OUTLINE FOR Independent Metropolis-Hastings Steps for Generalized Linear Models with Latent Gaussian Processes via Global Conditional Laplace Approximations

Matthew Simpson[1]

Abstract

KEY WORDS:

[1](to whom correspondence should be addressed) Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, themattsimpson@gmail.com

1 Outline

(Yes I know the title is long)

1.1 Basic Problem

Want to sample from a difficult posterior density. Suppose it has model parameter θ and latent process gaussian process y, and we observe data z. The target posterior distribution is $[\theta, y|z] \propto [z|y, \theta][y|\theta][\theta]$ where $[y|\theta]$ is a normal density. A common MCMC strategy for this class of models is a data augmentation Gibbs sampler, i.e. draw $\theta \sim [\theta|y, z]$ and $y \sim [y|\theta, z]$ iteratively. Often $[\theta|y, z]$ is a relatively easy to sample from, but $[y|\theta, z]$ is not a density of known form and so requires a Metropolis step.

1.2 Laplace approximations

There are three ways we can employ Laplace approximations here:

1. Global Laplace approximation as a joint Metropolis proposal for $[y, \theta|z]$. But θ is often non-normal in the posterior.

2. Local conditional Laplace approximation as a proposal for $[y|\theta, z]$, i.e. compute the Laplace approximation to $[y|\theta, z]$ every iteration of MCMC. But then we have to do numerical optimization to find y's conditional mode every iteration of the MCMC, which can be expensive.

3. Global conditional Laplace approximation as a proposal for $[y|\theta, z]$, i.e. compute the global Laplace approximation once, then compute the implied conditional distribution for $[y|\theta, z]$ every iteration. Often much cheaper because much of the computation can be pre-computed. (This is the main contribution)

1.3 Basic structure of the paper

1. Introduction
2. Describe the problem for latent Gaussian process models.
3. Describe Laplace approximations and introduce the "Global conditional Laplace approximation" (GCLA)
4. Give some intuition for when the GCLA will be good vs. a GLA or a LCLA (global LA and local conditional LA). Maybe a theorem that explains why/when GCLA works about as well as LCLA. Maybe another theorem that explains when no CLA should work well. (Note: theorems may not be worth the time because it's not obvious to me how to go about proving them right now)
5. 2-4 examples illustrating both good and bad, with some simulations. (It doesn't always work. In particular, when the data model is highly non-normal, it can be extremely poor).

1.4 PSO tie in?

It's possible that the best of our PSO algorithms is actually good at finding the posterior mode in some cases. Worth checking, and if so, we have another minor contribution (and a chance to cite the STAT paper we're pushing out).

2 The Examples

2.1 County population model

This illustrates when the GLA is bad, but the GCLA is good rather nicely.

2.2 Election model

This expands the class of models somewhat; not sure if it illustrates anything very nicely. I'll have to think about this one a bit.

2.3 Unemployment Rates model

(County or tract or even state level). This was a model we jettisoned awhile ago, but basically it's a GLMM with a LGP where the data model is a Beta distribution. The GCLA doesn't work here because the data model is too non-normal - basically the uncertainties associated with the unemployment rates are too high and the rates are too close to zero, so the Beta distributions are U-shaped. But if we but a sufficiently tight prior on sufficiently high levels of certainty, I think we can get the GCLA to work. So it illustrates when ANY conditional LA should work well and when it shouldn't.