# Robust Approximate Bayesian Inference with Synthetic Likelihood

Associate Professor Chris Drovandi

School of Mathematical Sciences
ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)
Queensland University of Technology

Joint work with Dr David Frazier

8th July 2019

# Outline

1. Background
2. Bayesian Synthetic Likelihood (BSL)
3. Robust BSL
4. Other BSL work and summary

# Bayesian Statistics

In Bayesian statistics we are interested in sampling from the posterior:

$$p(\theta|y) \propto p(y|\theta)p(\theta),$$

where $p(y|\theta)$ is the likelihood and $p(\theta)$ is the prior.

# Markov Chain Monte Carlo

Construct ergodic Markov chain with invariant distribution $p(\theta|y)$

A common MCMC algorithm is Metropolis Hastings (MH) MCMC, where proposals $\theta^*$ are accepted with probability

$$\min\left(1, \frac{p(y|\theta^*)p(\theta^*)q(\theta|\theta^*)}{p(y|\theta)p(\theta)q(\theta^*|\theta)}\right),$$

where $q(\cdot)$ is the proposal density.

For complex models, $p(y|\theta)$ may not be computationally feasible.

# Approximate Bayesian Computation

Approximate Bayesian computation (ABC[1]) is current state-of-the-art likelihood-free Bayesian method.

ABC prefers parameter values that simulates $x$ close to $y$, in terms of summary statistics $S(\cdot)$.

Targets posterior conditional on observed summary $p(\theta|s_y) \propto p(s_y|\theta)p(\theta)$ where $s_y = S(y)$.

Estimates $p(s_y|\theta)$ non-parametrically via simulation (Blum 2010).

Choice of summary function $S(\cdot)$ trade-off between information loss and dimensionality.

[1] Handbook of Approximate Bayesian Computation (2018). Eds: Sisson, Fan, Beaumont. CRC Press.

# Approximate Bayesian Computation

ABC Approximation of likelihood $p(s_y|\theta)$

- Simulate $n$ iid datasets, denoted $x_{1:n} = (x_1, ..., x_n)$, from the model based on $\theta$.
- Calculate $n$ sets of summary statistics, $s_{1:n} = (s_1, ..., s_n)$
- The intractable $p(s_y|\theta)$ is replaced with the estimated ABC likelihood,

$$\hat{p}_\epsilon(s_y|\theta) = \frac{1}{n}\sum_{i=1}^{n} K_\epsilon\left(\rho(s_y, s_i)\right).$$

  - $\rho(\cdot)$ is called the discrepancy function
  - $K_\epsilon(\cdot)$ is a kernel weighting function with bandwidth $\epsilon$
  - $\epsilon$ is called the ABC tolerance (bias/variance trade-off)

# Approximate Bayesian Computation

Disadvantages

- Highly sensitive to choice of tuning parameter $\epsilon$, $\rho(\cdot)$ and to a lesser extent $K_\epsilon(\cdot)$
- No standard way to select $\epsilon$ or $\rho(\cdot)$.
- Suffers from curse of dimensionality with respect to size of summary statistic

# Motivating Example - Fowler's Toads

Individual-based model of a species called Fowler's Toads (*Anaxyrus fowleri*) developed by Marchand et al 2017[1].

Model assumes that a toad hides in its refuge site in the daytime and moves to a randomly chosen foraging place at night.

GPS location data are collected on $n_t$ toads for $n_d$ days ($n_t = 66$ and $n_d = 63$ here). Denote matrix $\boldsymbol{Y}$.

$\boldsymbol{Y}$ is summarised to 4 sets comprising the moving distances for time lags of $1, 2, 4, 8$ days.

[1] Marchand et al 2017. A stochastic movement model reproduces patterns of site fidelity and long-distance dispersal in a population of Fowler's toads. Ecological Modelling.

# Fowler's Toads – Model

For each toad, first generate an overnight displacement, $\Delta y$. Displacement assumed to follow Lévy-alpha stable distribution family with parameters $\alpha$ and $\delta$.

Total returning probability is a constant $p_0$.

- Model 1: Random return to site.
- Model 2: Nearest return
- Model 3: Probability of return to site depends on distance (extra parameter $d_0$). Not considered in this talk.

Parameter of interest $\theta = (\alpha, \delta, p_0)$. Priors from Marchand et al: $\alpha \sim U(1, 2)$, $\delta \sim U(10, 100)$ and $p_0 \sim U(0, 1)$.

# Fowler's Toads – Summary Statistic Selection

Number of returns for all four time lags (defined as distance $< 10$m).

For the non-returns we consider log difference between adjacent $p$-quantiles with $p = 0, 0.1, \ldots, 1$ and also the median. Repeat for each time lag.

Roughly 40 statistics. Difficult for conventional ABC to deal with.

# Parametric Alternatives

It might be possible to overcome some drawbacks of ABC by using a parametric approximation to $p(s_y|\theta)$ instead of the non-parametric approximation used by ABC.

# Bayesian Synthetic Likelihood

The synthetic likelihood (SL) method of Wood 2010 uses a multivariate normal approximation: $p(s_y|\theta) \approx \mathcal{N}(s_y; \mu(\theta), \Sigma(\theta))$.

- Suitable when summary statistics are subject to the central limit theorem
- Transformations to multivariate normality of summary statistics
- Summary statistics from indirect inference
- Popular & convenient choice

We refer to a Bayesian version of SL as BSL (Price et al 2018[1]).

[1] Price et al 2018. Bayesian Synthetic Likelihood. JCGS.

# Bayesian Synthetic Likelihood

Basic method

- Simulate $n$ iid datasets from the model based on $\theta$
- Calculate the $n$ sets of summary statistics
- Calculate the sample mean, $\mu_n$, and sample covariance matrix, $\Sigma_n$, of the set of simulated summary statistics
- The BSL replacement likelihood is

$$\mathcal{N}(s_y; \mu_n(\theta), \Sigma_n(\theta)).$$

- Only tuning parameter is $n$ (we find weak dependence on this choice). Choose $n$ to maximise computational efficiency.

# BSL for Model Misspecification

We define the model as being compatible with the summary statistics if $\inf_\theta ||b(\theta) - b_0|| = 0$ where $b(\theta) = \text{plim} S(x)$ and $b_0 = \text{plim} S(y)$.

We have found that BSL performs inefficiently when compatability does not hold (e.g. model misspecification).

We are interested in modifications to the synthetic likelihood that allow BSL to run in such cases.

Also interested in identifying which statistics the model is not compatible with – might inform model refinement.

# Adjusted Synthetic Likelihoods

Consider two adjusted synthetic likelihoods for detecting incompatible summaries by introducing auxiliary variables $\gamma$ of dimension $\mathrm{dim}(s)$.

1. Variance Inflation (R-BSL-V):

$$V_n(\zeta) := \Sigma_n(\theta) + \begin{pmatrix} [\Sigma_n(\theta)]_{11}\gamma_1^2 & 0 & \cdots & 0 \\ 0 & [\Sigma_n(\theta)]_{22}\gamma_2^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & [\Sigma_n(\theta)]_{d_\eta d_\eta}\gamma_{d_\eta}^2 \end{pmatrix},$$

2. Mean Adjustment (R-BSL-M):

$$\phi_n(\zeta) = \mu_n(\theta) + \mathrm{diag}(\Sigma_n^{1/2}(\theta))\Gamma,$$

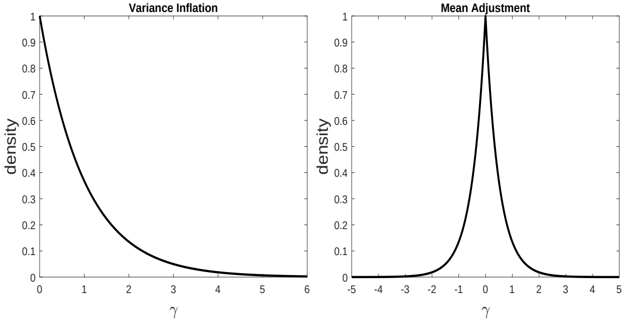
where $\Gamma$ is vector of $\gamma$'s.

# Priors for $\gamma$

We want to give the model a bit of extra flexibility, but not too much...

For variance inflation we take $\gamma \sim \text{Exp}(\mu)$. We set $\mu = 0.5$.

For mean adjustment we take $\gamma \sim \text{Laplace}(\mu = 0, b)$. We set $b = 0.5$.



# MCMC Sampling

Component-wise MCMC scheme.

Update $\theta|\gamma$ using standard Metropolis-Hastings.

Update each component of $\gamma|\theta$ using a slice sampler. Acceptance rate of 1, no additional tuning.

We don't seem to incur an additional penalty of sampling over a higher dimensional space.

# Theoretical Properties

If the model is compatible, what behavior should we expect from the R-BSL approach?

Under compatibility and other mild assumptions, we show that the posterior for $\Gamma$ converges to the prior.

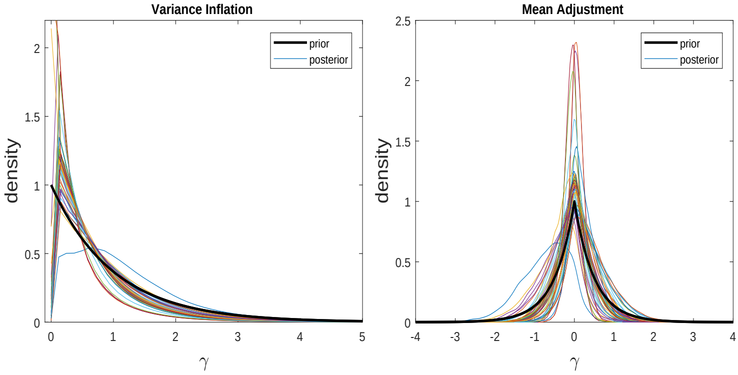
Thus incompatibility can be detected by departures from the prior.

# Toad: Results for Simulated Data from Model 2

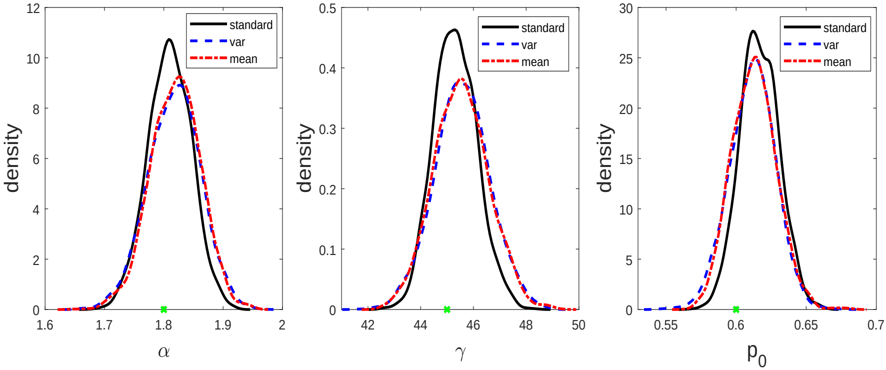
Consider model 2 as Marchand et al demonstrate that this model provides worse fit to data.

The MCMC acceptance rates for BSL, R-BSL-M and R-BSL-V are 11%, 9% and 22%, respectively ($n = 300$ for all methods).

Posteriors for $\gamma$:



# Toad: Results for Simulated Data from Model 2

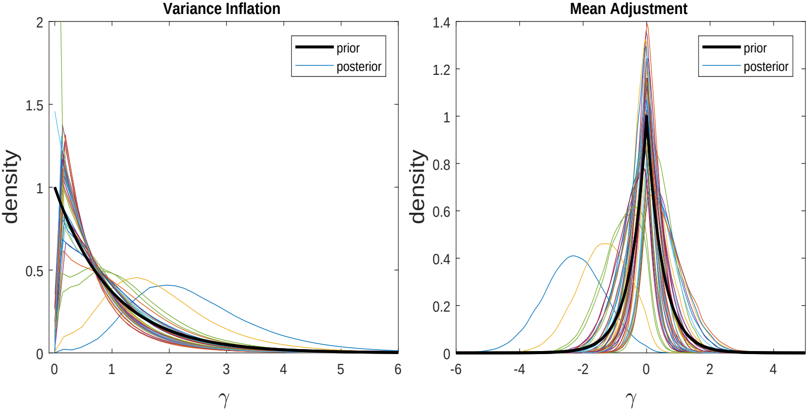



Univariate posterior distributions for the parameters when applying BSL (sold), R-BSL-V (dash) and R-BSL-M (dot-dash) to simulated data for the toad example. True parameter values are shown as crosses.

# Toad: Results for Real Data (Model 2)

MCMC acceptance rates of 9% (BSL with $n = 2000$), 7% (R-BSL-M with $n = 500$) and 15% (R-BSL-V with $n = 500$). R-BSL-V has roughly 1 order of magnitude computational efficiency gain over BSL.

Posteriors for $\gamma$:

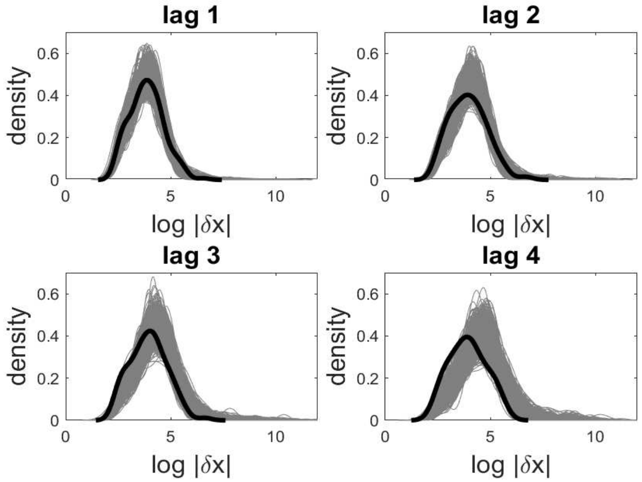


# Toad: Results for Real Data (Model 2)

Statistic with the largest incompatibility is the number of returns for lag 1. For R-BSL-V has 95% PI of (262, 346), the observed value is 234.

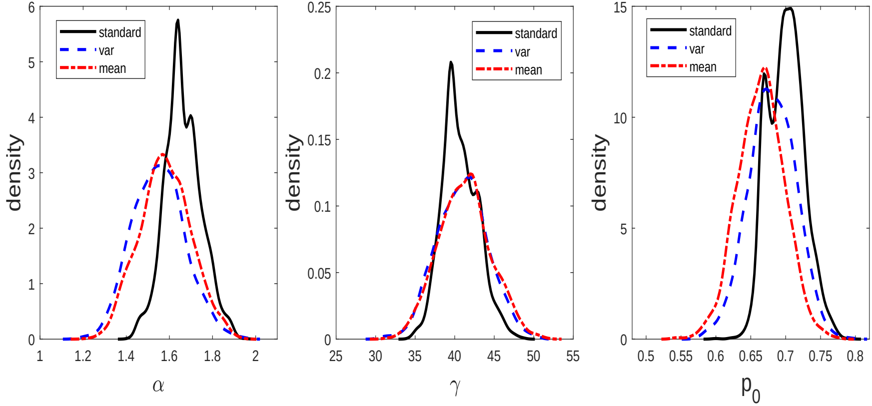
Other statistics showing some incompatibility are the first quantile differences of the non-returns for lags 4 and 8 days.

# Toad: Results for Real Data (Model 2)




Posterior predictive distributions of the log non-returns for the four lags.

# Toad: Results for Real Data (Model 2)




Univariate posterior distributions for the parameters when applying BSL (sold), R-BSL-V (dash) and R-BSL-M (dot-dash) to real data for the toad example.

# Summary and Future Work

Allows BSL to be run even when model cannot recover observed statistic.

Can identify incompatible summary statistics.

Variance inflation improves computational efficiency further, avoids numerical problems.

Updating auxiliary variables $\gamma$ seems to help prevent chain from getting stuck at overestimated synthetic likelihood.

Future Work: Extending this approach to other likelihood-free methods such as ABC.

# Other BSL Work

Using shrinkage covariance matrix estimators to reduce number of simulations (An et al 2019a, Ong et al 2018a).

Using semi-parametric density estimator to relax normality assumption (An et al 2019b).

Using VB instead of MCMC to speed things up (Ong et al 2018a, Ong et al 2018b).

Asymptotic properties of BSL (Frazier, Nott et al 2019).

We have an evolving R package for BSL:
`https://cran.r-project.org/web/packages/BSL/index.html`

# References

Price et al (2018). Bayesian synthetic likelihood. JCGS.

Frazier and Drovandi (2019). Robust Approximate Bayesian Inference with Synthetic Likelihood. https://arxiv.org/abs/1904.04551

Marchand et al 2017. A stochastic movement model reproduces patterns of site fidelity and long-distance dispersal in a population of Fowler's toads. Ecological Modelling.

Wood (2010). Statistical inference for noisy nonlinear ecological dynamic systems. Nature.

Handbook of Approximate Bayesian Computation (2018). Eds: Sisson, Fan, Beaumont. CRC Press.

Contact Details

Email: c.drovandi@qut.edu.au   Web: chrisdrovandi.weebly.com   Twitter: @chris_drovandi