Bayes Risk as an Alternative to Fisher Information in Determining Experimental Designs for Nonparametric Models

David S. Bayard, Roger Jelliffe, and Michael Neely

Laboratory of Applied Pharmacokinetics
Children's Hospital Los Angeles
Los Angeles, CA, USA

Supported by NIH Grants GM 068968 and HD 070886

Methods of Optimal Design

- **D-optimality** (traditional) maximizes the determinant of the Fisher Information Matrix (Fedorov 1972 [20], Silvey 1980 [19])
  - Intended for parameter estimation problems [11][12]
  - Modifications for robustness: ED, EID, API, etc.

- **Multiple Model Optimal (MMOpt) Design** (new) minimizes an overbound on the Bayes risk
  - Can be evaluated without numerical integration or Monte Carlo simulation/stochastic analysis
  - Intended for multiple model (MM) parametrizations which form the basis of the USC *BestDose* software
    - The multiple model representation corresponds to the support points in the nonparametric population model
  - Based on recent theoretical overbound on Bayes Risk (Blackmore et. al. 2008 [4])

- **Dynamic Model and Measurements**

  $\dot{x}(t) = f(x(t), d(t), \theta)$ State x, Input d, Parameter $\theta \in R^p$

  $\eta_i = h(x(t_i), \theta)$, System output at time t_i

  $y_i = \eta_i + \sigma_i n_i$, Noisy measurement at time t_i

  $n_i \sim N(0, 1)$, Gaussian measurement noise

  $\xi = \{t_1, ..., t_m\}$, Experiment design (optimal sampling)

- **D-Optimal Design**

  $\max_\xi \left|M\right|$

  where the Fisher Information Matrix M is given by,

  $$M(\theta, \xi) = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \left[\frac{\partial \eta_i}{\partial \theta} \frac{\partial \eta_i}{\partial \theta^T} \right] \Bigg|_{\theta = \overline{\theta}}$$

- Herein, $M(\theta, \xi)$ is assumed to be a function of θ

  (i.e., nonlinear problems)

D-Optimal Design for Nonlinear Problems

- D-Optimal design tells you where to place samples so that they are most sensitive to small changes in model parameter values
  - max |M|, where M is Fisher Information Matrix, and |(.)| is determinant
  - Useful tool has become standard in design of clinical experiments
- Problem with D-optimal design - Circular reasoning (Silvey 1980):
  - You need to know the true patient’s parameters before D-optimal design can be used to compute the best experiment to find these same parameters

- To robustify D-optimal design, an expectation is taken with respect to prior information (c.f., Chaloner [13]), Pronzato [14][15], Tod [16], D’Argenio [17]),

  ED: max E{ |M| }

  EID: min E{ 1/|M| }

  ELD (or API): max E{ log|M| }

- MMopt comparison will be made here relative to ED, EID and API. Not exhaustive - other experiment design criteria are possible (cf., Chaloner [13], Merle and Mentre [18])

Definition of ED, EID, API

- **Robust D-Optimal Designs**

  ED: $\arg\max_\xi \; E_\theta \left(\left|M\right| \right)$

  EID: $\arg\min_\xi \; E_\theta \left(\frac{1}{|M|} \right)$

  API: $\arg\max_\xi \; E_\theta \left(\, log\left|M\right| \right)$

  where,

  $\theta \in R^p$ - Parameter Vector

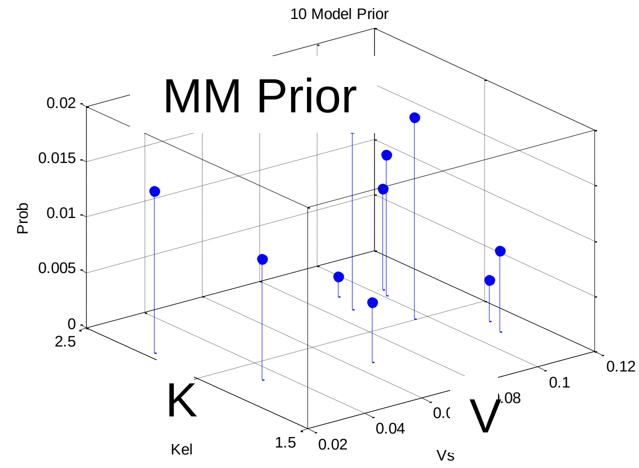
  $\eta = \{t_1, ..., t_m\}$ - Experiment design

  M - Fisher Information Matrix

- All above design metrics require Fisher Matrix M to be nonsingular, and hence **require at least p samples to be taken**, where p=# parameters

Multiple Model Optimal Design

- USC *BestDose* software [3] is based on a multiple model (MM) parametrization of the Bayesian prior (i.e., discrete support points in the nonparametric population model)
  - Nonparametric Maximum Likelihood (NPML) estimation of a population model has the form of a MM prior (Mallet [5], Lindsay [6]).
  - Software for population NPML modeling is available, e.g., NPEM (Schumitzky [7][11]), NPAG (Leary [8], Baek [9]), USC*PACK (Jelliffe [10]).



- Experiment design for MM (i.e., discrete) models is a subject found in classification theory
  - How do we sample the patient to find out which support point he best corresponds to?
  - Classifying patients is fundamentally different from trying to estimate patient's parameters

- Treating MM experiment design in the context of classification theory leads to the mathematical problem of minimizing Bayes risk (Duda et. al. [21])

Multiple Model Optimal Design (MMOpt)

- **Bayes Rule**

  $p(H_i|y,u) = \frac{p(y|H_i,u)p(H_i)}{p(y|u)}, \quad i = 1, ..., N$

  H_i - Hypothesis that model i corresponds to true subject

  u - Experiment design variable (to be optimized over)

- **Design Rule for MM Classifier**

  $p(H_j|y,u) = \max_i\{p(H_i|y,u)\}$ implies

  1. H_j is classified as TRUE
     (i.e., j'th model is classified as true subject)
  2. H_i for $i \neq j$ is classified as FALSE

- **Design Regions**

  MM classifier breaks y into N regions $R_i,\ i = 1, ..., N$
  such that H_j is classifed as TRUE when $y \in R_j$.

Multiple Model Experiment Design (MMED) (Cont'd)

- **Bayes Risk (i.e., Probability of MM Classifier Being Wrong)**

  $P(error) = \sum_i^N \sum_{j \neq i}^N P(y \in R_j, H_i | u)$

  $= \sum_i^N \sum_{j \neq i}^N P(y \in R_j | H_i) p(H_i)$

  $= \sum_i^N \sum_{j \neq i}^N \int_{R_j} P(y \in R_j | H_i) p(H_i) dy$

- **Result: (Blackmore et. al. 2008)**

  When performing hypothesis selection between an arbitrary number of hypotheses, for Gaussian observation distributions such that $p(Y|H_i) = N(\mu(i), \Sigma(i)),\ i = 1, ..., N$, the Bayes Risk is upper bounded as follows:

  $P(error) \leq \sum_i^N \sum_{j \neq i}^N P(H_i)^{\frac{1}{2}} P(H_j)^{\frac{1}{2}} e^{-k(i,j)}$

  where,

  $k(i,j) = \frac{1}{4}(\mu(j) - \mu(i))^T \left(\Sigma(i) + \Sigma(j) \right)^{-1} (\mu(j) - \mu(i)) + \frac{1}{2} \ln \frac{|\frac{1}{2}(\Sigma(i) + \Sigma(j))|}{\sqrt{|\Sigma(i)||\Sigma(j)|}}$

- In short, the MMED experiment design maximizes the probability that the true subject will be correctly identified (classified) among competing models

Simple 2-Model Example

- **Simple 2-Model Example**

$$y_i = \eta(t_i, a) + \sigma n_i$$
$$\eta(t, a) = e^{-at}$$
$$a = a_1 = 2 \text{ (fast model)}$$
$$a = a_2 = 0.25 \text{ (slow model)}$$
$$\sigma = 0.3$$
Prior: $p_1 = .5$, $p_2 = .5$

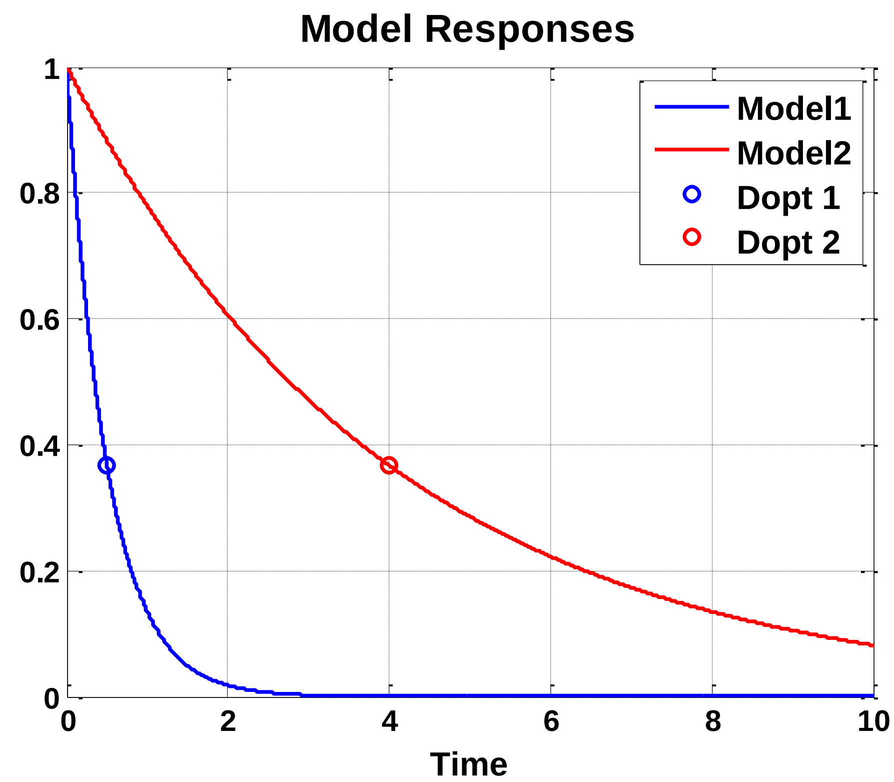
- **Fisher Information Matrix** M

$$\frac{\partial \eta}{\partial a} = -te^{-at}$$
$$M(a) = \frac{1}{\sigma^2}(te^{-at})^2$$
$$|M(a)| = t^2 e^{-2at}/\sigma^2$$

- **Individual D-Optimal Designs**

Fast Model: $t = 1/a_1 = .5$

Slow Model: $t = 1/a_2 = 4$



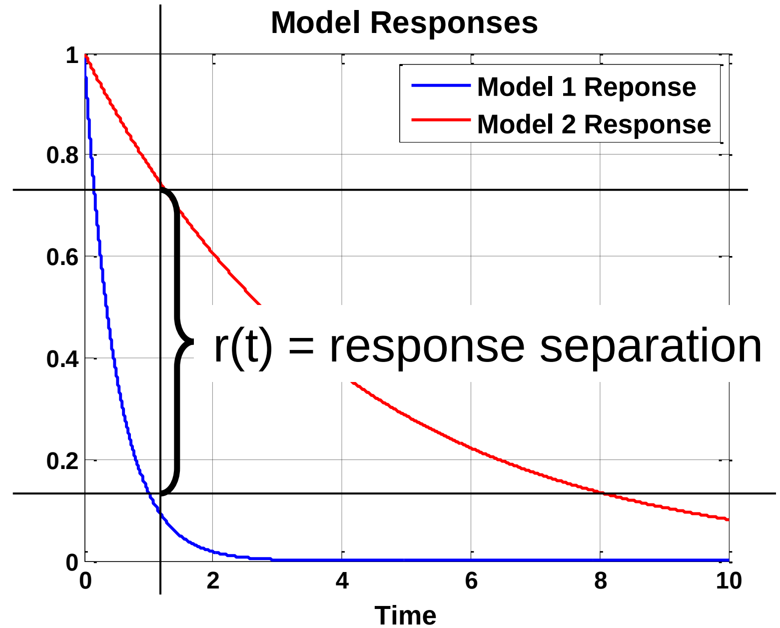
Model Response Separation r(t)

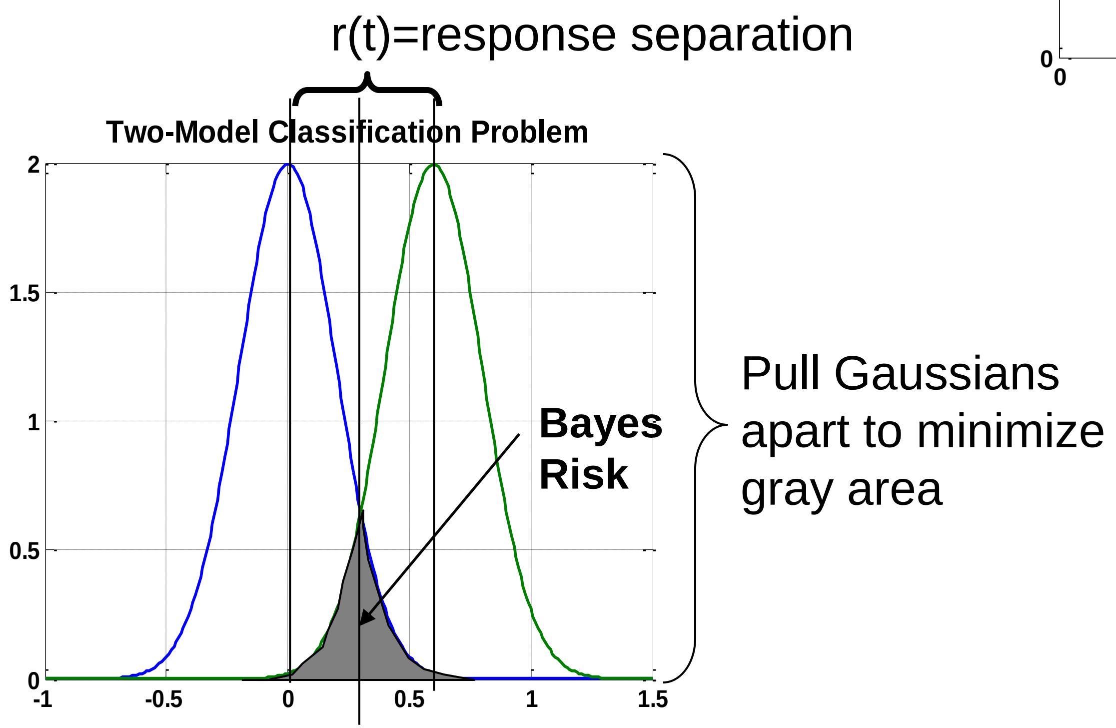
• **Model Response Separation r(t) is the separation between two model responses at a given time t**

$$r(t) = |\eta(t, a_1) - \eta(t, a_2)|$$

•**Defines natural statistic for discriminating between two models**

• **Bayes Risk is shown in gray area**





• Bayes Risk (gray area) decreases as response separation r(t) increases

• Models are best discriminated by sampling at a time t that maximizes r(t)

Bayes Optimal Design

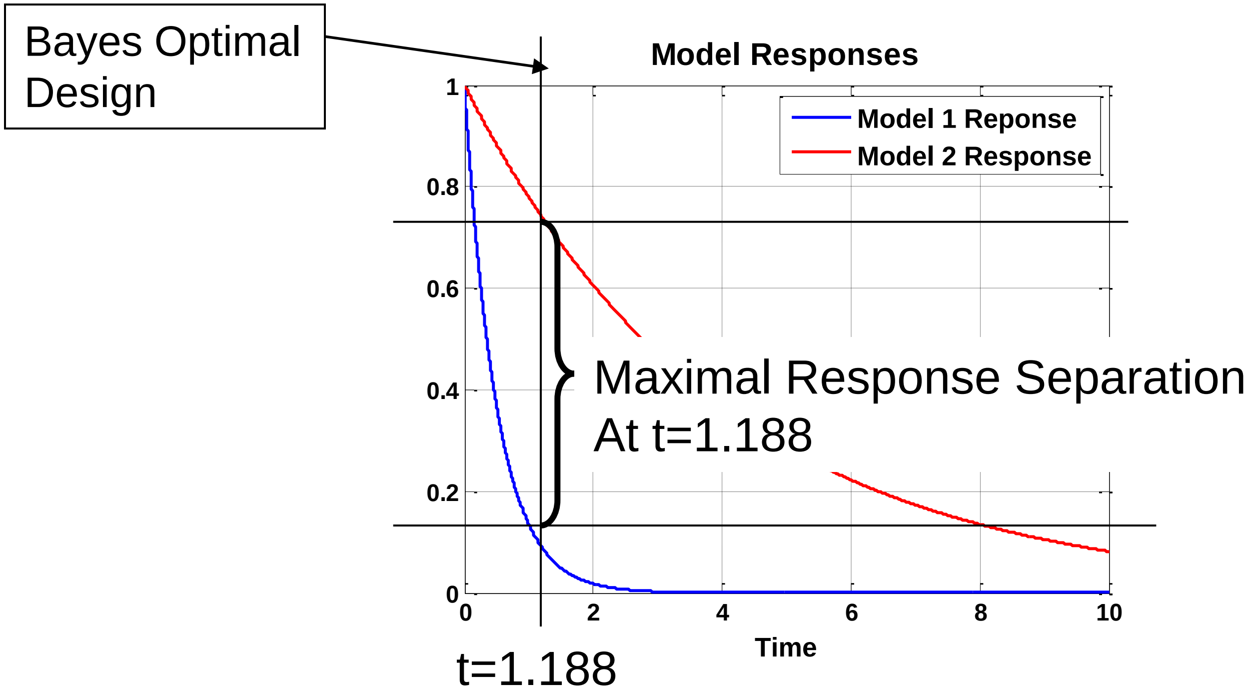
- Bayes Optimal Design minimizes Bayes Risk
- Bayes Risk = Prob(Type I error)+ Prob(Type II error)

$$= p_1 \int_{t_{cross}}^{\infty} p(y|a_1)dy + p_2 \int_{-\infty}^{t_{cross}} p(y|a_2)dy = 1 - \int_{-\infty}^{r(t)/2} N(0,\sigma^2)dy$$

where $r(t) = |\eta(t,a_1) - \eta(t,a_2)|$ is <u>response separation</u>

- Bayes Risk decreases monotonically with increasing $r(t)$
  - Minimized at time of maximal response separation $r(t)$

$$\arg\max_t \; r(t) = |\eta(t,a_1) - \eta(t,a_1)| = 1.188$$



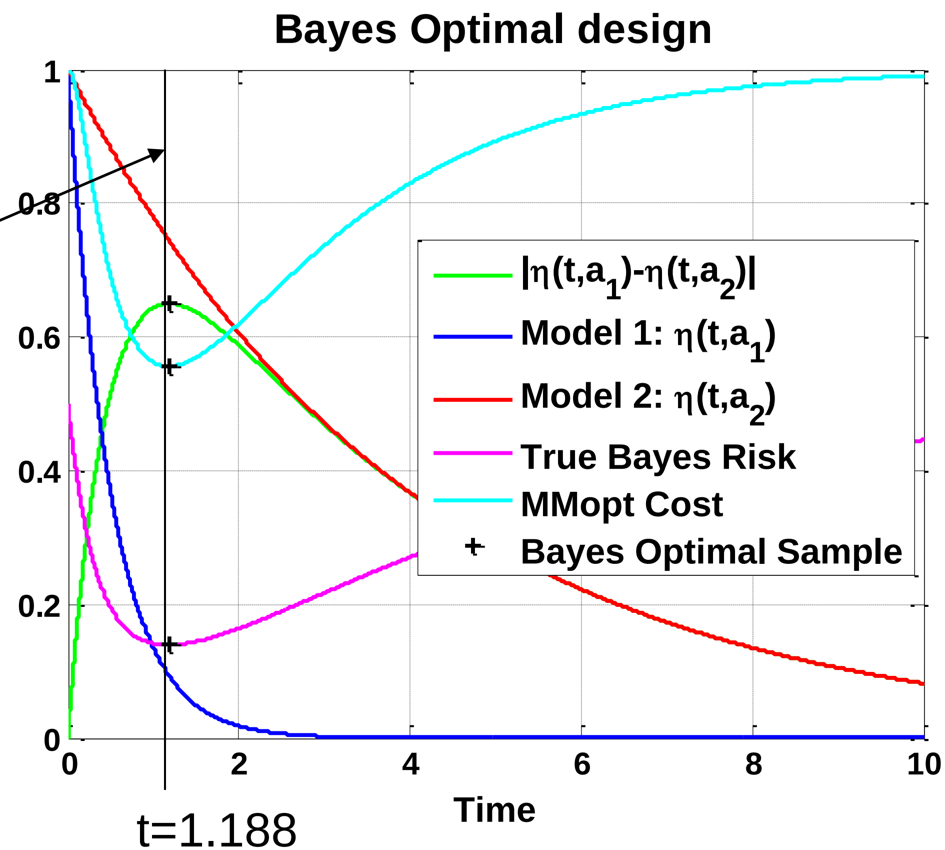
MMopt for 2-Model Example

$$\text{MMopt} = \arg\min_t \; p_1^{\frac{1}{2}} p_2^{\frac{1}{2}} e^{-k(1,2)} + p_2^{\frac{1}{2}} p_1^{\frac{1}{2}} e^{-k(2,1)} = \arg\min_t \; e^{-(\eta(t,a_1)-\eta(t,a_2))^2/(8\sigma^2)}$$

- Take negative-log since monotonic: $\text{MMopt} = \arg\max_t |\eta(t,a_1) - \eta(t,a_2)|$
- MMopt directly maximizes $r(t) = |\eta(t,a_1) - \eta(t,a_1)|$ response separation
- Hence, MMopt sample is at $t = 1.188$ which is same as Bayes Optimal Design

For 2-Model problem, MMopt gives identical solution as Bayes Optimal Design at t=1.188

Both designs maximize response separation (green curve)



ED Design for 2-Model Example

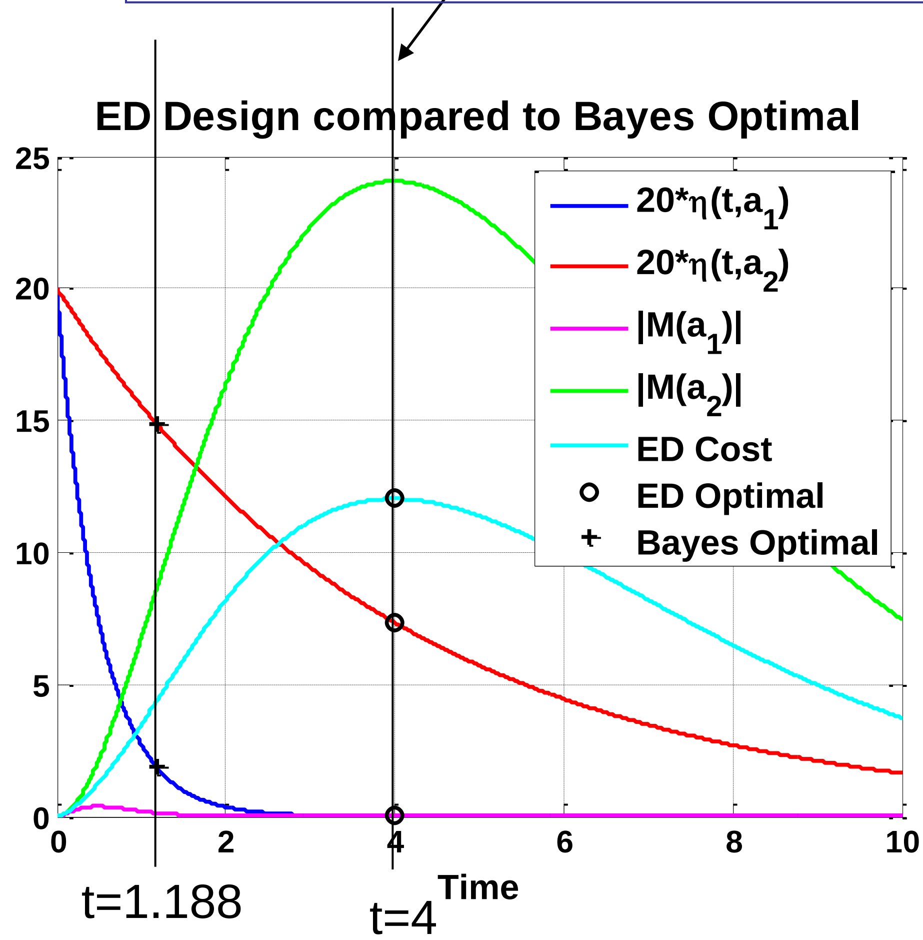
- **ED Optimal Design**

$$\text{ED information} = E[|M|]$$
$$= .5|M(a_1)| + .5|M(a_2)|$$
$$= .5\left(\frac{t^2 e^{-2a_1 t}}{\sigma^2}\right) + .5\left(\frac{t^2 e^{-2a_2 t}}{\sigma^2}\right)$$

ED Criterion is blind to model response separation:

$$\eta(t, a_1) - \eta(t, a_2)$$
$$= e^{-a_1 t} - e^{-a_2 t}$$

ED design at t=4 sec is very different from Bayes optimal design of t=1.188



Note: model responses have been rescaled to enable showing everything on one plot

2-Model Results

Design Metric	Sample Time	Bayes Risk
Bayes Optimal	1.188	0.1393
MMOpt	1.188	0.1393
ED	4.000	0.2701
EID	0.5600	0.1827
API	0.8890	0.1462

- MMopt has smallest Bayes Risk of all designs studied
- *Observation:* ED metric is blind to the underlying classification problem
  - Misses importance of maximizing the response difference $|\eta(t, a_1) - \eta(t, a_2)|$, which requires analyzing model responses jointly.
- Above observation applies to ED, EID, ELD, and any other information metric that simply takes an expectation over individual model information.
- As desired for classification, MMopt directly maximizes response difference $|\eta(t, a_1) - \eta(t, a_2)|$ and agrees exactly with Bayes optimal design for this problem

2-Models with Close Parameters

- **2-Model with Close Dynamics**

  $y_i = \eta(t_i, a) + \sigma n_i$

  $\eta(t, a) = e^{-at}$

  $a = a_1 = 1.1$ (fast model)

  $a = a_2 = 1$ (slow model)

  $\sigma = 0.3$

  Prior: $p_1 = .5$, $p_2 = .5$

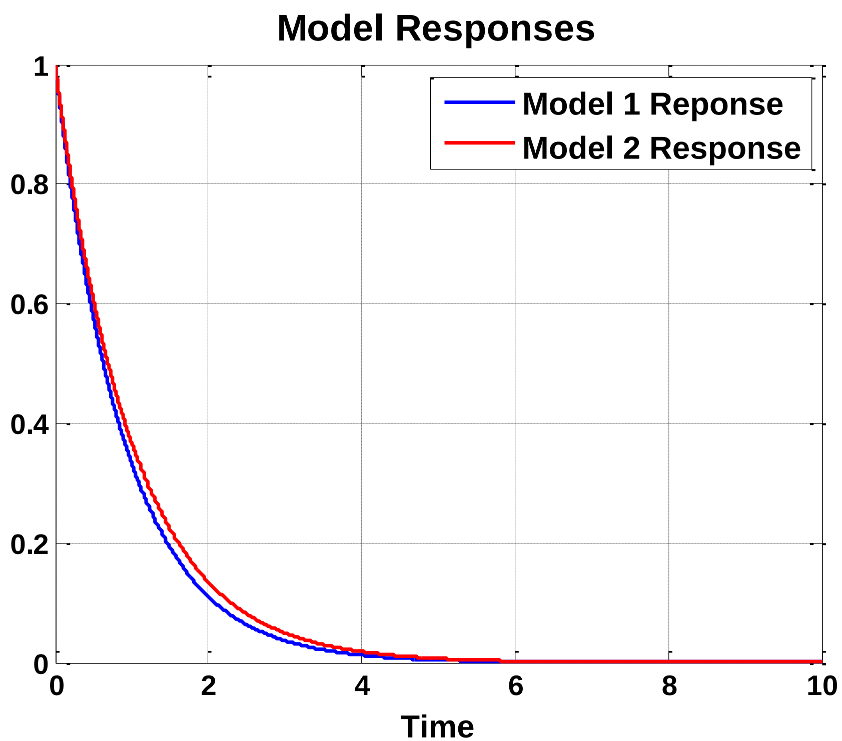
- **Individual D-Optimal Designs**

  Fast Model: $t = 1/a_1 = 0.9091$

  Slow Model: $t = 1/a_2 = 1$

  A-prior Mean Model: $t = 1/\overline{a} = 0.9524$

  $\overline{a} = (a_1 + a_2)/2$



Results for 2-Models with Close Parameters

Design Metric	Sample Time	Bayes Risk
Bayes Optimal	**0.9530**	**0.4767**
MMOpt	**0.9530**	**0.4767**
ED	**0.9570**	**0.4767**
EID	**0.9480**	**0.4767**
API	**0.9520**	**0.4767**
D-Optimal	**0.9524**	**0.4767**

- MMopt agrees exactly with Bayes optimal design for this problem
- Designs (ED, EID, API, D-Optimal) are close to Bayes Optimal design
- Not coincidence: for two model responses close enough to differ by a first-order approximation in a Taylor expansion:
  - Maximizing Model response separation is similar to maximizing Fisher information
  - Goals of classification and parameter estimation are very similar

- Result (see Appendix): For the 2-Model example, the D-Optimal design (designed for *a-prior* mean $\overline{a} = (a_1 + a_2)/2$) approaches the MMopt design (and Bayes Optimal Design) asymptotically as the *a-prior* parameter uncertainty $\Delta a = a_2 - a_1$ becomes small

4-Model Example

- Example where Bayes Optimal Design is too difficult to compute
- Demonstrates usefulness of MMopt design

- Two-Parameters a, b

$$y_i = \eta(t_i, a, b) + \sigma n_i$$

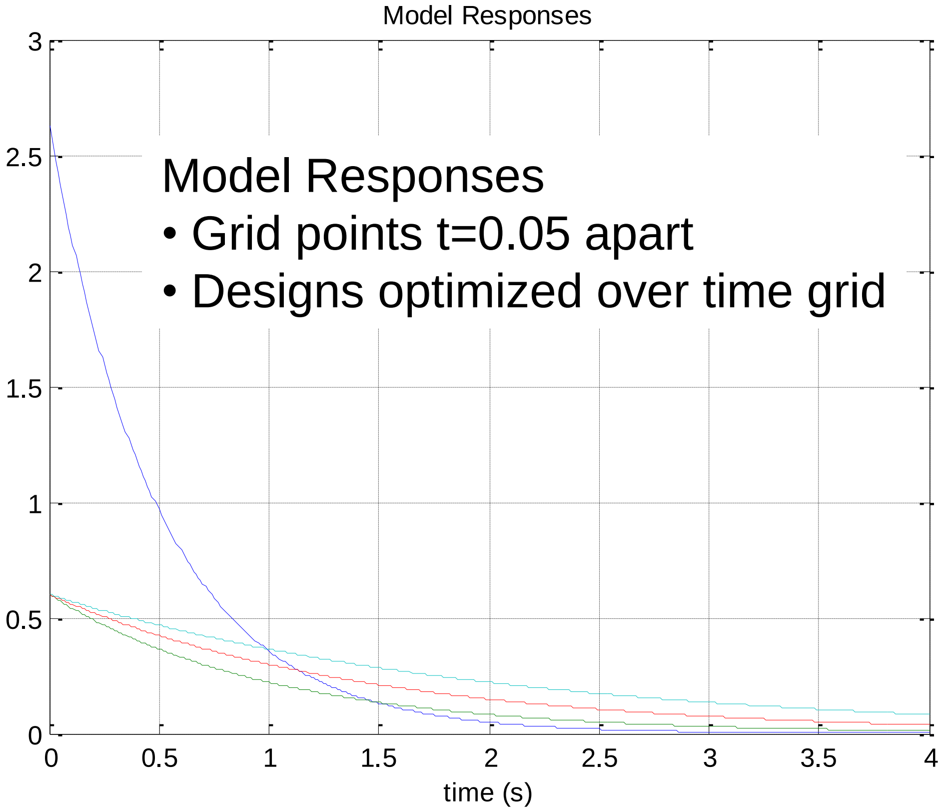
$$\eta(t, a, b) = be^{-at}$$

$$n_i \sim N(0, \sigma^2)$$

$$\sigma = 0.1$$

- Prior: $p_i = .25, \quad i = 1, ..., 4$

Model Parameters		
#	a	b
1	2	2.625
2	1	0.6
3	0.7	0.6
4	0.5	0.6



4-Model Results

Design Metric	2-Sample Times		Bayes Risk	Bayes Risk 99%Conf *
MMOpt	0.45	1.4	0.32839	+/- 0.00070
ED	0	0.8	0.37028	+/- 0.00070
EID	0	1	0.36044	+/- 0.00072
API	0	0.95	0.36234	+/- 0.00072

Design Metric	3-Sample Times			Bayes Risk	Bayes Risk 99% Conf *
MMOpt	0.45	1.4	1.4	0.28065	+/- 0.00067
ED	0	0.7	0.9	0.32048	+/- 0.00067
EID	0	0	1	0.36034	+/- 0.00072
API	0	0.85	0.105	0.3099	+/- 0.00069

- MMopt has smallest Bayes Risk of all designs studied
- Bayes optimal design is too difficult to compute for this problem

* evaluated based on Monte Carlo analysis using 1,400,000 runs per estimate

Experiment Design under Reparametrization

- **Typical PK Parameters**

  V - Volume of Distribution

  K - Elimination Rate

  C - Clearance ($C = VK$)

- **PK Parametrization 1: Volume and Elimination** (V, K)

  $\dot{x} = -Kx + d,$ Drug amount x, Dose d

  $\eta_i = \frac{x(t_i)}{V},$ Drug concentration at time t_i

- **PK Parametrization 2: Volume and Clearance** (V, C)

  $\dot{x} = -(C/V)x + d,$ Drug amount x, Dose d

  $\eta_i = \frac{x(t_i)}{V},$ Drug concentration at time t_i

- **Experiment Design under Nonlinear Reparametrization**

- ED and EID designs <u>are NOT invariant</u> under nonlinear reparametrization (see Appendix)

  - In PK example above, ED and EID give <u>different designs</u> depending on whether the (V, K) or (V, C) parametrization is used

- API and MMopt designs are <u>invariant</u> under nonlinear reparametrization (see Appendix)

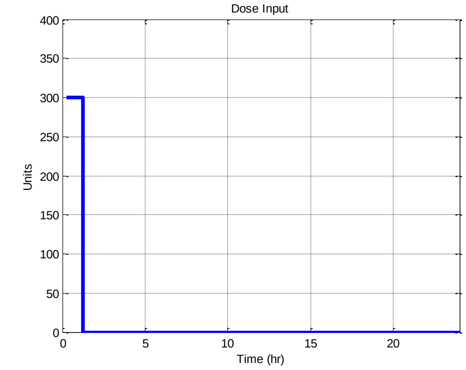
  - "log" in API metric enables invariance (decouples expectation of product into sum)
  - MMopt uses model responses directly rather than forming Fisher Matrix

PK Population Model with 10 Multiple Model Points - First-Order PK Model

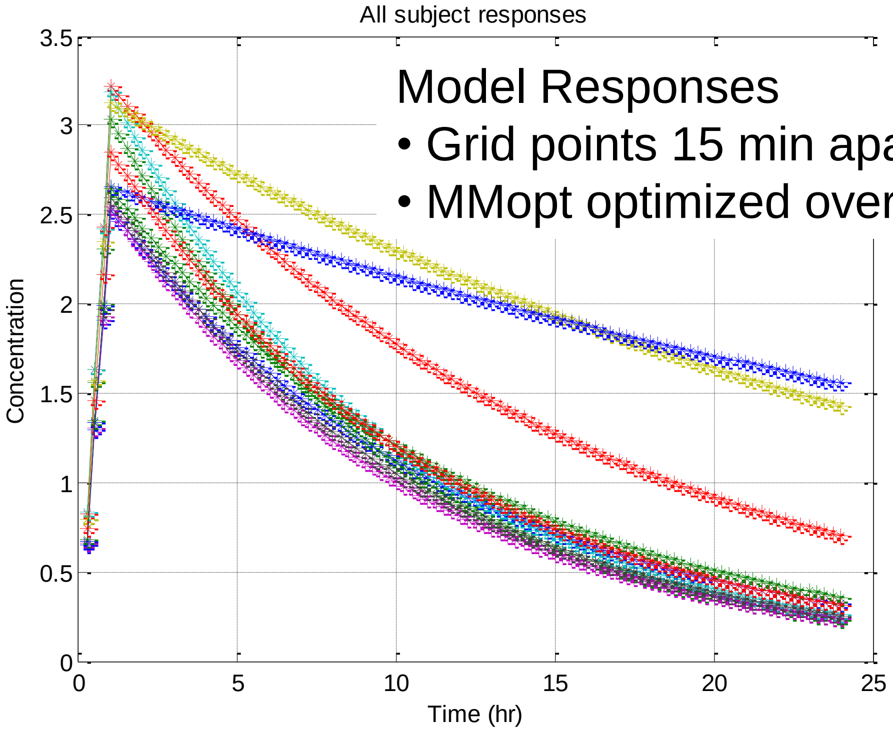
- **First-Order Model**

$$\dot{x} = -Kx + d$$
$$\eta_i = \frac{x(t_i)}{V}$$
$$y_i = \eta_i + \sigma_i n_i$$
$$n_i \sim N(0,1)$$
$$\sigma_i = 0.1$$

Dose input = 300 units for 1 hour, starting at time 0



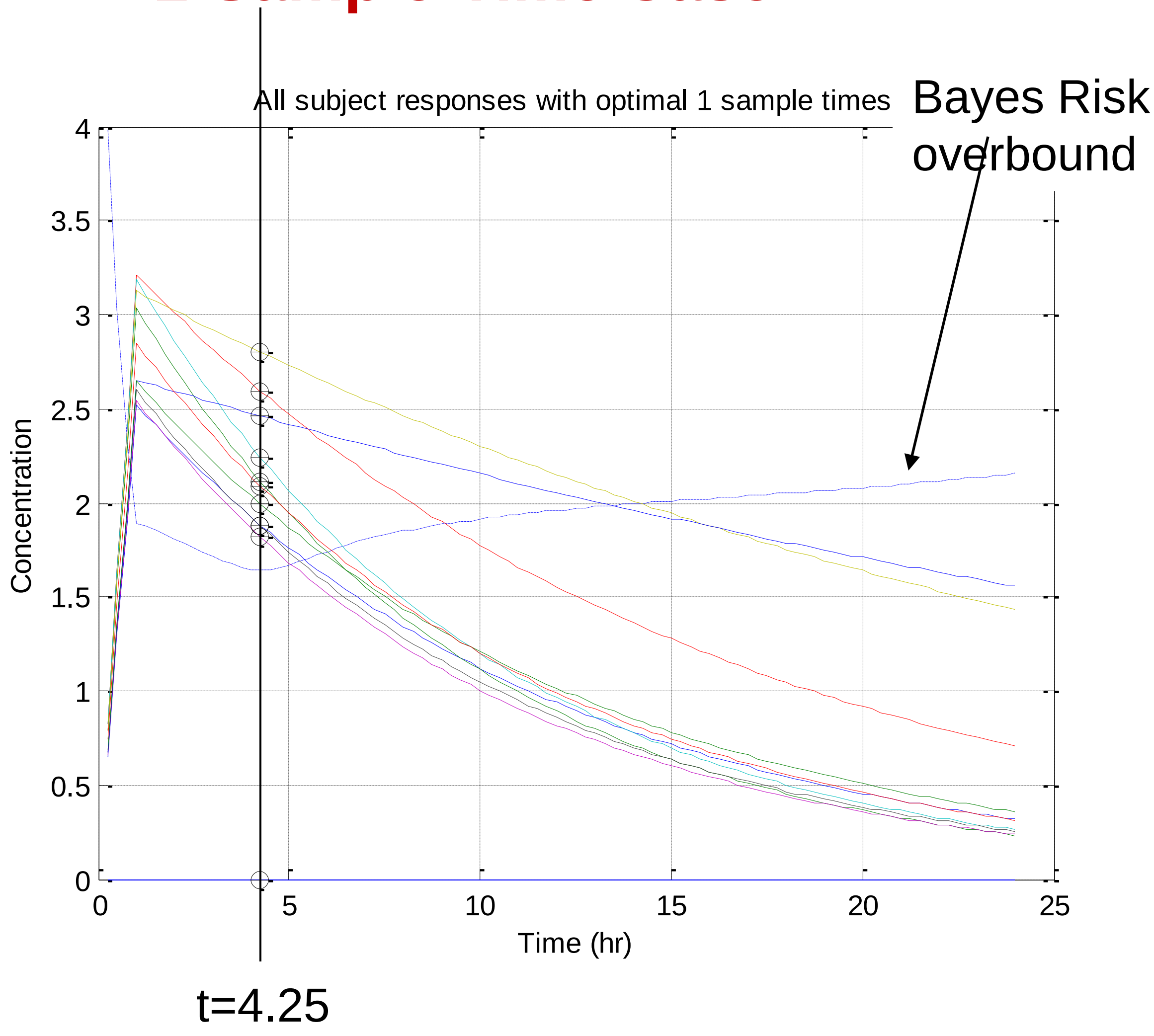
Model Parameters		
#	K	V
1	0.090088	113.7451
2	0.111611	93.4326
3	0.066074	90.2832
4	0.108604	89.2334
5	0.103047	112.1093
6	0.033965	94.3847
7	0.100859	109.8633
8	0.023174	111.7920
9	0.087041	108.6670
10	0.095996	100.3418



Model Responses
- Grid points 15 min apart
- MMopt optimized over time grid

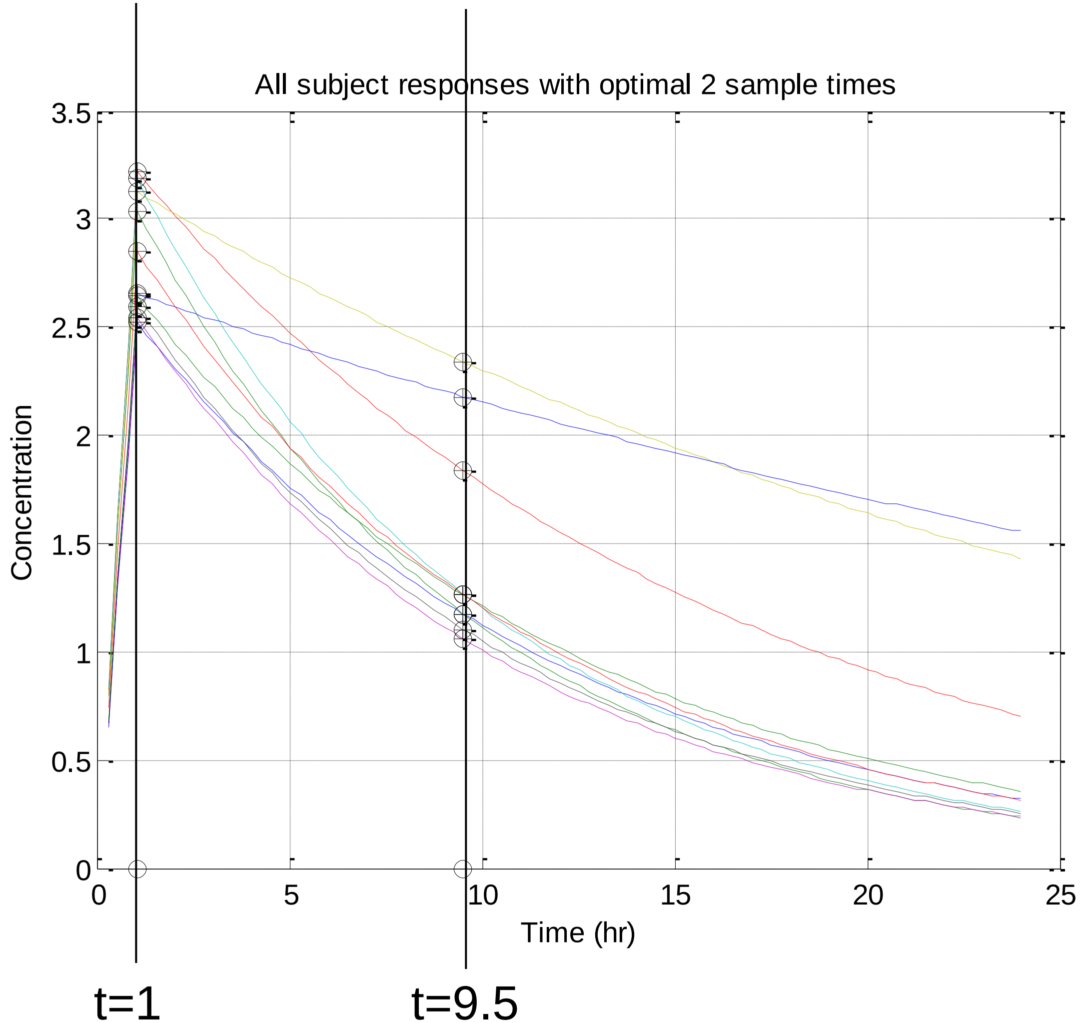
Example of MMopt Design with 10 Models 1-Sample-Time Case

- Optimal Sampling Times (hr)
  4.2500



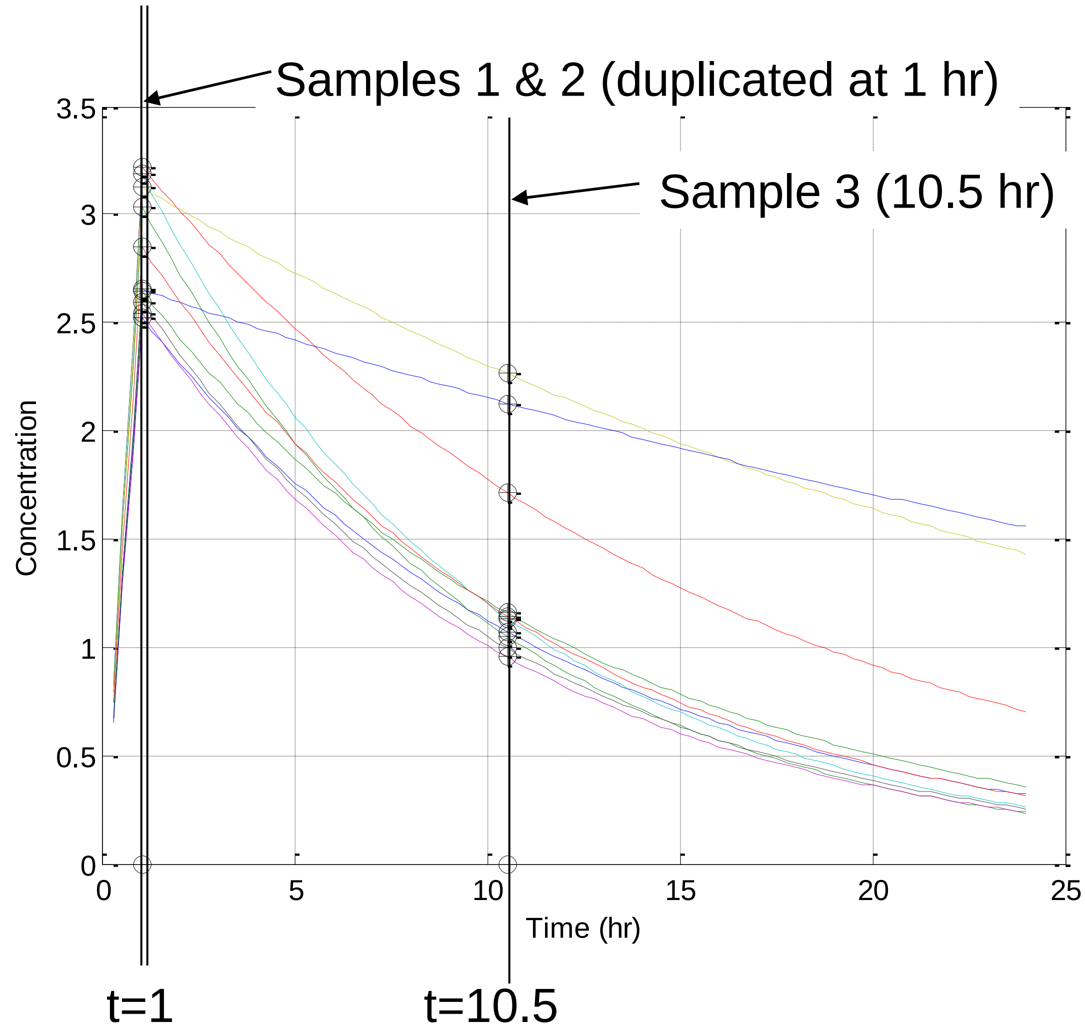
Example of MMopt Design with 10 Models
2-Sample-Time Case

- Optimal Sampling Times (hr)
  1.0000
  9.5000



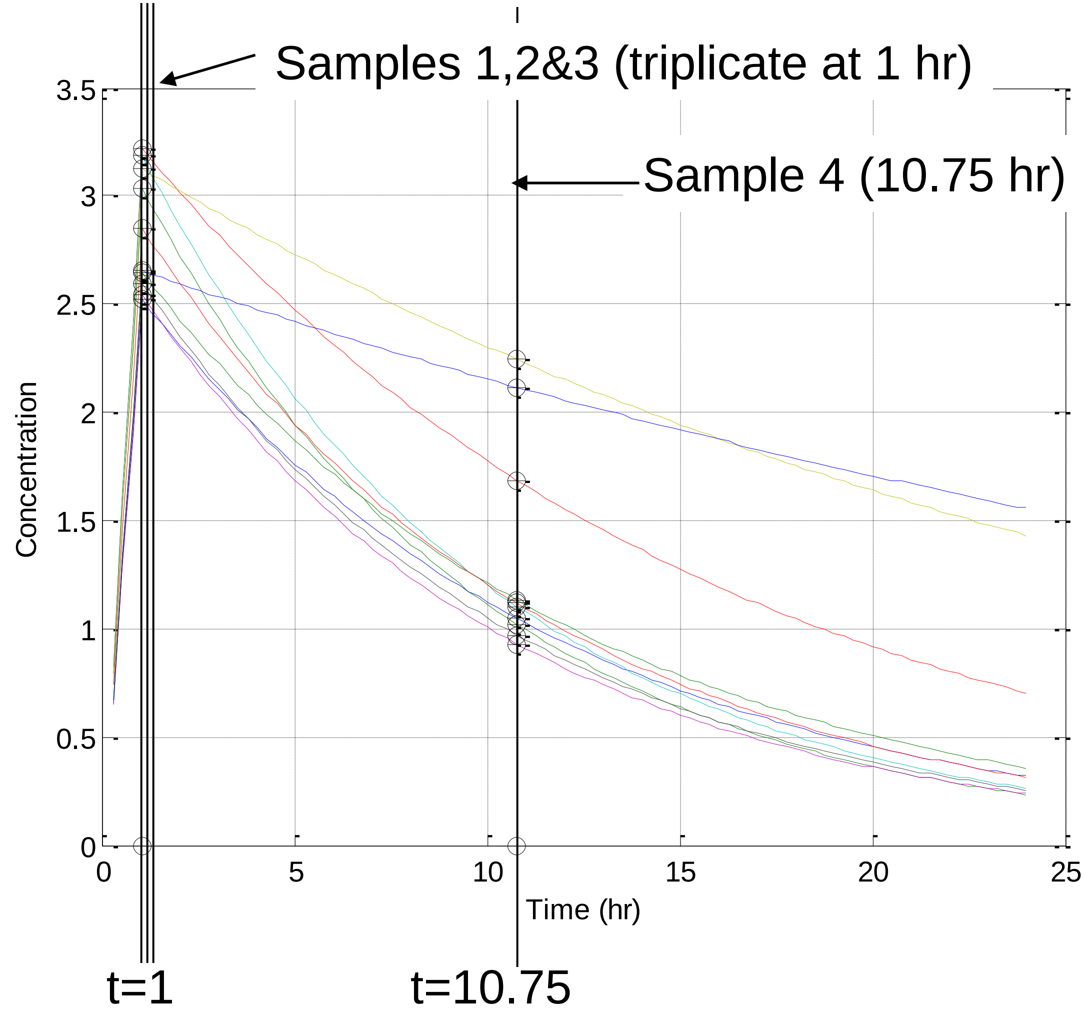
Example of MMopt Design with 10 Models: 3-Sample-Time Case

- Optimal Sampling Times (hr)
  - 1.0000
  - 1.0000
  - 10.5000



Example of MMopt Design with 10 Models: 4-Sample-Time Case

- Optimal Sampling Times (hr)
  1.0000
  1.0000
  1.0000
  10.7500



Comparison Table

	ED	EID	API	MMopt
Invariant under regular linear reparametrization*	**Yes**	**Yes**	**Yes**	**Yes**
Invariant under regular nonlinear reparametrization*	**No**	**No**	**Yes**	**Yes**
Allows taking fewer than p samples, p= # of parameters	**No**	**No**	**No**	**Yes**
Can handle heterogeneous model structures	**No**	**No**	**No**	**Yes**
Gives known optimal solution to 2-model example	**No**	**No**	**No**	**Yes**
Captures main elements of minimizing Bayes risk	**No**	**No**	**No**	**Yes**

*Proved in Appendix

Summary

- **Multiple Model Optimal Design (MMOpt) provides an alternative to Fisher-Information Matrix based design**
  - Particularly attractive for Nonparametric Models (MM discrete prior)
  - Based on true MM formulation of the problem (i.e., classification theory)
  - Has many advantages relative to ED, EID and API (see Table summary)
  - Based on recent theoretical overbound on Bayes Risk (Blackmore et. al. 2008)

- **Advantages of MMopt shown using simple 2-Model example**
  - Both Bayes Optimal design and MMopt maximize Model Response Separation
  - MMopt identical to Bayes Optimal Sample design for this problem
  - ED, EID and API did not perform well in terms of Bayes Risk unless a-priori parameter uncertainty is chosen small
  - Shown that goals of classification and parameter estimation become asymptotically similar as prior parameter uncertainty becomes vanishingly small

- **ED, EID and API designs do not explicitly consider model response separation**
  - Blind to the underlying classification problem
  - Not as well-suited for nonparametric models

- **MMopt captures essential elements of Bayes Risk minimization without the excessive computation**
  - Targeted to be included in a future release of the USC *BestDose* software

References (1/3)

[1] Bayard D, Jelliffe R, Schumitzky A, Milman M, Van Guilder M. Precision drug dosage regimens using multiple model adaptive control: Theory and application to simulated Vancomycin therapy. In: *Selected Topics in Mathematical Physics, Prof. R. Vasudevan Memorial Volume*. Madras: World Scientific Publishing Co.,1995.

[2] Schumitzky A. "Application of stochastic control theory to optimal design of dosage regimens," In: *Advanced methods of pharmacokinetic and pharmacodynamic systems analysis*. New York: Plenum Press; 1991:137-152

[3] USC *BestDose*, http://www.lapk.org

[4] Blackmore L, Rajamanoharan S, and Williams BC, "Active Estimation for Jump Markov Linear Systems," IEEE Trans. Automatic Control., Vol. 53, No. 10., pp. 2223-2236, Nov. 2008.

[5] Mallet A. "A maximum likelihood estimation method for random coefficient regression models," *Biometrika*. 1986;73:645-656.

[6] B. Lindsay, "The Geometry of Mixture Likelihoods: a General Theory," Ann. Statist. 11: 86-94, 1983.

[7] Schumitzky, "Nonparametric EM Algorithms for estimating prior distributions," Applied Mathematics and Computation, Vol. 45, Nol. 2, September 1991, Pages 143–157.

References (2/3)

[8] R. Leary, R. Jelliffe, A. Schumitzky, and M. Van Guilder. "An Adaptive Grid Non- Parametric Approach to Pharmacokinetic and Dynamic (PK/PD) Population Models." In Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium on, pp. 389-394. IEEE, 2001.

[9] Y. Baek, “An Interior Point Approach to Constrained Nonparametric Mixture Models,” Ph.D. Thesis, University of Washington, 2006.

[10] Jelliffe R, Schumitzky A, Bayard D, Van Guilder M, Leary RH. “The USC*PACK Programs for Parametric and Nonparametric Population PK/PD Modeling,” Population Analysis Group in Europe, Paris, France, June 2002.

[11] D.Z. D’Argenio, A. Schumitzky, and X. Wang, ADAPT 5 User’s Guide. Biomedical Simulation Resource, University of Southern California, 2009.

[12] D’Argenio DZ, “Optimal Sampling Times for Pharmacokinetic Experiments,” J. Pharmacokinetics and Biopharmaceutics, vol. 9, no. 6, 1981: 739-756

[13] K. Chaloner and I. Verdinelli, “Bayesian experimental design: A review,” Statistical Science, Vol. 10, No. 3, pp. 273-304, 1995.

[14] L. Pronzato and E. Walter, “Robust experiment design via stochastic approximation,” Mathematical Biosciences, Vol. 75, pp. 103-120, 1985

References (3/3)

[15] E. Walter and L. Pronzato, “Optimal experiment design for nonlinear models subject to large prior uncertainties,” American Journal of Physiology – Regulatory, Integrative and Comparative Physiology, 253:R530-R534, 1987.

[16] M. Tod and J-M Rocchisani, “Comparison of ED, EID, and API criteria for the robust optimization of sampling times in pharmacokinetics,” J. Pharmacokinetics and Biopharmaceutics, Vol. 25, No. 4, 1997.

[17] D.Z. D’Argenio, “Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments,” Mathematical Biosciences, Vol. 99, pp. 105-118, 1990.

[18] Y. Merle and F. Mentre, “Bayesian design critera: Computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model,” J. Pharmacokinetics and Biopharmaceutics, vol. 23, No. 1, 1995.

[19] S.D. Silvey, *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Chapman and Hall, London, 1980.

[20] V.V. Fedorov, *Theory of Optimal Experiments*. Academic Press, New York, 1972.

[21] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification. John Wiley & Sons, New York, 2001.

APPENDIX

Properties of ED-class Designs under Regular* Reparametrization (1/4)

- **Fisher Matrix for Parameter vector a**

$$M_a(a,\xi) = \sum_{i=1}^{m} \frac{1}{\sigma_i(a)^2} \left[\frac{\partial \eta(t_i,a)}{\partial a} \frac{\partial \eta(t_i,a)}{\partial a^T}\right]$$

where,

$y_i = \eta(t_i,a) + \sigma_i(a)n_i$, Noisy measurement at time t_i

$n_i \sim N(0,1)$, Gaussian measurement noise

$\xi = \{t_1, ..., t_m\}$, Experiment design (sampling)

- **Evaluate on $a = a_o$**

$$M_a(a_o,\xi) \triangleq M_a(a,\xi)\Big|_{a=a_o} = \sum_{i=1}^{m} \frac{1}{\sigma_i(a_o)^2} \left[\frac{\partial \eta(t_i,a)}{\partial a} \frac{\partial \eta(t_i,a)}{\partial a^T}\right]\Bigg|_{a=a_o}$$

- **Fisher Matrix for Parameter vector b**

$$M_b(b,\xi) = \sum_{i=1}^{m} \frac{1}{\sigma_i(b)^2} \left[\frac{\partial \eta(t_i,b)}{\partial b} \frac{\partial \eta(t_i,b)}{\partial b^T}\right]$$

- **Under Regular Reparameterization* $a = f(b)$ can show**

$$M_b(b,\xi) = T^T(b)\ M_a(f(b),\xi)\ T(b)$$

$T(b) \triangleq \frac{\partial f(b)}{\partial b^T}$, Differential matrix of transformation

Take determinant,

$$|M_b(b,\xi)| = J(b)^2 |M_a(f(b),\xi)|$$

$J(b) \triangleq |T(b)|$, Jacobian of transformation

*Mapping $a = f(b)$ is one-to-one and has continuous partial derivatives

Properties of Robust D-Optimal Designs under Regular Reparametrization (2/4)

- **Robust D-Optimal Designs for Parameter Vector a**

  ED: $\xi_a^{ED} \triangleq \arg\max_\xi \; E_a\left(\left|M_a(a,\xi)\right| \right)$

  EID: $\xi_a^{EID} \triangleq \arg\min_\xi \; E_a\left(\frac{1}{|M_a(a,\xi)|}\right)$

  API: $\xi_a^{API} \triangleq \arg\max_\xi \; E_a\left(\, log\left|M_a(a,\xi)\right| \right)$

- **Robust D-Optimal Designs under Regular Reparametrization $a = f(b)$**

- Fisher Matrix M assumed to be non-constant function of a (i.e., nonlinear problems)

- **Case I: Mapping $a = f(b)$ is nonlinear**

  $J(b)$ is non-constant function of b

- **Case II: Mapping $a = f(b) = Fb$ is linear**

  $J(b) = |F| = const$, where $F \in R^{p \times p}$ is a square invertible matrix

- Next slides will prove results in following table:

	ED	EID	ELD
Invariant under regular linear reparametrization	Yes	Yes	Yes
Invariant under regular nonlinear reparametrization	No	No	Yes

Properties of Robust D-Optimal Designs under Regular Reparametrization (3/4)

- **Case I: Mapping $a = f(b)$ is nonlinear**

- **ED Design**

$$\xi_b^{ED} \triangleq \arg\max_\xi \ E_b\left(\left|M_b(b,\xi)\right| \right) = \arg\max_\xi \ E_b\left(J(b)^2 |M_a(f(b),\xi)| \right)$$
$$= \arg\max_\xi \ E_a\left(J(f^{-1}(a))^2 |M_a(a,\xi)| \right) \neq \arg\max_\xi \ E_a\left(\left|M_a(a,\xi)\right| \right) = \xi_a^{ED}$$

Hence, $\xi_b^{ED} \neq \xi_a^{ED}$ and ED design is NOT invariant under regular nonlinear reparametrization

- **ED Design**

$$\xi_b^{EID} \triangleq \arg\min_\xi \ E_b\left(\frac{1}{\left|M_b(b,\xi)\right|} \right) = \arg\min_\xi \ E_b\left(\frac{1}{J(b)^2 |M_a(f(b),\xi)|} \right)$$
$$= \arg\min_\xi \ E_a\left(\frac{1}{J(f^{-1}(a))^2 |M_a(a,\xi)|} \right) \neq \arg\min_\xi \ E_a\left(\frac{1}{\left|M_a(a,\xi)\right|} \right) = \xi_a^{EID}$$

Hence, $\xi_b^{EID} \neq \xi_a^{EID}$ and EID design is NOT invariant under regular nonlinear reparametrization

- **API Design**

$$\xi_b^{API} \triangleq \arg\max_\xi \ E_b\left(log\left|M_b(b,\xi)\right| \right) = \arg\max_\xi \ E_b\left(log\left(J(b)^2 |M_a(f(b),\xi)|\right) \right)$$
$$= \arg\max_\xi \left(2E_b \ log \ J(b) + E_a \ log|M_a(a,\xi)| \right) = \arg\max_\xi \ E_a \ log|M_a(a,\xi)| = \xi_a^{API}$$

Hence, $\xi_b^{API} = \xi_a^{API}$ and API design IS invariant under regular nonlinear reparametrization

Properties of Robust D-Optimal Designs under Regular Reparametrization (4/4)

- **Case II: Mapping $a = f(b) = Fb$ is linear**

- ED Design

$$\xi_b^{ED} \triangleq \arg\max_\xi \ E_b\left(\left|M_b(b,\xi)\right| \right) = \arg\max_\xi \ E_b\left(J(b)^2 |M_a(f(b),\xi)| \right)$$
$$= \arg\max_\xi \ E_a\left(|F|^2 |M_a(a,\xi)| \right) = \arg\max_\xi \ E_a\left(\left|M_a(a,\xi)\right| \right) = \xi_a^{ED}$$

Hence, $\xi_b^{ED} = \xi_a^{ED}$ and ED design IS invariant under regular linear reparametrization

- EID Design

$$\xi_b^{EID} \triangleq \arg\min_\xi \ E_b\left(\frac{1}{\left|M_b(b,\xi)\right|} \right) = \arg\min_\xi \ E_b\left(\frac{1}{J(b)^2|M_a(f(b),\xi)|} \right)$$
$$= \arg\min_\xi \ E_a\left(\frac{1}{|F|^2|M_a(a,\xi)|} \right) = \arg\min_\xi \ E_a\left(\frac{1}{\left|M_a(a,\xi)\right|} \right) = \xi_a^{EID}$$

Hence, $\xi_b^{EID} = \xi_a^{EID}$ and EID design IS invariant under regular linear reparametrization

- API Design

$$\xi_b^{API} \triangleq \arg\max_\xi \ E_b\left(log\left|M_b(b,\xi)\right| \right) = \arg\max_\xi \ E_b\left(log\left(J(b)^2|M_a(f(b),\xi)|\right) \right)$$
$$= \arg\max_\xi \left(2E_b \ log|F| + E_a \ log|M_a(a,\xi)| \right) = \arg\max_\xi \ E_a \ log|M_a(a,\xi)| = \xi_a^{API}$$

Hence, $\xi_b^{API} = \xi_a^{API}$ and API design IS invariant under regular nonlinear reparametrization

Relation Between MMopt and Fisher Matrix in 2-Model Example with Close Parameters

- Assume model responses $\eta(t, a_1)$ and $\eta(t, a_2)$ can be expanded in a Taylor series about the mean parameter $\overline{a} = (a_1 + a_2)/2$

  $$\eta(t, a_2) = \eta(t, \overline{a}) + \frac{\partial \eta(t,a)}{\partial a}|_{a=\overline{a}}(a_2 - \overline{a}) + O(a_2 - \overline{a})^2$$

  $$\eta(t, a_1) = \eta(t, \overline{a}) + \frac{\partial \eta(t,a)}{\partial a}|_{a=\overline{a}}(a_1 - \overline{a}) + O(a_1 - \overline{a})^2$$

- Subtract Taylor expansions and rearrange using $\Delta a = a_2 - a_1$

  $$\eta(t, a_2) - \eta(t, a_1) = \frac{\partial \eta(t,a)}{\partial a}|_{a=\overline{a}}\Delta a + O(\Delta a)^2$$

  $$(\eta(t, a_2) - \eta(t, a_1))^2 = \left(\frac{\partial \eta(t,a)}{\partial a}|_{a=\overline{a}}\right)^2 (\Delta a)^2 + O(\Delta a)^3$$

  $$\frac{(\eta(t,a_2) - \eta(t,a_1))^2}{(\Delta a)^2} = \left(\frac{\partial \eta(t,a)}{\partial a}|_{a=\overline{a}}\right)^2 + O(\Delta a)$$

- Last equation relates MMopt Cost to D-Optimal cost, i.e.,

  $$-log \text{ MMopt Cost} = \text{const } |M(\overline{a})| + O(\Delta a)$$

- Result: For the 2-Model example, the D-Optimal design (designed for *a-prior* mean $\overline{a} = (a_1 + a_2)/2$) approaches the MMopt design (and Bayes Optimal Design) asymptotically as the *a-prior* parameter uncertainty $\Delta a = a_2 - a_1$ becomes small