Markov Chain Monte Carlo
Theory and practical applications

Randal Douc and Sylvain Le Corff

Télécom SudParis, Institut Polytechnique de Paris
randal.douc@telecom-sudparis.eu

Institut Mines-Télécom

Outline

Outline

Activities

Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$.

Definition (Invariant probability measure)

We say that $\pi \in \mathrm{M}_1(\mathsf{X})$ is an invariant probability measure for P if $\pi P = \pi$.

If $\pi P = \pi$, then $\pi P^n = \pi P^{n-1} = \ldots = \pi$.
Therefore, if $X_0 \sim \pi$ then $X_1 \sim \pi P = \pi$ and more generally, $X_n \sim \pi$.

Activities

Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$.

Definition (Invariant probability measure)

We say that $\pi \in \mathrm{M}_1(\mathsf{X})$ is an invariant probability measure for P if $\pi P = \pi$.

If $\pi P = \pi$, then $\pi P^n = \pi P^{n-1} = \ldots = \pi$.
Therefore, if $X_0 \sim \pi$ then $X_1 \sim \pi P = \pi$ and more generally, $X_n \sim \pi$.

Outline

Definition (Reversibility)

Let $\pi \in \mathrm{M}_1(\mathsf{X})$ and P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. We say that P is **π-reversible** if and only if (with infinitesimal notation)

$$\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x), \tag{1}$$

In other words, P is π-reversible iff for all measurable bounded or non-negative functions h on $\left(\mathsf{X}^2, \mathcal{X}^{\otimes 2}\right)$,

$$\iint_{\mathsf{X}^2} h(x, y)\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \iint_{\mathsf{X}^2} h(x, y)\pi(\mathrm{d}y)P(y, \mathrm{d}x). \tag{2}$$

Proposition

If the Markov kernel P is π-reversible, then it is π-invariant.

Definition (Reversibility)

Let $\pi \in \mathrm{M}_1(\mathsf{X})$ and P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. We say that P is π-reversible if and only if (with infinitesimal notation)

$$\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x), \quad (1)$$

In other words, P is π-reversible iff for all measurable bounded or non-negative functions h on $\left(\mathsf{X}^2, \mathcal{X}^{\otimes 2}\right)$,

$$\iint_{\mathsf{X}^2} h(x,y)\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \iint_{\mathsf{X}^2} h(x,y)\pi(\mathrm{d}y)P(y, \mathrm{d}x). \quad (2)$$

Proposition

If the Markov kernel P is π-reversible, then it is π-invariant.

Definition (Reversibility)

Let $\pi \in \mathrm{M}_1(\mathsf{X})$ and P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. We say that P is **π-reversible** if and only if (with infinitesimal notation)

$$\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x), \tag{1}$$

In other words, P is π-reversible iff for all measurable bounded or non-negative functions h on $\left(\mathsf{X}^2, \mathcal{X}^{\otimes 2}\right)$,

$$\iint_{\mathsf{X}^2} h(x,y)\pi(\mathrm{d}x)P(x,\mathrm{d}y) = \iint_{\mathsf{X}^2} h(x,y)\pi(\mathrm{d}y)P(y,\mathrm{d}x). \tag{2}$$

Proposition

If the Markov kernel P is π-reversible, then it is π-invariant.

Outline

Metropolis Hastings Algorithm

Input:n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n-1$
  1. Draw independently $Y_{t+1} \sim Q(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0,1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

Terminology:

- Q is the instrumental kernel or proposition kernel.
- The acceptance probability is usually chosen equal to $\alpha(x,y) = \alpha^{MH}(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ but other choices are possible.

Proposition

The Markov kernel associated to a MH algorithm is π-reversible.

Metropolis Hastings Algorithm

Input:n
Output: $X_0, \dots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n-1$
  1. Draw independently $Y_{t+1} \sim Q(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0,1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

Terminology:

- Q is the instrumental kernel or proposition kernel.
- The acceptance probability is usually chosen equal to $\alpha(x,y) = \alpha^{MH}(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ but other choices are possible.

Proposition

The Markov kernel associated to a MH algorithm is π-reversible.

Metropolis Hastings Algorithm

Input:n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n-1$
  1. Draw independently $Y_{t+1} \sim Q(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0,1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

Terminology:

- Q is the instrumental kernel or proposition kernel.
- The acceptance probability is usually chosen equal to $\alpha(x,y) = \alpha^{MH}(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ but other choices are possible.

Proposition

The Markov kernel associated to a MH algorithm is π-reversible.

Metropolis Hastings Algorithm

Input:n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n - 1$
  1. Draw independently $Y_{t+1} \sim Q(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0,1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

Terminology:

- Q is the instrumental kernel or proposition kernel.
- The acceptance probability is usually chosen equal to $\alpha(x,y) = \alpha^{MH}(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ but other choices are possible.

Proposition

The Markov kernel associated to a MH algorithm is π-reversible.

Metropolis Hastings Algorithm

Input:n

Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n-1$
  1. Draw independently $Y_{t+1} \sim Q(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0,1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

Terminology:

- Q is the instrumental kernel or proposition kernel.
- The acceptance probability is usually chosen equal to $\alpha(x,y) = \alpha^{MH}(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ but other choices are possible.

Proposition

The Markov kernel associated to a MH algorithm is π-reversible.

The independence sampler

1. If the proposition update is $Y_{t+1} \sim q(\cdot)$, then the proposed candidate is drawn irrespective of the current value of the Markov chain.
2. The proposition kernel is then $Q(x, \mathrm{d}y) = q(y)\lambda(\mathrm{d}y)$ where q is a density wrt λ on X, and in such case, the acceptance probability is $\alpha(x, y) = \min\left(\frac{\pi(y)q(x)}{\pi(x)q(y)}, 1\right)$
3. Such Metropolis-Hastings algorithm is called the Independence Sampler.

The independence sampler

1. If the proposition update is $Y_{t+1} \sim q(\cdot)$, then the proposed candidate is drawn irrespective of the current value of the Markov chain.
2. The proposition kernel is then $Q(x, \mathrm{d}y) = q(y)\lambda(\mathrm{d}y)$ where q is a density wrt λ on X, and in such case, the acceptance probability is $\alpha(x, y) = \min\left(\frac{\pi(y)q(x)}{\pi(x)q(y)}, 1\right)$
3. Such Metropolis-Hastings algorithm is called the Independence Sampler.

The independence sampler

1. If the proposition update is $Y_{t+1} \sim q(\cdot)$, then the proposed candidate is drawn irrespective of the current value of the Markov chain.
2. The proposition kernel is then $Q(x, \mathrm{d}y) = q(y)\lambda(\mathrm{d}y)$ where q is a density wrt λ on X, and in such case, the acceptance probability is $\alpha(x, y) = \min\left(\frac{\pi(y)q(x)}{\pi(x)q(y)}, 1\right)$
3. Such Metropolis-Hastings algorithm is called the Independence Sampler.

Random Walk MH algorithm

1. In this algorithm, the proposition update is $Y_{t+1} = X_k + \eta_k$ where $\eta_k \sim q(\cdot)$ where $q(u) = q(-u)$ for all $u \in \mathsf{X}$ and $\mathsf{X} = \mathbb{R}^p$.
2. In such case, the proposition kernel is $Q(x, \mathrm{d}y) = q(y - x)\lambda(\mathrm{d}y)$ and the acceptance probability is $\alpha(x, y) = \min\left(\frac{\pi(y)}{\pi(x)}, 1\right)$.
3. The associated algorithm is called the *(symmetric)* Random Walk Metropolis-Hasting.

Random Walk MH algorithm

1. In this algorithm, the proposition update is $Y_{t+1} = X_k + \eta_k$ where $\eta_k \sim q(\cdot)$ where $q(u) = q(-u)$ for all $u \in \mathsf{X}$ and $\mathsf{X} = \mathbb{R}^p$.
2. In such case, the proposition kernel is $Q(x, \mathrm{d}y) = q(y - x)\lambda(\mathrm{d}y)$ and the acceptance probability is $\alpha(x, y) = \min\left(\frac{\pi(y)}{\pi(x)}, 1\right)$.
3. The associated algorithm is called the *(symmetric)* Random Walk Metropolis-Hasting.

Random Walk MH algorithm

1. In this algorithm, the proposition update is $Y_{t+1} = X_k + \eta_k$ where $\eta_k \sim q(\cdot)$ where $q(u) = q(-u)$ for all $u \in \mathsf{X}$ and $\mathsf{X} = \mathbb{R}^p$.
2. In such case, the proposition kernel is $Q(x, \mathrm{d}y) = q(y - x)\lambda(\mathrm{d}y)$ and the acceptance probability is $\alpha(x, y) = \min\left(\frac{\pi(y)}{\pi(x)}, 1\right)$.
3. The associated algorithm is called the *(symmetric)* Random Walk Metropolis-Hasting.