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.

Mature Parasites

\[ M(t) \]

- Natural death: \( \gamma L(t) \)
- Maturation: \( \gamma L(t) \)
- Gain of immunity: \( \nu L(t) \)
- Loss of immunity: \( \mu I(t) \)

Juvenile Parasites

\[ L(t) \]

- Natural death: \( \nu L(t) \)
- Death due to immunity: \( \beta L(t)L(t) \)
- Gain of immunity: \( \nu L(t) \)
- Loss of immunity: \( \mu I(t) \)
The Model and Intractable Likelihood

- Deterministic form of the model

\[
\begin{align*}
\frac{dL}{dt} &= -\mu_L L - \beta IL - \gamma L, \\
\frac{dM}{dt} &= \gamma L - \mu_M M, \\
\frac{dl}{dt} &= \nu L - \mu_I I,
\end{align*}
\]

- \(\mu_m, \gamma\) fixed. \(\nu, \mu_L, \mu_I, \beta\) 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 \((\nu, \mu_L)\) 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 \((\nu \text{ and } \mu_L)\)???
Static Experimental Design

- Set-up: Have prior distribution \( p(\theta) \) for parameter of statistical model, with likelihood \( p(y|\theta, d) \).

- Wish to design for next \( n \) observations. Design variable: \( d = (d_1, \ldots, d_n) \).

- Define (general) utility function \( u(d, y, \theta) \). \( y \) is future data.

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

- Objective

\[
d^* = \arg \max_{d \in D} u(d).
\]

Too difficult to do directly
Motivating Example

Static Design

Design for Intractable Models

Examples

Ending

MCMC approach - Muller 1999

- Turn optimisation problem into simulation problem

\[ h(d, \theta, y) \propto u(d, y, \theta)p(y|d, \theta)p(\theta)p(d). \]

Admits \( \propto u(d)p(d) \) as marginal distribution

- Optimal design \( d^* \) is mode of marginal

- If marginal is flat, may be beneficial to sample from

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

for large \( J \). A marginal is \( \propto u(d)^J p(d) \)

- Muller (1999) propose MCMC for sampling.

\[ q(d^*, y^*, \theta^*|d, y, \theta) = p(y^*|\theta^*, d^*)p(\theta^*)q(d^*|d). \]

Utility functions and Computational issues

- Frequentist designs use Fisher information matrix. \( u(d, \theta) \)
  (utility independent of data, fun times!). Called robust or pseudo-Bayesian designs.

- Bayesian utilities typically \( u(d, y) \). Based on posterior, so independent of \( \theta \). Still need \( \theta \) for proposal (not so fun times!)

- Kullback-Leibler divergence \( u(d, y) = KL(p(\theta|y, d)||p(\theta)) \)

- Concentration of Posterior distribution
  - Set \( u(d, y) \) as entropy of \( p(\theta|y, d) \)
  - Posterior precision \( u(d, y) = 1/\text{det}(\text{var}(\theta|y, d)) \)

- \( u(d, y) \) 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... $d^{i-1}$ is current, with utility $u^{i-1} = u(d^{i-1}, \theta^{i-1}, y^{i-1})$
- Propose $d^* \sim q(d|d^{i-1})$, $\theta^* \sim p(\theta)$, $y^* \sim p(y|\theta^*, d^*)$.
  Compute $u^* = u(d^*, \theta^*, y^*)$
- Compute $\alpha = \min \left( 1, \frac{u^* p(d^*) q(d^{i-1}|d^*)}{u^{i-1} p(d^{i-1}) q(d^*|d^{i-1})} \right)$
- 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

\[ p(\theta, x|y, \epsilon) \propto g(y|x, \epsilon)p(x|\theta)p(\theta) \]

where \( g(y|x, \epsilon) \) is a weighting function. Popular choice 
\[ g(y|x, \epsilon) = 1(\rho(y, x) \leq \epsilon) \]

- How to choose \( \rho(y, x) \)?
  - Usually based on a set of low-dimensional set of summary statistics
  - Here use full data (low-dimensional designs)

- Choice of \( \epsilon \) trade-off between accuracy and efficiency (and Monte Carlo error)

The effect of the approximation
- Idea is that marginal \( p(\theta|y, \epsilon) \approx p(\theta|y) \)
- Errors from insufficient summaries and \( \epsilon > 0 \)

MCMC ABC and SMC ABC suitable for single data analysis

In context of static design, need ABC posterior at each iteration for a new $y$

ABC rejection amendable to multiple datasets from same model
ABC Rejection

1. Generate $\theta^i \sim p(\theta)$ for $i = 1, \ldots, N$
2. Simulate $x^i \sim p(y|\theta^i, d)$ for $i = 1, \ldots, N$
3. Compute discrepancies $\rho^i = \rho(y, x^i)$ for $i = 1, \ldots, N$, creating particles $\{\theta^i, \rho^i\}_{i=1}^N$
4. Sort the particle set via the discrepancy $\rho$
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 $\{\theta^i, x^i\}_{i=1}^N$ can be stored.

Discrepancy function:

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

(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.
  \[ P(S(t + \delta_t) = i - 1|S(t) = i) = b_1 i \delta_t + o(\delta_t) \]
  \[ b_1 \sim \log \text{normal}(-0.005, 0.01) \]

- 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

![Graph showing the number of infecteds over time]
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 points.
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

$t_{\text{min}} = 1$, $t_{\text{max}} = 300$ and $t_{\text{inc}} = 1$ days
Results

(a) death process  
(b) macroparasite process

Figure: Change in utility
Motivating Example

Results

Figure: Experimental design results for the macroparasite model for up to 4 design points. The solid line and dashed line results are based on Chris Drovandi Bayes on the Beach.
Results - Gain in utility

Figure: The expected utility (with error bars) of the designs $d_1 = (100)$, $d_2 = (50, 300)$, $d_3 = (50, 150, 300)$ and $d_4 = (25, 100, 175, 300)$ for the macroparasite model.
Stochastic DE - PK Model

\[ dX_t = \left( \frac{Dk_ake}{C_l} e^{-k_at} - k_eX_t \right) dt + \sigma dW_t \]

- 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 \( D = 4 \)
- \( t_{\text{min}} = 0.25, \ t_{\text{max}} = 12 \) and \( t_{\text{inc}} = 0.25 \) days
Results

Figure: PK Example.

(a) 1 design point

(b) 2 design points

(c) 3 design points

(d) 4 design points
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