BAYESIAN LEARNING

ECOLE CENTRALE LILLE

MASTER DATA SCIENCE

Bayesian learning

Authors:
Ayoub Youssoufi
Yassin El Hajj Chehade

Professor:
Rémi Bardenet

December 14, 2022

1 Introduction

Monte Carlo integration is a simple but rarely feasible method for estimating parameters using an assumed posterior distribution. The difficulty of Monte Carlo integration is that it requires that the posterior distribution can be directly drawn from. This is something that occurs only in a few special cases. The Bayesian Monte Carlo (BMC) method is very different in its efficiency and effectiveness in providing useful approximations for accurate inference in Bayesian applications. In this paper, we compare these two methods Simple Monte Carlo (SMC) and Bayesian Monte Carlo (BMC) in order to show that BMC outperform the classical SMC method.

2 Paper content

2.1 Simple Monte Carlo

In general, numerically evaluating the posterior expectation is challenging:

$$\hat{f_n} = \int f(x)p(x)dx$$

of a real-valued random variable X on a probability space. The expectation is approximated by taking the sample mean of independent observations Xi with the same distribution as X:

$$\hat{f_n} \approx \frac{1}{n}\sum Xi$$

if sampling from $p(x)$ is hard we use $q(x)$ to obtain the estimate :

$$\hat{f_n} \approx \frac{1}{n}\sum \frac{f(x)p(x)}{q(x)}q(x)$$

2.2 Bayesian Monte Carlo

Let's consider the prior over the function $p(f)$ nd makes inferences about f from a set of samples D giving the posterior distribution $p(f|D)$. Under the Gaussian process the prior posterior is Gaussian.

$$\mathsf{E}_{f|D}(f_n) = \int\int f(x)p(x)dxp(f|D)df = \int \hat{f}_p(x)p(x)dx$$

$$V_{f|D}(f_n) = \int\int Cov_D(f(x), f(x'))p(x)p(x')dxdx'$$

with :

$$\hat{f_n}(x) = k(x,x)K^{-}1f$$

and $Cov_D(f(x), f(x')) = k(x,x')_k(x,x)K^{-}1k(x,x')$
At the end, we infer the gaussian posterior f_n by its expectation and variance.

2.3 Optimal sample monte carlo

To improve efficiency of SMC, we generate independent samples from more-cleverly designed distributions. Recall the SMC formula with importance sampling $q(x)$:

$$\hat{f_D} \approx \frac{1}{n}\sum \frac{f(x)p(x)}{q(x)}$$

where : $q(x) > 0$ wherever $p(x) > 0$
Then the variance of this estimator is :

$$V(f_D) = \frac{1}{n}[\int \frac{f(x)^2 p(x)^2}{q(x)}dx - \hat{f_D}^2]$$

In order to optimize the result, we minimize the variance and the optimal $q(x)$ will be:

$$q^*(x) = \frac{|f(x)|p(x)}{\int |f(x')|p(x')dx'}$$

Hence, we will replace q^* in the above formula and we get the improved version of SMC.

2.4 Experiences and results:

Figure 1 shows the experience done in the paper where we measure $f(x)$ using the Gaussian distribution $p(x)$. We see that average squared error of the BMC decrease more than SMC when increasing the sample size. Therefore, the BMC outperforms the SMC. We also notice that the variance of the BMC is high for small sample size and it gets smaller when the sample size increase. Moreover the Optimal Importance simpler outperform the simple monte carlo as expected.

2.5 Implementation

We tried to implement the above experience with our function $f(x)$. We didn't get the same result as the author, especially because of the difficult computation that we approximated it and then it affects the final result. Also this bad result comes from the fact that our hyper-parameters are not tuned. The code is in the notebook "bayesian monte carlo".

2.6 Conclusion:

- BMC performs better than SMC. However, in case of high dimensional integrands, BMC requires thousands of samples to limit the variation of the variance which is not suitable for Gaussian Process as it is limited to only few samples.
- The BMC method requires that the distribution $p(x)$ can be evaluated. This contrasts with SMC where it only requires that samples can be drawn from some distribution $q(x)$.
- Despite the limitation of the approximation of complex integral using BMC. The BMC method outperform state-of-art classical methods.

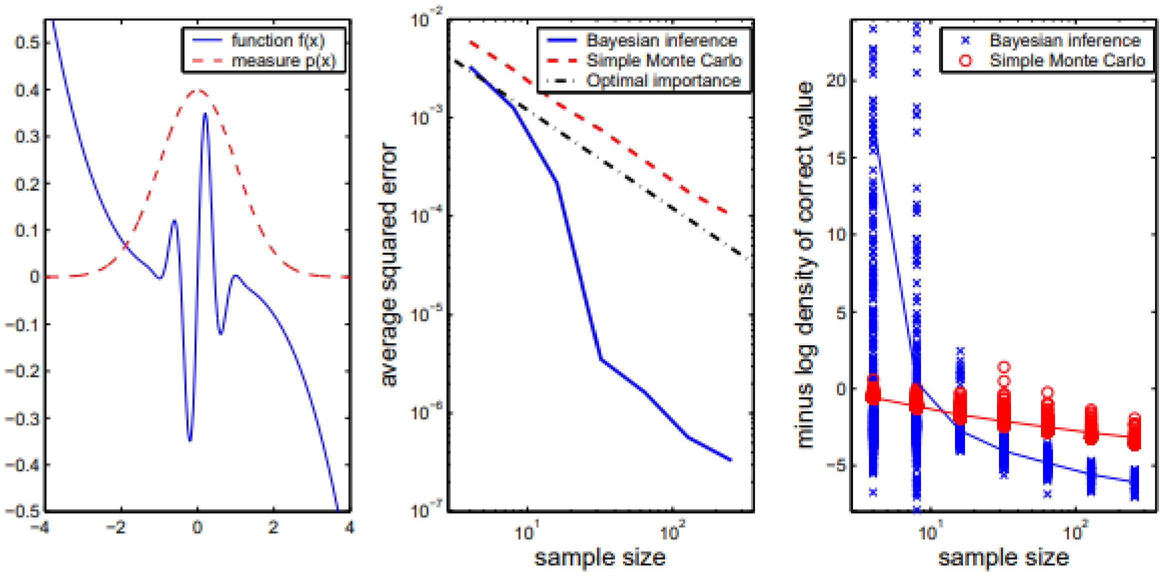

Figure 1: Comparison between Bayesian Monte Carlo and Simple Monte Carlo