Cadenas de Markov

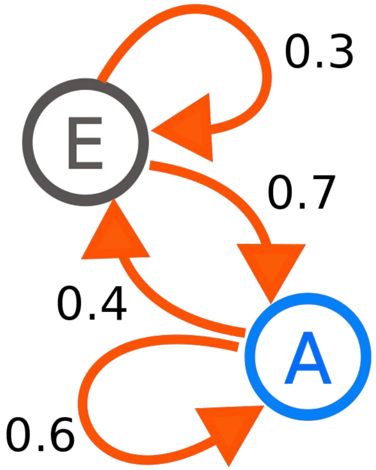
y MCMC

Cadenas de Markov

Las cadenas de Markov representan procesos estocásticos a tiempo discreto donde la probabilidad de cada estado futuro depende únicamente del estado actual, no de los estados pasados. Es decir:

$$P[X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n] = P[X_{n+1} = x_{n+1} | X_n = x_n]$$

A esto se le conoce como ***Propiedad de Markov***.



Source: WikiMedia

En una Cadena de Markov, un **estado** es una configuración particular o condición que el sistema puede adoptar.

Ejemplo: En un modelo de pronóstico del tiempo, los estados podrían ser "soleado", "nublado", "lluvioso".

Cadenas de Markov

Las **transiciones** son los cambios de un estado a otro. Estas definidas según la probabilidad de que el cambio respectivo suceda en un intervalo temporal.

Podemos definir una **matriz de transición**: tabla que representa las probabilidades de cambio de cada estado hacia todos los posibles estados siguientes. Ejemplo:

Actual \ Siguiente	Soleado	Nublado	Lluvioso
Soleado	0.5	0.3	0.2
Nublado	0.2	0.6	0.2
Lluvioso	0.1	0.4	0.5

También podemos crear un **diagrama de estados** de la cadena para representarla visualmente, incluyendo las probabilidades asociadas a cada cambio.

Cadenas de Markov

Estas cadenas se usan para **modelar** sistemas que se pueden describir por estados que cambian de forma aleatoria, siendo aplicadas en una amplia variedades de disciplinas.

- Predicción meteorológica, bioinformática, epidemiología, salud, genética, economía, etc.
- IA - machine learning (no supervisado), procesamiento de lenguaje natural, grafos, evaluación de software, redes neuronales,
- Juegos (de azar) y entretenimiento, música, modelos financieros, modelos de colas, ranking,
- etc...

La representación de un sistema (problema) mediante cadenas de Markov nos permite calcular la probabilidad de transición para cualquier secuencia de estados requerida, lo cual puede utilizarse tanto para hacer predicciones como tomar decisiones.

Cadenas de Markov con Monte Carlo

Las Cadenas de Markov Monte Carlo (MCMC, por sus siglas en inglés) son un conjunto de **algoritmos** utilizados para **muestrear de una distribución de probabilidad**.

Al combinar el método Monte Carlo (muestreo aleatorio) con las Cadenas de Markov, MCMC permite estimar distribuciones complejas que son difíciles de calcular directamente, o cuyo modelo tiene un costo computacional es muy elevado.

Son fundamentales en la *inferencia bayesiana*, donde se busca actualizar las creencias en función de la evidencia.
También tiene amplia aplicación en machine learning.

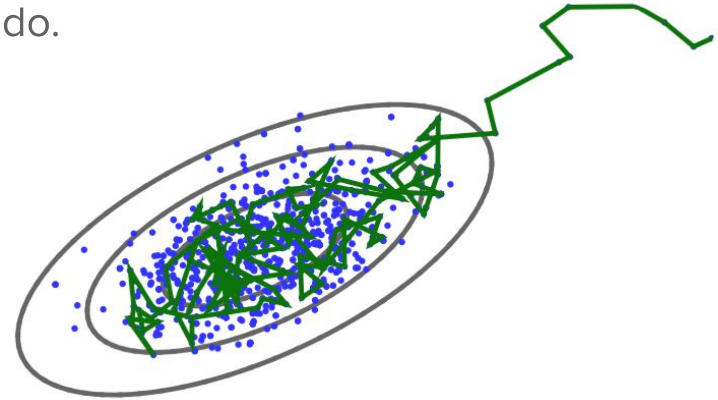


Ilustración MCMC. Source: UBC Wiki

MCMC: metodología

- MCMC genera una **serie de muestras** (estados de la cadena de Markov) que se utilizan para representar una distribución desconocida.
- El proceso comienza desde un **estado inicial** y luego realiza **transiciones** a nuevos estados basándose en ciertas **reglas** probabilísticas.
- Con **suficientes iteraciones**, la distribución de las muestras tiende a la distribución objetivo (la que se desea estimar).
- Se pueden descartar las primeras muestras para asegurar que la cadena ha alcanzado su estado estacionario (quemar la cadena).

Para samplear distribuciones de datos/modelos, así como para buscar mínimos/máximos en una distribución, el algoritmo MCMC usualmente utilizado es el de **Metropolis-Hastings**.

Metropolis-Hastings

Método MCMC para obtener una secuencia de muestras aleatorias de una distribución de probabilidad para la cual el muestreo directo es difícil. Funciona **proponiendo** una nueva muestra y **decidiendo** si aceptarla o rechazarla en función de la **probabilidad relativa a la muestra actual**.

Decisión

1. Se calcula la razón de aceptación comparando la densidad de probabilidad de la nueva muestra con la actual.

2. Si la nueva muestra es más probable, se acepta; si es menos probable, se acepta con una probabilidad igual a la razón de aceptación.

Metropolis-Hastings: procedimiento

Dado f (x) una función que sea proporcional a la distribución de probabilidad deseada P(x):

Inicio

- Elegir punto arbitrario x_t para la primera muestra

- Elegir una densidad de probabilidad arbitraria ***g***(x | y) que sugiere un candidato para el siguiente valor de muestra *x*, dado el valor de muestra anterior *y*. Una opción habitual es dejar que ***g***(x | y) sea una distribución gaussiana centrada en *y*, así los puntos cercanos son más propensos a ser visitados a continuación, convirtiendo la secuencia de muestras en una caminata aleatoria. La función ***g*** se denomina densidad de propuesta o distribución de saltos.

Metropolis-Hastings: procedimiento

En cada iteración t

Generar candidato x' para la siguiente muestra seleccionandolo de la distribución $\boldsymbol{g}(x' | x_t)$.

Calcular tasa de aceptación $\alpha = f (x') / f (x_t)$

Ya que f es proporcional a la densidad de P, podemos deducir que

$$\alpha = f (x') / f (x_t) = P (x') / P (x_t)$$

Aceptación o rechazo

- Generar número aleatorio uniforme $u \in [0 , 1]$

Si $u \leq \alpha$, entonces acepta el candidato de forma que $x_{t+1} = x'$

Si $u > \alpha$, entonces rechaza el candidato y coloca $x_{t+1} = x_t$ en su lugar.