Bayesian Quadrature for Ratios

Michael A. Osborne[†], Roman Garnett[^], Stephen J. Roberts[†], Christopher Hart[‡], Suzanne Aigrain[‡], Neale P. Gibson[‡]

[†] Engineering Science, University of Oxford; [^] Robotics Institute, Carnegie Mellon University; [‡] Physics, University of Oxford

Bayesian Quadrature (BQ) is a model-based approach to resolving non-analytic integrals, a common and important problem. The integrand is modelled using a Gaussian process (GP), such that samples of the integrand can be used to infer the value of the integral. Relative to common Monte Carlo (MC) approaches, this probabilistic approach permits more information to be gained from each integrand sample. This is useful when evaluating a sample is expensive, such as for large datasets.

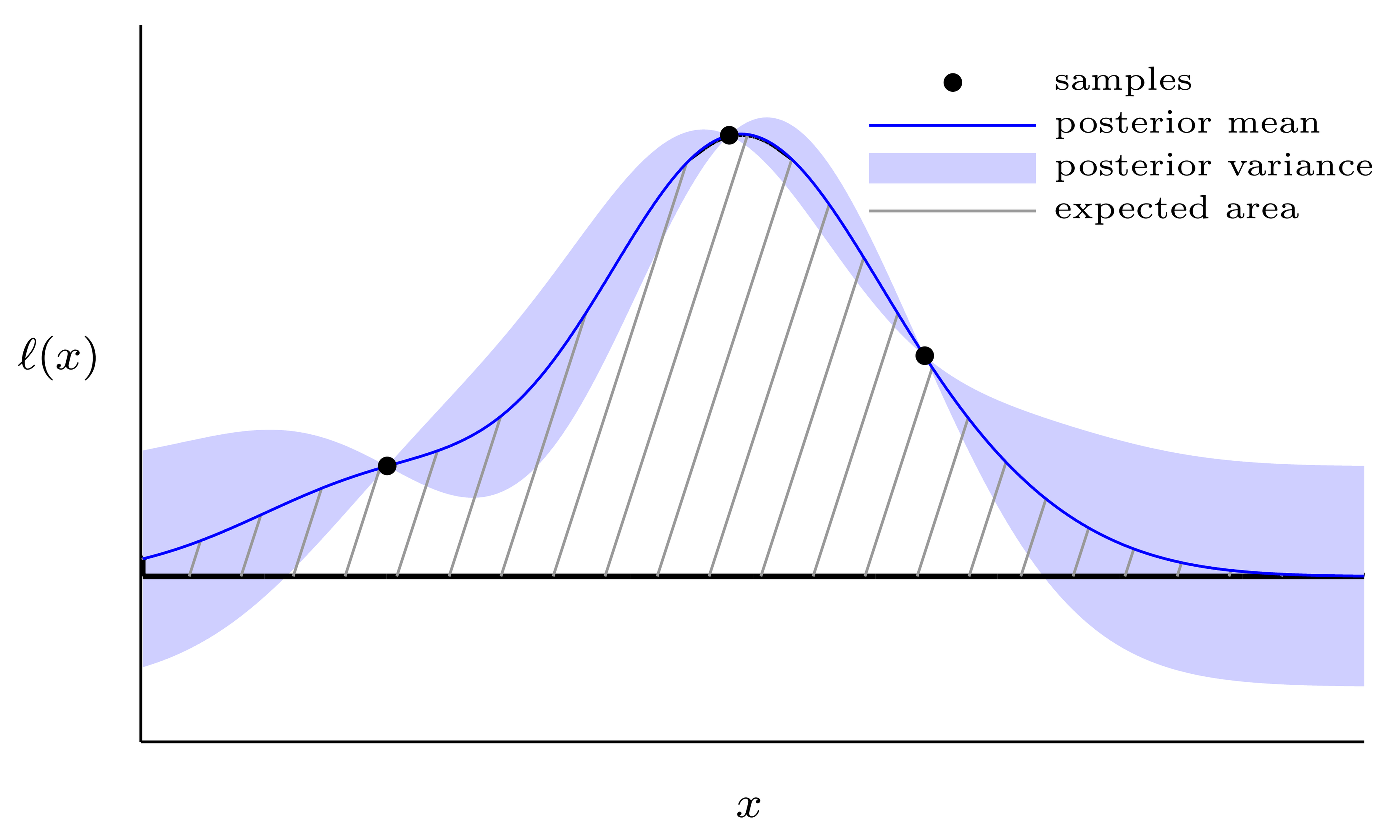

Figure 1*: A cartoon of Bayesian Quadrature.

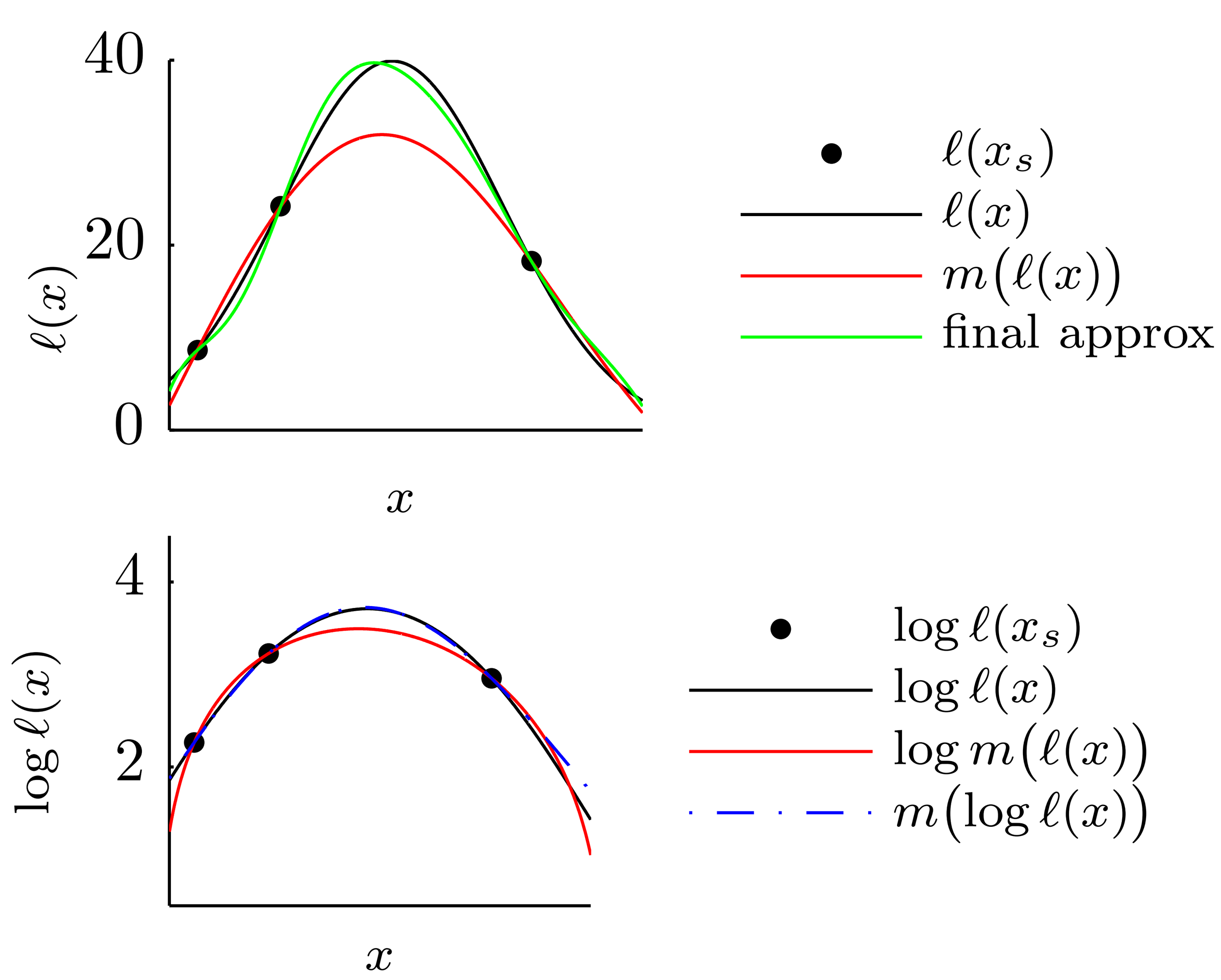

Figure 2*: Our approximate use of a GP over the logarithm of a probability density function $\ell(x)$ better captures the large dynamic range and non-negativity of such quantities than a GP over $\ell(x)$ itself.

(*) We thank David Duvenaud for the use of these figures.

However, a Gaussian process is a poor model for a probability distribution, as it is unable to enforce non-negativity. Additionally, the niceties that enable analytic inference for BQ break down if we try to estimate the ratio of two integrals with common terms, such as

$$p(f \mid y) = \frac{\int p(f \mid y, x)\, p(y \mid x)\, p(x)\, \mathrm{d}x}{\int p(y \mid x)\, p(x)\, \mathrm{d}x}$$

where we wish to model the correlations due to the shared term, $\ell(x) := p(y \mid x)$. Such ratios often occur when marginalising (or integrating over) the hyper-parameters of a probabilistic model.

We propose a modification of BQ, called Bayesian Quadrature for Ratios (BQR). We model the logarithm of the terms in our integrand using a GP, a more natural model that reflects the non-negativity of probability distributions. In order to effect approximate inference for a ratio of probabilistic integrals, we introduce a linearisation of the ratio, which we treat as a functional of the terms in the integrand.

We tested our approach to estimating the ratio of integrals, BQR, against a number of alternatives. Firstly, traditional MC; secondly, maximum likelihood (ML), which approximates our integrands as delta functions; thirdly, naïve BQ, NBQ, which acknowledges neither the correlation between integrands nor their non-negativity; and finally, correlated BQ, BQZ, which acknowledges such correlations but not non-negativity. Results show that MC estimates converge more slowly than BQ approaches and that ML produces good predictive means but is prone to under-estimating predictive variances. Constraining functions to be positive was less significant than modelling correlations, although both improved performance. BQR was the most broadly successful of tested methods.

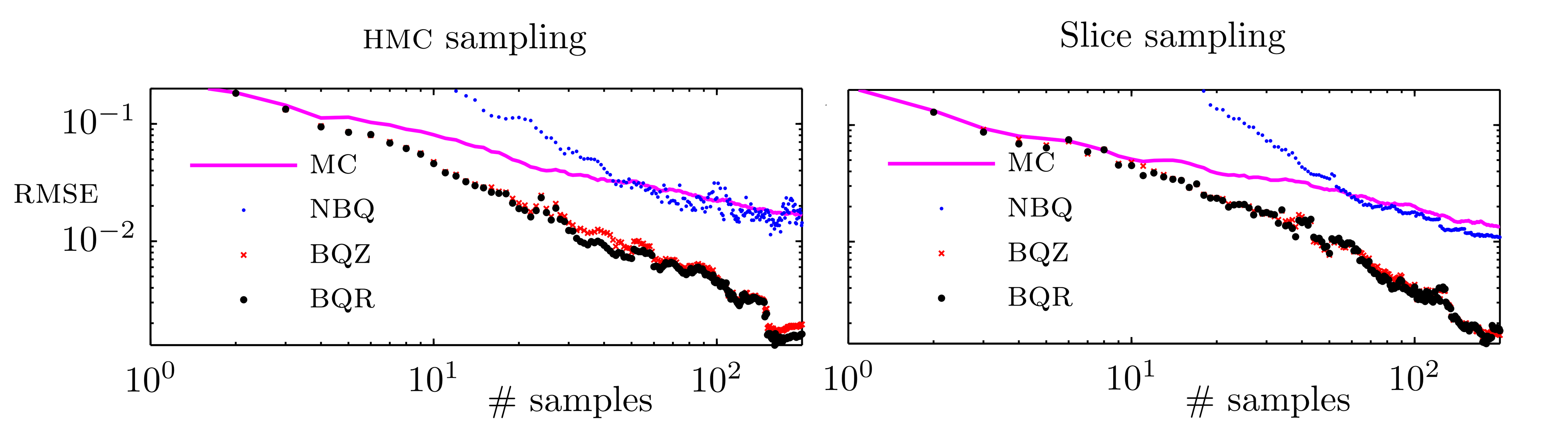

Figure 3: The RMSE between the true value of a ratio of analytic integrals (over mixtures of one-d Gaussians) and the estimates returned by different methods. All methods were given the same chains of samples, generated by both hybrid Monte Carlo (HMC) (left) and slice (right) samplers.

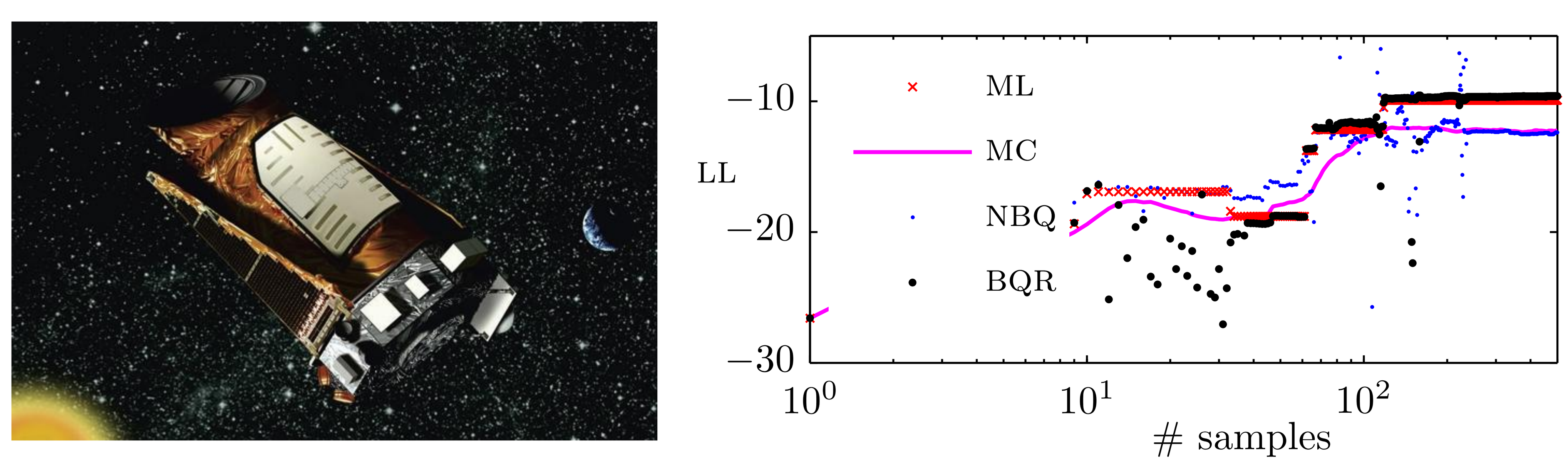

Figure 4: We attempted to regress flux from a star, using data from the Kepler mission. Above, the log-likelihood (LL) of held-out test data for predictions made by a GP, whose hyperparameters were marginalised using the methods shown. Samples were obtained using slice sampling.