A calculus for Markov chain Monte Carlo
studying approximations in algorithms

Rocco Caprio
joint work with Adam M Johansen

January 2024

WARWICK

https://arxiv.org/abs/2310.03853

Consider the problem of approximating an integral

$$I := \mu(f) = \int f(x)\mu(\mathrm{d}x)$$

where μ is a probability distribution and f is some function.

Consider the problem of approximating an integral

$$I := \mu(f) = \int f(x)\mu(\mathrm{d}x)$$

where μ is a probability distribution and f is some function.

Markov chain Monte Carlo algorithms are based on the construction of a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ with transition probabilities

$$\mathbb{P}(X_n \in A | X_{n-1} = x) = P_\mu(x, A)$$

μ here is the *invariant distribution* of the Markov Chain P_μ:

$$P_\mu(\mu, A) = \int \mu(\mathrm{d}x) P_\mu(x, A) = \mu(A)$$

If P_μ is good enough, we can conclude $I_n := n^{-1}\sum_{i=1}^n f(X_i) \to I$.

Algorithm 1 Hastings algorithm

Starting with $X_0 = x$. For $i = 1, \ldots, n$,

- Draw $Y \sim Q(X_{i-1}, \cdot)$
- Compute $r_\mu(X_{i-1}, Y) := \frac{\mu(Y)Q(Y,X_{i-1})}{\mu(X_{i-1})Q(X_{i-1},Y)}$.
- Set $X_i = Y$ with probability $g(r_\mu(X_{i-1}, Y))$, otherwise set $X_i = X_{i-1}$.

Where $g(x) = xg(1/x)$, $g(x) \leq 1$ is an acceptance/balancing function. With $g(x) = \min(1, x)$ this is the Metropolis-Hastings algorithm.

- In this work, we study some calculus-type methods and tools to compare Markov kernels with different invariant distributions, say P_μ and P_ν;

- In this work, we study some calculus-type methods and tools to compare Markov kernels with different invariant distributions, say P_μ and P_ν;
- In particular, we will check that *if the derivative in the invariant distribution of P is bounded*, then P_μ and P_ν *move alike* if μ and ν are close;

- In this work, we study some calculus-type methods and tools to compare Markov kernels with different invariant distributions, say P_μ and P_ν;
- In particular, we will check that *if the derivative in the invariant distribution of P is bounded*, then P_μ and P_ν *move alike* if μ and ν are close;
- This provides a natural theoretical framework to analyze approximation-based algorithms.

Markov families

There are usually many possible Markov chains for each given invariant distribution. We first need to restrict the space of Markov kernels.

There are usually many possible Markov chains for each given invariant distribution. We first need to restrict the space of Markov kernels.

Definition

A **family of Markov kernels** $\{P_\star\}$ is a collection of Markov kernels indexed by an open, convex subset of invariant distributions.

If for some $\mu, \nu \in \mathcal{P}(\mathsf{X})$, $P_\mu, P_\nu \in \{P_\star\}$, then the curve $\mu_t := (1-t)\mu + t\nu$ interpolating μ and ν satisfies $P_{\mu_t} \in \{P_\star\}$.

Your best friends are Markov families!

Example

The **Hastings** family

$$P_\mu(x,f) := \int f(y)Q(x,\mathrm{d}y)g(r_\mu(x,y)) + f(x)(1-\int Q(x,\mathrm{d}y)g(r_\mu(x,y)))$$

with $r_\mu(x,y) := \frac{\mu(y)Q(y,x)}{\mu(x)Q(x,y)}$, $g(x) = xg(1/x)$, $g(x) \leq 1$.

Many other choices are possible.
Gibbs, **MALA**, **Metropolis-in-Gibbs** families...

Derivative in the invariant distribution

Let $\{P_\star\}$ be a Markov family. We define the derivative of P by first fixing a starting distribution and a test function. Consider the functional $P_\cdot(\rho, f) : \mu \in \mathcal{P}(\mathsf{X}) \mapsto P_\mu(\rho, f) \in \mathbb{R}$ for some fixed (ρ, f).

Let $\{P_\star\}$ be a Markov family. We define the derivative of P by first fixing a starting distribution and a test function. Consider the functional $P_\cdot(\rho, f) : \mu \in \mathcal{P}(\mathsf{X}) \mapsto P_\mu(\rho, f) \in \mathbb{R}$ for some fixed (ρ, f).

Definition

The derivative of $P_\cdot(\rho, f)$ in the invariant distribution at $\mu \in \mathcal{P}_\lambda(\mathsf{X})$ is the functional $\partial_\pi P_\mu(\rho, f)[\cdot] : \mathcal{M}_{\lambda,0}(\mathsf{X}) \mapsto \mathbb{R}$ such that for all $\nu \in \mathcal{P}_\lambda(\mathsf{X})$,

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{\mu + t(\nu - \mu)}(\rho, f)\bigg|_{t=0} = \partial_\pi P_\mu(\rho, f)[\nu - \mu].$$

Where $\mathcal{M}_{\lambda,0}(\mathsf{X}) := \{\text{set of 0-mass measures with Lebesgue-density}\}$

Let $\{P_\star\}$ be a Markov family. We define the derivative of P by first fixing a starting distribution and a test function. Consider the functional $P_\cdot(\rho, f) : \mu \in \mathcal{P}(\mathsf{X}) \mapsto P_\mu(\rho, f) \in \mathbb{R}$ for some fixed (ρ, f).

Definition

The derivative of $P_\cdot(\rho, f)$ in the invariant distribution at $\mu \in \mathcal{P}_\lambda(\mathsf{X})$ is the functional $\partial_\pi P_\mu(\rho, f)[\cdot] : \mathcal{M}_{\lambda,0}(\mathsf{X}) \mapsto \mathbb{R}$ such that for all $\nu \in \mathcal{P}_\lambda(\mathsf{X})$,

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{\mu+t(\nu-\mu)}(\rho, f)\bigg|_{t=0} = \partial_\pi P_\mu(\rho, f)[\nu - \mu].$$

Where $\mathcal{M}_{\lambda,0}(\mathsf{X}) := \{\text{set of 0-mass measures with Lebesgue-density}\}$

The derivative of $P_\cdot(\rho, \cdot)$ is defined in the oblivious way: it is the operator

$$\partial_\pi P_\mu(\rho, \cdot)[\cdot] : f \in \{\text{some set}\} \mapsto \partial_\pi P_\mu(\rho, f)[\cdot]$$

.

For $\rho \in \mathcal{P}_\lambda(\mathsf{X})$, the action of the functional $\partial_\pi P_\mu(\rho, f)$ will be expressible in *integral form*: for some function $\partial_\pi P_\mu(\rho, f)(\cdot) : \mathsf{X} \mapsto \mathbb{R}$

$$\partial_\pi P_\mu(\rho, f)[\nu - \mu] = \int (\nu(y) - \mu(y))\partial_\pi P_\mu(\rho, f)(y)\lambda(\mathrm{d}y) \quad \text{and}$$
$$\mu(\partial_\pi P_\mu(\rho, f)(\cdot)) = 0.$$

$\partial_\pi P_\mu(\rho, f)(\cdot)$, if it exists, is the *first variation* of $P_\cdot(\rho, f)$. We call it the *density* of the derivative in the invariant distribution.

For $\rho \in \mathcal{P}_\lambda(\mathsf{X})$, the action of the functional $\partial_\pi P_\mu(\rho, f)$ will be expressible in *integral form*: for some function $\partial_\pi P_\mu(\rho, f)(\cdot) : \mathsf{X} \mapsto \mathbb{R}$

$$\partial_\pi P_\mu(\rho, f)[\nu - \mu] = \int (\nu(y) - \mu(y)) \partial_\pi P_\mu(\rho, f)(y) \lambda(\mathrm{d}y) \quad \text{and}$$
$$\mu(\partial_\pi P_\mu(\rho, f)(\cdot)) = 0.$$

$\partial_\pi P_\mu(\rho, f)(\cdot)$, if it exists, is the *first variation* of $P_\cdot(\rho, f)$. We call it the *density* of the derivative in the invariant distribution.

However, for $\rho \notin \mathcal{P}_\lambda(\mathsf{X})$ (e.g. $\rho = \delta_x$), $\partial_\pi P_\mu(\rho, f)$ *will have no density*:

$$\partial_\pi P_\mu(x, f)[\nu - \mu] = (\text{a } \textit{density} \text{ part}) + (\text{a } \textit{singular} \text{ part})$$

Getting to know $\partial_\pi P_\mu(\rho, f)$

- $\partial_\pi P_\mu(\rho, \cdot)$ describes how $P_\mu(\rho, \cdot)$ changes when we perturb the distribution μ infinitesimally by a 0-mass measure.

$\Rightarrow$ If $\partial_\pi P_\mu(\rho, \cdot)$ is *large*, $P(\rho, \cdot)$ is not *robust* to changes in the invariant.

Getting to know $\partial_\pi P_\mu(\rho, f)$

- $\partial_\pi P_\mu(\rho, \cdot)$ describes how $P_\mu(\rho, \cdot)$ changes when we perturb the distribution μ infinitesimally by a 0-mass measure.

  $\Rightarrow$ If $\partial_\pi P_\mu(\rho, \cdot)$ is *large*, $P(\rho, \cdot)$ is not *robust* to changes in the invariant.

- Interestingly, it always hold
$$\partial_\pi P_\mu(\mu, f)(y) = f(y) - P_\mu(y, f).$$
  (minus) the generator of P! So...

Getting to know $\partial_\pi P_\mu(\rho, f)$

- $\partial_\pi P_\mu(\rho, \cdot)$ describes how $P_\mu(\rho, \cdot)$ changes when we perturb the distribution μ infinitesimally by a 0-mass measure.

  $\Rightarrow$ If $\partial_\pi P_\mu(\rho, \cdot)$ is *large*, $P(\rho, \cdot)$ is not *robust* to changes in the invariant.

- Interestingly, it always hold

$$\partial_\pi P_\mu(\mu, f)(y) = f(y) - P_\mu(y, f).$$

  (minus) the generator of P! So...

Theorem (Ergodic Theorem with derivatives)

Suppose that $\{X_n\}_{n\in\mathbb{N}}$ is an aperiodic, μ-irreducible Markov Chain with transition probabilities P_μ and invariant distribution μ. If there exists some petite set C, some $b < \infty$ and a non-negative finite function f bounded on C such that

$$-\partial_\pi P_\mu(\mu, f)(x) \leq -1 + b1_C(x), \quad x \in \mathsf{X}$$

whenever such kernel derivative exists, then for all $x \in \mathsf{X}$, as $k \to \infty$,

$$\left\| P_\mu^k(x, \cdot) - \mu \right\|_{\mathrm{tv}} \to 0.$$

Consider the Hastings family. Let $\mathcal{W}_H := \{\rho \in \mathcal{P}_\lambda(\mathsf{X}) : \rho/\mu^2 \text{ is bounded}\}$ – and assume that g is differentiable.

Proposition

The Hastings kernel $P_\cdot(\rho, \cdot)$ is differentiable in the invariant distribution at μ for all $\rho \in \mathcal{W}_H$. The derivative $\partial_\pi P_\mu(\rho, \cdot)$ admits an integral representation, with its density given by

$$\begin{aligned}\partial_\pi P_\mu(\rho, f)(y) = &\int (f(y) - f(z))\frac{\rho(z)}{\mu(z)} g'(r_\mu(z, y))Q(y, \mathrm{d}z) \\ &- \frac{\rho(y)}{\mu(y)^2}\int (f(z) - f(y))q(z, y)g'(r_\mu(y, z))\mu(\mathrm{d}z).\end{aligned}$$

Example: the Hastings family

Consider the Hastings family. Let $\mathcal{W}_H := \{\rho \in \mathcal{P}_\lambda(\mathsf{X}) : \rho/\mu^2 \text{ is bounded}\}$ – and assume that g is differentiable.

Proposition

The Hastings kernel $P_\cdot(\rho, \cdot)$ is differentiable in the invariant distribution at μ for all $\rho \in \mathcal{W}_H$. The derivative $\partial_\pi P_\mu(\rho, \cdot)$ admits an integral representation, with its density given by

$$\partial_\pi P_\mu(\rho, f)(y) = \int (f(y) - f(z)) \frac{\rho(z)}{\mu(z)} g'(r_\mu(z, y)) Q(y, \mathrm{d}z) - \frac{\rho(y)}{\mu(y)^2} \int (f(z) - f(y)) q(z, y) g'(r_\mu(y, z)) \mu(\mathrm{d}z).$$

- This rules out Metropolis-Hastings but don't worry.

Proposition (cont.)

Under some regularity conditions, $P_\cdot(x, \cdot)$ is also differentiable for all $x \in \mathsf{X}$. $\partial_\pi P_\mu(x, \cdot)$ is given by

$$\partial_\pi P_\mu(x, f)[\nu - \mu] = \underbrace{\int (f(y) - f(x)) \frac{g'(r_\mu(x, y)) q(y, x)}{\mu(x)} (\nu - \mu)(\mathrm{d}y)}_{\textit{density part}}$$

$$\underbrace{- \frac{(\nu - \mu)(x)}{\mu(x)^2} \int (f(y) - f(x)) q(y, x) g'(r_\mu(x, y)) \mu(\mathrm{d}y)}_{\textit{singular part}} .$$

Fundamental theorem of MCMC Calculus

Theorem

For all ρ *such that* $P_{\cdot}(\rho, \cdot)$ *is differentiable at* $\mu_t := (1-t)\mu + t\nu$ *for all* $t \in [0,1]$

$$P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot) = \int_0^1 \partial_\pi P_{\mu_t}(\rho, \cdot)[\nu - \mu] \, \mathrm{d}t.$$

Bounded derivatives

We want to say *bounded derivative* and μ, ν *close* $\Rightarrow$ P_μ and P_ν *move alike*.

Bounded derivatives

We want to say *bounded derivative* and μ, ν *close* $\Rightarrow P_\mu$ and P_ν *move alike*.

Definition

We say that a differentiable Markov kernel $P_\cdot(\rho, \cdot)$ has a bounded derivative at μ towards ν if there exist constants $M_{1,\rho}, M_{2,\rho} < \infty$ such that

$$\left\| P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot) \right\|_{\mathrm{tv}} \leq M_{1,\rho} \|\mu - \nu\|_{\mathrm{tv}} + M_{2,\rho} \rho(|\mu - \nu|). \tag{1}$$

We want to say *bounded derivative* and μ, ν *close* $\Rightarrow$ P_μ and P_ν *move alike*.

Definition

We say that a differentiable Markov kernel $P_\cdot(\rho, \cdot)$ has a bounded derivative at μ towards ν if there exist constants $M_{1,\rho}, M_{2,\rho} < \infty$ such that

$$\left\| P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot) \right\|_{\mathrm{tv}} \leq M_{1,\rho} \|\mu - \nu\|_{\mathrm{tv}} + M_{2,\rho}\rho(|\mu - \nu|). \tag{1}$$

- $d(\mu, \nu) = \|\mu - \nu\|_{\mathrm{tv}} + \rho(|\mu - \nu|)$ is a metric, so this could be written as Lipschitz continuity w.r.t. d, with constant given by $\max(M_{1,\rho}, M_{2,\rho})$. However, (1) separates contributions.

We want to say *bounded derivative* and μ, ν *close* $\Rightarrow$ P_μ and P_ν *move alike*.

Definition

We say that a differentiable Markov kernel $P_\cdot(\rho, \cdot)$ has a bounded derivative at μ towards ν if there exist constants $M_{1,\rho}, M_{2,\rho} < \infty$ such that

$$\left\| P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot) \right\|_{\mathrm{tv}} \leq M_{1,\rho} \|\mu - \nu\|_{\mathrm{tv}} + M_{2,\rho} \rho(|\mu - \nu|). \tag{1}$$

- $d(\mu, \nu) = \|\mu - \nu\|_{\mathrm{tv}} + \rho(|\mu - \nu|)$ is a metric, so this could be written as Lipschitz continuity w.r.t. d, with constant given by $\max(M_{1,\rho}, M_{2,\rho})$. However, (1) separates contributions.
- When $\rho \in \mathcal{P}_\lambda(\mathsf{X})$, this definition is equivalent to the existence of a constant $M_\rho < \infty$ such that

  $$\left\| P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot) \right\|_{\mathrm{tv}} \leq M_\rho \|\mu - \nu\|_{\mathrm{tv}} .$$

  whereas if $\rho = \delta_x$, it becomes

  $$\left\| P_\mu(x, \cdot) - P_\nu(x, \cdot) \right\|_{\mathrm{tv}} \leq M_{1,x} \|\mu - \nu\|_{\mathrm{tv}} + M_{2,x} |\mu(x) - \nu(x)|.$$

Proposition

For Hastings kernels, if $P_\cdot(\rho, \cdot)$ is differentiable in $\mu_t = (1-t)\mu + t\nu$ for all $t \in [0,1]$ and a boundedness condition on the derivative holds, then $P_\cdot(\rho, \cdot)$ will have a bounded derivative in the invariant distribution at μ towards ν. The "mean-value" inequalities

$$\left\| P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot) \right\|_{\mathrm{tv}} \leq M_\rho \|\mu - \nu\|_{\mathrm{tv}}$$

$$\left\| P_\mu(x, \cdot) - P_\nu(x, \cdot) \right\|_{\mathrm{tv}} \leq M_{1,x} \|\mu - \nu\|_{\mathrm{tv}} + M_{2,x} |\mu(x) - \nu(x)|$$

hold, with some explicit values for the 'Lipschitz constants' $M_\rho, M_{1,x}, M_{2,x}$.

Proposition

For Hastings kernels, if $P_{\cdot}(\rho, \cdot)$ is differentiable in $\mu_t = (1-t)\mu + t\nu$ for all $t \in [0,1]$ and a boundedness condition on the derivative holds, then $P_{\cdot}(\rho, \cdot)$ will have a bounded derivative in the invariant distribution at μ towards ν. The "mean-value" inequalities

$$\left\|P_\mu(\rho, \cdot) - P_\nu(\rho, \cdot)\right\|_{\mathrm{tv}} \leq M_\rho \|\mu - \nu\|_{\mathrm{tv}}$$

$$\left\|P_\mu(x, \cdot) - P_\nu(x, \cdot)\right\|_{\mathrm{tv}} \leq M_{1,x} \|\mu - \nu\|_{\mathrm{tv}} + M_{2,x}|\mu(x) - \nu(x)|$$

hold, with some explicit values for the 'Lipschitz constants' $M_\rho, M_{1,x}, M_{2,x}$.

By considering $g_j(x) := (x + \cdots + x^j)/(1 + x + \cdots + x^j) \to \min(1, x)$ we can also obtain similar mean-value inequalities for the Metropolis-Hastings family.

- The Metropolis-Hastings kernel is an example of non-differentiable but Lipschitz mapping of its invariant distribution!

Example - Mean Value for Hastings kernels

These inequalities seem to be fairly tight!

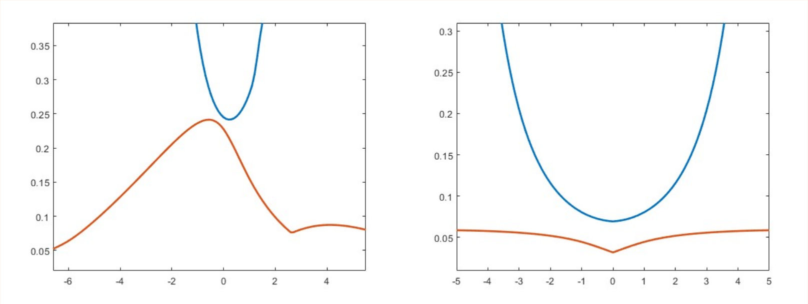

Figure: The mean value inequality $\left\| P_\mu(x,\cdot) - P_\nu(x,\cdot) \right\|_{\mathrm{tv}} \leq \|\mu - \nu\|_{\mathrm{tv}} M_{1,x} + |\mu(x) - \nu(x)| M_{2,x}$ for RW MH as a function of $x \in \mathsf{X}$

With a bounded derivative P_{μ_n} chain moves similarly to P_μ if μ_n and μ are close. An alternative is to minimize fluctuations P_{μ_n} around P_μ. Suppose that $\{\mu_n\}_{n\in\mathbb{N}} \in \mathcal{P}(\mathsf{X})$ are random, and that $n^{-1/2}[\mu_n - \mu](f) \Rightarrow N(0, v(f))$.

With a bounded derivative P_{μ_n} chain moves similarly to P_μ if μ_n and μ are close. An alternative is to minimize fluctuations P_{μ_n} around P_μ. Suppose that $\{\mu_n\}_{n\in\mathbb{N}} \in \mathcal{P}(\mathsf{X})$ are random, and that $n^{-1/2}[\mu_n - \mu](f) \Rightarrow N(0, v(f))$.

Proposition

If $P_\cdot(\rho, f)$ is differentiable, then

$$n^{-1/2}(P_{\mu_n}(\rho, f) - P_\mu(\rho, f)) \Rightarrow N\left(0, v(\partial_\pi P_\mu(\rho, f))\right).$$

The random fluctuations of $P_{\mu_n}(\rho, f)$ around $P_\mu(\rho, f)$ depend explicitly on the derivative $\partial_\pi P_\mu(\rho, f)$.

Applications

- The methods developed allow an easy comparison between Markov Chains of the same family with different invariant distributions.

- The methods developed allow an easy comparison between Markov Chains of the same family with different invariant distributions.
- It is natural to think of ν as an approximation of μ. In this sense we are investigating when an "approximated" Markov chain P_ν moves like the limiting P_μ, which is desirable.

- The methods developed allow an easy comparison between Markov Chains of the same family with different invariant distributions.
- It is natural to think of ν as an approximation of μ. In this sense we are investigating when an "approximated" Markov chain P_ν moves like the limiting P_μ, which is desirable.
- Roughly speaking (and under some other regularity conditions), if $P_\cdot(x, \cdot)$ has a bounded derivative in the invariant distribution at μ towards ν, and ν is somewhat close to μ, the "approximated" Markov chain P_ν will achieve the same asymptotic variance of P_μ *plus* an additional variability due the fluctuations of ν around μ.

Suppose that we can write $\mu = \Phi(\eta)$, for some transformation $\Phi : \mathcal{P}(\mathsf{X}) \mapsto \mathcal{P}(\mathsf{X})$.

Algorithm 2 Sequential MCMC

1. Simulate $Y'_{i+1} \sim P_\eta(Y'_i, \cdot)$ for $i = 0, \ldots, n$; set $\eta_n := n^{-1} \sum_{i=0}^{n} \delta_{Y'_i}$.
2. Simulate $Y_{i+1} \sim P_{\Phi(\eta_n)}(Y_i, \cdot)$ for $i = 0, \ldots, n$.

Typically Φ is the Boltzmann-Gibbs transformation associated to some Feynman-Kac model, and this algorithm is popular in the context of particle filtering where

$$\underbrace{\mu}_{\text{Predictive distribution at later time}} = \Phi(\underbrace{\eta}_{\text{Predictive distribution at earlier time}})$$

See Berzuini et al., 1997; Golightly and Wilkinson, 2006; Septier et al., 2009; Li et al., 2023; Finke, Doucet, and Johansen, 2020...

Denote with σ^2 the asymptotic variance achieved by the Markov chain P_μ and let $\mu_n := \Phi(\eta_n)$.

Theorem

Assume that

- *$P_\cdot$ has a bounded derivative in the invariant distribution at μ towards every μ_n;*
- *$\mu_n(x) \to \mu(x)$ for all $x \in \mathsf{X}$ and $n^{-1/2}[\mu_n - \mu](f) \Rightarrow N(0, v(f))$.*
- *P satisfies an uniform drift and minorization condition.*

Then,

$$n^{-1/2} \sum_{i=0}^{n} f(Y_i) - \mu(f) \Rightarrow N(0, \sigma^2(f) + v^2(f)).$$

Extends the theoretical guarantees for sMCMC of Finke, Doucet, and Johansen, 2020!

Conclusions

- We studied methods to compare Markov chains using different invariant distributions commonly used in Markov chain Monte Carlo using a calculus-type approach;

Conclusions

- We studied methods to compare Markov chains using different invariant distributions commonly used in Markov chain Monte Carlo using a calculus-type approach;
- Via the "mean-value" inequalities, we studied when we can expect two Markov chains with different but close invariant distributions to behave similarly;

Conclusions

- We studied methods to compare Markov chains using different invariant distributions commonly used in Markov chain Monte Carlo using a calculus-type approach;
- Via the "mean-value" inequalities, we studied when we can expect two Markov chains with different but close invariant distributions to behave similarly;
- Other than being of theoretical interest per se, these methods are naturally suited to the study of approximation-based algorithms. In general, whenever we can control how good the approximation of the invariant is, there should be room to use these tools;

Conclusions

- We studied methods to compare Markov chains using different invariant distributions commonly used in Markov chain Monte Carlo using a calculus-type approach;
- Via the "mean-value" inequalities, we studied when we can expect two Markov chains with different but close invariant distributions to behave similarly;
- Other than being of theoretical interest per se, these methods are naturally suited to the study of approximation-based algorithms. In general, whenever we can control how good the approximation of the invariant is, there should be room to use these tools;
- The methods and the strategy developed can be adapted to other contexts. A calculus with proposals and/or other distributions the kernel depends on?

Conclusions

- We studied methods to compare Markov chains using different invariant distributions commonly used in Markov chain Monte Carlo using a calculus-type approach;
- Via the "mean-value" inequalities, we studied when we can expect two Markov chains with different but close invariant distributions to behave similarly;
- Other than being of theoretical interest per se, these methods are naturally suited to the study of approximation-based algorithms. In general, whenever we can control how good the approximation of the invariant is, there should be room to use these tools;
- The methods and the strategy developed can be adapted to other contexts. A calculus with proposals and/or other distributions the kernel depends on?
- Many directions to explore, in terms of development of the theory (second order derivatives, Taylor-type theorems...) and/or employing these or similar results in interesting contexts.

References I

Ascolani, Filippo and Giacomo Zanella (2023). "Complexity of Gibbs samplers through Bayesian asymptotics". In: *arXiv:2304.06993*.

Berzuini, Carlo et al. (1997). "Dynamic conditional independence models and Markov chain Monte Carlo methods". In: *Journal of the American Statistical Association* 92, pp. 1403–1412.

Caprio, Rocco and Adam M Johansen (2023). "A calculus for Markov chain Monte Carlo: studying approximations in algorithms". In: *arXiv preprint arXiv:2310.03853*.

Del Moral, Pierre (2004). *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. New York: Springer-Verlag.

Finke, Axel, Arnaud Doucet, and Adam M Johansen (2020). "Limit theorems for sequential MCMC methods". In: *Advances in Applied Probability* 52, pp. 377–403.

Golightly, Andrew and Darren J Wilkinson (2006). "Bayesian sequential inference for nonlinear multivariate diffusions". In: *Statistics and Computing* 16, pp. 323–338.

Huber, Peter J and Elvezio M Ronchetti (2009). *Robust statistics*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Li, Qing et al. (2023). "An Adaptive and Scalable Multi-Object Tracker Based on the Non-Homogeneous Poisson Process". In: *IEEE Transactions on Signal Processing* 71, pp. 105–120.

Septier, François et al. (2009). "On MCMC-based particle methods for Bayesian filtering: Application to multitarget tracking". In: *2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 360–363.