Simulación Monte Carlo y Cadenas de Markov

Módulo 4 : Técnicas computacionales avanzadas para modelar fenómenos sociales
Concentración en Economía Aplicada y Ciencia de Datos
ITESM

5 de mayo de 2023

Método de Monte Carlo

Idea: Método para calcular el área bajo una curva. Es una solución estadística al problema de integración.

Método de Monte Carlo

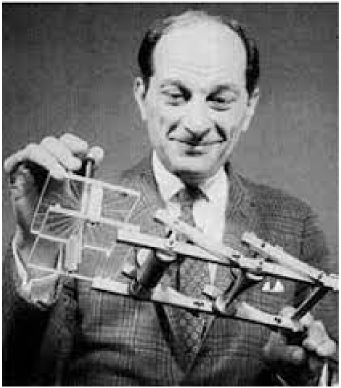


Figura: Stanisław Ulam

Método de Monte Carlo: La legenda (Zwanzig and Mahjani, 2019)

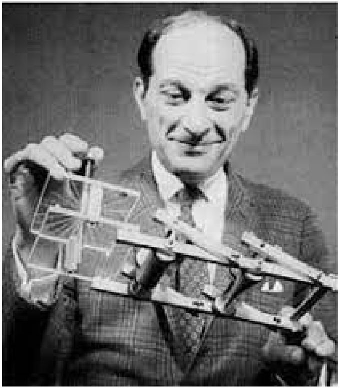


Figura: Stanisław Ulam

Método de Monte Carlo: La legenda(Zwanzig and Mahjani, 2019)

- ¿Cuál es la probabilidad de ganar el juego de solitario?
- Hacer los cálculos por combinatoria era muy largo.
- Decidió calcular la probabilidad al jugar el juego varias veces.
- Jugó 100 veces y contó el número de juegos ganados.
- Ulam estableció la idea que una integral[1] puede verse como **el valor esperado de una función**, y que este puede ser estimado estadísticamente.

[1] La probabilidad de ganar en este caso.

Método de Monte Carlo (MC)

El objetivo de los métodos de Monte Carlo es encontrar el valor de la integral:

$$\mu = \int f(x)dx < \infty$$

Principio general del método de MC

- Encontrar una factorización adecuada para $f(x)$,

$$f(x) = h(x)p(x) \tag{1}$$

  donde $p(x)$ es una densidad. μ puede ser interpretada como un valor esperado

$$\mu = Eh(x) = \int h(x)p(x)dx$$

Principio general del método de MC

- Generar una muestra iid de tamaño arbitrario de $p(x)$, *i.e.* $x_1, x_2, \ldots, x_n$. La aproximación de MC es el estimador muestral de la media que es dada por el promedio

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} h(x_i)$$

Algoritmo MC Independiente

Algorithm 1: MC independiente

1.- Factoriza $f(x) = h(x)p(x)$, donde $p(x)$ es una función de densidad.

2.- Genera una muestra $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ de la distribución $p(x)$

3.- Aproxima μ con el promedio muestral

$$\hat{\mu} = \frac{1}{n}(h(x^{(1)}) + \cdots + h(x^{(n)}))$$

Ejemplo : Algoritmo MC Independiente

Supongamos que queremos encontrar el valor de

$$\mu = \int_0^{0,4} 4x^5 \exp\left(-x^3\right) dx$$

Podemos utilizar el Algoritmo MC Independiente para calcular μ.

Usamos la distribución Weibull con parámetro de escala,λ, 1 y parámetro de forma, α, 3 para factorizar la función. La función de densidad de la distribución Weibull es:

$$\lambda\alpha(\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}$$

$$p(x) = 3x^2 e^{-x^3}, \text{ para } x \geq 0$$

y

$$h(x) = \frac{4}{3}x^3$$

```
meth1 <- function(N){

        x <- rweibull(N, 3, 1) # Generamos N va de la dist Weibull

        z <- x[x<0.4] # Truncamos los valores 0-0.4

        hz <- (4/3) * z^3

        return(sum(hz)/N)
        }

meth1(1000) # Implementamos la aproximacion
```

Método de Monte Carlo : Algoritmo Hit-Miss

Para una función f acotada en $0 \leq f(\theta) \leq M$, y $\theta \in [a, b]$, queremos calcular la integral

$$I = \int_a^b f(\theta) d\theta \tag{2}$$

Dicha gráfica queda inscrita en el rectángulo $R = [a, b] \times [0, M]$.

Método de Monte Carlo

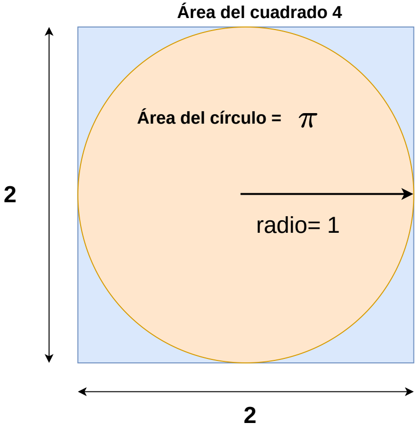
El algoritmo consiste en llenar el rectángulo $R = [a, b] \times [0, M]$ simulando una muestra $(\theta_1, \phi_1), \ldots, (\theta_N, \phi_N)$ de $p(\theta, \phi)$ de una función de densidad de una distribución uniforme sobre el rectángulo R y contando cuántos de estos valores caen bajo la curva $\phi = f(\theta)$.

El estimador $\hat{I}$ obtenido es un estimador insesgado de I.

La varianza del estimador es

$$Var(\hat{I}) = \frac{I}{N}\{M(b-a) - I\}$$

Cálculo de π por el método de Monte Carlo (Wilkinson and Allen, 1998)



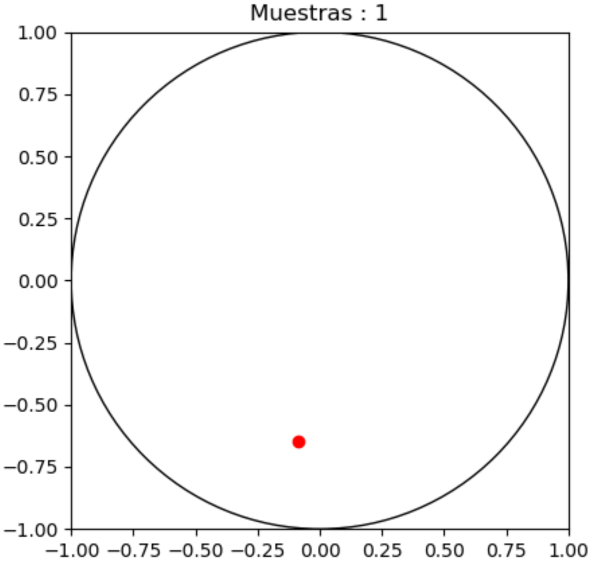
$$\frac{\text{Área del círculo}}{\text{Área del cuadrado}} = \frac{\pi(1)^2}{4} = \frac{\pi}{4}$$

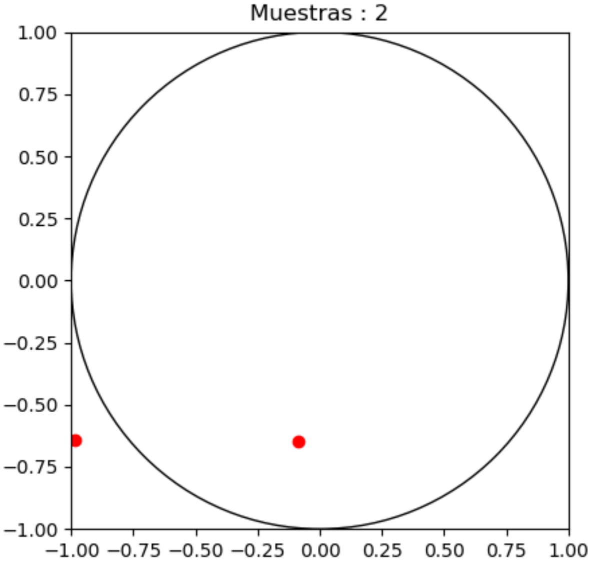
que puede ser descrita por la integral

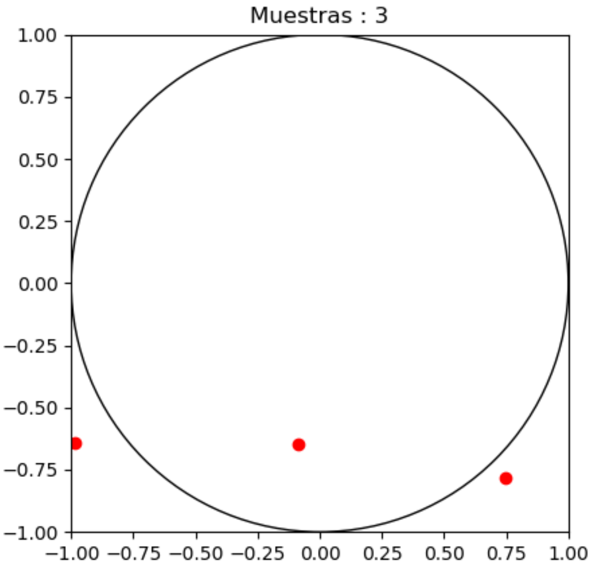
$$\int_0^1 \sqrt{1-x^2}dx = \frac{\pi}{4}$$

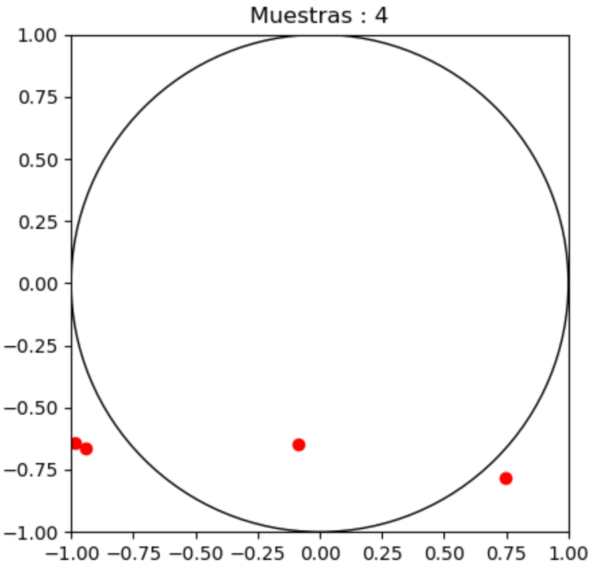
Para calcular π se generan parejas aleatorias de números (x_r, y_r), distribuidos uniformemente entre 0 y 1, y contamos cuántos de estos puntos caen dentro del círculo, esto es, si se cumple la igualdad

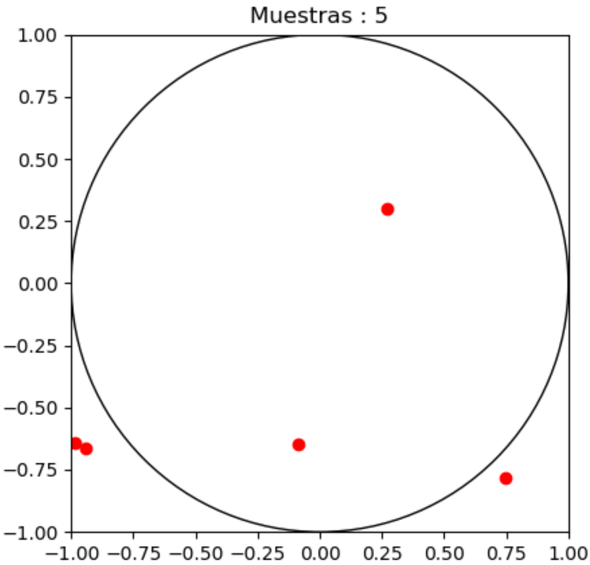
$$y_r^2 + x_r^2 \leq 1$$

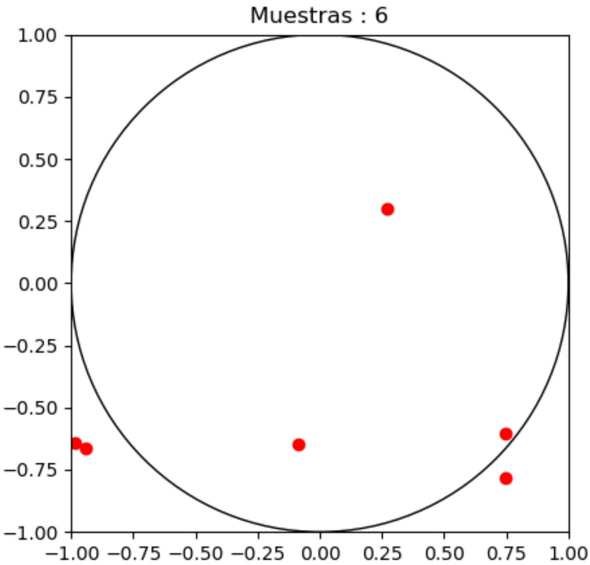

Muestras : 1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

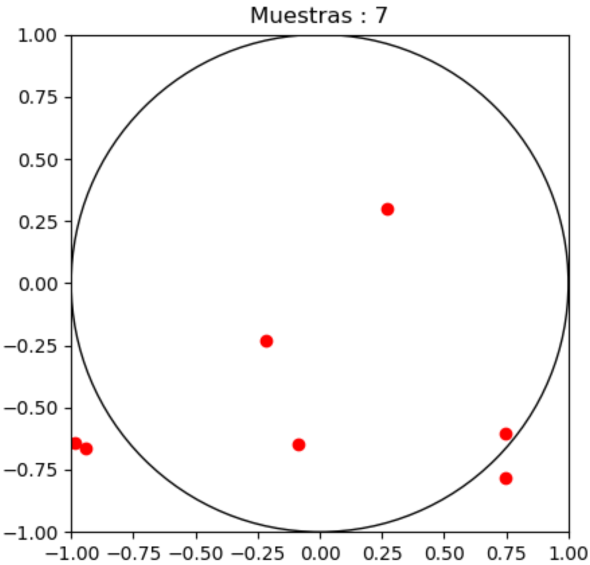

Muestras : 2
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

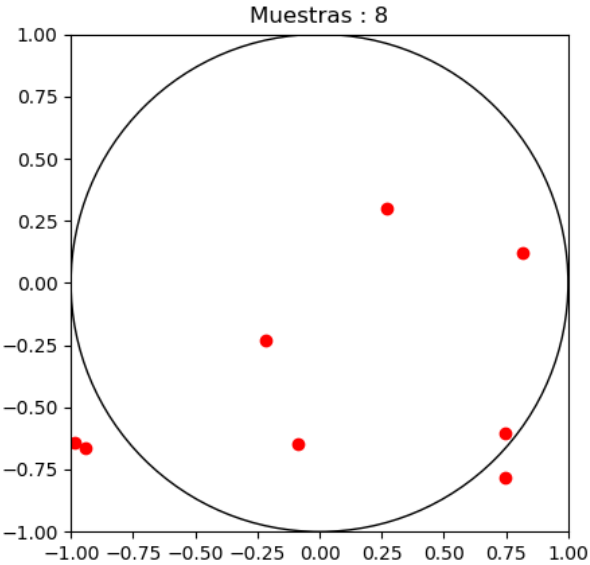

Muestras : 3
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

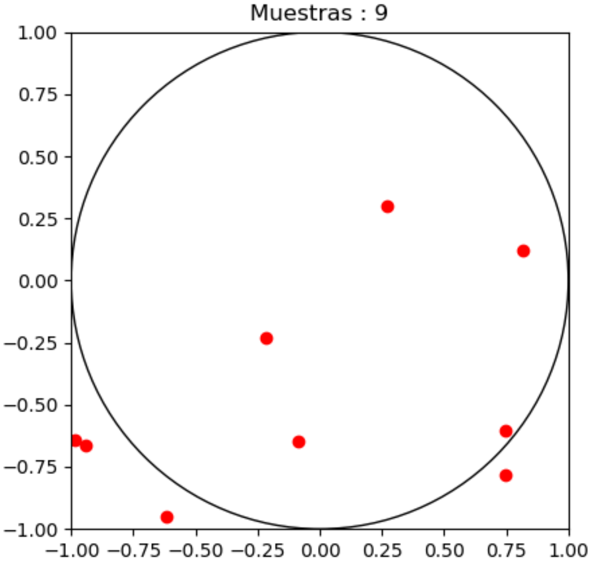

Muestras : 4
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

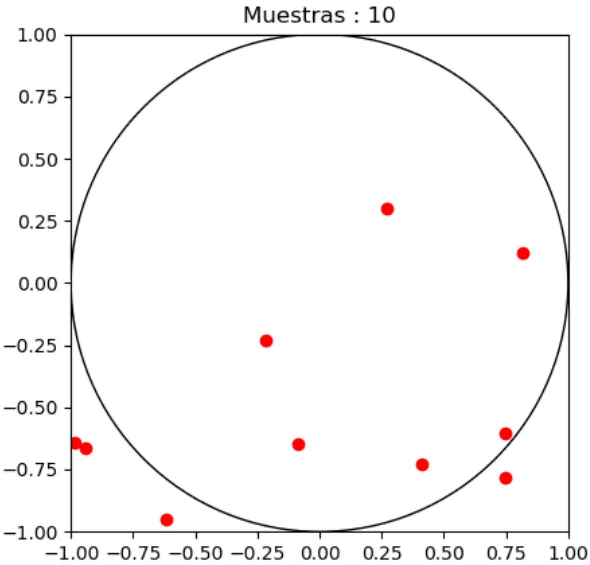

Muestras : 5
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

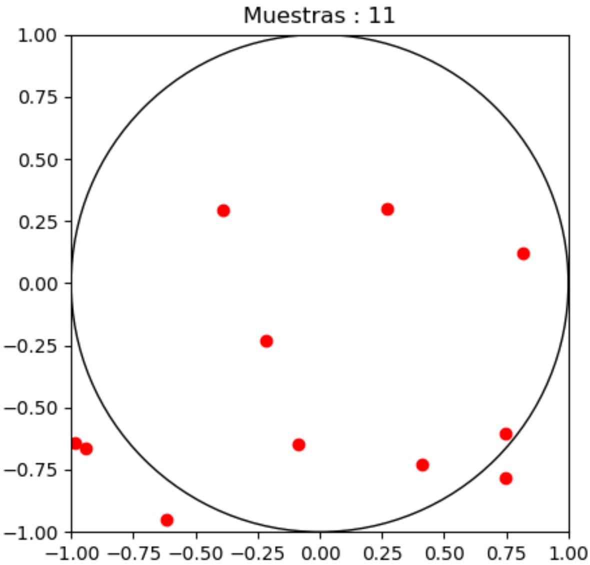

Muestras : 6
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

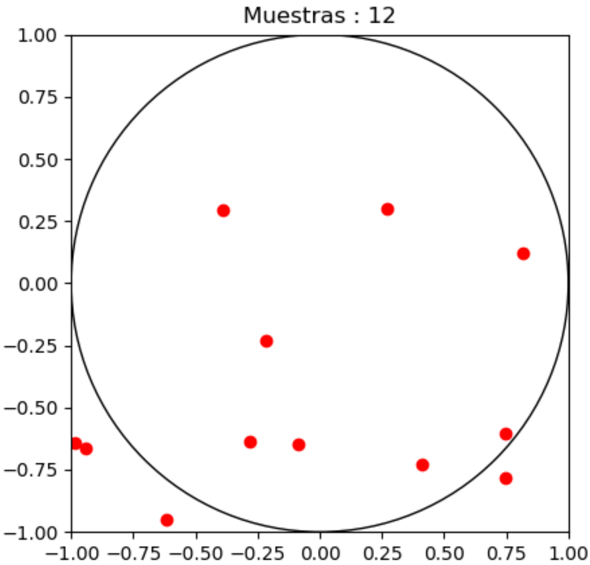

Muestras : 7
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

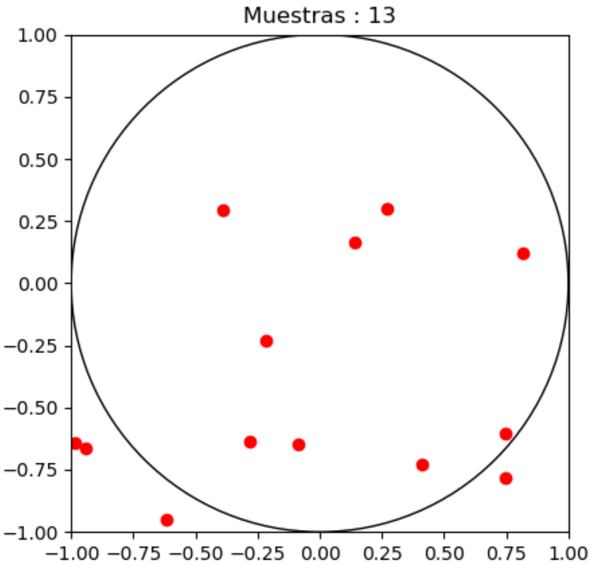

Muestras : 8
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

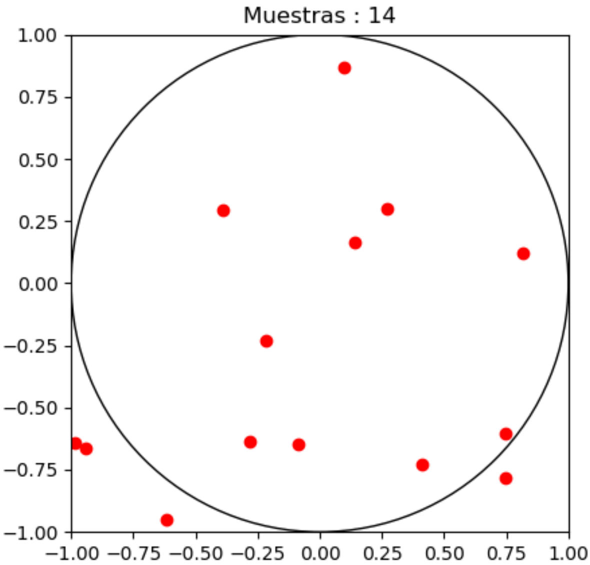

Muestras : 9
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

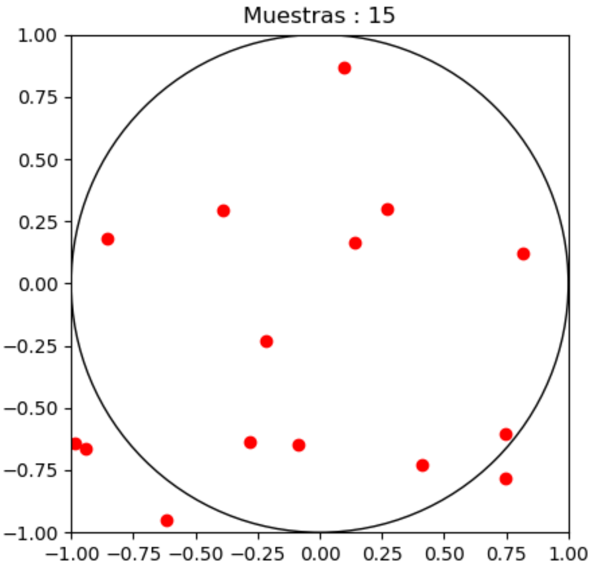

Muestras : 10
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

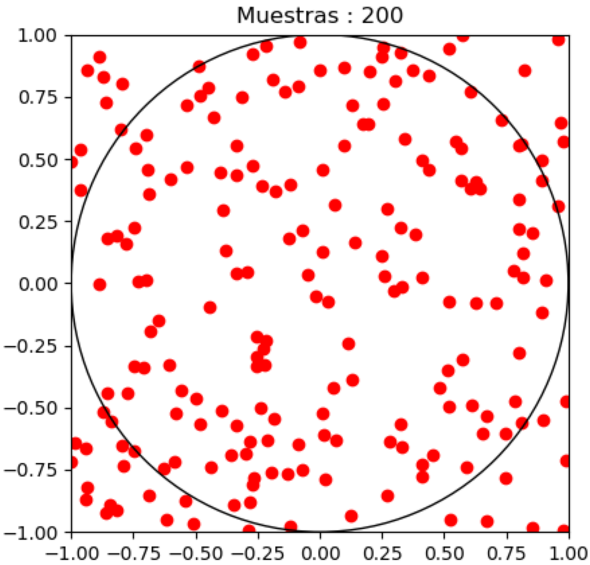

Muestras : 11
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

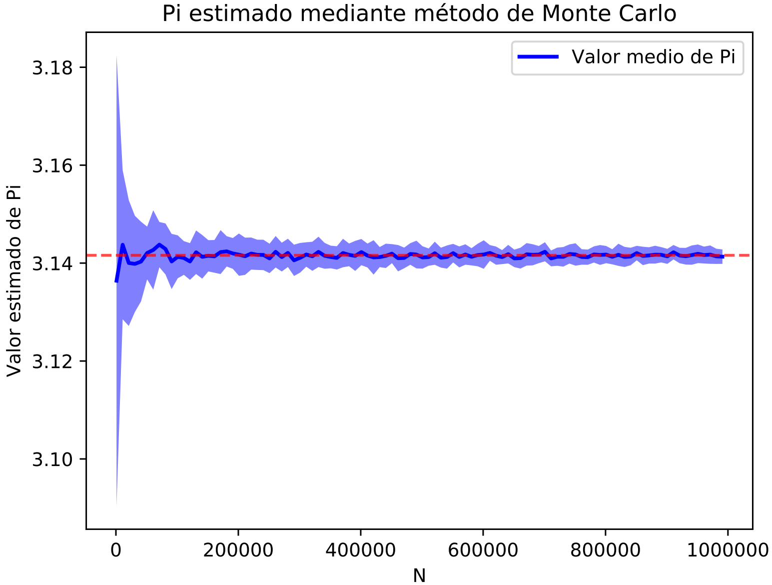

Muestras : 12
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00


Muestras : 13
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00


Muestras : 14
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00


Muestras : 15
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00


Muestras : 200
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00


Pi estimado mediante método de Monte Carlo
Valor medio de Pi
Valor estimado de Pi
3.18
3.16
3.14
3.12
3.10
0
200000
400000
600000
800000
1000000
N

```python
import random
import numpy as np
import matplotlib.pyplot as plt
import math

pi_mean = []
pi_std = []

for n in range(1000,1001000,10000):
    print(n)
    experimentos = []
    for _ in range(30):
        parejas = [(random.random(),random.random()) for i in range
    (n)]

        count = 0

        for x, y in parejas:
          if x * x + y * y < 1:
            count += 1

        pi = (count/n) * 4
        experimentos.append(pi)

    pi_mean.append(np.mean(experimentos))
    pi_std.append(np.std(experimentos))
```

Monte Carlo vía Cadenas de Markov (MCMC)

Consideremos nuevamente el problema de determinar el valor de la integral $\mu = \int f(x)dx$, donde factorizamos $f(x) = h(x)\pi(x)$,

$$\mu = E_\pi h(x) = \int_B h(x)\pi(x)dx, \quad B \subseteq R^d$$

Los métodos de Monte Carlo vía Cadenas de Markov (MCMC) son especialmente útiles para integrales sobre regiones de varias dimensiones, dado que la simulación de variables iid en varias dimensiones es altamente ineficiente.

La integral μ puede interpretarse como un valor esperado con respecto a una distribución estacionaria π de una Cadena de Markov.

Monte Carlo vía Cadenas de Markov (MCMC)

Consideremos nuevamente el problema de determinar el valor de la integral $\mu = \int f(x)dx$, donde factorizamos $f(x) = h(x)\pi(x)$,

$$\mu = E_\pi h(x) = \int_B h(x)\pi(x)dx, \quad B \subseteq R^d$$

Los métodos de Monte Carlo vía Cadenas de Markov (MCMC) son especialmente útiles para integrales sobre regiones de varias dimensiones, dado que la simulación de variables iid en varias dimensiones es altamente ineficiente.

La integral μ puede interpretarse como un valor esperado con respecto a una distribución estacionaria π de una Cadena de Markov.

¿?

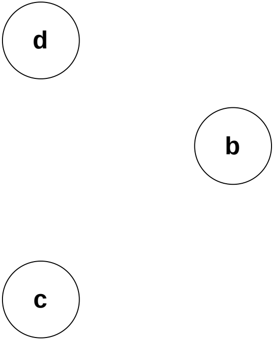
Propiedad de Markov

Definition

Propiedad de Markov. Un proceso estocástico discreto con estados finitos que cumple la propiedad de Markov si,

$$P(X_{t+1} = s | X_1 = s_1, \dots, X_t = s_t) = P(X_{t+1} = s | X_t = s_t)$$

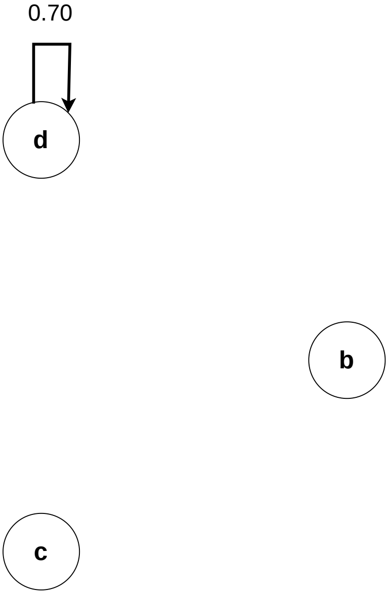
Cadenas de Markov



Estados:

d: Dormitorio
b: Bar
c: Comedor

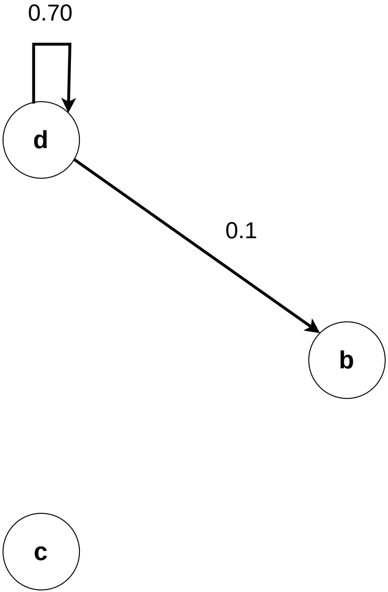
Ejemplo tomado de



La arista representa la posibilidad que un pasajero que está en el dormitorio en el tiempo t continúe en el dormitorio en el tiempo $t + 1$.

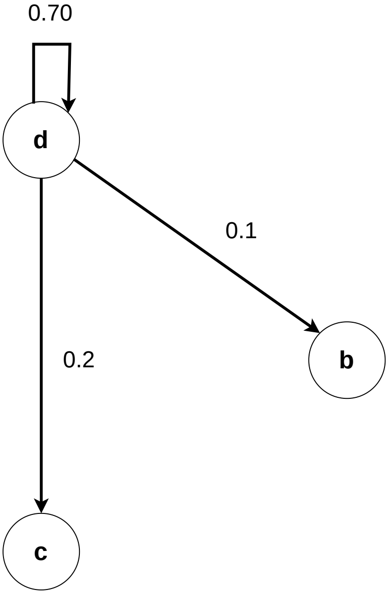
Se etiqueta con la probabilidad de que esto suceda (**Probabilidad de transición**)

$$P(X_{t+1} = d \mid X_t = d) = 0{,}7$$



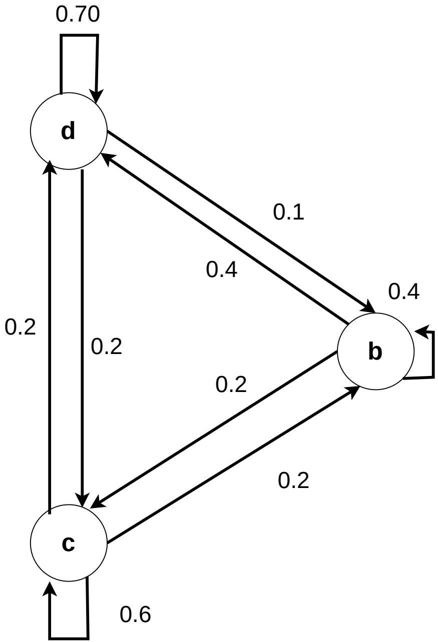
Arista para la transición del dormitorio al bar

$$P(X_{t+1} = b \mid X_t = d) = 0{,}1$$

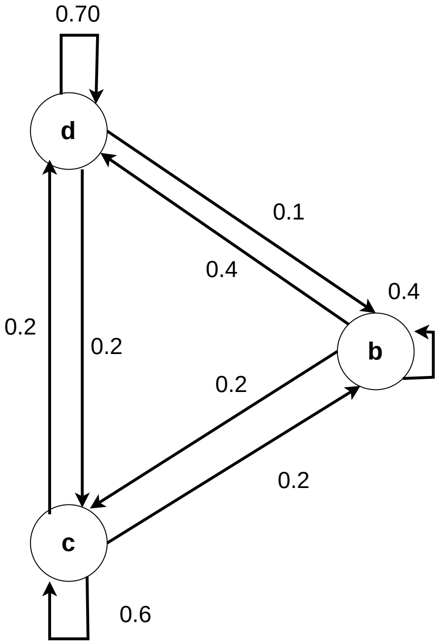


Arista para la transición del dormitorio al comedor

$$P(X_{t+1} = c \mid X_t = d) = 0{,}2$$



Agregamos todas las transiciones posibles y sus probabilidades.

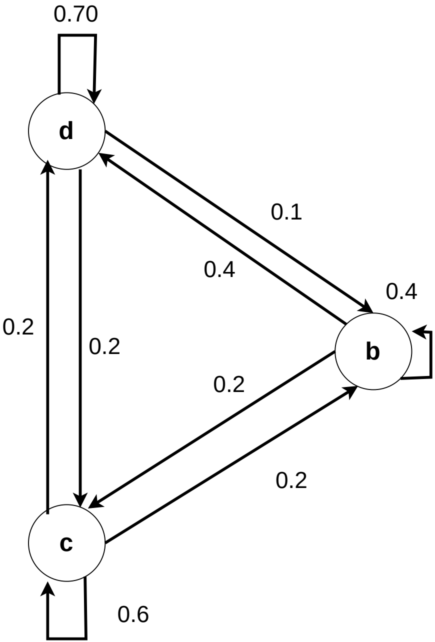


Las probabilidades de transición pueden representarse de manera matricial:

$$T = \begin{bmatrix} 0{,}7 & 0{,}1 & 0{,}2 \\ 0{,}4 & 0{,}4 & 0{,}2 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{bmatrix} \tag{3}$$

con entradas

$$p_{i,j} \equiv P(X_{t+1} = i \mid X_t = j)$$



Matriz de probabilidad de estado:

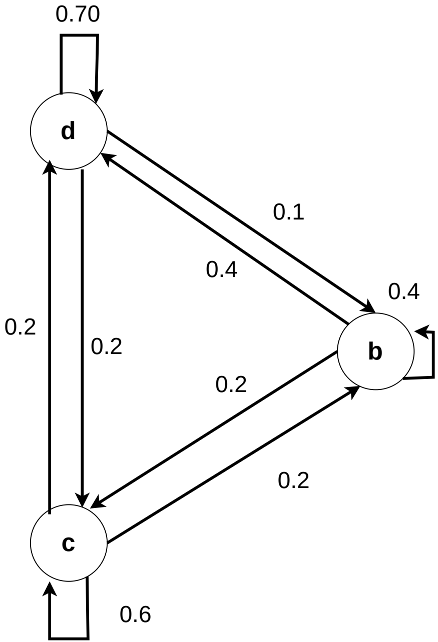
$$T = \begin{bmatrix} 0{,}7 & 0{,}1 & 0{,}2 \\ 0{,}4 & 0{,}4 & 0{,}2 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{bmatrix}$$

La dinámica para la distribución de probabilidad puede expresarse[a]

$$\begin{bmatrix} P(X_{t+1} = d) \\ P(X_{t+1} = b) \\ P(X_{t+1} = c) \end{bmatrix} = \begin{bmatrix} P(X_t = d) \\ P(X_t = b) \\ P(X_t = c) \end{bmatrix} T \tag{4}$$

$$P(X_{t+1}) = P(X_t)T \tag{5}$$

[a] $P(X_t)$ es un vector fila.



Si sabemos que el pasajero se encuentra en el dormitorio al tiempo $t = 0$:

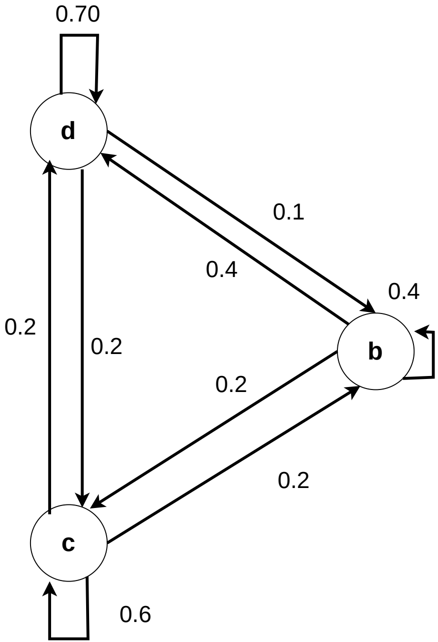
$$P(X_0) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = P_0$$

Al tiempo siguiente la distribución es:

$$P(X_1) = P_0 T = P_1$$

$$P(X_1) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0{,}7 & 0{,}1 & 0{,}2 \\ 0{,}4 & 0{,}4 & 0{,}2 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{bmatrix}$$

$$P(X_1) = \begin{bmatrix} 0{,}7 \\ 0{,}1 \\ 0{,}2 \end{bmatrix}$$



Para tiempos futuros

$$P(X_t) = P(X_{t-1})T \equiv P_t$$

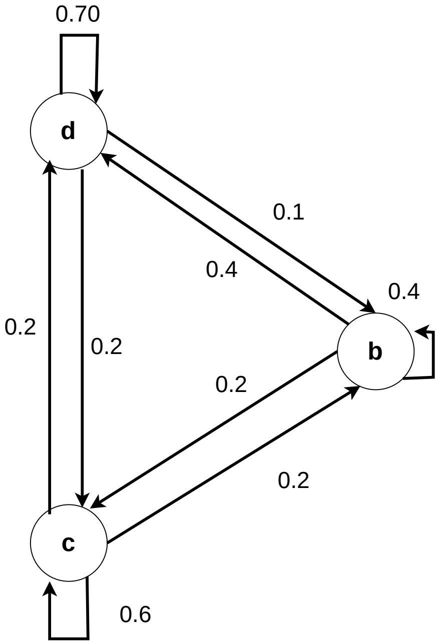
$$P1 = P_0 T$$

$$P2 = P_1 T$$

$$\begin{bmatrix} 0{,}7 \\ 0{,}1 \\ 0{,}2 \end{bmatrix} \begin{bmatrix} 0{,}7 & 0{,}1 & 0{,}2 \\ 0{,}4 & 0{,}4 & 0{,}2 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{bmatrix}$$

$$= \begin{bmatrix} 0{,}57 \\ 0{,}15 \\ 0{,}28 \end{bmatrix}$$

$$P_0 T T = P_0 T^2$$



Para tiempos futuros

$$P_t = P_0 T^t$$

Teorema Perron-Frobenius

Un proceso de Markov converge a un equilibrio estadístico único si cumple las siguientes condiciones

- **Conjunto finito de estados**: $S = \{1, 2, \ldots, K\}$
- **Reglas de transición fijas**: La matriz de transición no cambia en el tiempo.
- **Ergodicidad (accesibilidad entre estados)**: El sistema puede alcanzar cualquier estado desde cualquier estado a través de una serie de transiciones.
- **No ciclicidad**: El sistema no produce un ciclo a través de una secuencia de estados.

Comportamiento de largo plazo(Dobrow, 2016)

En procesos estocásticos y determinísticos frecuentemente nos interesa el comportamiento de largo plazo del sistema.

Martell recolectó datos diarios sobre riesgo de incendios del Canadian Forest Fire Weather Index en un periodo de 26 años para 15 estaciones climáticas.

Para la región de Ontario construyó una cadena de Markov de cinco estados de los cambios diarios en el indice.

El índice clasifica cada riesgo diario de incendio en el bosque en nil, low, moderate, high y extremo.

Comportamiento de largo plazo(Dobrow, 2016)

La matrix de transición es la siguiente

$$T = \begin{array}{c} \\ \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{pmatrix} 0{,}575 & 0{,}118 & 0{,}172 & 0{,}109 & 0{,}026 \\ 0{,}453 & 0{,}243 & 0{,}148 & 0{,}123 & 0{,}033 \\ 0{,}104 & 0{,}343 & 0{,}367 & 0{,}167 & 0{,}019 \\ 0{,}015 & 0{,}066 & 0{,}318 & 0{,}505 & 0{,}096 \\ 0{,}000 & 0{,}060 & 0{,}149 & 0{,}567 & 0{,}224 \end{pmatrix} \end{array} \tag{6}$$

Comportamiento de largo plazo(Dobrow, 2016)

A los administradores del bosque les interesa conocer el comportamiento de largo plazo. Esto es, la distribución de probabilidad de largo plazo del índice diario.

Con respecto al riesgo particular de cualquier día[2] **¿Cual es la probabilidad de riesgo de un día típico al comienzo de verano?**

Para responder a la pregunta, obtengamos las potencia de la matriz de transición para distintos t.

[2] Esto es, independientemente de la condición inicial o punto de partida.

Comportamiento de largo plazo(Dobrow, 2016)

$$T^2 = \begin{array}{c c} & \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{array}{l} \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} & \begin{pmatrix} 0{,}404 & 0{,}164 & 0{,}218 & 0{,}176 & 0{,}038 \\ 0{,}388 & 0{,}173 & 0{,}212 & 0{,}185 & 0{,}042 \\ 0{,}256 & 0{,}234 & 0{,}259 & 0{,}21 & 0{,}041 \\ 0{,}079 & 0{,}166 & 0{,}304 & 0{,}372 & 0{,}079 \\ 0{,}051 & 0{,}117 & 0{,}277 & 0{,}446 & 0{,}109 \end{pmatrix} \end{array} \quad (7)$$

Comportamiento de largo plazo(Dobrow, 2016)

$$T^3 = \begin{array}{r} \\ \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{pmatrix} 0{,}332 & 0{,}176 & 0{,}235 & 0{,}211 & 0{,}046 \\ 0{,}326 & 0{,}175 & 0{,}235 & 0{,}216 & 0{,}047 \\ 0{,}283 & 0{,}192 & 0{,}247 & 0{,}229 & 0{,}049 \\ 0{,}158 & 0{,}183 & 0{,}28 & 0{,}312 & 0{,}067 \\ 0{,}118 & 0{,}165 & 0{,}286 & 0{,}353 & 0{,}078 \end{pmatrix} \end{array} \tag{8}$$

Comportamiento de largo plazo(Dobrow, 2016)

$$T^5 = \begin{array}{c|c} & \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{array}{l} \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} & \begin{pmatrix} 0{,}282 & 0{,}18 & 0{,}248 & 0{,}239 & 0{,}051 \\ 0{,}279 & 0{,}18 & 0{,}248 & 0{,}241 & 0{,}052 \\ 0{,}273 & 0{,}181 & 0{,}25 & 0{,}244 & 0{,}052 \\ 0{,}235 & 0{,}183 & 0{,}259 & 0{,}266 & 0{,}057 \\ 0{,}217 & 0{,}183 & 0{,}264 & 0{,}277 & 0{,}06 \end{pmatrix} \end{array} \tag{9}$$

Comportamiento de largo plazo(Dobrow, 2016)

$$
T^{10} = \begin{array}{c} \\ \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array}
\begin{array}{c} \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\
\begin{pmatrix}
0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}053 \\
0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\
0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\
0{,}263 & 0{,}181 & 0{,}252 & 0{,}25 & 0{,}054 \\
0{,}262 & 0{,}181 & 0{,}252 & 0{,}251 & 0{,}054
\end{pmatrix} \end{array} \tag{10}
$$

Comportamiento de largo plazo(Dobrow, 2016)

$$T^{17} = \begin{array}{c c} & \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{array}{l} \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} & \begin{pmatrix} 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \end{pmatrix} \end{array} \tag{11}$$

Comportamiento de largo plazo(Dobrow, 2016)

$$T^{18} = \begin{array}{c} \\ \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{pmatrix} 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \\ 0{,}264 & 0{,}181 & 0{,}252 & 0{,}249 & 0{,}054 \end{pmatrix} \end{array} \tag{12}$$

Comportamiento de largo plazo(Dobrow, 2016)

- La evidencia numérica sugiere que las potencias de la matriz de transición están convergiendo a un límite.
- Además, las filas de las matrices, en el límite, son todas las mismas.
- El hecho que las filas de T^{17} son las mismas significa que la probabilidad del índice de incendio, después de 17 días, no depende del nivel actual de riesgo.
- Después de 17 días, el efecto de estado inicial se ha desvanecido, y no tiene efectos en la distribución del indice de incendio.

Comportamiento de largo plazo(Dobrow, 2016)

La distribución de largo plazo del índice de fuego es la siguiente:

Nil	Low	Moderate	High	Extreme
0.264	0.181	0.252	0.249	0.054

Ejercicio 1

Zhang et al. (2011) estudiaron los cambios en la distribución de humedales en Yinchuan, China. Los humedales son considerados entre los ecosistemas más importantes de la tierra.

Zhang et al. (2011) desarrollaron un modelo de Markov para dar seguimiento a los cambios anuales entre tipos de humedales.

Ejercicio 1

Utilizando imágenes y datos satelitales de los años 1991, 1999 y 2006, estimaron las distribuciones anuales de tipos de humedales en la región y estimaron la siguiente matriz de transición:

$$T = \begin{array}{c} \\ \text{River} \\ \text{Lake} \\ \text{Pond} \\ \text{Paddy} \\ \text{Non} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{River} & \text{Lake} & \text{Pond} & \text{Paddy} & \text{Non} \end{array} \\ \begin{pmatrix} 0{,}342 & 0{,}005 & 0{,}001 & 0{,}020 & 0{,}632 \\ 0{,}001 & 0{,}252 & 0{,}107 & 0{,}005 & 0{,}635 \\ 0{,}000 & 0{,}043 & 0{,}508 & 0{,}015 & 0{,}434 \\ 0{,}001 & 0{,}002 & 0{,}004 & 0{,}665 & 0{,}328 \\ 0{,}007 & 0{,}007 & 0{,}007 & 0{,}025 & 0{,}954 \end{pmatrix} \end{array} \tag{13}$$

El estado *Non* se refiere a regiones que no son humedales.

Ejercicio 1

- Utiliza la matriz de transición para encontrar la distribución estacionaria de los estados de humedales.
- ¿En cuántos años la distribución de humedales estará en estado estacionario?

Ejercicio 2

La siguiente matriz de transición representa la probabilidad de cambio de alcaldía de acuerdo a la encuesta intercensal de 2015.

Se construyó identificando los flujos de personas que en 2010 vivían en una alcaldía i y que en 2015 vivían en una alcaldía j.

$$T = \begin{array}{cc} & \begin{array}{cccccccccccccccc} 9002 & 9003 & 9004 & 9005 & 9006 & 9007 & 9008 & 9009 & 9010 & 9011 & 9012 & 9013 & 9014 & 9015 & 9016 & 9017 \end{array} \\ \begin{array}{c} 9002 \\ 9003 \\ 9004 \\ 9005 \\ 9006 \\ 9007 \\ 9008 \\ 9009 \\ 9010 \\ 9011 \\ 9012 \\ 9013 \\ 9014 \\ 9015 \\ 9016 \\ 9017 \end{array} & \begin{pmatrix} {,}974 & {,}002 & {,}0 & {,}006 & {,}001 & {,}002 & {,}0 & {,}0 & {,}001 & {,}0 & {,}001 & {,}001 & {,}002 & {,}004 & {,}004 & {,}001 \\ {,}0 & {,}944 & {,}001 & {,}002 & {,}001 & {,}01 & {,}001 & {,}001 & {,}004 & {,}002 & {,}012 & {,}004 & {,}01 & {,}004 & {,}002 & {,}001 \\ {,}0 & {,}001 & {,}984 & {,}001 & {,}0 & {,}0 & {,}0 & {,}0 & {,}004 & {,}0 & {,}001 & {,}0 & {,}001 & {,}004 & {,}004 & {,}0 \\ {,}002 & {,}002 & {,}0 & {,}983 & {,}001 & {,}001 & {,}0 & {,}0 & {,}001 & {,}0 & {,}001 & {,}0 & {,}002 & {,}004 & {,}001 & {,}002 \\ {,}001 & {,}003 & {,}0 & {,}003 & {,}957 & {,}015 & {,}0 & {,}0 & {,}002 & {,}002 & {,}001 & {,}0 & {,}006 & {,}004 & {,}002 & {,}004 \\ {,}0 & {,}003 & {,}0 & {,}001 & {,}004 & {,}977 & {,}0 & {,}0 & {,}001 & {,}004 & {,}002 & {,}001 & {,}003 & {,}002 & {,}001 & {,}001 \\ {,}001 & {,}003 & {,}0 & {,}001 & {,}0 & {,}004 & {,}97 & {,}0 & {,}007 & {,}0 & {,}008 & {,}001 & {,}003 & {,}001 & {,}001 & {,}0 \\ {,}0 & {,}001 & {,}0 & {,}0 & {,}0 & {,}001 & {,}0 & {,}987 & {,}001 & {,}002 & {,}001 & {,}004 & {,}001 & {,}0 & {,}001 & {,}0 \\ {,}001 & {,}002 & {,}002 & {,}001 & {,}001 & {,}002 & {,}002 & {,}0 & {,}972 & {,}0 & {,}002 & {,}0 & {,}006 & {,}003 & {,}003 & {,}001 \\ {,}0 & {,}004 & {,}0 & {,}001 & {,}0 & {,}009 & {,}0 & {,}001 & {,}001 & {,}975 & {,}002 & {,}003 & {,}001 & {,}001 & {,}0 & {,}001 \\ {,}0 & {,}008 & {,}0 & {,}001 & {,}001 & {,}002 & {,}003 & {,}001 & {,}003 & {,}001 & {,}967 & {,}003 & {,}006 & {,}002 & {,}002 & {,}0 \\ {,}001 & {,}005 & {,}0 & {,}001 & {,}0 & {,}002 & {,}0 & {,}003 & {,}001 & {,}003 & {,}006 & {,}973 & {,}002 & {,}001 & {,}001 & {,}0 \\ {,}001 & {,}013 & {,}001 & {,}002 & {,}004 & {,}008 & {,}001 & {,}0 & {,}01 & {,}002 & {,}006 & {,}001 & {,}934 & {,}009 & {,}007 & {,}001 \\ {,}004 & {,}004 & {,}0 & {,}006 & {,}002 & {,}004 & {,}0 & {,}0 & {,}002 & {,}0 & {,}002 & {,}0 & {,}012 & {,}952 & {,}006 & {,}005 \\ {,}008 & {,}002 & {,}003 & {,}002 & {,}001 & {,}001 & {,}0 & {,}0 & {,}005 & {,}0 & {,}002 & {,}001 & {,}009 & {,}009 & {,}955 & {,}002 \\ {,}001 & {,}002 & {,}0 & {,}006 & {,}005 & {,}005 & {,}0 & {,}0 & {,}001 & {,}0 & {,}002 & {,}0 & {,}005 & {,}005 & {,}001 & {,}966 \end{pmatrix} \end{array} \tag{14}$$

Ejercicio 2

Clave	Demarcación terrirorial
09002	Azcapotzalco
09003	Coyoacán
09004	Cuajimalpa de Morelos
09005	Gustavo A. Madero
09006	Iztacalco
09007	Iztapalapa
09008	La Magdalena Contreras
09009	Milpa Alta
090010	Álvaro Obregón
090011	Tláhuac
090012	Tlalpan
090013	Xochimilco
090014	Benito Juárez
090015	Cuauhtémoc
090016	Miguel Hidalgo
090017	Venustiano Carranza

Ejercicio 2

- Utiliza la matriz de transición para encontrar la distribución estacionaria.
- ¿En cuántos años la distribución probabilidad de cambio de alcaldía estará en estado estacionario?

Monte Carlo vía Cadenas de Markov (MCMC)

Consideremos nuevamente el problema de determinar el valor de la integral $\mu = \int f(x)dx$, donde factorizamos $f(x) = h(x)\pi(x)$,

$$\mu = E_\pi h(x) = \int_B h(x)\pi(x)dx, \quad B \subseteq R^d$$

Los métodos de Monte Carlo vía Cadenas de Markov (MCMC) son especialmente útiles para integrales sobre regiones de varias dimensiones, dado que la simulación de variables iid en varias dimensiones es altamente ineficiente.

La integral μ puede interpretarse como un valor esperado con respecto a una distribución estacionaria π de una Cadena de Markov.

Algoritmo de Metropolis (Hanada and Matsuura, 2022)

Supongamos una distribución de probabilidad $P(x)$

$$P(x) = \frac{e^{-S(x)}}{Z} \tag{15}$$

En física, la función $S(x)$ es llamada la acción y el factor de normalización Z es llamada la función de partición.

En nuestro contexto, podemos pensar S como una función de log verosimilitud.

Se asume que $S(x)$ es una función continua de una variable real x.

Algoritmo de Metropolis (Hanada and Matsuura, 2022)

Para el caso de una distribución normal[3], la acción es

$$S(x) = \frac{x^2}{2}$$

y la función de partición es

$$Z = \sqrt{2\pi}$$

Normalmente en aplicaciones sólo conocemos $S(x)$ y Z es desconocida.

[3] La distribución normal es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Para $\mu = 0$ y $\sigma = 1$, obtenemos las funciones de acción y partición propuestas $S(x)$

Algoritmo de Metropolis (Hanada and Matsuura, 2022)

En el Algoritmo de Metropolis comenzamos con un valor inicial $x^{(0)}$ y generamos una secuencia $x^{(1)}, x^{(2)}, \ldots, x^{(k)}, x^{(k+1)}$, siguiendo las siguientes reglas:

Algorithm 2: Algoritmo (General) de Metropolis

Escoge un número real Δx de forma aleatoria

Se propone $x' = x^{(k)} + \Delta x$ como candidato para $x^{(k)}$

Metropolis test: se acepta el candidato x' y el valor de x es actualizado a $x^{(k+1)} = x'$ con probabilidad $\text{mín}\left(1, e^{S(x^{(k)})-S(x')}\right)$.

En caso contrario, se rechaza al candidato x' y el valor de x permanece inalterado, esto es, $x^{(k+1)} = x^{(k)}$

Para el **Metropolis test**, se genera un número aleatorio r entre 0 y 1 y se acepta el candidato x' si $r < e^{S(x^{(k)})-S(x')}$.

Implementación del Algoritmo de Metropolis

```
import numpy as np

niter = 100000
x = 0

for it in range(niter):
    backup = x
    action_init = (1/2)*x**2
    dx = np.random.random()
    dx = (dx-0.5)*2.0
    x = x + dx
    action_final =  (1/2)*x**2

    ### Metropolis tes
    r = np.random.random()

    if np.exp(action_init - action_final) > r:
        ## Aceptamos el candidato
        naccept += 1
    else:
        ## Rechazamos el candidato
        x = backup
```

References I

Dobrow, R. P. (2016). *Introduction to stochastic processes with R*. John Wiley & Sons.

Hanada, M. and Matsuura, S. (2022). *MCMC from Scratch: A Practical Introduction to Markov Chain Monte Carlo*. Springer Nature.

Theodoridis, S. (2020). *Machine Learning. A Bayesian and Optimization Perspective*. Academic Press, second edition edition.

Wilkinson, B. and Allen, M. (1998). *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*. Prentice-Hall, Inc., USA.

Zhang, R., Tang, C., Ma, S., Yuan, H., Gao, L., and Fan, W. (2011). Using markov chains to analyze changes in wetland trends in arid yinchuan plain, china. *Mathematical and Computer Modelling*, 54(3-4):924–930.

Zwanzig, S. and Mahjani, B. (2019). *Computer intensive methods in statistics*. CRC Press.