Master's Programme in Data Science

VMBC Report

Tuomas Vuontisjarvi

September 27, 2025

University of Helsinki
Faculty of Science

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Data Science

Tekijä — Författare — Author

Tuomas Vuontisjarvi

Työn nimi — Arbetets titel — Title

VMBC Report

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Report	September 27, 2025	8

Tiivistelmä — Referat — Abstract

This report is about Variational Bayesian Monte Carlo (VBMC), a method for performing Bayesian inference with complex and computationally expensive black-box models. Key concepts related to the VMBC are explained to provide clear understanding of understanding. With those definitions explained, this report will explore the use cases for the algorithm and cases were the algorithm is unfit for, or if there are better methods available. Finally the usage of the algorithm is explored with the provided pyVMBC python package.

Avainsanat — Nyckelord — Keywords

layout, summary, list of references

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

1. Introduction

According to Acerbi (2018), a significant problem with probabilistic models that have expensive, black-box likelihoods is that the characteristics prevent the usage of standard techniques for Bayesian inference.

In order to address the problem of high computational cost, a novel sample efficient method has been introduced, called Variational Bayesian Monte Carlo (VMBC). Acerbi [2] claims that the VMBC solves the previously costly problem by combining variational inference with Bayesian quadrature, solving model posteriors efficiently and with a relatively small amount of sampling.

This report aims to explain the VMBC by exploring the key concepts behind it. Chapter three focuses on explaining all the relevant concepts and providing examples and on the way. The goal is to build a clear picture of the background domain and the issues researchers face.

Chapter four will explore the VMBC in more detail with the advantage of clearly defined concepts. The aim is to explain how it works and what kind of problems it solves. Also the chapter will consider problems that the VMBC is ill suited for.

Chapter five will dive deeper into the workings of the VMBC by using the python package pyVBMC provided by the author in another paper.

2. Key concepts and VMBC explained

i. Black-box models

ii. Bayesian inference

iii. Approximate inference methods

iv. Model posteriors

v. Gaussian process

vi. Monte Carlo algorithms

The VMBC is used for complex and computationally expensive black-box models. Acerbi [2][3] notes a few examples of such models, such as computational neuroscience, biology and big data models. The algorithm is a novel approximate inference method for learning about black-box models. A model is black-box model when there is no access to its inner processes. This means that the model can be viewed completely in terms of its inputs and outputs.

One way of learning about the model is Bayesian inference, which is a method for computing posterior distribution over parameters and the model evidence. However, since Bayesian inference is generally analytically intractable [2], statistical approximate inference methods are often used. These methods include Markov Chain Monte Carlo algorithms and variational inference.

As Acerbi [2] notes, existing methods of approximate inference, such as above examples, generally require knowledge about the model in order to produce approximate inference. When a method requires more knowledge about the model than just inputs and outputs, by definition it can't be applied to black-box models. Some methods can bypass this requirement when given a very large number of model evaluations.

A computationally expensive black-box model is a model where evaluating the model is time consuming, which means that there generally isn't access to large number of model evaluations. Therefore the existing methods for approximate Bayesian inference are unfit for computationally expensive black-box models. Expensive model is defined as one evaluation taking one second or more per evaluation[1]

From the point of view of the pyVBMC, a python package for performing VMBC model and posterior inference, the black-box model is provided as Python function, which calculates target log likelihood of the black-box model.

The VMBC produces a flexible approximate posterior distribution of the model parameters [2]. The posterior distribution is a joint probability distribution which describes how plausible each parameter is given the observed data. The posterior is expressed as $p(x \mid \mathcal{D})$, where $\mathcal{D}$ is the dataset or the evidence and $x \in \mathbb{R}^D$, which are the black-box model parameters.

The black-box model for the algorithm is expressed as $p(\mathcal{D} \mid x)$. This means that the input for the algorithm is a likelihood function which provides the log-likelihood for seeing the evidence with the given priors.

3. VMBC use cases and fallouts

4. pyVMBC examples

5. Conclusions

References

[1] Bobby Huggins,Chengkun Li, Marlon Tobaben, Mikko J. Aarnos and Luigi Acerbi. Pyvbmc: Efficient bayesian inference in python, 2018. https://arxiv.org/pdf/2303.09519.

[2] Luigi Acerbi. Variational bayesian monte carlo, 2018. https://arxiv.org/pdf/1810.05558.

[3] Luigi Acerbi. Variational bayesian monte carlo with noisy likelihoods, 2020. https://arxiv.org/pdf/2006.08655.