Numerical cost of the posterior Bayesian mean with a Langevin diffusion

Sébastien Gadat and Fabien Panloup and Pepito Pellegrini

Stochastic processes and statistical machine learning, II, Institut de Mathématiques de Toulouse, March 2019

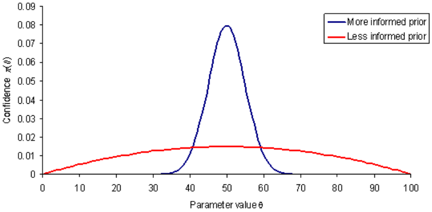
I - 1 Motivations : Learning with Bayesian Frequentist procedures

Consider a family of probability distributions

$$(\mathbb{P}_\theta)_{\theta\in\Theta} \qquad \Theta \subset \mathbb{R}^p.$$

We observe i.i.d. realizations $(X_i)_{1\leqslant i\leqslant N}$ sampled from $\mathbb{P}_{\theta_0}$.

- **Frequentist paradigm :** θ_0 exists as a hidden parameter to be recovered from the observations $(X_i)_{i\geqslant 1}$.
  Main typical tool : law of large number
- **Bayesian paradigm :** θ_0 is randomly picked with a probability π_0 over Θ that translates a prior knowledge on the parameter.



Statistical goal : recover some function of θ_0.

I - 1 Motivations : Learning with Bayesian Frequentist procedures

Bayesian paradigm : use the information brought by $(X_i)_{i \geqslant 1}$ to update our belief on Θ and compute a posterior distribution

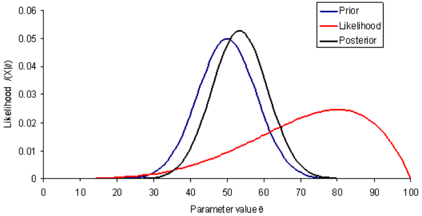
Main typical tools :

- Likelihood of the observations :

$$L_n(\theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i)$$

- Posterior distribution π_n obtained by the Bayes rule :

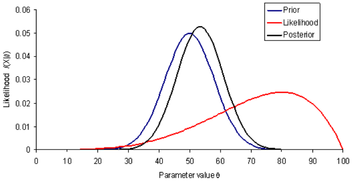
$$\pi_n(\theta) = \mathbb{P}[\theta | X_1, \ldots, X_n] = \frac{\mathbb{P}_\theta[X_1, \ldots, X_n]\pi_0[\theta]}{\mathbb{P}[X_1, \ldots, X_n]} \propto \pi_0(\theta) L_n(\theta)$$

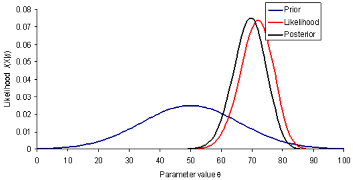


- The posterior distribution π_n is a random probability distribution over Θ
- Randomness is brought by the observations $X_1, \ldots, X_n$.

I - 1 Motivations : Learning with Bayesian Frequentist procedures

Bayesian learning : Expect a good behaviour of π_n to produce inference





The larger n, the better the information for θ_0, translated in π_n...
But not so easy to compute π_n...

- Purely Bayesian approaches : design some efficient (stochastic) algorithms to compute or approximate the posterior distribution π_n. Design a distribution q_{t_n} over Θ such that :

$$q_{t_n} \simeq \pi_n.$$

- Frequentist Bayes point of view : quantify the information brought by the concentration of π_n.

$$\pi_n \longrightarrow \delta_{\theta_0}?$$

I - 2 Cost of Bayesian learning

Two important questions :

- Question Q_1 :
  There is no reason to believe in an easy close formula for π_n. Bayesian computations are commonly using :
  - Markov Chains Metropolis Hastings procedures
  - Continuous time Langevin diffusions

  (q_t) such that

$$D(q_t, \pi_n) \leqslant \nu_t$$

- Question Q_2 :
  To recover any function $f(\theta_0)$, we need to quantify the amount of information brought by n observations

$$d(\pi_n, \delta_{\theta_0}) \leqslant \epsilon_n \longrightarrow 0 \quad \text{when} \quad n \longrightarrow +\infty$$

Key remark :
The budget constraint of n observations naturally limits the statistical accuracy in Q_2 we can expect...
There is no need to do too much computations in Q_1, with a too large t.

$$t_n = \inf\{t \geqslant 0 \,|\, \nu_t \lesssim \epsilon_n\}.$$

I - 2 Cost of Bayesian learning

In this talk :

- Question Q_1 : q_t will be the distribution at time t of a continuous time Markov process :

$$d\theta_t = \nabla_\theta[\log(\pi_0 L_n)](\theta_t)dt + dB_t \tag{1}$$

  Our estimator will be related to this S.D.E.

- Question Q_2 : The Bayesian estimator that translates the posterior contraction around θ_0 will be the posterior mean :

$$\hat{\theta}_n := \int_\Theta \theta d\pi_n(\theta). \tag{2}$$

Therefore, we need to mix several stories :

- sharp analysis of the behaviour of the posterior distribution (2) :

$$\mathbb{E}[|\hat{\theta}_n - \theta_0|^2] \leqslant \epsilon_n^2$$

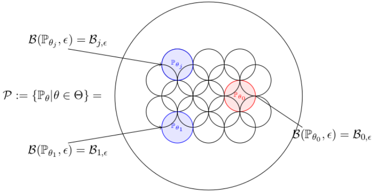
- ergodicity of the Langevin diffusion process (1) and Cesaro averaging :

$$\left| \frac{1}{t} \int_0^t \theta_s ds - \nu_\infty(I_d) \right| \leqslant \nu_t$$

I - 3 State of the art - Bayesian consistency

Not up-to-date state of the art :

- Bayesian consistency is an old story : Doob (1949) and Freedman, Le Cam and Schwartz, Ibragimov and Hasminskii' in the 60s-70s (positive results, no rates)
- Evidences that the situation is not so obvious with negative results of Freedman and Diaconis (1986).
- Key results of Barron (1988), Ghosal, Gosh and van der Vaart (2000) : tight conditions on the prior and on the complexity of $(\mathbb{P}_\theta)_{\theta\in\Theta}$.



- Castillo, van der Vaart, van Zanten, Nickl with Bernstein von Mises like theorems in various situations. Incidentally, results on the posterior mean

$$\hat{\theta}_n = \mathbb{E}_{\pi_n}[\theta].$$

I - 3 State of the art - Ergodicity of Markov processes

Not up-to-date state of the art :

- Ergodicity of Markov chains / processes : coupling arguments Doeblin (1940)
- Lyapunov type conditions : Hasminskii , Meyn-Tweedie (1970-1990)

$$LV \leqslant \beta - \alpha V$$

- Quantitative results with spectral approach / functional inequalities : Bakry and Ledoux, Cattiaux, ... (2000-.)

$$\int [f(x) - \nu_\infty(f)]^2 d\nu(x) \leqslant C_p \int |\nabla f|^2(x) d\nu(x)$$

- Link between functional approaches and Lyapunov one, additive functionals : Cattiaux, Chafai, Guillin, Zitt (2012).

II - 1 Bayesian consistency - Formulation of a result

Bayesian consistency translates

"the concentration of π_n near a Dirac mass at θ_0".

Naturally : result on the probability distributions $(\mathbb{P}_\theta)_{\theta\in\Theta}$, not one on $\theta \in \Theta$.

Θ finite

Introducing $\Lambda_n(\theta) = \frac{L_n(\theta)}{L_n(\theta_0)}$ and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, we remark that

$$\mathbb{E}[\Lambda_n(\theta) \,|\mathcal{F}_{n-1}] = \Lambda_{n-1}(\theta).$$

If ψ is a concave function, the Jensen inequality yields

$$\mathbb{E}[\psi(\Lambda_n(\theta)) \,|\mathcal{F}_{n-1}] \leqslant \psi(\Lambda_{n-1}(\theta)).$$

Take $\psi = \sqrt{}$ and obtain a quantitative result in terms of the Hellinger distance :

$$\mathbb{E}\left[\sqrt{\frac{L_n(\theta)}{L_n(\theta_0)}} \,|\mathcal{F}_{n-1}\right] \leqslant e^{-\frac{1}{2}d_H^2(\mathbb{P}_\theta,\mathbb{P}_{\theta_0})}\sqrt{\frac{L_{n-1}(\theta)}{L_{n-1}(\theta_0)}}.$$

Then use the sum is 1 :

$$\sum_{\theta\in\Theta} \pi_n(\theta) = 1 \Longrightarrow \pi_n(\theta_0) = \frac{1}{1 + \sum_{\theta\neq\theta_0} \pi_n(\theta)/\pi_n(\theta_0)}$$

and

$$\frac{\pi_n(\theta)}{\pi_n(\theta_0)} = \frac{\pi_0(\theta)}{\pi_0(\theta_0)} L_n(\theta).$$

II - 1 Bayesian consistency - Formulation of a result

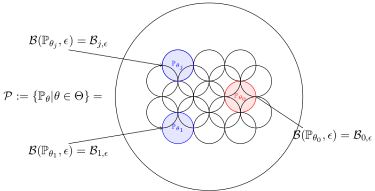
Θ finite

- Exponential concentration of $\pi_n(\theta_0) \longrightarrow 1$ at rate
$$e^{-n \inf_{\theta \neq \theta_0} d_H^2(\mathbb{P}_\theta, \mathbb{P}_{\theta_0})}$$
- Two important effects :
  - Size of Θ
  - Size of the prior $\pi_0(\theta_0)$

Θ infinite

Generalization not straightforward :

- Identifiability is needed :
$$\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2} \Longrightarrow \theta_1 = \theta_2$$
- Need to understand $B_{d_H}(\theta, \epsilon)$ and their number. The exponential contraction has to fight vs the number of balls (entropy bracketing)
- The prior mass of a ball around θ_0 is important.



II - 1 Bayesian consistency - Formulation of a result

Define $\mathcal{P} = (\mathbb{P}_\theta)_{\theta \in \Theta}$. Almost exact statement :

Theorem (Ghosal - Gosh - van der Vaart - 2000)

Assume that ϵ_n *is a sequence such that* $\epsilon_n \longrightarrow 0$ *and* $n\epsilon_n^2 \longrightarrow +\infty$ *with :*

- $\log N_{[\epsilon_n]}(\mathcal{P}, d_H) \leqslant n\epsilon_n^2$
- $\pi_0(B_{d_H}(\theta_0, \epsilon_n)) \geqslant e^{-n\epsilon_n^2 C}$.

Then a sufficiently large constant M *exists such that*

$$\pi_n(B_{d_H}(\theta_0, M\epsilon_n)) \longrightarrow 1 \quad \textit{when} \quad n \longrightarrow +\infty$$

in $\mathbb{P}_{\theta_0}$ *probability.*

Result translated to θ itself if we can prove that for a suitable α and c :

$$d_H(\mathbb{P}_\theta, \mathbb{P}_{\theta_0}) \geqslant c \wedge \|\theta - \theta_0\|^\alpha$$

Annoying fact : not enough for an upper bound of

$$\mathbb{E}_{\theta_0}[\|\hat{\theta}_n - \theta_0\|^2]$$

II - 2 Bayesian posterior mean - Tail behaviour ?

Consider $a > 0$ and the former sequence ϵ_n, the Jensen inequality leads to

$$\mathbb{E}_{\theta_0}[\|\hat{\theta}_n - \theta_0\|^2] \leqslant a^2\epsilon_n^2 + 2\int_0^{+\infty} \underbrace{(a\epsilon_n + r)}_{:=r_{a,n}} \mathbb{E}_{\theta_0}\left[\pi_n(\|\theta - \theta_0\| \geqslant a\epsilon_n + r)\right]dr$$

Need to produce an upper bound of the expectation of the posterior tail.
Approach of Castillo and van der Vaart (2012) [1] to obtain an upper bound of the quadratic loss.

- Introduce a family of tests $\phi_n^r \in \{0, 1\}$ such that

$$\mathbb{E}_{\theta_0}[\phi_n^r] \lesssim e^{-cnr_{a,n}^\beta} \qquad \text{and} \qquad \sup_{\theta:\|\theta-\theta_0\|\geqslant r_{a,n}} \mathbb{E}_\theta[(1-\phi_n^r)] \lesssim e^{-cnr_{a,n}^\beta}$$

- Exponential decay of the type I and type II errors (with a uniform control) :

$$\mathbb{P}_{\theta_0}[\phi_n^r = 1] \lesssim e^{-cnr_{a,n}^\beta} \qquad \text{and} \qquad \sup_{\theta:\|\theta-\theta_0\|\geqslant r_{a,n}} \mathbb{P}_\theta[\phi_n^r = 0] \lesssim e^{-cnr_{a,n}^\beta}.$$

- How to obtain this family of tests ? Use concentration inequalities.
  Main example : location model with log-concave densities
  U a convex function, $\Theta = \mathbb{R}^p$ and

$$\forall (x,\theta) \in \mathbb{R}^d \times \mathbb{R}^d \qquad \mathbb{P}_\theta(dx) \propto e^{-U(x-\theta)}dx$$

1. Proof slightly incorrected in [CvdV12] for $\mathbb{E}_{\theta_0}[\|\hat{\theta}_n - \theta_0\|^2]$

II - 2 Bayesian posterior mean - Tail behaviour ?

Introduce the (random) normalizing constant :

$$Z_n = \int_{\Theta} \frac{L_n(\theta)}{L_n(\theta_0)} \pi_0(\theta) d\theta$$

Use the Tonelly relationship and decompose the red term into three parts

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left[\pi_n(\|\theta-\theta_0\| \geqslant r_{a,n})\right] &\leqslant \mathbb{E}_{\theta_0}\left[\phi_n^r\right] + \mathbb{E}_{\theta_0}\left[\mathbf{1}_{Z_n \leqslant \delta_{r,n}}\right] \\
&+ \mathbb{E}_{\theta_0}\left[(1-\phi_n^r)\pi_n(\|\theta-\theta_0\| \geqslant r_{a,n})\mathbf{1}_{Z_n \geqslant \delta_{r,n}}\right] \\
&\leqslant \mathbb{P}_{\theta_0}\left[\phi_n^r = 1\right] + \mathbb{P}_{\theta_0}\left[Z_n \leqslant \delta_{n,r}\right] + \\
&+ \int_{\theta : \|\theta-\theta_0\| \geqslant r_{a,n}} \mathbb{E}_{\theta_0}\left[\mathbf{1}_{Z_n \geqslant \delta_{r,n}} \frac{(1-\phi_n^r)\frac{L_n(\theta)}{L_n(\theta_0)}}{Z_n}\right] \pi_0(\theta) d\theta, \\
&\leqslant \mathbb{P}_{\theta_0}\left[\phi_n^r = 1\right] + \mathbb{P}_{\theta_0}\left[Z_n \leqslant \delta_{n,r}\right] \\
&+ \delta_{n,r}^{-1} \int_{\theta : \|\theta-\theta_0\| \geqslant r_{a,n}} \mathbb{E}_{\theta_0}\left[(1-\phi_n^r)\frac{L_n(\theta)}{L_n(\theta_0)}\right] \pi_0(\theta) d\theta
\end{aligned}$$

Key remark : change of measure

$$\mathbb{E}_{\theta_0}\left[(1-\phi_n^r)\frac{L_n(\theta)}{L_n(\theta_0)}\right] = \mathbb{E}_{\theta}\left[(1-\phi_n^r)\right]$$

II - 3 Family of tests (ϕ_n^r) **Log-concave translation model in $\mathbb{R}^p$**

$$\mathbb{P}_\theta(x)dx = e^{-U(x-\theta)}dx$$

with

- U a convex function over $\mathbb{R}^p$
- U is $\mathcal{C}_L^1$: ∇U is a L Lipschitz function.

Define

$$m(\theta) = \mathbb{E}_\theta[X]$$

- As a translation model, $\theta \longmapsto \mathbb{P}_\theta$ is an injective map and the statistical model is therefore identifiable.
- Denote by $\bar{X}_n$ the empirical mean of the n sample $(X_1, \ldots, X_n)$ and define

$$\phi_n^r = \mathbf{1}_{|\bar{X}_n - m(\theta_0)| > \frac{r_{a,n}}{2}}.$$

- As a log-concave distribution, $\mathbb{P}_\theta$ satisfies a Poincaré inequality (Bobkov 1999) of constant C_U :

$$Var_\theta(f) \leqslant C_U \int \|\nabla f(x)\|^2 d\mathbb{P}_\theta(x)$$

- Concentration inequality then holds (Bobkov-Ledoux, 1997) :

$$\mathbb{P}_{\theta_0}\left[\phi_n^r = 1\right] \lesssim e^{-cn\frac{r_{a,n}^2}{C_U} \wedge \frac{r_{a,n}}{\sqrt{C_U}}} \quad \text{and} \quad \sup_{\theta : \|\theta - \theta_0\| \geqslant r_{a,n}} \mathbb{P}_\theta\left[\phi_n^r = 0\right] \lesssim e^{-cn\frac{r_{a,n}^2}{C_U} \wedge \frac{r_{a,n}}{\sqrt{C_U}}}$$

II - 4 Prior

In our translation model with log-concave density, the effect of the dimension p is null when looking at the complexity of the model (easy testing).
But... the dimension p acts on the size of $\delta_{n,r}$. Small fraud in this talk, details are skipped.
We can prove that

$$\mathbb{E}_{\theta_0}\left[\pi_n(\|\theta-\theta_0\| \geqslant r_{a,n})\right] \lesssim e^{-cn\frac{r_{a,n}^2}{C_U} \wedge \frac{r_{a,n}}{\sqrt{C_U}}}\left[1+e^{\log \pi_0^{-1}(B(\theta_0,\epsilon_n))}\right].$$

where $B(\theta_0,\epsilon_n)$ is the Euclidean ball centered at θ_0 of radius ϵ_n.

- When $r=0$, we need to design the sequence ϵ_n such that

$$\log \pi_0^{-1}(B(\theta_0,\epsilon_n)) \leqslant \frac{n\epsilon_n^2}{C_U}$$

- For a prior with continuous density π_0, the volume of $B(\theta_0,\epsilon)$ satisfies :

$$\log \pi_0^{-1}(B(\theta_0,\epsilon)) \lesssim p \log \epsilon^{-1} + \log(\Gamma(p/2+1)).$$

- We are led to the choice :

$$\epsilon_n = \sqrt{pC_U\frac{\log(n)}{n}}.$$

II - 5 Posterior mean - Log-concave translation model

Recall that

$$\mathbb{P}_\theta(x)dx = e^{-U(x-\theta)}dx$$

and

$$\hat{\theta}_n = \int_{\mathbb{R}^p} \theta d\pi_n(\theta)$$

Theorem

Assume that $U : \mathbb{R}^p \longrightarrow \mathbb{R}_+$ is convex and ∇U is L-Lipschitz. Consider a standard Gaussian prior π_0 over $\mathbb{R}^p$, then

$$\mathbb{E}_{\theta_0}[\|\hat{\pi}_n - \theta_0\|^2] \lesssim C_U p \frac{\log n}{n}$$

- Seems that we obtain the good convergence rate (up to the log term) . . .
- if C_U does not depend on p
- If we trust in the K.L.S. conjecture, why not ?

III - 1 Langevin diffusion

Question : how to sample π_n ?

Consider W a convex potential over $\Theta = \mathbb{R}^p$, $(B_t)_{t \geqslant 0}$ a p dimensional standard Brownian motion and the diffusion

$$d\theta_t = -\nabla W(\theta_t)dt + \sqrt{2}dB_t \quad \text{and} \quad \theta_0 \sim \mathbb{Q}_0 \tag{3}$$

- Under mild assumptions on W, we shall assume existence of trajectories.
- $(\theta_t)_{t \geqslant 0}$ is a Markov process and we have existence and uniqueness of the invariant measure as well.
- The invariant measure is a.c. w.r.t. Lebesgue measure. The associated density μ_∞ is given by the Gibbs field

$$\mu_\infty(\theta) \propto e^{-W(\theta)}$$

- Popular idea in Bayesian statistics : use (3) with the data-dependent potential :

$$W(\theta) = \log \pi_n(\theta)^{-1} = \log \pi_0(\theta)^{-1} + \sum_{i=1}^{n} \log U_\theta(X_i).$$

so that

$$\mu_\infty = \pi_{\mathbf{n}}.$$

III - 2 Ergodic behaviour of the Langevin diffusion

The semi-group being elliptic, for any $t > 0$, the law $\mathbb{Q}_t$ of θ_t is absolutely continuous w.r.t. the Lebesgue measure. We denote by q_t the density :

$$\forall \theta \in \Theta \qquad q_t(\theta) = \frac{d\mathbb{Q}_t(\theta)}{d\lambda(\theta)}.$$

Two approaches :

- Coupling with a Lyapunov function (à la Meyn-Tweedie) to obtain some Wasserstein or Total variation upper bounds

$$W_1(q_t, \mu_\infty) \leqslant \psi_{W_1}(t) \quad \text{or} \quad TV(q_t, \mu_\infty) \leqslant \psi_{TV}(t).$$

- Spectral approach with a functional inequality on μ_∞ to obtain some $\mathbb{L}^2$ or Ent results :

$$\|q_t - \mu_\infty\|_2^2 \leqslant \psi_{\mathbb{L}^2}(t) \quad \text{or} \quad Ent(q_t, \mu_\infty) \leqslant \psi_{Ent}(t)$$

Pro and cons of the two methods above :

- Lyapunov functions are easy to derive and M-T estimates can be obtained without too much computations
- Quantitative estimates obtained by coupling are overly pessimistic [2]
- Spectral approaches are sharp for some specific functions
- Obtaining functional inequalities is sometimes not so obvious

2. among other, bad scaling with the dimension

III - 2 Ergodic behaviour of the Langevin diffusion

Log-concave translation model

$$W_n(\theta) = \log \pi_0(\theta)^{-1} + \sum_{i=1}^{n} \log U(X_i - \theta)$$

- The second part of W_n is convex.
- The choice of π_0 is up to the user (at the moment, we do not need to choose an annoying heavy tail prior. [3].

If π_0 is chosen log-concave, we will obtain Poincaré inequalities on π_n.
Consequence : $\forall f \in \mathbb{L}^2(\pi_n)$:

$$\int_\Theta [\mathbb{E}_\vartheta[f(\theta_t)] - \pi_n(f)]^2 d\pi_n(\vartheta) \leqslant e^{-2\lambda_n t} \int_\Theta [f(\vartheta) - \pi_n(f)]^2 d\pi_n(\vartheta).$$

Our target is the posterior mean, *i.e.*,

$$\hat{\theta}_n = \pi_n(I) = \int_\Theta \theta d\pi_n(\theta)$$

obtained with $f = I$ ($f(\theta) = \theta$).

3. aka Exponential Weighted Aggregates for high dimensional regression

III - 3 Averaging along a trajectory of a Langevin diffusion

Given one trajectory, we use the convergence $\mathcal{L}(\theta_t) \longrightarrow \pi_n$ with $\tilde{\theta}_T$:

$$\tilde{\theta}_T = \frac{1}{T}\int_0^T \theta_s ds$$

Following arguments of Cattiaux, Chafai and Guillin 2012, we can prove the following result

Theorem

For any $\alpha > 1$ and any time $t > 0$:

$$\mathbb{E}[\|\tilde{\theta}_T - \hat{\theta}_n\|^2] \leqslant 10\alpha(J_0 \wedge 1)\sqrt{\mathbb{M}_4}\left[C_{W_n}\frac{\log T}{T} + T^{-\alpha}\right],$$

where

- $J_0 = \|m_0 - 1\|^2_{\mathbb{L}^2(\pi_n)}$ *where m_0 is the density of θ_0 w.r.t. π_n.*
- C_{W_n} *is the Poincaré constant associated to the distribution e^{-W_n}*
- $\mathbb{M}_4$ *is the fourth-order moment of the distribution π_n :*

$$\mathbb{M}_4 = \pi_n(I^4).$$

III - 3 Averaging along a trajectory of a Langevin diffusion

$$\tilde{\theta}_T = \frac{1}{T}\int_0^T \theta_s ds$$

$$\mathbb{E}[\|\tilde{\theta}_T - \hat{\theta}_n\|^2] \leqslant 10\alpha(J_0 \wedge 1)\sqrt{\mathbb{M}_4}\left[C_{W_n}\frac{\log T}{T} + T^{-\alpha}\right],$$

- $C_{W_n} = \lambda_n^{-1}$ quantifies the rate of convergence of θ_t towards the stationary regime.
- C_{W_n} is small when the potential function W_n has an important curvature.
- If m_0 is close to 1 (J_0 close to 0), good behaviour.
- We need an upper bound of $\mathbb{M}_4$.
- T quantifies the horizon of simulation.

Most of the objects above are sample dependent

III - 4 Fourth order moment

$$W_n(\theta) = \sum_{i=1}^{n} U(X_i - \theta) + \log(\pi_0^{-1}(\theta)) \qquad \text{and} \qquad \pi_n \propto e^{-W_n}.$$

Use the convexity of U and the Jensen inequality to prove the following result

Proposition

If U is convex and $\mathcal{C}_L^1$, if π_0 is Gaussian prior, then a constant C exists such that

$$\mathbb{M}_4 \leqslant C[1 + \| \arg\min W_n \|^4].$$

- A priori : $\mathbb{M}_4$ does not really increase with n.
- We can use other prior (here for the sake of convenience Gaussian)
- We only need to understand the sample dependent random variable

$$\| \arg\min W_n \|^4.$$

III - 5 Poincaré constant

$$W_n(\theta) = \sum_{i=1}^{n} U(X_i - \theta) + \log(\pi_0^{-1}(\theta)) \qquad \text{and} \qquad \pi_n \propto e^{-W_n}.$$

Use the Bakry-Emery result to state the following result

Proposition

*If U is **strongly convex** and π_0 is a log-concave prior, then*

$$C_{W_n} \lesssim \frac{1}{n}$$

Not straigthforward to extend the study to the simple convex situation...
Help of a Bobkov's result (AOP 1999) on log-concave distributions ?

Proposition

*If U is **convex** and π_0 a log-concave prior, then*

$$C_{W_n} \leqslant 432\mathbb{M}_2.$$

III - 6 Computational cost

Log-concave translation model

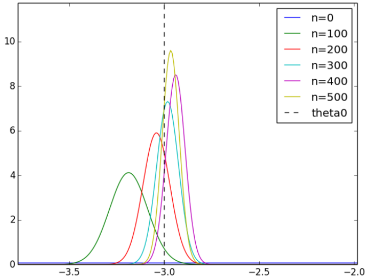
The horizon time T needed to obtain and admissible estimation should satisfy

$$\mathbb{E}[\|\tilde{\theta}_T - \hat{\theta}_n\|^2] \leqslant C_U p \frac{\log n}{n}.$$

We obtain that :

$$T_{n,p} \geqslant \frac{n}{p \log n} \times \frac{C_{W_n}\sqrt{M_4}}{C_U}.$$

Example of π_n in the Gaussian situation :



IV Frauds and on-going issues

- Understand the statistical properties of C_{W_n}
- $\tilde{\theta}_T$ is not tractable ... Urgent need to implement a discretization.
  - Euler scheme
  - Romberg scheme
  - Multi-level strategies
- Discretization is certainly carrying the main computational effort.
- On-line flow of observations $X_1, \ldots, X_n$.
  How to produce an on-line numerical scheme ?