Bayesian sequential design in pharmacokinetics

James McGree
Queensland University of Technology
james.mcgree@qut.edu.au

Collaborators: Chris Drovandi, Gentry White, Tony Pettitt
Acknowledgements: Australian Research Council Discovery Grant

20th June, 2016

Bayesian design problem

- Usually quantify experimental goals via a utility function $u(d)$
- Optimal design can be expressed as

$$d^* = \arg\max_{d \in \mathcal{D}} \int u(d, z) p(z|d) \mathrm{d}z,$$

$$d^* = \arg\max_{d \in \mathcal{D}} \sum_{m=1}^{K} p(m) \int u(d, z, m) p(z|m, d) \mathrm{d}z$$

Why do this to ourselves?

- Why make a difficult problem more difficult?
- What is FO, FOCE, nonlinearity?
- Appropriate to design under the planned estimation framework
- Model and parameter uncertainty are most rigoriously handled within Bayesian framework
- Inference framework more appropriate for complex models
- Wider variety of useful criteria/utility functions, for example, for model choice (mutual information)

Bayesian sequential design

- Adaptive decisions as new data are collected
- More robust to parameter and model uncertainty
- Natural to use Bayesian framework. Posterior becomes new prior
- Next decision obtained by looking forward to all future decisions (backward induction)
- Simplified by myopic design (one-at-a-time)
- Next design point $d_{t+1} = \arg\max U(d|y_{1:t}, d_{1:t})$. $y_{1:t}$ collected data at design $d_{1:t}$. U is utility function

Computational difficulties

- In sequential design, one needs to evaluate $u(d|y_{1:t}, d_{1:t})$

$$\int_z u(d, z|y_{1:t}, d_{1:t}) p(z|d, y_{1:t}, d_{1:t}) \mathrm{d}z$$

$$\sum_{m=1}^{K} p(m|y_{1:t}, d_{1:t}) \int_z u(d, z, m|y_{1:t}, d_{1:t}) p(z|d, m, y_{1:t}, d_{1:t}) \mathrm{d}z$$

- Then, need to find d that maximises $u(d|y_{1:t}, d_{1:t})$
- Hence, need to approximate or sample from a large number of posterior distributions for different priors, designs and data
- How can this be done efficiently?

SMC for one static model m

- Sample from sequence of targets
- Data annealing here

$$p_t(\theta_m|m, y_{1:t}, d_{1:t}) = f(y_{1:t}|m, \theta_m, d_{1:t})p(\theta_m|m)/Z_{m,t}, \text{ for } t = 1, \ldots, T.$$

$y_{1:t}$ (independent) data up to t, $d_{1:t}$ design points up to t, θ_m parameter for model $m = 1, \ldots, K$.

$$p(y_{1:t}|m, d_{1:t}) = Z_{m,t} = \int f(y_{1:t}|m, \theta_m, d_{1:t})p(\theta_m|m)d\theta_m.$$

- SMC: Generate a weighted sample (particles) for each target in the sequence via steps
  - Reweight: particles as data comes in (efficient)
  - Resample: when ESS small
  - Mutation: diversify duplicated particles (can be efficient)

SMC for one static model m (algorithm) Chopin (2002)

- Have current particles $\{W_t^i, \theta_t^i\}_{i=1}^N$ based on data $y_{1:t}$
- Re-weight step to included y_{t+1}

$$W_{t+1}^i \propto W_t^i f(y_{t+1}|\theta_t^i, d_{t+1}).$$

- Check effective sample size: $\text{ESS} = 1/\sum_{i=1}^N (W_{t+1}^i)^2$
- If ESS $> E$ (e.g. $E = N/2$) go back to re-weight step for next observation
- If ESS $< E$ do the following
- Resample proportional to weights. Duplicates good particles
- Mutation: Move all particles via MCMC kernel say R times (adaptive proposal)

SMC for multiple models

- Effectively run an SMC algorithm for each model $m = 1, \ldots, K$
- Have set of N particles for each model $\{W^i_{m,t}, \theta^i_{m,t}\}^N_{i=1}$.
- ESS for each model m
- resampling and within-model updates when required
- Design part: use data up to t, $y_{1:t}$, and particles of all models to compute the next design d_{t+1}

SMC Estimate of Evidence Del Moral et al (2006)

- It can be shown

$$Z_{t+1}/Z_t = f(y_{t+1}|y_{1:t}, d_{t+1}) = \int_\theta f(y_{t+1}|\theta, d_{t+1}) p(\theta|y_{1:t}, d_{1:t}) d\theta.$$

- Using SMC particles to approximate posterior at t gives estimator

$$Z_{t+1}/Z_t \approx \sum_{i=1}^{N} W_t^i f(y_{t+1}|\theta_t^i, d_{t+1}).$$

- Can then obtain approximation of Z_{t+1} through

$$\frac{Z_{t+1}}{Z_0} = \frac{Z_{t+1}}{Z_t}\frac{Z_t}{Z_{t-1}} \cdots \frac{Z_1}{Z_0}.$$

- Also gives estimate of posterior predictive probability of y_{t+1}

But what about random effects models?

- SMC requires the likelihood to be computed a large number of times
- However, computing the likelihood can be difficult for random effect models as, for example

$$f(y|\theta, d) = \int f(y|\theta, \beta, d)p(\beta|\mu, \Omega)\mathrm{d}\beta$$

- If model is nonlinear then generally analytically intractable
- Can be approximated
- Needs to be computationally efficient and unbiased
- SMC for random effects models?
- Efficient approximates of model evidence and predictive probabilities of random effect models....

Exact-Approximate SMC

- The (observed data) likelihood

$$f(y|\theta^{(i)}, d) = \int f(y|\beta, \theta^{(i)}, d)p(\beta|\mu^{(i)}, \Omega^{(i)})d\beta$$

- Can be estimated unbiasedly. For example, from McGree et. al (2015), for each particle $\theta^{(i)}$

$$f(y|\theta^{(i)}, d) = \frac{1}{Q}\sum_{j=1}^{Q} f(y|\beta^{(j)}, \theta^{(i)}, d) \quad (1)$$

where $\beta^{(j)} \sim p(\mu^{(i)}, \Omega^{(i)}), \;\; j = 1, \ldots, Q.$

- SMC with unbiased estimate of likelihood $\rightarrow$ an exact-approximate algorithm! (Duan and Fulop 2013)
- Andrieu and Roberts (2009) for MCMC and Tran et al. (2014) for importance sampling.

Bayesian A-optimality

For a single model, this can be achieved by maximising the following:

$$u(d|y_{1:t}, d_{1:t}) = 1/\text{trace VAR}[\theta|d, y_{1:t}, d_{1:t}].$$

This is extended to the case of K models by maximising the inverse of the sum of the traces of the posterior variances for all K models. That is,

$$u(d|y_{1:t}, d_{1:t}) = 1/\sum_{l=1}^{K} \log \text{trace VAR}[\theta_l|d, y_{1:t}, d_{1:t}, M = l].$$

Other utilities are also available for parameter estimation (KLD, Bayesian D-optimality, etc).

Utility estimation in sequential design

Expected utility of d is given by $u(d|y_{1:t}, d_{1:t}) =$

$$\sum_{m=1}^{K} p(m|y_{1:t}, d_{1:t}) \int_z u(d, z, m|y_{1:t}, d_{1:t}) p(z|m, d, y_{1:t}, d_{1:t}) \mathrm{d}z$$

For each $\theta^i_{m,t}$, simulate $z^i_{m,t}$. Then, MC integration yields:

$$u(d|y_{1:t}, d_{1:t}) \approx \sum_{m=1}^{K} p(m|y_{1:t}, d_{1:t}) \sum_{i=1}^{N} W^i_{m,t} u(d, z^i_{m,t}, m|y_{1:t}, d_{1:t}).$$

The $u(d, z^i_{m,t}, m|y_{1:t}, d_{1:t})$ is evaluated via importance sampling where $z^i_{m,t}$ (and d) are supposed observed data.

Application - Pharmacokinetics

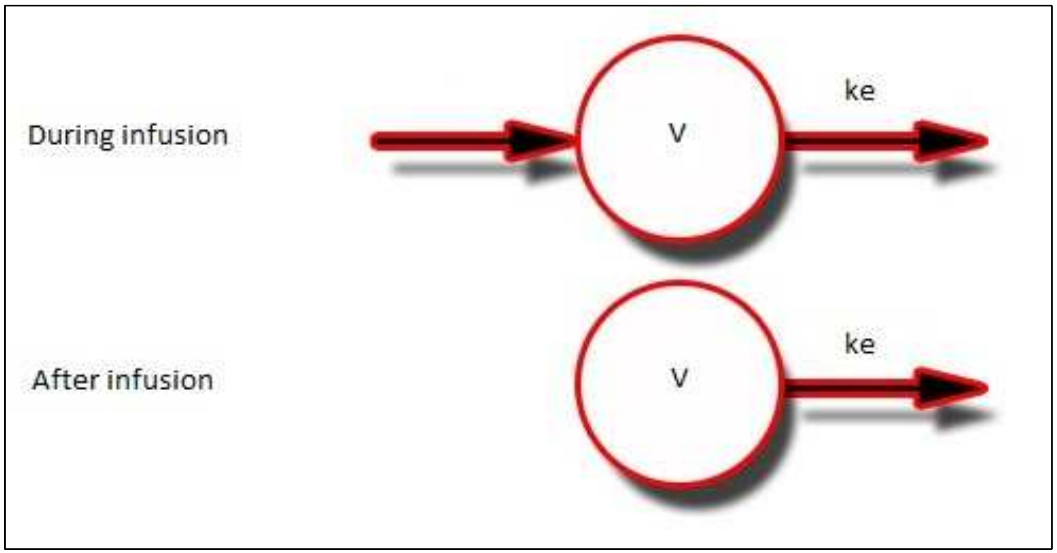

Figure: One compartment infusion model

One compartment infusion - Pharmacokinetics

- For subject t with design $d_t = (d_{1t}, d_{2t})$, define

$$y_t \sim MVN(g(\beta_t, d_t), \delta I),$$
$$\beta_t \sim MVN(\mu, \Omega),$$

- Here β_t is random effect for tth subject

$$g(\beta_t, d_t) = \begin{cases} \frac{D}{Tinf}\frac{1}{k_t v_t}(1 - \exp(-k d_t)), & \text{for } d_t \leq Tinf \\ \frac{D}{Tinf}\frac{1}{k_t v_t}(1 - \exp(-k_t Tinf))\exp(-k(t - Tinf)), & \text{else} \end{cases}$$

where $(k_t, v_t) = \exp(\beta_t + \mu)$

Priors: $\mu \sim MVN(0, \Sigma)$, for Σ known.

$\Omega \sim InvWish(\Psi, \nu)$, for Ψ and ν known

$\log \delta \sim N(a, b)$, for a and b known,

- Design objective is to learn about parameters: $\theta = (\mu, \Omega, \delta)$.

One compartment infusion - Pharmacokinetics

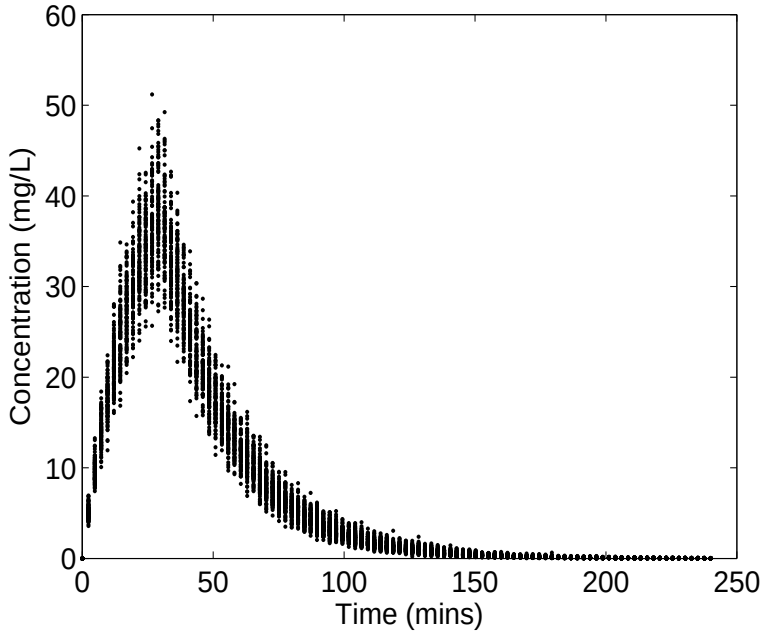

Figure: Prior predictive plot for one-compartment infusion model

One compartment infusion - Pharmacokinetics

- Computationally expensive to implement search algorithm.
- Consider discretised design space (mins since start of infusion of length $Tinf = 30$ mins):

$$\begin{bmatrix} 6 & 15 \\ 15 & 30 \\ 30 & 45 \\ 45 & 60 \\ 60 & 120 \\ 120 & 180 \\ 180 & 240 \\ 240 & 300 \\ 300 & 360 \end{bmatrix}$$

- Design is found via Bayesian A-optimality and random design.

One compartment infusion - Pharmacokinetics

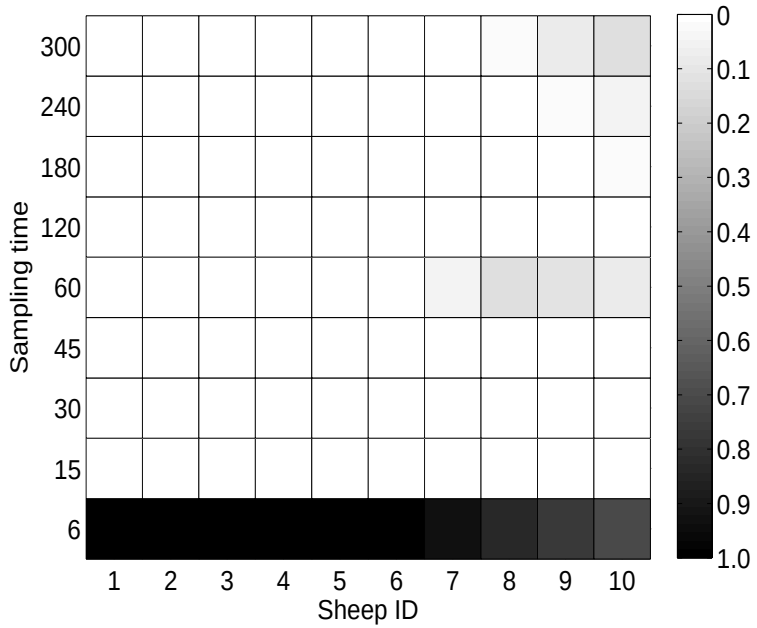

Figure: Selected A-optimal designs for one-compartment infusion model in simulation study.

One compartment infusion - Pharmacokinetics

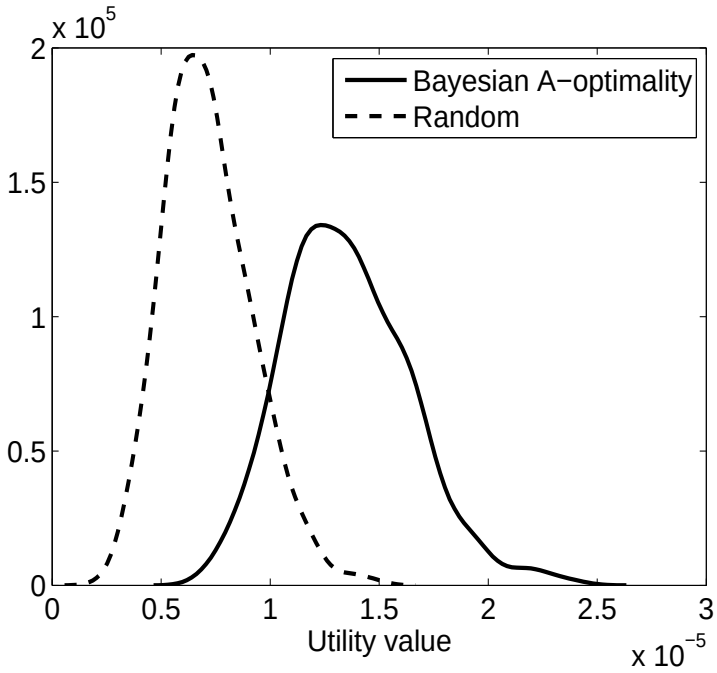

Figure: Utility values for the 500 simulated trials for the A-optimality and random utility.

Discussion

- Developed a framework to efficiently undertake Bayesian design in PK settings
- Framework is highly computational - GPU made this work possible in a reasonable amount of time
- Also considered design for 1cpt and 2cpt models (not shown here)
- Framework should be useful in general sequential setting (GLMMs)?

Related Bayesian design work

MCMC framework

- The so called 'Mueller algorithm' (Mueller, 1999) with extensions (Amzal et al., 2006)
- GLMs (Weir, et al., 2007 and McGree et al., 2012)
- Accelerated life test (Weaver et al., 2016)

SMC framework

- Estimation for GLMs (Drovandi, McGree and Pettitt, 2013, Azadi et al., 2014)
- Model discrimination for GLMs and GNLMs (Drovandi, McGree and Pettitt, 2014)
- Model discrimination and estimation for GLMs and GNLMs (McGree, 2016)

ABC framework

- Intractable likelihoods (Drovandi and Pettitt, 2014, Price et al., 2016)

Future Bayesian design

Further extensions to mixed effects settings

- Model discrimination?
- Dual purpose designs - model discrimination and estimation?

Static designs (high dimensional problems)

- Need fast search algorithms - ACE (Overstall and Woods, 2015)?
- Need fast posterior approximations - Expectation propagation (Minka, 2005), Variational approximations (Nott et al. , 2013)?

Key References

- Andrieu and Roberts (2009). *Annals of Statistics*, **37**, 697-725.
- Chopin (2002). *Biometrika*, **89**:539-551.
- Drovandi, McGree, and Pettitt (2014). *Journal of Computational and Graphical Statistics*, **23**:3-24.
- Drovandi, McGree, and Pettitt (2013). *Computational Statistics & Data Aanalysis*, **57**:320-335.
- Del Moral, Doucet and Jasra (2006). *Journal of the Royal Statistics Society: Series B*, **68**:411-436.
- McGree et al. (2012). *Journal of Statistical Planning and Inference*, **142**, 1480-1492.
- McGree, White, Drovandi, and Pettitt (2015). *Statistics and Computing*. To appear.
- McGree (2016). *Computational Statistics & Data Aanalysis*. Accepted for publication.
- Nott et al . (2013). ArXiv:1307.7962 [stat.ME]
- Overstall and Woods (2015) ArXiv:1501.00264 [stat.ME]
- Tran et al. (2014). ArXiv:1402.6035 [stat.ME]