Riemann manifold Hamiltonian Monte Carlo methods

Pierre Boyeau   Baptiste Kerléguer

École Normale Supérieure Paris-Saclay

10 janvier 2018

Plan

1. Context
   - Hamiltonian Monte Carlo
   - Variété Riemannienne
   - Riemann Manifold Hamiltonian Monte Carlo
2. Application: Régression Logistique Bayésienne

Plan

Hamiltonian Monte Carlo Classique

Idée: Introduction d'une variable auxiliaire $\boldsymbol{p}$:
Log-proba négative cible: Hamiltonien:

$$H(\boldsymbol{\theta}, \boldsymbol{p}) = -\mathcal{L}(\boldsymbol{\theta}) + \frac{1}{2}\log(2\pi)^2|\boldsymbol{M}| + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p}$$

avec M une matrice de masse

Équation d'évolution :

$$\frac{d\boldsymbol{\theta}}{d\tau} = \frac{\partial H}{\partial \boldsymbol{p}} \quad (1)$$

$$\frac{d\boldsymbol{p}}{d\tau} = -\frac{\partial H}{\partial \boldsymbol{\theta}} \quad (2)$$

Variété Riemannienne

Idée: Prendre en compte des arguments de géométrie Riemannienne pour *sampler* p: $M \to \boldsymbol{\theta}\ G(\boldsymbol{\theta})$.
Nouvelle équation d'évolution:

$$\frac{d\boldsymbol{\theta}}{d\tau} = G(\boldsymbol{\theta})^{-1}\boldsymbol{p} \quad (3)$$

$$\frac{d\boldsymbol{p}}{d\tau} = \nabla_{\theta}\mathcal{L}(\boldsymbol{\theta}) - \frac{1}{2}\mathrm{tr}\left(G(\boldsymbol{\theta})^{-1}\frac{\partial G(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) + \frac{1}{2}\boldsymbol{p}^{T}(\boldsymbol{\theta})^{-1}\frac{\partial G(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta})^{-1}\boldsymbol{p} \quad (4)$$

$\Rightarrow$ Calcul de $G(\boldsymbol{\theta})^{-1}$ nécessaire.

Choix de $G(\boldsymbol{\theta})$

Notre choix de $G(\theta)$ en considérant le problème actuel

$$G(\beta) = \boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X} + \alpha^{-1} \boldsymbol{I}$$

Pour obtenir une variété de courbure constante:

$$G = \boldsymbol{X}^T \boldsymbol{X} + \alpha^{-1} \boldsymbol{I}$$

RMHMC

On suppose qu'on a $\boldsymbol{\theta}_n$. Gibbs sampling:

1. Sampler $\boldsymbol{p}_{n+1}|\boldsymbol{\theta_n} \sim \mathcal{N}(0, G(\boldsymbol{\theta}_n))$
2. Le *proposal* $(\boldsymbol{\theta}_*, \boldsymbol{p}_*)$ est obtenu à l'aide de l'intégrateur de Verlet.
   $\boldsymbol{\theta}_*$ accepté avec probabilité:

$$min(1, \exp[H(\boldsymbol{\theta}_n, \boldsymbol{p}_{n+1}) - H(\boldsymbol{\theta}_*, \boldsymbol{p}_*)])$$

Interprétation et remarques

Mise en place

Algorithme

Require: $G(0)$ $-\mathcal{L}(\theta)$ p^0 θ^0

for $i = 0 : N - 1$ **do**

sample p^{i+1} selon $\mathcal{N}(O, G^i)$

for $j = 1 : Nb_{leapfrogs}$ **do**

$p_{tempo} = p_\tau$

for $k = 1 : N_{pointfixe}$ **do**

$p\left(\tau + \frac{\epsilon}{2}\right) = p\left(\tau\right) - \frac{\epsilon}{2}\nabla_\theta H\left(\theta(\tau), p_{tempo}\right)$

end for

$p\left(\tau + \frac{\epsilon}{2}\right) = p_{tempo}$

$\theta_{tempo} = \theta(\tau)$

for $k = 1 : N_{pointfixe}$ **do**

$\theta(\tau+\epsilon) = \theta(\tau)+\frac{\epsilon}{2}\left[G^{-1}\left(\theta(\tau)\right) + G^{-1}\left(\theta_{tempo}\right)\right] p\left(\tau + \frac{\epsilon}{2}\right)$

end for

$\theta(\tau + \epsilon) = \theta_{tempo}$

$p(\tau + \epsilon) = p\left(\tau + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2}\nabla_\theta H\left(\theta\left(\tau + \frac{\epsilon}{2}\right), p\left(\tau + \frac{\epsilon}{2}\right)\right)$

end for

end for

Plan

Régression Logistique Bayésienne

Modèle:

1. Prior gaussien: $\beta \sim \mathcal{N}(0, \alpha I)$
2. Loi $Y|\beta$: modèle de régression logistique classique

Tenseur métrique choisi: information de Fisher
$G(\beta) = \boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X} + \alpha^{-1} I$
Travail réalisé: Etude comparée de RMHMC et de HMC sur cette tâche et sur données simulées.

Performances

Difficultés rencontrées

Conclusion sur RMHMC

S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth.
Hybird Monte Carlo.
Physics Letters B, 195(2):216–222, 1987.

M. Girolami and B. Calderhead.
Riemann manifold Langevin and Hamiltonian Monte Carlo methods.
The Royal Statistical Society, 73(1):1–37, 2011.