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

¹School of Mathematical Sciences
Queensland University of Technology
Australia

²ARC 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. (e.g., Foot and mouth disease)

![Graph showing infection dynamics](http://animalia-life.club)

Source: http://animalia-life.club
Experimental design in Epidemiology

- Spread of a disease within a herd of cows (e.g., Foot and mouth disease)

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

Source: http://animalia-life.club
Experimental design in Epidemiology

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

- 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, \ldots, 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** - \( d^* = \arg \max_d u(d) \), where

\[
u(d) = \sum_{m=1}^{K} p(m) \int_y u(d, y, m)p(y|d, m)dy.
\]
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** - \( d^* = \arg \max_d u(d) \), where

\[
u(d) = \sum_{m=1}^{K} p(m) \int_y u(d, y, m) p(y|d, m)dy.
\]

- \( u(d, y, m) \) is some measure of information gained from \( d \) given model \( m \) and observed data \( y \).
Background
Bayesian experimental designs

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

But it can be approximated, e.g., Monte Carlo integration,

\[
\begin{align*}
    u(d) & \approx \sum_{m=1}^{K} p(m) \frac{1}{B} \sum_{b=1}^{B} u(d, y_{mb}, m),
\end{align*}
\]

where \( y_{mb} \sim p(y|\theta_{mb}, m, d) \) and \( \theta_{mb} \sim p(\theta|m) \).
In most cases, $u(d)$ cannot be solved analytically.

But it can be approximated, e.g., 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 $\theta_{mb} \sim p(\theta|m)$.

Here, some posterior inference is required to evaluate $u(d, y_{mb}, m)$. 

Background
Bayesian experimental designs

- In most cases, $u(d)$ cannot be solved analytically.
- But it can be approximated, e.g., 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 $\theta_{mb} \sim p(\theta|m)$.

- Here, some posterior inference is required to evaluate $u(d, y_{mb}, m)$.
- Hence, $K \times B$ posterior distributions need to be approximated or sampled from to approximate $u(d)$. 

Computationally challenging task,
▶ approximating the expected utility;
▶ maximising the utility.
In most cases, $u(d)$ cannot be solved analytically.

But it can be approximated, e.g., Monte Carlo integration,

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

where $y_{mb} \sim p(y | \theta_{mb}, m, d)$ and $\theta_{mb} \sim p(\theta | m)$.

Here, some posterior inference is required to evaluate $u(d, y_{mb}, m)$.

Hence, $K \times B$ posterior distributions need to be approximated or sampled from to approximate $u(d)$.

Computationally challenging task,

- approximating the expected utility;
- maximising the utility.
Total entropy

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

\[
u_T(d, y, m) = \int_\theta p(\theta|m, y, d) \log p(y|\theta, m, d)d\theta - \log p(y|d).
\]
Total entropy

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

\[
u_T(d, y, m) = \int_\theta p(\theta|m, y, d) \log p(y|\theta, m, d) d\theta - \log p(y|d).
\]

- In general, total entropy is a computationally challenging utility.
- **Limited use** (Borth, 1975 and McGree, 2017).
The **total entropy** utility function can be defined as follows:

\[
    u_T(d, y, m) = \int_\theta p(\theta|m, y, d) \log p(y|\theta, m, d) d\theta - \log p(y|d).
\]

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

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

\[
u_T(d, y, m) = \int_\theta p(\theta|m, y, d) \log p(y|\theta, m, d) d\theta - \log p(y|d).
\]

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

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

\[ u_T(d, y, m) = \int_\theta p(\theta|m, y, d) \log p(y|\theta, m, d) d\theta - \log p(y|d). \]

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

- Estimation (KLD) (McGree, 2017).

\[ u_P(d, y, m) = \int \theta p(\theta|m, y, d) \log p(y|\theta, m, d) d\theta - \log p(y|m, d). \]

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

\[ u_M(d, y, m) = -\log p(m|y, d). \]
Other utilities

- Estimation (KLD) (McGree, 2017).

  \[ u_\mathcal{P}(\mathbf{d}, \mathbf{y}, m) = \int_{\theta} p(\theta|m, \mathbf{y}, \mathbf{d}) \log p(\mathbf{y}|\theta, m, \mathbf{d}) d\theta - \log p(\mathbf{y}|m, \mathbf{d}). \]

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

  \[ u_\mathcal{M}(\mathbf{d}, \mathbf{y}, m) = -\log p(m|\mathbf{y}, \mathbf{d}). \]

- Difficult to evaluate \( p(\mathbf{y}|\theta, m, \mathbf{d}) \) for models with intractable likelihoods.

- Hence, \( p(\theta|m, \mathbf{y}, \mathbf{d}) \), \( \log p(\mathbf{y}|m, \mathbf{d}) \) and \( \log p(m|\mathbf{y}, \mathbf{d}) \) are more difficult to approximate.
Other utilities

- Estimation (KLD) (McGree, 2017).

\[ u_P(d, y, m) = \int_{\theta} p(\theta|m, y, d) \log p(y|\theta, m, d) d\theta - \log p(y|m, d). \]

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

\[ u_M(d, y, m) = -\log p(m|y, d). \]

- Difficult to evaluate \( p(y|\theta, m, d) \) for models with intractable likelihoods.

- Hence, \( p(\theta|m, y, d) \), \( \log p(y|m, d) \) and \( \log p(m|y, d) \) 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 - \( p(y|\theta, m, d) \)

Source: Figure 2 (Wood, 2010)
Synthetic likelihood approach

- Counts will be observed from our experiments.
Synthetic likelihood approach

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

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

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

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

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:

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

where \( \theta_b \sim p(\theta|m) \).
Approximating utility functions

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

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

where \( \theta_b \sim p(\theta|m) \).

- Also for \( p(y|d) \)

\[
\hat{p}(y|d) = \sum_{m=1}^{K} \hat{p}(y|m, d)p(m).
\]
Approximating utility functions

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

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

where $\theta_b \sim p(\theta|m)$.

- 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.
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 $\theta_b \sim p(\theta), \, b = 1, \ldots, B$ (equal weights).
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:

\[
\{\theta_b, W_b\}_{b=1}^B.
\]
Approximating utility functions

- **Estimation**:
  \[ \hat{u}_P(d, y, m) = \sum_{b=1}^{B} W_b \log \hat{p}(y|\theta_b, m, d) - \log \hat{p}(y|m, d). \]

- **Discrimination**:
  \[ \hat{u}_M(d, y, m) = \log \hat{p}(m|y, d). \]

- **Total entropy**:
  \[ \hat{u}_T(d, y, m) = \sum_{b=1}^{B} W_b \log \hat{p}(y|\theta_b, m, d) - \log \hat{p}(y|d). \]
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 $\Delta t$ are

- a **Susceptible** becomes an **Infectious** individual

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

- an **Infectious** individual gets **Recovered**

$$
P[s, i - 1 | s, i] = \alpha i \Delta t + \mathcal{O}(\Delta t).
$$
SEIR model

The probabilities of possible events in the next time period $\Delta_t$ are

- a Susceptible becomes an Exposed individual

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

- an Exposed individual becomes an Infectious individual

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

- an Infectious individual gets Recovered

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

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