

BAYESIAN ADAPTIVE INDEPENDENCE SAMPLING WITH LATENT VARIABLES

Christian M. Davey¹, Jonathan M. Keith¹, Christian P. Robert^{2,3}

¹School of Mathematical Sciences, Monash University, Australia

²Research Centre in Decision Mathematics, Université Paris-Dauphine, France

³Department of Statistics, University of Warwick, United Kingdom

The Bayesian Adaptive Independence Sampler (BAIS) [5] is a population MCMC algorithm consisting of N independence samplers run in parallel, all with a common proposal density. The proposal is a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ that adapts at each iteration to progressively improve sampler performance. Adaptation is achieved by drawing parameters μ and Σ from a posterior distribution $p(\mu, \Sigma | x_1, \dots, x_N)$ using a Bayesian model applied to the current N independence samples (x_1, \dots, x_N) . The distribution to be sampled is thus a product of $p(\mu, \Sigma | x_1, \dots, x_N)$ and N independent replications of the target $f(x)$. Hence, the adaptation can be continued indefinitely.

Independence samplers work best when the proposal approximates the target distribution. Consequently, BAIS is not an efficient sampling strategy for multimodal distributions that are poorly approximated by a multivariate normal. Here we generalise BAIS to allow the proposal to be a finite mixture of K multivariate normal distributions. This is done by augmenting each independence sampler with a latent variable, a strategy we call Bayesian Adaptive Independence Sampling with Latent variables (BAIS+L).

In this work we describe the significant practical and computational problems that arise in implementing BAIS+L, and discuss exact and approximate strategies for solving these problems.

An Approximate Sampler

The Bayesian Adaptive Independence Sampler with Latent variables (BAIS+L), outlined in Algorithm 1 updates K means μ_k and covariances matrices Σ_k following [2, pp. 86–87]. As in [2] μ_0 is the prior mean, ν_0 , and Λ_0 , are the prior degrees of freedom and prior scale on the distribution of each Σ_k , respectively, and κ_0 is the prior number of observations of the scale of each Σ_k . An exact acceptance ratio α for this approach involves an intractable quantity. Therefore, BAIS+L uses an approximate ratio, which leads to the stationary distribution being an approximation of the intended target.

Algorithm 1 BAIS+L

```

1: for Iteration  $t = 1$  to  $t = T$  do
2:   for Mixture component  $k = 1$  to  $k = K$  do
3:     Compute  $o_k$  then  $\bar{x}_k$  then  $\mathbf{S}_k$ .
4:     Generate  $\Sigma_k$  then  $\mu_k$ .
5:   end for
6:   Generate  $(\pi_1, \dots, \pi_K)$ .
7:   for Sampling chain  $n = 1$  to  $n = N$  do
8:     Propose  $w \sim \text{Categorical}(\pi_1, \dots, \pi_K)$ 
9:     propose  $y \sim \mathcal{N}(\mu_w, \Sigma_w)$ .
10:    Compute  $\alpha$  and generate  $u \sim \mathcal{U}(0, 1)$ .
11:    if  $u \leq \alpha$  then
12:      Set  $[x_n^{(t)}, z_n^{(t)}] = (y, w)$ .
13:    else
14:      Set  $[x_n^{(t)}, z_n^{(t)}] = [x_n^{(t-1)}, z_n^{(t-1)}]$ .
15:    end if
16:  end for
17: end for

```

$$o_k = \sum_{n=1}^N I_k [z_n^{(t-1)}] \quad \bar{x}_k = \frac{1}{o_k} \sum_{n=1}^N x_n^{(t-1)} I_k [z_n^{(t-1)}]$$

$$\mathbf{S}_k = \sum_{n=1}^N I_k [z_n^{(t-1)}] [x_n^{(t-1)} - \bar{x}_k] [x_n^{(t-1)} - \bar{x}_k]^T$$

$$\alpha = \min \left\{ 1, \frac{f(y)}{f[x_n^{(t-1)}]} \cdot \frac{\sum_{k=1}^K \pi_k \mathcal{N}[x_n^{(t-1)} | \mu_k, \mathbf{S}_k]}{\sum_{k=1}^K \pi_k \mathcal{N}(y | \mu_k, \mathbf{S}_k)} \right\}$$

$$\pi = (\pi_1, \dots, \pi_K) \sim \text{Dirichlet}(o_1 + 1, \dots, o_K + 1)$$

$$\mu_k \sim \mathcal{N} \left(\frac{\kappa_0}{\kappa_0 + o_k} \mu_0 + \frac{o_k}{\kappa_0 + o_k} \bar{x}_k, \frac{\Sigma_k}{\kappa_0 + o_k} \right)$$

$$\Sigma_k \sim \text{Inv-W}_{\nu_0 + o_k} \left[\Lambda_0 + \mathbf{S}_k + \frac{\kappa_0 o_k (\bar{x}_k - \mu_0) (\bar{x}_k - \mu_0)^T}{\kappa_0 + o_k} \right],$$

where $I_k(a) = 1$ if $a = k$ and 0 otherwise.

Lemma

A population of samples, generated with a Metropolis-Hastings [6, 3] sampler that updates each element of the population using a transition kernel, with parameters dependent only on the remaining elements and a standard Metropolis-Hastings acceptance ratio, will have the desired target as its stationary distribution [1].

An Exact Sampler

The Bayesian Adaptive Mixture Independence Sampler (BAMIS), outlined in Algorithm 2, is justified the preceding Lemma, thus avoiding the need to introduce approximations, as in BAIS+L.

Algorithm 2 BAMIS

```

1: for Iteration  $t = 1$  to  $t = T$  do
2:   for Sampling chain  $n = 1$  to  $n = N$  do
3:     for Mixture component  $k = 1$  to  $k = K$  do
4:       Compute  $o_k$  then  $\bar{x}_k$  then  $\mathbf{S}_k$ .
5:       Generate  $\Sigma_k$  and  $\mu_k$  as for BAIS+L.
6:     end for
7:     Generate  $(\pi_1, \dots, \pi_K)$  as for BAIS+L.
8:     Propose  $w \sim \text{Categorical}(\pi_1, \dots, \pi_K)$ .
9:     Propose  $y \sim \mathcal{N}(\mu_w, \Sigma_w)$ .
10:    Compute  $\alpha$  as for BAIS+L.
11:    Generate  $u \sim \mathcal{U}(0, 1)$ .
12:    if  $u \leq \alpha$  then
13:      Set  $[x_n^{(t)}, z_n^{(t)}] = (y, w)$ .
14:    else
15:      Set  $[x_n^{(t)}, z_n^{(t)}] = [x_n^{(t-1)}, z_n^{(t-1)}]$ .
16:    end if
17:  end for
18: end for

```

$$o_k = \sum_{i=1}^{n-1} I_k [z_i^{(t)}] + \sum_{i=n+1}^N I_k [z_i^{(t-1)}]$$

$$\bar{x}_k = \frac{1}{o_k} \left\{ \sum_{i=1}^{n-1} x_i^{(t)} I_k [z_i^{(t)}] + \sum_{i=n+1}^N x_i^{(t-1)} I_k [z_i^{(t-1)}] \right\}$$

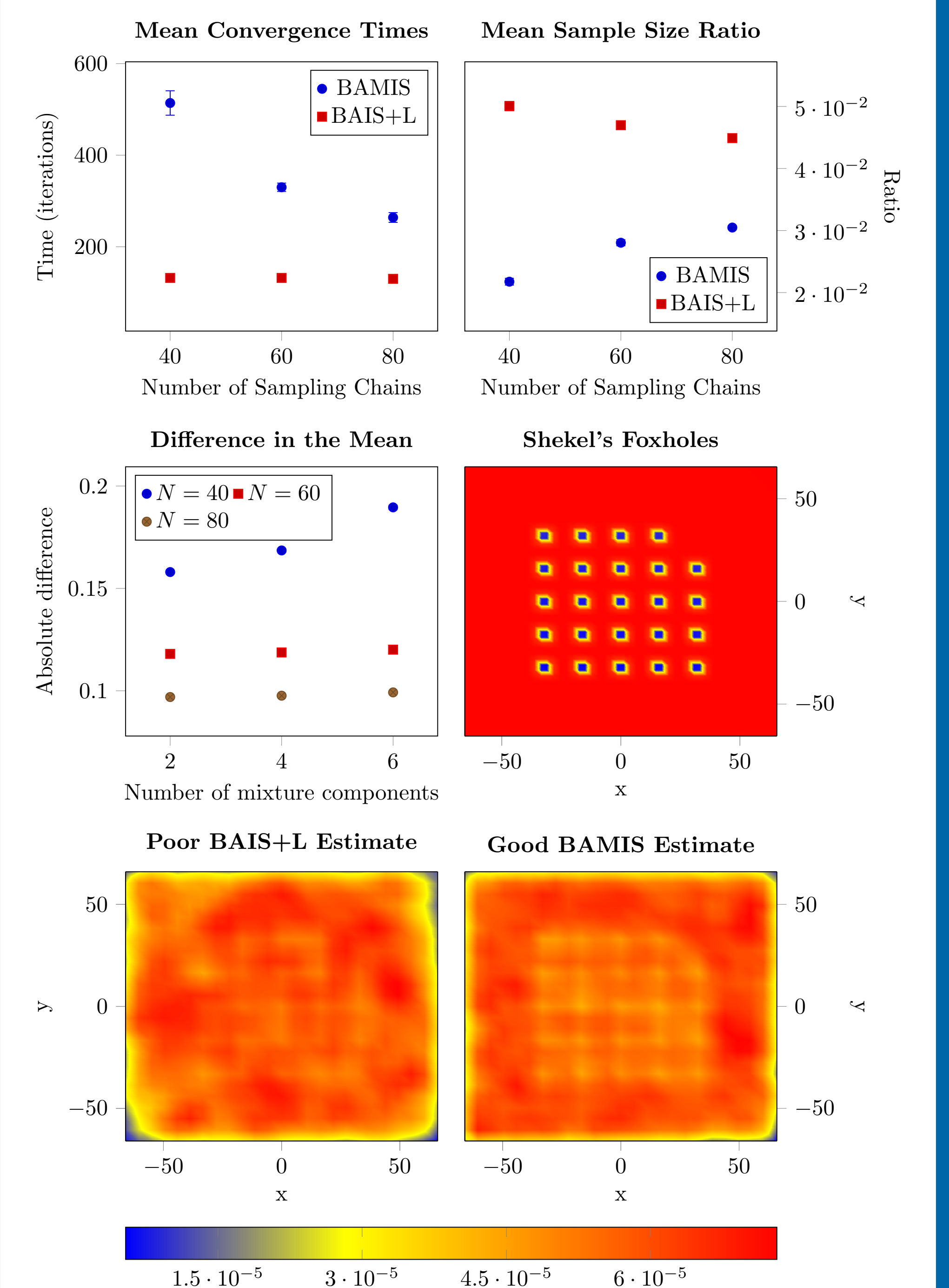
$$\mathbf{S}_k = \sum_{i=1}^{n-1} I_k [z_i^{(t)}] [x_i^{(t)} - \bar{x}_k] [x_i^{(t)} - \bar{x}_k]^T + \sum_{i=n+1}^N I_k [z_i^{(t-1)}] [x_i^{(t-1)} - \bar{x}_k] [x_i^{(t-1)} - \bar{x}_k]^T$$

Comparing Approaches

Scatter plots of the mean convergence time (top-left) and the mean ratio of the effective number of samples to the total number of samples (top-right) of a 5D quartic function [4, 7] on $\mathbf{x} \in [-1.28, 1.28]^5$,

$$f(\{x_1, \dots, x_5\}) = \sum_{d=1}^5 (ix_d^2 + \eta), \quad \eta \sim \mathcal{U}(0, 1),$$

with $K = 6$, suggest that BAIS+L produces faster convergence and mixing than BAMIS but that this improvement diminishes with N .



The difference between BAIS+L and BAMIS estimates of the first four central moments in each dimension (\bar{X} , middle-left) decreased with N , especially for large K . This suggests that the stationary distributions of the two samplers approach each other as N increases.

The bottom two contour plots show kernel density estimates of Shekel's foxholes [4, 7] without the top-right hole on $\mathbf{x} \in [-65.536, 65.536]^2$ (true contour, middle-right). We generated these examples with the default `kde2d` function from R's [8] `MASS` library [9] from a poor BAIS+L run with $K = 6, N = 40$ and a good BAMIS run with $K = 4, N = 80$. The reduction in density around the edges of the sampled domain is due to the use of a normal kernel in the density estimation.

A cursory glance over multiple such simulations suggested that BAIS+L and BAMIS performed comparably, with a larger N generally producing a clearer representation of the target for any K with each sampler (not shown).

References

- [1] J. Besag et al. "Bayesian computation and stochastic systems". In: *Stat. Sci.* 10.1 (1995), pp. 3–66.
- [2] A. Gelman et al. *Bayesian data analysis*. Second Ed. Chapman & Hall, 2004.
- [3] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (1970), pp. 97–109.
- [4] L. Ingber and B. Rosen. "Genetic algorithms and very fast simulated reannealing: a comparison". In: *Math. Comput. Model.* 16.11 (1992), pp. 87–100.
- [5] J. M. Keith, D. P. Kroese, and G. Y. Sofronov. "Adaptive independence samplers". In: *Stat. Comput.* 18.4 (2008), pp. 409–420.
- [6] N. Metropolis et al. "Equation of state calculations by fast computing machines". In: *J. Chem. Phys.* 21.6 (1953), pp. 1087–1092.
- [7] R. H. Storn and K. Price. "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces". In: *J. Global Optim.* 11 (1997), pp. 341–359.
- [8] R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [9] W. N. Venables and B. D. Ripley. *Modern applied statistics with S*. Fourth Ed. ISBN 0-387-95457-0. New York: Springer, 2002.