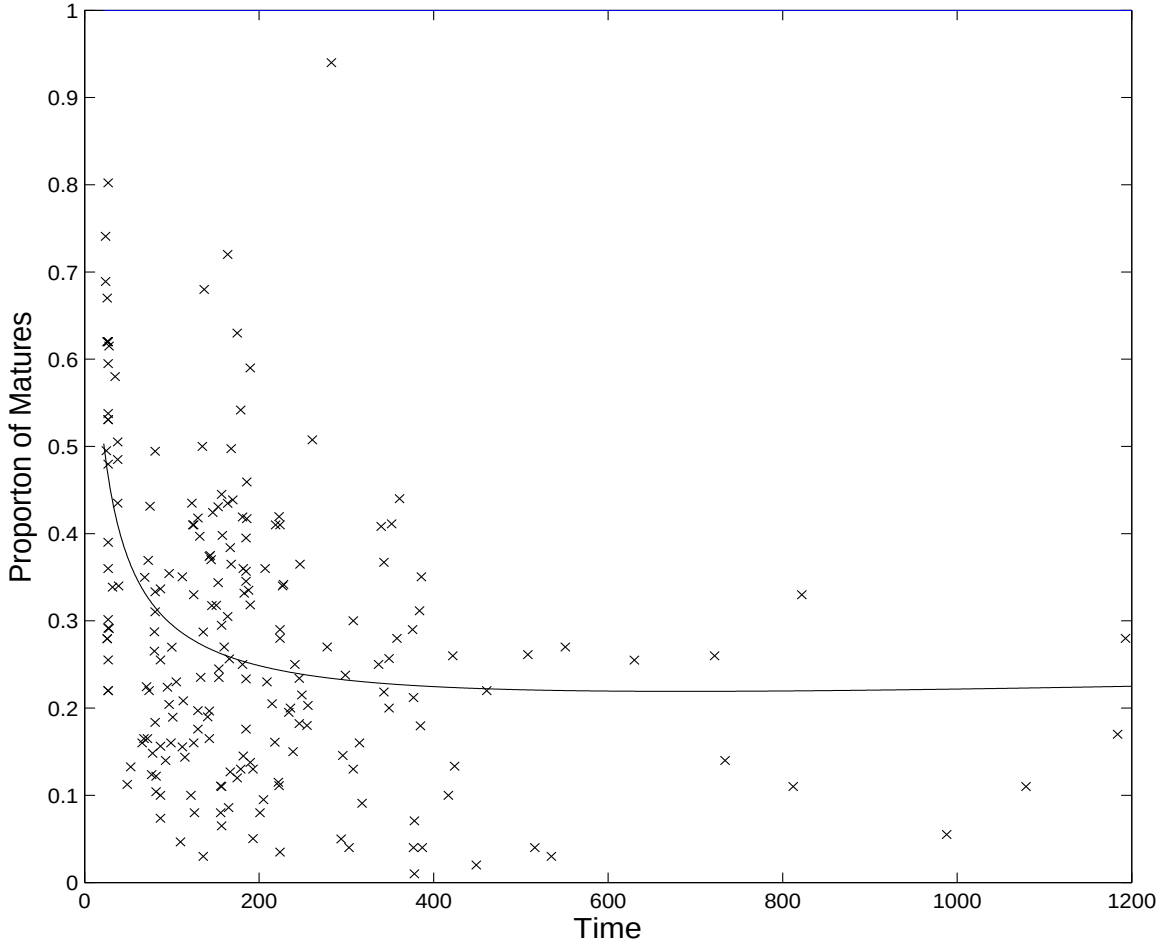
Bayesian Experimental Design for Models with Intractable Likelihoods

Chris Drovandi (me) and Tony Pettitt

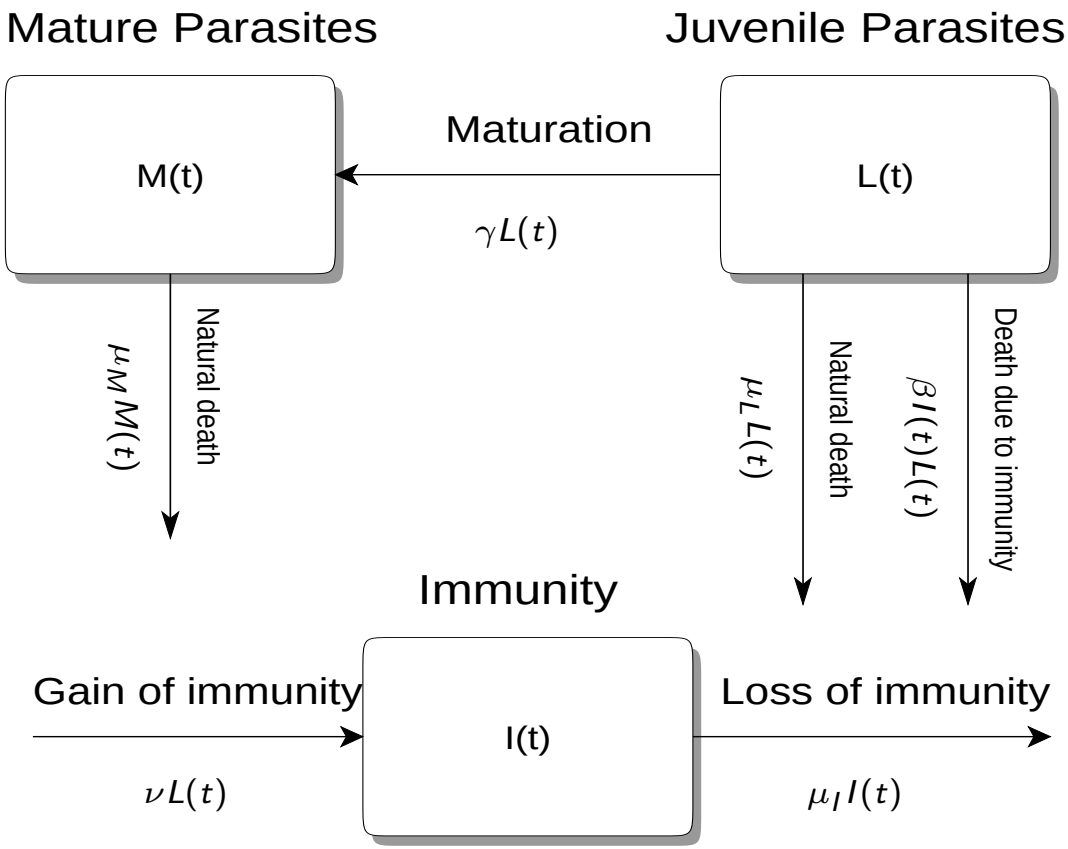
November 6, 2012

Macroparasite Immunity

- Estimate parameters of a Markov process model explaining macroparasite population development with host immunity
- 212 hosts (cats) $i = 1, \ldots, 212$. Each cat injected with l_i juvenile *Brugia pahangi* larvae (approximately 100 or 200).
- At time t_i host is sacrificed and the number of matures are recorded
- Host assumed to develop an immunity
- Three variable problem: $M(t)$ matures, $L(t)$ juveniles, $I(t)$ immunity.
- Only $L(0)$ and $M(t_i)$ is observed for each host


Proporton of Matures
Time
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
200
400
600
800
1000
1200

Trivariate Markov Process of Riley et al (2003)



The Model and Intractable Likelihood

- Deterministic form of the model

$$\frac{dL}{dt} = -\mu_L L - \beta I L - \gamma L,$$
$$\frac{dM}{dt} = \gamma L - \mu_M M,$$
$$\frac{dI}{dt} = \nu L - \mu_I I,$$

- μ_m, γ fixed. ν, μ_L, μ_I, β require estimation
- Likelihood based inference appears intractable.
  - Drovandi et al (2011) fit the model to data using approximate Bayesian computation

The problem

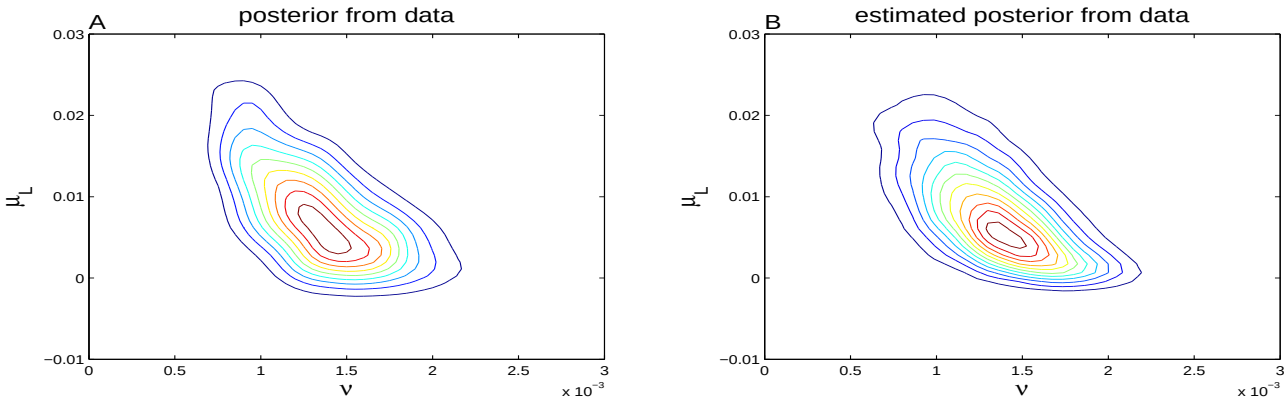


Figure : (a) bivariate density estimate of (ν, μ_L) obtained from Drovandi et al (2011). (b) bivariate density estimate based on samples from the parametric approximation of the posterior in (a).

- Posterior becomes the prior for future experiments (figure above)
- How to choose the sacrifice times and initial larvae injection to gain most information about parameters (ν and μ_L)???

Static Experimental Design

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

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

- 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

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

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

$$h(\mathbf{d}, \theta_1, \ldots, \theta_J, \mathbf{y_1}, \ldots, \mathbf{y_J}) \propto p(\mathbf{d}) \prod_{j=1}^{J} u(\mathbf{d}, \mathbf{y_j}, \theta_j) p(\mathbf{y_j}|\mathbf{d}, \theta_j) p(\theta_j),$$

  for large J. A marginal is $\propto u(\mathbf{d})^J p(\mathbf{d})$
- Muller (1999) propose MCMC for sampling.
  $q(\mathbf{d}^*, \mathbf{y}^*, \theta^*|\mathbf{d}, \mathbf{y}, \theta) = p(\mathbf{y}^*|\theta^*, \mathbf{d}^*) p(\theta^*) q(\mathbf{d}^*|\mathbf{d})$.
- Amzal et al (2006). Use SMC by gradually increasing J.

Utility functions and Computational issues

- Frequentist designs use Fisher information matrix. $u(\mathbf{d}, \boldsymbol{\theta})$ (utility independent of data, fun times!). Called robust or pseudo-Bayesian designs.
- Bayesian utilities typically $u(\mathbf{d}, \mathbf{y})$. Based on posterior, so independent of $\boldsymbol{\theta}$. Still need $\boldsymbol{\theta}$ for proposal (not so fun times!)
- Kullback-Leibler divergence $u(\mathbf{d}, \mathbf{y}) = KL(p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})||p(\boldsymbol{\theta}))$
- Concentration of Posterior distribution
  - Set $u(\mathbf{d}, \mathbf{y})$ as entropy of $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$
  - Posterior precision $u(\mathbf{d}, \mathbf{y}) = 1/\det(\text{var}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}))$
- $u(\mathbf{d}, \mathbf{y})$ must usually be approximated (also not so cool...)

Bayesian designs more difficult computationally: Sampling over large design space + parameter + future data (hard slog)

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 $\alpha = \min\left(1, \frac{u^* p(\mathbf{d}^*) q(\mathbf{d^{i-1}}|\mathbf{d}^*)}{u^{i-1} p(\mathbf{d^{i-1}}) q(\mathbf{d}^*|\mathbf{d^{i-1}})}\right)$
- Likelihood functions cancel as per ABC MCMC (Marjoram et al (2003))
- Now simply need utility function that does not require likelihood evaluation

Utility for intractable likelihoods

- Fisher information not always helpful here (It's not Bayesian anyway!). Can be done for Markov processes (see Pagendam and Ross (2012)), but expensive.
- Bayesian utilities based on posterior.
- Approximate true posterior via approximate Bayesian computation (later)

Utility

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

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 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 data 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

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. Discard $(1 - \alpha)N$ of the particles with the highest discrepancy. Effectively $\epsilon = \rho^{\alpha N}$

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:

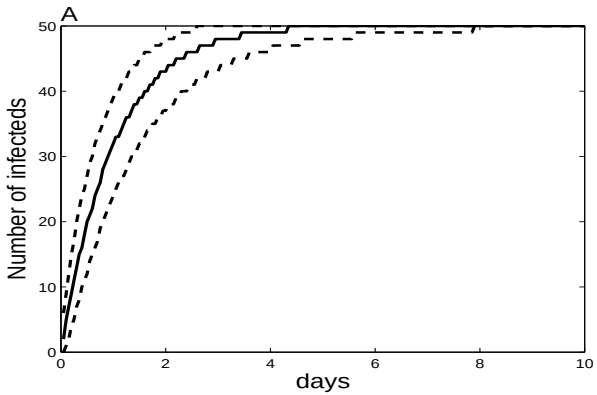
$$\rho(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^{D} \frac{|y_i - x_i|}{\text{std}_{p(\theta)}(x_i)}, \tag{1}$$

Discretising the Design Space

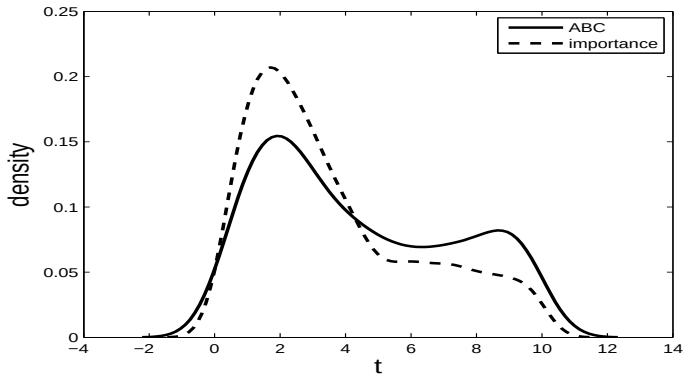
- Discretise design (time) space: $t_{min}, t_{max}, t_{inc}$
- Prior simulations get recorded at each design point in design space
- Prior simulations are done before Muller algorithm
- Metropolis-Hastings sample over discrete design space

Example - pure death process

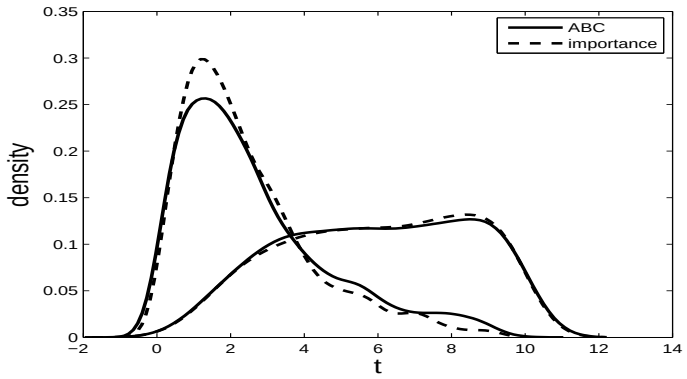
- SI Model, pure death, closed population.
  $P(S(t+\delta_t) = i-1 | S(t) = i) = b_1 i \delta_t + o(\delta_t)$.
  $b_1 \sim$ log normal$(-0.005, 0.01)$
- Binomial likelihood. Allows comparison of likelihood-free (100,000 prior simulations), likelihood-based designs
- $t_{\min} = 0.01$, $t_{\max} = 10$ and $t_{\text{inc}} = 0.01$.
- Likelihood-free: ABC posterior. Likelihood-based: Posterior via importance sampling



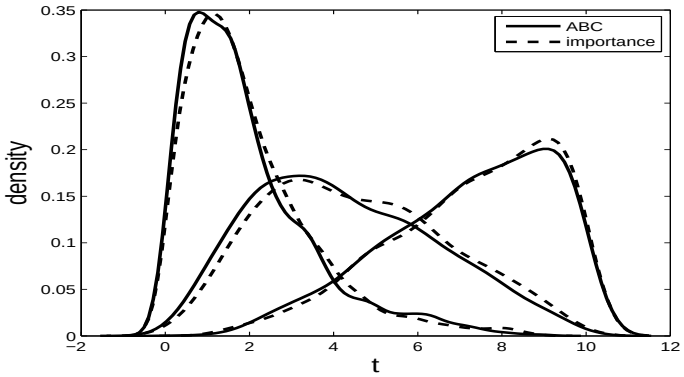
Results



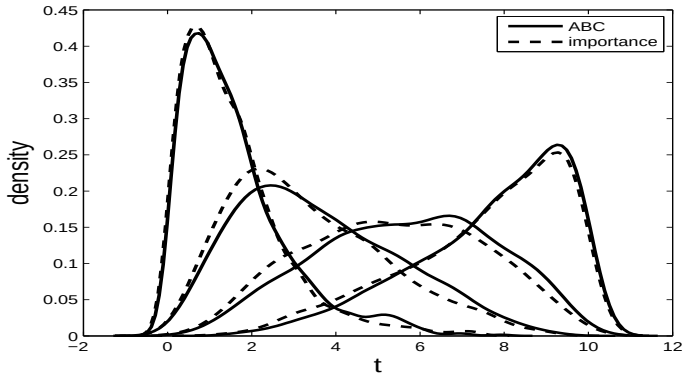
(a) 1 design point



(b) 2 design points



(c) 3 design points

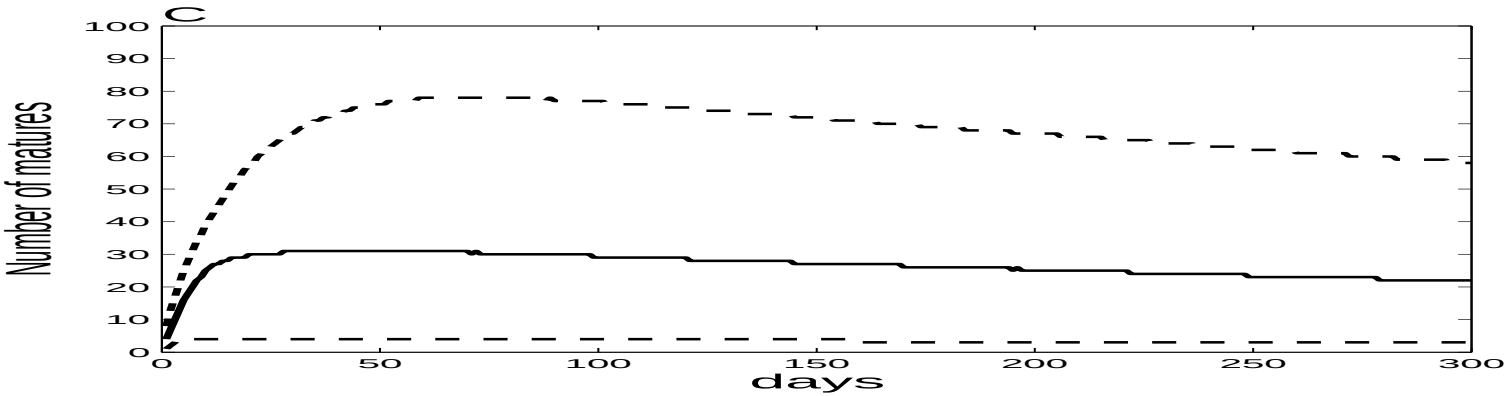


(d) 4 design points

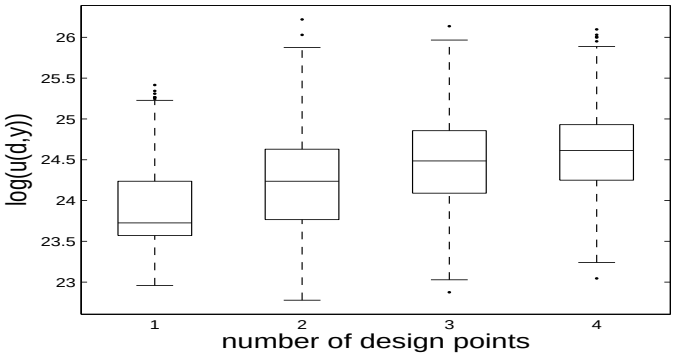
Figure : Experimental design results for death model for up to 4 design

Macroparasite Population Evolution

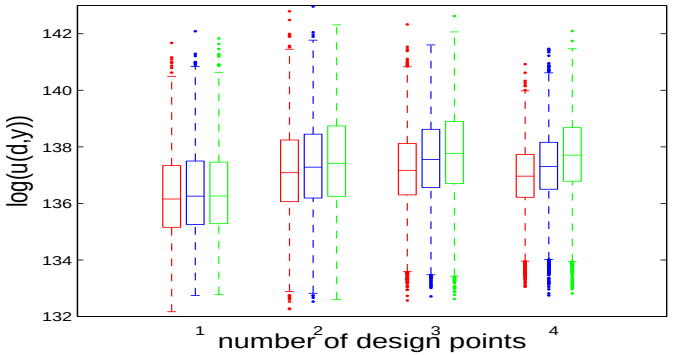
- Prior on $(\nu, \sqrt{\mu_L})$ is bivariate normal with mean (0.0013,0.0854), variance of $(1.0469 \times 10^{-7}, 0.0012)$ and a correlation of -0.6865.
- Lot's of prior predictive variability: trial 200,000 and 800,000 prior simulations
- $t_{\min} = 1$, $t_{\max} = 300$ and $t_{\text{inc}} = 1$ days



Results



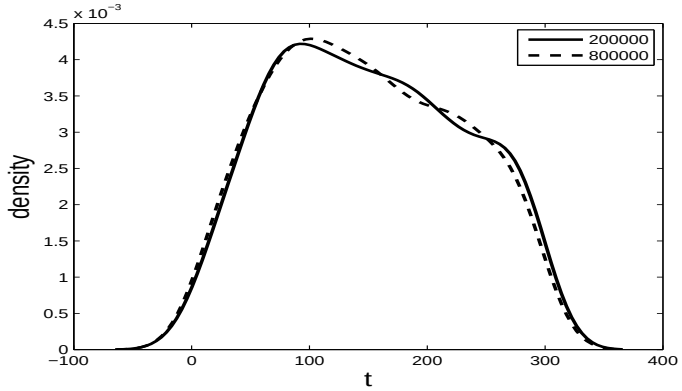
(a) death process



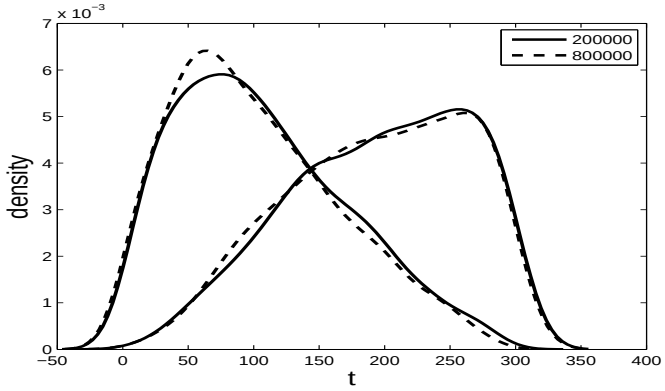
(b) macroparasite proces

Figure : Change in utility

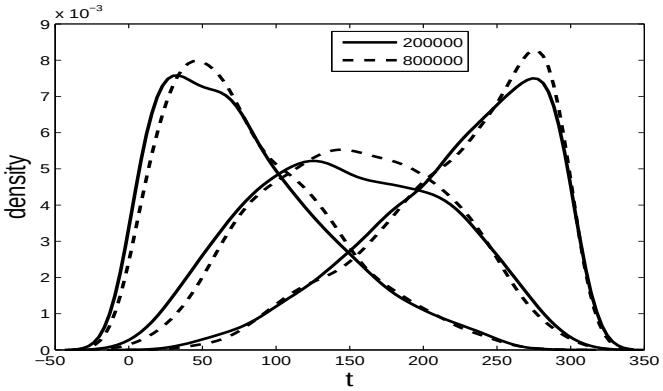
Results



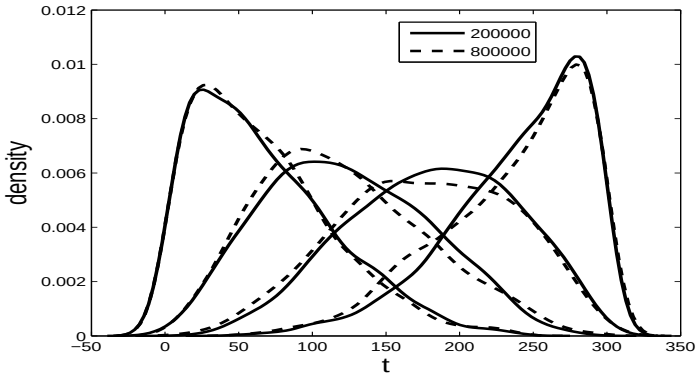
(a) 1 design point



(b) 2 design points



(c) 3 design points



(d) 4 design points

Figure : Experimental design results for the macroparasite model for up

Results - Gain in utility

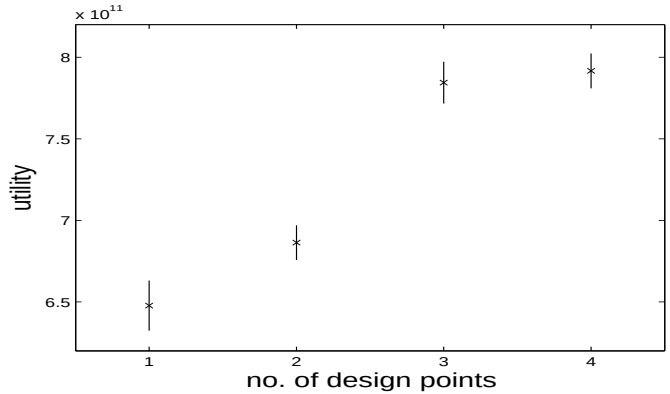
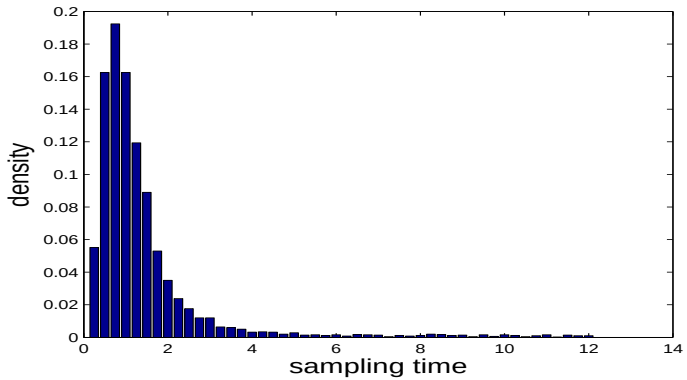


Figure : The expected utility (with error bars) of the designs $\mathbf{d_1} = (100)$, $\mathbf{d_2} = (50, 300)$, $\mathbf{d_3} = (50, 150, 300)$ and $\mathbf{d_4} = (25, 100, 175, 300)$ for the macroparasite model.

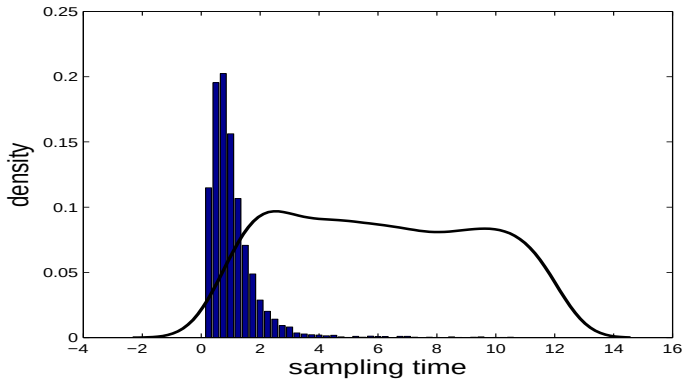
Stochastic DE - PK Model

- $dX_t = \left(\frac{D k_a k_e}{C_l} e^{-k_a t} - k_e X_t \right) dt + \sigma dW_t$
- Log-normal priors on k_a, k_e and C_l
- Exact simulation not available. Approximate simulation via Euler's method
- 200,000 prior simulations with $D = 4$
- $t_{\min} = 0.25$, $t_{\max} = 12$ and $t_{\text{inc}} = 0.25$ days

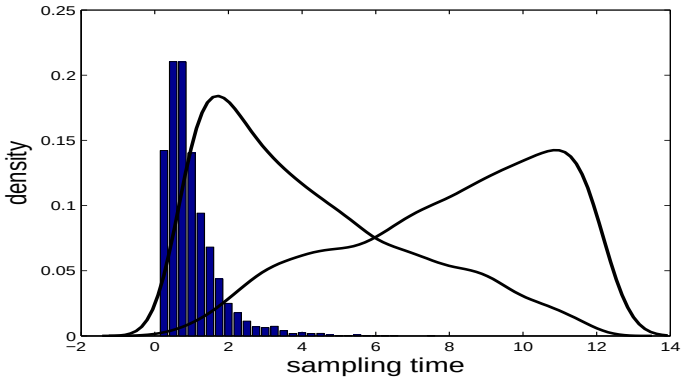
Results



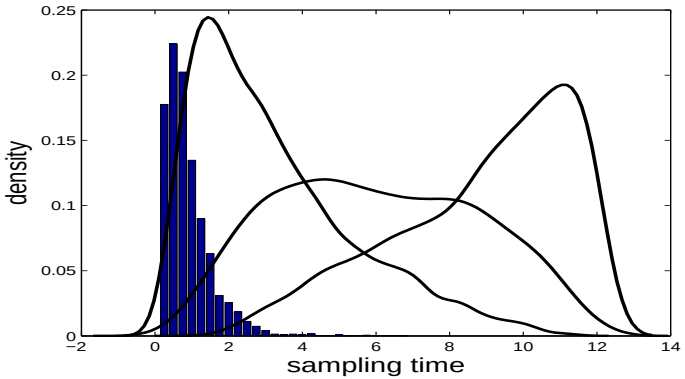
(a) 1 design point



(b) 2 design points



(c) 3 design points



(d) 4 design points

Figure : PK Example.

Limitations and Future Work

Limitations

- Only suitable for low dimensional designs. Bayesian high dimensional designs difficult even with likelihood!
- 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

- Macroparasite Example: Extend to choose either 100 or 200 initial larvae
- Extend to higher dimensional design (ABC rejection might die)
- Model discrimination for intractable likelihoods
- Indirect inference?

References

- Drovandi and Pettitt (2012). 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.
- Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011). Approximate Bayesian computation using indirect inference. Journal of the Royal Statistical Society: Series C (Applied Statistics), 60(3):503-524.