Particle filter approach to Bayesian sequential design

J. M. McGree[1]
with C. C. Drovandi and A. N. Pettitt

[1]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.

Drovandi, C.C., McGree, J.M. and Pettitt, A.N. (2011), and Gramacy, R.B. and Polson, N.G. (2011).

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.

Chang, H.H. and Ying, Z. (2009), Dror, H.A. and Steinberg, D.M. (2008), Loeppky, J.L., Moore, L.M. and Williams, B.J. (2010), Müller, P. Berry, D.A., Grieve, A.P., Smith, M. and Krams, M. (2007), Pavel, H. and Miroslav, S. (2010) and Rosenberger, W.F. and Haines, L.M. (2002).

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.

Babb, J., Rogatko, A., Zacks, S. (1998), O'Quigley, J., Pepe, M. and Fisher, L. (1990), Whitehead, J. and Brunier, H. (1995) and Zhou, Y., Whitehead, J., Korhonen, P. and Mustonen, M. (2008).

Model

- For simplicity, consider a single regressor, dose;
- Assume dose affects the mean of the distribution for the ith observation, $E[Y_i] = g^{-1}(\eta_i)$, through link function $g(.)$.
- Consider a flexible model for the predictor, η_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 ith subject, $\boldsymbol{\theta} = (\theta_0, \theta_1, \lambda)$ represents the parameters of the model and we restrict $\lambda \in (0, 1]$.

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

Houede, N., Thall, P.F., Nguyen, H., Paoletti, X. and Kramar, A. (2010).

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 $\boldsymbol{\theta} = (\theta_0^T, \theta_1^T, \lambda^T)$, the target dose for the linear, log and power predictor model can be calculated as:

$$
\begin{aligned}
D^* &= \frac{\eta^* - \theta_0^T}{\theta_1^T} + 1, \\
D^* &= \exp\left(\frac{\eta^* - \theta_0^T}{\theta_1^T}\right) \\
D^* &= \left(\lambda^T \frac{\eta^* - \theta_0^T}{\theta_1^T} + 1\right)^{1/\lambda^T}.
\end{aligned}
$$

Priors

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

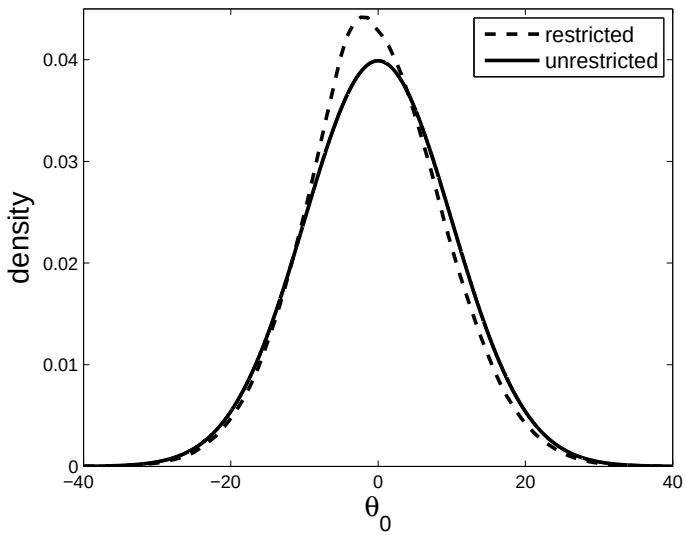
$$\pi(\theta_0, \theta_1) = N(\theta_0; 0, 100)N(\theta_1; 0, 100).$$

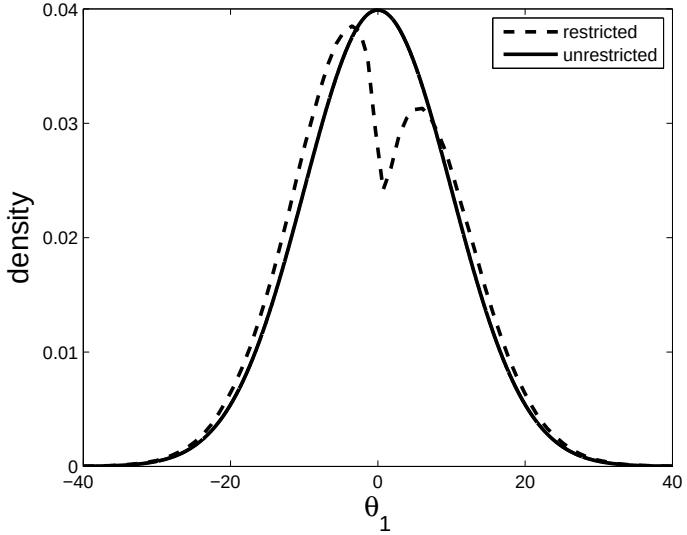
- This reflects our lack of knowledge of the parameters apriori.
- A slightly more restrictive prior is placed on the power predictor model to avoid imaginary components

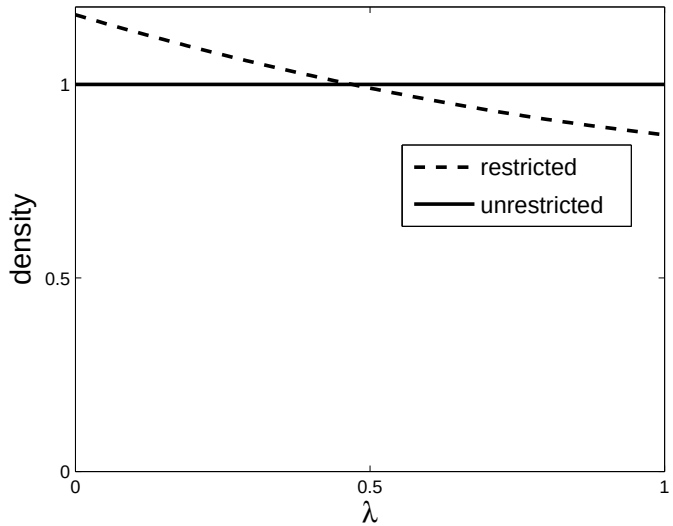
$$\pi(\theta_0, \theta_1, \lambda) = N(\theta_0; 0, 100)N(\theta_1; 0, 100)U(0, 1)1(\lambda\frac{\eta^* - \theta_0}{\theta_1} + 1 > 0),$$

where $1(.)$ denotes the indicator function.

Induced priors







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

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

- Weighted sample of $\boldsymbol{\theta}$ can be used to estimate, for example, expectations such as $E_p(g(\boldsymbol{\theta})) \approx \sum_{k=1}^N W^k g(\boldsymbol{\theta}^k)$.
- 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:

$$\pi_t(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{D}_{1:t}) = \frac{f(\mathbf{y}_{1:t}|\boldsymbol{\theta}, \mathbf{D}_{1:t})\pi(\boldsymbol{\theta})}{Z_t}, \text{ for } t = 1, \ldots, T,$$

  where T the number of subjects, $\mathbf{y}_{1:t}$ is the data up to subject t and $\mathbf{D}_{1:t}$ are the corresponding doses.
- Z_t is the normalizing constant, and can usually be ignored for parameter estimation.
- The initial importance distribution for π_1 is given by prior $\pi(\boldsymbol{\theta})$.

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.

Kitagawa, G. (1962) and Liu, J.S., Chen, R. and Wong, W.H. (1998).

Mutation step in sequential Monte Carlo

- 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 π_t.

Amzal, B., Bois, F.Y., Parent, E. and Robert, C.P. (2006), Chopin, N. (2002) and Del Moral, P., Doucet, A. and Jasra, A. (2006).

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 $d \in \mathbf{D_a}$;
  - 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 $U(d, z)$.

McGree, J.M., Drovandi, C.C., Thompson, M.H., Eccleston, J.A., Duffull, S.B., Mengersen, K., Pettitt, A.N. and Goggion, T. (2011).

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 $U(d, z)$.

$$f(z|\mathbf{y}_{1:t}, d) = \int_{\boldsymbol{\theta}} f(z|\boldsymbol{\theta}, d)\pi(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{D}_{1:t})d\boldsymbol{\theta}.$$

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

- $f(z|\mathbf{y}_{1:t}, d)$ 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

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

Utility functions

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

Atkinson, A. C. and Donev, A. N. (1992), Chaloner, K. and Verdinelli, I. (1995), Haines, L., Perevozskaya, I. and Rosenberger, W. (2003) and Kullback, S. and Leibler, R. A. (1951).

Example

Setup:

- Number of particles, $N = 1000$;
- Threshold for $ESS = 0.75$;
- Resampling technique: Systematic resampling;
- Total number of subjects: 100;
- $\mathbf{D_a} = \{0.1, 0.2, \ldots, 1.0\}$
- 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 $\boldsymbol{\theta}^T = (-0.25, 5, 1)$;
- $p^* = 0.2$ such that $D^* = 0.773$.

Model	Utility	Mean	Median	Std	IQR	RMSE
Linear	Post. var-cov	0.79	0.78	0.08	0.10	0.08
	Post. var	0.78	0.78	0.05	0.07	0.05
	Hybrid	0.78	0.78	0.05	0.07	0.05
Log	Post. var-cov	0.79	0.79	0.07	0.09	0.07
	Post. var	0.76	0.76	0.07	0.09	0.07
	Hybrid	0.76	0.77	0.06	0.08	0.06
Power	Post. var-cov	0.79	0.79	0.07	0.10	0.07
	Post. var	0.77	0.77	0.06	0.08	0.06
	Hybrid	0.77	0.78	0.06	0.08	0.06

Estimation of the MTD

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

Model	Utility	Mean	Median	Std	IQR	RMSE
Linear	Post. var-cov	-0.11	-0.10	0.16	0.21	0.35
	Post. var	0.00	0.01	0.11	0.15	0.22
	Hybrid	0.00	0.01	0.12	0.15	0.23
Log	Post. var-cov	0.22	0.22	0.06	0.07	0.06
	Post. var	0.21	0.20	0.04	0.06	0.04
	Hybrid	0.21	0.21	0.04	0.06	0.04
Power	Post. var-cov	0.15	0.14	0.05	0.08	0.07
	Post. var	0.19	0.18	0.12	0.07	0.12
	Hybrid	0.18	0.18	0.05	0.07	0.05

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

Amzal, B., Bois, F.Y., Parent, E. and Robert, C.P. (2006). *Journal of the American Statistical Assocation*. **101**, 773-785.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*.

Babb, J., Rogatko, A., Zacks, S. (1998). *Statistics in Medicine*. **17**, 1103-1400.

Chaloner, K. and Verdinelli, I. (1995). *Statistical Science*. **10**, 273-304.

Chang, H.H. and Ying, Z. (2009). *The Annals of Statistics*. **37**, 1466-1488.

Chopin, N. (2002). *Biometrika*. **89**, 539-551.

Del Moral, P., Doucet, A. and Jasra, A. (2006). *Journal of the Royal Statistical Society: Series B*. **68**, 411-436.

Dror, H.A. and Steinberg, D.M. (2008). *Journal of the American Statistical Association*. **103**, 288-298.

Drovandi, C.C., McGree, J.M. and Pettitt, A.N. (2011). *Statistics in Medicine*. Submitted for publication. Available at: `http://eprints.qut.edu.au/`.

Gramacy, R.B. and Polson, N.G. (2011). *Journal of Computational and Graphical Statistics*. **20**, 102-118.

Haines, L., Perevozskaya, I. and Rosenberger, W. (2003). *Biometrics*. **59**, 591-600.

Houede, N., Thall, P.F., Nguyen, H., Paoletti, X. and Kramar, A. (2010). *Biometrics*. **66**, 532-540.

Kitagawa, G. (1962). *Journal of Computational and Graphical Statistics*. **5**, 1-25.

Kullback, S. and Leibler, R. A. (1951). *The Annals of Mathematical Statistics*. **22**, 79-86.

Liu, J.S., Chen, R. and Wong, W.H. (1998). *Journal of the American Statistical Association*. **93**, 1022-1031.

Loeppky, J.L., Moore, L.M. and Williams, B.J. (2010). *Journal of Statistical Planning and Inference*. **140**, 1452-1464.

McGree, J.M., Drovandi, C.C., Thompson, M.H., Eccleston, J.A., Duffull, S.B., Mengersen, K., Pettitt, A.N. and Goggion, T. (2011). *Journal of Statistical Planning and Inference*. Submitted for publication. Available at: `http:eprints.qut.edu.au/43760/2011`.

Müller, P. Berry, D.A., Grieve, A.P., Smith, M. and Krams, M. (2007). *Journal of Statistical Planning and Inference*. **137**, 3140-3150.

O'Quigley, J., Pepe, M. and Fisher, L. (1990). *Biometrics*. **46**, 33-48.

Pavel, H. and Miroslav, S. (2010). *Neurocomputing*. **73**, 3284-3290.

Rosenberger, W.F. and Haines, L.M. (2002), *Statistics in Medicine*. **21**, 2757-2770.

Whitehead, J. and Brunier, H. (1995). *Statistics in Medicine*. **14**, 885-893.

Zhou, Y., Whitehead, J., Korhonen, P. and Mustonen, M. (2008). *Biometrics*. **64**, 299-308.