ADVANCED STOCHASTIC PROCESSES - 8° LECTURE

The hypothesis of positive recurrence is the most difficult to be checked. However, there are many probability tools that can be adopted, for example generating functions:

<u>Example</u>: Consider $(S_n)_{n\geq 0}$ the RW on $\mathbb{Z}$ s.t.

$$S_0 = 0 \;,\quad S_n = \sum_{k=1}^{n} X_k \;,\quad \mathbb{P}(X_k = +1) = \mathbb{P}(X_k = -1) = \tfrac{1}{2} \;,\; \forall k \in \mathbb{N}$$

(called SSRW)

We have already checked that

S is recurrent $\left(\sum_{n=0}^{+\infty} p_{0 0}^{(n)} = +\infty \right)$.

Positive or null recurrent? For $z \in \mathbb{C}$, $|z|<1$, consider

$$\varphi(z) = \sum_{n\geq 0} \mathbb{P}_0(X_{2n} = 0)\, z^{2n} = 1 + \sum_{k\geq 1} \sum_{n\geq 1} \mathbb{P}_0(\tau_0^k = 2n)\, z^{2n}$$

$$= 1 + \sum_{k\geq 1} \mathbb{E}_0\left(z^{\tau_0^k}\right) \overset{*}{=}$$

Note that $\tau_0^k = \tau_0 + (\tau_0^2 - \tau_0) + \dots + (\tau_0^k - \tau_0^{k-1})$. By strong Markov property

$$\varphi(z) \overset{*}{=} \sum_{k\geq 0} \left(\mathbb{E}_0\left(z^{\tau_0}\right)\right)^k = \underline{\frac{1}{1 - \mathbb{E}_0\left(z^{\tau_0}\right)}}$$

On the other hand, by explicit computation:

$$\varphi(z) = \sum_{n\geq 0} p_{00}^{(n)} z^{2n} = \sum_{n\geq 0} \binom{2n}{n} \left(\frac{z}{2}\right)^{2n} = \sum_{n\geq 0} \binom{-\frac{1}{2}}{n} (-z^2)^n = (1 - z^2)^{-\frac{1}{2}} = \frac{1}{\sqrt{1 - z^2}}$$

$$\binom{2n}{n} = \binom{-\frac{1}{2}}{n}(-4)^n \qquad \text{generalized binomial theorem } (*)$$

$$\Longrightarrow \mathbb{E}_0\left(z^{\tau_0}\right) = 1 - \sqrt{1 - z^2}$$

$$\Longrightarrow \mathbb{E}_0(\tau_0) = \lim_{z\to 1} \frac{d}{dz} \mathbb{E}_0\left(z^{\tau_0}\right) = \lim_{z\to 1} \frac{z}{\sqrt{1 - z^2}} = +\infty$$

In conclusion, the SSRW on $\mathbb{Z}$ is null recurrent.

$(*)$ $\forall z \in \mathbb{C}, |z|<1$ and $a \in \mathbb{R}$: $(1+z)^a = \sum_{n\geq 0} \binom{a}{n} z^a$, where $\binom{a}{n} = \dfrac{a(a-1)\cdots(a-n+1)}{n!}$

Ergodic Theorem

Let X be a MC on S irreducible and positive recurrent (or ergodic)

$\rightarrow \exists$ unique invariant distribution π.

- As a main result, we will show that, $\forall x \in S$, $\pi(x)$ approximates the proportion of time spent in x by the MC: $N_n(x)/n \approx \pi(x)$.
- This result can be strengthen to show that the average $\pi(f)$ of a function $f: S \to \mathbb{R}$ w.r.t. π can be well approximated by taking the time average along the chain: $\sum_{k=1}^{n} f(X_k)/n \approx \pi(f)$

Theorem [ergodic]:

Let X be a ergodic MC with (unique) stationary distribution π. and $f: S \to \mathbb{R}$ s.t. $\pi(|f|) < \infty$. Then, $\forall \mu \in \mathcal{P}(S)$ initial measure,

$$(1) \qquad \frac{\sum_{k=1}^{n} f(X_k)}{n} \xrightarrow[n \to \infty]{} \pi(f) \qquad \mathbb{P}_\mu\text{-a.s.}$$

In particular, for $f = \mathbb{1}_{\{x\}}$, $x \in S$ fixed, it holds

$$(2) \qquad \frac{N_n(x)}{n} \xrightarrow[n \to \infty]{} \pi(x) \qquad \mathbb{P}_\mu\text{-a.s.}$$

Notice that the above statement can be seen as a strong law of large number for sum of (correlated) random variables.
Indeed, for given $x \in S$, let $Y_k := \mathbb{1}_{\{X_k = x\}}$, $\forall k \geq 0$, so that $Y_k \sim Be(\mu P^k(x))$, $\forall k \geq 0$ and w.r.t. $\mathbb{P}_\mu$, and set $S_n := \sum_{k=1}^{n} Y_k \equiv N_n(x)$. Then (i) $\iff$ SLLN for $(Y_k)_{k \geq 0}$ w.r.t. $\mathbb{P}_\mu$

Proof [Ergodic Theorem]: We claim that

Proof [Ergodic Theorem]: We claim that

$$\text{(c)} \quad \lim_{n\to\infty} \sum_{k=1}^{n} \frac{f(X_k)}{N_n(x)} = \mu_x(f), \quad \mathbb{P}_\mu\text{- a.s.}$$

Applying the claim to $f=1$, we get

$$\lim_{n\to\infty} \frac{n}{N_n(x)} = \sum_{y\in S} \mu_x(y) = \mathbb{E}_x(\tau_x) < \infty, \quad \mathbb{P}_\mu\text{-a.s.}, \text{ and then}$$

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{f(X_k)}{n} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} f(X_k)/N_n(x)}{n/N_n(x)} \underset{\mathbb{P}_\mu\text{-a.s.}}{=} \frac{\mu_x(f)}{\mathbb{E}_x(\tau_x)} = \sum_{y\in S} \underbrace{\frac{\mu_x(y)}{\mathbb{E}_x(\tau_x)}}_{\text{"}\pi(y)\text{"}} f(y)$$

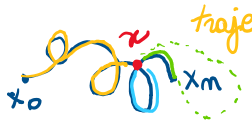
$$= \pi(f)$$

hence concluding the proof of The Theorem.

To prove the claim, first notice that

$$\tau_x^{N_x(n)} \le n < \tau_x^{N_x(n)+1}$$

it takes some time.
A picture may help.



trajectory of lenght n can be splitted in paths among visits to x, plus a remaining term which is not yet arrived to x

- Consider first $\sum_{k=1}^{\tau_x^m} \frac{f(X_k)}{m}$, with $m = N_x(n)$ for convenience

$$\circledast \quad \sum_{k=1}^{\tau_x^m} \frac{f(X_k)}{m} = \sum_{k=1}^{\tau_x} \frac{f(X_k)}{m} + \sum_{k=\tau_x+1}^{\tau_x^2} f(X_k) + \dots + \sum_{k=\tau_x^{m-1}+1}^{\tau_x^m} f(X_k)$$

Let us define: $Y_j = \sum_{k=\tau_x^{j-1}+1}^{\tau_x^j} f(X_k)$, $\forall\, j = 1,\dots,m$

and $\tau_x^0 = 0$ by convention

By the strong Markov property $(Y_j)_{j=1,\dots,m}$ are iid with average

$$\mathbb{E}_x(Y_j) = \mathbb{E}_x\left(\sum_{k=1}^{\tau_x} f(X_k)\right) = \mathbb{E}_x\left(\sum_{y\in S} f(y) \sum_{k=1}^{\tau_x} \mathbb{1}_{\{X_k=y\}}\right)$$

$$= \sum_{y\in S} f(y)\, \underbrace{\mathbb{E}_x\left(N_{\tau_x}(y)\right)}_{\mu_x(y)} = \sum_{y\in S} f(y)\mu_x(y) = \mu_x(f)$$

Note: Y is independent of Y, Y ... but it may have a

$\mu_x(y)$

Note: Y_1 is independent of $Y_2, \dots, Y_m$, but it may have a different distribution since the initial measure is $\mu \neq \delta_x$. However it is just one term of the sum in ⊛ and it will not give contribution as it is divided by $m = N_{T_x}(n) \xrightarrow[n\to\infty]{} \infty$.

In conclusion, by the SLLN, we get $(\forall \mu \in \mathcal{P}(S))$

$$\frac{\sum_{k=1}^{T_x^m} f(X_k)}{m} = \frac{\sum_{k=1}^{m} Y_k}{m} \xrightarrow[m\to\infty]{} \mu_x(f), \quad \mathbb{P}_\mu\text{-a.s.}$$

- At last, notice that

$$\underbrace{\frac{\sum_{k=1}^{T_x^n} f(X_k)}{N_x(n)}}_{\downarrow\, \mathbb{P}_\mu\text{-a.s.}\ \mu_x(f)} \le \sum_{k=1}^{n} \frac{f(X_k)}{N_x(n)} \le \underbrace{\sum_{k=1}^{T_x^{n+1}} \frac{f(X_k)}{N_x(n)}}_{\downarrow\, \mathbb{P}_\mu\text{-a.s.}\ \mu_x(f)}$$

$\frac{N_x(n)}{N_x(n+1)} \xrightarrow[n\to\infty]{} 1 \quad \mathbb{P}_\mu\text{-a.s}$

concluding the proof of the claim and of the Theorem. #

The ergodic theorem provides the theoretical basis of Monte Carlo Markov chains (MCMC). Let us first define reversibility:

Def. [Reversible MC]:

Let $(X_n)_{n\geq 0}$ be a MC on S with tran. matrix P, and $\pi \in \mathcal{P}(S)$.

Then X is reversible w.r.t. π if $\forall x, y \in S$

$$\boxed{\pi(x) P_{x,y} = \pi(y) P_{y,x}}$$

detailed balance condition

Remark: if X is reversible w.r.t π, then:

1. π is invariant distr: $\pi P(x) = \sum_y \pi(y) P_{y,x} = \sum_y \pi(x) P_{x,y} = \pi(x)$
2. The evolution of the MC backward in time has the same transition probabilities: $\forall\, m \in \mathbb{N}$ fixed, let $(\tilde{X}_k)_{k \le m}$ s.t. $\tilde{X}_k := X_{m-k}$ (notice that k can be taken negative)

$\tilde{X}_k := X_{n-k}$ (notice that k can be taken negative)

Then $\mathbb{P}_\pi(\tilde{X}_{k+1}=y \mid \tilde{X}_k = x) \underset{\text{Bayes and 1.}}{=} \dots = \dfrac{\pi(y)P_{y,x}}{\pi(x)} = P_{x,y}$

3. It holds the cycle condition:

$$P_{x_0,x_1} \cdot P_{x_1,x_2} \cdots P_{x_{n-1}x_n} \cdot P_{x_n x_0} = P_{x_n,x_{n-1}} \cdots \cdot P_{x_1,x_0} \cdot P_{x_0,x_n}$$

Example:

Let $(S_n)_{n\geq 0}$ be a random walk on $\mathbb{T}_L = \dfrac{\mathbb{Z}}{L\mathbb{Z}}$ (L odd)

s.t. $\mathbb{P}(S_n = k \mid S_{n-1} = j) = \begin{cases} p & \text{if } k=j+1 \\ 1-p & \text{if } k=j-1 \\ 0 & \text{otherwise} \end{cases}$

Then, it is easy to verify that $\exists!$ invariant distr. $\pi = \mathcal{U}(\mathbb{T}_L)$

(indeed: $p\cdot\frac{1}{L} + (1-p)\cdot\frac{1}{L} = \frac{1}{L}$)

However, the detailed balance condition $\frac{1}{L}p \overset{?}{=} \frac{1}{L}(1-p)$ holds only if $p=\frac{1}{2}$, hence (S_n) is reversible only if $p=\frac{1}{2}$, which is intuitive as if $p\neq\frac{1}{2}$, the dynamic is drifted in one direction ($p>\frac{1}{2}$ ↻, $p<\frac{1}{2}$ ↺)

Markov Chain Monte Carlo (MCMC)

It is a method to sample a distribution π on S finite, or to compute the average value of function $\varphi: S \to \mathbb{R}$.

It is relevant when S is very large (e.g. $S = E^N \to$ voter model) so that π is defined up to a normalization constant, namely

$$\pi(x) = c\, h(x) \quad , \quad c^{-1} = \sum_{x\in S} h(x) \quad \text{large sum difficult to compute} \quad ,$$

and very efficient in terms of algorithmic time.

Idea: The SLLN for iid r.v. $(X_n)_{n\in\mathbb{N}}$ with common law π allows to approximate the average of φ w.r.t. π with an allows to approximate the average of φ w.r.t. π with an empirical average of $(\varphi(X_n))_{n\in\mathbb{N}}$:

this holds for a arbitrary space S with φ bounded

$$(a) \quad \sum_{x\in S}\varphi(x)\pi(x) = \mathbb{E}_{\pi}(\varphi(X)) \approx \frac{1}{n}\sum_{k=1}^{n}\varphi(X_k)$$

However, we should be able to sample $X_n \sim \pi$, for given π.

Goal: If we are able to construct an ergodic MC $(X_n)_{n\geq 0}$ on S with invariant measure π, we can use the ergodic theorem to get the approximation (a) along the MC.

In practise, we construct ergodic MC that are reversible w.r.t. π with a two step procedure:

1° STEP: Consider an arbitrary irreducible transition matrix Q, s.t. $q_{x,y}>0 \Leftrightarrow q_{y,x}>0$. If $X_n = x$, with probability $q_{x,y}$ we propose the transition to y. PROPOSAL

2° STEP: We accept the transition with prob. $\alpha(x,y) \in [0,1]$, so that

$$P_{x,y} = \mathbb{P}(X_{n+1}=y \mid X_n = x) = \begin{cases} q_{x,y}\cdot\alpha(x,y) & \text{if } x\neq y \\ 1-\sum_{z\neq x} P_{x,z} & \text{if } x=y \end{cases}$$

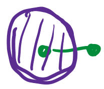
(ACCEPTANCE)

We only have to find $\alpha(x,y)$ to obtain reversibility w.r.t π, which is the following:

$$\underbrace{\forall x\neq y \text{ s.t. } q_{x,y}\neq 0}_{\text{otherwise, trivial identity}} : \quad \boxed{\underbrace{\pi(x)}_{=\cancel{c}\cdot h(x)}\alpha(x,y)q_{x,y} = \underbrace{\pi(y)}_{=\cancel{c}\,h(y)}\alpha(y,x)q_{y,x}}$$

$$\Longrightarrow \quad \alpha(x,y) = \frac{\pi(y)\,q_{y,x}}{\pi(x)\,q_{x,y}}\cdot\alpha(y,x) = \frac{h(y)\,q_{y,x}}{h(x)\,q_{x,y}}\,\alpha(y,x) \quad (b)$$

Example: The hard-core model

- Consider a finite graph $G=(V,E)$ with $V=\{1,2,\dots,N\}$
- At each vertex $i\in V$ is associated a value $x_i\in\{0,1\}$ that tell us if the vertex is free ($x_i=0$) or occupied ($x_i=1$)
- The whole configuration of particles on G is described by
  $x=(x_i)_{i\in V}\in\{0,1\}^V=:S$
- The allowed (or feasible) configurations are a subset
  $F\subset S$: $F=\{x\in S : x_i\cdot x_j=0 \ \forall\, i\sim j\}$
  where $i\sim j$ means that $(i,j)\in E$.
  Equivalently $x\in F$ if particles are not in neighboring vertices (hard constraint : )
- Let $\pi\in\mathcal{P}(S)$ s.t. $\pi(x)=\frac{1}{|F|}\mathbb{1}_F(x)$

Goal: Approximate π by constructing a MC (on F) which is irreducible and reversible w.r.t. π.

Construction:

1. Start with $x\in F$ (for example, the empty configuration)
   $\hookrightarrow X_0\sim\delta_x$ (e.g. $x=\underline{0}=(0,\dots,0)$)

2. Given $X_n=x\in F$, we construct X_{n+1} in two steps:
   i. Choose a vertex $i\in V$ uniformly at random (with prob. $\frac{1}{N}$) and propose to update $X_n(i)$ (update only at site i).

   ii. * if $x_j=0 \ \forall\, j\sim i$, then set $X_{n+1}(i)=\begin{cases}1-x_i & \text{with prob } \frac12\\ x_i & \text{with prob } \frac12\end{cases}$
   (and $X_{n+1}(j)=x_j \ \forall\, j\neq i$)
   * if $x_j=1$ for some $j\sim i$, then set $X_{n+1}=X_n=x$.

* if $x_j = 1$ for some $j \sim i$, then set $X_{n+1} = X_n = x$.

In conclusion, $P_{x,y} = \begin{cases} 0 & \text{if } d(x,y) = \#\{i : x_i \neq y_i\} \geq 2 \\ & \text{or if } y \notin F \\ \frac{1}{2N} & \text{if } d(x,y) = 1 \text{ and } y \in F \\ 1 - \sum_{z \neq x} P_{x,z} & \text{if } y = x \end{cases}$

The described MC is irreducible (check!) and reversible w.r.t π.

Indeed: $\pi(x) P_{x,y} \overset{?}{=} \pi(y) P_{y,x}$

if $d(x,y) \geq 2$ $\qquad 0 = 0$

if $d(x,y) = 1$ $\quad \frac{1}{|F|} 1\!\!1_F(x) \cdot \frac{1}{2|V|} 1\!\!1_F(y) = \frac{1}{F} 1\!\!1_F(y) \cdot \frac{1}{2|V|} 1\!\!1_F(x)$ ✓

There are many solution to (b). For example:

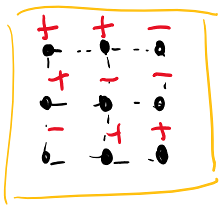
Hastings - Metropolis algorithm $\longrightarrow \alpha(x,y) = \min\left\{1, \frac{h(y) q_{y,x}}{h(x) q_{x,y}}\right\}$

↓

Example: Ising model on a lattice $\Lambda = \{0, \dots, N-1\}^d$

↳ ferromagnetism in cristals

$|\Lambda| = N^d$



$\Lambda \subset \mathbb{Z}^2$

General assumptions
- The sites of Λ are where atoms are placed
- each atom has a magnetic spin $x(i) \in \{+1, -1\}$, $\forall i \in \Lambda$
- neighboring atoms in Λ interact

$i = (i_1, \dots, i_d)$ and $j = (j_1, \dots, j_d)$ are neighbors if $\sum_{\ell=1}^{d} |i_\ell - j_\ell| = 1$. Then write $i \sim j$

Let $S = \{+1, -1\}^{|\Lambda|}$ (so that $|S| = 2^{N^d}$) $\longrightarrow$ compare with voter model

The Ising probability measure on Λ on S is defined by the density

The Ising probability measure on Λ on S is defined by the density

Gibbs probability measure $\pi(x) = \dfrac{e^{-\beta H(x)}}{Z_\beta}$, where $\begin{cases} \beta > 0 \text{ is a fixed parameter} & (\text{inverse temperature}) \\ H(x) = -\sum_{i \in \Lambda} \sum_{j \sim i} x(i)\, x(j) & (\text{energy of the state } x) \end{cases}$

and $Z_\beta = \sum_{x \in S} e^{-\beta H(x)}$ (normalization costant, called partition function)

While is difficult to get the precise value of Z_β, following the Metropolis procedure it is easy to construct an ergodic MC to sample from π: * Let $q_{xy} = \begin{cases} \frac{1}{|\Lambda|} & \text{if } \sum_{i \in \Lambda} |x(i) - y(i)| = 1 \\ 0 & \text{otherwise} \end{cases}$, $\forall x, y \in S$

Exercise

Determine $\alpha(x,y)$ to get a reversible measure w.r.t. π.