The Robust Quasi Monte Carlo Method

Emmanuel Gobet[1], Matthieu Lerasle[2] and David Métivier

École Polytechnique, CMAP[1] and ENSAE, CREST[2]

Agenda

Monte Carlo method

Goal : estimate $\mu = \int_{[0,1]^d} f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \mathbb{E}\left(f(\boldsymbol{U})\right) = \mathbb{E}\left(X\right)$

- $\boldsymbol{U}$ r.v. uniformly distributed
- $f \in \mathbb{L}_2[0,1)^d$

Monte Carlo method

Goal : estimate $\mu = \int_{[0,1]^d} f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \mathbb{E}\left(f(\boldsymbol{U})\right) = \mathbb{E}\left(X\right)$

- $\boldsymbol{U}$ r.v. uniformly distributed
- $f \in \mathbb{L}_2[0,1)^d$

Usual Monte-Carlo method :

1. Sample i.i.d. $X = f(\boldsymbol{U})$
2. Empirical Mean estimate $\overline{X}_M = \dfrac{1}{M}\sum_{i=1}^{M} X_i$

$\rightarrow$ Asymptotic confidence interval (Central Limit Theorem)

Can we do better ?

Wish list

1. Estimators $\hat{\mu}_M$ for a wide class of $X = f(U)$, typically $X \in \mathbb{L}_2[0,1)^d$.
2. Find the associated **non-asymptotic** confidence interval
3. Convergence **faster than Monte Carlo** i.e. $o(M^{-1/2})$

Can we do better ?

Wish list

1. Estimators $\hat{\mu}_M$ for a wide class of $X = f(U)$, typically $X \in \mathbb{L}_2[0,1)^d$.
2. Find the associated **non-asymptotic** confidence interval
3. Convergence **faster than Monte Carlo** i.e. $o(M^{-1/2})$

Our solutions : combining

a. Robust mean estimators
b. Randomized Quasi-Monte Carlo methods

$\rightarrow$ Optimal confidence interval for $f \in \mathbb{L}_2[0,1)^d$

What is the best we can achieve ?

Confidence interval $\mathbb{P}\left(|A_M(f) - \mu| < g(\delta) = \epsilon\right) \geq 1 - \delta$

- $A_M(f)$ generic approximating algorithm with M sample
- Error ϵ, Confidence $1 - \delta$
- $f \in W$ normed linear space of functions defined on a domain $[0, 1)^d$

Definition 1 (Minimal probabilistic MC error at uncertainty δ)

$$g_M^{\text{opt}}(\delta, W) := \sup_{\|f\|_W \leq 1} \inf\{\epsilon > 0 \mid \exists\, (\epsilon, \delta) - \text{approximating algorithm } A_M\}$$

What is the best we can achieve ?

Confidence interval $\mathbb{P}\left(|A_M(f) - \mu| < g(\delta) = \epsilon\right) \geq 1 - \delta$

- $A_M(f)$ generic approximating algorithm with M sample
- Error ϵ, Confidence $1 - \delta$
- $f \in W$ normed linear space of functions defined on a domain $[0, 1)^d$

Definition 1 (Minimal probabilistic MC error at uncertainty δ)

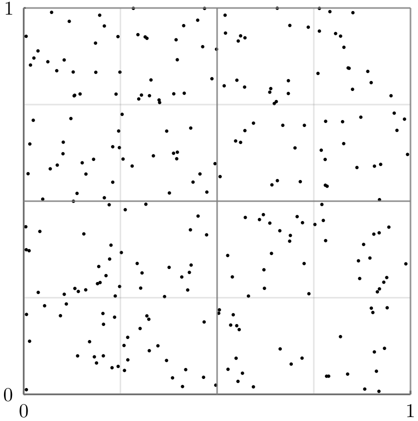
$$g_M^{\text{opt}}(\delta, W) := \sup_{\|f\|_W \leq 1} \inf\{\epsilon > 0 \mid \exists\, (\epsilon, \delta) - \text{approximating algorithm } A_M\}$$

Theorem 2 (Kunsch & Rudolf 2019)

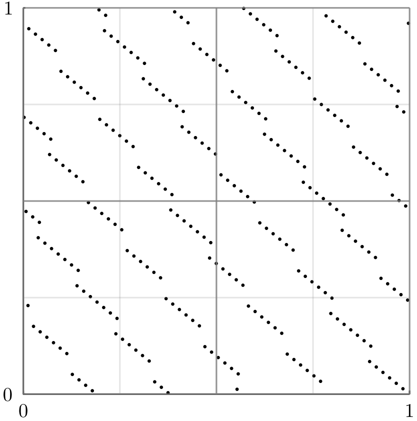
For Sobolev spaces , i.e. $f \in \mathbb{L}_p$ + *Mixed Smoothness "regularity"* $r \geq 0$
For $p \geq 2$ *and* $\log \delta^{-1} < M$

$$g_M^{\text{opt}}(\delta, W_p^{\text{mix},r}([0,1]^d)) \succeq \frac{1}{M^r}\sqrt{\frac{\log \delta^{-1}}{M}}$$

Quasi Monte Carlo

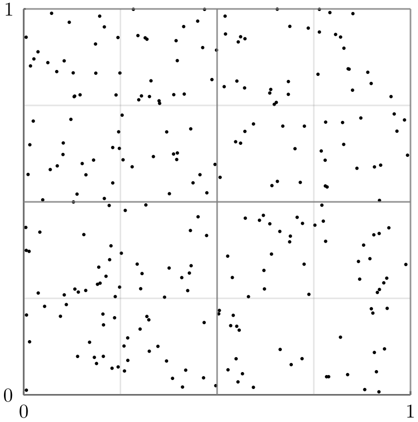


i.i.d. samples $U_i \sim \mathcal{U}([0,1]^d)$
error $= \mathcal{O}(N^{-1/2})$

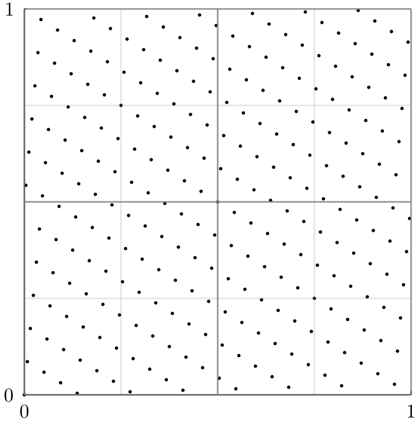


Irrational translation on the torus :
$u_i = \left(\mathrm{Frac}(i\sqrt{2}), \mathrm{Frac}(i\sqrt{3})\right)$

Quasi Monte Carlo



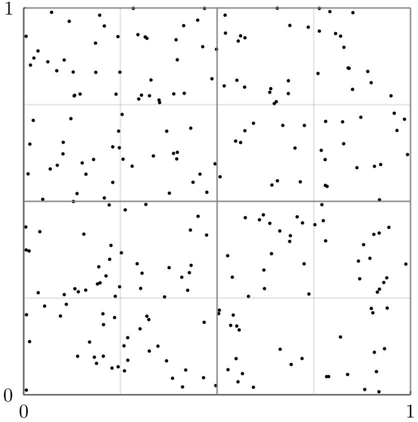
i.i.d. samples $U_i \sim \mathcal{U}([0,1]^d)$
error $= \mathcal{O}(N^{-1/2})$



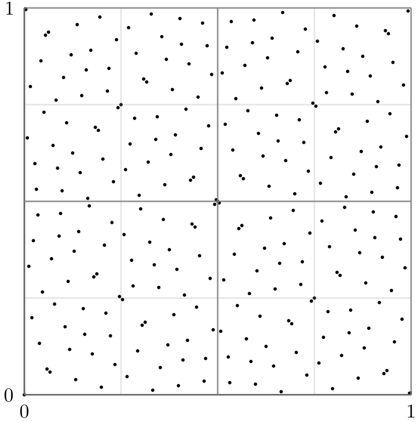
Lattice
error $= \mathcal{O}(\log(N)^d N^{-1})$

⚠ QMC deterministic error bound : Strong condition on f + intractable

Quasi Monte Carlo



i.i.d. samples $U_i \sim \mathcal{U}([0,1]^d)$
error $= \mathcal{O}(N^{-1/2})$



$(t=0, m=8, d=2)$-net
error $= \mathcal{O}(\log(N)^d N^{-1})$

⚠ QMC deterministic error bound : Strong condition on f + intractable

Randomized Quasi Monte Carlo (RQMC)

⚠ QMC error bounds are really difficult to compute in practice

→ Randomization to get confidence intervals

Randomized Quasi Monte Carlo (RQMC)

⚠ QMC error bounds are really difficult to compute in practice

→ Randomization to get confidence intervals

- The deterministic QMC sequence $(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N)$ is randomized $(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N)$
  - Still has good QMC properties
  - $\boldsymbol{R}_i \sim \mathcal{U}([0,1]^d)$ for all $i \in \{1, \ldots, N\}$

  ⚠ The $\boldsymbol{R}_m$ $m = 1, \ldots, N$ are **NOT** independent.

$\Rightarrow$ $\overline{\mu}_N = \dfrac{1}{N} \displaystyle\sum_{m=1}^{N} f(\boldsymbol{R}_m)$ is **unbiased**

Randomized Quasi Monte Carlo (RQMC)

⚠ QMC error bounds are really difficult to compute in practice

→ Randomization to get confidence intervals

- The deterministic QMC sequence $(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N)$ is randomized $(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N)$
  - Still has good QMC properties
  - $\boldsymbol{R}_i \sim \mathcal{U}([0,1]^d)$ for all $i \in \{1, \dots, N\}$
  - ⚠ The $\boldsymbol{R}_m$ $m = 1, \dots, N$ are **NOT** independent.

$\Rightarrow$ $\overline{\mu}_N = \dfrac{1}{N} \displaystyle\sum_{m=1}^{N} f\left(\boldsymbol{R}_m\right)$ is **unbiased**

Examples of randomization methods

- Cranley-Patterson rotation : random shift

Randomized Quasi Monte Carlo (RQMC)

⚠ QMC error bounds are really difficult to compute in practice

→ Randomization to get confidence intervals

- The deterministic QMC sequence $(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N)$ is randomized $(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N)$
  - Still has good QMC properties
  - $\boldsymbol{R}_i \sim \mathcal{U}([0,1]^d)$ for all $i \in \{1, \dots, N\}$

  ⚠ The $\boldsymbol{R}_m$ $m = 1, \dots, N$ are **NOT** independent.

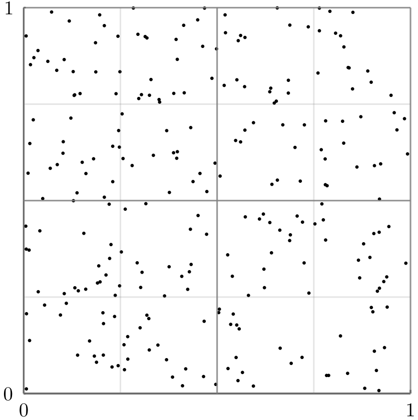
$$\Rightarrow \overline{\mu}_N = \frac{1}{N} \sum_{m=1}^{N} f(\boldsymbol{R}_m) \text{ is } \textbf{unbiased}$$

Examples of randomization methods

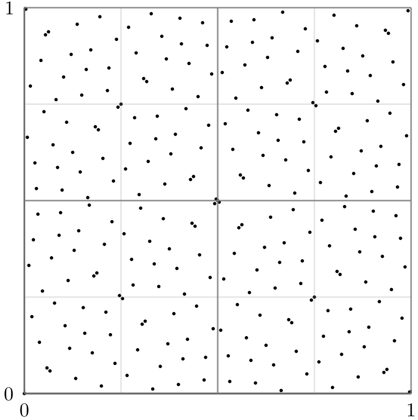
- Cranley-Patterson rotation : random shift
- Nested Uniform Scrambling for (t, m, d)-net [Owen, 1995].

  → Preserve the properties of (t, m, d)-net.
  - A lot of theoretical convergence guarantees

Examples

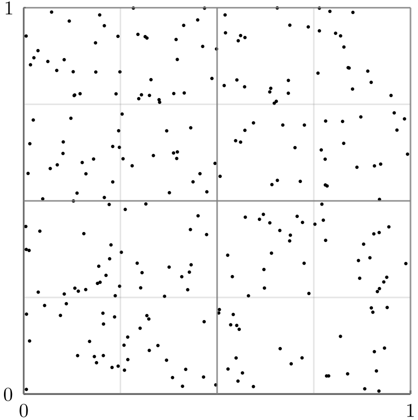


i.i.d. samples $U_i \sim \mathcal{U}([0,1]^d)$

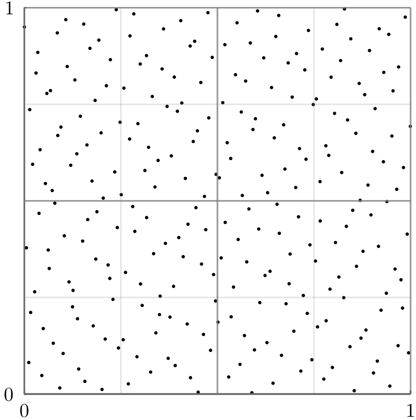


$(t = 0, m = 8, d = 2)$-net

Examples



i.i.d. samples $U_i \sim \mathcal{U}([0,1]^d)$



Scrambled

$(t = 0, m = 8, d = 2)$-net

Variance of scrambled nets

Theorem 3 (Variance of scrambled nets (Owen 1995, 1997, 1998))

If $\overline{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{R}_i)$ *with scrambled* (t, m, d)*-nets in base* b*, then*

- *If* $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \leq b^{t/2}\left(\frac{b+1}{b-1}\right)^{d/2}\sigma$
- *If* $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \to 0$ *as* $N \to \infty$

Variance of scrambled nets

Theorem 3 (Variance of scrambled nets (Owen 1995, 1997, 1998))

If $\overline{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{R}_i)$ *with scrambled* (t, m, d)*-nets in base* b*, then*

- *If* $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \leq b^{t/2}\left(\frac{b+1}{b-1}\right)^{d/2}\sigma$
- *If* $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \to 0$ *as* $N \to \infty$
- *If in addition* $f \in W_p^{\text{mix},1}$ $\sigma_N = \mathcal{O}\left(\frac{\log(N)^{(d-1)/2}}{N^{3/2}}\right)$.

- If $f \in W_p^{\text{mix},r}$, $\sigma_N = \mathcal{O}(N^{-r-1/2}(\log N)^{d(r+1)/2})$ **[Dick 2011]**

Variance of scrambled nets

Theorem 3 (Variance of scrambled nets (Owen 1995, 1997, 1998))

If $\overline{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{R}_i)$ *with scrambled* (t, m, d)*-nets in base* b*, then*

- *If* $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \leq b^{t/2}\left(\frac{b+1}{b-1}\right)^{d/2}\sigma$
- *If* $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \to 0$ *as* $N \to \infty$
- *If in addition* $f \in W_p^{\text{mix},1}$ $\sigma_N = \mathcal{O}\left(\frac{\log(N)^{(d-1)/2}}{N^{3/2}}\right)$.

- If $f \in W_p^{\text{mix},r}$, $\sigma_N = \mathcal{O}(N^{-r-1/2}(\log N)^{d(r+1)/2})$ **[Dick 2011]**
- Unbounded $f \propto x^{-1/2+\nu}$ with $0 < \nu < 1/2$. $\sigma_N = \mathcal{O}\left(N^{-1/2-\nu}\right)$ [Gobet, Lerasle, M. 2022]

Take home : Scrambled RQMC is ALWAYS (asymptotically) faster than MC

Confidence intervals for Randomized Quasi Monte Carlo

Strategy and notations :

- Computational budget : $M = n \times N$ points
- n independent mini-batches $\underbrace{\boldsymbol{R}_1^{(1)}, \ldots, \boldsymbol{R}_N^{(1)}}_{\mu_N^{(1)}}, \ldots, \underbrace{\boldsymbol{R}_1^{(n)}, \ldots, \boldsymbol{R}_N^{(n)}}_{\mu_N^{(n)}}$

$$\overline{\mu}_N^{(i)} := \frac{1}{N} \sum_{j=1}^{N} f(\boldsymbol{R}_j^{(i)}) \text{ with } \sigma_N = o(N^{-1/2}) \text{ (variance reduction)}$$

Confidence intervals for Randomized Quasi Monte Carlo

Strategy and notations :

- Computational budget : $M = n \times N$ points
- n independent mini-batches $\underbrace{\boldsymbol{R}_1^{(1)}, \dots, \boldsymbol{R}_N^{(1)}}_{\mu_N^{(1)}}, \dots, \underbrace{\boldsymbol{R}_1^{(n)}, \dots, \boldsymbol{R}_N^{(n)}}_{\mu_N^{(n)}}$

$$\overline{\mu}_N^{(i)} := \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{R}_j^{(i)}) \text{ with } \sigma_N = o(N^{-1/2}) \text{ (variance reduction)}$$

- Empirical Mean (EM) : $\overline{\mu}_M = \overline{\mu}_{N,n} := \frac{1}{n}\sum_{i=1}^{n} \overline{\mu}_N^{(i)}$.

Confidence intervals $\quad \mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq g(\delta)\right) \leq \delta$?

Goal : Optimal K.R. bounds $g_M^{\text{opt}}(\delta, W_p^{\text{mix},0} = \mathbb{L}_2) \asymp \sqrt{\log \delta^{-1}/M}$

Asymptotic bounds for $\mathbb{L}_2$ random variables

Proposition 1 (Central Limit Theorem)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)}) \in \mathbb{L}_2$ *then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \leq \sigma_N \frac{\Phi^{-1}(1-\delta)}{\sqrt{n}}\right) \to 1-\delta \ ,$$

where $\Phi^{-1}(1-\delta) \overset{\delta\to 0}{\sim} \sqrt{2\log(1/\delta)}$ *is the standard Gaussian quantile.*

- \+ Bound $\propto \sigma_N \to 0$ as $N \to +\infty$.
- \- Bound valid only as $n \to +\infty$ (especially true for $f(U)$ far from being Gaussian)
- One would prefer n small and N large to benefit from variance reduction

Asymptotic bounds for $\mathbb{L}_2$ random variables

Proposition 1 (Central Limit Theorem)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{R}_j^{(i)}) \in \mathbb{L}_2$ *then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \leq \sigma_N \frac{\Phi^{-1}(1-\delta)}{\sqrt{n}}\right) \to 1-\delta \ ,$$

where $\Phi^{-1}(1-\delta) \overset{\delta\to 0}{\sim} \sqrt{2\log(1/\delta)}$ *is the standard Gaussian quantile.*

+ Bound $\propto \sigma_N \to 0$ as $N \to +\infty$.
- Bound valid only as $n \to +\infty$ (especially true for $f(U)$ far from being Gaussian)

- One would prefer n small and N large to benefit from variance reduction

$\Rightarrow$ Non-asymptotic bounds (finite n)

Finite sample bounds for bounded random variables

Proposition 2 (Hoeffding's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)}) \leq c_N$ *a.s. then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq c_N \sqrt{\frac{\log(2/\delta)}{n}}\right) \leq \delta$$

Finite sample bounds for bounded random variables

Proposition 2 (Hoeffding's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)}) \leq c_N$ *a.s. then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq c_N\sqrt{\frac{\log(2/\delta)}{n}}\right) \leq \delta$$

- \+ Finite sample bound
- \- Requires boundedness assumptions.
- \- σ_N is replaced by $c_N = \|\overline{\mu}_N^{(i)}\|_\infty$.
- \- A priori, for QMC sequences, $c_N \nrightarrow 0$ with N

Similar for Bennett, Bernstein etc. one never gets something $\propto \sigma_N$

Finite sample bounds under finite second moment

Proposition 3 (Chebyshev's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)})$ *has a finite second moment then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq \frac{\sigma_N}{\sqrt{\delta n}}\right) \leq \delta \qquad \forall \delta \in (0,1).$$

- \+ Non asymptotic + good w.r.t. σ_N.
- \- Bad Confidence level $1/\sqrt{\delta}$ compared to CLT-level $\sqrt{\log(1/\delta)}$.

Finite sample bounds under finite second moment

Proposition 3 (Chebyshev's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)})$ *has a finite second moment then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq \frac{\sigma_N}{\sqrt{\delta n}}\right) \leq \delta \qquad \forall \delta \in (0,1).$$

+ Non asymptotic + good w.r.t. σ_N.
- Bad Confidence level $1/\sqrt{\delta}$ compared to CLT-level $\sqrt{\log(1/\delta)}$.

Take home : The Empirical Mean is

either good w.r.t. δ **BUT** strong conditions on f + sub-optimal w.r.t. σ_N

or good w.r.t. σ_N + $f \in \mathbb{L}_2$ **BUT** bad behavior w.r.t. δ.

Finite sample bounds under finite second moment

Proposition 3 (Chebyshev's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)})$ *has a finite second moment then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq \frac{\sigma_N}{\sqrt{\delta n}}\right) \leq \delta \qquad \forall \delta \in (0,1).$$

+ Non asymptotic + good w.r.t. σ_N.
- Bad Confidence level $1/\sqrt{\delta}$ compared to CLT-level $\sqrt{\log(1/\delta)}$.

Take home : The Empirical Mean is

either good w.r.t. δ **BUT** strong conditions on f + sub-optimal w.r.t. σ_N

or good w.r.t. σ_N + $f \in \mathbb{L}_2$ **BUT** bad behavior w.r.t. δ.

Objective to achieve optimally : we want mean estimators with bounds

- $\propto$ rate σ_N, (to benefit from QMC)
- $\propto$ the optimal $\sqrt{\log 1/\delta}$ rate

Robust mean estimator

Definition 4 (δ-Sub-Gaussian "Robust" mean estimator)

Given $\delta \in (0,1)$, $\widehat{\mu}_n(X_1,\ldots,X_n)$ is a δ-Sub-Gaussian mean estimators, if

$$\mathbb{P}\left(|\widehat{\mu}_n - \mu| \geq L\sigma\sqrt{\frac{\log(2/\delta)}{n}}\right) \leq \delta.$$

for some $\sqrt{2} \leq L < +\infty$.

Robust mean estimator

Definition 4 (δ-Sub-Gaussian "Robust" mean estimator)

Given $\delta \in (0,1)$, $\widehat{\mu}_n(X_1, \dots, X_n)$ is a δ-Sub-Gaussian mean estimators, if

$$\mathbb{P}\left(|\widehat{\mu}_n - \mu| \geq L\sigma\sqrt{\frac{\log(2/\delta)}{n}} \right) \leq \delta.$$

for some $\sqrt{2} \leq L < +\infty$.

- Optimal robust estimator has $L = \sqrt{2}$
- $\hat{\mu}_n$ **depends on the confidence level** δ
- There are conditions like $n \geq \log c\delta^{-1}$

Robust mean estimator

Definition 4 (δ-Sub-Gaussian “Robust” mean estimator)

Given $\delta \in (0,1)$, $\widehat{\mu}_n(X_1,\ldots,X_n)$ is a δ-Sub-Gaussian mean estimators, if

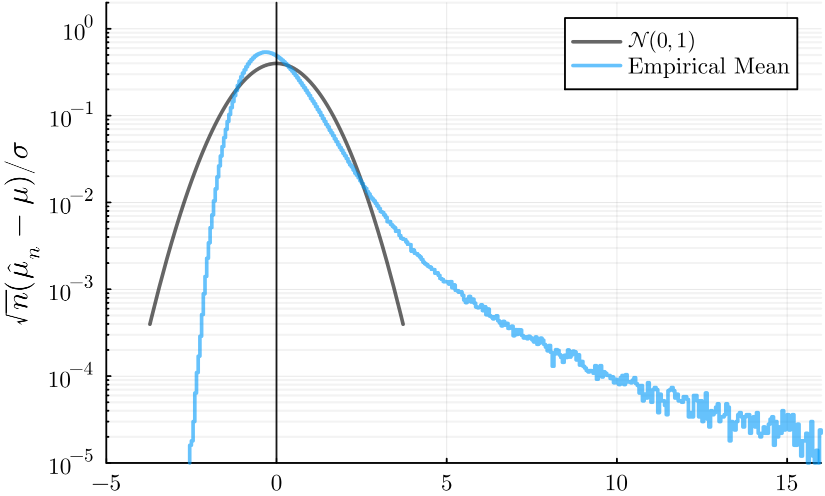
$$\mathbb{P}\left(|\widehat{\mu}_n - \mu| \geq L\sigma\sqrt{\frac{\log(2/\delta)}{n}}\right) \leq \delta.$$

for some $\sqrt{2} \leq L < +\infty$.

- Optimal robust estimator has $L = \sqrt{2}$
- $\hat{\mu}_n$ **depends on the confidence level** δ
- There are conditions like $n \geq \log c\delta^{-1}$
- Famous examples : Median of Mean, Trimmed Mean
- [Catoni, 2012] build estimator quasi optimal $L_n \approx \sqrt{2}$ (as $n \to \infty$).
- Very active topics : plenty of other estimators in the sea

Robust Mean estimation : Heavy tail example

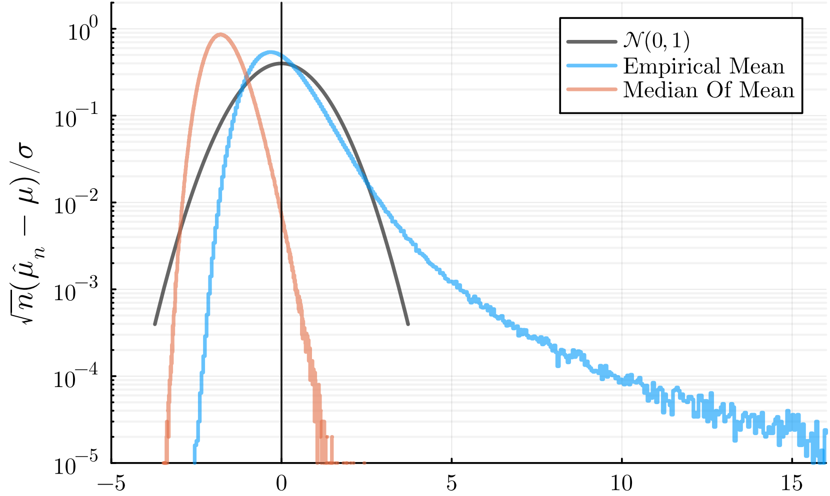
Pareto distribution $\mathrm{pdf}(x) = \nu x^{-(\nu+1)}$. $\nu > k \Rightarrow$ Finite k-moment



$\nu = 2.5, \delta \simeq 0.1\%, n = 112$

Robust Mean estimation : Heavy tail example

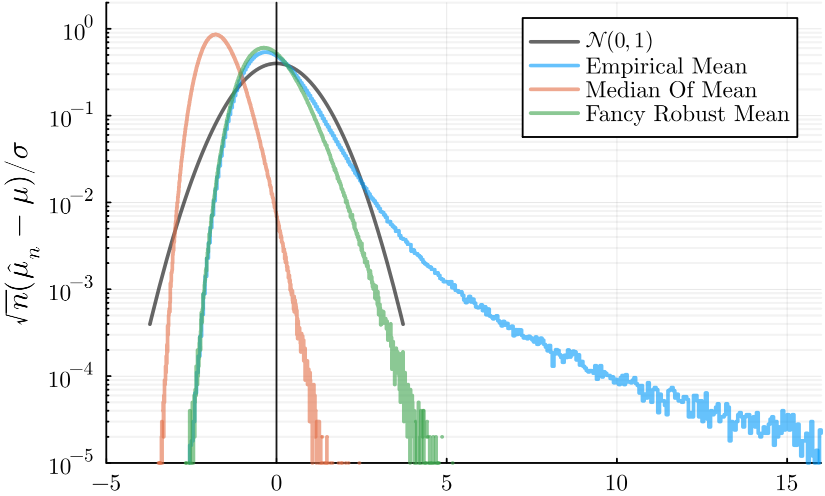
Pareto distribution $\mathrm{pdf}(x) = \nu x^{-(\nu+1)}$. $\nu > k \Rightarrow$ Finite k-moment



$\nu = 2.5$, $\delta \simeq 0.1\%$, $n = 112$

Robust Mean estimation : Heavy tail example

Pareto distribution $\mathrm{pdf}(x) = \nu x^{-(\nu+1)}$. $\nu > k \Rightarrow$ Finite k-moment



$\nu = 2.5$, $\delta \simeq 0.1\%$, $n = 112$

Example of construction : Z-estimators

- EM $\overline{\mu}_n$ solves w.r.t θ the equation $\sum_i (X_i - \theta) = 0$
- Z-estimator are defined as the zero of

$$\mathcal{R}_{n,\psi}(\theta) = \sum_{i=1}^{n} \psi\left(\alpha(X_i - \theta)\right),$$

  where the influence function $\psi : \mathbb{R} \to \mathbb{R}$ is an antisymmetric non-decreasing function and $\alpha \in \mathbb{R}$ is a tuning parameter.
- Examples :
  - $\psi(x) = x \Rightarrow$ EM,
  - $\psi(x) = \mathrm{sign}(x) \Rightarrow$ median estimator.

Huber 1964 influence function :

$$\psi(x) = \begin{cases} 1 & \text{if } x > 1 \\ x & \text{if } |x| \leq 1 \\ -1 & \text{if } x < -1 \end{cases}$$

Z-estimator with Catoni influence function :

$$\psi(x) = \begin{cases} \log(1 + x + x^2/2) & \text{if } x \leq 0 \\ -\log(1 - x + x^2/2) & \text{if } x > 0. \end{cases}$$

Still sensitive to very large values but with a logarithmic impact.

Z-estimator with Catoni influence function :

$$\psi(x) = \begin{cases} \log(1 + x + x^2/2) & \text{if } x \leq 0 \\ -\log(1 - x + x^2/2) & \text{if } x > 0. \end{cases}$$

Still sensitive to very large values but with a logarithmic impact.

Theorem 5 (Catoni estimator [Catoni, 2012])

Let $\delta \in (0,1)$ be such that $n > 2\log(2/\delta)$.

$$\textit{\textbf{Set}} \quad \alpha = \sqrt{\frac{2\log(2/\delta)}{n\sigma^2\left(1 + \frac{2\log(2/\delta)}{n-2\log(2/\delta)}\right)}}.$$

$$\textit{\textbf{Then,}} \quad \mathbb{P}\left(|\widehat{\mu}_n - \mu| < \sigma\sqrt{2}\sqrt{\frac{\log(2/\delta)}{n}\frac{1}{1-2\log(2/\delta)/n}}\right) \geq 1 - \delta.$$

- Does require knowledge of σ.
+ Nearly optimal
+ Small bias

Median-of-Means like estimator

Theorem 6 (Median-of-Means estimator [Devroye et al., 2016])

- *Split the n-sample in k blocks $J_1, \ldots, J_k$*
- *Set $\overline{X}_{J_i}$ the empirical mean over the block J_i*
- *Set*

$$\widehat{\mu}_n = \text{median}\left(\overline{X}_{J_1}, \ldots, \overline{X}_{J_k}\right).$$

Assume that $m = n/k$ is integer and that the size of each block is m. The MoM with $k = \lceil 8 \log(1/\delta) \rceil \leq n$ is a δ-sub-Gaussian estimator :

$$\mathbb{P}\left(|\widehat{\mu}_n - \mu| \leq \sigma\sqrt{\frac{32\log(1/\delta)}{n}}\right) \geq 1 - \delta$$

+ No knowledge of σ.
- Far from optimal

Theorem 7 (Lee-Valiant estimator [Lee and Valiant, 2020])

For a given $n \geq \log \delta^{-1}$*, define the* MoM *computed on* $k = \log(\delta^{-1})$ *groups (*k *integer). The Lee-Valiant estimator is then defined as*

$$\widehat{\mu}_n = \textsc{MoM} + \frac{1}{n}\sum_{i=1}^{n}(X_i - \textsc{MoM})(1 - \min(\alpha(X_i - \textsc{MoM})^2, 1))$$

where α *solves* $\displaystyle\sum_{i=1}^{n} \min(\alpha(X_i - \textsc{MoM})^2, 1) = \frac{1}{3}\log\left(\frac{1}{\delta}\right)$

Then, $\quad \mathbb{P}\left(|\widehat{\mu}_n - \mu| < \sigma(1 + o(1))\sqrt{\frac{2\log(1/\delta)}{n}}\right) \geq 1 - \delta.$

- \+ No knowledge of σ.
- \+ Nearly optimal
- \+ Very good in practice
- \- The $o(1)$ term is not much tractable.

Optimal Confidence intervals

- Computational budget : $M = n \times N$ points
- n independent mini-batches $(\bar{\mu}_N^{(i)}, i = 1, \dots, n)$
- $\overline{\mu}_N^{(i)} := \frac{1}{N} \sum_{j=1}^N f(\boldsymbol{R}_j^{(i)})$ where $(\boldsymbol{R}_1^{(i)}, \dots, \boldsymbol{R}_N^{(i)})$ is a scrambled net

Theorem 8 (Optimal confidence intervals)

$\hat{\mu}_{N,n} = \mathrm{RobustMeanEstimator}(\bar{\mu}_N^{(1)}, \dots, \bar{\mu}_N^{(n)})$ *and* $n \gtrsim \log 1/\delta$

$$\mathbb{P}\left(|\hat{\mu}_{N,n} - \mu| > L\sqrt{N}\sigma_N \sqrt{\frac{\log 1/\delta}{M}} \right) \leqslant \delta$$

Optimal Confidence intervals

- Computational budget : $M = n \times N$ points
- n independent mini-batches $(\overline{\mu}_N^{(i)}, i = 1, \ldots, n)$
- $\overline{\mu}_N^{(i)} := \frac{1}{N} \sum_{j=1}^{N} f(\boldsymbol{R}_j^{(i)})$ where $(\boldsymbol{R}_1^{(i)}, \ldots, \boldsymbol{R}_N^{(i)})$ is a scrambled net

Theorem 8 (Optimal confidence intervals)

$\hat{\mu}_{N,n} = \mathrm{RobustMeanEstimator}(\bar{\mu}_N^{(1)}, \ldots, \bar{\mu}_N^{(n)})$ *and* $n \gtrsim \log 1/\delta$

$$\mathbb{P}\left(|\hat{\mu}_{N,n} - \mu| > L\sqrt{N}\sigma_N \sqrt{\frac{\log 1/\delta}{M}} \right) \leqslant \delta$$

- For all $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \leq$ Cst at **finite** $N \to$ Optimal non asymptotic K.R. bound
- For $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \to 0$, $N \to \infty$

Optimal Confidence intervals

- Computational budget : $M = n \times N$ points
- n independent mini-batches $(\overline{\mu}_N^{(i)}, i = 1, \ldots, n)$
- $\overline{\mu}_N^{(i)} := \frac{1}{N} \sum_{j=1}^{N} f(\boldsymbol{R}_j^{(i)})$ where $(\boldsymbol{R}_1^{(i)}, \ldots, \boldsymbol{R}_N^{(i)})$ is a scrambled net

Theorem 8 (Optimal confidence intervals)

$\hat{\mu}_{N,n} = \mathrm{RobustMeanEstimator}(\bar{\mu}_N^{(1)}, \ldots, \bar{\mu}_N^{(n)})$ *and* $n \gtrsim \log 1/\delta$

$$\mathbb{P}\left(|\hat{\mu}_{N,n} - \mu| > L\sqrt{N}\sigma_N \sqrt{\frac{\log 1/\delta}{M}} \right) \leqslant \delta$$

- For all $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \leq$ Cst at **finite** $N \rightarrow$ Optimal non asymptotic K.R. bound
- For $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \rightarrow 0$, $N \rightarrow \infty$
- For singular $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N = \mathcal{O}(N^{-\nu})$

Optimal Confidence intervals

- Computational budget : $M = n \times N$ points
- n independent mini-batches $(\overline{\mu}_N^{(i)}, i = 1, \dots, n)$
- $\overline{\mu}_N^{(i)} := \frac{1}{N} \sum_{j=1}^N f(\boldsymbol{R}_j^{(i)})$ where $(\boldsymbol{R}_1^{(i)}, \dots, \boldsymbol{R}_N^{(i)})$ is a scrambled net

Theorem 8 (Optimal confidence intervals)

$\hat{\mu}_{N,n} = \mathrm{RobustMeanEstimator}(\bar{\mu}_N^{(1)}, \dots, \bar{\mu}_N^{(n)})$ *and* $n \gtrsim \log 1/\delta$

$$\mathbb{P}\left(|\hat{\mu}_{N,n} - \mu| > L\sqrt{N}\sigma_N \sqrt{\frac{\log 1/\delta}{M}} \right) \leqslant \delta$$

- For all $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \leq$ Cst at **finite** $N \to$ Optimal non asymptotic K.R. bound
- For $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N \to 0$, $N \to \infty$
- For singular $f \in \mathbb{L}_2$, $\sqrt{N}\sigma_N = \mathcal{O}(N^{-\nu})$
- For $f \in W_p^{\mathrm{mix}, r>0}$, $\sqrt{N}\sigma_N = \mathcal{O}(\log N^{d(r+1)/2})/N^r \to$ Lower K.R. bound up to log terms

Settings [Owen 2013]

Test function :

$$f(\boldsymbol{x}) = f_{\boldsymbol{\beta},\boldsymbol{h}}(\boldsymbol{x}) = \prod_{j=1}^{d}(1+\beta_j h_j(x_j))$$

such that $\mu = 1$ and $\sigma = 1/2$ with

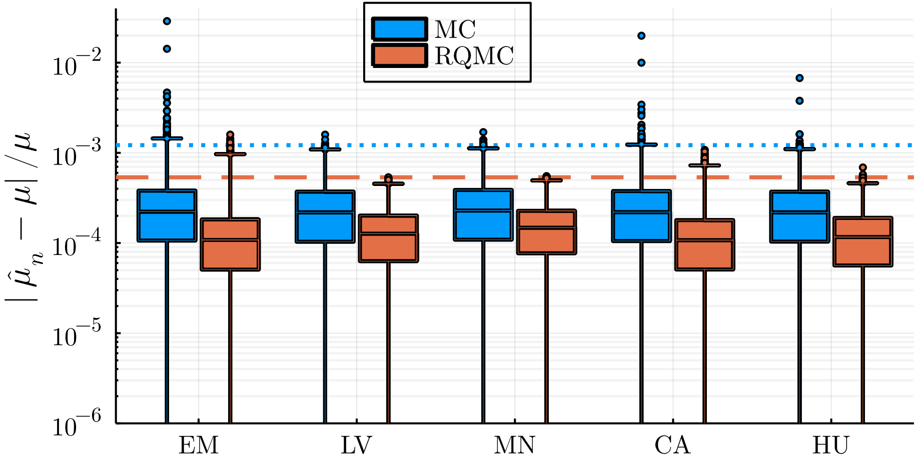
$$h_\nu(x) = \frac{\sqrt{2\nu}(2\nu+1)}{|1-2\nu|}\left(\frac{1}{x^{\frac{1}{2}-\nu}} - \frac{2}{2\nu+1}\right).$$

- $\nu > 1/2$: f a bit smooth (non-singular)
- $0 < \nu < 1/2$: f singular and $\mathbb{L}_2(0,1]^d$

The $\boldsymbol{\beta}$ control how each dimension contributes :

$$\beta_j = \frac{\beta}{\log(1+j)},$$

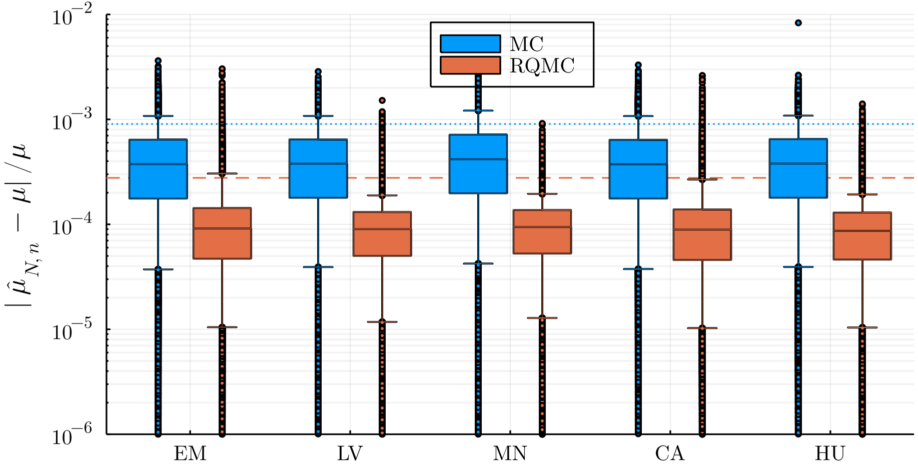
Monte Carlo vs Robust Randomised Quasi Monte Carlo



f **singular**, $\delta \simeq 0.1\%$, $n = 56$, $N = 2^{15}$, $d = 17$, $\nu = 0.05$

The whiskers show the estimated quantile $1 - \delta$

Monte Carlo vs Robust Randomised Quasi Monte Carlo



f **smooth** and **bounded**, $\delta \simeq 5\%$, $n = 12$, $N = 2^{16}$, $d = 17$, $\nu = 2 \times 10^5$

The whiskers show the estimated quantile $1 - \delta$

Conclusion

Take home :

- Use good (R)QMC methods !
- Forget about Empirical Mean far from Gaussian
- RQMC + Robust estimators achieve optimal error bound

+ Julia Packages `RobustMeans.jl` + `QuasiMonteCarlo.jl`

Julia is great for **collaborative/interdisciplinary** scientific computing (and fast+readable)

Conclusion

Take home :

- Use good (R)QMC methods !
- Forget about Empirical Mean far from Gaussian
- RQMC + Robust estimators achieve optimal error bound

+ Julia Packages `RobustMeans.jl` + `QuasiMonteCarlo.jl`
Julia is great for **collaborative/interdisciplinary** scientific computing (and fast+readable)

Perspective

- Extension to $f : [0, 1]^d \mapsto \mathbb{R}^q$
  - ? What are sub Gaussian/Robust estimator ? Few results [Lugosi and Mendelson, 2019].
  - ? Is RQMC + Robust still the way to go ?
- Contaminated dataset

Conclusion

Take home :

- Use good (R)QMC methods !
- Forget about Empirical Mean far from Gaussian
- RQMC + Robust estimators achieve optimal error bound

+ Julia Packages `RobustMeans.jl` + `QuasiMonteCarlo.jl`
Julia is great for **collaborative/interdisciplinary** scientific computing (and fast+readable)

Perspective

- Extension to $f : [0,1]^d \mapsto \mathbb{R}^q$
  - ? What are sub Gaussian/Robust estimator ? Few results [Lugosi and Mendelson, 2019].
  - ? Is RQMC + Robust still the way to go ?
- INRAE Contaminated dataset

Thank you !

References I

Catoni, O. (2012).
Challenging the empirical mean and empirical variance : A deviation study.
Annales de l'I.H.P. Probabilités et statistiques, 48(4) :1148–1185.

Chamakh, L., Gobet, E., and Liu, W. (2021).
Orlicz norms and concentration inequalities for β-heavy tailed random variables.
In minor revision for Bernoulli.

Devroye, L., Lerasle, M., Lugosi, G., and Oliveira, R. I. (2016).
Sub-Gaussian mean estimators.
The Annals of Statistics, 44(6) :2695–2725.

Lee, J. C. H. and Valiant, P. (2020).
Optimal Sub-Gaussian Mean Estimation in \$\mathbb{R}\$.
arxiv.

Lugosi, G. and Mendelson, S. (2019).
Risk minimization by median-of-means tournaments.
Journal of the European Mathematical Society, 22(3) :925–965.

References II

Owen, A. B. (1995).
Randomly Permuted (t,m,s)-Nets and (t, s)-Sequences.
In Niederreiter, H. and Shiue, P. J.-S., editors, Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Lecture Notes in Statistics, pages 299–317, New York, NY. Springer.

Finite sample bounds for bounded random variables

Proposition 4 (Bennett's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{R}_j^{(i)}) \leq c_N$ *a.s. then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq \epsilon\right) \leq f(\epsilon) \quad \left(\Leftrightarrow \mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq g(\delta)\right) \leq \delta\right)$$

with $\quad f(\varepsilon) = 2\exp\left(-\frac{n\sigma_N^2}{c_N^2} h\left(\frac{c_N \varepsilon}{\sigma_N^2}\right)\right), \quad h(u) = (1+u)\log(1+u) - u.$

Finite sample bounds for bounded random variables

Proposition 4 (Bennett's inequality)

Assume that $\overline{\mu}_N^{(i)} = \frac{1}{N}\sum_{j=1}^N f(\boldsymbol{R}_j^{(i)}) \leq c_N$ *a.s. then*

$$\mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq \epsilon\right) \leq f(\epsilon) \quad \left(\Leftrightarrow \mathbb{P}\left(|\overline{\mu}_{N,n} - \mu| \geq g(\delta)\right) \leq \delta\right)$$

with $\quad f(\varepsilon) = 2\exp\left(-\frac{n\sigma_N^2}{c_N^2} h\left(\frac{c_N\varepsilon}{\sigma_N^2}\right)\right), \quad h(u) = (1+u)\log(1+u) - u.$

If $n\sigma_N^2 \to 0$ and c_N constant, for fixed δ, $g(\delta) \asymp \dfrac{\log(1/\delta)}{n\log\left(\frac{\log(1/\delta)}{n\sigma_N^2}\right)}$

- Requires boundedness assumptions.
- Slower than the asymptotic bound

$$\frac{[n\log(1/n\sigma_N^2)]^{-1}}{\sigma_N/\sqrt{n}} = \frac{1}{(n\sigma_N^2)^{1/2}\log(1/n\sigma_N^2)} \to \infty, \qquad \text{when } n\sigma_N^2 \to 0.$$

Finite sample bounds under exponential moments

We relax the boundedness assumption on f into subexponential tail assumptions, at the price of logarithmic factor.

- Assume the following Bernstein's conditions :

$$\mathbb{E}\left(|\overline{\mu}_N^{(i)} - \mu|^q\right) \leq \frac{q!}{2}\sigma_N^2 c_N^{q-2}, \qquad \text{for any integer} \quad q \geq 2,$$

  where $c_N > 0$.
- Thus $\overline{\mu}_N^{(i)}$ has finite exponential moments and c_N is an upper bound on the Orlicz norm $\|\mu_N^{(i)}\|_{\psi_1} \lesssim c_N$. No reason that c_N goes to 0.

Proposition 5 (Bernstein's inequality)

$$f(\varepsilon) = 2\exp\left(-\frac{n\varepsilon^2}{2(\sigma_N^2 + c_N\varepsilon/3)}\right).$$

  - Asymptotically, we do not benefit from LD properties.
- [Chamakh et al., 2021] deals with β-heavy-tailed distribution (including log-normal distributions) $\rightarrow$ same issue

Finite sample bounds for β-heavy-tailed $\overline{\mu}_N^{(i)}$. [Chamakh et al., 2021] deals with β-heavy-tailed distribution (including log-normal distributions).

- Assume that $\exists \beta > 1$ s.t. the Orlicz norm $\|f(\boldsymbol{U})\|_\beta^{\mathtt{HT}}$ is finite, where

$$\|f(\boldsymbol{U})\|_\beta^{\mathtt{HT}} := \inf\left\{c > 0 : \ \mathbb{E}\left(\Psi_\beta\left(|f(\boldsymbol{U})|/c\right)\right) \le 1\right\},$$
$$\text{with} \quad \Psi_\beta(x) := \exp((\ln(x+1))^\beta) - 1.$$

  Then

$$f(\varepsilon) = 2\exp\left(-\left(\ln\left(1 + \frac{\varepsilon n}{c\left(\sigma_N\sqrt{n} + \|\overline{\mu}_N^{(1)}\|_\beta^{\mathtt{HT}}\Psi_{1/\beta}(n)\right)}\right)\right)^\beta\right).$$

  Here the sequence $\Psi_{1/\beta}(n) \to +\infty$ as $n \to +\infty$ but slowlier than any polynomial.
- Asymptotically, we do not benefit from LD properties when $\sigma_N^2 n \to 0$.