Bayesian Designs for Michaelis-Menten kinetics

John Matthews and Gilly Allcock
Department of Statistics
University of Newcastle upon Tyne

j.n.s.matthews@ncl.ac.uk
References etc. on http://www.mas.ncl.ac.uk/~njnsm/talks/titles.htm

Enzymology

- Many biochemical reactions would, of their own accord, proceed at a rate that is far too slow to be of use.
- Enzymes are natural catalysts which greatly increase the rate of reaction.

$$\begin{array}{c} \text{Substrate} \\ \text{e.g. glucose - 6 - phosphate} \end{array} \xrightarrow{Enzyme \quad (e.g. G6Pase)} \begin{array}{c} \text{Product} \\ \text{e.g. glucose} \end{array}$$

Michaelis-Menten equation

For many enzymes the rate of reaction is determined by the *Michaelis-Menten* equation

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

Here V_{max} is the maximum rate at which substrate is turned into product and K_M is the *Michaelis* parameter, the substrate concentration at which the rate of reaction is 50% of its maximum.

Enzymologists are interested in the values of these parameters, and also in derived quantities such as the specificity constant V_{max}/K_M.

Parameter Estimation

- The enzymologist observes the values of v, v_i, at a series of substrate concentrations, s_i, i=1,..,n.
- Parameters are estimated by fitting the Michaelis-Menten equation to these data
- Will start with the model

$$v_i = \frac{V_{\max} s_i}{K_M + s_i} + e_i$$

with e_i a residual with zero mean and constant variance.

- Substantial history to fitting this model, and also some concerns over the use of this model (Ruppert, Cressie and Carroll, 1989; Nelder, 1991; also Cornish-Bowden 1995)

Design Problem

- How should the experimenter choose the substrate concentrations?
- Some work on this: Currie (1982) in *Biometrics*, also Duggleby (1979) and Endrenyi & Chan (1981) in enzymology literature
- Depends on the aims of the experiment
- Will be assumed that the aim is to estimate the parameter(s) and to do this with maximal precision.
- Will not consider studies where the aim is to differentiate between different types of reaction.

Expected Information matrix

For the above model the expected information matrix is proportional to

$$s^{-2}NM = s^{-2}N\begin{pmatrix} \sum_{j=1}^{m} h_j \dfrac{s_j^2}{(K_M + s_j)^2} & \\ -V_{\max} \sum_{j=1}^{m} h_j \dfrac{s_j^2}{(K_M + s_j)^3} & V_{\max}^2 \sum_{j=1}^{m} h_j \dfrac{s_j^2}{(K_M + s_j)^4} \end{pmatrix}$$

We assume that N observations are made at m distinct substrate concentrations. The number of observations at s_j is Nh_j, where $h_j \geq 0$, $\sum h_j = 1$.

Locally D-optimal design

The log of the determinant of the above can be written as the log of :

$$\Delta = \left(\sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M + s_j)^2} \right) \left(\sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M + s_j)^4} \right) - \left(\sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M + s_j)^3} \right)^2$$

where terms not involving the design points $x = (\boldsymbol{s}, \boldsymbol{h})$ have been omitted

Depends on K_M (though not on V_{max}).

For m=2 writing $y_j = s_j/(K_M + s_j)$ gives the above as

$$h_1 h_2 y_1^2 y_2^2 (y_1 - y_2)^2$$

The optimal design has $h_1 = h_2 = ½$ and $y_1 = ½$ and $y_2 = 1$, i.e. $s_1 = K_M$, $s_2 = \infty$.

(Currie, Duggleby, Endrenyi)

Bayesian D-Optimal design

Find design by maximising $E_{prior}(\log \det(s^{-2}NM))$

Specify knowledge about K_M through a prior.

Objective factors into $f(N)+f(\sigma)+f(V_{max})+f(K_M, \text{design})$

So no need to specify a prior for V_{max}, only marginal for K_M

Convenient to assume prior has finite support on K_L, K_U. These to be specified by investigator.

Some parsimony achieved by scaling: write $s_j = K_U t_j$, $K_M = K_U k$ (with $K_L/K_U = k_L < k < 1$).

Two priors: 1. k uniform over its range 2. $\log k$ uniform over its range.

Optimal Bayesian 2-point design

A bit of an indulgence, but analytical progress can be made here.

Designs all give equal weight to both points.

Larger concentration is at infinity

Smaller concentration t_1 is at the solution to $\mathrm{E}_p\left(\frac{k-t_1}{k+t_1}\right)=0$

An approximate solution is therefore $t_1 = \mathrm{E}_p(k)$, which fits with locally optimal solution. Also, Jensen's inequality shows that in fact $t_1 \leq \mathrm{E}_p(k)$.

For prior 1, t_1 is 0.397 ($k_L = 0$); for prior 2, $t_1 = \sqrt{k_L}$.

Optimal Bayesian designs

- Search numerically for optimal design for $m = 3, 4, \ldots$
- Use NAG software for quadrature and optimisation.
- Search for $0 \le t_j \le T$, and $h_j \ge 0$, $\sum h_j = 1$, where T is just some 'large' scaled concentration, arbitrarily set at 10 (sensitivity to choice can be explored)

Optimal designs

k_L	t								h			
Uniform on k												
0	0.02	0.39	10				0.02	0.49	0.49			
Uniform on log k												
10^{-2}	0.04	0.33	10				0.26	0.30	0.44			
10^{-5}	4.4E-5	4.9E-4	3.8E-3	2.9E-2	2.9E-1	10	0.12	0.10	0.11	0.14	0.23	0.30

All of these can be confirmed to be optimal from the ‘derivative’ plots

$d(t) = \mathrm{E}_p\,[trM(x^*)^{-1}m(t,k)]$, is ≤ 2 if x^* is optimal and = 2 only at points in x^*

Alternative criteria

There may be interest in simply finding designs which are good for estimating K_M or alternatively V_{max}/K_M.

For former, criterion is to minimise

$$\log\left[\sum_{j=1}^{m} h_j \frac{t_j^2}{(k+t_j)^2}\right] - \log f(k;x,m)$$

where $f(.;.)$ denotes the determinant in the preceding criterion.

Locally optimum design (k=1), gives $s_1 = K_M/\sqrt{2}$, $s_2 = \infty$; $h_1 = 1/\sqrt{2}$.

For specificity ratio, V_{max}/K_M optimal designs are based on

$$\log\left[\sum_{j=1}^{m} h_j \frac{t_j^4}{(k+t_j)^4}\right]$$

$$-\log\left[\sum_{j=1}^{m} h_j \frac{t_j^2}{(k+t_j)^2} \sum_{j=1}^{m} h_j \frac{t_j^4}{(k+t_j)^4} - \left(\sum_{j=1}^{m} h_j \frac{t_j^3}{(k+t_j)^3}\right)^2\right]$$

Locally optimum design has same design points as for K_M but different weights.

s_j	$K_M/\sqrt{2}$	∞
K_M	$1/\sqrt{2}$	$1-1/\sqrt{2}$
V_{max}/K_M	$\frac{1}{2}(1+1/\sqrt{2})$	$\frac{1}{2}(1-1/\sqrt{2})$

Optimal Designs

k_L	t						h					
Optimal designs for $\mathrm{var}(\hat{K}_M)$												
10^{-2}	0.029	0.269	10.0				0.451	0.319	0.230			
10^{-5}	4.0E-5	4.5E-4	3.6E-3	2.7E-2	2.4E-1	10	0.22	0.17	0.16	0.16	0.18	0.12

- Designs need greater weight at lower concentrations than for D-optimal designs.
- Intuitively reasonable as the relative importance of information about V_{max} is less important.

Why the point at 10?

- In all designs found so far, some weight has been given to a point at the upper limit of the range for the (scaled) substrate concentrations.
- This gives information about $V_{\max}$: essential even when interest is focussed solely on K_M.
- Also, designs apply to all priors on ($V_{\max}$, K_M), including those with very specific prior knowledge about $V_{\max}$.
- If there is good prior knowledge about $V_{\max}$, why the point at 10?

Answer is that criterion

$$E_p\left(\log\det\left[s^{-2}N\begin{pmatrix} \sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M+s_j)^2} & \\ -V_{\max}\sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M+s_j)^3} & V_{\max}^2 \sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M+s_j)^4} \end{pmatrix}\right]\right)$$

does not take prior information into account in the *analysis*. To do so requires criterion to be modified to:

$$E_p\left(\log\det\left[s^{-2}N\begin{pmatrix} \sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M+s_j)^2} & \\ -V_{\max}\sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M+s_j)^3} & V_{\max}^2 \sum_{j=1}^{m} h_j \frac{s_j^2}{(K_M+s_j)^4} \end{pmatrix}+R\right]\right)$$

R^{-1} being the dispersion matrix of the prior.

Prior Precision Matrix, *R*

Reasonable to take the priors for $V_{\max}$ and K_M to be independent.

$$R = \begin{pmatrix} 1/\operatorname{var}(V_{\max}) & \\ 0 & 1/\operatorname{var}(K_M) \end{pmatrix} = \begin{pmatrix} s_V^{-2} & \\ 0 & K_U^{-2}\operatorname{var}(k)^{-1} \end{pmatrix}$$

Optimal design now depends on s and N. However, write R^* as:

$$R^* = \frac{s^2}{N} R = \begin{pmatrix} 1/(Nl) & \\ 0 & s^2/(N\operatorname{var}(K_M)) \end{pmatrix} = \begin{pmatrix} 1/(Nl) & \\ 0 & s^2 K_U^{-2}\operatorname{var}(k)^{-1}/N \end{pmatrix}$$

where $l = s_V^2 \big/ s^2$ is the prior variance of $V_{\max}$ in units of the RMS. New criterion is expectation over prior of log of

$$\Delta = \left(\sum_{j=1}^{m} h_j \frac{t_j^2}{(k+t_j)^2} + \frac{1}{Nl} \right)\left(\sum_{j=1}^{m} h_j \frac{t_j^2}{(k+t_j)^4} + \frac{1}{Nl\tilde{V}^2\operatorname{var}(k)} \right) - \left(\sum_{j=1}^{m} h_j \frac{t_j^2}{(k+t_j)^3} \right)^2$$

where $\tilde{V}$ is $V_{\max}$ scaled by its prior SD.

Prior specification

Prior for V_{max} is $N(V_0 s_V, s_V^2)$, {so for $\tilde{V}$ is $N(V_0,1)$}.

Prior for k is either the prior of the associated uniform disn. or improper, var(k)$^{-1}$=0.

Note that if improper prior used for k then objective function does not depend on V, except through 1/(Nl), so expectation is a one-dimensional integral.

Designs obtained for Improper Prior

All have N=5

$\kappa = 0.01$		
t	***h***	***l***
0.036	0.27	10
0.33	0.31	
10	0.42	
0.036	0.33	1
0.34	0.37	
10	0.31	
0.037	0.48	0.1
5.0	0.52	

$\kappa = 0.001$		
t	***h***	***l***
0.004	0.20	10
0.04	0.16	
0.30	0.28	
10	0.37	
0.004	0.23	1
0.04	0.20	
0.32	0.33	
10	0.24	
0.006	0.46	0.1
0.18	0.54	

A Glimpse of other error models

Ruppert *et al*. (1989) discussed constant variance assumption and a weighting/transformation approach.

Nelder (1991) suggested application of extended quasi-likelihood to explore models with $Var(y) = s^{2} m^{V}$ with a data-determined value for z (in Nelder's example a value between 1 and 2 was obtained)

We explore the cases z = 1 and 2

Information matrix is $s^{-2} \dfrac{\partial m}{\partial b}^{T} diag(m_i^{V}) \dfrac{\partial m}{\partial b}$

Information matrix

$$
\begin{pmatrix}
V_{\max}^{-V} \sum_{j=1}^{m} h_j \dfrac{s_j^{2-V}}{(K_M + s_j)^{2-V}} & \\
-V_{\max}^{1-V} \sum_{j=1}^{m} h_j \dfrac{s_j^{2-V}}{(K_M + s_j)^{3-V}} & V_{\max}^{2-V} \sum_{j=1}^{m} h_j \dfrac{s_j^{2-V}}{(K_M + s_j)^{4-V}}
\end{pmatrix}
$$

D-Optimal designs for z = 1 and searching using T = 10 gives

k_L	***t***						***h***			
Uniform on k										
0		0.16	10			0.5	0.5			
Uniform on log k										
10^{-2}	0.02	0.16	10		0.32	0.22	0.46			
10^{-5}	3.0E-5	5.7E-4	7.7E-3	1.1E-1	10	0.16	0.13	0.15	0.23	0.33

For z = 2 determinant of information matrix becomes

$$\sum_{j=1}^{m} h_j \sum_{j=1}^{m} h_j \frac{1}{(k+t_j)^2} - \left(\sum_{j=1}^{m} h_j \frac{1}{(k+t_j)} \right)^2 = \sum_{j=1}^{m} h_j (z_j - \bar{z}_w)^2$$

$$z_j = (k+t_j)^{-1}$$

This is maximised by a two-point design, with concentrations at 0 and T, equally weighted (for any prior)

Some Efficiencies

Duggleby suggested equal numbers of observations at each of

$$K_M/4,\ K_M/2,\ K_M,\ 2K_M,\ 4K_M.$$

What is efficiency of this design?

We have a prior for k, and it seems reasonable to use the mean of the prior to compare Bayesian designs with Duggleby's design. Scaling this suggests comparing optimal designs with t_j's equal to:

$$\bar{k}/4,\ \bar{k}/2,\ \bar{k},\ 2\bar{k},\ 4\bar{k}$$

Criterion is $\exp(E/p)$ where p is no. parameters and

$$E = \mathrm{E}_p\left[\log\det(M(x^*)^{-1})\right] - \mathrm{E}_p\left[\log\det(M(x_{Duggleby})^{-1})\right]$$

k_L	D-optimal	K_M
Uniform		
0	0.51	0.4
Log-Uniform		
10^{-2}	0.52	0.38
10^{-5}	0.26	0.08

Amended designs

k_L	D-optimal	K_M
Uniform		
0	0.81	0.73
Log-Uniform		
10^{-2}	0.85	0.74
10^{-5}	0.39	0.14

General remarks

- Optimal designs can have few points
- Reliant on idea that there is a single purpose behind the study
- Using a prior distribution increases the number of points in the design, as a ‘hedge’ against the uncertainty around the values of the parameters