General Bayesian optimal design of experiments

Antony Overstall

University of Southampton
Southampton, United Kingdom

MODA - 19th June 2026

Thank you

Thank you to Werner and the conference organisers!

- Introduction: layout notation
- Bayesian optimal design: including for misspecified models.
- General Bayesian inference
- General Bayesian optimal design
- General Bayesian optimal design for linear models
- Further examples and discussion

Joint work with **Jacinta Holloway-Brown** (University of Adelaide, Australia) and **James McGree** (Queensland University of Technology, Australia)

Introduction

Summary

- Approach is motivated by designing a clinical trial for a child vaccine for indigenous Australians.
- The response is time-to-gastroenteritis.
- Common classical approach is to use Cox proportional hazards model - doesn't require specific distributional assumptions.
- How can one apply Cox model with a Bayesian approach?
- Can use general Bayesian inference - doesn't require specific distributional assumptions.
- This talk will introduce a decision-theoretic optimal design framework for general Bayesian inference.

Aim of the experiment

Suppose the aim of the experiment is to learn the relationship between

$$\begin{array}{rl} \text{response} & y \\ \text{controllable variables} & \mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{X} \subset \mathbb{R}^k. \end{array}$$

- The experiment consists of n runs.
- The ith run consists of specifying controllable variables $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik}) \in \mathbb{X}$ and observing response y_i, for $i = 1, \ldots, n$.
- Let $\mathbf{y} = (y_1, \ldots, y_n)$ be the $n \times 1$ vector of responses.
- Let X be the $n \times k$ design matrix with ith row given by $\mathbf{x}_i$, for $i = 1, \ldots, n$.
- The data $\mathbf{y}$ and X are used to estimate the relationship between y and $\mathbf{x}$.
- The research question is: how should the design $X \in \mathbb{X}^n$ be chosen to best estimate this relationship?

Bayesian optimal design

Parametric statistical inference

- The responses are realisations of an unknown probability distribution $\mathcal{T}(X)$ depending on design X.
- Assume a statistical model: a probability distribution $\mathcal{S}(\mathbf{t}; X)$ depending on parameters $\mathbf{t}$ and design X.
- Assume there exist *true parameter values* $\boldsymbol{\theta}_T$, such that $\mathcal{S}(\boldsymbol{\theta}_T; X) = \mathcal{T}(X)$.
- Estimating the relationship between y and $\mathbf{x}$, now reduces to estimating $\boldsymbol{\theta}_T$.

- The Bayesian approach involves evaluating the *Bayesian* posterior distribution, $\mathcal{B}(\mathbf{y}; X)$, of $\boldsymbol{\theta}_T$ with pdf

$$\pi_B(\mathbf{t}|\mathbf{y}, X) \propto s(\mathbf{y}; \mathbf{t}, X)\pi_T(\mathbf{t}),$$

where

$$\begin{aligned} s(\cdot; \mathbf{t}, X) &\quad : \quad \text{joint pdf/pf of } \mathcal{S}(\mathbf{t}; X); \\ \pi_T(\cdot) &\quad : \quad \text{prior pdf of prior distribution, } \mathcal{P}_T. \end{aligned}$$

- $\mathcal{P}_T$ represents information known about $\boldsymbol{\theta}_T$ before observing $\mathbf{y}$.

Bayesian decision-theoretic optimal design of experiments

- *Bayesian optimal design* (BOD) for brevity.
- Specify a utility function

$$u_B(\mathbf{t}, \mathbf{y}, X)$$

giving the gain in information from estimating parameter values $\mathbf{t}$ using data $\mathbf{y}$ and X, with Bayesian posterior $\mathcal{B}(\mathbf{y}; X)$.

- Expected Bayesian utility is

$$U_B(X) = \mathrm{E}_{\mathcal{P}_T} \left\{ \mathrm{E}_{\mathcal{S}(\boldsymbol{\theta}_T; X)} \left[u_B(\boldsymbol{\theta}_T, \mathbf{y}, X) \right] \right\}.$$

- Bayesian optimal design is given by maximising $U_B(X)$ over $\mathbb{X}^n$.

Shannon information (SH)

$$u_B^{(SH)}(\mathbf{t}, \mathbf{y}, X) = \log \pi_B\left(\mathbf{t} | \mathbf{y}, X\right).$$

Another example is negative squared error (NE).

Advantages and disadvantages of Bayesian optimal design

Advantages:

1. Can take account of experimental aim through choice of utility function.
2. Accounts for all assumed sources of uncertainty.

Disadvantages:

1. Bayesian optimal design not available in closed form and requires numerical methods.
2. The model, given by $\mathcal{S}(\cdot; X)$, is misspecified.

   There do not exist $\boldsymbol{\theta}_T$ such that $\mathcal{S}(\boldsymbol{\theta}_T; X) = \mathcal{T}(X)$.

Bayesian optimal design under misspecified models

- The Bayesian posterior under a misspecified model provides coherent inference for $\boldsymbol{\theta}_{SI}$: the parameter values that minimise the KL divergence between $\mathcal{T}(X)$ and $\mathcal{S}(\mathbf{t}; X)$.
- Expected Bayesian utility is

$$U_B(X) = \mathrm{E}_{\mathcal{P}_T}\left\{\mathrm{E}_{\mathcal{S}(\boldsymbol{\theta}_T; X)}\left[u_B(\boldsymbol{\theta}_T, \mathbf{y}, X)\right]\right\}.$$

- The inner expectation is the average utility under $\mathbf{y} \sim \mathcal{S}(\mathbf{t}; X) \neq \mathcal{T}(X)$ for any $\mathbf{t}$.
- The "outer" expectation is with respect to $\boldsymbol{\theta}_T$ which do not exist!
- Could replace $\boldsymbol{\theta}_T$ in the utility by $\boldsymbol{\theta}_{SI}$: but what should the "inner" expectation be taken with respect to?
  - If we use the statistical model, then $\boldsymbol{\theta}_{SI} = \boldsymbol{\theta}_T$ and the KL divergence is zero.
- Bayesian design only makes sense if the model is true.

Bayesian optimal design under misspecified models

- Expected Bayesian utility is

$$U_B(X) = \mathrm{E}_{\mathcal{P}_T}\left\{\mathrm{E}_{\mathcal{S}(\boldsymbol{\theta}_T;X)}\left[u_B(\boldsymbol{\theta}_T, \mathbf{y}, X)\right]\right\}.$$

- Overstall & McGree (2022) replaced $\boldsymbol{\theta}_T$ by $\boldsymbol{\theta}_{SI}$ and $\mathcal{S}(\cdot; X)$ by a designer distribution.
- We go further and consider a more general approach to statistical inference.

General Bayesian inference

- General Bayesian inference allows learning from data, $\mathbf{y}$ and X, without necessarily specifying a probability distribution $\mathcal{S}(\mathbf{t}, X)$.
- Specify a loss function $\ell(\mathbf{t}; \mathbf{y}, X)$.
- The loss identifies desirable parameters values for data $\mathbf{y}$ and X, i.e. desirable parameters have small loss.
- A classical approach would be to find the M-estimates:

$$\hat{\boldsymbol{\theta}}_\ell = \arg\min_{\mathbf{t}} \ell(\mathbf{t}; \mathbf{y}, X).$$

- Alternatively, General Bayesian inference proceeds by evaluating the General Bayesian posterior distribution, $\mathcal{G}_\ell(\mathbf{y}; X)$, with pdf

$$\pi_\ell(\mathbf{t}|\mathbf{y}, X) \propto \exp\left[-w\ell(\mathbf{t}; \mathbf{y}, X)\right] \pi_\ell(\mathbf{t}).$$

Target parameter values

$$\pi_\ell(\mathbf{t}|\mathbf{y}, X) \propto \exp\left[-w\ell(\mathbf{t}; \mathbf{y}, X)\right] \pi_\ell(\mathbf{t}).$$

- The General Bayesian posterior is targeting/estimating $\boldsymbol{\theta}_\ell$: the values of $\mathbf{t}$ that minimise the expected loss:

  $$L(\mathbf{t}; X) = \mathrm{E}_{\mathcal{T}(X)}\left[w\ell(\mathbf{t}; \mathbf{y}, X)\right],$$

  under the true distribution of $\mathcal{T}(X)$ (Bissiri et al, 2016).
- $\pi_\ell(\cdot)$ is the pdf of the prior distribution, $\mathcal{P}_\ell$, for $\boldsymbol{\theta}_\ell$.

Calibration weight

$$\pi_\ell(\mathbf{t}|\mathbf{y}, X) \propto \exp\left[-w\ell(\mathbf{t}; \mathbf{y}, X)\right] \pi_\ell(\mathbf{t}).$$

- The General Bayesian posterior depends on calibration weight $w > 0$.
- If $w \to 0$, then $\mathcal{G}_\ell(\mathbf{y}; X) \to \mathcal{P}_\ell$.
- If $w \to \infty$, then $\mathcal{G}_\ell(\mathbf{y}; X)$ converges to a point mass at $\hat{\boldsymbol{\theta}}_\ell$.
- The calibration weight controls the rate of learning.
- Its specification is significant.
- Example: Maintaining operational characteristics; specify w so that

$$\mathrm{tr}\left[\hat{\mathrm{var}}_{\mathcal{G}_\ell(\mathbf{y};X)}(\boldsymbol{\theta}_\ell)\right] = \mathrm{tr}\left[\hat{\mathrm{var}}\left(\hat{\boldsymbol{\theta}}_\ell\right)\right].$$

General Bayesian optimal design

Expected Bayesian utility

- The idea is to create an decision-theoretic optimal design framework for General Bayesian inference.
- Called *General Bayesian optimal design* (GBOD) for brevity.

- General Bayesian optimal design starts with specification of utility function
$$u_G(\mathbf{t}, \mathbf{y}, X)$$
giving the gain in information from estimating parameter values $\mathbf{t}$ using data $\mathbf{y}$ and X, with General Bayesian posterior $\mathcal{G}_\ell$ for $\boldsymbol{\theta}_\ell$.
- Can use the same utility functions as Bayesian optimal design, e.g. negative squared error (NE) or Shannon information (SH).

Ideal scenario

- Inside the expected utility we would like to evaluate $u_G(\mathbf{t}, \mathbf{y}, X)$ at $\mathbf{t} = \boldsymbol{\theta}_\ell$ where

$$\boldsymbol{\theta}_\ell = \mathrm{E}_{\mathcal{T}(X)}\left[w\ell(\mathbf{t}; \mathbf{y}, X)\right]$$

  are the true target parameter values.
- Then the expected General Bayesian utility would be

$$U_G(X) = \mathrm{E}_{\mathcal{T}(X)}\left\{u_G\left[\boldsymbol{\theta}_\ell, \mathbf{y}, X\right]\right\}.$$

Utility

- However, we do not know $\mathcal{T}(X)$.
- Instead we specify a *designer distribution*, denoted $\mathcal{D}(\mathbf{r}; X)$, as an approximation to $\mathcal{T}(X)$.
- $\mathbf{r}$ are hypervariables which allow more than one approximation to $\mathcal{T}(X)$.
- $\mathbf{r} \sim \mathcal{P}_D$.
- The difference between $\mathcal{D}(\mathbf{r}; X)$ and a statistical model $\mathcal{S}(\mathbf{t}; X)$ will be addressed in 2 slides time.

Expected General Bayesian utility

- For a given designer distribution, $\mathcal{D}(\mathbf{r}; X)$, we can define target parameter values

$$\boldsymbol{\theta}_{\ell,D}(\mathbf{r}, X) = \arg\min_{\mathbf{t}} \mathrm{E}_{\mathcal{D}(\mathbf{r};X)} \left[w\ell(\mathbf{t}; \mathbf{y}, X) \right].$$

- The idea is that there exists some $\mathbf{r}$ such that $\boldsymbol{\theta}_{\ell,D}(\mathbf{r}, X) \approx \boldsymbol{\theta}_{\ell}$.
- The utility is now evaluated at $\mathbf{t} = \boldsymbol{\theta}_{\ell,D}(\mathbf{r}, X)$.
- The expected General Bayesian utility is

$$U_G(X) = \mathrm{E}_{\mathcal{P}_{\mathcal{D}}} \left(\mathrm{E}_{\mathcal{D}(\mathbf{r};X)} \left\{ u_G \left[\boldsymbol{\theta}_{\ell,D}(\mathbf{r}, X), \mathbf{y}, X \right] \right\} \right)$$

- General Bayesian optimal design is given by maximising $U_G(X)$ over $\mathbb{X}^n$.

Difference between $\mathcal{D}(\mathbf{r}; X)$ and a statistical model $\mathcal{S}(\mathbf{t}; X)$

- If we are able to specify designer distribution $\mathcal{D}(\mathbf{r}; X)$, why can't we just use this as the statistical model?
- $\mathcal{D}(\mathbf{r}; X)$ will never be "fitted" to the data $\mathbf{y}$ and X.
- Therefore, we can be much more flexible in the specification of $\mathcal{D}(\mathbf{r}; X)$ than $\mathcal{S}(\mathbf{t}; X)$.
- $\mathcal{D}(\mathbf{r}; X)$ does not need to address aims of experiment.

General Bayesian optimal design for linear models

Bayesian (and classical) optimal design for linear models

- Specify $\mathcal{S}(\mathbf{t}, s^2; X)$ such that $y_1, \ldots, y_n$ are independent with

$$y_i \sim \mathrm{N}\left(\mathbf{f}(\mathbf{x}_i)^T \mathbf{t}, s^2\right),$$

  where $\mathbf{f}(\mathbf{x})$ is a regression function and $s^2 > 0$ is nuisance parameter.
- Assume there exist $\mathbf{t} = \boldsymbol{\theta}_T$ and $s^2 = \sigma_T^2$, such that $\mathcal{S}(\boldsymbol{\theta}_T, \sigma_T^2; X) = \mathcal{T}(X)$.
- Assume uniform prior distributions for $\boldsymbol{\theta}_T$ and σ_T^2.
- Bayesian optimal design under Shannon information is given by maximising

$$\tilde{U}_B^{(SH)}(X) \quad = \quad \log |F^T F|, \quad \text{(D-optimality)}$$

  where F is $n \times p$ model matrix with ith row $\mathbf{f}(\mathbf{x}_i)^T$, for $i = 1, \ldots, n$.

Squared error loss

- For General Bayesian inference, consider the squared error loss.

$$\ell_{SS}(\mathbf{t}; \mathbf{y}, X) = \sum_{i=1}^{n} \left(y_i - \mathbf{f}(\mathbf{x}_i)^T \mathbf{t}\right)^2 .$$

- The General Bayesian posterior is given by

$$\pi_{SS}(\mathbf{t}|\mathbf{y}, X) \propto \exp\left[-w \sum_{i=1}^{n} \left(y_i - \mathbf{f}(\mathbf{x}_i)^T \mathbf{t}\right)^2\right] \pi_{SS}(\mathbf{t}).$$

- Under a uniform prior, the General Bayesian posterior, $\mathcal{G}(\mathbf{y}; X)$, is

$$\mathrm{N}\left(\hat{\boldsymbol{\theta}}_{SS}(\mathbf{y}; X), \frac{1}{2w}(F^T F)^{-1}\right),$$

where the M-estimates

$$\hat{\boldsymbol{\theta}}_{SS}(\mathbf{y}; X) = (F^T F)^{-1} F^T \mathbf{y},$$

are the usual least squares estimates.

Calibration weight

- Using the operational characteristics concept from earlier.

$$\begin{aligned} \mathrm{var}_{\mathcal{G}_{SS}(\mathbf{y};X)}(\boldsymbol{\theta}_{SS}) &= \mathrm{var}\left[\hat{\boldsymbol{\theta}}_{SS}(\mathbf{y},X)\right], \\ \frac{1}{2w}(F^T F)^{-1} &= \tau^2 (F^T F)^{-1}. \end{aligned}$$

- Hence $w = 1/2\tau^2$ where τ^2 is the true response variance (assumed finite).
- Use

$$w = \frac{1}{2\tilde{\tau}^2},$$

where $\tilde{\tau}^2$ is an estimate of the true response variance, τ^2, under the unique-treatment model (Gilmour & Trinca, 2012).
- The unique-treatment model estimates τ^2 by using repeated treatments.

- For the designer distribution, $\mathcal{D}(\mathbf{r}; X)$, we assume

$$y_i = \mu(\mathbf{x}_i) + \epsilon_i.$$

- $\epsilon_1, \ldots, \epsilon_n$ are independent
- ϵ_i has a scale mixture of normal distributions, i.e. $\epsilon_i \sim \mathrm{N}(0, \kappa)$ where κ is a random variable.
- The hyper-variables are $\mathbf{r} = (\mu(\cdot), \kappa)$.
- The target parameter values are

$$\begin{aligned} \boldsymbol{\theta}_{SS,D}(\mathbf{r}; X) &= \arg\min_{\mathbf{t}} \mathrm{E}_{\mathcal{D}(\mathbf{r};X)} \left[\frac{1}{2\tilde{\sigma}^2} \ell_{SS}(\mathbf{t}; \mathbf{y}, X) \right] \\ &= \left(F^T F\right)^{-1} F^T \boldsymbol{\mu}, \end{aligned}$$

where $\boldsymbol{\mu} = (\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_n))$.

Expected General Bayesian utilities

The expected General Bayesian utility under Shannon information is (up to some constants)

$$\tilde{U}_G^{(SH)}(X) = -ph(d) + \log|F^T F|.$$

where

$$d = \text{Number of repeated treatments}$$

and

$$h(d) = \begin{cases} \psi\left(\frac{d}{2}\right) - \log d + \frac{d}{d-2} & \text{if } d > 2; \\ \infty & \text{if } d = 0, 1, 2. \end{cases}$$

Expected General Bayesian utilities

$$\tilde{U}_G^{(SH)}(X) = -ph(d) + \log|F^T F|.$$

- The objective function is a modification of objective function for D-optimality.
- The modification rewards increasing d, i.e. the amount of replication.
- If $d = 0, 1, 2$, then $\tilde{U}_G^{(SH)}(X) = -\infty$: this ensures that the calibration weight can be specified and the General Bayesian posterior exists.
- Do not need to specify $\mathcal{P}_{\mathcal{D}}$; the distribution for $\mathbf{r} = (\mu(\cdot), \kappa)$.
- Can also do this with negative squared error utility
  - Result: modified A-optimal objective function ensuring $d > 0$.
  - $\epsilon_1, \ldots, \epsilon_n$ do not need to be normally distributed.

- Consider a numerical example from Gilmour & Trinca (2011).
- Here, $n = 16$, $k = 3$, $\mathbf{f}(\cdot)$ implements an intercept, main & quadratic effects, and two-way interactions, giving $p = 10$.
- All designs have points in $\{0, \pm 1\}$.
- Design efficiencies:

	Design	
	GBOD	BOD (D-optimal)
GBOD	100%	0%
BOD (D-optimal)	85%	100%
d	6	0

- The difference between the Bayesian and general Bayesian posteriors is similar to the difference between maximum likelihood and least squares estimates.
- Functionally, they are the same, but the former requires an assumption of normality for the responses.
- The variance of both of these estimates is

$$\tau^2(F^T F)^{-1}.$$

- The general Bayesian posterior has the same variance but replaces τ^2 by a model-independent estimate (as part of specifying the calibration weight).
- The choice of estimate drives the reward for replicated design points.

Concluding remarks

Concluding remarks

- Summary
- Further examples
- Current and future work

Summary 1

- General Bayesian inference is an extension of Bayesian inference.
- Specification of likelihood is replaced by specification of loss function.
- Target parameters are those that minimise expectation (with respect to true data-generating distribution) of loss function.

Summary 2

- General Bayesian optimal design is a decision-theoretic framework for General Bayesian inference.
- Unknown true data-generating distribution is replaced by designer distribution: flexible representations of true data-generating distribution.
- Having to specify the designer distribution is a hurdle - but it does not need to be fitted to responses.
- Design will depend on specification of w. But specification of w is not a problem unique to design.

Further examples

- Considered linear model with L^2 loss (Tou & Wu, 2015) instead of squared error.
- Count responses with negative quasi-log-likelihood as loss function.
- Time-to-event responses with negative Cox log-partial-likelihood as loss function.
- For more details, see Overstall, Holloway-Brown & McGree (2026).
- Approach has been applied to designing a clinical trial for a child vaccine for indigenous Australians (McGree, et al 2025).

References

- Bissiri, P., Holmes, C. and Walker, S. (2016) A general framework for updating prior beliefs. *JRSSB* 78, 1103-1130.
- Gilmour, S. and Trinca, L. (2012) Optimum design of experiments for statistical inference (with discussion). *JRSSC* 61, 345-401.
- McGree, J.M., Overstall, A.M., Jones, M. and Mahar, R.K. (2025) A general Bayesian approach to design adaptive clinical trials with time-to-event outcomes. *Statistics in Medicine* **24**.
- Overstall, A.M., and McGree, J.M. (2022) Bayesian decision-theoretic design of experiments under an alternative model *Bayesian Analysis* 17, 1021-1041.
- Overstall, A.M., Holloway-Brown, J. and McGree, J.M. (2026) Gibbs optimal design of experiments. *Statistical Science* **To appear**.
- Tuo, R. and Wu, C. (2015) Efficient calibration for imperfect computer models. *Annals of Statistics* 43, 2331–2352.