Bayesian Experimental Design for Stochastic Biological Models with Computational Demanding Likelihoods

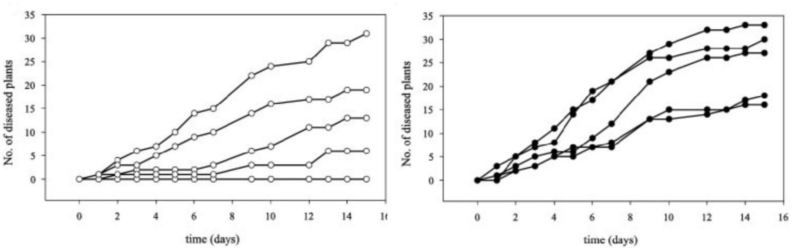
Dr Chris Drovandi
Queensland University of Technology
c.drovandi@qut.edu.au

Collaborators: Tony Pettitt, Mitchell O'Sullivan
Acknowledgements: Australian Research Council Discovery Grant and organisers of BISP8

June 7, 2013

Motivating Example

Kleczkowski et al (1996); Gibson et al (2004)



- Two competing models for each dataset (left treatment applied)
- When should samples be taken to estimate parameters or discriminate between models?

Models proposed for data

Markov Model (in deterministic DE form)

$$\frac{dI}{dt} = (r_p + r_s I)e^{-at}(N - I)$$

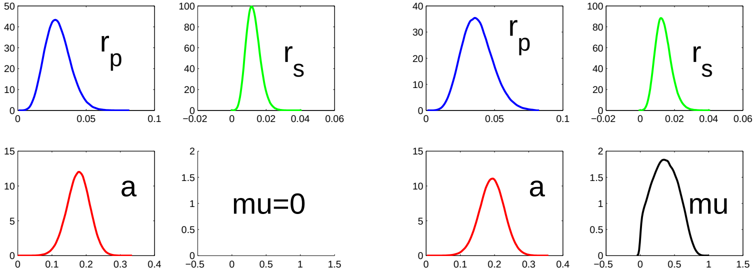
- Likelihood is expensive (if evaluated many times). $\boldsymbol{\theta} = (r_p, r_s, a)$.

Non-Markov Model (in deterministic DE form)

$$\frac{dE}{dt} = (r_p + r_s I)e^{-at}(N - I - E)$$

- An exposed (E) plant becomes infectious (I) after μ units of time. $\boldsymbol{\theta} = (r_p, r_s, a, \mu)$
- Likelihood intractable
- Easy to simulate from both models.
- Models are nested

The problem



Priors for future experiment with no treatment, Markov (left), non-Markov (right)

Models fitted via (auxiliary variable) MCMC.

- Posterior becomes the prior for future experiments (figure above)
- Objective: Determine optimal sampling times for estimating parameters and discriminating between models

Static Experimental Design

- Set-up: Have prior distribution $p(\boldsymbol{\theta})$ for parameter of statistical model, with likelihood $p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d})$.
- Wish to design for next n observations. Design variable: $\mathbf{d} = (d_1, \ldots, d_n)$.
- Define (general) utility function $u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})$. $\mathbf{y}$ is future data.

$$u(\mathbf{d}) = E_{\boldsymbol{\theta},\mathbf{y}}[u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})] = \int_{\mathbf{y}} \int_{\boldsymbol{\theta}} u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{y},$$

- Model uncertainty can be included ($m = 1, \ldots, K$ models)

$$u(\mathbf{d}) = \sum_{m=1}^{K} \int_{\mathbf{y}} \int_{\boldsymbol{\theta}} u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}, m) p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_{\mathbf{m}}, m) p(\boldsymbol{\theta}_{\mathbf{m}}|m) p(m) d\boldsymbol{\theta} d\mathbf{y},$$

- Objective

$$\mathbf{d}^* = \arg\max_{\mathbf{d} \in \mathcal{D}} u(\mathbf{d}).$$

Too difficult to do directly

MCMC approach - Muller 1999

- Turn optimisation problem into simulation problem

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}) p(\boldsymbol{\theta}).$$

  Admits $\propto u(\mathbf{d})$ as marginal distribution
- Optimal design $\mathbf{d}^*$ is mode of marginal
- If marginal is flat, may be beneficial to sample from

$$h(\mathbf{d}, \boldsymbol{\theta}_\mathbf{1}, \ldots, \boldsymbol{\theta}_\mathbf{J}, \mathbf{y}_\mathbf{1}, \ldots, \mathbf{y}_\mathbf{J}) \propto \prod_{j=1}^{J} u(\mathbf{d}, \mathbf{y}_\mathbf{j}, \boldsymbol{\theta}_\mathbf{j}) p(\mathbf{y}_\mathbf{j}|\mathbf{d}, \boldsymbol{\theta}_\mathbf{j}) p(\boldsymbol{\theta}_\mathbf{j}),$$

  for large J. A marginal is $\propto u(\mathbf{d})^J$
- Muller (1999) proposes MCMC for sampling. $q(\mathbf{d}^*, \mathbf{y}^*, \boldsymbol{\theta}^*|\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}^*|\boldsymbol{\theta}^*, \mathbf{d}^*) p(\boldsymbol{\theta}^*) q(\mathbf{d}^*|\mathbf{d})$.
- Easy to extend to model uncertainty

Bayesian utility functions

- Based on posterior, so independent of $\boldsymbol{\theta}$, $u(\mathbf{d}, \mathbf{y})$. Still need $\boldsymbol{\theta}$ for proposal in above MCMC

Parameter Estimation Design

- Kullback-Leibler divergence $u(\mathbf{d}, \mathbf{y}) = KL(p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})||p(\boldsymbol{\theta}))$. Mutual info between $\boldsymbol{\theta}$ and $\mathbf{y}$.
- Concentration of Posterior distribution
  - Set $u(\mathbf{d}, \mathbf{y})$ as entropy of $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$
  - Posterior precision $u(\mathbf{d}, \mathbf{y}) = 1/\det(\text{var}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}))$

Bayesian utility functions (cont...)

Model Discrimination Design

- For nested models focus on estimating ‘extra’ parameters.
- General Bayesian model discrimination utility is mutual information between M and $\mathbf{y}$. Drovandi et al 2013 JCGS show that

$$u(\mathbf{d}, \mathbf{y}, m) = \log p(m|\mathbf{y}, \mathbf{d}).$$

Even if likelihood available then $u(\mathbf{d}, \mathbf{y})$ and $p(m|\mathbf{y}, \mathbf{d})$ must still be estimated.

Muller Algorithm

- A closer look at the Muller algo... $\mathbf{d^{i-1}}$ is current, with utility $u^{i-1} = u(\mathbf{d^{i-1}}, \boldsymbol{\theta}^{\mathbf{i-1}}, \mathbf{y^{i-1}})$
- Propose $\mathbf{d}^* \sim q(\mathbf{d}|\mathbf{d^{i-1}})$, $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$, $\mathbf{y}^* \sim p(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{d}^*)$. Compute $u^* = u(\mathbf{d}^*, \boldsymbol{\theta}^*, \mathbf{y}^*)$
- Compute $\alpha = \min\left(1, \frac{u^* q(\mathbf{d^{i-1}}|\mathbf{d}^*)}{u^{i-1} q(\mathbf{d}^*|\mathbf{d^{i-1}})}\right)$
- Likelihood functions cancel as per ABC MCMC (Marjoram et al (2003))
- Now simply need utility function that does not require likelihood evaluation

Bayesian utilities in presence of intractable likelihoods

- Bayesian utilities based on posterior.
- Approximate true posterior via ABC (soon)

Parameter Estimation

- KLD between prior and ABC posterior (normal approximation)
- Concentration of ABC posterior (e.g. entropy or precision). Precision straightforward to calculate:

$$u(\mathbf{d}, \mathbf{y}) = 1/\det(\mathrm{Var}_{\theta|\mathbf{y},\mathbf{d}}(\theta)).$$

Model Discrimination

- KLD or concentration of ABC posterior of extra parameters.
- For mutual info use ABC posterior model probability

Approximate Bayesian Computation

- Simulation based method that does not involve likelihood evaluations
- Involves a joint 'approximate' posterior distribution

$$p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y}, \epsilon) \propto g(\mathbf{y}|\mathbf{x}, \epsilon)p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- where $g(\mathbf{y}|\mathbf{x}, \epsilon)$ is a weighting function. Popular choice $g(\mathbf{y}|\mathbf{x}, \epsilon) = 1(\rho(\mathbf{y}, \mathbf{x}) \leq \epsilon)$
- How to choose $\rho(\mathbf{y}, \mathbf{x})$?
  - Usually based on a set of low-dimensional set of summary statistics
  - Here use full discrete data (low-dimensional designs)
- Choice of ϵ trade-off between accuracy and efficiency (and Monte Carlo error)

The effect of the approximation

- Idea is that marginal $p(\boldsymbol{\theta}|\mathbf{y}, \epsilon) \approx p(\boldsymbol{\theta}|\mathbf{y})$
- Errors from insufficient summaries and $\epsilon > 0$

ABC Algorithms

- Many algorithms available: ABC rejection (Beaumont et al 2002), MCMC ABC (Marjoram et al (2003)), SMC ABC (Sisson et al (2007), CD+TP (2011))
- MCMC ABC and SMC ABC suitable for single dataset analysis
- In context of static design, need ABC posterior at each iteration for a new $\mathbf{y}$
- ABC rejection amendable to multiple datasets from same model
- Easy to extend ABC rejection to include model uncertainty (Grelaud et al 2009)

ABC Rejection

1. Generate $\boldsymbol{\theta}^{\mathbf{i}} \sim p(\boldsymbol{\theta})$ for $i = 1, \ldots, N$
2. Simulate $\mathbf{x}^{\mathbf{i}} \sim p(\mathbf{y}|\boldsymbol{\theta}^{\mathbf{i}}, \mathbf{d})$ for $i = 1, \ldots, N$
3. Compute discrepancies $\rho^i = \rho(\mathbf{y}, \mathbf{x}^{\mathbf{i}})$ for $i = 1, \ldots, N$, creating particles $\{\boldsymbol{\theta}^{\mathbf{i}}, \rho^i\}_{i=1}^N$
4. Sort the particle set via the discrepancy ρ
5. Calculate $\epsilon = \rho^{\lfloor \alpha N \rfloor}$ (where $\lfloor \cdot \rfloor$ denotes the floor function). The ABC posterior samples consist of the set $\{\boldsymbol{\theta}^{\mathbf{i}} | \rho^i \leq \epsilon\}_{i=1}^N$

Steps 1 and 2 are independent of data and $\{\boldsymbol{\theta}^{\mathbf{i}}, \mathbf{x}^{\mathbf{i}}\}_{i=1}^N$ can be stored.

Discrepancy function:

$$\rho(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^{D} \frac{|y_i - x_i|}{\text{std}_{p(\boldsymbol{\theta})}(x_i)}, \quad (1)$$

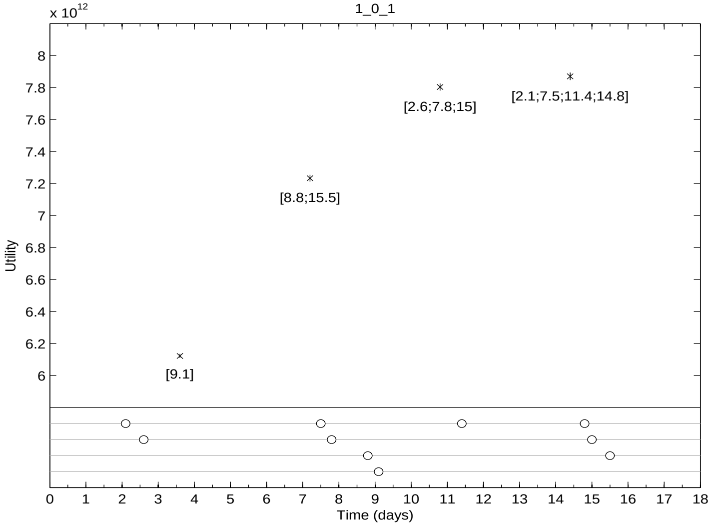
Discretising the Design Space

- Discretise design (time) space: $t_{min}, t_{max}, t_{inc}$
- Prior simulations get recorded at each design point in design space
- Prior simulations are done before Muller algorithm
- Metropolis-Hastings sample over discrete design space

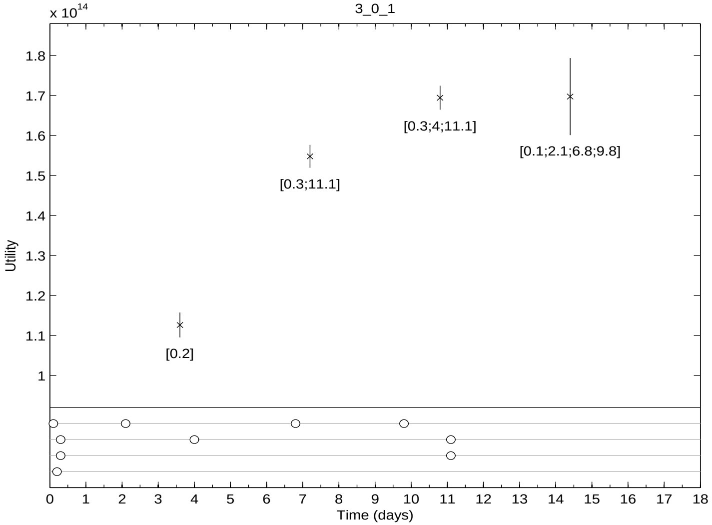
Algorithm Settings

- Run Muller algorithm for 100,000 iterations
- 100,000 prior simulations recorded at time interval $t_{inc} = 0.1$.
- Determine ϵ based on 200th smallest discrepancy.
- $J = 5$ for better identification of model
- Non-parametric estimate of multivariate density function of $\mathbf{d}$, then maximise. (see Cook et al (2008))

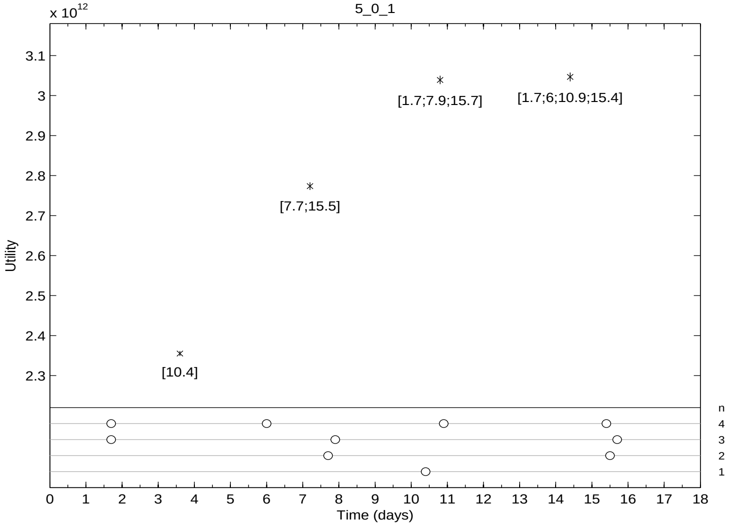
Parameter Estimation (Markov model)



Parameter Estimation (non-Markov model)



Parameter Estimation (non-Markov model excluding μ)



Parameter Estimation (non-Markov model just μ)

RESULTS STILL TO COME

Model Discrimination

RESULTS STILL TO COME

Limitations and Future Work

Limitations

- Only suitable for low dimensional designs and low number of sampling times.
- Initial condition needs to be fixed, or only a few options
- ABC rejection needs informative priors $\rightarrow$ but so does Bayesian design in general

Future Work

- Extend to larger number of sampling times (EP-ABC Algorithm)

Key References

- Drovandi and Pettitt (2013). Bayesian experimental design for models with intractable likelihoods. http://eprints.qut.edu.au/53924/.
- Cook, A. R., Gibson, G. J., and Gilligan, C. A. (2008). Optimal observation times in experimental epidemic processes. Biometrics, 64(3):860-868.
- Muller, P. (1999). Simulation based optimal design. In Bayesian statistics 6: proceedings of the Sixth Valencia International Meeting.
- Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. Genetics, 162(4):2025-2035.
- Gibson, G. J., Kleczkowski, A., and Gilligan, C. A. (2004) Bayesian analysis of botanical epidemics using stochastic compartmental models. PNAS, 101(3): 12120-12124.