Dual-purpose Bayesian design for parameter estimation and model discrimination of models with intractable likelihoods

M. B. Dehideniya [1] C. C. Drovandi [1,2] J. M. McGree [1,2]

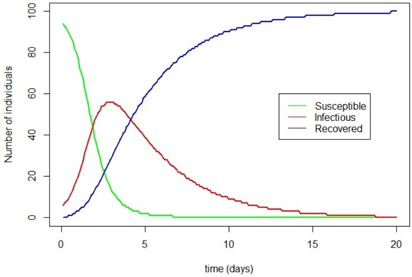
[1]School of Mathematical Sciences
Queensland University of Technology
Australia

[2]ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)

Bayes on the Beach - 2017

Experimental design in Epidemiology

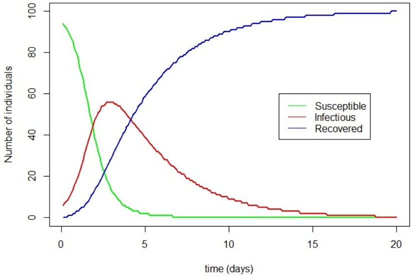
- Spread of a disease within a herd of cows.(eg.Foot and mouth disease)



Source : http://animalia-life.club

Experimental design in Epidemiology

- Spread of a disease within a herd of cows.(eg.Foot and mouth disease)

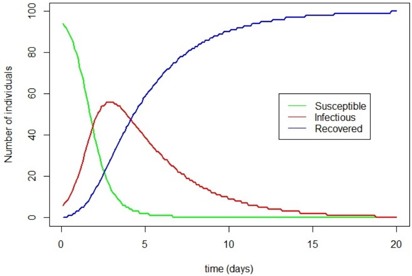


Source : http://animalia-life.club

- Competing models - SIR (Orsel et al., 2007) and SEIR (Backer et al., 2012)

Experimental design in Epidemiology

- Spread of a disease within a herd of cows.(eg.Foot and mouth disease)



Source : http://animalia-life.club

- Competing models - SIR (Orsel et al., 2007) and SEIR (Backer et al., 2012)

- Not practical to continuously observe the process.
- A set of distinct observational times $\{t_1, t_2, ..., t_n\}$ - Design.

Background

Bayesian experimental designs

- Consider Bayesian design,
  - due to the availability of important utilities (total entropy).
  - to appropriately handle uncertainty about models and parameters.

Background

Bayesian experimental designs

- Consider Bayesian design,
  - due to the availability of important utilities (total entropy).
  - to appropriately handle uncertainty about models and parameters.
- Assume Bayesian inference will be undertaken upon observing data.

Background

Bayesian experimental designs

- Consider Bayesian design,
  - due to the availability of important utilities (total entropy).
  - to appropriately handle uncertainty about models and parameters.
- Assume Bayesian inference will be undertaken upon observing data.
- **Optimal design** - $\boldsymbol{d}^* = \arg\ \max_{\boldsymbol{d}} u(\boldsymbol{d})$, where

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

Background

Bayesian experimental designs

- Consider Bayesian design,
  - due to the availability of important utilities (total entropy).
  - to appropriately handle uncertainty about models and parameters.
- Assume Bayesian inference will be undertaken upon observing data.
- **Optimal design** - $\boldsymbol{d}^* = \arg\ \max_{\boldsymbol{d}} u(\boldsymbol{d})$, where

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

- $u(\boldsymbol{d}, \boldsymbol{y}, m)$ is some measure of information gained from $\boldsymbol{d}$ given model m and observed data $\boldsymbol{y}$.

Background

Bayesian experimental designs

- In most cases, $u(\boldsymbol{d})$ cannot be solved analytically.

- In most cases, $u(\boldsymbol{d})$ cannot be solved analytically.
- But it can be approximated, e.g., Monte Carlo integration,

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

where $\boldsymbol{y}_{mb} \sim p(\boldsymbol{y}|\boldsymbol{\theta}_{mb}, m, \mathbf{d})$ and $\boldsymbol{\theta}_{mb} \sim p(\boldsymbol{\theta}|m)$.

Background

Bayesian experimental designs

- In most cases, $u(\boldsymbol{d})$ cannot be solved analytically.
- But it can be approximated, e.g., Monte Carlo integration,

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

  where $\boldsymbol{y}_{mb} \sim p(\boldsymbol{y}|\boldsymbol{\theta}_{mb}, m, \mathbf{d})$ and $\boldsymbol{\theta}_{mb} \sim p(\boldsymbol{\theta}|m)$.
- Here, some posterior inference is required to evaluate $u(\boldsymbol{d}, \boldsymbol{y}_{mb}, m)$.

Background

Bayesian experimental designs

- In most cases, $u(\boldsymbol{d})$ cannot be solved analytically.
- But it can be approximated, e.g., Monte Carlo integration,

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

  where $\boldsymbol{y}_{mb} \sim p(\boldsymbol{y}|\boldsymbol{\theta}_{mb}, m, \mathbf{d})$ and $\boldsymbol{\theta}_{mb} \sim p(\boldsymbol{\theta}|m)$.
- Here, some posterior inference is required to evaluate $u(\boldsymbol{d}, \boldsymbol{y}_{mb}, m)$.
- Hence, $K \times B$ posterior distributions need to be approximated or sampled from to approximated $u(\boldsymbol{d})$.

Background

Bayesian experimental designs

- In most cases, $u(\boldsymbol{d})$ cannot be solved analytically.
- But it can be approximated, e.g., Monte Carlo integration,

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

  where $\boldsymbol{y}_{mb} \sim p(\boldsymbol{y}|\boldsymbol{\theta}_{mb}, m, \mathbf{d})$ and $\boldsymbol{\theta}_{mb} \sim p(\boldsymbol{\theta}|m)$.
- Here, some posterior inference is required to evaluate $u(\boldsymbol{d}, \boldsymbol{y}_{mb}, m)$.
- Hence, $K \times B$ posterior distributions need to be approximated or sampled from to approximated $u(\boldsymbol{d})$.
- **Computationally challenging task**,
  - approximating the expected utility;
  - maximising the utility.

Total entropy

- The **total entropy** utility function can be defined as follows:

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

Total entropy

- The **total entropy** utility function can be defined as follows:

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

- In general, total entropy is a computationally challenging utility.
- **Limited use** (Borth, 1975 and McGree, 2017).

Total entropy

- The **total entropy** utility function can be defined as follows:

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

- In general, total entropy is a computationally challenging utility.
- **Limited use** (Borth, 1975 and McGree, 2017).
- Difficult to evaluate $p(\boldsymbol{y}|\boldsymbol{\theta}, m, \boldsymbol{d})$ for models with intractable likelihoods.

Total entropy

- The **total entropy** utility function can be defined as follows:

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

- In general, total entropy is a computationally challenging utility.
- **Limited use** (Borth, 1975 and McGree, 2017).
- Difficult to evaluate $p(\boldsymbol{y}|\boldsymbol{\theta}, m, \boldsymbol{d})$ for models with intractable likelihoods.
- Hence, $p(\boldsymbol{\theta}|m, \boldsymbol{y}, \boldsymbol{d})$ and $\log p(\boldsymbol{y}|\boldsymbol{d})$ are more difficult to approximate.

Total entropy

- The **total entropy** utility function can be defined as follows:

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

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

Other utilities

- Estimation (KLD) (McGree, 2017).

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

- Model discrimination (Drovandi, McGree, Pettitt, 2014).

$$u_M(\boldsymbol{d}, \boldsymbol{y}, m) = -\log p(m|\boldsymbol{y}, \boldsymbol{d}).$$

Other utilities

- Estimation (KLD) (McGree, 2017).

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

- Model discrimination (Drovandi, McGree, Pettitt, 2014).

$$u_M(\boldsymbol{d}, \boldsymbol{y}, m) = -\log p(m|\boldsymbol{y}, \boldsymbol{d}).$$

- Difficult to evaluate $p(\boldsymbol{y}|\boldsymbol{\theta}, m, \boldsymbol{d})$ for models with intractable likelihoods.
- Hence, $p(\boldsymbol{\theta}|m, \boldsymbol{y}, \boldsymbol{d})$, $\log p(\boldsymbol{y}|m, \boldsymbol{d})$ and $\log p(m|\boldsymbol{y}, \boldsymbol{d})$ are more difficult to approximate.

Other utilities

- Estimation (KLD) (McGree, 2017).

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

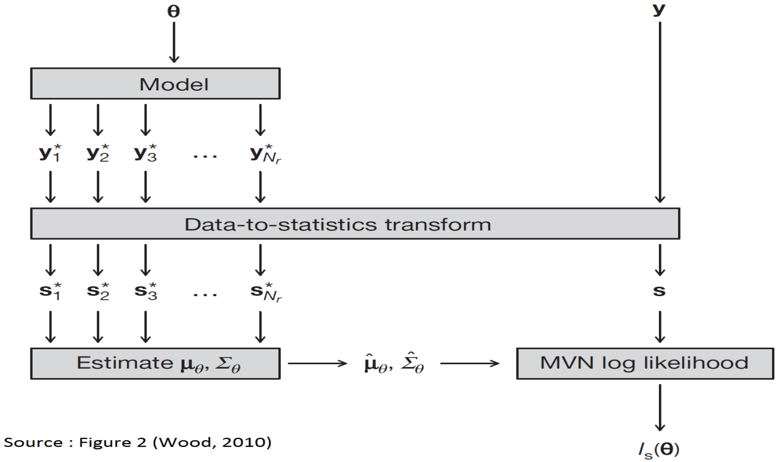
- Model discrimination (Drovandi, McGree, Pettitt, 2014).

$$u_M(\boldsymbol{d}, \boldsymbol{y}, m) = -\log p(m|\boldsymbol{y}, \boldsymbol{d}).$$

- Difficult to evaluate $p(\boldsymbol{y}|\boldsymbol{\theta}, m, \boldsymbol{d})$ for models with intractable likelihoods.
- Hence, $p(\boldsymbol{\theta}|m, \boldsymbol{y}, \boldsymbol{d})$, $\log p(\boldsymbol{y}|m, \boldsymbol{d})$ and $\log p(m|\boldsymbol{y}, \boldsymbol{d})$ are more difficult to approximate.
- Motivates the use of a **synthetic likelihood approach** more generally than just with total entropy.

Synthetic likelihood approach

- Wood (2010) approach - $p(\boldsymbol{y}|\boldsymbol{\theta}, m, \boldsymbol{d})$



Source : Figure 2 (Wood, 2010)

QUT ACEMS

Synthetic likelihood approach

- Counts will be observed from our experiments.

Synthetic likelihood approach

- Counts will be observed from our experiments.
- Extension for **discrete** data.

Synthetic likelihood approach

- Counts will be observed from our experiments.
- Extension for **discrete** data.
- In our case, no summary statistics are considered (mean and variance of simulated data).

Synthetic likelihood approach

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

Synthetic likelihood approach

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

$$p_{SL}(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\theta}, m, \boldsymbol{d}) = p(\boldsymbol{y}_1 - c < \boldsymbol{Y}_1 < \boldsymbol{y}_1 + c, \ldots, \boldsymbol{y}_p - c < \boldsymbol{Y}_p < \boldsymbol{y}_p + c),$$

where $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_p) \sim N(\hat{\mu}(\boldsymbol{\theta}, m, \boldsymbol{d}), \hat{\Sigma}(\boldsymbol{\theta}, m, \boldsymbol{d}))$, c $= 0.5$.

Approximating utility functions

- **Marginal likelihood** can be approximated as follows:

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

where $\boldsymbol{\theta}_b \sim p(\boldsymbol{\theta}|m)$.

Approximating utility functions

- **Marginal likelihood** can be approximated as follows:

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

  where $\boldsymbol{\theta}_b \sim p(\boldsymbol{\theta}|m)$.

- Also for $p(y|d)$

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

Approximating utility functions

- **Marginal likelihood** can be approximated as follows:

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

  where $\boldsymbol{\theta}_b \sim p(\boldsymbol{\theta}|m)$.

- Also for $p(y|d)$

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

- Then, **posterior model probabilities**:

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

Approximating utility functions

- Employ **importance sampling** for approximating posterior distributions.

Approximating utility functions

- Employ **importance sampling** for approximating posterior distributions.
- Use prior as importance distribution.

Approximating utility functions

- Employ **importance sampling** for approximating posterior distributions.
- Use prior as importance distribution.
- Sample $\boldsymbol{\theta}_b \sim p(\boldsymbol{\theta})$, $b = 1, \ldots, B$ (equal weights).

Approximating utility functions

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

$$\{\boldsymbol{\theta}_b, W_b\}_{b=1}^{B}.$$

Approximating utility functions

- **Estimation**:

$$\hat{u}_P(d, \boldsymbol{y}, m) = \sum_{b=1}^{B} W_m^b \log \hat{p}(\boldsymbol{y}|\boldsymbol{\theta}_b, m, \boldsymbol{d}) - \log \hat{p}(\boldsymbol{y}|m, \boldsymbol{d}).$$

- **Discrimination**:

$$\hat{u}_M(\boldsymbol{d}, \boldsymbol{y}, m) = \log \hat{p}(m|\boldsymbol{y}, \boldsymbol{d}).$$

- **Total entropy**:

$$\hat{u}_T(\boldsymbol{d}, \boldsymbol{y}, m) = \sum_{b=1}^{B} W_m^b \log \hat{p}(\boldsymbol{y}|\boldsymbol{\theta}_b, m, \boldsymbol{d}) - \log \hat{p}(\boldsymbol{y}|\boldsymbol{d}).$$

SIR model

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

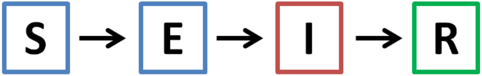
- a **S**usceptible becomes an **I**nfectious individual

$$P\left[s-1, i+1 | s, i\right] = \frac{\beta\, s\, i}{N} \Delta_t + \mathcal{O}(\Delta_t),$$

- an **I**nfectious individual gets **R**ecovered

$$P\left[s, i-1 | s, i\right] = \alpha i \, \Delta_t + \mathcal{O}(\Delta_t).$$

SEIR model



The probabilities of possible events in the next time period Δ_t are

- a **S**usceptible becomes an **E**xposed individual

$$P\left[s-1, e+1, i \mid s, e, i\right] = \frac{\beta\, s\, i}{N} \Delta_t + \mathcal{O}(\Delta_t),$$

- an **E**xposed individual becomes an **I**nfectious individual

$$P\left[s, e-1, i+1 \mid s, e, i\right] = \alpha_I e\, \Delta_t + \mathcal{O}(\Delta_t),$$

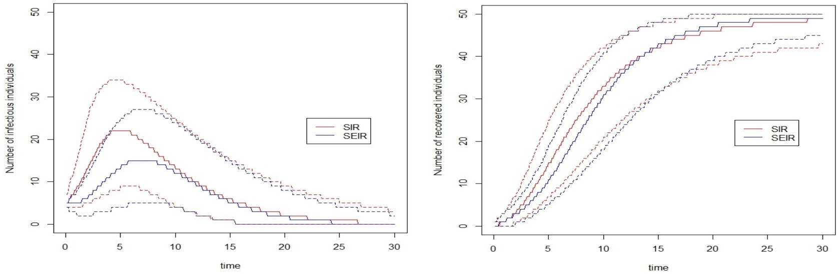
- an **I**nfectious individual gets **R**ecovered ,

$$P\left[s, e, i-1 \mid s, e, i\right] = \alpha_R i\, \Delta_t + \mathcal{O}(\Delta_t).$$

Application

Prior predictive distribution under SIR and SEIR models (prior for SEIR model taken from Backer et al., 2012).

- SEIR : $\beta \sim LN(0.44, 0.16^2)$, $\alpha_I \sim G(25.55, 0.02)$, $\alpha_R \sim G(7.25, 0.04)$.
- SIR: $\beta \sim LN(-0.09, 0.19^2)$, $\alpha \sim G(10.30, 0.02)$.

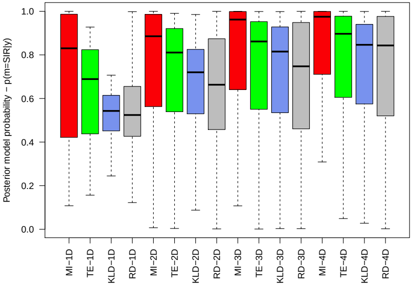


Optimal designs

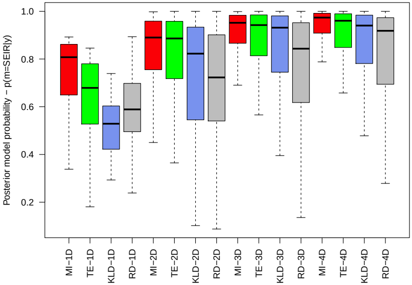
- Refined coordinate exchange algorithm (Dehideniya et al., 2017)

Utility Function	Optimal design $\boldsymbol{d}^*$	$U(\boldsymbol{d}^*)$
KLD	(11.6)	0.91
	(9.4, 19.1)	1.27
	(7.4, 14.2, 27.1)	1.47
	(7.3, 10.9, 16.4, 27.1)	1.60
MI	(3.1)	-0.43
	(4.1, 16)	-0.34
	(0.7, 4.1, 18.4)	-0.30
	(0.7, 4.1, 10.1, 25.3)	-0.28
TE	(7.0)	0.97
	(6.7, 17.5)	1.56
	(6.5, 13.5, 27.1)	1.81
	(5.5, 10.8, 16.3, 27.1)	1.97

Performance of optimal designs in model discrimination

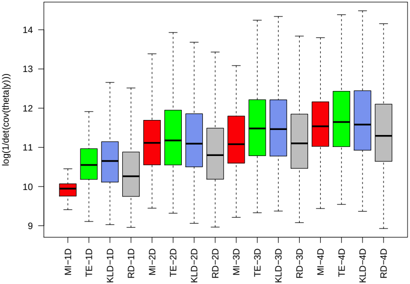


(a) SIR model

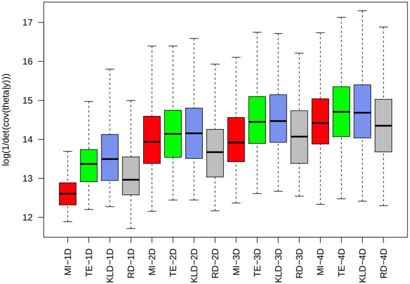


(b) SEIR model

Performance of optimal designs in parameter estimation



(a) SIR model



(b) SEIR model

Discussion

- Approach to design experiments for models with intractable likelihoods.

Discussion

- Approach to design experiments for models with intractable likelihoods.
- Flexible in that a variety of utility functions can be efficiently estimated.

Future research

- Is the normal approximation reasonable, in general? (Other distributions were considered.)

Future research

- Is the normal approximation reasonable, in general? (Other distributions were considered.)
- How small can the sample size (no. of individuals) be?

Future research

- Is the normal approximation reasonable, in general? (Other distributions were considered.)
- How small can the sample size (no. of individuals) be?
- How to extend this method for high dimensional Bayesian design problems for models with intractable likelihoods?
  - Suitable posterior approximations.
  - Possible computational resources (GPU).

Key references

- Borth, D. M., 1975. A total entropy criterion for the dual problem of model discrimination and parameter estimation. Journal of the Royal Statistical Society. Series B (Methodological) 37 (1), 77-87.
- Backer, J., Hagenaars, T., Nodelijk, G., and van Roermund, H. (2012). Vaccination against foot-and-mouth disease i: Epidemiological consequences. Preventive Veterinary Medicine 107, 27 - 40.
- Dehideniya, M. B., Drovandi, C. C., and McGree, J. M. (2017). Optimal Bayesian design for discriminating between models with intractable likelihoods in epidemiology. Technical report, Queensland University of Technology.
- Drovandi, C. C., McGree, J. M., Pettitt, A. N., 2014. A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design. Journal of Computational and Graphical Statistics 23 (1), 3-24.
- McGree, J. M. , 2017. Developments of the total entropy utility function for the dual purpose of model discrimination and parameter estimation in Bayesian design. Computational Statistics & Data Analysis, 113, 207-225.
- Orsel, K., de Jong, M., Bouma, A., Stegeman, J., and Dekker, A. (2007). The effect of vaccination on foot and mouth disease virus transmission among dairy cows. Vaccine 25, 327 - 335.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466, 1102-1104.