Parameter Estimation Using Only Model Simulations

Associate Professor Chris Drovandi

School of Mathematical Sciences
QUT Centre for Data Science
ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)

26th June 2020

Why Statistical Modelling?

Gain insight into underlying biology, physics etc of a real process.

Perform hypothesis testing (e.g. is a new drug effective?).

Use the model to make predictions and perform “what-if" scenarios.

Generative (or mechanistic) models: model underlying mechanisms responsible for generating the data.

Why Parameter Estimation?

Models typically have unknown parameters (θ). For models to be most useful we need values for the parameters.

Natural approach: collect data from real system and estimate the parameters based on the data (sometimes referred to as “calibration").

Introduction to Bayesian Statistics

The Bayesian approach treats θ (vector) as a random variable.

Information about θ before data collection encapsulated in prior distribution $p(\theta)$.

Combine with the information we obtain about θ from the data y quantified by the likelihood function $p(y|\theta)$. Using Bayes' rule

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta),$$

where $p(\theta|y)$ is called the posterior distribution.

Simple Example

Consider tossing a two-sided coin 10 times with unknown probability θ of getting a head.

Assume that in 10 tosses we get $y = (1, 1, 1, 1, 0, 0, 1, 1, 1, 1)$ (i.e. 8 heads).

What is the posterior distribution for θ?

Simple Example (Cont...)

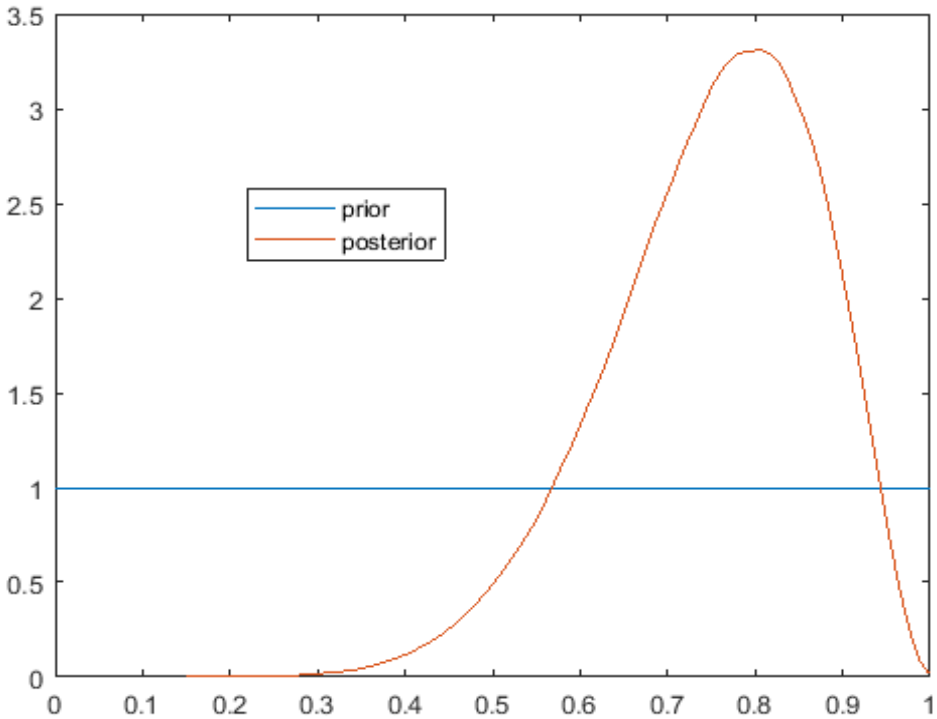
The number of "successes" out of a fixed number of trials has a binomial distribution. Thus the likelihood function is

$$p(y|\theta) = \binom{10}{8}\theta^8(1-\theta)^2.$$

A vague prior on θ may be uniform over (0,1), $p(\theta) = 1$, $0 < \theta < 1$.

Simple Example (Cont...)

We can show that $p(\theta|y)$ is Beta$(9, 3)$ where Beta(α, β) is the beta distribution with parameters α and β.



Bayesian Computational Algorithms

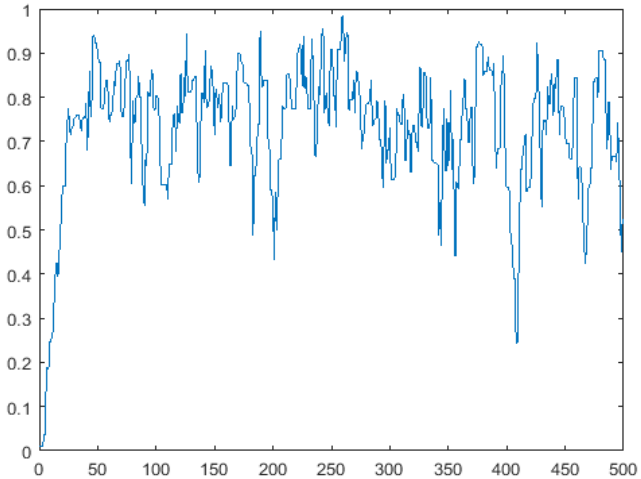
In most cases we don't know the form of the posterior distribution.

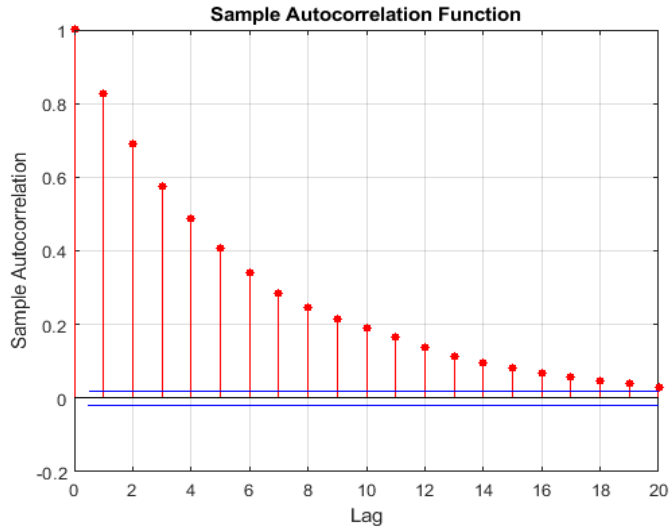
Computational algorithms (such as Markov chain Monte Carlo) have been designed to generate 'samples' from the posterior distribution.

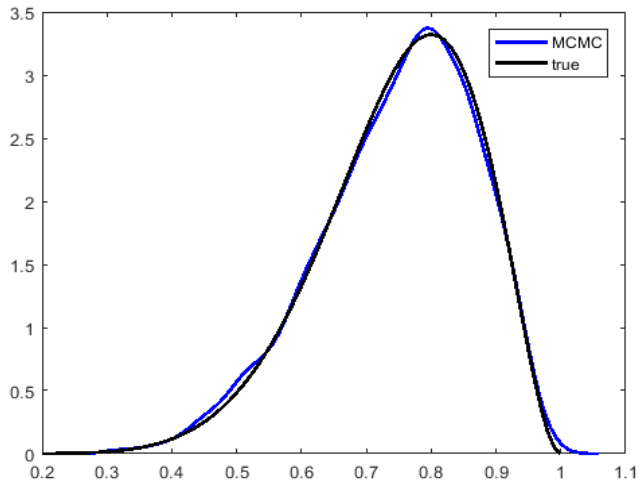
We can use these samples to estimate posterior quantities.

However, standard methods still require evaluation of the likelihood function.

MCMC on coin tossing example







Break for questions

Can't Compute the Likelihood?

For many complex models the likelihood function is not available.

But, often we can simulate/run the model for any given value of θ to produce (pseudo) data.

Can we estimate θ based only on model simulations?

Motivating Example – Collective Cell Spreading

Stochastic models of collective cell spreading are important for understanding processes such as cancer spread.

Parameter estimation is important for investigating the impact of drug treatments.

Drug treatments can be more effective during different phases of the cell cycle.

Here we consider a stochastic model of collective cell spreading[1], where cells move on a 2D hexagonal lattice and go through four phases.

[1] Simpson et al (2018). Physica A. 510:375-386.

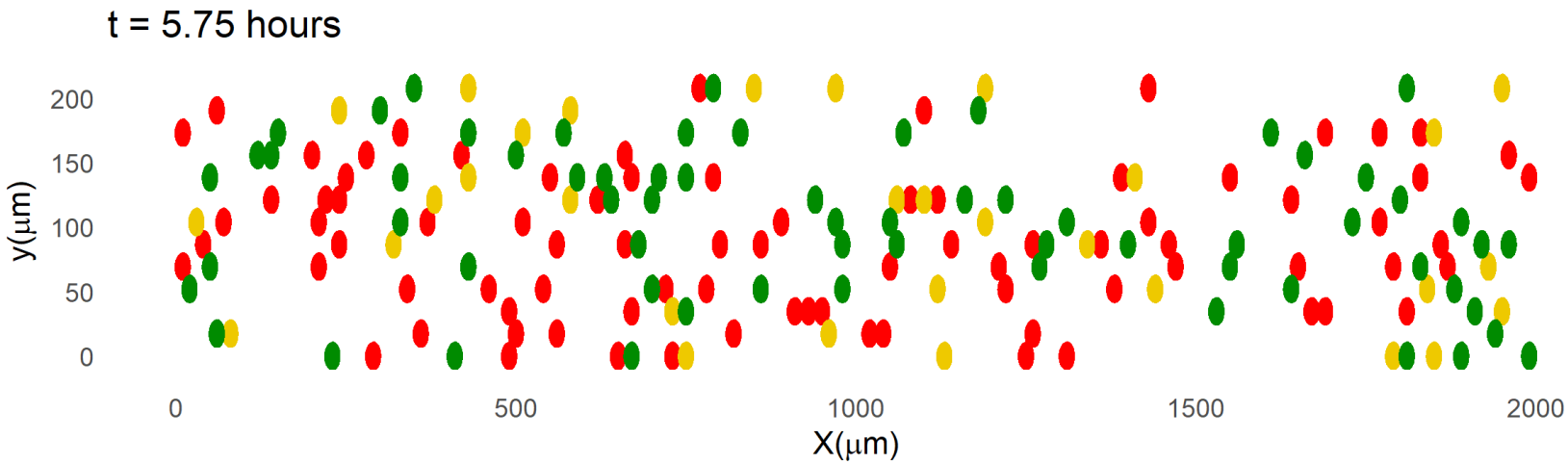
Motivating Example – Simulation Model

Algorithm 1 Collective Cell Spreading Model

1: Compute number of cells (Nred, Nyellow, Ngreen) in domain(x,y).
2: Compute $ar = P_r \times Nred$, $tr = K_{ry} \times Nred$
3: Compute $ay = P_y \times Nyellow$, $ty = K_{yg} \times Nyellow$
4: Compute $ag = P_g \times Ngreen$, $tg = K_{gr} \times Ngreen$
5: Compute $a0 = ar + ay + ag + tr + ty + tg$ and $R = u(0,1)$
6: **If** $R \leq \frac{ar}{a0}$ **then** Red cell migrates
7: **else if** $R \leq \frac{ar+ay}{a0}$ **then** Yellow cell migrates
8: **else if** $R \leq \frac{ar+ay+ag}{a0}$ **then** Green cell migrates
9: **else if** $R \leq \frac{ar+ay+ag+tr}{a0}$ **then** Red to Yellow cell transition
10: **else if** $R \leq \frac{ar+ay+ag+tr+ty}{a0}$ **then** Yellow to Green cell transition
11: **else** Green cell division and transition into red cells
12: Compute $\tau = \frac{1}{a0} \times log(\frac{1}{u(0,1)})$ and Set $t = t + \tau$
13: repeat until $t < t_{stop}$

QUT Centre for Data Science ACEMS

Motivating Example



Motivating Example

Here we consider tracking a certain number of cells for 48 hours.

The likelihood function is not available, but model simulation is relatively straightforward.

How can we estimate the parameters?

Approximate Bayesian Computation

Let's start with Bayes' rule:

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

Here we can't compute $p(y|\theta)$ directly, consider:

$$p(y|\theta) = \int_x 1(x = y)p(x|\theta)dx,$$

where x is data simulated from the model at θ. We can estimate this integral by taking a single simulation (or multiple).

Approximate Bayesian Computation

In general, it's not feasible to match x and y exactly, so we introduce a distance function ρ and tolerance ϵ: $1(\rho(y, x) \leq \epsilon)$.

Further, it is hard to match datasets if they are high dimensional. Introduce a summary statistic function $S(\cdot)$ believed to carry most information in the original data.

This results in the ABC posterior:

$$p_\epsilon(\theta, x|y) \propto 1(\rho(S(y), S(x)) \leq \epsilon)p(x|\theta)p(\theta),$$

In essence, we replace likelihood evaluation with model simulation and ask "is it close enough to the observed data?".

ABC Approximation

How approximate is ABC?

If $S(\cdot)$ is a sufficient statistic and $\epsilon \to 0$ then ABC is 'exact'.

However:

1. Generally not feasible to simulate perfect 'matches', i.e. require $\epsilon > 0$.
2. Most models do not have a low dimensional sufficient statistic, must resort to summary statistic (information loss).

These are the two sources of error in ABC.

Selecting Summary Statistics

These two sources of approximation are conflicting.

Ideally, we would like $S(\cdot)$ to be high dimensional, reducing information loss. But harder to find close matches in high dimension.

If $S(\cdot)$ is low dimensional, then we reduce the effect of ϵ, but might lose too much information.

General principle: Choose a summary statistic as low dimensional as possible, yet retaining as much information as possible (not easy in practice).

ABC Sampling

How do we generate samples from the ABC posterior?

Simplest method is ABC rejection[2].

1. Draw $\theta_i \sim p(\cdot)$ and simulate $x_i \sim p(\cdot|\theta_i)$ for $i = 1, \ldots, M$.
2. Compute discrepancy $\rho_i = \rho(S(y), S(x_i))$. Produces collection $\{\theta_i, \rho_i\}_{i=1}^{M}$
3. Keep $N = \alpha \times M$ of θ_i with smallest ρ_i (this defines the ϵ)

Choice of α trade off between accuracy and Monte Carlo error.

[2] Beaumont et al (2002). Genetics 162:2025-2035.

ABC Sampling

ABC rejection can be inefficient if (ABC) posterior different to the prior.

Can embed ABC into more efficient algorithms such as MCMC[1] and sequential Monte Carlo (SMC)[2].

SMC samples a sequence of distributions with decreasing ABC tolerance $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_T$. The proposal distribution for θ improves at each iteration.

[1] Marjoram et al (2003). PNAS 100:15324-15328.
[2] Sisson et al (2007). PNAS 104:1760-1765.

ABC coin tossing

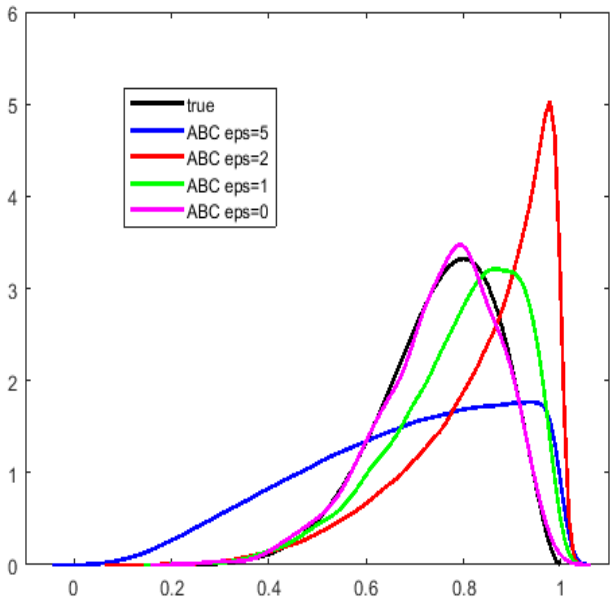
Let's return to the coin tossing example to illustrate the concepts.

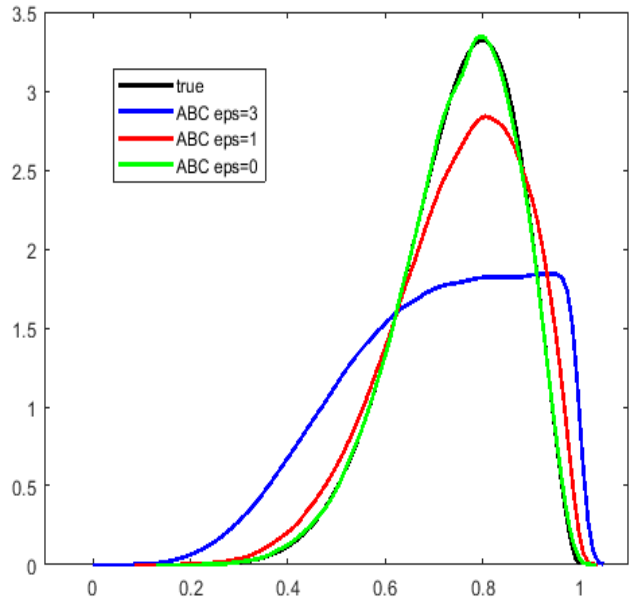
Clearly ABC is not really needed here, as the likelihood is completely tractable.

Here we run ABC rejection with:

- $M = 1$ million
- α = 50%, 10%, 1%, 0.1%.
- $S(y) = y$ and $S(y) = \text{sum}(y)$ (both are sufficient).
- ρ is sum of absolute differences.

ABC coin tossing results





(left) full data as summary statistic, (right) number of heads as summary statistic

Back to motiving example

Here we work with simulated data and treat is as “observed":

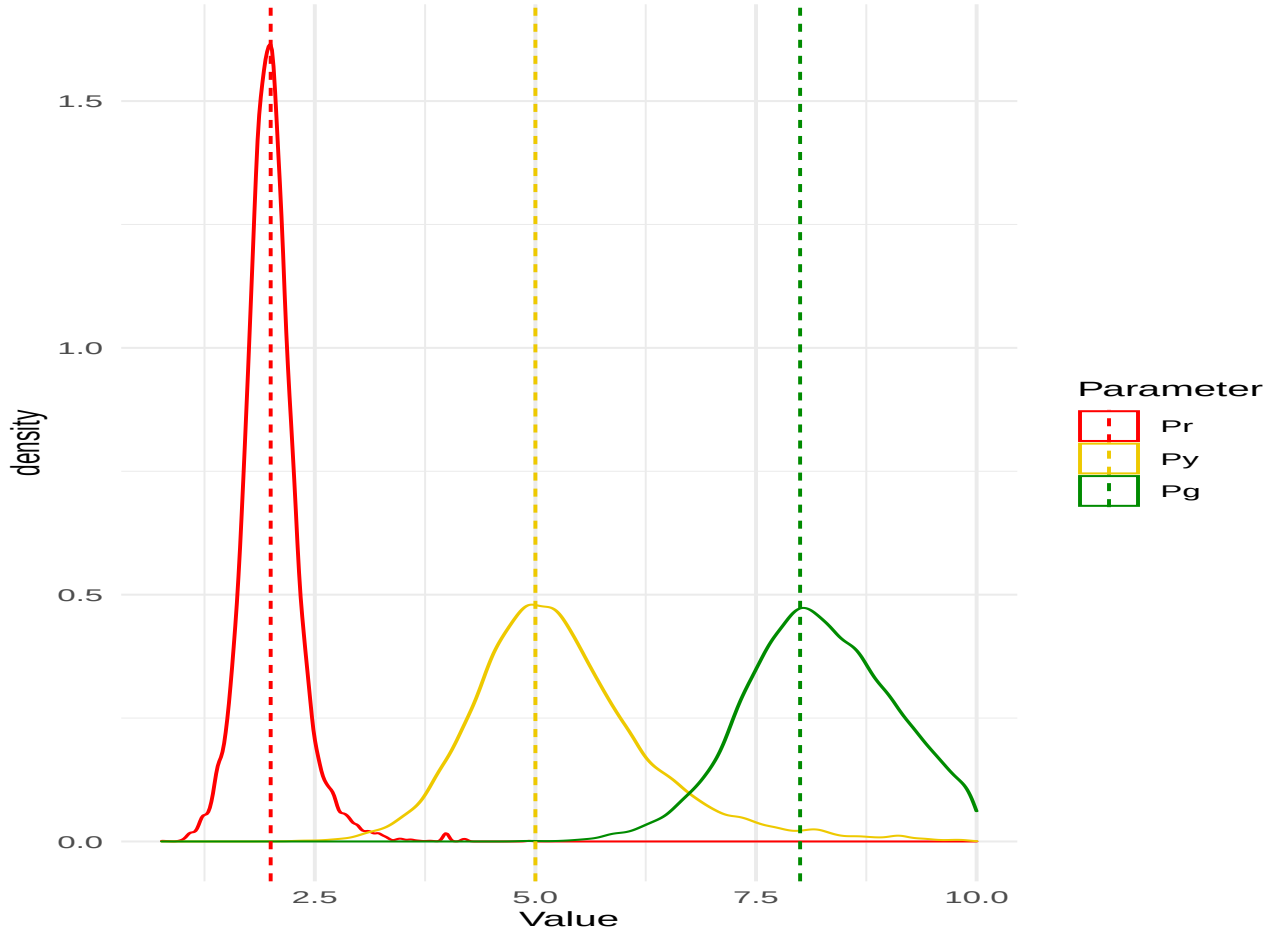
- Transition rates: $K_{ry} = 0.04, K_{yg} = 0.08, K_{gr} = 0.17$ (assumed known)
- Motility rates: $P_r = 2, P_y = 5, P_g = 8$ (to be estimated)

Summary statistics: Average distance (over 20 tracked cells) travelled in red, yellow, and green phases of cell cycle; S_r, S_y, S_g respectively.

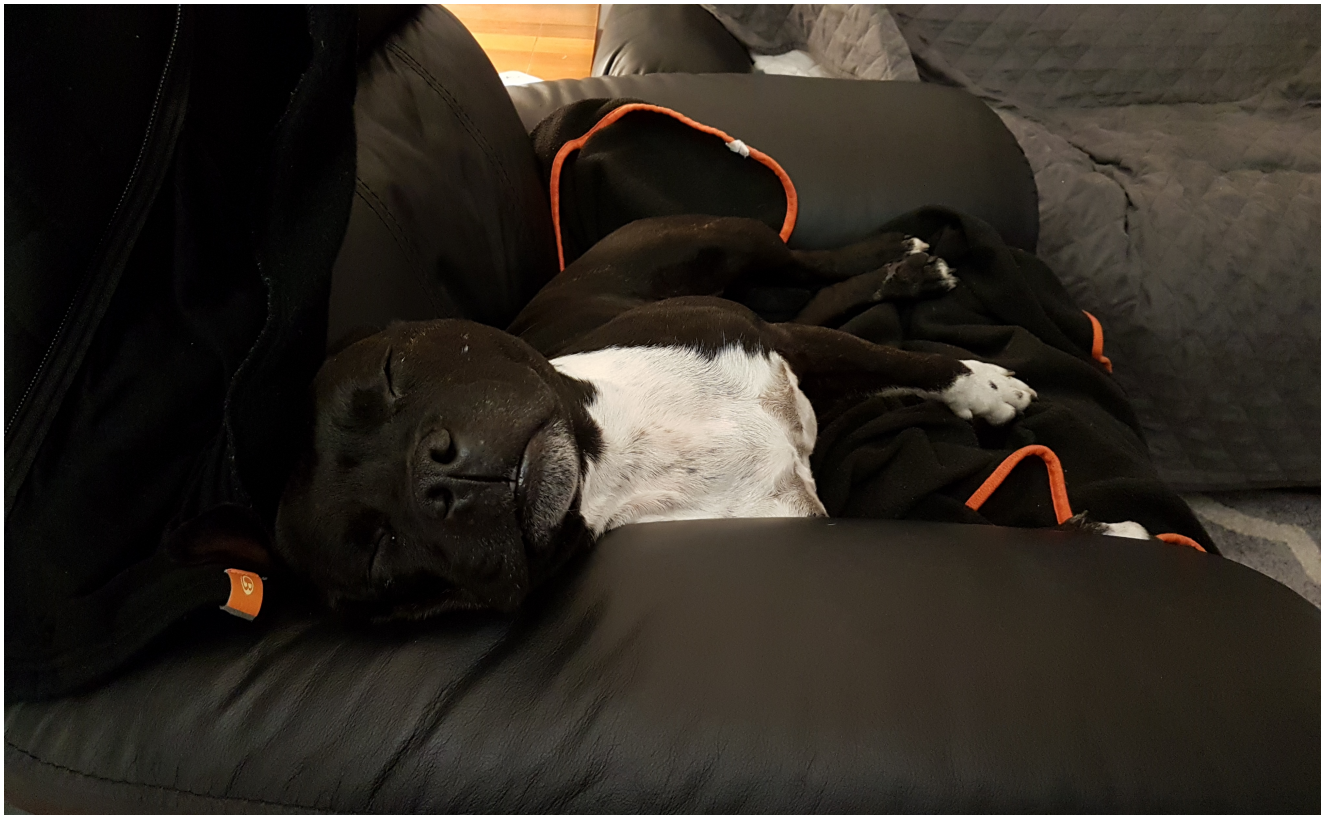
Distance function: Euclidean Distance

SMC ABC (ABC rejection too inefficient)

Motivating Example results



Break for questions



ABC Drawbacks

Highly sensitive to choice of tuning parameters ϵ, $\rho(\cdot)$.

No standard way to select ϵ or $\rho(\cdot)$.

Suffers from curse of dimensionality with respect to size of summary statistic.

Bayesian Synthetic Likelihood

An alternative approach that overcomes some limitations of ABC is Bayesian synthetic likelihood (BSL)[1][2].

BSL makes the assumption that the distribution of $p(S(x)|\theta) = \mathcal{N}(\mu(\theta), \Sigma(\theta))$ is multivariate normal.

We can estimate $\mu(\theta)$ and $\Sigma(\theta)$ via simulation.

[1] Wood (2010). Nature. 466:1102.

[2] Price et al (2018). Journal of Computational and Graphical Statistics. 27:1-11

Estimating the Synthetic Likelihood

Basic method

- Simulate n iid datasets from the model based on θ
- Calculate the n sets of summary statistics
- Calculate the sample mean, μ_n, and sample covariance matrix, Σ_n, of the set of simulated summary statistics
- The BSL replacement likelihood is

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

Pros and Cons

Advantages of BSL:

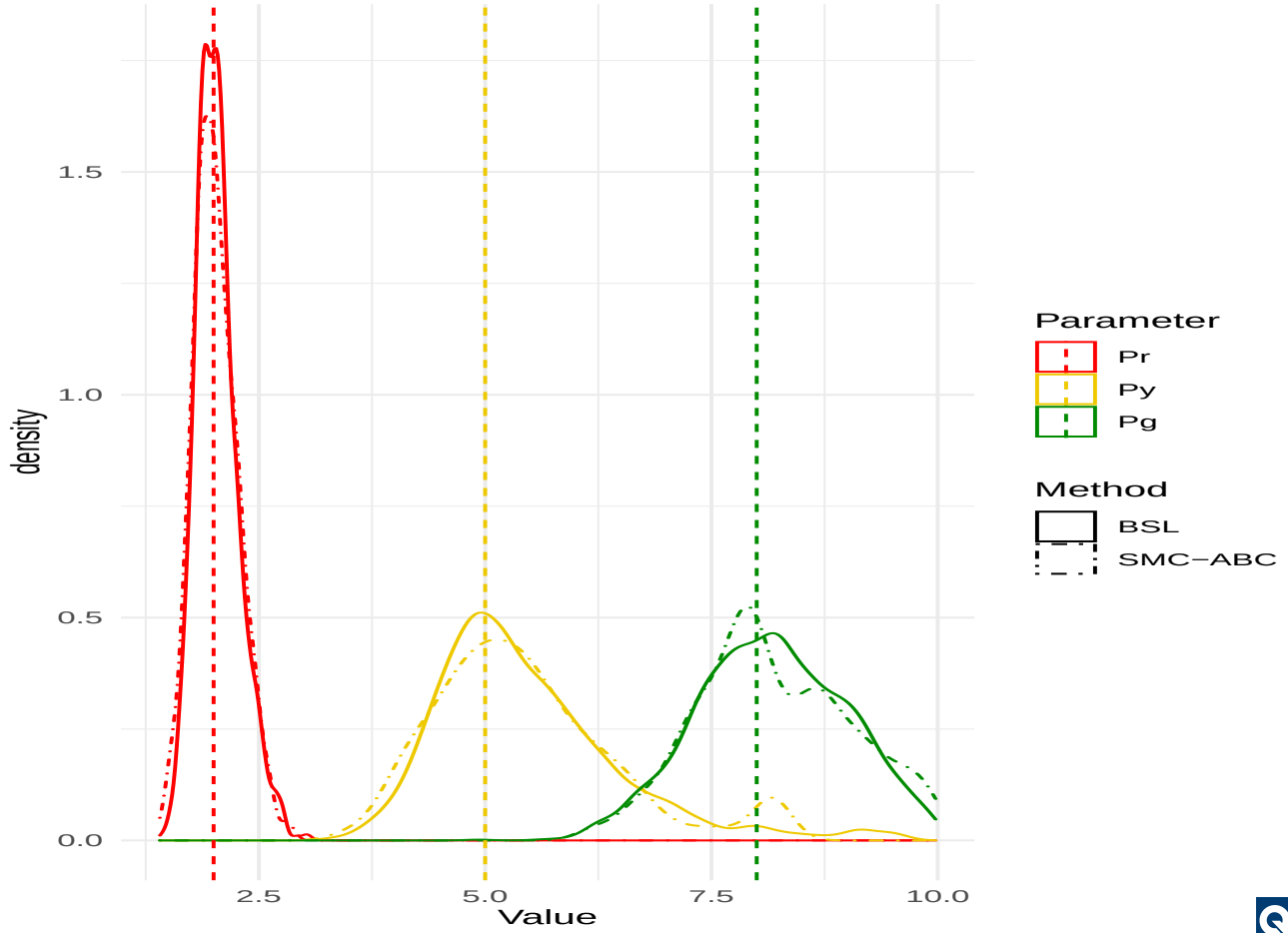
- No choice of ϵ and $\rho(\cdot)$ needed.
- Only tuning parameter is n (we find weak dependence on this choice). Choose n to maximise computational efficiency.
- Due to normality assumption, it scales better to high dimensional summary statistic.

Disadvantages of BSL:

- Strong normality assumption may not be reasonable.
- Still suffers from the curse of dimensionality with respect to summary statistic dimension.

Back to Motivating Example

Running MCMC BSL with $n = 12$.



BSL Extensions

Semiparametric estimators to relax normality assumption.

Covariance shrinkage estimation to reduce number of model simulations.

Decorrelation transformations to make covariance shrinkage estimation even more effective.

Robustness to model misspecification.

Theoretical properties of BSL.

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

Challenges

Scaling to high dimensional summary statistic and parameter

Handling expensive model simulators

Model selection

Model misspecification

References

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

Some of my research

Drovandi and Pettitt (2011) Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics.

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

An et al (2020) Robust Bayesian Synthetic Likelihood via a Semi-Parametric Approach. Statistics and Computing.

Priddle et al (2020) Efficient Bayesian synthetic likelihood with whitening transformations. https://arxiv.org/abs/1909.04857

Frazier and Drovandi (2020). Robust Approximate Bayesian Inference with Synthetic Likelihood. https://arxiv.org/abs/1904.04551

Contact Details

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

QUT Centre for Data Science ACEMS