Gaussian approximation and Output Analysis for high-dimensional MCMC

Ardjen Pengel

TU Delft

Delft University of Technology, The Netherlands

April 26, 2024

Joint work with

- Jun Yang; University of Copenhagen

- Zhou Zhou; University of Toronto

Talk outline

- Brief introduction Markov Chain Monte Carlo
- Gaussian approximation results
- Uncertainty Quantification high-dimensional MCMC

Markov Chain Monte Carlo

- Goal: Simulate π, a probability distribution of interest, and often interested in estimating

$$\pi(f) := \mathbb{E}_{\pi}[f(X)] = \int f(x)\pi(dx)$$

with

$$f : \mathbb{R}^n \to \mathbb{R}^d$$

Markov Chain Monte Carlo

- Goal: Simulate π, a probability distribution of interest, and often interested in estimating

$$\pi(f) := \mathbb{E}_\pi[f(X)] = \int f(x)\pi(dx)$$

with

$$f : \mathbb{R}^n \to \mathbb{R}^d$$

- MCMC methods construct a Markov chain $X = (X_t)_{t\in\mathbb{N}}$ such that the long-run behavior of the process is described by π
- Has applications in fields ranging from statistics to physics
- We consider the high-dimensional setting where both n and d can be large

Introduction MCMC

- Markov chain $X = (X_t)_{t \in \mathbb{N}}$ with stationary distribution π
- Convergence to stationary distribution

$$\sup_{B \in \mathcal{B}} |\mathbb{P}_{\pi_0}(X_T \in B) - \pi(B)| \to 0 \text{ as } T \to \infty.$$

Introduction MCMC

- Markov chain $X = (X_t)_{t\in\mathbb{N}}$ with stationary distribution π
- Convergence to stationary distribution

$$\sup_{B\in\mathcal{B}} |\mathbb{P}_{\pi_0}(X_T \in B) - \pi(B)| \to 0 \text{ as } T \to \infty.$$

- Ergodic LLN

$$\hat{\pi}_T(f) := \frac{1}{T}\sum_{k=1}^{T} f(X_k) \xrightarrow{a.s.} \int f(x)\pi(dx) =: \pi(f), \text{ as } T \to \infty.$$

Introduction MCMC

- Markov chain $X = (X_t)_{t \in \mathbb{N}}$ with stationary distribution π
- Convergence to stationary distribution

$$\sup_{B \in \mathcal{B}} |\mathbb{P}_{\pi_0}(X_T \in B) - \pi(B)| \to 0 \text{ as } T \to \infty.$$

- Ergodic LLN

$$\hat{\pi}_T(f) := \frac{1}{T} \sum_{k=1}^{T} f(X_k) \xrightarrow{a.s.} \int f(x)\pi(dx) =: \pi(f), \text{ as } T \to \infty.$$

- MCMC is approximate inference

MCMC Convergence Diagnostics

Markov chain $X = (X_t)_{t \in \mathbb{N}}$ with stationary distribution π

Practical Questions

- **Q1:** When is it reasonable to assume that our sampling algorithm is in equilibrium?
- **Q2:** How long do we need to run our sampling algorithm?

MCMC Convergence Diagnostics

Markov chain $X = (X_t)_{t\in\mathbb{N}}$ with stationary distribution π

Practical Questions

- **Q1:** When is it reasonable to assume that our sampling algorithm is in equilibrium?
- **Q2:** How long do we need to run our sampling algorithm?
- **Q3:** How does the answer of Q1 and Q2 depend on the dimension?

MCMC Convergence Diagnostics

Markov chain $X = (X_t)_{t \in \mathbb{N}}$ with stationary distribution π

Practical Questions

- **Q1:** When is it reasonable to assume that our sampling algorithm is in equilibrium?
- **Q2:** How long do we need to run our sampling algorithm?
- **Q3:** How does the answer of Q1 and Q2 depend on the dimension?

Theoretical Answers

- **A1:** Quantitative bounds of convergence to stationarity

MCMC Convergence Diagnostics

Markov chain $X = (X_t)_{t \in \mathbb{N}}$ with stationary distribution π

Practical Questions

- **Q1:** When is it reasonable to assume that our sampling algorithm is in equilibrium?
- **Q2:** How long do we need to run our sampling algorithm?
- **Q3:** How does the answer of Q1 and Q2 depend on the dimension?

Theoretical Answers

- **A1:** Quantitative bounds of convergence to stationarity
- **A2:** Uncertainty Quantification and Termination criteria for simulation output

Markov chain CLT

Let X be polynomially ergodic of sufficiently high order.
Then for all $f : E \to \mathbb{R}^d$ with $\pi(\|f\|^{2+\varepsilon}) < \infty$

- CLT

$$\frac{1}{\sqrt{T}} \sum_{k=1}^{T} (f(X_k) - \mu(f)) \xrightarrow{w} \mathcal{N}_d(0, \Sigma_f).$$

- FCLT

$$\left(\frac{1}{\sqrt{T}} \sum_{k=0}^{\lfloor Tt \rfloor} (f(X_k) - \mu(f)) \right)_t \xrightarrow{w} \Sigma_f^{1/2} W \text{ as } n \to \infty,$$

with

$$\Sigma_f = \sum_{k=0}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_k)) + \sum_{k=0}^{\infty} \mathrm{Cov}_\pi(f(X_k), f(X_0))$$

W is a d-dimensional Brownian motion

Motivating example: Uncertainty Quantification

We need to estimate asymptotic variance Σ_f

- Confidence ellipsoid for uncertainty quantification

$$C_T = \{\theta \in \mathbb{R}^d : T(\hat{\pi}_T(f) - \theta)^\top \hat{\Sigma}_T^{-1}(\hat{\pi}_T(f) - \theta) < q_\alpha\}$$

Motivating example: Uncertainty Quantification

We need to estimate asymptotic variance Σ_f

- Confidence ellipsoid for uncertainty quantification

$$C_T = \{\theta \in \mathbb{R}^d : T(\hat{\pi}_T(f) - \theta)^\top \hat{\Sigma}_T^{-1}(\hat{\pi}_T(f) - \theta) < q_\alpha\}$$

- Stopping rules: Terminate simulation when confidence ellipsoid has desired volume.

$$T_1(\varepsilon) = \inf\{T > 0 : \mathrm{Vol}(C(T))^{1/d} + 1_{\{T > T^*(\varepsilon,d,n)\}} \leq \varepsilon\}.$$

Motivating example: Uncertainty Quantification

We need to estimate asymptotic variance Σ_f

- Confidence ellipsoid for uncertainty quantification

$$C_T = \{\theta \in \mathbb{R}^d : T(\hat{\pi}_T(f) - \theta)^\top \hat{\Sigma}_T^{-1}(\hat{\pi}_T(f) - \theta) < q_\alpha\}$$

- Stopping rules: Terminate simulation when confidence ellipsoid has desired volume.

$$T_1(\varepsilon) = \inf\{T > 0 : \mathrm{Vol}(C(T))^{1/d} + 1_{\{T > T^*(\varepsilon,d,n)\}} \leq \varepsilon\}.$$

- Analysis of $T_1(\varepsilon)$ and $\hat{\Sigma}_f$ requires refinements FCLT

Gaussian approximation (GA)

On some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we can construct X and Brownian motion W such that

$$\left\| \sum_{t=0}^{T} [f(X_t) - \pi(f)] - \Sigma_f^{1/2} W(T) \right\| = O(\psi_T)$$

almost surely or in probability.

$$\limsup_{T \to \infty} \frac{1}{\psi_T} \left\| \sum_{t=0}^{T} [f(X_t) - \pi(f)] - \Sigma_f^{1/2} W(T) \right\| < C$$

almost surely or in probability.

- Weak GA: convergence in probability
- Strong GA: convergence almost surely

Sequential Gaussian approximation

On some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we can construct X and Brownian motion W such that

$$\sup_{0<s\leq T} \left\| \sum_{t=0}^{\lfloor s \rfloor} [f(X_t) - \pi(f)] - \Sigma_f^{1/2} W(s) \right\| = O_P(\psi_T)$$

- Implies convergence rate FCLT

$$d_{PL}\left(\frac{1}{\sqrt{T}} \sum_{k=0}^{\lfloor Ts \rfloor} [f(X_k) - \pi(f)], W \right) \approx O\left(\frac{\psi_T}{\sqrt{T}} \right)$$

Output Analysis: Current Results

- Existing MCMC GA rates are not dimension-dependent [1][2][3]
- Empirical findings: multivariate rates not appropriate for output analysis in high-dimensional setting

1 J. Flegal and G. Jones (2010). "Batch means and spectral variance estimators in Markov chain Monte Carlo". In: *The Annals of Statistics* 38.2, pp. 1034–1070

2 F. Merlevède, E. Rio, et al. (2015). "Strong approximation for additive functionals of geometrically ergodic Markov chains". In: *Electronic Journal of Probability* 20

3 A. Banerjee and D. Vats (2022). "Multivariate strong invariance principles in Markov chain Monte Carlo". In: *arXiv preprint arXiv:2211.06855*

4 P. W. Glynn and W. Whitt (1992). "The asymptotic validity of sequential stopping rules for stochastic simulations". In: *The Annals of Applied Probability* 2.1, pp. 180–198

5 D. Vats, J. M. Flegal, and G. L. Jones (2019). "Multivariate output analysis for Markov chain Monte Carlo". In: *Biometrika* 106.2, pp. 321–337

Output Analysis: Current Results

- Existing MCMC GA rates are not dimension-dependent [1,2,3]
- Empirical findings: multivariate rates not appropriate for output analysis in high-dimensional setting
- Analysis of termination criteria does not take dimension into account [4] [5].

---

1 J. Flegal and G. Jones (2010). "Batch means and spectral variance estimators in Markov chain Monte Carlo". In: *The Annals of Statistics* 38.2, pp. 1034–1070

2 F. Merlevède, E. Rio, et al. (2015). "Strong approximation for additive functionals of geometrically ergodic Markov chains". In: *Electronic Journal of Probability* 20

3 A. Banerjee and D. Vats (2022). "Multivariate strong invariance principles in Markov chain Monte Carlo". In: *arXiv preprint arXiv:2211.06855*

4 P. W. Glynn and W. Whitt (1992). "The asymptotic validity of sequential stopping rules for stochastic simulations". In: *The Annals of Applied Probability* 2.1, pp. 180–198

5 D. Vats, J. M. Flegal, and G. L. Jones (2019). "Multivariate output analysis for Markov chain Monte Carlo". In: *Biometrika* 106.2, pp. 321–337

Output Analysis: Current Results

- Existing MCMC GA rates are not dimension-dependent [1][2][3]
- Empirical findings: multivariate rates not appropriate for output analysis in high-dimensional setting
- Analysis of termination criteria does not take dimension into account [4] [5].
- How does the GA convergence rate depend on the dimensions n, d of the target distribution and feature vector respectively?

---

1 J. Flegal and G. Jones (2010). "Batch means and spectral variance estimators in Markov chain Monte Carlo". In: *The Annals of Statistics* 38.2, pp. 1034–1070

2 F. Merlevède, E. Rio, et al. (2015). "Strong approximation for additive functionals of geometrically ergodic Markov chains". In: *Electronic Journal of Probability* 20

3 A. Banerjee and D. Vats (2022). "Multivariate strong invariance principles in Markov chain Monte Carlo". In: *arXiv preprint arXiv:2211.06855*

4 P. W. Glynn and W. Whitt (1992). "The asymptotic validity of sequential stopping rules for stochastic simulations". In: *The Annals of Applied Probability* 2.1, pp. 180–198

5 D. Vats, J. M. Flegal, and G. L. Jones (2019). "Multivariate output analysis for Markov chain Monte Carlo". In: *Biometrika* 106.2, pp. 321–337

MCMC: Foster-Lyapunov drift conditions

Geometric drift condition: Let $V : \mathbb{R}^n \to \mathbb{R}^+, \lambda \in (0, 1), 0 < b < \infty$, and

$$\mathbb{E}[V(X_{t+1})|X_t] \leq \lambda V(X_t) + b1_C(X_t)$$

- V is the energy function: takes low values in high π–probability regions and high values in low π–probability regions.

MCMC: Foster-Lyapunov drift conditions

Geometric drift condition: Let $V : \mathbb{R}^n \to \mathbb{R}^+, \lambda \in (0,1), 0 < b < \infty$, and

$$\mathbb{E}[V(X_{t+1})|X_t] \leq \lambda V(X_t) + b1_C(X_t)$$

- V is the energy function: takes low values in high π–probability regions and high values in low π–probability regions.
- Drift towards set C: $\mathbb{E}_x \tau_C \leq V(x)$ for $x \notin C$.

MCMC: Foster-Lyapunov drift conditions

Geometric drift condition: Let $V : \mathbb{R}^n \to \mathbb{R}^+, \lambda \in (0,1), 0 < b < \infty$, and

$$\mathbb{E}[V(X_{t+1})|X_t] \leq \lambda V(X_t) + b1_C(X_t)$$

- V is the energy function: takes low values in high π–probability regions and high values in low π–probability regions.
- Drift towards set C: $\mathbb{E}_x \tau_C \leq V(x)$ for $x \notin C$.

Polynomial drift condition: Let $\eta \in (0,1)$, $0 < b, c < \infty$, and

$$\mathbb{E}[V(X_{t+1})|X_t] \leq V(X_t) - cV(X_t)^\eta + b1_C(X_t)$$

MCMC: Minorization condition

One-step minorization

$$P(x, dy) = \mathbb{P}(X_{t+1} \in dy | X_t = x) \geq \alpha\nu(dy), \;\; x \in C$$

MCMC: Minorization condition

One-step minorization

$$P(x, dy) = \mathbb{P}(X_{t+1} \in dy | X_t = x) \geq \alpha\nu(dy), \;\; x \in C$$

- Splitting transition kernel: $P(x, dy) = \alpha\nu(dy) + (1 - \alpha)R(x, dy)$
- Every time the process is in C with probability α move independent of its past

MCMC: Minorization condition

One-step minorization

$$P(x, dy) = \mathbb{P}(X_{t+1} \in dy | X_t = x) \geq \alpha\nu(dy), \; x \in C$$

- Splitting transition kernel: $P(x, dy) = \alpha\nu(dy) + (1 - \alpha)R(x, dy)$
- Every time the process is in C with probability α move independent of its past

m_0-step minorization

$$P^{m_0}(x, dy) \geq \alpha\nu(dy), \; x \in C$$

- Every time the process is in C with probability α: $X_{t+m_0} \perp\!\!\!\perp \mathcal{F}_t$

Convergence Complexity MCMC

- How do the convergence properties of MCMC scale with n, d?

6 J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

7 Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Convergence Complexity MCMC

- How do the convergence properties of MCMC scale with n, d?
- Family of drift and minorisation conditions that remain stable as dimension n grows.

$$\limsup_{n\to\infty} \lambda_n < 1 \quad \text{as } n \to \infty.$$

- Gives dimension-dependent convergence rates to stationarity[6][7]

6 J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

7 Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Convergence Complexity MCMC

- How do the convergence properties of MCMC scale with n, d?
- Family of drift and minorisation conditions that remain stable as dimension n grows.

$$\limsup_{n\to\infty} \lambda_n < 1 \quad \text{as } n \to \infty.$$

- Gives dimension-dependent convergence rates to stationarity[6][7]

Central Question

How does the dimension affect the uncertainty quantification?

---

6 J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

7 Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Moment conditions

Let $f : \mathbb{R}^n \to \mathbb{R}^d$ with

- (A1): $\sup_{i \in \{1,\cdots,d\}} \pi(|f_i|^{p+\epsilon}) < \infty$ with $p > 2$
- (A2): $\sup_{i \in \{1,\cdots,d\}} \pi(e^{t|f_i|}) < \infty$ with $t > 0$

Weak Gaussian Approximation

Theorem

Assume that X satisfies a stable geometric drift condition

$$PV \leq \lambda_n V + b_n 1_C,$$

and a one-step minorization

Weak Gaussian Approximation

Theorem

Assume that X satisfies a stable geometric drift condition

$$PV \leq \lambda_n V + b_n 1_C,$$

and a one-step minorization Then we have

$$\left| \sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| = O_P\left(\psi_n^2 \left(\frac{\sigma_d}{\sigma_0} \right)^{1/2} d \log(d)\, T^{1/p} \right)$$

where

$$\psi_n = \alpha_n^{-1/p} \left(\frac{b_n}{\alpha_n(1-\lambda_0)} \right)^{1/p^2}$$

- (F)CLT requires for large p: $d = o(\sqrt{T})$

Strong Gaussian Approximation

Theorem

Assume that X satisfies a stable geometric drift condition

$$PV \leq \lambda_n V + b_n 1_C,$$

and an m_0-step minorization. Then we have

$$\left| \sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T \right| = O_P\left(\psi_n^2 \left(\frac{\sigma_d}{\sigma_0} \right)^{1/2} d \log(d) \, T^{\frac{1}{4} + \frac{1}{4(p-1)}} \right)$$

where

$$\psi_n = \alpha_n^{-1/p} m_0 \left(\frac{2b_n}{\alpha_n(1-\lambda_0)} \right)^{1/p^2}$$

- (F)CLT requires for large p: $d = o(T^{1/4})$

Gaussian Approximation Results

Theorem

Assume that X satisfies an m_0–step minorization and a polynomial drift condition

$$PV \leq V - cV^{\eta} + b1_C,$$

Then

Gaussian Approximation Results

Theorem

Assume that X satisfies an m_0–step minorization and a polynomial drift condition

$$PV \leq V - cV^{\eta} + b1_C,$$

Then

$$\left|\sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = O_P\left(\psi_n^2 \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} d\log(d)\, T^{\frac{1}{4}+\frac{1}{4(p_0-1)}}\right),$$

where

$$p_0 = \begin{cases} \frac{pq(\eta)}{p+q(\eta)+\varepsilon}, & \textit{if } \frac{2p}{3p-2} < \eta \leq a(p), \\ p, & \textit{if } \eta > a(p), \\ q(\eta) - \bar{\epsilon}, & \textit{if } \eta > 1/2 \ \ \textit{and A2 holds}, \end{cases}$$

with $q(\eta) = \eta/(1-\eta)$.

Gaussian Approximation Results

Theorem

Assume that X satisfies an m_0–step minorization and a polynomial drift condition

$$PV \leq V - cV^{\eta} + b1_C,$$

Then

$$\left|\sum_{t=1}^{T} f(X_t) - T\pi(f) - \Sigma_f^{1/2} W_T\right| = O_P\left(\psi_n^2 \left(\frac{\sigma_d}{\sigma_0}\right)^{1/2} d \log(d)\, T^{\frac{1}{4}+\frac{1}{4(p_0-1)}}\right),$$

where

$$p_0 = \begin{cases} \frac{pq(\eta)}{p+q(\eta)+\varepsilon}, & \textit{if } \frac{2p}{3p-2} < \eta \leq a(p), \\ p, & \textit{if } \eta > a(p), \\ q(\eta) - \bar{\epsilon}, & \textit{if } \eta > 1/2 \quad \textit{and A2 holds}, \end{cases}$$

with $q(\eta) = \eta/(1-\eta)$.

$$\psi_n = \alpha^{-1/p_0} m_0^{q/p_0^2} \left(1 + \frac{b}{c\alpha} + \frac{\upsilon_c + b}{1-\alpha}\right)^{1/p_0^2}$$

Gaussian approximation results; Extensions

- One-step minorization attains optimal KMT bound in sample size; results are applicable to
  - Gaussian and hierarchical models[8]
  - Bayesian probit models[9]

[8] J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

[9] Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Gaussian approximation results; Extensions

- One-step minorization attains optimal KMT bound in sample size; results are applicable to
  - Gaussian and hierarchical models[8]
  - Bayesian probit models[9]
- Approximation results also valid for continuous-time processes

8 J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

9 Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Gaussian approximation results; Extensions

- One-step minorization attains optimal KMT bound in sample size; results are applicable to
  - Gaussian and hierarchical models[8]
  - Bayesian probit models[9]
- Approximation results also valid for continuous-time processes

8 J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

9 Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Gaussian approximation results; Extensions

- One-step minorization attains optimal KMT bound in sample size; results are applicable to
  - Gaussian and hierarchical models[8]
  - Bayesian probit models[9]
- Approximation results also valid for continuous-time processes
- Dimension dependence can be improved

$$\psi_d = \sqrt{d}\pi(\|f\|^p)^{1/p}$$

  - Sparsity coordinates π
  - Bayesian regularisation

8 J. Yang and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500

9 Q. Qin and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347

Estimation of asymptotic variance

- Asymptotic variance

$$\Sigma_f = \sum_{k=0}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_k)) + \sum_{k=0}^{\infty} \mathrm{Cov}_\pi(f(X_k), f(X_0))$$

- Batch means: divide simulation output into k_T batches of size ℓ_T. Compute means

$$\hat{\pi}^j_{\ell_T}(f) = \frac{1}{\ell_T} \sum_{(j-1)\ell_T}^{i\ell_T} f(X_t), \text{ for } j = 1, \cdots, k_T.$$

$$\hat{\Sigma}^{BM}_T = \frac{\ell_T}{k_T - 1} \sum_{j=1}^{k_T} (\hat{\pi}^j_{\ell_T}(f) - \hat{\pi}_T(f))(\hat{\pi}^j_{\ell_T}(f) - \hat{\pi}_T(f))^\top$$

- How to choose ℓ_T?

Analysis of BM estimator

Assume that $\ell_{T,d} \to \infty$ such that $\ell_{T,d} \asymp \sqrt{T}\psi_{T,d,n}$ then

- Consistency $\hat{\Sigma}_T \xrightarrow{p} \Sigma_f$ as $T \to \infty$

Analysis of BM estimator

Assume that $\ell_{T,d} \to \infty$ such that $\ell_{T,d} \asymp \sqrt{T}\psi_{T,d,n}$ then

- Consistency $\hat{\Sigma}_T \xrightarrow{p} \Sigma_f$ as $T \to \infty$
- Larger approximation error $\psi_{T,d}$ requires larger batch size $\ell_{T,d}$
- Convergence rate GA corresponds to decay of autocovariance
- Asymptotic Normality of $\hat{\Sigma}_T$

Analysis of BM estimator

Assume that $\ell_{T,d} \to \infty$ such that $\ell_{T,d} \asymp \sqrt{T}\psi_{T,d,n}$

	one-step minorisation	**multi-step minorisation**
exponential drift	$\sqrt{d}\,T^{\frac{1}{2}+\frac{1}{p}}$	$\sqrt{d}\,T^{\frac{3}{4}+\frac{1}{4(p-1)}}$
polynomial drift	$\sqrt{d}\,T^{\frac{1}{2}+\frac{1}{p_0}}$	$\sqrt{d}\,T^{\frac{3}{4}+\frac{1}{4(p_0-1)}}$

Table: Batch size ℓ_T multivariate setting

Gaussian approximation

- Different convergence rates
  - Strong Gaussian approximation: $\bar{\psi}_d = d^{23/4}$ and $\bar{\Psi}_T = T^{1/p}$
  - Partial sum Gaussian approximation :

$$\sup_{0<s\leq T} \left\| \sum_{t=0}^{\lfloor s \rfloor} [f(X_t) - \pi(f)] - \Sigma_f^{1/2} W(s) \right\| = O_P(\psi_n \psi_d^* \Psi_T^*)$$

  with $\psi_d^* = d^{3/4}$ and $\Psi_T^* = T^{1/4}$

- Both can be used to give quantitative convergence bounds of termination criteria [10]

[10] P. W. Glynn and W. Whitt (1992). "The asymptotic validity of sequential stopping rules for stochastic simulations". In: *The Annals of Applied Probability* 2.1, pp. 180–198

MCMC Termination Criteria

- Termination time

$$T_1(\varepsilon) = \inf\{T > 0 : Vol(C(T))^{1/d} + 1_{\{T > T^*(\varepsilon,d,n)\}} \leq \varepsilon\}$$

- Choose simulation threshold T^*

	one-step minorisation	**multi-step minorisation**
geometric drift	$\left(\frac{1}{\varepsilon}\right)^{\frac{4p}{(p-2)}(1+\bar{\delta})}$	$\left(\frac{1}{\varepsilon}\right)^{\frac{8(p-1)}{(p-2)}(1+\bar{\delta})}$
polynomial drift	$\left(\frac{1}{\varepsilon}\right)^{\frac{4p_0}{(p_0-2)}(1+\bar{\delta})}$	$\left(\frac{1}{\varepsilon}\right)^{\frac{8(p_0-1)}{(p_0-2)}(1+\bar{\delta})}$

Table: Dependence of T^* on precision ε for any $\bar{\delta} > 0$

- For high-dimensional setting: multiplicative factor $(\psi_n d^{3/2} \psi_d)^2$ for large p

MCMC Termination criteria

Theorem

Consider stopping rule

$$T_1(\varepsilon) = \inf\{T > 0 : Vol(C(T))^{1/d} + 1_{\{T > T^*(\varepsilon,d,n)\}} \leq \varepsilon\}$$

with

$$C_T = \{\theta \in \mathbb{R}^d : T(\hat{\pi}_T(f) - \theta)^\top \hat{\Sigma}_T^{-1}(\hat{\pi}_T(f) - \theta) < q_\alpha\}$$

MCMC Termination criteria

Theorem

Consider stopping rule

$$T_1(\varepsilon) = \inf\{T > 0 : Vol(C(T))^{1/d} + 1_{\{T > T^*(\varepsilon,d,n)\}} \leq \varepsilon\}$$

with

$$C_T = \{\theta \in \mathbb{R}^d : T(\hat{\pi}_T(f) - \theta)^\top \hat{\Sigma}_T^{-1}(\hat{\pi}_T(f) - \theta) < q_\alpha\}$$

Then we have as $\varepsilon \downarrow 0$

1. *Asymptotic validity of the resulting confidence set*

$$\mathbb{P}_\pi(C(T_1(\varepsilon)) \ni \pi(f)) \to 1 - \alpha.$$

Bibliography I

Banerjee, A. and D. Vats (2022). "Multivariate strong invariance principles in Markov chain Monte Carlo". In: *arXiv preprint arXiv:2211.06855*.

Flegal, J. and G. Jones (2010). "Batch means and spectral variance estimators in Markov chain Monte Carlo". In: *The Annals of Statistics* 38.2, pp. 1034–1070.

Glynn, P. W. and W. Whitt (1992). "The asymptotic validity of sequential stopping rules for stochastic simulations". In: *The Annals of Applied Probability* 2.1, pp. 180–198.

Merlevède, F., E. Rio, et al. (2015). "Strong approximation for additive functionals of geometrically ergodic Markov chains". In: *Electronic Journal of Probability* 20.

Qin, Q. and J. P. Hobert (2019). "Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression". In: *The Annals of Statistics* 47.4, pp. 2320–2347.

Vats, D., J. M. Flegal, and G. L. Jones (2019). "Multivariate output analysis for Markov chain Monte Carlo". In: *Biometrika* 106.2, pp. 321–337.

Yang, J. and J. S. Rosenthal (2023). "Complexity results for MCMC derived from quantitative bounds". In: *The Annals of Applied Probability* 33.2, pp. 1459–1500.