Seminar Paper

Laplace Approximation

Department of Statistics
Ludwig-Maximilians-Universität München

Renate Schmeidler

Munich, Month Dayth, Year

Supervised by Dr Ludwig Bothmann

Abstract

Laplace approximation is a technique that is used especially in Bayesian statistics to calculate approximate values for posterior densities, means, and variances. The approach is analytical and is based on a Taylor polynomial approximation. The core idea is to replace the probability density function under consideration with an appropriate Gaussian. In some cases, the approximation will be bad; nevertheless, the Laplace approximation is an important tool for statistical analysis.
In this paper, we give an overview of the theoretical background and the methodical approach and point out some illustrating examples.

Contents

1 Introduction

In computing posterior densities or estimators for the posterior mean or variance Bayesian statisticians are normally faced with the need of calculating integrals. Many of these integrals are not solvable in an exact and analytical way. In such cases, it is necessary to resort to numerical and approximation methods.
A number of such numerical or approximation methods have been developed, often enhanced several times to present. Among the best known are the MCMC and the Laplace approximation. In this paper, we focus on the Laplace approximation.

The Laplace approximation is an analytical approach for calculating an approximate value for an integral which can be brought to a certain form. It has more requirements than the MCMC, but when it is applicable, it is typically much faster than the MCMC (see, e. g., Azevedo-Filho and Shachter, 1994, example in section 4).

At its origins, the Laplace approximation dates back to Pierre-Simon Laplace. In the second half of the 20^{th} century, a series of developments and further enhancements were made. In Section 2, we will take a brief look at some of them. Then, in Section 3, the mathematical background of the Laplace approximation is explained in a broad outline. Section 4 elucidates the single basic steps in applying the Laplace approximation.
After this theoretical framework, in Section 5, we will have a look at some illustrating examples, where we will compare several known probability functions with their Laplace approximation.

2 Related Work

In 1774 Pierre-Simon Laplace presented a new approach for the approximate calculation of an integral of a special form (published as "Mémoire sur la probabilité des causes par les évènemens", full text see e.g. Stigler's English translation, Laplace, 1986).
Today, the method he proposed is referred to as "Laplace method". Already Laplace placed this task in the context of statistics, and indeed Laplace approximation, which is based on an application and special enhancement of Laplace method, proved helpful especially in Bayesian statistics.
Therefore, in the 20^{th} century, scientists addressed various issues around the Laplace method or rather the Laplace approximation. E. g., Tierney and Kadane (1986) presented the method of Laplace approximation applied to the issue of approximating posterior means and variances for positive functions of parameters and applied to the issue of approximating marginal posterior densities. In the case of these means and variances, they could prove a better limiting behavior of the approximation, i. e., from $\mathcal{O}(n^{-1})$ to $\mathcal{O}(n^{-2})$. The technique of the fully exponential Laplace approximation they had used could be extended to expectations and variances of nonpositive functions by Tierney et al. (1989).
Azevedo-Filho and Shachter (1994) summarized in their paper the scientific development in relation to Laplace' method approximations and suggested an application for probabilistic inference in belief networks. Shun and McCullagh (1995) addressed the problem that usual Laplace approximation does not ensure a good approximation when the di-

mension p of the considered integral increases with the sample size n.
In the second decade of the 21^{th} century, the R-INLA-Project inspired the broad application and further enhancement of the Laplace approximation. Even in the recent time, scientific work in the context of the Laplace approximation is carried out; e. g. McDonald and Campbell (2024) proposed in their paper a probabilistic numerical framework that shall provide some sort of diagnostic tool whether the Laplace approximation in a special case will fit well. Katsevich (2024) addressed the problem that Laplace approximation has issues in high-dimensional spaces and developed some enhancement. Kasprzak et al. (2025) concentrated in their article on error bounds for the Laplace approximation for a Bayesian posterior.

Finally, it should be mentioned that the Laplace approximation by now has become firmly established in reference books or scientific textbooks such as Murphy (2022), Murphy (2023) or Krause and Hübotter (2025).

3 Theoretical Background of the Laplace Approximation

Let us start with an insight into the mathematical background of the Laplace approximation. Thereby we will only clarify the line of reasoning and omit the exact mathematical details. For more information and a deeper understanding we will give references.

3.1 Laplace method

(For this part, see the very good explanation in the paper "Laplace's Method Approximations for Probabilistic Inference in Belief Networks with Continuous Variables". Azevedo-Filho and Shachter, 1994).
The Laplace approximation is based on the Laplace method. The Laplace method yields an approximate value for an integral of the following form:

$$\int_a^b b(t) \exp\left(-nr(t)\right)\, dt$$

As a requirement, the function r should be differentiable twice and should have a unique minimum at a point $\hat{t}$ where $\hat{t} \in (a, b)$. The function b should be differentiable; $b(\hat{t}) \neq 0$ should hold. n should be a large positive number. (The approximation for the integral will be the better, the larger n).
The core idea is to substitute the functions r and b by Taylor polynomials which are expanded at the minimum point $\hat{t}$ von r:
The function r is substituted by the Taylor polynomial of degree 2 expanded at $\hat{t}$:

$$r(t) \approx r(\hat{t}) + r'(\hat{t}) \cdot (t - \hat{t}) + \frac{1}{2} \cdot (t - \hat{t})^2$$

In this expression, the linear term $r'(\hat{t}) \cdot (t - \hat{t})$ equals 0, because $\hat{t}$ is the minimum point of r and thus, $r'(\hat{t}) = 0$. So we get

$$r(t) \approx r(\hat{t}) + \frac{1}{2} \cdot (t - \hat{t})^2$$

The function b is replaced by its Taylor polynomial of degree 1 expanded at $\hat{t}$:

$$b(t) \approx b(\hat{t}) + b'(\hat{t}) \cdot (t - \hat{t})$$

Due to these approximations, we get:

$$\int_a^b b(t) \exp(-nr(t))\, dt \approx \int_a^b \left(b(\hat{t}) + b'(\hat{t})\right)(t - \hat{t}) \cdot \exp\left(-n \cdot \left(r(\hat{t}) + \frac{1}{2} \cdot (t - \hat{t})^2\right)\right) dt$$

By further transformation (for details, see Appendix A.1), this term can be written in the following form:

$$\underbrace{C \cdot \int_a^b \exp\left(-n \cdot \frac{1}{2} \cdot (t - \hat{t})^2\right) dt}_{\text{(unnormalized) Gaussian}} + \underbrace{G \cdot \int_a^b t \cdot \exp\left(-n \cdot \frac{1}{2} \cdot (t - \hat{t})^2\right) dt}_{\text{computable integral}}$$

and thus we can compute an approximate value for the integral $\int_a^b b(t) \exp(-nr(t))\, dt$.
Thereby, the computation of

$$C \cdot \int_a^b \exp\left(-n \cdot \frac{1}{2} \cdot (t - \hat{t})^2\right) dt$$

will involve evaluating a Gaussian.
The approximation of the integral will be better with a larger size n.

(Explanation taken from Azevedo-Filho and Shachter, 1994, for more details see there.)

3.2 Method applied to Bayesian statistics

(For this part, see, e. g., "Probabilistic Machine Learning: An Introduction", Section 4.6.8.2, p. 152-153 (Murphy, 2022), or "Probabilistic Artificial Intelligence", Section 5.1, p. 83-84, (Krause and Hübotter, 2025).)

Let us presume the following situation: During our investigations in Bayesian statistics we have worked out the posterior density $p(\boldsymbol{\theta}|\mathcal{D})$ of the parameter vector $\boldsymbol{\theta}$ for the model we have chosen. But $p(\boldsymbol{\theta}|\mathcal{D})$ is rather complex and therefore we want to approximate it by a (multivariate) Gaussian.

Our original posterior density $p(\boldsymbol{\theta}|\mathcal{D})$ is in the following form:

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{Z} \cdot p(\boldsymbol{\theta}, \mathcal{D})$$

where Z is the normalization constant.

In a first step, we bring $p(\boldsymbol{\theta}|\mathcal{D})$ to a form that contains an exponential function:

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{1}{Z} \cdot \exp\left(-\Big(-\log\left(p\left(\boldsymbol{\theta}, \mathcal{D}\right)\right)\Big)\right)$$

For abbreviation we denote $f(\boldsymbol{\theta}) := -\log(p(\boldsymbol{\theta}, \mathcal{D})$.

Now, analogously to the Laplace method in 3.1, we approximate f by the 2^{th}-order Taylor polynomial around the mode $\hat{\boldsymbol{\theta}}$ of f. In usual case, we will need multivariate calculus for doing that, of course; thus, we get

$$f(\boldsymbol{\theta}) \approx f(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \boldsymbol{g} + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \boldsymbol{H} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

In this formula, $\boldsymbol{g}$ denotes the gradient of f at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{H}$ denotes the Hessian at $\hat{\boldsymbol{\theta}}$.
In our context, $\boldsymbol{g} = \mathbf{0}$ is true, because $\hat{\boldsymbol{\theta}}$ is the mode of f. Thus, we get

$$f(\boldsymbol{\theta}) \approx f(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \boldsymbol{H} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

With the backward transformation $p(\boldsymbol{\theta}|\mathcal{D}) = \exp(-f(\boldsymbol{\theta}))$ we get

$$p(\boldsymbol{\theta}|\mathcal{D}) \approx \exp f(\hat{\boldsymbol{\theta}}) \cdot \exp \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \boldsymbol{H} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

Since $f(\hat{\boldsymbol{\theta}})$ is constant and $\exp \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \boldsymbol{H} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ is the kernel of the Gaussian distribution with mean $\hat{\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{H^{-1}}$ we find the Laplace approximation $\hat{p}(\boldsymbol{\theta}|\mathcal{D})$ as

$$\hat{p}(\boldsymbol{\theta}|\mathcal{D}) := \mathcal{N}(\hat{\boldsymbol{\theta}}, \boldsymbol{H^{-1}})$$

(Derivation taken from Murphy (2022), p. 152-153, and Krause and Hübotter (2025), p. 83-84).

3.3 Background: Bayesian Central Limit Theorem

In the integral considered in the Laplace method in 3.1, there is a parameter n. This n should be a large number because then the approximation will be good. In this light, it seems strange that in 3.2 no such n appears (or $n = 1$ holds, which one would not refer to as "a large number"). Can we really be sure that the approximation $\hat{p}(\boldsymbol{\theta}|\mathcal{D})$ we found in 3.2 is a good approximation?

For this question, let us go back to the transformation in 3.2

$$p(\boldsymbol{\theta}|\mathcal{D}) = \exp\left(-(-\log(p(\boldsymbol{\theta}|\mathcal{D})))\right)$$

Instead of this, we write now

$$p(\boldsymbol{\theta}|\mathcal{D}) = \exp\left(-n(-\frac{1}{n}\log(p(\boldsymbol{\theta}|\mathcal{D})))\right)$$

Then let us do the Taylor expansion with

$$f(\boldsymbol{\theta}) := -\frac{1}{n}\log(p(\boldsymbol{\theta}|\mathcal{D}))$$

Finally, we will get the same formula for $\hat{p}(\boldsymbol{\theta}|\mathcal{D})$ as in 3.2, because n and $\frac{1}{n}$ will reduce.

But now, the question might arise whether the effect of a large n will be eliminated by the downsizing effect of $\frac{1}{n}$.

At this point, the Bernstein - von Mises Theorem (also known as Bayesian Central Limit Theorem) gives us a theoretical background that under certain requirements (referred to as "Laplace regularity") and mild assumptions the asymptotic normality of the posterior holds (see Azevedo-Filho and Shachter, 1994).
In this context, the number n denotes the sample size of the sample $\mathcal{D}$.

Thus, the Laplace approximation is a analytical and deterministic method when applied to a single posterior function. But let us consider a series of posteriors, each of them being the prior of the next one, and each having a larger sample size n than the previous one. When we apply the Laplace approximation to each of them, then the Laplace approximation will tend to be more and more exact, when the requirements of the Bernstein - von Mises Theorem hold. So, in this case, the Laplace approximation can be seen in some respects a bit like a repetitive approximation method.

4 Methodical Approach of Laplace Approximation

4.1 The Necessary Single Steps

Now we shall look at the specific steps necessary to compute the Laplace approximation for some known function h. Therefore, for the basic case, let us assume that we have a density function h, (e.g. some posterior $p(\theta|D)$ that we came across during Bayesian analysis) and we want to find the Laplace approximation for it.
h may be a multivariate function. For the algorithm, the function values of h must be positive. In mathematical notation, we note $h : \Theta \to \mathbb{R}^+, \boldsymbol{\theta} \mapsto \mathbb{R}$, with $\Theta \subseteq \mathbb{R}^d$. As another requirement of the algorithm, the function h must be continuously differentiable twice. In the context here, we omit additional specific mathematical requirements for h or for Θ at this point.
For calculating the Laplace approximation of h, perform the following steps:

(A) Preparation

In the very first step, it is necessary to form the negative logarithm $\theta \mapsto -log(h(\theta))$ of the function h. Let us denote $f : \theta \mapsto f(\theta) := -log(h(\theta))$.

(B) Optimization
Find the local minimum $\hat{\theta}$ of f. (It is required that $\hat{\theta}$ is in the interior of Θ, otherwise the algorithm might not perform correctly.)
Remark: Note that the minimum of f corresponds to the maximum of h.

(C) Hessian
Compute the Hessian $H = \nabla^2 f|_{\theta=\hat{\theta}}$ of h evaluated at the point $\theta = \hat{\theta}$.

(D) Inverse Hessian
Calculate the inverse H^{-1} of the matrix H.

(E) Result
The Laplace approximation of the density function f is given by $\mathcal{N}(\hat{\theta}, -H^{-1})$.

4.2 Possible Issues

Computing the Hessian and calculating the inverse of it may be very complex, when the dimension of the parameter space is high. In many real-world usage scenarios, it will be necessary to resort to numerical methods when performing (C) or (D). Even for finding the minimum in (B) in much cases numerical methods will be preferred.

In (A), there are problems if the function h has negative values. Tierney et al. (1989) addressed this in their paper.

5 Some Basic Illustrating Examples

For illustration, we will apply the Laplace approximation to some probability density function which we know exactly. Thus, we can compare the original function (in the graphics: red, solid line) with the Laplace approximation (in the graphics: blue, dashed line).

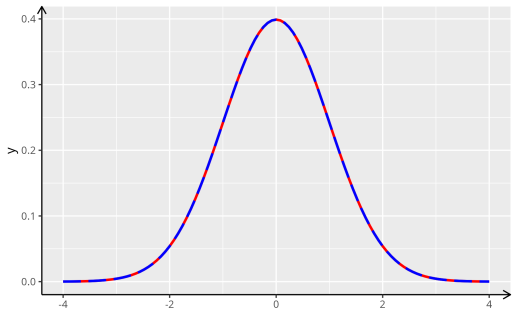
The Laplace approximation approximates an arbitrary probability distribution p by a Gaussian distribution. Therefore, we expect that the goodness-of-fit will be the better the more p is similar to a Gaussian.
For example, the goodness-of-fit will become worse when the distribution p is

- heavy-tailed
- skewed
- multimodal

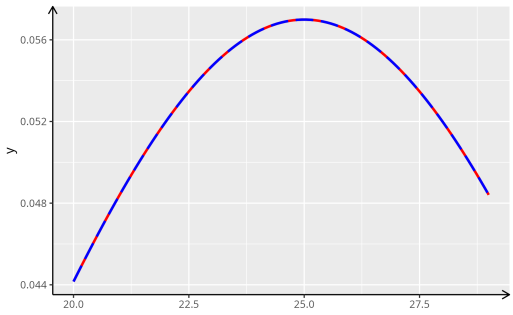
We illustrate this using some distributions p, where we know the properties, e. g. Student's t distribution or the gamma distribution. (Of course, in real-life scenarios, you would normally not do a Laplace approximation for these distributions.)

5.1 Gaussian Distribution

For Gaussian distributions, the Laplace approximation is exact. That is not surprising because of the way the Laplace approximation is constructed. Let us just verify this for two Gaussian distributions with mean μ and standard deviation σ, see Figure 1.



Gaussian distribution,
with $\mu = 0$ and $\sigma = 1$

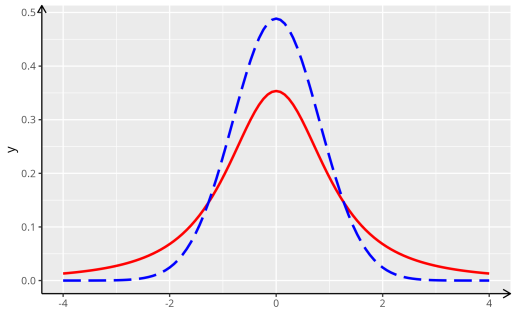


Gaussian distribution,
with $\mu = 25$ and $\sigma = 7$,
picture detail

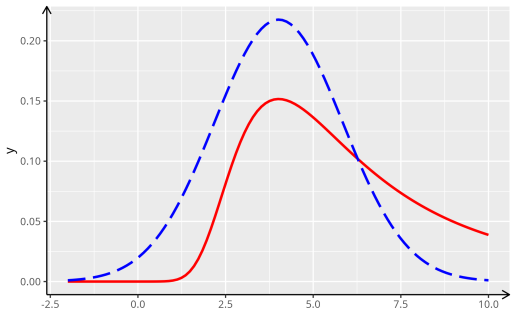
Figure 1: Laplace approximation for Gaussian distribution

5.2 Student's t-Distribution

For the t distribution, the Laplace approximation is bad when the degrees of freedom are very few. It becomes even worse if the non-centrality parameter δ does not equal zero. See Figure 2.



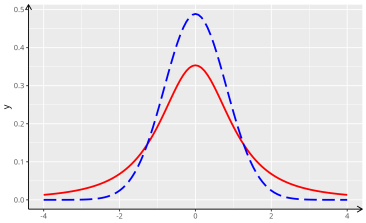
Student's t-distribution,
with $df = 2$ and $\delta = 0$



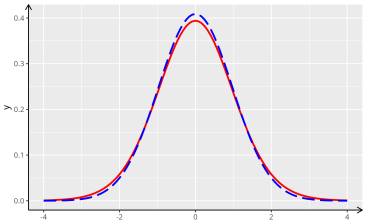
Student's t-distribution,
with $df = 2$ and $\delta = 5$

Figure 2: Laplace approximation for Student's t distribution with few degrees of freedom

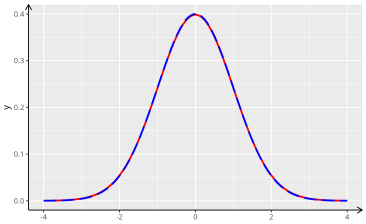
When the degrees of freedom increase, Student's t-distribution gets closer to a normal distribution. Thus, we expect an increasing goodness-of-fit of the Laplace approximation when we raise the degrees of freedom. Verify this with the example shown in Figure 3.



Student's t-distribution, with $df = 2$ and $\delta = 0$



Student's t-distribution, with $df = 20$ and $\delta = 0$

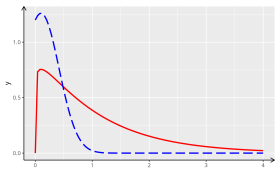


Student's t-distribution, with $df = 200$ and $\delta = 0$

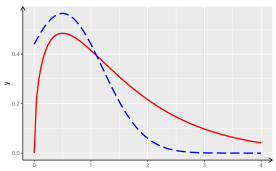
Figure 3: Laplace approximation for Student's t distribution with different degrees of freedom

5.3 Gamma distribution

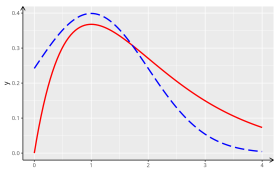
The shape of a Gamma distribution depends on the two parameters a ("shape") and λ ("rate"). It is clearly different from the shape of a Gaussian, because it is not symmetric, only defined for positive $x-$values and skewed positive.
Thus, we do not expect the Laplace approximation to fit very well. Indeed, for $a \leq 1$ the Laplace approximation will not work at all because in this case the Gamma function does not have a maximum.
The graphics of Figure 4 illustrate the Laplace approximation for four examples of Gamma distributions with different shape parameters a and with rate parameter $\lambda = 1$ in each example.



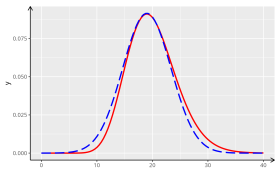
Gamma distribution, with $a = 1.1$



Gamma distribution, with $a = 1.5$



Gamma distribution, with $a = 2$



Gamma distribution, with $a = 20$

Figure 4: Laplace approximation for Gamma distributions with different shape parameters

5.4 Multimodal Distributions

The Laplace method in its basic form is not appropriate for distributions p that have more than one mode. The Gaussian $\hat{p}$ that we fit by the Laplace approximation will concentrate around one of the peaks. It is not sure that it will be the highest peak. The other peaks will not be taken into account.

To illustration, see Figure 5 and Figure 6.

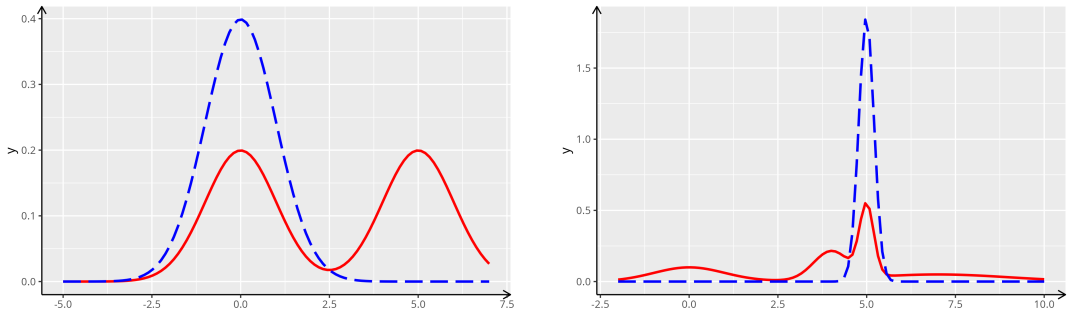


Figure 5: Laplace approximation applied to multimodal functions

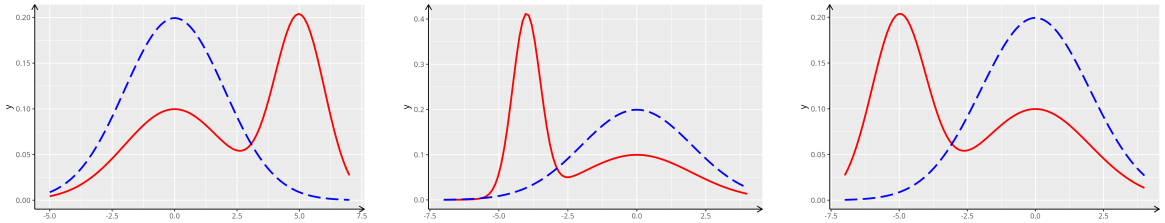


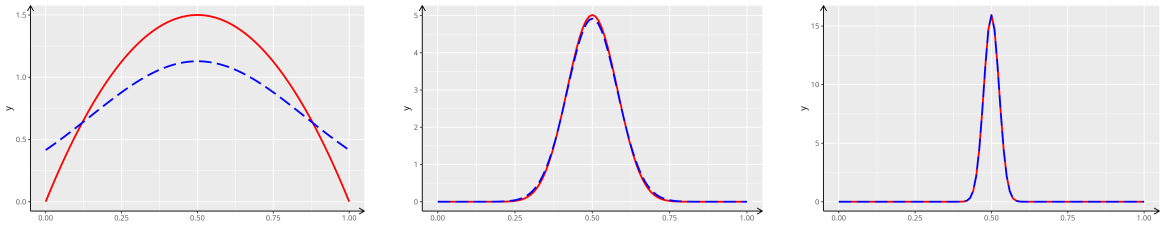
Figure 6: The goodness-of-fit can be rather bad for multimodal functions.

5.5 Beta Distribution

The beta distribution $B(\alpha, \beta)$ depends on two parameters. If $\alpha = \beta$, the beta distribution will be symmetric. In the context of a beta binomial model, the sum $\alpha + \beta$ corresponds to the number n of trials (i. e. to the sample size).
Thus, we expect that the goodness-of-fit becomes better when

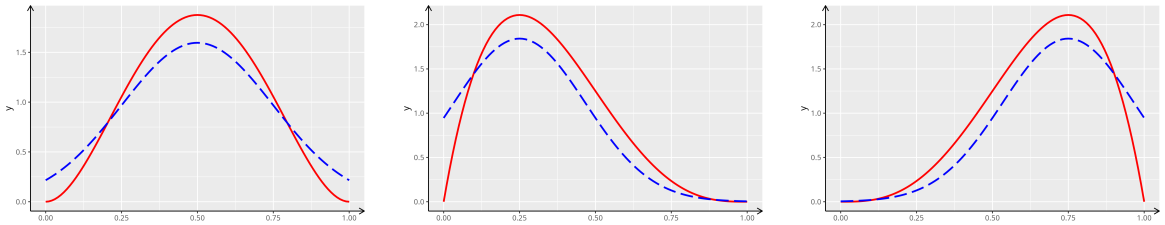
- the sum $\alpha + \beta$ is bigger
- α are more similar.

You might verify this in checking Figure 7, Figure 8 and Figure 9.



Beta distribution, with $\alpha = 2$, $\beta = 2$ | Beta distribution, with $\alpha = 20$, $\beta = 20$ | Beta distribution, with $\alpha = 200$, $\beta = 200$

Figure 7: symmetric beta distributions; $\alpha = \beta$

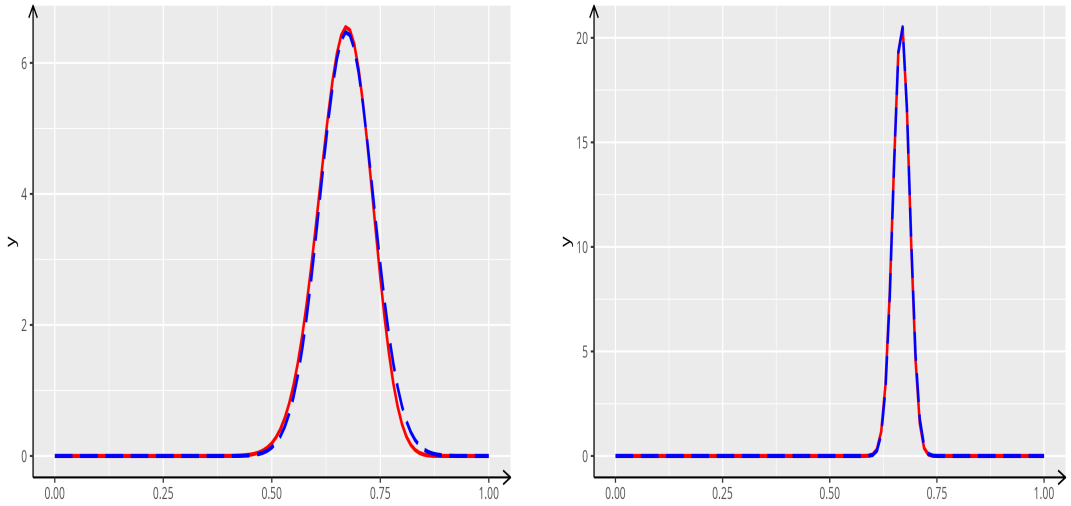


Beta distribution,
$\alpha = 3$, $\beta = 3$

Beta distribution,
$\alpha = 2$, $\beta = 4$

Beta distribution,
$\alpha = 4$, $\beta = 2$

Figure 8: one symmetric and two nonsymmetric beta distributions; $\alpha + \beta = \text{const}$ here.



Beta distribution,
$\alpha = 40$, $\beta = 20$

Beta distribution,
$\alpha = 400$, $\beta = 200$

Figure 9: nonsymmetric beta distributions, $\alpha : \beta = \text{const}$, increasing $\alpha + \beta$

A Appendix

A.1 Transforming the Integral in 3.1

$$\int_a^b b(t)\exp(-nr(t))\,dt \approx$$

$$\approx \int_a^b \Big(b(\hat{t}) + b'(\hat{t})(t-\hat{t})\Big)\cdot\exp\left(-n\cdot\left(r(\hat{t})+\frac{1}{2}\cdot(t-\hat{t})^2\right)\right)\,dt$$

$$= \int_a^b \Big(b(\hat{t}) + b'(\hat{t})t - b'(\hat{t})\hat{t}\Big)\cdot\exp\left((-n\cdot)r(\hat{t})+(-n)\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

$$= \int_a^b \Big(b(\hat{t}) - b'(\hat{t})\hat{t} + b'(\hat{t})t\Big)\cdot\exp\big(-n\cdot r(\hat{t})\big)\cdot\exp\Big(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\Big)\,dt$$

$$= \exp\left(-n\cdot r(\hat{t})\right)\cdot\int_a^b \Big(b(\hat{t}) - b'(\hat{t})\hat{t} + b'(\hat{t})t\Big)\cdot\exp\left(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

The last transformation was feasible, because $\hat{t}$ is constant.
With $K := \exp\big(-n\cdot r(\hat{t})\big)$, we can continue:

$$\int_a^b b(t)\exp(-nr(t))\,dt \approx$$

$$\approx K\cdot\int_a^b \Big(b(\hat{t}) - b'(\hat{t})\hat{t} + b'(\hat{t})t\Big)\cdot\exp\left(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

$$= K\cdot\int_a^b \Big(b(\hat{t}) - b'(\hat{t})\hat{t}\Big)\cdot\exp\left(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

$$+ K\cdot\int_a^b b'(\hat{t})\cdot t\cdot\exp\left(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

$$= K\cdot\Big(b(\hat{t}) - b'(\hat{t})\hat{t}\Big)\int_a^b \exp\left(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

$$+ K\cdot b'(\hat{t})\int_a^b t\cdot\exp\left(-n\cdot\frac{1}{2}\cdot(t-\hat{t})^2\right)\,dt$$

With $C := K \cdot \left(b(\hat{t}) - b'(\hat{t})\hat{t}\right)$ and $G := K \cdot b'(\hat{t})$, we get

$$\int_a^b b(t)\exp(-nr(t))\,dt \approx \underbrace{C \cdot \int_a^b \exp\left(-n \cdot \frac{1}{2} \cdot (t-\hat{t})^2\right) dt}_{\text{(unnormalized) Gaussian}} + \underbrace{G \cdot \int_a^b t \cdot \exp\left(-n \cdot \frac{1}{2} \cdot (t-\hat{t})^2\right) dt}_{\text{computable}}$$

and thus, we can compute an approximate value for the integral.

B Electronic appendix

Code and figures are provided in electronic form.
See
`https://github.com/RenateSchmeidlerLMUCampus/ProbabilisticML_LaplaceApproximation`

C Acknowledgements

I used AI for a brainstorming at the beginning and, at several passages in the text, for finding appropriate translations to English.

References

Azevedo-Filho, A. and Shachter, R. D. (1994). Laplace's method approximations for probabilistic inferencein belief networks with continuous variables, *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, UAI'94, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 28–36. event-place: Seattle, WA.

Kasprzak, M. J., Giordano, R. and Broderick, T. (2025). How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences. arXiv:2209.14992 [math].
URL: *http://arxiv.org/abs/2209.14992*

Katsevich, A. (2024). Improved dimension dependence in the Bernstein von Mises Theorem via a new Laplace approximation bound. arXiv:2308.06899 [math].
URL: *http://arxiv.org/abs/2308.06899*

Krause, A. and Hübotter, J. (2025). Probabilistic Artificial Intelligence. Version Number: 1.
URL: *https://arxiv.org/abs/2502.05244*

Laplace, P. S. (1986). Memoir on the Probability of the Causes of Events, *Statistical Science* **1**(3): 364–378. Publisher: Institute of Mathematical Statistics.
URL: *http://www.jstor.org/stable/2245476*

McDonald, S. and Campbell, D. (2024). A probabilistic diagnostic for Laplace approximations: Introduction and experimentation. arXiv:2411.01697 [stat].
URL: *http://arxiv.org/abs/2411.01697*

Murphy, K. P. (2022). *Probabilistic Machine Learning: An introduction*, MIT Press.
URL: *http://probml.github.io/book1*

Murphy, K. P. (2023). *Probabilistic Machine Learning: Advanced Topics*, MIT Press.
URL: *http://probml.github.io/book2*

Shun, Z. and McCullagh, P. (1995). Laplace Approximation of High Dimensional Integrals, *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(4): 749–760. Publisher: [Royal Statistical Society, Oxford University Press].
URL: *http://www.jstor.org/stable/2345941*

Tierney, L. and Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistical Association* **81**(393): 82–86. Number: 393.
URL: *http://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478240*

Tierney, L., Kass, R. E. and Kadane, J. B. (1989). Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions, *Journal of the American Statistical Association* **84**(407): 710–716.
URL: *https://www.tandfonline.com/doi/full/10.1080/01621459.1989.10478824*

Declaration of authorship

I hereby declare that the report submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the Thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the report as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future Theses submitted. Further rights of reproduction and usage, however, are not granted here. This paper was not previously presented to another examination board and has not been published.

Location, date

Name