
Title

Gian Carlo Diluvi
STAT 520C Final Project
April 2023

1 Introduction

The Bayesian statistical framework provides practitioners with a principled means to obtain finite-sample uncertainty quantification guarantees. Specifically, the posterior distribution encodes the uncertainty around the unobserved quantities given the observations and a prior distribution. Posterior credible intervals (CIs) or more general regions allow practitioners to quantify the accuracy of point estimates.

The quality of credible intervals can be measured by the proportion of times they contain the “true” value of the parameter. This is known as the frequentist coverage, and Bayesian credible intervals are asymptotically exact in this sense as a consequence of the Bernstein-von Mises theorem. Another way to assess the quality of an interval is via its *Bayesian* coverage, i.e., the proportion of times the interval contains the parameter value that generated the data and itself is a realization from the prior distribution. The Bayesian coverage of an interval leverages the fact that the prior distribution quantifies the uncertainty around the parameter before observing data.

In class, we showed that credible intervals attain nominal Bayesian coverage (i.e., their coverage is equal to the credibility level) for any sample size and without any regularity conditions. This follows from the definition of Bayesian coverage and credible intervals. However, this result assumes that we have access to *exact* credible intervals. For all but the simplest of models, however, the posterior distribution is intractable and has to be approximated numerically. In this setting, the resulting credible intervals are approximate and their Bayesian coverage need not be exact.

Variational inference (VI) is a scalable framework to learn posterior distributions. Succinctly, VI casts inference as an optimization problem by approximating the posterior with an element of a family of candidate approximations that minimizes some divergence to the posterior. The quality of the approximation depends on the flexibility of the family of approximations, the divergence being minimized, and whether the optimization problem can be solved reliably. A simple instantiation of VI consists on finding the best Gaussian approximation to the posterior distribution (a framework called Gaussian VI). In this case, the optimization problem is usually amenable to off-the-shelf stochastic optimization algorithms for many common divergences.

In this work, I study the Bayesian coverage of credible intervals of Gaussian VI when minimizing the Kullback-Leibler (KL) divergence. Specifically, I focus on the impact of minimizing the reverse KL (from approximation to target) versus the forward KL (from target to approximation). The former is known to produce approximations that underestimate the posterior variance and viceversa. Through two simulation studies, I conclude that the forward KL tends to have better Bayesian coverage when

the posterior has heavier-than-Gaussian tails, which is a common occurrence (e.g., in Bayesian logistic regression, see Section 3.2.)

2 Background

3 Experiments

3.1 Cauchy

3.2 Logistics regression

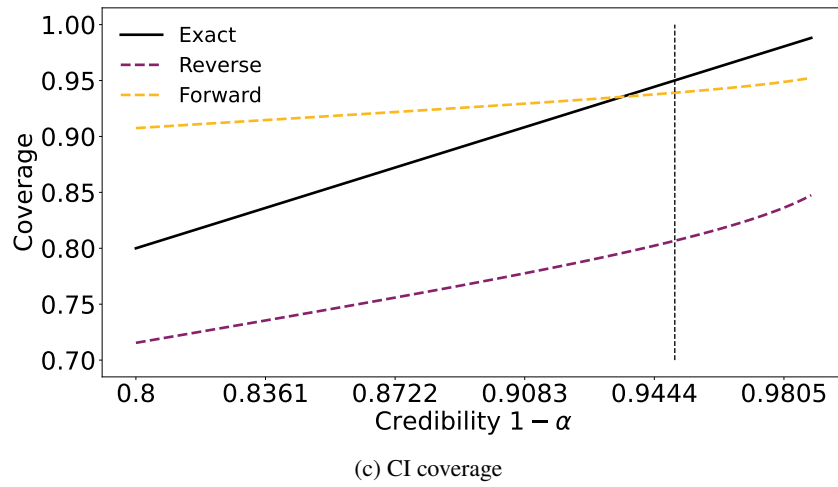
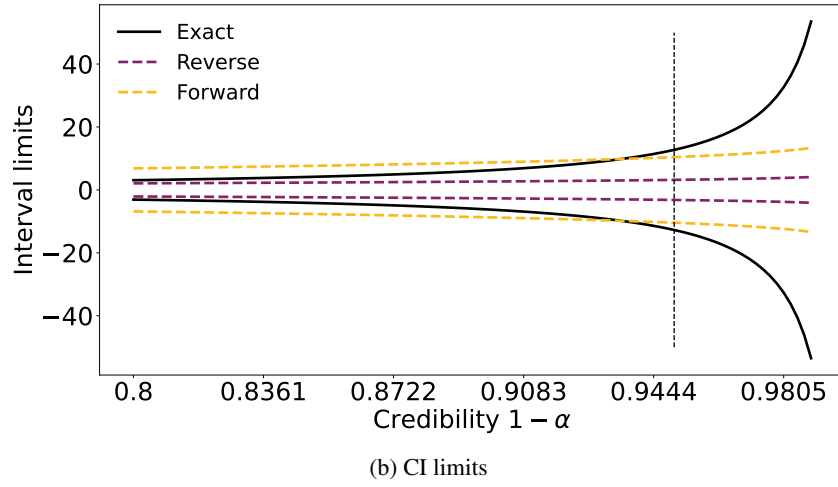
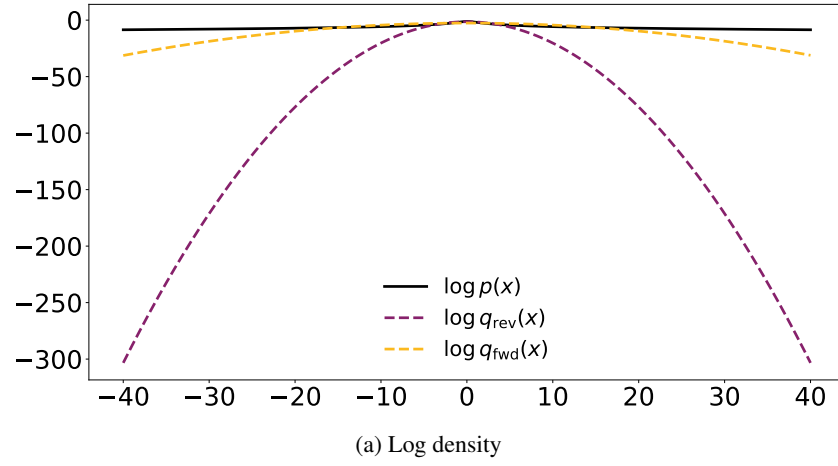


Figure 1: Results on the Cauchy distribution.

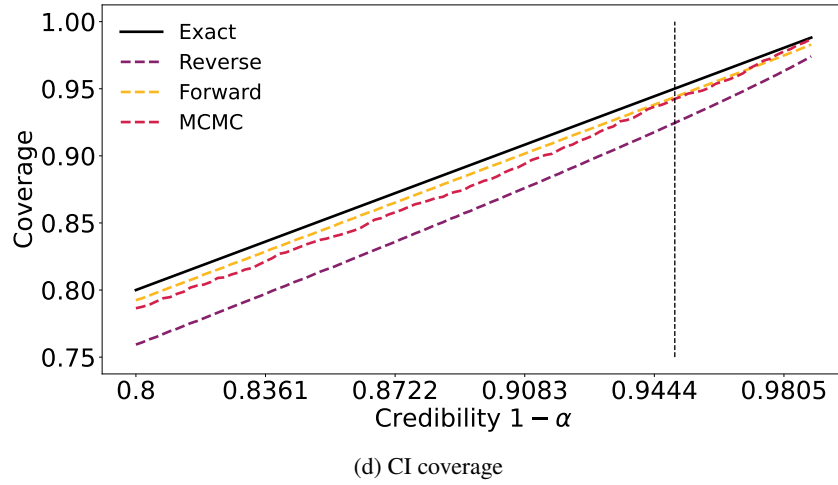
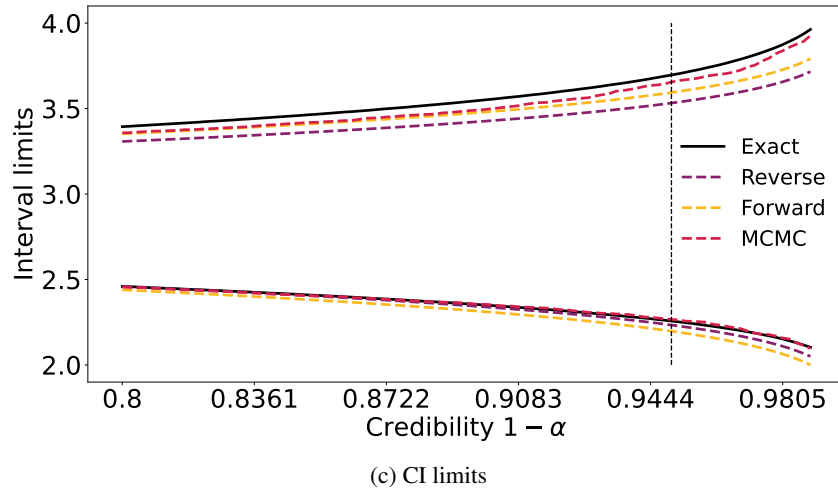
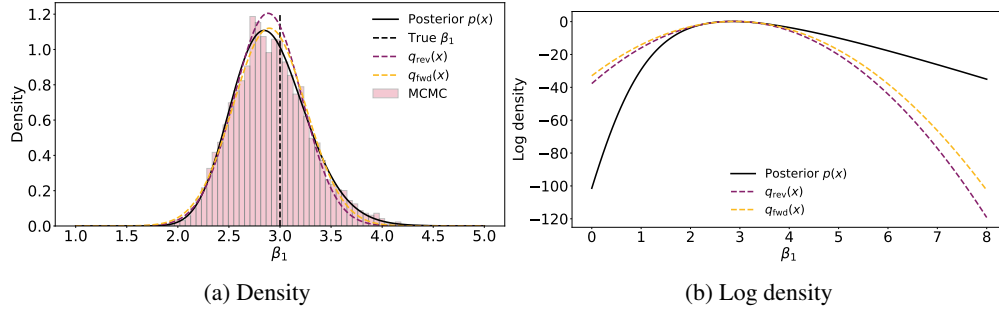


Figure 2: Results on the logistic regression example.

References