Analytic Approximations in Bayesian Inference

Daniel García (dg416@ic.ac.uk)
Supervisor: Alastair Young

Imperial College London

Problem Overview

• Bayesian inference on a scalar parameter of interest is based on its posterior distribution

$$\pi(\psi|y) \propto \int \pi(\psi,\chi)\exp\{l(\psi,\chi)\}d\chi.$$

• This distribution is often intractable and non-trivial to approximate (curse of dimensionality).

• MCMC methods are often used to bypass this problem, but don't solve it completely.

• Accurate analytic approximations of $\pi(\psi|y)$ derived from Laplace's method provide an alternative route to posterior inference that has certain advantages over the former.

Objectives

• Detailed examination of analytic approaches to Bayesian inference about a scalar parameter of interest.

• Analysis of their accuracy in **high-dimensional** settings and, if appropriate, provide refinements for these situations.

Main Results

• The posterior distribution function can be approximated with *relative* error of order $O(n^{-3/2})$ by

$$\mathbb{P}(\Psi \le \psi|y) \doteq \begin{cases} \Phi(r(\psi)) + \phi(r(\psi))\left(r(\psi)^{-1} - q(\psi)^{-1}\right) & (1) \\ \Phi(r^*(\psi)) & (2) \end{cases}$$

• The quantities $r(\psi)$, $r^*(\psi)$ and $q(\psi)$ involve only derivatives of the log-likelihood up to order 2, which may need to be computed numerically.

• They have been shown to be remarkably accurate if the parameter dimension is small, but their performance in high-dimensional contexts is unknown.

Applications

Why are analytic approximations desirable? Two possible applications:

• **Simulation.** Approximation (2) provides the following simple simulation scheme:

1. Generate $z \sim N(0,1)$;
2. Solve $r^*(\psi) = z$ (often numerically).

$\rightarrow$ It gives independent samples of the posterior and does not require tuning of parameters.

$\rightarrow$ It does not provide perfect sampling asymptotically, as all samples are taken from the same approximated distribution.

• **Sensitivity analysis on the prior.** Approximations (1) and (2) depend on the prior only through $\pi(\hat\psi,\hat\chi)/\pi(\psi,\hat\chi_\psi)$, so they can be used to assess the dependence of the posterior distribution on π.

An Example

• We have 3 observations from p normal populations with common variance and different means.

• The common variance σ^2 is of interest and the p means $\mu = (\mu_1,\dots,\mu_p)$ are nuisance parameters.

• The prior is $\pi(\sigma^2,\mu) \propto \sigma^{-2}$.

• The error of approximation (2) increases considerably with p:

p	0.01	0.05	0.1
5	0.012	0.059	0.116
50	0.017	0.075	0.141
500	0.046	0.159	0.264
1000	0.079	0.234	0.359