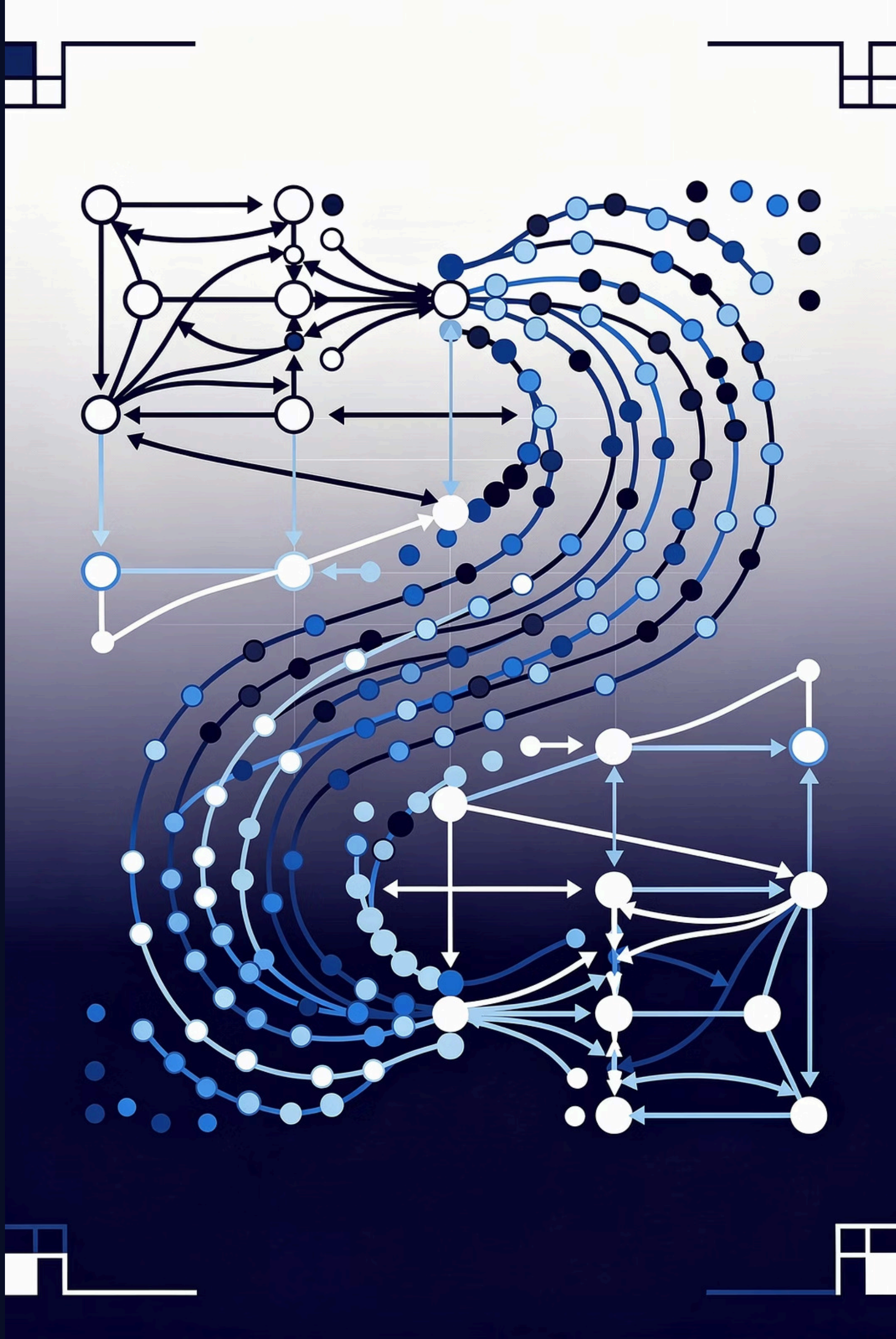
Simulación Monte Carlo y Cadenas de Markov

MÓDULO 4 FEB-JUN 2026

Técnicas computacionales avanzadas para modelar fenómenos sociales. Concentración en Economía Aplicada y Ciencia de Datos, ITESM-SF, Ciudad de México.



Método de Monte Carlo: Fundamentos

Idea Central

Método estadístico para calcular el área bajo una curva mediante simulación. Es una solución probabilística al problema de integración numérica.

Supongamos que existe $M > 0$ tal que $0 \leq f(\theta) \leq M$ para todo $\theta \in [a, b]$. Queremos calcular la integral:

$$I = \int_a^b f(\theta)d\theta$$

El valor de la integral corresponde al área bajo la curva $\phi = f(\theta)$ para $\theta \in [a, b]$, inscrita en el rectángulo $R = [a, b] \times [0, M]$.

Estimación por Monte Carlo

La función de densidad de una distribución uniforme sobre el rectángulo R se define como:

$$p(\theta, \phi) = \frac{1}{M(b-a)} I_R(\theta, \phi)$$

Simulación

Se genera una muestra $(\theta_1, \phi_1), \ldots, (\theta_n, \phi_n)$ de $p(\theta, \phi)$

Conteo

Se cuentan cuántos valores caen bajo la curva $\phi = f(\theta)$

Estimador

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

La varianza del estimador está dada por:

$$\mathrm{Var}(\hat{I}) = \frac{1}{N}\{M(b-a) - I\}$$

Cálculo de π por Monte Carlo

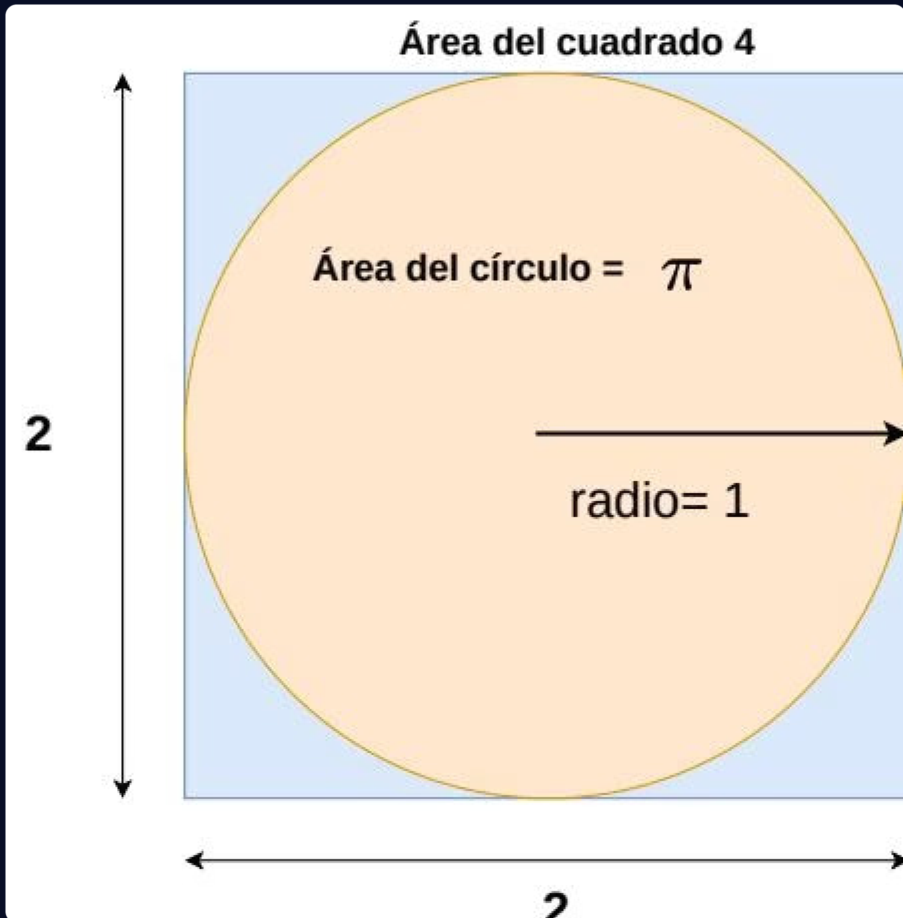


Diagrama mostrando un cuadrado de lado 2 que contiene un círculo de radio 1. Área del cuadrado = 4, Área del círculo = π.

La relación entre las áreas se expresa como:

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

Esto puede describirse mediante la integral:

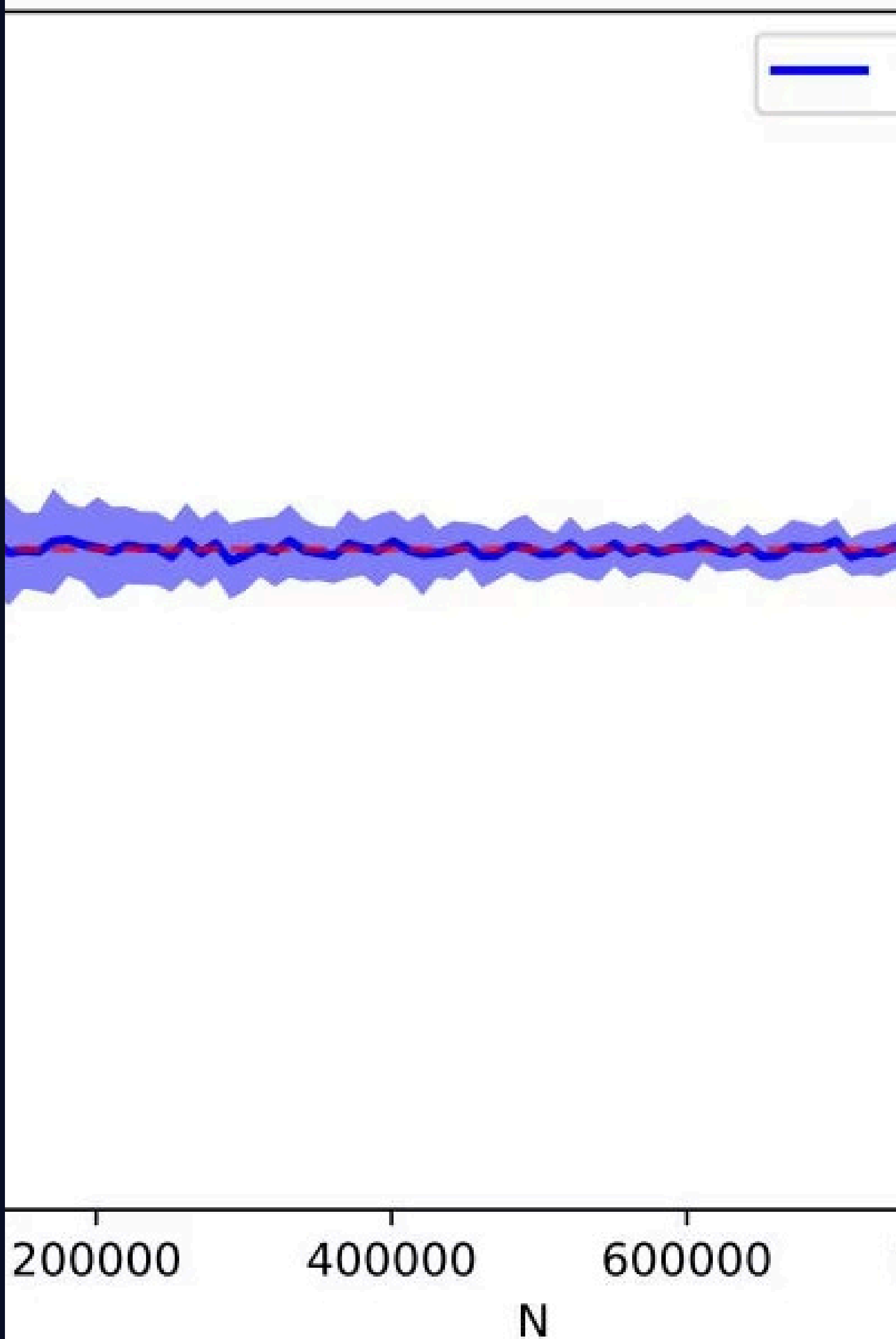
$$\int_0^1 \sqrt{1 - x^2}dx = \frac{\pi}{4}$$

Se generan parejas aleatorias (x_r, y_r) distribuidas uniformemente entre 0 y 1, contando cuántos puntos caen dentro del círculo: $y_r^2 + x_r^2 \leq 1$

Convergencia del Método Monte Carlo

La gráfica ilustra cómo el valor estimado de π converge al valor real (línea roja punteada ≈ 3.14159) conforme aumenta el número de iteraciones N. Las fluctuaciones iniciales disminuyen significativamente después de 100,000 iteraciones.

3.14	3.1415	3.1415
N = 100,000	N = 600,000	N = 1,000,000
Primera aproximación estable	Convergencia mejorada	Estimación final precisa



Inferencia Bayesiana y MCMC

En un contexto bayesiano con modelo muestral $Y \sim p(y|\theta)$ y distribución inicial $p(\theta)$, la distribución posterior se calcula:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(\theta')p(y|\theta')d\theta'}$$

01

Desafío Computacional

El cálculo analítico de $p(\theta|y)$ es complicado en la mayoría de los casos, incluso descartando el denominador como factor de escalamiento.

02

Técnicas Alternativas

Existen métodos como aproximación de Laplace y cuadratura, pero las técnicas MCMC son las más utilizadas por la capacidad de cómputo disponible.

03

Idea de MCMC

Construir una cadena de Markov fácil de simular mediante muestreo, cuya distribución de equilibrio corresponda a la distribución posterior de interés.

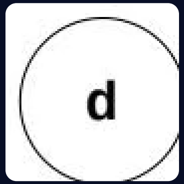
Cadenas de Markov: Definición y Ejemplo

Definición: Proceso estocástico discreto con estados finitos que cumple la propiedad de Markov:

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

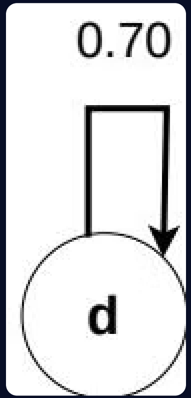
Estados del Sistema

- **d:** Dormitorio
- **b:** Bar
- **c:** Comedor

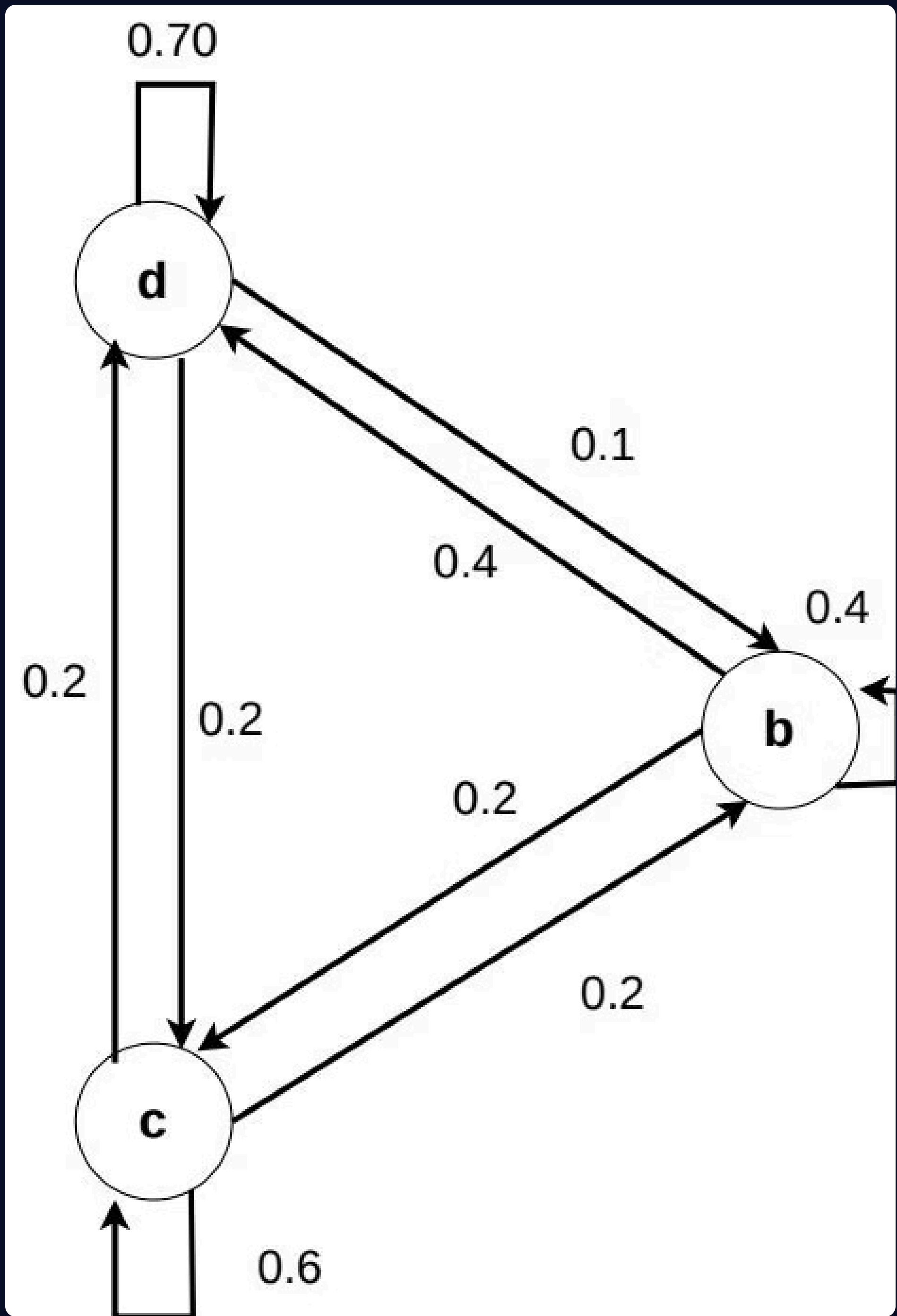


Las aristas representan probabilidades de transición entre estados. Por ejemplo, la probabilidad de permanecer en el dormitorio es:

$$P(X_{t+1} = d \mid X_t = d) = 0.7$$



Matriz de Transición y Dinámica



Las probabilidades de transición se representan matricialmente con entradas $p_{i,j} \equiv P(X_{t+1} = i \mid X_t = j)$:

$$T = \begin{bmatrix} 0.7 & 0.4 & 0.2 \\ 0.1 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$$

La dinámica de la distribución de probabilidad se expresa como:

$$P(X_{t+1}) = TP(X_t)$$

Si el pasajero está en el dormitorio al tiempo t=0, entonces $P(X_0) = [1, 0, 0]^T$. Al tiempo siguiente: $P(X_1) = TP_0 = [0.7, 0.1, 0.2]^T$. Para tiempos futuros: $P_t = T^t P_0$

Teorema Perron-Frobenius y Algoritmo Metropolis

1

Conjunto Finito

Estados S = {1, 2, ..., K}

2

Transición Fija

Matriz no cambia en el tiempo

3

Ergodicidad

Accesibilidad entre todos los estados

4

No Ciclicidad

Sin ciclos repetitivos de estados

El algoritmo de Metropolis construye una cadena donde la distribución propuesta q(·|xn-1) cambia con el tiempo. Generar una nueva muestra depende del estado previo. El valor de aceptación se calcula como:

$$\alpha(x|x_{n-1}) = \min\left\{1, \frac{p(x)}{p(x_{n-1})}\right\}$$

Si u ~ U(0,1) y u ≤ α(x|xn-1), entonces xn = x; de lo contrario, xn = xn-1.

Ejemplo: Distribución Normal con Varianza Conocida

Sea $\theta \sim \text{normal}(\mu, \tau^2)$ y $\{y_1, \ldots, y_n \mid \theta\} \sim$ i.i.d. $\text{normal}(\theta, \sigma^2)$. La distribución posterior de θ es $\text{normal}(\mu_n, \tau_n^2)$ donde:

$$\mu_n = \bar{y}\frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}} + \mu\frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$$

$$\tau_n^2 = \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$$

Parámetros del Ejemplo

- $\sigma^2 = 1$, $\tau^2 = 10$, $\mu = 5$, $n = 5$
- $y = \{9, 37, 10, 18, 9, 16, 11, 60, 10, 33\}$
- Resultado: $\mu_n = 10.03$, $\tau_n^2 = 0.2$
- $p(\theta|y) = \text{dnorm}(10.03, 0.44)$

Proporción de Aceptación

Para estabilidad numérica, se calcula log r:

$$\log r = \sum_{i=1}^{n}[\log \text{dnorm}(y_i, \theta^*, \sigma) - \log \text{dnorm}(y_i, \theta^{(s)}, \sigma)] + \log \text{dnorm}(\theta^*, \mu, \tau) - \log \text{dnorm}(\theta^{(s)}, \mu, \tau)$$

El valor propuesto se acepta si $\log u < \log r$.

Referencias: Theodoridis, S. (2020). Machine Learning: A Bayesian and Optimization Perspective. Academic Press. | Wilkinson, B. and Allen, M. (1998). Parallel Programming. Prentice-Hall.