Stability of Markov Chains & Approximate MCMC methods

Pierre Alquier

ENSAE ParisTech

Workshop
"Markov chain Monte Carlo (MCMC) methods"
26/11/2014

Metropolis-Hastings Algorithm

- Objective : simulate from $p(\theta)$.

Metropolis-Hastings Algorithm

- Objective : simulate from $p(\theta)$.
- Usually : $p(\theta) = \pi(\theta|x) \propto \pi(\theta)\mathcal{L}_\theta(x)$ known up to a constant.

Metropolis-Hastings Algorithm

- Objective : simulate from $p(\theta)$.
- Usually : $p(\theta) = \pi(\theta|x) \propto \pi(\theta)\mathcal{L}_\theta(x)$ known up to a constant.

Metropolis-Hastings Algorithm (MH)

- arbitraty θ_0,
- given θ_n,
  1. draw $t_{n+1} \sim q(\cdot|\theta_n)$,
  2. $\theta_{n+1} = \begin{cases} t_{n+1} \text{ with proba. } a(\theta_n, t_{n+1}) \\ \theta_n \text{ otherwise.} \end{cases}$

$$a(\theta, t) = \frac{p(t)q(\theta|t)}{p(\theta)q(t|\theta)} \wedge 1.$$

Potential Problem with MH (1/2)

In some situations, the computation of the acceptance ratio $a(\theta_n, t_{n+1})$ is too slow.

Example 1 : "Big Data"

$x = (x_1, \dots, x_n)$ iid with n **very** large,

$$p(\theta) \propto \pi(\theta) \prod_{i=1}^{n} f_\theta(x_i).$$

A. Korattikara, Y. Chen & M. Welling (2014). Austerity in MCMC Land : Cutting the Metropolis-Hastings Budget. *Proceedings of ICML 2014.*

Potential Problem with MH (2/2)

Example 2 : Exponential Random Graph Model (ERGM)

Given a set of nodes $\{1, \dots, n\}$, and x a graph on these nodes represented by the adjacency matrix $x_{i,j} = 1 \Leftrightarrow$ "i and j are connected", and $s(x)$ be a vector of statistics. We define :

$$f_\theta(x) = \frac{\exp(\theta^T s(x))}{\sum_y \exp(\theta^T s(x))} = \frac{\exp(\theta^T s(x))}{Z(\theta)}.$$

Then

$$a(\theta, t) = \frac{\pi(t)\exp(t^T s(x)) Z(\theta) q(\theta|t)}{\pi(\theta)\exp(\theta^T s(x)) Z(t) q(t|\theta)}.$$

A. Caimo & N. Friel (2011). Bayesian Inference for Exponential Random Graphs Model. *Social Networks* 33 :41–55.

Noisy MCMC Algorithm

In this talk, we focus on the following algorithm :

Noisy MCMC Algorithm

- arbitraty θ_0,
- given θ_n,
  1. draw $t_{n+1} \sim q(\cdot)$,
  2. $\theta_{n+1} = \begin{cases} t_{n+1} \text{ with proba. } \hat{a}(\theta_n, t_{n+1}) \\ \theta_n \text{ otherwise,} \end{cases}$

where $\hat{a}(\theta, t)$ is any approximation of $a(\theta, t)$.

Some Remarks

- There is no reason for this algorithm to target the objective $p(\theta)$.

Some Remarks

- There is no reason for this algorithm to target the objective $p(\theta)$.
- Conditons on $\hat{a}(\theta, t)$ such that we sample from a distribution that is "not too far from $p(\theta)$" ?

Some Remarks

- There is no reason for this algorithm to target the objective $p(\theta)$.
- Conditons on $\hat{a}(\theta, t)$ such that we sample from a distribution that is "not too far from $p(\theta)$" ?
- When $P \simeq P'$ (transition kernels), what can we say about their asymptotic distributions ? $\rightarrow$ *stability of Markov Chains* theory,

N. V. Kartashov (1996). *Strong Stable Markov Chains*, VSP, Utrecht.

Some Remarks

- There is no reason for this algorithm to target the objective $p(\theta)$.
- Conditons on $\hat{a}(\theta, t)$ such that we sample from a distribution that is "not too far from $p(\theta)$" ?
- When $P \simeq P'$ (transition kernels), what can we say about their asymptotic distributions ? $\rightarrow$ *stability of Markov Chains* theory,

N. V. Kartashov (1996). *Strong Stable Markov Chains*, VSP, Utrecht.

- This is a general idea in computational statistics : if your task is beyond your computational power, solve a simpler task and hope (prove ?) that the two solutions are not so different. E.g. : the LASSO.

Overview of the Talk

Total Variation (TV) Distance bewteen Kernels

Reminder :

$$\|\pi - \pi'\|_{\mathrm{TV}} = \sup_{A \text{ event}} |\pi(A) - \pi'(A)| = \frac{1}{2} \int |\pi(\theta) - \pi'(\theta)| \mathrm{d}\theta.$$

Total Variation (TV) Distance bewteen Kernels

Reminder :

$$\|\pi - \pi'\|_{\mathrm{TV}} = \sup_{A \text{ event}} |\pi(A) - \pi'(A)| = \frac{1}{2}\int |\pi(\theta) - \pi'(\theta)| \mathrm{d}\theta.$$

Definition - TV for Kernels

$$\|P - P'\|_{\mathrm{TV}} = \sup_x \|P(x,\cdot.) - P'(x,\cdot)\|_{\mathrm{TV}}.$$

Reminder : Convergence of Markov Chains

Ergodicity

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \xrightarrow[n\to\infty]{} 0.$$

Reminder : Convergence of Markov Chains

Ergodicity

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \xrightarrow[n\to\infty]{} 0.$$

Geometric ergodicity, $\rho < 1$

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \leq C(\theta_0)\rho^n.$$

Reminder : Convergence of Markov Chains

Ergodicity

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \xrightarrow[n\to\infty]{} 0.$$

Geometric ergodicity, $\rho < 1$

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \leq C(\theta_0)\rho^n.$$

Comes with CLT :

$$\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} g(\theta_i) - \mathbb{E}_{T\sim\pi}[g(T)]\right\} \xrightarrow[n\to\infty]{d.} \mathcal{N}(0, \sigma_g^2).$$

S. P. Meyn & R. L. Tweedie (1993). *Markov Chains and Stochastic Stability*, Springer.

Reminder : Convergence of Markov Chains

Ergodicity

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \xrightarrow[n\to\infty]{} 0.$$

Geometric ergodicity, $\rho < 1$

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \leq C(\theta_0)\rho^n.$$

Uniform ergodicity, $C \geq 1, \rho < 1$

$$\forall \theta_0, \quad \|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \leq C\rho^n.$$

Stability for Uniformly Ergodic Chains

Theorem

Assume that P is uniformly ergodic : $\|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \leq C\rho^n$. Then

$$\|\delta_{\theta_0} P^n - \delta_{\theta_0} (P')^n\|_{\mathrm{TV}} \leq K(C,\rho) \|P - P'\|_{\mathrm{TV}},$$

$$K(C,\rho) = \hat{n} + \frac{C\rho^{\hat{n}}}{1-\rho} \text{ and } \hat{n} = \left\lceil \frac{\log(1/C)}{\log(\rho)} \right\rceil .$$

A. Yu. Mitrophanov (2005). Sensitivity and Convergence of Uniformly Ergodic Markov Chains. *Journal of Applied Probability* 42 :1003–1014.

Refined Version

Assume that P is uniformly ergodic : $\|\delta_{\theta_0} P^n - \pi\|_{\mathrm{TV}} \leq C\rho^n$.
Then

$$\|p_0 P^n - p_0'(P')^n\|_{\mathrm{TV}} \leq \begin{cases} \|p_0 - p_0'\|_{\mathrm{TV}} + n\|P - P'\|_{\mathrm{TV}} \\ \qquad \text{when } n \leq \hat{n}, \\ \\ C\rho^n \|p_0 - p_0'\|_{\mathrm{TV}} \\ \qquad + \left(\hat{n} + C\frac{\rho^{\hat{n}} - \rho^n}{1-\rho}\right)\|P - P'\|_{\mathrm{TV}} \\ \qquad \text{when } n > \hat{n}. \end{cases}$$

Consequence for Noisy MCMC

Reminder : $a(\theta, t)$ and approximation $\hat{a}(\theta, t)$. Note that $\hat{a}(\theta, t)$ might be based on additional Monte Carlo simulations, in this case, we should write $\hat{a}(\theta, t, S)$ where S stands for these simulations.

Corollary

- There is a function $\delta(\theta, t)$ such that

$$\mathbb{E}_S \left| a(\theta, t) - \hat{a}(\theta, t, S) \right| \leq \delta(\theta, t).$$

- The kernel P associated with $a(\theta, t)$ is uniformly ergodic with constants C, ρ.

Then $\|\delta_{\theta_0} P^n - \delta_{\theta_0} \hat{P}^n\|_{\mathrm{TV}} \leq 2K(C, \rho) \sup_{\theta} \int q(\mathrm{d}t|\theta)\delta(\theta, t).$

Intractable Likelihood / ERGM

$$f_\theta(x) = \frac{\exp(\theta^T s(x))}{Z(\theta)}, \; a(\theta, t) = \frac{\pi(t)\exp(t^T s(x)) q(\theta)}{\pi(\theta)\exp(\theta^T s(x)) q(t)} \frac{Z(\theta)}{Z(t)} \wedge 1$$

and we cannot compute Z.

Intractable Likelihood / ERGM

$$f_\theta(x) = \frac{\exp(\theta^T s(x))}{Z(\theta)}, \; a(\theta, t) = \frac{\pi(t)\exp(t^T s(x))q(\theta)}{\pi(\theta)\exp(\theta^T s(x))q(t)} \frac{Z(\theta)}{Z(t)} \wedge 1$$

and we cannot compute Z. However,

$$\mathbb{E}_{x\sim f_t}\left(\frac{\exp(\theta^T s(x))}{\exp(t^T s(x))}\right) = \sum_x \frac{\exp(\theta^T s(x))}{\exp(t^T s(x))} \frac{\exp(t^T s(x))}{Z(t)} = \frac{Z(\theta)}{Z(t)}$$

Intractable Likelihood / ERGM

$$f_\theta(x) = \frac{\exp(\theta^T s(x))}{Z(\theta)}, \; a(\theta, t) = \frac{\pi(t)\exp(t^T s(x))q(\theta)}{\pi(\theta)\exp(\theta^T s(x))q(t)} \frac{Z(\theta)}{Z(t)} \wedge 1$$

and we cannot compute Z. However,

$$\mathbb{E}_{x\sim f_t}\left(\frac{\exp(\theta^T s(x))}{\exp(t^T s(x))}\right) = \sum_x \frac{\exp(\theta^T s(x))}{\exp(t^T s(x))} \frac{\exp(t^T s(x))}{Z(t)} = \frac{Z(\theta)}{Z(t)}$$

so we can draw $S_N = (x_1, \dots, x_N)$ iid from f_t (feasible) and

$$\hat{a}(\theta, t, S_N) = 1 \wedge \frac{\pi(t)\exp(t^T s(x))q(\theta)}{\pi(\theta)\exp(\theta^T s(x))q(t)} \frac{1}{N}\sum_{i=1}^{N} \frac{\exp(\theta^T s(x_i))}{\exp(t^T s(x_i))}.$$

Comments

$$\hat{a}(\theta, t, S_N) = 1 \wedge \frac{\pi(t)\exp(t^T s(x))q(\theta)}{\pi(\theta)\exp(\theta^T s(x))q(t)} \frac{1}{N}\sum_{i=1}^{N} \frac{\exp(\theta^T s(x_i))}{\exp(t^T s(x_i))}.$$

- notation : P the original kernel, $\hat{P}_N$ the approx.

Comments

$$\hat{a}(\theta, t, S_N) = 1 \wedge \frac{\pi(t)\exp(t^T s(x))q(\theta)}{\pi(\theta)\exp(\theta^T s(x))q(t)} \frac{1}{N}\sum_{i=1}^{N} \frac{\exp(\theta^T s(x_i))}{\exp(t^T s(x_i))}.$$

- notation : P the original kernel, $\hat{P}_N$ the approx.
- for $N = 1$, this algorithm is known as the exchange algorithm and is known to be exact.

I. Murray, Z. Ghahramani & D. MacKay (2006). MCMC for Doubly-Intractable Distributions. *Proceedings of the 22nd Conference on Uncertainty and Artificial Intelligence*, AUAI Press.

Comments

$$\hat{a}(\theta, t, S_N) = 1 \wedge \frac{\pi(t)\exp(t^T s(x))q(\theta)}{\pi(\theta)\exp(\theta^T s(x))q(t)} \frac{1}{N}\sum_{i=1}^{N} \frac{\exp(\theta^T s(x_i))}{\exp(t^T s(x_i))}.$$

- notation : P the original kernel, $\hat{P}_N$ the approx.
- for $N = 1$, this algorithm is known as the exchange algorithm and is known to be exact.

I. Murray, Z. Ghahramani & D. MacKay (2006). MCMC for Doubly-Intractable Distributions. *Proceedings of the 22nd Conference on Uncertainty and Artificial Intelligence*, AUAI Press.

- Better mixing when $N > 1$, but in this case the algorithm is no longer exact. Mitrophanov's theorem will tell us how to calibrate N to reach a given accuracy.

Noisy MCMC for ERGM

Corollary

Assume that

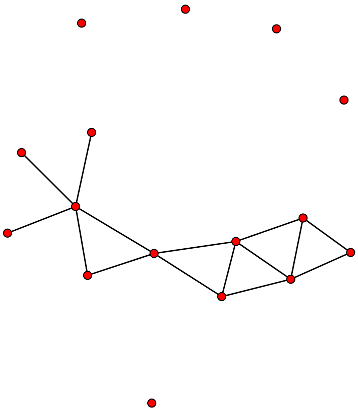
- the parameter space is bounded : $\sup_{\theta\in\Theta}\|\theta\| = \mathcal{T} < \infty$,
- there is a constant $c > 0$ such that

$$c \leq \pi(\theta), q(\theta|t) \leq 1/c.$$

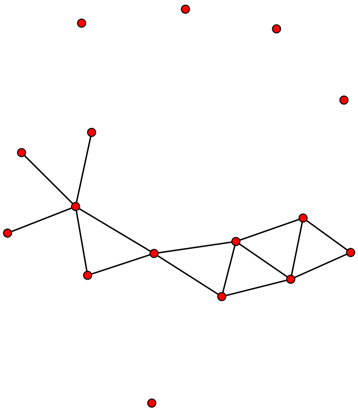
Then : $$\|\delta_{\theta_0}P^n - \delta_{\theta_0}\hat{P}_N^n\|_{\mathrm{TV}} \leq \frac{\mathcal{C}(\mathcal{T},c,s)}{\sqrt{N}}, \mathcal{C}(\mathcal{T},c,s) \text{ known.}$$

P. Alquier, N. Friel, R. G. Everitt & A. Boland (2014). Noisy Monte-Carlo : Convergence of Markov Chains with Approximate Transition Kernels. *Statistics and Computing*, to appear.

Simulations : Florentine Family Business Dataset



Simulations : Florentine Family Business Dataset



$$s(x) = (s_1(x), s_2(x))$$

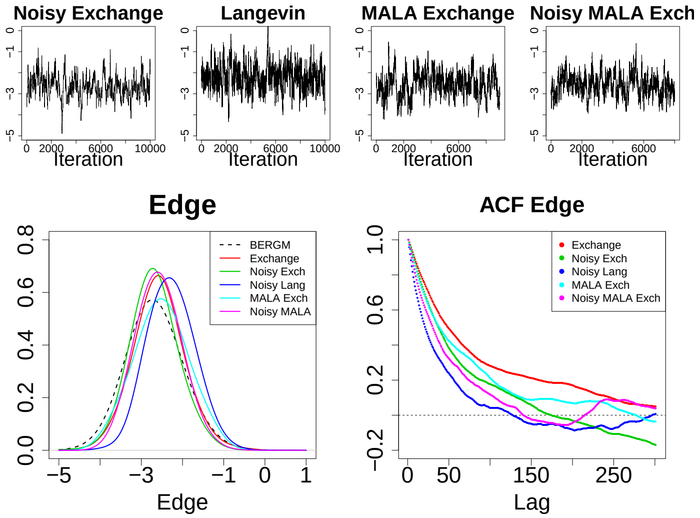
- $s_1(x)$ number of edges,
- $s_2(x)$ number of 2-stars.

Numerical Results

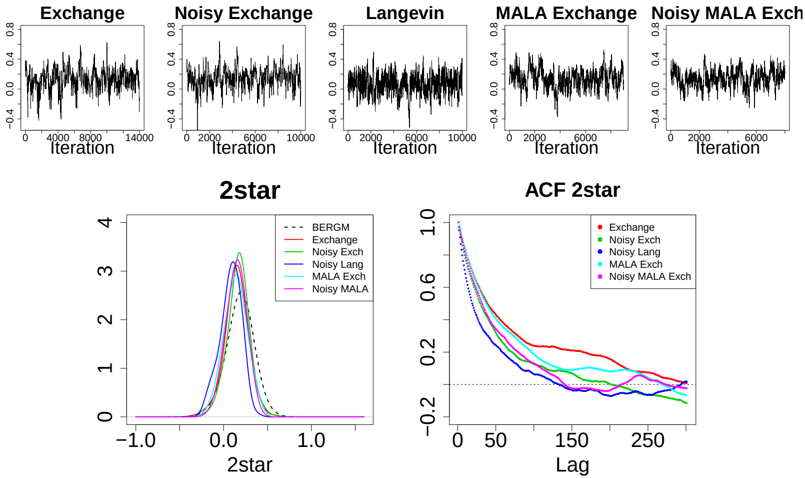
Method	Edge Mean	Edge SD	2-star Mean	2-star SD
BERGM	-2.675	0.647	0.188	0.155
Exchange	-2.573	0.568	0.146	0.133
Noisy Exchange	-2.686	0.526	0.167	0.122
Noisy Langevin	-2.281	0.513	0.081	0.119
MALA Exchange	-2.518	0.62	0.136	0.128
Noisy MALA	-2.584	0.498	0.144	0.113

Table : Posterior means and standard deviations.

Chains, density and ACF plot for the edge statistic.



Chains, density and ACF plot for the 2-star stat.



Austerity in MCMC Land

$x = (x_1, \dots, x_n)$ iid with n **very** large, $p(\theta) \propto \pi(\theta) \prod_{i=1}^n f_\theta(x_i)$.

A. Korattikara, Y. Chen & M. Welling (2014). Austerity in MCMC Land : Cutting the Metropolis-Hastings Budget. *Proceedings of ICML 2014*.

MH :

1. draw $t_{n+1} \sim q(\cdot|\theta_n)$, $U \sim \mathcal{U}[0,1]$,
2. $\theta_{n+1} =$
$$\begin{cases} t_{n+1} \text{ when } U \leq b(\theta_n, t_{n+1}) = \frac{\pi(t_{n+1}) \prod_{i=1}^n f_{t_{n+1}}(x_i) q(\theta_n|t_{n+1})}{\pi(\theta_n) \prod_{i=1}^n f_{\theta_n}(x_i) q(t_{n+1}|\theta_n)} \\ \theta_n \text{ otherwise.} \end{cases}$$

Austerity in MCMC Land

$x = (x_1, \dots, x_n)$ iid with n **very** large, $p(\theta) \propto \pi(\theta) \prod_{i=1}^n f_\theta(x_i)$.

A. Korattikara, Y. Chen & M. Welling (2014). Austerity in MCMC Land : Cutting the Metropolis-Hastings Budget. *Proceedings of ICML 2014*.

MH :

1. draw $t_{n+1} \sim q(\cdot|\theta_n)$, $U \sim \mathcal{U}[0,1]$,
2. $\theta_{n+1} =$
$$\begin{cases} t_{n+1} \text{ when } U \leq b(\theta_n, t_{n+1}) = \frac{\pi(t_{n+1}) \prod_{i=1}^n f_{t_{n+1}}(x_i) q(\theta_n|t_{n+1})}{\pi(\theta_n) \prod_{i=1}^n f_{\theta_n}(x_i) q(t_{n+1}|\theta_n)} \\ \theta_n \text{ otherwise.} \end{cases}$$

Idea : to test the hypothesis $U \leq b(\theta_n, t_{n+1})$ with a given confidence level instead.

Testing $U \leq b(\theta, t)$

$$U \leq \frac{\pi(t) \prod_{i=1}^{n} f_t(x_i) q(\theta|t)}{\pi(\theta) \prod_{i=1}^{n} f_\theta(x_i) q(t|\theta)}$$

$$\mathbf{H}_0 : \quad \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_t(x_i)}{f_\theta(x_i)} \geq \frac{1}{n} \log \left[U \frac{\pi(\theta) q(t|\theta)}{\pi(t) q(\theta|t)} \right] =: V.$$

Testing $U \leq b(\theta, t)$

$$U \leq \frac{\pi(t) \prod_{i=1}^{n} f_t(x_i) q(\theta|t)}{\pi(\theta) \prod_{i=1}^{n} f_\theta(x_i) q(t|\theta)}$$

$$\mathbf{H}_0 : \quad \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_t(x_i)}{f_\theta(x_i)} \geq \frac{1}{n} \log \left[U \frac{\pi(\theta) q(t|\theta)}{\pi(t) q(\theta|t)} \right] =: V.$$

We draw $(y_1, \dots, y_B)$ iid in $\{x_1, \dots, x_n\}$, $B \ll n$,

$$\frac{1}{B} \sum_{i=1}^{B} \log \frac{f_t(y_i)}{f_\theta(y_i)} \begin{cases} \geq V + c \Rightarrow & \text{accept } \mathbf{H}_0, \\ \leq V - c \Rightarrow & \text{reject } \mathbf{H}_0, \\ \in]V - c, V + c[\Rightarrow & \text{start again, increase } B. \end{cases}$$

We choose c so that the type 1 and type 2 errors are $\leq \alpha$ fixed.

Theoretical Analysis

R. Bardenet, A. Doucet & C. Holmes (2014). Towards Scaling up Markov Chain Monte Carlo : an Adaptive Subsampling Approach. *Proceedings of ICML 2014*.

They calibrate c through Audibert's empirical Bernstein's inequality and obtain :

Theorem

Denote P_α the approximate kernel with level $\alpha > 0$. Assume that the original kernel P is uniformly ergodic (C, ρ). Then

$$\|\delta_{\theta_0} P^n - \delta_{\theta_0} P_\alpha^n\|_{\mathrm{TV}} \leq \alpha K'(C, \rho).$$

Theoretical Analysis

R. Bardenet, A. Doucet & C. Holmes (2014). Towards Scaling up Markov Chain Monte Carlo : an Adaptive Subsampling Approach. *Proceedings of ICML 2014.*

They calibrate c through Audibert's empirical Bernstein's inequality and obtain :

Theorem

Denote P_α the approximate kernel with level $\alpha > 0$. Assume that the original kernel P is uniformly ergodic (C, ρ). Then

$$\|\delta_{\theta_0} P^n - \delta_{\theta_0} P_\alpha^n\|_{\mathrm{TV}} \leq \alpha K'(C, \rho).$$

However, limitation : they prove that after the burn-in period, we tend to need $B \simeq n/2$ at each step.

Limitations

$$\|\delta_{\theta_0} P^n - \delta_{\theta_0} \hat{P}^n\|_{\mathrm{TV}} \leq 2K(C, \rho) \sup_{\theta} \int q(\mathrm{d}t|\theta)\delta(\theta, t).$$

Limitations

$$\|\delta_{\theta_0} P^n - \delta_{\theta_0} \hat{P}^n\|_{\mathrm{TV}} \leq 2K(C, \rho) \sup_{\theta} \int q(\mathrm{d}t|\theta)\delta(\theta, t).$$

- First, in many situations, $\int q(\mathrm{d}t|\theta)\delta(\theta, t) = \infty$. (However, in some cases, $\delta(\theta, t) = \delta$ can be made arbitraty small as in the examples above).

Limitations

$$\|\delta_{\theta_0} P^n - \delta_{\theta_0} \hat{P}^n\|_{\mathrm{TV}} \leq 2K(C, \rho) \sup_{\theta} \int q(\mathrm{d}t|\theta)\delta(\theta, t).$$

- First, in many situations, $\int q(\mathrm{d}t|\theta)\delta(\theta, t) = \infty$. (However, in some cases, $\delta(\theta, t) = \delta$ can be made arbitraty small as in the examples above).
- Uniformly ergodic chains are *rare* when one uses MH algorithm. Example : in the ERGM model, we need a bounded Θ...

Beyond Uniformly Ergodic Markov chains

Reminder : a criterion for geometric ergodicity.
$\exists \lambda \in (0,1), \exists b < \infty, \exists V : \Theta \to [1,\infty], \exists \mathcal{C} \subset \Theta$ such that

$$\forall \theta \in \Theta, \int P(\theta, \mathrm{d}\theta') V(\theta') \leq \lambda V(\theta) + b \mathbf{1}_{\theta \in \mathcal{C}}.$$

Beyond Uniformly Ergodic Markov chains

Reminder : a criterion for geometric ergodicity.
$\exists\lambda \in (0,1), \exists b < \infty, \exists V : \Theta \to [1,\infty], \exists \mathcal{C} \subset \Theta$ such that

$$\forall \theta \in \Theta, \int P(\theta, \mathrm{d}\theta')V(\theta') \leq \lambda V(\theta) + b\mathbf{1}_{\theta\in\mathcal{C}}.$$

This *drift condition*, together with "usual" assumptions (irreducibility...) ensure that P is V-uniformly ergodic :

$$\|\delta_{\theta_0} P^n - \pi\|_V \leq C\rho^n V(\theta_0)$$

and this turn out to be equivalent to geometric ergodicity !

Beyond Uniformly Ergodic Markov chains

Reminder : a criterion for geometric ergodicity.
$\exists \lambda \in (0,1), \exists b < \infty, \exists V : \Theta \to [1,\infty], \exists \mathcal{C} \subset \Theta$ such that

$$\forall \theta \in \Theta, \int P(\theta, \mathrm{d}\theta') V(\theta') \leq \lambda V(\theta) + b \mathbf{1}_{\theta \in \mathcal{C}}.$$

This *drift condition*, together with "usual" assumptions (irreducibility...) ensure that P is V-uniformly ergodic :

$$\|\delta_{\theta_0} P^n - \pi\|_V \leq C \rho^n V(\theta_0)$$

and this turn out to be equivalent to geometric ergodicity ! This is satisfied in many practical situations,

G. O. Roberts & J. S. Rosenthal (2004). General State Space Markov Chains and MCMC Algorithms. *Probability Surveys* 1 :20–71.

Stability of Geometrically Ergodic Markov Chains

Theorem

Sequence of kernels $\|\hat{P}_N - P\|_{\mathrm{TV}} \to 0$, $\exists V(\cdot) \geq 1$:

- P is V-uniformly ergodic ;
- $\exists N_0 \in \mathbb{N}, 0 < \delta < 1, L > 0, \forall N \geq N_0,$

$$\int V(\theta)\hat{P}_N(\theta_0, \mathrm{d}\theta) \leq \delta V(\theta_0) + L.$$

Then for N large enough, $\hat{P}_N$ is geometrically ergodic with limiting distribution π_N and $\|\pi_N - \pi\| \xrightarrow[N\to\infty]{} 0$.

D. Ferré, L. Hervé & J. Ledoux (2013). Regular Perturbations of V-Geometrically Ergodic Markov Chains. *Journal of Applied Probability* **50** :184–194.

More improvements

Recent works on non-asymptotic bounds for MCMC based on the Ricci's curvature of P :

A. Joulin & Y. Ollivier (2010). Curvature, Concentration and Error Estimates for Markov Chain Monte Carlo. *The Annals of Probability* **38** :2418–2442.

A. Drumus & E. Moulines (2014). Quantitative Bounds of Convergence for Geometrically Ergodic Markov Chains in the Wasserstein Distance with Application to the Metropolis Adjusted Langevin Algorithm. *Statistics and Computing*, to appear.

Stability using these tools :

N. S. Pillai & A. Smith (2014). Ergodicity of Approximate MCMC Chains with Applications to Large Data Sets. *Preprint* arXiv :1405.0182.