VARIABLE TRANSFORMATION

One variable

EXAMPLE

$$X \in \Omega_X = \{1, \ldots, 6\}$$

$$\forall x \in \Omega_X, P(X = x) = \frac{1}{6}$$

We have a dice, and we define:

$$Y = g(X) = |X - 3|$$

What's the distribution of *Y*?

$$P(Y = y) = \sum_{x \in \Omega_X / x = g^{-1}(y)} P(X = x)$$

CONTINUOUS VARIABLES

Suppose *Y* is an **increasing** function of *X*, then ...

$$F(Y = y) = P(Y \leq y) = P(g(X) \leq y) =$$
$$= P(X \leq g^{-1}(y)) = F(X = g^{-1}(y))$$

Taking the derivative

$$f(Y = y) = \frac{d}{dy} F(X = g^{-1}(y)) = f(X = g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$$

Similar for the decreasing case. In general:

$$f(Y = y) = f(X = g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

CONTINUOUS VARIABLES: EXAMPLE

$$X \sim Exp(\lambda) \rightarrow f_X(x) = \lambda e^{-\lambda x}$$

$$y = g(x) = kX,\ f_Y(y)??$$

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)$$

$$g^{-1}(y) = \frac{y}{k},\ \frac{d}{dy}g^{-1}(y) = \frac{1}{k}$$

$$f_Y(y) = \lambda e^{-\lambda\frac{y}{x}}\frac{1}{k} = \frac{\lambda}{k}e^{-\frac{\lambda}{k}y}$$

Therefore:

$$X \sim Exp(\lambda) \rightarrow kX \sim Exp(\frac{\lambda}{k})$$

TWO VARIABLE CASE

$$f_{X,Y}(x,y)$$

$$Z = g_Z(X,Y);\ T = g_T(X,Y)$$

$$X = l_X(Z,T);\ Y = l_Y(Z,T)$$

$$|J| = \left|\frac{\partial(X,Y)}{\partial(Z,T)}\right| = \begin{vmatrix} \frac{\partial X}{\partial Z} & \frac{\partial X}{\partial T} \\ \frac{\partial Y}{\partial Z} & \frac{\partial Y}{\partial T} \end{vmatrix}$$

$$f_{Z,T}(z,t) = f_{X,Y}(l_X(z,t), l_Y(z,t))||J||$$

Two variables: Example

$$X \sim Exp(\lambda_1) \rightarrow f_X(x) = \lambda_1 e^{-\lambda_1 x}$$

$$Y \sim Exp(\lambda_2) \rightarrow f_Y(y) = \lambda_2 e^{-\lambda_2 y}$$

$$X \perp Y$$

$$f_{XY}(x, y) = \lambda_1 \lambda_2 e^{-\lambda_1 x - \lambda_2 y}$$

$$Z = X + Y \rightarrow f_Z(z) = ?$$

$$Z = X + Y$$
$$T = X$$

$$X = T$$
$$Y = Z - T$$

$$||J|| = 1$$

$$f_{ZT}(z, t) = f_{XY}(t, z - t)||J|| = \lambda_1 \lambda_2 e^{-\lambda_1 t - \lambda_2 (z - t)}$$

$$= \lambda_1 \lambda_2 e^{-(\lambda_1 - \lambda_2) t} e^{-\lambda_2 z}$$

TWO VARIABLES: EXAMPLE

$Z = X + Y \rightarrow f_Z(z) = ?$

$$\begin{aligned} Z &= X + Y \\ T &= X \end{aligned} \qquad \begin{aligned} X &= T \\ Y &= Z - T \end{aligned}$$

$f_{ZT}(z,t) = \lambda_1 \lambda_2 e^{-(\lambda_1 - \lambda_2)t} e^{-\lambda_2 z}$

$Y \geq 0 \rightarrow Z \geq T \rightarrow 0 \leq t \leq z$

$f_Z(z) = \int_{\Omega_T} f_{ZT}(z,t)dt$

$f_Z(z) = \int_0^z \lambda_1 \lambda_2 e^{-\lambda_2 z} e^{-(\lambda_1 - \lambda_2)t} dt \; = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2}(e^{-\lambda_1 z} - e^{-\lambda_2 z})$

Hypoexponential distribution

$$E[X+Y] = E[Z] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}$$

SAMPLING BASED APPROXIMATIONS

Monte Carlo Methods

APPROACHING THE EXPECTATION

$$X \sim Exp(\lambda_1) \rightarrow f_X(x) = \lambda_1 e^{-\lambda_1 x} \qquad X \perp Y$$
$$Y \sim Exp(\lambda_2) \rightarrow f_Y(y) = \lambda_2 e^{-\lambda_2 y} \qquad E[X+Y] =?$$

1.- Sample X and Y

2.- For each sample, get Z=X+Y

3.- Approach E[Z] with the sample mean

EXAMPLE: $\lambda_1 = 3, \lambda_2 = 2; E[X+Y] = 0.8\hat{3}$

10 samples → 1.168

100 samples → 0.850

1000 samples → 0.816

10000 samples → 0.835

100000 samples → 0.833

1000000 samples → 0.832

Sampling based approaches are great …

… if we can sample the distributions!

$$f(x) = e^{e^{-(x-\mu)/\beta}}$$

Sampling general Distributions

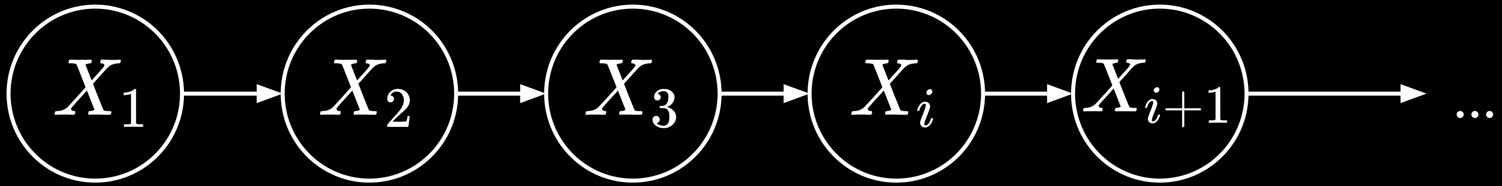
Markov Chain Monte Carlo

MARKOVIAN PROPERTY

The conditional probability distribution of **future states** of the process **depends only upon** the **present state**

MARKOV CHAIN

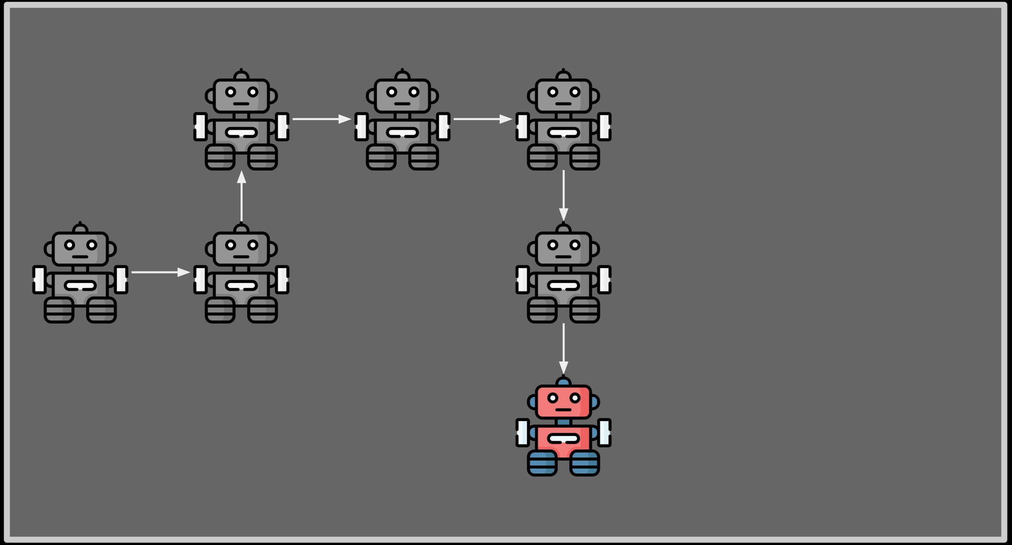
Stochastic process → Infinite sequence of random variables



Due to the Markovian property:

$$\forall i \;\; P(X_{i+1} | X_1, X_2 \ldots, X_i) = P(X_{i+1} | X_i)$$

MARKOV CHAIN: DISCRETE EXAMPLE



Random movement in discrete steps, according to:

$$X_i \in \Omega_X = \{\uparrow, \downarrow, \leftarrow, \rightarrow\}$$

$$\forall a, b \in \Omega_X \;\; P(X_{i+1} = a | X_i = b)$$

X_{i+1}

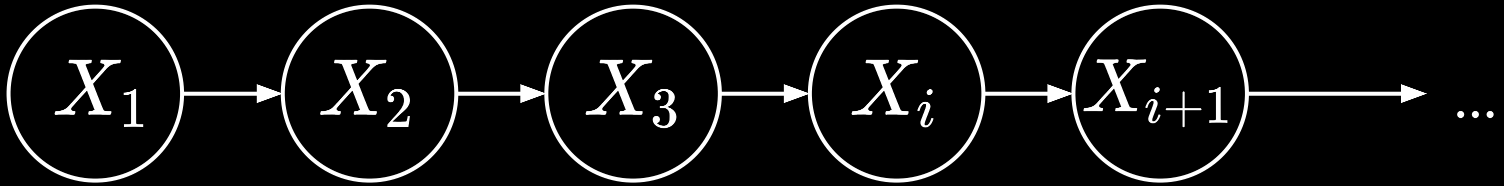
X_i

	↑	↓	←	→
↑	0.7	0.1	0.1	0.1
↓	0.1	0.7	0.1	0.1
←	0.1	0.1	0.7	0.1
→	0.1	0.1	0.1	0.7

If the transition probabilities are **time-independent**, then we say the chain is **homogeneous**

MARKOV CHAIN

Stochastic process → Infinite sequence of random variables



Due to the Markovian property:

$$\forall i \;\; P(X_{i+1} | X_1, X_2 \ldots, X_i) = P(X_{i+1} | X_i)$$

STATIONARY DISTRIBUTION

If, for *i* big enough, $P(X_{i+1}) = P(X_i) = \pi(X)$

We say that the Markov chain has a **stationary distribution**, $\pi(X)$

$$x_0\,, x_1\,, \ldots\ldots\,, \underset{\text{Sampling of } \pi(X)}{\underrightarrow{x_i\,, x_{i+1}\,, \ldots \qquad\qquad\qquad}}$$

If P is the transition probability matrix, then:

$$\pi^T = \pi^T P$$

DEFINITIONS AND PROPERTIES

Ergodic or irreducible: If it is possible to move between any two given states

Periodic: If it is possible to return to the same state at given intervals

If a Markov chain is **ergodic** and **aperiodic**, then it has a **unique stationary distribution**

SAMPLING ANY DISTRIBUTION

How do we sample this distribution?

$$f(x) = \begin{cases} 1 - |1 - x|, & 0 < x < 2 \\ 0, & \text{in other case} \end{cases}$$

Find a Markov chain with a unique stationary distribution such that:

$$\pi(x) = f(x)$$

Then, sample the Markov chain until convergence

MARKOV CHAIN MONTE CARLO

Metropolis-Hasting algorithm

METROPOLIS - HASTING ALGORITHM

We need two elements for our algorithm:

A mechanism to propose new values

$$g(x^*|x_{t-1})$$

A criterion to accept or reject them

$$\rho = \frac{f(x^*)g(x^*|x_{t-1})}{f(x_{t-1})g(x_{t-1}|x^*)} \quad p = \min\{1, \rho\}$$

At each step *t*

1.- Propose a new value x^* using *g* and the previous value x_{t-1}

2.- With probability p, $x_t = x^*$, $x_t = x_{t-1}$ otherwise

METROPOLIS - HASTING ALGORITHM

Two typical approaches for the proposal distribution:

$$g(x^*|x_{t-1}) = g(x^*)$$

$$g(x^*|x_{t-1}) = h(x^* - x_{t-1})$$ with *h* symmetric

In the latter case:

$$g(x^*|x_{t-1}) = g(x_{t-1}|x^*) \rightarrow \rho = \frac{f(x^*)}{f(x_{t-1})}$$

EXAMPLE

$$f(x) = \begin{cases} 1 - |1 - x|, & 0 < x < 2 \\ 0, & \text{in other case} \end{cases}$$

Proposal distribution

$$g(x^*|x_{t-1}) = g(x^*) = Unif(0, 2)$$

$$g(x^*) = g(x_{t-1}) \rightarrow \rho = \frac{f(x^*)}{f(x_{t-1})}$$

* We use *g* also to generate x_0

Probabilistic Modeling
and
Bayesian NEtworKs

Probability Distributions

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

INFORMATIKA FAKULTATEA
FACULTAD DE INFORMÁTICA

Konputazio Ingeniaritza eta Sistema Adimentsuak Unibertsitate Masterra

Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes