Simulation - Lecture 6 - Markov Chain Monte Carlo

Lecture version: Monday 24th February, 2020, 17:54

Robert Davies

Part A Simulation and Statistical Programming

Hilary Term 2020

Recap and overview of MCMC

- Our aim is to estimate $\theta = \mathbb{E}_p(\phi(X))$ where X is a random variable on Ω with pmf (or pdf) p.
- Up to this point we have based our estimates on independent and identically distributed draws from either p itself, or some proposal distribution with pmf/pdf q.
- In MCMC we simulate a correlated sequence $X_0, X_1, X_2, ...$ such that X_t is approximately distributed from p for t large, and rely on the usual estimate

$$\hat{\theta}_n = \frac{1}{n} \sum_{t=0}^{n-1} \phi(X_t).$$

- We will suppose the space of states of X is discrete (finite or countable).
- But it should be kept in mind that MCMC methods are applicable to continuous state spaces, and in fact one of the most versatile and widespread classes of Monte Carlo algorithms currently.

Outline

Recap on Markov chains

Markov Chain Monte Carlo

Markov chains

- From Part A Probability.
- Let $(X_t)_{t=0,1,\dots}$ be a homogeneous Markov chain of random variables on Ω with starting distribution $X_0 \sim p^{(0)}$ and transition matrix $P = (P_{ij})_{i,j\in\Omega}$ with

$$P_{i,j} = \mathbb{P}(X_{t+1} = j | X_t = i).$$

  for $i, j \in \Omega$.
- We write $(X_t)_{t=0,1,\dots} \sim \mathrm{Markov}(p^{(0)}, P)$
- Denote by $P^{(n)}_{i,j}$ the n-step transition probabilities

$$P^{(n)}_{i,j} = \mathbb{P}(X_{t+n} = j | X_t = i)$$

  and by $p^{(n)}(i) = \mathbb{P}(X_n = i)$.
- Recall that a Markov chain, or equivalently its transition matrix P, is irreducible if and only if, for each pair of states $i, j \in \Omega$ there is n such that $P^{(n)}_{i,j} > 0$.

Markov chains

- π is a stationary, or invariant distribution of P, if π verifies

$$\pi_j = \sum_{i \in \Omega} \pi_i P_{ij}$$

  for all $j \in \Omega$.
- If $p^{(0)} = \pi$ then

$$p^{(1)}(j) = \sum_{i \in \Omega} p^{(0)}(i) P_{i,j},$$

  so $p^{(1)}(j) = \pi(j)$ also. Iterating, $p^{(t)} = \pi$ for each $t = 1, 2, \ldots$ in the chain, so the distribution of X_t doesn't change with t, it is stationary.
- If P is irreducible, and has a stationary distribution π, then this stationary distribution π is unique.
- The Markov chain is aperiodic if $P_{i,j}^{(n)} > 0$ for all sufficiently large n beyond a threshold n

Ergodic Theorem for Markov chains

Theorem (Theorems 6.1, 6.2, 6.3 of Part A Probability)

Let $(X_t)_{t=0,1,\dots} \sim \mathrm{Markov}(\lambda, P)$ *be an irreducible Markov chain on a discrete state space* Ω*. Assume it admits a stationary distribution* π*. Then, for any initial distribution* λ

$$\frac{1}{n}\sum_{t=0}^{n-1} \mathbb{I}(X_t = i) \to \pi(i) \quad \textit{almost surely, as } n \to \infty.$$

That is

$$\mathbb{P}\left(\frac{1}{n}\sum_{t=0}^{n-1} \mathbb{I}(X_t = i) \to \pi(i)\right) = 1.$$

Additionally, if the chain is aperiodic, then for all $i \in \Omega$

$$\mathbb{P}(X_n = i) \to \pi(i)$$

Ergodic Theorem for Markov chains

Corollary

Let $(X_t)_{t=0,1,\ldots} \sim \mathrm{Markov}(\lambda, P)$ be an irreducible Markov chain on a discrete state space Ω. Assume it admits a stationary distribution π. Let $\phi : \Omega \to \mathbb{R}$ be a bounded function, X a discrete random variable on Ω with pmf π and $\theta = \mathbb{E}_\pi[\phi(X)] = \sum_{i \in \Omega} \phi(i)\pi(i)$. Then, for any initial distribution λ

$$\frac{1}{n}\sum_{t=0}^{n-1} \phi(X_t) \to \theta \quad \text{almost surely, as } n \to \infty.$$

That is

$$\mathbb{P}\left(\frac{1}{n}\sum_{t=0}^{n-1} \phi(X_t) \to \theta\right) = 1.$$

Ergodic Theorem for Markov chains

Proof (non-examinable). Assume wlog $|\phi(i)| < 1$ for all $i \in \Omega$. Let $A \subset \Omega$ and $V_i(n) = \sum_{t=0}^{n-1} \mathbb{I}(X_t = i)$.

$$
\begin{aligned}
\left| \frac{1}{n} \sum_{t=0}^{n-1} \phi(X_t) - \theta \right| &= \left| \frac{1}{n} \sum_{t=0}^{n-1} \sum_{i \in \Omega} \phi(i) \mathbb{I}(X_t = i) - \sum_{i \in \Omega} \phi(i) \pi(i) \right| \\
&= \left| \sum_{i \in \Omega} \phi(i) \left(\frac{1}{n} V_i(n) - \pi(i) \right) \right| \\
&\leq \sum_{i \in A} \left| \frac{V_i(n)}{n} - \pi(i) \right| + \sum_{i \notin A} \left| \frac{V_i(n)}{n} - \pi(i) \right| \\
&\leq \sum_{i \in A} \left| \frac{V_i(n)}{n} - \pi(i) \right| + \sum_{i \notin A} \left(\frac{V_i(n)}{n} + \pi(i) \right) \\
&= \sum_{i \in A} \left| \frac{V_i(n)}{n} - \pi(i) \right| + \sum_{i \in A} \left(\pi(i) - \frac{V_i(n)}{n} \right) + 2 \sum_{i \notin A} \pi(i) \\
&\leq 2 \sum_{i \in A} \left| \frac{V_i(n)}{n} - \pi(i) \right| + 2 \sum_{i \notin A} \pi(i)
\end{aligned}
$$

where in line 5 we've used $\sum_{i \notin A} \frac{V_i(n)}{n} = 1 - \sum_{i \in A} \frac{V_i(n)}{n} = \sum_i \pi(i) - \sum_{i \in A} \frac{V_i(n)}{n}$.

Ergodic Theorem for Markov chains

Proof (continued). Let $\epsilon > 0$, and take A finite such that $\sum_{i \notin A} \pi(i) < \epsilon/4$. For $N \in \mathbb{N}$, Define the event

$$E_N = \left\{ \sum_{i \in A} \left| \left(\frac{V_i(n)}{n} - \pi(i) \right) \right| < \epsilon/4 \text{ for all } n \geq N \right\}.$$

As $\mathbb{P}(\frac{V_i(n)}{n} \to \pi(i)) = 1$ for all $i \in \Omega$ and A is finite, the event E_N must occur for some N hence $\mathbb{P}(\cup E_N) = 1$. It follows that, for any $\epsilon > 0$

$$\mathbb{P}\left(\exists N \text{ such that for all } n \geq N, \quad \left| \frac{1}{n} \sum_{t=0}^{n-1} \phi(X_t) - \theta \right| < \epsilon \right) = 1.$$

Reversible Markov chains

- In a reversible Markov chain we cannot distinguish the direction of simulation from inspection of a realization of the chain and its reversal, even with knowledge of the transition matrix.
- A Markov chain $(X_t)_{t\geq 0} \sim \mathrm{Markov}(\pi, P)$ is reversible iff

$$\mathbb{P}(X_0, X_1, \ldots, X_n) = \mathbb{P}(X_n, X_{n-1}, \ldots, X_0)$$

  for any $n \geq 0$.
- We also say that P is reversible with respect to π, or π-reversible
- Most MCMC algorithms are based on reversible Markov chains.

Reversible Markov chains

- It seems clear that a Markov chain will be reversible if and only if $\mathbb{P}(X_{t-1} = j | X_t = i) = P_{i,j}$, so that any particular transition occurs with equal probability in forward and reverse directions.

Theorem

Let P be a transition matrix. If there is a probability mass function π on Ω such that π and P satisfy the detailed balance condition

$$\pi(i)P_{i,j} = \pi(j)P_{j,i} \qquad \text{for all pairs } i, j \in \Omega,$$

then

(I) $\pi = \pi P$, so π is stationary for P and

(II) the chain $(X_t) \sim \mathrm{Markov}(\pi, P)$ is reversible.

Reversible Markov chains

- Proof of (I): sum both sides of detailed balance equation over $i \in \Omega$. Now $\sum_i P_{j,i} = 1$ so $\sum_i \pi(i) P_{i,j} = \pi(j)$.
- Proof of (II), we have π a stationary distribution of P so $\mathbb{P}(X_t = i) = \pi(i)$ for all $t = 1, 2, \ldots$ along the chain. Then

$$
\begin{aligned}
\mathbb{P}(X_{t-1} = j | X_t = i) &= \mathbb{P}(X_t = i | X_{t-1} = j) \frac{\mathbb{P}(X_{t-1} = j)}{\mathbb{P}(X_t = i)} \text{ (Bayes rule)} \\
&= P_{j,i} \pi(j) / \pi(i) \text{ (stationarity)} \\
&= P_{i,j} \text{ (detailed balance).}
\end{aligned}
$$

Outline

Recap on Markov chains

Markov Chain Monte Carlo

Markov chain Monte Carlo method (discrete case)

Let X be a discrete random variable with pmf p on Ω, ϕ a bounded function on Ω and $\theta = \mathbb{E}_p[\phi(X)]$. Consider a homogeneous Markov chain $(X_t)_{t=0,1,\dots} \sim \mathrm{Markov}(\lambda, P)$ with initial distribution λ on Ω and transition matrix P, such that P is irreducible, and admits p as invariant distribution. Then, for any initial distribution λ the MCMC estimator

$$\widehat{\theta}_n^{\texttt{MCMC}} = \frac{1}{n} \sum_{i=1}^{n-1} \phi(X_t)$$

is (weakly and strongly) consistent

$$\widehat{\theta}_n^{\texttt{MCMC}} \to \theta \text{ almost surely as } n \to \infty$$

and, if the chain is aperiodic

$$X_t \to p \text{ in distribution as } t \to \infty.$$

Markov chain Monte Carlo method

- Proof follows directly from the ergodic theorem and corollary
- Note that the estimator is biased, as $\lambda \neq p$ (otherwise we would use a standard Monte Carlo estimator)
- For t large, we have $X_t \overset{d}{\simeq} X$
- In order to implement the MCMC algorithm we need, for a given target distribution p, to find an irreducible (and aperiodic) transition matrix P with admits p as invariant distribution
- Most MCMC algorithms use a transition matrix P which is reversible with respect to p

Markov chain Monte Carlo method

- We can construct a Markov Chain Monte Carlo estimator for θ if we can find a transition matrix P that is irreducible and satisfies the detailed balance equations
- **Next time:** The Metropolis-Hastings algorithm provides a generic way to obtain such P for any target distribution p