Proceso estocástico

Un *proceso estocástico* es una familia de variables aleatorias $\{X(t)\}_{t\in T}$, donde T es el *conjunto de índices* o *espacio de parámetro* del proceso. El conjunto de valores tomados por cada variable aleatoria $X(t)$ se lo denomina *espacio de estados* S.

El espacio de parámetros puede ser continuo (por ej., $T = (a, b) \subset \mathbb{R}$) o discreto (por ej., $T = \mathbb{N}$). Por otra parte, el proceso se denomina discreto o continuo de acuerdo a si el espacio de estados S es discreto o continuo.

Para poder describir de manera completa un proceso estocástico, es necesario dar las distribuciones de probabilidad conjuntas de las variables aleatorias $\{X(t_1), X(t_2), \cdots, X(t_n)\}$ para toda tupla $\{t_1, t_2, \cdots, t_n\} \subset T$ y todo $n \in \mathbb{N}$. En el caso de un proceso discreto, se debe poder escribir

$$p(x_1, t_1, x_2, t_2, \cdots, x_n, t_n) = P\big(X(t_1) = x_1, X(t_2) = x_2, \cdots, X(t_n) = x_n\big).$$

En el caso de un proceso continuo, se debe dar la densidad de probabilidad conjunta, la cual podemos escribir como

$$f(x_1, t_1, x_2, t_2, \cdots, x_n, t_n).$$

En general, no es sencilla la descripción completa de un proceso estocástico de esta manera. Por ello se utilizan ciertas simplificaciones que facilitan dicha descripción. Algunas simplificaciones que vemos: procesos estacionarios; procesos con incrementos independientes; procesos de Markov.

En todos los procesos estocásticos es válida la ecuación de Chapman-Kolmogorov:

$$p(x_1, t_1, \cdots, x_{k-1}, t_{k-1}, x_{k+1}, t_{k+1}, \cdots, x_n, t_n) = \sum_{\mathbf{x_k}\in S} p(x_1, t_1, \cdots, x_{k-1}, t_{k-1}, \mathbf{x_k}, \mathbf{t_k}, x_{k+1}, t_{k+1}, \cdots, x_n, t_n)$$

para procesos discretos, y

$$f(x_1, t_1, \cdots, x_{k-1}, t_{k-1}, x_{k+1}, t_{k+1}, \cdots, x_n, t_n) = \int_{-\infty}^{+\infty} f(x_1, t_1, \cdots, x_{k-1}, t_{k-1}, \mathbf{x_k}, \mathbf{t_k}, x_{k+1}, t_{k+1}, \cdots, x_n, t_n)\, d\mathbf{x_k}$$

para procesos continuos.

Proceso estacionario

Se dice que un proceso estocástico $\{X(t)\}_t$ es *estacionario* sii para todo $n \in \mathbb{N}$, toda tupla $\{t_1, t_2, \cdots, t_n\} \subset T$ y todo Δt tal que $\{t_1 + \Delta t, t_2 + \Delta t, \cdots, t_n + \Delta t\} \subset T$ se tiene

$$f(x_1, t_1 + \Delta t, x_2, t_2 + \Delta t, \cdots, x_n, t_n + \Delta t) = f(x_1, t_1, x_2, t_2, \cdots, x_n, t_n)$$

si el proceso es continuo, y

$$p(x_1, t_1 + \Delta t, x_2, t_2 + \Delta t, \cdots, x_n, t_n + \Delta t) = p(x_1, t_1, x_2, t_2, \cdots, x_n, t_n)$$

si es discreto.
Básicamente: $\{X(t)\}_t$ es estacionario si no se lo puede distinguir del proceso $\{X(t+\Delta t)\}_t$.

Incrementos independientes

Se dice que un proceso estocástico $\{X(t)\}_t$ tiene *incrementos independientes* sii para $t_1, t_2, t_3, t_4 \in T$

$$[t_1, t_2) \cap [t_3, t_4) = \emptyset \Rightarrow (X(t_2) - X(t_1)) \text{ y } (X(t_4) - X(t_3)) \text{ son v.a. independientes}$$

Intuitivamente: los cambios del proceso estocástico en partes "separadas" del espacio de índices T son independientes.

Se dice que un proceso estocástico $\{X(t)\}_t$ tiene *incrementos estacionarios* sii para todo $t_1, t_2 \in T$ y para todo Δt tal que $t_1 + \Delta t, t_2 + \Delta t \in T$, las variables aleatorias

$$(X(t_2) - X(t_1)) \text{ y } (X(t_2 + \Delta t) - X(t_1 + \Delta t))$$

tienen la misma distribución de probabilidades.

Procesos de Markov

Se dice que un proceso estocástico $\{X(t)\}_{t\in T}$ es un *proceso de Markov* sii para todo conjunto de valores $t_1 \leq t_2 \leq \cdots \leq t_{n-1} \leq t_n$, $t_i \in T$, y para todo $x_1, x_2, \cdots, x_n \in S$ se cumple

$$f\left(x_n, t_n | x_{n-1}, t_{n-1}, \cdots, x_2, t_2, x_1, t_1\right) = f\left(x_n, t_n | x_{n-1}, t_{n-1}\right)$$

si el proceso es continuo, y

$$p\left(x_n, t_n | x_{n-1}, t_{n-1}, \cdots, x_2, t_2, x_1, t_1\right) = p\left(x_n, t_n | x_{n-1}, t_{n-1}\right)$$

si el proceso es discreto.
Intuitivamente: el proceso depende del estado más reciente y no de toda la *historia*.

En los procesos de Markov, la ecuación de Chapman-Kolmogorov tiene una forma muy simple: si $t_1 \leq t_2 \leq t_3$,

$$p(x_3, t_3 | x_1, t_1) = \sum_{x_2 \in S} p(x_3, t_3 | x_2, t_2) p(x_2, t_2 | x_1, t_1)$$

para procesos discretos, y

$$f(x_3, t_3 | x_1, t_1) = \int_{-\infty}^{+\infty} f(x_3, t_3 | x_2, t_2) f(x_2, t_2 | x_1, t_1)\, dx_2$$

para procesos continuos.

Problema 1 de la guía 6

La caminata aleatoria (random walk) simétrica Supongamos el siguiente juego: se lanza una moneda en forma reiterada y el jugador gana 1 peso si sale cara y pierde 1 peso si sale ceca.

1. Se define el proceso estocástico $\{X_n : n \in \mathbb{N}^0\}$ donde X_n es el dinero que tiene el jugador al cabo de n jugadas (se supone $X_0 = 0$). Definir el recorrido de X_n y el espacio de estados del proceso (válido para todo valor de n).
2. Obtener la distribución de probabilidades de X_n.
3. Expresar la variable X_n como una suma de variables aleatorias independientes e idénticamente distribuidas y calcular $E(X_n)$ y $V(X_n)$.

Nota: Se llama caminata aleatoria porque, en vez de un juego, se puede pensar en la caminata de un individuo que se mueve un paso a la izquierda o a la derecha de acuerdo al resultado de lanzar una moneda.

El siguiente código de *Octave* simula el proceso (se muestran `M` *realizaciones del proceso* y `N` instantes posteriores al inicial):

```
N = 100; M = 10;
for i=1:M
    X(1,i) = 0;
    for k=1:N
        Rk = rand;
        Yk = -(Rk<1/2)+(Rk>=1/2);
        X(k+1,i) = X(k,i)+Yk;
    end
end
plot(X,'o-')
```

Respuesta:
El proceso tiene las siguientes características:

- Se trata de un proceso de Markov. En efecto, el dinero del jugador luego de n jugadas no depende de toda la historia de apuestas, sino solamente del dinero que tiene luego de la jugada $n-1$ y el que gane o pierda en la jugada n:

$$P\big(X_n = x_n \big| X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots, X_1 = x_1\big) = P\big(X_n = x_n \big| X_{n-1} = x_{n-1}\big), \tag{1}$$

  donde asumido que *los lanzamientos de monedas son experimentos independientes.* Más aún,

$$P\big(X_n = x_n \big| X_{n-1} = x_{n-1}\big) = \begin{cases} \frac{1}{2} & \text{si } x_n = x_{n-1} \pm 1, \\ 0 & \text{para todo otro caso.} \end{cases} \tag{2}$$

- Definamos el proceso estocástico $\{Y_n\}_{n\in\mathbb{N}}$ como

$$Y_n = X_n - X_{n-1} \qquad n \in \mathbb{N}. \tag{3}$$

  Es decir, Y_n es la cantidad ganada en la jugada n-ésima. Por lo tanto, $Y_n \in \{-1, +1\}$. Es fácil ver que

$$X_n = \sum_{k=1}^{n} Y_k. \tag{4}$$

Dado que los lanzamientos de monedas son independientes e idénticos, las Y_n son variables aleatorias i.i.d. con

$$P\big(Y_n = \pm 1\big) = \frac{1}{2}, \tag{5}$$

$$E\left[Y_n\right] = (-1)\frac{1}{2} + (+1)\frac{1}{2} = 0, \tag{6}$$

$$\mathrm{Var}\left[Y_n\right] = E\left[Y_n^2\right] - \left(E\left[Y_n\right]\right)^2 = E\left[Y_n^2\right] = (-1)^2\frac{1}{2} + (+1)^2\frac{1}{2} = 1. \tag{7}$$

Luego,

$$E\left[X_n\right] = E\left[\sum_{k=1}^{n} Y_k\right] = \sum_{k=1}^{n} E\left[Y_k\right] = 0, \tag{8}$$

$$\mathrm{Var}\left[X_n\right] = \mathrm{Var}\left[\sum_{k=1}^{n} Y_k\right] = \sum_{k=1}^{n} \mathrm{Var}\left[Y_k\right] = \sum_{k=1}^{n} \mathrm{Var}\left[Y_1\right] = n, \tag{9}$$

donde en la última ecuación hemos usado el hecho que la varianza de una suma de variables aleatorias independientes es igual a la suma de las varianzas de cada variable aleatoria.

- El proceso X_n tiene incrementos independientes. En efecto, sean $i \le j < m \le n \in \mathbb{N}$. Luego

$$X_j - X_i = \sum_{k=i+1}^{j} Y_k, \qquad X_n - X_m = \sum_{l=m+1}^{n} Y_l. \tag{10}$$

Dado que los lanzamientos de las monedas son independientes, $\{Y_{m+1}, Y_{m+2}, \cdots, Y_n\}$ son independientes entre sí y de $\{Y_{i+1}, Y_{i+2}, \cdots, Y_j\}$. Por lo tanto, $(X_n - X_m)$ es independiente de $(X_j - X_i)$.

- El proceso X_n tiene incrementos estacionarios. Sean $i < j, m \in \mathbb{N}$. Luego

$$X_j - X_i = \sum_{k=i+1}^{j} Y_k, \qquad X_{j+m} - X_{i+m} = \sum_{l=i+m+1}^{j+m} Y_l = \sum_{h=i+1}^{j} Y_{l+m}. \tag{11}$$

Dado que las variables Y_k son i.i.d., $\sum_{k=i+1}^{j} Y_k$ tiene la misma distribución que $\sum_{h=i+1}^{j} Y_{l+m}$.

- Llamemos $\{G_n\}_{n\in\mathbb{N}}$ al proceso estocástico

$$G_n = \begin{cases} 1 & \text{si el jugador gana en la } n\text{-ésima jugada,} \\ 0 & \text{si el jugador pierde en la } n\text{-ésima jugada.} \end{cases} \tag{12}$$

Dado que los lanzamientos de monedas son independientes e idénticos, las G_n son variables aleatorias i.i.d. con $G_n \sim \mathrm{Bernoulli}(1/2)$. Es fácil ver que

$$Y_n = 2\,G_n - 1. \tag{13}$$

Por lo tanto,

$$X_n = 2\sum_{k=1}^{n} G_k - n. \tag{14}$$

Definamos

$$H_n = \sum_{k=1}^{n} G_k. \tag{15}$$

Sabemos que la suma de variables aleatorias i.i.d. Bernoulli es una variable aleatoria binomial, es decir, $H_n \sim \text{Binomial}(n, 1/2)$. Por tanto,

$$\begin{aligned} P\big(X_n = x\big) &= P\big(2H_n - n = x\big) = P\big(H_n = \frac{n+x}{2}\big) \\ &= \begin{cases} \binom{n}{\frac{n+x}{2}} \left(\frac{1}{2}\right)^n & \text{si } \frac{n+x}{2} = 0, 1, \cdots, n \\ 0 & \text{en todo otro caso} \end{cases} \\ &= \begin{cases} \binom{n}{\frac{n+x}{2}} \left(\frac{1}{2}\right)^n & \text{si } x = -n, -n+2, \cdots, n-2, n \\ 0 & \text{en todo otro caso.} \end{cases} \end{aligned} \tag{16}$$

Es decir que el recorrido de X_n es $\{-n, -n+2, \cdots, n-2, n\}$.

- A pesar que ya hemos demostrado que tiene incrementos independientes y estacionarios, el proceso X_n no es estacionario. En efecto,

$$P\big(X_2 = 1\big) = 0 \neq \frac{1}{2} = P\big(X_1 = 1\big), \tag{17}$$

por lo que no se cumple que X_{n+h} y X_n tengan la misma distribución ($n, h = 1$ en el ejemplo).

Finalmente, el código en *Octave* se entiende de la siguiente manera:

- `X` será una matriz con `N` filas y `M` columnas que se llenará gracias a los dos lazos `for ... end` entre las líneas **04-08** y **02-09**.
- Cada columna de `X` tendrá una *realización* del proceso estocástico X_n, esto es, una posible historia de juego.
- Cada realización se simula recordando que $X_n = \sum_{k=1}^{n} Y_k$. La línea **07** escribe exactamente esta ecuación, donde `X(k+1,i)` es X_k en la `i`-ésima realización. La diferencia entre k y `k+1` se debe a que *Octave* indexa los arreglos empezando de 1 (ver línea **03**).
- Las líneas **05-06** simulan la ganancia Y_k (`Yk`) en la `i`-ésima realización. La función `rand` de *Octave* simula un experimento en el que se obtiene un número con distribución $U(0, 1)$. Si `Rk`$= R_k \sim U(0, 1)$, entonces

$$P\big(R_k < \frac{1}{2}\big) = P\big(R_k \geq \frac{1}{2}\big) = \frac{1}{2}. \tag{18}$$

Ahora bien:

  - `(Rk<1/2)` da 1 si se cumple la condición entre paréntesis y 0 en caso contrario;
  - `(Rk>=1/2)` da 1 si se cumple la condición entre paréntesis y 0 en caso contrario.

Por lo tanto, `Yk` será -1 con probabilidad $1/2$ y $+1$ con igual probabilidad.

- La línea **10** simplemente grafica las realizaciones. Ver Fig. 1.

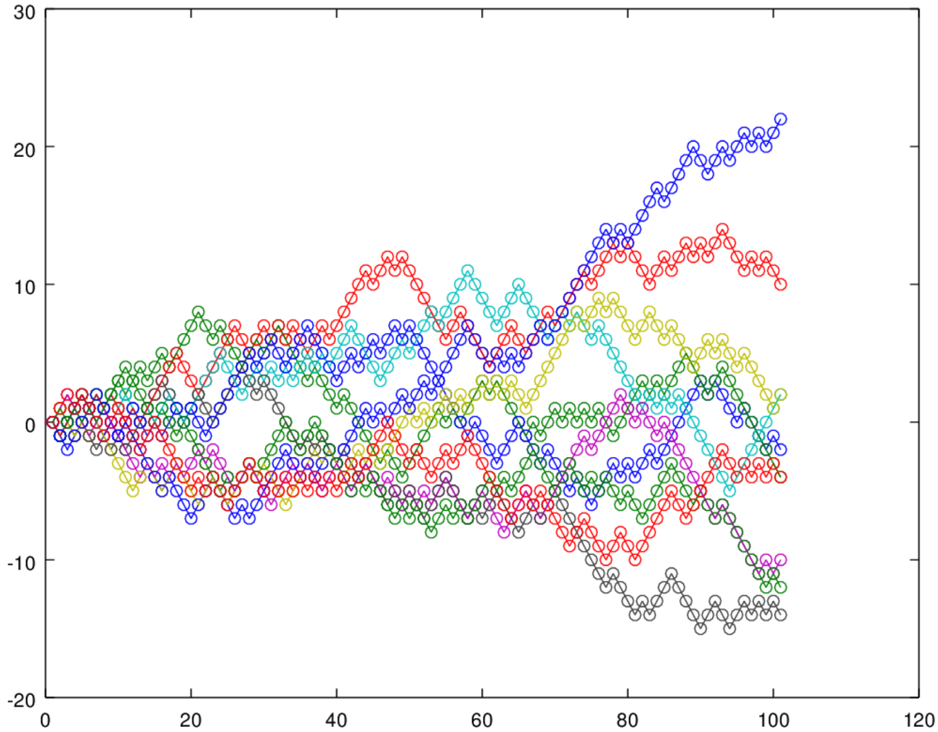


Figura 1: Diez realizaciones del proceso estocástico X_n en el ejercicio 1 de la guía 6. El eje de las abscisas corresponde a $n-1$, el índice. Las ordenadas corresponden a la cantidad ganada por el jugador (\$). Cada realización está graficada con un color diferente.

Proceso de conteo

Considere un proceso estocástico $\{T(k)\}_{k\in\mathbb{N}^0}$ con las siguientes características:

1. $T_0 = 0$.
2. $k \leq m \Rightarrow T_k \leq T_m$.

Este tipo de proceso estocástico puede utilizarse para modelar los instantes de ocurrencia de ciertos *eventos*, como ser, arribos de paquetes en una red informática, llegada de clientes en una cola, emisión de una partícula, etc. El espacio de estadios puede ser discreto o continuo.
A cada proceso que marca los instantes de ocurrencia de eventos, se le puede asociar un *proceso de conteo* $\{N(t)\}_{t\geq 0}$ definido como:

$$N(t) = \max\{k : T_k \leq t\}.$$

Es decir: $N(t)$ cuenta la cantidad de eventos hasta el instante t. Se puede observar que $N(t)$ tiene las siguientes características:

1. $N(0) = 0$.
2. $N(t) \in \mathbb{N}^0$.
3. $s \leq t \Rightarrow N(s) \leq N(t)$.
4. $N(t) - N(s)$ es la cantidad de eventos en el intervalo $(s, t]$.

Proceso de Poisson

Un proceso estocástico $\{N(t)\}_{t\in\mathbb{R}^{\geq 0}}$ es un *proceso de Poisson* con tasa $\lambda > 0$ sii es un proceso de conteo que satisface las siguientes condiciones:

1. Tiene incrementos independientes.
2. Los incrementos son estacionarios.
3. La probabilidad de que exactamente un evento ocurra en un intervalo de longitud Δt es $\lambda\Delta t + o(\Delta t)$.
4. La probabilidad de que más de un evento ocurra en un intervalo de longitud Δt es $o(\Delta t)$.

La condición 3 dice que λ es la *tasa de generación de eventos*. El proceso es el mismo se lo mire en el intervalo en que se lo mire: esto está garantizado por la condición 2. La condición 1 impide que haya reglas para "secuencias de rachas" del estilo "a muchos eventos en un intervalo, siguen muchos eventos en el intervalo siguiente". Finalmente, la condición 4 nos dice que no se dan dos eventos simultáneamente.
A partir de la definición, se puede verificar que $N(t)$ es un proceso de Markov. Más aún, se puede demostrar que $N(t) \sim \text{Poisson}(\lambda t)$, es decir,

$$P\big(N(t) = k\big) = \frac{(\lambda t)^k}{k!}e^{-\lambda t} \qquad k = 0, 1, 2, \cdots$$

Sea $\{T(k)\}_{k\in\mathbb{N}^0}$ el proceso estocástico que modela los instantes de ocurrencia de eventos asociados al proceso de Poisson $\{N(t)\}_{t\in\mathbb{R}^{\geq 0}}$. También se puede demostrar que, para todo $k \in \mathbb{N}^0$, $(T(k+1) - T(k))$ son variables aleatorias i.i.d. con distribución exponencial de parámetro λ. Es decir, los tiempos entre eventos son variables exponenciales independientes.

Problema 8 de la guía 6

Suponga que en cierto banco se atiende, en promedio durante una parte del día, a cuatro clientes cada seis minutos según un proceso de Poisson. Calcular la probabilidad de que:

1. puedan atenderse a seis o más clientes en seis minutos;
2. se empleen más de tres minutos en atender a un cliente;
3. el tiempo de atención a un cliente esté comprendido entre dos y cuatro minutos;
4. el tiempo que insuma atender 10 clientes sea menor a 10 minutos.

Respuesta:
Llamemos $N(t)$ al proceso de Poisson que se corresponde con la cantidad de clientes atendida a partir de cierto instante de tiempo que tomamos como punto de partida y llamamos $t = 0$. La tasa del proceso de Poisson es

$$\lambda = \frac{4 \text{ clientes}}{6 \text{ minutos}} = \frac{2}{3}\,\text{min}^{-1}. \tag{19}$$

Luego, $N(6) \sim \text{Poisson}(4)$ y la probabilidad de atender a seis o más clientes en seis minutos es

$$P\big(N(6) \geq 6\big) = 1 - P\big(N(6) \leq 5\big) = 1 - \sum_{k=0}^{5} \frac{4^k}{k!} e^{-4} \approx 1 - 0.7851 = 0.2149. \tag{20}$$

Definamos T como el tiempo de atención de un cliente. Recordando que el tiempo entre eventos de un proceso Poisson tiene una distribución exponencial cuyo parámetro es igual a la tasa del proceso, tenemos que $T \sim \text{Expo}(2/3)$, siendo T medido en minutos. Por lo tanto, para las preguntas 2 y 3 tenemos:

$$P\big(T > 3\big) = 1 - F_T(3) = 1 - \left(1 - e^{-\frac{2}{3}\cdot 3}\right) = e^{-2} \approx 0.1353, \tag{21}$$

$$\begin{aligned} P\big(2 < T < 4\big) = P\big(2 < T \leq 4\big) &= F_T(4) - F_T(3) = \left(1 - e^{-\frac{2}{3}\cdot 4}\right) - \left(1 - e^{-\frac{2}{3}\cdot 2}\right) \\ &= e^{-\frac{4}{3}} - e^{-\frac{8}{3}} \approx 0.1941. \end{aligned} \tag{22}$$

Para la última pregunta, llamemos T_k al tiempo de atención de k clientes. Es fácil ver que (mirar la definición de proceso de conteo)

$$T_k < t \Leftrightarrow N(t) \geq k. \tag{23}$$

Por lo tanto,

$$\begin{aligned} P\big(T_{10} < 10\big) = P\big(N(10) \geq 10\big) &= 1 - P\big(N(10) \leq 9\big) = 1 - \sum_{k=0}^{9} \frac{\left(\frac{20}{3}\right)^k}{k!} e^{-\frac{20}{3}} \\ &\approx 1 - 0.8626 = 0.1373. \end{aligned} \tag{24}$$

Cadenas de Markov

Las *cadenas de Markov* son procesos de Markov en los cuales el espacio de estados es discreto: $S = \{s_1, s_2, s_3, \cdots\}$. Nosotros nos concentraremos en cadenas cuyo espacio de parámetro o conjunto de índices es también discreto: $T = \mathbb{N}^0$. Una cadena de Markov $\{X(n)\}_{n\in\mathbb{N}^0}$ queda completamente descripta por

1. La distribución de probabilidades del estado inicial:

$$p_j(0) = P\big(X(0) = s_j\big) \qquad \forall s_j \in S.$$

   Es claro $\sum_j p_j(0) = 1$.

2. Las probabilidades de transición entre cada par de estados en cada instante de tiempo:

$$p_{ij}(n) = P\big(X(n+1) = s_j | X(n) = s_i\big) \qquad \forall s_i, s_j \in S.$$

   Dado que la cadena de Markov debe pasar de un estado a otro en cada instante de tiempo, $\sum_j p_{ij}(n) = 1$ para todo i.

A partir de estos datos y, utilizando la ecuación de Chapman-Kolomogorov, se puede encontrar la distribución de probabilidades de $X(n)$ para todo n. En efecto, por la ecuación de Chapman-Kolmogorov

$$p_j(n+1) = \sum_i p_{ij}(n)p_i(n).$$

Una forma compacta de escribir la ecuación de Chapman-Kolmogorov es introduciendo la matriz de probabilidades de transición

$$\mathbb{P}(n) = \begin{pmatrix} p_{11}(n) & p_{12}(n) & \cdots \\ p_{21}(n) & p_{22}(n) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Las filas de esta matriz suman 1. Si definimos el vector de probabilidades de estado $\vec{p}(n) = \begin{pmatrix} p_1(n) & p_2(n) & \cdots \end{pmatrix}(n)$, luego

$$\vec{p}(n+1) = \vec{p}(n)\mathbb{P}(n) = \vec{p}(0)\prod_{k=0}^{n}\mathbb{P}(k),$$

donde hemos usado inducción en la última igualdad.
Se dice que la cadena de Markov es *homogénea* si $p_{ij}(n) = p_{ij}(k)$ para todo k, n; es decir, las probabilidades de transición no cambian con el tiempo. En este caso, $\mathbb{P}$ es constante y

$$\vec{p}(n+1) = \vec{p}(0)\mathbb{P}^{n+1}.$$

Tipos de estados en cadenas homogéneas

Si $P\big(X(n+m)=s_j|\,X(n)=s_i\big)>0$ para algún m, entonces se dice que el estado s_j es *accesible* desde el estado s_i. Si todos los estados de una cadena de Markov son accesibles de todos los otros estados, se dice que la cadena de Markov es *irreducible*.
Un estado s_i es *absorbente* si $p_{ii}=1$. Un conjunto de estados C es *cerrado* si la probabilidad de salir a estados fuera de C es cero.
Definamos $q_{ij}(n)$ como la *probabilidad de primer transición* del estado s_i al estado s_j en n pasos, es decir,

$$q_{ij}(n)=P\big(X(n)=s_j,\,X(m)\neq s_j \text{ para } 0<m<n|\,X(0)=s_i\big).$$

Por otro lado, la *probabilidad de primer transición* del estado s_i al s_j está dada por $q_{ij}=\sum_n q_{ij}(n)$. A partir de esto, surge la siguiente clasificación de estados:

- Un estado s_j es persistente si $q_{jj}=1$ (seguro que la cadena regresa al mismo estado).
  - Un estado persistente es nulo si el *tiempo medio de recurrencia* es infinito: $\sum_n nq_{jj}(n)=\infty$.
  - Un estado persistente es no-nulo si el *tiempo medio de recurrencia* es finito: $\sum_n nq_{jj}(n)<\infty$.
- Un estado s_j es *transitorio* si $q_{jj}<1$ (no es seguro que la cadena regrese al mismo estado).
- Un estado s_j es *periódico* con período T_0 si $P\big(X(n)=s_j\big)X(0)=s_j=0$ salvo para $n=kT_0$.

Teorema: Si una cadena de Markov es irreducible, entonces todos los estados son del mismo tipo: todos transitorios, todos persistentes nulos o todos persistentes no-nulos. Todos los estados son o aperiódicos o periódicos con el mismo período.

Teorema: En una cadena de Markov finita (espacio de estados finito), es imposible que todos los estados sean transitorios. Más aún, si es irreducible, entonces todos los estados son todos persistentes no-nulos.

Corolario: Una cadena de Markov finita irreducible tiene una *distribución estacionaria de probabilidades*, es decir, existe

$$\vec{\pi}=\lim_{n\to\infty}\vec{p}(n),$$

y ese límite es independiente de la distribución de probabilidades del estado original $\vec{p}(0)$.

Dado que

$$\vec{\pi}=\lim_{n\to\infty}\vec{p}(n)=\lim_{n\to\infty}\vec{p}(n+1)=\lim_{n\to\infty}\vec{p}(n)\mathbb{P}=\vec{\pi}\mathbb{P}.$$

Es decir, $\vec{\pi}$ puede encontrarse como el autovector a izquierda de la matriz $\mathbb{P}$ correspondiente al autovalor 1.

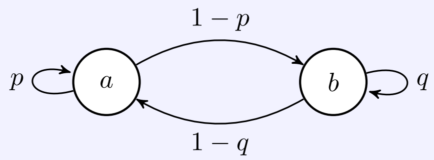
Diagrama de estados de una cadena de Markov

Se trata de un grafo dirigido y con aristas etiquetadas. Cada nodo del grafo se corresponde a un estado posible de la cadena de Markov. La etiqueta de cada arista dirigida indica la probabilidad de transición de un estado hacia otro.
En el ejemplo de la figura siguiente, hay dos estados, a y b. Luego

$$P\big(X(n+1) = a |\, X(n) = a\big) = p, \; P\big(X(n+1) = b |\, X(n) = a\big) = 1 - p,$$

$$P\big(X(n+1) = b |\, X(n) = b\big) = q, \; P\big(X(n+1) = a |\, X(n) = b\big) = 1 - q.$$

Obsérvese que la suma de las etiquetas de todas las aristas (flechas) que salen de un nodo debe ser 1.



Ejercicio 13 de la guía 6

Tres supermercados S_1, S_2 y S_3 compiten por los clientes. Una investigación determina que al comenzar el mes de agosto los tres supermercados tienen igual cantidad de clientes. Al finalizar el mes se observa que:

1. S_1 conserva el 80 % de sus clientes y gana el 10 % y el 2 % de los clientes de S_2 y S_3 respectivamente.
2. S_2 conserva el 70 % de sus clientes y gana el 14 % y el 8 % de los clientes de S_1 y S_3 respectivamente.
3. S_3 conserva el 90 % de sus clientes y gana el 6 % y el 20 % de los clientes de S_1 y S_2 respectivamente.

Sea P es la matriz cuadrada de elementos p_{ij}, donde p_{ij} es la probabilidad de que un cliente del supermercado S_i se pase al supermercado S_j al cabo de un mes.

1. Construya la matriz de transición $\mathbb{P}$, con los datos del problema.
2. Si $\mathbb{P}^{(n)}$ es la matriz cuyo elemento de la posición i, j indica la proporción de clientes que se pasaron del supermercado S_i al S_j al cabo de n meses (bajo el supuesto de que estas proporciones permanecen invariables mes a mes), determine qué porcentaje de clientes se pasaron de S_2 al S_3 al cabo de 2 meses.
3. Sea $\vec{p}(0) = (1/3\ 1/3\ 1/3)$ el vector fila cuyos elementos indican la proporción de clientes que tenía inicialmente cada supermercado. El producto $A\mathbb{P}^{(n)}$ indica la proporción de clientes de cada supermercado al cabo de n meses. Calcule qué proporción de clientes tiene cada supermercado al cabo de un año, dos años, tres años y cuatro años.
4. ¿Qué puede concluir acerca de la proporción de clientes de cada supermercado a largo plazo?
5. Verifique que la respuesta del punto d) también puede obtenerse resolviendo el sistema de ecuaciones $x\mathbb{P} = 1.x$, donde $x = (x_1\ x_2\ x_3)$ y $x_1 + x_2 + x_3 = 1$, es decir, x es un autovector fila a izquierda de la matriz $\mathbb{P}$ correspondiente al autovalor 1 .

Respuesta:

Tomemos una persona al azar de entre el universo de clientes de supermercados. Luego, sea $\{X(n)\}_{n\in\mathbb{N}^0}$ el proceso estocástico que describe de qué supermercado es cliente dicha persona durante el mes n-ésimo. El espacio de estados del proceso es $S = \{s_1, s_2, s_3\}$, donde

$$s_i = [\text{cliente del supermercado } S_i] \qquad i = 1, 2, 3. \tag{25}$$

Dado que, por el enunciado,

$$P\big(X(n+1) = s_j |\, X(n) = s_i, X(n-1) = s_k, \cdots\big) = P\big(X(n+1) = s_j |\, X(n) = s_i\big),$$

sabemos que $X(n)$ es un proceso de Markov. Más aún, como que el espacio de estados S es discreto, se trata de una cadena de Markov. La Fig. 2 muestra el diagrama de estados de la cadena.

A partir del diagrama, es fácil escribir la matriz de probabilidades de transición:

$$\mathbb{P} = \begin{pmatrix} 0.80 & 0.14 & 0.06 \\ 0.10 & 0.70 & 0.20 \\ 0.02 & 0.08 & 0.90 \end{pmatrix}, \tag{26}$$

donde la primer, segunda y tercer fila corresponden a los estados s_1, s_2 y s_3, respectivamente.

La matriz $\mathbb{P}^{(n)}$ de probabilidades de transición en n pasos se puede escribir en función de la matriz de transición $\mathbb{P}$ como

$$\mathbb{P}^{(n)} = \mathbb{P}^n. \tag{27}$$

En particular,

$$\mathbb{P}^{(2)} = \mathbb{P}^2 = \begin{pmatrix} 0.6552 & 0.2148 & 0.1300 \\ 0.1540 & 0.5200 & 0.3260 \\ 0.0420 & 0.1308 & 0.8272 \end{pmatrix}. \tag{28}$$

La fracción de clientes que pasan del supermercado S_2 al S_3 en dos meses, se puede obtener como $\mathbb{P}^{(2)}(2,3) = 0.3260$.

Si los clientes se distribuyen uniformemente entre los tres supermercados al inicio, $\vec{p}(0) = (1/3\ 1/3\ 1/3)$. Luego de un año, la distribución será:

$$\vec{p}(12) = \vec{p}(0)\mathbb{P}^{12} \approx \begin{pmatrix} 0.1906 & 0.2428 & 0.5666 \end{pmatrix}. \tag{29}$$

Y luego de 2 y 3 años:

$$\vec{p}(24) = \vec{p}(0)\mathbb{P}^{24} \approx \begin{pmatrix} 0.1786 & 0.2389 & 0.5825 \end{pmatrix}, \tag{30}$$

$$\vec{p}(36) = \vec{p}(0)\mathbb{P}^{36} \approx \begin{pmatrix} 0.1777 & 0.2386 & 0.5836 \end{pmatrix}. \tag{31}$$

Para hacer estas cuentas, hemos utilizado un programa de computadora. Más abajo mostramos cómo hacerlo de otra manera. Lo que es importante es notar que los valores aparentan converger. De hecho, la cadena de Markov es irreducible y existe un estado estacionario. El mismo puede ser obtenido planteando

$$\vec{\pi} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} = \vec{\pi}\mathbb{P} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} 0.80 & 0.14 & 0.06 \\ 0.10 & 0.70 & 0.20 \\ 0.02 & 0.08 & 0.90 \end{pmatrix}. \tag{32}$$

Esto da a lugar a tres ecuaciones con tres incógnitas:

$$0.80\pi_1 + 0.10\pi_2 + 0.02\pi_3 = \pi_1 \tag{33}$$
$$0.14\pi_1 + 0.70\pi_2 + 0.08\pi_3 = \pi_2 \tag{34}$$
$$0.06\pi_1 + 0.20\pi_2 + 0.90\pi_3 = \pi_3 \tag{35}$$

Obsérvese que si se suman m.a.m. las Ecs. (33)-(35), se obtiene:

$$\pi_1 + \pi_2 + \pi_3 = \pi_1 + \pi_2 + \pi_3. \tag{36}$$

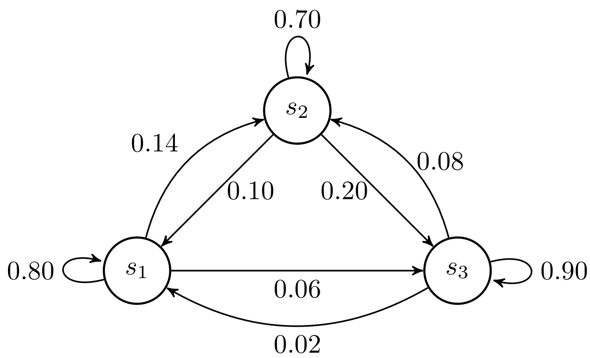


Figura 2: Diagrama de estados de la cadena de Markov del Ejercicio 13 de la guía 6.

Esto implica que las Ecs. (33)-(35) son linealmente dependientes. Para poder obtener una única solución, reemplazamos una de las ecuaciones por la ecuación referente a la normalización del vector $\vec{\pi}$:

$$0.80\pi_1 + 0.10\pi_2 + 0.02\pi_3 = \pi_1 \tag{37}$$
$$0.14\pi_1 + 0.70\pi_2 + 0.08\pi_3 = \pi_2 \tag{38}$$
$$\pi_1 + \pi_2 + \pi_3 = 1 \tag{39}$$

De la Ec. (37), obtenemos

$$\pi_1 = \frac{1}{2}\pi_2 + \frac{1}{10}\pi_3. \tag{40}$$

Reemplazando esta ecuación en la (38),

$$\frac{14}{100}\left(\frac{1}{2}\pi_2 + \frac{1}{10}\pi_3\right) + \frac{70}{100}\pi_2 + \frac{8}{100}\pi_3 = \pi_2 \Rightarrow \pi_2 = \frac{47}{115}\pi_3. \tag{41}$$

Substituyendo esto en la Ec. (40), tenemos

$$\pi_1 = \frac{47}{230}\pi_3 + \frac{1}{10}\pi_3 = \frac{35}{115}\pi_3. \tag{42}$$

Finalmente, substituyendo las Ecs. (41)-(42) en la Ec. (39),

$$\frac{35}{115}\pi_3 + \frac{47}{115}\pi_3 + \pi_3 = \frac{35+47+115}{115}\pi_3 = 1 \Rightarrow \tag{43}$$

$$\Rightarrow \pi_3 = \frac{115}{197} \approx 0.5838,\ \pi_2 = \frac{47}{197} \approx 0.2386,\ \pi_1 = \frac{35}{197} \approx 0.1777. \tag{44}$$

Obsérvese que estos números coinciden bastante bien con los obtenidos cuando se buscó la distribución de probabilidades de luego de 3 años.

Extra: Autovalores y autovectores de la matriz de transición

Asumamos que la matriz $\mathbb{P}$ es diagonalizable. En este caso, la matriz de transición se puede escribir como

$$\mathbb{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \tag{45}$$

donde $\mathbf{V}$ es la matriz de autovectores y $\mathbf{\Lambda}$ es la matriz diagonal de autovalores. Luego,

$$\mathbb{P}^n = \mathbf{V}\mathbf{\Lambda}^n\mathbf{V}^{-1}. \tag{46}$$

Es fácil verificar que uno de los autovalores es igual 1. En efecto, dado que las filas de la matriz de probabilidad de transición suman 1,

$$\mathbb{P}\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}. \tag{47}$$

Por lo tanto, podemos escribir

$$\mathbb{P} = \mathbf{V}\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & 1\end{pmatrix}\mathbf{V}^{-1}, \tag{48}$$

$$\mathbb{P}^n = \mathbf{V}\begin{pmatrix}\lambda_1^n & 0 & 0\\ 0 & \lambda_2^n & 0\\ 0 & 0 & 1\end{pmatrix}\mathbf{V}^{-1}. \tag{49}$$

A partir de estas ecuaciones se puede encontrar $\vec{p}(n)$ para cualquier n. Haciendo algunas cuentas (o utilizando un programa matemático como *Octave*):

$$\vec{p}(n) = \vec{p}(0)\mathbb{P}^n = \vec{p}(0)\mathbf{V} \begin{pmatrix} \left(\frac{7}{10} - \frac{\sqrt{7}}{25}\right)^n & 0 & 0 \\ 0 & \left(\frac{7}{10} + \frac{\sqrt{7}}{25}\right)^n & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{V}^{-1}, \tag{50}$$

donde, por ej.,

$$\mathbf{V} \approx \begin{pmatrix} +0.5152 & -0.9147 & 1 \\ -0.8368 & -0.1908 & 1 \\ +0.1852 & +0.3564 & 1 \end{pmatrix}, \ \mathbf{V}^{-1} \approx \begin{pmatrix} +0.3698 & -0.8590 & +0.4892 \\ -0.6908 & -0.2230 & +0.9138 \\ +0.1777 & +0.2386 & +0.5838 \end{pmatrix}. \tag{51}$$

Cuando $n \to \infty$, tenemos

$$\begin{aligned} \lim_{n\to\infty} \vec{p}(n) &= \vec{p}(0)\mathbf{V} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{V}^{-1} = \vec{p}(0) \begin{pmatrix} * & * & 1 \\ * & * & 1 \\ * & * & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{V}^{-1} \\ &= \begin{pmatrix} * & * & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{V}^{-1} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \mathbf{V}^{-1} \\ &\approx \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} * & * & * \\ * & * & * \\ +0.1777 & +0.2386 & +0.5838 \end{pmatrix} \\ &\approx \begin{pmatrix} +0.1777 & +0.2386 & +0.5838 \end{pmatrix}, \end{aligned} \tag{52}$$

donde hemos usado el hecho que la suma de las componentes de $\vec{p}(0)$ es 1. Obsérvese que hemos llegado a la misma probabilidad estacionaria, aunque por un camino distinto. La clave fue que la matriz $\mathbb{P}$ tiene uno solo autovalor igual a 1 y los demás son menores a 1 en módulo.