Particle filter approach to Bayesian sequential design

J. M. McGree

with C. C. Drovandi and A. N. Pettitt

Queensland University of Technology
Mathematical Sciences
james.mcgree@qut.edu.au

August 29, 2011
Introduction

- Particle filter approach for sequential design;
- Computationally convenient - Information from new data is incorporated via a re-weighting step;
- We consider a flexible parametric model;
- Explore ‘fully Bayesian’ utility functions including a proposed hybrid utility;
- Application: Estimation of the maximum tolerated dose (MTD) in phase I clinical trials.

Outline of Sequential design

- Use currently available data to make informed decisions (e.g. dose selection);
- Decision based on optimizing some utility function;
- Useful approach for designs for nonlinear models in the presence of parameter and model uncertainty;
- The incorporation of parameter and model uncertainty is achieved within a Bayesian framework;
- The sequential nature of how the data is observed lends itself sequential Monte Carlo (SMC);
- Methodology for general discrete observation processes. However, focus on binary outcomes.

The design of phase I clinical trials

- Early stage clinical trial;
- Doses are generally administered to subjects or cohorts;
- Aim is to (choose the next dose to precisely) estimate the maximum tolerated dose (MTD);
- Next dose is chosen using all available information.

Model

- For simplicity, consider a single regressor, dose;
- Assume dose affects the mean of the distribution for the $i$th observation, $E[Y_i] = g^{-1}(\eta_i)$, through link function $g(.)$.
- Consider a flexible model for the predictor, $\eta_i$, of the form

$$
\eta_i = \theta_0 + \theta_1 \frac{D_i^\lambda - 1}{\lambda},
$$

where $D_i$ is the dose assigned to the $i$th subject, $\theta = (\theta_0, \theta_1, \lambda)$ represents the parameters of the model and we restrict $\lambda \in (0, 1]$.

- ‘Power predictor model’ includes special cases (1) $\lambda = 1$, $\eta_i$ is affected linearly with $D_i$ (‘linear predictor model’), and (2) As $\lambda \to \infty$, $\eta_i$ is affected linearly with $\log D_i$ (‘log predictor model’).
Target dose

- Target dose, $D^*$: The dose which produces a probability equal to $p^*$ of experiencing an adverse event ($P(Y = 1|D^*) = p^*$).
- We refer to this as the MTD.
- Given $\eta^* = g(p^*)$, and true values of the parameters $\theta = (\theta_0^T, \theta_1^T, \lambda^T)$, the target dose for the linear, log and power predictor model can be calculated as:

$$D^* = \eta^* - \theta_0^T \theta_1^T + 1,$$

$$D^* = \exp \left( \eta^* - \theta_0^T \theta_1^T \right),$$

$$D^* = \left( \lambda^T \eta^* - \theta_0^T \theta_1^T + 1 \right)^{1/\lambda^T}.$$
Priors

- For linear and log predictor models, we place independent and uninformative normal priors on the parameters

\[ \pi(\theta_0, \theta_1) = N(\theta_0; 0, 100)N(\theta_1; 0, 100). \]

- This reflects our lack of knowledge of the parameters apriori.

- A slightly more restrictive prior is placed on the power predictor model to avoid imaginary components

\[ \pi(\theta_0, \theta_1, \lambda) = N(\theta_0; 0, 100)N(\theta_1; 0, 100)U(0, 1)1(\lambda \frac{\eta^* - \theta_0}{\theta_1} + 1 > 0), \]

where \(1(.)\) denotes the indicator function.
Induced priors

Figure: Induced priors.
Steps in sequential design for dose finding studies

1. Update current information;
   - SMC
     - Re-weight step;
     - Re-sample step;
     - Propagation step (when needed).

2. Dose selection: Find the dose that optimizes some utility function;
   - Variety of utility functions available;
   - Evaluate/optimize within the SMC framework.
Re-weight step in sequential Monte Carlo

Re-weight via importance sampling.

- Target distribution $p(\theta)$ difficult to sample from directly;
- Assume there is another distribution $h(\theta)$ (importance distribution) which is straightforward to sample from;
- Draw $N$ samples from $h(.)$, $\theta^1, \ldots, \theta^N$, and weight these samples to reflect $p(\theta)$

$$w^k = \frac{p(\theta^k)}{h(\theta^k)} \text{, } W^k = \frac{w^k}{\sum_{j=1}^N w^j}.$$ 

- Weighted sample of $\theta$ can be used to estimate, for example, expectations such as $E_{p}(g(\theta)) \approx \sum_{k=1}^N W^k g(\theta^k)$.
- The efficiency of a weighted sample can be measured via the effective sample size ($\approx 1/\sum_{k=1}^N (W^k)^2$).
Re-weight step in sequential Monte Carlo

In the dose finding application, our sequence of target distributions is:

\[
\pi_t(\theta|y_{1:t}, D_{1:t}) = \frac{f(y_{1:t}|\theta, D_{1:t})\pi(\theta)}{Z_t}, \text{ for } t = 1, \ldots, T,
\]

where \( T \) the number of subjects, \( y_{1:t} \) is the data up to subject \( t \) and \( D_{1:t} \) are the corresponding doses.

\( Z_t \) is the normalizing constant, and can usually be ignored for parameter estimation.

The initial importance distribution for \( \pi_1 \) is given by prior \( \pi(\theta) \).
Re-sample step in sequential Monte Carlo

- As we move through the sequence of target distributions, the importance weights become more variable and skewed;
- To prevent this, a re-sampling step is performed when $ESS$ becomes undesirably ‘small’;
- This will duplicate promising particles and discard those with negligible weight;
- Re-sampling techniques include multinomial, residual and systematic resampling.

Once a resample step has been performed, some particles may be duplicated many times;

The particle set can be diversified via a mutation step;

Particles are propagated via an MCMC kernel that is stationary for $\pi_t$. 

---

Utility functions

- Importance sampling is also used in dose selection to obtain a weighted sample for the target distribution involving:
  - All current data;
  - Proposed dose $d \in D_a$;
  - Possible value of the response, $z \in \{0, 1\}$.
- A weighted sample must therefore be obtained for every combination of $(d, z)$.
- Denote the utility for the dose $d$ resulting in observation $z$ as $U(d, z)$.

---

Utility functions

- Utility of proposed dose, $U(d)$, can be obtained by taking the expectation of $U(d, z)$ wrt the distribution of $z$ given $d$.
- The posterior predictive distribution of $z$ is used to find $U(d)$, the expectation of $U(d, z)$.

$$f(z|y_{1:t}, d) = \int_{\theta} f(z|\theta, d)\pi(\theta|y_{1:t}, D_{1:t})d\theta.$$  

Then, $U(d) = \sum_{z \in \{0,1\}} f(z|y_{1:t}, d)U(d, z)$.

- $f(z|y_{1:t}, d)$ can be approximated by summing the importance weights, after taking into account $(d, z)$.
- $U(d, z)$ is also approximated using our weighted samples from the posterior.
- Then, the dose for the $t + 1$th subject is $D_{t+1}$ such that

$$D_{t+1} = \arg \max_d \hat{U}(d).$$

James McGree
Bayesian sequential design
August 29, 2011
15/21
Utility functions

- A variety of utility functions are available. Examples include:
  - Kullback-Leibler divergence utility;
    - Next dose selected maximizes the information gain on $\theta$;
    - Evaluating utility involves estimating the normalizing constant;
    - One advantage of SMC - straightforward to estimate normalizing constant.
  - Posterior variance-covariance:
    $$U(d, z) = -\det[\text{Var}(\theta|y_{1:t}, z, D_{1:t}, d)];$$
  - Posterior variance: $U(d, z) = -\text{Var}(D^*|y_{1:t}, z, D_{1:t}, d);$  
  - Posterior IQR: $U(d, z) = -\text{IQR}(D^*|y_{1:t}, z, D_{1:t}, d);$  
  - Hybrid utility: Mixture of posterior variance-covariance and posterior variance.

---

Example

Setup:

- Number of particles, $N = 1000$;
- Threshold for $ESS = 0.75$;
- Resampling technique: Systematic resampling;
- Total number of subjects: 100;
- $D_a = \{0.1, 0.2, \ldots, 1.0\}$
- Utilities: Posterior variance-covariance, posterior variance and hybrid;
- Simulated 500 dose finding trials.
Estimation of the MTD

- Suppose the linear predictor model is true;
- Parameter configuration $\theta^T = (-0.25, 5, 1)$;
- $p^* = 0.2$ such that $D^* = 0.773$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Utility</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>IQR</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Post. var-cov</td>
<td>0.79</td>
<td>0.78</td>
<td>0.08</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Post. var</td>
<td>0.78</td>
<td>0.78</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.78</td>
<td>0.78</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Log</td>
<td>Post. var-cov</td>
<td>0.79</td>
<td>0.79</td>
<td>0.07</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Post. var</td>
<td>0.76</td>
<td>0.76</td>
<td>0.07</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.76</td>
<td>0.77</td>
<td>0.06</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>Power</td>
<td>Post. var-cov</td>
<td>0.79</td>
<td>0.79</td>
<td>0.07</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Post. var</td>
<td>0.77</td>
<td>0.77</td>
<td>0.06</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.77</td>
<td>0.78</td>
<td>0.06</td>
<td>0.08</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Estimation of the MTD

- Suppose the log predictor model is true;
- Parameter configuration $\theta^T = (1, 3, 0)$;
- $p^* = 0.02$ such that $D^* = 0.196$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Utility</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>IQR</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Post. var-cov</td>
<td>-0.11</td>
<td>-0.10</td>
<td>0.16</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Post. var</td>
<td>0.00</td>
<td>0.01</td>
<td>0.11</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.00</td>
<td>0.01</td>
<td>0.12</td>
<td>0.15</td>
<td>0.23</td>
</tr>
<tr>
<td>Log</td>
<td>Post. var-cov</td>
<td>0.22</td>
<td>0.22</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Post. var</td>
<td>0.21</td>
<td>0.20</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.21</td>
<td>0.21</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Power</td>
<td>Post. var-cov</td>
<td>0.15</td>
<td>0.14</td>
<td>0.05</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Post. var</td>
<td>0.19</td>
<td>0.18</td>
<td>0.12</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.18</td>
<td>0.18</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Discussion

- Presented SMC methodology for sequential design in a discrete data setting;
- Considered a flexible parametric model robust against model uncertainty;
- Explored current and proposed ‘fully-Bayesian’ utility functions - Overall hybrid gave the best results;
- The combination of the power predictor model and the hybrid utility produced robust estimates of the MTD;
- Other issues: No safety considerations or stopping rule - approaches available in the literature, and easily included.
- Future work: How to handle model uncertainty in a more rigorous manner.
References


