The Laplace Approximation

By
Clément ADANDÉ,
Emmanuel NYANDU KAGARABI
Hermann Michael TITCHO

African Institute for Mathematical Sciences
SOUTH AFRICA

April 11, 2024

I. Introduction

In Bayesian statistics, available knowledge about parameters (θ) in a statistical model is updated with the information in observed data $(\mathcal{D})$ using Bayes' theorem :

$$\boxed{p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta}}.$$

However, then quantity $\int p(\mathcal{D}|\theta)p(\theta)d\theta$ is often **intractable**, except for simple special cases. To overcome this problem, one can use some methods like **Grid Approximation**, **Laplace Approximation**, **Variational Approximation**, **Markov Chain Monte Carlo (MCMC) Approximation**. So, here, we are focus on Laplace Approximation.

Main idea : $q(z) \approx p(z) = \mathcal{N}\left(z|z_0, A^{-1}\right)$

The **Laplace Approximation** is a simple and widely used framework which aims to find a **Gaussian approximation** to a probability density defined over a set of continuous variables. The mean equal to the **MAP** solution and precision equal to the observed **Fisher information**.

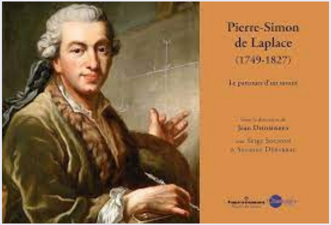


FIGURE: Pierre-Simon, Marquis de Laplace(23 March 1749–5 March 1827)

II. Development

II.1 Laplace approximation in 1d

Let's consider a **single continuous variable** z with distribution

$$p(z) = \frac{1}{Z} f(z) \text{ where } Z = \int f(z) dz$$

is the **normalisation coefficient** (**unknown**). The **goal** is to find a Gaussian approximation $q(z)$ which is centred on a mode of the distribution $p(z)$.

Steps

1. Find the mode of $p(z)$ (ie a point z_0 s.t. $p'(z_0) = 0$.)
2. Calculate the 2nd derivative (One may need Taylor expansion of $\log f(z)$ centred on z_0).
3. Write the Gaussian $q(z) \approx p(z) = \mathcal{N}\left(z|z_0, A^{-1}\right)$

Let's go

A Gaussian distribution has the property that its logarithm is a quadratic function of the variables. Let's consider now the Taylor expansion of $\ln f(z)$ centered on the mode z_0.

$$\ln f(z) \approx \ln f(z_0) + \frac{d}{dz}\left[\ln f(z_0)\right](z - z_0) + \frac{1}{2!}\frac{d^2}{dz^2}[\ln f(z_0)](z - z_0)^2$$

where

$$\frac{d}{dz}\left[\ln f(z_0)\right] = [\ln[f(z_0)]' = \ln'[f(z_0)]f'(z_0) = 0$$

Then,

$$\ln f(z) \approx \ln[f(z_0)] + 0 - \frac{1}{2}\left[-\frac{d^2}{dz^2}[\ln f(z_0)](z - z_0)^2\right]$$

$$\Rightarrow \ln f(z) \approx \ln[f(z_0)] - \frac{1}{2}A(z - z_0)^2 \text{ where } A = -\frac{d^2}{dz^2}[\ln f(z_0)]$$

$$\Rightarrow \exp\{\ln f(z)\} \approx \exp\left\{\ln[f(z_0)] - \frac{1}{2}A(z-z_0)^2\right\}$$

$$\Rightarrow f(z) \approx \exp\{\ln[f(z_0)]\} \cdot \exp\left\{-\frac{1}{2}A(z-z_0)^2\right\}$$

$$\Rightarrow f(z) \approx f(z_0)\exp\left\{-\frac{1}{2}A(z-z_0)^2\right\}$$

We obtain a normalized distribution $q(z)$ by making use of the standard result for the normalization of a gaussian. In fact,

$$
\begin{aligned}
Z &= \int f(z)dz \\
&= \int f(z_0) \exp\left\{-\frac{1}{2}A(z-z_0)^2\right\} dz \\
&= f(z_0) \int \exp\left\{-\frac{1}{2}A(z-z_0)^2\right\} dz \\
&= f(z_0) \int \exp\left\{-\frac{1}{2}(A^{-1})^{-1}(z-z_0)^2\right\} dz \\
&= f(z_0) \int \exp\left\{-\frac{1}{2}\frac{(z-z_0)^2}{A^{-1}}\right\} dz \\
&= f(z_0)\sqrt{2\pi A^{-1}}
\end{aligned}
$$

Then,

$$
\begin{aligned}
q(z) \approx p(z) &= \frac{1}{Z} f(z) \\
&= \frac{1}{f(z_0)\sqrt{2\pi A^{-1}}} f(z_0) \exp\left\{-\frac{1}{2} A (z-z_0)^2\right\} \\
&= \left(\frac{A}{\sqrt{2\pi}}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{(z-z_0)^2}{A^{-1}}\right\} \\
&= \frac{1}{\sqrt{2A^{-1}}} \exp\left\{-\frac{1}{2}\frac{(z-z_0)^2}{A^{-1}}\right\} \\
&= \mathcal{N}\left(z | z_0, A^{-1}\right)
\end{aligned}
$$

Example of Laplace approximation in 1d

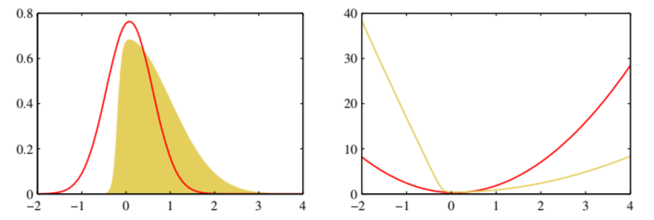


FIGURE: Illustration of the Laplace approximation applied to the distribution $p(z) \propto \exp(-z^2/2)\sigma(20z+4)$ where $\sigma(z)$ is the logistic sigmoid function defined by $\sigma(z) = \frac{1}{(1+e^{-z})}$. The left plot shows the normalized distribution $p(z)$ in yellow, together with the Laplace approximation centred on the mode z_0 of $p(z)$ in red. The right plot shows the negative logarithms of the corresponding curves. ([1, p.215])

Let's consider the same assumptions over $p(z)$, $f(z)$, Z_0 and $q(z)$ as before however, we are now in N-dimensional space. In fact, at the stationary point z_0, the gradient $\nabla f(x)$ will vanish. That's to say

$$\nabla_f(z_0) = 0.$$

By using Taylor expansion of $\ln f(z)$ at $z = z_0$, we have :

$$\ln f(z) \approx \ln f(z_0) + \nabla \ln f(z_0)(z - z_0) + \frac{1}{2!}(z - z_0)^T \nabla^2 \ln f(z_0)(z - z_0).$$

Then,

$$\ln f(z) \approx \ln f(z_0) + \frac{1}{2}(z - z_0)^T \nabla^2 \ln f(z_0)(z - z_0)$$

$$\Rightarrow \ln f(z) \approx \ln f(z_0) - \left[-\frac{1}{2}(z - z_0)^T \nabla^2 \ln f(z_0)(z - z_0)\right]$$

Let's say,

$$A = -\nabla^2 \ln f(z_0)$$

$$\Rightarrow \ln f(z) \approx \ln f(z_0) - \frac{1}{2}(z - z_0)^T A(z - z_0)$$

$$\Rightarrow \exp\{\ln f(z)\} \approx \exp\left\{\ln f(z_0) - \frac{1}{2}(z - z_0)^T A(z - z_0)\right\}$$

$$\Rightarrow f(z) \approx f(z_0) \exp\left\{-\frac{1}{2}(z - z_0)^T A(z - z_0)\right\}$$

As before, we obtain a normal distribution $q(z)$ by making use of the standard result for a normalized multivariate Gaussian :

$$\begin{aligned} Z &= \int f(z)dz \\ &= \int f(z_0) \exp\left\{-\frac{1}{2}(z - z_0)^T A(z - z_0)\right\} dz \\ &= f(z_0) \int \exp\left\{-\frac{1}{2}(z - z_0)^T A(z - z_0)\right\} dz \\ &= f(z_0) \int \exp\left\{-\frac{1}{2}(z - z_0)^T \left(A^{-1}\right)^{-1} (z - z_0)\right\} dz \\ &= f(z_0)\sqrt{|2\pi A^{-1}|} \end{aligned}$$

Then,

$$
\begin{aligned}
q(z) \approx p(z) &= \frac{1}{Z} f(z) \\
&= \frac{1}{f(z_0)\sqrt{(2\pi)^N |A^{-1}|}} f(z_0) \exp\left\{-\frac{1}{2}(z - z_0)^T A (z - z_0)\right\} \\
&= \left(\frac{A}{(2\pi)^N}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(z - z_0)^T \left(A^{-1}\right)^{-1} (z - z_0)\right\} \\
&= \frac{1}{\sqrt{(2\pi)^N |A^{-1}|}} \exp\left\{-\frac{1}{2}(z - z_0)^T \left(A^{-1}\right)^{-1} (z - z_0)\right\}
\end{aligned}
$$

$$
q(z) \approx p(z) = \mathcal{N}\left(z | z_0, A^{-1}\right)
$$

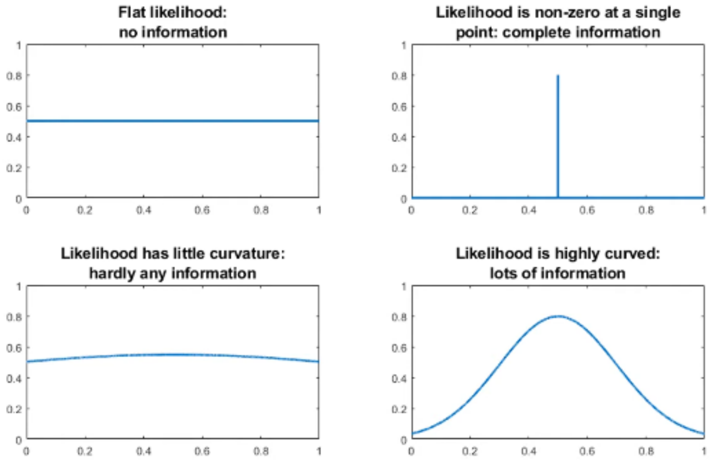
Where $|A|$ denotes the determinant of A. This gaussian distribution will be well defined provided its precision matrix, given by A, is positive definite.

II. 3 Fisher Information Matrix

- θ is parameter that models a distribution $p(X|\theta)$
- Fisher Information Matrix :

$$F(\theta) := \mathbb{E}_{p(X|\theta)}\left[\nabla \log p(X|\theta) \nabla \log p(X|\theta)^T\right]$$

- $F(\theta)$ measures the amount of information X carries about θ
- $F(\theta)$ measures the amount of curvature of the log-likelihood surface at its peak


Flat likelihood:
no information
Likelihood is non-zero at a single
point: complete information
Likelihood has little curvature:
hardly any information
Likelihood is highly curved:
lots of information
0
0.2
0.4
0.6
0.8
1

- $F(\theta)$ is the negative expected Hessian of the log likelihood [1] :

$$F(\theta) = -\mathbb{E}_{p(X|\theta)}\left[H_{\log p(X|\theta)}\right]$$

- In Laplace Approximation, $F(\hat{\theta})$ is the covariance of the fitted Gaussian

Implementation

See `Laplace_Apprximation_project.ipynb`

1. https ://agustinus.kristia.de/techblog/2018/03/11/fisher-information/

Laplace approximation framework

- Simple but widely used framework,
- Aims to find a Gaussian approximation to a probability density defined over a set of continuous variables, defined for both univariate and multivariate, directly applicable only to real variables.
- Method aims specifically at problems in which the distribution is unimodal, We can find the mode using a standard optimization method,
- Many distributions encountered in practice are multimodal :There will be different approximations according to which mode considered,

References : ([1][2][3])

- Christopher M Bishop.
  Pattern recognition and machine learning.
  Springer, 2006.

- Kevin P Murphy.
  Probabilistic machine learning : An introduction.
  MIT press, 2022.

- Kevin P Murphy.
  Probabilistic machine learning : Advanced topics.
  MIT press, 2023.

Acknowledgement

Thank you for your attention and participation !