A Dirichlet Form approach to MCMC Optimal Scaling

Giacomo Zanella, Wilfrid S. Kendall, and Mylène Bédard.

g.zanella@warwick.ac.uk, w.s.kendall@warwick.ac.uk,
mylene.bedard@umontreal.ca

Supported by EPSRC Research Grants EP/D002060, EP/K013939

LMS Durham Symposium on Stochastic Analysis

12th July 2017

Warwick Statistics

Introduction

Introduction to Markov chain Monte Carlo (MCMC)

General reference: Brooks et al. (2011) MCMC Handbook.

Suppose x represents an unknown (and therefore random!) parameter, and y represents data depending on the unknown parameter, joint probability density $p(x, y)$.

Introduction to Markov chain Monte Carlo (MCMC)

General reference: Brooks et al. (2011) MCMC Handbook.

Suppose x represents an unknown (and therefore random!) parameter, and y represents data depending on the unknown parameter, joint probability density $p(x, y)$.

Joint probability density

↓

$$p(x, y)$$

Introduction to Markov chain Monte Carlo (MCMC)

General reference: Brooks et al. (2011) MCMC Handbook.

Suppose x represents an unknown (and therefore random!) parameter, and y represents data depending on the unknown parameter, joint probability density $p(x, y)$.

Conditional density ↓   Joint probability density ↓

$$p(x|y) = \frac{p(x,y)}{Z}$$

↑

Norming constant Z can be hard to compute!

Introduction to Markov chain Monte Carlo (MCMC)

General reference: Brooks et al. (2011) MCMC Handbook.

Suppose x represents an unknown (and therefore random!) parameter, and y represents data depending on the unknown parameter, joint probability density $p(x, y)$.

Conditional density ↓ — Joint probability density ↓

$$p(x|y) = \frac{p(x,y)}{Z}$$

↑ Build Markov chain with this as equilibrium (no need to know Z)

↑ Norming constant Z can be hard to compute!

Introduction to Markov chain Monte Carlo (MCMC)

General reference: Brooks et al. (2011) MCMC Handbook.

Suppose x represents an unknown (and therefore random!) parameter, and y represents data depending on the unknown parameter, joint probability density $p(x, y)$.

Conditional density → $p(x|y)$; Joint probability density → $p(x,y)$

$$p(x|y) = \frac{p(x, y)}{Z}$$

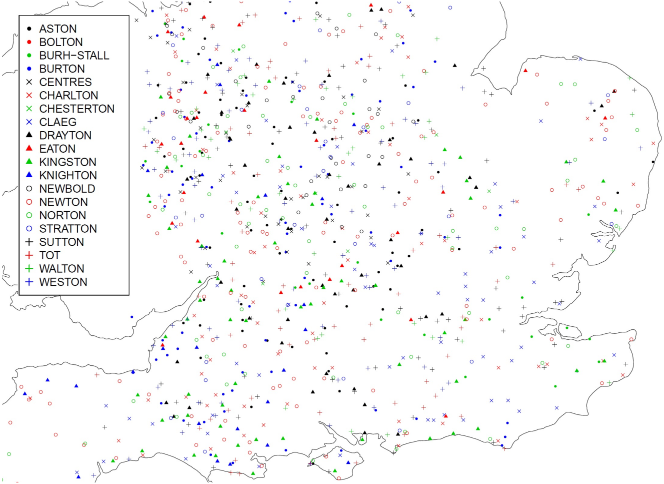
Build Markov chain with this as equilibrium (no need to know Z) ← $p(x|y)$

Norming constant Z can be hard to compute! ← Z

Simulate Markov chain till approximate equilibrium.

Example: MCMC for Anglo-Saxon statistics

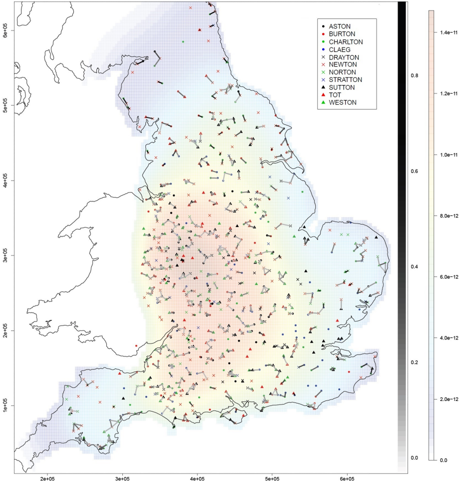
Some historians conjecture, Anglo-Saxon placenames cluster by *dissimilar* names. Zanella (2015, 2016) uses MCMC: data provides some support, resulting in useful clustering.

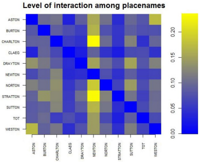


Warwick Statistics

Example: MCMC for Anglo-Saxon statistics

Some historians conjecture, Anglo-Saxon placenames cluster by *dissimilar* names. Zanella (2015, 2016) uses MCMC: data provides some support, resulting in useful clustering.





MCMC and optimal scaling

MCMC idea

Goal: estimate $E = \mathbb{E}_\pi[h(X)]$.

MCMC idea

Goal: estimate $E = \mathbb{E}_\pi[h(X)]$.

Method: simulate ergodic Markov chain with stationary distribution π: use empirical estimate $\hat{E}_n = \frac{1}{n}\sum_{n=n_0}^{n_0+n} h(X_n)$.

MCMC idea

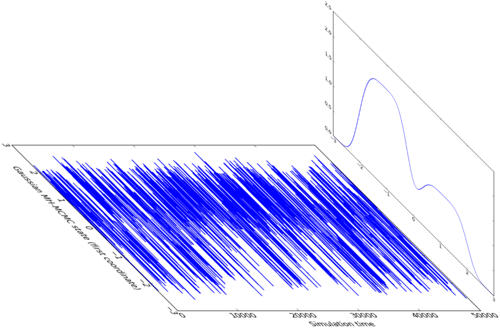
Goal: estimate $E = \mathbb{E}_\pi[h(X)]$.

Method: simulate ergodic Markov chain with stationary distribution π: use empirical estimate $\hat{E}_n = \frac{1}{n}\sum_{n=n_0}^{n_0+n} h(X_n)$.

(Much easier to apply theory if chain is reversible.)

MCMC idea

Goal: estimate $E = \mathbb{E}_\pi[h(X)]$.
Method: simulate ergodic Markov chain with stationary distribution π: use empirical estimate $\hat{E}_n = \frac{1}{n}\sum_{n=n_0}^{n_0+n} h(X_n)$.
(Much easier to apply theory if chain is reversible.)
Theory: $\hat{E}_n \to E$ almost surely.



Varieties of MH-MCMC

Here is the famous Metropolis-Hastings recipe for drawing from a distribution with density f:

```
Propose Y using conditional density q(y|x);
Accept/Reject move from X to Y,
   based on ratio f(Y) q(X|Y) / f(X) q(Y|X)
```

Varieties of MH-MCMC

Here is the famous Metropolis-Hastings recipe for drawing from a distribution with density f:

```
Propose Y using conditional density q(y|x);
Accept/Reject move from X to Y,
   based on ratio f(Y) q(X|Y) / f(X) q(Y|X)
```

Options:

1. Independence sampler: proposal $q(y|x) = q(y)$ doesn't depend on x;

Varieties of MH-MCMC

Here is the famous Metropolis-Hastings recipe for drawing from a distribution with density f:

```
Propose Y using conditional density q(y|x);
Accept/Reject move from X to Y,
   based on ratio f(Y) q(X|Y) / f(X) q(Y|X)
```

Options:

1. Independence sampler: proposal $q(y|x) = q(y)$ doesn't depend on x;
2. Random walk (RW MH-MCMC): proposal $q(y|x) = q(y - x)$ behaves as a random walk;

Varieties of MH-MCMC

Here is the famous Metropolis-Hastings recipe for drawing from a distribution with density f:

```
Propose Y using conditional density q(y|x);
Accept/Reject move from X to Y,
   based on ratio f(Y) q(X|Y) / f(X) q(Y|X)
```

Options:

1. Independence sampler: proposal $q(y|x) = q(y)$ doesn't depend on x;
2. Random walk (RW MH-MCMC): proposal $q(y|x) = q(y - x)$ behaves as a random walk;
3. MALA MH-MCMC: proposal $q(y|x) = q(y - x - \lambda \operatorname{grad} \log f)$ drifts towards high target density f.

Varieties of MH-MCMC

Here is the famous Metropolis-Hastings recipe for drawing from a distribution with density f:

```
Propose Y using conditional density q(y|x);
Accept/Reject move from X to Y,
   based on ratio f(Y) q(X|Y) / f(X) q(Y|X)
```

Options:

1. Independence sampler: proposal $q(y|x) = q(y)$ doesn't depend on x;
2. Random walk (RW MH-MCMC): proposal $q(y|x) = q(y - x)$ behaves as a random walk;
3. MALA MH-MCMC: proposal $q(y|x) = q(y - x - \lambda \operatorname{grad} \log f)$ drifts towards high target density f.

We shall focus on RW MH-MCMC with Gaussian proposals.

Gaussian RW MH-MCMC

Simple `Python` code for Gaussian RW MH-MCMC, using `normal` and `exponential` from `Numpy`:

Gaussian RW MH-MCMC

Simple `Python` code for Gaussian RW MH-MCMC, using `normal` and `exponential` from `Numpy`:

Propose multivariate Gaussian step;

Gaussian RW MH-MCMC

Simple `Python` code for Gaussian RW MH-MCMC, using `normal` and `exponential` from `Numpy`:

Propose multivariate Gaussian step;

Test whether to accept proposal by comparing exponential random variable with log MH ratio;

Gaussian RW MH-MCMC

Simple `Python` code for Gaussian RW MH-MCMC, using `normal` and `exponential` from `Numpy`:

- Propose multivariate Gaussian step;
- Test whether to accept proposal by comparing exponential random variable with log MH ratio;
- Implement step if accepted (vector addition).

Gaussian RW MH-MCMC

Simple Python code for Gaussian RW MH-MCMC, using normal and exponential from Numpy:

- Propose multivariate Gaussian step;
- Test whether to accept proposal by comparing exponential random variable with log MH ratio;
- Implement step if accepted (vector addition).

```
while not mcmc.stopped():
  z = normal(0, tau, size=mcmc.dim)
  if exponential() > mcmc.phi(mcmc.x + z)-mcmc.phi(mcmc.x):
    mcmc.x += z
mcmc.record_result()
```

Gaussian RW MH-MCMC

Simple `Python` code for Gaussian RW MH-MCMC, using `normal` and `exponential` from `Numpy`:

- Propose multivariate Gaussian step;
- Test whether to accept proposal by comparing exponential random variable with log MH ratio;
- Implement step if accepted (vector addition).

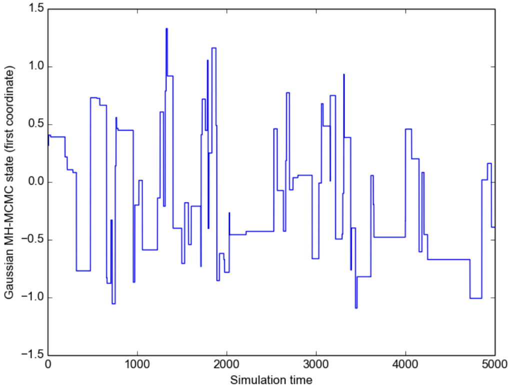
```
while not mcmc.stopped():
  z = normal(0, tau, size=mcmc.dim)
  if exponential() > mcmc.phi(mcmc.x + z)-mcmc.phi(mcmc.x):
    mcmc.x += z
mcmc.record_result()
```

What is best choice of scale / standard deviation `tau`?

RW MH-MCMC with Gaussian proposals

(smooth target, marginal $\propto \exp(-x^4)$)

Target is given by 10 *i.i.d.* coordinates.



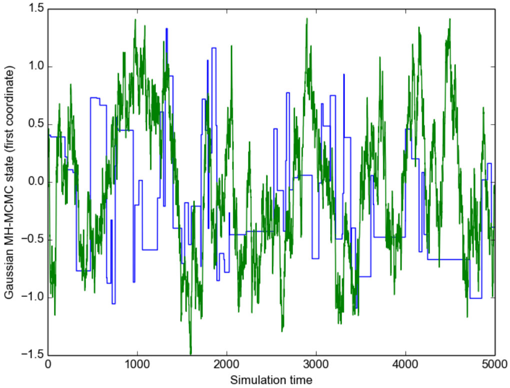
Scale parameter for proposal: $\tau = 1$ is too large!

Acceptance ratio 1.7%

RW MH-MCMC with Gaussian proposals

(smooth target, marginal $\propto \exp(-x^4)$)

Target is given by 10 *i.i.d.* coordinates.



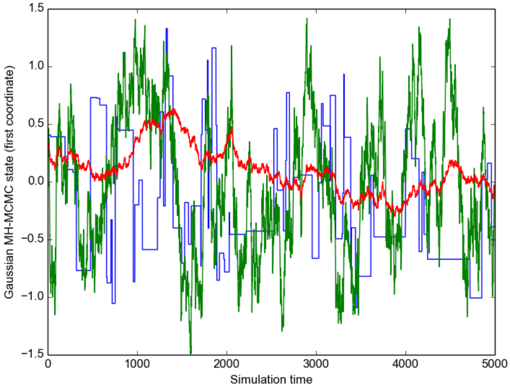
Scale parameter for proposal: $\tau = 0.1$ is better.

Acceptance ratio 76.5%

RW MH-MCMC with Gaussian proposals

(smooth target, marginal $\propto \exp(-x^4)$)

Target is given by 10 *i.i.d.* coordinates.



Scale parameter for proposal: $\tau = 0.01$ is too small.

Acceptance ratio 98.5%

MCMC Optimal Scaling: classic result (I)

RW MH-MCMC on $(\mathbb{R}^d, \pi^{\otimes d})$

$\pi(dx_i) = e^{-\phi(x_i)} dx_i$; MH acceptance rule $A^{(d)} = 0$ or 1.

$$\underline{X}_0^{(d)} = (\quad X_1 \quad , \ldots , \quad X_d \quad) \qquad X_i \overset{iid}{\sim} \pi$$

$$\underline{X}_1^{(d)} = (X_1 + A^{(d)} W_1, \ldots , X_d + A^{(d)} W_d) \qquad W_i \overset{iid}{\sim} N(0, \sigma_d^2)$$

Warwick Statistics

MCMC Optimal Scaling: classic result (I)

RW MH-MCMC on $(\mathbb{R}^d, \pi^{\otimes d})$

$\pi(dx_i) = e^{-\phi(x_i)} dx_i$; MH acceptance rule $A^{(d)} = 0$ or 1.

$$\underline{X}_0^{(d)} = (\quad X_1 \quad , \ldots , \quad X_d \quad) \qquad X_i \overset{iid}{\sim} \pi$$

$$\underline{X}_1^{(d)} = (X_1 + A^{(d)} W_1, \ldots , X_d + A^{(d)} W_d) \qquad W_i \overset{iid}{\sim} N(0, \sigma_d^2)$$

Questions: **(1)** complexity as $d \uparrow \infty$? **(2)** optimal σ_d?

MCMC Optimal Scaling: classic result (I)

RW MH-MCMC on $(\mathbb{R}^d, \pi^{\otimes d})$

$\pi(dx_i) = e^{-\phi(x_i)} dx_i$; MH acceptance rule $A^{(d)} = 0$ or 1.

$$\underline{X}_0^{(d)} = (\quad X_1 \quad , \ldots , \quad X_d \quad) \qquad X_i \overset{iid}{\sim} \pi$$

$$\underline{X}_1^{(d)} = (X_1 + A^{(d)} W_1, \ldots , X_d + A^{(d)} W_d) \qquad W_i \overset{iid}{\sim} N(0, \sigma_d^2)$$

Questions: **(1)** complexity as $d \uparrow \infty$? **(2)** optimal σ_d?

Theorem (Roberts, Gelman and Gilks, 1997)

Given $\sigma_d^2 = \frac{\sigma^2}{d}$, *Lipschitz* ϕ', *and finite* $\mathbb{E}_\pi[(\phi')^8]$, $\mathbb{E}_\pi[(\phi'')^4]$

$$\{X^{(d)}_{\lfloor td \rfloor,1}\}_t \Rightarrow Z \quad \textit{where} \quad \mathrm{d}Z_t = s(\sigma)^{\frac{1}{2}} \, \mathrm{d}B_t + \tfrac{1}{2} s(\sigma) \, \phi'(Z_t) \, \mathrm{d}t \, .$$

Warwick Statistics

MCMC Optimal Scaling: classic result (I)

RW MH-MCMC on $(\mathbb{R}^d, \pi^{\otimes d})$

$\pi(dx_i) = e^{-\phi(x_i)} dx_i$; MH acceptance rule $A^{(d)} = 0$ or 1.

$$\underline{X}_0^{(d)} = (\quad X_1 \quad , \ldots , \quad X_d \quad) \qquad X_i \overset{iid}{\sim} \pi$$

$$\underline{X}_1^{(d)} = (X_1 + A^{(d)} W_1, \ldots , X_d + A^{(d)} W_d) \qquad W_i \overset{iid}{\sim} N(0, \sigma_d^2)$$

Questions: **(1)** complexity as $d \uparrow \infty$? **(2)** optimal σ_d?

Theorem (Roberts, Gelman and Gilks, 1997)

Given $\sigma_d^2 = \frac{\sigma^2}{d}$, *Lipschitz* ϕ', *and finite* $\mathbb{E}_\pi[(\phi')^8]$, $\mathbb{E}_\pi[(\phi'')^4]$

$$\{X^{(d)}_{\lfloor td \rfloor,1}\}_t \Rightarrow Z \quad \textit{where} \quad \mathrm{d}Z_t = s(\sigma)^{\frac{1}{2}}\,\mathrm{d}B_t + \frac{1}{2} s(\sigma)\,\phi'(Z_t)\,\mathrm{d}t\,.$$

Answers: **(1)** mix in $O(d)$ steps; **(2)** $\sigma_{\max} = \arg\max_\sigma s(\sigma)$.

MCMC Optimal Scaling: classic result (II)

Optimization: maximize $s(\sigma)$!

Given $\mathcal{I} = \mathbb{E}_\pi[\phi'(X)^2]$ and normal CDF Φ,

$$s(\sigma) = \sigma^2\, 2\Phi(-\frac{\sigma\sqrt{\mathcal{I}}}{2}) = \sigma^2 A(\sigma) = \frac{4}{\mathcal{I}}\,(\Phi^{-1}(\frac{A(\sigma)}{2}))^2 A(\sigma)$$

So σ_{max} maximized by choosing asymptotic acceptance rate $A(\sigma_{\text{max}}) = \arg\max_{A\in[0,1]} \left\{(\Phi^{-1}(\frac{A}{2}))^2 A\right\} \approx 0.234$

MCMC Optimal Scaling: classic result (II)

Optimization: maximize $s(\sigma)$!

Given $\mathcal{I} = \mathbb{E}_\pi[\phi'(X)^2]$ and normal CDF Φ,

$$s(\sigma) \;=\; \sigma^2\, 2\Phi(-\tfrac{\sigma\sqrt{\mathcal{I}}}{2}) = \sigma^2 A(\sigma) \;=\; \tfrac{4}{\mathcal{I}}\, (\Phi^{-1}(\tfrac{A(\sigma)}{2}))^2 A(\sigma)$$

So σ_{max} maximized by choosing asymptotic acceptance rate $A(\sigma_{\text{max}}) = \arg\max_{A\in[0,1]} \left\{(\Phi^{-1}(\tfrac{A}{2}))^2 A\right\} \approx 0.234$

MCMC Optimal Scaling: classic result (II)

Optimization: maximize $s(\sigma)$!

Given $\mathcal{I} = \mathbb{E}_\pi[\phi'(X)^2]$ and normal CDF Φ,

$$s(\sigma) = \sigma^2\, 2\Phi(-\tfrac{\sigma\sqrt{\mathcal{I}}}{2}) = \sigma^2 A(\sigma) = \tfrac{4}{\mathcal{I}}\left(\Phi^{-1}(\tfrac{A(\sigma)}{2})\right)^2 A(\sigma)$$

So σ_{max} maximized by choosing asymptotic acceptance rate $A(\sigma_{\text{max}}) = \arg\max_{A\in[0,1]}\left\{(\Phi^{-1}(\tfrac{A}{2}))^2 A\right\} \approx 0.234$

Strengths:

- Establish complexity as $d \to \infty$;
- Practical information on how to tune proposal;
- Does not depend on ϕ (CLT-type universality).

MCMC Optimal Scaling: classic result (II)

Optimization: maximize $s(\sigma)$!

Given $\mathcal{I} = \mathbb{E}_\pi[\phi'(X)^2]$ and normal CDF Φ,

$$s(\sigma) \;=\; \sigma^2\, 2\Phi(-\tfrac{\sigma\sqrt{\mathcal{I}}}{2}) = \sigma^2 A(\sigma) \;=\; \tfrac{4}{\mathcal{I}}\,(\Phi^{-1}(\tfrac{A(\sigma)}{2}))^2 A(\sigma)$$

So σ_{max} maximized by choosing asymptotic acceptance rate $A(\sigma_{\text{max}}) = \arg\max_{A\in[0,1]}\left\{(\Phi^{-1}(\tfrac{A}{2}))^2 A\right\} \approx 0.234$

Strengths:

- Establish complexity as $d \to \infty$;
- Practical information on how to tune proposal;
- Does not depend on ϕ (CLT-type universality).

Some weaknesses that we will address: (there are others)

- Convergence of marginal rather than joint distribution
- Strong regularity assumptions: Lipschitz g', finite $\mathbb{E}\left[(g')^8\right]$, $\mathbb{E}\left[(g'')^4\right]$.

Warwick Statistics

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);
- Non-identically distributed independent target coordinates (Bédard, 2007);

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);
- Non-identically distributed independent target coordinates (Bédard, 2007);
- Gibbs random fields (Breyer and Roberts, 2000);

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);
- Non-identically distributed independent target coordinates (Bédard, 2007);
- Gibbs random fields (Breyer and Roberts, 2000);
- Infinite dimensional random fields (Mattingly, Pillai and Stuart, 2012);

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);
- Non-identically distributed independent target coordinates (Bédard, 2007);
- Gibbs random fields (Breyer and Roberts, 2000);
- Infinite dimensional random fields (Mattingly, Pillai and Stuart, 2012);
- Markov chains on a hypercube (Roberts, 1998);

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);
- Non-identically distributed independent target coordinates (Bédard, 2007);
- Gibbs random fields (Breyer and Roberts, 2000);
- Infinite dimensional random fields (Mattingly, Pillai and Stuart, 2012);
- Markov chains on a hypercube (Roberts, 1998);
- Adaptive MCMC; adjust online to optimize acceptance probability (Andrieu and Thoms, 2008; Rosenthal, 2011).

MCMC Optimal Scaling: classic result (III)

There is a wide range of extensions: for example,

- Langevin / MALA, for which the magic acceptance probability is 0.574 (Roberts and Rosenthal, 1998);
- Non-identically distributed independent target coordinates (Bédard, 2007);
- Gibbs random fields (Breyer and Roberts, 2000);
- Infinite dimensional random fields (Mattingly, Pillai and Stuart, 2012);
- Markov chains on a hypercube (Roberts, 1998);
- Adaptive MCMC; adjust online to optimize acceptance probability (Andrieu and Thoms, 2008; Rosenthal, 2011).

All these build on the s.d.e. approach of Roberts, Gelman and Gilks (1997); hence regularity conditions tend to be severe (but see Durmus et al., 2016).

Dirichlet forms and optimal scaling

Dirichlet forms and MCMC 1

Definition of Dirichlet form

A (symmetric) Dirichlet form $\mathcal{E}$ on a Hilbert space H is a closed bilinear function $\mathcal{E}(u, v)$, defined / finite for any $u, v \in \mathcal{D} \subseteq H$, which satisfies:

1. $\mathcal{D}$ is a dense linear subspace of H;
2. $\mathcal{E}(u, v) = \mathcal{E}(v, u)$ for $u, v \in \mathcal{D}$, so $\mathcal{E}$ is symmetric;
3. $\mathcal{E}(u) = \mathcal{E}(u, u) \geq 0$ for $u \in \mathcal{D}$;
4. $\mathcal{D}$ is a Hilbert space under the ("Sobolev") inner product $\langle u, v \rangle + \mathcal{E}(u, v)$;
5. If $u \in \mathcal{D}$ then $u_* = (u \wedge 1) \vee 0 \in \mathcal{D}$, moreover $\mathcal{E}(u_*, u_*) \leq \mathcal{E}(u, u)$.

Relate to Markov process if (quasi)-regular.

Dirichlet forms and MCMC 1

Definition of Dirichlet form

A (symmetric) Dirichlet form $\mathcal{E}$ on a Hilbert space H is a closed bilinear function $\mathcal{E}(u, v)$, defined / finite for any $u, v \in \mathcal{D} \subseteq H$, which satisfies:

1. $\mathcal{D}$ is a dense linear subspace of H;
2. $\mathcal{E}(u, v) = \mathcal{E}(v, u)$ for $u, v \in \mathcal{D}$, so $\mathcal{E}$ is symmetric;
3. $\mathcal{E}(u) = \mathcal{E}(u, u) \geq 0$ for $u \in \mathcal{D}$;
4. $\mathcal{D}$ is a Hilbert space under the ("Sobolev") inner product $\langle u, v \rangle + \mathcal{E}(u, v)$;
5. If $u \in \mathcal{D}$ then $u_* = (u \wedge 1) \vee 0 \in \mathcal{D}$, moreover $\mathcal{E}(u_*, u_*) \leq \mathcal{E}(u, u)$.

Relate to Markov process if (quasi)-regular.
Regular Dirichlet form for locally compact Polish E:
$\mathcal{D} \cap C_0(E)$ is $\mathcal{E}^{\frac{1}{2}}$-dense in $\mathcal{D}$, uniformly dense in $C_0(E)$.

Dirichlet forms and MCMC 2

Two examples

1. Dirichlet form obtained from (re-scaled) RW MH-MCMC:

$$\mathcal{E}_d(h) \quad = \quad \frac{d}{2}\,\mathbb{E}\left[\left(h(\underline{X}_1^{(d)}) - h(\underline{X}_0^{(d)})\right)^2\right].$$

($\mathcal{E}_d$ can be viewed as the Dirichlet form arising from speeding up the RW MH-MCMC by rate d.)

Dirichlet forms and MCMC 2

Two examples

1. Dirichlet form obtained from (re-scaled) RW MH-MCMC:

$$\mathcal{E}_d(h) \quad = \quad \frac{d}{2}\,\mathbb{E}\left[\left(h(\underline{X}_1^{(d)}) - h(\underline{X}_0^{(d)})\right)^2\right].$$

   ($\mathcal{E}_d$ can be viewed as the Dirichlet form arising from speeding up the RW MH-MCMC by rate d.)

2. Heuristic “infinite-dimensional diffusion” limit of this form under scaling:

$$\mathcal{E}_\infty(h) \quad = \quad \frac{s(\sigma)}{2}\,\mathbb{E}_{\pi^{\otimes\infty}}\left[|\nabla h|^2\right].$$

Dirichlet forms and MCMC 2

Two examples

1. Dirichlet form obtained from (re-scaled) RW MH-MCMC:

$$\mathcal{E}_d(h) \quad = \quad \frac{d}{2}\,\mathbb{E}\left[\left(h(\underline{X}_1^{(d)}) - h(\underline{X}_0^{(d)})\right)^2\right].$$

   ($\mathcal{E}_d$ can be viewed as the Dirichlet form arising from speeding up the RW MH-MCMC by rate d.)

2. Heuristic "infinite-dimensional diffusion" limit of this form under scaling:

$$\mathcal{E}_\infty(h) \quad = \quad \frac{s(\sigma)}{2}\,\mathbb{E}_{\pi^{\otimes\infty}}\left[|\nabla h|^2\right].$$

   Under mild conditions this is:
   closable $\surd$, Dirichlet $\surd$, quasi-regular $\surd$.

Dirichlet forms and MCMC 2

Two examples

1. Dirichlet form obtained from (re-scaled) RW MH-MCMC:

$$\mathcal{E}_d(h) \quad = \quad \frac{d}{2}\,\mathbb{E}\left[\left(h(\underline{X}_1^{(d)}) - h(\underline{X}_0^{(d)})\right)^2\right].$$

   ($\mathcal{E}_d$ can be viewed as the Dirichlet form arising from speeding up the RW MH-MCMC by rate d.)

2. Heuristic "infinite-dimensional diffusion" limit of this form under scaling:

$$\mathcal{E}_\infty(h) \quad = \quad \frac{s(\sigma)}{2}\,\mathbb{E}_{\pi^{\otimes\infty}}\left[|\nabla h|^2\right].$$

   Under mild conditions this is:
   closable $\surd$, Dirichlet $\surd$, quasi-regular $\surd$.
   Can we deduce that the RW MH-MCMC scales to look like the "infinite-dimensional diffusion",
   by showing that $\mathcal{E}_d$ "converges" to $\mathcal{E}_\infty$?

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if

   (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   - (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   - (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   - (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   - (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
2. Mosco (1994) introduces stronger conditions;

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   - (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   - (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
2. Mosco (1994) introduces stronger conditions;
   - (M1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h$ **weakly** in H;

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   - (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   - (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
2. Mosco (1994) introduces stronger conditions;
   - (M1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h$ **weakly** in H;
   - (M2) For every $h \in H$ there are $h_n \to h$ **strongly** in H such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   - (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   - (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
2. Mosco (1994) introduces stronger conditions;
   - (M1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h$ **weakly** in H;
   - (M2) For every $h \in H$ there are $h_n \to h$ **strongly** in H such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
3. Mosco (1994, Theorem 2.4.1, Corollary 2.6.1): conditions (M1) and (M2) imply convergence of associated resolvent operators,

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   (Γ1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
2. Mosco (1994) introduces stronger conditions;
   (M1) $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h$ **weakly** in H;
   (M2) For every $h \in H$ there are $h_n \to h$ **strongly** in H such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.
3. Mosco (1994, Theorem 2.4.1, Corollary 2.6.1): conditions (M1) and (M2) imply convergence of associated resolvent operators,

   and indeed of associated semigroups.

Useful modes of convergence for Dirichlet forms

1. Gamma-convergence; $\mathcal{E}_n$ "Γ-converges" to $\mathcal{E}_\infty$ if
   - (Γ1) $\mathcal{E}_\infty(h) \le \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h \in H$;
   - (Γ2) For every $h \in H$ there are $h_n \to h \in H$ such that $\mathcal{E}_\infty(h) \ge \limsup_n \mathcal{E}_n(h_n)$.
2. Mosco (1994) introduces stronger conditions;
   - (M1) $\mathcal{E}_\infty(h) \le \liminf_n \mathcal{E}_n(h_n)$ whenever $h_n \to h$ **weakly** in H;
   - (M2) For every $h \in H$ there are $h_n \to h$ **strongly** in H such that $\mathcal{E}_\infty(h) \ge \limsup_n \mathcal{E}_n(h_n)$.
3. Mosco (1994, Theorem 2.4.1, Corollary 2.6.1): conditions (M1) and (M2) imply convergence of associated resolvent operators,

   and indeed of associated semigroups.
4. Sun (1998) gives further conditions which imply weak convergence of the associated processes: these conditions are implied by existence of a finite constant C such that $\mathcal{E}_n(h) \le C(|h|^2 + \mathcal{E}(h))$ for all $h \in H$.

Results and methods of proofs

Results

Theorem (Zanella, Bédard and WSK, 2016)

Consider the Gaussian RW MH-MCMC based on proposal variance σ^2/d with target $\pi^{\otimes d}$, where $\mathrm{d}\pi = f\,\mathrm{d}x = e^{-\phi}\,\mathrm{d}x$. Suppose $\mathcal{I} = \int_\infty^\infty |\phi'|^2 f\,\mathrm{d}x < \infty$ (finite Fisher information), and $|\phi'(x+v) - \phi'(x)| < \kappa \max\{|v|^\gamma, |v|^\alpha\}$ for some $\kappa > 0$, $0 < \gamma < 1$, and $\alpha > 1$.

Let $\mathcal{E}_d$ be the corresponding Dirichlet form scaled as above. $\mathcal{E}_d$ Mosco-converges to $\mathbb{E}\left[1 \wedge \exp(\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I}))\right]\mathcal{E}_\infty$, so corresponding L^2 semigroups also converge.

Results

Theorem (Zanella, Bédard and WSK, 2016)

Consider the Gaussian RW MH-MCMC based on proposal variance σ^2/d with target $\pi^{\otimes d}$, where $\mathrm{d}\pi = f\,\mathrm{d}x = e^{-\phi}\,\mathrm{d}x$. Suppose $\mathcal{I} = \int_{\infty}^{\infty} |\phi'|^2 f\,\mathrm{d}x < \infty$ (finite Fisher information), and $|\phi'(x+v) - \phi'(x)| < \kappa \max\{|v|^{\gamma}, |v|^{\alpha}\}$ for some $\kappa > 0$, $0 < \gamma < 1$, and $\alpha > 1$.

Let $\mathcal{E}_d$ be the corresponding Dirichlet form scaled as above. $\mathcal{E}_d$ Mosco-converges to $\mathbb{E}\left[1 \wedge \exp(\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I}))\right]\mathcal{E}_\infty$, so corresponding L^2 semigroups also converge.

Corollary

Suppose in the above that ϕ' is globally Lipschitz. The correspondingly scaled processes exhibit weak convergence.

Methods of proof 1: a CLT result

Lemma (A conditional CLT)

Under the conditions of the Corollary, almost surely (in $\underline{x}$ with invariant measure $\pi^{\otimes\infty}$) the log Metropolis-Hastings ratio converges weakly (in $\underline{W}$) as follows as $d \to \infty$:

$$\log\left(\prod_{i=1}^{d} \frac{f(x_i + \frac{\sigma W_i}{\sqrt{d}})}{f(x_i)}\right) = \sum_{i=1}^{d}\left(\phi(x_i + \frac{\sigma W_i}{\sqrt{d}}) - \phi(x_i)\right) \Rightarrow \mathcal{N}(-\tfrac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I}).$$

Warwick Statistics

Methods of proof 1: a CLT result

Lemma (A conditional CLT)

Under the conditions of the Corollary, almost surely (in $\underline{x}$ with invariant measure $\pi^{\otimes\infty}$) the log Metropolis-Hastings ratio converges weakly (in $\underline{W}$) as follows as $d \to \infty$:

$$\log\left(\prod_{i=1}^{d} \frac{f(x_i + \frac{\sigma W_i}{\sqrt{d}})}{f(x_i)}\right) = \sum_{i=1}^{d}\left(\phi(x_i + \frac{\sigma W_i}{\sqrt{d}}) - \phi(x_i)\right) \Rightarrow \mathcal{N}(-\tfrac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I}).$$

We may use this to deduce the asymptotic acceptance rate of the RW MH-MCMC sampler.

Warwick Statistics

Key idea for CLT

Use exact Taylor expansion techniques:

$$\sum_{i=1}^{d} \left(\phi(x_i + \tfrac{\sigma W_i}{\sqrt{d}}) - \phi(x_i) \right) \quad =$$

$$\sum_{i=1}^{d} \phi'(x_i) \tfrac{\sigma W_i}{\sqrt{d}} + \sum_{i=1}^{d} \tfrac{\sigma W_i}{\sqrt{d}} \int_0^1 \left(\phi'(x_i + \tfrac{\sigma W_i}{\sqrt{d}} u) - \phi'(x_i) \right) \mathrm{d}u \,.$$

Key idea for CLT

Use exact Taylor expansion techniques:

$$\sum_{i=1}^{d}\left(\phi(x_i+\tfrac{\sigma W_i}{\sqrt{d}})-\phi(x_i)\right) \quad =$$

$$\sum_{i=1}^{d}\phi'(x_i)\frac{\sigma W_i}{\sqrt{d}}+\sum_{i=1}^{d}\frac{\sigma W_i}{\sqrt{d}}\int_0^1\left(\phi'(x_i+\tfrac{\sigma W_i}{\sqrt{d}}u)-\phi'(x_i)\right)\,\mathrm{d}u\,.$$

Condition implicitly on $\underline{x}$ for first 2.5 steps.

1. First summand converges weakly to $\mathcal{N}(0,\sigma^2\mathcal{I})$.

Key idea for CLT

Use exact Taylor expansion techniques:

$$\sum_{i=1}^{d} \left(\phi(x_i + \tfrac{\sigma W_i}{\sqrt{d}}) - \phi(x_i)\right) =$$

$$\sum_{i=1}^{d} \phi'(x_i)\tfrac{\sigma W_i}{\sqrt{d}} + \sum_{i=1}^{d} \tfrac{\sigma W_i}{\sqrt{d}} \int_0^1 \left(\phi'(x_i + \tfrac{\sigma W_i}{\sqrt{d}} u) - \phi'(x_i)\right) \mathrm{d}u\,.$$

Condition implicitly on $\underline{x}$ for first 2.5 steps.

1. First summand converges weakly to $\mathcal{N}(0, \sigma^2 \mathcal{I})$.
2. Decompose variance of second summand to deduce $\operatorname{Var}\left[\sum_{i=1}^{d} \tfrac{\sigma W_i}{\sqrt{d}} \int_0^1 \left(\phi'(x_i + \tfrac{\sigma W_i}{\sqrt{d}} u - \phi'(x_i)\right) \mathrm{d}u\right] \to 0.$

Warwick Statistics

Key idea for CLT

Use exact Taylor expansion techniques:

$$\sum_{i=1}^{d} \left(\phi(x_i + \tfrac{\sigma W_i}{\sqrt{d}}) - \phi(x_i) \right) \quad =$$

$$\sum_{i=1}^{d} \phi'(x_i) \frac{\sigma W_i}{\sqrt{d}} + \sum_{i=1}^{d} \frac{\sigma W_i}{\sqrt{d}} \int_0^1 \left(\phi'(x_i + \tfrac{\sigma W_i}{\sqrt{d}} u) - \phi'(x_i) \right) \mathrm{d}u \,.$$

Condition implicitly on $\underline{x}$ for first 2.5 steps.

1. First summand converges weakly to $\mathcal{N}(0, \sigma^2 \mathcal{I})$.
2. Decompose variance of second summand to deduce $\mathrm{Var}\left[\sum_{i=1}^{d} \frac{\sigma W_i}{\sqrt{d}} \int_0^1 \left(\phi'(x_i + \frac{\sigma W_i}{\sqrt{d}} u - \phi'(x_i) \right) \mathrm{d}u \right] \to 0$.
3. Use Hoeffding's inequality then absolute expectations: $\mathbb{E}\left[\sum_{i=1}^{d} \frac{\sigma W_i}{\sqrt{d}} \int_0^1 \left(\phi'(x_i + \frac{\sigma W_i}{\sqrt{d}} u - \phi'(x_i) \right) \mathrm{d}u \right] \to -\frac{1}{2}\sigma^2 \mathcal{I}$.

Methods of proof 2: establishing condition (M2)

For every $h \in L^2(\pi^{\otimes\infty})$, find $h_n \to h$ (strongly) in $L^2(\pi^{\otimes\infty})$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.

1. Sufficient to consider case $\mathcal{E}_\infty(h) < \infty$.

Methods of proof 2: establishing condition (M2)

For every $h \in L^2(\pi^{\otimes\infty})$, find $h_n \to h$ (strongly) in $L^2(\pi^{\otimes\infty})$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.

1. Sufficient to consider case $\mathcal{E}_\infty(h) < \infty$.
2. Find sequence of smooth cylinder functions h_n with "compact cylindrical support", such that $|\mathcal{E}_\infty(h) - \mathcal{E}_\infty(h_n)| \leq 1/n$.

Methods of proof 2: establishing condition (M2)

For every $h \in L^2(\pi^{\otimes\infty})$, find $h_n \to h$ (strongly) in $L^2(\pi^{\otimes\infty})$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.

1. Sufficient to consider case $\mathcal{E}_\infty(h) < \infty$.
2. Find sequence of smooth cylinder functions h_n with "compact cylindrical support", such that $|\mathcal{E}_\infty(h) - \mathcal{E}_\infty(h_n)| \leq 1/n$.
3. Using smoothness *etc*, $\mathcal{E}_m(h_n) \to \mathcal{E}_\infty(h_n)$ as $m \to \infty$.

Methods of proof 2: establishing condition (M2)

For every $h \in L^2(\pi^{\otimes\infty})$, find $h_n \to h$ (strongly) in $L^2(\pi^{\otimes\infty})$ such that $\mathcal{E}_\infty(h) \geq \limsup_n \mathcal{E}_n(h_n)$.

1. Sufficient to consider case $\mathcal{E}_\infty(h) < \infty$.
2. Find sequence of smooth cylinder functions h_n with "compact cylindrical support", such that $|\mathcal{E}_\infty(h) - \mathcal{E}_\infty(h_n)| \leq 1/n$.
3. Using smoothness *etc*, $\mathcal{E}_m(h_n) \to \mathcal{E}_\infty(h_n)$ as $m \to \infty$.
4. Subsequences . . . .

Methods of proof 3: establishing condition (M1)

If $h_n \to h$ weakly in $L^2(\pi^{\otimes\infty})$, show $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$.
Detailed stochastic analysis involves:

1. Set $\Psi_n(h) = \sqrt{\frac{n}{2}}(h(\underline{X}_0^{(n)}) - h(\underline{X}_1^{(n)}))$.

Warwick Statistics

Methods of proof 3: establishing condition (M1)

If $h_n \to h$ weakly in $L^2(\pi^{\otimes\infty})$, show $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$.
Detailed stochastic analysis involves:

1. Set $\Psi_n(h) = \sqrt{\frac{n}{2}}(h(\underline{X}_0^{(n)}) - h(\underline{X}_1^{(n)}))$.
2. Integrate against test function $\xi(\underline{X}_{1:N}, \underline{W}_{1:N})\mathbb{I}(U < a(\underline{X}_{1:N}, \underline{W}_{1:N}))$ for ξ smooth, compact support, U a Uniform$(0,1)$ random variable. Apply Cauchy-Schwarz.

Methods of proof 3: establishing condition (M1)

If $h_n \to h$ weakly in $L^2(\pi^{\otimes\infty})$, show $\mathcal{E}_\infty(h) \leq \liminf_n \mathcal{E}_n(h_n)$.
Detailed stochastic analysis involves:

1. Set $\Psi_n(h) = \sqrt{\frac{n}{2}}(h(\underline{X}_0^{(n)}) - h(\underline{X}_1^{(n)}))$.
2. Integrate against test function $\xi(\underline{X}_{1:N}, \underline{W}_{1:N})\mathbb{I}(U < a(\underline{X}_{1:N}, \underline{W}_{1:N}))$ for ξ smooth, compact support, U a Uniform$(0,1)$ random variable. Apply Cauchy-Schwarz.
3. Use integration by parts, careful analysis and conditions on ϕ'.

Doing even better

Durmus et al. (2016) introduce L^p mean differentiability:

Doing even better

Durmus et al. (2016) introduce L^p mean differentiability: there is $\dot{\phi}$ such that, for some $p > 2$, some $\alpha > 0$,

$$\begin{aligned} \phi(X+u) - \phi(X) &= \left(\dot{\phi}(X) + R(X,u)\right) u\,, \\ \mathbb{E}\left[|R(X,u)|^p\right]^{1/p} &= o(|u|^\alpha)\,. \end{aligned}$$

Doing even better

Durmus et al. (2016) introduce L^p mean differentiability: there is $\dot{\phi}$ such that, for some $p > 2$, some $\alpha > 0$,

$$\begin{aligned} \phi(X+u) - \phi(X) &= (\dot{\phi}(X) + R(X,u))\, u\,, \\ \mathbb{E}\left[|R(X,u)|^p\right]^{1/p} &= o(|u|^\alpha)\,. \end{aligned}$$

Also $\mathcal{I} = \mathbb{E}\left[|\dot{\phi}|^2\right] < \infty$.

Warwick Statistics

Doing even better

Durmus et al. (2016) introduce L^p mean differentiability: there is $\dot{\phi}$ such that, for some $p > 2$, some $\alpha > 0$,

$$\begin{aligned} \phi(X+u) - \phi(X) &= \left(\dot{\phi}(X) + R(X,u)\right) u\,, \\ \mathbb{E}\left[|R(X,u)|^p\right]^{1/p} &= o(|u|^\alpha)\,. \end{aligned}$$

Also $\mathcal{I} = \mathbb{E}\left[|\dot{\phi}|^2\right] < \infty$.

Durmus et al. (2016) obtain optimal scaling results when $p > 4$, and $\mathbb{E}\left[|\dot{\phi}|^6\right] < \infty$,

Doing even better

Durmus et al. (2016) introduce L^p mean differentiability: there is $\dot{\phi}$ such that, for some $p > 2$, some $\alpha > 0$,

$$
\begin{aligned}
\phi(X+u) - \phi(X) &= \left(\dot{\phi}(X) + R(X,u)\right) u\,, \\
\mathbb{E}\left[|R(X,u)|^p\right]^{1/p} &= o(|u|^\alpha)\,.
\end{aligned}
$$

Also $\mathcal{I} = \mathbb{E}\left[|\dot{\phi}|^2\right] < \infty$.

Durmus et al. (2016) obtain optimal scaling results when $p > 4$, and $\mathbb{E}\left[|\dot{\phi}|^6\right] < \infty$,

L^p mean differentiability applies straightforwardly to the Zanella, Bédard and WSK (2016) argument *mutatis mutandis*: the regularity conditions can be weakened even more at least for vague convergence.

Warwick Statistics

Conclusion

Warwick Statistics

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;
- Combine with L^p mean differentiability to obtain further relaxation of regularity conditions;

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;
- Combine with L^p mean differentiability to obtain further relaxation of regularity conditions;
- *Soft* argument for $\frac{1}{2}$variance $+$ mean ≈ 0 implied by $\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I})$;

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;
- Combine with L^p mean differentiability to obtain further relaxation of regularity conditions;
- *Soft* argument for $\frac{1}{2}$variance + mean ≈ 0 implied by $\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I})$;
- MALA generalization (exercise in progress);

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;
- Combine with L^p mean differentiability to obtain further relaxation of regularity conditions;
- *Soft* argument for $\frac{1}{2}$variance + mean ≈ 0 implied by $\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I})$;
- MALA generalization (exercise in progress);
- Need to explore development beyond *i.i.d.* targets; *e.g.* can regularity be similarly relaxed in more general random field settings?

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;
- Combine with L^p mean differentiability to obtain further relaxation of regularity conditions;
- *Soft* argument for $\frac{1}{2}$variance + mean ≈ 0 implied by $\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I})$;
- MALA generalization (exercise in progress);
- Need to explore development beyond *i.i.d.* targets; *e.g.* can regularity be similarly relaxed in more general random field settings?
- Apply to discrete Markov chain cases? (*c.f.* Roberts, 1998);

Conclusion

- The Dirichlet form approach allows significant relaxation of conditions required for optimal scaling results;
- Combine with L^p mean differentiability to obtain further relaxation of regularity conditions;
- *Soft* argument for $\frac{1}{2}$variance + mean ≈ 0 implied by $\mathcal{N}(-\frac{1}{2}\sigma^2\mathcal{I}, \sigma^2\mathcal{I})$;
- MALA generalization (exercise in progress);
- Need to explore development beyond *i.i.d.* targets; *e.g.* can regularity be similarly relaxed in more general random field settings?
- Apply to discrete Markov chain cases? (*c.f.* Roberts, 1998);
- Investigate applications to Adaptive MCMC.

References:

- Andrieu, Christophe and Johannes Thoms (2008). "A tutorial on adaptive MCMC". In: *Statistics and Computing* 18.4, pp. 343–373.
- Bédard, Mylène (2007). "Weak convergence of Metropolis algorithms for non-I.I.D. target distributions". In: *Annals of Applied Probability* 17.4, pp. 1222–1244.
- Breyer, L A and Gareth O Roberts (2000). "From Metropolis to diffusions : Gibbs states and optimal scaling". In: *Stochastic Processes and their Applications* 90.2, pp. 181–206.
- Brooks, Stephen P, Andrew Gelman, Galin L Jones and Xiao-Li Meng (2011). *Handbook of Markov Chain Monte Carlo*. Boca Raton: Chapman & Hall/CRC, pp. 592+xxv.

Durmus, Alain, Sylvain Le Corff, Eric Moulines and Gareth O Roberts (2016). "Optimal scaling of the Random Walk Metropolis algorithm under L^p mean differentiability". In: *arXiv* 1604.06664.

Geyer, Charlie (1999). "Likelihood inference for spatial point processes". In: *Stochastic Geometry: likelihood and computation*. Ed. by Ole E Barndorff-Nielsen, WSK and M N M van Lieshout. Boca Raton: Chapman & Hall/CRC. Chap. 4, pp. 79–140.

Hastings, W K (1970). "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57, pp. 97–109.

Mattingly, Jonathan C., Natesh S. Pillai and Andrew M. Stuart (2012). "Diffusion limits of the random walk metropolis algorithm in high dimensions". In: *Annals of Applied Probability* 22.3, pp. 881–890.

- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller and Edward Teller (1953). "Equation of state calculations by fast computing machines". en. In: *Journal Chemical Physics* 21.6, pp. 1087–1092.
- Mosco, Umberto (1994). "Composite media and asymptotic Dirichlet forms". In: *Journal of Functional Analysis* 123.2, pp. 368–421.
- Roberts, Gareth O (1998). "Optimal Metropolis algorithms for product measures on the vertices of a hypercube". In: *Stochastics and Stochastic Reports* June 2013, pp. 37–41.
- Roberts, Gareth O, A Gelman and W Gilks (1997). "Weak Convergence and Optimal Scaling of Random Walk Algorithms". In: *The Annals of Applied Probability* 7.1, pp. 110–120.

Warwick Statistics

Roberts, Gareth O and Jeffrey S Rosenthal (1998). "Optimal scaling of discrete approximations to Langevin diffusions.". In: *J. R. Statist. Soc. B* 60.1, pp. 255–268.

Rosenthal, Jeffrey S (2011). "Optimal Proposal Distributions and Adaptive MCMC". In: *Handbook of Markov Chain Monte Carlo* 1, pp. 93–112.

Sun, Wei (1998). "Weak convergence of Dirichlet processes". In: *Science in China Series A: Mathematics* 41.1, pp. 8–21.

Thompson, Elizabeth A (2005). "MCMC in the analysis of genetic data on pedigree". In: *Markov chain Monte Carlo: Innovations and Applications*. Ed. by WSK, Faming Liang and Jian-Sheng Wang. Singapore: World Scientific. Chap. 5, pp. 183–216.

Zanella, Giacomo (2015). "Bayesian Complementary Clustering, MCMC, and Anglo-Saxon Placenames". PhD Thesis. University of Warwick.

Zanella, Giacomo (2016). “Random Partition Models and Complementary Clustering of Anglo-Saxon Placenames”. In: *Annals of Applied Statistics* 9.4, pp. 1792–1822.

Zanella, Giacomo, Mylène Bédard and WSK (2016). “A Dirichlet Form approach to MCMC Optimal Scaling”. In: *arXiv* 1606.01528, 22pp.

```
Version 1.34    (Wed Jul 12 06:27:58 2017 +0100)
================================================
commit d81469c5f14c484aa363616e3d701c6e3fbb1141
Author: Wilfrid Kendall <W.S.Kendall@warwick.ac.uk>

    Made it explicit that Lp mean differentiability still doesn't cover weak
    without extra regularity: need to beat this!
```

Warwick Statistics