Bayesian design for models with intractable likelihoods: A synthetic likelihood approach

James McGree
Associate Professor of Statistics
School of Mathematical Sciences
Queensland University of Technology

james.mcgree@qut.edu.au | www.jamesmcgree.com
@j_mcgree

Joint work with Dehideniya, M. and Drovandi, C.

July 31, 2018

Outline of talk

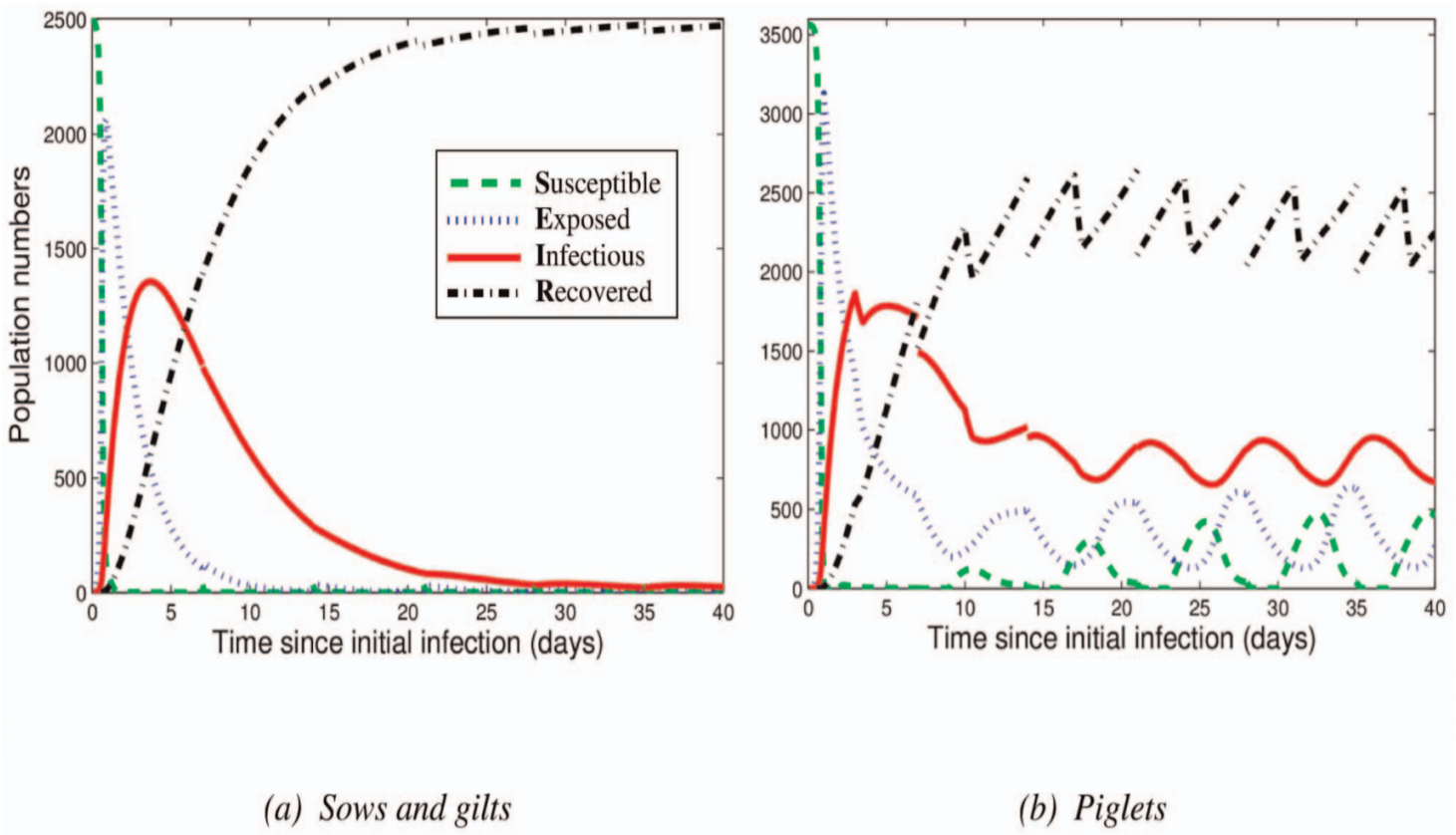
- Types of models/experiments
- Bayesian inference and design (in brief, lit review omitted)
- Utility functions - Total entropy
- Synthetic likelihood
- Approximating utility functions
- Locating Bayesian designs (omitted)
- Illustrative example (in brief)
- Motivating example
- Conclusions and further work

Models from epidemiology

- Foot and mouth disease: Highly contagious disease, high impact on cattle, pig and sheep.
- Outbreaks in the UK in 2001 and 2007, China 2005, and Japan in 2010
- Large impact on animal welfare (culling), and large economic impact through trade restrictions
- Literature shows competing models for the spread of the disease (Orsel et al., 2007:SIR and Backer et al., 2012: SEIR)
- Interest in finding a preferred model, and finding precise estimates of parameters

Models from epidemiology

Virus dynamics within swine farms (Reynolds et al., 2014)



(a) *Sows and gilts* (b) *Piglets*

Design question

- Design question: when to observe the process
- Not practical to continuously observe the process
- Expensive to collect many data points
- Consider Bayesian design due to the availability of important utilities (total entropy), and appropriately handle uncertainty (particularly model uncertainty)
- Assume Bayesian inference will be undertaken upon observing data (and in deriving designs)
- Framework yields particular challenges

Background - Bayesian inference

- When interested in estimating θ

$$p(\theta|y,d) \propto p(\theta)p(y|\theta,d),$$

where $p(\theta)$ is prior and $p(y|\theta,d)$ is the likelihood.

- When interested in model choice, suppose K models are being considered, with $m = 1,\ldots,K$

$$p(\theta_m|y,m,d) = \frac{p(\theta_m|m)p(y|\theta_m,m,d)}{Z_m},$$

where $Z_m = \int_{\theta_m} p(\theta_m|m)p(y|\theta_m,m,d)\mathrm{d}\theta_m$.

- $Z_m \propto p(m|y,d)$, so pick the model with the largest Z_m.

Background

- Two main challenges:
  1. Approximating the expected utility;
  2. Maximising the utility.
- Focus on (1) for models with intractable likelihoods
- Maximise expected utility $d^* = \arg\ \max_d u(d)$, where

$$u(d) = \sum_{m=1}^{K} p(m) \int_y u(d, y, m) p(y|d, m) \mathrm{d}y.$$

- $u(d, y, m)$ is some measure of information gained from d given model m and observed data y.
- Importantly, $u(d, y, m)$ is typically a function of $p(\theta_m|y, m, d)$.

Background

- $u(d)$ typically cannot be solved analytically
- Can be approximated using Monte Carlo integration

$$u(d) \approx \sum_{m=1}^{K} p(m) \frac{1}{B} \sum_{b=1}^{B} u(d, y_{mb}, m),$$

  where $y_{mb} \sim p(y|\theta_{mb}, m, d)$ and $\theta_{mb} \sim p(\theta|m)$.
- Typical Bayesian utility is the KLD between the prior and the posterior.
- Hence, B posterior distributions need to be approximated or sampled from to approximated $u(d)$.
- Computationally challenging task.

Total entropy

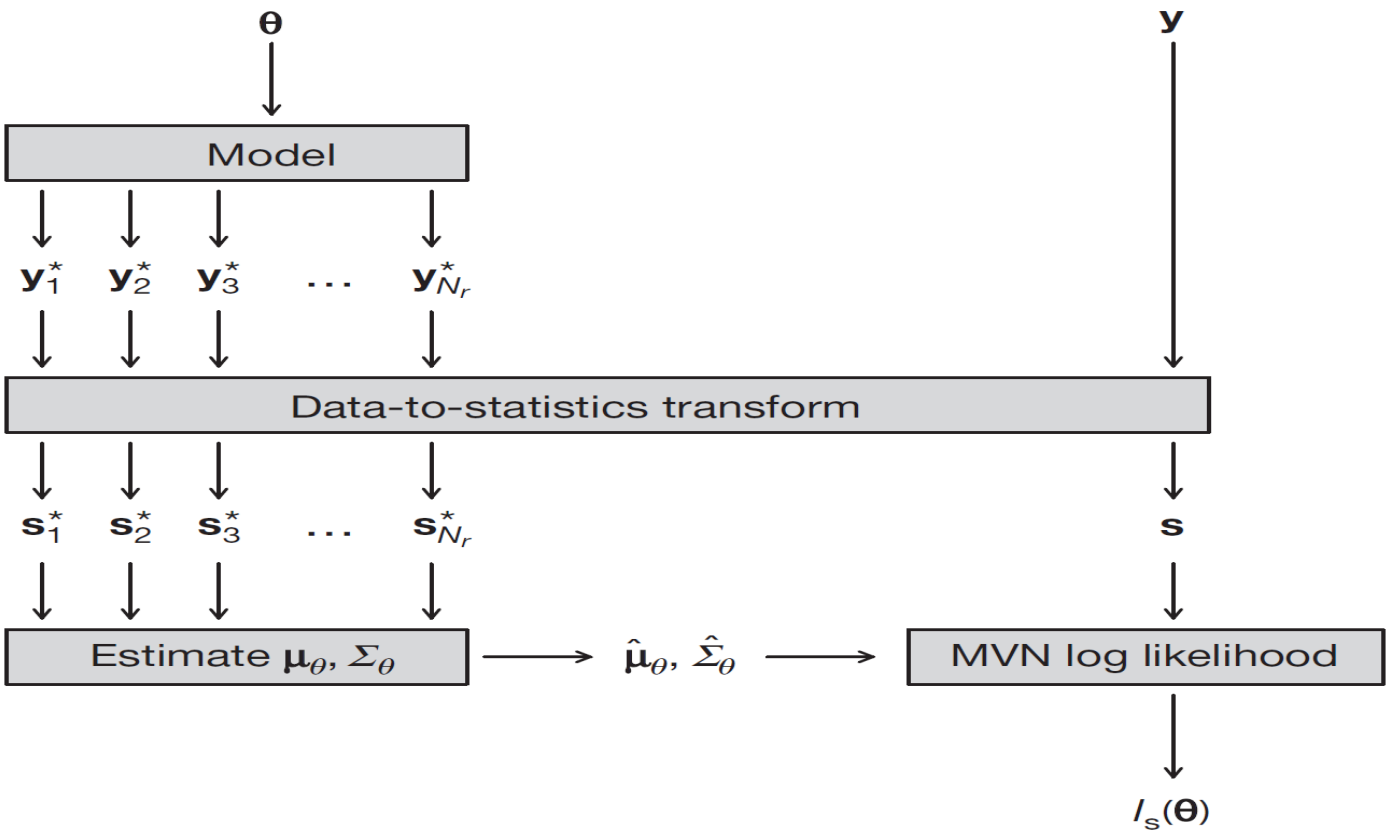
- The total entropy utility function can be defined as follows

$$u_T(d, y, m) = \int_\theta p(\theta|m, y, d) \log p(y|\theta, m, d) \mathrm{d}\theta - \log p(y|d)$$

- Difficult to approximate $\log p(y|d)$
- Computationally difficult to efficiently approximate $p(\theta|m, y, d)$
- Difficult to evaluate $\log p(y|\theta, m, d)$ for models with intractable likelihoods
- Motivates the use of a synthetic likelihood approach....
- In general, total entropy is a computationally challenging utility
- Limited use (Borth, 1975 and McGree, 2017)

Synthetic likelihood approach

- Wood (2010) approach



Synthetic likelihood approach

- Counts will be observed from our experiments
- Extension for discrete data
- In our case, no summary statistics (mean and variance of simulated data)
- Idea: Discretise the Normal distribution via continuity correction
- Likelihood for discrete data is thus:

$$p_{SL}(Y = y|\theta, m, d) = p(y_1 - c < Y_1 < y_1 + c, \ldots, y_p - c < Y_p < y_p + c),$$

where $(Y_1, \ldots, Y_p) \sim N(\hat{\mu}(\theta, m, d), \hat{\Sigma}(\theta, m, d))$, c = 0.5

Approximating utility functions

- Marginal likelihood can be approximated as follows:

$$\hat{p}(y|m,d) = \sum_{b=1}^{B} p_{SL}(y|\theta_b, m, d),$$

  where (eg) $\theta_b \sim p(\theta)$.
- Also for $p(y|d)$

$$\hat{p}(y|d) = \sum_{m=1}^{K} \hat{p}(y|m,d)p(m).$$

- Then posterior model probabilities:

$$\hat{p}(m|y,d) = \frac{\hat{p}(y|m,d)p(m)}{\hat{p}(y|d)}$$

Approximating utility functions

- Employ importance sampling for approximating posterior distributions
- Use prior as importance distribution
- Sample $\theta_b \sim p(\theta)$, $b = 1, \ldots, B$ (equal weights)
- Update weights via synthetic likelihood to yield W_b; the normalised importance weights
- $p(\theta|y, m, d)$ can be approximated by the particle set:

$$\{\theta_b, W_b\}_{b=1}^{B}$$

- Total entropy:

$$\hat{u}_T(d, y_{mb}, m) = \sum_{b=1}^{B} W_m^b \log \hat{p}(y_{mb}|\theta_{mb}, m, d) - \log \hat{p}(y_{mb}|d)$$

Example 1

Death model

- Probability that an infection occurs in the next time period Δ_t at time t with s susceptibles:

$$P(S(t+\Delta_t) = s-1 | S(t) = s) = bs\Delta_t + O(\Delta_t)$$

Susceptibles to infecteds (SI) model

- Probability that an infection occurs in the next time period Δ_t at time t with s susceptibles:

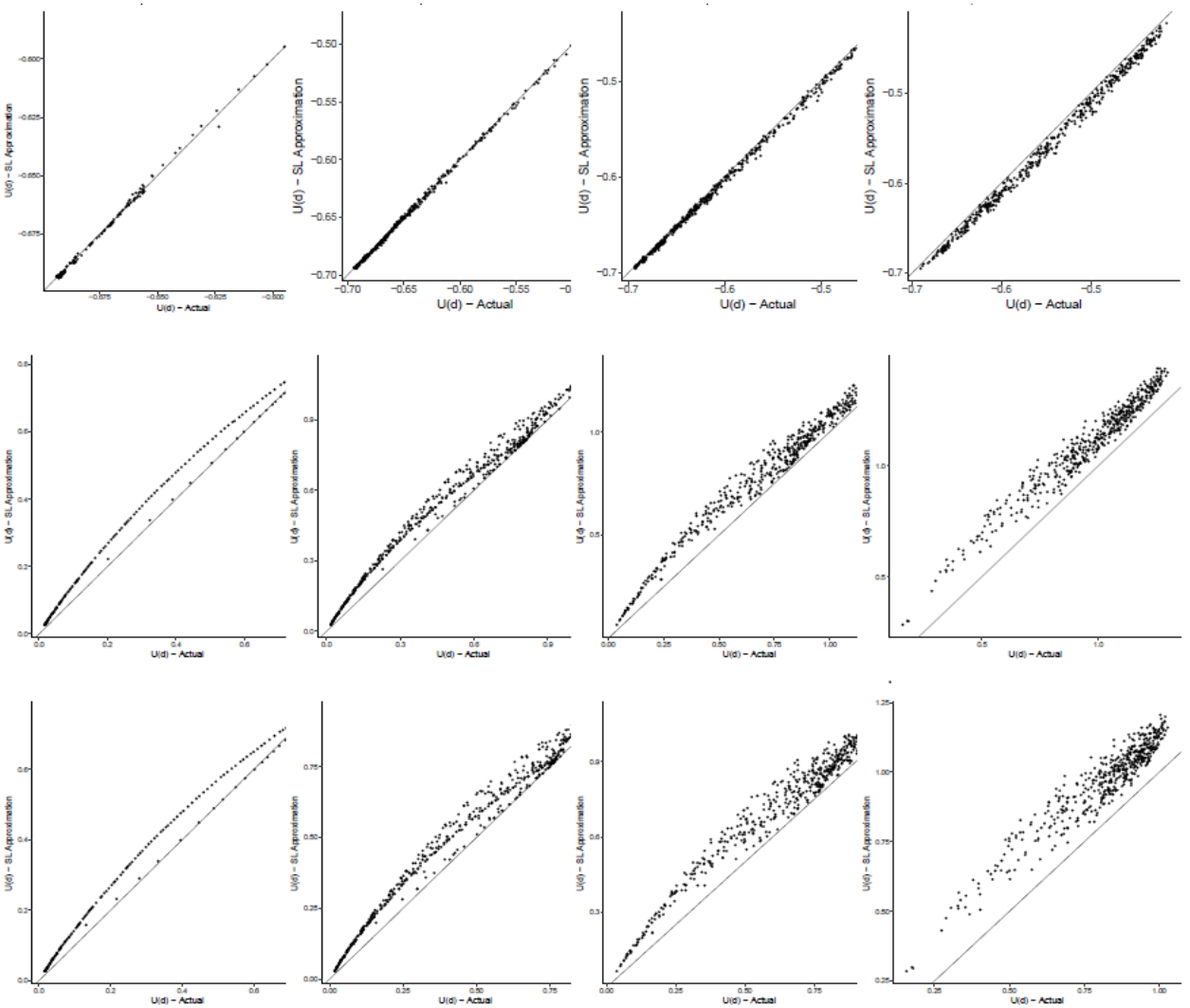
$$P(S(t+\Delta_t) = s-1 | S(t) = s) = (b_1 + b_2(N-s))s\Delta_t + O(\Delta_t)$$

Example 1

MI

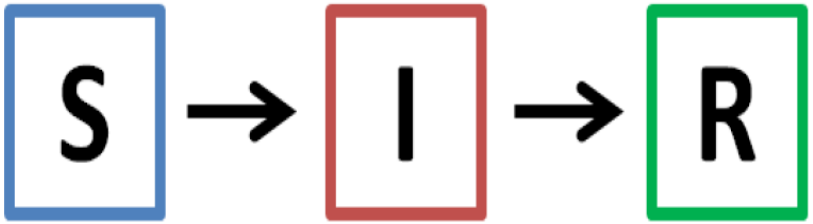
KLD

TOTAL



1 design point | 2 design points | 4 design points | 8 design points

Example 2



SIR model Given that at time t there are s susceptibles and i infectious individuals in a closed population of size N, the probabilities of possible events in the next Δ_t time step are:

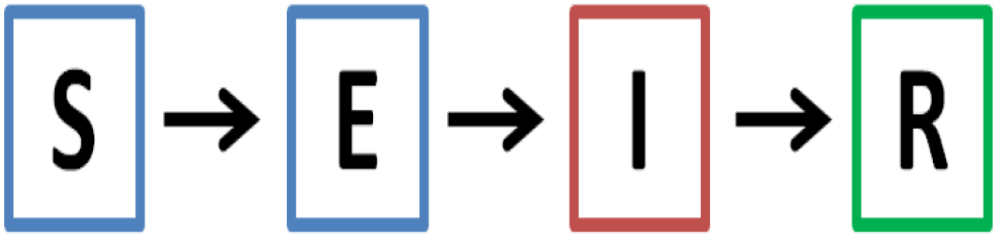
- A susceptible becomes infectious:

$$P(s-1, i+1|s, i) = \frac{\beta si}{N}\Delta_t + O(\Delta_t)$$

- An infectious individual recovers

$$P(s, i-1|s, i) = \alpha i \Delta_t + O(\Delta_t)$$

Example 2



SEIR model As above but now have e exposed individuals, the probabilities of possible events in the next Δ_t time step are:

- A susceptible becomes exposed:

$$P(s-1, e+1, i|s, e, i) = \frac{\beta si}{N}\Delta_t + O(\Delta_t)$$

- An exposed individual becomes infectious

$$P(s, e-1, i+1|s, e, i) = \alpha_I e\Delta_t + O(\Delta_t)$$

- An infectious individual recovers

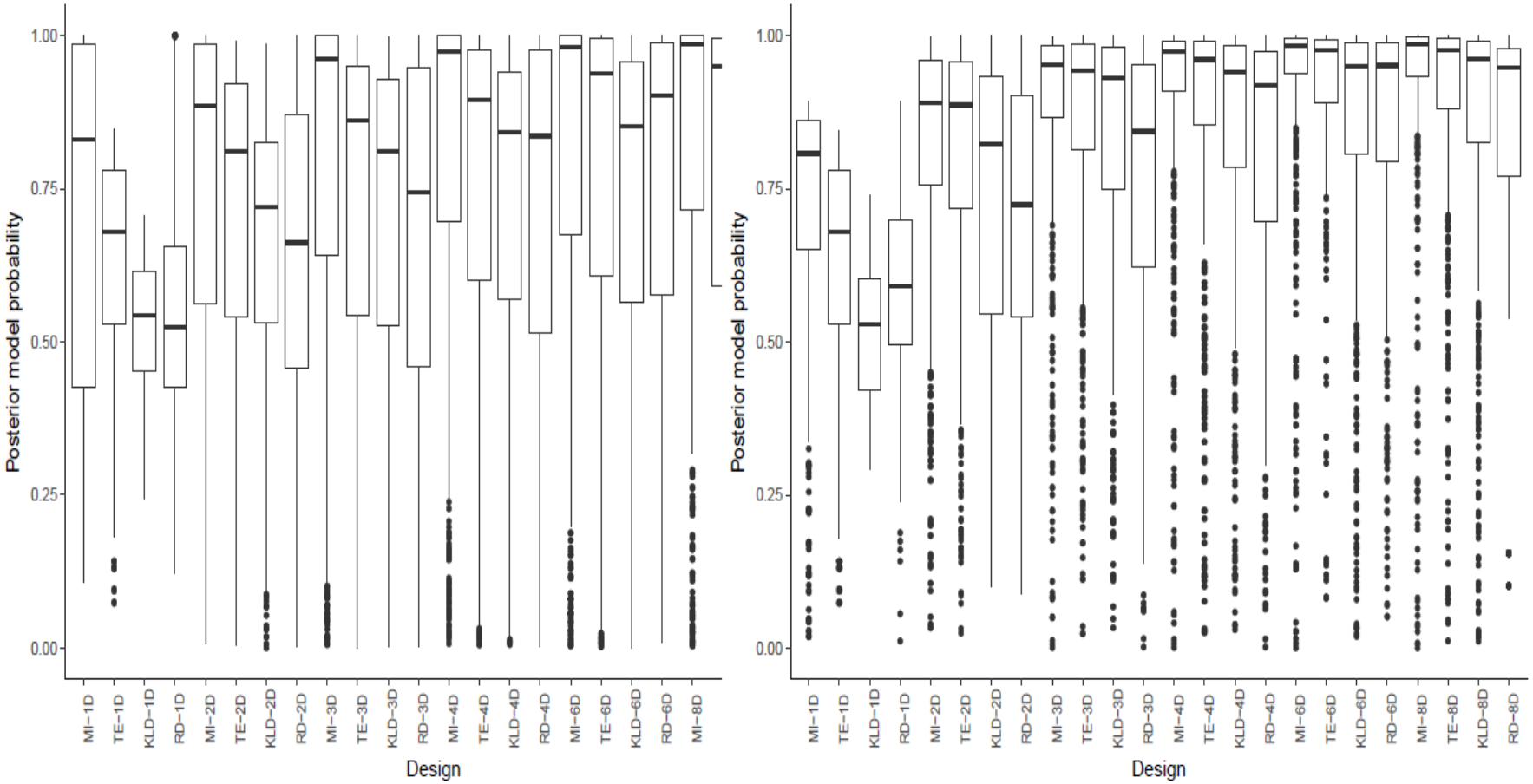
$$P(s, e, i-1|s, e, i) = \alpha_R i\Delta_t + O(\Delta_t)$$

Example 2

Table 2: Optimal designs derived under different utility functions.

Utility function	Optimal design $\mathbf{d}^*$	$U(\mathbf{d}^*)$
Mutual Information	(3.1)	-0.43 (0.02)
	(4.1, 16.0)	-0.34 (0.02)
	(0.7, 4.1, 18.4)	-0.30 (0.02)
	(0.7, 4.1, 10.1, 25.3)	-0.28 (0.02)
	(0.7, 3.1, 5.3, 6.5, 10.1, 25.3)	- 0.27 (0.02)
	(0.7, 2.9, 4.1, 5.3, 6.3, 6.5, 10.1, 25.4)	-0.27 (0.02)
Total Entropy	(7.0)	0.97 (0.02)
	(6.7, 17.5)	1.56 (0.03)
	(6.5, 13.5, 27.1)	1.81 (0.03)
	(5.5, 10.8, 16.3, 27.1)	1.97 (0.03)
	(4.1, 7.0, 10.8, 14.2, 18.8, 27.1)	2.16 (0.03)
	(4.1, 7.0, 10.8, 12.9, 15.2, 17.5, 21.7, 27.3)	2.30 (0.04)
KLD	(11.6)	0.91 (0.02)
	(9.4, 19.1)	1.26 (0.03)
	(7.4, 14.2, 27.1)	1.47 (0.03)
	(7.3, 10.9, 16.4, 27.1)	1.60 (0.03)
	(7.3, 10.7, 14.2, 17.8, 21.7, 28.2)	1.79 (0.03)
	(7.3, 10.7, 12.8, 15.0, 17.4, 21, 23.8, 28.2)	1.94 (0.03)

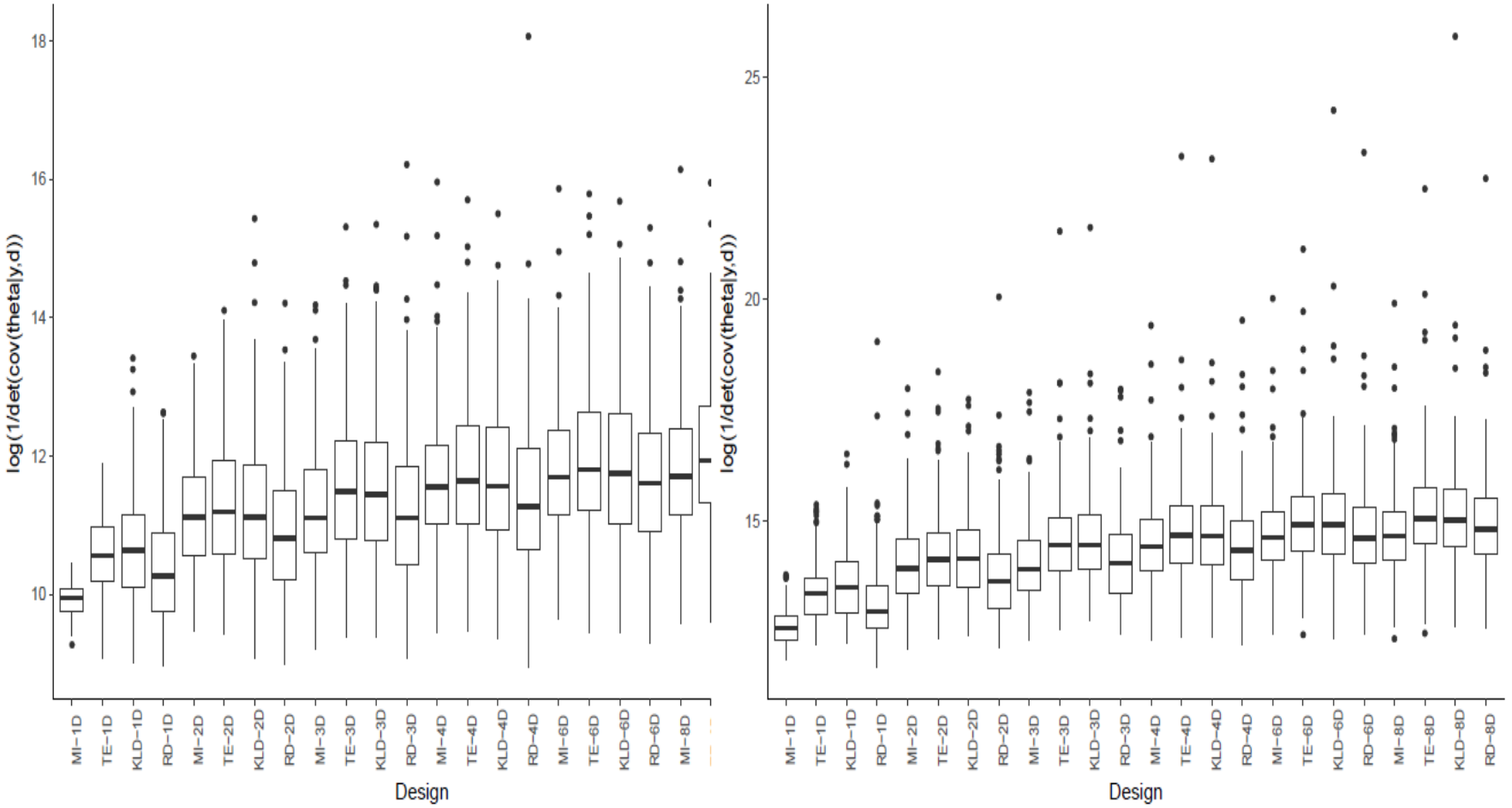
Example 2 - Results for SIR model



(a) SIR model

(b) SEIR model

Example 2 - Results for SIR model



(a) SIR model (b) SEIR model

Future research

- Proposed approach to design experiments for models with intractable likelihoods
- Flexible in that a variety of utility functions can be efficiently estimated
- Is the Normal approximation reasonable, in general? Other distributions were considered....
- How small can the sample size (no. of ind) be? Trialled sample size of 25 with reasonable results.
- Extensions for higher dimensional design problems?
- E.g: Use suitable importance distribution (i.e: not the prior)

Selected references

- Backer, Hagenaars, Nodelijk and Roermund (2012). Preventive Veterinary Medicine, 107, 27-40.
- Borth (1975). JRSS(B), 37, 77-87.
- Box and Hill (1967). Technometrics, 9, 57-71.
- Hill (1978). Technometrics, 20, 15-21.
- McGree (2017). CSDA, 113, 207-225.
- Wood (2010). Nature, 466, 1102-1104