Non-asymptotic error estimates for the Laplace approximation in Bayesian inverse problems

Tapio Helin **Remo Kretschmann**

LUT University

13th International Conference on
Monte Carlo Methods and Applications
19 August 2021

Structure

Introduction

Central error estimate

Explicit error estimate

Perturbed linear problems with Gaussian prior

Focus

Approximation of posterior distribution of Bayesian inverse problem by **Gaussian distribution** according to Laplace's method.

Laplace's method

1. Replace log-posterior density by second order Taylor approximation around MAP estimate
2. Renormalise

Examples for use of Laplace approximation

- When **sampling** posterior distribution is **too expensive**.
- Inverse problems that are **close to linear problem**.

Motivation

- **Asymptotic properties** of Laplace approximation in small noise or large data limit have been studied extensively.
- In practice, one is often interested in **quantifying approximation error** for **given noise level**.

Problems

Nonlinearity of problem or **high problem dimension** may cause **large approximation error** even for low noise level.

Goal

Understand and **quantify influence of**

1. **nonlinearity of forward mapping**,
2. **problem dimension**

on Laplace approximation error.

Recent results

- **Asymptotic behaviour** of Laplace approximation **in context of inverse problems**: **Error in Hellinger distance** converges in order of noise level [Schillings, Sprungk, and Wacker 2020].
- **Bernstein–von Mises theorem** for inverse problems when **problem dimension tends to infinity** with certain rate as noise level tends to zero [Lu 2017].

Contribution

Main results

Non-asymptotic error estimates in total variation distance for Laplace approximation in Bayesian inverse problems:

1. Central error estimate
2. Error estimate that **makes explicit** influence of **non-Gaussianity of likelihood**, **non-Gaussianity of prior**, and **problem dimension**
3. Error estimate for **perturbed linear problems** with Gaussian prior that **makes explicit** influence of **nonlinear perturbation**

Total variation distance

Definition

Total variation distance between two probability measures μ and ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ defined by

$$\begin{aligned} d_{\mathsf{TV}}(\mu, \nu) &= \frac{1}{2} \int_{\mathbb{R}^d} \left| \frac{\mathrm{d}\nu}{\mathrm{d}\lambda} - \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \right| \mathrm{d}\lambda \\ &= \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\nu(A) - \mu(A)|, \end{aligned}$$

where λ denotes Lebesgue measure on $\mathbb{R}^d$.

Total variation error of Laplace approximation is **measure of non-Gaussianity** of posterior distribution.

Structure

Set-up

For $\varepsilon > 0$, recover $x \in \mathbb{R}^d$ from noisy measurement $y \in \mathbb{R}^d$, where

$$y = G(x) + \sqrt{\varepsilon}\eta.$$

- Nonlinear **forward mapping** G
- **Random noise** $\eta \in \mathbb{R}^d$ with standard normal distribution

$$\eta \sim \mathcal{N}(0, I_d)$$

- **Prior distribution**

$$\mu(\mathrm{d}x) = \exp(-R(x))\mathrm{d}x$$

- **Posterior distribution** given by Bayes' formula as

$$\mu^y(\mathrm{d}x) \propto \exp\left(-\frac{1}{2\varepsilon}|y - G(x)|^2 - R(x)\right)\mathrm{d}x$$

Laplace approximation

Assumption

$$I(x) := \frac{1}{2}|y - G(x)|^2 + \varepsilon R(x)$$

has **unique minimiser** $\hat{x} \in \mathbb{R}^d$, $I \in C^2(\mathbb{R}^d, \mathbb{R})$, and Hessian $(\mathrm{H}I)(\hat{x})$ is **positive definite**.

Then, **Laplace approximation** of μ^y defined as

$$\mathcal{L}_{\mu^y} := \mathcal{N}(\hat{x}, \varepsilon\Sigma),$$

where $\Sigma := (\mathrm{H}I(\hat{x}))^{-1}$. This way,

$$\mu^y(\mathrm{d}x) \propto \exp\left(-\frac{1}{\varepsilon} I(x)\right) \mathrm{d}x,$$

$$\mathcal{L}_{\mu^y}(\mathrm{d}x) \propto \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_\Sigma^2\right) \mathrm{d}x.$$

Assumptions

Define $\Phi(x) := \frac{1}{2}|y - G(x)|^2$, so that $I(x) = \Phi(x) + \varepsilon R(x)$.

Bounds on log-likelihood and log-prior density

$\Phi, R \in C^3(\mathbb{R}^d, \mathbb{R})$ and there exists $K > 0$ such that

$$\max\left\{\|D^3\Phi(x)\|_\Sigma, \|D^3 R(x)\|_\Sigma\right\} \leq K \quad \text{for all } x \in \mathbb{R}^d,$$

where $\|D^3\Phi(x)\|_\Sigma := \sup\left\{|D^3\Phi(x)(h_1, h_2, h_3)| : \|h_j\|_\Sigma \leq 1\right\}$.

Quadratic bound on log-posterior density

There exists $0 < \delta \leq 1$ such that

$$I(x) - I(\hat{x}) \geq \frac{\delta}{2}\|x - \hat{x}\|_\Sigma^2 \quad \text{for all } x \in \mathbb{R}^d.$$

Want to estimate $d_{\mathrm{TV}}(\mu^y, \mathcal{L}_{\mu^y})$ in terms of K, δ, d, and ε.

Central error estimate

Theorem

Under previous assumptions on Φ, R, and I, we have

$$d_{\mathrm{TV}}(\mu^y, \mathcal{L}_{\mu^y}) \leq E_1(r_0; K) + E_2(r_0; \delta) \quad \text{for all } r_0 \geq 0,$$

where

$$E_1(r_0; K) := (2\varepsilon)^{-\frac{d}{2}} \frac{2}{\Gamma\left(\frac{d}{2}\right)} \int_0^{r_0} f(r) r^{d-1} \exp\left(-\frac{1}{2\varepsilon} r^2\right) \mathrm{d}x,$$

$$f(r) := \exp\left(\frac{(1+\varepsilon)K}{6\varepsilon} r^3\right) - 1,$$

and

$$E_2(r_0; \delta) := \delta^{-\frac{d}{2}} \frac{\Gamma\left(\frac{d}{2}, \frac{\delta}{2\varepsilon} r_0^2\right)}{\Gamma\left(\frac{d}{2}\right)}.$$

Optimal choice of r_0

Proposition

Optimal choice of r_0 in previous estimate is either 0 or satisfies

$$\exp\left(\frac{(1+\varepsilon)K}{6\varepsilon}r_0^3\right) - 1 - \exp\left(\frac{1-\delta}{2\varepsilon}r_0^2\right) = 0.$$

Second error term can be written as

$$E_2(r_0;\delta) = (2\varepsilon)^{-\frac{d}{2}}\frac{2}{\Gamma\left(\frac{d}{2}\right)}\int_{r_0}^{\infty} r^{d-1}\exp\left(-\frac{\delta}{2\varepsilon}r^2\right)\mathrm{d}x.$$

Structure

Explicit error estimate

Theorem

Suppose that previous assumptions on Φ, R, and I hold. If K, δ, ε, and d satisfy

$$\frac{2}{e\delta^{\frac{d}{2}+\frac{3}{2}}}\exp\left(-\frac{1}{8}\left(\frac{6\delta^{\frac{3}{2}}}{(1+\varepsilon)\varepsilon^{\frac{1}{2}}K}\right)^{\frac{2}{3}}\right)\leq\frac{(1+\varepsilon)\varepsilon^{\frac{1}{2}}Kd^{\frac{3}{2}}}{6\delta^{\frac{3}{2}}}\leq\frac{1}{8},$$

then

$$d_{\mathrm{TV}}(\mu^y,\mathcal{L}_{\mu^y})\leq C(1+\varepsilon)\varepsilon^{\frac{1}{2}}K\Gamma_d,$$

where

$$C:=\frac{2}{3}\sqrt{2}e\quad\text{and}\quad\Gamma_d:=\frac{\Gamma\left(\frac{d}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}.$$

- Note that $\Gamma_d\asymp\left(\frac{d}{2}\right)^{\frac{3}{2}}$ as $d\to\infty$.

Asymptotic behaviour as problem dimension $d \to \infty$

Index K_d, δ_d, and ε_d by $d \in \mathbb{N}$.

Corollary

Suppose that previous assumptions hold for all $d \in \mathbb{N}$. If $\delta_d \leq e^{-1/2}$, $\varepsilon_d \leq 1$,

$$\varepsilon_d^{\frac{1}{2}} K_d \to 0, \quad \text{and} \quad \varepsilon_d^{\frac{1}{2}} K_d d^{\frac{3}{2}} \leq 3 \left(\frac{\delta_d}{-8 \ln \delta_d} \right)^{\frac{3}{2}}$$

for all $d \in \mathbb{N}$, then for every $C > \frac{2}{3}\sqrt{2}e$ there exists $N \in \mathbb{N}$ such that

$$d_{\mathrm{TV}}(\mu^y, \mathcal{L}_{\mu^y}) \leq C \varepsilon_d^{\frac{1}{2}} K_d d^{\frac{3}{2}}$$

for all $d \geq N$.

Structure

Perturbed linear problems with Gaussian prior

- **Forward mapping** G given by linear mapping with small nonlinear perturbation of **size** $\tau \geq 0$,

$$G_\tau(x) = Ax + \tau F(x),$$

where $A \in \mathbb{R}^{d \times d}$ and $F \in C^3(\mathbb{R}^d)$.

- **Gaussian prior distribution** $\mu = \mathcal{N}(m_0, \Sigma_0)$

Assumption

There exists $\tau_0 > 0$, such that for all $\tau \in [0, \tau_0]$,

$$I_\tau(x) = \frac{1}{2}|Ax + \tau F(x) - y_\tau|^2 + \frac{\varepsilon}{2}\|x - m_0\|^2_{\Sigma_0}$$

has **unique minimiser** $\hat{x}_\tau$ with $(\mathrm{H}I_\tau)(\hat{x}_\tau) > 0$. Furthermore, y_τ, $\hat{x}_\tau$, and $\Sigma_\tau := \mathrm{H}I_\tau(\hat{x}_\tau)^{-1}$ **converge** as $\tau \to 0$ with $\lim_{\tau \to 0} \Sigma_\tau > 0$.

Assumptions

Let $B(M) \subset \mathbb{R}^d$ denote a closed Euclidean ball with radius M around the origin.

Bounds on nonlinear perturbation

There exist $C_0, \ldots, C_3 > 0$ and $M > 0$ such that

$$\|D^j F(x)\|_{\Sigma_\tau} \leq C_j, \quad j = 0, \ldots, 3,$$

for all $x \in \mathbb{R}^d$ and $\tau \in [0, \tau_0]$, and

$$D^3 F \equiv 0 \quad \text{on } \mathbb{R}^d \setminus B(M).$$

Error estimate for perturbed linear problems

Theorem

Under the previous assumptions, there exists $\tau_1 \in (0, \tau_0]$ such that

$$d_{\mathrm{TV}}(\mu^{y_\tau}, \mathcal{L}_{\mu^{y_\tau}}) \leq C\Gamma_d(1+\varepsilon)\varepsilon^{\frac{1}{2}} \left(V(\tau)\tau + \frac{W}{2}\tau^2 \right)$$

for all $\tau \in [0, \tau_1]$, where

$$C := \frac{2}{3}\sqrt{2}e, \quad \Gamma_d := \frac{\Gamma\left(\frac{d}{2} + \frac{3}{2}\right)}{\Gamma\left(\frac{d}{2}\right)},$$

$$V(\tau) := C_3(\|A\| M + |y_\tau|) + 3C_2 \left\| A \Sigma_\tau^{\frac{1}{2}} \right\|,$$

$$W := C_3 C_0 + 3C_2 C_1.$$

Moreover, $\{V(\tau)\}_{\tau \in [0, \tau_1]}$ is bounded.

Conclusion

Laplace approximation in Bayesian inverse problems

- Have non-asymptotic bound for total variation error.
- Under certain conditions, total variation error depends **linearly** on non-Gaussianity of likelihood and prior.
- For perturbed linear problems, total variation error depends **linearly** on size of nonlinear perturbation.

Outlook

- Estimate error in Wasserstein distance to achieve better or no dependence on problem dimension.

References

T. Helin and R. Kretschmann.
Non-asymptotic error estimates for the Laplace approximation in Bayesian inverse problems.
Preprint, 2020, arXiv:2012.06603.

C. Schillings, B. Sprungk, and P. Wacker.
On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems.
Numerische Mathematik, 145 (2020), pp. 915–971.

Y. Lu.
On the Bernstein–von Mises theorem for high dimensional nonlinear Bayesian inverse problems.
Preprint, 2017, arXiv:1706.00289.