A Probabilistic Tool for Assessing the Laplace Approximation: Review, Optimization, and GPU Implementation

Shaun McDonald, David Campbell, Haoxuan Zhou

SFU

Abstract

Many dynamic statistical models are characterized by a joint distribution with potentially high-dimensional nuisance parameters. To avoid the computational cost of numerically marginalizing over such parameters, we can instead use the Laplace approximation for the integral. This approximation is exact if and only if the distribution is Gaussian in the nuisance parameters; therefore, its effectiveness depends intimately on the true shape of the distribution. Here, we review a probabilistic numeric tool for assessing the quality of the Laplace approximation. This diagnostic tool recasts the approximation problem in a Bayesian framework, modelling the distribution with a Gaussian process prior and evaluating it at a coarse grid of points to obtain a posterior from which inferences can be made. We will discuss approaches for optimizing the tool, choosing evaluation points and other parameters to maximize our ability to detect non-Gaussianity. Our methods will be demonstrated on a variety of distributions, showing the effects of shape and dimension on the optimization. Finally, we will implement our methods using a GPU, highlighting possible performance advantages over the more conventional CPU.

1. Introduction

- Consider model $f(X,t,\theta)$ for data X, nuisance parameters t, and structural parameters θ
- Inference requires integrating out t
- For high-dimensional t, integral may be intractable and MCMC may be inefficient
- Using second-order Taylor approximation of $\log f$ about MLE $\hat{t}$ gives *Laplace approximation* (L.A.)

$$\int f(Y,t,\theta)dt \approx f(Y,\hat{t},\theta)\int \exp\left[-\frac{1}{2}(t-\hat{t})' H(Y,\hat{t},\theta)(t-\hat{t})\right]dt,$$

where H=Hessian of $\log f$ w.r.t. t

- L.A.is tractable and fast since R.H.S. integrand is Gaussian kernel
- Exact if f is Gaussian – if not, results may be invalid
- Can use *probabilistic integration* to create diagnostic tool to assess quality of L.A.

2. Probabilistic Solver

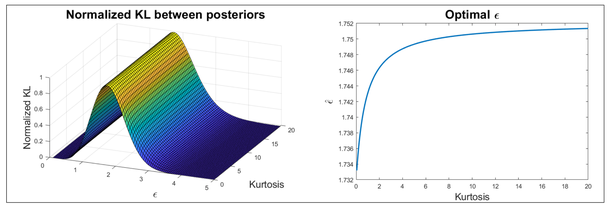
- Assume one-dimensional t here and ignore X and θ
- To assess L.A., use *probabilistic numerics* – recast numerical problem as Bayesian one, using probability theory to quantify uncertainty
- Jointly model *full function* $f(t)$ and *target* function $\int f(t)dt$ with Gaussian process prior
- Condition on evaluations of f at interrogation points $\boldsymbol{s} = [s_1, s_2, s_3]$ to get posterior
- $\boldsymbol{s} = [\hat{t} - \epsilon\delta, \hat{t}, \hat{t} + \epsilon\delta]$, where δ is the standard deviation of L.A. and ϵ is tbd
- Prior mean $m_t^0(t)$ for full function is Gaussian kernel (i.e. assume L.A. is true)
- Obtain prior mean $m^0(t)$ for target by integrating from s_1 to s_3
- Obtain covariances by convoluting and integrating squared exponential kernel with hyperparameters λ (length-scale) and α (precision)
- Following Chkrebtii et al. (2016) and Zhou (2017), set $\lambda = 1.5\epsilon\delta$ and $\alpha = 1/(\epsilon\delta)$
- Posterior for target function is $\mathcal{N}\left(m^1(s_3), C^1(s_3,s_3)\right)$
- If L.A. is true, mean doesn't update and posterior is $\mathcal{N}\left(m^0(s_3), C^1(s_3,s_3)\right)$
- For most powerful diagnostic, pick ϵ to maximize *KL divergence* between these:

$$\hat{\epsilon} = \text{argmax}_\epsilon \frac{\left(m^1(s_3) - m^0(s_3)\right)^2}{2C^1(s_3,s_3)} = \text{argmax}_\epsilon \frac{\text{bias}^2}{2 \cdot \text{variance}}$$

3. One-Dimensional Experiments

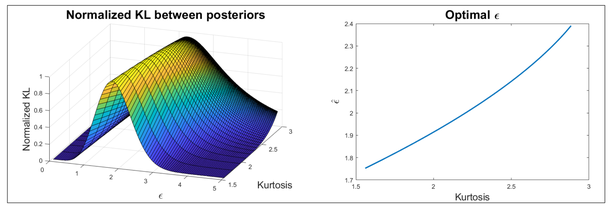
T-distribution

- Optimize for various ν values



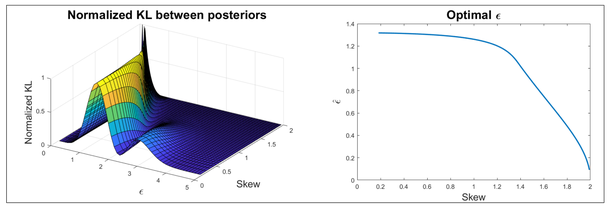
Gaussian mixture

- $0.5\mathcal{N}(0,1) + 0.5\mathcal{N}(0,\sigma^2)$; optimize for various $\sigma^2 > 1$



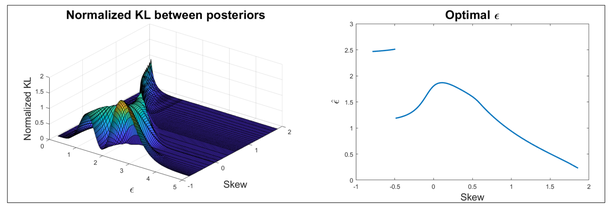
Gamma distribution

- Fix scale parameter, optimize for various shape values



Weibull distribution

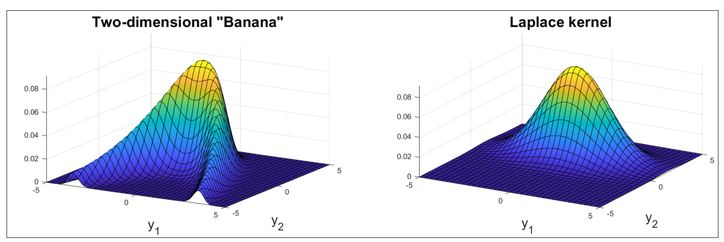
- Fix scale parameter, optimize for various shape values

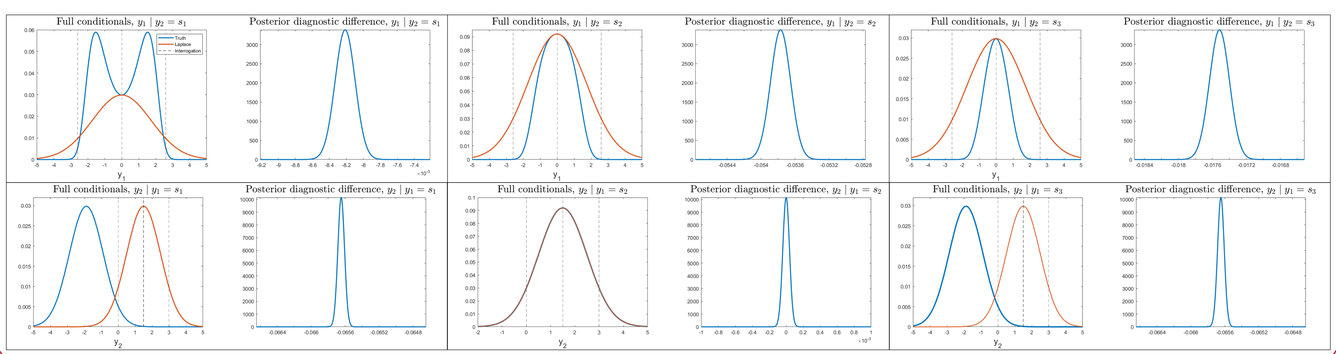


- For Gamma and Weibull with shape parameter ≤ 2, optimal ϵ iff $s_1 = 0$
- For highly left-skewed Weibulls, optimal ϵ around 2.5
- Otherwise, taking ϵ between 1.5 and 1.75 is reasonable

4. Two-Dimensional Banana-Shaped Distribution

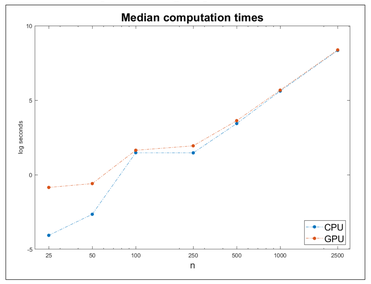
- 2-D Gaussian with one coordinate "twisted": $\left(y_1, y_2 - 0.5(y_1^2 - 3)\right) \sim \mathcal{N}\left(0, \text{diag}(3,1)\right)$ (Haario et al. 1999)
- 2-D L.A.
- Three 1-D diagnostics for each coordinate: conditional distributions given each point in $\boldsymbol{s}$ for other coordinate
- Use $\epsilon = 1.5$ rule of thumb
- Assess quality of L.A. and dependence all at once by looking at differences between posteriors
- Differentiating between shift and shape effects may be difficult





5. GPU Tests

- Add *n-2* independent components to banana
- Conditionals of each coordinate given $2n-1$ combinations of points in other directions for $2n^2 - n$ solvers in total
- Repeated runs on different types of processor – can GPU's speed things up?



- Overhead/computation speed tradeoff means CPU wins in our experiments
- However, extrapolation suggests GPU will win for larger dimensions

6. Discussion

- Probabilistic numerics provides powerful, easy, reasonably fast diagnostic for L.A.
- Can't be analytically optimized due to dependence on f, but we have shape-dependent rules of thumb
- Future work:
  - Different covariance structures with sparsity
  - More detailed high-dimensional joint modeling
  - Folding into larger optimization tasks, e.g. MLE's in state-space models

Acknowledgements

Chkrebtii, O. A., Campbell, D. A., Calderhead, B., & Girolami, M. A. (2016). Bayesian Solution Uncertainty Quantification for Differential Equations. *Bayesian Analysis, 11*(4), 1239-1267.

Haario, H., Saksman, E., & Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics, 14*(3), 375-395.

Hennig, P., Osborne, M. A., & Girolami, M. (2015). Probabilistic numerics and uncertainty in computations. *Royal Society A, 471.*

Zhou, H. (2017). Bayesian Integration for Assessing the Quality of the Laplace Approximation. Master's Thesis, Simon Fraser University.

I would also like to thank Lucas Wu for providing the template for this poster.