Probability and Stochastic Processes I I - Lecture 6e

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC632024.html

2024

VI.6 Continuous State Space Markov Chains

- here $T = \mathbb{N}_0$ and $\mathcal{S}$ with σ-algebra $\mathcal{C}$ (could be $\mathcal{S} = \mathbb{R}^k, \mathcal{C} = \mathcal{B}^k$)

Definition VI.11 A *Markov kernel* $P : \mathcal{S} \times \mathcal{C} \to [0, 1]$ for a time homogeneous, Markov process with state space $\mathcal{S}$ and initial probability measure ν on $(\mathcal{S}, \mathcal{C})$ satisfies (i) $P(s, \cdot)$ is a probability measure on $\mathcal{C}$ for every $s \in \mathcal{S}$ and (ii) $P(\cdot, C) : (\mathcal{S}, \mathcal{C}) \to (\mathbb{R}^1, \mathcal{B}^1)$ for every $C \in \mathcal{C}$.

- so $P(\cdot, C)$ is a random variable and $P(s, C)$ is interpreted as the conditional probability that the next state is in C given that the current state is s

- note when $\mathcal{S} \subset \mathbb{Z}$, then $P(\cdot, C)$ is given by $P(i, \{j\}) = p_{ij}$

- so the time homogeneous Markov process with state space $\mathcal{S}$ and Markov kernel P is given by $\{X_n : n \in \mathbb{N}_0\}$ where $X_0 \sim \nu$ and $X_n \mid X_0, \ldots, X_{n-1} \sim P(X_{n-1}, \cdot)$

- such Markov processes arise in many statistical problems where there is a (posterior) distribution on a continuous space we want to sample from but can only find a Markov chain Monte Carlo algorithm to do this approximately

Example VI.11

- $\{X_n : n \in \mathbb{N}_0\}$ is given by $X_0 = 0$ (ν is degenerate a 0) and $P(x, \cdot) = N(x/2, 1)$ so $X_1 \sim N(0, 1)$ and

$$\begin{aligned} P(X_2 \leq x_2) &= \int_{-\infty}^{\infty} P(X_2 \leq x_2 \,|\, X_1)(x_1)\, P_{X_1}(dx_1) \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{x_2 - x_1/2}{1}\right) \varphi(x_1)\, dx_1 \end{aligned}$$

but with $Z \sim N(0, 1)$ independent of X_1, then $X_2 = Z + X_1/2 \sim N(0, 1 + 1/4) = N(0, 5/4), X_3 = Z + X_2/2 \sim N(0, 1 + 5/16) = N(0, 21/16)$, etc. ■

- note - as long as it is easy to generate from ν and $P(s, \cdot)$ for each s, then it is easy to simulate the process

- we need a generalization of the concept of irreducibility because in the continuous case $P(s, \{x\}) = 0$ for all s, x

- note

$$\begin{aligned} P^{(2)}(s, C) &= P(X_2 \in C \mid X_0 = s) \\ &= \int_{\mathcal{S}} P(X_2 \in C \mid X_0 = s, X_1 = s')P(s, ds') \\ &= \int_{\mathcal{S}} P(X_2 \in C \mid X_1 = s')P(s, ds') \text{ by MP} \\ &= \int_{\mathcal{S}} P(s', C)P(s, ds') \text{ by TH} \end{aligned}$$

and $P^{(n)}$ is obtained similarly

Definition VI.12 A MC with state space $\mathcal{S}$ is called *ϕ-irreducible*, where ϕ is a measure on $\mathcal{C}$, if for all $s \in \mathcal{S}$, whenever $\phi(C) > 0$ then $P^{(n)}(s, C) > 0$ for some n.

- typically $\phi =$ volume measure and in the discrete case if $\phi =$ counting measure this is the same definition in that case

Proposition VI.35 If there is $\delta > 0$ s.t. $P(s, \cdot)$ has a positive density on $[s - \delta, s + \delta]$ for every $s \in \mathcal{S} = \mathbb{R}^1$, then the MC with this kernel is ϕ-irreducible where ϕ is volume measure.

Proof: Suppose $C \in \mathcal{B}^1$ and $\phi(C) > 0$. Then $\exists m \in \mathbb{N}$ s.t. $\phi(C \cap [-m, m]) > 0$. Choose n s.t. $s - n\delta < -m < m < s + n\delta$. Then since $P(s, \cdot)$ has positive density on $[s - \delta, s + \delta]$ this implies $P^{(n)}(s, \cdot)$ has positive density on $[s - n\delta, s + n\delta] > 0$ (see * below). But $C \cap [-m, m] \subset [s - n\delta, s + n\delta]$ so

$$P^{(n)}(s, C) \geq P^{(n)}(s, C \cap [-m, m]) > 0.$$

■

* if $P(s,\cdot)$ has positive density on $[s-\delta, s+\delta]$, then so does $P^{(2)}(s,\cdot)$ on $[s-2\delta, s+2\delta]$ since

$$
\begin{aligned}
& P^{(2)}(s,[s-2\delta, s+2\delta]) \\
= \; & \int_{-\infty}^{\infty} P(x,[s-2\delta, s+2\delta])\, P(s,dx) \\
= \; & \int_{-\infty}^{\infty} \left(\int_{s-2\delta}^{s+2\delta} f(z\,|\,x)dz \right) P(s,dx) \\
\geq \; & \int_{s-\delta}^{s+\delta} \left(\int_{s-2\delta}^{s+2\delta} f(z\,|\,x)dz \right) f(x\,|\,s)\,dx > 0
\end{aligned}
$$

since $f(z\,|\,x) > 0$ for all $z \in [x-\delta, x+\delta]$ and this implies $\int_{s-2\delta}^{s+2\delta} f(z\,|\,x)dz > 0$ whenever $[x-\delta, x+\delta] \subset [s-2\delta, s+2\delta]$ and this holds whenever $x \in [s-\delta, s+\delta]$ and note $f(x\,|\,s) > 0$ there

Definition VI.13 A MC with state space $\mathcal{S}$ has *period* b if $\mathcal{S}$ can be decomposed into mutually disjoint subsets $\mathcal{S} = \mathcal{S}_0 \cup \cdots \cup \mathcal{S}_{b-1}$ s.t. $P(s, S_i) = 1$ whenever $s \in S_{i-1}$ and $i = [(i-1)+1] \bmod b$. If $b = 1$, then the MC is said to be *aperiodic*.

Proposition VI.36 If there is $\delta > 0$ s.t. $P(s, \cdot)$ has a positive density on $[s-\delta, s+\delta]$ for every $s \in \mathcal{S} = \mathbb{R}^1$, then a ϕ-irreducible MC, where ϕ is volume measure, is aperiodic.

Proof: Suppose b is the period and let $s \in \mathcal{S}_i$ where $\text{vol}(\mathcal{S}_i) > 0$. Also, let $m \in \mathbb{N}$ be s.t. $\text{vol}(\mathcal{S}_i \cap [-m, m]) > 0$. Choose n_0 s.t. $s - n_0\delta < -m < m < s + n_0\delta$. Then $P(s, \cdot)$ has positive density on $[s-\delta, s+\delta]$ which implies $P^{(n)}(s, \cdot)$ has positive density on $[s - n\delta, s + n\delta] > 0$ for every $n \geq n_0$ and in particular for $n = n_0$ and $n = n_0 + 1$. But if $b \geq 2$ it is not possible that $b|n_0$ and $b|n_0 + 1$ and so $b = 1$. ■

Definition VI.14 A MC with state space $\mathcal{S}$ has *stationary distribution* π if $\Pi(C) = \int_{\mathcal{S}} P(s, C)\, \Pi(ds)$ for all $C \in \mathcal{C}$.

- as in the discrete case this implies that when $\nu = \pi$ then $X_n \sim \pi$ for all n

- also if $P(s, C)$ has density, $f(x \,|\, s)$ then π is stationary when

$$\pi(x) = \int_{\mathcal{S}} f(x \,|\, s)\pi(s)\, ds$$

- if the Markov kernel $P(s, \cdot)$ has density $f(x \,|\, s)$ then it is *reversible wrt* π whenever $f(x \,|\, s)\pi(s) = f(s \,|\, x)\pi(x)$ and then

$$\int_{\mathcal{S}} f(x \,|\, s)\pi(s)\, ds = \int_{\mathcal{S}} f(s \,|\, x)\pi(x)\, ds = \pi(x)$$

and so π is stationary

Proposition VI.37 If a MC is ϕ-irreducible, aperiodic and has stationary distribution Π, then

$$\lim_{n\to\infty} P(X_n \in C \,|\, X_0 = s) = \Pi(C)$$

for all $s \in \mathcal{S}$ excepting possibly a set having Π measure 0.

Proof: fact

Example VI.11 *(continued)*

- recall $f(x \,|\, s)$ is the $N(s/2, 1)$ density so $f(s \,|\, x)$ is the $N(x/2, 1)$ density
- so for reversibility we want

$$\begin{aligned} \exp(-(x-s/2)^2/2)\pi(s) &= \exp(-(x/2-s)^2/2)\pi(x) \text{ or} \\ \frac{\pi(s)}{\pi(x)} &= \exp\left(-\frac{(x/2-s)^2}{2} + \frac{(x-s/2)^2}{2}\right) \\ &= \exp\left(\frac{3}{8}(x^2-s^2)\right) \end{aligned}$$

which is satisfied by $\pi = N(0, 4/3)$

- also the chain satisfies Propositions VI.35 and VI.36 (normal densities are positive everywhere) so Prop. VI.37 applies ■

Example VI.12 *Markov Chain Monte Carlo*

- suppose we want to sample from a distribution π and we can't find a good (e.g., rejection or inversion) algorithm to do this directly

- we try to create a MC, that can simulated from, s.t. π is its unique stationary distribution so that X_n is approximately distributed π for large n

- let $q(s,\cdot)$ be a proposal density for each $s \in \mathcal{S}$ and suppose q is symmetric, i.e., $q(s,x) = q(x,s)$

- define $P(s,\cdot)$ by

$$\begin{aligned} P(s, C\backslash\{s\}) &= \int_{C\backslash\{s\}} q(s,x) \min\left(1, \frac{\pi(x)}{\pi(s)}\right) dx \\ P(s,\{s\}) &= 1 - \int_{\mathcal{S}} q(s,x) \min\left(1, \frac{\pi(x)}{\pi(s)}\right) dx \end{aligned}$$

and note

$$\int_{\mathcal{S}} q(s,x) \min\left(1, \frac{\pi(x)}{\pi(s)}\right) dx \le \int_{\mathcal{S}} q(s,x)\, dx = 1$$

so $P(s,.)$ is a probability measure

- the chain may stay at state s with possibly positive probability

- this is known as the *Metropolis algorithm*

1. given $X_{n-1} = s$ generate $Y_n \sim q(s, \cdot)$ stat. ind. of $U_n \sim U(0,1)$

2. put

$$X_n = \begin{cases} Y_n & \text{if } U_n \le \pi(Y_n)/\pi(X_{n-1}) \\ X_{n-1} & \text{otherwise} \end{cases}$$

- to see that this works

$$\begin{aligned} & P(X_n \in C\backslash\{s\} \,|\, X_{n-1} = s) \\ = \ & P(Y_n \in C\backslash\{s\}, U_n \le \pi(Y_n)/\pi(s) \,|\, X_{n-1} = s) \\ = \ & \int_{C\backslash\{s\}} \int_0^{\min(1,\pi(x)/\pi(s))} q(s,x)\, du dx \\ = \ & \int_{C\backslash\{s\}} q(s,x) \min(1, \pi(x)/\pi(s))\, du dx \end{aligned}$$

Proposition VI.38 *(Continuous MCMC Convergence)* If $\pi(s) > 0$ for every $s \in \mathcal{S}$ and there is $\delta > 0$ s.t. $q(s,x) > 0$ whenever $|x - s| < \delta$, then $X_n \xrightarrow{d} \pi$.

Proof: We have, for $s \neq x$

$$\begin{aligned} & \pi(s)q(s,x)\min\left(1, \pi(x)/\pi(s)\right) \\ = & \left\{ \begin{array}{ll} \pi(s)q(s,x) & \text{if } \pi(x)/\pi(s) > 1 \\ \pi(x)q(s,x) & \text{if } \pi(x)/\pi(s) \leq 1 \end{array} \right. \\ = & \ \pi(x)q(x,s)\min\left(1, \pi(s)/\pi(x)\right) \end{aligned}$$

and so the chain is reversible and π is a stationary distribution for the chain. Also, $\pi(s)q(s,x)\min\left(1, \pi(x)/\pi(s)\right) > 0$ whenever $|x - s| < \delta$ and so $P(s, \cdot)$ has positive density when$|x - s| < \delta$ and so the chain is ϕ-irreducible ($\phi =$ volume) and aperiodic. The result then follows from Prop. VI.37. ■

Exercise VI.18 Text 4.7.11

Exercise VI.19 Text 4.8.3