Bayesian statistics: efficiency of marginalising over discrete latent parameters

Kate Zhang
kate.zhang@student.unimelb.edu.au
Supervisors: Damjan Vukcevic,
with assistance from Jeffrey Pullin and Lyle Gurrin

Introduction

Discrete Latent Parameters

Which is more efficient?

- Sample discrete parameters using techniques such as Gibbs Sampling
- Marginalise out the discrete parameter

For Bayesian model involving discrete variables, it is possible to sample a value of the discrete parameter at each iteration using methods such as Gibbs Sampling. With algorithms that do not support sampling discrete parameters, an alternative is to marginalise out the discrete parameters from the likelihood. In this project, we are interested in analysing which approach is more computationally efficient.

Methods

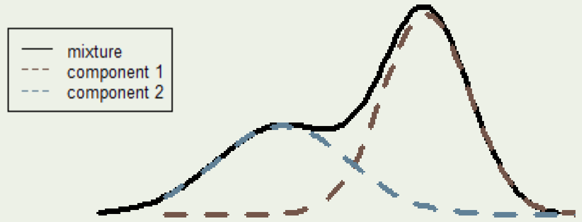
1. Write the discrete and marginalised Bayesian models in JAGS
2. Simulate data sets
3. Estimate the parameters with JAGS
4. Record computation time and analyse result

Two-component normal mixture model

We are using a simple mixture model of two normal distributions. Suppose we have two independent normal distributions $Y_1 \sim N(\mu_1, \sigma_1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$. Then we have the mixture model Y:

$$Y = Z \cdot Y_1 + (1 - Z) \cdot Y_2$$

Where z is the discrete latent variable where $z = 0$ or 1, with $Z \sim Ber(p)$, p is the mixing component.



JAGS models and marginalisation

We have the unmarginalised likelihood:

$$\Pr(y|\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p_1, p_2, z) = \prod_{i=1}^{I} (p_{z_i} \varphi_{z_i}(y))$$

Marginalised likelihood:

$$\Pr(y|\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p_1, p_2) = \prod_{i=1}^{I} (p_1 \varphi_1(y) + p_2 \varphi_2(y))$$

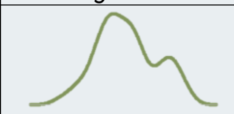
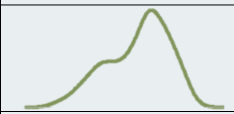
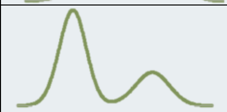
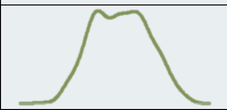
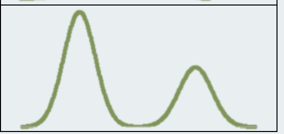
With these likelihoods we wrote six JAGS models:

Model 1(m1)	Model 2(m2)	Model 3(m3)
- Not marginalised	- Marginalised with function dnormmix() in JAGS - Requires the mix module	- Marginalised manually with the "zeros trick"

Model 1(m1r)	Model 2(m2r)	Model 3(m3r)
Model 1 but with a reduced set of available samplers	Model 2 but with a reduced set of available samplers	Model 3 but with a reduced set of available samplers

*Models with reduced sets are only allowed to use Slice and Dirichlet sampling

Data sets

	Sample size	Density
data1	350	
data2	800	
data3	350	
data4	950	
data5	350	

Inference for *z* via conditioning

For the marginalised models, we are able to infer the discrete latent variable through conditional probability:

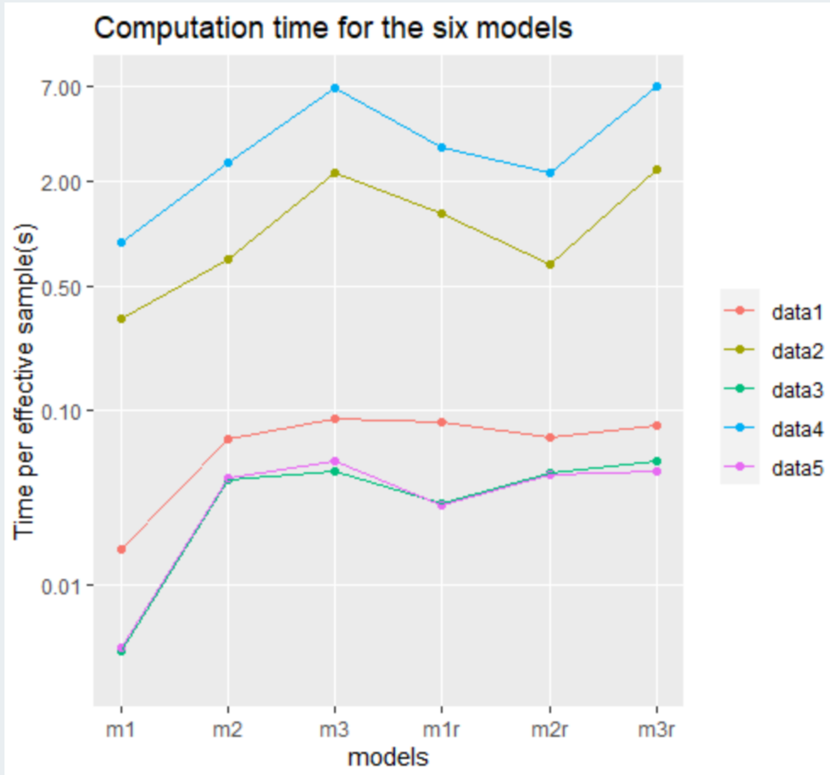
$$\Pr(z = k|\mu_1, \mu_2, \sigma_1, \sigma_2, p_1, p_2, y) = \frac{p_k \varphi_k(y)}{p_1 \varphi_1(y) + p_2 \varphi_2(y)}$$

Where $k = \{1,2\}$(either component 1 or component 2), p_1, p_2 are mixing components and $p_1 + p_2 = 1$

Through the fact that $E(Z = 2) = 1 - E(Z = 1)$, we can derive that:

$$\mathrm{E}(Z) = -E(Z = 1) + 2.$$

Result and analysis



Six models led to similar estimates of the parameter means and quantiles for each data set but varied in the computation time required to produce one effective sample of the least converged continuous parameter.

m1 took significantly shorter time than the other models for all data sets, as it's benefiting from a more efficient sampling algorithm.

Looking at the models with the reduced set, the marginalised model(m2r) appears to be slightly more efficient than the discrete model(m1r) in most cases. However, for data sets where the two components are clearly identifiable, marginalisation tends to lose its advantage (data4 and data5). This inconsistency across the data sets is mainly due to the way Slice sampling interacts with the density of our data sets.

With the marginalised models, model 3 is always less efficient than model 2, attributing to the fact that the "zeros trick" tends to force JAGS into using an inefficient sampling strategy.