Computational Statistics  Nov 17, 2021

## Stochastic Processes

or $X(t) = X_t$

$X_t$ is a r.v.

A stochastic process $\{X_t : t \in T\}$ is a collection of random variables indexed by $t$

- $X_t$ takes values in the state space $\mathcal{X}$
- $T$ is the index set (i.e. $\mathbb{R}$, $\mathbb{N}$, ..)
- Ex include stock prices, weather, IID ~~sq var~~ $X_1, \dots, X_n \dots$
- Recall: for $X_1, \dots, X_n$ the joint density is given by

$$f(x_1, \dots, x_n) = f(x_1)\, f(x_2 | x_1)\, f(x_3 | x_1, x_2) \cdots f(x_n | x_1, \dots, x_{n-1})$$

$$= \prod_{i=1}^{n} f(x_i \mid \text{past } i\text{'s})$$

## Markov Chains

**Def** $\{X_n : n \in T\}$ is a Markov Chain

if $\mathbb{P}(X_n = x \mid X_0, \dots, X_{n-1}) = \mathbb{P}(X_n = x \mid X_{n-1})$

for all $n \in T$ and $x \in \mathcal{X}$.

$$\Rightarrow \quad f(x_n \mid x_{n-1} \dots x_0) = f(x_n \mid x_{n-1})$$

$$\Rightarrow \quad f(x_1, x_2, \dots x_n) = f(x_1)\, f(x_2 | x_1)\, f(x_3 | x_2) \cdots f(x_n | x_{n-1})$$

Questions to answer:

1. When does a MC achieve "equilibrium"? Does it at all?
2. Estimate parameters controlling the MC
3. Can we construct a MC that converges to a specified equilibrium? i.e. $X_n \rightsquigarrow F$, some given distribution

## Transition Probability

**Def:** $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ are the transition probabilities.

If $p_{ij}$ does not depend on $n$, called a <u>homogeneous MC.</u>

The matrix $P$ with elements $P_{ij} = p_{ij}$ is known as the transition matrix.

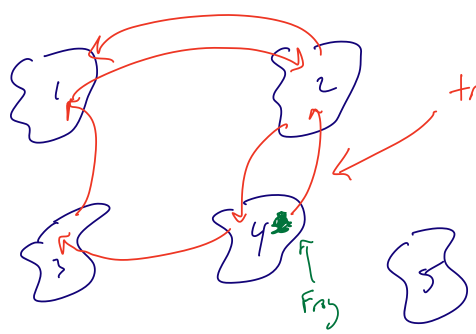
Two properties of the transition probabilities:

① $p_{ij} \geq 0$

② $\sum_j p_{ij} = 1$ ( Typo in book).

Each row of $P$ is a prob. mass function.

"states" : 1, 2, ...




transition probabilities

$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$

n-step transition probability: $\mathbb{P}(X_{m+n} = j \mid X_m = i) = p_{ij}(n)$

Theorem (Chapman-Kolmogorov) The n-step transition probabilities satisfy:

$$p_{ij}(m+n) = \sum_k p_{ik}(m)\, p_{kj}(n) = \big(P(m)\, P(n)\big)_{ij}$$

$\Rightarrow \quad P(2) = P \cdot P = P^2$

$\Rightarrow \quad P(3) = P^3$

$\Rightarrow \quad P(n) = P^n$

This means that if at time 0, my probability of being in state $i$ is $\mu_i$, and define

$$\mu(0) = (\mu_1(0)\, \mu_2(0) \dots \mu_N(0))$$

$\Rightarrow \quad \mu(1) = \mu(0)\, P,$

$\Rightarrow \quad \mu(n) = \mu(0)\, P^n \quad \leftarrow$ matrix vector multiplication.

Question: As $n \to \infty$, is $\mu_i(n) > 0$? Or is $p_{ij} > 0$ for all $i$?

Def: state $i$ reaches state $j$ ($j$ is accessible from $i$) if $p_{ij}(n) > 0$ for some $n$

$\Rightarrow \quad i \to j$

$\Rightarrow \quad$ if $i \to j$ and $j \to i$, then $i \leftrightarrow j$ "communicate"

<u>Thm</u>

1. $i \leftrightarrow i$
2. $i \leftrightarrow j \Rightarrow j \leftrightarrow i$
3. $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.
4. The state space $\mathcal{X}$ can be written as a disjoint union of <u>classes</u> $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \dots$ where $i, j$ communicate iff $i, j \in \mathcal{X}_k$.

<u>Def</u>: If all states communicate, then the chain is <u>irreducible</u>.

<u>Closed</u>: set of states is closed if the chain enters but never leaves.

Closed set with a single state: <u>an absorbing state</u>.

Recurrent/persistent state : $\mathbb{P}(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) = 1$

Transient : else.

<u>Stationarity</u> $\pi$ is a stationary (or invariant) distribution if $\pi = \pi P$.

$\Rightarrow$ $\pi$ is a row eigenvector of $P$

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

$\Rightarrow$ with eigenvalue 1.

Idea : Draw $X_0$ from $\pi$, a stationary distribution of $P$.

Next, draw $X_1 \sim \pi P$.

Notationally : $X_1 \sim \mu_1 = \mu_0 P = \pi P = \pi$

$\Rightarrow$ If $X_2 \sim \mu_2 = \mu_1 P = \mu_0 P^2 = \pi P = \pi$

$\Rightarrow$ that $X_2 \sim \pi$

When a chain has distribution $\pi$, it will forever.

Def A Markov Chain has limiting distribution $\pi$ if

$$P^n \to \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots & & & \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix}$$

$$\Rightarrow \mu_0 P^n = \pi$$

$$(\mu_1 \cdots \mu_N) \begin{pmatrix} \pi_1 & \cdots & \pi_N \\ \pi_1 & & \vdots \\ \vdots & & \vdots \\ \pi_1 & & \pi_N \end{pmatrix} = \left(\pi_1 \sum \mu_j \quad \pi_2 \sum \mu_j \quad \cdots \right)$$

$$= (\pi_1 \quad \pi_2 \quad \cdots \quad \pi_N).$$

Detailed Balance $\pi$ satisfies detailed balance if for all $i,j$

$$\pi_i P_{ij} = \underbrace{P_{ji}\pi_j}_{\mathbb{P}(X_{n+1}=i,\, X_n=j)}$$

$$\underbrace{\mathbb{P}(X_n=i)\,\mathbb{P}(X_{n+1}=j \mid X_n=i)}_{\mathbb{P}(X_{n+1}=j,\, X_n=i)}$$

<u>Thm</u> If $\pi$ satisfies detailed balance, then $\pi$ is a stationary distribution.

<u>Proof</u>: Detailed balance says $\pi_i p_{ij} = p_{ji} \pi_j$

We need to show that $\pi P = \pi$. The $j^{th}$ element of $\pi P = (\pi P)_j = \sum_{k=1}^{N} \pi_k P_{kj} = \sum_{k=1}^{N} P_{jk} \pi_j = \pi_j \sum_{k=1}^{N} P_{jk}$

$= \pi_j$. ✓

## <u>Markov Chain Monte Carlo</u> (MCMC)

Goal: Estimate an integral $\mathbb{E}(h(X)) = \int h(x) f(x)\, dx$.

Idea: Construct a Markov Chain $X_1, X_2, \ldots$ whose stationary distribution is $f$

$\Rightarrow \quad X_n \sim F = \int f$

We're specifying $\pi$, and trying to find $P$ such that $\pi = \pi P$.

If this can be done, then under certain assumptions

$$\frac{1}{N} \sum_{i=1}^{N} h(X_i) \xrightarrow{\mathbb{P}} \mathbb{E}(h(X)).$$

For example: Draw from posterior in Bayesian calculation: $f(\theta | x) = \dfrac{\mathcal{L}(\theta)\, f(\theta)}{C}$ $\quad \int \mathcal{L}(\theta)\, f(\theta)\, d\theta$

## Specific Algorithm   Metropolis - Hastings.

Listed as one of top 10 algorithms of $20^{th}$ century.
(along with FFT, FMM, QR, Fortran)

Goal: Draw samples from $X$ with density $f$.

**M-H Algorithm**

⓪ Choose $X_0$ arbitrarily. Assuming that we have generated $X_0, .., X_i$ :

① Generate $Y$ from density $q(y|X_i)$

proposal or candidate value

$q$ is a density that is easy to draw from : proposal distribution

Ex: $q(y|x) \sim N(x, \sigma^2)$.

② Evaluate $r = r(X_i, Y)$ where

$$r(x,y) = \min\left\{ \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)} , 1 \right\}$$

③ Set $X_{i+1} = \begin{cases} Y & \text{with probability } r \\ X_i & \text{with probability } 1-r \end{cases}$

Completely opaque algorithm, look at specific example first before understanding why it works.

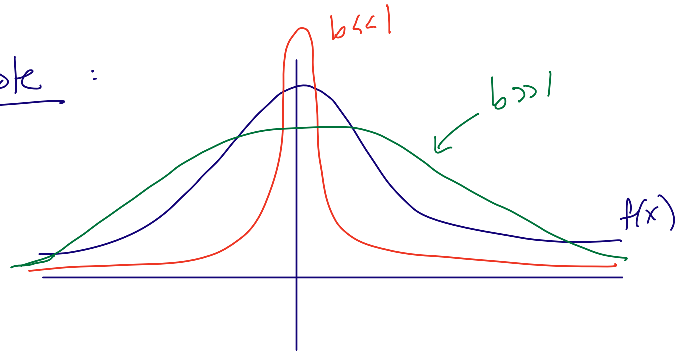
Ex: Draw from Cauchy distribution $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

Take $q(y|x) = \frac{1}{\sqrt{2\pi}\, b} e^{-(y-x)^2/2b^2}$.

So then
$$r(x,y) = \min\left\{ \frac{f(y)}{f(x)}\frac{q(x|y)}{q(y|x)}, 1 \right\}$$
$$= \min\left\{ \frac{1+x^2}{1+y^2}\frac{e^{-(x-y)^2/2b^2}}{e^{-(y-x)^2/2b^2}}, 1 \right\}$$
$$= \min\left\{ \frac{1+x^2}{1+y^2}, 1 \right\}$$

So the algorithm reduces to following:
$$X_{i+1} = \begin{cases} Y \sim N(X_i, b^2) & \text{with probability } r(X_i, Y) \\ X_i & \text{with prob. } 1 - r(X_i, Y) \end{cases}$$

Note:




Why does this algorithm work at all?

Short answer: We enforce detailed balance in the chain, therefore guaranteeing the existence of a stationary distribution.

<u>Recall:</u> $p_{ij}\pi_i = p_{ji}\pi_j$

Continuous version of detailed balance:

$$p_{ij} \to p(x,y) \approx \mathbb{P}(X_{n+1}=y \mid X_n = x)$$

$$\pi_i \to f(x) \approx \mathbb{P}(X_n \approx x).$$

The function $f$ is a stationary distribution if

$$f(y) = \int p(x,y)\, f(x)\, dx$$

$\Rightarrow$ Detailed Balance then means that

$$f(x)\, p(x,y) = f(y)\, p(y,x)$$

If this equation holds, then just integrate each side to show that $f$ is a stationary distribution.

Using the construction of the M-H algorithm, show that detailed balance is satisfied, and therefore $f$ is the stationary distribution.

Consider $x, y$ (i.e. $x = X_i$, and $y = Y$, the proposal value).

Either $f(x)\, q(y|x) < f(y)\, q(x|y)$

or $f(x)\, q(y|x) > f(y)\, q(x|y)$ $(*)$

Without loss of generality, assume that $(*)$ holds.

and we then have:

$$\frac{f(y)\, q(x|y)}{f(x)\, q(y|x)} \geq 1$$

and therefore $r(x,y) = \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)}$.

(And obviously $r(y,x) = \min\left\{ \underbrace{\frac{f(x)\, q(y|x)}{f(y)\, q(x|y)}}_{\geq 1}, 1 \right\} = 1$.)

Next, compute the transition probabilities:

$p(x,y) = \mathbb{P}(x \to y)$ and requires that

(i) generate $y$

(ii) accept $y$

$$\Rightarrow \quad p(x,y) = q(y|x) \cdot r(x,y) = \cancel{q(y|x)} \cdot \frac{f(y)}{f(x)} \frac{q(x|y)}{\cancel{q(y|x)}}$$

$$= \frac{f(y)}{f(x)} q(x|y)$$

$$\Rightarrow \quad f(x)\, p(x,y) = f(y)\, q(x|y)$$

On the other hand, $p(y,x) = \mathbb{P}(y \to x)$ and requires:

(i) generate $x$

(ii) accept $x$

$$\Rightarrow \quad p(y,x) = q(x|y)\, r(y,x) = q(x|y)$$

And therefore $\quad f(x)\, p(x,y) = f(y)\, p(y,x)$. ✓

This is detailed balance.

---

Monte Carlo methods:

$$= \int h(x) f(x)\, dx \approx \underbrace{\frac{1}{N} \sum_{j=1}^{N} h(X_i)}_{I}$$

where $X_i \sim$ samples from $f$

$$\mathbb{E}(I) = \int h f$$

$$\mathrm{Var}(I) \propto \frac{1}{N} \quad \Rightarrow \quad \mathrm{std}(I) \sim \frac{1}{\sqrt{N}}$$