# Generalised Ensemble Kalman Inversion

Owen Jackson Imke Botha Jennifer Flegg

University of Melbourne

## Background

Markov Chain Monte Carlo (MCMC) is a well-known Exact Bayesian Inference technique for sampling from a posterior distribution. However, in practice it requires the likelihood distribution to be 'easy' to sample from, which is not always the case, particularly in high dimensional settings.

Ensemble Kalman Techniques refer to a series of Monte Carlo methods designed for these high dimensional settings, imposing a restrictive assumption of additive Gaussian likelihoods, with a known covariance matrix. Another likelihood free technique is Approximate Bayesian Computing (ABC), which works for general likelihoods, but has so far performed poorly in high dimensional settings.

## Generalised Ensemble Kalman Inversion

Duffield and Singh (2022) proposed an algorithm called Generalised Ensemble Kalman Inversion (GEKI) which removes the Gaussian assumption behind previous Ensemble Kalman techniques.

For example, in the Ensemble Kalman Inversion (EKI) algorithm, we assume that $p(y|x^{(i)}) \sim N(\mathcal{H}(x^{(i)}), R)$ where $\mathcal{H}$ is a deterministic operator (not necessarily linear) and $R$ is a known covariance matrix. By contrast, the GEKI algorithm only requires that we are able to simulate synthetic data $y_{l-1}^{(i)} \sim p(\cdot|x_{l-1}^{i})$.

## Purpose

GEKI offers a promising alternative to EKI and ABC. The purpose of this project is to investigate the performance of the algorithm in two settings

- A multivariate normal model with unknown mean and noise
- A malaria transmission model using real world data of malaria cases in Afghanistan

Of particular interest is the ability of GEKI to estimate the noise parameter in these models.

## GEKI Algorithm

Inputs

- $y$, a single draw from the likelihood using the true (unknown) parameters
- $\{\lambda_l\}_{l=0}^{L}$, a sequence of inverse temperatures

Instead of providing a sequence of inverse temperatures, they can be selected adaptively using pseudo particle weights, as described in [2].

The algorithm proceeds as follows

1. Initialize particles: $x_0^{(i)} \sim p(x)$
2. Generate synthetic data: $y_{l-1}^{(i)} \sim p(\cdot|x_{l-1}^{(i)})$
3. While $\lambda_l < 1$ do:
4. Calculate the change in inverse temperature: $h_l = \lambda_l - \lambda_{l-1}$
5. Calculate covariance matrices: $C_{l-1}^{xx}, C_{l-1}^{yy}, C_{l-1}^{xy}, C_{l-1}^{yx}, C_{l-1}^{y|x}$ as described in [2]. $C_{l-1}^{y|x}$ is the counterpart of the known covariance matrix $R$ in the EKI algorithm.
6. Generate perturbations: $\eta_l^{(i)} \sim N(0, C_{l-1}^{y|x})$
7. Update particles: $x_l^{(i)} = x_{l-1}^{(i)} + H(y - y_{l-1}^{(i)} - \eta_l^{(i)})$ where $H = C_{l-1}^{xy}(C_{l-1}^{yy} + (h_l^{-1} - 1)C_{l-1}^{y|x})^{-1}$ acts as the Kalman gain matrix.
8. Return ensemble

The algorithm is constructed to be asymptotically unbiased in the linear Gaussian case [2].

## Multivariate Normal Model

We assume that we have access to a single observation from a multivariate normal distribution $y \sim N(\alpha x, \sigma^2 I)$ where $\alpha$ is a scalar and $x$ is a known vector. The unknown parameters of the model are $\theta = (\alpha, \sigma^2)$. Since the likelihood is tractable we can use MCMC as a basis for comparison. I draw from the samples from the priors $\alpha \sim N(0, 5^2)$ and $\log(\sigma^2) \sim N(2, 1^2)$.

## Results

GEKI and MCMC both do well at estimating the mean parameter. The plots below show the EKI and MCMC marginal posteriors of $\alpha$ when $\alpha = 2$ and $\sigma = 2$.

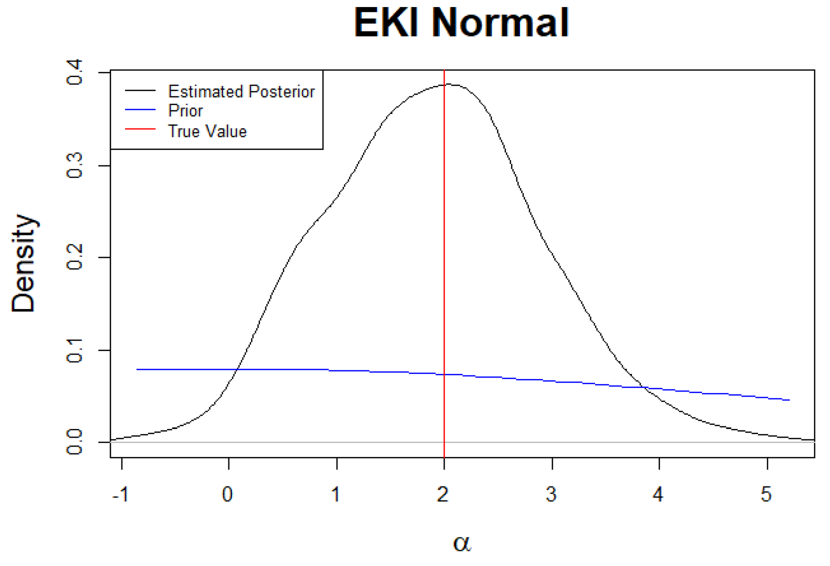


Figure 1. Marginal Posterior of $\alpha$ for GEKI

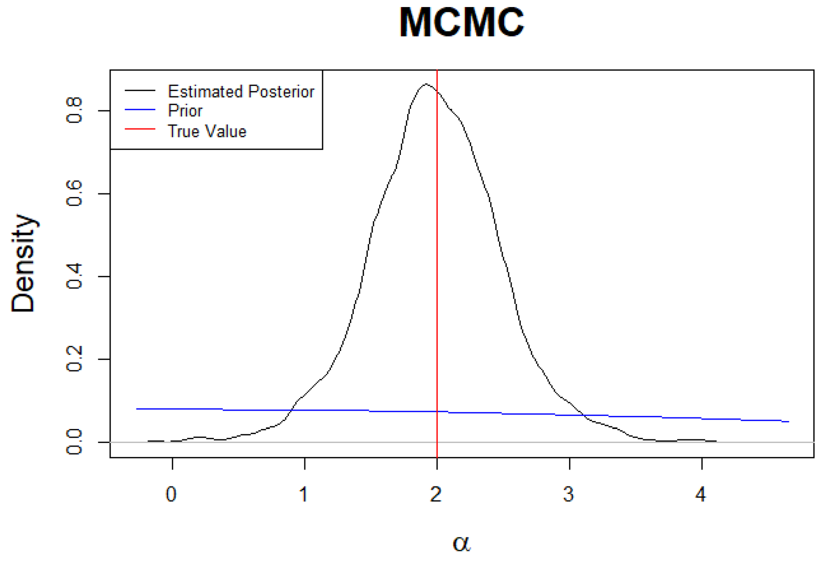


Figure 2. Marginal Posterior of $\alpha$ for MCMC

However, GEKI does poorly at estimating the noise parameter, remaining anchored to its prior distribution. By contrast, MCMC is able to estimate both parameters. The plots below show the EKI and MCMC marginal posteriors of $\log(\sigma^2)$ when $\alpha = 2$ and $\sigma = 8$.

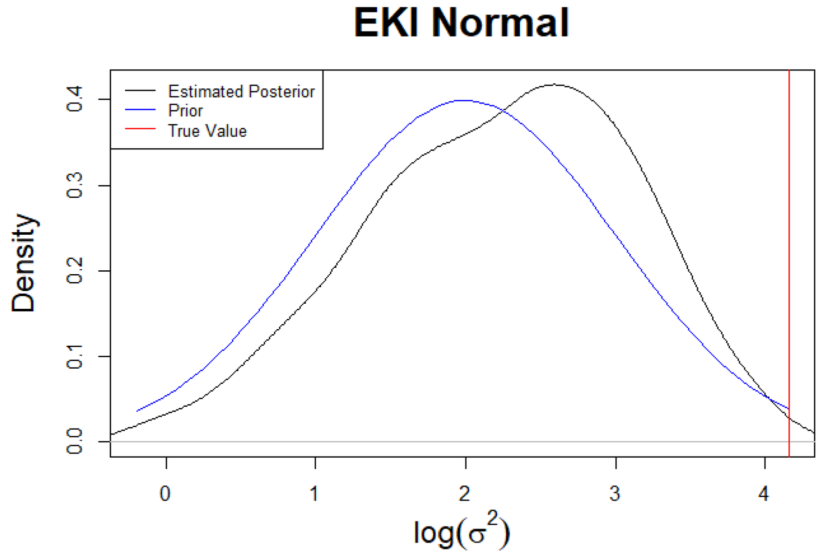


Figure 3. Marginal Posterior of $\log(\sigma^2)$ for GEKI

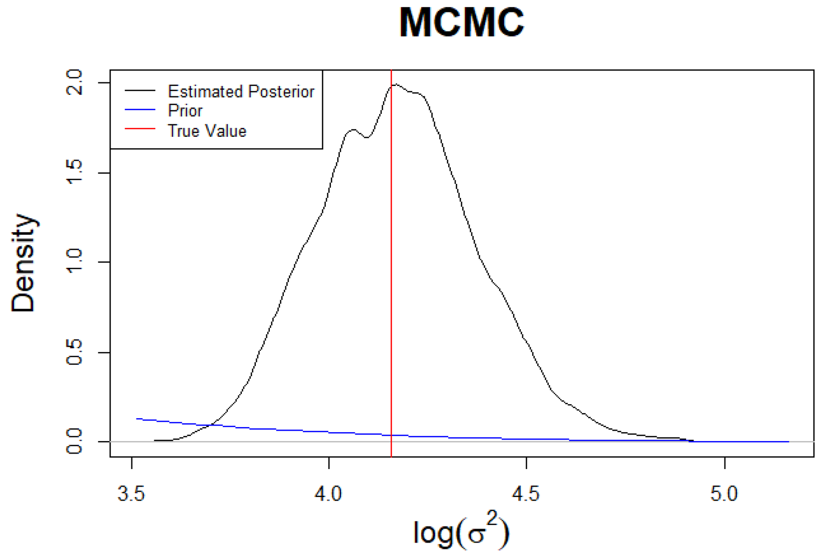


Figure 4. Marginal Posterior of $\log(\sigma^2)$ for MCMC

One explanation for this phenomenon is that the GEKI algorithm relies on a cross covariance between the parameters and the data, but $\sigma^2$ and $y$ are theoretically uncorrelated. Since $\sigma^2$ provides no signal about the location of the data, the particles are unable to change from their initial distribution.

While beyond the scope of this project, one would expect this phenomenon to be observed in similar parameters that affect variance and kurtosis only (e.g. the degrees of freedom in a t-distribution). Improvements to the algorithm which may be able to address this issue are also beyond the scope of this project.

## Malaria Transmission Model

Malaria is still one of the world's deadliest diseases and accurately predicting the effects of proposed interventions is of crucial importance.

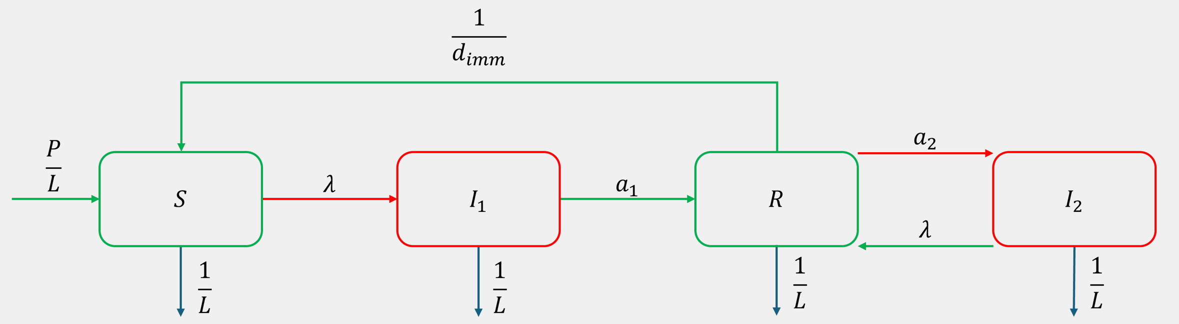


Figure 5. State transitions in the Malaria Model

The Malaria transmission model proposed in **needs citation** is governed by a system of differential equations involving a series of known parameters and four unknown parameters of interest: $d_{in}, \phi, \eta_0$ and $\sigma^2$ [1]. However, since the normal model demonstrated poor results for $\sigma^2$ we take $\theta = (d_{in}, \phi, \eta_0)$. For priors we have $d_{in} \sim \text{Half-Normal}(2^2)$ and $\phi, \eta_0 \sim \text{Logit-Normal}(0, 1)$.

## Results

Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Phasellus libero enim, gravida sed erat sit amet, scelerisque congue diam. Fusce dapibus dui ut augue pulvinar iaculis.

| First column | Second column | Third column | Fourth |
|---|---|---|---|
| Foo | 13.37 | 384,394 | $\alpha$ |
| Bar | 2.17 | 1,392 | $\beta$ |
| Baz | 3.14 | 83,742 | $\delta$ |
| Qux | 7.59 | 974 | $\gamma$ |

Table 1. A table caption.

Donec quis posuere ligula. Nunc feugiat elit a mi malesuada consequat. Sed imperdiet augue ac nibh aliquet tristique. Aenean eu tortor vulputate, eleifend lorem in, dictum urna. Proin auctor ante in augue tincidunt tempor. Proin pellentesque vulputate odio, ac gravida nulla posuere efficitur. Aenean at velit vel dolor blandit molestie. Mauris laoreet commodo quam, non luctus nibh ullamcorper in. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos.

Nulla varius finibus volutpat. Mauris molestie lorem tincidunt, iaculis libero at, gravida ante. Phasellus at felis eu neque suscipit suscipit. Integer ullamcorper, dui nec pretium ornare, urna dolor consequat libero, in feugiat elit lorem euismod lacus. Pellentesque sit amet dolor mollis, auctor urna non, tempus sem.

## References

[1] Amani A. Alahmadi, Jennifer A. Flegg, Davis G. Cochrane, Christopher C. Drovandi, and Jonathan M. Keith. A comparison of approximate versus exact techniques for bayesian parameter inference in nonlinear ordinary differential equation models. *Royal Society Open Science*, 7, 2022.

[2] Samuel Duffield and Sumeetpal S. Singh. Ensemble kalman inversion for general likelihoods. *Statistics and Probability Letters*, 187, 2022.

## Acknowledgments