

Central Limit Theorem for ergodic averages of Markov chains & the comparison of sampling algorithms for heavy-tailed distributions

Aleksandar Mijatović
Department of Statistics



Joint work with Miha Brešar and Gareth Roberts

18 December 2025

Part II: Unbiased sampling for heavy-tailed distributions

EPSRC

Engineering and Physical Sciences
Research Council

EP/W006227/1 & EP/V009478/1

Motivation: Sampling from heavy-tailed targets on $\mathbb{R}^d$

Let $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for large $|x|$ and some $\nu \in (0, \infty)$

Objective: **estimate** $\pi(g) = \int_{\mathbb{R}^d} g(x)\pi(x)dx$ for $g : \mathbb{R}^d \rightarrow \mathbb{R} \in L^1(\pi)$

This problem appears in **practical** applications [Gel+08; GLM18] including logistic regression with t -distribution or Cauchy priors.

Run **sampling algorithm** $X = (X_n)_{n \in \mathbb{N}}$ (i.e. **Markov chain with invariant law π**):

$$\text{estimator} \quad S_N(g) = \frac{1}{N} \sum_{k=1}^N g(X_k) \longrightarrow \pi(g) \quad \text{converges a.s. as } N \uparrow \infty.$$

What is the **confidence** interval for $S_N(g)$ when $g \in L^2(\pi)$?

How fast does X **converge** to π ?

L-drift condition $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$

$V : \mathcal{X} \rightarrow [1, \infty)$ cts, $\limsup_{n \rightarrow \infty} V(X_n) = \infty$ and $b, \ell_0 \in [0, \infty)$

(i) $\varphi : (0, 1] \rightarrow \mathbb{R}_+$ continuous, $r\varphi(1/r) \downarrow 0$ as $r \uparrow \infty$ and

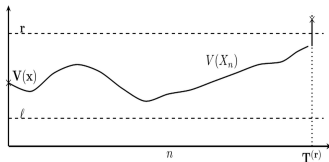
$$P(1/V)(x) - 1/V(x) \leq \varphi(1/V)(x) \quad \text{for } |x| \text{ large}$$

(ii) For $\Psi : [1, \infty) \rightarrow [1, \infty)$ submultip., any $\ell > \ell_0$, $\exists C_\ell > 0$ s.t.

$\mathbb{P}_x(T^{(r)} < S_{(\ell)}) \geq C_\ell / \Psi(r)$ for all $\ell + 1 < V(x) < r < \infty$ where

$T^{(r)} := \inf\{n \in \mathbb{N} : V(X_n) > r\}$ & $S_{(\ell)} := \inf\{n \in \mathbb{N} : V(X_n) < \ell\}$

Intuition: • Ψ quantifies the **tendency** for X to explore the tails:



- in practice (ii) implied by *drift cond* $P(\Psi \circ V)(x) \geq \Psi \circ V(x)$ for $|x|$ big
- φ quantifies the **stickiness** of X in the tail and **duration** of its return

Applications to (SIM-ALG) samplers

We can now apply our theory to the following families of algorithms:

((SIM-ALG)) **RWM** with **finite (fv-RWM)** and **infinite (iv-RWM)** variance proposals, **MALA**, Unadjusted Langevin algorithm (**ULA**) and the Stereographic Projection Sampler (**SPS**).

We also apply our theory to the **independence sampler (IS)** in order to compare IS to the samplers in (SIM-ALG).

Markov chain X with a given target π

Kernel $Q(x, dy) = q(x, y)dy$ proposes move for X from state x , accepted with probability

$$\alpha(x, y) = \begin{cases} \min\left\{\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right\}, & \text{if } \pi(x)q(x, y) > 0, \\ 1, & \text{if } \pi(x)q(x, y) = 0. \end{cases}$$

Markov transition kernel P for X is

$$P(x, dy) = p(x, y)dy + r(x)\delta_x(dy),$$

where δ_x is a Dirac delta at x and

$$p(x, y) := \begin{cases} \alpha(x, y)q(x, y), & x \neq y, \\ 0, & x = y, \end{cases} \quad \& \quad r(x) := \int_{\mathbb{R}^d} (1 - \alpha(x, y))q(x, y)dy.$$

Markov chain X with a given target π

Kernel $Q(x, dy) = q(x, y)dy$ proposes move for X from state x , accepted with probability

$$\alpha(x, y) = \begin{cases} \min\left\{\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right\}, & \text{if } \pi(x)q(x, y) > 0, \\ 1, & \text{if } \pi(x)q(x, y) = 0. \end{cases}$$

Markov transition kernel P for X is

$$P(x, dy) = p(x, y)dy + r(x)\delta_x(dy),$$

where δ_x is a Dirac delta at x and

$$p(x, y) := \begin{cases} \alpha(x, y)q(x, y), & x \neq y, \\ 0, & x = y, \end{cases} \quad \& \quad r(x) := \int_{\mathbb{R}^d} (1 - \alpha(x, y))q(x, y)dy.$$

Assume from now on that density π is proportional to

$$x \mapsto \ell(|x|)/|x|^{\nu+d}, \quad \text{for all large } |x| \text{ and some } \nu \in (0, \infty),$$

where $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfies $0 < \underline{\lim}_{r \rightarrow \infty} \ell(r) \leq \overline{\lim}_{r \rightarrow \infty} \ell(r) < \infty$.

Random-Walk Metropolis proposals with **finite variance**

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $\nu \in (0, \infty)$.

Law π has **ν -moments**:

$$\pi(|\cdot| > r) \approx r^{-\nu} \quad \text{as } r \rightarrow \infty$$

Random-Walk Metropolis proposals with **finite variance**

Target measure $\pi(x) \approx (1 + |x|)^{-(d+v)}$ for $v \in (0, \infty)$.

Law π has **v-moments**:

$$\pi(|\cdot| > r) \approx r^{-v} \quad \text{as } r \rightarrow \infty$$

fv-RWM chain at $X_n = x \in \mathbb{R}^d$ **proposes** $Y \sim Q(x, \cdot) = x + N(0, h\text{Id}_d)$.

Random-Walk Metropolis proposals with **finite variance**

Target measure $\pi(x) \approx (1 + |x|)^{-(d+v)}$ for $v \in (0, \infty)$.

Law π has **v-moments**:

$$\pi(|\cdot| > r) \approx r^{-v} \quad \text{as } r \rightarrow \infty$$

fv-RWM chain at $X_n = x \in \mathbb{R}^d$ **proposes** $Y \sim Q(x, \cdot) = x + N(0, h\text{Id}_d)$.

The move $x \rightarrow Y$ is **accepted** with probability $\alpha(x, Y) = \min \left\{ \frac{\pi(Y)}{\pi(x)}, 1 \right\}$.

Random-Walk Metropolis proposals with **finite variance**

Target measure $\pi(x) \approx (1 + |x|)^{-(d+v)}$ for $v \in (0, \infty)$.

Law π has **v-moments**:

$$\pi(|\cdot| > r) \approx r^{-v} \quad \text{as } r \rightarrow \infty$$

fv-RWM chain at $X_n = x \in \mathbb{R}^d$ **proposes** $Y \sim Q(x, \cdot) = x + N(0, h\text{Id}_d)$.

The move $x \rightarrow Y$ is **accepted** with probability $\alpha(x, Y) = \min \left\{ \frac{\pi(Y)}{\pi(x)}, 1 \right\}$.

For fv-RWM chain X targeting π , the **L(V, φ , Ψ)** conditions hold:

$$\mathbf{V}(x) = |x|, \quad \varphi(1/r) = 1/r^3 \quad \text{and} \quad \Psi(r) = r^{v+2+\varepsilon}.$$

fv-RWM chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ -drift condition holds:

$$\mathbf{V}(x) = |x|, \varphi(1/r) = 1/r^3 \text{ and } \Psi(r) = r^{\nu+2+\epsilon}.$$

Theorem (fv-RWM)

- ① **[CLT-necessary condition]** Fix measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.
- (Ia) Let $g(x) \geq |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. If $2s \in (\nu - 2, \nu)$, then the CLT for $S_n(g)$ **fails**.
 - (Ib) For bounded g , if $\nu < 2$, then CLT for $S_n(g)$ **fails**.

fv-RWM chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+v)}$

$L(\mathbf{V}, \varphi, \Psi)$ -drift condition holds:

$$V(x) = |x|, \varphi(1/r) = 1/r^3 \text{ and } \Psi(r) = r^{v+2+\varepsilon}.$$

Theorem (fv-RWM)

① **[CLT-necessary condition]** Fix measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

(Ia) Let $g(x) \geq |x|^s$ for some $2s \in (0, v)$ and all $|x|$ sufficiently large. If $2s \in (v - 2, v)$, then the CLT for $S_n(g)$ **fails**.

(Ib) For bounded g , if $v < 2$, then CLT for $S_n(g)$ **fails**.

② **[Lower bounds]** Fix $x \in \mathbb{R}^d$. For $p \geq 0$, let $f_p(y) = 1 \vee |y|^p$.

(IIa) For any $p \in [0, v)$ and every $\varepsilon > 0$, $\exists c_p \in (0, \infty)$ such that

$$c_p n^{-\frac{v-p}{2}-\varepsilon} \leq \|P^n(x, \cdot) - \pi\|_{f_p} \quad \text{for all } n \in \mathbb{N} \setminus \{0\}.$$

(IIb) For any $p \in [1, v)$ and every $\varepsilon > 0$, $\exists c_{\mathcal{W}, p} \in (0, \infty)$ such that

$$c_{\mathcal{W}, p} n^{-\frac{v-p}{2p}-\varepsilon} \leq \mathcal{W}_p(P^n(x, \cdot), \pi) \quad \text{for all } n \in \mathbb{N} \setminus \{0\}.$$

Simulations: fv-RWM with Gaussian proposals

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for large $|x|$ and some $\nu > 0$

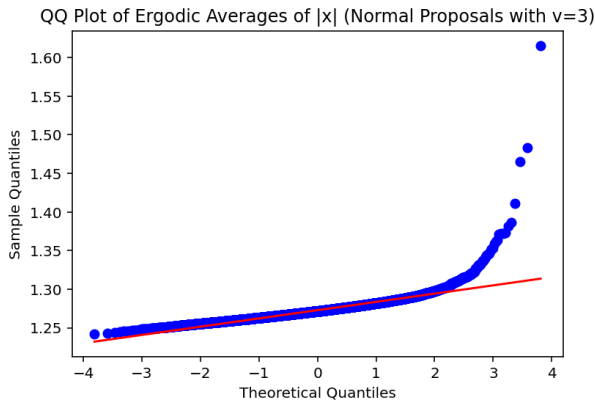
Consider an estimator $S_N(|\cdot|)$ for $\int_{\mathbb{R}^d} |x| \pi(dx)$

By above result the CLT for $S_N(|\cdot|)$ fails if $\nu - 2 < 2 \cdot 1 < 3$

Simulations: fv-RWM with Gaussian proposals

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for large $|x|$ and some $\nu > 0$
Consider an estimator $S_N(|\cdot|)$ for $\int_{\mathbb{R}^d} |x| \pi(dx)$

By above result the CLT for $S_N(|\cdot|)$ fails if $\nu - 2 < 2 \cdot 1 < 3$



Random-Walk Metropolis proposals with **infinite variance**

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $\nu \in (0, \infty)$.

iv-RWM chain X at state $x \in \mathbb{R}^d$ **proposes** $y \sim Q(x, \cdot) = x + ht(\eta)$,
where t -distribution $t(\eta)$ with $\eta \in (0, 2)$ degrees of freedom has density
proportional to

$$x \mapsto (1 + |x|^2/\eta)^{-(d+\eta)/2}$$

Random-Walk Metropolis proposals with **infinite variance**

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $\nu \in (0, \infty)$.

iv-RWM chain X at state $x \in \mathbb{R}^d$ **proposes** $y \sim Q(x, \cdot) = x + ht(\eta)$, where t -distribution $t(\eta)$ with $\eta \in (0, 2)$ degrees of freedom has density proportional to

$$x \mapsto (1 + |x|^2/\eta)^{-(d+\eta)/2}$$

For chain X targeting π the $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with:

$$\mathbf{V}(x) = |x|, \varphi(1/r) = 1/r^{1+\eta} \text{ and } \Psi(r) = r^{\nu+\eta+\varepsilon}$$

For RWM chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$, the $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold:

$$\mathbf{V}(x) = |x|, \varphi(1/r) = 1/r^{1+\eta} \text{ and } \Psi(r) = r^{\nu+\eta+\varepsilon}.$$

Theorem (RWM: proposals with $\eta \in (0, 2)$ moments)

① **[CLT-necessary condition]** Fix measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

(la) Let $g(x) \geq |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. If $2s \in (\nu - \eta, \nu)$, the CLT for $S_n(g)$ **fails**.

(lb) Let g be bounded. If $\nu - \eta < 0$, the CLT for $S_n(g)$ **fails**.

For RWM chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$, the $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold:

$$\mathbf{V}(x) = |x|, \varphi(1/r) = 1/r^{1+\eta} \text{ and } \Psi(r) = r^{\nu+\eta+\varepsilon}.$$

Theorem (RWM: proposals with $\eta \in (0, 2)$ moments)

① **[CLT-necessary condition]** Fix measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

(Ia) Let $g(x) \geq |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. If $2s \in (\nu - \eta, \nu)$, the CLT for $S_n(g)$ **fails**.

(Ib) Let g be bounded. If $\nu - \eta < 0$, the CLT for $S_n(g)$ **fails**.

② **[Lower bounds]** Fix $x \in \mathbb{R}^d$. For $p \geq 0$ let $f_p(y) = 1 \vee |y|^p$.

(IIa) For any $p \in [0, \nu)$ and every $\varepsilon > 0$, $\exists c_p \in (0, \infty)$ such that

$$c_p n^{-\frac{\nu-p}{\eta}-\varepsilon} \leq \|P^n(x, \cdot) - \pi\|_{f_p} \quad \text{for all } n \in \mathbb{N} \setminus \{0\}.$$

(IIb) For any $p \in [1, \nu)$ and every $\varepsilon > 0$, $\exists c_{\mathcal{W},p} \in (0, \infty)$ such that

$$c_{\mathcal{W},p} n^{-\frac{\nu-p}{\eta p}-\varepsilon} \leq \mathcal{W}_p(P^n(x, \cdot), \pi) \quad \text{for all } n \in \mathbb{N} \setminus \{0\}.$$

How do heavy-tailed proposals improve performance?

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for some $\nu \in (0, \infty)$. Hence

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\} \approx \min \left\{ 1, \left(\frac{|x|}{|y|} \right)^{d+\nu} \right\}$$

Proposals:

Finite variance fv-RWM: $\varphi(1/r) = 1/r^3$, $\Psi(r) = r^{\nu+2+\varepsilon}$

$t(\eta)$ -distribution ($\eta \in (0, 2)$) iv-RWM: $\varphi(1/r) = 1/r^{1+\eta}$, $\Psi(r) = r^{\nu+\eta+\varepsilon}$.

How do heavy-tailed proposals improve performance?

Target measure $\pi(x) \approx (1 + |x|)^{-(d+v)}$ for some $v \in (0, \infty)$. Hence

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\} \approx \min \left\{ 1, \left(\frac{|x|}{|y|} \right)^{d+v} \right\}$$

Proposals:

Finite variance fv-RWM: $\varphi(1/r) = 1/r^3$, $\Psi(r) = r^{v+2+\varepsilon}$

$t(\eta)$ -distribution ($\eta \in (0, 2)$) iv-RWM: $\varphi(1/r) = 1/r^{1+\eta}$, $\Psi(r) = r^{v+\eta+\varepsilon}$.

Tendency to explore tails (Ψ): improvement depends on tail-decay exponent v of π due to Metropolis step.

How do heavy-tailed proposals improve performance?

Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for some $\nu \in (0, \infty)$. Hence

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\} \approx \min \left\{ 1, \left(\frac{|x|}{|y|} \right)^{d+\nu} \right\}$$

Proposals:

Finite variance fv-RWM: $\varphi(1/r) = 1/r^3$, $\Psi(r) = r^{\nu+2+\varepsilon}$

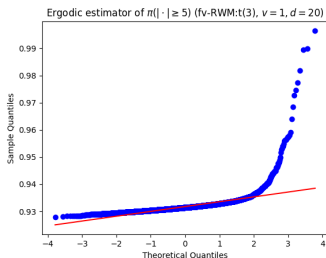
$t(\eta)$ -distribution ($\eta \in (0, 2)$) iv-RWM: $\varphi(1/r) = 1/r^{1+\eta}$, $\Psi(r) = r^{\nu+\eta+\varepsilon}$.

Tendency to explore tails (Ψ): improvement depends on tail-decay exponent ν of π due to Metropolis step.

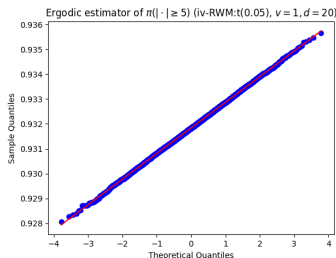
Duration of returns (φ): substantial improvement because **all** large jumps “towards the origin” are accepted

fv-RWM (Gaussian proposal) and iv-RWM ($t(\eta)$ -distribution proposal) in dimension $d = 20$

- (a) Finite variance proposals: $\varphi(1/r) = 1/r^3$ and $\Psi(r) = r^{\nu+2+\varepsilon}$
(b) t -distribution $t(\eta)$ proposals: $\varphi(1/r) = 1/r^{1+\eta}$ and $\Psi(r) = r^{\nu+\eta+\varepsilon}$.



(a) $\nu < 2$: CLT **fails**



(b) $\nu > \eta$: CLT **holds**

Figure: QQ-plots of $5 \cdot 10^4$ ergodic averages $S_n(g)$ for $\nu = 1$ and the function $g(x) = \mathbb{1}\{|x| \geq 5\}$ after $n = 10^8$ steps. In (b), $\eta = 0.05 < 2$

Metropolis adjusted Langevin algorithm (MALA)

Proposal is a discretisation of a Langevin diffusion

$$dY_t = \nabla \log(\pi)(Y_t)dt + \sqrt{2}dW_t$$

with invariant measure π .

Metropolis adjusted Langevin algorithm (MALA)

Proposal is a discretisation of a Langevin diffusion

$$dY_t = \nabla \log(\pi)(Y_t)dt + \sqrt{2}dW_t$$

with invariant measure π . For step-size discretisation $h > 0$, the proposal transition kernel is given by:

$$Q_h(x, \cdot) := N(x + h\nabla \log \pi(x), 2hId), \quad x \in \mathbb{R}^d.$$

The MALA chain X has a Metropolis kernel P with proposal Q_h .

Metropolis adjusted Langevin algorithm (MALA)

Proposal is a discretisation of a Langevin diffusion

$$dY_t = \nabla \log(\pi)(Y_t)dt + \sqrt{2}dW_t$$

with invariant measure π . For step-size discretisation $h > 0$, the proposal transition kernel is given by:

$$Q_h(x, \cdot) := N(x + h\nabla \log \pi(x), 2hId), \quad x \in \mathbb{R}^d.$$

The MALA chain X has a Metropolis kernel P with proposal Q_h .

For chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for large $|x|$ and some $\nu > 0$, the $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with:

$$\mathbf{V}(x) = |x|, \quad \varphi(1/r) = 1/r^3 \text{ and } \Psi(r) = r^{\nu+2+\varepsilon}.$$

Metropolis adjusted Langevin algorithm (MALA)

Proposal is a discretisation of a Langevin diffusion

$$dY_t = \nabla \log(\pi)(Y_t)dt + \sqrt{2}dW_t$$

with invariant measure π . For step-size discretisation $h > 0$, the proposal transition kernel is given by:

$$Q_h(x, \cdot) := N(x + h\nabla \log \pi(x), 2hId), \quad x \in \mathbb{R}^d.$$

The MALA chain X has a Metropolis kernel P with proposal Q_h .

For chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for large $|x|$ and some $\nu > 0$, the $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with:

$$\mathbf{V}(x) = |x|, \varphi(1/r) = 1/r^3 \text{ and } \Psi(r) = r^{\nu+2+\varepsilon}.$$

MALA $\mathbf{L}$ -drift conditions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ satisfied with **the same** $\mathbf{V}, \varphi, \Psi$ as both **fv-RWM** and **ULA**, implying the same conclusions.

MALA chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$, $|x|$ large

The **L**-drift conditions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with:

$$V(x) = |x|, \varphi(1/r) = 1/r^3 \text{ and } \Psi(r) = r^{\nu+2+\varepsilon}.$$

Theorem (MALA)

- ① **[CLT-necessary condition]** Fix measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.
- (la) Let $g(x) \geq |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. If $2s \in (\nu - 2, \nu)$, then the CLT for $S_n(g)$ **fails**.
 - (lb) For bounded g , if $\nu < 2$, then CLT for $S_n(g)$ **fails**.

MALA chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$, $|x|$ large

The **L**-drift conditions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with:

$$V(x) = |x|, \varphi(1/r) = 1/r^3 \text{ and } \Psi(r) = r^{\nu+2+\varepsilon}.$$

Theorem (MALA)

① **[CLT-necessary condition]** Fix measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

(Ia) Let $g(x) \geq |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. If $2s \in (\nu - 2, \nu)$, then the CLT for $S_n(g)$ **fails**.

(Ib) For bounded g , if $\nu < 2$, then CLT for $S_n(g)$ **fails**.

② **[Lower bounds]** Fix $x \in \mathbb{R}^d$. For $p \geq 0$, let $f_p(y) = 1 \vee |y|^p$.

(IIa) For any $p \in [0, \nu)$ and every $\varepsilon > 0$, $\exists c_p \in (0, \infty)$ such that

$$c_p n^{-\frac{\nu-p}{2}-\varepsilon} \leq \|P^n(x, \cdot) - \pi\|_{f_p} \quad \text{for all } n \in \mathbb{N} \setminus \{0\}.$$

(IIb) For any $p \in [1, \nu)$ and every $\varepsilon > 0$, $\exists c_{\mathcal{W}, p} \in (0, \infty)$ such that

$$c_{\mathcal{W}, p} n^{-\frac{\nu-p}{2p}-\varepsilon} \leq \mathcal{W}_p(P^n(x, \cdot), \pi) \quad \text{for all } n \in \mathbb{N} \setminus \{0\}.$$

MALA simulations

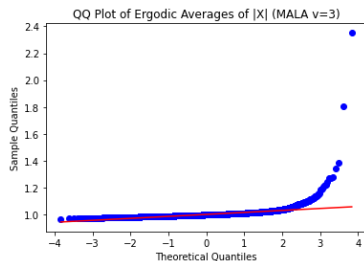
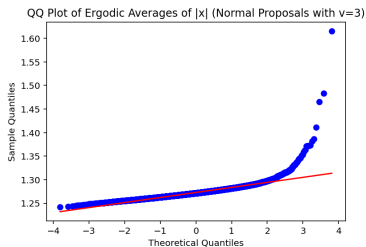
Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $|x|$ large and some $\nu > 0$
Consider an estimator $S_N(|x|)$ for first abs. moment $\pi(|x|)$.

By above result the CLT for $S_N(|x|)$ **fails** if $\nu - 2 < 2$

MALA simulations

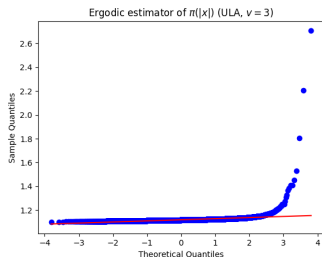
Target measure $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $|x|$ large and some $\nu > 0$
Consider an estimator $S_N(|x|)$ for first abs. moment $\pi(|x|)$.

By above result the CLT for $S_N(|x|)$ **fails** if $\nu - 2 < 2$

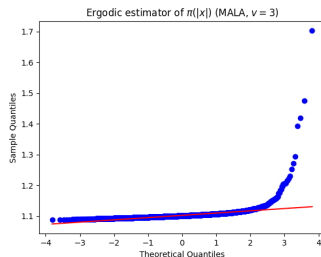


QQ-plots for MALA and RWM with Gaussian proposals. Gradient-based proposals **do not help** with CLTs.

Comparison of ULA and MALA



(a) ULA



(b) MALA

Figure: QQ-plots of 10^4 ergodic averages $S_n(g)$ for the function $g(x) = |x|$ after $n = 5 \cdot 10^7$ steps of the ULA chain (resp. MALA chain) targeting the t -distribution. The QQ-plots in panels (3a) and (3b) show that the CLTs fail, which agrees our results since $\nu - 2 = 1 < 2 = 2s$.

Stereographic Projection Sampler (SPS)

Let $\mathbb{S}^d := \{(\mathbf{y}, z) \in \mathbb{R}^d \times \mathbb{R} : |\mathbf{y}|^2 + z^2 = 1\}$ be the unit sphere in $\mathbb{R}^{d+1}$.

A stereographic projection is a bijection $\text{SP} : \mathbb{S}^d \setminus \{(\mathbf{0}, 1)\} \rightarrow \mathbb{R}^d$ given by

$$\text{SP}(\mathbf{y}, z) = \frac{1}{1-z}\mathbf{y} = \mathbf{x} \in \mathbb{R}^d \quad \text{with Jacobian } J_{\text{SP}}(\mathbf{x}) \propto (1 + |\mathbf{x}|^2)^d.$$

Moreover, the inverse $\text{SP}^{-1} : \mathbb{R}^d \rightarrow \mathbb{S}^{d-1} \setminus \{(\mathbf{0}, 1)\}$ is given by

$$\mathbf{y} = \frac{2}{|\mathbf{x}|^2 + 1}\mathbf{x} \quad \& \quad z = \frac{|\mathbf{x}|^2 - 1}{|\mathbf{x}|^2 + 1}.$$

Stereographic Projection Sampler (SPS)

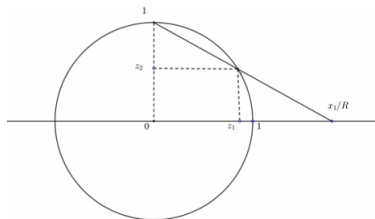
Let $\mathbb{S}^d := \{(\mathbf{y}, z) \in \mathbb{R}^d \times \mathbb{R} : |\mathbf{y}|^2 + z^2 = 1\}$ be the unit sphere in $\mathbb{R}^{d+1}$.

A stereographic projection is a bijection $\text{SP} : \mathbb{S}^d \setminus \{(\mathbf{0}, 1)\} \rightarrow \mathbb{R}^d$ given by

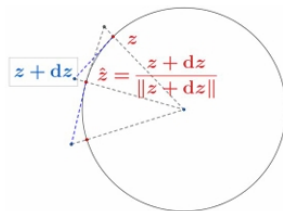
$$\text{SP}(\mathbf{y}, z) = \frac{1}{1-z}\mathbf{y} = \mathbf{x} \in \mathbb{R}^d \quad \text{with Jacobian } J_{\text{SP}}(\mathbf{x}) \propto (1 + |\mathbf{x}|^2)^d.$$

Moreover, the inverse $\text{SP}^{-1} : \mathbb{R}^d \rightarrow \mathbb{S}^{d-1} \setminus \{(\mathbf{0}, 1)\}$ is given by

$$\mathbf{y} = \frac{2}{|\mathbf{x}|^2 + 1}\mathbf{x} \quad \& \quad z = \frac{|\mathbf{x}|^2 - 1}{|\mathbf{x}|^2 + 1}.$$



(a) [YŁR24, Fig. 2]



(b) [YŁR24, Fig. 3]

Stereographic Projection Sampler (SPS) [YLR24]

Algorithm: SPS

Input Current state be $X_n = x \in \mathbb{R}^d$

- Set $\mathbf{z} := \text{SP}^{-1}(x)$
- Sample independently $\tilde{\mathbf{z}} \sim N(0, h^2 \text{Id}_{d+1})$
- Set $\mathbf{z}' := \tilde{\mathbf{z}} - \langle \mathbf{z}, \tilde{\mathbf{z}} \rangle \mathbf{z}$ and $\hat{\mathbf{z}} := \frac{\mathbf{z} + \mathbf{z}'}{|\mathbf{z} + \mathbf{z}'|}$
- The proposal $\hat{X} := \text{SP}(\hat{\mathbf{z}})$

Output Set $X_{n+1} = \hat{X}$ with probability $1 \wedge \frac{\pi(\hat{X})(1+|\hat{X}|^2)^d}{\pi(x)(1+|x|^2)^d}$;
Otherwise, set $X_{n+1} = x$

Stereographic Projection Sampler (SPS) [YŁR24]

Algorithm: SPS

Input Current state be $X_n = x \in \mathbb{R}^d$

- Set $\mathbf{z} := \text{SP}^{-1}(x)$
- Sample independently $\tilde{\mathbf{z}} \sim N(0, h^2 \text{Id}_{d+1})$
- Set $\mathbf{z}' := \tilde{\mathbf{z}} - \langle \mathbf{z}, \tilde{\mathbf{z}} \rangle \mathbf{z}$ and $\hat{\mathbf{z}} := \frac{\mathbf{z} + \mathbf{z}'}{|\mathbf{z} + \mathbf{z}'|}$
- The proposal $\hat{X} := \text{SP}(\hat{\mathbf{z}})$

Output Set $X_{n+1} = \hat{X}$ with probability $1 \wedge \frac{\pi(\hat{X})(1+|\hat{X}|^2)^d}{\pi(x)(1+|x|^2)^d}$;
Otherwise, set $X_{n+1} = x$

Remark. If

$$\sup_{x \in \mathbb{R}^d} \pi(x)(|x|^2 + 1)^d < \infty,$$

then SPS is **uniformly ergodic** [YŁR24] and ergodic estimator for any function in $L^2(\pi)$ satisfies a CLT

SPS algorithm in the heavy tailed regime: $0 < \nu < d$

Recall $\pi(x) \propto \frac{\ell(|x|)}{|x|^{\nu+d}}$ (with $0 < \underline{\lim}_{r \rightarrow \infty} \ell(r) \leq \overline{\lim}_{r \rightarrow \infty} \ell(r) < \infty$)

Under $0 < \nu < d$, we have

$$\sup_{x \in \mathbb{R}^d} \pi(x)(|x|^2 + 1)^d = \infty.$$

SPS algorithm in the heavy tailed regime: $0 < \nu < d$

Recall $\pi(x) \propto \frac{\ell(|x|)}{|x|^{\nu+d}}$ (with $0 < \underline{\lim}_{r \rightarrow \infty} \ell(r) \leq \overline{\lim}_{r \rightarrow \infty} \ell(r) < \infty$)

Under $0 < \nu < d$, we have

$$\sup_{x \in \mathbb{R}^d} \pi(x)(|x|^2 + 1)^d = \infty.$$

Upper bounds and sufficient conditions: apply [DGM08] with

$$PV - V \leq -\phi \circ V \quad \text{outside of a compact}$$

for concave $\phi(r) = r^{1-(d-\nu)/(2\gamma)}$ and $V(x) = |x|^{2\gamma}$ (with $\gamma \uparrow d/2$).

SPS algorithm in the heavy tailed regime: $0 < \nu < d$

Recall $\pi(x) \propto \frac{\ell(|x|)}{|x|^{\nu+d}}$ (with $0 < \underline{\lim}_{r \rightarrow \infty} \ell(r) \leq \overline{\lim}_{r \rightarrow \infty} \ell(r) < \infty$)

Under $0 < \nu < d$, we have

$$\sup_{x \in \mathbb{R}^d} \pi(x)(|x|^2 + 1)^d = \infty.$$

Upper bounds and sufficient conditions: apply [DGM08] with

$$PV - V \leq -\phi \circ V \quad \text{outside of a compact}$$

for concave $\phi(r) = r^{1-(d-\nu)/(2\gamma)}$ and $V(x) = |x|^{2\gamma}$ (with $\gamma \uparrow d/2$).

The $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with:

$$\mathbf{V}(x) = |x|^{2\gamma}, \quad \varphi(1/r) = 1/r^{(d-\nu)/2\gamma} \quad \text{and} \quad \Psi(r) = r^{d/2\gamma} \quad (\text{with } 0 < \gamma \ll 1)$$

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $|x|$ large

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with: $V(x) = |x|^{2\gamma}$, $\varphi(1/r) = 1/r^{(d-\nu)/2\gamma}$, $\Psi(r) = r^{d/2\gamma}$

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $|x|$ large

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with: $\mathbf{V}(x) = |x|^{2\gamma}$, $\varphi(1/r) = 1/r^{(d-\nu)/2\gamma}$, $\Psi(r) = r^{d/2\gamma}$

Theorem

Assume $\nu \in (0, d)$. Then SPS chain X satisfies:

- ① **[CLT-characterisation]** Fix a measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.
 - (la) Let $g(x) = |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. Then the CLT **holds** for $S_n(g)$ if $2s \in (0, \nu - d/2)$ and **fails** if $2s \in (\nu - d/2, \nu)$.
 - (lb) If g bounded, CLT **holds** for $S_n(g)$ if $\nu > d/2$ and **fails** if $\nu < d/2$.

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$ for $|x|$ large

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ hold with: $\mathbf{V}(x) = |x|^{2\gamma}$, $\varphi(1/r) = 1/r^{(d-\nu)/2\gamma}$, $\Psi(r) = r^{d/2\gamma}$

Theorem

Assume $\nu \in (0, d)$. Then SPS chain X satisfies:

- ① **[CLT-characterisation]** Fix a measurable $g : \mathbb{R}^d \rightarrow \mathbb{R}$.
 - (Ia) Let $g(x) = |x|^s$ for some $2s \in (0, \nu)$ and all $|x|$ sufficiently large. Then the CLT **holds** for $S_n(g)$ if $2s \in (0, \nu - d/2)$ and **fails** if $2s \in (\nu - d/2, \nu)$.
 - (Ib) If g bounded, CLT **holds** for $S_n(g)$ if $\nu > d/2$ and **fails** if $\nu < d/2$.

- ② **[Matching bounds]** Fix $x \in \mathbb{R}^d$. Set $f_p(y) = 1 \vee |y|^p$.

- (IIa) For any $p \in [0, \nu)$ and $\varepsilon > 0$, $\exists c_p, C_p \in (0, \infty)$ such that

$$c_p n^{-\frac{\nu-p}{d-\nu}-\varepsilon} \leq \|P^n(x, \cdot) - \pi\|_{f_p} \leq C_p n^{-\frac{\nu-p}{d-\nu}+\varepsilon} \quad \forall n \in \mathbb{N}.$$

- (IIb) For $p \in [1, \nu)$ and $\varepsilon > 0$, $\exists c_{\mathcal{W},p}, C_{\mathcal{W},p} \in (0, \infty)$ such that

$$c_{\mathcal{W},p} n^{-\frac{\nu-p}{p(d-\nu)}-\varepsilon} \leq \mathcal{W}_p(P^n(x, \cdot), \pi) \leq C_{\mathcal{W},p} n^{-\frac{\nu-p}{p(d-\nu)}+\varepsilon} \quad \forall n \in \mathbb{N}.$$

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

SPS exhibits **one-jump phenomenon** in Ψ -tail exploration:

$$\mathbb{P}_x(T^{(r)} < S_{(\ell)}) \approx P(x, \{V \geq r\}) \quad \text{as } r \rightarrow \infty.$$

fv-RWM & iv-RWM, MALA and ULA all exhibit **many-jumps phenomenon** in Ψ -tail exploration.

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

SPS exhibits **one-jump phenomenon** in Ψ -tail exploration:

$$\mathbb{P}_x(T^{(r)} < S_{(\ell)}) \approx P(x, \{V \geq r\}) \quad \text{as } r \rightarrow \infty.$$

fv-RWM & iv-RWM, MALA and ULA all exhibit **many-jumps phenomenon** in Ψ -tail exploration.

For targets π with **finite d moments** (i.e. $\nu > d$), SPS is the optimal choice: exponential convergence & CLTs hold for all $g \in L^2(\pi)$.

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

SPS exhibits **one-jump phenomenon** in Ψ -tail exploration:

$$\mathbb{P}_x(T^{(r)} < S_{(\ell)}) \approx P(x, \{V \geq r\}) \quad \text{as } r \rightarrow \infty.$$

fv-RWM & iv-RWM, MALA and ULA all exhibit **many-jumps phenomenon** in Ψ -tail exploration.

For targets π with **finite d moments** (i.e. $\nu > d$), SPS is the optimal choice: exponential convergence & CLTs hold for all $g \in L^2(\pi)$.

For heavy-tailed targets π with **infinite d -th moment** (i.e. $0 < \nu < d$), iv-RWM(η) with $\eta \in (0, \min\{2, d - \nu\})$ **outperforms** SPS.

SPS chain X targeting $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

SPS exhibits **one-jump phenomenon** in Ψ -tail exploration:

$$\mathbb{P}_x(T^{(r)} < S_{(\ell)}) \approx P(x, \{V \geq r\}) \quad \text{as } r \rightarrow \infty.$$

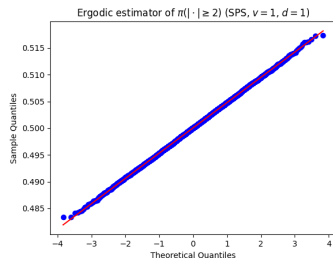
fv-RWM & iv-RWM, MALA and ULA all exhibit **many-jumps phenomenon** in Ψ -tail exploration.

For targets π with **finite d moments** (i.e. $\nu > d$), SPS is the optimal choice: exponential convergence & CLTs hold for all $g \in L^2(\pi)$.

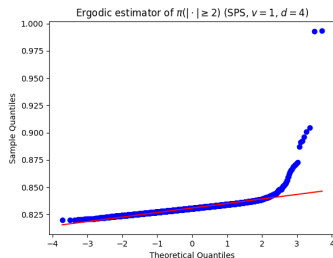
For heavy-tailed targets π with **infinite d -th moment** (i.e. $0 < \nu < d$), iv-RWM(η) with $\eta \in (0, \min\{2, d - \nu\})$ **outperforms** SPS.

SPS exhibits **fast- Ψ slow- φ** behaviour as Metropolis correction rejects jumps “towards the origin”. In contrast, iv-RWM(η) is **slow- Ψ & fast- φ** .

Success and failure of SPS



(a)



(b)

Figure: QQ-plots of $5 \cdot 10^3$ ergodic averages $S_n(g)$ for the function $g(x) = \mathbb{1}_{[2, \infty)}(|x|)$ after $n = 6 \cdot 10^6$ steps of the SPS targeting t -distribution $t(1)$ in dimension $d = 1$ (resp. $d = 4$). The QQ-plot in panel (5a) (resp. (5b)) shows that CLT holds (resp. fails), which agrees with our theory since $v = 1 > 1/2 = d/2$ (resp. $v = 1 < 2 = d/2$).

Comparison of (SIM-ALG) samplers in $\mathbb{R}^d$ targeting $\pi(|\cdot| > r) \approx r^{-\nu}$ as $r \uparrow \infty$ for some $\nu > 0$

Algorithm	CLT	ρ (TV)	β	Jump	Comment
fv-RWM	$\nu > 2$	$\nu/2$	2	many	Robust, slow, some CLT failure
iv-RWM ($\eta \in (0, 2)$)	$\nu > \eta$	ν/η	η	many	Huge gain on fvRWM; robust option for $\eta \ll 2$
ULA/MALA	$\nu > 2$	$\nu/2$	2	many	Despite gradients, no gain on fvRWM
SPS ($\nu < d$)	$\nu > d/2$	$\frac{\nu}{d-\nu}$	$d - \nu$	single	Fast yet fragile : best in (SIM-ALG) if $\nu > d$ [YLR24], worst if $\nu < d-2$
IS ($k > \nu$)	$\nu > k/2$	$\frac{\nu}{k-\nu}$	$k - \nu$	single	Extremely fragile : if $\nu = k$ exact tail simulation; if $k \neq \nu$ impractical

Notation. CLT: ergodic estimator $S_N(\mathbb{1}_A)$ of $\pi(A)$; ρ (TV): decay rate $\|P^n - \pi\|_{\text{TV}} \approx n^{-\rho}$; β : for φ in **L**-drift conditions and a rate-optimal V : $V(x)\varphi(1/V(x)) \approx |x|^{-\beta}$; **Jump**: many- or single-jump tail exploration.

Comparison of (SIM-ALG) samplers in $\mathbb{R}^d$ targeting $\pi(|\cdot| > r) \approx r^{-\nu}$ as $r \uparrow \infty$ for some $\nu > 0$

Algorithm	CLT	ρ (TV)	β	Jump	Comment
fv-RWM	$\nu > 2$	$\nu/2$	2	many	Robust, slow, some CLT failure
iv-RWM ($\eta \in (0, 2)$)	$\nu > \eta$	ν/η	η	many	Huge gain on fvRWM; robust option for $\eta \ll 2$
ULA/MALA	$\nu > 2$	$\nu/2$	2	many	Despite gradients, no gain on fvRWM
SPS ($\nu < d$)	$\nu > d/2$	$\frac{\nu}{d-\nu}$	$d - \nu$	single	Fast yet fragile : best in (SIM-ALG) if $\nu > d$ [YLR24], worst if $\nu < d - 2$
IS ($k > \nu$)	$\nu > k/2$	$\frac{\nu}{k-\nu}$	$k - \nu$	single	Extremely fragile : if $\nu = k$ exact tail simulation; if $k \neq \nu$ impractical

Notation. **CLT**: ergodic estimator $S_N(\mathbb{1}_A)$ of $\pi(A)$; **ρ (TV)**: decay rate $\|P^n - \pi\|_{\text{TV}} \approx n^{-\rho}$; **β** : for φ in **L**-drift conditions and a rate-optimal V : $V(x)\varphi(1/V(x)) \approx |x|^{-\beta}$; **Jump**: many- or single-jump tail exploration.

Remark If $\nu > d$ (resp. $\nu > k$), then SPS (resp. IS) is uniformly ergodic [YLR24] (resp. [MT96, Thm 2.1]) and the CLT holds.

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ -drift calculus for sampling algorithms

Question. How to choose the best sampling algorithm for a heavy-tailed distribution π ?

Functions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ yielding *sharp* bounds for alg with kernel P satisfy

$$\bar{\pi}(r) := \pi(\mathbf{V} \geq r) \approx 1/(r\varphi(1/r)\Psi(r)), \quad \text{as } r \rightarrow \infty.$$

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ -drift calculus for sampling algorithms

Question. How to choose the best sampling algorithm for a heavy-tailed distribution π ?

Functions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ yielding *sharp* bounds for alg with kernel P satisfy

$$\bar{\pi}(r) := \pi(\mathbf{V} \geq r) \approx 1/(r\varphi(1/r)\Psi(r)), \quad \text{as } r \rightarrow \infty.$$

If CLT holds for ergodic averages under algorithm P , then

$$r \mapsto \frac{1}{\Psi(G_1^{-1}(\sqrt{r}))} \approx \frac{\bar{\pi}(G_1^{-1}(\sqrt{r}))}{\sqrt{r}} \in L_{\text{loc}}^1(\infty),$$

where G_1^{-1} is the inverse of the function $G_1 : r \mapsto 1/(r\varphi(1/r))$

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ -drift calculus for sampling algorithms

Question. How to choose the best sampling algorithm for a heavy-tailed distribution π ?

Functions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ yielding *sharp* bounds for alg with kernel P satisfy

$$\bar{\pi}(r) := \pi(\mathbf{V} \geq r) \approx 1/(r\varphi(1/r)\Psi(r)), \quad \text{as } r \rightarrow \infty.$$

If CLT holds for ergodic averages under algorithm P , then

$$r \mapsto \frac{1}{\Psi(G_1^{-1}(\sqrt{r}))} \approx \frac{\bar{\pi}(G_1^{-1}(\sqrt{r}))}{\sqrt{r}} \in L_{\text{loc}}^1(\infty),$$

where G_1^{-1} is the inverse of the function $G_1 : r \mapsto 1/(r\varphi(1/r))$

Answer. Choose the algorithm P with φ such that

$$r \mapsto r\varphi(1/r) \quad \text{decays to 0 \textbf{slowest} as } r \rightarrow \infty$$

$\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ -drift calculus for sampling algorithms

Question. How to choose the best sampling algorithm for a heavy-tailed distribution π ?

Functions $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ yielding *sharp* bounds for alg with kernel P satisfy

$$\bar{\pi}(r) := \pi(\mathbf{V} \geq r) \approx 1/(r\varphi(1/r)\Psi(r)), \quad \text{as } r \rightarrow \infty.$$

If CLT holds for ergodic averages under algorithm P , then

$$r \mapsto \frac{1}{\Psi(G_1^{-1}(\sqrt{r}))} \approx \frac{\bar{\pi}(G_1^{-1}(\sqrt{r}))}{\sqrt{r}} \in L_{\text{loc}}^1(\infty),$$

where G_1^{-1} is the inverse of the function $G_1 : r \mapsto 1/(r\varphi(1/r))$

Answer. Choose the algorithm P with φ such that

$$r \mapsto r\varphi(1/r) \quad \text{decays to 0 \textbf{slowest} as } r \rightarrow \infty$$

$$\iff r \mapsto \varphi(r) \text{ grows \textbf{fastest} at 0} \quad (\text{recall } P(1/V) - 1/V \leq \varphi(1/V))$$

Conclusions

Vanilla algs (RWM, MALA, ULA), ubiquitous in applications, may yield estimators with **infinite asymptotic variance** if target π is heavy-tailed.

Conclusions

Vanilla algs (RWM, MALA, ULA), ubiquitous in applications, may yield estimators with **infinite asymptotic variance** if target π is heavy-tailed.

The **L**-drift framework $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ elucidates behaviour of samplers:

- φ – how “sticky” the chain is in the tails (**duration** of return times);
- Ψ – **tendency** to leave the compacts.

Conclusions

Vanilla algs (RWM, MALA, ULA), ubiquitous in applications, may yield estimators with **infinite asymptotic variance** if target π is heavy-tailed.

The **L**-drift framework $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ elucidates behaviour of samplers:

- φ – how “sticky” the chain is in the tails (**duration** of return times);
- Ψ – **tendency** to leave the compacts.

Best choice in (SIM-ALG) for heavy-tailed $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

- Random-Walk Metropolis with *heavy-tailed proposals* for $0 < \nu < d$;
- Stereographic-Projection Sampler when the tail index satisfies $\nu > d$.

Conclusions

Vanilla algs (RWM, MALA, ULA), ubiquitous in applications, may yield estimators with **infinite asymptotic variance** if target π is heavy-tailed.

The **L**-drift framework $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ elucidates behaviour of samplers:

- φ – how “sticky” the chain is in the tails (**duration** of return times);
- Ψ – **tendency** to leave the compacts.

Best choice in (SIM-ALG) for heavy-tailed $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

- Random-Walk Metropolis with *heavy-tailed proposals* for $0 < \nu < d$;
- Stereographic-Projection Sampler when the tail index satisfies $\nu > d$.

Algorithm-design insight: efficient samplers should be **Ψ -slow** (many excursion into tails: slow growth of Ψ at ∞) and **φ -fast** (short return times from each excursion: fast growth of φ at 0).

Conclusions

Vanilla algs (RWM, MALA, ULA), ubiquitous in applications, may yield estimators with **infinite asymptotic variance** if target π is heavy-tailed.

The **L-drift** framework $\mathbf{L}(\mathbf{V}, \varphi, \Psi)$ elucidates behaviour of samplers:

- φ – how “sticky” the chain is in the tails (**duration** of return times);
- Ψ – **tendency** to leave the compacts.

Best choice in (SIM-ALG) for heavy-tailed $\pi(x) \approx (1 + |x|)^{-(d+\nu)}$

- Random-Walk Metropolis with *heavy-tailed proposals* for $0 < \nu < d$;
- Stereographic-Projection Sampler when the tail index satisfies $\nu > d$.

Algorithm-design insight: efficient samplers should be **Ψ -slow** (many excursion into tails: slow growth of Ψ at ∞) and **φ -fast** (short return times from each excursion: fast growth of φ at 0).

A **new sampler** based on these principles, better than any in (SIM-ALG), as well as Hamiltonian Monte Carlo, is coming **soon!**

- [DGM08] Randal Douc, Arnaud Guillin, and Eric Moulines. “Bounds on regeneration times and limit theorems for subgeometric Markov chains”. *Ann. Inst. Henri Poincaré Probab. Stat.* 44 (2008).
- [Gel+08] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. “A weakly informative default prior distribution for logistic and other regression models”. *Ann. Appl. Stat.* 2 (2008).
- [GLM18] Joyee Ghosh, Yingbo Li, and Robin Mitra. “On the use of Cauchy prior distributions for Bayesian logistic regression”. *Bayesian Anal.* 13 (2018).
- [MT96] K. L. Mengersen and R. L. Tweedie. “Rates of convergence of the Hastings and Metropolis algorithms”. *Ann. Statist.* 24 (1996).
- [YŁR24] Jun Yang, Krzysztof Łatuszyński, and Gareth O. Roberts. “Stereographic Markov chain Monte Carlo”. *Ann. Statist.* 52 (2024).

Paper available on arXiv, see the link below the video.

Thank you for watching!!