STA 447/2006 Lecture 5

Additional MC topic: Markov chain Monte Carlo (MCMC).

Goal: sample from target distribution π on S

(assume discrete for simplicity)

- S extremely large / infinite.
- We know π up to normalization factors.

($\pi(x) \propto f(x)$ for some known f).

Algorithmically: Construct an MC $(X_t)_{t \geq 0}$ with some init distribution, irreducible, aperiodic

stationary distribution π

Then $\mathbb{P}(X_t = j) \to \pi_j$

(and also SLLN for MCs).

Metropolis – Hastings algorithm.

- Start with a proposal distribution

$q(x, \cdot)$ for each $x \in S$.

- Define a new transition kernel.

$$p(x,y) = \min\left\{ q(x,y),\ q(y,x)\frac{\pi(y)}{\pi(x)} \right\}.$$

for every $x, y \in S$.

Operationally, at $X_t = x$, we sample $Y_t \sim q(X_t, \cdot)$

$$X_{t+1} = \begin{cases} Y_t & \text{w.p. } \min\left\{1, \dfrac{q(Y_t, X_t)\cdot \pi(Y_t)}{q(X_t, Y_t)\cdot \pi(X_t)}\right\} \\ X_t & \text{otherwise.} \end{cases}$$

- Key observation:

$$p(x,y)\cdot \pi(x) = \min\left\{ q(x,y)\cdot \pi(x),\ q(y,x)\cdot \pi(y) \right\}$$
$$= \pi(y)\cdot p(y,x)$$

which verifies that p is reversible w.r.t. π.

Martingales.

Idea: "fair gambling".

Def. A martingale is a real-valued stochastic process satisfying.

(i) $\mathbb{E}[|X_n|] < +\infty \quad (\forall n)$.

(ii) $\mathbb{E}[X_{n+1} \mid X_1, X_2, \cdots, X_n] = X_n. \quad (\forall n)$.

($(X_n)_{n\geq 1}$ may not be Markov.

Cond. distribution of X_{n+1} can depend on $X_1 \cdots X_n$

But cond. mean needs to be exactly X_n).

---

eg. SRW: $X_n = \sum_{i=1}^{n} \varepsilon_i$ where $\varepsilon_i \overset{iid}{\sim} \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2}. \end{cases}$

$$\mathbb{E}[|X_n|] \leq \sum_{i=1}^{n} \mathbb{E}[|\varepsilon_i|] \leq n.$$

$$\mathbb{E}[X_{n+1} \mid X_1, \cdots X_n] = X_n.$$

(more generally, partial sum process of iid. r.v.s

$X_n = \sum_{i=1}^{n} Y_i$ where $\mathbb{E}[|Y_i|] < +\infty$, $\mathbb{E}[Y_i] = 0$)

e.g. Let $(X_n)_{n\ge 0}$ be 1D SRW

Let $M_n = X_n^2 - n$.

(Idea: want to study $(X_n^2)_{n\ge0}$, but it's not MG.

a necessary condition to make it MG

$$\mathbb{E}[M_n] = \mathbb{E}[X_n^2] - n = 0$$

"$-n$" correction term to make it MG).

- $\mathbb{E}[|M_n|] \le \mathbb{E}[X_n^2] + n \le 2n < +\infty$.

- $\mathbb{E}[M_{n+1} \mid X_1, X_2, \dots, X_n]$

$$= \mathbb{E}[(X_n + \varepsilon_{n+1})^2 - (n+1) \mid X_1, \dots X_n].$$

$$= (X_n^2 + 1) - (n+1) + 2X_n \cdot \underbrace{\mathbb{E}[\varepsilon_{n+1} \mid X_1 \dots X_n]}_{=0}$$

$$= X_n^2 - n = M_n.$$

Notation: Given a stochastic process $(X_n)_{n\ge0}$.

$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n) =$ "information contained in $X_0, X_1, \dots X_n$".

We use $\mathbb{E}[X_{n+1} | \mathcal{F}_n]$ as a shorthand notation for $\mathbb{E}[X_{n+1} | X_0, X_1, X_2 \cdots X_n]$.

Fact For $0 \le m < n$

$$\mathbb{E}[X_n | \mathcal{F}_m] \left(= \mathbb{E}[X_n | X_0, X_1 \cdots, X_m] \right) = X_m.$$

for $(X_n)_{n \ge 0}$ MG.

Proof: $\mathbb{E}[X_{n+2} | X_0, X_1 \cdots X_n]$

$$= \mathbb{E}\Big[\underbrace{\mathbb{E}[X_{n+2} | X_0, X_1, \cdots X_{n+1}]}_{= X_{n+1}} \Big| X_0 X_1 \cdots X_n\Big] = X_n$$

(then apply induction)

(Taking $m = 0$. $\mathbb{E}[X_n] = \mathbb{E}[X_0]$).

Major theme (for following few lectures)

$$\mathbb{E}[X_T] \overset{?}{=} \mathbb{E}[X_0]$$

for some r.v. T that may depend on the process $(X_n)_{n \ge 0}$.

This cannot be true in general

e.g. $(X_n)_{n\geq 0}$ SRW.

$$T = \arg\max \{ X_t : 0 \leq t \leq 100 \}.$$

Some algebra shows that $\mathbb{E}[X_T] \geq c > 0$ (for $c > 0$)

Problem: at round 65, we don't know if $T = 65$.

We only observe the value of T after 100 rounds.

Stopping time:

(Idea: You know it when the time is reached.

Def. A non-negative-integer r.v. T is a stopping time if the event $\{T = n\}$ is determined by $X_0, X_1, \dots X_n$ for any $n = 0, 1, 2, \dots$

(measurable in $\mathcal{F}_n$)

Remark: def also extends to cts-time processes.
We require $\{T \leq t\}$ is determined by $(X_s)_{0\leq s\leq t}$.
(for any $t \geq 0$).

Examples of stopping times.

- Any deterministic time.
- Hitting times.

  eg. $T = \inf\{t \geq 0 : X_t = x\}$

  eg. $T = \inf\{t \geq 0 : X_t \in A\}$

  eg. $T = T_x^{(k)}$ (k-th visit to x).
- $T = \inf\{t \geq 2 : X_{t-2} = 5\}$ $(= T_5 + 2)$.
- If T_1 and T_2 are both stopping times.

— $\min(T_1, T_2)$ ✓

— $\max(T_1, T_2)$ ✓

— $T_1 + T_2$ ✓ Yes for discrete time

— $T_1 \times T_2$ ✗. No for cts time.

— $T_1 - T_2$ (assuming $T_1 \geq T_2$ w.p. 1). ✗.

Recall MC. "Strong Markov property".

$(X_n)_{n\geq 0}$ DTMC, T is a stopping time.

Conditionally on $T = t$, $\underset{x_0}{\overset{\shortparallel}{X_0}}, \underset{x_1}{\overset{\shortparallel}{X_1}}, X_2, \dots \underset{x_t}{\overset{\shortparallel}{X_t}}$.

Then $(X_n)_{n\geq T}$ conditionally is a fresh new MC starting from x_t.

Back to MGs.

Goal: $\mathbb{E}[X_T] \overset{?}{=} \mathbb{E}[X_0]$.

Another non-example:

$(X_n)_{n\geq 0}$ SRW.

$$T = \inf\{t \geq 0 : X_t = 5\}$$

Since SRW is recurrent, $\mathbb{P}(T < +\infty) = 1$.

$$5 = \mathbb{E}[X_T] \neq \mathbb{E}[X_0] = 0.$$

Problematic from gambling perspective:

- SRW is null recurrent, $\mathbb{E}[T] = +\infty$.
- Process can grow unboundedly before T.
  (but we only have finite \$\$\$)

"Easy case":

Suppose the stopping time T satisfies

$$\mathbb{P}(T \leq m) = 1 \quad \text{for some } m < +\infty$$

Lemma. If $(X_n)_{n\geq 0}$ is MG, T as above

then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

(Optional stopping lemma - bounded case)

Proof of the lemma:

$$\mathbb{E}[X_T] - \mathbb{E}[X_0] = \mathbb{E}[X_T - X_0]$$

$$= \mathbb{E}\left[\sum_{k=1}^{T} (X_k - X_{k-1})\right]$$

$$= \mathbb{E}\left[\sum_{k=1}^{m} (X_k - X_{k-1}) \cdot 1_{k \le T}\right]$$

$$= \sum_{k=1}^{m} \mathbb{E}\left[(X_k - X_{k-1}) 1_{k \le T}\right].$$

k-th term $= \mathbb{E}\left[(X_k - X_{k-1}) 1_{k \le T}\right]$

$$= \mathbb{E}\left[\mathbb{E}\left[(X_k - X_{k-1}) 1_{k \le T} \mid \mathcal{F}_{k-1}\right]\right]$$

"seems to involve information at time k".

Indeed, it does not.

$$1_{k \le T} = 1 - \sum_{j=0}^{k-1} 1_{\{T=j\}}$$

By stopping time, $\{T=j\}$ determined by $(X_t)_{0 \le t \le j}$

So $1_{\{k \le T\}}$ is determined by $(X_t)_{0 \le t \le k-1}$.

So we get

$$k\text{-th term} = \mathbb{E}\Big[1_{k \le T} \cdot \underbrace{\mathbb{E}[X_k - X_{k-1} \mid \mathcal{F}_{k-1}]}_{=0}\Big] = 0.$$

Relaxing the boundedness condition on T:

Thm (Optional stopping). T is stopping time $\mathbb{P}(T < +\infty) = 1$.

$(X_n)_{n \ge 0}$ is MG, T is stopping

(i) $\mathbb{E}[|X_T|] < +\infty$

(ii) $\lim_{n \to +\infty} \mathbb{E}[|X_n| 1_{T > n}] = 0$.

(truncation err $\to 0$).

then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Useful Corollary. Assuming $\mathbb{P}(T < +\infty) = 1$.

Suppose that $\exists\, B > 0$ s.t.

$$\mathbb{P}(|X_n| \le B, \ \forall n = 0, 1, \cdots, T) = 1$$

then OST holds true.

Proof: verifying conditions.

(i) $|X_T| \le B$ (a.s.) so $\mathbb{E}[|X_T|] < +\infty$

(ii). $\mathbb{E}[|X_n| \cdot 1_{T>n}] \le B \cdot \mathbb{E}[1_{T>n}]$

$= B \cdot \mathbb{P}(T > n)$

Since we assume $\mathbb{P}(T < +\infty) = 1$,

we have $\lim_{n \to +\infty} \mathbb{P}(T > n) = 0$.

Proof: "truncation arguments."

$T_m = \min\{T, m\}$ for any $m > 0$.

T_m is also a stopping time. $0 \le T_m \le m$.

By lemma, $\mathbb{E}[X_{T_m}] = \mathbb{E}[X_0]$ $(\forall m)$.

$$X_{T_m} = X_T 1_{\{T \le m\}} + X_m 1_{\{T > m\}}$$

$$X_T = X_T 1_{\{T \le m\}} + X_T 1_{\{T > m\}}.$$

error terms

$$\left|\mathbb{E}[X_{T_m}] - \mathbb{E}[X_T]\right|$$

$$\leq \underbrace{\mathbb{E}\left[|X_m| \mathbb{1}_{T>m}\right]}_{\to 0 \text{ following condition (ii)}} + \underbrace{\mathbb{E}\left[|X_T| \cdot \mathbb{1}_{T>m}\right]}$$

$|X_T| \cdot \mathbb{1}_{T>m} \leq |X_T|$.

$\mathbb{E}[|X_T|] < +\infty$

and $\mathbb{P}(|X_T| \mathbb{1}_{T>m} \to 0) = 1$

By DCT,

$\mathbb{E}[|X_T| \cdot \mathbb{1}_{T>m}] \to 0$.

So

$$\left|\mathbb{E}[X_{T_m}] - \mathbb{E}[X_T]\right| \to 0$$

$$\shortparallel$$

$$\left|\mathbb{E}[X_0] - \mathbb{E}[X_T]\right|$$

This concludes the proof of OST

---

e.g. Gambler's ruin.

- symmetric case $\qquad X_{n+1} = X_n + \varepsilon_{n+1}$

where $\varepsilon_{n+1} \overset{iid}{\sim} \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2. \end{cases}$

$X_0 = a \in [0, c]$. $\qquad (X_t)_{t\geq 0}$ is MG.

$T =$ hitting time of $\{0, c\}$

$|X_n|$ is bounded by c up to time T.

So we can use OST

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] = a.$$

$$\parallel$$

$$c \cdot \mathbb{P}(X_T = c) + 0 \cdot \mathbb{P}(X_T = 0).$$

So $\mathbb{P}(X_T = c) = \frac{a}{c}$.

· Asymmetric case. $\quad \varepsilon_n \overset{iid}{\sim} \begin{cases} 1 & \text{w.p. } p \\ -1 & \text{w.p. } (1-p). \end{cases}$

$(p \neq \frac{1}{2})$.

Construct a martingale (easily verified)

$$Y_n = \left(\frac{1-p}{p}\right)^{X_n} \quad (\text{for } n \geq 0).$$

$|Y_n| \leq \max\left\{1, \left(\frac{1-p}{p}\right)^c\right\}$ up to time T.

By OST.

$$\mathbb{E}[Y_T] = \mathbb{E}[Y_0] = \left(\frac{1-p}{p}\right)^a$$

$$\shortparallel$$

$$\left(\frac{1-p}{p}\right)^c \cdot \mathbb{P}(X_T = c) + 1 \cdot \mathbb{P}(X_T = 0).$$

Solve for $\mathbb{P}(X_T = c)$.