbb{E}(E_t \mid \mathcal{F}_s) &= e^{t\lambda(1-e^{-q})} e^{-qN_s} \mathbb{E}\left(e^{-q(N_t-N_s)} \mid \mathcal{F}_s\right) \\ &= e^{t\lambda(1-e^{-q})} e^{-qN_s} e^{-(t-s)\lambda(1-e^{-q})} \\ &= E_s. \end{aligned}$$

Una consecuencia del cálculo de la transformada de Laplace de N_t es que podemos calcular la esperanza y la varianza de N_t : si definimos

$$\varphi_t(q) = \mathbb{E}(e^{-qN_t}) = e^{-t\lambda(1-e^{-q})},$$

y derivamos bajo el signo integral cuando $q > 0$, se obtiene

$$\varphi_t(q) (-\lambda t e^{-q}) = \frac{\partial \varphi_t(q)}{\partial q} = -\mathbb{E}(N_t e^{-qN_t}).$$

Luego el teorema de convergencia monótona nos permite ver que

$$\lambda t = \mathbb{E}(N_t).$$

Al derivar dos veces se tiene que para $q > 0$:

$$\varphi_t(q) (\lambda t e^{-q})^2 + \varphi_t(q) \lambda t e^{-q} = \frac{\partial^2 \varphi_t(q)}{\partial q^2} = \mathbb{E}(N_t^2 e^{-qN_t}),$$

por lo que el teorema de convergencia monótona nos dice que

$$\mathbb{E}(N_t^2) = (\lambda t)^2 + \lambda t.$$

Finalmente, se obtiene

$$\text{Var}(N_t) = \lambda t.$$

Continuaremos ahora con el modelo de las reclamaciones en una compañía de seguros: es natural que los montos de los siniestros son independientes entre si y de los tiempos en que acontecen. Por lo tanto se plantéa el siguiente modelo completo:

- (1) Los tiempos en que acontecen los siniestros tienen como proceso de contéo asociado a un proceso de Poisson N . Sea λ su intensidad.
- (2) Los montos de los siniestros son variables aleatorias independientes e idénticamente distribuidas ξ_1, ξ_2, \dots

Sea $S_n = \xi_1 + \dots + \xi_n$ la sucesión de sumas parciales asociada a los montos de los siniestros. Entonces, al tiempo t , el monto total que ha sido reclamado a la compañía de seguros es

$$X_t = S_{N_t} = \sum_{i \leq N_t} \xi_i.$$

Al proceso X se le conoce con el nombre de **proceso de Poisson compuesto**. A la distribución de ξ_i se le conoce como distribución de salto de X y al parámetro de N se le conoce como intensidad de X . El siguiente código permite simular al proceso de Poisson compuesto cuando la distribución de salto es la de 100,000 veces una Beta de parámetros 1/2 y 1/2. La idea es modelar una situación en la que es más probable que los montos de los siniestros sean o muy pequeños o muy grandes (hasta el tope de 100,000 que podría ser la suma asegurada). El resultado se puede apreciar en la Figura 1

```
lambda<-10 # Intensidad
s<-rexp(1,lambda) # s representa el tiempo del primer siniestro
x<-100000*rbeta(1,1/2,1/2) # x representa el monto del primer siniestro
T<-c(0,s) # El vector T irá acumulando los tiempos en que van
            ocurriendo siniestros
X<-c(0,x) # X irá acumulando los montos de los siniestros
N<-0 # N nos dirá cuantos siniestros ocurren hasta el tiempo
      1
while(s<1){ # Mientras no se haya sobrepasado el instante 1
  N<-N+1 # Aumentamos el número de siniestros
  s<-s+rexp(1,lambda) # Vemos el tiempo en el que ocurre el siguiente
                      siniestro
  x<-100000*rbeta(1,1/2,1/2) # Calculamos su monto
  T<-c(T,s) # Aumentamos un evento temporal
  X<-c(X,tail(X,1)+x) # Agregamos el monto de la reclamación
}
plot(T,X)
```

LISTING 4.3. PoissonCompuesto.R

Se pueden hacer cálculos paralelos a los del proceso de Poisson en el caso Poisson compuesto. Por ejemplo, calculemos la media y varianza de X_t : si $\mu = \mathbb{E}(S_1)$ y $\sigma^2 = \text{Var}(S_1)$ son finitas entonces

$$\mathbb{E}(X_t^2) = \sum_{n=0}^{\infty} \mathbb{E}(S_n^2 \mathbf{1}_{(N_t=n)}) = \sum_{n=0}^{\infty} (n\sigma^2 + n^2\mu^2) \mathbb{P}(N_t = n) = \mathbb{E}(\sigma^2 N_t + \mu^2 N_t^2) < \infty,$$

ya que las variables Poisson tienen momentos de cualquier orden. La media está dada por

$$\mathbb{E}(X_t) = \sum_{n=0}^{\infty} \mathbb{E}(S_n \mathbf{1}_{(N_t=n)}) = \sum_{n=0}^{\infty} n\mu \mathbb{P}(N_t = n) = \lambda t \mu = \mathbb{E}(N_t) \mathbb{E}(S_1).$$

Así, también se tiene que

$$\begin{aligned} \text{Var}(X_t) &= \mathbb{E}\left(\mathbb{E}\left((S_{N_t} - \lambda\mu t)^2 \mid N_t\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left((S_{N_t} - N_t\mu)^2 + (N_t\mu - \lambda\mu t)^2 \mid N_t\right)\right) \\ &= \mathbb{E}\left(N_t\sigma^2 + \mu^2(N_t - \lambda t)^2\right) \\ &= \sigma^2\lambda t + \mu^2\lambda t. \end{aligned}$$

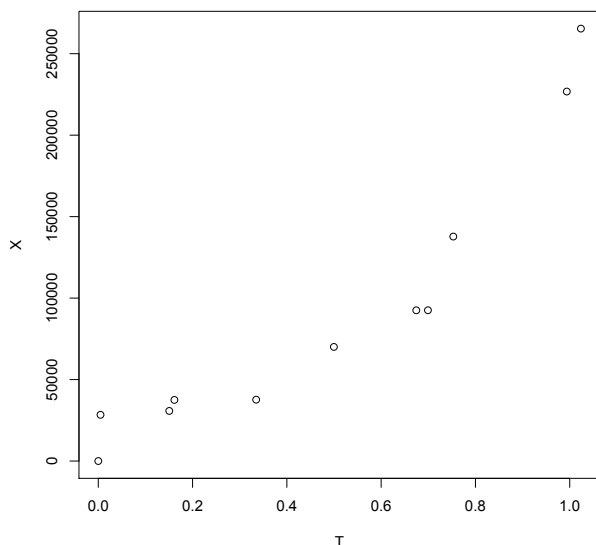


FIGURA 1. Trayectoria de un proceso de Poisson compuesto

Las igualdades anteriores también se pueden interpretar como un caso particular de las identidades de Wald.

Continuaremos ahora con un análisis del Teorema 4.1. Primero probaremos que un proceso de conteo con tiempos interarribo independientes y exponenciales de parámetro λ tiene incrementos independientes y estacionarios.

Sean S_1, S_2, \dots exponenciales independientes de parámetro λ , T las sumas parciales asociadas dadas por $T_0 = 0, T_n = S_1 + \dots + S_n$ y $N = (N_t, t \geq 0)$ el proceso de conteo asociado dado por

$$N_t = \sum n \mathbf{1}_{T_n \leq t < T_{n+1}}.$$

Consideremos al proceso N^t dado por $N_s^t = N_{t+s} - N_t$. Este proceso es un proceso de conteo. Sus tiempos interarribo son

$$S_1^t = T_{N_t+1} - t, S_2^t = S_{N_t+2}, S_3^t = S_{N_t+3}, \dots$$

Al descomponer respecto al valor de N_t vemos que

$$\begin{aligned}
 & \mathbb{P}(S_1^t > t_1, S_2^t > t_2, \dots, S_n^t > t_n) \\
 &= \sum_m \mathbb{P}(N_t = m, S_{m+1} - t > t_1, S_{m+2} > t_2, \dots, S_{m+n} > t_n) \\
 &= \sum_m \mathbb{P}(T_m \leq t < T_m + S_{m+1}, T_m + S_{m+1} - t > t_1, S_{m+2} > t_2, \dots, S_{m+n} > t_n) \\
 &= \sum_m \mathbb{P}(T_m \leq t < T_m + S_{m+1}, T_m + S_{m+1} - t > t_1) \mathbb{P}(S_{m+2} > t_2, \dots, S_{m+n} > t_n).
 \end{aligned}$$

Al condicionar sobre el valor de T_m , su independencia con S_{m+1} y utilizar la propiedad de pérdida de memoria para S_1 , vemos que

$$\begin{aligned}
 & \mathbb{P}(T_m \leq t < T_m + S_{m+1}, S_{m+1} - t > t_1) \\
 &= \mathbb{E}(\mathbb{E}(\mathbf{1}_{T_m \leq t < T_m + S_{m+1}} \mathbf{1}_{T_m + S_{m+1} - t > t_1} \mid T_1)) \\
 &= \mathbb{E}(\mathbf{1}_{T_m \leq t} e^{-\lambda(t-T_n)}) e^{-\lambda t_1}.
 \end{aligned}$$

Sin embargo, al utilizar $t_1 = 0$, nos damos cuenta de que

$$\mathbb{P}(N_t = m) = \mathbb{P}(T_m \leq t < T_m + S_{m+1}) = \mathbb{E}(\mathbf{1}_{T_m \leq t} e^{-\lambda(t-T_n)}),$$

por lo cual se deduce que

$$\mathbb{P}(S_1^t > t_1, S_2^t > t_2, \dots, S_n^t > t_n) = \sum_m \mathbb{P}(N_t = m) e^{-\lambda t_1} e^{-\lambda t_2} \dots e^{-\lambda t_n}.$$

Se deduce que los tiempos de salto de N^t son exponenciales independientes de parámetro λ . Se puede concluir que N^t tiene las mismas distribuciones finito-dimensionales que N , es decir, que si $0 = t_0 \leq t_1 \leq \dots$ entonces

$$\mathbb{P}(N_{t_1} = m_1, \dots, N_{t_n} = m_n) = \mathbb{P}(N_{t_1}^t = m_1, \dots, N_{t_n}^t = m_n).$$

(Esto se verifica al expresar los eventos anteriores en términos de los tiempos interarribo.) En particular, N^t tiene incrementos estacionarios. Ahora veamos que N^t tiene incrementos independientes. Para esto, consideremos al evento

$$A = \{N_{s_1} = k_1, \dots, N_{s_m} = k_m\},$$

donde $s_1, \dots, s_n \leq t$. Al descomponer sobre el valor de N_t se obtiene

$$\begin{aligned}
 & \mathbb{P}(A, N_t = m, S_1^t > t_1, S_2^t > t_2, \dots, S_n^t > t_n) \\
 &= \sum_m \mathbb{P}(A, T_m \leq t < T_m + S_{m+1}, T_m + S_{m+1} - t > t_1) \mathbb{P}(S_{m+2} > t_2, \dots, S_{m+n} > t_n).
 \end{aligned}$$

Sin embargo, como el conjunto $A \cap \{N_t = m\}$ se escribe en términos de S_1, \dots, S_m y de $\{N_t = m\}$, podemos repetir el argumento anterior para verificar que

$$\mathbb{P}(A, T_m \leq t < T_m + S_{m+1}, T_m + S_{m+1} - t > t_1) = \mathbb{P}(A \cap \{N_t = m\}) e^{-\lambda t_1}.$$

Por lo tanto,

$$\begin{aligned}\mathbb{P}(A, N_t = m, S_1^t > t_1, S_2^t > t_2, \dots, S_n^t > t_n) \\ = \mathbb{P}(A, N_t = m) \mathbb{P}(S_1^t > t_1, S_2^t > t_2, \dots, S_n^t > t_n).\end{aligned}$$

Lo anterior implica que N_s^t es independiente de N_{s_1}, \dots, N_{s_m} si $s_1, \dots, s_m \leq t$, y por inducción se puede probar entonces que N tiene incrementos independientes. Así, N es un proceso de contéo con incrementos independientes y estacionarios.

Consideremos ahora un proceso de contéo con incrementos independientes y estacionarios y veamos que se trata de un proceso de renovación con tiempos interarribo exponenciales. De la propiedad de incrementos independientes y estacionarios, podemos inmediatamente verificar que T_1 tiene una distribución exponencial. En efecto, notemos que

$$\begin{aligned}\mathbb{P}(T_1 > t + s) &= \mathbb{P}(N_t = 0, N_{t+s} - N_t = 0) \\ &= \mathbb{P}(N_t = 0) \mathbb{P}(N_s = 0) = \mathbb{P}(T_1 > t) \mathbb{P}(T_1 > s)\end{aligned}$$

Así, o T_1 es infinito casi seguramente, lo cual dice que N es idénticamente cero, ó es exponencial de algún parámetro $\lambda > 0$.

Ahora sólo debemos ver la razón por la que los tiempos interarribo son independientes y con la misma distribución que el primero. Para esto, analizaremos dos implicaciones de nuestras hipótesis conocidas como la propiedad de Markov y de Markov fuerte.

PROPOSICIÓN 4.2 (Propiedad de Markov del proceso de Poisson). *El proceso N^t dado por $N_s^t = N_{t+s} - N_t$ es un proceso de Poisson con las mismas distribuciones finito-dimensionales que N y es independiente de \mathcal{F}_t .*

DEMOSTRACIÓN. Lo primero es notar que N^t es un proceso de Lévy y de contéo. Para ver que tiene las mismas distribuciones finito-dimensionales que N , notamos que

$$\begin{aligned}& \left(N_{t_1}^t, N_{t_2}^t - N_{t_1}^t, \dots, N_{t_n}^t - N_{t_{n-1}}^t \right) \\ &= \left(N_{t+t_1} - N_t, N_{t+t_2} - N_{t+t_1}, \dots, N_{t+t_n} - N_{t+t_{n-1}} \right) \\ &\stackrel{d}{=} \left(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}} \right).\end{aligned}$$

Al considerar $f(x_1, \dots, x_n) = (x_1, x_1 + x_2, \dots, x_1 + \dots + x_n)$, vemos que entonces

$$\begin{aligned}(N_{t_1}^t, N_{t_2}^t, \dots, N_{t_n}^t) &= f(N_{t_1}^t, N_{t_2}^t - N_{t_1}^t, \dots, N_{t_n}^t - N_{t_{n-1}}^t) \\ &\stackrel{d}{=} f(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}) \\ &= (N_{t_1}, N_{t_2}, \dots, N_{t_n}).\end{aligned}$$

Ahora sólo se debe probar que N^t es independiente de \mathcal{F}_t . Sin embargo, si $0 = s_0 \leq s_1 \leq s_m \leq t = t_0 \leq t_1 \leq t_n$ entonces las variables

$$(N_{s_1} - N_{s_0}, \dots, N_{s_m} - N_{s_{m-1}})$$

y

$$(N_{t_1} - N_{t_0}, \dots, N_{t_m} - N_{t_{m-1}}) = (N_{t_1}^t, \dots, N_{t_m-t}^t - N_{t_{m-1}-t}^t)$$

son independientes. Eso implica que

$$(N_{s_1}, \dots, N_{s_m}) \quad \text{y} \quad (N_{t_1}^t, \dots, N_{t_m-t}^t - N_{t_{m-1}-t}^t)$$

son independientes y que por lo tanto, si definimos a los π -sistemas

$$\mathcal{C}_1 = \{N_{s_1} \in A_1, \dots, N_{s_m} \in A_m\}$$

y

$$\mathcal{C}_2 = \{N_{t_1}^t \in B_1, \dots, N_{t_n}^t \in B_n\}$$

se sigue que

$$\mathcal{M}_1 = \{A \in \mathcal{F}_t : A \perp B \text{ para todo } B \in \mathcal{C}_2\}$$

contiene a \mathcal{C}_1 . Puesto que \mathcal{M}_1 es un λ -sistema y $\sigma(\mathcal{C}_1) = \mathcal{F}_s$ entonces

$$\mathcal{F}_s \subset \mathcal{M}_1.$$

Por lo tanto el λ -sistema

$$\mathcal{M}_2 = \{B \in \mathcal{F}_t : A \perp B \text{ para todo } A \in \mathcal{F}_s\}$$

contiene al π -sistema \mathcal{C}_2 . Esto implica que $\sigma(N_s^t : s \geq 0) \subset \mathcal{M}_2$ y que por lo tanto N^t es independiente de \mathcal{F}_t . \square

Ahora, haremos una extensión adicional de la propiedad anterior.

PROPOSICIÓN 4.3 (Propiedad de Markov fuerte). *Si $T : \Omega \rightarrow [0, \infty)$ es una variable aleatoria tal que $\{T \leq t\} \in \mathcal{F}_t$ para toda $t \geq 0$ y*

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t \forall t\}$$

entonces el proceso N^T dado por $N_t^T = N_{T+t} - N_T$ es un proceso de Lévy y de contéo con las mismas distribuciones finito-dimensionales que N e independiente de \mathcal{F}_T .

Veamos cómo se aplica la propiedad de Markov fuerte: puesto que los tiempos interarribo de N^{T_n} son $T_{n+1} - T_n, T_{n+2} - T_{n+1}, \dots$ y son independientes de \mathcal{F}_{T_n} , se deduce que $T_{n+1} - T_n$ es exponencial e independiente de $T_1, T_2 - T_1, \dots, T_n - T_{n-1}$. Esto termina la prueba de que un proceso de contéo y de Lévy es el proceso de contéo asociado a un proceso de renovación exponencial.

PRUEBA DE LA PROPIEDAD DE MARKOV FUERTE. Para cada $n \geq 1$, definimos T^n igual a $(k+1)/2^n$ si $T \in [k/2^n, (k+1)/2^n)$. Formalmente, podemos escribir

$$T^n = \lceil 2^n T \rceil / 2^n.$$

Entonces T^n es una sucesión de variables aleatorias que decrecen hacia T . Además, notemos que

$$\begin{aligned} \{T^n = k + 1/2^n\} &= \{k/2^n \leq T < (k+1)/2^n\} \\ &= \{T < (k+1)/2^n\} \setminus \{T < k/2^n\} \in \mathcal{F}_{(k+1)/2^n}. \end{aligned}$$

Por el mismo argumento, si $A \in \mathcal{F}_T$ entonces

$$A \cap \{T^n = k/2^n\} \in \mathcal{F}_{(k+1)/2^n}.$$

Así, por la propiedad de Markov, vemos que

$$\begin{aligned} \mathbb{P}(A, T^n = k + 1/2^n, N_{t_1}^{T^n} = k_1, \dots, N_{t_m}^{T^n} = k_m) \\ = \mathbb{P}(A, T^n = k + 1/2^n) \mathbb{P}(N_{t_1} = k_1, \dots, N_{t_m} = k_m). \end{aligned}$$

Al sumar sobre k , vemos que

$$\mathbb{P}(A, N_{t_1}^{T^n} = k_1, \dots, N_{t_m}^{T^n} = k_m) = \mathbb{P}(A) \mathbb{P}(N_{t_1} = k_1, \dots, N_{t_m} = k_m).$$

Como N es continuo por la derecha y T^n decrece a T , vemos que conforme $n \rightarrow \infty$: $\mathbb{P}(A, N_{t_1}^T = k_1, \dots, N_{t_m}^T = k_m) = \mathbb{P}(A) \mathbb{P}(N_{t_1} = k_1, \dots, N_{t_m} = k_m)$. Se concluye que N^T es un proceso de contéo y de Lévy con las mismas distribuciones finito-dimensionales que N . Por lo tanto, sus tiempos interarribo tienen la mismas distribuciones conjuntas, pues si \tilde{S} son los tiempos interarribo de N^T entonces

$$\begin{aligned} \mathbb{P}(\tilde{S}_1 > s_1, \tilde{S}_2 > s_2, \dots, \tilde{S}_n > s_n) \\ = \mathbb{P}(\tilde{T}_1 > s_1, \tilde{T}_2 > s_1 + s_2, \dots, \tilde{T}_n > s_1 + \dots + s_n) \\ = \mathbb{P}(N_{s_1}^T \leq 0, N_{s_1+s_2}^T \leq 1, \dots, N_{s_1+\dots+s_n}^T \leq n-1) \\ = \mathbb{P}(N_{s_1} \leq 0, N_{s_1+s_2} \leq 1, \dots, N_{s_1+\dots+s_n} \leq n-1) \\ = \mathbb{P}(S_1 > s_1, S_2 > s_2, \dots, S_n > s_n). \end{aligned}$$

□

PROPOSICIÓN 4.4. *El proceso de Poisson compuesto es un proceso de Lévy y satisface la propiedad de Markov siguiente: si $\mathcal{F}_t^X = \sigma(X_r : r \leq t)$ entonces el proceso X^t dado por $X_s^t = X_{t+s} - X_t$ es un proceso de Poisson compuesto con la misma intensidad y distribución de salto que X y es independiente de \mathcal{F}_t^X .*

DEMOSTRACIÓN. El proceso X comienza en cero y tiene trayectorias que son constantes por pedazos, en particular, continuas por la derecha y con límites por la izquierda.

Para ver que X tiene incrementos independientes y estacionarios, notemos que si $0 = t_0 \leq t_1 \leq \dots \leq t_n$ y $A_i \in \mathcal{B}_{\mathbb{R}}$ para $i = 1, \dots, n$, entonces

$$\begin{aligned} & \mathbb{P}(X_{t_i} - X_{t_{i-1}} \in A_i, i = 1, \dots, n) \\ &= \sum_{0=k_0 \leq k_1 \leq \dots \leq k_n} \mathbb{P}(S_{k_i} - S_{k_{i-1}} \in A_i, N_{t_i} = k_i, i = 1, \dots, n) \\ &= \sum_{0=k_0 \leq k_1 \leq \dots \leq k_n} \prod_{i=1}^n \mathbb{P}(S_{k_i} - S_{k_{i-1}} \in A_i) \mathbb{P}(N_{t_i} = k_i, i = 1, \dots, n), \end{aligned}$$

puesto que las variables $\{S_i : i = 0, 1, \dots\}$ y $\{N_t : t \geq 0\}$ son independientes. Si $k_2 \geq k_1$, entonces $S_{k_2} - S_{k_1}$ tiene la misma distribución que $S_{k_1 - k_2}$, entonces

$$\begin{aligned} & \sum_{0=k_0 \leq k_1 \leq \dots \leq k_n} \prod_{i=1}^n \mathbb{P}(S_{k_i} - S_{k_{i-1}} \in A_i) \mathbb{P}(N_{t_i} = k_i, i = 1, \dots, n) \\ &= \sum_{0=k_0 \leq k_1 \leq \dots \leq k_n} \prod_{i=1}^n \mathbb{P}(S_{k_i - k_{i-1}} \in A_i) \prod_{i=1}^n \mathbb{P}(N_{t_i} - N_{t_{i-1}} = k_i - k_{i-1}) \\ &= \sum_{0 \leq j_1, \dots, j_n} \prod_{i=1}^n \mathbb{P}(S_{j_i} \in A_i, N_{t_i - t_{i-1}} = j_i) \\ &= \prod_{i=1}^n \mathbb{P}(X_{t_i - t_{i-1}} \in A_i). \end{aligned}$$

Si utilizamos el caso $n = 2$, con $A_1 = \mathbb{R}$, podemos concluir que $X_{t_2} - X_{t_1}$ y $X_{t_2 - t_1}$ tienen la misma distribución, de donde la igualdad

$$\mathbb{P}(X_{t_i} - X_{t_{i-1}} \in A_i, i = 1, \dots, n) = \prod_{i=1}^n \mathbb{P}(X_{t_i - t_{i-1}} \in A_i)$$

nos permite concluir la independencia de los incrementos y por lo tanto, que el proceso Poisson compuesto es un proceso de Lévy.

La propiedad de Markov es válida más generalmente para procesos de Lévy y tiene una demostración muy similar a la del proceso de Poisson. \square

Veamos una aplicación adicional del Teorema 4.1.

PROPOSICIÓN 4.5. *Si N^1 y N^2 son dos procesos de Poisson independientes de intensidad λ^1 y λ^2 entonces $N^1 + N^2$ es un proceso de Poisson de intensidad $\lambda^1 + \lambda^2$.*

En efecto, $N^1 + N^2$ es un proceso de contéo y de Lévy y su primer salto, que es igual a $S_1^1 \wedge S_2^1$, tiene distribución exponencial de parámetro $\lambda^1 + \lambda^2$.

CAPÍTULO 5

Procesos de Markov constantes por pedazos

En este capítulo analizaremos una clase de procesos estocásticos que generaliza al proceso de Poisson y tiene similitudes con las cadenas de Markov: los procesos de Markov a tiempo continuo con valores en un conjunto a lo más numerable E . Comenzaremos con un estudio del proceso de Poisson, al expresar su propiedad de Markov de forma que se parezca a la de las cadenas de Markov. Luego, se introducirá un segundo ejemplo, el de los procesos de nacimiento puro. Finalmente, comenzaremos el estudio de procesos con trayectorias constantes por pedazos y daremos una descripción probabilística de estos al introducir parámetros que determinan a un proceso de Markov: la distribución inicial y la matriz infinitesimal. Luego, estudiaremos sus probabilidades de transición mediante unas ecuaciones diferenciales que satisfacen y que están ligadas con su matriz infinitesimal: las ecuaciones backward y forward de Kolmogorov. Finalmente, veremos cómo las ecuaciones backward nos permiten estudiar a las distribuciones invariantes de los procesos de Markov a tiempo continuo.

1. El proceso de Poisson como proceso de Markov

El proceso de Poisson toma valores en los naturales. Como lo hemos definido, siempre comienza en cero, a diferencia de las cadenas de Markov que podemos comenzar en cualquier parte de su espacio de estados. Recordemos que si N es un proceso de Poisson de parámetro λ , al tener incrementos independientes y estacionarios, podemos escribir, para $t_0 < t_1 < t_2 < \dots < t_n$

$$\begin{aligned} \mathbb{P}(N_{t_1} = k_1, N_{t_2} = k_2, \dots, N_{t_n} = k_n) \\ = \mathbb{P}(N_{t_1} = k_1) \mathbb{P}(N_{t_2 - t_1} = k_2 - k_1) \dots \mathbb{P}(N_{t_n - t_{n-1}} = k_n - k_{n-1}). \end{aligned}$$

Por lo tanto, si definimos

$$P_t(i, j) = \mathbb{P}(N_t = j - i) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

vemos que

$$\mathbb{P}(N_{t_1} = k_1, N_{t_2} = k_2, \dots, N_{t_n} = k_n) = P_{t_1}(0, k_1) P_{t_2 - t_1}(k_1, k_2) \cdot P_{t_n - t_{n-1}}(k_{n-1}, k_n).$$

La expresión anterior ya es muy paralela a la que vemos en cadenas de Markov; la entrada i, j de la matriz (infinita) P_t se puede interpretar como la probabilidad de

ir de i a j en t unidades de tiempo. Esto además nos dice cómo podríamos definir a un proceso estocástico que fuera como el proceso de Poisson pero que comenzara en k : simplemente sustituimos $P_{t_1}(0, k_1)$ por $P_{t_1}(k, k_1)$. Sin embargo, no tenemos un proceso estocástico que satisfaga lo anterior; demos una construcción de él.

Sea N un proceso de Poisson de parámetro λ . Si $t > 0$, hemos visto que el proceso N^t dado por $N_s^t = N_{t+s} - N_t$ es un proceso de Poisson independiente de N_t . Por lo tanto:

$$\begin{aligned} & \mathbb{P}(N_{t+s_1} = k_1, \dots, N_{t+s_n} = k_n | N_t = k) \\ &= \frac{\mathbb{P}(N_{t+s_1} = k_1, \dots, N_{t+s_n} = k_n, N_t = k)}{\mathbb{P}(N_t = k)} \\ &= \frac{\mathbb{P}(N_{s_1} = k_1 - k, \dots, N_{s_n} = k_n - k) \mathbb{P}(N_t = k)}{\mathbb{P}(N_t = k)} \\ &= \mathbb{P}(N_{s_1} + k = k_1, \dots, N_{s_n} + k = k_n). \end{aligned}$$

Vemos entonces que, la distribución condicional del proceso de Poisson en tiempos posteriores a t condicionalmente a $N_t = k$ es la misma que la del proceso $k + N$. Así, es natural definir al proceso de Poisson que comienza en k como $k + N$, pues este proceso tiene trayectorias como las del proceso de Poisson (aumenta de uno en uno en ciertos tiempos aleatorios) pero comienza en k . Además:

$$\mathbb{P}(N_{t+s_1} = k_1, \dots, N_{t+s_n} = k_n | N_t = k) = P_{t_1}(k, k_1) P_{t_2-t_1}(k_1, k_2) \cdot P_{t_n}(k_{n-1}, k_n)$$

Recordemos que en el caso de cadenas de Markov, se satisfacen las ecuaciones de Chapman-Kolmogorov. En este contexto, las ecuaciones se pueden expresar como sigue:

$$P_{t+s}(i, k) = \sum_j P_s(i, j) P_t(j, k).$$

EJERCICIO 5.1. Al utilizar el teorema del binomio, pruebe directamente que la ecuación anterior se satisface. Dé además un argumento probabilístico, basado en condicionar con lo que sucede al tiempo s , para probar dicha ecuación.

La diferencia con las cadenas de Markov y el proceso de Poisson es que, al ser el segundo un proceso de Markov a tiempo continuo, en lugar de tener una sola matriz cuyas potencias nos permiten describir al proceso, tenemos toda una colección de matrices ($P_t, t \geq 0$). Uno de los objetivos de este capítulo es mostrar como, con una definición adecuada, podemos generar a todas las matrices ($P_t, t \geq 0$) mediante una sola matriz, la llamada matriz infinitesimal o matriz de tasas de transición. La idea es la siguiente: si x es un número, podemos interpolar a la sucesión $x^n, n \in \mathbb{N}$ mediante la función exponencial: $x^n = e^{n \log x}$. El lado derecho ya tiene sentido si $n \geq 0$ y no sólo si $n \in \mathbb{R}$. En cierto sentido, si $P_t = e^{tQ}$, podríamos interpretar a Q como el logaritmo de P_1 . Para ver cómo podríamos obtener a Q , notemos que para

obtener a $\log x$ si conocemos a $f(t) = e^{t \log x}$, simplemente calculamos $\log x = f'(0)$ y además $f'(t) = f'(0) f(t)$. En el caso de nuestra matriz P_t , vemos que

$$\begin{aligned} \frac{d}{dt} P_t(i, j) &= \frac{d}{dt} e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} \\ &= \frac{1}{(j-i)!} \left[e^{-\lambda t} (j-i) t^{j-i-1} \lambda^{j-i} \mathbf{1}_{j \geq i+1} - \lambda e^{-\lambda t} (\lambda t)^{j-i} \right] \end{aligned}$$

por lo que al evaluar en cero se obtiene:

$$\left. \frac{d}{dt} \right|_{t=0} P_t(i, j) = \begin{cases} -\lambda & j = i \\ \lambda & j = i + 1 \\ 0 & j \neq i, i + 1 \end{cases}.$$

EJERCICIO 5.2. Sea

$$Q(i, j) = \begin{cases} -\lambda & j = i \\ \lambda & j = i + 1 \\ 0 & j \neq i, i + 1 \end{cases}.$$

Pruebe directamente que

$$(5) \quad \frac{d}{dt} P_t(i, j) = Q P_t(i, j) = P_t Q(i, j),$$

donde $Q P_t$ es el producto de las matrices Q y P_t .

A las ecuaciones (diferenciales) (5) se les conoce como **ecuaciones de Kolmogorov**. Esbochemos una interpretación y deducción probabilística de dichas ecuaciones. Para calcular $P_{t+h}(0, j)$, podemos descomponer respecto del valor de N_h para obtener

$$P_{t+h}(i, j) = \mathbb{P}(N_{t+h} = j) = \mathbb{P}(N_h = i) \mathbb{P}(N_t = j - i).$$

Por otra parte, al utilizar explícitamente la distribución Poisson, vemos que

$$\lim_{h \rightarrow 0} \frac{1 - \mathbb{P}(N_h = 0)}{h} = \lambda \lim_{h \rightarrow 0} \frac{\mathbb{P}(N_h = 1)}{h} = \lambda \quad \lim_{h \rightarrow 0} \frac{\mathbb{P}(N_h \geq 2)}{h} = 0.$$

La interpretación es que es muy probable que haya cero saltos en un intervalo de longitud h , hay probabilidad proporcional al tamaño del intervalo de que haya un sólo salto, y muy improbable que haya más de dos saltos. Se concluye que

$$\begin{aligned} \frac{\partial P_t(i, k)}{\partial t} &= \lim_{h \rightarrow 0} \frac{P_h(i, i) P_t(i, k) - P_t(i, k)}{h} \\ &\quad + \lim_{h \rightarrow 0} \frac{P_h(i, i+1) P_t(i+1, k)}{h} \\ &= P_t(i+1, k) \lambda - P_t(i, j) \lambda = P_t Q(i, k) \end{aligned}$$

A esta ecuación se le denomina **ecuación hacia atrás de Kolmogorov** (también llamada, aún en español, ecuación backward) y fué obtenida al descomponer a

N_{t+h} respecto del valor de N_h . Ésta implica automáticamente la **ecuación hacia adelante de Kolmogorov**

$$\frac{\partial P_t(i, k)}{\partial t} = QP_t(i, k)$$

al utilizar la relación $P_t(i+1, j) = P_t(i, j-1)$.

2. El proceso de nacimiento puro

El proceso de nacimiento puro es una generalización del proceso de Poisson que puede presentar un fenómeno interesante: la explosión. Daremos una definición de dicho proceso paralela a la del proceso de Poisson.

DEFINICIÓN. Sean $q_0, q_1, \dots \in (0, \infty)$. Un **proceso de nacimiento puro** es el proceso de contéo cuyos tiempos interarribo conforman una sucesión de variables exponenciales independientes S_1, S_2, \dots de parámetros $\lambda_0, \lambda_1, \dots$

La interpretación del proceso de nacimiento puro es que λ_i representa la tasa a la que un nuevo individuo se agrega a una población cuando esta tiene i elementos. Así, el proceso de nacimiento puro se construye al definir a los tiempos de salto

$$0 = T_0 \quad T_n = S_1 + \dots + S_n \quad T_\infty = \sum_{n=0}^{\infty} S_i$$

y al proceso de contéo asociado

$$N_t = \sum_{n \in \mathbb{N}} n \mathbf{1}_{T_n \leq t < T_{n+1}}.$$

Claramente, cuando $q_i = \lambda$ para toda i el proceso que hemos construido es el proceso de Poisson. Sin embargo, cuando tenemos q_i dependientes de i se presenta un nuevo fenómeno. En efecto, en el caso del proceso de Poisson, sabemos que $T_n/n \rightarrow 1/\lambda = \mathbb{E}(S_1) > 0$ por la ley fuerte de los grandes números y por lo tanto $T_\infty = \infty$ casi seguramente. Sin embargo, en el caso del proceso de nacimiento puro puede suceder que $T_\infty < \infty$, en cuyo caso:

$$\lim_{t \rightarrow T_\infty -} N_t = \infty.$$

Decimos entonces que ha ocurrido la explosión (en este caso demográfica) del proceso de nacimiento puro y es natural definir

$$N_t = \infty \mathbf{1}_{T_\infty \leq t} + \sum_{n \in \mathbb{N}} n \mathbf{1}_{T_n \leq t < T_{n+1}}$$

pues con la definición anterior el proceso se vuelve cero despues de que la población haya explotado.

Puesto que el proceso de nacimiento puro admite una construcción sencilla, se puede dar un criterio explícito para la explosión, el cual además prueba que la explosión se da o con probabilidad cero o con probabilidad uno.

PROPOSICIÓN 5.1. $T_\infty = \infty$ casi seguramente si $\mathbb{E}(T_\infty) = \sum_i 1/q_i = \infty$. Si $\sum_i 1/q_i < \infty$ entonces $T_\infty < \infty$ casi seguramente.

DEMOSTRACIÓN. Si $\sum_i 1/q_i < \infty$ entonces podemos aplicar el teorema de convergencia monótona a la sucesión T_n para deducir que

$$\mathbb{E}(T_\infty) = \lim_n \mathbb{E}(T_n) = \lim_n \sum_{m \leq n} 1/q_m = \sum_n 1/q_n < \infty,$$

por lo que $T_\infty < \infty$ casi seguramente.

Por otra parte, si $\sum_i 1/q_i = \infty$, entonces podemos aplicar el teorema de convergencia acotada para ver que

$$\mathbb{E}(e^{-T_\infty}) = \lim_n \mathbb{E}(e^{-T_n}) = \lim_n \prod_{i=1}^n \frac{q_i}{1+q_i}.$$

Ahora dividiremos en dos partes nuestro análisis: si existe $\mu > 0$ tal que $q_i \leq \mu$ para una cantidad infinita de índices i entonces

$$\prod_{i=1}^n \frac{q_i}{1+q_i} \leq \left(\frac{1}{1+1/\mu} \right)^n \rightarrow 0$$

conforme $n \rightarrow \infty$. Si por otra parte $q_i \rightarrow \infty$ entonces utilizamos el hecho de que $q_i \log(1+1/q_i) \rightarrow 1$ conforme $n \rightarrow \infty$ y por lo tanto

$$\prod_{i=1}^n \frac{q_i}{1+q_i} = e^{-\sum_{i=1}^n \log(1+1/q_i)} \leq e^{-C \sum_{i=1}^n 1/q_i} \rightarrow 0.$$

Vemos que en cualquier caso

$$\mathbb{E}(e^{-T_\infty}) = 0,$$

lo cual nos dice que $T_\infty = \infty$ casi seguramente. \square

Para analizar la propiedad de Markov del proceso de nacimiento puro, necesitamos definirlo cuando comienza en $i \in \mathbb{N}$. Una definición posible es que se trata del proceso $(N_{T_i+t}, t \geq 0)$. Este proceso va tomando los valores sucesivos $i, i+1, i+2, \dots$ y los tiempos que permanece en cada uno de los estados son S_i, S_{i+1}, \dots . Pues este proceso tiene la misma estructura probabilística que i más un proceso de conteo cuyos tiempos interarribo son exponenciales independientes de parámetros $\lambda_i, \lambda_{i+1}, \dots$, o sea, que una definición equivalente es que un proceso de nacimiento puro con parámetros $\lambda_0, \lambda_1, \dots$ que comienza en i es i más un proceso de nacimiento puro con parámetros $\lambda_i, \lambda_{i+1}, \dots$. Con estos preliminares podemos enunciar la propiedad de Markov del proceso de nacimiento puro.

PROPOSICIÓN 5.2. Sean N un proceso de nacimiento puro de parámetros $\lambda_0, \lambda_1, \dots$ y $s \geq 0$. Condicionalmente a $N_s = i$, el proceso estocástico $(N_{t+s} - i, t \geq 0)$ es un proceso de nacimiento puro de parámetros $\lambda_i, \lambda_{i+1}, \dots$ y es condicionalmente independiente de $\mathcal{F}_s = \sigma(N_r : r \leq t)$ dado $N_s = i$.

En otras palabras, condicionalmente a $N_s = i$, el proceso $N_{t+s}, t \geq 0$ es un proceso de nacimiento puro que comienza en i .

DEMOSTRACIÓN. El proceso $N_{t+s} - i, t \geq 0$ es un proceso de conteo cuyos tiempos interarribo son $T_{N_s+1} - s, S_{N_s+2}, S_{N_s+3}, \dots$. Denotémoslos como S_0^s, S_1^s, \dots . Debemos ver que condicionalmente a $N_s = i$, estos tiempos interarribo son exponenciales, independientes y de parámetros $\lambda_i, \lambda_{i+1}, \dots$. En efecto:

$$\begin{aligned} \mathbb{P}(N_s = i, S_0^s > s_0, \dots, S_n^s > s_n) \\ = \mathbb{P}(T_i \leq s < T_i + S_i, s_0 + s < T_i + S_i, s_1 < S_{i+1}, \dots, s_n < S_{i+n}). \end{aligned}$$

Al condicionar por T_i , utilizar la independencia de las variables T_i, S_i, \dots, S_{i+n} y la propiedad de pérdida de memoria de la distribución exponencial vemos que

$$\begin{aligned} \mathbb{P}(T_i \leq s < T_i + S_i, s_0 + s < T_i + S_i, s_1 < S_{i+1}, \dots, s_n < S_{i+n}). \\ = \mathbb{E}(\mathbf{1}_{T_i \leq s} e^{-\lambda_i} (s - T_i)) e^{-\lambda_i s_1} \dots e^{-\lambda_{i+n} s_n}. \end{aligned}$$

Sin embargo, al poner $s_0 = \dots = s_n = 0$, vemos que

$$\mathbb{E}(\mathbf{1}_{T_i \leq s} e^{-\lambda_i} (s - T_i)) = \mathbb{P}(N_s = i)$$

y por lo tanto, al condicionar por $N_s = i$, vemos que N^s es un proceso de nacimiento puro que comienza en i .

Falta demostrar que N^s es condicionalmente independiente de \mathcal{F}_s dado que $N_s = i$, esto es, que para todo $A \in \mathcal{F}_s$, se tiene que

$$\begin{aligned} \mathbb{P}(N_{t_1}^s = k_1, \dots, N_{t_n}^s = k_n, A | N_s = i) \\ = \mathbb{P}(N_{t_1}^s = k_1, \dots, N_{t_n}^s = k_n | N_s = i) \mathbb{P}(A | N_s = i). \end{aligned}$$

Por clases monótonas, basta verificarlo cuando $A = \{N_{s_1} = j_1, \dots, N_{s_m} = j_m\}$ con $s_1 \leq \dots \leq s_m \leq t$ y $j_1 \leq \dots \leq j_m \leq i$. Para esto, seguimos el mismo razonamiento anterior al notar que

$$A \cap \{N_s = i\} = \{T_{j_1} \leq s_1 < T_{j_1+1}, \dots, T_{j_m} \leq s < T_{j_m+1}\} \cap \{T_i \leq s < T_{i+1}\}.$$

□

EJERCICIO 5.3. Haga un programa en R que simule al proceso de nacimiento puro que comienza en 1 si $\lambda_i = i\lambda$ para algún $\lambda > 0$. ¿Este proceso explota? Haga otro programa que simule el caso $\lambda_i = \lambda i^2$ y diga si ocurre explosión o no.

La propiedad de Markov se puede interpretar de manera similar a la del proceso de Poisson. Dados los parámetros $\mathbf{q} = (\lambda_0, \lambda_1, \dots)$, sea $P_t(i, j)$ la probabilidad de que un proceso de nacimiento puro de parámetro \mathbf{q} que comienza en i se encuentre en j al tiempo t . Entonces:

$$\begin{aligned} \mathbb{P}(N_{s+t_1} = k_1, \dots, N_{s+t_n} = k_n | N_s = k) \\ = P_{t_1}(k, k_1) P_{t_2-t_1}(k_1, k_2) \dots P_{t_n-t_{n-1}}(k_{n-1}, k_n). \end{aligned}$$

Aquí es más difícil obtener una expresión explícita para $P_t(i, j)$. Ésto se puede lograr cuando por ejemplo $\lambda_i = i\lambda$ para alguna $\lambda > 0$ (que es el caso del **Proceso de Yule**, básico en la teoría de procesos de ramificación a tiempo continuo). Sin embargo, podemos argumentar por qué son válidas las ecuaciones hacia atrás de Kolmogorov: Si h es pequeño y nuestro proceso de nacimiento puro comienza en i , será raro que haya más de dos saltos en un intervalo de tiempo pequeño. Lo que se afirma es que, exactamente como en el caso de procesos de Poisson,

$$\lim_{h \rightarrow 0} \frac{1 - P_h(i, i)}{h} = \lambda_i \lim_{h \rightarrow 0} \frac{P_h(i, i+1)}{h} = \lambda_i \quad \lim_{h \rightarrow 0} \frac{\sum_{j \geq 2} P_h(i, j)}{h} = 0.$$

Por lo tanto, las ecuaciones hacia atrás de Kolmogorov serían

$$\frac{\partial}{\partial t} P_t(i, j) = \lambda_i P_t(i+1, j) - \lambda_i P_t(i, j).$$

Vemos entonces que la matriz infinitesimal o de tasas de transición estaría dada por

$$Q(i, j) = \begin{cases} -\lambda_i & i = j \\ \lambda_i & j = i+1 \\ 0 & \text{en otro caso} \end{cases}$$

y que las ecuaciones hacia atrás se pueden escribir como

$$\frac{\partial}{\partial t} P_t(i, j) = Q P_t(i, j).$$

3. Matrices infinitesimales y construcción de procesos de Markov

Comenzaremos por describir a un tercer ejemplo de proceso de Markov a tiempo continuo. Este ejemplo se basa en calcular el mínimo de variables exponenciales independientes.

Comenzaremos con un ejemplo con espacio de estados E finito. Para cada $x, y \in E$ con $x \neq y$ sea $\lambda_{x,y} \geq 0$. Sean $E_{i,x,y}, \geq 1, x, y \in E, x \neq y$ variables aleatorias exponenciales independientes, tal que $E_{i,x,y}$ tiene parámetro $\lambda_{x,y}$. A $\lambda_{x,y}$ lo interpretaremos como la tasa a la que deajo el estado x para acceder al estado y y supondremos que para cada $x \in E$ existe $y \in E$ distinta de x tal que $\lambda_{x,y} > 0$. Si $x \in E$, definamos a

$$T_1 = \min_{y \neq x} E_{1,x,y} \quad \text{donde } T_1 = E_{1,x,X_1}.$$

Recursivamente, definiremos a

$$T_{n+1} = \min_{y \neq X_n} E_{n+1,X_n,y} \quad \text{donde } T_{n+1} = E_{1,X_n,X_{n+1}}.$$

Definamos finalmente, al proceso a tiempo continuo de interés:

$$Z_t = X_n \quad \text{si } t \in [T_n, T_{n+1}).$$

Hagamos un ejemplo en R para ilustrar la definición. Comencemos con $E = \{1, 2\}$, por lo que necesitamos definir a dos cantidades positivas $\lambda_{1,2}$ (que denotaremos por λ_1 y digamos que es igual a 2) y $\lambda_{2,1}$ (denotada λ_2 y digamos igual a 1). Si queremos comenzar nuestra construcción con $x = 1$, podemos utilizar el siguiente código.

```
# Simulemos una cadena de Markov en tiempo continuo en $E={1,2}$
# l(1) es la tasa a la que dejo $1$ (para ir a 2) y viceversa
l=c(2,1);
# x es el estado actual, comenzamos en
x=1;
X=c(x);
# T contendr'a los tiempos en que cambia la cadena
T=0
# n es la cantidad de pasos
n=20;
for(i in 1:n)
{
  #Cambio de estado
  x=3-x;
  #Agrego el nuevo estado
  X=c(X,x)
  # Simulo la exponencial de la tasa adecuada y la agrego a mis tiempos
  T=c(T,tail(T,1)+rexp(1,l[x]));
}
# Grafico la trayectoria
plot(T,X,type="s")
```

LISTING 5.1. QCadena2Estados.R

Una exponencial de parámetro λ tiene la misma distribución que una exponencial de parámetro 1 dividida entre λ , por lo cual en el ciclo es equivalente utilizar el comando $T=c(T,\text{tail}(T,1)+\text{rexp}(1)/l[x]);$.

Un ejemplo un poco más complicado lo obtenemos si $E = \{1, 2, 3\}$ y ponemos las tasas de transición $\lambda_{1,2} = 2$, $\lambda_{2,1} = \lambda_{2,3} = 1$, $\lambda_{3,1} = 1/3$ y todas las demás igual a cero. En este caso, nuestro primer impulso podría ser el utilizar el siguiente código:

```
# Simulemos una cadena de Markov en tiempo continuo en $E={1,2,3}$
# La matriz de tasas de transición
L=matrix(0,3,3)
L[1,2]=2
L[2,1]=1
L[2,3]=1
L[3,1]=1/3
# Estado inicial
x=1
# Vector de estados
X=c(x)
# Vector de tiempos
T=c(0)
# N'umero de pasos
n=20
```

```

for(i in 1:20)
{
# Genero las exponenciales
e=rexp(3)/L[x,]
# El nuevo tiempo de salto
t=min(e)
T=c(T,tail(T,1)+t)
# El nuevo estado
x=which.min(e)
X=c(X,x)
}
plot(T,X,type="s")

```

LISTING 5.2. QCadena3Estados.R

Hay una manera de realizar la simulación anterior con menos variables aleatorias. En vez de las 3 exponenciales, podemos utilizar una exponencial y una uniforme. La idea es utilizar la proposición siguiente para ver que el mínimo de variables exponenciales es exponencial.

PROPOSICIÓN 5.3. *Sean T_1, \dots, T_n variables exponenciales independientes de parámetros respectivos $\lambda_1, \dots, \lambda_n$. Sea $T = \min_i T_i$. Entonces, con probabilidad 1, existe un único índice K tal que $T = T_K$. T y K son independientes, T tiene distribución exponencial de parámetro $\lambda = \lambda_1 + \dots + \lambda_n$ y $\mathbb{P}(K = k) = \lambda_k / \lambda$.*

DEMOSTRACIÓN. Calculemos:

$$\begin{aligned}
 \mathbb{P}(T \geq t, T_j > T_k \text{ para toda } j \neq k) &= \mathbb{P}(T_k \geq t, T_j > T_k \text{ para toda } j \neq k) \\
 &= \mathbb{E} \left(\mathbf{1}_{T_k \geq t} \prod_{j \neq k} e^{-\lambda_k T_k} \right) \\
 &= \int_t^\infty \lambda_k e^{-\lambda_k s} \prod_{j \neq k} e^{-\lambda_j s} ds \\
 &= \frac{\lambda_k}{\lambda} e^{-\lambda t}.
 \end{aligned}$$

Al evaluar en $t = 0$ vemos que

$$\mathbb{P}(T_j > T_k \text{ para toda } j \neq k) = \frac{\lambda_k}{\lambda}$$

por lo que al utilizar que los eventos anteriores son ajenos conforme variamos a k y su unión es la probabilidad de que el mínimo de T_1, \dots, T_n se alcance para un sólo índice, vemos que

$$\mathbb{P}(\text{Existe un único } k \in \{1, \dots, n\} \text{ tal que } T = T_k) = \sum_k \frac{\lambda_k}{\lambda} = \frac{\lambda}{\lambda} = 1.$$

Al sumar sobre k , vemos que

$$\mathbb{P}(T \geq t) = e^{-\lambda t},$$

lo cual implica que T es exponencial de parámetro λ

Puesto que con probabilidad 1 hay un único índice (aleatorio) en el que se alcanza el mínimo, digamos K tal que $T_K = T$, vemos que

$$\mathbb{P}(K = k, T \geq t) = \frac{\lambda_k}{\lambda} e^{-\lambda t} = \mathbb{P}(K = k) \mathbb{P}(T \geq t). \quad \square$$

Así, nuestro proceso a tiempo continuo se puede simular también mediante el siguiente código.

```
# Simulemos una cadena de Markov en tiempo continuo en $E=\{1,2,3\}$
# La matriz de tasas de transición
L=matrix(0,3,3)
L[1,2]=2
L[2,1]=1
L[2,3]=1
L[3,1]=1/3
# Tasas totales de salto
l=rowSums(L)
# Matriz de transición
P=L/rowSums(L)
# Estado inicial
x=1
# Vector de estados
X=c(x)
# Vector de tiempos
T=c(0)
# N\numero de pasos
n=20
for(i in 1:20)
{
  # Genero la exponencial
  e=rexp(1,1[x])
  # El nuevo tiempo de salto
  T=c(T,tail(T,1)+e)
  # El nuevo estado
  x=sample(3,1,prob=P[x,])
  X=c(X,x)
}
plot(T,X,type="s")
```

LISTING 5.3. QCadena3EstadosConMinimo.R

Continuemos con el análisis de la cadena general. A partir del código, vemos que los parámetros se pueden reducir a una matriz de transición P con ceros en la diagonal y un vector l de tasas totales de salto. Explícitamente,

$$l(x) = \sum_y \lambda_{x,y} \quad \text{y} \quad P_{x,y} = \frac{\lambda_{x,y}}{l(x)}.$$

Además, es posible construir nuestra cadena mediante una sucesión de variables exponenciales estándar independientes ξ_1, ξ_2, \dots entre sí y de una sucesión iid U_1, U_2, \dots de variables uniformes. En efecto, podemos definir a $X_0 = x$, $T_0 = x$

y a $T_1 = S_1 = \xi_1/l(X_0)$. Luego, utilizamos a la variable uniforme U_1 para escoger a un elemento aleatorio X_1 de E tal que $\mathbb{P}(X_1 = y) = P_{X_0, y}$. Finalmente, continuamos este procedimiento de manera recursiva: si ya tenemos definidos a X_0, \dots, X_n y a T_0, \dots, T_n en términos de S_1, \dots, S_n y U_1, \dots, U_n entonces definimos $S_{n+1} = \xi_{n+1}/l(X_n)$, $T_{n+1} = T_n + S_{n+1}$ y utilizamos a las variables U_{n+1} y X_n para construir a X_{n+1} de tal manera que $\mathbb{P}(X_{n+1} = y | X_n = x) = P_{x, y}$. Finalmente, recordemos que $Z_t = X_n$ si $T_n \leq t < T_{n+1}$. Resulta ser que Z es un proceso de Markov a tiempo continuo. Formalmente, sea $P_t(x, y)$ la probabilidad de que $Z_t = y$ cuando $Z_0 = x$ y probemos que

$$\begin{aligned} \mathbb{P}_{x_0}(X_{t_1} = x_1, \dots, X_{t_n} = y_n) \\ = P_{t_1-t_0}(x_0, x_1) P_{t_2-t_1}(x_1, x_2) \cdots P_{t_n-t_{n-1}}(x_{n-1}, x_n) \end{aligned}$$

donde \mathbb{P}_x es la medida de probabilidad que rige a Z cuando comienza en x y $0 = t_0 < t_1 < \dots < t_n$. El argumento es parecido a cuando probamos que el proceso de Poisson (o más generalmente el de nacimiento y muerte) es un proceso de Markov, se basa en notar que el proceso posterior a s , condicionalmente a $Z_s = x$ tiene la misma construcción probabilística que Z comenzando en x . Esto es, que sus tiempos y lugares de salto se pueden obtener a partir de una sucesión de variables uniformes y exponenciales estándar independientes. Concentrémonos mejor en las aplicaciones de este hecho.

La primera aplicación es a la obtención de las ecuaciones de Chapman-Kolmogorov. En efecto, vemos que

$$\begin{aligned} P_{t+s}(x, z) &= \mathbb{P}_x(Z_{t+s} = z) \\ &= \sum_y \mathbb{P}_x(Z_s = y, Z_{t+s} = z) \\ &= \sum_y P_s(x, y) P_t(y, z). \end{aligned}$$

Seguido de esto, obtendremos las ecuaciones de Kolmogorov como en el caso de procesos de procesos de Poisson. Se afirma que

$$\lim_{h \rightarrow 0} \frac{1 - \mathbb{P}_x(N_h = 0)}{h} = l(x), \quad \lim_{h \rightarrow 0} \frac{\mathbb{P}_x(N_h = 1), Z_t = y}{h} = l(x) P_{x, y}$$

y

$$\lim_{h \rightarrow 0} \frac{\mathbb{P}_x(N_h \geq 2)}{h} = 0.$$

En efecto, puesto que bajo \mathbb{P}_x el primer salto de N ocurre en una variable exponencial de parámetro $l(x)$, entonces

$$\lim_{h \rightarrow 0} \frac{1 - \mathbb{P}_x(N_h = 0)}{h} = \lim_{h \rightarrow 0} \frac{1 - e^{-l(x)h}}{h} = l(x).$$

Por otra parte, vemos que

$$\begin{aligned}\{N_h \geq 2\} &= \{S_1 + S_2 \leq h\} \\ &= \left\{ \frac{\xi_1}{l(X_0)} + \frac{\xi_2}{l(X_1)} \leq h \right\} \\ &\subset \left\{ \frac{\xi_1}{\lambda} + \frac{\xi_2}{\lambda} \leq h \right\}.\end{aligned}$$

El último evento corresponde a que un proceso de Poisson de parámetro λ , digamos N^λ tenga más de dos saltos en el intervalo $[0, t]$, que ya hemos estimado. Se concluye que

$$0 \leq \limsup_{h \rightarrow 0} \frac{\mathbb{P}_x(N_h \geq 2)}{h} \leq \lim_{h \rightarrow 0} \frac{\mathbb{P}_x(N_h^\lambda \geq 2)}{h} = 0.$$

Finalmente, veamos cómo ocuparnos de $\mathbb{P}_x(N_h, X_h = y) = 1$. Puesto que U_1 es independiente de ξ_1, ξ_2 , entonces

$$\begin{aligned}\frac{1}{h} \mathbb{P}_x(N_h = 1, Z_h = y) &= \frac{1}{h} \mathbb{P}_x\left(\frac{\xi_1}{l(x)} \leq h < \frac{\xi_1}{l(x)} + \frac{\xi_2}{l(y)}\right) P_{x,y} \\ &= P_{x,y} \frac{1}{h} \mathbb{P}_x\left(\frac{\xi_1}{l(x)} \leq h < \frac{\xi_1}{l(x)} + \frac{\xi_2}{\lambda}\right) \\ &= P_{x,y} \frac{1}{h} \int_0^h l(x) e^{-l(x)s} e^{-l(y)(h-s)} ds \\ &= P_{x,y} l(x) e^{-l(y)h} \frac{1}{h} \int_0^h e^{-s[l(y)-l(x)]} ds \\ &\rightarrow_{h \rightarrow 0} l(x) P_{x,y}.\end{aligned}$$

Podemos entonces deducir que

$$\lim_{h \rightarrow 0} \frac{P_h(x, x) - 1}{h} = -l(x) \quad \text{y para } y \neq x \quad \lim_{h \rightarrow 0} \frac{P_h(x, y)}{h} = P_{x,y} l(x).$$

Llámemósle Q a la matriz

$$Q_{x,y} = \begin{cases} -l(x) & y = x \\ l(x) P_{x,y} & y \neq x \end{cases}.$$

A esta matriz se le conoce como **matriz infinitesimal** del proceso de Markov Z .

Al utilizar la finitud de E para intercambiar derivadas y sumas y las ecuaciones de Chapman-Kolmogorov se obtiene

$$\begin{aligned} \frac{d}{dt} P_t(x, z) &= \lim_{h \rightarrow 0} \frac{P_{t+h}(x, z) - P_t(x, z)}{h} \\ &= \lim_{h \rightarrow 0} P_t(x, z) \frac{P_h(z, z) - 1}{h} + \sum_{y \neq z} P_t(x, y) \lim_{h \rightarrow 0} \frac{1}{h} P_h(y, z) \\ &= \sum_{y \neq z} P_t(x, y) l(y) P_{y, z} - P_t(x, y) l(y) \\ &= (P_t Q)_{x, y}. \end{aligned}$$

Estas son las **ecuaciones forward de Kolmogorov**. De igual manera, al utilizar la finitud de E se obtienen las **ecuaciones backward de Kolmogorov**

$$\frac{d}{dt} P_t(x, z) = (Q P_t)_{x, y}.$$

Como ya se ha comentado, la anterior ecuación diferencial matricial es parecida a la ecuación diferencial $f'(t) = \lambda f(t)$ (con $f(t) = P_t$ y $\lambda = Q$) cuya única solución es $f(t) = e^{\lambda t}$. Aún en el caso matricial se le puede dar sentido a esta ecuación diferencial al definir la exponencial de la matriz tQ por medio de

$$e^{tQ} = \sum_n \frac{t^n Q^n}{n!}.$$

Entonces, se deduce que

$$P_t = e^{tQ}.$$

La pregunta que sigue es si es posible realizar la construcción que acabamos de explorar en espacio de estados numerable. La respuesta es básicamente que sí, salvo que se debe prestar atención al fenómeno de explosión. Otra pregunta importante es si cualquier cadena de Markov a tiempo continuo se puede construir como lo acabamos de hacer. La respuesta es básicamente afirmativa de nuevo cuidandonos de la explosión. La siguiente sección aborda este segundo cuestionamiento.

4. Descripción probabilística de procesos de Markov constantes por pedazos

En esta sección definiremos a las cadenas de Markov a tiempo continuo y analizaremos su estructura y su comportamiento a tiempos grandes. Se realizará por lo tanto un estudio paralelo al de las cadenas de Markov a tiempo discreto.

Sea E un conjunto a lo más numerable, $\Delta \notin E$ algún punto que denominaremos **cementerio** y a donde mandaremos las trayectorias de un proceso de Markov cuando explote, y sea C el conjunto de funciones $f : [0, \infty) \rightarrow E \cup \{\Delta\}$ que satisfacen:

- (1) si $f(t) \in E$ entonces existe $\delta > 0$ tal que $f(s) = f(t)$ para $s \in [t, t + \delta]$,
- (2) si $f(t) = \Delta$ entonces $f(s) = \Delta$ para toda $s \geq t$ y
- (3) si $t_1 < t_2 < \dots$, $t_n \rightarrow t < \infty$ y $f(t_{n+1}) \neq f(t_n)$ entonces $f(t) = \Delta$.

DEFINICIÓN. Una **cadena de Markov a tiempo continuo** con espacio de estados E es un proceso estocástico $X = (X_t, t \geq 0)$ tal que

- (1) X_t toma valores en $E \cup \{\Delta\}$
- (2) para todo $\omega \in \Omega$, la trayectoria $t \mapsto X_t(\omega)$ es un elemento del conjunto de funciones C ,
- (3) existe una colección de matrices estocásticas $(P_t, t \geq 0)$ indexadas por los elementos de E tales que si $0 = t_0 < t_1 < \dots < t_n$ entonces:

$$\begin{aligned} \mathbb{P}(X_0 = x, X_{t_1} = x_1, \dots, X_{t_n} = x_n) \\ = \mathbb{P}(X_0 = x) P_{t_1 - t_0}(x_0, x_1) P_{t_2 - t_1}(x_1, x_2) \cdots P_{t_n - t_{n-1}}(x_{n-1}, x_n). \end{aligned}$$

Denotaremos por \mathbb{P}_x a \mathbb{P} condicionada por $X_0 = x$. A la colección de matrices estocásticas $(P_t, t \geq 0)$ les llamamos **probabilidades de transición**.

Comenzamos por verificar una propiedad familiar de las probabilidades de transición.

PROPOSICIÓN 5.4. *Las probabilidades de transición satisfacen las ecuaciones de Chapman-Kolmogorov*

$$P_{t+s} = P_t P_s.$$

DEMOSTRACIÓN. Al descomponer $\mathbb{P}_x(X_{t+s} = z)$ por el valor que toma X_t se obtiene

$$\mathbb{P}_x(X_{t+s} = z) = \sum_y \mathbb{P}_x(X_t = y, X_{t+s} = z) = \sum_y P_t(x, y) P_s(y, z),$$

de lo cual se sigue que P_{t+s} es el producto de las matrices P_t y P_s . □

PROPOSICIÓN 5.5 (Propiedad de Markov). *Si X es una cadena de Markov a tiempo continuo y $X_s^t = X_{t+s}$, entonces X^t también es una cadena de Markov a tiempo continuo con las mismas probabilidades de transición que X . X^t es independiente de $X_s, s \leq t$ condicionalmente a X_t .*

DEMOSTRACIÓN. Sean $t_1 < t_2 < \dots$. Entonces

$$\begin{aligned} \mathbb{P}_x(X_0^t = x_0, X_{t_1}^t = x_1, \dots, X_{t_n}^t = x_n) \\ = \mathbb{P}_x(X_{t+0} = x_0, X_{t+t_1} = x_1, \dots, X_{t+t_n} = x_n) \\ = P_t(x, x_0) P_{t_1}(x_0, x_1) \cdots P_{t_n - t_{n-1}}(x_{n-1}, x_n). \end{aligned}$$

Puesto que $P_t(x, x_0) = \mathbb{P}_x(X_0^t = x_0)$, vemos que X^t es una cadena de Markov a tiempo continuo con las mismas probabilidades de transición que X .

Para ver la independencia, hacemos un cálculo similar: sean $0 < s_1 < \dots < s_m \leq t$ y $0 < t_1 < \dots < t_n$. Al utilizar la definición de cadena de Markov, obtenemos

$$\begin{aligned} \mathbb{P}_x(X_{s_1} = x_1, \dots, X_{s_m} = x_m, X_0^t = y_0, X_{t_1}^t = y_1, \dots, X_{t_n}^t = y_n) \\ = P_{s_1}(x, x_1) P_{s_2-s_1}(x_1, x_2) \cdots P_{s_m-s_{m-1}}(x_{m-1}, x_m) \\ P_{t-s_m}(x_m, y_0) P_{t_1}(x_1, y_0) P_{t_2-t_1}(y_1, y_2) \cdots P_{t_n-t_{n-1}}(y_{n-1}, y_n). \end{aligned}$$

Al condicional por $X_t = y_0$ vemos que

$$\begin{aligned} \mathbb{P}_x(X_{s_1} = x_1, \dots, X_{s_m} = x_m, X_0^t = y_0, X_{t_1}^t = y_1, \dots, X_{t_n}^t = y_n \mid X_t = y_0) \\ \mathbb{P}_x(X_{s_1} = x_1, \dots, X_{s_m} = x_m \mid X_0^t = y_0) \mathbb{P}_{y_0}(X_{t_1} = y_1, \dots, X_{t_n} = y_n) \end{aligned}$$

por lo que X^t es independiente de X_s , $s \leq t$ condicionalmente a X_t . \square

Consideremos a los tiempos aleatorios

$$T_0 = 0, \quad T_{n+1} = \inf \{t \geq T_n : X_t \neq X_{T_n}\} \quad \text{y} \quad \zeta = \lim_{n \rightarrow \infty} T_n$$

con la convención $\inf \emptyset = \infty$.

PROPOSICIÓN 5.6. T_n y ζ son tiempos de paro respecto de la filtración canónica.

Existen tres categorías para las trayectorias en términos de estos tiempos aleatorios:

Absorción: Cuando existe n tal que $T_n < \infty = T_{n+1}$, en cuyo caso $X_t = X_{T_n}$ para toda $t \geq T_n$,

Explosión: cuando $\zeta < \infty$ y

Movimiento perpetuo: cuando $T_n < \infty$ para toda n y $\zeta = \infty$.

PROPOSICIÓN 5.7. Bajo \mathbb{P}_x , T_1 es exponencial de parámetro $c(x) \in [0, \infty)$. Si $c(x) > 0$ entonces las variables X_{T_1} y T_1 son independientes.

DEMOSTRACIÓN. Al utilizar la propiedad de Markov, vemos que

$$\begin{aligned} \mathbb{P}_x(T_1 > t + s) &= \mathbb{P}_x(T_1 > s, X_s = x, T_1(X^s) > t) \\ &= \mathbb{P}_x(T_1 > s) \mathbb{P}_x(T_1 > t) \end{aligned}$$

y por lo tanto, bajo \mathbb{P}_x , T_1 es exponencial.

Por otra parte

$$\begin{aligned} \mathbb{P}_x(\mathbf{1}_{T_1 > t} X_{T_1} = y) \\ = \mathbb{P}_x\left(\mathbf{1}_{T_1 > t}, X_t = x, X_{T_1(X^t)}^t = y\right) \\ = \mathbb{P}_x(\mathbf{1}_{T_1 > t}) \mathbb{P}_x(X_{T_1} = y) \end{aligned}$$

por lo que T_1 es independiente de $X \circ \theta_{T_1}$. \square

A $c(x)$ la interpretamos como la tasa a la que dejamos el estado x . Definamos ahora

$$P_{x,y} = \begin{cases} 0 & c(x) = 0 \\ \mathbb{P}_x(X_{T_1} = y) & c(x) \neq 0 \end{cases} \quad \text{y} \quad \alpha(x, y) = c(x) P_{x,y}.$$

A α se le conoce como la matriz de tasas de transición y la interpretación de $\alpha(x, y)$ es la tasa a la que dejamos el estado x para pasar al estado y . La matriz de tasas de transición es el parámetro que nos permitirá caracterizar a la familia Markoviana. Para verificar por qué, es necesario extender la propiedad de Markov.

TEOREMA 5.1 (Propiedad de Markov fuerte). *Sea T un tiempo de paro finito. Entonces el proceso X^T dado por $X_t^T = X_{T+t}$ es una cadena de Markov con las mismas probabilidades de transición que X que es independiente de $X_s, s \leq t$ condicionalmente a X_T y a $T \leq t$.*

La prueba es un tanto más complicada del nivel que se quiere para este curso y no se dará.

El teorema anterior nos permitirá caracterizar a la cadena de Markov en tiempo continuo en términos de la matriz de tasas de transición α , o equivalentemente, de c y P . Sea Z el proceso estocástico a tiempo discreto definido por

$$Z_n = X_{T_n}$$

si $T_n < \infty$. En el caso absorbente, definimos $Z_{n+m} = Z_n$ para toda $m \geq 1$ si $T_n < \infty = T_{n+1}$.

TEOREMA 5.2. *El proceso Z es una cadena de Markov de matriz de transición P que comienza en x bajo \mathbb{P}_x . Si $c(x) > 0$ para toda $x \in E$, condicionalmente a Z , las variables S_1, S_2, \dots con $S_i = T_i - T_{i-1}$ son independientes y exponenciales de parámetros $c(Z_0), c(Z_1), \dots$*

DEMOSTRACIÓN. Al utilizar el lema de clases de Dynkin, vemos que es suficiente verificar que

$$(6) \quad \begin{aligned} & \mathbb{P}_x(Z_1 = x_1, \dots, Z_n = x_n, S_1 > t_1, \dots, S_n > t_n) \\ &= P_{x,x_1} \cdots P_{x_{n-1},x_n} e^{-c(x)t_1} e^{-c(x_1)t_2} \cdots e^{-c(x_{n-1})t_n}. \end{aligned}$$

Esto se sigue por inducción al utilizar la propiedad de Markov fuerte. La base inductiva es la Proposición 5.7. Por otra parte, si suponemos válida la ecuación (6) vemos que al aplicar la propiedad de Markov fuerte al instante T_n y la Proposición 5.7 se sigue que

$$\begin{aligned} & \mathbb{P}_x(Z_1 = x_1, \dots, Z_{n+1} = x_{n+1}, S_1 > t_1, \dots, S_{n+1} > t_{n+1}) \\ &= \mathbb{P}_x(Z_1 = x_1, \dots, Z_n = x_n, S_1 > t_1, \dots, S_n > t_n) e^{-c(x_n)t_{n+1}} P_{x_n,x_{n+1}} \end{aligned}$$

puesto que Z_{n+1} es el primer estado al que salta X^{T_n} y S_{n+1} es el tiempo que tarda en realizar X^{T_n} su primer salto. \square

Dada una función $\alpha : E \times E \rightarrow [0, \infty)$ tal que $c(x) = \sum_y \alpha(x, y) < \infty$, podemos definir a $P_{x,y} = \alpha(x, y) / c(x)$ cuando $c(x) > 0$ y a $P_{x,y} = \delta_{x,y}$ cuando $c(x) = 0$ y preguntarnos cuándo existe una cadena de Markov a tiempo continuo cuya matriz de tasas de transición sea α . Nos abocaremos ahora a verificar que se puede construir una cadena de Markov cuya matriz de tasas de transición sea α . En efecto, sean $\tilde{S}_1, \tilde{S}_2, \dots$ variables aleatorias exponenciales de parámetro 1 y Z una cadena de Markov con matriz de transición P que comienza en X . Ahora definamos

$$T_0 = 0, \quad y \quad T_{n+1} = T_n + \tilde{S}_n / c(Z_n).$$

Consideremos al proceso X definido mediante

$$\tilde{X}_t = Z_n \quad \text{si} \quad t \in [T_n, T_{n+1}).$$

Definimos a \mathbb{P}_x como \mathbb{P} condicionada por $Z_0 = x$. Se afirma que \mathbb{P}_x es una familia Markoviana cuya matriz de tasas de transición es α . Por definición, bajo la medida de probabilidad \mathbb{P}_x es válida la ecuación (6).

TEOREMA 5.3. *La colección $(\mathbb{P}_x)_{x \in E}$ es una familia Markoviana con matriz de tasas de transición α .*

De hecho, hemos verificado este teorema en la sección anterior y la prueba en este caso en el que el espacio de estados es posiblemente infinito es muy similar.

5. Las ecuaciones backward y forward de Kolmogorov

Tal como la propiedad de Markov y de Markov fuerte nos llevan a relaciones de recurrencia para probabilidades que deseamos calcular, en tiempo continuo nos llevan a ecuaciones diferenciales. Una de ellas es la ecuación backward de Kolmogorov. Sea $(\mathbb{P}_x, x \in E)$ una familia markoviana con **probabilidades de transición** $P_t(x, y) = \mathbb{P}_x(X_t = y)$. Estas probabilidades de transición satisfacen las ecuaciones de Chapman-Kolmogorov

$$P_{t+s}(x, z) = \sum_y P_s(x, y) P_t(y, z).$$

TEOREMA 5.4 (Ecuaciones backward de Kolmogorov). *Para cualquier $x, y \in E$, las probabilidades de transición satisfacen la ecuación backward de Kolmogorov*

$$\frac{\partial}{\partial t} P_t(x, y) = \sum_{z \in E} \alpha(x, z) P_t(x, z) - P_t(x, y).$$

Dada la matriz de tasas de transición, definiremos a la matriz infinitesimal Q mediante:

$$Q_{x,y} = \begin{cases} \alpha(x, y) & x \neq y \\ -c(x) & x = y \end{cases}.$$

Entonces la ecuación backward de Kolmogorov se puede escribir como la ecuación diferencial para la matriz P_t

$$\frac{\partial}{\partial t} P_t = Q P_t.$$

Esto explica la conexión con ecuaciones diferenciales: las probabilidades de transición de una familia markoviana satisfacen una ecuación diferencial. Veremos que en el caso de espacio de estados finito, la teoría clásica de ecuaciones diferenciales lineales nos permite verificar que existe una única solución para la ecuación backward de Kolmogorov y por lo tanto nos da una manera de obtener, a veces explícitamente pero inclusive también numéricamente, a las probabilidades de transición de la familia Markoviana.

DEMOSTRACIÓN. Heurísticamente, la prueba es una aplicación de la propiedad de Markov fuerte. Sin embargo, necesitamos una verisón que también depende del tiempo. Específicamente, notemos que si $s \leq t$

$$P_t(x, y) = \mathbb{P}_x(X_t = y) = \mathbb{E}_x(P_{t-s}(X_s) y).$$

Podemos por lo tanto pensar que por la propiedad de Markov fuerte aplicada al tiempo de paro $\sigma = t \wedge T_1$ se satisface

$$(7) \quad P_t(x, y) = \mathbb{E}_x(P_{t-\sigma}(X_\sigma, y))$$

para $t > 0$. Esto es en efecto cierto pero no se ha demostrado y se sigue del hecho de que podemos aproximar al tiempo de paro σ por $\sigma^n = \lceil \sigma 2^n \rceil / 2^n$ y tomar el límite conforme $n \rightarrow \infty$ para verificar (7). Veamos cómo se aplica dicha ecuación. De (7) se deduce que:

$$P_t f(x) = \mathbb{E}_x(P_{t-\sigma} f(X_\sigma)) = f(x) e^{-\alpha(x)t} + \int_0^t \sum_y e^{-\alpha(x)s} \alpha(x, y) P_{t-s} f(y) ds.$$

Al multiplicar por $e^{\alpha(x)t}$ de ambos lados se obtiene

$$e^{\alpha(x)t} P_t f(x) = f(x) + \int_0^t e^{\alpha(x)s} \sum_y \alpha(x, y) P_s f(y) ds.$$

Finalmente, la expresión del lado derecho muestra que el lado izquierdo es derivable y por la continuidad del integrando vemos que

$$\frac{\partial}{\partial t} P_t f(x) + c(x) P_t f(x) = \sum_y \alpha(x, y) P_t f(y). \quad \square$$

Ahora recordemos que en espacio de estados finito, digamos de cardinalidad n , estas ecuaciones backward admiten una solución en términos de la matriz infinitesimal Q y esto nos permitirá introducir a las ecuaciones forward. Cuando I es finito, debemos resolver el sistema de ecuaciones

$$\frac{\partial}{\partial t} P_t = Q P_t.$$

Este es un sistema lineal y si Q fuera de tamaño 1×1 , tendría como única solución a la función exponencial. Lo mismo sucede en el caso $n \times n$, si definimos a la matriz

$$e^{tQ} = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!}$$

para cualquier $t \in \mathbb{R}$. La convergencia de la serie se sigue pues la sucesión

$$\sum_{n=0}^N \frac{t^n Q^n}{n!}.$$

es de Cauchy cuando se utiliza la norma

$$\|Q\| = \max \{\|Qx\| : x \in \mathbb{R}^n, \|x\| = 1\}.$$

En efecto, puesto que esta norma es submultiplicativa, se sigue que:

$$\sup_{n \geq m} \left\| \sum_{k=0}^n \frac{t^k Q^k}{k!} - \sum_{k=0}^m \frac{t^k Q^k}{k!} \right\| \leq \sum_{k=m+1}^{\infty} \frac{|t|^k \|Q\|^k}{k!} \rightarrow_{m \rightarrow \infty} 0.$$

Ahora veremos que e^{tQ} , $t \geq 0$ es la única solución a las ecuaciones backward y forward de Kolmogorov:

$$\frac{\partial}{\partial t} e^{tQ} = Q e^{tQ} \quad \text{y} \quad \frac{\partial}{\partial t} e^{tQ} = e^{tQ} Q.$$

Además, satisfacen las ecuaciones de Chapman-Kolmogorov

$$e^{(s+t)Q} = e^{sQ} e^{tQ}.$$

En efecto, Chapman-Kolmogorov se sigue de la definición de la exponencial de una matriz. Por otra parte, podemos derivar término a término la serie de potencias (cuyo radio de convergencia, con la norma matricial, es infinito) para obtener

$$\frac{\partial}{\partial t} e^{tQ} = \sum_{k=1}^{\infty} \frac{t^{k-1} Q^k}{(k-1)!} = \sum_{k=0}^{\infty} \frac{t^k Q^{k+1}}{k!} = \begin{cases} Q e^{tQ} \\ e^{tQ} Q \end{cases},$$

lo cual muestra que se satisfacen las ecuaciones backward y forward. Además $e^{0Q} = \text{Id}$. Para la unicidad de la solución a estas últimas, supongamos que P_t , $t \geq 0$ es una colección de matrices en \mathbb{R}^n que satisface las ecuaciones backward (el argumento para las forward es similar) y tal que $P_0 = \text{Id}$. Notemos que la inversa de e^{tQ} es e^{-tQ} . Entonces

$$\frac{\partial}{\partial t} e^{-tQ} P_t = -Q e^{-tQ} P_t + e^{-tQ} Q P_t = -Q e^{-tQ} P_t + Q e^{-tQ} P_t = 0.$$

Por lo tanto $e^{-tQ} P_t$ es constante y como la constante es $P_0 = \text{Id}$, vemos que $P_t = e^{tQ}$.

6. Distribuciones invariantes

Ahora pasaremos al estudio de las distribuciones invariantes para familias markovianas. La liga entre el tiempo continuo y discreto nos lo proporciona el siguiente resultado que se sigue de las ecuaciones backward de Kolmogorov.

DEFINICIÓN. Decimos que una distribución ν en E (identificada con la colección numérica $\nu_x = \nu(\{x\})$) es **invariante** para una familia Markoviana si

$$\sum_x \nu_x P_t(x, y) = \nu_y.$$

En otras palabras, la distribución ν es invariante si la distribución de X_t bajo $\mathbb{P}_\nu = \sum_x \nu_x \mathbb{P}_x$ es igual a la de X_0 .

TEOREMA 5.5. *Una medida de probabilidad ν tal que $\sum_x \nu_x c(x) < \infty$ es invariante para X si y sólo si $c\nu = (c_x \nu_x, x \in E)$ es invariante para la cadena asociada.*

DEMOSTRACIÓN. Por la ecuación backward de Kolmogorov y el teorema de Fubini-Tonelli se sigue que

$$(8) \quad \sum_x \nu_x P_t(x, z) = \sum_x \nu_x P_0(x, z) + \int_0^t \sum_y \sum_x \nu_x \alpha(x, y) [P_s(y, z) - P(x, z)] ds.$$

Así, ν es invariante si y sólo si la integral del lado derecho es igual a 0 para cualquier t . Escribamos a $\alpha(x, y) = c(x) P(x, y)$, donde P es la matriz de transición de la cadena asociada. Puesto que $\sum_x c_x \nu_x$ y $t \mapsto P_t(x, y)$ es continua, podemos aplicar el teorema de convergencia dominada para concluir que el integrando en el lado derecho de (8) es continuo. Por lo tanto, ν es invariante si y sólo si

$$\begin{aligned} 0 &= \sum_y \sum_x \nu_x \alpha(x, y) [P_0(y, z) - P_0(x, z)] \\ &= \sum_y \sum_x c(x) \nu_x P_{x, y} [\mathbf{1}_{y=z} - \mathbf{1}_{x=z}] = \sum_x c(x) \nu_x P_{x, z} - c_z \nu_z \end{aligned}$$

En otras palabras, $c\nu$ es invariante para la cadena asociada. □

Recordemos que en el teorema fundamental de convergencia para cadenas de Markov (en tiempo discreto) la periodicidad juega un rol importante. Ahora veremos que en tiempo continuo, en cierto sentido el proceso ya es periódico.

PROPOSICIÓN 5.8. *$P_t(x, y) > 0$ para alguna $t > 0$ si y sólo si $P_t(x, y) > 0$ para toda $t > 0$. En particular $P_t(x, x) > 0$ para toda $t \geq 0$.*

DEMOSTRACIÓN. El caso particular es simple:

$$P_t(x, x) \geq \mathbb{P}_x(T_1 > t) > 0.$$

Por otra parte, si $y \neq x$ y para la cadena asociada se accede de x a y entonces existen $x_0, \dots, x_n \in E$ tales que $x_0 = x$, $x_n = y$ y $x_{k+1} \neq x_k$ para los cuales

$$P_{x,x_1} \cdots P_{x_{n-1},y} > 0.$$

En particular, se tiene que $c(x_i) > 0$ para $i < n$.

Si S_1, \dots, S_{n+1} son exponenciales de parámetro 1 independientes entonces

$$P_t(x, y) \geq \mathbb{P} \left(\sum_{k \leq n} \frac{S_k}{c(x_{k-1})} \leq t < \sum_{k \leq n+1} \frac{S_k}{c(x_{k-1})} \right) P(x_0, x_1) \cdots P_{x_{n-1}, y} > 0.$$

(Sólo se debe tener cuidado si $c(y) = 0$.)

Finalmente, si de x no se accede a y para la cadena asociada Z entonces

$$\mathbb{P}_x(X_t \neq y \text{ para toda } t \geq 0) = \mathbb{P}_x(Z_n \neq y \text{ para toda } n \geq 0) = 1. \quad \square$$

Una familia markoviana es **irreducible** si $\mathbb{P}_x(X_t = y) > 0$ para toda $t > 0$ y toda $y \in E$.

PROPOSICIÓN 5.9. *Si la cadena asociada a una familia markoviana irreducible es recurrente entonces no hay explosión.*

Lo anterior nos dice que los conjuntos $\{t \geq 0 : X_t = y\}$ y $\{n \in \mathbb{N} : Z_n = y\}$ o son ambos acotados o ambos no acotados para familias markovianas irreducibles. En el primer caso hablamos de **transitoriedad** y en el segundo de **recurrencia**.

DEMOSTRACIÓN. Sólo hay que notar que si $\mathbb{P}_x(Z_n = x \text{ i.o.})$ entonces o Z_n se absorbe en x (que sucede si y sólo si $c(x) > 0$ y no es compatible con la irreducibilidad de la cadena) ó $c(x) > 0$ y

$$\sum_n \frac{1}{c(Z_n)} \geq \infty / c(x) = \infty$$

\mathbb{P}_x -casi seguramente, en cuyo caso, al condicionar con Z , vemos que no hay explosión. (Recordemos que si τ_i son exponenciales independientes de parámetro λ_i entonces $\sum \tau_i = \infty$ casi seguramente si y sólo si $\sum 1/\lambda_i = \infty$.) \square

TEOREMA 5.6. *Si (\mathbb{P}_x) es una familia markoviana irreducible entonces son equivalentes:*

- (1) *Existe una única distribución invariante ν para la familia que satisface $\nu_x > 0$ para toda $x \in E$ y para cualquier distribución inicial μ :*

$$\lim_{t \rightarrow \infty} \sum_x |\mathbb{P}_\nu(X_t = y) - \nu_y| = 0.$$

- (2) *Para alguna $h > 0$, la sucesión de variables aleatorias $(X_{nh}, n \in \mathbb{N})$ es una cadena de Markov positivo recurrente.*

En caso contrario, no existe ninguna distribución invariante y $\mathbb{P}_x(X_t = y) \rightarrow 0$ conforme $t \rightarrow \infty$.

DEMOSTRACIÓN. Sólo demostraremos la equivalencia. (La prueba completa se puede verificar en el libro de Kallenberg.)

Sea $h > 0$. Notemos que $(X_{nh}, n \geq 0)$ es una cadena de Markov con matriz de transición $P_h(x, y)$, $x, y \in E$. En efecto, vemos que

$$\mathbb{P}_x(X_h = x_1, \dots, X_{nh} = x_n) = P_h(x, x_1) P_h(x_1, x_2) \cdots P_h(x_{n-1}, x_n).$$

Si para alguna h , dicha cadena de Markov es positivo recurrente, entonces al ser irreducible y aperiódica, existe una única distribución invariante ν_h . Por otra parte, la cadena de Markov $X_{nh/2}, n \geq 0$ debe también ser positivo recurrente pues su tiempo de primer retorno está acotado por dos veces el tiempo de primer retorno de $X_{nh}, n \geq 0$, el cual es integrable. Así, existe una única distribución invariante para $X_{nh/2}$, digamos $\nu_{h/2}$ pero como ésta también es invariante para X_{nh} , vemos que $\nu_{h/2} = \nu_h$. Escribamos por lo tanto $\nu = \nu^h$. Generalizando, vemos que para cualquier racional no-negativo q , la distribución de X_{qh} bajo \mathbb{P}_ν es ν y, al aproximar a cualquier $t > 0$ por la derecha por reales de la forma qh , vemos que ν es invariante para la familia markoviana. Para mostrar la convergencia en variación, notemos que, de acuerdo al teorema fundamental de convergencia para cadenas de Markov, se tiene que

$$\sum_x |P_{nh}(x, y) - \nu_y| \rightarrow 0$$

conforme $n \rightarrow \infty$. Por lo tanto, al escribir a t (de manera única) en la forma $nh + r$ con $n \in \mathbb{N}$ y $0 \leq r < h$, las ecuaciones de Chapman-Kolmogorov y la invariancia de ν nos dicen que

$$\sum_x |P_t(x, y) - \nu_y| \leq \sum_x \sum_y |P_{nh}(x, z) - \nu_z| P_r(z, y) \rightarrow 0.$$

Por lo tanto, el teorema de convergencia dominada nos permite afirmar que

$$\lim_{t \rightarrow \infty} \sum_x |\mathbb{P}_\nu(X_t = y) - \nu_y| = 0.$$

Por otra parte, si existe una distribución invariante ν para la familia markoviana, entonces ν es una distribución invariante para X_{nh} , lo que implica que esta es positivo recurrente para cualquier $h > 0$. \square

Finalmente, pasamos a la relación entre el comportamiento asintótico de la probabilidad de transición y los tiempos medios de recurrencia. Sea

$$\tilde{T}^y = \min \{t > T_1 : X_t = y\}.$$

TEOREMA 5.7. Si y no es absorbente entonces

$$\lim_{t \rightarrow \infty} P_t(x, y) = \frac{\mathbb{P}_x(T^y < \infty)}{c(y) \mathbb{E}_y(\tilde{T}^y)}.$$

DEMOSTRACIÓN. Sólo podremos probarlo en el caso transitorio y positivo recurrente. En el caso nulo recurrente, tendremos la convergencia en el sentido de Cesàro.

Primero nos concentraremos en el caso $x = y$. Si y es transitorio entonces $\mathbb{E}_y(\tilde{T}^y) = \infty$ y por lo tanto el enunciado es válido. Si por otra parte y es positivo recurrente y nos concentramos en su clase de comunicación, esta será irreducible y sabemos que $P_t(x, y)$ converge a ν_y donde ν es la distribución invariante única (en la clase de comunicación de y). Así, los tiempos medios de ocupación

$$L_t = \frac{1}{t} \int_0^t \mathbf{1}_{X_s=y} ds$$

satisfacen:

$$\mathbb{E}_x(L_t) = \frac{1}{t} \int_0^t \mathbb{P}_x(X_s = y) ds = \frac{1}{t} \int_0^t P_s(x, y) ds \rightarrow_{t \rightarrow \infty} \nu_y.$$

Por otra parte, si $\tilde{T}_n^y = \tilde{T}^y + \tilde{T}_{n-1}^y(X^{T^y})$ representa al tiempo del enésimo retorno de la cadena al estado y , la propiedad de Markov fuerte nos dice que \tilde{T}_n^y es una caminata aleatoria. Como $\tilde{T}^y(X^{T_1})$ se puede acotar en términos del tiempo de visita a y por la cadena X_{T_1+nh} , $n \geq 0$, que es finito por ser positivo recurrente, vemos que $\mathbb{E}_x(\tilde{T}^y) < \infty$, por lo que podemos aplicar la ley fuerte de los grandes números y deducir que bajo \mathbb{P}_y se tiene que $\tilde{T}_n^y/n \rightarrow \mathbb{E}_y(\tilde{T}_n^y)$. Por esto, observamos que

$$\frac{L_{\tilde{T}_n^y}}{\tilde{T}_n^y} = \frac{\xi_1 + \cdots + \xi_n}{\tilde{T}_n^y} \rightarrow \frac{1}{c(y) \mathbb{E}_x(T_y)}$$

donde $\xi_i = T_1 \circ \theta_{\tilde{T}_n^y}$ son variables exponenciales de parámetro $c(y)$ (a la cuales también les aplicamos la ley fuerte de los grandes números). Finalmente, por convergencia dominada vemos que

$$\mathbb{E}_x(L_t) \rightarrow \frac{1}{c(y) \mathbb{E}_x(T_y)},$$

lo cual prueba el resultado en este caso. □