Large deviations for MCMC:
The surprisingly curious case of the Metropolis-Hastings algorithm

Pierre Nyquist

Department of Mathematical Sciences
Chalmers & University of Gothenburg

Algorithms seminar
Warwick, February 23, 2024

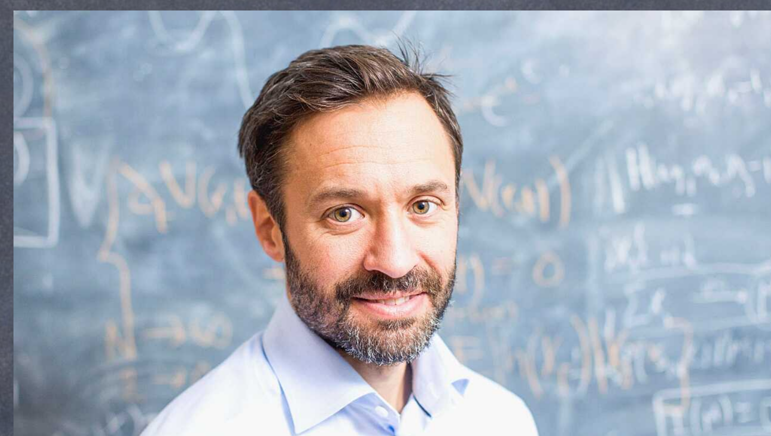
joint work with Federica Milinanni (+ others)

@PierreNyq
https://people.kth.se/~pierren/

Milinanni, N. - A large deviation principle for the empirical measures of Metropolis-Hastings chains.
Stochastic Process and their Applications, 170 (2024).

Milinanni, N. - Large deviations for certain Metropolis-Hastings chains: Existence of suitable Lyapunov functions *
Preprint, arXiv next week.

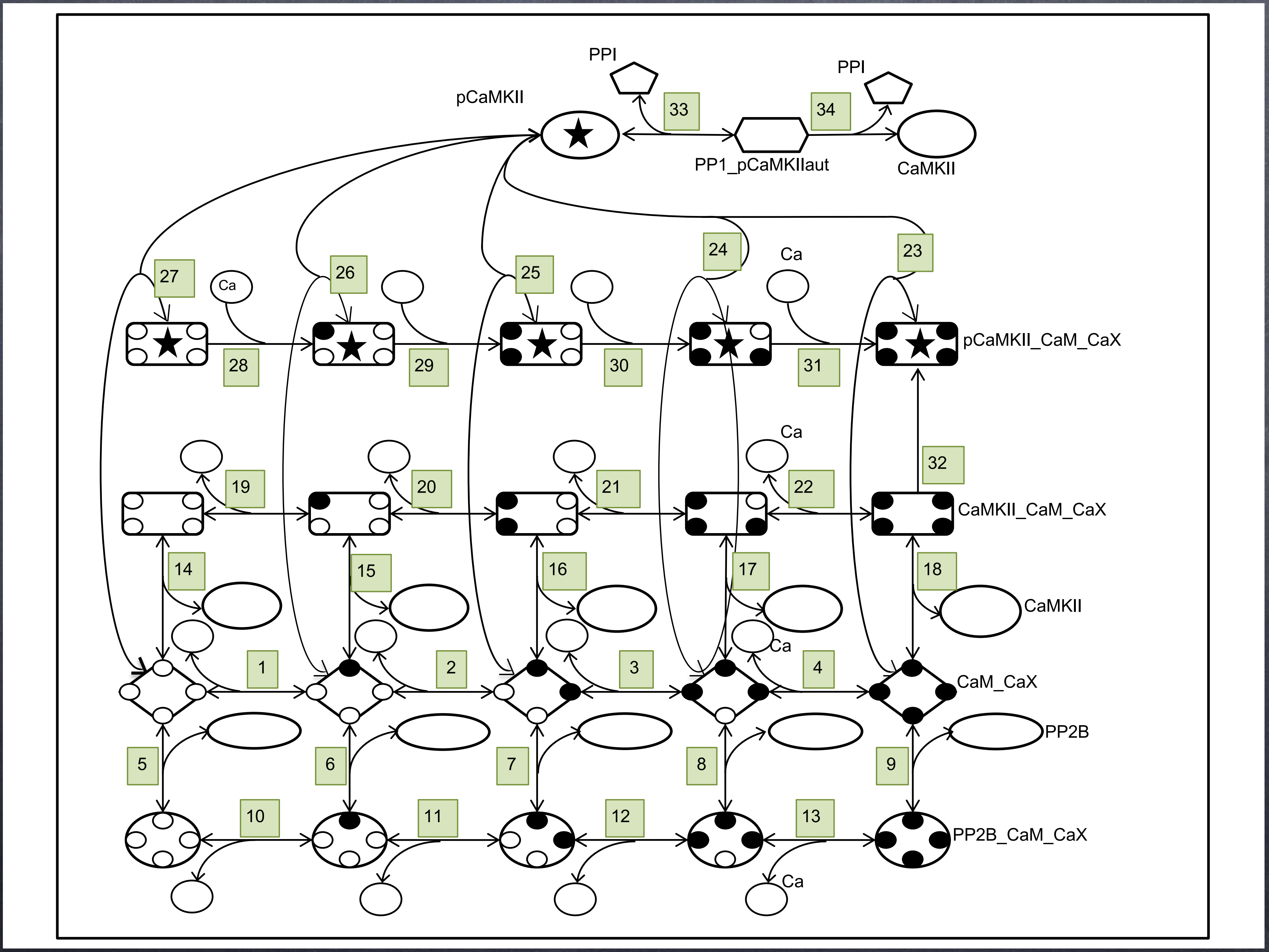
* Prelim title.




I. Introduction

Starting point: Subcellular pathway models in neuroscience



Main question: How to sample from a distribution π know only up to a normalising constant?

$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R}\,.$$

Example I: Bayesian inference. Posterior distributions on the form

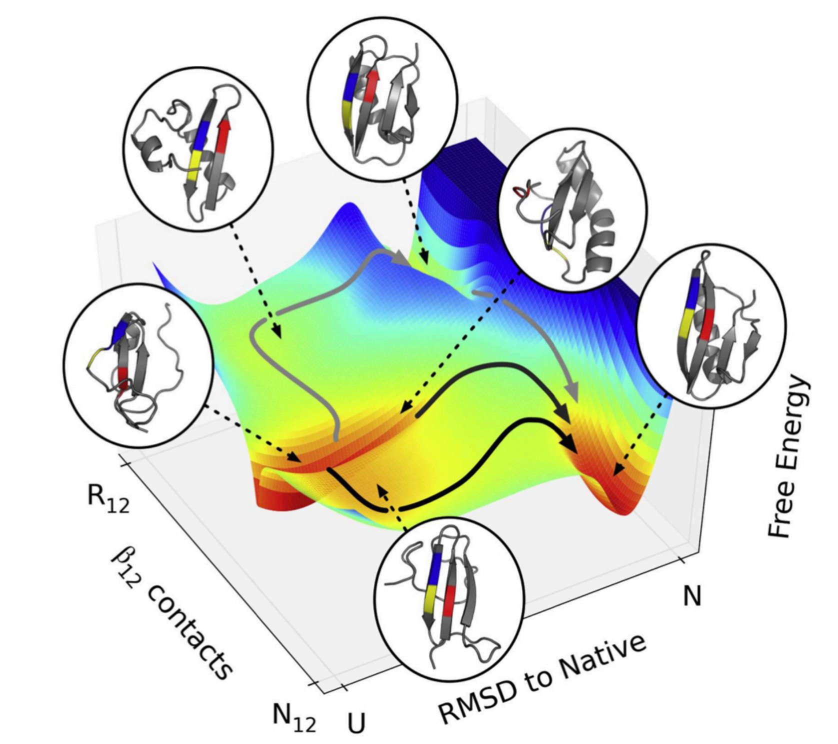
$$\pi(\xi) \propto \pi_0(\xi) L(\mathbf{x}_{1:n} \,|\, \xi),$$

with unknown normalising constant $Z = \int \pi_0(\xi) L(\mathbf{x}_{1:n} \,|\, \xi) d\xi$.

Main question: How to sample from a distribution π know only up to a normalising constant?

$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R} .$$

Example II: Computational chemistry. Compute thermodynamic properties with respect to the Gibbs measure $\propto e^{-U}$.



Source: Schwantes, Shukla, Pande Biophysical Journal, 2016.

Main question: How to sample from a distribution π know only up to a normalising constant?

$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R}\,.$$

Example III: Counting problems. Determine the number of objects in a large (finite) class that satisfy certain constraints.

Ex: Number of binary contingency tables with row and column sums $\mathbf{r} = (r_1, \ldots, r_m)$ and $\mathbf{c} = (c_1, \ldots, c_n)$ . .

$$\left|\mathscr{X}*\right| = \left|\left\{\mathbf{x} \in \{0,1\}^{m+n} : \sum_{i=1}^{m} x_{i,j} = c_j, j = 1,\ldots,n, \ \sum_{j=1}^{n} x_{i,j} = r_i, i = 1,\ldots,m\right\}\right|.$$

Main question: How to sample from a distribution π know only up to a normalising constant?

$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R}\,.$$

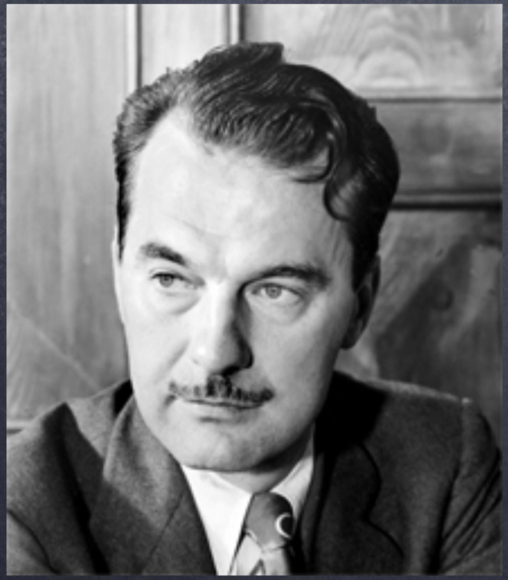
Idea: Construct a Markov process with π as invariant measure.

<u>Main question:</u> How to sample from a distribution π know only up to a normalising constant?

$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R}\,.$$

<u>Idea:</u> Construct a Markov process with π as invariant measure.

(Metropolis et al. 1953, Hastings 1970.)

Main question: How to sample from a distribution π know only up to a normalising constant?

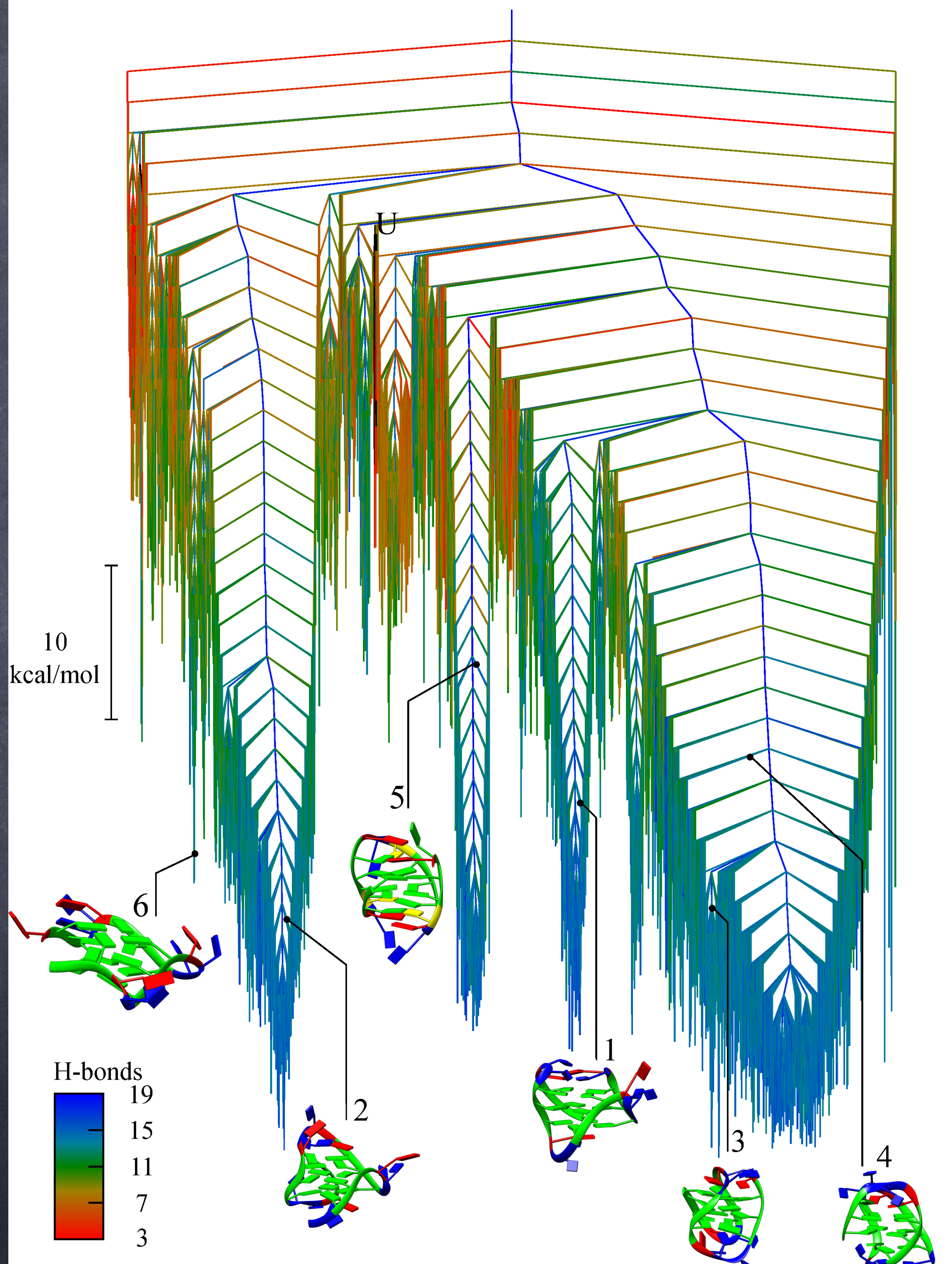
$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R}\,.$$

Idea: Construct a Markov process with π as invariant measure.

(Metropolis et al. 1953, Hastings 1970.)

Infinitely many possibilities. How to choose?

Main hindrance: Poor communication / complex energy landscape.


U
10
kcal/mol
5
6
2
1
3
4
H-bonds
19
15
11
7
3

Main question: How to sample from a distribution π know only up to a normalising constant?

$$\pi(y) \propto \exp\{-U(y)\}, \quad U : S \to \mathbb{R}\,.$$

Idea: Construct a Markov process with π as invariant measure.

(Metropolis et al. 1953, Hastings 1970.)

Infinitely many possibilities. How to choose?

Main hindrance: Poor communication / complex energy landscape.

Q: How to analyse the efficiency of MCMC methods?

Performance analysis of MCMC methods:

Q: How to analyse the efficiency of MCMC methods?

Performance analysis of MCMC methods:

Q: How to analyse the efficiency of MCMC methods?

Step 1: Ergodicity of the underlying process $X = \{X(t)\}_{t \geq 0}$.

Beyond ergodicity, tools include empirical observations, spectral properties ("2nd eigenvalue information"), asymptotic variance and functional inequalities (Poincaré, log-Sobolev).

Performance analysis of MCMC methods:

Q: How to analyse the efficiency of MCMC methods?

Step 1: Ergodicity of the underlying process $X = \{X(t)\}_{t\geq 0}$.

Beyond ergodicity, tools include empirical observations, spectral properties ("2nd eigenvalue information"), asymptotic variance and functional inequalities (Poincaré, log-Sobolev).

In practice: approximation of π built on the empirical measure

$$\eta_T = \frac{1}{T}\int_0^T \delta_{X(t)}(\,\cdot\,)dt\,.$$

Under ergodicity $\eta_T \to \pi$.

Performance analysis of MCMC methods:

Q: How to analyse the efficiency of MCMC methods?

Step 1: Ergodicity of the underlying process $X = \{X(t)\}_{t \geq 0}$.

Beyond ergodicity, tools include empirical observations, spectral properties ("2nd eigenvalue information"), asymptotic variance and functional inequalities (Poincaré, log-Sobolev).

In practice: approximation of π built on the empirical measure

$$\eta_T = \frac{1}{T} \int_0^T \delta_{X(t)}(\,\cdot\,)dt\,.$$

Under ergodicity $\eta_T \to \pi$.

Empirical measure large deviations: Relates directly to the behaviour of η_T as $T \to \infty$. So far (severely) underutilised.

II. Primer on Large deviations

Large deviation principle:

A sequence $\{X_n\}_n$ of random elements satisfy the large deviation principle (LDP), with rate function $I : \mathscr{X} \to [0,\infty]$, and speed n if

$$
\begin{aligned}
-\inf_{x \in G^\circ} I(x) &\leq \liminf_n \frac{1}{n} \log P(X_n \in G^\circ) \\
&\leq \limsup_n \frac{1}{n} \log P(X_n \in \bar{G}) \leq -\inf_{x \in \bar{G}} I(x) \,.
\end{aligned}
$$

Large deviation principle:

A sequence $\{X_n\}_n$ of random elements satisfy the **large deviation principle** (LDP), with **rate function** $I : \mathscr{X} \to [0,\infty]$, and speed n if

$$-\inf_{x\in G^\circ} I(x) \leq \liminf_n \frac{1}{n} \log P(X_n \in G^\circ)$$
$$\leq \limsup_n \frac{1}{n} \log P(X_n \in \bar{G}) \leq -\inf_{x\in\bar{G}} I(x)\,.$$

Gist: For measurable $G \subset \mathscr{X}$,

$$P(X_n \in G) \approx \mathbf{exp}\left\{ -n \inf_{x\in G} I(x) \right\}.$$

Large deviation principle:

A sequence $\{X_n\}_n$ of random elements satisfy the large deviation principle (LDP), with rate function $I : \mathscr{X} \to [0,\infty]$, and speed n if

$$
\begin{aligned}
-\inf_{x\in G^\circ} I(x) &\le \liminf_n \frac{1}{n}\log P(X_n \in G^\circ) \\
&\le \limsup_n \frac{1}{n}\log P(X_n \in \bar{G}) \le -\inf_{x\in\bar{G}} I(x)\,.
\end{aligned}
$$

Gist: For measurable $G \subset \mathscr{X}$,

$$P(X_n \in G) \approx \mathbf{exp}\left\{-n\inf_{x\in G} I(x)\right\}.$$

Minimisers of I characterise how events occur.

Large deviation principle:

A sequence $\{X_n\}_n$ of random elements satisfy the large deviation principle (LDP), with rate function $I : \mathscr{X} \to [0,\infty]$, and speed n if

$$- \inf_{x\in G^\circ} I(x) \le \liminf_n \frac{1}{n} \log P(X_n \in G^\circ)$$

$$\le \limsup_n \frac{1}{n} \log P(X_n \in \bar{G}) \le - \inf_{x\in \bar{G}} I(x) .$$

Gist: For measurable $G \subset \mathscr{X}$,

$$P(X_n \in G) \approx \exp\left\{ -n \inf_{x\in G} I(x) \right\} .$$

Minimisers of I characterise how events occur.

Example: Schilders theorem

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Example: Schilders theorem

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2 \ : ||x||^2<1\}$ in $[0,T]$?

Example: Schilders theorem

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2 \ : ||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D?

Example: Schilders theorem

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2 \ :||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D?

Ans: $\{X^\epsilon\}_{\epsilon>0}$ satisfies LDP with rate function

$$I(\varphi)=\frac{1}{2}\int_0^T ||\dot{\varphi}(s)||^2 ds;\quad \varphi\in AC([0,T]:\mathbb{R}^2),\ \varphi(0)=0.$$

Roughly:

$$P(X^\epsilon \text{ leaves } D)\approx \exp\left\{-\frac{1}{\epsilon}\inf_{\varphi}\left\{I(\varphi):\ \varphi(0)=0,\ \exists\tau\in[0,T] \text{ s.t. } \varphi(\tau)\in\partial D\right\}\right\}$$

Example: Schilders theorem

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2\ :\ ||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D?

Ans: $\{X^\epsilon\}_{\epsilon>0}$ satisfies LDP with rate function

$$I(\varphi)=\frac{1}{2}\int_0^T ||\dot\varphi(s)||^2\,ds;\quad \varphi\in AC([0,T]:\mathbb{R}^2),\ \varphi(0)=0.$$

Roughly:

$$P(X^\epsilon \text{ leaves } D)\approx \exp\left\{-\frac{1}{\epsilon}\inf_{\varphi}\{I(\varphi):\ \ \varphi(0)=0,\ \exists\tau\in[0,T] \text{ s.t. } \varphi(\tau)\in\partial D\}\right\}$$

Example: Schilders theorem

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2 \ : ||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D?

Ans: $\{X^\epsilon\}_{\epsilon>0}$ satisfies LDP with rate function

$$I(\varphi)=\frac{1}{2}\int_0^T ||\dot{\varphi}(s)||^2 ds; \quad \varphi\in AC([0,T]:\mathbb{R}^2),\ \varphi(0)=0.$$

Roughly:

$$P(X^\epsilon \text{ leaves } D)\approx \exp\left\{-\frac{1}{\epsilon}\inf_{\varphi}\{I(\varphi):\ \varphi(0)=0,\ \exists\tau\in[0,T] \text{ s.t. } \varphi(\tau)\in\partial D\}\right\}$$

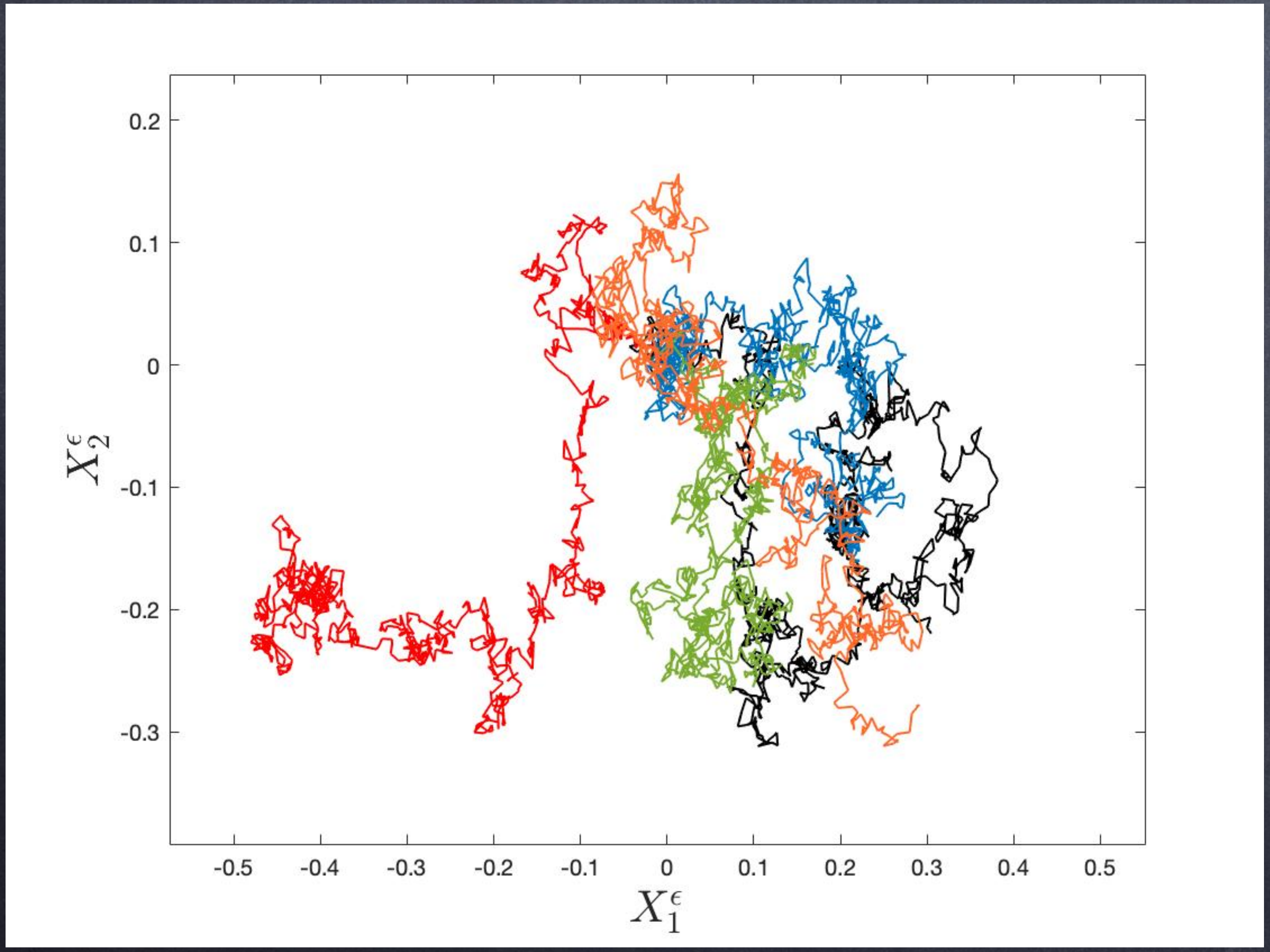
Solution $\varphi(s)=(C_1 s, C_2 s)$ where $C_1^2+C_2^2=1/T^2$. Linear towrds ∂D, reach at T.

Example: Schilders theorem (cont'd)

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2\ :\ ||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D? LDP: Linear towards ∂D, reach at T.



5 of 100K trajectories

$\epsilon=0.044$

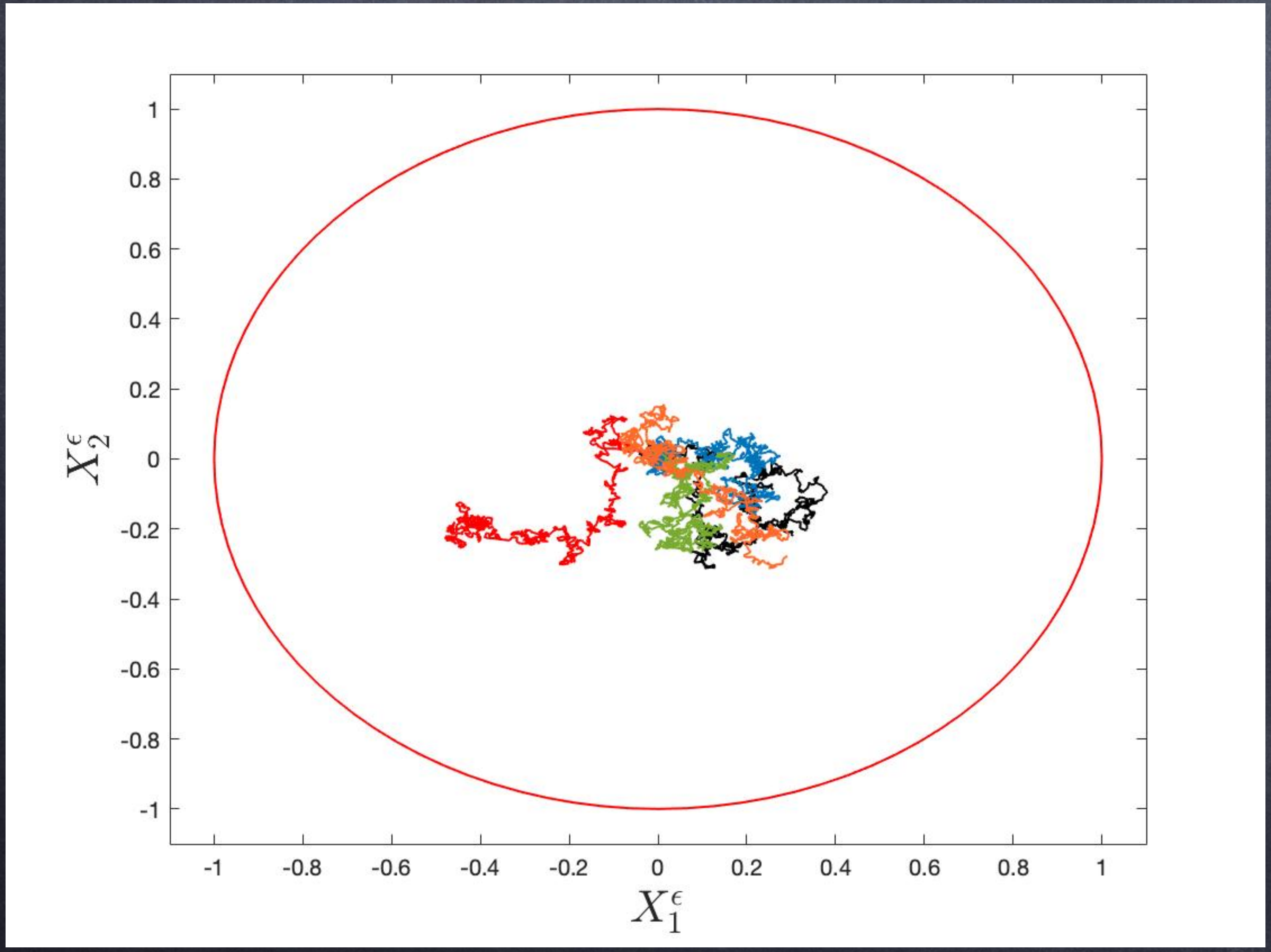
Probability $\approx 10^{-5}$

Example: Schilders theorem (cont'd)

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$,
$X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2 \;:\; ||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D? LDP: Linear towards ∂D, reach at T.



5 of 100K trajectories

$\epsilon=0.044$

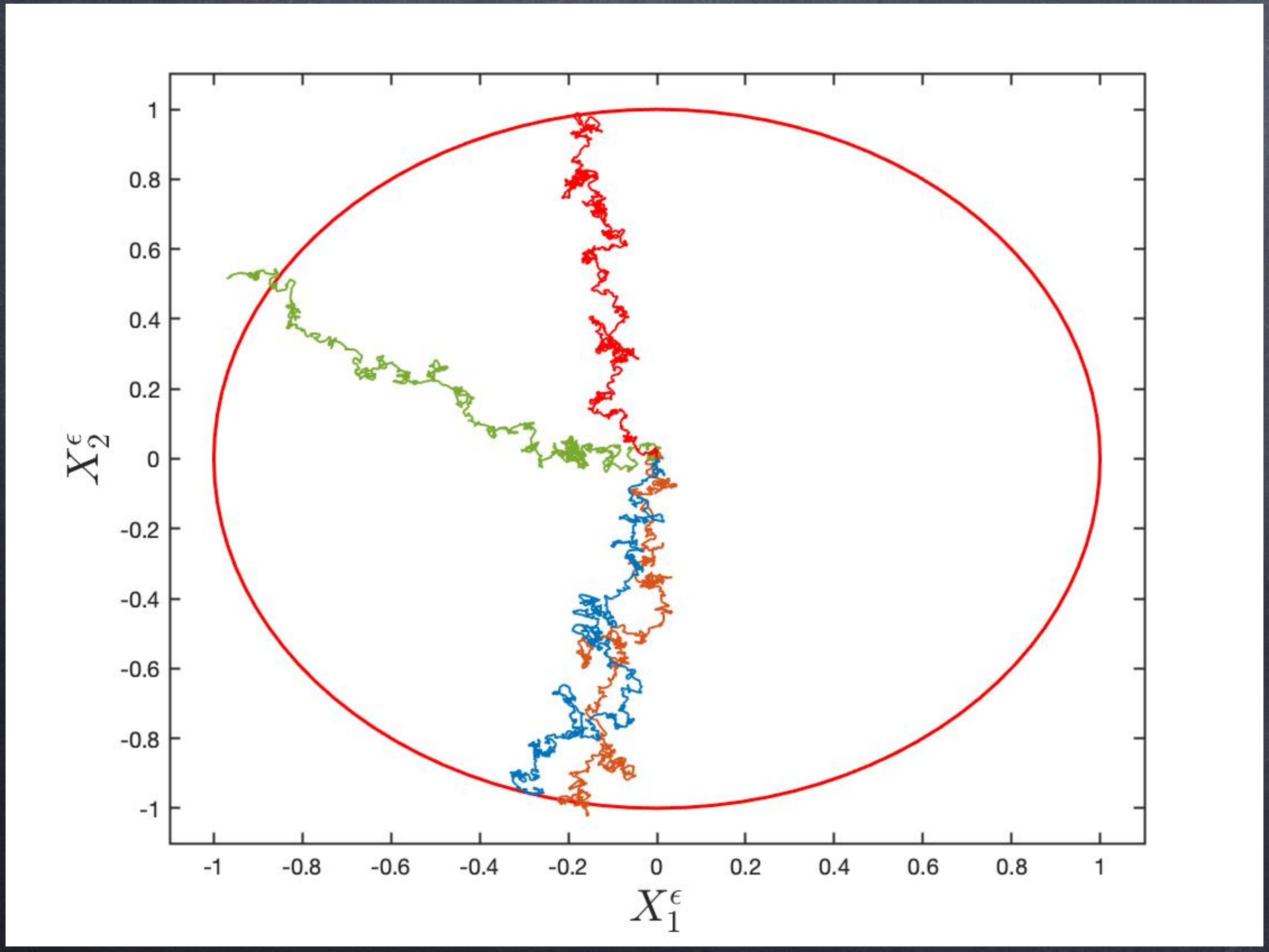
Probability $\approx 10^{-5}$

Example: Schilders theorem (cont'd)

Consider scaled BM: $\{B(t)\}_{t\in[0,T]}$ standard BM in $\mathbb{R}^2$, $B(0)=0$, $\epsilon>0$, $X^\epsilon(t)=\sqrt{\epsilon}B(t)$.

Q1: Probability X^ϵ leaves $D=\{x\in\mathbb{R}^2 \ : ||x||^2<1\}$ in $[0,T]$?

Q2: How does X^ϵ exit D? LDP: Linear towards ∂D, reach at T.



4 of 100K left D

$\epsilon = 0.044$

Probability $\approx 10^{-5}$

LDP: Empirical measures of a Markov chain

A sequence $\{X_n\}_n$ of random elements satisfy the **large deviation principle** (LDP), with **rate function** $I : \mathscr{X} \to [0,\infty]$, and speed n if

$$- \inf_{x \in G^\circ} I(x) \le \liminf_n \frac{1}{n} \log P(X_n \in G^\circ)$$
$$\le \limsup_n \frac{1}{n} \log P(X_n \in \bar{G}) \le - \inf_{x \in \bar{G}} I(x) \,.$$

LDP: Empirical measures of a Markov chain

A sequence $\{X_n\}_n$ of random elements satisfy the large deviation principle (LDP), with rate function $I : \mathscr{X} \to [0,\infty]$, and speed n if

$$- \inf_{x \in G^\circ} I(x) \le \liminf_n \frac{1}{n} \log P(X_n \in G^\circ)$$
$$\le \limsup_n \frac{1}{n} \log P(X_n \in \bar{G}) \le - \inf_{x \in \bar{G}} I(x) \, .$$

Consider a Markov chain $\{Y_n\}_{n \ge 0}$.

Define corresponding sequence of empirical measures:

$$L_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{X_i}, \quad n \ge 1.$$

Empirical measure LDP: LDP for $\{L_n\}_{n \ge 1}$.

?

III. Large deviations and Monte Carlo

Large deviations and Monte Carlo methods

Large deviations used extensively in the analysis and design of rare-event methods. Relies on process-level LDP's.

Large deviations and Monte Carlo methods

Large deviations used extensively in the analysis and design of rare-event methods. Relies on process-level LDP's.

Bucklew - Introduction to rare event simulation.
Springer-Verlag, 2004

Dupuis, Wang - Subsolutions of an Isaacs equation and efficient schemes for importance sampling.
Math. Oper. Res. 32(3), 723-757, 2007

Budhiraja, Dupuis - Analysis and approximation of rare events: Representations and weak convergence methods.
Springer, 2019.

Rhee et al. -Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound Poisson processes.
Math. Oper. Res. 44(3), 919-942, 2019.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Dupuis et al.- On the infinite swapping limit for parallel tempering.
SIAM Multiscale Model. Simul., 10(3):986-1022, 2012.

Rey-Bellet, Spiliopoulos - Irreversible Langevin samplers and variance reduction: A large deviations approach.
Nonlinearity, 28(7):2081, 2015.

Bierkens - Non-reversible Metropolis-Hastings.
Stat. Comput., 26(6):1213-1228, 2016.

Doll, Dupuis, N. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms.
Appl. Math. Optim., 78(1):103-144, 2018.

Bierkens, N., Schlottke - Large deviations for the empirical measure of the zig-zag process.
Ann. Appl. Probab., 31(6):2811-2843, 2021.

Dupuis, Wu - Analysis and optimization of certain parallel tempering Monte Carlo methods in the low temperature limit.
Multiscale Model. Simul., 20(1):220-249, 2022.

Large deviations and Monte Carlo methods

Rising interest in the use of LDPs for MCMC methods.
Empirical measure LDP's the right thing to study.

Interested in using LD approach for:

- Metropolis-adjusted Langevin algorithm (MALA),
- Random walk Metropolis (RWM),
- ABC-MCMC.

Metropolis-Hastings the foundational building block.

→ (Surprisingly!) Many (theoretical) questions remain open.

Metropolis-Hastings:

- State space $S \subseteq \mathbb{R}^d$
- Proposal distribution $J(\cdot\,|\,x)$, $x \in S$
- For a state x and proposal y, define the acceptance probability

$$\omega(x,y) = \min\left\{1, \frac{\pi(y)J(x\,|\,y)}{\pi(x)J(y\,|\,x)}\right\}.$$

- Metropolis-Hastings algorithm: Given $X_i = x_i$,

i) Generate a proposal $Y_{i+1} \sim J(\cdot\,|\,x_i)$.

ii) Set

$$X_{i+1} = \begin{cases} Y_{i+1}, & \text{w. probability } \omega(x_i, Y_{i+1}) \\ x_i, & \text{w. probability } 1-\omega(x_i, Y_{i+1})\,. \end{cases}$$

Metropolis-Hastings:

- State space $S \subseteq \mathbb{R}^d$
- Proposal distribution $J(\,\cdot\,|\,x)$, $x \in S$
- For a state x and proposal y, define the acceptance probability

$$\omega(x,y) = \min\left\{1, \frac{\pi(y)J(x\,|\,y)}{\pi(x)J(y\,|\,x)}\right\}.$$

- Metropolis-Hastings algorithm: Generate Markov chain w. kernel

$$K(x,dy) = a(x,dy) + r(x)\delta_x(dy),$$

where

$$a(x,dy) = \min\left\{1, \frac{\pi(y)J(x\,|\,y)}{\pi(x)J(y\,|\,x)}\right\} J(dy\,|\,x), \qquad r(x) = 1 - \int_S a(x,dy)\,.$$

Metropolis-Hastings:

- State space $S \subseteq \mathbb{R}^d$
- Proposal distribution $J(\cdot\,|\,x)$, $x \in S$
- For a state x and proposal y, define the acceptance probability

$$\omega(x,y) = \min\left\{1, \frac{\pi(y)J(x\,|\,y)}{\pi(x)J(y\,|\,x)}\right\}.$$

- Metropolis-Hastings algorithm: Generate Markov chain w. kernel

$$K(x,dy) = a(x,dy) + r(x)\delta_x(dy),$$

where

$$a(x,dy) = \min\left\{1, \frac{\pi(y)J(x\,|\,y)}{\pi(x)J(y\,|\,x)}\right\} J(dy\,|\,x), \qquad r(x) = 1 - \int_S a(x,dy).$$

Q: What about empirical measure large deviations for MH chains?

IV. Large deviations for MH chains

Large deviations for Metropolis-Hastings chains:

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u>0} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathscr{P}(S).$$

Large deviations for Metropolis-Hastings chains:

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u>0} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathscr{P}(S) .$$

Q: What about MH chains?

Large deviations for Metropolis-Hastings chains:

Let $\mathfrak{X}$ be a compact metric space and let $\lambda(dx)$ be a probability measure on $\mathfrak{X}$. Let $X_0, X_1, X_2, \cdots$ be a stationary Markov process whose state space is $\mathfrak{X}$, with $X_0 = x$, having transition probability function $\pi(x, dy)$ about which we assume:

1. $\pi(x, dy) = \pi(x, y)\lambda(dy)$,

2. there exist constants a and A such that $0 < a \leqq \pi(x, y) \leqq A < \infty$ for all $x \in \mathfrak{X}$ and almost all (λ-measure) $y \in \mathfrak{X}$,

3. for any function $u(y) \in L_1(\lambda)$,

$$\int_{\mathfrak{X}} \pi(x, y)u(y)\lambda(dy)$$

is a continuous function of x.

(Donsker, Varadhan '75)

Large deviations for Metropolis-Hastings chains:

Let $\mathbf{X}$ be a compact metric space and let $\lambda(dx)$ be a probability measure on $\mathbf{X}$. Let $X_0, X_1, X_2, \cdots$ be a stationary Markov process whose state space is $\mathbf{X}$, with $X_0 = x$, having transition probability function $\pi(x, dy)$ about which we assume:

1. $\pi(x, dy) = \pi(x, y)\lambda(dy)$,

2. there exist constants a and A such that $0 < a \leqq \pi(x, y) \leqq A < \infty$ for all $x \in \mathbf{X}$ and almost all (λ-measure) $y \in \mathbf{X}$,

3. for any function $u(y) \in L_1(\lambda)$,

$$\int_{\mathbf{X}} \pi(x, y) u(y) \lambda(dy)$$

is a continuous function of x.

(Donsker, Varadhan '75)

Kontoyiannis, Meyn '05)

(i) The Markov process $\mathbf{\Phi}$ is ψ-irreducible, aperiodic, and it satisfies condition (DV3) with some Lyapunov function $V : \mathsf{X} \to [1, \infty)$;

(ii) There exists $T_0 > 0$ such that, for each $r < \|W\|_\infty$, there is a measure β_r with $\beta_r(e^V) < \infty$ and $\mathsf{P}_x\{\Phi(T_0) \in A,\ \tau_{C_W^c(r)} > T_0\} \le \beta_r(A)$ for all $x \in C_W(r)$, $A \in \mathcal{B}$.

(DV3+)

Large deviations for Metropolis-Hastings chains:

Let $\mathfrak{X}$ be a compact metric space and let $\lambda(dx)$ be a probability measure on $\mathfrak{X}$. Let $X_0, X_1, X_2, \cdots$ be a stationary Markov process whose state space is $\mathfrak{X}$, with $X_0 = x$, having transition probability function $\pi(x, dy)$ about which we assume:

1. $\pi(x, dy) = \pi(x, y)\lambda(dy)$,

2. there exist constants a and A such that $0 < a \leqq \pi(x, y) \leqq A < \infty$ for all $x \in \mathfrak{X}$ and almost all (λ-measure) $y \in \mathfrak{X}$,

3. for any function $u(y) \in L_1(\lambda)$,

$$\int_{\mathfrak{X}} \pi(x, y) u(y) \lambda(dy)$$

is a continuous function of x.

(Donsker, Varadhan '75)

Kontoyiannis, Meyn '05)

$$\left.\begin{array}{l}\text{(i) The Markov process } \boldsymbol{\Phi} \text{ is } \psi\text{-irreducible, aperiodic, and it satisfies} \\ \quad \text{condition (DV3) with some Lyapunov function } V : X \to [1, \infty); \\ \text{(ii) There exists } T_0 > 0 \text{ such that, for each } r < \|W\|_\infty \text{, there is a measure } \beta_r \\ \quad \text{with } \beta_r(e^V) < \infty \text{ and } \mathsf{P}_x\{\Phi(T_0) \in A,\ \tau_{C_W^c(r)} > T_0\} \leq \beta_r(A) \text{ for all} \\ \quad x \in C_W(r),\ A \in \mathcal{B}.\end{array}\right\} \textbf{(DV3+)}$$

Condition 6.3 *The transition kernel p satisfies the following transitivity condition. There exist positive integers l_0 and n_0 such that for all x and ζ in S,*

$$\sum_{i=l_0}^{\infty} \frac{1}{2^i} p^{(i)}(x, dy) \ll \sum_{j=n_0}^{\infty} \frac{1}{2^j} p^{(j)}(\zeta, dy), \qquad (6.7)$$

where $p^{(k)}$ denotes the k-step transition probability.

(Dupuis, Liu '15; Budhiraja, Dupuis '19)

Large deviations for Metropolis-Hastings chains:

Need new conditions adapted to MH-type dynamics.

Main issue: Rejection part $r(x)\delta_x(dy)$ in K.

Large deviations for Metropolis-Hastings chains:

Need new conditions adapted to MH-type dynamics.

Main issue: Rejection part $r(x)\delta_x(dy)$ in K.

One possible set of assumptions:

A.1) Target π equivalent to λ (Lebesgue) on S, has cont. density.

A.2) Proposal $J(\cdot\,|\,x) \ll \pi$ for all $x \in S$. Density is cont. and bounded and $J(y\,|\,x) > 0$ for all $(x, y) \in S^2$.

A.3) There exists a suitable Lyapunov-type function associated with K (for non-compactness)

Large deviations for Metropolis-Hastings chains:

Need new conditions adapted to MH-type dynamics.

Main issue: Rejection part $r(x)\delta_x(dy)$ in K.

One possible set of assumptions:

A.1) Target π equivalent to λ (Lebesgue) on S, has cont. density.

A.2) Proposal $J(\cdot\,|\,x) \ll \pi$ for all $x \in S$. Density is cont. and bounded and $J(y\,|\,x) > 0$ for all $(x, y) \in S^2$.

A.3) There exists a suitable Lyapunov-type function associated with K (for non-compactness)

Large deviations for Metropolis-Hastings chains:

Theorem (Milinanni, N. '24a): Under assumptions (A.1), (A.2), (A.3), the empirical measures $\{L_n\}_{n\geq 1}$ associated with the MH chain $\{X_i\}_{i\geq 0}$ satisfy an LDP with rate function

$$I(\mu) = \inf_{\gamma\in A(\mu)} R(\gamma \,||\, \mu \otimes K), \quad \mu \in \mathscr{P}(S)\,.$$

$$A(\mu) = \{\gamma \in \mathscr{P}(S^2) : [\gamma]_1 = [\gamma]_2 = \mu\}\,.$$

$$R(\mu \,||\, \nu) = \begin{cases} \int_S \log\left(\frac{d\mu}{d\nu}\right) d\mu, & \mu \ll \nu, \\ +\infty, & \text{otherwise}\,. \end{cases}$$

Large deviations for Metropolis-Hastings chains:

Theorem (Milinanni, N. '24a): Under assumptions (A.1), (A.2), (A.3), the empirical measures $\{L_n\}_{n\geq 1}$ associated with the MH chain $\{X_i\}_{i\geq 0}$ satisfy an LDP with rate function

$$I(\mu) = \inf_{\gamma \in A(\mu)} R(\gamma \,||\, \mu \otimes K), \quad \mu \in \mathscr{P}(S)\,.$$

$$A(\mu) = \{\gamma \in \mathscr{P}(S^2) : [\gamma]_1 = [\gamma]_2 = \mu\}\,.$$

$$R(\mu \,||\, \nu) = \begin{cases} \int_S \log\left(\frac{d\mu}{d\nu}\right) d\mu, & \mu \ll \nu, \\ +\infty, & \text{otherwise}\,. \end{cases}$$

Idea: Use rate function to gauge efficiency / compare alg's.
"Larger = better"

Toy example: IMH

Toy example (WIP): Independent MH sampler

Proposal distribution $J(\,\cdot\,|x) = f(\,\cdot\,)$, $\forall x \in S$.

Q: For a given target, can we find the "best" sampling dist.?

Toy example (WIP): Independent MH sampler

Proposal distribution $J(\cdot \,|\, x) = f(\cdot)$, $\forall x \in S$.

Q: For a given target, can we find the "best" sampling dist.?

Take $\pi \sim N(0,1)$, $f \sim N(m, s^2)$. Rate function $I(\cdot) = I_f(\cdot) = I(\cdot\,; m, s)$

Toy example (WIP): Independent MH sampler

Proposal distribution $J(\cdot \,|\, x) = f(\cdot)$, $\forall x \in S$.

Q: For a given target, can we find the "best" sampling dist.?

Take $\pi \sim N(0,1)$, $f \sim N(m, s^2)$. Rate function $I(\cdot) = I_f(\cdot) = I(\cdot\,; m, s)$

"Ideal": find optimal $(m*, s*)$ for all (relevant) $\mu \in \mathscr{P}(S)$:

$$I(\mu; m^*, s^*) \geq I(\mu; m, s), \quad \forall \mu, m, s \,.$$

Toy example (WIP): Independent MH sampler

Proposal distribution $J(\,\cdot\,|x) = f(\,\cdot\,)$, $\forall x \in S$.

Q: For a given target, can we find the "best" sampling dist.?

Take $\pi \sim N(0,1)$, $f \sim N(m, s^2)$. Rate function $I(\,\cdot\,) = I_f(\,\cdot\,) = I(\,\cdot\,; m, s)$

"Ideal": find optimal $(m*, s*)$ for all (relevant) $\mu \in \mathscr{P}(S)$:

$$I(\mu; m^*, s^*) \geq I(\mu; m, s), \quad \forall \mu, m, s\,.$$

Reality: Numerical comparison of lower bound for a given μ.

Lower bound for the rate function:

$$I_f(\mu) \geq -\log\left(1 - \frac{1}{2}\iint \min\left\{\frac{f(x)}{\pi(x)}, \frac{f(y)}{\pi(y)}\right\}\left(\sqrt{\mu(x)\pi(y)} - \sqrt{\mu(y)\pi(x)}\right)^2 dxdy\right)$$

Toy example (WIP): Independent MH sampler

Proposal distribution $J(\,\cdot\,|x) = f(\,\cdot\,)$, $\forall x \in S$.

Q: For a given target, can we find the "best" sampling dist.?

Take $\pi \sim N(0,1)$, $f \sim N(m, s^2)$. Rate function $I(\,\cdot\,) = I_f(\,\cdot\,) = I(\,\cdot\,; m, s)$

Lower bound for the rate function:

$$I_f(\mu) \geq -\log\left(1 - \frac{1}{2}\iint \min\left\{\frac{f(x)}{\pi(x)}, \frac{f(y)}{\pi(y)}\right\}\left(\sqrt{\mu(x)\pi(y)} - \sqrt{\mu(y)\pi(x)}\right)^2 dxdy\right)$$

Toy example (WIP): Independent MH sampler

Proposal distribution $J(\cdot \mid x) = f(\cdot)$, $\forall x \in S$.

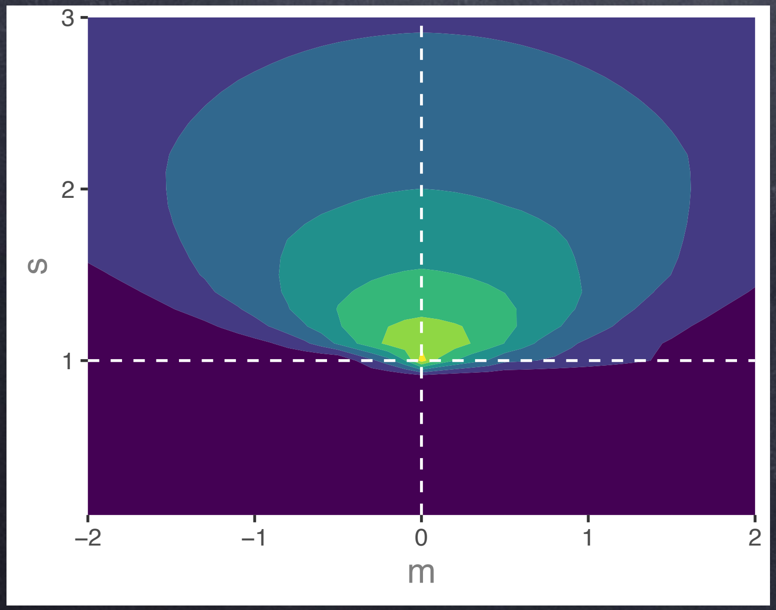
Q: For a given target, can we find the "best" sampling dist.?

Take $\pi \sim N(0,1)$, $f \sim N(m, s^2)$. Rate function $I(\cdot) = I_f(\cdot) = I(\cdot\,; m, s)$

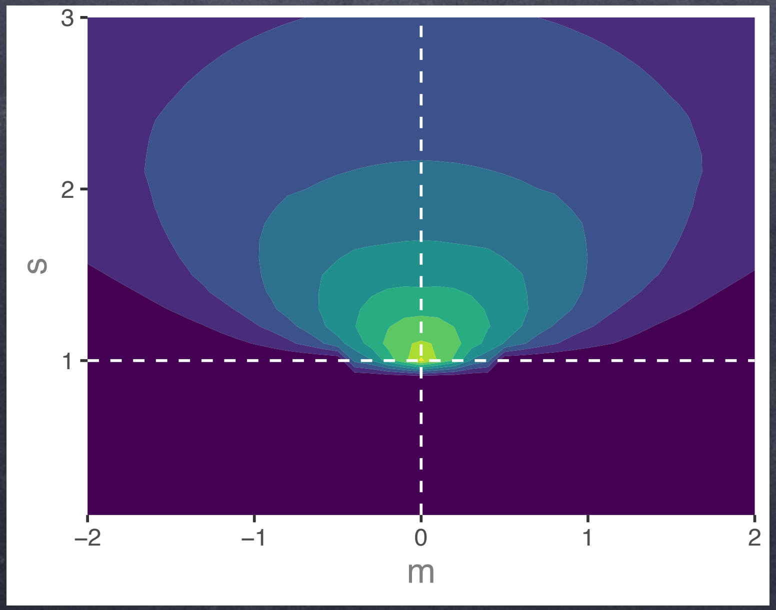
Lower bound for the rate function:

$$I_f(\mu) \geq -\log\left(1 - \frac{1}{2}\iint \min\left\{\frac{f(x)}{\pi(x)}, \frac{f(y)}{\pi(y)}\right\}\left(\sqrt{\mu(x)\pi(y)} - \sqrt{\mu(y)\pi(x)}\right)^2 dxdy\right)$$

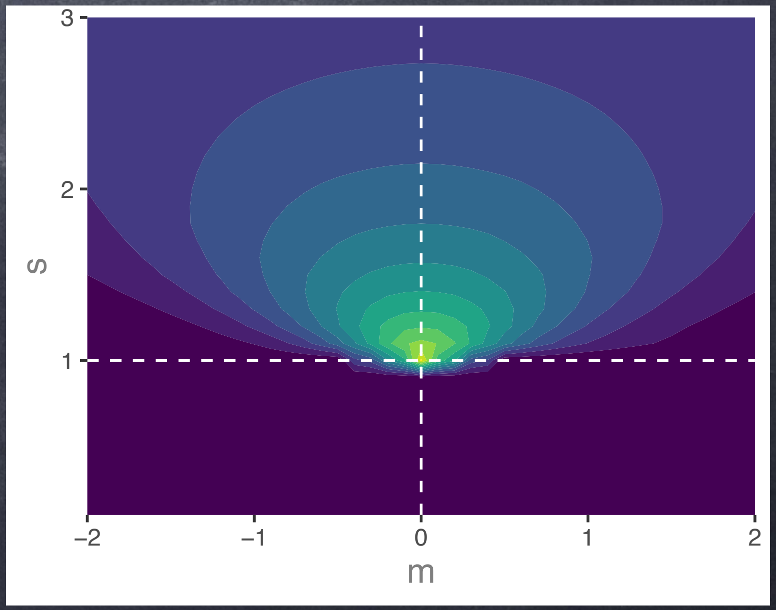
$\mu \sim N(1,2)$



$\mu \sim Gamma(3,5)$



$\mu \sim Uni(0,1)$



Toy example (WIP): Independent MH sampler

Proposal distribution $J(\cdot\,|x) = f(\cdot)$, $\forall x \in S$.

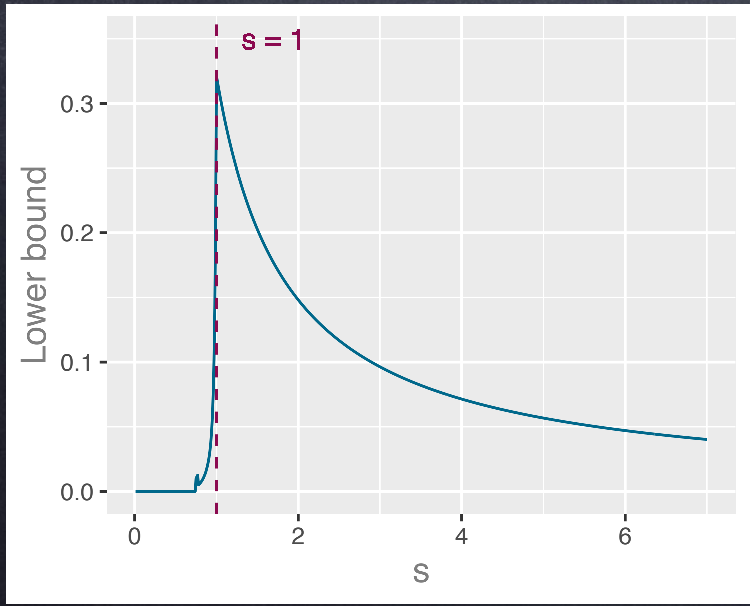
Q: For a given target, can we find the "best" sampling dist.?

Take $\pi \sim N(0,1)$, $f \sim N(m, s^2)$. Rate function $I(\cdot) = I_f(\cdot) = I(\cdot\,; m, s)$

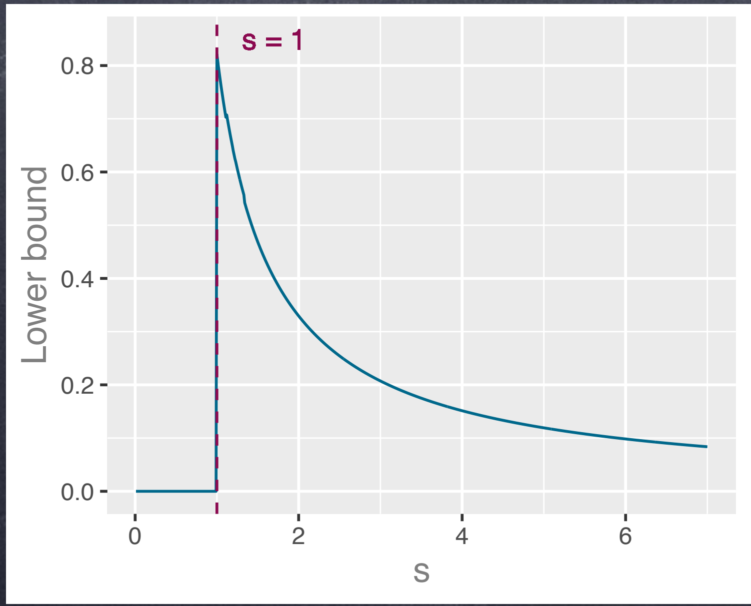
Lower bound for the rate function:

$$I_f(\mu) \geq -\log\left(1 - \frac{1}{2}\iint \min\left\{\frac{f(x)}{\pi(x)}, \frac{f(y)}{\pi(y)}\right\}\left(\sqrt{\mu(x)\pi(y)} - \sqrt{\mu(y)\pi(x)}\right)^2 dxdy\right)$$

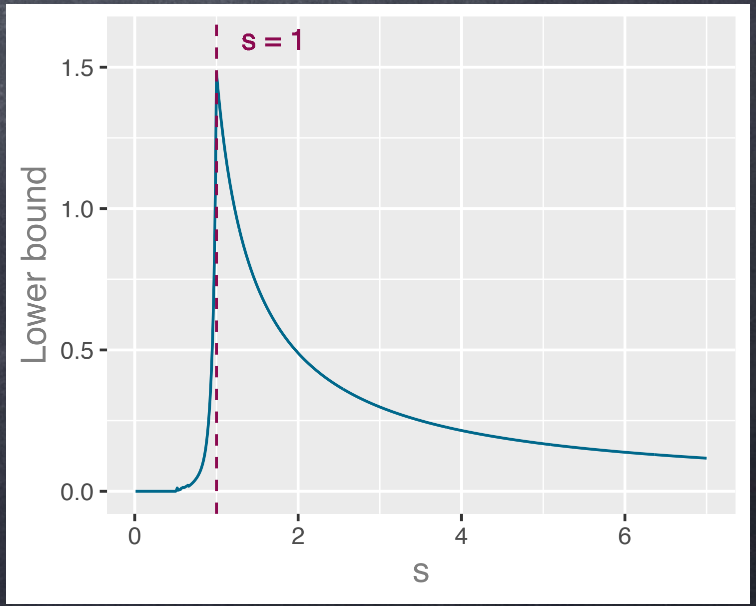
$\mu \sim N(1,2)$



$\mu \sim Gamma(3,5)$



$\mu \sim Uni(0,1)$



Large deviations for Metropolis-Hastings chains:

Theorem (Milinanni, N. '24a): Under assumptions (A.1), (A.2), (A.3), the empirical measures $\{L_n\}_{n\geq 0}$ associated with the MH chain $\{X_i\}_{i\geq 0}$ satisfy an LDP with rate function

$$I(\mu) = \inf_{\gamma \in A(\mu)} R(\gamma \,||\, \mu \otimes K), \quad \mu \in \mathscr{P}(S)\,.$$

$$A(\mu) = \{\gamma \in \mathscr{P}(S^2) : [\gamma]_1 = [\gamma]_2 = \mu\}\,.$$

$$R(\mu \,||\, \nu) = \begin{cases} \int_S \log\left(\frac{d\mu}{d\nu}\right) d\mu, & \mu \ll \nu, \\ +\infty, & \text{otherwise}. \end{cases}$$

Idea: Use rate function to gauge efficiency / compare alg's.
"Larger = better"

Large deviations for Metropolis-Hastings chains:

Theorem (Milinanni, N. '24a): Under assumptions (A.1), (A.2), (A.3), the empirical measures $\{L_n\}_{n\geq 0}$ associated with the MH chain $\{X_i\}_{i\geq 0}$ satisfy an LDP with rate function

$$I(\mu) = \inf_{\gamma \in A(\mu)} R(\gamma \,||\, \mu \otimes K), \quad \mu \in \mathscr{P}(S)\,.$$

Proof strategy: Establish variational upper & lower bounds:

$$\limsup_{n\to\infty} \;(\liminf_{n\to\infty}) -\frac{1}{n} \log E\left[e^{-nF(L_n)}\right] \leq (\;\geq\;) \inf_{\mu \in \mathscr{P}(S)} \big(F(\mu) + I(\mu)\big)$$

Relies on stochastic control and weak convergence methods.

Large deviations for Metropolis-Hastings chains:

I. Variational representation: For F bounded, cont.,

$$-\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] = \inf_{\{\bar{\mu}_i^n\}} E\left[F(\bar{L}_n) + \frac{1}{n}\sum_{i=1}^{n} R(\bar{\mu}_i^n \,||\, K(\bar{X}_i^n, \cdot\,)\right].$$

$\bar{\mu}_i^n$: cond. distribution of $\bar{X}_i^n$ given $\sigma(\bar{X}_1^n, \ldots, \bar{X}_{n-1}^n)$.

$\bar{L}^n(\,\cdot\,) = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{\bar{X}_i^n}(\,\cdot\,)$: controlled empirical measure.

Large deviations for Metropolis-Hastings chains:

I. Variational representation: For F bounded, cont.,

$$-\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] = \inf_{\{\bar{\mu}_i^n\}} E\left[F(\bar{L}_n) + \frac{1}{n}\sum_{i=1}^{n} R(\bar{\mu}_i^n \;||\; K(\bar{X}_i^n, \cdot\,)\right].$$

$\bar{\mu}_i^n$: cond. distribution of $\bar{X}_i^n$ given $\sigma(\bar{X}_1^n, \ldots, \bar{X}_{n-1}^n)$.

$\bar{L}^n(\,\cdot\,) = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{\bar{X}_i^n}(\,\cdot\,)$: controlled empirical measure.

II. Variational upper bound:

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] \geq \inf_{\mu\in\mathcal{P}(S)} \big(F(\mu) + I(\mu)\big)$$

"Easy" direction. Show Feller property for K. Rest from Budhiraja & Dupuis.

Large deviations for Metropolis-Hastings chains:

III. Variational lower bound:

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] \leq \inf_{\mu\in\mathscr{P}(S)} \big(F(\mu) + I(\mu)\big)$$

Large deviations for Metropolis-Hastings chains:

III. Variational lower bound:

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] \le \inf_{\mu\in\mathscr{P}(S)} \big(F(\mu)+I(\mu)\big)$$

Difficult part: construction of near-optimal controls $\{\bar{\mu}_i^n\}_{i=1}^n$.

Key property in Budhiraja & Dupuis: $I(\nu)<\infty$ guarantees $\nu \ll \pi$.

Large deviations for Metropolis-Hastings chains:

III. Variational lower bound:

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] \le \inf_{\mu\in\mathscr{P}(S)} \big(F(\mu)+I(\mu)\big)$$

Difficult part: construction of near-optimal controls $\{\bar{\mu}_i^n\}_{i=1}^n$.

Key property in Budhiraja & Dupuis: $I(\nu) < \infty$ guarantees $\nu \ll \pi$.

Not true for MH; due to rejection part $r(x)\delta_x(dy)$ in K.

Large deviations for Metropolis-Hastings chains:

III. Variational lower bound:

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(L_n)}\right] \leq \inf_{\mu\in\mathscr{P}(S)} \left(F(\mu)+I(\mu)\right)$$

Difficult part: construction of near-optimal controls $\{\bar{\mu}_i^n\}_{i=1}^n$.

Key property in Budhiraja & Dupuis: $I(\nu) < \infty$ guarantees $\nu \ll \pi$.

Not true for MH; due to rejection part $r(x)\delta_x(dy)$ in K.

Idea: Take $\nu \in \mathscr{P}(S)$ s.t. $I(\nu) < \infty$. Show existence of ν^* s.t.:

(i) arbitrarily close to ν,

(ii) $I(\nu^*) \leq I(\nu) + \epsilon$,

(iii) $\nu^* \ll \pi$.

Condition (A.3) needed to show tightness of controls.

?

V. On condition (A.3): Existence of a suitable Lyapunov function

(is it ever satisfied?)

Existence of Lyapunov function I:

Condition (A.3): There exists a function $U : S \to [0,\infty)$ such that

a) $\inf_{x \in S} \left\{ U(x) - \log \int_S e^{U(y)} K(x, dy) \right\} > -\infty$.

b) For each $M < \infty$, the following set is relatively compact in S:

$$\left\{ x \in S : U(x) - \log \int_S e^{U(y)} K(x, dy) \le M \right\}.$$

c) For every compact $A \subset S$, there exists $C_A < \infty$ such that

$$\sup_{x \in A} U(x) \le C_A.$$

Existence of Lyapunov function I:

Condition (A.3): There exists a function $U : S \to [0,\infty)$ such that

a) $\inf_{x \in S} \left\{ U(x) - \log \int_S e^{U(y)} K(x, dy) \right\} > -\infty.$

b) For each $M < \infty$, the following set is relatively compact in S:

$$\left\{ x \in S : U(x) - \log \int_S e^{U(y)} K(x, dy) \leq M \right\}.$$

c) For every compact $A \subset S$, there exists $C_A < \infty$ such that

$$\sup_{x \in A} U(x) \leq C_A.$$

Note: For compact S condition is trivially satisfied.

Henceforth: $S = \mathbb{R}^d$.

Existence of Lyapunov function II:

Condition (A.3): Part (b) critical part,

b) For each $M < \infty$, the following set is relatively compact in S:

$$\left\{ x \in S : U(x) - \log \int_S e^{U(y)} K(x, dy) \leq M \right\}.$$

Existence of Lyapunov function II:

Condition (A.3): Part (b) critical part,

b) For each $M < \infty$, the following set is relatively compact in S:

$$\left\{ x \in S : U(x) - \log \int_S e^{U(y)} K(x, dy) \leq M \right\}.$$

Proposition (Milinanni, N., 24b): (A.3b) is equivalent to

$$\lim_{|x| \to \infty} \int_S a(x, y) dy = 1,$$

and

$$\lim_{|x| \to \infty} \int_S e^{U(y) - U(x)} a(x, y) dy = 0.$$

(where: $a(x, dy) = \min \left\{ 1, \frac{\pi(y) J(x \,|\, y)}{\pi(x) J(y \,|\, x)} \right\} J(dy \,|\, x)$)

Existence of Lyapunov function III: Independent MH

Existence of Lyapunov function III: Independent MH

Proposal distribution $J(\,\cdot\,|x) = f(\,\cdot\,)$, $\forall x \in S$.

$$\Rightarrow\ a(x,y) = \min\left\{1, \frac{\pi(y)f(x)}{\pi(x)f(y)}\right\} f(y)\ ,\ \forall x \in S.$$

Consider target and proposal on the form

$$\pi(x) \propto e^{-\eta|x|^{\alpha}}, \quad f(y) \propto e^{-\gamma|x|^{\beta}}.$$

Existence of Lyapunov function III: Independent MH

Proposal distribution $J(\cdot\,|x) = f(\cdot)$, $\forall x \in S$.

$$\Rightarrow\ a(x,y) = \min\left\{1, \frac{\pi(y)f(x)}{\pi(x)f(y)}\right\} f(y)\ ,\ \forall x \in S.$$

Consider target and proposal on the form

$$\pi(x) \propto e^{-\eta|x|^{\alpha}}, \quad f(y) \propto e^{-\gamma|x|^{\beta}}.$$

Proposition (Milinanni, N., 24b): (A.3) is satisfied iff either of the following hold:

i) $\alpha = \beta,\ \ \eta > \gamma,$

ii) $\alpha \geq \beta$.

Gist: Target has lighter tails than proposal. Same as for UE/GE.

Existence of Lyapunov function IV: MALA

Existence of Lyapunov function IV: MALA

Proposal distribution:

$$J(y|x) \propto \exp\left\{-\frac{1}{2\varepsilon}\left|y - x - \frac{\varepsilon}{2}\nabla\log\pi(x)\right|^2\right\}, \quad \varepsilon > 0.$$

Consider target on the form

$$\pi(x) \propto e^{-\eta|x|^\alpha}.$$

Existence of Lyapunov function IV: MALA

Proposal distribution:

$$J(y|x) \propto \exp\left\{-\frac{1}{2\varepsilon}\left|y - x - \frac{\varepsilon}{2}\nabla \log \pi(x)\right|^2\right\}, \quad \varepsilon > 0.$$

Consider target on the form

$$\pi(x) \propto e^{-\eta|x|^\alpha}.$$

Proposition (Milinanni, N., 24b): (A.3) is satisfied iff either of the following hold:

i) $\alpha = 2, \quad \varepsilon\eta < 2,$

ii) $1 < \alpha < 2.$

Existence of Lyapunov function V: RWM

Existence of Lyapunov function V: RWM

Proposal distribution $J(y|x) = \hat{J}(y-x) = \hat{J}(x-y)$.

$$\Rightarrow \; a(x,y) = \min\left\{ 1, \frac{\pi(y)}{\pi(x)} \right\} \hat{J}(y-x) \;, \; \forall x \in S.$$

Existence of Lyapunov function V: RWM

Proposal distribution $J(y|x) = \hat{J}(y-x) = \hat{J}(x-y)$.

$$\Rightarrow a(x,y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}\hat{J}(y-x) \;,\; \forall x \in S.$$

Proposition (Milinanni, N., 24b): For the RWM algorithm, there does not exist a function U satisfying condition (A.3), regardless of the choice of π.

LDP for MH chains: LDPs for IMH and MALA chains

Theorem (Milinanni, N., 24b):

Consider a target on the form

$$\pi(x) \propto e^{-\eta|x|^{\alpha}}.$$

LDP for MH chains: LDPs for IMH and MALA chains

Theorem (Milinanni, N., 24b):

Consider a target on the form

$$\pi(x) \propto e^{-\eta|x|^{\alpha}}.$$

i) For IMH, with proposal on the form $f(y) \propto e^{-\gamma|x|^{\beta}}$, if either $\alpha = \beta$ and $\eta > \gamma$, or $\alpha > \beta$, the empirical measures of the underlying MH chain satisfy an LDP.

LDP for MH chains: LDPs for IMH and MALA chains

Theorem (Milinanni, N., 24b):

Consider a target on the form

$$\pi(x) \propto e^{-\eta |x|^{\alpha}}.$$

i) For IMH, with proposal on the form $f(y) \propto e^{-\gamma |x|^{\beta}}$, if either $\alpha = \beta$ and $\eta > \gamma$, or $\alpha > \beta$, the empirical measures of the underlying MH chain satisfy an LDP.

ii) For MALA, with proposal

$$J(y \mid x) \propto \exp\left\{ -\frac{1}{2\varepsilon} \left| y - x + \frac{\varepsilon \eta \alpha}{2} |x|^{\alpha-2} x \right|^{2} \right\}, \quad \varepsilon > 0.$$

if either $\alpha = 2$ and $\varepsilon \eta < 2$, or $\alpha \in (1,2)$, the empirical measures of the underlying MH chain satisfy an LDP.

LDP for MH chains: LDPs for IMH and MALA chains

Theorem (Milinanni, N., 24b):

Consider a target on the form

$$\pi(x) \propto e^{-\eta|x|^{\alpha}}.$$

i) For IMH, with proposal on the form $f(y) \propto e^{-\gamma|x|^{\beta}}$, if either $\alpha = \beta$ and $\eta > \gamma$, or $\alpha > \beta$, the empirical measures of the underlying MH chain satisfy an LDP.

ii) For MALA, with proposal

$$J(y|x) \propto \exp\left\{-\frac{1}{2\varepsilon}\left|y - x + \frac{\varepsilon\eta\alpha}{2}|x|^{\alpha-2}x\right|^2\right\}, \quad \varepsilon > 0.$$

if either $\alpha = 2$ and $\varepsilon\eta < 2$, or $\alpha \in (1,2)$, the empirical measures of the underlying MH chain satisfy an LDP.

Q: When should we expect an LDP to hold for MH chains?

LDP for MH chains: A conjecture

Q: When should we expect an LDP to hold for MH chains?

LDP for MH chains: A conjecture

Q: When should we expect an LDP to hold for MH chains?

I. Comparison of (A.3) and drift condition: Standard drift cond. for V: $\lambda \in (0,1)$, $b < \infty$,

$$\int_S V(y)K(x,dy) \leq \lambda V(x) + bI\{x \in C\}.$$

For $U = logV$ drift condition becomes

$$U(x) - \log \int_S e^{U(y)}K(x,dy) \geq -\log\left(\lambda + e^{-U(x)}bI\{x \in C\}\right).$$

$\Rightarrow$ the Lyapunov function V gives rise to U satisfying (A.3a).

LDP for MH chains: A conjecture

Q: When should we expect an LDP to hold for MH chains?

I. Comparison of (A.3) and drift condition: Standard drift cond. for V: $\lambda \in (0,1)$, $b < \infty$,

$$\int_S V(y)K(x,dy) \leq \lambda V(x) + bI\{x \in C\}.$$

For $U = \log V$ drift condition becomes

$$U(x) - \log \int_S e^{U(y)}K(x,dy) \geq -\log\left(\lambda + e^{-U(x)}bI\{x \in C\}\right).$$

$\Rightarrow$ the Lyapunov function V gives rise to U satisfying (A.3a).

II. Previous LDP results: Typically for geometrically ergodic chains (e.g., Kontoyiannis & Meyn '03, '05).

LDP for MH chains: A conjecture

III. Results for IMH, MALA, RWM:

LDP for MH chains: A conjecture

III. Results for IMH, MALA, RWM:

		Assumption (A.3)	Geometric ergodicity
IMH	$\alpha = \beta, \eta > \gamma,$ or $\alpha \geq \beta$.	✓	✓
	otherwise	✗	✗
MALA	$\alpha = 2,\ \varepsilon\eta < 2,$ or $\alpha \in (1,2)$.	✓	✓
	$\alpha = 1$	✗	(✓)
	otherwise	✗	✗
RWM	tails as in [MT96]	✗	✓
	otherwise	✗	✗

LDP for MH chains: A conjecture

III. Results for IMH, MALA, RWM:

		Assumption (A.3)	Geometric ergodicity
IMH	$\alpha = \beta, \eta > \gamma,$ or $\alpha \geq \beta$.	✓	✓
	otherwise	✗	✗
MALA	$\alpha = 2,\ \varepsilon\eta < 2,$ or $\alpha \in (1,2)$.	✓	✓
	$\alpha = 1$	✗	(✓)
	otherwise	✗	✗
RWM	tails as in [MT96]	✗	✓
	otherwise	✗	✗

Current (abstract) LDP: (A.3b) the restrictive condition.

Conjecture: (A.3b) too strict, geometric ergodicity enough.

On-going/future work

On-going/future work

Alternative representations for the rate function.

Similar to work by D-V; relation to Dirichlet forms...

On-going/future work

Alternative representations for the rate function.

Similar to work by D-V; relation to Dirichlet forms…

In-depth study of RWM and non-reversible setting.

Compare to recent work by Andi et al.

Generalise the finite-state examples by Bierkens '16.

On-going/future work

Alternative representations for the rate function.

- Similar to work by D-V; relation to Dirichlet forms...

In-depth study of RWM and non-reversible setting.

- Compare to recent work by Andi et al.
- Generalise the finite-state examples by Bierkens '16.

Examine connection LDP $\Leftrightarrow$ geometric ergodicity.

On-going/future work

Alternative representations for the rate function.

Similar to work by D-V; relation to Dirichlet forms...

In-depth study of RWM and non-reversible setting.

Compare to recent work by Andi et al.

Generalise the finite-state examples by Bierkens '16.

Examine connection LDP ⇔ geometric ergodicity.

Extend LDP approach to other types/classes of algorithms.

On-going/future work

Alternative representations for the rate function.

Similar to work by D-V; relation to Dirichlet forms...

In-depth study of RWM and non-reversible setting.

Compare to recent work by Andi et al.

Generalise the finite-state examples by Bierkens '16.

Examine connection LDP $\Leftrightarrow$ geometric ergodicity.

Extend LDP approach to other types/classes of algorithms.

Examine high-dimensional limit/optimal scaling using LD/rate function.

Thank you!

@PierreNyq

https://people.kth.se/~pierren/

Bonus material

Spectral properties: Concern the convergence rate of transition probabilities. Easy to come up with examples of processes with large spectral gap but fast convergence of time averages.

Ex. (Rosenthal '03): $$P = \begin{pmatrix} \epsilon & 1-\epsilon \\ 1-\epsilon & \epsilon \end{pmatrix}.$$

Empirical measure converges rapidly to (1/2,1/2). Spectral gap suggest very slow convergence.

Reversibility of MH and MH-like algorithms often good:

+ Neat mathematical theory: self-adjoint transition operator, spectrum is real, geometric ergodicity gives CLT for L^2 functions...

+ Local condition; helps with implementation.

- Leads to random-walk behaviour. Pot. slow convergence and high computational cost per iteration.

Reversibility of MH and MH-like algorithms often good:

+ Neat mathematical theory: self-adjoint transition operator, spectrum is real, geometric ergodicity gives CLT for L^2 functions...

+ Local condition; helps with implementation.

- Leads to random-walk behaviour. Pot. slow convergence and high computational cost per iteration.

Non-reversible processes avoid RW behaviour by introducing auxiliary variables (e.g. velocity).

Reversibility of MH and MH-like algorithms often good:

+ Neat mathematical theory: self-adjoint transition operator, spectrum is real, geometric ergodicity gives CLT for L^2 functions...

+ Local condition; helps with implementation.

- Leads to random-walk behaviour. Pot. slow convergence and high computational cost per iteration.

Non-reversible processes avoid RW behaviour by introducing auxiliary variables (e.g. velocity).

Continuous-time MCMC methods introduced to have such non-reversible processes. Based on piecewise deterministic Markov processes (PDMPs).

Large deviations for Metropolis-Hastings chains:

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u \in \mathscr{D}^+(L)} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathscr{P}(S).$$

Large deviations for Metropolis-Hastings chains:

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u \in \mathcal{D}^+(L)} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathcal{P}(S).$$

DV-like results typically rely on the following properties:

<u>Large deviations for Metropolis-Hastings chains:</u>

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u \in \mathscr{D}^+(L)} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathscr{P}(S).$$

DV-like results typically rely on the following properties:

DV.I) The semigroup is Feller continuous and strongly continuous.

Large deviations for Metropolis-Hastings chains:

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u \in \mathscr{D}^+(L)} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathscr{P}(S).$$

DV-like results typically rely on the following properties:

DV.1) The semigroup is Feller continuous and strongly continuous.

DV.2) There is a reference measure such that transition probabilities are abs. cont. w.r.t. this measure (transitivity assump.)

Large deviations for Metropolis-Hastings chains:

Empirical measure large deviations for Markov processes dates back to work by Donsker and Varadhan ('75-'76)

Covers many (well-behaved) Markov processes, rate function on variational form:

$$I(\mu) = -\inf_{u \in \mathscr{D}^+(L)} \int \log \frac{Ku}{u} d\mu, \quad \mu \in \mathscr{P}(S).$$

DV-like results typically rely on the following properties:

DV.1) The semigroup is Feller continuous and strongly continuous.

DV.2) There is a reference measure such that transition probabilities are abs. cont. w.r.t. this measure (transitivity assump.)