DTMC and Applications

<u>Def</u>: $(X_n)_{n\in\mathbb{Z}_{\geq 0}}$ is DTMC if it satisfies Markovian property: $P(X_{n+1}=i_{n+1} \mid X_k = i_k, 1\le k\le n) = P(X_{n+1}=i_{n+1} \mid X_n = i_n)$

<u>Rmk</u>: i) The equation can be extended:

$P(X_{n+1}\in A_{n+1} \mid X_k \in A_k, 1\le k\le n) =$

$P(X_{n+1}\in A_{n+1} \mid X_n\in A_n)$. for $A_k \subset S$, $1\le k\le n+1$.

ii) A: past. B: present. C: future.

$\Rightarrow P(C\mid BA) = P(C\mid B)$. depend on present.

$P(AC\mid B) = P(A\mid B)P(C\mid B)$. A, C have the same role.

A DTMC is specifyed by:

i) State space S (discrete). at most countable.

ii) Initial dist. λ_0. $X_0\sim\lambda_0$.

iii) Prob. transition rules: $P=(P_{ij})$.

(1) <u>Metroplis Method</u>:

To simulate a dist. λ on finite states space S. One way is using Limit Thm to approxi.

On a connected graph. Denote $N(j)$ is set of neighbour points of j. $d(i)$ is $\# N(i)$

① Run a random Walk on the graph:

$$P_{rw}(i,j) = \begin{cases} 1/d(i) \,. & j \in N(i) \\ 0 \,. & j \notin N(i) \end{cases}$$

$\pi_{rw}(i) = d(i) / \sum_{j \in S} d(j)$, is stationary dist.

② <u>Modification</u>:

Note $\pi_{rw}(i) \propto d(i)$. if We want: $\pi(i) \propto f(i) d(i)$

$$\text{Set}: \ p(i,j) = \begin{cases} \min\{1, f(j)/f(i)\}/d(i) \,. & j \in N(i) \\ 1 - \sum_{j \in N(i)} p(i,j) \,. & j = i \\ 0 & \text{. otherwise} \end{cases}$$

<u>Pf</u>: $\pi(i) p(i,j) \propto \min\{f(i), f(j)\}$. By symmetry:

$\pi(i) p(i,j) = \pi(j) p(j,i) \Rightarrow \pi P = \pi$. stationary.

<u>Interpret</u>: 1) Choose a neighbour of i w.p. $1/d(i)$

2) If $f(j) \geq f(i)$. Then accept j.

3) If $f(j) < f(i)$. Then reject w.p. $1 - \frac{f(j)}{f(i)}$

<u>Rmk</u>: We don't need to know $(\pi(i))$. exactly.

Avoid Normalization.

(2) <u>Simulated Annealing</u>:

Target: Find $i \in S$. st. $c(i) = \min_S c(k)$. where $c(i)$ is the cost. function. def on set of nodes S in a graph. $|S| < \infty$.

Def: prob. dist. $\alpha_T = \{\alpha_T(i) \mid i \in S\}$ on S is:

$\alpha_T(i) = d(i) e^{-c(i)/T} / G(T)$. $G(T) = \sum_S d(i) e^{-c(i)/T}$

Rmk: i) Choose $f(i) = e^{-c(i)/T}$ in Metroplis Metod

$\pi(i) \propto d(i) f(i) \Rightarrow \pi(i) = \alpha_T(i)$.

ii) T stands for temperature.

Denote: $S^* = \{i^* \in S \mid c(i^*) = \min_S c(k)\}$. With a dist.

α^* on S^* is: $\alpha^*(i) = d(i) / \sum_{S^*} d(i)$. if $i \in S^*$.

$\alpha^*(i) = 0$. if $i \notin S^*$.

Rmk: We only put positive prob. on optimals.

prop. $\alpha_T \xrightarrow{d} \alpha^*$. as $T \downarrow 0$.

Pf: Check $\alpha_T(i) \to \alpha^*(i)$: divide $e^{-\min c(i)/T}$. Set $T \downarrow 0$

procedure: 1°) Simulate the stationary α_{T_1} by Metroplis Method. $A_{T_1} = (A_{T_1}(i,j))_{S \times S}$ is its prob. transition matrix. def by (1) ②.

2°) When dist. is approaching α_{T_1} nearly. Decrease the temperature to $T_2 < T_1$.

Then approxi. π_{T_2} by A_{T_2} prob. matrix.

3') Repeat these steps. Let $T_n \to 0$ $(n \to \infty)$

Next. Suppose we have $X_0 \sim \nu_0$. We will cool the temperature at each step. i.e. $P(X_n = i \mid X_{n-1} = j) = A_{T_{n-1}}(j, i)$

i.e. $X_n \sim \nu_0 A_{T_0} A_{T_1} \cdots A_{T_{n-1}}$ $\quad T_n \downarrow 0$.

<u>Denote:</u> $A^{(n)} = \prod_{0}^{n-1} A_{T_k}$. $\quad \nu^{(n)} = \nu_0 A^{(n)}$.

$r = \min_{i \in S_c} \max_{j \in S} \ell(i,j)$, ℓ is distance. and denote:

$S_c = \{ i \mid c(j) > c(i),\ \text{exist some } j \in N(i) \}$

$L = \max_{i \in S} \max_{j \in N(i)} | c(j) - c(i) |$. max local fluctuation.

<u>Rmk:</u> S_c is set of points which is not local max. Then r is min radius of $i \in S_c$. st. contain all $j \in S$

$\Rightarrow$ Our goal is simulating π^*:

<u>Thm.</u> (Main Thm)

For cooling schedule $(T_i)_{i \geq 0}$. satisfies:

i) $T_{n+1} < T_n$. $\forall n \geq 0$. $\quad$ ii) $T_n \to 0$ $(n \to \infty)$

iii) $\sum_k e^{-rL/T_{rk-1}} = \infty$. Then, we have:

$$\| A^{(n)}(i, \cdot) - \pi^* \| = \sup_{s \in S} | A^{(n)}(i, s) - \pi^*(s) | \xrightarrow{n \to \infty} 0. \ \forall i$$

<u>Pf:</u> It's application of theory of time-inhomogeneous Markov chain. will be proven later.

Rmk: We can schedule proper temperatures

$(T_n) = (\gamma / \log n)$. $\gamma \geq rL$.

(3) Ergodicity of

Time-inhomogenous MC:

Consider a time-inhomogenous Markov Chain (X_n).

From X_n to X_{n+1}, it has different prob. transition

matrix P_n. Denote: $P^{(m,n)} = \prod_{m}^{n-1} P_k$.

Def: i) (X_n) is strongly ergodic if $\exists$ dist. π^*

s.t. $\lim_{n\to\infty} \sup_{i\in S} \| P^{(m,n)}(i,\cdot) - \pi^* \| = 0$. $\forall m$

ii) (X_n) is weakly ergodic if:

$\lim_{n\to\infty} \sup_{i,j} \| P^{(m,n)}(i,\cdot) - P^{(m,n)}(j,\cdot) \| = 0$. $\forall m$

Rmk: i) "Weak ergodic" is kind of "loss of memory" after a long time. But: it's unnecessary to converge to some dist.

ii) "$\forall m$" is because we don't want some prob. matrix P_i to determine the whole convergence. (like seq converge)

① Ergodic Coefficient:

Denote: $\delta(P) = \sup_{i,j} \| P(i,\cdot) - P(j,\cdot) \|$. for prob matrix P.

Rmk: $\delta(P)=0 \Rightarrow$ Next move doesn't depend on current state. i.e. lose memory totally.

Lemma $\delta(PQ) \le \delta(P)\delta(Q)$. $\forall P, Q$. prob. matrix.

Pf: Set $A = \{k \mid (PQ)(i,k) > (PQ)(j,k)\}$.

$$\forall i,j \in S: \sum_{k\in S} ((PQ)_{ik} - (PQ)_{jk})^+ = \sum_{k\in A} (\square - \square)$$

$$= \sum_{k\in A}\sum_{l\in S} P_{il}Q_{lk} - P_{jl}Q_{lk}$$

$$= \sum_{l\in S} (P_{il} - P_{jl}) \sum_{k\in A} Q_{lk}$$

$$\le \sum_{S} (P_{il}-P_{jl})^+ \sup_{l} \sum_{k\in A} Q_{lk} - (P_{il}-P_{jl})^- \inf_{l} \sum_{A} Q_{lk}$$

$$\overset{\sum P_{il} - P_{jl} = 0}{=} \sum_{l\in S} (P_{il}-P_{jl})^+ (\sup_{l} \sum_{A} Q_{lk} - \inf_{l} \sum_{A} Q_{lk})$$

$$\le \sum_{S} (P_{il}-P_{jl})^+ \delta(Q) \le \delta(P)\delta(Q).$$

Lemma. For dist. λ, μ. and prob matrix P. We have:

$$\|(\lambda-\mu)P\| \le \|\lambda-\mu\|\delta(P).$$

Pf. $\forall M \subseteq S$. $|\sum_{M} (\lambda-\mu)P(i)| = |\sum_{i\in M}\sum_{t\in S} (\lambda-\mu)_t P_{ti}|$

$$= |\sum_{S} (\lambda-\mu)_t \sum_{i\in M} P_{ti}| \quad (\text{By } \sum(\lambda-\mu)_t = 0)$$

$$\le \sum_{t\in S} (\lambda-\mu)_t^+ (\sup_{t} \sum_{M} P_{ti} - \inf_{t} \sum_{M} P_{ti})$$

$$\le \sum_{S} (\lambda-\mu)_t^+ \delta(P) = \|\lambda-\mu\|\delta(P).$$

Lemma If $\exists$ column of Q. prob. matrix. whose entries $\ge r$.

Then: $\delta(a) \le 1-n$.

Pf. $\delta(a) = \sup_{i,j} \sum_S (a_{ik} - a_{jk})^+ = \sup_{i,j} \sum_A (a_{ik} - a_{jk})$

$\le \sup_{i,j} \sum_A a(i,k) - n \le 1-n$

where suppose that column has index i_0.

WLOG. $i_0 \in A = \{a(i,k) > a(j,k)\}$. Otherwise: consider A^c. since $\sum a_{ik} - a_{jk} = 0$.

prop. If $\exists (n_k) \uparrow \infty$. $\sum (1 - \delta(P^{(n_k, n_{k+1})})) = \infty$. Then: (X_n) is weakly ergodic.

Pf. By $1-x \le e^{-x}$. We have: $\prod_m \delta(P^{(n_k, n_{k+1})}) = 0$. $\forall m$.

$\Rightarrow \delta(P^{(n_k, n_L)}) \le \prod_{i \ge k}^{L-1} \delta(P^{(n_i, n_{i+1})}) \to 0$. $(L \to \infty)$

So: $\delta(P^{(m,n)}) \lesssim \delta(P^{(m, n_k)}) \prod_{i \ge k}^{L(n)} \delta(P^{(n_i, n_{i+1})}) \xrightarrow{n \to \infty} 0$

prop. If (X_n) is weakly ergodic. $\exists (\pi_k)_{k \ge 0}$ stationary dist. for $(P_k)_{k \ge 0}$. s.t. $\sum \| \pi_k - \pi_{k+1} \| < \infty$. Then: (X_n) is strongly ergodic. $\pi^* = \lim_n \pi_n$.

Pf. 1) $\sum \| \pi_k - \pi_{k+1} \| < \infty \Rightarrow \lim \pi_n = \pi^*$ exist.

$\| P^{(k,m)}(i,\cdot) - \pi^*(\cdot) \| \le \| P^{(k,m)}(i,\cdot) - \pi_l P^{(l,m)} \|$

$+ \| \pi_l P^{(l,m)} - \pi_m \| + \| \pi_m - \pi^* \| =: A_1 + A_2 + A_3$

2) $A_1 \le \| P^{(k,l)}(i,\cdot) - \pi_l(\cdot) \| \delta(P^{(l,m)})$

$\to 0$. (as $m \to \infty$). $\forall l$.

3') A_2 : $\pi_l P^{(l,m)} - \pi_m = \pi_l p_l \cdot P^{(l+1,m)} - \pi_m$

$$= \pi_l \cdot P^{(l+1,m)} - \pi_m$$

$$= \pi_{l+1} P^{(l+1,m)} - \pi_m + (\pi_l - \pi_{l+1}) P^{(l+1,m)}.$$

$$= \cdots = \sum_{l}^{m-1} (\pi_k - \pi_{k+1}) P^{(k+1,m)}$$

$$\Rightarrow A_2 \lesssim \sum_{l}^{m-1} \| \pi_k - \pi_{k+1} \| \to 0 \ (m \to \infty). \ (\delta(p) \le 2).$$

4') $A_3 \to 0$ $(m \to \infty)$. by 1'). directly.

② Proof of Main Thm:

<u>Denote</u>: Node i_0 is called center. if : $\max_{j \in S} \ell(i_0, j)$

$$= r = \min_{S_c} \max_{S} \ell(i,j) \ , \quad D = \max_i d(i).$$

1') Show : $\sum_{k \ge 1} (1 - \delta(P^{(kr-r,\,kr)})) = \infty$. $P_n =: A_{T_n}$.

<u>Pf</u>: For $i \ne j$: $P_{ij}(n) = \min\{1, e^{-\ell(j)+\ell(i)/T_n}\}/d(i) \ge e^{-L/T_n}/D$.

$$P_{i_0,i_0}(n) = 1 - \sum_{j \in N(i)} P_{ij}(n) \xrightarrow{n \to \infty} 1 - \sum_{\substack{j \in N(i) \\ \ell(j) \le \ell(i)}} 1/d(i)$$

$$= \sum_{\substack{j \in N(i) \\ \ell(j) > \ell(i)}} 1/d(i) > 0 \quad \text{by } i_0 \in S_c.$$

$\Rightarrow P_{i_0,i_0}(n) \ge e^{-L/T_n}/D$. for $\forall$ large n.

So : $P^{(m-r,m)}(i, i_0) \ge e^{-rL/T_m}/D^{r}$. $\forall i \in S$. $\forall m$.

By Lemma: $(1 - \delta(P^{(m-r,m)})) \ge n$.

2') Show : $\sum \| \pi_n - \pi_{n+1} \| < \infty$. where $\pi_n(i) = \dfrac{d(i) e^{-\ell(i)/T_n}}{G(T_n)}$

for strongly ergodic.

Lemma: i) $i \in S^* \Rightarrow \pi_n(i) < \pi_{n+1}(i)$. $\forall n$.

ii) $i \notin S^* \Rightarrow \exists \tilde{n}_i$. s.t. $\pi_{n+1}(i) < \pi_n(i)$. $\forall n \geq \tilde{n}_i$.

Pf: i) is directly check.

ii) $f_i(T) =: e^{-c(i)/T} / G(T)$.

$$f_i'(T) = \lambda(T) \left(\sum_S d(k) e^{-c(k)/T} (c(i) - c(k)) \right)$$

$$\geq \lambda(T) \left(\lambda_1 e^{-\frac{c(i)}{T}} + \lambda_2 e^{-\frac{c(m)}{T}} \right).$$

where $\lambda(T) \geq 0$. $\lambda_1 = \sum_{c(k) \geq c(i)} d(k)(c(i) - c(k))$

and $\lambda_2 = \sum_{c(k) < c(i)} d(k)(c(i) - c(k))$. $c(m) = \min_{\substack{k \\ c(k) < c(i)}} c(k)$

by $i \notin S^*$. $(\{k \mid c(k) < c(i)\} \neq \emptyset)$

$\Rightarrow$ for sufficient small T. ($T_n \downarrow$ as $n \uparrow$)

Return to pf:

Choose $\hat{n} = \max_{i \notin S^*} \{\tilde{n}(i)\}$. $\sum_{\hat{n}}^{\infty} \| \pi_n - \pi_{n+1} \| = \sum_{\hat{n}}^{\infty} \sum_{i \in S} (\pi_{n+1} - \pi_n)_i^+$

$= \sum_{i \in S^*} \sum_{\hat{n}} (\pi_{n+1}(i) - \pi_n(i))^+$. by monotone Lemma.

$$\Rightarrow \sum_{\hat{n}}^{\infty} \| \pi_n - \pi_{n+1} \| \leq \sum_{i \in S^*} \sum_{\hat{n}} (\pi_{n+1}(i) - \pi_n(i))$$

$$= \sum_{S^*} (\pi^*(i) - \pi_{\hat{n}}(i)) \leq \pi^*(S^*) = 1$$

(4) Card Shuffling:

Q: How close is the desk of cards being random after n shufflings?

i) <u>Rule</u>: Take the top card and insert it into a random position of the desk. equally likely. (May be top again)

ii) <u>Markov Chain</u>:

State space: S_{52} permutation of desk of cards. so that $|S_{52}| = 52!$

initial dist.: $\lambda_0(\beta) = 1$. $\beta = \{$ card k on position $k\}$

($\downarrow$ position 1 → position 52)

<u>Rmk</u>: Stationary λ is: $\lambda(\alpha) = 1/52!$. $\forall \alpha \in S_{52}$.

and note: (X_n) is irred. aperiodic. MC.

$\Rightarrow \|\lambda_n - \lambda\| \to 0$. $(n \to \infty)$

① Cost of random time:

We will find a random time T. st. $X_T \sim Uni(52!)$.

<u>Pef</u>: i) U_n is the position of top card move at the n^{th} shuffling. $U_n \overset{i.i.d}{\sim} Unif(52)$.

ii) $T_1 = \inf\{n \geq 1 \mid U_n = 52\}$. $T_k = \inf\{n > T_k \mid U_n \geq 53-k\}$.

<u>Rmk</u>: i) T_k means the time when the top card is inserted below card 52 (so it's in a random position). At time $T = T_{51}+1$. We insert card 52 to a random position. Then we have a desk shuffled randomly!

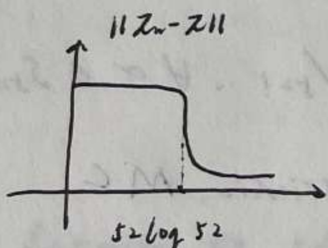
ii) We don't know T_k exactly happen, but know $T = T_{51}+1$, since the top card is 52 at time T_{51}.

For finding $E(T)$:

$T_1 \sim Geo(1/52)$, $T_2 - T_1 \sim Geo(\frac{2}{52})$ ... $T_k - T_{k-1} \sim Geo(\frac{k}{52})$

$T = \sum_2^{52} (T_k - T_{k-1}) + T_1 \Rightarrow E(T) = 52 (\sum_1^{52} 1/k) \sim 52 \log 52$.

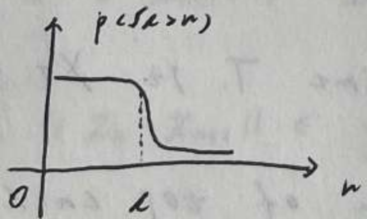
③ Threshold Phenomenon:



$n \uparrow$, $\|\mathcal{Z}_n - \mathcal{Z}\|$ will have a abrupt decrease from nearly 1 to nearly 0 at $n = [52 \log 52]$.

Rmk: Other example:

$X_n \overset{i.i.d}{\sim} N(1,1)$, $S_d = \sum_1^d X_n$. Then near $n = d$:



$p(S_d > n)$ will decrease at order $O(\sqrt{d})$, if d is large enough.

Denote: $A(n) = \|\mathcal{Z}_n - \mathcal{Z}\|$.

Def: Random time T is strong stationary time

if: i) T is stopping time

ii) X_T indept with T.

iii) $X_T \sim \mathcal{Z}$, stationary dist.

Lemma. If T is strong stationary time for (X_n).

Then: $\|\pi_n - \pi\| \leq p(T \geq n)$. $\forall n$.

Pf. 1) Show: $p(T \leq n, X_n \in A) = p(T \leq n)\pi(A)$.

$$LHS = \sum_{k=1}^{n} p(T=k, X_n \in A)$$

$$= \sum_k \sum_i p(T=k, X_k = i, X_n \in A)$$

$$= \sum_k \sum_i p^{n-k}(i, A)\, p(T=k, X_T = i)$$

$$= p(T \leq n)\, \pi(A).$$

2) $\forall A \subseteq S$. $\pi(A) - \pi_n(A) =$

$$\pi(A) - p(X_n \in A, T \leq n) - p(X_n \in A, T > n)$$

$$= \pi(A)\, p(T > n) - p(X_n \in A, T > n)$$

$$\Rightarrow |\pi(A) - \pi_n(A)| \leq \max\{\square, \square\} \leq p(T \geq n)$$

Thm. $\Delta(d\log d + cd) \leq p(T > d\log d + cd) \leq e^{-c}$. $\forall c \geq 0$. $d = 52$.

Pf: $T \sim \mathrm{Geo}(1/d) \oplus \cdots \oplus \mathrm{Geo}(d/d)$

(Famous dist. in Coupons collector Problem)

$$\{T > n\} = \bigcup_1^d \{k^{th} \text{ coupon isn't collected at time } n\}$$

$$\triangleq \bigcup_1^d A_k.$$

$$\Rightarrow p(T > n) \leq \sum p(A_k) = \sum_1^d \left(\frac{d-1}{d}\right)^n \leq d e^{-n/d}$$

Rmk: Fix c. if d is large enough, then cd is small relative to $d\log d$

So it will have an abrupt decrease near $d\log d$ when c changes a lot.

Thm. $k(d) = d\log d - c_d d$. $(c_d) \uparrow \infty$ as $d \to \infty$.

Then: $\| \pi^{(d)}_{k(d)} - \pi^{(d)} \| \xrightarrow{d\to\infty} 1$. index "$(d)$" means

dist. π. is relative to desk size d.

Pf: We want to find event $A \subseteq S$. st.

$\pi^{(d)}_{k(d)}(A)$ is large (≈ 1). $\pi^{(d)}(A)$ is small.

Consider: $A_{d,a} = \{$ Card $d-a+1 \sim d$ are

still in their origin order $\}$

$\pi(A_{d,a}) = 1/a!$ (permutate first a cards)

Note: $A_{d,a} \supseteq \{$ Card $d-a+1$ isn't at top yet $\}$

$\Rightarrow \pi_k(A_{d,a}) \geq p(U > k)$.

U is number of shufflings required for

card $d-a+1$ to rise at top.

$U \sim \mathrm{Geo}(a/d) \oplus \cdots \oplus \mathrm{Geo}(d-1/d)$

$$\begin{cases} E(U) = d(\log d - \log a + o(1)) \\ \mathrm{Var}(U) \leq d^2\left(\frac{1}{a^2} + \cdots\right) = O\left(\frac{1}{a}\right)d^2 \quad \left(\mathrm{Var}(\mathrm{Geo}(p)) = \frac{1-p}{p^2}\right) \end{cases}$$

So: $p(U > k(d)) = p(U - E(U) > -d(c_d - \log a_d + o(1)))$

$\geq 1 - \mathrm{Var}(U)/d^2(\square)^2$

Choose $a_d = e^{c_d/2}$. ($a = a_d$ depend on d)

Rmk: By the two Thms. above. We know:

$(\; d\log d \;)$ a subtle interval of $d\log d$

contains a abrupt decrease!