Parameter-dependent Optimal Designs

John Matthews
University of Newcastle
j.n.s.matthews@ncl.ac.uk

May 2010

Introduction

- Experiments are performed to learn about some system

Introduction

- Experiments are performed to learn about some system
- Designing such experiments requires knowledge of the system

Introduction

- Experiments are performed to learn about some system
- Designing such experiments requires knowledge of the system
- This is an inevitable tension in experimental design

Introduction

- Experiments are performed to learn about some system
- Designing such experiments requires knowledge of the system
- This is an inevitable tension in experimental design
- Satisfactory resolution requires judgment and prior knowledge

Introduction

- Experiments are performed to learn about some system
- Designing such experiments requires knowledge of the system
- This is an inevitable tension in experimental design
- Satisfactory resolution requires judgment and prior knowledge
- Bayesian ideas are useful but need not entail a Bayesian analysis

Introduction

- Experiments are performed to learn about some system
- Designing such experiments requires knowledge of the system
- This is an inevitable tension in experimental design
- Satisfactory resolution requires judgment and prior knowledge
- Bayesian ideas are useful but need not entail a Bayesian analysis
- This talk illustrates some aspects of this using Michaelis-Menten experiments

Main example - Michaelis Menten studies

Rate of reaction in enzyme kinetics determined by the *Michaelis Menten* equation.

$$v = \frac{V_{\max} s}{K + s}$$

Main example - Michaelis Menten studies

Rate of reaction in enzyme kinetics determined by the *Michaelis Menten* equation.

$$v = \frac{V_{\max} s}{K + s}$$

where

- s is the substrate concentration, v the rate of reaction

Main example - Michaelis Menten studies

Rate of reaction in enzyme kinetics determined by the *Michaelis Menten* equation.

$$v = \frac{V_{\max} s}{K + s}$$

where

- s is the substrate concentration, v the rate of reaction
- K is the *Michaelis* parameter;

Main example - Michaelis Menten studies

Rate of reaction in enzyme kinetics determined by the *Michaelis Menten* equation.

$$v = \frac{V_{\max} s}{K + s}$$

where

- s is the substrate concentration, v the rate of reaction
- K is the *Michaelis* parameter; $V_{\max}$ the maximum rate of reaction

Main example - Michaelis Menten studies

Rate of reaction in enzyme kinetics determined by the *Michaelis Menten* equation.

$$v = \frac{V_{\max} s}{K + s}$$

where

- s is the substrate concentration, v the rate of reaction
- K is the *Michaelis* parameter; $V_{\max}$ the maximum rate of reaction
- interest focusses on K and other quantities, such as $V_{\max}/K$

Statistical model

- Conduct experiment to estimate parameters

Statistical model

- Conduct experiment to estimate parameters
- Observe rate v_i at concentrations s_i, $i = 1, \ldots, N$

Statistical model

- Conduct experiment to estimate parameters
- Observe rate v_i at concentrations s_i, $i = 1, \ldots, N$
- Assume model

$$v_i = \frac{V_{\max} s_i}{K + s_i} + \epsilon_i$$

with the ϵ_i being independent, $\mathrm{N}(0, \sigma^2)$

Statistical model

- Conduct experiment to estimate parameters
- Observe rate v_i at concentrations s_i, $i = 1, \ldots, N$
- Assume model

$$v_i = \frac{V_{\max} s_i}{K + s_i} + \epsilon_i$$

with the ϵ_i being independent, $\mathrm{N}(0, \sigma^2)$

- Not wholly realistic but will serve for present purposes.

Statistical model

- Conduct experiment to estimate parameters
- Observe rate v_i at concentrations s_i, $i = 1, \ldots, N$
- Assume model

$$v_i = \frac{V_{\max} s_i}{K + s_i} + \epsilon_i$$

  with the ϵ_i being independent, $\mathrm{N}(0, \sigma^2)$
- Not wholly realistic but will serve for present purposes.
- Substantial history to fitting this model exists.

Design Problem

Design Problem

- How should the substrate concentrations be chosen?

Design Problem

- How should the substrate concentrations be chosen?
- Depends on aim of experiment

Design Problem

- How should the substrate concentrations be chosen?
- Depends on aim of experiment
- Suppose we wish to estimate K as precisely as possible.

Expected Information

Expected information matrix can be written as $\mathscr{I} = N\sigma^{-2}\mathscr{M}$, where

Expected Information

Expected information matrix can be written as $\mathscr{I} = N\sigma^{-2}\mathscr{M}$, where

$$\mathscr{M} = \begin{pmatrix} \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^2} & \\ -V_{\max} \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^3} & V_{\max}^2 \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^4} \end{pmatrix}$$

Expected Information

Expected information matrix can be written as $\mathscr{I} = N\sigma^{-2}\mathscr{M}$, where

$$\mathscr{M} = \begin{pmatrix} \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^2} & \\ -V_{\max} \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^3} & V_{\max}^2 \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^4} \end{pmatrix}$$

- Usual continuous design framework: N observations at n *distinct* concentrations s_j;

Expected Information

Expected information matrix can be written as $\mathscr{I} = N\sigma^{-2}\mathscr{M}$, where

$$\mathscr{M} = \begin{pmatrix} \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^2} & \\ -V_{\max} \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^3} & V_{\max}^2 \sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^4} \end{pmatrix}$$

- Usual continuous design framework: N observations at n *distinct* concentrations s_j;
- at s_j make $N\eta_j$ observations: $\eta_j > 0$, $\sum \eta_j = 1$

D-optimal

Let us seek a D-optimal design - i.e. maximise $\det(\mathscr{I})$.

D-optimal

Let us seek a D-optimal design - i.e. maximise $\det(\mathscr{I})$.
Equivalent to maximising $\Delta = \det(\mathscr{M})$

$$\Delta = \left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^2}\right)\left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^4}\right) - \left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^3}\right)^2$$

D-optimal

Let us seek a D-optimal design - i.e. maximise $\det(\mathscr{I})$.
Equivalent to maximising $\Delta = \det(\mathscr{M})$

$$\Delta = \left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^2}\right)\left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^4}\right) - \left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^3}\right)^2$$

Depends on design $\xi = (\mathbf{s}, \eta)$ and K but not $V_{\max}$, nor N nor σ^2.

D-optimal

Let us seek a D-optimal design - i.e. maximise $\det(\mathscr{I})$.
Equivalent to maximising $\Delta = \det(\mathscr{M})$

$$\Delta = \left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^2}\right)\left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^4}\right) - \left(\sum_{j=1}^{n} \eta_j \frac{s_j^2}{(K+s_j)^3}\right)^2$$

Depends on design $\xi = (\mathbf{s}, \eta)$ and K but not $V_{\max}$, nor N nor σ^2.

Locally optimal design

Locally D-optimal design: $m = 2$, write $y_j = s_j/(K + s_j)$.

$$\Delta = \eta_1 \eta_2 y_1^2 y_2^2 (y_1 - y_2)^2$$

Optimal design $\eta_j = \frac{1}{2}$, $y_1 = \frac{1}{2}$, $y_2 = 1$ i.e. $s_1 = K$, $s_2 = \infty$

'Bayesian' optimal design

K is unknown so find design by maximising $\mathbb{E}_{\pi}(\log\det(\mathscr{I}))$

'Bayesian' optimal design

K is unknown so find design by maximising $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ for prior $\pi(K, V_{\max})$

'Bayesian' optimal design

K is unknown so find design by maximising $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ for prior $\pi(K, V_{\max})$

Objective factors as

$$f(N) + f(\sigma^2) + f(V_{\max}) + f(K, \xi)$$

'Bayesian' optimal design

K is unknown so find design by maximising $\mathbb{E}_{\pi}(\log \det(\mathscr{I}))$ for prior $\pi(K, V_{\max})$
Objective factors as

$$f(N) + f(\sigma^2) + f(V_{\max}) + f(K, \xi)$$

So expectation involving ξ depends only on the K-marginal of π.

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:
  1. $s_j = K_U t_j$

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:
  1. $s_j = K_U t_j$
  2. $K = K_U \kappa$

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:
  1. $s_j = K_U t_j$
  2. $K = K_U \kappa$
  3. $\kappa_L = K_L/K_U < \kappa < 1$

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:
  1. $s_j = K_U t_j$
  2. $K = K_U \kappa$
  3. $\kappa_L = K_L / K_U < \kappa < 1$
- Assume either
  1. κ uniform on its range or

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:
  1. $s_j = K_U t_j$
  2. $K = K_U \kappa$
  3. $\kappa_L = K_L/K_U < \kappa < 1$
- Assume either
  1. κ uniform on its range or
  2. $\log \kappa$ uniform on its range

Prior for K

- Convenient to assume finite support on $[K_L, K_U]$
- Some parsimony in specification if we scale:
  1. $s_j = K_U t_j$
  2. $K = K_U \kappa$
  3. $\kappa_L = K_L/K_U < \kappa < 1$
- Assume either
  1. κ uniform on its range or
  2. $\log \kappa$ uniform on its range

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_{\pi}(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$
t_1 is solution of:

$$\mathbb{E}_\pi\left(\frac{\kappa - t_1}{\kappa + t_1}\right) = 0$$

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$
t_1 is solution of:

$$\mathbb{E}_\pi\left(\frac{\kappa - t_1}{\kappa + t_1}\right) = 0$$

So $t_1 \approx \mathbb{E}(\kappa)$.

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$
t_1 is solution of:

$$\mathbb{E}_\pi\left(\frac{\kappa - t_1}{\kappa + t_1}\right) = 0$$

So $t_1 \approx \mathbb{E}(\kappa)$. Also Jensen's inequality $\Rightarrow t_1 \leq \mathbb{E}(\kappa)$

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$
t_1 is solution of:

$$\mathbb{E}_\pi\left(\frac{\kappa - t_1}{\kappa + t_1}\right) = 0$$

So $t_1 \approx \mathbb{E}(\kappa)$. Also Jensen's inequality $\Rightarrow t_1 \leq \mathbb{E}(\kappa)$
$\kappa \sim$Unif. $\kappa_L = 0$, $t_1 \approx 0.397$:

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$
t_1 is solution of:

$$\mathbb{E}_\pi\left(\frac{\kappa - t_1}{\kappa + t_1}\right) = 0$$

So $t_1 \approx \mathbb{E}(\kappa)$. Also Jensen's inequality $\Rightarrow t_1 \leq \mathbb{E}(\kappa)$
$\kappa \sim$Unif. $\kappa_L = 0$, $t_1 \approx 0.397$: for $\log\kappa \sim$Unif., $t_1 = \sqrt{\kappa_L}$

Optimal 'Bayesian' design, $m = 2$

- Need to maximise $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ numerically
- Can find optimal 'Bayesian' two-point design explicitly.
- A bit of an indulgence but... (also often need a lot prior uncertainty to persuade $m > p$)

Two-point 'Bayesian' design

$\eta_1 = \eta_2 = \frac{1}{2}$
$t_2 = \infty$
t_1 is solution of:

$$\mathbb{E}_\pi\left(\frac{\kappa - t_1}{\kappa + t_1}\right) = 0$$

So $t_1 \approx \mathbb{E}(\kappa)$. Also Jensen's inequality $\Rightarrow t_1 \leq \mathbb{E}(\kappa)$
κ ~Unif. $\kappa_L = 0$, $t_1 \approx 0.397$: for $\log\kappa$ ~Unif., $t_1 = \sqrt{\kappa_L}$

Finding Optimal 'Bayesian' Design, $m > 2$

- Search Numerically, successively $m = 3, 4, \ldots$

Finding Optimal 'Bayesian' Design, $m > 2$

- Search Numerically, successively $m = 3, 4, \ldots$
- Search for $t_j \in [0, T]$ and $\eta_j \geq 0$, $\sum \eta_j = 1$

Finding Optimal 'Bayesian' Design, $m > 2$

- Search Numerically, successively $m = 3, 4, \ldots$
- Search for $t_j \in [0, T]$ and $\eta_j \geq 0$, $\sum \eta_j = 1$
- T some arbitrary 'large' scaled concentration - herein $T = 10$

Optimal 'Bayesian' Design

		$\kappa \sim$ Unif	$\kappa_L = 0$
t	0.02	0.39	10
η	0.02	0.49	0.49

Optimal 'Bayesian' Design

		$\kappa \sim$ Unif	$\kappa_L = 0$
t	0.02	0.39	10
η	0.02	0.49	0.49
		$\log \kappa \sim$ Unif	$\kappa_L = 10^{-2}$
t	0.04	0.33	10
η	0.26	0.30	0.44

Optimal 'Bayesian' Design

		$\kappa \sim$ Unif	$\kappa_L = 0$			
t	0.02	0.39	10			
η	0.02	0.49	0.49			
		$\log\kappa \sim$ Unif	$\kappa_L = 10^{-2}$			
t	0.04	0.33	10			
η	0.26	0.30	0.44			
		$\log\kappa \sim$ Unif	$\kappa_L = 10^{-5}$			
t	4.4e-5	4.9e-4	3.8e-3	0.029	0.29	10
η	0.12	0.10	0.11	0.14	0.23	0.30

The point at T

- All designs ascribe some weight (quite a lot usually) to $t = T = 10$

The point at T

- All designs ascribe some weight (quite a lot usually) to $t = T = 10$
- Gives information about $V_{\max}$

The point at T

- All designs ascribe some weight (quite a lot usually) to $t = T = 10$
- Gives information about $V_{\max}$
- Essential even when interest in focussed on K

The point at T

- All designs ascribe some weight (quite a lot usually) to $t = T = 10$
- Gives information about V_{max}
- Essential even when interest in focussed on K
- Applies whatever the prior for V_{max}

The point at T

- All designs ascribe some weight (quite a lot usually) to $t = T = 10$
- Gives information about $V_{\max}$
- Essential even when interest in focussed on K
- Applies whatever the prior for $V_{\max}$
- So what happens if $\pi(V_{\max})$ is very concentrated?

Is the criterion right?

- The criterion $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ was introduced rather casually - what is its provenance?

Is the criterion right?

- The criterion $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ was introduced rather casually - what is its provenance?
- Formal Bayesian design was introduced by Lindley (1956) using utility functions

Is the criterion right?

- The criterion $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ was introduced rather casually - what is its provenance?
- Formal Bayesian design was introduced by Lindley (1956) using utility functions
    - Specific utilities matched to aims of the experiment should be identified

Is the criterion right?

- The criterion $\mathbb{E}_\pi(\log \det(\mathscr{I}))$ was introduced rather casually - what is its provenance?
- Formal Bayesian design was introduced by Lindley (1956) using utility functions
  - Specific utilities matched to aims of the experiment should be identified
  - For inference about a parameter θ Lindley made extensive use of Shannon information

Is the criterion right?

- The criterion $\mathbb{E}_\pi(\log\det(\mathscr{I}))$ was introduced rather casually - what is its provenance?
- Formal Bayesian design was introduced by Lindley (1956) using utility functions
  - Specific utilities matched to aims of the experiment should be identified
  - For inference about a parameter θ Lindley made extensive use of Shannon information
  - This work was later formailised by Bernardo (1979)

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$
- Perform an experiment ξ and observe data $\mathbf{x}$

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$
- Perform an experiment ξ and observe data $\mathbf{x}$
- Information now available about θ is $\mathscr{I}_1 = \mathbb{E}_{\theta|\mathbf{x},\xi}(\log \pi(\theta|\mathbf{x},\xi))$

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$
- Perform an experiment ξ and observe data $\mathbf{x}$
- Information now available about θ is $\mathscr{I}_1 = \mathbb{E}_{\theta|\mathbf{x},\xi}(\log \pi(\theta|\mathbf{x},\xi))$
- Choose ξ to maximise $\mathscr{I}_1 - \mathscr{I}_0$:

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$
- Perform an experiment ξ and observe data $\mathbf{x}$
- Information now available about θ is $\mathscr{I}_1 = \mathbb{E}_{\theta|\mathbf{x},\xi}(\log \pi(\theta|\mathbf{x},\xi))$
- Choose ξ to maximise $\mathscr{I}_1 - \mathscr{I}_0$:but this depends on $\mathbf{x}$

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$
- Perform an experiment ξ and observe data $\mathbf{x}$
- Information now available about θ is $\mathscr{I}_1 = \mathbb{E}_{\theta|\mathbf{x},\xi}(\log \pi(\theta|\mathbf{x},\xi))$
- Choose ξ to maximise $\mathscr{I}_1 - \mathscr{I}_0$:but this depends on $\mathbf{x}$
- So choose ξ to maximise $\mathbb{E}_{\mathbf{x}|\xi}(\mathscr{I}_1 - \mathscr{I}_0)$,

Summary of Lindley's approach

- Information about θ in prior $\pi(\theta)$ is $\mathscr{I}_0 = \mathbb{E}_\theta(\log \pi(\theta))$
- Perform an experiment ξ and observe data $\mathbf{x}$
- Information now available about θ is $\mathscr{I}_1 = \mathbb{E}_{\theta|\mathbf{x},\xi}(\log \pi(\theta|\mathbf{x},\xi))$
- Choose ξ to maximise $\mathscr{I}_1 - \mathscr{I}_0$:but this depends on $\mathbf{x}$
- So choose ξ to maximise $\mathbb{E}_{\mathbf{x}|\xi}(\mathscr{I}_1 - \mathscr{I}_0)$, which amounts to

$$\int \log \pi(\theta|\mathbf{x},\xi)\pi(\theta,\mathbf{x}|\xi)d\theta d\mathbf{x} = \int \log \pi(\theta|\mathbf{x},\xi)\pi(\mathbf{x}|\theta,\xi)\pi(\theta)d\theta d\mathbf{x}$$

Approximating posterior i

- Previous integrals are awkward - some approximations useful

Approximating posterior i

- Previous integrals are awkward - some approximations useful
- Consider only cases where $\pi(\theta|\mathbf{x}, \xi)$ depends on data only through the MLE, $\hat{\theta}$

Approximating posterior i

- Previous integrals are awkward - some approximations useful
- Consider only cases where $\pi(\theta|\mathbf{x},\xi)$ depends on data only through the MLE, $\hat{\theta}$
- Commonly used approximation is $\pi(\theta|\mathbf{x},\xi) = N(\hat{\theta}, \mathscr{I}(\hat{\theta})^{-1})$

Approximating posterior i

- Previous integrals are awkward - some approximations useful
- Consider only cases where $\pi(\theta|\mathbf{x},\xi)$ depends on data only through the MLE, $\hat{\theta}$
- Commonly used approximation is $\pi(\theta|\mathbf{x},\xi) = N(\hat{\theta}, \mathscr{I}(\hat{\theta})^{-1})$
- From this we obtain $\mathbb{E}_{\mathbf{x}|\xi}(\mathscr{J}_1 - \mathscr{J}_0) \approx \frac{1}{2}\mathbb{E}_{\theta}(\log\det\mathscr{I}(\theta))$

Approximating posterior ii

- This approximation for the posterior overlooks the contribution of the prior

Approximating posterior ii

- This approximation for the posterior overlooks the contribution of the prior
- An alternative is $\pi(\theta|\mathbf{x}, \xi) = N(\hat{\theta}, [\mathscr{I}(\hat{\theta}) + R]^{-1})$

Approximating posterior ii

- This approximation for the posterior overlooks the contribution of the prior
- An alternative is $\pi(\theta|\mathbf{x},\xi) = N(\hat{\theta}, [\mathscr{I}(\hat{\theta}) + R]^{-1})$
- R is the prior precision matrix

Approximating posterior ii

- This approximation for the posterior overlooks the contribution of the prior
- An alternative is $\pi(\theta|\mathbf{x},\xi) = N(\hat{\theta}, [\mathscr{I}(\hat{\theta}) + R]^{-1})$
- R is the prior precision matrix
- This leads to $\mathbb{E}_{\mathbf{x}|\xi}(\mathscr{I}_1 - \mathscr{I}_0) \approx \frac{1}{2}\mathbb{E}_\theta[\log\det(\mathscr{I}(\theta) + R)]$

Approximating posterior ii

- This approximation for the posterior overlooks the contribution of the prior
- An alternative is $\pi(\theta|\mathbf{x}, \xi) = N(\hat{\theta}, [\mathscr{I}(\hat{\theta}) + R]^{-1})$
- R is the prior precision matrix
- This leads to $\mathbb{E}_{\mathbf{x}|\xi}(\mathscr{J}_1 - \mathscr{J}_0) \approx \frac{1}{2}\mathbb{E}_\theta[\log\det(\mathscr{I}(\theta) + R)]$

Application to Michaelis Menten example i

- The criterion is now:

$$\mathbb{E}_\theta[\log\det(\mathscr{I}(\theta) + R)] = \mathbb{E}_\theta[\log\det(N\sigma^{-2}\mathscr{M} + R)]$$

Application to Michaelis Menten example i

- The criterion is now:

$$\mathbb{E}_\theta[\log\det(\mathscr{I}(\theta) + R)] = \mathbb{E}_\theta[\log\det(N\sigma^{-2}\mathscr{M} + R)]$$

- Now N and σ^2 do not 'scale out' of the problem

Application to Michaelis Menten example i

- Write $A_u = \sum \eta_j s_j^2/(K+s_j)^{2+u}$ for $u = 0, 1, 2$, so

Application to Michaelis Menten example i

- Write $A_u = \sum \eta_j s_j^2/(K + s_j)^{2+u}$ for $u = 0, 1, 2$, so

$$\mathscr{M} = \begin{pmatrix} A_0 & \\ -V_{\max} A_1 & V_{\max}^2 A_2 \end{pmatrix}$$

- Also we have

$$R = \begin{pmatrix} \text{var}_\pi(V_{\max})^{-1} & \\ 0 & \text{var}_\pi(K)^{-1} \end{pmatrix} = \begin{pmatrix} R_V & \\ 0 & K_U^{-2} R_\kappa \end{pmatrix}$$

Application to Michaelis Menten example ii

Transforming to κ and t_j and evaluating the determinant, the part that determines the design is:

Application to Michaelis Menten example ii

Transforming to κ and t_j and evaluating the determinant, the part that determines the design is:

$$\log\left[\left(\frac{R_V\sigma^2}{N}+A_0\right)\left(\frac{R_\kappa\sigma^2}{NV_{\max}^2}+A_2\right)-A_1^2\right]$$

whereas before it was simply

$$\log\left[A_0A_2-A_1^2\right]$$

Some simplifications

- Assume σ^2 is known - bit of nonsense but allows illustration to proceed

Some simplifications

- Assume σ^2 is known - bit of nonsense but allows illustration to proceed
- Assume prior variance of $V_{\max}$ is σ_V^2 and set $\lambda = \sigma_V^2/\sigma^2$.

Some simplifications

- Assume σ^2 is known - bit of nonsense but allows illustration to proceed
- Assume prior variance of $V_{\max}$ is σ_V^2 and set $\lambda = \sigma_V^2/\sigma^2$.
- Write $\tilde{V} = V_{\max}/\sigma_V$

Some simplifications

- Assume σ^2 is known - bit of nonsense but allows illustration to proceed
- Assume prior variance of $V_{\max}$ is σ_V^2 and set $\lambda = \sigma_V^2/\sigma^2$.
- Write $\tilde{V} = V_{\max}/\sigma_V$
- Need to maximise

$$\mathbb{E}_{\kappa,\tilde{V}} \log\left[\left(\frac{1}{N\lambda} + A_0\right)\left(\frac{1}{N\lambda\tilde{V}^2 \text{var}(\kappa)} + A_2\right) - A_1^2\right]$$

with prior for κ as before and for $\tilde{V} \sim N(V_0, 1)$

Some examples

Take $V_0 = 10$ and $N = 5$ and assume $\log \kappa$ is uniform and $\kappa_L = 0.01$.

Some examples

Take $V_0 = 10$ and $N = 5$ and assume $\log \kappa$ is uniform and $\kappa_L = 0.01$.

λ	**t**	η		
	0.04	0.27	*0.04*	*0.26*
10	0.33	0.31	*0.33*	*0.30*
	10	0.42	*10*	*0.44*

Some examples

Take $V_0 = 10$ and $N = 5$ and assume $\log \kappa$ is uniform and $\kappa_L = 0.01$.

λ	**t**	η		
	0.04	0.27	*0.04*	*0.26*
10	0.33	0.31	*0.33*	*0.30*
	10	0.42	*10*	*0.44*
	0.04	0.33		
1	0.34	0.37		
	10	0.30		

Some examples

Take $V_0 = 10$ and $N = 5$ and assume $\log \kappa$ is uniform and $\kappa_L = 0.01$.

λ	**t**	η		
	0.04	0.27	*0.04*	*0.26*
10	0.33	0.31	*0.33*	*0.30*
	10	0.42	*10*	*0.44*
	0.04	0.33		
1	0.34	0.37		
	10	0.30		
	0.04	0.48		
0.1	5.0	0.52		

Observations

Observations

- Effect of prior is correctly accommodated provided correct objective function is used.

Observations

- Effect of prior is correctly accommodated provided correct objective function is used.
- Need to be careful with what we mean by 'correct'

Observations

- Effect of prior is correctly accommodated provided correct objective function is used.
- Need to be careful with what we mean by 'correct'
- Role of prior, at least strictly, is to allow *your* analysis to update *your* beliefs

Observations

- Effect of prior is correctly accommodated provided correct objective function is used.
- Need to be careful with what we mean by 'correct'
- Role of prior, at least strictly, is to allow *your* analysis to update *your* beliefs
- For design, you are generating data that others may wish to use.

Observations

- Effect of prior is correctly accommodated provided correct objective function is used.
- Need to be careful with what we mean by 'correct'
- Role of prior, at least strictly, is to allow *your* analysis to update *your* beliefs
- For design, you are generating data that others may wish to use. Inappropriately prescriptive priors could lead to data being collected which is of narrow use.

Observations

- Effect of prior is correctly accommodated provided correct objective function is used.
- Need to be careful with what we mean by 'correct'
- Role of prior, at least strictly, is to allow *your* analysis to update *your* beliefs
- For design, you are generating data that others may wish to use. Inappropriately prescriptive priors could lead to data being collected which is of narrow use.
- The issue of a prior for design and a prior for analysis is discussed in Cox and Reid *The Theory of the Design of Experiments*

References

1. Bernardo JM (1979) Expected information as expected utility. *Annals of Statistics*, 7, 686-690.
2. Challoner K and Verdinelli I (1995) Bayesian experimental design: a review, *Statistical Science*, 10, 273-304.
3. Cox DR & Reid N (2000) *The Theory of the Design of Experiments* Chapman and Hall/CRC, Boca Raton.
4. Lindley DV (1956) On a measure of the information provided by an experiment, *Annals of Mathematical Statistics*, 27, 986-1005.
5. Matthews JNS & Allcock GC (2004) Optimal designs for Michaelis Menten kinetic studies. *Statistics in Medicine*, 23, 477-91.