Pablo Ferrari, DM-UBA

Tópicos de Procesos estocasticos

Primer cuatrimestre 2018

Actualizadas el 3 de abril de 2018

Definición: Proceso estocástico Dado un conjunto I de índices, un proceso estocástico es una familia de variables aleatorias $(X_i : i \in I)$.

Cadenas de Markov

Esta sección está baseada en el libro de Durrett [2].

Ejemplo. La ruina del jugador A cada jugada gano 1 peso con probabilidad $p = 0{,}4$ o pierdo 1 con proba $1 - p = 0{,}6$. Me retiro cuando mi fortuna llega a N (mi decisión) o a 0 (decisión del casino).

$X_n =$ fortuna en el instante n. Es una variable aleatoria cuya distribución depende del pasado. Propiedad de Markov: si $0 < i < N$,

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p$$

Decimos que $X_0, X_1, \dots$ es una *cadena de Markov* con *matriz de transición* $p(i, j)$ si

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p(i, j)$$

"dado el presente $\{X_n = i\}$, el futuro $\{X_{n+1} = j\}$ es independiente del pasado $\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}$"

Ejercicio Demuestre que si X_n es una cadena de Markov con matriz $p(i, j)$, entonces

$$P(X_{n+1} = j | X_n = i) = p(i, j)$$

La matriz de transición de la cadena ruina del jugador para $N = 5$ es

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0{,}6 & 0 & 0{,}4 & 0 & 0 & 0 \\ 0 & 0{,}6 & 0 & 0{,}4 & 0 & 0 \\ 0 & 0 & 0{,}6 & 0 & 0{,}4 & 0 \\ 0 & 0 & 0 & 0{,}6 & 0 & 0{,}4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

El *espacio de estados* en este caso es $\{0, \dots, 5\}$; i, j son llamados estados (de la cadena).

En forma de grafo:

Urna de Ehrenfest N bolas distribuídas en dos urnas. En cada paso elijo una bola al azar y la paso a la otra urna.

$X_n =$ número de bolas en la primera urna.

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{N - i}{N}.$$

$$P(X_{n+1} = i - 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{i}{N}.$$

Para $N = 5$ la matriz es

$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 4/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 2/5 & 0 \\ 0 & 0 & 0 & 4/5 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Pregunta: cuantas bolas habrá en la primera urna si repetimos el experimento muchas veces?

Propiedades de la matriz de transición

1) $p(i,j) \geq 0$ (entradas no negativas)

2) $\sum_j p(i,j) = 1$ para todo i. (filas suman 1)

En este caso diremos que p es una matriz *estocástica.*

Construcción de una cadena de markov a partir de una matriz estocástica Para cada matriz p con esas propiedades hay una cadena de Markov que la tiene como matriz de transición.

Una forma de demostrar esto es construir la cadena como función de variables aleatorias uniformes. Sean $U_1, U_2, \dots$ una sucesión de variables uniformes en $[0,1]$ independientes.

Para cada estado x defina una partición $(J(x,y) : y$ en el espacio de estados$)$ del intervalo $[0,1]$ tal que la longitud $|J(x,y)|$ del intervalo $J(x,y)$ sea igual a $p(x,y)$. Esto es posible porque como $\sum_y p(x,y) = 1$, entonces los intervalos J se pueden elegir disjuntos con longitud total 1. Observe que como U_n es uniforme en $[0,1]$, entonces

$$P(U_n \in J(x,y)) = p(x,y) \tag{1}$$

la probabilidad de que la uniforme pertenezca a un intervalo es la longitud del mismo.

Fije $X_0 = x_0$ dado y defina iterativamente

$$X_{n+1} = \sum_y y \mathbf{1}\{U_{n+1} \in J(X_n, y)\}. \tag{2}$$

Como los intervalos J forman una partición, esa suma tiene siempre un único término no nulo y vale que

$$X_{n+1} = y \qquad \text{si y sólo si} \qquad U_{n+1} \in J(X_n, y). \tag{3}$$

Lema 1 *El proceso definido por* (2) *es Markov con matriz* $p(x,y)$.

Demostración

$$\begin{aligned}
P(X_{n+1} &= y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = 0) \\
&= P(U_{n+1} \in J(X_n, y) | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = 0) \qquad \text{por (3)} \\
&= P(U_{n+1} \in J(x, y) | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = 0) \\
&= P(U_{n+1} \in J(x,y)) = p(x,y)
\end{aligned}$$

porque el evento en el condicionante depende de $x_0, U_1, \dots, U_n$ mientras que el evento $\{U_{n+1} \in J(x,y)\}$ depende sólo de U_{n+1} que es independiente de las otras uniformes. La última identidad viene de (1). □

Cadena del clima 1= lluvia, 2 = sol.

$$p = \begin{pmatrix} 0{,}6 & 0{,}4 \\ 0{,}2 & 0{,}8 \end{pmatrix}$$

cual es la proporción de días lluviosos a largo plazo?

Mobilidad social 1 = clase baja, 2 = clase media, 3 = clase alta

$$\begin{pmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{pmatrix}$$

Preferencia de marcas

Inventario Política de inventario s, S. Cuando la existencia de mercadería al final del día es menor o igual a s, ordenamos nueva mercadería para llevarlo a S, que llega temprano al día siguiente.

$X_n :=$ mercadería en existencia al final del día n.

$D_n :=$ demanda del día n.

Notación $x^+ = \max\{x, 0\}$.

$$X_{n+1} = \begin{cases} (X_n - D_{n+1})^+ & \text{si } X_n > s \\ (S - D_{n+1})^+ & \text{si } X_n \leq s \end{cases} \tag{4}$$

Ejercicio: Para $s = 1$, $S = 5$ y D_n toma los valores $0, 1, 2, 3$ con probabilidades respectivas $0{,}3, 0{,}4, 0{,}2, 0{,}1$. Escriba la matriz de transición en este caso.

Ramificación. Galton Watson Cada individuo de la n-ésima generacion produce k nuevos individuos con probabilidad p_k.

$X_n =$ número de individuos en la generación n.

$p(0, 0) = 0$ (absorbida en 0); $p(i, j) = P(Y_1 + \cdots + Y_i = j)$, donde $P(Y_m = k) = p_k$ (dado)

Wright-Fisher Una población tiene N individuos que tienen dos tipos posibles A o a. Para obtener los tipos de la generación siguiente se muestrea al azar de la generación presente. Sea $X_n =$ número de individuos de tipo A en el instante n. Entonces

$$p(i, j) = P(B(N, i/N) = j) = \binom{N}{j} (i/N)^j (1 - i/N)^{N-j}.$$

donde B es una variable aleatoria binomial con los parámetros indicados. De hecho, podemos pensar el número de A obtenido es el resultado de N ensayos independientes Bernoulli con parámetro i/N.

Observe que $p(0, 0) = p(N, N) = 1$, por lo tanto 0 y N son estados *absorbentes*. Si la cadena llega a un estado absorbente decimos que la cadena se *fija*.

Cual es la probabilidad que la cadena se fije en un número finito de pasos?

Se puede tambien pensar en *mutaciones*: Con probabilidad u cada A se transforma en a y con probabilidad v cada a se transforma en A. La probabilidad de producir un A en la $(n+1)$-ésima generación dado que hay i A en la n-ésima es

$$\rho_i = v\frac{N - i}{N} + u\frac{i}{N}$$

La matriz en ese caso queda

$$p(i, j) = P(B(N, \rho_i) = j) = \binom{N}{j} \rho_i^j (1 - \rho_i)^{N-j}$$

Si u y v son positivos, nos preguntamos si la población se estabiliza en algún valor.

Probabilidades de transición a varios pasos

Vimos que $p(i, j) = P(X_{n+1} = j | X_n = i)$.

Nos interesa calcular que pasa después de varios pasos:

$$p^m(i, j) := P(X_{n+m} = j | X_n = i)$$

Veamos qué pasa en el ejemplo de mobilidad social con estados 1,2,3:

$$\begin{pmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{pmatrix}$$

Si mis padres son clase media 2, cual es la probabilidad que yo sea clase baja 1 y mis hijos clase alta?

$$p(2,1)p(1,3)$$

Veamos: calculalemos preliminarmente

$$\begin{aligned}
&P(X_1 = 1, X_2 = 3|X_0 = 2)\\
&= \frac{P(X_1 = 1, X_2 = 3, X_0 = 2)}{P(X_0 = 2)}\\
&= \frac{P(X_2 = 3|X_0 = 2, X_1 = 1)\, P(X_1 = 1, X_0 = 2)}{P(X_0 = 2)}\\
&= \frac{P(X_2 = 3|X_1 = 1)\, P(X_2 = 1|X_0 = 2)P(X_0 = 2)}{P(X_0 = 2)}\\
&= p(2,1)p(1,3).
\end{aligned}$$

Cual es la probabilidad que mis hijos sean clase 1 dado que mis padres son clase 2?

$$\begin{aligned}
P(X_2 = 3|X_0 = 2) &= \sum_{j=1}^{3} P(X_1 = j, X_2 = 3|X_0 = 2) &(5)\\
&= \sum_{j=1}^{3} p(2,j)p(j,3). &(6)
\end{aligned}$$

Es la entrada $(2,3)$ de la matriz $p \times p$ (producto matricial).

Ecuaciones de Chapman-Kolmogorov

$$p^{n+m}(x,y) = \sum_z p^n(x,z)p^m(z,y), \qquad 0 \le k \le n.$$

Corolario 2 *Las probabilidades de transición a m pasos son iguales a la potencia m de la matriz de transición p. Es decir*

$$p^m(i,j) = (p \times \cdots \times p)(i,j) \qquad\qquad (m \text{ veces}).$$

Demostración Es un corolario de Chapman-Kolmogorov.

$$p^{m+1}(i,j) = \sum_k p^m(i,k)p^1(k,j). \qquad \square$$

Demostración [de Chapman-Kolmogorov] Usando la partición $\{\{X_n = z\}, z\}$, por probabilidad total,

$$\begin{aligned}
p^{m+n}(i,j) &= \sum_z P(X_{n+m} = y, X_n = z|X_0 = x)\\
&= \sum_z \frac{P(X_{n+m} = y, X_n = z, X_0 = x)}{P(X_0 = x)}\\
&= \sum_z \frac{P(X_{n+m} = y|X_n = z, X_0 = x)P(X_n = z, X_0 = x)}{P(X_0 = x)}\\
&= \sum_z P(X_{n+m} = y|X_n = z)P(X_n = z|X_0 = x) \qquad \text{por Markov}\\
&= \sum_z p^n(x,z)p^m(z,y). \qquad \square
\end{aligned}$$

Ejemplo: ruina del jugador. Ver Durrett.

Clasificación de estados

Notación: $P_x(A) = P(A|X_0 = x)$. Defina

$$T_y = \text{mín}\{n \geq 1 : X_n = y\}$$

es el instante de la primera visita a y después del tiempo cero. Defina:

$$\rho_{yy} := P_y(T_y < \infty),$$

la probabilidad que la cadena vuelva a y después del tiempo 0.

Queremos probar que la probabilidad que la cadena vuelva a y por lo menos dos veces es ρ_{yy}^2.

Decimos que T es un *tiempo de parada* si el evento $\{T = n\}$ es función del vector $(X_0, \ldots, X_n\}$ (no depende del futuro).

T_y es un tiempo de parada: para $n \geq 1$,

$$\{T_y = n\} = \{X_1 \neq y, \ldots, X_{n-1} \neq y, X_n = y\}$$

La propiedad de Markov vale en tiempos de parada:

Teorema 3 (Propiedad fuerte de Markov) *Si T es un tiempo de parada y $X_T = y$, entonces $X_{T+k}, k \geq 0$ tiene la misma distribución que $X_k, k \geq 0$ con $X_0 = y$.*

Demostración Vamos a demostrar preliminarmente que

$$P(X_{T+1} = z|T = n, X_T = y) = p(y, z) \tag{7}$$

Sea V_n el conjunto de vectores $(x_0, \ldots, x_n)$ tal que si $(X_0, \ldots, X_n) \in V_n$, entonces $T = n$ y $X_T = y$. En términos de eventos

$$\{T = n, X_T = y\} = \{(X_0, \ldots, X_n) \in V_n\}$$

Calculemos

$$\begin{aligned} &P(X_{T+1} = z, X_T = y, T = n) \\ &\quad = \sum_{x \in V_n} P(X_{n+1} = z, X_0 = x_0, \ldots, X_n = x_n) \\ &= \sum_{x \in V_n} P(X_{n+1} = z|X_0 = x_0, \ldots, X_n = x_n) P(X_0 = x_0, \ldots, X_n = x_n) \end{aligned}$$

Como $x \in V_n$, $T = n$ y $x_n = y$. Por lo tanto, usando la propiedad de Markov, lo de arriba es

$$p(y, z) \sum_{x \in V_n} P(X_0 = x_0, \ldots, X_n = x_n) = p(y, z) P(T = n, X_T = y)$$

y dividiendo por la última probabilidad, obtenemos (7).

Si sumamos (8) en n obtenemos

$$P(X_{T+1} = z|X_T = y) = p(y, z), \tag{8}$$

etcétera. $\square$

Definamos $T_y^1 = T_y$ y para $k \geq 2$,

$$T_y^k = \text{mín}\{n > T_y^{k-1} : X_n = y\}$$

el instante de la k-ésima visita a y.

Usando la propiedad fuerte de Markov, la probabilidad que X_n visite por lo menos $k + 1$ veces el estado y dado que lo visitó k veces es ρ_{yy}. Usando inducción, obtenemos

$$P(T_y^k < \infty) = \rho_{yy}^k.$$

Esto divide los estados en dos clases:

1) $\rho_{yy} < 1$. La probabilidad de visitar exactamente k veces y es una variable aleatoria geométrica con parametro $1 - \rho_{yy} > 0$. Por lo tanto finita. En este caso el estado y es llamado *transitorio.*

2) $\rho_{yy} = 1$. En este caso la cadena empezando por y va a visitar el estado y infinitas veces. Decimos que y es *recurrente.*

Ejemplos: en la **ruina del jugador** los estados 0 y 5 son recurrentes y 1,2,3,4 son estados transitorios.

En **mobilidad social** todos los estados son recurrentes porque hay una probabilidad de al menos 0,1 de alcanzar un estado empezando de cualquier otro estado. Por lo tanto,

$$P(T_x > n) \leq (0{,}9)^n \tag{9}$$

que se va a 0 cuando $n \to \infty$.

Esto se puede generalizar:

Lema 4 *Si $P_x(T_y \leq k) \geq \alpha > 0$ para todo estado x, entonces*

$$P_x(T_y > nk) \leq (1 - \alpha)^n$$

Identificando estados recurrentes y transitorios Decimos que x *comunica* con y si $\rho_{xy} := P_x(T_y < \infty) > 0$. En este caso escribimos $x \to y$.

Lema 5 *Si $x \to y$ e $y \to z$ entonces $x \to z$.*

Demostración Si $x \to y$, entonces hay un m tal que $p^m(x, y) > 0$. Por la misma razón, hay un n tal que $p^n(y, z) > 0$. Como por Chapman-Kolmogorov $p^{m+n}(x, z) \geq p^m(x, y)p^n(y, z)$, concluimos que $x \to z$. $\square$

Lema 6 *Si $\rho_{xy} > 0$ pero $\rho_{yx} < 1$, entonces x es transitorio.*

Demostración Como $\rho_{xy} > 0$, existe un número natural k y estados $y_1, \ldots, y_k$, todos distintos de x y de y tales que $p(x, y_1)p(y_1, y_2)\ldots p(y_k, y) > 0$. Entonces

$$P_x(T_x = \infty) = p(x, y_1)p(y_1, y_2)\ldots p(y_k, y)(1 - \rho_{yx}) > 0$$

que implica que x es transitorio. $\square$

Lema 7 *Si x es recurrente y $\rho_{xy} > 0$ entonces $\rho_{yx} = 1$.*

Demostración Por absurdo, si $\rho_{yx} < 1$ entonces el lema anterior implicaría que x es transitorio. $\square$

Dada una matriz de transición $p(x, y)$, decimos que un conjunto de estados A es *cerrado* si no se puede salir, es decir si $p(x, y) = 0$ para todo $x \in A$ e $y \in A^c$.

Un conjunto B de eventos es *irreducible* si cualquier par de estados $x, y \in B$ comunican ($x \to y$ e $y \to x$).

Teorema 8 (Importante) *Si un conjunto finito de estados C es cerrado e irreducible, entonces todos los estados en C son recurrentes.*

Veremos la demostración enseguida. Como consecuencia del Teorema 8, tenemos

Teorema 9 *Si el espacio de estados S es finito, entonces*

$$S = T\dot{\cup}R_1\dot{\cup}\ldots\dot{\cup}R_k, \tag{10}$$

donde T es un conjunto de estados transitorios y los R_i son conjuntos cerrados e irreducibles de estados recurrentes.

Demostración Sea

$$T := \{x \in S : \exists y \text{ con } x \to y \text{ pero } y \not\to x\} \tag{11}$$

el conjunto de estados transitorios por el teorema anterior. Vamos a ver que los restantes estados son recurrentes.

Sea $x \in S \setminus T$ y defina $C_x := \{y : x \to y\}$. Como x no es transitorio, vale que los y en C_x satisfacen tambien $y \to x$.

C_x es cerrado. Para verlo note que si $y \in C_x$ y $y \to z$, entonces $x \to z$ y $z \in C_x$.

C_x irreducible. Para verlo, tomemos $y, z \in C_x$. Como $y \to x$ y $x \to z$, tenemos $y \to z$.

Como C_x es cerrado e irreducible, entonces todos sus estados son recurrentes, por el teorema precedente.

Denote $R_1 = C_x$ e iterativamente elija $w \in S \setminus (T \cup R_1)$ (si hay, si no, listo) y repita el procedimiento. □

Demostración del Teorema 8 Basta demostrar los lemas siguientes:

Lema 10 *Si x es recurrente y $x \to y$ entonces y es recurrente.*

Lema 11 *Un conjunto finito y cerrado de estados tiene por lo menos un estado recurrente.*

Recurrencia y número esperado de visitas Defina

$$N(y) := \sum_{n\geq 1} \mathbf{1}\{X_n = y\}, \tag{12}$$

el número de visitas a y en tiempos positivos. Sacando esperanzas,

$$E_x N(y) = \sum_{n\geq 1} p^n(x, y) \tag{13}$$

Defina $T_y^1 = T_t$ e iterativamente

$$T_y^k := \min\{n > T_y^{k-1} : X_n = y\}$$

el instante de la k-ésima visita a y. Recordemos que $\rho_{xy} = P_x(T_y < \infty)$. Por la probabilidad fuerte de Markov,

$$P_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}.$$

Lema 12

$$E_x N(y) = \frac{\rho_{xy}}{1 - \rho_{yy}} \tag{14}$$

Demostración Sabemos que para X entera no negativa vale

$$EX = \sum_{k\geq 1} P(X \geq k)$$

(para demostrarlo escriba $X = \sum_{k\geq 1} \mathbf{1}\{X \geq k\}$ y saque esperanzas). Además,

$$\{N(y) \geq k\} = \{T_y^k < \infty\}$$

Entonces

$$E_x N(y) = \sum_{k\geq 1} P_x(N(y) \geq k) = \rho_{xy} \sum_{k\geq 1} \rho_{yy}^{k-1} = \frac{\rho_{xy}}{1-\rho_{yy}}. \quad \square$$

Teorema 13 *y es recurrente si y sólo si*

$$\sum_{n\geq 0} p^n(y,y) = E_y N(y) = \infty.$$

Demostración La primera igualdad sigue del lema 13. La segunda del Lema 14 y del hecho que la esperanza es igual a ∞ si y sólo si $\rho_{yy} = 1$. $\square$

Demostración del Lema 10. Supongamos x recurrente y $\rho_{xy} > 0$. Por el Lema 7 tenemos $\rho_{yx} > 0$. Entonces existen j y ℓ tales que $p^j(y,x) > 0$, $p^\ell(x,y) > 0$ y

$$p^{j+k+\ell}(y,y) \geq p^j(y,x) p^k(x,x) p^\ell(x,y) \tag{15}$$

y así

$$\sum_k p^{j+k+\ell}(y,y) \geq p^j(y,x) \sum_k p^k(x,x) p^\ell(x,y) = \infty \tag{16}$$

porque x es recurrente. Pero eso implica que $\sum_m p^m(y,y) = \infty$. Teorema 13 implica que y es recurrente. $\square$

Demostración del Lema 11. Si todos los estados en C son transitorios, $E_x N(y) < \infty$ para todo x e y.

$$\infty > \sum_{y\in C} E_x N(y) = \sum_{y\in C} \sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty} \sum_{y\in C} p^n(x,y) = \sum_{n=1}^{\infty} 1 = \infty. \tag{17}$$

La contradicción demuestra el lema. $\square$

Distribuciones estacionarias

Vamos a estudiar el comportamiento asintótico de cadenas de Markov. En particular vamos a demostrar que si p es irreducible y aperiódica, entonces

$$p^n(x,y) \to \pi(y)$$

π se va a llamar distribución estacionaria.

Estados iniciales aleatorios Supongamos que X_0 es una variable aleatoria con distribución q:

$$P(X_0 = k) = q(k), \qquad \sum_k q(k) = 1.$$

Calculemos la probabilidad de X_n tomando en cuenta la aleatoriedad de X_0:

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_n = j, X_0 = i) \\ &= \sum_i P(X_n = j | X_0 = i) P(X_0 = i) \end{aligned}$$

Es decir,

$$P(X_n = j) = \sum_i q(i) p^n(i,j)$$

Multiplicamos el vector fila q por la matriz p^n.

Clima Supongamos que inicialmente la probabilidad de 1 (sol) es 0,3 y la de 2 (lluvia) es 0,7.

Multiplicando el vector $(0{,}3, 0{,}7)$ por la matriz de la cadena clima

$$(0{,}3, 0{,}7)\begin{pmatrix} 0{,}6 & 0{,}4 \\ 0{,}2 & 0{,}8 \end{pmatrix} = (0{,}32, 0{,}68)$$

La proba de sol al día siguiente es 0,32.

Mobilidad social Si inicialmente la distribución es $q = (0{,}5, 0{,}2, 0{,}3)$, entonces

$$qp = (0{,}47, 0{,}32, 0{,}21)$$

La proporción de gente en la clase media salta de 0,2 a 0,32 después de un paso.

Distribuciónes estacionarias Si $qp = q$ decimos que q es una *distribución estacionaria. Ecuaciones de balance*:

$$q(j) = \sum_i q(i)p(i,j), \qquad i \in S.$$

Para encontrar una medida estacionaria hay que resolver ese sistema de ecuaciones lineares con incógnitas $q(i)$ y coeficientes $p(i,j)$.

Clima Para encontrar la medida estacionaria planteamos las ecuaciones:

$$(\pi_1, \pi_2) = \begin{pmatrix} 0{,}6 & 0{,}4 \\ 0{,}2 & 0{,}8 \end{pmatrix} = (\pi_1, \pi_2)$$

Es decir

$$\begin{aligned} 0{,}6\pi_1 + 0{,}2\pi_2 &= \pi_1 \\ 0{,}4\pi_1 + 0{,}8\pi_2 &= \pi_2 \end{aligned}$$

cuya solución es $(\pi_1, \pi_2) = (1/3, 2/3)$.

Dos estados Cuando el espacio de estados es $\{1, 2\}$ y la matriz p está dada por

$$p = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

la solución de las ecuaciones da

$$\pi_1 = \frac{b}{a+b}, \qquad \pi_2 = \frac{a}{a+b}$$

El flujo de 1 a 2 es el mismo que el de 2 a 1.

Mobilidad social La solución de $\pi p = \pi$ da

$$\pi = (0{,}468085, 0{,}340425, 0{,}191489).$$

Teorema 14 (Álgebra linear) *Suponga que p es una matriz irreducible de tamaño $k \times k$. Entonces hay una única solución π a las ecuaciones de balance $\pi p = \pi$ con $\sum_x \pi_x = 1$. La solucion satisface $\pi_x > 0$ para todo x.*

Demostración Vamos a dar una demostración probabilista de este teorema. □

Periodicidad Estamos interesados en calcular el límite en n de $p^n(x,y)$. Cuando y es transitorio esa probabilidad se va a cero. Nos restringimos entonces a los estados recurrentes.

Ehrenfest con $N = 3$: p es

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

Mientras p^2 es igual a

$$p^2 = \begin{pmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 7/9 & 0 & 2/9 \\ 2/9 & 0 & 7/9 & 0 \\ 0 & 2/3 & 0 & 1/3 \end{pmatrix}$$

Tendremos que para n impar p^n tendrá las mismas entradas positivas que p mientras que para n par p^n tendrá las mismas entradas positivas que p^2.

Este problema puede ocurrir en general para múltiplos de otros números.

Cadena renovación $S = \{0, 1, 2, \dots\}$

$$\begin{aligned} p(x+1, x) &= 1, \quad x \geq 1 \\ p(0, k-1) &= f_k. \end{aligned}$$

con $\sum_k f_k = 1$. Si $X_0 = 0$ y $f_5 = f_{15} = 1/2$, entonces la cadena volverá a 0 solamente en momentos múltiplos de 5.

El *período de un estado* x es el mayor número que divide todos los $n \geq 1$ tales que $p^n(x,x) > 0$, es decir,

$$\text{Período de } x = \text{MCD}\,(I_x), \quad \text{donde } I_x := \{n \geq 1 : p^n(x,x) > 0\}$$

En el ejemplo anterior,

$$I_x = \{2, 4, 6, \dots\}, \qquad \text{para todo } x.$$

es decir que el mcd es 2. En la cadena renovación $I_x = \{5, 10, 15, \dots\}$ y el período es 5 para todo x.

Cadena triángulo-cuadrado $S = \{(-1,-1), (-1,0), (0,0), (1,0), (1,-1), (0,-1)\}$

$$p = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0{,}5 & 0 & 0{,}5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Ver grafo de adyacencias. Desde 0 vuelve en tres pasos con probabilidad 1/2 o en 4 pasos con la misma probabilidad. Así 3 y 4 están en I_0.

Lema 15 *I_x es cerrado para la suma.*

Demostración Si $i, j \in I_x$, entonces $p^i(x,x) > 0$ y $p^j(x,x) > 0$. Pero por Chapman-Kolmogorov, $p^{i+j}(x,x) \geq p^i(x,x)p^j(x,x)$, que implica $i + j \in I_x$. $\square$

Usando el lema vemos que

$$I_0 = \{3, 4, 6, 7, 8, 9, 10, 11, \dots\}$$

Una vez que tenemos 3 enteros seguidos, tendremos todos los siguientes. Aquí 3 es el mínimo de I_0.

En la cadena renovación, con $f_5 = f_{12} = 1/2$ tendremos que $5, 12 \in I_0$, de tal forma que usando el lema,

$$I_0 = \{5, 10, 12, 15, 17, 20, 22, 24, 25, 27, 29, 30, 32, 34, 35, 36, 37, 39, 40, 41, 42, 43, \dots\}$$

Cuando hay 5 números consecutivos, tenemos el resto.

Lema 16 *Si x tiene período 1, entonces I_x contiene todos los enteros menos un número finito.*

Demostración Por el Lema 15, I_x es cerrado para la suma. Como x tiene período 1, $mcd(I_x) = 1$. Un resultado de teoría de números, Teorema 1.1 en Bremaud [1], página 418, implica que I_x contiene a todos los naturales menos un número finito. Ver también Haggstrom [4], página 26. $\square$

Una cadena es *aperiódica* si todos sus estados tienen período 1.

Lema 17 *Si $p(x, x) > 0$, entonces x tiene período 1.*

Demostración Si $p(x, x) > 0$, entonces $1 \in I_x$, por lo tanto el mcd $I_x = 1$. $\square$

Ejemplos: clima, mobilidad social, etc.

Lema 18 *Si $\rho_{xy} > 0$ y $\rho_{yx} > 0$, entonces x e y tienen el mismo período.*

Demostración Suponga que c y d son los períodos de x e y, respectivamente. Sean k y m tales que $p^k(x, y) > 0$ y $p^m(y, x) > 0$; son finitos por hipótesis. Como

$$p^{k+m}(x, x) \geq p^k(x, y)p^m(y, x) > 0$$

$k + m$ es múltiplo de c. Sea $\ell \in I_y$, es decir tal que $p^\ell(y, y) > 0$. Como

$$p^{k+\ell+m}(x, x) \geq p^k(x, y)p^\ell(y, y)p^m(y, x) > 0,$$

$k + \ell + m \in I_x$ y $k + \ell + m$ tiene que ser múltiplo de c. Pero como $k + m$ ya es múltiplo de c, también ℓ tiene que ser múltiplo de c. Como d es el mcd(I_y), tenemos que $d \geq c$. El mismo argumento muestra que $c \geq d$. $\square$

Inventario $S = \{1, 2, 3, 4, 5\}$

$$p = \begin{pmatrix} 0 & 0 & 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0 & 0 & 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}3 & 0{,}4 & 0{,}3 & 0 & 0 & 0 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 & 0 & 0 \\ 0 & 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 & 0 \\ 0 & 0 & 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \end{pmatrix}$$

Como $p(x, x) > 0$ para $x = 2, 3, 4, 5$, esos estados son aperiódicos. Como la cadena es irreducible, los otros dos estados también son aperiódicos.

Basket $S = \{1, 2, 3, 4\}$

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

1 y 4 son aperiódicos. Como la cadena es irreducible, todos los estados son aperiódicos.

Comportamiento asintótico

Teorema 19 *Sea p una matriz irreducible, aperiódica y con una medida invariante π. Entonces*

$$\lim_n p^n(x,y) = \pi(y)$$

Ejemplo 20 Sea $(X_n : n \in \mathbb{N})$ una cadena de Markov en $S = \{1,2\}$ y matriz

$$p = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \tag{18}$$

Calculemos $p^2(1,1) = P_1(X_2 = 1)$:

$$P(X_2^1 = 1) = p(1,1)p(1,1) + p(1,2)p(2,1) = a^2 + (1-a)(1-b), \tag{19}$$

que es el primer elemento de la matriz

$$p^2 = \begin{pmatrix} a^2 + (1-a)(1-b) & a(1-a) + (1-a)b \\ (1-b)a + b(1-b) & b^2 + (1-b)(1-a) \end{pmatrix} \tag{20}$$

Para valores grandes de n usamos la notación $P(n) = P(X_n^1 = 1)$. Así,

$$\begin{aligned} P(n) &= P(n-1)p(1,1) + (1 - P(n-1))p(2,1) \\ &= P(n-1)a + (1 - P(n-1))(1-b) \\ &= P(n-1)(a+b-1) + (1-b). \end{aligned} \tag{21}$$

Substituyendo $P(n-1)$ en función de $P(n-2)$, obtenemos

$$\begin{aligned} P(n) &= [P(n-2)(a+b-1) + (1-b)](a+b-1) + (1-b) \\ &= P(n-2)(a+b-1)^2 + (1-b)(a+b-1) + (1-b). \end{aligned} \tag{22}$$

Iterando,

$$P(n) = P(0)(a+b-1)^n + (1-b)\sum_{k=0}^{n-1}(a+b-1)^k. \tag{23}$$

Como $P(0) = 1$, sumando la serie geométrica tenemos

$$\begin{aligned} P(n) &= (a+b-1)^n + \frac{(1-b)[1-(a+b-1)^n]}{(1-a)+(1-b)} \\ &= \frac{1-b}{(1-a)+(1-b)} + (a+b-1)^n \frac{1-a}{(1-a)+(1-b)}. \end{aligned} \tag{24}$$

Observe que $P(n)$ converge exponencialmente rápido a

$$\frac{1-b}{(1-a)+(1-b)}. \tag{25}$$

Analogamente se puede ver que $p^n(i,j)$ converge exponencialmente rápido a las coordenadas de la matriz

$$\begin{pmatrix} \dfrac{1-b}{(1-a)+(1-b)} & \dfrac{1-a}{(1-a)+(1-b)} \\ \\ \dfrac{1-b}{(1-a)+(1-b)} & \dfrac{1-a}{(1-a)+(1-b)} \end{pmatrix} \tag{26}$$

Este es uno de los pocos ejemplos donde se puede probar este límite a mano.

Decimos que una medida $\mu(x) \geq 0$ es una *medida estacionaria* si satisface las ecuaciones de balance $\mu p = \mu$. Si S es finito, μ se puede normalizar para obtener una distribución estacionaria.

Teorema 21 *Si p es irreducible y recurrente entonces existe una medida estacionaria μ para p con $\mu(x) > 0$ para todo x.*

Sea $N_n(y) := \sum_{i=1}^n \mathbf{1}\{X_n = y\}$, el número de visitas a y hasta el instante n.

Teorema 22 (Promedios temporales) *Si p es irreducible y recurrente, entonces*

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{E_y T_y}, \qquad c.s.$$

Teorema 23 *Si p es irreducible y tiene una distribución estacionaria π, entonces*

$$\pi(y) = 1/E_y T_y.$$

Teorema 24 *Sea p es irreducible con una distribución estacionaria π. Sea $f : S \to \mathbb{R}$ tal que*

$$\sum_x |f(x)|\pi(x) < \infty.$$

Entonces

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_x f(x)\pi(x), \quad a.s.$$

Inventario Un negocio tiene un potencial de ventas de $0, 1, 2, 3$ unidades cada día, con probabilidades $0{,}3, 0{,}4, 0{,}2, 0{,}1$, respectivamente. Cada noche son ordenadas nuevas unidades que están disponibles la mañana siguiente. La venta de una unidad produce un lucro de \$12 pero almacenar una unidad non vendida cuesta \$2 por noche. Como es imposible vender 4 unidades en un día, nunca hay más de 3 unidades cada mañana.

Si llamamos D_n la demanda en el día n y usamos la política (S, s), completar las existencias hasta S si quedan s o menos unidades en el depósito, llamando X_n al número de unidades al final del día n, tendremos

$$X_{n+1} = (S - D_{n+1})^+ \,\mathbf{1}\{X_n \le s\} + (X_n - D_{n+1})^+ \,\mathbf{1}\{X_n > s\}$$

Política 2,3. Ordenamos para completar 3 cuando quedan 2 o menos unidades. Empezamos así cada día con 3 unidades.

La matriz de transición es

$$p = \begin{pmatrix} 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \end{pmatrix}$$

La distribución estacionaria es $\pi = (0{,}1, 0{,}2, 0{,}4, 0{,}3)$. Si terminamos el día con k unidades, quiere decir que vendimos $3 - k$ y almacenamos k. Nuestro ingreso asintótico por día es (usando $f(k) = 12k$),

$$\sum_k f(k)\,\pi(k) = 36(0{,}1) + 24(0{,}2) + 12(0{,}4) = 13{,}2 \text{ pesos por día}$$

y el costo por almacenamiento (usando $g(k) = 2k$),

$$\sum_k g(k)\,\pi(k) = 2(0{,}2) + 4(0{,}4) + 6(0{,}3) = 3{,}8 \text{ pesos por día}$$

El lucro neto es $13{,}2 - 3{,}8 = 9{,}4$ pesos por día.

Política 1,3. La matriz es

$$p = \begin{pmatrix} 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}3 & 0{,}4 & 0{,}3 & 0 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \end{pmatrix}$$

La distribución estacionaria es $\pi = (19/110, 30/110, 40/110, 21/110)$.

Haciendo las cuentas da un lucro neto asintótico de \$9,6128.

Política 0,3. La matriz es

$$p = \begin{pmatrix} 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \\ 0{,}7 & 0{,}3 & 0 & 0 \\ 0{,}3 & 0{,}4 & 0{,}3 & 0 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}3 \end{pmatrix}$$

La distribución estacionaria es $\pi = (343/1070, 300/1070, 280/1070, 147/1070)$. Haciendo las cuentas, el lucro es \$9,40.

La política 1,3 es óptima.

Ejemplos especiales

Cadenas doblemente estocásticas Decimos que una matriz estocástica es *doblemente estocástica* si la suma de cada una de sus columnas es 1:

$$\sum_x p(x, y) = 1.$$

Teorema 25 *Una matriz p de $N \times N$ es doblemente estocástica si y sólo si la distribución uniforme $\pi(x) = 1/N$ es una distribución estacionaria para p.*

Demostración Si π es uniforme,

$$\sum_x \pi(x)p(x, y) = \frac{1}{N}\sum_x p(x, y) = \frac{1}{N} = \pi(y).$$

es decir, π satisface las ecuaciones de balance. Reciprocamente, si la distribución uniforme es estacionaria para p, vale la segunda igualdad, que implica que p es doblemente estocástica. □

Ejemplo 26 Paseo aleatorio simétrico reflejado en un segmento.

$$p = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

Esa matriz es claramente doblemente estocástica. Por lo tanto su distribución invariante es uniforme en el segmento. La matriz es irreducible y aperiódica.

Ejemplo 27 Juego de tablero circular. Tablero circular con 6 espacios $\{0, 1, 2, 3, 4, 5\}$ periódico (el 0 sigue al 5). Considere un dado que tiene tres faces 1, dos faces 2 y una face 3. $q(1) = 3/6$, $q(2) = 2/6$, $q(3) = 1/6$. Defina

$$p(x, y) = q(z), \qquad \text{si } y = (x + z) \mod 6.$$

Matriz de transición:

$$p = \begin{pmatrix} 0 & 1/2 & 1/3 & 1/6 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 1/6 & 0 \\ 0 & 0 & 0 & 1/2 & 1/3 & 1/6 \\ 1/6 & 0 & 0 & 0 & 1/2 & 1/3 \\ 1/3 & 1/6 & 0 & 0 & 0 & 1/2 \\ 1/2 & 1/3 & 1/6 & 0 & 0 & 0 \end{pmatrix}$$

Se ve que la matriz es doblemente estocástica, así que su distribución invariante es uniforme. $\pi(x) = 1/6$ para todo x. La matriz es irreducible y aperiódica. En particular $p^3(x, y) > 0$ para todo x, y.

Ejemplo 28 *Monopolio para matemáticos.* Tenemos ahora 40 casilleros $\{0, 1, \ldots, 39\}$ con condiciones periódicas y para avanzar lanzamos dos dados. Si $r(k)$ es la probabilidad que la suma de dos dados tenga como resultado k, la matriz de transición es

$$p(x, y) = r(z), \qquad \text{si } y = (x + z) \mod 40.$$

Cada fila de la matriz es igual a la anterior pero rotada una unidad. Por eso es doblemente estocástica. Verifique que $p^4(x, y) > 0$ para todo x, y.

Balance detallado Decimos que π satisface la condición de *balance detallado* si

$$\pi(x)p(x, y) = \pi(y)p(y, x)$$

Si sumamos en x obtenemos que π satisface las ecuaciones de balance:

$$\sum_x \pi(x)p(x, y) = \pi(y) \sum_x p(y, x) = \pi(y).$$

Ejemplo 29 *Una matriz que no satisface balance detallado.* $S = \{1, 2, 3\}$

$$p = \begin{pmatrix} 0{,}5 & 0{,}5 & 0 \\ 0{,}3 & 0{,}1 & 0{,}6 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{pmatrix}$$

Como es irreducible y aperiódica, tiene una única distribución estacionaria π. Pero π no puede satisfacer balance detallado porque $p(1, 3) = 0$ pero $p(3, 1) = 0{,}2 \neq 0$.

Ejemplo 30 *Cadenas de nacimiento y muerte.* El espacio de estados es un intervalo entero $\{\ell, \ell + 1, \ldots, r - 1, r\}$ y es imposible saltar más de una unidad:

$$p(x, y) = 0, \qquad \text{si } |x - y| > 1.$$

Las probabilidades de transición son

$$\begin{aligned} p(x, x + 1) &= p_x, \quad \text{para } x < r \\ p(x, x - 1) &= q_x, \quad \text{para } x > \ell \\ p(x, x) &= 1 - p_x - q_x, \quad \text{para } \ell \leq x \leq r \end{aligned}$$

y $p(x, y) = 0$ en los otros casos.

Buscamos π que satisfaga balance detallado $\pi(x)p_x = \pi(x+1)q_{x+1}$, para $\ell \leq x < r$. π tiene que satisfacer

$$\pi(x + 1) = \frac{p_x}{q_{x+1}} \pi(x)$$

Empezando en $x = \ell$ e iterando obtenemos

$$\pi(\ell + i) = \frac{p_{\ell+i-1} \cdots p_\ell}{q_{\ell+i} \cdots q_{\ell+1}} \pi(\ell)$$

Ejemplo 31 *Ehrenfest con $N = 3$.*

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

Para N genérico tenemos

$$p_x = p(x, x + 1) = (N - x)/N, \qquad q_{x+1} = p(x + 1, x) = (x + 1)/N$$

Resolviendo balance detallado obtenemos

$$\pi(x) = 2^{-N}\frac{N(N-1)\dots N-x+1}{x(x-1)\dots 1} = 2^{-N}\binom{N}{x}$$

Verifiquemos. Balance detallado es equivalente a

$$(N-x)\binom{N}{x} = (x+1)\binom{N}{x+1}$$

Acabamos de demostrar el lema siguiente.

Lema 32 *Toda cadena de nacimiento y muerte con transiciones positivas tiene una medida m que satisface balance detallado. Si $\sum_x m(x) < \infty$, entonces la medida de probabilidad π dada por $\pi(x) := m(x)/\sum_y m(y)$ satisface balance detallado..*

Ejemplo 33 *Paseos aleatorios en grafos.* Un grafo $G = (V, E)$, $V =$ son vértices y $E \subset \{\{u, v\} : u, v \in V\}$ es el conjunto de *aristas*, es un subconjunto de los pares de vertices no ordenados. $A(u, v) = 1$ si $\{u, v\} \in E$, es decir, si hay una arista conectando u y v; en este caso decimos que u y v son *vecinos*. Se trata de grafo no orientado. El *grado* $d(v)$ del vertice v es el número de sus vecinos:

$$d(v) := \sum_v A(u, v).$$

Definimos

$$p(u, v) := \frac{A(u, v)}{d(u)}$$

es una matriz de transición. Para cada constante c, $\pi(u) := cd(u)$ satisface balance detallado:

$$\pi(u)p(u, v) = cA(u, v) = cA(v, u) = \pi(v)p(v, u).$$

Para grafos con un número finito de vértices, si tomamos $c = 1/(\text{suma de los grados})$, obtenemos una distribución estacionaria.

Ejemplo 34 *Paseo del caballo de ajedrez.* Las posibles movidas de un caballo de ajedrez inducen un grafo cuyos vértices son las 64 casillas del tablero de ajedrez y las aristas conectan casillas que se pueden alcanzar entre sí por saltos del caballo. Los vértices tienen grado entre 2 (en los rincones) hasta 8 (en el centro):

2	3	4	4	4	4	3	2
3	4	6	6	6	6	4	3
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
3	4	6	6	6	6	4	3
2	3	4	4	4	4	3	2

La suma de los grados da 336. La distribución estacionaria es $\pi(x) = d(x)/336$.

Reversibilidad

Sea $p(x, y)$ una matriz de transión con una distribución estacionaria $\pi(x)$. Sea $(X_0, X_1, \dots)$ una realización de la cadena de Markov con distribución inicial π, es decir $X_0 \sim \pi$.

Teorema 35 *Fije n. Considere el proceso $Y_m = X_{n-m}$, para $0 \le m \le n$. Entonces Y_m es Markov con matriz de transición*

$$p^*(x, y) = P(Y_{m+1} = y | Y_m = x) = \frac{\pi(y)p(y, x)}{\pi(x)}$$

Además π es una distribución estacionaria para la matriz p^*.

Demostración

$$\begin{aligned}
&P(Y_{m+1} = y_{m+1} | Y_m = y_m, Y_{m-1} = y_{m-1}, \dots, Y_0 = y_0)\\
&= \frac{P(Y_{m+1} = y_{m+1}, Y_m = y_m, Y_{m-1} = y_{m-1}, \dots, Y_0 = y_0)}{P(Y_m = y_m, Y_{m-1} = y_{m-1}, \dots, Y_0 = y_0)}\\
&= \frac{P(X_{n-(m+1)} = y_{m+1}, X_{n-m} = y_m, X_{n-(m-1)} = y_{m-1}, \dots, X_n = y_0)}{P(X_{n-m} = y_m, X_{n-(m-1)} = y_{m-1}, \dots, X_n = y_0)}\\
&= \frac{\pi(y_{m+1})p(y_{m+1}, y_m)P(X_{n-(m-1)} = y_{m-1}, \dots, X_n = y_0 | X_{n-m} = y_m)}{\pi(y_m)P(X_{n-(m-1)} = y_{m-1}, \dots, X_n = y_0 | X_{n-m} = y_m)}\\
&= \frac{\pi(y_{m+1})p(y_{m+1}, y_m)}{\pi(y_m)} = p^*(y_m, y_{m+1}). \qquad \square
\end{aligned}$$

Note que p^* es una matriz de transición:

$$\sum_y p^*(x,y) = \sum_y \frac{\pi(y)}{\pi(x)} p(y,x) = \frac{\pi(x)}{\pi(x)} = 1.$$

porque π es estacionaria para p. Finalmente,

$$\sum_x \pi(x) p^*(x,y) = \sum_x \pi(x) \frac{\pi(y)}{\pi(x)} p(y,x) = \pi(y),$$

lo que demuestra que π es estacionaria para p^*.

Cuando π satisface las ecuaciones de balance detallado:

$$p^*(x,y) = \frac{\pi(y)}{\pi(x)} p(y,x) = p(x,y).$$

La película yendo para atrás o para adelante tiene la misma distribución.

Algoritmo Metrópolis-Hastings

El objetivo es generar muestras de una distribución π dada. Para eso vamos a construir una cadena de Markov cuya distribución estacionaria es π. Empezamos con una matriz de transición arbitraria $q(x,y)$ que será usada para *proponer* un salto. El salto será aceptado con probabilidad

$$r(x,y) = \min\Big\{\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\Big\}$$

Definimos así la matriz de transición

$$p(x,y) = q(x,y) r(x,y).$$

Lema 36 *La distribución π es reversible para p.*

Demostración Supongamos que $\pi(y)q(y,x) > \pi(x)q(x,y)$. En este caso,

$$\begin{aligned}
\pi(x)p(x,y) &= \pi(x)q(x,y)1\\
\pi(y)p(y,x) &= \pi(y)q(y,x)\frac{\pi(x)q(x,y)}{\pi(y)q(y,x)} = \pi(x)q(x,y).
\end{aligned}$$

Es decir que se satisfacen las ecuaciones de balance detallado. $\square$

Usando los teoremas de convergencia se pueden obtener muestras aproximadas de π, o las esperanzas en relación a π de funciones objetivo. Por ejemplo:

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to_n \sum_x f(x)\pi(x).$$

Distribución geométrica Supongamos que queremos generar muestras de la distribución geométrica $\pi(x) = \theta^x(1-\theta)$, $x = 0, 1, 2, \ldots$, con $\theta < 1$. Elegimos q como el paseo aleatorio simétrico $q(0,0) = q(x, x+1) = q(x+1, x) = 1/2$, $x \geq 0$. Como q es simétrica, $r(x,y) = \text{mín}\{\pi(y)/\pi(x), 1\}$. Como $\pi(x) > \pi(x+1)$, tenemos para $x \geq 1$:

$$p(x, x-1) = \frac{1}{2} \quad p(x, x+1) = \frac{\theta}{2} \quad p(x,x) = \frac{1-\theta}{2}.$$

Cuando $x = 0$, $\pi(-1) = 0$, así que

$$p(0,-1) = 0 \quad p(0,1) = \frac{\theta}{2} \quad p(0,0) = 1 - \frac{\theta}{2}.$$

Balance detallado:

$$\pi(x)p(x, x+1) = \theta^x(1-\theta)\frac{\theta}{2} = \pi(x+1)p(x+1, x).$$

Binomial Consideremos ahora Binomial(N, θ). Es decir $\pi(x) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$, $x \in \{0, \ldots, N\}$.

$$q(x,y) = \frac{1}{N+1}, \quad y \in \{0, \ldots, N\}, \quad \text{uniforme}$$

Como q es simétrica r usa el cociente de las probabilidades:

$$r(x,y) = \text{mín}\Big\{1, \frac{\pi(y)}{\pi(x)}\Big\}$$

Modelo de Ising $\Lambda = \{-L, \ldots, L\}^2$. Cada punto $x \in \Lambda$ tiene un spin en $\{-1, 1\}$. Un estado del sistema es una función $\xi : \Lambda \to \{-1, 1\}$. $\xi \in \{-1, 1\}^\Lambda$, que llamamos *configuración*. Dos puntos $x, y \in \Lambda$ son *vecinos* si $\|x - y\| = 1$.

Parámetro β temperatura inversa.

$$\pi(\xi) := \frac{1}{Z}\exp\Big(\beta \sum_{\{x,y\}\subset\Lambda:\|x-y\|=1} \xi(x)\xi(y)\Big)$$

donde Z es la normalización para que π sea una probabilidad. Las condiciones de frontera pueden ser fijadas de diferentes maneras. En nuestra definición ignoramos lo que pasa fuera de Λ.

Configuraciones con máxima probabilidad son las que tienen todos los spins iguales. "Todo uno" y "todo menos uno".

Z es difícil de calcular. Por eso es difícil obtener muestras de π. Por suerte Metropolis-Hastings depende de los cocientes, que no dependen de Z.

Defina ξ^x por

$$\xi^x(x) = -\xi(x) \text{ y } \xi^x(z) = \xi(z) \text{ para todo } z \neq x$$

Proponemos

$$q(\xi, \xi') = \begin{cases} \frac{1}{|\Lambda|}, & \text{si } \xi' = \xi^x, \text{ para algún } x \in \Lambda \\ 0, & \text{en los otros casos.} \end{cases}$$

Cambiar un spin elegido al azar en Λ. La probabilidad de transición queda

$$p(\xi, \xi^x) = q(\xi, \xi^x)\,\text{mín}\Big\{1, \frac{\pi(\xi^x)}{\pi(\xi)}\Big\}$$

Como $\xi^x(x) = -\xi(x)$,

$$\frac{\pi(\xi^x)}{\pi(\xi)} = \exp\Big(-2\beta \sum_{y\in\Lambda:\|x-y\|=1} \xi(x)\xi(y)\Big)$$

Baño caliente o heat bath. Esta cadena se usa para espacios de estados vectoriales. Como en el caso del modelo de Ising, el espacio de estados es A^Λ, donde A es un alfabeto finito; en Ising $A = \{-1, 1\}$. Hay una distribución π sobre S que queremos aproximar. Proponemos la siguiente matriz de transición. Para cada $x \in \Lambda$, definimos

$$p(\xi, \xi') = \frac{1}{|\Lambda|} \frac{\pi(\xi')}{\sum_{\eta:\eta(y)\in\mathbb{N}_x(\xi)} \pi(\eta)}, \qquad \xi' \in N_x(\xi) := \{\eta : \eta(y) = \xi(y), \forall y \neq x\}$$

En palabras: elegimos una coordenada x uniformemente en Λ, borramos el valor de esa coordenada y lo reemplazamos por un nuevo valor elegido con la distribución π, condicionada a los valores de las restantes coordenadas.

Como los denominadores de $p(\xi, \xi')$ coinciden para $\xi' \in N_x(\xi)$, para verificar balance detallado

$$\pi(\xi)p(\xi, \xi') = \pi(\xi')p(\xi', \xi)$$

basta ver que los numeradores lo hacen: $\pi(\xi)\pi(\xi') = \pi(\xi')\pi(\xi)$.

Modelo de Potts es una extensión del modelo de Ising a spins con un número finito de valores. $S = A^\Lambda$, A un alfabeto finito. La Hamiltoniana es

$$H(\xi) = - \sum_{\{x,y\}\subset\Lambda:\|x-y\|=1} \mathbf{1}\{\xi(x) = \xi(y)\}$$

y la distribución π está dada por

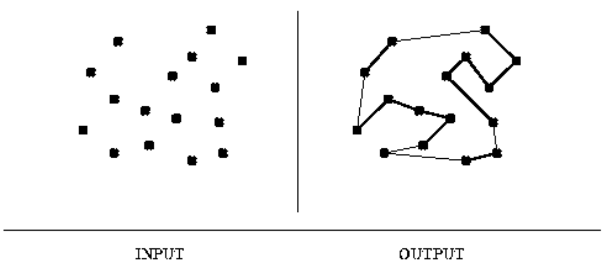
$$\pi(\xi) = \frac{1}{Z} \exp\big(-\beta H(\xi)\big)$$

Y así, para $\xi' \in N_x(\xi)$, tenemos

$$q(\xi, \xi') = \frac{\exp\Big(\beta \sum_{y:\|y-x\|=1} \mathbf{1}\{\xi'(x) = \xi(y)\}\Big)}{\sum_{a\in A} \exp\Big(\beta \sum_{y:\|y-x\|=1} \mathbf{1}\{a = \xi(y)\}\Big)}$$

Recocido Simulado o Simulated annealing. El algoritmo de Metropolis-Hastings puede ser usado para minimizar funciones complicadas.

El problema del agente viajero. Un agente que vive en la ciudad 0 tiene que visitar todas las ciudades $\{1, \ldots, N\}$ y volver a 0. El costo de viajar de la ciudad i a la ciudad j es $c(i, j)$. Se trata de encontrar un recorrido que minimice el costo total. El espacio de estados es el conjunto de permutaciones de $(1, \ldots, N)$ y a cada permutación $x = (i_1, i_2, \ldots, i_N)$ se le asigna el costo $H(x) = \sum_{k=0}^{N} c(i_k, i_{k+1})$, con la convención $i_{N+1} = i_0 = 0$. Se trata de encontrar un recorrido x que minimice H.



Para eso se empieza con una cadena de Markov q con espacio de estados $\{$permutaciones de $(1, \ldots, N)\}$. Por ejemplo, elegir dos ciudades sucesivas en la permutación y permutarlas.

Para cada $\beta > 0$ consideramos la distribución

$$\pi_\beta(x) := \frac{1}{Z} \exp\big(-\beta H(x)\big).$$

Cuanto menor $H(x)$, mayor la probabilidad que le asigna π_β a x. Además, para $\beta = \infty$ esa distribución le dá peso sólo a los mínimos absolutos de H.

Sea p_β la matriz de Metrópolis-Hastings asociada a π_β y a q:

$$p_\beta(x, y) := q(x, y) \min\Big\{1, \frac{\pi_\beta(y)}{\pi_\beta(x)}\Big\}$$

Régimen de enfriamiento: $\beta(n)$: se cambia el valor de β a medida que avanza el tiempo.

$$\beta(n) \to \infty, \quad \text{cuando } n \text{ va a infinito}$$

Despues consideramos una cadena de Markov no homogenea

$$P(X_{n+1} = y | X_n = x) = p_{\beta(n)}(x, y).$$

Eligiendo convenientemente la velocidad con la que β va a infinito, X_n converge a un mínimo de H.

Demostración de los teoremas de convergencia

Lema 37 *Si p admite una distribución estacionaria π, entonces los estados y con $\pi(y) > 0$ son recurrentes.*

Demostración Sabemos que

$$E_x N(y) = \sum_{n=1}^{\infty} p^n(x, y)$$

Entonces

$$\sum_x \pi(x) E_x N(y) = \sum_x \pi(x) \sum_{n=1}^{\infty} p^n(x, y)$$

Cambiando el orden de sumas y usando $\pi p^n = \pi$, tenemos

$$= \sum_{n=1}^{\infty} \sum_x \pi(x) p^n(x, y) = \sum_{n=1}^{\infty} \sum_x \pi(y) = \infty.$$

porque $\pi(y) > 0$. Por otro lado

$$E_x N(y) = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

Así

$$\infty = \sum_x \pi(x) \frac{\rho_{xy}}{1 - \rho_{yy}} \le \frac{1}{1 - \rho_{yy}}.$$

Esto demuestra que $\rho_{yy} = 1$, es decir y recurrente. $\square$

Teorema 38 *Sea p una matriz irreducible, aperiódica, con una distribución invariante π. Entonces*

$$\lim_n p^n(x, y) = \pi(y)$$

Demostración Esta demostración fué propuesta inicialmente por Doeblin. Considere una cadena de Markov en $S \times S$ con matriz $\bar{p}$ dada por

$$\bar{p}((x_1, y_2), (x_2, y_2)) = p(x_1, x_2) p(y_1, y_2)$$

Las marginales son cadenas independientes con matriz p. Llamemos (X_n, Y_n) la cadena con matriz $\bar{p}$.

1 Si p es aperiódica e irreducible, entonces $\bar{p}$ es irreducible.

Veamos. Como p es irreducible, existen K y L tales que $p^K(x_1,x_2) > 0$ y $p^L(y_1,y_2) > 0$. Como x_2 e y_2 tienen período 1, para M suficientemente grande $p^{L+M}(x_2,x_2) > 0$ y $p^{K+M}(y_2,y_2) > 0$. Entonces

$$\begin{aligned}
&\bar{p}^{K+L+M}((x_1,y_1),(x_2,y_2)) \\
&= p^{K+L+M}(x_1,x_2)\, p^{K+L+M}(y_1,y_2) \\
&\geq p^K(x_1,x_2)p^{L+M}(x_2,x_2)\, p^L(y_1,y_2)p^{K+M}(y_2,y_2) > 0.
\end{aligned}$$

2 Sea

$$T := \text{mín}\{n \geq 0 : X_n = Y_n\}, \quad \text{instante de encuentro de las marginales}$$

Entonces $P(T < \infty) = 1$.

Observe que T es un tiempo de parada para (X_n, Y_n):

$$\{T = n\} = \{X_1 \neq Y_1, \ldots, X_{n-1} \neq Y_{n-1}, X_n = Y_n\}.$$

Veamos. Como la evolución de las marginales es independiente, la distribución $\bar{\pi}(a,b) = \pi(a)\pi(b)$ es estacionaria para $\bar{p}$.

Por el lema anterior, todos los estados son recurrentes para $\bar{p}$. En particular el estado (x,x) es recurrente y

$$T_{(x,x)} := \text{mín}\{n \geq 0 : (X_n, Y_n) = (x,x)\}$$

satisface $P(T_{(x,x)} < \infty) = 1$. Como $T \leq T_{(x,x)}$, tenemos

$$P(T < \infty) = 1.$$

3

$$P(X_n = y, T \leq n) = P(Y_n = y, T \leq n) \tag{27}$$

$$\begin{aligned}
P(X_n = y, T \leq n) &= \sum_{m=1}^{n} \sum_x P(X_n = y, X_m = x, T = m) \\
&= \sum_{m=1}^{n} \sum_x P(X_n = y | X_m = x, T = m) P(X_m = x, T = m) \\
&= \sum_{m=1}^{n} \sum_x P(X_n = y | X_m = x) P(X_m = x, T = m) \qquad \text{Markov fuerte} \\
&= \sum_{m=1}^{n} \sum_x P(Y_n = y | Y_m = x) P(Y_m = x, T = m) \\
&= P(Y_n = y, T \leq n)
\end{aligned}$$

4

$$\sum_y |p^n(x,y) - \pi(y)| \leq 2P(T > n) \to 0. \tag{28}$$

Para verlo escribimos

$$P(X_n = y) = P(X_n = y, T \leq n) + P(X_n = y, T > n)$$

y lo mismo para Y_n y usando (27) tenemos

$$\begin{aligned}
|P(X_n = y) - P(Y_n = y)| &= |P(X_n = y, T > n) - P(Y_n = y, T > n)| \\
&\leq P(X_n = y, T > n) + P(Y_n = y, T > n)
\end{aligned}$$

Sumando en y:

$$\sum_y |P(X_n = y) - P(Y_n = y)| \leq 2P(T > n) \to 0.$$

Para probar (28) escribimos

$$\begin{aligned}
\sum_y |P_x(X_n = y) - \pi(y)| &= \sum_y |P_x(X_n = y) - \sum_z \pi(z) P_z(Y_n = y)| \\
&\leq \sum_y \sum_z \pi(z) |P_x(X_n = y) - P_z(Y_n = y)| \\
&\leq \sum_z \pi(z) 2P_{(x,z)}(T > n).
\end{aligned}$$

que se va a cero por convergencia dominada. □

Existencia de distribuciones estacionarias Sea

$$T_x := \text{ínf}\{n \geq 1 : X_n = x\},$$

un tiempo de parada porque

$$\{T_x = n\} = \{X_1 \neq x, \ldots, X_{n-1} \neq x, X_n = x\}.$$

Teorema 39 *Supongamos que p es irreducible y recurrente. Sea $x \in S$ y defina*

$$\mu_x(y) := \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

Entonces μ_x es una medida estacionaria con $0 < \mu_x(y) < \infty$ para todo y.

Esto es llamado el "truco del ciclo". $\mu_x(y)$ es el número esperado de visitas a y en el intervalo temporal $\{0, 1, \ldots, T_x - 1\}$. $\mu_x p$ es como calcular ese mismo número pero empezando del instante 1. Es decir, el número esperado de visitas a y en $\{1, \ldots, T_x\}$. Como $X_{T_x} = X_0 = x$, en los dos casos hay exactamente una visita a x y el mismo número de visitas a y. Por eso $\mu_x = \mu_x p$.

Demostración Sea

$$\bar{p}_n(x, y) = P_x(X_n = y, T_x > n)$$

de donde

$$\mu_x(y) = \sum_{n \geq 0} \bar{p}_n(x, y).$$

Intercambiando sumas,

$$\sum_y \mu_x(y) p(y, z) := \sum_{n=0}^{\infty} \sum_y \bar{p}_n(x, y) p(y, z)$$

Caso 1: $z \neq x$.

$$\begin{aligned}
\sum_y \bar{p}_n(x, y) p(y, z) &= \sum_y P_x(X_n = y, T_x > n, X_{n+1} = z) \\
&= P_x(T_x > n + 1, X_{n+1} = z) = \bar{p}_{n+1}(x, z).
\end{aligned}$$

Sumando en n tenemos

$$\sum_{n=0}^{\infty} \sum_y \bar{p}_n(x, y) p(y, z) = \sum_{n=0}^{\infty} \bar{p}_{n+1}(x, z) = \mu_x(z)$$

porque $\bar{p}_0(x, z) = 0$.

Caso 2: $z = x$.

$$\sum_y \bar{p}_n(x,y)p(y,x) = \sum_y P_x(X_n = y, T_x > n, X_{n+1} = x) = P_x(T_x = n+1).$$

Sumando en n tenemos

$$\sum_{n=0}^{\infty}\sum_y \bar{p}_n(x,y)p(y,x) = \sum_{n=0}^{\infty} P_x(T_x = n+1) = 1 = \mu_x(x)$$

porque $P_x(T_x = 0) = 0$.

Veamos que $\mu_x(y) < \infty$.

$$1 = \mu_x(x) = \sum_z \mu_x(z)p^n(z,x) \geq \mu_x(y)p^n(y,x).$$

Como la cadena es irreducible, existe un n con $p^n(y,x) > 0$ de donde podemos concluir $\mu_x(y) < \infty$.

Veamos que $\mu_x(y) > 0$. Es inmediato si $y = x$. Para $y \neq x$ por irreducibilidad, existe un camino $x = y_0, \dots, y_{K-1}, y_K = y$ todos distintos, tal que

$$p(x,y_1)\dots p(y_{K-1},y) > 0$$

Por lo tanto

$$P_x(X_K = y, T_x > K) > 0$$

y por lo tanto $\mu_x(y) > 0$. □

Teorema 40 *Suponga p irreducible y recurrente. Sea $N_n(y)$ el número de visitas a y en tiempos menores o iguales a n. Entonces*

$$\lim_n \frac{N_n(y)}{n} = \frac{1}{E_y(T_y)}$$

Demostración Supongamos que empezamos en y. Por la propiedad fuerte de Markov, los tiempos entre retornos, $t_1, t_2, \dots$ son independientes e identicamente distribuídos. Definiendo

$$R(k) := \min\{n \geq 1 : N_n(y) = k\},$$

el instante de la k-ésima visita a y, tenemos

$$\lim_k \frac{R(k)}{k} = E_y(T_y) \leq \infty.$$

por una versión de la ley fuerte de grandes números para variables no negativas que no requiere segundo momento finito. Ver Durrett [3], Teorema 2.4.1 cuando el primer momento es finito y el Teorema 2.4.5 cuando el primer momento es infinito.

Si en lugar de empezar en y, empezamos en otro estado, $t_1 < \infty$ y $t_2, t_3, \dots$ son iid y la ley de grandes números sigue valiendo.

De la definición de $R(k)$ tenemos

$$\frac{R(N_n(y))}{N_n(y)} \leq \frac{n}{N_n(y))} < \frac{R(N_n(y)+1)}{N_n(y)+1}\frac{N_n(y)+1}{N_n(y)}$$

Como $N_n(y)$ se va a infinito con n, ambos extremos de la desigualdad convergen a E_yT_y, de donde concluimos que

$$\lim_n \frac{n}{N_n(y)} = E_yT_y. \qquad \square$$

Teorema 41 *Si p es irreducible y tiene una distribución estacionaria π, entonces*

$$\pi(y) = \frac{1}{E_y T_y}.$$

Demostración Si X_0 tiene distribución π, por el teorema anterior,

$$\lim_n \frac{N_n(y)}{n} = \frac{1}{E_y T_y}.$$

Por otro lado, como $N_n(y) \leq n$, se puede demostrar (convergencia dominada) que

$$\lim_n \frac{E_\pi N_n(y)}{n} = \frac{1}{E_y T_y}. \tag{29}$$

Pero

$$E_\pi N_n(y) = \sum_x \pi(x) \sum_{k=1}^n p^k(x,y) = \sum_{k=1}^n \sum_x \pi(x) p^k(x,y) = n\pi(y).$$

o sea, que el límite en (29) es igual a $\pi(y)$, con lo que concluímos. □

Teorema 42 *Si p es irreducible, tiene una distribución estacionaria π y $\sum_x |f(x)|\pi(x) < \infty$, entonces*

$$\frac{1}{n} \sum_{m=1}^n f(X_m) \to \sum_x f(x)\pi(x).$$

Demostración Fijemos $X_0 = x$ y T_k el instante de la $k$ésima visita a x, $T_0 = 0$.

Por Markov fuerte, las variables

$$Y_k = \sum_{T_{k-1}+1}^{T_k} f(X_m)$$

son iid. Por el truco del ciclo,

$$EY_k = \sum_y \mu_x(y) f(y)$$

Por la ley de grandes números para iid:

$$\frac{1}{L} \sum_{m=1}^{T_L} f(X_m) = \frac{1}{L} \sum_{k=1}^{L} Y_k \to_L \sum_y \mu_x(y) f(y).$$

Tomando $L = N_n(x) = \max\{k : T_k \leq n\}$ e ignorando la contribución del último ciclo $(N_n(x), n]$

$$\frac{1}{n} \sum_{m=1}^n f(X_m) \approx \frac{N_n(x)}{n} \frac{1}{N_n(x)} \sum_{k=1}^{N_n(x)} Y_k$$

Usando el teorema anterior y la ley de grandes números,

$$\to \frac{1}{E_x T_x} \sum_y \mu_x(y) f(y) = \sum_y \pi(y) f(y). \qquad \square$$

Distribuciones de salida

Colegio de 2 años 60 % de los alumnos del año 1 pasan al año 2, 20 % repiten y 15 % abandonan. De los alumnos del año 2, 70 % se reciben, 20 % repiten y 10 % abandonan.

Qué fracción de los ingresantes se reciben en algún momento?

Espacio de estados $S = \{1, 2, G, A\}$ $G =$ se reciben. $A =$ abandonan.

$$p = \begin{pmatrix} 0{,}25 & 0{,}6 & 0 & 0{,}15 \\ 0 & 0{,}2 & 0{,}7 & 0{,}1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Sea $h(x)$ la probabilidad que un estudiante que está en el estado x se reciba. Tenemos $h(G) = 1$ y $h(A) = 0$. Para calcular los otros dos valores condicionamos al primer salto:

$$\begin{aligned} h(1) &= 0{,}25\, h(1) + 0{,}6\, h(2) \\ h(2) &= 0{,}2\, h(2) + 0{,}7 \end{aligned}$$

Solución

$$h(2) = 7/8, \quad h(1) = 0{,}7$$

Tenis Si un juego de tenis está en 3-3, 4-3 o 3-4, para que uno de los jugadores gane, tiene que sacarle al otro una diferencia de dos puntos.

El servidor (quien saca) tiene probabilidad 0,6 de ganar el punto. Cual es la probabilidad de que el servidor gane si el juego (a) está empatado, (b) si está un punto atrás, (c) si está un punto adelante? El espacio de estados $S = \{-2, -1, 0, 1, 2\}$.

$$p = \left(\begin{array}{c|ccc|c} 1 & 0 & 0 & 0 & 0 \\ \hline 0{,}6 & 0 & 0{,}4 & 0 & 0 \\ 0 & 0{,}6 & 0 & 0{,}4 & 0 \\ 0 & 0 & 0{,}6 & 0 & 0{,}4 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

-2 y 2 son absorventes.

Si $h(x)$ es la probabilidad de que el servidor gane cuando el resultado actual es x, condicionando al primer paso tenemos

$$h(x) = \sum_y p(x, y)h(y), \qquad h(2) = 1, \quad h(-2) = 0.$$

Esto nos pone tres ecuaciones con tres incógnitas:

$$\begin{aligned} h(1) &= 0{,}6 + 0{,}4h(0) \\ h(0) &= 0{,}6h(1) + 0{,}4h(-1) \\ h(-1) &= 0{,}6h(0) \end{aligned}$$

Resolviendo el sistema, obtenemos $h(0) = 0{,}6923$, lo que resuelve la pregunta (a). Para el caso (b) $x = 1$ y para el caso (c) $x = -1$.

Teorema 43 *Sea S finito. Sean $a, b \in S$ y $C = S \setminus \{a, b\}$. Suponga que $h(a) = 1$, $h(b) = 0$ y*

$$h(x) = \sum_{y \in S} p(x, y)h(y), \qquad x \in C.$$

Sea $V_z = \min\{n : X_n = z\}$ y $T = \min\{V_a, V_b\}$. Si $P_x(T < \infty) > 0$ para todo $x \in C$, entonces $h(x) = P_x(V_a < V_b)$.

Demostración Vimos que $P_x(T < \infty) = 1$ para todo x. Además, como h es armónica, $h(x) =$

$\sum_y p(x,y)h(y)$. Por la propiedad de Markov

$$\begin{aligned}
h(x) &= E_x h(X_{T\wedge n}) = E_x\big(h(X_n)\mathbf{1}\{T>n\}\big) \\
&\quad + \sum_{m=1}^{n} E_x(h(X_m)[\mathbf{1}\{T=m, X_m=a\} + \mathbf{1}\{T=m, X_m=b\}]) \\
&= E_x\big(h(X_n)\mathbf{1}\{T>n\}\big) + \sum_{m=1}^{n} E_x(h(X_m)\mathbf{1}\{V_a=m, V_a<V_b\}\big) \\
&\to_n \sum_{n=1}^{\infty} E_x \mathbf{1}\{V_a=m, V_a<V_b\} = P(V_a<V_b),
\end{aligned}$$

En la segunda igualdad usamos $h(a)=1$ y $h(b)=0$ y $\{T=n, X_m=a\}=\{m=V_a<V_b\}$. $\square$

Ruina del jugador, caso simétrico Dos jugadores tiran dos monedas, si son iguales A se lleva las monedas, si son diferentes, se las lleva B. La fortuna de A aumenta 1 con probabilidad 1/2 y disminuye 1 con la misma probabilidad. Supongamos que inicialmente A tiene 15 monedas y B tiene 10. El juego termina cuando uno de los dos se queda con todas las monedas.

Cual es la probabilidad que A gane el juego? La respuesta: 15/25.

Sea X_n = número de monedas de A después de jugar n veces. Como el juego es equitativo, $E_x X_n = x$. Sea $V_y = \min\{n : X_n = y\}$. Entonces

$$x = NP_x(V_N < V_0) + 0P_x(V_N > V_0).$$

que resolviendo da $P_x(V_N < V_0) = x/N$ para $0 \le x \le N$.

Para demostrarlo, condicionando al primer paso tenemos

$$h(x) = \frac{1}{2}h(x+1) + \frac{1}{2}h(x-1).$$

es decir,

$$h(x+1) - h(x) = h(x) - h(x-1)$$

inclinación constante. Como $h(0)=0$ y $h(N)=N$, tenemos $h(x)=x/N$.

Wright-Fisher sin mutación X_n es el número de genes tipo A.

$$p(x,y) = \binom{N}{y}\Big(\frac{x}{N}\Big)^y\Big(\frac{N-x}{N}\Big)^{N-y}$$

Si definimos $h(x)=x/N$, como el número medio de éxitos es x/N, tenemos

$$h(x) = \sum_y p(x,y)h(y)$$

Tomando $a=N$ y $b=0$, tenemos $h(a)=1$ y $h(b)=0$. Como $P_x(V_a \wedge V_b < \infty) > 0$, obtenemos

$$P_x(V_N < V_0) = x/N$$

es decir que la probabilidad que todos los genes queden en A es igual a la fracción inicial de genes en A.

Ruina del jugador, caso asimétrico En cada jugada gano 1 con probabilidad p y pierdo con probabilidad $q=1-p$. $p \neq 1/2$. Paro de jugar si la fortuna llega a N. La probabilidad de salir ganando del casino es

$$h(x) = P_x(V_N < V_0), \quad h(0)=0,\ h(N)=1.$$

Condicionando al primer paso:

$$h(x) = ph(x+1) + qh(x-1), x \in \{1, \dots, N-1\}.$$

Obtenemos

$$h(x+1) - h(x) = \frac{q}{p}(h(x) - h(x-1))$$

Llamando $c = h(1) - h(0)$,

$$h(x) - h(x-1) = \Big(\frac{q}{p}\Big)^{x-1}$$

Sumando en x

$$1 = h(N) - h(0) = \sum_{x=1}^{N}(h(x) - h(x-1)) = c\sum_{x=1}^{N}\Big(\frac{q}{p}\Big)^{x-1}$$

Llamando $\theta = q/p$, concluímos

$$h(x) = \frac{\theta^x - 1}{\theta^N - 1}.$$

Ruleta con 18 rojos, 18 negros y 2 verdes (0 y 00). Si apostamos \$1 a rojo, ganamos con probabilidad $p = 18/38$ y perdemos con probabilidad $q = 20/38$. Si llevamos \$50 con la esperanza de retirarnos con \$100. Cual es la probabilidad de salir ganando?

Aquí $\theta = q/p = 20/18$.

$$P_{50}(V_{100} < V_0) = \frac{\theta^{50} - 1}{\theta^{100} - 1} = \frac{(20/18)^{50} - 1}{(20/18)^{100} - 1} = 0{,}005128$$

Desde el punto de vista del casino: Para el casino $p = 20/38$. Suponga que el casino empieza con un capital de $x = 100$. La probabilidad que el capital del casino llegue a N antes de la bancarrota es

$$\frac{(18/20)^{100} - 1}{(18/20)^N - 1} \to_N (1 - (18/20)^{100}) \approx 1 - 2{,}656 \times 10^{-5}$$

Si el capital inicial es \$200, la probabilidad de quiebra se eleva al cuadrado (tiene que perder 100 y después otros 100).

Vemos que si $p > 1/2$ tenemos que $q/p < 1$ y

$$P_x(V_0 = \infty) = 1 - \Big(\frac{q}{p}\Big)^x \quad \text{y} \quad P_x(V_0 < \infty) = \Big(\frac{q}{p}\Big)^x \tag{30}$$

Tiempos de salida

Colegio de 2 años Espacio de estados $S = \{1, 2, G, A\}$, $G =$ se reciben, $A =$ abandonan.

$$p = \begin{pmatrix} 0{,}25 & 0{,}6 & 0 & 0{,}15 \\ 0 & 0{,}2 & 0{,}7 & 0{,}1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Cuanto tiempo pasa en promedio un estudiante en el colegio?

$g(x) =$ tiempo medio que le queda a un estudiante que está en el estado x. $g(G) = g(A) = 0$. Condicionando al primer paso:

$$g(1) = 1 + 0{,}25g(1) + 0{,}6g(2)$$
$$g(2) = 1 + 0{,}2g(2)$$

Se le suma 1 porque gasta un año para pasar de un estado a otro.

La solución es

$$g(1) = 2{,}3333$$

Tenis $X_n =$ diferencia de puntos entre el servidor y el contrincante. El espacio de estados es $\{2,1,0,-1,-2\}$ y la matriz es

$$p = \left(\begin{array}{c|ccc|c} 1 & 0 & 0 & 0 & 0 \\ \hline 0{,}6 & 0 & 0{,}4 & 0 & 0 \\ 0 & 0{,}6 & 0 & 0{,}4 & 0 \\ 0 & 0 & 0{,}6 & 0 & 0{,}4 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

Sea $g(x) =$ tiempo esperado para terminar el juego dado que empezamos con diferencia de puntos x. Tenemos

$$g(-2) = g(2) = 0.$$

Condicionando al primer paso y observando que en cada paso el tiempo se incrementa una unidad, tenemos

$$g(x) = 1 + \sum_{y\in S} p(x,y)g(y)$$

Si llamamos $r(x,y)$ la restricción de la matriz a los estados $\{1,0,-1\}$, tenemos

$$g(x) - \sum_{y\in S} r(x,y)g(y) = 1, \qquad x \in \{1,0,-1\}.$$

Si denotamos $\mathbf{1}$ el vector columna de dimensión 3, con 1 en todas las coordenadas, tenemos

$$(I-r)g = \mathbf{1} \qquad \text{restringiendo } g \text{ a } x \in \{1,0,-1\}.$$

obtenemos la solución

$$g = (1-r)^{-1}\mathbf{1}$$

Otra manera de ver lo mismo. Recordemos que N_y es el número de visitas a y para tiempos $n \geq 0$. Entonces,

$$E_x N(y) = \sum_{n=0}^{\infty} r^n(x,y) = (I-r)^{-1}(x,y)$$

Si llamamos T a la duración del juego, tenemos $T = \sum_y N(y)$ y

$$E_x T = \big((I-r)^{-1}\mathbf{1}\big)(x)$$

Volviendo al tenis,

$$I - r = \begin{pmatrix} 1 & -0{,}4 & 0 \\ -0{,}6 & 1 & -0{,}4 \\ 0 & -0{,}6 & 1 \end{pmatrix}; \qquad (I-r)^{-1} = \frac{1}{13}\begin{pmatrix} 19 & 10 & 4 \\ 15 & 25 & 10 \\ 9 & 15 & 19 \end{pmatrix}$$

De donde concluímos que $E_0 T = (15+25+10)/13 = 3{,}48$, la suma del número esperado de visitas al 1, 0 y -1, respectivamente, empezando en 0.

Teorema 44 *Consideremos una cadena de Markov con espacio de estados finito $S = A \cup C$ con $A, C \neq \emptyset$. Sea $T := \inf\{n \geq 0 : X_n \in A\}$. Supongamos $P_x(T < \infty) > 0$ para todo $x \in C$ y que g es una función que satisface*

$$g(a) = 0, \qquad a \in A$$
$$g(x) = 1 + \sum_{y\in S} p(x,y)g(y), \qquad x \in C$$

Entonces,

$$g(x) = E_x T = ((I-r)^{-1}\mathbf{1})(x), \tag{31}$$

donde $r(x,y) = p(x,y)\mathbf{1}\{x,y \in C\}$ es la matriz p restringida a C y $\mathbf{1}$ es el vector columna indexado por C con 1 en todas las coordenadas.

Demostración Vimos que $E_x T < \infty$ para todo $x \in C$. Como $g(a) = 0$ para $a \in A$, la ecuación para $g(x)$, $x \in C$, se puede escribir

$$\begin{aligned}
g(x) &= 1 + \sum_{y \in C} r(x,y) g(y) \\
&= 1 + \sum_{y \in C} r(x,y) \Big(1 + \sum_{z \in C} r(y,z) g(z) \Big) \\
&= \sum_{k=0}^{n-1} \sum_{z \in C} r^k(x,z) + \sum_{z \in C} r^n(x,z) g(z), \qquad n \geq 1, && (32) \\
&= \sum_{k=0}^{n-1} P_x(T > k) + E_x(g(X_n) \mathbf{1}\{T > n\}) && (33) \\
(&= E_x(T \wedge n) + E_x g(X_{T \wedge n}))
\end{aligned}$$

Como

$$\sum_{k=0}^{n-1} P_x(T > k) \underset{n}{\longrightarrow} \sum_{k=0}^{\infty} P_x(T > k) = E_x T$$

y como S es finito $\sup_x g(x) < \infty$, $E_x(g(X_n)\mathbf{1}\{T > n\}) \leq \|g\|_\infty P_x(T > n) \to_n 0$, que implica (33) converge a $E_x T$ con n. Sacando el límite en (32) y (33) concluimos (31). $\square$

Tiempo de espera hasta dos caras Sea T el tiempo de espera hasta ver dos caras seguidas al lanzar una moneda. Introducimos una cadena de Markov con espacio de estados $\{0, 1, 2\}$ que representa el número de caras en las dos últimas jugadas. Paramos cuando $X_n = 2$. La matriz es

$$p = \begin{pmatrix} 0{,}5 & 0{,}5 & 0 \\ 0{,}5 & 0 & 0{,}5 \\ 0 & 0 & 1 \end{pmatrix}$$

La matriz restringida es

$$r = \begin{pmatrix} 0{,}5 & 0{,}5 \\ 0{,}5 & 0 \end{pmatrix}$$

Aplicando el teorema

$$I - r = \begin{pmatrix} 0{,}5 & -0{,}5 \\ -0{,}5 & 1 \end{pmatrix} \qquad (I - r)^{-1} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

y $E_0 T = 4 + 2 = 6$.

Tiempo de espera hasta ceca-cara Sea T el tiempo de espera hasta ver ceca-cara (01) al lanzar una moneda. Introducimos una cadena de Markov con espacio de estados $\{00, 01, 10, 11\}$ que representa el resultado de las dos últimas jugadas. Paramos cuando $X_n = 01$. La matriz es

$$p = \begin{pmatrix} 0{,}5 & 0{,}5 & 0 & 0 \\ 0 & 0 & 0{,}5 & 0{,}5 \\ 0{,}5 & 0{,}5 & 0 & 0 \\ 0 & 0 & 0{,}5 & 0{,}5 \end{pmatrix}$$

Eliminando la columna y fila de 01, la matriz restringida queda

$$r = \begin{pmatrix} 0{,}5 & 0 & 0 \\ 0{,}5 & 0 & 0 \\ 0 & 0{,}5 & 0{,}5 \end{pmatrix}$$

Aplicando el teorema

$$I - r = \begin{pmatrix} 0{,}5 & 0 & 0 \\ -0{,}5 & 1 & 0 \\ 0 & -0{,}5 & 0{,}5 \end{pmatrix} \qquad (I - r)^{-1}\mathbf{1} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$$

Esto nos dice que $E_{00}T = 2$, $E_{10}T = 2$, $E_{11}T = 4$. Como las dos primeras jugadas tienen probabilidad 1/4 de dar cualquiera de las 4 posibilidades y $E_{01}T = 0$, tenemos

$$ET = 2 + \frac{1}{4}(0 + 2 + 2 + 4) = 4.$$

Duración de juegos honestos Consideremos la ruina del jugador con $p(x, x+1) = p(x+1, x) = \frac{1}{2}$. Sea $T := \min\{n : X_n \notin (0, N)$. Entonces

$$E_x T = x(N - x). \tag{34}$$

En el caso de las monedas, $N = 25$ y $x = 15$, el juego va a durar $15\,10 = 150$ en promedio. Si $N = 50$ y $x = 30$ el juego va a durar $30\,20 = 600$.

Hay dos maneras de probar esto. Una es verificando: $g(x) = E_x T$ satisface $g(0) = g(N) = 0$ y

$$g(x) = 1 + \frac{1}{2}g(x+1) + \frac{1}{2}g(x-1)$$

Si substituimos por $g(x) = x(N - x)$ vemos que ok.

Deduciendo a respuesta. Partiendo de la ecuación, obtenemos

$$g(x+1) - g(x) = -2 + g(x) - g(x-1)$$

Fijando $g(0) - g(1) = c$, obtenemos

$$g(k) - g(k-1) = c - 2(k-1)$$

Usando $g(0) = 0$ y sumando concluimos que $c = N - 1$ y que $g(x) = x(N - x)$.

Duración de juegos no honestos Consideremos la ruina del jugador con $p(x, x+1) = p\; p(x+1, x) = (1 - p) = q$. Dejamos al lector verificar que

$$E_x T = \frac{x}{q-p} - \frac{N}{q-p}\frac{1-(q/p)^x}{1-(q/p)^N}$$

Esto implica que si $p < q$ el juego tiene deriva hacia la izquierda y

$$\frac{N}{1-(q/p)^N} \to_N 0 \qquad \text{y} \qquad g(x) \to_N \frac{x}{q-p}$$

De hecho, la ganancia promedio por jugada es $p - q$ (que es negativo), por lo tanto necesitamos un tiempo del orden $|x/(p-q)|$ para perder x pesos.

Cuando $p > q$, vale que $(q/p)^N \to 0$ y tenemos

$$g(x) \approx \frac{N-x}{p-q}[1-(q/p)^x] + \frac{x}{p-q}(q/p)^x.$$

Vimos que la probabilidad de tocar 0 antes que N tiende a $1-(q/p)^x$. El valor $\frac{N-x}{p-q}$ es el tiempo promedio que va a necesitar la cadena para tocar N. El valor $\frac{x}{p-q}$ es el tiempo promedio que va a necesitar la cadena para tocar 0, *condicionado* a que va a tocar 0, asintoticamente en N.

Espacios con infinitos estados

Recurrencia no es suficiente para garantizar la existencia de una distribución estacionaria.

Paseo aleatorio reflejado Cadena de nacimiento en $\{0,1,\dots\}$ con transiciones

$$\begin{aligned} p(x,x+1) &= p, \quad i \ge 0,\\ p(x+1,x) &= 1-p, \quad i \ge 0,\\ p(0,0) &= 1-p \end{aligned}$$

Balance detallado:

$$p\pi(x) = (1-p)\pi(x+1), \qquad i \ge 0.$$

De donde

$$\pi(x+1) = \pi(x)\frac{p}{1-p}$$

y fijando $\pi(0) = c$ tenemos

$$\pi(x) = c\Big(\frac{p}{1-p}\Big)^x$$

Hay tres casos. Cuando $p < 1/2$ tenemos $\theta := p/(1-p) < 1$ y

$$\sum_{x\ge 0}\theta^x = \frac{1}{1-\theta}$$

Esto implica que que $c = 1-\theta$ y

$$\pi(x) = (1-\theta)\theta^x, \qquad x \ge 0.$$

geométrica de parámetro θ.

Cuando $p > 1/2$ todos los estados son transitorios. Efectivamente, vimos en (30) que en este caso, $P_x(T_0 < \infty) = ((1-p)/p)^x < 1$.

Cuando $p = 1/2$ todos los estados son recurrentes pero $E_0T_0 = \infty$. Vimos que

$$P_x(V_N < V_0) = \frac{x}{N} \underset{N\to\infty}{\longrightarrow} 0$$

Por lo tanto, $P_x(V_0 < V_N) \to 1$ y todo x es recurrente.

Veamos el tiempo promedio para retornar a 0. Si $V = \min\{n \ge 1 : X_n \notin \{0,N\}\}$, tenemos que $V \le T_0$. Por otro lado en (34) calculamos $E_1V = N-1$. Como esto vale para todo N, concluimos que $E_1T_0 \ge E_1V \to_N \infty$. Condicionando al primer paso:

$$E_0T_0 = \frac{1}{2}1 + \frac{1}{2}E_1T_0 = \infty$$

Decimos que x es *recurrente positivo* si $E_xT_x < \infty$.

Si x es recurrente pero no recurrente positivo, decimos que x es *recurrente nulo*.

Resumiendo el caso del paseo aleatorio reflejado, tenemos:

$p < 1/2$ implica 0 es recurrente positivo,
$p = 1/2$ implica 0 es recurrente nulo,
$p > 1/2$ implica 0 es transitorio.

Recurrencia positiva y distribución estacionaria Recordemos que si la cadena tiene una distribución estacionaria π, entonces

$$\pi(x) = \frac{1}{E_xT_x}$$

Por lo tanto, si $E_xT_x = \infty$, tendríamos $\pi(x) = 0$, es decir que si hay una distribución estacionaria π con $\pi(x) > 0$, entonces x es recurrente positivo.

Teorema 45 *Para una cadena irreducible, son equivalentes*

i. Algún estado es recurrente positivo.
ii. Hay una distribución estacionaria π.
iii. Todos los estados son recurrentes positivos.

Demostración (i) implica (ii): Sea x recurrente positivo. La masa de la medida μ_x es

$$\sum_y \mu_x(y) = \sum_{n=0}^{\infty} \sum_y P_x(X_n = y, T_x > n) = \sum_{n=0}^{\infty} P_x(T_x > n) = E_x T_x < \infty.$$

Esto implica que $\pi(y) = \mu_x(y)/E_x T_x$ es una distribución estacionaria.

(ii) implica (iii): Irreducibilidad implica $\pi(y) > 0$ para todo y y por el Teorema 41 $\pi(y) = 1/E_y T_y$, es decir que $E_y T_y < \infty$ para todo y.

(iii) implica (i): es trivial. □

Proceso de ramificación Z_n = tamaño de una población en el instante n. Cada individuo tiene un número aleatorio de hijos distribuidos como una variable aleatoria $\xi \geq 0$ con media $E\xi = \mu < \infty$ y distribución $P(\xi = j) = p_j$. Sean $\xi_{n,k}, n, k \geq 1$ iid con la misma distribución de ξ. Aquí $\xi_{n,k}$ es el número de hijos que tiene el k-ésimo individuo vivo en el instante n. Definimos $Z_0 = 1$ y para $n \geq 1$,

$$Z_n = \sum_{k=1}^{Z_{n-1}} \xi_{n,k}$$

Condicionando a Z_{n-1},

$$EZ_n = E(E(Z_n | Z_{n-1})) = \mu E Z_{n-1} = \mu^2 E Z_{n-2} = \mu^n.$$

Teorema subcrítico *Si $\mu < 1$ entonces*

$$P(Z_n \geq 1, \ \text{para todo } n \geq 0) = 0$$

Dem Por la desigualdad de Markov

$$P(Z_n \geq 1) \leq EZ_n = \mu^n \underset{n\to\infty}{\longrightarrow} 0, \qquad \text{si } \mu < 1.$$

Como $\{Z_n \geq 1\} \nearrow \{Z_n > 0, \ \text{para todo } n \geq 0\}$, podemos concluir. □

Teorema crítico *Si $\mu = 1$ y $P(\xi = 1) < 1$, entonces*

$$P(Z_n \geq 1, \ \text{para todo } n \geq 0) = 0$$

Teorema supercrítico *Si $\mu > 1$, entonces*

$$P(Z_n \geq 1, \ \text{para todo } n \geq 0) > 0$$

Defina la función $\phi : [0,1] \to [0,1]$ por $\phi(0) = p_0$ y para $s \in (0,1]$,

$$\phi(s) = \sum_{j \geq 0} s^j p_j$$

ϕ es continua en $[0,1]$ y si $p_0 + p_1 < 1$ (que asumimos, si no, se trata de un paseo aleatorio que ya vimos), para $s \in (0,1)$,

$$\phi'(s) = \sum_{j \geq 1} j s^{j-1} p_j > 0$$
$$\phi''(s) = \sum_{j \geq 2} j(j-1) s^{j-2} p_j > 0$$

O sea que ϕ es estrictamente creciente y estrictamente convexa en el intervalo $(0,1)$. Además $\lim_{s\nearrow 1}\phi'(s)=\mu$.

Defina $\theta_n = P(Z_n = 0|Z_0 = 1)$. Como $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$, tenemos $\theta_n \le \theta_{n+1}$. Además $\theta_0 = 0$, $\theta_n \le 1$. Por lo tanto $\theta_n \nearrow \theta_\infty \le 1$.

Condicionando a la primera generación,

$$\begin{aligned}\theta_n &= \sum_{j\ge 0} P(Z_n = 0|Z_1 = j)P(Z_1 = j|Z_0 = 1)\\ &= \sum_{j\ge 0} \left(P(Z_{n-1} = 0|Z_0 = 1)\right)^j p_j\\ &= \phi(\theta_{n-1}).\end{aligned}$$

La segunda igualdad se explica así: la probabilidad que el proceso se haya extinguido en el instante n sabiendo que hay j individuos en el instante 1 es igual a la probabilidad que cada una de las familias de los j individuos vivos en el instante 1 se haya extinguido en el instante n. Para concluir observe que las j familias evolucionan independientemente, y hay $n-1$ generaciones entre el instante 1 y el n.

Sacando límites, vemos que $\theta_\infty = \phi(\theta_\infty)$, un punto fijo de ϕ.

Si ρ es un punto fijo,

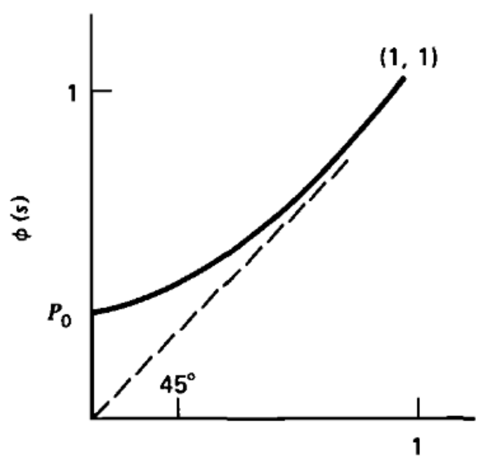
$$\rho = \sum_{k\ge 0} \rho^k p_k \ge \rho^0 p_0 = P(X_1 = 0) = \theta_1.$$

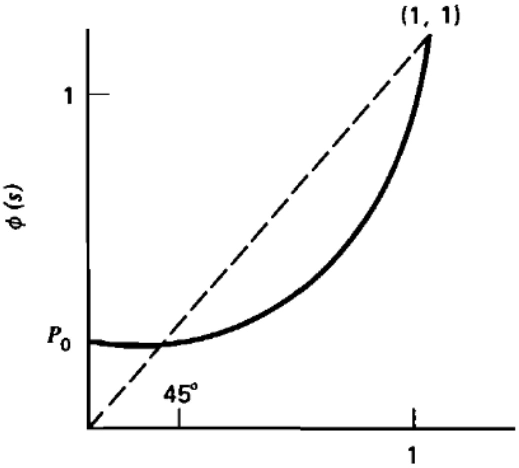
Como $\theta_1 \le \rho$ y si $\theta_n \le \rho$ entonces $\theta_{n+1} = \phi(\theta_n) \le \phi(\rho) = \rho$, tenemos que $\theta_n \le \rho$ para todo n. Es decir que θ_n converge al menor de los puntos fijos.

Dem del teorema crítico *Si $\phi'(1) = \mu = 1$ y $p_1 < 1$, como ϕ es estrictamente convexa $\phi(s) > s$ para $s \in (0,1)$ y ϕ tiene 1 como único punto fijo. Por lo tanto $\theta_n \to 1$.* □

Dem del teorema supercrítico *Si $\phi'(1) = \mu > 1$, entonces hay un único $\rho < 1$ tal que $\phi(\rho) = \rho$. Para ver esto, observe que $\phi(0) = p_0 \ge 0$, $\phi(1) = 1$ y $\phi'(1) = \mu > 0$, lo que implica que hay un único punto fijo ρ menor que 1. Unicidad es consecuencia de la estricta convexidad de ϕ. Por lo tanto $\theta_n \nearrow \rho < 1$.* □

Distinguimos dos casos: A la izquierda $\phi(s) > s$ para todo $s \in (0,1)$ y a la derecha $\phi(s) = s$ para algún





$s \in (0,1)$. En la figura de la izquierda $\phi'(1) \le 1$ y en la de la derecha $\phi'(1) > 1$.

Ejemplo. *Considere que la distribución del número de hijos es Poisson con parámetro λ. Es decir*

$$p_j = \frac{e^{-\lambda}\lambda^j}{j!}$$

La generadora de momentos es

$$\phi(s) = \exp(\lambda(s-1))$$

Por lo que la ecuación para el punto fijo es

$$\rho = \exp(\lambda(\rho - 1))$$

Relación con recurrencia *Existe (pero no vamos a probar) el siguiente teorema.*

Teorema 46 *Si $\mu = 1$ y $\sigma^2 = V\xi > 0$, entonces $P(X_n > 0) \sim \frac{2}{n\sigma^2}$. En particular $E_1T_0 = \infty$.*

Modifiquemos el proceso de ramificación para que resurja en 1 cada vez que muere: $p(0,1) = 1$. En ese caso tendremos

$\mu < 1$ implica 0 es recurrente positivo,
$\mu = 1$ implica 0 es recurrente nulo,
$\mu > 1$ implica 0 es transitorio.

Aplicación a una fila *Un cajero atiende a un cliente por minuto. En cada minuto llegan k clientes con probabilidad a_k, $k \geq 0$. Las llegadas en minutos diferentes son independientes. Sea X_n el número de personas esperando cuando parte el n-ésimo cliente. La matriz es*

$$p(0,k) = a_k, \qquad p(x, x-1+k) = a_k, \qquad p(x,y) = 0, \text{ en los otros casos.}$$

Si pensamos que los clientes que llegan en el instante en que es atendido el cliente i son los hijos de i, entonces la fila da lugar a un proceso de ramificación. Si ese proceso de ramificación se extingue después de haber tenido ℓ hijos, entonces la fila que empezó con 0 clientes en el instante ℓ se encuentra en el estado 0.

Como en ramificación el número de hijos hasta la extinción es finito con probabilidad uno si $\mu \leq 1$ y su esperanza es $\sum_n EZ_n = \sum_n \mu^n < \infty$, para $\mu < 1$, tenemos que la fila satisface

$\mu < 1$ implica 0 es recurrente positivo,
$\mu = 1$ implica 0 es recurrente nulo,
$\mu > 1$ implica 0 es transitorio.

Procesos de Poisson

Distribución exponencial *$T \sim$Exponencial(λ) si $P(T \leq t) = (1 - e^{-\lambda t})\mathbf{1}\{t \geq 0\}$.*

Densidad $f_T(t) = \lambda e^{-\lambda t}\mathbf{1}\{t \geq 0\}$

$ET = \frac{1}{\lambda}$

$ET^2 = \frac{2}{\lambda^2}$

$VT = ET^2 - (ET)^2 = \frac{1}{\lambda^2}$

Si $S \sim$ Exponencial(1), entonces $T = S/\lambda \sim$Exponencial(λ).

Falta de memoria:

$$P(T > t + s | T > s) = P(T > t).$$

Carrera de exponenciales: $T \sim$ Exponencial(λ), $S \sim$ Exponencial(μ), independientes, entonces $\min\{S,T\} \sim$ Exponencial$(\lambda + \mu)$

Quien gana?

$$\begin{aligned} P(S < T) &= \int_0^\infty f_S(s) P(T > s) ds = \int_0^\infty \mu e^{-\mu s} e^{-\lambda s} ds \\ &= \frac{\mu}{\lambda + \mu} \int_0^\infty (\lambda + \mu) e^{-(\lambda + \mu)s} ds = \frac{\mu}{\lambda + \mu} \end{aligned}$$

Lema 47 $T_i \sim$ *Exponencial*(λ_i) *independientes,* $V := \min\{T_1, \dots, T_n\}$, $I :=$ *índice del menor de los* Y_i*:*

$$I = i \quad \Leftrightarrow \quad Y_i = \min\{Y_1, \dots, Y_j\}$$

Entonces I *y* V *son independientes con distribución*

$$P(I = i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}, \quad V \sim \text{Exponencial}(\lambda_1 + \cdots + \lambda_n)$$

Demostración Denotemos $\lambda = \lambda_1 + \cdots + \lambda_n$

$$P(I = i, V > t) = \int_t^\infty \lambda_i e^{-\lambda_i s} \prod_{j \neq i} e^{-\lambda_j s} ds = \frac{\lambda_i}{\lambda} \int_t^\infty \lambda e^{-\lambda s} ds = \frac{\lambda_i}{\lambda} e^{-\lambda t} \tag{35}$$

como eso es el producto de una función de i *por una función de* t*, tenemos que* I *y* V *son independientes.*

Calculemos las distribuciones marginales. Poniendo $t = 0$*, obtenemos*

$$P(I = i) = P(I = i, V > 0) = \frac{\lambda_i}{\lambda}$$

y sumando (35) *sobre* $i \in \{1, \dots, n\}$ *obtenemos*

$$P(V > t) = \sum_i P(I = i, V > t) = \frac{\lambda}{\lambda} e^{-\lambda t} = e^{-\lambda t}. \qquad \square$$

Proposición 48 τ_i *independientes Exponencial*(λ)*. Entonces* $T_n := \tau_1 + \cdots + \tau_n \sim$*Gama*$(n, \lambda)$*. Es decir,*

$$f_{T_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \mathbf{1}\{t \geq 0\}.$$

Distribución Poisson *Una variable aleatoria* $X \sim$ *Poisson*(λ) *si*

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \qquad n = 0, 1, 2, \dots$$

Lema 49

$$E\big(X(X-1)\dots(X-k+1)\big) = \lambda^k$$

De donde sigue que $VX = \lambda$.

Demostración $X(X-1)\dots(X-k+1) = 0$ *si* $X \leq k-1$*. Así*

$$\begin{aligned} E\big(X(X-1)\dots(X-k+1)\big) &= \sum_{j=k}^\infty e^{-\lambda} \frac{\lambda^j}{j!} j(j-1)\dots(j-k+1) \\ &= \lambda^k \sum_{j=k}^\infty e^{-\lambda} \frac{\lambda^{j-k}}{(j-k)!} = \lambda^k \end{aligned}$$

Usando $VX = E(X(X-1)) + EX - (EX)^2$ *concluimos*

$$VX = \lambda^2 + \lambda - \lambda^2 = \lambda. \qquad \square$$

Lema 50 $X_i \sim$*Poisson*(λ_i) *independientes, entonces*

$$X_1 + \cdots + X_k \sim \text{Poisson}(\lambda_1 + \cdots + \lambda_k)$$

Definición del proceso de Poisson *Veremos dos definiciones. Proceso de Poisson de tasa* λ.

Definición *Sean $\tau_1, \tau_2, \dots$ iid Exponencial(λ). Sea $T_0 = 0$ y $T_n := \tau_1 + \cdots + \tau_n$. Defina $N(s) := \max\{n : T_n \le s\}$.*

τ_n tiempos entre llegadas a un banco.

T_n = instante de la n-ésima llegada.

$N(t)$ es el número de llegadas hasta el instante t

$N(s) = k$ si y sólo si $T_k \le s < T_{k+1}$.

Lema 51 *$N(s) \sim$ Poisson(λ) (como variable aleatoria).*

Demostración $N(s) = k$ si y sólo si $T_k \le s < T_{k+1}$.

$$\begin{aligned}
P(N(s) = n) &= \int_0^s f_{T_n}(t) P(\tau_{n+1} > s - t) dt \\
&= \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda(s-t)} dt, \qquad \text{(porque } T_n \sim \text{Gama}(n, \lambda)) \\
&= \frac{(\lambda t)^n}{(n-1)!} e^{-\lambda s} \int_0^s t^{n-1} dt = e^{-\lambda s} \frac{(\lambda s)^n}{n!}. \qquad \square
\end{aligned}$$

Lema 52 *$N(t+s) - N(s)$, $t \ge 0$ es un proceso de Poisson de tasa λ independiente de $N(r)$, $0 \le r \le s$.*

Demostración *Si llamamos $\tilde\tau_1, \tilde\tau_2, \dots$ los tiempos entre llegadas después de s, condicionando a $\{N(s) = n, T_n = u\}$ con $u \in [0, s)$, tendremos $\tilde\tau_1 = T_{n+1} - s$ y su distribución se puede calcular así:*

$$\begin{aligned}
P(\tilde\tau_1 > t | T_n = u, N(s) = n) &= P(T_{n+1} - s > t | T_n = u, N(s) = n) \\
&= P(T_{n+1} - s > t | T_n = u, T_{n+1} > s) \\
&= P(\tau_{n+1} > t + s - u | \tau_{n+1} > s - u, T_n = u) \\
&= P(\tau_{n+1} > t + s - u | \tau_{n+1} > s - u) = e^{-\lambda t}
\end{aligned}$$

donde la segunda igualdad se deduce de la independencia entre T_n y τ_{n+1} y la tercera por falta de memoria de la exponencial. Como, por probabilidad total tenemos

$$\begin{aligned}
&P(\tilde\tau_1 > t) \\
&= \sum_{n \ge 0} \int_0^s P(T_{n+1} - s > t | T_n = u, N(s) = n) P(N(s) = n) f_{T_n | N(s) = n}(u) du \\
&= e^{-\lambda t} \sum_{n \ge 0} \int_0^s P(N(s) = n) f_{T_n | N(s) = n}(u) du = e^{-\lambda t}.
\end{aligned}$$

podemos concluir que $\tilde\tau_1 \sim$ Exponencial(λ).

Para los intervalos entre llegadas sucesivos se hace el mismo cálculo, condicionando a $\{N(s) = n, T_n = u\}$ tenemos $\tilde\tau_1 = T_{n+1} - s = \tau_{n+1} - (s - u)$ y $\tilde\tau_j = \tau_{n+j}$ para $j \ge 2$. Por ejemplo:

$$\begin{aligned}
P(\tilde\tau_2 > t_2, \tilde\tau_1 > t_1) &= P(\tau_{n+1} > t_1 + s - u, \tau_{n+2} > t_2 | \tau_{n+1} > s - u) \\
&= e^{-\lambda t_1} e^{-\lambda t_2}.
\end{aligned}$$

Para ver que $\tilde\tau_1, \tilde\tau_2, \dots$ es independiente de $N(t), t \le s$, basta ver que el cálculo de la distribución de $\tilde\tau_j$ depende de $N(t), t \le s$ solamente a traves de $T_{N(s)}$ y vimos que en realidad es independiente de $T_{N(s)}$. $\square$

Lema 53 *El proceso de Poisson $N(t)$ tiene incrementos independientes: si $t_0 < t_1 < \ldots < t_n$ entonces*

$$N(t_1) - N(t_0), \dots, N(t_n) - N(t_{n-1}) \text{ son independientes.}$$

Demostración *El Lema 52 dice que $N(t_n) - N(t_{n-1})$ es independiente de $N(r), r \le t_{n-1}$ y por lo tanto de $N(t_1) - N(t_0), \dots, N(t_{n-1}) - N(t_{n-2})$. Concluya usando inducción.* $\square$

Segunda definición de proceso de Poisson

Teorema 54 *$(N(s) : s \geq 0)$ es un proceso de Poisson si y sólo si se satisfacen las tres condiciones siguientes*
(i) $N(0) = 0$,
(ii) $N(t_s) - N(s) \sim$ Poisson(λt)
(iii) $N(t)$ tiene incrementos independientes.

Demostración *Los lemas 52 y 53 demuestran [$\Rightarrow$]. Para ver la vuelta, sea T_n el instante de la n-ésima llegada y $\tau_j = T_j - T_{j-1}$ ($T_0 = 0$).*

$$P(\tau_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

Calculemos la distribución de τ_2:

$$\begin{aligned} P(\tau_2 > t|\tau_1 = s) &= P(\text{no hay llegadas en } (s, t+s]|\tau_1 = s) \\ &= P(N(t+s) - N(s) = 0|N(r) = 0, r < s; N(s) = 1) \\ &= P(N(t+s) - N(s) = 0) = e^{-\lambda t} \end{aligned}$$

por incrementos independientes. Siga por inducción. $\square$

La distribución Binomial con p chico aproxima la Poisson *Sea $X_n \sim$ Binomial$(n, \lambda/n)$, es decir $p_n = \lambda/n$.*

Sabemos que

$$\lim_n P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \geq 0.$$

es decir que aproxima la distribución Poisson.

Suma de Bernoulli con p chico aproxima Poisson. Cotas.

Teorema 55 *Para $1 \leq m \leq n$ sean*

$$\begin{aligned} X_{m,n} &\sim \text{Bernoulli}(p_{m,n}), \quad \text{independientes.} \\ Y_{m,n} &\sim \text{Poisson}(p_{m,n}), \quad \text{independientes.} \end{aligned}$$

Sea

$$\begin{aligned} S_n &:= X_{1,n} + \cdots + X_{n,n}; \\ Z_n &:= Y_{1,n} + \cdots + Y_{n,n} \end{aligned}$$

Entonces

$$\|S_n - Z_n\| := \frac{1}{2}\sum_k |P(S_n = k) - P(Z_n = k)| \leq \sum_{m=1}^n p_{m,n}^2.$$

Demostración *Comparemos primero $X \sim$ Bernoulli(p) con $Y \sim$Poisson(p):*

$$\begin{aligned} &2\|X - Y\| = \sum_k |P(X = k) - P(Y = k)| \\ &= |P(X = 0) - P(Y = 0)| + |P(X = 1) - P(Y = 1)| + \sum_{k\geq 2} |P(X = k) - P(Y = k)| \\ &= |1 - p - e^{-p}| + |p - pe^{-p}| + |0 - (1 - (1+p)e^{-p})| \\ &\leq e^{-p} - 1 + p + p - pe^{-p} + 1 - e^{-p} - pe^{-p} \\ &= 2p(1 - e^{-p}) \leq 2p^2, \end{aligned} \tag{36}$$

donde usamos $1 - p \le e^{-p} \le 1$ en las dos desigualdades.

Queda como ejercicio demostrar que

$$\|S_n - Z_n\| \le \sum_{m=1}^{n} \|X_{m,n} - Y_{m,n}\| \le p_{m,n}^2,$$

por (36). $\square$

Note que $Z_n \sim$ Poisson$(p_{1,n} + \cdots + p_{n,n})$.

Esa aproximación es útil si el máximo de los $p_{m,n}$ es chico:

$$2\sum_{m=1}^{n} p_{m,n}^2 \le 2(\max_m p_{m,n}) \sum_{m=1}^{n} p_{m,n}. \qquad \square$$

Proceso de Poisson no homogeneo *Decimos que $(N(s) : s \ge 0)$ es un proceso de Poisson no homogeneo de tasa $\lambda(r), r \ge 0$ si se satisfacen las tres condiciones siguientes*
(i) $N(0) = 0$.
(ii) $N(t) - N(s) \sim$ Poisson$\int_s^t \lambda(r)dr$.
(iii) $N(t)$ tiene incrementos independientes.

La definición con los tiempos inter-llegadas exponenciales independientes no funciona. Los τ_i no son ni exponenciales ni independientes.

$$P(\tau_1 > t) = P(N(t) = 0) = \exp\Big(\int_0^t \lambda(r)dr\Big) = e^{-\mu(t)},$$

donde

$$\mu(t) := \int_0^t \lambda(r)dr.$$

Entonces

$$f_{\tau_1}(t) = \lambda(t)e^{-\mu(t)}.$$

y la densidad conjunta de τ_1, τ_2 es

$$f_{\tau_1,\tau_2}(t,s) = \lambda(t)e^{-\mu(t)}\lambda(t+s)e^{\mu(s)-\mu(t)}$$

que no factorizan si $\lambda(r)$ no es constante.

Proceso de Poisson compuesto Restaurante en la ruta. *Autos llegan a un parador de acuerdo a un proceso de Poisson de parámetro λ por hora. En el i-ésimo auto hay Y_i personas, con distribución independiente del instante de su llegada. Cuantos clientes espera el restaurante entre las 11:30 y las 14:30, el horario de almuerzo?*

Mensajes electrónicos en un servidor *Mensajes llegan a un servidor de mails de acuerdo a un Proceso de Poisson de parámetro λ por minuto. El tamaño del i-ésimo mensaje es una variable aleatoria Y_i independiente del instante de llegada. Cual es el tamaño total de los mensajes llegados en 10 minutos?*

Definimos

$$S(t) = Y_1 + \cdots + Y_{N(t)}$$

Lema 56 *Sea N una variable aleatoria entera no negativa y $S = Y_1 + \cdots + Y_N$, con Y_i iid, independientes de N. Entonces*
(i) Si $E|Y_i| < \infty$, $EN < \infty$, entonces $ES = EY_1 EN$. Wald.
(ii) Si $EY_i^2 < \infty$, $EN^2 < \infty$, entonces $VS = EN\,VY_1 + VN\,(EY_1)^2$.
(iii) Si $N \sim$ Poisson(λ), entonces $VS = \lambda EY_1^2$.

Demostración *(i) es la identidad de Wald, ya demostrada.*

(ii) Como $VX = EX^2 - (EX)^2$*, tenemos*

$$E(S^2|N = n) = E(X_1 + \cdots + X_n)^2 = nVY_1 + (nEY_1)^2$$

De donde

$$\begin{aligned} ES^2 &= \sum_n E(S^2|N = n)P(N = n) = \sum_n (nVY_1 + (nEY_1)^2)P(N = n) \\ &= VY_1 \sum_n nP(N = n) + (EY_1)^2 \sum_n n^2 P(N = n) \\ &= VY_1\, EN + (EY_1)^2\, EN^2 \end{aligned}$$

Como $ES = EN\, EY_1$*, tenemos*

$$\begin{aligned} VS &= ES^2 - (ES)^2 = VY_1\, EN + (EY_1)^2\, EN^2 - (EN)^2\, (EY_1)^2 \\ &= VY_1\, EN + (EY_1)^2\, (EN^2 - (EN)^2) \end{aligned}$$

que demuestra (ii).

(iii) Usamos que $EX^2 = VX + (EX)^2$ *y que* $EN = VN = \lambda$ *para obtener (iii).* □

Cerveza *En un bar los clientes llegan de acuerdo a un proceso de Poisson de tasa 81 y que cada cliente gasta \$8 con un desvío standard de \$6. La facturación promedio por día del bar es* $81\,\$8 = \$648$*. La variancia de la facturación es*

$$81(6^2 + 8^2) = 8100.$$

El desvío standard del día es \$90, *a comparar con la esperanza \$648.*

Adelgazamiento *Supongamos que* Y_i *son variables discretas.*

Sea

$$N_j(t) := \sum_{i=1}^{N(t)} \mathbf{1}\{Y_i = j\}.$$

número de $i \le N(t)$ *tales que* $Y_i = j$*.*

Teorema 57 $N_j(t)$ *son Procesos de Poisson de tasa* $\lambda_j = \lambda P(Y_1 = j)$*, independientes.*

Demostración *Asuma primero que* $P(Y_i = 1) = p = 1 - P(Y_i = 2)$*.*

Por la independencia de incrementos del proceso de Poisson $N(t)$ *y la independencia de* Y_i *es claro que los incrementos del vector* $(N_1(t), N_2(t)), t \ge 0$ *son independientes. Como es claro que* $N_i(0) = 0$*, falta ver que*

$$X_i := N_i(t + s) - N_i(s), \quad i = 1, 2,$$

son independientes y Poisson. Sea $X = X_1 + X_2 = N(t + s) - N(s)$*.*

$$\begin{aligned} P(X_1 = j, X_2 = k) &= P(X = j + k)P(X_1 = j, X_2 = k|X = j + k) \\ &= e^{\lambda t} \frac{(\lambda t)^{j+k}}{(j + k)!} \frac{(j + k)!}{j!k!} p^j (1 - p)^k \\ &= e^{\lambda pt} \frac{(\lambda pt)^j}{j!}\, e^{\lambda(1-p)t} \frac{(\lambda(1 - p)t)^k}{j!} \end{aligned}$$

Para el caso general se usa la multinomial con $p_j = P(Y_1 = j)$ *y con* $X = X_1 + \cdots + X_m$ *se obtiene*

$$\begin{aligned} &P(X_1 = k_1, \ldots, X_m = k_m) \\ &= P(X = k_1 + \cdots + k_m)P(X_1 = k_1, \ldots, X_m = k_m|X = k_1 + \cdots + k_m) \\ &= \prod_{\ell=1}^{m} e^{\lambda p_\ell t} \frac{(\lambda p_\ell t)^{k_\ell}}{k_\ell!}. \qquad \square \end{aligned}$$

Este resultado se extiende al caso no homogeneo.

Teorema 58 *Si $N(t)$ es un proceso de Poisson de parámetro λ y $\tilde{N}(t)$ es el proceso obtenido al aceptar cada punto de $N(t)$ en s con probabilidad $p(s)$, independientemente de los otros, entonces $\tilde{N}(t)$ es un proceso de Poisson no homogeneo con tasa $\lambda p(s)$.*

Fila $M|G|\infty$. *Supongamos que las llamadas a una central llegan de acuerdo a un proceso de Poisson de tasa λ. La i-ésima llamada dura Y_i, una variable aleatoria con distribución acumulada G, es decir $P(Y_i \le t) = G(t)$, y esperanza $EY_i < \infty$. Asumimos que Y_i son independientes.*

Fijemos un instante $u > 0$. Llamamos $p(s)$ = probabilidad que una llamada llegada en el instante s esté en curso en el instante u; es decir $p(s) = 1 - G(u - s)$ para $s \le u$ y $p(s) = 0$ para $s > u$. Observe que $p(s)$ depende de u.

Si llamamos $\tilde{N}(t)$ = número de llamadas iniciadas antes de t y en curso en el instante u, tendremos que $\tilde{N}(u)$ es el número total de llamadas en curso en el instante u y por el teorema, $\tilde{N}(u)$ es una variable Poisson con media

$$E\tilde{N}(u) = \int_0^u \lambda p(s)ds = \int_0^u \lambda(1 - G(u-s))ds = \lambda \int_0^u (1 - G(r))dr$$

Tomando límite cuando $u \to \infty$, el número medio de llamadas en curso converge a

$$\lambda \int_0^\infty (1 - G(r))dr = \lambda EY_1.$$

Asumimos que el sistema empezó vacío, pero el límite es el mismo si empezamos con un número finito de llamadas.

Superposición

Teorema 59 *Suponga que $N_1(t), \dots, N_k(t)$ son procesos de Poisson independientes con tasas $\lambda_1, \dots, \lambda_k$, respectivamente. Entonces $N(t) := N_1(t) + \dots + N_k(t)$ es un proceso de Poisson de tasa $\lambda := \lambda_1 + \dots + \lambda_k$.*

Demostración *Es claro que $N(t)$ satisface $N(0) = 0$ y tiene incrementos independientes. Lo único que falta probar es que $N(T + s) - N(s)$ es Poisson de parámetro λ. Pero esto sigue del hecho que suma de variables aleatorias Poisson independientes es Poisson.* $\square$

Carrera de procesos de Poisson. *Sean $N_i(t)$ procesos de Poisson independientes de tasas λ_i. Cual es la probabilidad que las seis primeras llegadas de $N_1(t)$ ocurran antes que las 4 primeras llegadas de $N_2(t)$?*

Definiendo $N(t)$ como $N_1(t) + N_2(t)$ y $N_i(t)$ como adelgazamiento de $N(t)$ tenemos que cada llegada de $N(t)$ es de tipo 1 (pertenece a $N_1(t)$) con probabilidad $p = \lambda_1/(\lambda_1 + \lambda_2)$.

Aquí no importan los instantes de llegadas, sino el orden y tipo de llegada. Tenemos

$$\{\text{6 tipo 1 antes de 4 tipo 2}\} = \{\text{al menos 6 tipo 1 en las primeras 9 llegadas}\}$$
$$\{\text{4 tipo 2 antes de 6 tipo 1}\} = \{\text{al menos 4 tipo 2 en las primeras 9 llegadas}\}$$

Por lo tanto la probabilidad que buscamos es

$$\binom{9}{6}p^6(1-p)^3 + \binom{9}{7}p^7(1-p)^2 + \binom{9}{8}p^8(1-p) + \binom{9}{9}p^9$$

Si $\lambda_1 = \lambda_2$ tenemos $p = 1/2$ y esa probabilidad vale 0,273.

Condicionamiento *Sean $T_1, T_2, \dots$ los instantes de llegadas de un proceso de Poisson de tasa λ. Sean $U_1, \dots, U_n$ iid Uniforme$[0, t]$. Sean $V_1, \dots, V_n$ los estadísticos de orden de las uniformes:*

$$V_1 < \ldots < V_n \qquad y \qquad \{V_1, \dots, V_n\} = \{U_1, \dots, U_n\}.$$

El vector $(V_1, \dots, V_n)$ está distribuido uniformemente en $\{0 < t_1 < \ldots < t_n < t\}$. Su densidad es

$$f(v_1, \dots, v_n) = \frac{n!}{t^n}\mathbf{1}\{0 < t_1 < \ldots < t_n < t\}.$$

Teorema 60 *Si condicionamos a $N(t) = n$ entonces $(T_1, \dots, T_n)$ tiene la misma distribución que $(V_1, \dots, V_n)$ (como vectores). Por lo tanto $\{T_1, \dots, T_n\}$ tiene la misma distribución que $\{U_1, \dots, U_n\}$ (como conjuntos de puntos).*

Demostración *Hagamos el cálculo para $n = 3$. Estamos condicionando a $\{N(t) = 3\}$. La densidad de (T_1, T_2, T_3) condicionada es*

$$\begin{aligned}
&\text{“}P(T_1 = t_1, T_2 = t_2, T_3 = t_3, \tau_4 > t - t_3 \mid N(t) = 3)\text{”} \\
&= \frac{\lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \lambda e^{-\lambda(t_3 - t_2)} e^{-\lambda(t - t_3)}}{e^{-\lambda t}(\lambda t)^3/3!} \mathbf{1}\{0 < t_1 < t_2 < t_3 < t\} \\
&= \frac{3!}{t^3} \mathbf{1}\{0 < t_1 < t_2 < t_3 < t\}
\end{aligned}$$

O sea que la distribución de (T_1, T_2, T_3) condicionada a $N(t) = 3$ es uniforme en el conjunto $\{(t_1, t_2, t_3) : 0 < t_1 < t_2 < t_3 < t\}$. Ese conjunto tiene volumen $t^3/3!$ porque $\{(t_1, t_2, t_3) : 0 < t_1, t_2, t_3 < t\}$ tiene volumen t^3 y ese es uno de los $3!$ posibles órdenes entre las coordenadas.

El mismo razonamiento sirve para demostrar que la densidad de $(T_1, \dots, T_n)$ dado $N(t) = n$ está dada por $n!/t^n$. $\square$

Es decir que si $N(t) = n$, la posición de los puntos es la misma que la de n uniformes en $[0, t]$.

Teorema 61 *Si $s < t$ y $0 \le m \le n$, entonces*

$$P(N(s) = m \mid N(t) = n) = \binom{n}{m} \Big(\frac{s}{t}\Big)^m \Big(1 - \frac{s}{t}\Big)^{n-m}.$$

Es decir la distribución de $N(s)$ dado $N(t) = n$ es Binomial$(n, s/t)$.

Demostración

$$P(N(s) = m | N(t) = n) = P\Big(\sum_{i=1}^{n} \mathbf{1}\{U_i \le s\} = m\Big) \tag{37}$$

donde $U_i \sim$ Uniforme$[0, t]$ independientes, lo que implica $\mathbf{1}\{U_i \le s\} \sim$ Bernoulli(s/t) independientes. $\square$

Construcción de procesos de Poisson en $\mathbb{R}^d$

Sea $\lambda > 0$ y considere

1. *Una partición $\mathcal{J}$ de $\mathbb{R}^d$ ($\cup_{A \in \mathcal{J}} A = \mathbb{R}^d$ y $A \cap B = \emptyset$ para $A, B \in \mathcal{J}$). Asuma A Boreliano acotado con medida de Lebesgue $l(A) < \infty$ para todo $A \in \mathcal{J}$.*
2. *Una familia de variables aleatorias independientes $Y_A \sim$ Poisson$(\lambda l(A))$.*
3. *Una familia de variables aleatorias independientes $(U_{A,j}, j \ge 1, A \in \mathcal{J})$, con $U_{A,j}$ uniformemente distribuída en A. Esta familia es independiente de la familia (Y_A).*

Defina el conjunto aleatorio

$$S := \bigcup_{A \in \mathcal{J}} \{U_{A,j} : j \le Y_A\} \tag{38}$$

S es llamado proceso de Poisson *de tasa λ.*

Definición formal de proceso puntual *El proceso de Poisson que construimos es un caso particular de proceso puntual. Considere los subconjuntos de puntos de $\mathbb{R}^d$ localmente finitos:*

$$\mathcal{M} = \{s \subset \mathbb{R}^d : |s \cap C| < \infty, \text{ para todo conjunto } C \subset \mathbb{R}^d \text{ acotado}\}$$

donde $|*|$ es el cardinal de $*$. Considere la siguiente familia de subconjuntos de $\mathcal{M}$

$$\begin{aligned}\mathcal{A} := \big\{\{s \in \mathcal{M} \,:\, |s \cap B_i| = b_i, i = 1, \dots, n\},\\ B_1, \dots, B_n \text{ Borelianos acotados}, b_i, n \geq 0\big\}\end{aligned}$$

$\mathcal{A}$ es cerrada para intersecciones. Sea $\sigma(\mathcal{A})$ la menor sigma-álgebra que contiene a $\mathcal{A}$.

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad.

Un proceso puntual *es una función $S : \Omega \to \mathcal{M}$ tal que $\{S \in D\} \in \mathcal{F}$ para todo $D \in \sigma(\mathcal{A})$. Es decir S es una variable aleatoria con valores en $\mathcal{M}$.*

Observe que $N(B) := |S \cap B|$ es una variable aleatoria con valores en $\{0, 1, 2, \dots\}$. La distribución de S está caracterizada por la distribución del "vector" aleatorio

$$\big(N(B) : B \text{ Boreliano acotado}\big)$$

Llamaremos μ la distribución de un proceso puntual S:

$$\mu(D) := P(S \in D) \text{ para } D \in \sigma(\mathcal{A})$$

Teorema 62 *Sea $\mathcal{A}$ una familia de conjuntos de $\mathcal{M}$ cerrada para intersecciones. Si μ, ν son distribuciones de procesos puntuales en $\mathcal{M}$ y $\mu(D) = \nu(D)$ para todo $D \in \mathcal{A}$, entonces $\mu = \nu$.*

Hay una clase menor de eventos que determinan una probabilidad en $\mathcal{M}$.

Teorema 63 *Sea*

$$\mathcal{A}_0 := \big\{\{S \in \mathcal{M} \,:\, N_S(B) = 0\}, B \text{ Boreliano acotado}\big\} \tag{39}$$

Si μ, ν son probabilidades definidas en $\mathcal{M}$ y $\mu(D) = \nu(D)$ para todo $D \in \mathcal{A}_0$, entonces $\mu = \nu$.

Si μ es la distribución de un proceso puntual S, los números $\mu(S \in \mathcal{M} \,:\, N_S(B) = 0)$ son llamados probabilidades vacías *de μ.*

Más detalles en Apéndice B de Moeller y Waagepetersen [5].

Propiedades del proceso de Poisson

Teorema 64 *La distribución de un proceso de Poisson S no depende de la partición $\mathcal{J}$ usada para construirlo.*

La demostración de ese teorema se basa en el lema siguiente.

Lema 65 *Sea S el proceso de Poisson (38). Entonces $P(S \cap B = \emptyset) = e^{-\lambda l(B)}$, para todo Boreliano B.*

Demostración

$$\begin{aligned}P(S \cap B = \emptyset) &= P\Big(\bigcap_{A \in \mathcal{J}} \{S \cap B \cap A = \emptyset\}\Big)\\ &= \prod_{A \in \mathcal{J}} P(S \cap B \cap A = \emptyset)\end{aligned} \tag{40}$$

esto porque los eventos $\{S \cap B \cap A = \emptyset\}$ son independientes porque los $A \in \mathcal{J}$ son disjuntos. Usemos que $Y_A \sim$ Poisson$(\lambda l(A))$ para escribir

$$P(S \cap B \cap A = \emptyset) = \sum_{n=0}^{\infty} P(S \cap B \cap A = \emptyset | Y_A = n) e^{-\lambda l(A)} \frac{(\lambda l(A))^n}{n!} \tag{41}$$

Como $S \cap A = \{U_{i,1}, \dots, U_{i,Y_A}\}$, son uniformes independientes en A,

$$
\begin{aligned}
P(S \cap B \cap A = \emptyset | Y_A = n) &= P(B \cap \{U_{A,1}, \dots, U_{A,n}\} = \emptyset) \\
&= P(U_{A,1} \notin B) \dots P(U_{A,n} \notin B) \\
&= \Big(1 - \frac{l(B\ \cap A)}{l(A)}\Big)^n
\end{aligned}
$$

Así (41) es igual a

$$
P(S \cap B \cap A = \emptyset) = e^{-\lambda l(A)} \sum_{n=0}^{\infty} \Big(1 - \frac{l(B \cap A)}{l(A)}\Big)^n \frac{(\lambda l(A))^n}{n!} = e^{-\lambda l(B \cap A)}.
$$

Substituyendo en (40), estamos. $\square$

Demostración del Teorema 64 *El lema anterior muestra que las probabilidades vacías son independientes de la partición usada para definir el proceso.* $\square$

Vamos a usar la notación $N(B) = |S \cap B|$, es la medida de conteo aleatoria asociada a S.

Corolario 66 *(a) Para cada Boreliano acotado B,*

$$
P(N(B) = k) = \frac{e^{-\lambda l(B)}(\lambda l(B))^k}{k!}
$$

(b) Dado que $|S \cap B| = n$, la distribución de $S \cap B$ es la misma que la de n puntos uniformemente distribuidos en B.

(c) Sean $B_1, \dots, B_L$ Borelianos acotados disjuntos. Entonces $N(B_1), \dots, N(B_L)$ son independientes, Poisson con medias $\lambda l(B_1), \dots, \lambda l(B_L)$ respectivamente.

Demostración *Sea $\mathcal{J}$ una partición tal que $B \in \mathcal{J}$. Entonces (a) y (b) son consecuencias de la construcción: Y_B, el número de puntos de S en B es Poisson($\lambda l(B)$) y una vez decidido que $Y_B = n$, los puntos estan distribuidos uniformemente en B. Si elegimos la partición de tal manera que $B_1, \dots, B_L \in \mathcal{J}$, obtenemos (c).* $\square$

Proyecciones *Sea M la medida de conteo asociada a un proceso de Poisson de tasa 1 en $\mathbb{R}^2$. Queremos construir un proceso de Poisson de tasa λ en $\mathbb{R}$ en función de M.*

Para cada Boreliano $B \subset \mathbb{R}$ define

$$
N(B) := M(B \times [0, \lambda]), \qquad B \text{ Boreliano en } \mathbb{R}. \tag{42}
$$

es decir, el número de puntos de N en B es igual al número de puntos de M en $B \times [0, \lambda]$. Esto corresponde a proyectar los puntos de $M(\cdot)$ de la faja $\mathbb{R} \times [0, \lambda]$ sobre $\mathbb{R}$.

Notación $N_t = N([0, t]$.

Lema 67 *El proceso $N_t := N[0, t]$, definido en (42) es un proceso de Poisson en $\mathbb{R}$ de parámetro λ.*

Demostración *Es inmediato que tiene incrementos independientes y que $N_0 = 0$, Basta probar que para intervalos disjuntos $I_1, \dots, I_n$:*

$$
P(N(I_i) = k_i\,,\, i = 1, \dots, n) = \prod_{i=1}^{n} \frac{e^{-\lambda |I_i|}(\lambda |I_i|)^{k_i}}{k_i!},
$$

que sigue de la definición (42) y del hecho que M es Poisson en $\mathbb{R}^2$. $\square$

Lema 68 *Sea I un intervalo finito. El evento "dos puntos de M son proyectados en un único punto de I" tiene probabilidad cero.*

Demostración *Podemos asumir que $I = [0,1]$. Consideramos una partición de I en intervalos de longitud δ: $I_n^\delta = (n\delta, (n+1)\delta]$. La probabilidad que dos puntos sean proyectados en el mismo está acotada por arriba por la probabilidad que haya dos puntos en el mismo intervalo, que está dada por*

$$P(\cup_{n=1}^{|I|/\delta}\{M(I_n^\delta \times [0,\lambda]) \geq 2\}) \leq \sum_{n=1}^{|I|/\delta} P(M(I_n^\delta \times [0,\lambda]) \geq 2) \leq \frac{|I|}{\delta} o(\delta).$$

Esto se va a cero con $\delta \to 0$. $\square$

Acoplamiento *Considere tasas λ_1 y λ_2 y defina los procesos*

$$N_i(B) = M(B \times [0,\lambda_i]), \quad B \text{ Boreliano en } \mathbb{R}. \tag{43}$$

Lema 69 *Si $\lambda_1 \geq \lambda_2$, entonces*

$$N_1(B) \geq N_2(B),$$

para todo $B \subset \mathbb{R}$.

Demostración *Como $B \times [0,\lambda_1] \supset B \times [0,\lambda_2]$, tenemos*

$$N_1(B) = M(B \times [0,\lambda_1]) \geq M(B \times [0,\lambda_2]) = N_2(B). \quad \square$$

Superposición y adelgazamiento *Los hombres llegan a un cajero de acuerdo a un proceso de Poisson de tasa $p\lambda$; las mujeres a tasa $(1-p)\lambda$. Estos dos procesos son independientes.*

Construimos los procesos como proyección del proceso M bidimensional. Definimos

$$N_1(I) = M(I \times [0, p\lambda]); \quad N_2(I) = M(I \times [p\lambda, \lambda]).$$

Ya vimos que el proceso N definido por

$$N(I) = N_1(I) + N_2(I) = M([0,\lambda] \times I)$$

es Poisson de tasa λ.

Si T_n son los tiempos de llegada de N, marcamos cada uno de esos instantes como 1 o 2 de acuerdo a la faja de donde fueron proyectados:

$$G_i = G(T_i) = \begin{cases} 1 & \text{si } T_i \text{ viene de la faja 1} \\ 2 & \text{si } T_i \text{ viene de la faja 2.} \end{cases}$$

Proposición 70 *G_i son iid con marginales*

$$P(G_i = 1) = p\,; \qquad P(G_i = 2) = 1 - p$$

Teorema 71 *Sea $M(\cdot)$ un proceso de Poisson en $\mathbb{R}^2$ de tasa 1. Sean $S_1, S_2, \ldots$ los tiempos de ocurrencia de llegadas en la banda $[0,\lambda]$. Sean $W_1, W_2, \ldots$ las segundas coordenadas de esos tiempos. Entonces, $(S_{i+1} - S_i)_{i\geq 1}$ son iid exponencial (λ) y $(W_i)_{i\geq 1}$ son iid Uniforme$[0,\lambda]$. Además $(S_{i+1} - S_i)_{i\geq 1}$ y $(W_i)_{i\geq 1}$ son independientes.*

Procesos no homogeneos. *Sea $\lambda : \mathbb{R} \to \mathbb{R}^+$ no negativa continua por pedazos.*

Queremos construir un proceso de Poisson en $\mathbb{R}$ con tasa $\lambda(t)$. Es decir, un proceso N_t tal que, definiendo $\mu(t) = \int_0^t \lambda(r)dr)$,

$$P(N_t - N_{t+s}) = k) = \frac{e^{-(\mu(t)-\mu(s))}(\mu(t) - \mu(s))^k}{k!}$$

Consideramos un proceso bidimensional $M(\cdot)$ y definimos

$$N_t := M(\Lambda([0,t])), \tag{44}$$

donde $\Lambda([0,t]) = \{(s,y) \in \mathbb{R}^2 : s \in [0,t] \text{ y } y \leq \lambda(s)\}$. Es decir, N_t es el proceso obtenido al proyectar los puntos de $M(\cdot)$ que caen debajo de la curva $\lambda(s)$.

Lema 72 *El proceso N_t es un proceso no homogeneo de tasa $\lambda(\cdot)$.*

Acoplamiento

Proposición 73 *Sean $\lambda_1(t)$ y $\lambda_2(t)$ funciones que satisfacen*

$$\lambda_1(t) \leq \lambda_2(t) \text{ para todo } t$$

Entonces los procesos $N_1(\cdot)$ y $N_2(\cdot)$ definidos por (44) son Poisson con tasas $\lambda_1(t)$ y $\lambda_2(t)$ respectivamente y además para todo t

$$N_1([s,t]) \leq N_2([s,t]). \qquad \text{para todo } s < t.$$

Procesos de renovación

Sean t_n variables aleatorias independientes identicamente distribuídas positivas con distribución acumulada

$$F(t) = P(t_i \leq t)$$

y sea

$$T_n = t_1 + \cdots + t_n$$

Sea

$$N(t) = \max\{n : T_n \leq t\}$$

Ejemplos:

Lámparas *t_i duración de la i-ésima lámpara, T_n instante en que se quema la n-ésima lámpara, $N(t)$ número de lámparas que se quemaron hasta el instante t, $N(t) = n$ implica que la $(n+1)$-ésima lámpara está prendida.*

Vuelta a x en una cadena de Markov *X_n cadena de Markov con $X_0 = x$, T_n instante de la n-ésima vuelta a x, etc.*

Proceso de Poisson unidimensional *En este caso $t_i \sim$ Exponencial(λ).*

Teorema 74 *Sea $\mu = Et_i$ el tiempo medio entre llegadas. Si $P(t_i > 0) > 0$, entonces*

$$\lim_t \frac{N(t)}{t} = \frac{1}{\mu}, \qquad c.s.$$

Proceso de Poisson *En el proceso de Poisson $Et_i = 1/\lambda$, por lo tanto en ese caso el teorema dice que $\frac{N(t)}{t} \to \lambda$.*

Para probar el teorema usamos la ley fuerte de grandes números que vimos en Proba:

Teorema 75 (LGN) *Sean* $x_1, x_2, \ldots$ *iid con* $Ex_i = \mu$ *y sea* $S_n = x_1 + \cdots + x_n$. *Entonces*

$$\lim_n \frac{S_n}{n} = \mu, \qquad c.s.$$

Demostración del Teorema 74 Tomando $x_i = t_i$ tenemos que $S_n = T_n$. La LGN dice que $T_n/n \to \mu$. *Por definición*

$$T_{N(t)} \leq t < T_{N(t)+1}$$

Dividiendo por $N(t)$*, tenemos*

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{T_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)}$$

Por la LGN los lados izquierdo y derecho convergen a μ. *Por lo tanto* $t/N(t) \to \mu$ *y* $N(t)/t \to 1/\mu$. $\square$

Teorema 76 *Si* $Et_i < \infty$*, entonces*

$$\lim_t \frac{T_{N(t)+1} - t}{t} = 0. \tag{45}$$

$$\lim_t \frac{t_{N(t)+1}}{t} = 0. \tag{46}$$

La demostración de este teorema usa acoplamiento a un nivel superior al de este curso.

Renovación con recompensa *Supongamos que en el instante* T_i *ganamos una recompensa* r_i. *Asumimos que los pares* (t_i, r_i) *son iid. La recompensa al tiempo* t *se define por*

$$R(t) = \sum_{i=1}^{N(t)} r_i$$

Teorema 77

$$\lim_t \frac{R(t)}{t} = \frac{Er_i}{Et_i}, \qquad c.s.$$

Demostración *Multiplicando y dividiendo por* $N(t)$,

$$\frac{R(t)}{t} = \Big(\frac{1}{N(t)} \sum_{i=1}^{N(t)} r_i\Big) \frac{N(t)}{t} \to Er_i \frac{1}{Et_i},$$

donde usamos LGN para las iid r_i *y para* $N(t)$. $\square$

Costo medio de tener un auto *El tiempo de vida de un auto es una variable aleatoria continua con densidad* h. *Compro un auto nuevo cuando el viejo se rompe o llega a* T *años de viejo. Un auto nuevo cuesta* A *pesos. Si el auto viejo se rompe antes de* T*, hay que gastar* B *pesos para arreglarlo. Cuanto cuesta tener un auto por unidad de tiempo? (no consideramos lo que ganamos al vender el auto viejo).*

La duración del ciclo t_i *tiene esperanza*

$$Et_i = \int_0^T th(t)dt + T\int_T^\infty h(t)dt$$

El costo del ciclo i *tiene esperanza*

$$Er_i = A + B\int_0^T h(t)dt$$

Así, el costo por unidad de tiempo es

$$\frac{Er_i}{Et_i} = \frac{A + B\int_0^T h(t)dt}{\int_0^T th(t)dt + T\int_T^\infty h(t)dt}$$

Supongamos que el tiempo de vida es uniforme en $[0, 10]$: $h(t) = \frac{1}{10}\mathbf{1}\{t \in [0, 10]\}$. *Que* $A = 10$ *y* $B = 3$. *Para* $T \in [0, 10]$, *tenemos*

$$Er_i = 10 + 3\frac{T}{10} = 10 + 0{,}3T$$
$$Et_i = \int_0^T \frac{t}{10}dt + T\Big(1 - \frac{T}{10}\Big) = \frac{T^2}{20} + T - \frac{T^2}{10} = T - 0{,}05T^2$$

O sea

$$\frac{Er_i}{Et_i} = \frac{10 + 0{,}3T}{T - 0{,}05T^2}$$

Minimizando en relación a T *obtenemos* $T_{\text{mín}} = 8{,}83$. *Esto significa que la política de cambiar el auto si no se rompió a los 8,83 años minimiza el costo del auto por unidad de tiempo.*

Renovación alternada *Sean* $s_1, s_2, \dots$ *iid con distribución* F *y esperanza* μ_F *y* $u_1, u_2, \dots$ *iid con distribución* G *y esperanza* μ_G. *En cada ciclo una máquina trabaja* s_i *y está en reparación* u_i. *Es un ciclo alternado, donde estamos en uno de dos estados 1 (máquina en funcionamiento) o 2 (máquina en reparación).*

Teorema 78 *En un proceso de renovación alternada, el tiempo medio que pasa en el estado 1 es*

$$\frac{\mu_F}{\mu_F + \mu_G}.$$

Demostración *Ciclo dura* $t_i = s_i + u_i$. *Son iid con media* $Et_i = \mu_F + \mu_G$. *Pensamos que la recompensa en el ciclo* i *es el tiempo que pasa funcionando, es decir* $r_i = s_i$. *El tiempo en funcionamiento hasta el instante* t *es*

$$R(t) + \text{mín}\{s_{N(t)+1}, t - T_{N(t)}\}$$

Observe que el mínimo del segundo sumando es menor o igual a $s_{N(t)+1} \le t_{N(t)+1}$. *Asumiendo que* $t_{N(t)+1}/t \to 0$ *(que no estoy probando), tenemos que la proporción de tiempo en funcionamiento*

$$\frac{R(t)}{t} = \frac{Er_i}{Et_i} = \frac{\mu_F}{\mu_F + \mu_G}. \qquad \square$$

Inspección aleatoria de lámparas *Las lámparas tienen un tiempo de vida con distribución* F *y media* μ_F. *Un electricista inspecciona la lámpara en instantes de un proceso de Poisson*(λ). *Si la lámpara está quemada, la cambia. A que tasa son reemplazadas las lámparas? Qué proporción de tiempo está prendida la lámpara? Qué proporción de visitas son inútiles (porque la lámpara está funcionando)?*

La duración esperada del ciclo es $Et_i = \mu_F + \frac{1}{\lambda}$. *En cada ciclo se reemplaza una lámpara;* $r_i \equiv 1$. *Llamando* $R(t)$ *el número de lámparas cambiadas hasta* t, *tenemos*

$$\frac{R(t)}{t} \to \frac{Er_i}{Et_i} = \frac{1}{\mu_F + \frac{1}{\lambda}}$$

La lámpara está prendida en el ciclo i *por* $r_i = s_i$ *unidades de tiempo. Entonces, la fracción de tiempo que está prendida es*

$$\frac{\mu_F}{\mu_F + \frac{1}{\lambda}}$$

Si llamamos $V(t)$ *el número de visitas del electricista, tenemos* $V(t)/t \to \lambda$. *Así que la proporción de visitas en las cuales reemplaza una lámpara es*

$$\frac{N(t)}{V(t)} \to \frac{1/(\mu_F + \frac{1}{\lambda})}{\lambda} = \frac{1/\lambda}{\mu_F + \frac{1}{\lambda}}.$$

Y la proporción de visitas que encuentran la lámpara funcionando es su complemento:

$$1 - \frac{N(t)}{V(t)} \to \frac{\mu_F}{\mu_F + \frac{1}{\lambda}}.$$

Edad y residuo *Sean $t_1, t_2, \dots$ los tiempos entre llegadas, $T_n = t_1 + \cdots + t_n$ el instante de la n-ésima renovación y $N(t)$ el número de renovaciones hasta el instante t. Definimos*

$$\text{Edad: } A(t) = t - T_{N(t)}, \qquad \text{Residuo: } Z(t) = T_{N(t)+1} - t$$

En el proceso $N(t+s) - N(t)$, el instante de la primera llegada es $Z(t)$ y los instantes siguientes tienen distribución independiente de $Z(t)$. Por lo tanto si $Z(t)$ converge a Z cuando t tiende a infinito, tendremos

$$N(t+s) - N(t) \xrightarrow[t\to\infty]{D} \tilde{N}(s) \qquad \text{como proceso}$$

y $\tilde{N}(s)$ es un proceso estacionario:

$$\tilde{N}(s) \overset{D}{=} \tilde{N}(t+s) - \tilde{N}(t) \qquad \text{como proceso}$$

Los tiempos de llegadas para $\tilde{N}(s)$ tienen la misma distribución que $Z, Z+T_1, Z+T_2, \dots$.

Caso discreto *Sean T_m los instantes de llegadas de $N(t)$. Defina*

$$V_m = \mathbf{1}\{m \in \{T_0, T_1, \dots\}\}$$

V_m indica los instantes de renovación. Tenemos

$$\begin{aligned} A_n &= \min\{n - m : m \le n,\ V_m = 1\}, \\ Z_n &= \min\{m - n : m \ge n,\ V_m = 1\} \end{aligned}$$

Ejemplo:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
V_n	1	0	0	0	1	0	0	1	1	0	0	0	0	1
A_n	0	1	2	3	0	1	2	0	0	1	2	3	4	0
Z_n	0	3	2	1	0	2	1	0	0	4	3	2	1	0

Vemos que en cada ciclo entre dos unos de V_n aparecen los mismos valores de A_n y Z_n con el orden reverso. Eso quiere decir que a largo plazo $A_n = i$ va a aparecer tantas veces como $Z_n = i$:

$$\lim_n \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}\{A_m = i\} = \lim_n \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}\{Z_m = i\}$$

y sacando esperanzas,

$$\lim_n \frac{1}{n} \sum_{m=1}^{n} P(A_m = i) = \lim_n \frac{1}{n} \sum_{m=1}^{n} P(Z_m = i)$$

Z_n es una cadena de Markov con transiciones

$$\begin{aligned} p(0,j) &= f_{j+1}, && j \ge 0 \\ p(i,i-1) &= 1, && i \ge 1 \\ p(i,j) &= 0, && \text{en los otros casos} \end{aligned}$$

donde $f_j = P(t_i = j)$. La cadena es irreducible si $f_i > 0$ para infinitos i; si hay un K tal que $f_i = 0$ para $i > K$, entonces la cadena es irreducible en $\{0, \dots, K\}$.

La cadena es siempre recurrente en la clase de estados irreducibles.

Para definir una medida estacionaria usamos la construcción del ciclo. Entre dos visitas a 0, la cadena visita i si el salto es mayor o igual a i. Entonces

$$\mu(i) = P(t_1 > i)$$

Como

$$\sum_{i=0}^{\infty} \mu(i) = Et_1,$$

Habrá una distribución estacionaria π si $Et_1 < \infty$. En este caso:

$$\pi(i) = \frac{P(t_1 > i)}{Et_1}$$

Entonces, como aplicación de uno de los teoremas de convergencia, tenemos:

Teorema 79 *Suponga que $Et_1 < \infty$ y que el máximo común divisor de $\{k : f_k > 0\}$ es 1. Entonces.*

$$\lim_n P(Z_n = i) = \frac{P(t_1 > i)}{Et_1}$$

Caso general

Teorema 80

$$I := \lim_t \frac{1}{t}\int_0^t \mathbf{1}\{A_s > x, Z_s > y\}ds = \frac{1}{Et_1}\int_{x+y}^{\infty} P(t_i > z)dz$$

Demostración *Es fácil de ver que*

$$\{s \in [T_{i-1}, T_i] : A_s > x, Z_s > y\} = [T_{i-1} + x, T_i - y]$$

cuya longitud es $(t_i - (x+y))^+$. Así:

$$r_i := \int_{T_{i-1}}^{T_i} \mathbf{1}\{A_s > x, Z_s > y\}ds = (t_i - (x+y))^+$$

y como $(t_i - (x+y))^+ \leq t_i$, tenemos

$$I = \frac{\sum_{i=1}^{N(t)} r_i + O(t_{N(t)+1})}{t} \underset{t\to\infty}{\longrightarrow} \frac{Er_1}{Et_1}$$

Sólo falta calcular Er_1:

$$\begin{aligned} Er_1 &= E(t_i - (x+y))^+ = E(t_i\,\mathbf{1}\{t_i \geq x+y\}) \\ &= \int_o^{\infty} P(t_i\,\mathbf{1}\{t_i \geq x+y\} > z)dz = \int_{x+y}^{\infty} P(t_i > z)dz \end{aligned}$$

porque $(t_i - (x+y))^+$ es una variable no negativa. □

A partir de ese teorema obtenemos las distribuciones marginales de la edad y el residuo.

El teorema dice que si s es uniformemente distribuído entre 0 y t, entonces cuando $t \to \infty$ el vector $(A(s), Z(s))$ converge en distribución a un vector (A, Z) con densidad

$$P(A > x, Z > y) = \frac{1}{Et_1}\int_{x+y}^{\infty} P(t_i > z)dz \tag{47}$$

cuya densidad se obtiene diferenciando esa expresión en x e y:

$$f_{A,Z}(a,z) = \frac{1}{Et_1} f_{t_1}(a+z) \tag{48}$$

Poniendo $y = 0$ en (47) obtenemos la marginal de A

$$f_A(a) = \frac{1}{Et_1}P(t_1 > a), \qquad a \geq 0$$

y analogamente, poniendo $x = 0$ obtenemos la misma marginal para Z:

$$f_Z(z) = \frac{1}{Et_1}P(t_1 > z), \qquad z \geq 0$$

La esperanza de A es

$$EA = \int_0^{\infty} P(A > x)dx = \int_0^{\infty}\int_x^{\infty} f_A(x)dx \tag{49}$$

$$= \frac{1}{Et_1}\int_0^{\infty}\Big(\int_x^{\infty} P(t_1 > a)da\Big)dx = \frac{Et_1^2}{2Et_1}. \tag{50}$$

Veamos. Usando partes, para $X \geq 0$

$$\begin{aligned} EX^2 &= \int_0^\infty x^2 f(x)dx \\ &= \int_0^\infty 2x \int_x^\infty f(a)da \\ &= \int_0^\infty 2 \int_x^\infty P(X > a)dadx \end{aligned}$$

Caso exponencial. *Si $t_1 \sim$ exponencial(λ), la densidad conjunta es*

$$\frac{e^{-\lambda(x+y)}}{1/\lambda} = \lambda e^{-\lambda a} e^{-\lambda z}$$

En este caso la edad y el residuo son independientes exponenciales de parámetro λ.

Caso uniforme$[0, b]$. *La densidad conjunta se obtiene de (48) para $a, z > 0$ y $a + z < b$ tenemos*

$$f_{A,Z}(a, z) = \frac{1/b}{b/2} = \frac{2}{b^2}.$$

Las marginales se obtienen de igualar z a cero en (47):

$$\frac{(b - x)/b}{b/2} = \frac{2}{b}\Big(1 - \frac{x}{b}\Big)$$

Paradoja de la inspección *Sea $L(t) = A(t) + Z(t)$ el tiempo de vida de la lámpara en uso en el instante t. Usando (49) tenemos que el límite $L = \lim_t L(t)$ tiene esperanza*

$$EL = \frac{Et_1^2}{Et_1} > Et_1.$$

porque $Et_1^2 \geq (Et_1)^2$. Esto es una paradoja porque el promedio de los tiempos de vida converge a Et_1:

$$\frac{t_1 + \cdots + t_n}{n} \to Et_1.$$

Por eso

$$\frac{t_1 + \cdots + t_{N(t)}}{N(t)} \to Et_1.$$

Pero al hacer el promedio del tiempo de vida de la lámpara en el instante t, contamos t_i durante todo el intervalo $[T_i, T_i + t_i]$, es decir que cada t_i contribuye a la integral con t_i^2:

$$\frac{1}{t}\int_0^t L(t)dt \approx \frac{N(t)}{t}\frac{1}{N(t)}\sum_{i=1}^{N(t)} t_i\, t_i \to \frac{Et_i^2}{Et_i}.$$

Cadenas de Markov a tiempo continuo

Vamos a considerar procesos de Markov con tiempo $t \in \mathbb{R}^+$. En lugar de condicionar a todo el pasado, condicionamos a un conjunto finito de tiempos en el pasado. Decimos que un proceso X_t es de Markov si

$$P(X_{t+s} = y \,|\, X_s = x,\, x_{s_n} = i_n, \ldots, x_{s_0} = i_0) = P(X_t = y \,|\, X_0 = x)$$

para todo $n \geq 0$, estados $i_0, \ldots, i_n$ y tiempos $s_0 < \ldots < s_n$.

Ejemplo *Sea Y_n una cadena de Markov con matriz de transición $u(i, j)$. Sea $N(t)$ un proceso de Poisson de tasa λ. Sea*

$$X_t = Y_{N(t)}$$

Es decir que X_t hace los mismos saltos que Y_n pero en los instantes de las llegadas de un proceso de Poisson. La Markovianidad de ese proceso sigue de la falta de memoria de la exponencial y de la propiedad de Markov de Y_n.

Definimos las probabilidades de transición

$$p_t(x,y) := P(X_t = y|X_0 = x)$$

En el ejemplo Poisson tenemos

$$\begin{aligned} p_t(x,y) &= \sum_n P(N(t)=n)P(Y_n = y|Y_0 = x) \\ &= \sum_n \frac{e^{-\lambda t}(\lambda t)^n}{n!} u^n(x,y) \end{aligned}$$

Teorema 81 (Chapman Kolmogorov)

$$p_{t+s}(x,y) = \sum_z p_t(x,z)p_s(z,y)$$

Demostración

$$\begin{aligned} P(X_t = y|X_0 = x) &= \sum_z P(X_t = y, X_s = z|X_0 = x) \\ &= \sum_z P(X_t = y|X_s = z, X_0 = x)P(X_s = z|X_0 = x). \\ &= \sum_z P(X_t = y|X_s = z)P(X_s = z|X_0 = x). \square \end{aligned}$$

Tasas de salto *Definimos*

$$q(x,y) = \lim_{h\to 0} \frac{p_h(x,y)}{h}$$

si el límite existe, llamamos tasas de salto *de x a y.*

En el caso del Poisson:

$$P(N(h)\geq 2) = 1 - P(N(h)=0) - P(N(h)=1) = \lambda o(h) + o(h)$$

Donde $o(h)/h \to_h 0$.

Es decir que en ese caso

$$\begin{aligned} q(x,y) &= \lim_{h\to 0}\Big(\frac{P(N(h)=1)u(x,y)}{h} + \frac{o(h)}{h}\Big) \\ &= \lim_{h\to 0} \lambda e^{-\lambda h}u(x,y) \ = \ \lambda u(x,y), \end{aligned}$$

que se puede interpretar: "saltamos de x a tasa λ y una vez que saltamos, vamos a y con proba $u(x,y)$."

Ejemplos Proceso de Poisson. *$N(t)$ tiene tasas $q(x,x+1) = \lambda$, y $q(x,z) = 0$ si $z \neq x+1$.*

Fila $M|M|s$. *Llegadas Poisson(λ), servicios Exponencial(μ), s servidores*

$$\begin{aligned} q(x,x+1) &= \lambda, \quad x \geq 0 \\ q(x,x-1) &= \min\{s,x\}\mu, \quad x \geq 1 \end{aligned}$$

Proceso de ramificación. *Cada individuo muere a tasa μ y crea un nuevo individuo a tasa λ.*

$$\begin{aligned} q(x,x+1) &= \lambda x, \quad x \geq 0 \\ q(x,x-1) &= \mu x, \quad x \geq 1 \end{aligned}$$

Cuando $\mu = 0$ se llama Yule. *Nacimiento puro.*

Construcción *Sea* $\lambda_x = \sum_y q(x,y)$. *Es la tasa de salto a partir de* x. *Informalmente podemos pensar que cuando el proceso está en* x, *espera un tiempo exponencial de parámetro* λ_x *y cuando decide saltar, lo hace a* y *con probabilidad* $q(x,y)/\lambda_x$.

Estirando exponenciales. *Sea*

$$r(x,y) = \frac{q(x,y)}{\lambda_x}$$

Matriz de ruteo. *Sea* Y_n *la cadena de Markov con probabilidades de transición* $r(x,y)$. *Es el* esqueleto *discreto de* X_t.

Sean $\tau_1, \tau_2, \ldots$ *iid Exponencial*(1).

Sea $t_1 = \tau_0/\lambda(Y_0)$ *y en general*

$$t_n = \frac{\tau_{n-1}}{\lambda_{Y_{n-1}}}$$

O sea que, dado que $Y_{n-1} = y$, t_n *es una variable exponencial de parámetro* λ_y.

Definimos

$$T_n = t_1 + \cdots + t_n$$

es el instante en que la cadena salta a la posición Y_n.

En símbolos, si ponemos $T_0 = 0$ *tenemos*

$$X(t) = Y_n \qquad \text{para } T_n \le t < T_{n+1}.$$

Así, construimos $X(t)$ *para todo* $0 \le t \le T_n$. *Si* $T_n \to \infty$, *pudimos construir la cadena para* $t \in [0, \sup_n T_n)$.

Proposición 82 *El proceso* $X(t)$ *así construido es Markov con tasas* q.

Demostración *Ejercicio.*

Construcción usando procesos de Markov bi-dimensionales *Dadas las tasas* $q(x,y)$ *queremos construir un proceso* X_t *que satisfaga*

$$P(X_{t+h} = y \mid X_t = x) = hq(x,y) + o(h). \tag{51}$$

Recordemos

$$\lambda_x = \sum_y q(x,y),$$

es la tasa de salida *de* x.

Consideremos un proceso de Poisson bi-dimensional $M(\cdot)$. *For cada estado* x *consideramos una partición del intervalo* $I_x = [0, \lambda_x]$ *en Borelianos* $B(x,y)$ *de longitud* $q(x,y)$.

Fijemos $X_0 = x_0$, *un estado arbitrario. Sea* T_1 *el primer evento del proceso* $M(\cdot)$ *en la banda* $[0,\infty) \times I_{x_0}$:

$$T_1 = \text{ínf}\{t > 0 : M([0,t] \times I_{x_0}) > 0\}.$$

Defina $x_1 = y$, *el único estado que satisface*

$$\begin{aligned} &\text{ínf}\{t > 0 : M([0,t] \times I_{x_0}) > 0\} \\ &= \text{ínf}\{t > 0 : M([0,t] \times B(x_0,y)) > 0\} \end{aligned} \tag{52}$$

es decir, x_1 *está determinado or el intervalo* $B(x_0,x_1)$ *que realiza el ínfimo.*

Continuamos iterativamente, asumimos que T_{n-1} *y* x_{n-1} *están determinados y definimos*

$$T_n = \text{ínf}\{t > T_{n-1} : M((I_{x_{n-1}} \times T_{n-1}, t] > 0\}.$$

y $x_n = y$ si y sólo si

$$\begin{aligned} &\inf\{t > T_{n-1} : M((T_{n-1}, t] \times I_{x_{n-1}}) > 0\} && (53) \\ &= \inf\{t > 0 : M((T_{n-1}, t] \times B(x_{n-1}, y) > 0)\}. && (54) \end{aligned}$$

Definición 83 Defina $T_\infty = \sup_n T_n$ y

$$X_t = x_n, \text{ si } t \in [T_n, T_{n+1}), \quad \text{para } t \in [0, \infty) \tag{55}$$

Así, para cada realización del proceso de Poisson bidimensional M, construimos una realización del proceso $(X_t : t \in [0, T_\infty))$. T_n es el n-ésimo instante de salto y x_n es el n-ésimo estado visitado por el proceso.

Proposición 84 *El proceso $(X_t : t \in [0, T_\infty))$ definido en (53) satisface (51).*

Demostración *Por definición,*

$$\begin{aligned} &P(X_{t+h} = y \mid X_t = x) && (56) \\ &= P\big\{M((t, t+h] \times B(x, y)) = 1\big\} + P(\textit{otras cosas}), && (57) \end{aligned}$$

donde el evento {otras cosas} está contenido en el evento

$$\{M((t, t+h] \times [0, \lambda_x]) \geq 2\},$$

el proceso de Poisson M contiene dos o más puntos en el rectángulo $[0, \lambda_x] \times (t, t+h]$. Por definición de $M(\cdot)$, tenemos

$$\begin{aligned} P(M((t, t+h] \times [0, \lambda_x]) \geq 2) &= o(h) \quad \textit{and} && (58) \\ P(M((t, t+h] \times B(x, y) = 1) &= hq(x, y) + o(h). && (59) \end{aligned}$$

Esto demuestra la proposición. □

El proceso es Markov:

Proposición 85 *El proceso (X_t) $(X_t : t \in [0, T_\infty))$ definido en (53) satisface*

$$P(X_{t+u} = y \mid X_t = x, X_{s_n} = x_n, \ldots, X_{s_0} = x_0) = P(X_{t+u} = y \mid X_t = x).$$

para todo $n \geq$ y $s_0 < \ldots < s_n < t$, $x_0, \ldots, x_n$.

Demostración *El proceso $(X_s : s \geq t)$ después de t depende solamente del proceso M en la región posterior a t y del valor de X_t at time t, una función de $M((0, t) \times \mathbb{R}^+)$. Por eso, dado $X_t = x$, X_{t+u} no depende del evento $\{X_{s_n} = x_n, \ldots, X_{s_0} = x_0\}$.* □

Ejemplo 86 Ejemplo fila con dos servidores y espacio limitado de espera. Espacio de estados $\mathcal{X} = \{0, 1, 2\}$. X_t es el número de clientes en el sistema en el instante t. Los clientes llegan a tasa λ y los servicios son exponenciales a tasa μ. Las tasas son:

$$\begin{aligned} q(0, 1) = q(1, 2) &= \lambda && (60) \\ q(1, 0) = \mu; \quad q(2, 1) &= 2\mu && (61) \\ q(x, y) = 0, &\text{ en los otros casos.} && (62) \end{aligned}$$

Los intervalos usados en la construcción son los siguientes:

$$\begin{aligned} B(0, 1) = B(1, 2) &= [0, \lambda] && (63) \\ B(1, 0) = [\lambda, \lambda + \mu]; &\ B(2, 1) = [0, 2\mu]. && (64) \end{aligned}$$

Todas las tasas están acotadas por $\max\{\lambda + \mu, 2\mu\}$.

Nacimiento puro X_t *en* $\{0,1,2,\dots\}$ *con tasas*

$$q(x,x+1) = \lambda x \tag{65}$$
$$q(x,y) = 0, \text{ en los otros casos.} \tag{66}$$

La tasa de llegadas en el instante t es proporcional al número de llegadas hasta ese instante. Los intervalos son

$$B(x,x+1) = [0,\lambda x) \tag{67}$$

Explosiones *Si el espacio de estados es finito, entonces* $T_\infty = \infty$. *Es decir, en cada intervalo acotado hay un número finito de saltos. Cuando el espacio es infinito, la situación es diferente.*

Considere nacimiento puro con tasas

$$q(x,y) = \begin{cases} 2^x, & \text{si } y = x+1 \\ 0, & \text{si no} \end{cases} \tag{68}$$

La construcción usando el proceso de Poisson es la siguiente. Defina $T_0 = 0$ *y recursivamente*

$$T_n = \inf\{t > T_{n-1} : M(([0,2^n] \times T_{n-1}, t]) = 1\} \tag{69}$$

y defina

$$X_t \ := \ n, \quad \text{para } t \in [T_n, T_{n+1}]. \tag{70}$$

Si $x_0 = 0$, *el* n*-ésimo salto ocurre en el instante*

$$T_n = \sum_{i=0}^{n} \tau_i,$$

donde los τ_i *son independientes, exponenciales con media* $E\tau_i = 2^{-i}$, *para* $i \geq 0$. *Así,*

$$ET_n \ = \ \sum_{i=0}^{n} 2^{-i} \ \leq \ 2, \quad \text{for all } n. \tag{71}$$

Define $T_\infty = \sup_n T_n$. *We prove now that* T_∞ *is a finite random variable. Since* T_n *is an increasing sequence,*

$$P(T_\infty > t) \ = \ P(\cup_n \{T_n > t\}) \ = \ \lim_n P(T_n > t) \leq \lim_n \frac{ET_n}{t} \ \leq \ \frac{2}{t} \tag{72}$$

because $\{T_n > t\}$ *is an increasing sequence of events and then Markov inequality. We have proved that* $P(T_\infty > t)$ *goes to zero as* $t \to \infty$, *which implies that* T_∞ *is a finite random variable.*

Hence, for all $t > T_\infty$, *the process performs infinitely many jumps before time* t.

Definición 87 We say that the process X_t explodes if

$$P(\lim_{n\to\infty} T_n < \infty) > 0.$$

After a finite random time T_∞ the process is not formally defined. But we can define an explosive process by adding a new state called ∞ with transition rates $q(\infty, x) = 0$ for all $x \in E$.

If there are no explosions, that is, if

$$P(\lim_{n\to\infty} T_n < \infty) = 0,$$

then, the rates $q(x,y)$ *define univoquely a process which can be constructed as in Proposition 84.*

Properties *Given that at time T_n the process X_t is at state x, the time elapsed up to the next jump is an exponentially distributed random variable with mean $1/\lambda_x$; when the process decides to jump, it does so to state y with probability*

$$p(x,y) = \frac{q(x,y)}{\lambda_x}. \tag{73}$$

These properties are proven in the next two theorems.

Teorema 88 *For a continuous time Markov process X_t,*

$$P(T_{n+1} - T_n > t \mid X_{T_n} = x) = e^{-t\lambda_x}. \tag{74}$$

Demostración *We prove the theorem for the case when the rates λ_x are uniformly bounded by $\lambda \in (0,\infty)$. The general case can be proven using finite approximations. We use the representation of $M(\cdot)$ in the strip $[0,\infty] \times [0,\lambda]$ of Theorem 71.*

It is clear that the set $(T_n)_n$ is contained in the set $(S_n)_n$ defined in Theorem 71. Indeed, given $x_0 \in \mathcal{X}$, we can define

$$\begin{aligned} T_n &= \text{mín}\{S_k > T_{n-1} : W_k < \lambda_{x_{n-1}}\} \\ K_n &= \{k : S_k = T_n\} \\ x_n &= \{y : W_{K_n} \in I(x_{n-1}, y)\}. \end{aligned} \tag{75}$$

This definition is a consequence of the representation of the bi-dimensional Poisson process of Theorem (71) and the construction of the Markov process using the Poisson process summarized in Definition 55.

The distribution of $T_{n+1} - T_n$, conditioned to $X_{T_n} = x$ is given by

$$\begin{aligned} &P(T_{n+1} - T_n > t \mid X_{T_n} = x) \\ &= \frac{P(T_{n+1} - T_n > t, X_{T_n} = x)}{P(X_{T_n} = x)}. \end{aligned} \tag{76}$$

Conditioning on the possible values K_n may assume, the denominator can be written as

$$\sum_k P(T_{n+1} - T_n > t, X_{T_n} = x, K_n = k)$$

$$= \sum_k P(T_{n+1} - S_k > t, X_{S_k} = x, K_n = k)$$

$$= \sum_k \sum_\ell P(S_{k+\ell} - S_k > t, W_{k+1} > \lambda_x, \dots, W_{k+\ell-1} > \lambda_x,$$

$$W_{k+\ell} < \lambda_x)\, P(X_{S_k} =$$

$$= e^{-\lambda_x} \sum_k P(X_{S_k} = x, K_n = k) = e^{-\lambda_x},$$

which finishes the proof. □

Definición 89 The discrete process $(Y_n : n \in \mathbb{N})$ defined by $Y_n = X_{T_n}$ is called *skeleton* of the (continuous time) process $(X_t : t \in \mathbb{R}^+)$.

Teorema 90 *The skeleton (Y_n) of a continuous time process (X_t) is a Markov chain with transitions probability $p(x,y)$ given by* (73).

Demostración *We want to prove that*

$$P(X_{T_{n+1}} = y \mid X_{T_n} = x) = p(x,y). \tag{77}$$

We use again the construction (75). *Partitioning according with the possible values of K_n:*

$$P(X_{T_{n+1}} = y, X_{T_n} = x) \;=\; \sum_k P(X_{T_{n+1}} = y, X_{T_n} = x, K_n = k) \tag{78}$$

By construction, the event $\{X_{T_{n+1}} = y, X_{T_n} = x, K_n = k\}$ is just

$$\bigcup_{l\geq 1} \{W_{k+1} > \lambda_x, \dots, W_{k+\ell-1} > \lambda_x, W_{k+\ell} \in B(y|x), X_{S_k} = x, K_n = k\},$$

where we used the convention that for $l = 1$ the event $\{W_{k+1} > \lambda_x, \dots, W_{k+\ell-1} > \lambda_x, W_{k+\ell} \in B(y|x)\}$ is just $\{W_{k+1} \in B(y|x)\}$. By independence between (W_k) and (S_k), expression (78) equals

$$\sum_k \sum_{l\geq 1} P(W_{k+1} > \lambda_x, \dots, W_{k+\ell-1} > \lambda_x, W_{k+\ell} \in B(y|x)) \times P(X_{S_k} = x, K_n = k)$$

But,

$$\sum_{l\geq 1} P(W_{k+1} > \lambda_x, \dots, W_{k+\ell-1} > \lambda_x, W_{k+\ell} \in B(y|x)) = \frac{q(x,y)}{\lambda_x}. \tag{79}$$

Hence, (78) equals

$$\frac{q(x,y)}{\lambda_x} \sum_k P(X_{S_k} = x, K_n = k) \;=\; \frac{q(x,y)}{\lambda_x} P(X_{T_n} = x), \tag{80}$$

which implies (77). $\square$

Kolmogorov equations *It is useful to use the following matrix notation. Let Q be the matrix with entries*

$$q(x,y) \qquad \text{se } x \neq y \tag{81}$$

$$q(x,x) = -\lambda_x \;=\; -\sum_{y\neq x} q(x,y). \tag{82}$$

and P_t be the matrix with entries

$$p_t(x,y) \;=\; P(X_t = y \mid X_0 = x).$$

Con esta notación, las ecuaciones de Chapman-Kolmogorov dicen

$$P_{t+s} \;=\; P_t P_s. \tag{83}$$

for all $s, t \geq 0$.

Proposición 91 (Kolmogorov equations) *The following identities hold*

$$P'_t = QP_t \quad \textit{(Kolmogorov Backward equations)}$$

$$P'_t = P_t Q \quad \textit{(Kolmogorov Forward equations)}$$

for all $t \geq 0$, where P'_t is the matrix having as entries $p'_t(x,y)$ the derivatives of the entries of the matrix P_t.

Demostración Backward equations. *Using Chapman-Kolmogorov,*

$$\begin{aligned} p_{t+h}(x,y) - p_t(x,y) &= \sum_z p_h(x,z)p_t(z,y) - p_t(x,y) \\ &= (p_h(x,x) - 1)p_t(x,y)) + \sum_{z\neq x} p_h(x,z)p_t(z,y). \end{aligned}$$

Dividing by h and taking h to zero we obtain $p'_t(x,y)$ in the left hand side. To compute the right hand side, observe that

$$p_h(x,x) = 1 - \lambda_x h + o(h).$$

Hence

$$\lim_{h\to 0} \frac{p_h(x,x)-1}{h} = -\lambda_x = q(x,x).$$

Analogously, for $x \neq y$

$$p_h(x,y) = q(x,y)h + o(h)$$

and

$$\lim_{h\to 0} \frac{p_h(x,y)}{h} = q(x,y).$$

This shows the Kolmogorov Backward equations. The forward equations are proven analogously. To start, use Chapman-Komogorov to write

$$p_{t+h}(x,y) = \sum_z p_t(x,z)p_h(z,y). \ \square$$

Las ecuaciones backward dicen $P_t' = QP_t$*. Si* P_t *fuera un número, tendríamos*

$$P_t = e^{Qt}$$

Formalmente podemos definir la matriz

$$e^{Qt} = \sum_{n\geq 0} \frac{(Qt)^n}{n!} = \sum_{n\geq 0} Q^n \frac{t^n}{n!}$$

y diferenciar para obtener

$$\frac{d}{dt}e^{Qt} = \sum_{n\geq 1} Q^n \frac{t^{n-1}}{(n-1)!} = Q\sum_{n\geq 1} Q^{n-1}\frac{t^{n-1}}{(n-1)!} = QP_t$$

A pesar que en general el producto no es conmutativo, tenemos que $QP_t = P_tQ$*. Para verlo de otra manera,*

$$QP_t = \sum_{n\geq 0} QQ^{n-1}\frac{t^{n-1}}{(n-1)!} = \sum_{n\geq 0} Q^{n-1}\frac{t^{n-1}}{(n-1)!}Q = P_tQ.$$

Ejemplos Proceso de Poisson *Sabemos que*

$$p_t(x,y) = e^{-\lambda t}\frac{(\lambda t)^{y-x}}{(y-x)!}. \tag{84}$$

Si diferenciamos la expresión (84)*, obtenemos*

$$\begin{aligned} p_t'(x,y) &= -\lambda e^{-\lambda t}\frac{(\lambda t)^{y-x}}{(y-x)!} + \lambda e^{-\lambda t}\frac{(\lambda t)^{y-x-1}}{(y-x-1)!} \\ &= \lambda p_t(x+1,y) - \lambda p_t(x,y). \end{aligned}$$

que son justamente las ecuaciones backward.

Proceso con estados $\{1,2\}$ *Hay sólo dos tasas de salto* $q(1,2)=\lambda$, $q(2,1)=\mu$*. La matriz de tasas es*

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

Las ecuaciones para atrás en forma de matriz dicen $P_t' = QP_t$:

$$\begin{pmatrix} p_t'(1,1) & p_t'(1,2) \\ p_t'(2,1) & p_t'(2,2) \end{pmatrix} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}\begin{pmatrix} p_t(1,1) & p_t(1,2) \\ p_t(2,1) & p_t(2,2) \end{pmatrix}$$

Como $p_t(i,2) = 1 - p_t(i,1)$*, basta calcular* $p_t(i,1)$*. Tenemos*

$$p_t'(1,1) = -\lambda p_t(1,1) + \lambda p_t(2,1) = -\lambda(p_t(1,1) - p_t(2,1)) \tag{85}$$
$$p_t'(2,1) = -\mu p_t(2,1) + \mu p_t(1,1) = \mu(p_t(1,1) - p_t(2,1)) \tag{86}$$

restando las ecuaciones tenemos

$$[p_t(1,1) - p_t(2,1)]' = -(\lambda + \mu)(p_t(1,1) - p_t(2,1))$$

Como $p_0(1,1) = 1$ *y* $p_0(2,1) = 0$*, tenemos*

$$p_t(1,1) - p_t(2,1) = e^{-(\lambda+\mu)t}$$

Substituyendo en (85) *e integrando,*

$$\begin{aligned} p_t(1,1) &= p_0(1,1) + \frac{\lambda}{\mu+\lambda} e^{-(\lambda+\mu)s}\Big|_0^t = \frac{\lambda}{\mu+\lambda}(1 + e^{-(\lambda+\mu)s}) \\ p_t(2,1) &= p_0(2,1) + \frac{\lambda}{\mu+\lambda} e^{-(\lambda+\mu)s}\Big|_0^t = \frac{\mu}{\mu+\lambda}(1 - e^{-(\lambda+\mu)s}) \end{aligned}$$

Es decir que el proceso converge exponencialmente rápido al equilibrio.

Proceso de Yule *Cada partícula se divide en 2 a tasa* β*. Es decir*

$$q(x, x+1) = \beta x$$

Adivinamos que

$$p_t(1,x) = e^{-\beta t}(1 - e^{-\beta t})^{x-1}, \quad x \geq 1.$$

Es decir, una geométrica$(e^{-\beta t})$*. Queda como ejercicio probar que esa expresión satisface las ecuaciones para adelante*

$$p_t'(1,x) = -\beta x p_t(1,x) + \beta(x-1)p_t(1,x-1).$$

Recurrence and transience

Definición 92 Sea

$$T^{x,y} = \text{ínf}\{t > T_1 : X_t^x = y\}$$

be the first time the process starting at x hits y. The exigency $t > T_1$ is posed to avoid $T^{x,x} \equiv 0$.

Definición 93 We say that a state x is

$$\text{transient, if } P(T^{x,x} = \infty) > 0; \tag{87}$$
$$\text{null recurrent, if } P(T^{x,x} = \infty) = 0 \text{ and } ET^{x,x} = \infty; \tag{88}$$
$$\text{positive recurrent, if } ET^{x,x} < \infty. \tag{89}$$

If the state space is finite, there are no null recurrent states.

Definición 94 We say that a process is irreducible if for all states x, y, the probability to hit y starting from x in a finite amount of time is positive:

$$P(T^{x,y} < \infty) > 0.$$

Teorema 95 *A process* (X_t) *is irreducible if and only if its skeleton* (Y_n) *is irreducible.*

A state x *is recurrent (respectively transient) for the process* (X_t) *if and only if* x *is recurrent (respectively transient) for the skeleton* (Y_n).

If the exit rates are uniformly bounded from below and above, that is, if

$$0 < \text{ínf}_x \lambda_x; \ \sup_x \lambda_x < \infty,$$

then x *is null recurrent for* (X_t)*, if and only if* x *is null recurrent for the skeleton* (Y_n).

Demostración *The proof is straightforward.* $\square$

Remark. It is clear that in the finite state space case the hypotheses of Theorem (95) *are automatically satisfied.*

The first part of the theorem says that a state is recurrent (respectively transient) for the continuous time process if and only if it is recurrent (respectively transient) for its skeleton. Hence, recurrence and transience are equivalent for a process and its skeleton when the state space is finite.

When the state space is infinite it is possible to find examples of continuous time processes (violating the conditions of Theorem 95) and its skeleton with qualitative different behavior. We see now some of these examples.

Ejemplo 96 In this example we present a process having null recurrent states which are positive recurrent for the skeleton. Consider the rates $q(x,0) = 1/2^x$, $q(x,x+1) = 1/2^x$. Hence $p(x,0) = 1/2$, $p(x,x+1) = 1/2$ and the skeleton is positive recurrent, because the return time to the origin is given by a geometric random variable with parameter $1/2$. On the other hand, since the mean jump time of each state x is 2^x,

$$ET^{00} = \sum 2^x 1/2^{x+1} = \infty.$$

Ejemplo 97 A simple (however explosive) example for which the states are positive recurrent for the continuous time process but null recurrent for its skeleton is given by the following rates: $q(x,0) = 1$, $q(x,x+1) = x^2$, $q(x,y) = 0$ otherwise. The transition probabilities of the skeleton are given by $p(x,0) = 1/(1+x^2)$ and $p(x,x+1) = x^2/(1+x^2)$. The mean return time to the origin of the skeleton is given by

$$\sum_x x\left(\frac{1}{1+x^2}\right)\prod_{y=1}^{x-1}\left(\frac{y^2}{1+y^2}\right).$$

We let the reader the proof that this sum is infinity. The mean return time to the origin for the continuous process is given by

$$\sum_x \left(\sum_{y=1}^{x}\frac{1}{y^2}\right)\left(\frac{1}{1+x^2}\right)\left(\prod_{y=1}^{x-1}\left(\frac{y^2}{1+y^2}\right)\right).$$

We let the reader the proof that this sum is finite.

Invariant measures

Definición 98 We say that π is an *stationary distribution* for (X_t) if

$$\sum_x \pi(x)p_t(x,y) = \pi(y) \quad \text{Balance equations} \tag{90}$$

$$\sum_x \pi(x) = 1 \tag{91}$$

that is, if the distribution of the initial state is given by π, then the distribution of the process at time t is also given by π for any $t \geq 0$. We also use the term *stationary measure* to refer to a measure satisfying the balance equations but it is not a probability.

Teorema 99 *A distribution π is stationary for a process with rates $q(x,y)$ if and only if*

$$\sum_x \pi(x)q(x,y) = \pi(y)\sum_z q(y,z). \tag{92}$$

Condition (92) *can be interpreted as a* flux condition*: the entrance rate under π to state y is the same as the exit rate from y. For this reason the equations* (92) *are called* balance equations.

Demostración *Assume π stationary, then (90) read*

$$\pi P_t = \pi.$$

Differentiating we get

$$\begin{aligned} 0 &= \sum_x \pi(x) p_t'(x,y) = \sum_x \pi(x) \sum_z p_t(x,z) Q(z,y) \\ &= \sum_z \sum_x \pi(x) p_t(x,z) Q(z,y) = \sum_z \pi(z) Q(z,y) \end{aligned}$$

where we have applied the forward equations. This proves (92). Reciprocally, equations (92) say

$$\pi Q = 0.$$

Applying Kolmogorov backwards equations we get

$$(\pi P_t)' = \pi P_t' = \pi Q P_t = 0;$$

In other words, if the initial state of a process is chosen accordingly to the law π, the law of the process at any future time t is still π. This is because P_0 is the identity matrix and $\pi P_0 = \pi$. □

Proceso clima *Hay 3 estados: sol, nublado, lluvia. El tiempo que está en sol es exponencial $1/3$ de donde pasa a nublado, queda un tiempo exponencial $1/4$ cuando empieza a llover y llueve un tiempo exponencial 1, cuando vuelve a sol. La matriz de tasas es*

$$\begin{pmatrix} -1/3 & 1/3 & 0 \\ 0 & -1/4 & 1/4 \\ 1 & 0 & -1 \end{pmatrix}$$

$\pi Q = \pi$ equivale a las ecuaciones

$$\begin{array}{llll} -\frac{1}{3}\pi(1) & & +\pi(3) & = 0 \\ \frac{1}{3}\pi(1) & -\frac{1}{4}\pi(2) & & = 0 \\ & \frac{1}{4}\pi(2) & -\pi(3) & = 0 \end{array}$$

La solución es $\pi = (\frac{3}{8}, \frac{4}{8}, \frac{1}{8})$. Podríamos haber obtenido esto con procesos de renovación con recompensa. La fracción de tiempo que estamos en cada estado es el conciente entre el tiempo medio en el estado y el tiempo medio del ciclo.

Harris recurrencia y convergencia exponencial al equilibrio *En el siguiente teorema probamos simultaneamente la existencia y la convergencia a velocidad exponencial bajo una condición que es conocida como* recurrencia de Harris. *Defina*

$$\gamma(z) := \min_i Q(i,z), \qquad \gamma := \sum_z \gamma(z) \tag{93}$$

Teorema 100 *Sea X_t un proceso de Markov con tasas Q. Si $\gamma > 0$, entonces (X_t) tiene una única distribución estacionaria π. Además, el proceso converge a π en variación total, a velocidad exponencial con coeficiente γ:*

$$\sup_x \frac{1}{2} \sum_z |\pi(z) - P_t(x,z)| < e^{-\gamma t}.$$

Demostración *Primero vamos a demostrar que si hay una distribución estacionaria π, entonces vale la convergencia en variación total. Después demostraremos la existencia y unicidad de π.*

Sin pérdida de generalidad asumimos que el espacio de estados es $\{1, \dots, K\} \subset \mathbb{N}$, para algún $K > 0$.

Convergencia. *Construiremos dos cadenas con estados iniciales distintos con saltos governados por los puntos del mismo proceso de Poisson M y la misma partición $B(i,j)$, elegida de manera de controlar el tiempo de encuentro de las marginales.*

Construya una familia de intervalos sucesivos $J(z)$ disjuntos de tamaño

$$|J(z)| = \gamma(z).$$

Para eso, denote $J(z) = [a(z), b(z))$ e imponga $a(1) = 0$ y $a(z+1) = b(z)$. La longitud del intervalo $\dot\cup_z J(z) = [0, b(K))$ es igual a $b(K) = \gamma$.

Para cada estado i construya una partición del intervalo $[\gamma, \lambda_i + \gamma(i))$ con intervalos disjuntos sucesivos $I(i, z)$ de tamaño

$$|I(i, z)| = q(i, z) - \gamma(z);$$

para eso asuma $I(i, z) = [a(i, z), b(i, z))$ con $b(i, z) - a(i, z) = q(i, z) - \gamma(z)$ y $a(i, 1) = \gamma$, $b(i, z) = a(i, z+1)$. Tendremos $b(i, K) = \lambda_i + \gamma(i)$.

Defina

$$B(i, z) = J(z) \cup I(i, z)$$

Como $B(i, z)$ tiene medida $q(i, z)$, la construcción de X_t usando el proceso de Poisson bidimensional M se puede hacer con estas particiones.

Vamos a realizar (X_t, Y_t) con estado inicial $(X_0, Y_0) = (x, y)$ y cada marginal es governada por los puntos de M y la partición B recién construída.

Si el proceso en el instante $t-$ se encuentra en el estado (i, j) y M contiene al punto (t, u), con u en el intervalo

$$\cup_z J(z) = [0, \gamma),$$

entonces ambos procesos coalescen. Más precisamente, hay tres casos:

(a) $u \in J(z)$ para $z \notin \{i, j\}$, en ese caso ambas marginales saltan a z;

(b) $u \in J(i)$, en ese caso la segunda marginal salta a i y la primera marginal se queda en i;

(c) $u \in J(j)$, en ese caso la primera marginal salta a j y la segunda marginal se queda en j.

Esto se puede hacer porque por debajo de γ la partición $B(i, z)$ no depende de x:

$$B(i, j) \cap [0, \gamma) = B(i', j) \cap [0, \gamma), \quad \text{para todo } j.$$

El instante de coalescencia se denota τ:

$$\tau = \inf\{t > 0 : M([0, t] \times [0, \gamma)) > 0\}.$$

Por lo tanto,

$$t > \tau \quad \text{implica} \quad X_t = Y_t, \tag{94}$$

además

$$\tau \sim \text{ Exponencial}(\gamma) : \quad P(\tau > t) = e^{-\gamma t}. \tag{95}$$

Para concluir, escribimos

$$\begin{aligned}
\sum_z |P_t(y, z) - P_t(x, z)| &= \sum_z |P(X_t = z) - P(Y_t = z)| \\
&= \sum_z |E(\mathbf{1}\{X_t = z\} - \mathbf{1}\{Y_t = z\})| \\
&\le E\Big(\sum_z |\mathbf{1}\{X_t = z\} - \mathbf{1}\{Y_t = z\}|\Big) \\
&= 2E\mathbf{1}\{X_t \neq Y_t\} = 2P(X_t \neq Y_t) \\
&\le 2P(\tau > t) = 2e^{-\gamma t}, \text{ usando (94) y (95).}
\end{aligned}$$

Si el estado inicial de Y_t es aleatorio con distribución estacionaria π,

$$\begin{aligned}\sum_z |\pi(z) - P_t(x,z)| &= \sum_z \Big| \sum_y \pi(y)P_t(y,z) - \sum_y \pi(y)P_t(x,z)\Big| \\ &\leq \sum_y \pi(y) \sum_z |P_t(y,z) - P_t(x,z)| \\ &\leq \sum_y \pi(y) 2e^{-\gamma t} = 2e^{-\gamma t}\end{aligned}$$

Esto demuestra la convergencia a velocidad exponencial a la distribución estacionaria.

Existencia de la distribución estacionaria. Simulación perfecta *Denote $X^x_{[s,t]}$ el proceso que tiene una condición inicial $x_s = x$ en el instante s y utiliza los puntos de M en la banda $[s,t] \times [0,\infty)$. Esa construcción es invariante por traslaciones:*

$$(X^x_{[s,s+t]}, t \geq 0) \text{ tiene la misma distribución que } (X^x_{[0,t]}, t \geq 0)$$

Vamos a sacar el límite cuando s se va a menos infinito. Sea

$$\tau(t) := \sup\{s < t : M([s,t] \times [0,\infty)) > 0\}$$

En el instante $\tau(t)$, el proceso asume un valor aleatorio Z con distribución ρ dada por

$$\rho(x) = \frac{\gamma(x)}{\gamma}$$

independiente del pasado y desde ahí evoluciona de acuerdo a los puntos de M en $[\tau(t), t] \times \mathbb{R}^+$:

$$X^x_{[s,t]} = X_{[\tau(t),t]} = \sum_z \mathbf{1}\{Z = z\} X^z_{[\tau(t),t]}, \quad \text{para todo } x,\ s \leq \tau(t).$$

Si definimos

$$Z_t := X_{[\tau(t),t]}, \quad t \in \mathbb{R}$$

tenemos que Z_t es Markov con tasas Q y es estacionaria: $P(Z_t = z)$ no depende de t. Por lo tanto, llamando π a la distribución de Z_t, vale que π es invariante para Q. $\square$

Ejemplo. *Estados $\{1,2,3\}$.*

$$Q = \begin{pmatrix} -3 & 1 & 2 \\ 4 & -5 & 1 \\ 2 & 3 & -5 \end{pmatrix} \tag{96}$$

$$(\gamma(z)) = (2,1,1) \qquad \gamma = 4.$$

$$(J(z)) = ([0,2),[2,3),[3,4))$$

$$(Q(i,z) - \gamma(z)) = \begin{pmatrix} * & 0 & 1 \\ 2 & * & 0 \\ 0 & 2 & * \end{pmatrix}$$

$$I(i,j) = \begin{pmatrix} * & \emptyset & [4,5) \\ [4,6) & * & \emptyset \\ \emptyset & [4,6) & * \end{pmatrix}$$

La existencia de una única distribución estacionaria y convergencia bajo la hipótesis de recurrencia positiva se demuestra como en el caso discreto. La diferencia es que no existe el problema de la periodicidad.

Teorema 101 *Si el proceso Markoviano de salto (X_t) es irreducible y tiene una distribución estacionaria π, entonces*

$$\lim_{t\to\infty} p_t(x,y) = \pi(y).$$

Demostración *Como $p_t(i,j) > 0$, tenemos que para todo $h > 0$ la cadena discreta con matriz de transición $p_h(x,y)$ es recurrente positiva y aperiódica y tiene distribución estacionaria π. Por lo tanto*

$$\lim_n p_{nh}(i,j) = \pi(j).\square$$

Skeletons *Assume the process (X_t) and its discrete skeleton Y_n defined by $Y_n = X_{T_n}$ are irreducible and positive recurrent. An invariant measure ν for Y_n must satisfy the balance equations*

$$\sum_x \nu(x)p(x,y) = \nu(y)$$

The invariant measure π for X_t must satisfy

$$\sum_x \pi(x)q(x,y) = \pi(y)\lambda_y$$

This implies that

(a) If ν is the invariant measure for Y_n and $\sum_z \frac{\nu(z)}{\lambda_z} < \infty$, then the measure π defined by

$$\pi(x) = \frac{\nu(x)}{\lambda_x}\Big(\sum_z \frac{\nu(z)}{\lambda_z}\Big)^{-1} \tag{97}$$

is invariant for X_t.

(b) If π is invariant for X_t and $\sum_z \nu(z)\lambda_z < \infty$, then

$$\nu(x) = \pi(x)\lambda_x\Big(\sum_z \pi(z)\lambda_z\Big)^{-1} \tag{98}$$

is invariant for Y_n.

As a corollary we have that if the exit rates do not depend on the state, then a measure is invariant for the continuous time process if and only if it is invariant for the discrete time one. That is, if $\lambda_x = \lambda_y$ for all $x, y \in \mathcal{X}$, then $\pi(x) = \nu(x)$ for all $x \in \mathcal{X}$.

Birth and death process *A birth and death process represents the growth (or extinction) of a population. The value X_t represents the number of alive individuals of the population at time t. The rates of birth and death depend only on the number of alive individuals. That is,*

$$q(x,x+1) \;=\; \lambda_x \quad\text{and}\quad q(x,x-1) = \mu_x, \tag{99}$$

where λ_x, μ_x are families of non-negative parameters. We use the balance equations (92) to look for conditions under which the process admits an invariant measure. We look for a vector π satisfying the equations

$$\pi(0)q(0,1) \;=\; \pi(1)q(1,0) \tag{100}$$

$$\pi(x)\left(q(x,x-1)+q(x,x+1)\right) \;=\; \pi(x-1)q(x-1,x)+\pi(x+1)q(x+1,x), \tag{101}$$

for $x \geq 1$. From where,

$$\pi(x+1)\mu_{x+1} - \pi(x)\lambda_x \;=\; 0, \tag{102}$$

and

$$\pi(x+1) \;=\; \frac{\lambda_x}{\mu_{x+1}}\pi(x), \quad x \geq 0.$$

Hence, for all $x \geq 1$

$$\pi(x) \;=\; \frac{\lambda_0 \dots \lambda_{x-1}}{\mu_1 \dots \mu_x}\pi(0). \tag{103}$$

It is clear that $\pi(x)$ *so constructed satisfies* (92). *To satisfy* (91) *we need* $\pi(0) > 0$. *Hence, there will be a solution if*

$$\sum_{x \geq 1} \frac{\lambda_0 \dots \lambda_{x-1}}{\mu_1 \dots \mu_x} < \infty. \tag{104}$$

If (104) *is satisfied, then we can define*

$$\pi(0) \;=\; \left(\frac{\lambda_0 \dots \lambda_{x-1}}{\mu_1 \dots \mu_x}\right),$$

and $\pi(x)$ *inductively by* (103).

Fila $M|M|\infty$

$$q(x, x+1) = \lambda, \qquad q(x+1, x) = \mu(x+1)$$

En este caso $\pi(x) = e^{-\lambda/\mu}(\lambda/\mu)^x/x!$.

Fila $M|M|1$

$$q(x, x+1) = \lambda, \qquad q(x+1, x) = \mu$$

En este caso $\pi(x) = (\lambda/\mu)^x(1 - \lambda/\mu)$.

Martingalas

Pensemos que M_n *es la fortuna de un jugador después del* n*-ésimo juego y* X_n *el resultado del* n*-ésimo juego.*

Decimos que un proceso $(M_0, M_1, \dots)$ *es* adaptado *a un proceso* $(X_0, X_1, \dots)$ *si* M_n *es función de* $X_n, \dots, X_0, M_0$. *Un proceso adaptado* M_n *es una* martingala *respecto a* $(X_0, X_1, \dots)$ *si para todo* $n \geq 0$ *tenemos* $E|M_n| < \infty$ *y*

$$E(M_{n+1} - M_n | X_n = x_n, \dots, X_0 = x_0, M_0 = m_0) = 0$$

Bajo ese condicionamiento, M_n *es una función de* $x_n, \dots, x_0, m_0$.

La condición $E|M_n| < \infty$ *es para garantizar que existe la esperanza condicional.*

La segunda condicion: dado el pasado hasta n*, el lucro medio a obtener en el* $(n+1)$ *juego es nulo.*

Escribimos

$$A_\nu = \{X_n = x_n, \dots, X_0 = x_0, M_0 = m_0\}, \qquad \nu = (x_n, \dots, x_0, m_0) \tag{105}$$

Paseos aleatorios X_i *iid,* $EX_i = \mu$. $S_n = S_0 + X_1 + \dots + X_n$. $M_n = S_n - \mu n$ *es una martingala porque*

$$E(M_{n+1} - M_n | A_\nu) = EX_{n+1} - \mu = 0$$

Lema 102 *Si* M_n *es una martingala, entonces* $E(M_{n+k} - M_n | A_\nu) = 0$ *para todo* $k \geq 0$.

Demostración *Inducción. Supongamos que vale para* k. *Veamos para* $k+1$:

$$\begin{aligned} E(M_{n+k+1} - M_n | A_\nu) &= E(M_{n+k+1} - M_{n+k} | A_\nu) + E(M_{n+k} - M_n | A_\nu) \\ &= \sum_{\tilde{\nu}} E(M_{n+k+1} - M_{n+k} | A_{\tilde{\nu}} A_\nu) P(A_{\tilde{\nu}} | A_\nu) \end{aligned}$$

por la hipótesis inductiva y probabilidad total, donde

$$A_{\tilde{\nu}} = \{X_{n+k} = x_{n+k}, \dots, X_{n+1} = x_{n+1}\}, \qquad \tilde{\nu} = (x_{n+k}, \dots, x_{n+1})$$

Concluimos porque la esperanza condicional a un paso es cero por la definición de martingala. $\square$

Decimos que M_n *es una* supermartingala *si*

$$E(M_{n+1} - M_n|A_\nu) \leq 0$$

Decimos que M_n *es una* submartingala *si*

$$E(M_{n+1} - M_n|A_\nu) \geq 0$$

En el paseo aleatorio, si $\mu \leq 0$ *tenemos* S_n *es supermartingala y si* $\mu \geq 0$, S_n *es submartingala.*

Teorema 103 *Sea* X_n *cadena de Markov y* $f(x, n)$ *una función tal que*

$$f(x, n) = \sum_y p(x, y) f(y, n + 1)$$

Entonces, $M_n = f(X_n, n)$ *es una martingala en relación a* X_n*. En particular, si* $h(x) = \sum_y p(x, y)h(y)$*, entonces* $h(X_n)$ *es una martingala.*

Demostración

$$E(f(X_{n+1}, n + 1)|A_\nu) = \sum_y p(x_n, y) f(y, n + 1) = f(x_n, n). \qquad \square$$

Ruina del jugador X_i *paseo aleatorio* $P(X_i = 1) = p$, $P(X_i = -1) = 1 - p$, $p \notin \{0, \frac{1}{2}, 1\}$. $S_n = S_0 + X_1 + \cdots + X_n$.

$$M_n = \Big(\frac{1-p}{p}\Big)^{S_n}$$

es una martingala con respecto a X_n.

Usando el teorema con $h(x) = ((1 - p)/p)^x$ *hay que ver que es armónica:* $h = Ph$*:*

$$\begin{aligned}\sum_y p(x, y)h(y) &= p\Big(\frac{1-p}{p}\Big)^{x+1} + (1-p)\Big(\frac{1-p}{p}\Big)^{x-1} \\ &= (1-p)\Big(\frac{1-p}{p}\Big)^{x} + p\Big(\frac{1-p}{p}\Big)^{x} = \Big(\frac{1-p}{p}\Big)^{x}\end{aligned}$$

Paseo aleatorio simétrico Y_i *iid con*

$$P(Y_i = 1) = P(Y_i = -1) = 1/2$$

$$X_n = X_0 + Y_1 + \cdots + Y_n.$$

$M_n = X_n^2 - n$ *es una martingala con relación a* X_n.

Para verlo usamos el teorema con $f(x, n) = x^2 - n$ *y*

$$\begin{aligned}E(f(X_{n+1}, n + 1)|A_\nu) &= \frac{1}{2}(x_n + 1)^2 + \frac{1}{2}(x_n - 1)^2 - (n + 1) \\ &= x_n^2 - n.\end{aligned}$$

Productos de variables independientes $X_i \geq 0$ *independientes con* $EX_i = 1$*. Entonces*

$$M_n = M_0 X_1 \dots X_n$$

es una martingala en relación a X_n. Para verlo

$$E(M_{n+1} - M_n|A_\nu) = M_n(\nu)E(X_{n+1} - 1|A_\nu) = 0$$

Martingala exponencial Y_i *iid con* $\phi(\theta) = E\exp(\theta Y_1) < \infty$. $S_n = S_0 + Y_1 + \cdots + Y_n$. *Entonces*

$$M_n = \frac{\exp(\theta S_n)}{(\phi(\theta))^n}$$

es una martingala en relación a Y_n. En particular, si $\phi(\theta) = 1$, $\phi(\theta S_n)$ es una martingala.

Para verlo, escribimos $X_i = \frac{\exp(\theta Y_i)}{\phi(\theta)}$ y tenemos $M_n = M_0 X_1 \dots X_n$ y nos reducimos al ejemplo anterior.

Propiedades

Desigualdad de Jensen. *Si ϕ es convexa, entonces $E\phi(X) \geq \phi(EX)$, $E(\phi(X)|A) \geq \phi(E(X|A))$.*

Demostración. *Si X toma los valores a y b con probabilidades p y $1-p$, tenemos que $EX = ap+b(1-p)$, $\phi(EX) = \phi(ap+b(1-p)) \leq \phi(a)p+\phi(b)(1-p) = E\phi(X)$. La desigualdad es por la definición de función convexa. La extensión al caso general discreto queda como ejercicio.*

Lema 104 *M_n martingala ϕ convexa, entonces $\phi(M_n)$ submartingala. Si M_n es submartingala y ϕ es convexa no decreciente, entonces $\phi(M_n)$ es submartingala.*

Demostración *Para la primera parte, por Jensen*

$$E(\phi(M_{n+1})|A_\nu) \geq \phi(E(M_{n+1}|A_\nu)) = \phi(M_n(\nu))$$

El segundo enunciado se demuestra igual con $\geq$ en lugar de $=$. $\square$

Como x^2 es convexa, M_n martingala implica que M_n^2 es submartingala.

Lema 105 *Si M_n es martingala, entonces*

$$E(M_{n+1}^2|A_\nu) - (M_n(\nu))^2 = E((M_{n+1} - M_n)^2|A_\nu)$$

Demostración

$$\begin{aligned} E((M_{n+1} - M_n)^2|A_\nu) &= E(M_{n+1}^2|A_\nu) - 2E(M_{n+1}M_n|A_\nu) + E(M_n^2|A_\nu) \\ &= E(M_{n+1}^2|A_\nu) - E(M_n^2|A_\nu) \end{aligned}$$

porque $E(M_{n+1}M_n|A_\nu) = M_n(\nu)E(M_{n+1}|A_\nu) = (M_n(\nu))^2$. $\square$

Lema 106 (incrementos ortogonales) *Si M_n es una martingala y $0 \leq i \leq j \leq k < n$, entonces*

$$E[(M_n - M_k)M_j] = 0, \qquad y \qquad E[(M_n - M_k)(M_j - M_i)] = 0$$

Observe que si definimos $E(XY)$ como producto interno de las variables aleatorias X, Y con segundo momento finito, el lema dice que los incrementos son ortogonales en ese espacio.

Demostración *El segundo resultado sigue del primero restando. Para demostrar el primero, denote*

$$A_\nu = \{X_k = x_k, \dots, X_0 = x_0, M_0 = m_0\}; \qquad \nu = (x_k, \dots, x_0, m_0). \tag{106}$$

Escriba

$$\begin{aligned} E[(M_n - M_k)M_j] &= \sum_\nu E[(M_n - M_k)M_j|A_\nu]P(A_\nu) \\ &= \sum_\nu M_j(\nu)E[M_n - M_k|A_\nu]P(A_\nu) = 0 \end{aligned}$$

La primera igualdad es probabilidad total. La segunda es porque M_j es una función de ν en el evento A_ν, por lo tanto es constante y puede salir fuera de la esperanza. La tercera es por la propiedad de martingala. $\square$

Corolario 107

$$E(M_n - M_0)^2 = \sum_{k=1}^{n} E(M_k - M_{k-1})^2$$

Demostración

$$\begin{aligned} E(M_n - M_0)^2 &= E\Big(\sum_{k=1}^{n}(M_k - M_{k-1})\Big)^2 \\ &= \sum_{k=1}^{n} E(M_k - M_{k-1})^2 + 2 \sum_{1\le j<k\le n} E[(M_k - M_{k-1})(M_j - M_{j-1})] \end{aligned}$$

pero el segundo término es nulo por el lema anterior. $\square$

Estrategias y tiempos de parada *En un juego desfavorable nuestra fortuna va a decrecer.*

Teorema 108 *Si M_n es una supermartingala y $m \le n$ entonces $EM_m \ge EM_n$.*

Demostración *Basta probar que $EM_k \ge EM_{k+1}$. Usando la notación* (106)*,*

$$E[M_{k+1} - M_k] = \sum_x E[M_{k+1} - M_k | A_\nu] P(A_\nu)$$

porque M_n es supermartingala. $\square$

Multiplicando por -1 obtenemos el siguiente teorema:

Teorema 109 *Si M_n es una submartingala y $m \le n$ entonces $EM_m \le EM_n$.*

Como una martingala es sub y super martingala, también obtenemos.

Teorema 110 *Si M_n es una martingala y $m \le n$ entonces $EM_m = EM_n$.*

No se puede ganar en un juego desfavorable

Estrategia duplicar *Suponga que jugamos al rojo en una ruleta. Empezamos apostando \$1. Si perdemos, apostamos \$2. Si ganamos, volvemos a apostar \$1. Así, cada vez que ganamos, nuestra fortuna aumenta \$1.*

Pero si vemos nuestra ganancia en un instante dado. Por ejemplo después de jugar 4 veces, se puede ver que nuestra ganancia media es 0.

Procesos predecibles *Vamos a definir H_n la apuesta en el instante n. H_n es* predecible *si se puede determinar sabiendo $X_{n-1}, \dots, X_0, M_0$.*

Pensemos por un momento que H_m es el stock que tengo entre los instantes $m-1$ y m y M_n es el precio por unidad de stock en el instante n. Nuestra fortuna en el instante n es

$$W_n = W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1})$$

porque el cambio de fortuna entre los instantes $m-1$ y m es el stock por el precio del stock: $H_m(M_m - M_{m-1})$

Para formular la estrategia duplicar en este contexto, sea $X_m = 1$ si la m-ésima moneda es cara y -1 si es ceca. Sea $M_n = X_1 + \cdots + X_n$ el lucro neto de un jugador que apuesta \$1 cada jugada. H_n es la estrategia duplicar.

Teorema 111 *Sea M_n una supermartingala en relación a X_n, H_n predecible y $0 \leq H_n \leq c_n$, una constante que puede depender de n. Entonces*

$$W_n := W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1})$$

es una supermartingala.

Pedimos $H_n \geq 0$ porque sólo se aceptan apuestas positivas. La condición $\leq c_n$ es para garantizar que las esperanzas sean finitas.

Demostración *El incremento de fortuna entre n y $n+1$ es*

$$W_{n+1} - W_n = H_{n+1}(M_{n+1} - M_n)$$

Usamos la notación (106) con $k = n$ y observamos que H_{n+1} es constante en A_ν. Entonces

$$E(H_{n+1}(M_{n+1} - M_n)|A_\nu) = H_{n+1}E(M_{n+1} - M_n|A_\nu) \leq 0.$$

que demuestra que W_n es supermartingala. $\square$

El mismo resultado vale para submartingalas. Para martingalas también vale, pero no es necesario suponer que $H_n \geq 0$, basta $E|H_n| \leq c_n$.

Recordemos la definición de tiempo de parada. T es un tiempo de parada para X_n si $\{T = n\} = \{(X_n, \dots, X_0, M_0) \in B\}$ para algún Boreliano. Es decir $T = n$ puede ser determinado sabiendo los resultados de los primeros n juegos y la fortuna inicial.

Teorema 112 *Sea M_n una supermartingala y T es un tiempo de parada en relación a X_n. Entonces el proceso $M_{T\wedge n}$ es una supermartingala. En particular, $EM_{T\wedge n} \leq M_0$.*

Demostración *Considere la estrategia de apostar \$1 hasta un tiempo de parada T. Así tenemos $H_m = \mathbf{1}\{T \geq m\}$. Para ver que H_m es predecible,*

$$\{H_m = 0\} = \{T \geq m\}^c = \{T \leq m-1\} = \cup_{k=1}^{m-1}\{T = k\}.$$

Como $\{T = k\}$ puede sed determinado por $M_0, X_0, \dots, X_k$, H_m puede ser determinado por $M_0, X_0, \dots, X_{m-1}$, lo que demuestra que H_m es predecible. Si nuestra fortuna inicial es $W_0 = M_0$, nuestra fortuna en el instante n es

$$W_n = M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}) = M_{T\wedge n}$$

porque si $T > n$ tenemos que $H_m = 0$ para $m > T$ y no se suman y si $n \leq T$ entonces sumamos hasta n. Es decir, sumamos hasta el mínimo.

Por el teorema 111, W_n es una supermartingala. En particular $E_{T\wedge n} \leq M_0$. $\square$

Aplicaciones *Sea $X_1, X_2, \dots$ iid con $P(X_i = 1) = P(X_i = -1) = 1/2$. Sea $S_n = X_1 + \cdots + X_n$ y $\tau = \text{mín}\{n : X_n \notin (a,b)\}$. Como S_n es una martingala y τ es un tiempo de parada, nos gustaría decir*

$$x = E_x S_\tau = aP_x(S_\tau = a) + bP_x(1 - P(S_\tau = a))$$

y resolviendo la ecuación, concluir que

$$P_x(S_\tau = a) = \frac{b-x}{b-a}.$$

Fórmula correcta, pero hay que tener cuidado.

Lo que hicimos es

$$\begin{aligned} x = E_x(S_{\tau\wedge n}) &= \lim_n E_x(S_{\tau\wedge n}) \quad \text{(por el teorema, constante en } n\text{)} \\ \text{“}=\text{”}\ E_x(\lim_n S_{\tau\wedge n}) &= E_x S_\tau \quad \text{(ojo! límite adentro de la integral)} \\ &= aP_x(S_\tau = a) + b(1 - P_x(S_\tau = a)) \quad \text{(probabilidad de salida)} \end{aligned}$$

Ejemplo. Martingala mala. *Supongamos que $x = 1$, $V_a = \min\{n : S_n = a\}$ y $T = V_0$. Como S_n es recurrente, $P_1(T < \infty) = 1$ y*

$$E_1 S_T = 0 \neq 1 = E_1 S_{T\wedge n}, \quad \text{para todo } n.$$

Un buen contraejemplo de que no siempre vale $\lim EX_n = E \lim X_n$.

Ruina del jugador *Sean X_i iid con*

$$P(X_i = 1) = p, \quad P(X_i = -1) = q = 1 - p, \quad \frac{1}{2} < p < 1.$$

Sea $h(x) = (q/p)^x$. Vimos que $M_n := h(S_n)$ es una martingala. Tenemos que $\tau := \min\{n : S_n \notin (a, b)\}$ es un tiempo de parada y vimos que $P_x(\tau < \infty) = 1$.

Aceptando poner el límite dentro de la esperanza, tendríamos

$$\left(\frac{p}{q}\right)^x = \left(\frac{p}{q}\right)^a P_x(S_\tau = a) + \left(\frac{p}{q}\right)^b [1 - P_x(S_\tau = a)] \tag{107}$$

y resolviendo, tendríamos

$$P_x(S_\tau = a) = \frac{(q/p)^b - (q/p)^x}{(q/p)^b - (q/p)^a}$$

Demostración de (107)*:*

$$\begin{aligned} (q/p)^x &= E_x M_{\tau\wedge n} \\ &= (q/p)^a P_x(\tau \le n, S_\tau = a) + (q/p)^b P_x(\tau \le n, S_\tau = b) \\ &\quad + E_x((q/p)^{S_n}, \tau > n) \end{aligned}$$

Como $\{\tau \le n\} \nearrow \{\tau < \infty\}$ y $P(\tau < \infty) = 1$,

$$\begin{aligned} P_x(\tau \le n, S_\tau = a) &\nearrow P_x(S_\tau = a) \\ P_x(\tau \le n, S_\tau = b) &\nearrow P_x(S_\tau = b) \end{aligned}$$

Como $(q/p)^{S_n} \le (q/p)^a$ en el conjunto $\{\tau > n\}$, tenemos

$$E_x((q/p)^{S_n}, \tau > n) \le (q/p)^a P_x(\tau > n) \to_n 0, \quad \text{porque } P(\tau < \infty) = 1.$$

lo que demuestra (107).

Teorema 113 *Si M_n es una martingala y T un tiempo de parada con $P(T < \infty) = 1$ y $|M_{T\wedge n}| \le K$ para alguna constante K, entonces $EM_T = EM_0$.*

Demostración

$$EM_0 = EM_{T\wedge n} = E(M_T, T \le n) + E(M_n, T > n)$$

El módulo del segundo término está acotado por $KP(T > n)$ y para el primero escriba

$$|E(M_T, T \leq n) - E(M_T)| = |E(M_T, T > n)| \leq KP(T > n). \quad \square$$

Paseo aleatorio continuo a la izquierda *Sean X_i variables iid enteras con*

$$EX_i > 0, \quad P(X_i \geq -1) = 1, \quad P(X_i = -1) > 0.$$

Sea $\phi(\theta) = E\exp(\theta X_i)$ y sea

$$0 > \alpha := \text{ solución de } \phi(\alpha) = 1.$$

Para ver que existe ese α note que (i) $\phi(0) = 1$ y

$$\phi'(\theta) = \frac{d}{d\theta} E e^{\theta X_i} = E(X_i e^{\theta X_i}), \quad \text{y así} \quad \phi'(0) = EX_i > 0,$$

de donde sigue que $\phi(\theta) < 1$ para θ negativo chico. (ii) Si $\theta < 0$, tenemos $\phi(\theta) \geq e^{-\theta} P(X_i = -1) \to \infty$ con $\theta \to -\infty$.

Para el α elegido $M_n := \exp(\alpha S_n)$ es una martingala:

$$E(M_{n+1} - M_n | A_\nu) = M_n(\nu)(E\exp(\alpha X_n) - 1) = 0.$$

Teorema 114 *Sea S_n un paseo aleatorio continuo a la izquierda con media positiva. Sea $a < x$ y $V_a = \min\{n : S_n = a\}$. Entonces,*

$$P_x(V_a < \infty) = e^{\alpha(x-a)}.$$

Demostración *Informalmente:*

$$e^{\alpha x} = E(\exp(S_{V_a}), V_a < \infty) = \exp(\alpha a) P(V_a < \infty) \tag{108}$$

y podemos concluir. Para demostrar (108) escribimos:

$$e^{\alpha x} = E_0 \exp(\alpha S_{V_a \wedge n}) = e^{\alpha a} P_0(V_a \leq n) + E_0(\exp(\alpha S_n), V_a > n)$$

Pero en $\{V_a = \infty\}$ tenemos $S_n/n \to EX_i > 0$, por lo tanto $S_n \to \infty$ c.s. y $\exp(\alpha S_n) \to 0$ c.s.. Como $\exp(\alpha S_n) \leq 1$, convergencia dominada implica que $E\exp(\alpha S_n)$ también se va a 0, con lo que concluimos la demostración de (108). $\square$

Convergencia

Teorema 115 *Si $X_n \geq 0$ es una supermartingala, entonces $X_\infty = \lim_n X_n$ c.s. existe y $EX_\infty \leq EX_0$.*

Lema 116 *Sea $X_n \geq 0$ una supermartingala y $\lambda > 0$.*

$$P(\max_n X_n > \lambda) \leq \frac{EX_0}{\lambda}.$$

Demostración *Sea $T = \min\{n \geq 0 : X_n > \lambda\}$. Por el Teorema 112,*

$$\begin{aligned} EX_0 &\geq E(X_{T \wedge n}) = E(X_T, T \leq n) + E(X_n, T > n) \\ &\geq \lambda P(T \leq n) \end{aligned}$$

Es decir, $P(T \leq n) \leq EX_0/\lambda$. Como $\{T \leq n\}$ es una sucesión no creciente de eventos que converge a $\{T < \infty\} = \{\max_n X_n > \lambda\}$, podemos concluir. $\square$

Demostración *[Demostración del Teorema 115] Sean $a < b$, $S_0 = 0$ y defina los tiempos de parada*

$$\begin{aligned} R_k &:= \min\{m \geq S_{k-1} : X_m \leq a\} \\ S_k &:= \min\{m \geq R_k : X_m \geq b\} \end{aligned}$$

Usando el lema anterior con la supermartingala $\tilde{X}_n := X_{R_k+n}$ *y* $\lambda = b$,

$$P(S_k < \infty | R_k < \infty) \leq \frac{E\tilde{X}_0}{b} = \frac{EX_{R_k}}{b} \leq \frac{a}{b}.$$

Iterando, vemos que

$$P(S_k < \infty) \leq \left(\frac{a}{b}\right)^k. \tag{109}$$

Como la cota es sumable, por Borel Cantelli, X_n *cruza desde* a *hasta* b *solamente un número finito de veces.*

Defina

$$Y = \liminf_n X_n \quad y \quad Z = \limsup_n X_n.$$

Para probar que los límites son iguales con probabilidad 1, razonamos por absurdo. Supongamos $P(Y < Z) > 0$, *entonces hay números* $a < b$ *tales que* $P(Y < a < b < Z) > 0$. *Pero en este caso* X_n *cruza desde abajo de* a *hasta arriba de* b *infinitas veces con probabilidad positiva. Como eso tiene probabilidad cero, la convergencia queda demostrada.*

Para ver que $EX_\infty \leq EX_0$, *observe que para cualquier número positivo* K *tenemos*

$$EX_0 = EX_n \geq E(X_n \wedge K) \underset{n}{\to} E(X_\infty \wedge K)$$

Para ver la convergencia sea $\tilde{X}_n = X_n \wedge K$

$$\begin{aligned} |E(X_n \wedge K) - E(X_\infty \wedge K)| &= |E(\tilde{X}_n - \tilde{X}_\infty)| && (110) \\ &\leq E|\tilde{X}_n - \tilde{X}_\infty| && (111) \\ &= E(|\tilde{X}_n - \tilde{X}_\infty| && (112) \end{aligned}$$

por lema abajo. Por convergencia monótona

$$E(X_\infty \wedge K) \underset{K}{\nearrow} EX_\infty. \qquad \square$$

Lema 117 $X_n \to 0$ *c.s. y* $|X_n| \leq K$ *para todo* n, *entonces* $EX_n \to 0$.

Demostración

$$\begin{aligned} |EX_n| \leq E|X_n| &= E(|X_n|\mathbf{1}\{|X_n| > \varepsilon\}) + E(|X_n|\mathbf{1}\{|X_n| \leq \varepsilon\}) && (113) \\ &\leq KP(|X_n| > \varepsilon) + \varepsilon && (114) \end{aligned}$$

y sacando el límite primero en n *y después en* ε *estamos.* $\square$

Urna de Polya. *Una urna tiene bolas rojas y verdes. En el instante 0 hay* k *bolas con al menos una bola de cada color. En el instante* n *sacamos una bola al azar y la retornamos a la urna con otra del mismo color. Sea* X_n *la fracción de bolas rojas en el instante* n. *Veremos que* X_n *es una martingala. Observe que en el instante* n *hay* $n+k$ *bolas en la urna. Así,* $R_n = (n+k)X_n$ *es el número do bolas rojas en la urna después de la* n*-ésima extracción y*

$$P(R_{n+1} = R_n + 1) = X_n, \qquad P(R_{n+1} = R_n) = 1 - X_n$$

Denotando $A_\nu = \{X_n = x_n, \ldots, X_0 = x_0\}$, *tenemos*

$$\begin{aligned} E(X_{n+1}|A_\nu) &= \frac{R_n + 1}{n+k+1}\frac{R_n}{n+k} + \frac{R_n}{n+k+1}\Big(1 - \frac{R_n}{n+k}\Big) \\ &= \frac{1}{n+k+1}\frac{R_n}{n+k} + \frac{n+k}{n+k+1}\frac{R_n}{n+k} = \frac{R_n}{n+k} = X_n. \end{aligned}$$

Como $X_n \geq 0$, *el Teorema 115 implica* $X_n \to X_\infty$, a.s.

Queda como ejercicio demostrar que si $k = 2$, *con una bola de cada color, entonces*

$$P\Big(X_n = \frac{j}{n+2}\Big) = \frac{1}{n+1}, \qquad j \in \{1, \ldots, n+1\},$$

(sugiero usar inducción). Sacando límite, X_n *converge en distribución a Uniforme*$[0,1]$. *Como consecuencia,* $X_\infty \sim$ *Uniforme*$[0,1]$.

Proceso de ramificación. *Recordemos que* Z_n *es el número de individuos en el instante* n *y que en cada paso los individuos presentes crean independientemente nuevos individuos con media* $\mu \in (0,\infty)$. *Si* $p(x,y)$ *es la matriz de transición de esta cadena y definimos* $h(x,n) = x/\mu^n$ *vemos que*

$$\sum_y p(x,y)h(y,n+1) = \frac{1}{\mu^{n+1}}\sum_y p(x,y)y = \frac{\mu x}{\mu^{n+1}} = h(x,n)$$

Por lo tanto, por el Teorema 103, $W_n = Z_n/\mu^n$ *es una martingala.*

Subcrítico $\mu < 1$. *Por Chevichev,*

$$P(Z \geq 1) \leq EZ_n = \mu^n EW_n = \mu^n EW_0 = \mu^n EZ_0 \underset{n}{\to} 0$$

Crítico *Sea* p_k *la probabilidad que un individuo tenga* k *hijos. Si* $\mu = 1$ *y* $p_1 < 1$, *entonces* $P(Z_n > 0) \to_n 0$. *Veamos: Cuando* $\mu = 1$ *tenemos que* Z_n *es una martingala y por lo tanto converge casi seguramente. Como* Z_n *es entera, si* $Z_n(\omega)$ *converge, lo tiene que hacer a algún* $j = j(\omega) \in \mathbb{N}$. *Pero esto implica que existe un* $n(\omega)$ *tal que* $Z_n(\omega) = j$ *para* $n \geq n(\omega)$, *pero esto tiene probabilidad 0 porque* $p_1 < 1$.

Supercrítico *Si* $\mu > 1$ *entonces* $Z_n/\mu^n \to_n W$.

Probaremos que si $\mu > 1$ *entonces* $P(Z_n > 0$ *para todo* $n) > 0$. *Consideramos el proceso modificado* S_n, *que cuenta en cada paso la reproducción de uno de los individuos. Si* $S_n > 0$ *entonces* $S_{n+1} = S_n - 1 + Y_{n+1}$, *donde* $P(Y_n = k) = p_k$. *Definiendo* $T_0 := \min\{n : S_n = 0\}$ *tenemos*

$$\{Z_n > 0 \text{ para todo } n\} = \{T_0 = \infty\}.$$

Como $Y_n \geq 0$, *tenemos que* S_n *es un paseo aleatorio continuo a la izquierda con incrementos* $X_m := -1 + Y_m$ *y* $EX_m = \mu - 1 > 0$ *cuando* $\mu > 1$. *El Teorema 114 dice que*

$$P_1(T_0 < \infty) = e^\alpha$$

con $\alpha < 0$ *solución de* $E\exp(\alpha X_i) = 1$. *Poniendo* $\rho = e^\alpha$, *esto significa*

$$1 = \sum_k p_k \rho^{k-1} \qquad \text{que es lo mismo que } \rho = \sum_k p_k \rho^k$$

que es el mismo resultado que encontramos antes: la probabilidad de extinción es el punto fijo en $[0,1)$ *de* ϕ.

Si podemos demostrar que $P(W > 0) > 0$, *entonces sabremos que la población crece exponencialmente. El teorema de Kesten-Stigum dice que*

$$P(W > 0) > 0 \quad \text{si y solo si} \quad \sum_k p_k k \log k < \infty.$$

Ese teorema es difícil de demostrar.

Se puede demostrar

Teorema 118 *Si* $\sum_k kp_k > 1$ *y* $\sum_k k^2 p_k < \infty$, *entonces* $P(W = 0) = \rho$.

Paseos aleatorios

Contando caminos *Un* ***camino*** *de longitud n es un vector $(s_0, s_1, \ldots, s_n)$,*

$$s_k = x_1 + \cdots + x_k$$

donde los incrementos $x_i \in \{-1, 1\}$.

Hay 2^n caminos de longitud n. Si $s_0 = 0$ y $s_n = x$, entonces los a incrementos positivos y los b incrementos negativos deben satisfacer:

$$a + b = n, \qquad a - b = x.$$

Es decir:

$$a = \frac{n+x}{2}, \qquad b = \frac{n-x}{2}.$$

Así, $N_{n,x}$ el número de caminos de longitud n que van de 0 a x es

$$N_{n,x} = \binom{a+b}{a} = \binom{a+b}{b}$$

Consideraremos $N_{n,x} = 0$ cuando no se puede alcanzar x en n pasos.

Ejemplo Elecciones. *Supongamos que en una elección el candidato A saca a votos y el candidato B saca b votos, con $a > b$ (es decir A gana la elección).*

Cual es la probabilidad que durante todo el escrutinio A esté por delante de B?

Podemos representar la ventaja de A por un camino: cada vez que sale un voto para A sumamos 1 y cada vez que sale un voto para B restamos 1. O sea que $x_i = 1$ si el i-ésimo voto computado sale para A y $x_i = -1$ en caso que sea para B. La ventaja de A después de computar el k-ésimo voto es

$$s_k = x_1 + \cdots + x_k$$

A lidera todo el escrutinio si para todo $0 < k \le n$,

$$s_1 > 0, s_2 > 0, \ldots, s_k > 0.$$

Asumimos que todos los posibles caminos de tamaño n que terminan en $a - b$ son igualmente probables. (todas las permutaciones de los votos son igualmente probables)

Principio de reflexión

Considere puntos espacio-temporales (k, x) y (n, y) con $0 \le k < n$, $x > 0$, $y > 0$.

El punto ***reflejado*** *de (k, x) es $(k, -x)$*

Consideraremos caminos que van de (k, x) a (n, y).

Principio de reflexión *El número de caminos que van de (k, x) a (n, y) que toca o cruza el eje de las absisas es igual al número de caminos que van de $(k, -x)$ a (n, y).*

Dem *Considere un camino $x = s_k, s_{k+1}, \ldots, s_n = y$ que toque el eje de las absisas. Sea T el primer instante en que eso sucede:*

$$T = \min\{i \in [k, n] : s_i = 0\}$$

El camino

$$-x = -s_k, -s_{k+1}, \ldots, -s_{T-1}, 0, s_{T+1}, \ldots, s_n = y$$

va de $(k, -x)$ a (n, y).

Como las secciones $(k, x), \ldots, (t, 0)$ y $(k, -x), \ldots, (t, 0)$ son reflejadas una de la otra, existe una biyección entre esos dos pedazos. Esto implica que el número de caminos es el mismo. $\square$

***Lema del escrutinio* Sean n *y* x *enteros positivos. Hay exactamente* $\frac{x}{n} N_{n,x}$ *caminos* $(s_1, \ldots, s_n = x)$ *desde el origen* $(0,0)$ *a* (n, x) *tal que* $s_1 > 0, \ldots, s_n > 0$.**

Dem *Claramente hay tantos caminos admisibles como caminos desde $(1,1)$ a (n,x) que no tocan el eje de las absisas. Por el lema de la reflexión, ese número es*

$$N_{n-1,x-1} - N_{n-1,x+1} = \binom{a+b-1}{a-1} - \binom{a+b-1}{a}$$

con a y b satisfaciendo que $a+b=n$ y $a-b=x$. Una cuenta muestra que ese número es igual a $\frac{x}{n}N_{n,x}$.
$\square$

Paseo aleatorio como proceso estocástico *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes con distribución*

$$P(X_i = 1) = \frac{1}{2}, \qquad P(X_i = -1) = \frac{1}{2}.$$

*Se define **paseo aleatorio** al proceso*

$$S_n = X_1 + \cdots + X_n, \quad n \geq 0$$

La probabilidad que el paseo esté en x en el instante n es

$$p_{n,x} = P(S_n = x) = \binom{n}{\frac{n+x}{2}} 2^{-n}$$

(se interpreta como 0 si $\frac{n+x}{2}$ no es un entero entre 0 y n.)

*Una **vuelta al origen** ocurre en el instante $2k$ si $S_{2k} = 0$. La vuelta sólo puede ocurrir en instantes pares.*

Definimos $u_{2k} := P(S_{2k} = 0)$.

$$u_{2k} = \binom{2k}{k} 2^{-2k}$$

Aproximación de Stirling del factorial:

$$\lim_n \frac{n!}{\sqrt{2\pi n}(n/e)^n} = 1$$

Ejercicio *Use la aproximación de Stirling para probar que*

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}$$

Eso quiere decir que

$$\lim_{k\to\infty} u_{2k}\sqrt{\pi k} = 1$$

El TCL nos dice que

$$\lim_n P(S_n \leq r\sqrt{n}) = \phi(r)$$

donde ϕ es la función de distribución acumulada de la Normal standard.

*El **primer retorno al origen** ocurre en el instante $2k$ si*

$$S_1 \neq 0, \ldots, S_{2k-1} \neq 0, S_{2k} = 0$$

y su probabilidad se denota f_{2k}.

Lema Las probabilidades u_{2k} y f_{2k} se relacionan por

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \cdots + f_{2n} u_0$$

Dem *Use el teorema de la probabilidad total.* $\square$

Sea $T := \min\{n > 0 : S_n = 0\}$ instante del primer retorno al origen.

Lema *Sea $n > 0$, entonces*

$$P(T > 2n) = P(S_{2n} = 0)$$

Dem *Por simetría,*

$$P(T > 2n) = P(S_1 > 0, \dots, S_{2n} > 0) + P(S_1 < 0, \dots, S_{2n} < 0)$$
$$= 2P(S_1 > 0, \dots, S_{2n} > 0)$$

Por el teorema de la probabilidad total:

$$P(S_1 > 0, \dots, S_{2n} > 0) = \sum_{x \geq 1} P(S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2x)$$

Por el lema de reflexión,

$$P(S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2x)$$
$$= 2^{-2n}(N_{2n-1,2x-1} - N_{2n-1,2x+1}) = \frac{1}{2}(p_{2n-1,2x-1} - p_{2n-1,2x+1})$$

Sumando (telescopicamente),

$$\sum_{x \geq 1} \frac{1}{2}(p_{2n-1,2x-1} - p_{2n-1,2x+1}) = \frac{1}{2} p_{2n-1,1} = \frac{1}{2} u_{2n}$$

Esa es la probabilidad que el camino sea siempre positivo. Sumando la proba que el camino sea siempre negativo obtenemos el lema. □

Máximo *El* ***máximo*** M_n *está definido por*

$$M_n(S_0, S_1, \dots) = \max\{S_0, \dots, S_n\}$$

Lema *Sea y un entero tal que $n \geq y > 0$. La probabilidad de un camino de $(0,0)$ a $(2n,0)$ con un máximo mayor o igual a y es igual a $p_{2n,2y} = P(S_{2n} = 2y)$.*

Dem *Queremos calcular $P(M_{2n} \geq y, S_{2n} = 0)$. El número de caminos de $(0,0)$ a $(2n,0)$ que tocan o cruzan y es igual al número de caminos de $(0,y)$ a $(2n,y)$ que tocan 0. Por el Lema de reflexión, ese número es igual a $N_{2n,2y}$. Multiplicando por 2^{-2n}, obtenemos*

$$P(M_{2n} \geq y, S_{2n} = 0) = p_{2n,2y}. \qquad \square$$

Observe que

$$p_{2n,2y} = \binom{2n}{\frac{2n+2y}{2}} = \binom{2n}{n+y}$$

Lema

$$\lim_{n \to \infty} P\big(M_{2n} \geq b\sqrt{2n} \mid S_{2n} = 0\big) = e^{-2b^2}$$

Dem *Dividiendo la expresión obtenida para $p_{2n,2y}$ por $p_{2n,0} = \binom{2n}{n}2^{-2n}$, cancelan los $(2n)!$ y las potencias de 2 y obtenemos*

$$P(M_{2n} \geq y | S_{2n} = 0) = \frac{p_{2n,2y}}{p_{2n,0}} = \frac{n!\, n!}{(n-y)!\,(n+y)!}$$
$$= \frac{n(n-1)\dots(n-y+1)}{(n+y)(n+y-1)\dots(n+1)}$$

dividiendo cada uno de los términos del denominador por el correspondiente término del numerador, obtenemos

$$= \Big(\Big(1 + \frac{y}{n}\Big)\Big(1 + \frac{y}{n-1}\Big)\dots\Big(1 + \frac{y}{n-y+1}\Big)\Big)^{-1}$$

Substituyendo $y = b\sqrt{2n}$, y

$$= \Big(\Big(1 + \frac{b\sqrt{2}}{\sqrt{n}}\Big)\Big(1 + \frac{b\sqrt{2}}{\sqrt{n} - \frac{1}{\sqrt{n}}}\Big)\dots\Big(1 + \frac{b\sqrt{2}}{\sqrt{n} - \frac{b\sqrt{2}+1}{\sqrt{n}}}\Big)\Big)^{-1}$$

$$\sim \Big(1 + \frac{b\sqrt{2}}{\sqrt{n}}\Big)^{-b\sqrt{2}\sqrt{n}} \to e^{-2b^2} \qquad \square$$

Ley de grandes números *X_n tiene esperanza cero. Por lo tanto, por independencia y estacionariedad tenemos:*

$$\lim_n \frac{S_n}{n} = \frac{1}{2}$$

Incrementos independientes

$$\textit{Las variables } S_n \textit{ y } S_{n+m} - S_n \textit{ son independientes} \tag{115}$$

porque dependen de pedazos disjuntos de las variables X_i.

Por el TCL: $\frac{S_n}{\sqrt{n}} \to$ Normal$(0,1)$.

Además

$$S_{n+m} - S_n \sim S_m, \tag{116}$$

porque ambas variables son sumas de m variables independientes Bernoulli$(\frac{1}{2})$.

Con el mismo argumento vamos a probar que las distribuciones finito dimensionales convergen a las distribuciones finito dimensionales de un Movimiento Browniano *(a ser definido y construido).*

Defino un nuevo proceso a tiempo continuo:

$$B_n(t) := \sqrt{n} S_{\lfloor tn \rfloor}$$

Se estudian tiempos grandes de orden n en espacios grandes de orden $\sqrt{n}$. Esto se llama ***rescalamiento difusivo****.*

Por el TCL, para cada t fijo $B_n(t)$ converge en distribución a una variable aleatoria $B(t)$ con distribución normal:

$$\lim_{n\to\infty} B_n(t) \stackrel{D}{=} B(t) \sim \textit{Normal}(0,t)$$

Teorema 119 *para cualquier conjunto de k instantes $0 \le t_1 < \ldots < t_k < \infty$*

$$\lim_n (B_n(t_1), \ldots, B_n(t_k)) \stackrel{D}{=} (B(t_1), \ldots, B(t_k))$$

Donde $(B(t_1), \ldots, B(t_k))$ es un vector Gaussiano con covarianzas

$$cov(B(t_i), B(t_j)) = t_i, \textit{ si } t_i < t_j.$$

Un vector aleatorio $Y = (Y_1, \ldots, Y_n)$ es Gaussiano *si todas sus combinaciones lineales son normales: para cualquier conjunto de números reales $a_1, \ldots, a_n$,*

$$\sum_{i=1}^n a_i Y_i \quad \textit{tiene distribución normal multivariada.}$$

Un vector Gaussiano esta caracterizado por sus covarianzas $Cov(G_i, G_j)$, $i, j = 1, \ldots, k$.

Esta definición es equivalente a la existencia de una matriz A de $n \times m$ y un vector n dimensional b tal que $Y \stackrel{D}{=} A^T Z + b$, donde el vector $Z = (Z_1, \ldots, Z_k)$ consiste de normales standard independientes.

Demostración del Teorema. *Como B_n tiene incrementos independientes y estacionarios, por independencia y el teorema central del límite, vale que*

$$\lim_n (B_n(t_1), B_n(t_2) - B_n(t_1), \ldots, B_n(t_k) - B_n(t_{k-1}))$$

$$\stackrel{D}{=} (G_1, \ldots, G_k))$$

con $G_i \sim$ Normal$(0, t_i - t_{i-1})$ independientes. Definiendo

$$B(t_i) := G_1 + \cdots + G_i,$$

concluimos que

$$\lim_n (B_n(t_1), B_n(t_2), \ldots, B_n(t_k)) \stackrel{D}{=} (B(t_1), B(t_2), \ldots, B(t_k)),$$

que es un vector Gaussiano (esto se ve inmediatamente usando la definicion).

Calculemos las covarianzas: para $i < j$,

$$Cov(B(t_i), B(t_j)) = Cov(B(t_i), B(t_i) + G_{i+1} + \cdots + G_j) = VB(t_i)$$

porque $B(t_i) = G_1 + \cdots + G_i$ es independiente de $G_{i+1}, \dots, G_j$. Finalmente

$$VB(t_i) = VG_1 + \cdots + VG_i = t_i. \quad \square$$

Decimos que un proceso $B(t)$ es Gaussiano *si sus distribuciones finito-dimensionales son vectores Gaussianos.*

Deducimos que ***si existe un proceso*** *Gaussiano $B(t)$ con covarianzas $cov(B(t), B(t+s)) = t$, entonces las distribuciones finito-dimensionales de $B_n(t)$ convergen a las de $B(t)$.*

Preliminares: Vectores Gaussianos

Decimos que un vector $X = (X_1, \dots, X_n)$ es Gaussiano *si existe una matriz de $n \times m$ dimensiones A y un vector n-dimensional b tal que $X^t = AZ + b$, donde Z es un vector m-dimensional con coordenadas* Normal$(0, 1)$ *independientes.*

El vector Z es un vector Gaussiano standard.

La matriz de covarianzas *está dada por*

$$Cov(X) = E[(X - EX)(X - EX)^t] = AA^t$$

Es decir, $Cov(X_i, X_j) = \sum_k a_{ik} a_{jk}$.

Lema 120 *Si A es ortogonal de dimensión $d \times d$, es decir $AA^t = I$ (identidad) y X es un vector Normal standard de dimensión d, entonces AX también es un vector Normal standard de dimensión d.*

Demostración *X tiene densidad*

$$f(x_1, \dots, x_d) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2}$$

donde $|\cdot|$ es la norma Euclidea. La densidad de AX es $f(A^{-1}x) \det |A^{-1}|$. El determinante es 1 y como matrices ortogonales preservan la norma Euclidea, la densidad de X es invariante bajo A. $\square$

Corolario 121 *Sean $X_i \sim N(0, \sigma^2)$ independientes. Entonces $X_1 - X_2$ y $X_1 + X_2$ son $N(0, 2\sigma^2)$ independientes.*

Demostración *$(X_1/\sigma, X_2/\sigma)^t$ es un vector Gaussiano standard. La matriz*

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

es ortogonal y

$$AX^t = \big((X_1 + X_2)/(\sqrt{2}\sigma), (X_1 - X_2)/(\sqrt{2}\sigma)\big)$$

tiene entonces coordenadas independientes normales standard. $\square$

El resultado que necesitamos es un caso particular del lema anterior que se puede demostrar directamente:

Lema 122 *Sean $U, V \sim$ Normal$(0,1)$ independientes. Defina*

$$X := \frac{V}{2} + \frac{U}{2}, \qquad Y := \frac{V}{2} - \frac{U}{2} \tag{117}$$

Entonces X y Y son independientes con distribución $N(0,1/2)$.

Demostración Sea

$$g(u,v) = \Big(\frac{v}{2} + \frac{u}{2}, \frac{v}{2} - \frac{u}{2}\Big)$$

cuya inversa es

$$g^{-1}(x,y) = (x+y, x-y)$$

y |Jacobiano| = 2.

$$\begin{aligned} f_{X,Y}(x,y) &= 2f_{U,V}(g^{-1}(x,y)) = 2f_{U,V}(x+y,x-y) \\ &= \frac{1}{\pi}\exp\Big(-\frac{(x+y)^2}{2} - \frac{(x-y)^2}{2}\Big) \\ &= \frac{1}{\pi}\exp\Big(-x^2-y^2\Big) \\ &= \frac{1}{\sqrt{2\pi\frac{1}{2}}}\exp\Big(\frac{-x^2}{2\frac{1}{2}}\Big)\frac{1}{\sqrt{2\pi\frac{1}{2}}}\exp\Big(\frac{-y^2}{2\frac{1}{2}}\Big) \end{aligned}$$

Que es lo que queríamos demostrar. □

Proposición 123 *Si X e Y son vectores Gaussianos d dimensionales con $EX = EY$ y $CovX = CovY$, entonces X e Y tienen la misma distribución.*

Corolario 124 *Sea $X = (X_1, \dots, X_n)$ un vector Gaussiano. Si $Cov(X_i, X_j) = 0$ para todo $i \neq j$, entonces X_i son independientes.*

Lema 125 *Si $Z \sim N(0,1)$ y $z > 0$, entonces*

$$P(Z > z) \leq \frac{1}{z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}$$

Demostración

$$P(Z > z) = \int_z^\infty e^{-x^2/2}dx \leq \int_z^\infty \frac{x}{z}e^{-x^2/2}dx = \frac{1}{z}e^{-z^2/2}. \quad \square$$

Movimiento Browniano

Un proceso $B(t), t \geq 0$ ($t \in \mathbb{R}$) es un movimiento Browniano standard *si*

0. $B(0) = 0$.

1. $B(t)$ tiene incrementos estacionarios e independientes.

2. $B(t)$ tiene distribución Normal $N(0,t)$

3. $t \mapsto B(t)$ es una función continua, casi seguramente.

Ojo, la continuidad no sigue de las distribuciones finito-dimensionales. Por ejemplo, si $B(t)$ es MB y U es uniforme en $[0,1]$, definiendo $\tilde{B}(t) := B(t)$ para $t \neq U$ y $B(t) = 0$ para $t = U$, vemos que las distribuciones finito-dimensionales de $\tilde{B}(t)$ son las mismas que las de $B(t)$ pero $\tilde{B}(t)$ no es continuo.

En general, llamamos Movimiento Browniano con coeficiente de difusión $c^2 > 0$ al proceso $\tilde{B}(t) := cB(t)$. Tiene las mismas propiedades que el Browniano con la diferencia que $\tilde{B}(t) \sim N(0, c^2)$.

Construcción de Levy del Movimiento Browniano *(Mörters-Peres, Toth)*

Vamos a construir primero en los diádicos. *Sea*

$$\mathcal{D}_n = \{\tfrac{k}{2^n} : 0 \le k \le 2^n,\ k \in \mathbb{Z}\}$$

Vamos a muestrear $B(t)$ para t tiempo diádico en $\mathcal{D}_n$ e interpolamos linealmente. Después probamos que el límite cuando $n \to \infty$ existe y tiene las propiedades del movimiento Browniano. Definimos

$$\mathcal{D} = \cup_n \mathcal{D}_n$$

Sean Z_t, $t \in \mathcal{D}$ una colección iid Normal$(0,1)$.

Defina $B(t)$ para t en $\mathcal{D}_0 = \{0,1\}$:

$$B(0) = 0, \quad B(1) = Z_1.$$

Para cada n definimos variables $B(d)$, $d \in \mathcal{D}_n$ tales que

1) Para cada $r < s < t$ en $\mathcal{D}_n$, $B(t) - B(s) \sim N(0, t-s))$ y es independiente de $B(s) - B(r)$.

2) los vectores $(B(d) : d \in \mathcal{D}_n)$ y $(Z_t : t \in \mathcal{D} \setminus \mathcal{D}_n)$ son independientes.

Ya lo vimos para $\mathcal{D}_0 = \{0,1\}$. Procedemos por inducción:

Asumimos que $B(t)$ está definido para $t \in \mathcal{D}_{n-1}$ y que (1) y (2) valen para $\mathcal{D}_{n-1}$.

Para $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ defina:

$$B(d) = \frac{B(d - 2^{-n}) + B(d + 2^{-n})}{2} + \frac{Z_d}{2^{(n+1)/2}}$$

O sea, d es el punto medio del intervalo $[d - 2^{-n}, d + 2^{-n}]$ cuyos extremos están en $\mathcal{D}_{n-1}$ y $B(d)$ es una normal cuya media es el promedio de los valores definidos en esos extremos y varianza $2^{-(n+1)}$.

El primer sumando es función de $(Z_s : s \in \mathcal{D}_{n-1})$ y por lo tanto independiente de $(Z_d : d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1})$. Esto demuestra la propiedad (2) para $\mathcal{D}_n$.

Además, como $\frac{B(d-2^{-n}) - B(d+2^{-n})}{2}$ también es función de $(Z_s : s \in \mathcal{D}_{n-1})$, tenemos que es independiente de $\frac{Z_d}{2^{(n+1)/2}}$. Por lo tanto su suma

$$B(d) - B(d - 2^{-n}) \quad \text{y su diferencia } B(d + 2^{-n}) - B(d) \tag{118}$$

son independientes y $N(0, 2^{-n})$ por el Lema 122.

De hecho, todos los incrementos $B(d) - B(d - 2^{-n})$ para $d \in \mathcal{D}_n \setminus \{0\}$ son independientes. Para verlo, basta probar que son independientes 2 a 2, porque son vectores Gaussianos. Hay dos posibilidades: o bien son incrementos sucesivos como en (118) o están separados por un $d' \in \mathcal{D}_{n-1}$. Como los incrementos a la izquierda y derecha de d' están construídos con conjuntos disjuntos de variables Z_t, son independientes y eso concluye el paso inductivo.

Los procesos interpolados *Defina*

$$F_0(t) = \begin{cases} Z_1, & \text{si } t = 1 \\ 0, & \text{si } t = 0 \\ \text{interpolación lineal} & \text{entre esos valores} \end{cases}$$

$$F_n(t) = \begin{cases} 2^{-(n+1)/2} Z_t, & \text{si } t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1} \\ 0 & \text{si } t \in \mathcal{D}_{n-1} \\ \text{interpolación lineal} & \text{entre puntos consecutivos de } \mathcal{D}_n \end{cases}$$

Esas funciones son continuas y por la definición de B en los diádicos,

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d), \quad d \in \mathcal{D}_n$$

(ejercicio). Por otro lado, por el Lema 125, para $c > 0$,

$$P(|Z_d| \geq c\sqrt{n}) = 2P(Z_d \geq c\sqrt{n}) \leq \frac{2}{c\sqrt{n2\pi}} \exp\Big(\frac{-c^2 n}{2}\Big) \leq \exp\Big(\frac{-c^2 n}{2}\Big),$$

para $n \geq 2/(c^2\pi)$ (tal que $2/(c\sqrt{n2\pi}) \leq 1$).

Sea $A_n = \{$existe $d \in \mathcal{D}_n$ tal que $|Z_d| \geq c\sqrt{n}\}$.

$$\begin{aligned} P(A_n) &\leq \sum_{d\in\mathcal{D}_n} P(|Z_d| \geq c\sqrt{n}) \\ &\leq (2^n + 1)\exp\Big(\frac{-c^2 n}{2}\Big) \end{aligned}$$

para $n \geq 2/(c^2\pi)$. Por lo tanto, si $c > \sqrt{2\log 2}$,

$$\sum_n P(A_n) < \infty.$$

Por Borel Cantelli, casi seguramente existe un N (aleatorio, que depende de $(Z_d : d \in \mathcal{D})$) tal que para todo $n \geq N$ y $d \in \mathcal{D}_n$ tenemos $|Z_d| < c\sqrt{n}$. Por lo tanto para $n \geq N$,

$$\|F_n\|_\infty < c\sqrt{n}2^{-n/2}$$

Esto implica que casi seguramente la serie

$$B(t) = \sum_n F_n(t)$$

es uniformemente convergente en $t \in [0,1]$. Por lo tanto el límite es continuo y lo llamamos $(B(t) : t \in [0,1])$.

Lema 126 *El límite $(B(t) : t \in [0,1])$ tiene incrementos estacionarios e independientes y satisface $B(t) \sim N(0,t)$.*

Demostración *Esto es una consecuencia de las propiedades de $(B(t) : t \in \mathcal{D} \cap [0,1])$. Veamos.*

Sea $t_k \in \mathcal{D}$ una sucesión de diádicos que converge a $t \in [0,1]$. Como $B(t)$ es continuo, las variables aleatorias $B(t_k)$ convergen casi seguramente a $B(t)$. Como $B(t_k) \sim N(0,t_k)$ y $B(t_k)$ converge en distribución a $N(0,t)$,

$$P(B(t_k) \leq b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi t_k}} e^{x^2/2t_k} dx \underset{k\to\infty}{\to} \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi t}} e^{x^2/2t} dx$$

tenemos que $B(t) \sim N(0,t)$. El mismo argumento sirve para probar incrementos independientes y estacionarios. Sean $s_1 < \ldots < s_n$ y $s_{1,k} \leq \cdots \leq s_{n,k}$ diádicos que convergen a s_i cuando $k \to \infty$. Por la continuidad de $B(t)$,

$$\big(B(s_{2,k}) - B(s_{1,k}), \ldots, B(s_{n,k}) - B(s_{n-1,k})\big) \underset{k\to\infty}{\to} \big(B(s_2) - B(s_1), \ldots, B(s_n) - B(s_{n-1})\big)$$

Además

$$\begin{aligned} &P(\cap_{i=1}^{n-1}\{B(s_{i+1,k}) - B(s_{i,k}) \leq b_i\}) \\ &= \prod_{i=1}^{n-1} P(B(s_{i+1,k}) - B(s_{i,k}) \leq b_i) \\ &= \prod_{i=1}^{n-1} P(N(0, s_{i+1,k} - s_{i,k}) \leq b_i) \\ &= \prod_{i=1}^{n-1} P(N(0, s_{i+1} - s_i) \leq b_i) \end{aligned}$$

Lo que implica que el vector $\big(B(s_2)-B(s_1),\dots,B(s_n)-B(s_{n-1})\big)$ tiene coordenadas independientes Gaussianas con varianza igual al tamaño del intervalo. $\square$

Construcción del proceso para cualquier $t\in\mathbb{R}^+$***.*** *Sean $\big((B^i(s):s\in[0,1]):i=1,2,\dots\big)$ copias independientes del Browniano con tiempos en $[0,1]$, donde B^i usa $(Z_d:d\in i+\mathcal{D})$, normales asociadas a los diádicos de $[i,i+1]$.*

Para $t\ge 0$ defina

$$B(t)=B^{[t]}(t-[t])+\sum_{i=0}^{[t]-1}B^{[t]}(1)$$

Con esto terminamos la definición del Browniano $(B(t):t\ge 0)$ como una función de $(Z_d,\, d\in\cup_i(i+\mathcal{D}))$, las normales standard asociadas a los diádicos de $\mathbb{R}$. $\square$

Propiedades de las trayectorias

Lema (invariancia bajo rescalamiento) Sea $B(t)$ ***Browniano y*** $a>0$***. Entonces el proceso*** $X(t)$ ***definido por*** $X(t)=\frac{1}{a}B(a^2t)$ ***es tambien un Browniano standard.***

Dem *Se ve inmediatamente que el proceso es Gaussiano y que tiene trayectorias continuas y $\tilde{B}(0)=0$. Queda como ejercicio ver que tiene las covarianzas correctas.* $\square$

Lema. *$B(t)$ es Browniano standard. Entonces el proceso $X(t)$ definido por*

$$\begin{cases}0 & \text{para } t=0\\ tB(1/t) & \text{para } t>0\end{cases}$$

es tambien Browniano standard.

Dem *Se ve inmediatamente que el proceso es Gaussiano y que tiene trayectorias continuas en todo $t>0$ y $X(0)=0$. Que los incrementos son independientes y Gaussianos es claro. Basta ver que son estacionarios.*

$$\begin{aligned}V(X(t+h)-X(t))&=V((t+h)B(1/(t+h))-tB(1/t))\\&=h^2VB(1/(t+h))+t^2V(B(1/(t+h))-tB(1/t))\\&=\frac{h^2}{t+h}+t^2\frac{t+h-t}{t(t+h)}\\&=\frac{th^2+t^2h}{t(t+h)}=\frac{th(t+h)}{t(t+h)}=h.\end{aligned}$$

Queda como ejercicio ver que es continuo en 0 (esto dá un poco más de trabajo). $\square$

Lema. *$B(t)$ es Browniano standard. Entonces*

$$\lim_t\frac{B(t)}{t}=0,\ \textit{casi seguramente}$$

Dem *Use la inversión temporal para escribir*

$$\lim_t\frac{B(t)}{t}=\lim_t\frac{tB(1/t)}{t}=\lim_t B(1/t)=0\ \textit{casi seguramente},$$

por la continuidad de las trayectorias en 0. $\square$

Observación sobre la convergencia casi segura *La construcción de Levy produce un isomorfismo $\varphi:Z\mapsto B$, donde $Z=(Z_d,\, d\in\mathcal{D})$ es la familia de iid normales standard indexadas por los diádicos y $B=(B(t):t\in[0,1])$ es la trayectoria del Browniano. Si llamamos A al evento "B tiene trayectorias continuas", la afirmación "B tiene trayectorias continuas casi seguramente" es $P(\phi(Z)\in A)=1$. Los dos lemas precedentes se refieren a funciones $\gamma:B\mapsto X$, donde $X=(X(t):t\in[0,1])$ tiene las propiedades* *del Browniano. En particular $\phi^{-1}(X)$ es una colección $\tilde{Z}$ de iid normales standard indexadas por los diádicos; $\tilde{Z} = \phi^{-1}(\gamma(\phi(Z)))$. Así $P(X \in A) = P(\phi(\tilde{Z}) \in A) = P(\phi(Z) \in A) = P(B \in A)$. O sea que propiedades casi seguras para B son también propiedades casi seguras para $\gamma(B)$.*

No Diferenciabilidad

Teorema 127 *Para todo $0 < a < b < \infty$ el movimiento Browniano no es monótono en $[a, b]$, casi seguramente.*

Demostración *Por absurdo. Supongamos que para todo $a \le s < t \le b$ tenemos $B(s) \le B(t)$. Sean $a = a_1 \le \cdots \le a_{n+1} = b$. Cada incremento $B(a_i) - B(a_{i+1})$ tiene que tener el mismo signo. Pero como los incrementos son independientes, la probabilidad que tengan el mismo signo es $2/2^n$, lo que demuestra que la probabilidad que $[a, b]$ sea un intervalo de monotonicidad debe ser cero. Por lo tanto, la probabilidad que un intervalo con extremos racionales sea de monotonicidad es cero. Como todo intervalo con extremos reales distintos contiene un intervalo con extremos racionales distintos, estamos.* $\square$

Vimos que $B(t)/t \to 0$ cuando $t \to \infty$. Es decir que casi seguramente el Browniano es sublineal. Ahora veremos que crece más rápido que $\sqrt{t}$.

Proposición para el movimiento Browniano $B(t)$ vale

$$\limsup_n \frac{B(n)}{\sqrt{n}} = +\infty, \quad \liminf_n \frac{B(n)}{\sqrt{n}} = -\infty \tag{119}$$

Para probar esa proposición necesitamos un teorema érgodico.

Intercambiabilidad *Sean $X = (X_1, X_2, \dots)$ iid. Una* permutación finita *$\pi : \mathbb{N} \to \mathbb{N}$ es una biyección tal que $\pi(n) = n$ para n suficientemente grande.*

Para un evento $A = \{(X_1, X_2, \dots) \in B\}$ y una permutación finita π, defina

$$\pi A := \{(X_{\pi(1)}, X_{\pi(2)}, \dots) \in B\}$$

Decimos que un evento $A = \{(X_1, X_2, \dots) \in B\}$ es intercambiable *para X si A es invariante por permutaciones finitas, es decir $\pi A = A$ para toda permutación π finita.*

Ejemplos:

$A = \{\lim_n (1/n) \sum_{i=1}^n X_i = 0\}$,

$A = \{(X_n, \dots, X_{n+k}) \in C$, *infinitas veces*$\}$.

Aproximación por eventos cilíndricos *Sea $\mathcal{F}_n$ el conjunto de los eventos que dependen de las primeras n coordenadas:*

$$\mathcal{F}_n = \{(X_1, \dots, X_n) \in B : B \in \mathcal{B}_n\}$$

donde $\mathcal{B}_n$ son los Borelianos de $\mathbb{R}^n$. Proposición 2.33 en Breiman dice que todo evento A puede ser aproximado por eventos A_n en $\mathcal{F}_n$ en el sentido

$$P(A \Delta A_n) \to 0.$$

Lema (ley 0-1 de Hewitt-Savage) Sean $X = (X_1, X_2, \dots)$ iid. Si A es intercambiable para X, entonces $P(A) \in \{0, 1\}$.

Dem *Sea A intercambiable y $A_n \in \mathcal{F}_n$ un evento que depende de las primeras n coordenadas que aproxima a A, es decir tal que $P(A_n \Delta A) \to 0$ y, por lo tanto, $P(A_n) \to P(A)$.*

Escribimos $A_n = \{(X_1, \dots, X_n) \in C_n\}$ y definimos $\tilde{A}_n = \{(X_{n+1}, \dots, X_{2n}) \in C_n\}$.

Sea π la permutación que intercambia las primeras n coordenadas con las segundas n, es decir $(\pi(1), \pi(2), \dots) = (n+1, \dots, 2n, 1, \dots, n, 2n+1, \dots)$. Entonces,

$$P(\tilde{A}_n) = P(\pi A_n)$$

y usando $\pi A = A$,

$$P(\tilde{A}_n \Delta A) = P(\pi A_n \Delta \pi A) = P(A_n \Delta A) \to 0$$

Eso implica que $P(A_n \cap \tilde{A}_n) \to P(A)$.

Por otro lado, como X es una sucesión de iid y A_n y $\tilde{A}_n$ dependen de pedazos disjuntos del vector X,

$$P(A_n \cap \tilde{A}_n) = P(A_n)P(\tilde{A}_n) \to P(A)^2$$

Concluímos que $P(A) \in \{0,1\}$. $\square$

Dem de la proposición *Fije $c > 0$ y observe que por el Lema de Fatou,*

$$P(B(n) > c\sqrt{n},\ i.v.) \geq \limsup_n P(B(n) > c\sqrt{n}) \tag{120}$$

($\limsup_n P(A_n) = \limsup_n E1_{A_n} \leq E \limsup_n 1_{A_n} = E1_{\limsup_n A_n} = P(A_n\ i.v.)$)

Como $B(n) \sim$ Normal$(0,n)$, tenemos

$$P(B(n) > c\sqrt{n}) = P(B(1) > c) > 0. \tag{121}$$

Sea $X_n = B(n) - B(n-1)$ y note que

$$\{B(n) > c\sqrt{n},\ i.v.\} = \Big\{\sum_{i=1}^{n} X_i > c\sqrt{n},\ i.v.\Big\}$$

es un evento intercambiable, con probabilidad positiva por (120) y (121). Hewitt-Savage implica que tiene probabilidad 1. Tomando la intersección sobre todos los enteros c, probamos el $\limsup$ en (119). El otro límite se prueba igual. $\square$

Observación *La sigma álgebra intercambiable no es equivalente a la sigma álgebra terminal: por ejemplo, para variables X_i tomando valores en $\{-1,1\}$ el evento $A := \{X_1 + \cdots + X_n = 0,\ i.v.\}$ es intercambiable pero no está en la sigma álgebra terminal: la sucesión $\{1,-1,1,-1,\dots\} \in A$ pero si cambiamos la primera coordenada, la sucesión $\{-1,-1,1,-1,\dots\} \notin A$.*

Defina las derivadas superior e inferior de una función f por

$$D^* f(t) := \limsup_{h \searrow 0} \frac{f(t+h) - f(t)}{h},$$

$$D_* f(t) := \liminf_{h \searrow 0} \frac{f(t+h) - f(t)}{h}$$

Teorema Para cada $t \geq 0$ el movimiento Browniano no es diferenciable en t casi seguramente. Además $D^*B(t) = +\infty$ y $D_*B(t) = -\infty$.

Dem *Sea B Browniano standard. Defina $X(t) = tB(1/t)$ que también es Browniano standard. Entonces*

$$\begin{aligned} D^*X(0) &\geq \limsup_n \frac{X(1/n) - X(0)}{1/n} \geq \limsup_n \sqrt{n}X(1/n) \\ &= \limsup_n \frac{B(n)}{\sqrt{n}} = \infty \end{aligned}$$

por la proposición.

Analogamente, $D_*X(0) = -\infty$, por lo que X no es diferenciable en 0.

Sea $(B(t) : t \geq 0)$ un Browniano standard y defina $Y(s) = B(t+s) = B(t)$ que es un Browniano standard. No diferenciabilidad de X en cero es equivalente a no diferenciabilidad de B en t. $\square$

Convergencia del paseo aleatorio al Movimiento Browniano

Empezamos con un lema útil. Ya lo vimos en el caso del paseo aleatorio.

Lema 128 *Sea $B(t)$ movimiento Browniano y $a < 0 < b$. Sea $T = \text{ínf}\{t > 0 : B(t) \notin (a, b)\}$. Entonces*

$$P(B(T) = a) = \frac{b}{|a| + b}, \quad P(B(T) = b) = \frac{|a|}{|a| + b}$$
$$ET = |a|b.$$

Demostración

Veamos primero que la esperanza ET es finita.

$$\begin{aligned} P(T > j) &= P(|B(s)| < |a| \vee b, \text{ para todo } s \in [0, j]) \\ &= P(|B(i)| < |a| \vee b, \text{ para todo } i \in \{1, \dots, j\}) \\ &\le P(B|(j) - B(j-1)| < |a| \vee b) \dots P(B(1) - B(0) < |a| \vee b) \\ &= [P(|Z| < |a| \vee b)]^j \end{aligned}$$

con $Z \sim$ Normal$(0, 1)$. Entonces,

$$ET = \int_0^\infty P(T > t)dt \le \sum_{k=1}^\infty P(T > k) \le \frac{\text{constante}}{P(|Z| > |a| \vee b)} < \infty.$$

Usando el Teorema 113 con la martingala $B(t)$, como ET es finita,

$$0 = EB(T) = aP(B(T) = a) + bP(B(T) = b)$$

y además, $P(B(T) = a) + P(B(T) = b) = 1$, podemos concluir la primera parte.

La segunda parte es análoga al mismo resultado para el paseo aleatorio en la región $\{0, \dots, N\}$: si llamamos T al tiempo de salida de $\{1, \dots, N-1\}$, vimos en (34) que $E_x(T) = x(N - x)$, pero la demostración es más sofisticada. El segundo Lema de Wald enunciado en el Teorema 2.48 en Morters y Peres que dice que si T es un tiempo de parada para $B(t)$ con esperanza finita, entonces

$$ET = E(B(T)^2)$$

de donde se deduce

$$ET = E(B(T)^2) = \frac{a^2 b}{|a| + b} + \frac{|a|b^2}{|a| + b} = |a|b. \quad \square$$

Incrustamiento de Skorohod (embedding)

Teorema de incrustamento de Skorohod *Cualquier variable aleatoria real X centrada con segundo momento finito se puede expresar como el valor del Browniano en un tiempo de parada T que depende de la distribución de X.*

Teorema 129 *Sea $(B(t) : t \ge 0)$ el movimiento Browniano standard. Sea X una variable aleatoria con $EX = 0$ y $EX^2 < \infty$. Entonces existe un tiempo de parada T para $B(t)$ tal que $B(T)$ tiene la misma distribución de X y $ET = EX^2$.*

Ejemplo *Sean $a < 0 < b$ y X con distribución $P(X = a) = b/(b - a)$, $P(X_i = b) = -a/(b - a)$ (es la única variable que toma valores a, b que tiene esperanza 0). Por el teorema anterior, para el tiempo de parada $T = \text{ínf}\{t : B(t) \notin (a, b)\}$, la variable $B(T)$ tiene la misma distribución de X y $ET = -ab$ es finita. Esto demuestra el teorema para esta variable. Hay varias construcciones para construir los tiempos de parada para cualquier variable.*

Paseo rescalado en la trayectoria del Browniano *Vamos a considerar las variables X_i iid con $P(X_i = 1) = 1/2 = P(X = -1)$, es decir $a = -1$, $b = 1$ en el ejemplo anterior.*

El paseo aleatorio es la variable

$$S_k = X_1 + \cdots + X_k$$

y el paseo interpolado es

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)(S_{\lfloor t \rfloor + 1} - S_{\lfloor t \rfloor}).$$

Esta es una función $S \in C[0,\infty)$. Definimos la siguiente sucesión de funciones aleatorias $(S_n^* : n \geq 1)$ con $S_n^* \in C[0,\infty)$ por

$$S_n^*(t) := \frac{S(nt)}{\sqrt{n}}, \quad t \in [0,1].$$

Lema 130 *Sea $(B(t) : t \geq 0)$ el movimiento Browniano standard. Sea X una variable aleatoria con $EX = 0$ y $EX^2 = 1$. Entonces existe una sucesión de tiempos de parada*

$$0 = T_0 \leq T_1 \leq T_2 \leq \dots$$

para $B(t)$ tal que, definiendo $S_n := B(T_n)$,

a) $(B(T_n) :: n \geq 0)$ tiene la distribución de un paseo aleatorio con incrementos X.

b) La sucesión $(S_n^* : n \geq 0)$ dada por $S_n^*(t) = \frac{S(nt)}{\sqrt{n}}$ construída con este paseo aleatorio satisface

$$\lim_{n\to\infty} P\Big(\sup_{0\leq t\leq 1} \Big| \frac{B(nt)}{\sqrt{n}} - \frac{S(nt)}{\sqrt{n}} \Big| > \varepsilon \Big) = 0, \qquad \textit{para todo } \varepsilon > 0.$$

Esto es convergencia en probabilidad en el espacio $C[0,1]$ con la norma del supremo.

Observación: escribiendo $B_n^*(t) = \frac{B(nt)}{\sqrt{n}}$, tenemos que $B_n^* \in C(0,1)$ y $S_n^* \in C(0,1)$. Denotando $\|f\| = \sup_{t\in[0,1]} |f(t)$, tenemos que la convergencia en (b) se puede escribir como

$$\lim_{n\to\infty} P(\|B_n^* - S_n^*\| > \varepsilon) = 0.$$

Pero observe que ni B_n^* ni S_n^* convergen! Sólo la diferencia converge en este lema.

Demostración parcial *Usando en incrustamiento de Skorohod, definimos T_1 el tiempo de parada con $ET_1 = 1$ y $B(T_1)$ con la misma distribución que X. Por la propiedad fuerte de Markov,*

$$(B_2(t) : t \geq 0) := (B(T_1 + t) - B(T_1) : t \geq 0)$$

es el movimiento Browniano y es independiente de lo ocurrido antes de T_1, en particular de $(T_1, B(T_1)$ y por simetría $B(T_1)$ tiene la misma distribución que X_1.

Definimos T_2' para el Browniano B_2 de la misma manera que T_1 para B. Así, $T_1 + T_2$ es un tiempo de parada para B y $B(T_2)$ tiene la misma distribución que $X_1 + X_2$. Iterando, obtenemos $T_1 \leq T_2 \leq \dots$ con $S_n = B(T_n)$ es paseo aleatorio incrustado y $ET_n = n$. Esto prueba el punto (a).

Para el punto (b) hay que usar la ley de grandes números para $T_n = \sum_{i=1}^n (T_i - T_{i-1})$, variables independientes con el primer momento finito:

$$\lim_n T_n = \lim_n \frac{1}{n} \sum_{i=1}^n (T_i - T_{i-1}) = 1.$$

Los detalles están en el Lema 5.24 de Morters y Peres. □

Principio de invariancia de Donsker

Teorema 131 *En el espacio $C[0,1]$ de la funciones continuas con la norma del supremo, la sucesión*

$$\Big(\Big(\frac{S(nt)}{\sqrt{n}} : t \in [0,1] \Big) : n \geq 1 \Big)$$

converge en distribución al movimiento Browniano $(B(t) : t \in [0,1])$.

La demostración de este teorema es análoga (pero más delicada) que la demostración que convergencia en probabilidad implica convergencia en distribución. Hay que usar el item (b) del lema anterior para probar que para cada conjunto K cerrado de $C[0,1]$,

$$P(S_n^* \in K) \to P(B \in K)$$

usando también que $B_n^* \sim B$, por la invariancia de escala.

Convergencia casi segura de trayectorias por subsucesiones *Definamos*

$$B^n(t) := \sqrt{n}B(t/n)$$

que es un Browniano standard por el lema de cambio de escala. Fijemos n y definamos ahora T_k^n usando la trayectoria $B^n(t)$

$$T_k^n := \text{ínf}\{t > T_{k-1}^n : |B^n(t) - B^n(T_{k-1}^n)| = 1\}$$

Entonces

$$(B^n(T_k^n), k \geq 0) \text{ tiene la misma distribución que } (S_k, k \geq 0)$$

Definamos

$$\tilde{S}^n(t) := B(T_{[nt]}^n/n) + (t - T_{[nt]}^n/n)(B(T_{[(n+1)t]}^n/n) - B(T_{[nt]}^n/n))$$

Por las definiciones tenemos que T_k^n/n es la primera vez que el Browniano $B(t)$ hace k cambios de magnitud $1/\sqrt{n}$:

$$T_k^n = \text{ínf}\{t > T_{k-1}^n : |B(t/n) - B(T_{k-1}^n/n)| = 1/\sqrt{n}\}$$

Para cada n fijo:

$$(\tilde{S}^n(t), t \in [0,1]) \text{ tiene la misma distribución que } \Big(\frac{S(nt)}{\sqrt{n}}, t \in [0,1]\Big)$$

La evolución de esos dos procesos cuando n cambia es muy diferente: mientras que para t fijo $\frac{S(nt)}{\sqrt{n}}$ no converge en probabilidad (ejercicio), $\tilde{S}^n(t)$ converge casi seguramente por subsucesiones a una variable Gaussiana $B(t)$ de varianza t y media 0, como lo demuestra el siguiente teorema:

Teorema Sea** $(B(t),\, t \in [0,1])$ **movimiento Browniano y** $(\tilde{S}^n(t),\, t \in [0,1])$ **como definido arriba en función de** B**. Entonces

$$\lim_{n\to\infty} P\big(\sup_{t\in[0,1]} |\tilde{S}^n(t) - B(t)| > \varepsilon\big) = 0.$$

Es decir** $(B_n(t) : t \in [0,1])$ **converge en probabilidad a** $(B(t) : t \in [0,1])$ **en la norma del supremo. En particular, existe una sucesión** n_j **tal que

$$\lim_j \sup_{t\in[0,1]} |\tilde{S}^{n_j}(t) - B(t)| = 0, \qquad \textit{casi seguramente}.$$

***Dem** Como $T_k - T_{k-1}$ son iid con $E(T_k - T_{k-1}) = 1$, por la ley débil de grandes números tenemos para cada t fijo*

$$\lim_n \frac{T_{[nt]}}{n} = t, \quad \text{en probabilidad.}$$

Si pudieramos probar que para alguna subsucesión

$$W_n := \sup_{t\in[0,1]} \Big|\frac{T_{[nt]}}{n} - t\Big| \to 0, \quad \text{casi seguramente,}$$

entonces, por la continuidad del Browniano, tendriamos para la misma subsucesión

$$\sup_{t\in[0,1]} |B_n(t) - B(t)| \to 0, \quad \text{casi seguramente,}$$

Se puede probar que $W_n \to 0$ en probabilidad, es una versión de la ley del logaritmo iterado. □

Referencias

[1] *Pierre Brémaud.* Markov chains, *volume 31 of* Texts in Applied Mathematics. *Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.*

[2] *Richard Durrett.* Essentials of stochastic processes. *Springer Texts in Statistics. Springer, New York, second edition, 2012.*

[3] *Rick Durrett.* Probability: theory and examples. *Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.*

[4] *Olle Häggström.* Finite Markov chains and algorithmic applications, *volume 52 of* London Mathematical Society Student Texts. *Cambridge University Press, Cambridge, 2002.*

[5] *Jesper Møller and Rasmus Plenge Waagepetersen.* Statistical inference and simulation for spatial point processes, *volume 100 of* Monographs on Statistics and Applied Probability. *Chapman & Hall/CRC, Boca Raton, FL, 2004.*